Deep Learning Approaches for Automated Diagnosis of COVID-19 Using Imbalanced Training CXR Data

Ajay Sharma(✉) and Pramod Kumar Mishra

Department of Computer Science, Institute of Science, Banaras Hindu University, Varanasi 221005, India
{ajay.sharma17,mishra}@bhu.ac.in

Abstract. Due to the exponential raise of COVID-19 worldwide, it is important that the artificial intelligence community address to analyze CXR images for early classification of COVID-19 patients. Unfortunately, it is very difficult to collect data in such epidemic situations, which is essential for better training of deep convolutional neural networks. To address the limited dataset challenge, the author makes use of a deep transfer learning approach. The presence of limited number of COVID-19 samples may lead to biased learning due to class imbalance. To resolve class imbalance, we propose a new class weighted loss function that reduces biasness and improves COVID-19 sensitivity. Classification and preprocessing are two concrete components of this study. For classification, we compare five pre-trained deep neural networks architectures i.e. DenseNet169, InceptionResNetV2, MobileNet, Vgg19 and NASNetMobile as a baseline to achieve transfer learning. This study is conducted using two fused datasets where samples are collected from four heterogeneous data resources. Based on number of classes we make four different classification scenarios to compare five baseline architectures in two stages. These scenarios are COVID-19 vs non- COVID-19, COVID-19 vs Pneumonia vs Normal, COVID-19 pneumonia vs Viral vs Bacterial pneumonia vs Normal and COVID19 vs Normal vs Virus + Bacterial pneumonia. The primary goal of this study is to improve COVID-19 sensitivity. Experimental outcomes show that DenseNet169 achieves the highest accuracy and sensitivity for COVID-19 detection with score of 95.04% and 100% for 4-class classification and 99.17% and 100% for 3 class-classification.

Keywords: Deep learning · Chest X-ray · COVID-19 classification · Pneumonia classification · Image processing · Imbalanced data

1 Introduction

Coronavirus disease 2019 (COVID-19) [1], was declared a global pandemic by WHO just in less than four months when 3.3 million confirmed and 238,000 deaths were reported as of 2nd May 2020. COVID-19 disease was originated by SARSCoV-2, till February 7 2021, 105,394,301 confirmed cases and with 2,302,302 deaths were reported

I. Woungang et al. (Eds.): ANTIC 2021, CCIS 1534, pp. 1–20, 2022.
https://doi.org/10.1007/978-3-030-96040-7_36

globally to WHO [2]. Due to lack of sufficient treatment and its extremely contagious nature, it is essential to prevent its spreading and to develop different methodologies that can early classify findings of COVID-19. Worldwide, many scientists of medicine, clinical study, Artificial intelligence and others are trying to build up different approaches to easily categorize and to prevent the spreading of COVID-19 and any such pandemic in future. Initial testing indicates that reverse-transcriptase polymerase chain reaction (RT-PCR) has very low sensitivity for COVID-19 [3] but radiological examinations are found more useful in assessment and diagnosis of disease evolution. According to most of clinical studies worldwide, exposed that lung infection has been seen in COVID-19 patients [28]. For lung-related infection CT (chest-computed tomography) is utmost effective technique for COVID-19 diagnosis but it is expensive. Screening done by CT on COVID-19 patients showed more sensitivity [4] as compared to initial testing technique [3, 10] RT-PCR. But, due to unexpected surge in COVID-19 pervasiveness it becomes difficult to make routine use of CT because of portability issue and higher cost than CXR. Currently, CXR modality is considered a standard tool to detect infection attacks of COVID-19. In past CXR images are widely used to analyze the lungs to diagnose abscesses, pneumonia, tuberculosis [21], lung inflammation and enlarged lymph nodes [9]. Also, radiological examinations disclosed that because of similarity in viral pneumonia and COVID-19 [28], some patients of COVID-19 infection were identified as pneumonia. Thus, early detection and isolation of COVID-19 cases is significant to prevent it from [33] spreading.

In order to make more proficient future healthcare systems, more research should be carried out on medical images and radiological examinations to develop better diagnostic systems that assist doctors in early diagnosis. In regard of COVID-19, radiological examinations have undoubtedly shown certain abnormal patches in CXR [28] of patients having COVID-19 and pneumonia. Accordingly, numerous CNN models [6] have been explored actively for the classification of CXR [13–15, 50] and CT images [7, 12, 28] having COVID-19 symptoms. There exist several popular deep learning models that have been proven very useful for numerous radiological applications for classification of pneumonia and tuberculosis [16, 22, 27]. However, CNN models have successfully analyzed CT and video endoscopy [7] more efficiently compared to radiologists. Several studies [5, 8] have shown that deep learning models developed for image analysis tasks has surpassed performance label of radiologists. Initially, Wang et al. [7] proposed VGG16 model especially for classification of pneumonia in lungs. Following this, Rajpurkar et al. [8] developed a system based on deep CNN model named DenseNet201 to distinguish pneumonia among various pathologies using CXR images. CheXNet methodology was later proposed using 121 layered CNN architecture by making the largest ChestX-ray14 dataset [11] dataset holding more than 100,000 frontal-view X-ray images. Comparative study has successfully shown that proposed model gains superior label of performance compared to radiologists based on F1-score metric. Another CheXpert [9] model proposed for diagnosing 14 different pathologies from CXR radiographs outperforms the results shown by three radiologists [9, 10]. These studies motivate researchers to develop new models or use existing models for the diagnosis of COVID-19 infection in CXR radiographs.

The research contribution by Wang and Wong [28] proposed deep learning-based model named COVID-Net for diagnosis of COVID-19 in CXR with 80% sensitivity. Following this, numerous deep learning applications including AlexNet, Inception, GoogleNet, VGG16, MobileNetV2, VGG19, ResNet50, InceptionResNetV2, and many more have been actively explored for drawing successful [22] conclusions for COVID-19 from CXR images. But, due to limited availability of COVID-19 Chest X-ray samples as compared to other similar characteristics pathology like pneumonia may resulted in biased learning [29]. Although, motivated by success of these models in the early classification of COVID-19 cases by finding certain abnormalities in CXR images enforce us to explore further deep learning methodologies handle such epidemic data in future [50, 51]. Major contribution of this paper is to develop a CNN architecture appropriate for handling limited imbalanced training dataset based on better image processing methods.

The whole structure of this article is defined as follows: In Sect. 2, we discuss the existing contributions of research studies. In Sect. 3 we discuss dataset collection following this in Sect. 4 we discuss current methodology that has been used to develop reliable model using image preprocessing and weighted loss approach. In Sect. 5 we discuss results obtained by the proposed methodology and attention visualization using Grad-CAM. Finally, Sect. 6 ends with a conclusion.

2 Literature Review

With the advancement in artificial intelligence and deep learning, it has been widely applicable for various radiological examinations on CXR images for analyzing respiratory diseases in last decade. Inspired by early success for interpretation of radiological examinations successfully using deep learning, various research studies have been conducted for identification of COVID infected patients by CXR images [12–14]. Majority of these studies use similar COVID-19 dataset collected by Cohen et al. in combination with subset of various pneumonia datasets. For instance, the research study proposed in [15] author used three CNN-based models for detection of COVID-19, these baseline architectures are InceptionResNetV2, InceptionV3 and ResNet50. Experiments had been performed on dataset containing only 50 COVID-19 and 50 normal patients. Result highlights that ResNet50 achieves good accuracy of 98%. Another deep learning pre-trained architecture VGG16 [17] was used to categorize COVID-19 against control and pneumonia classes. The study was carried out on balanced dataset containing 132 COVID-19, 132 pneumonia and 132 controls images. This achieves 100% sensitivity for identifying COVID-19 whereas less for other two. In [18] by Abdullahi et al. employed a pretrained deep learning model for classification of COVID-19 pneumonia, against BP, VP and control CXR examples with different classification strategies. This study experimented AlexNet model on various classification strategies for comparison of COVID-19 class, these are normal vs COVID-19, normal vs bacterial pneumonia, normal vs viral pneumonia and bacterial pneumonia vs COVID-19 whereas in 3- way (normal vs COVID-19 vs. bacterial pneumonia) is compared and in 4-way classification all are considered a different label.

In another study [19], Xception is employed as baseline architecture for detection of COVID-19. This transfer learning approach make use of 500 normal, 127 COVID-19, and

500 pneumonia samples collected from different heterogeneous sources. The proposed model achieves an accuracy value of 97% for covid-19 detection. In [20], the author tested seven CNN architectures using imbalanced dataset containing 25 COVID infected and 50 normal patients. DenseNet121 and VGG19 pretrained models present best results with F1-score of 0.91 and 0.89 for COVID-19 and normal samples. In [23], four different CNN architectures (DenseNet-121, SqueezeNet, ResNet50 and ResNet18) were used to facilitate transfer learning. For the experimental study 5, 000 no-finding and pneumonia images and 184 COVID-19 samples has been collected. The study claims 98% recall and 93% specificity. In [24], five deep learning models (DenseNet169, ResNet, ResNet-v2, Inception-v3, Inception and NASNetLarge) selected as baseline. Resampling and entropy-based approach is used to handle data imbalance. NesNetLarge achieves an accuracy of 98% and 96% and recall of 90% and 91% in 2 and 3 class problem.

Another approach was employed in [25] to mitigate data imbalance by generating duplicate samples for better training. With the help of generative adversarial network (GAN) similar samples of CXR images were produced. This study uses dataset with a corpus of 1, 124 normal and 403 COVID-19 images. With augmentation author presents that accuracy improved from 85% to 95% using VGG16 network as backbone. Another study in [26] was conducted for diagnosing COVID-19 in binary and multiclass problem. The comparative study used similar dataset as in [19] but make comparison between covid-19, control and covid-19, pneumonia and control samples. Modified DarkNet model employed as baseline model with cross-validation ($k = 5$) achieves 98% and 87% accuracy score in 2-way and multi-way classification. In [30], GAN model has been employed for increasing the corpus of data with data augmentation. Dataset used to study contains 307 images having 4 different classes: COVID-19, normal, VP and BP. AlexNet, Restnet18, GoogleNet were the CNN architectures experimented for classification. GoogleNet achieved improved accuracy score of 99%.

In [1], found new way to train models based on patch-wise approach. Analogous, concept of transfer learning with ResNet18 has experimented as classification network model and to achieve segmentation FC-DenseNet is used. The study was carried out using a corpus of COVID-19 (180), BP (54) and VP (20), normal (191) and tuberculosis (57). This patch-wise study attained an accuracy score 89%. Specifically, wong and wang et al. introduced DL based COVID-Net [31] framework in early times. Experiments comprised on standard dataset comprising of 13975 CXR images, including classes pneumonia, normal and 266 COVID examples. COVID-Net model achieves better recall and accuracy score of 91% and 93.3% for COVID-19 cases as compared to VGG19 and ResNet50. Another COVID-19 detection model named COVID-AID [32] built on 121 layered DenseNet model investigated on covid-chestxray-dataset [37] for covid cases. It distinguishes COVID-19 infection cases with 100% sensitivity among VP, BP and control cases. In [33], another DeTrac (Decompose, Transfer and Compose) architecture based on deep and shallow transfer learning (VGG19, ResNet, AlexNet, GoogLeNet and squeezeNet) to classify COVID-19 (105) against normal (80) and SARS (11) cases. DeTrac achieves an accuracy of 93.1% for covid-19 detection.

3 Dataset Description

For this research study, we COVID-19 CXR dataset used in above studies but with few updated COVID-19 samples. In this study, we make two separate datasets COVIDx and COVIDy by utilizing data from different heterogeneous CXR datasets. COVIDx dataset makes by fusion of three popular publicly available datasets i.e. COVID-19 chest Xray-dataset [37], RSNA (Radiological Society of North America [34]) and USNLM (U.S. national library of medicine [36]). Covid chest X-ray dataset is a publicly available database of CXR images collected by Cohen et al. [37] related to MERS, ARDS, SARS, COVID-19 pneumonia, viral pneumonia etc. from various resources accessible at different public domains [37]. Another NIH CXR14 [36] dataset is used for collecting pneumonia and normal CXR images. For more robust training and testing another NLM(MC) dataset is used for collecting more normal images. The complete description of the number of images sampled for making fused dataset COVIDx is shown in Table 1. Further in the second stage, we prepare another dataset COVIDy where we use the same subset of samples from COVIDx frontal-view chest X-ray images for Covid-19 and normal but new chest-Xray-pneumonia dataset [35] is used for extracting images of CXR viral /bacterial pneumonia labeled differently for more robust model development. The aim for making this dataset is to identify how better developed deep learning model can identify Covid-19 pneumonia when bacterial pneumonia (BP) and viral pneumonia (VP) are included differently. Table 1 represents a detailed description about the datasets used for collecting CXR images, their respective class, number of samples present in a particular class, and total number of samples.

Table 1. Class summary and data resources related to fused datasets.

	Dataset used	Class	#	Total
COVIDx	Covid CXR [37]	COVID-19	193	**193**
		Pneumonia	38	**520**
	RSNA [34]	Pneumonia	482	
		Normal	443	**523**
	NLM(MC) [36]	Normal	80	
COVIDy	COVIDx	Covid-19	193	**193**
		Normal	173	**173**
	CXR Pneumonia [35]	VP	122	**122**
		BP	113	**113**

4 Proposed Contributions

In the following subsections, we are going to discuss methodology and proposed contributions for better identification of COVID-19 CXR samples. In the first subsection, we

discuss basic initial preprocessing of an image and the class imbalance approach used. In Further subsections, the two-stage classification Process, classification architecture and evaluation metrics have been discussed.

4.1 Image Preprocessing

Computer-aided diagnosis of medical images needs identification of an essential region of interest whereas image preprocessing helps to reduce time and space complexity of an algorithm as well as error rate. Basic preprocessing steps involved are as follows:

CXR images for this study are collected from different heterogeneous data sources. These sources may represent heterogeneity in shape, size, datatype, scanning condition, acquisition condition, range and postprocessing etc. CXR images are normalized first to ensure mean value to zero and standard deviation to one [38]. Further, their data types from uint8/uint16 converted to uniform format float32. Also, images are reshaped to 320 * 320 * 3 so that the classification process can be achieved rapidly.

Histogram equalization (HE) is used to enhance contrast by adjusting intensity values. Initially, the images from RGB are converted to L*a*b* color space. L*a*b* model not only represents colors of CMYK and RGB but also expresses color alleged by human eye. Medical image analysis prefers L*a*b* because it compensates inequality in color distribution of the RGB color model. L* is luminance component that is taken into consideration by HE [39]. Conversion from RGB to L*a*b* directly not possible so it is processed as i.e., RGB-XYZ-L*a*b* where XYZ is another color space. This resulted preprocessed image helps classification network to focus on more relevant features present within lungs. Due to implementation of above algorithm, there may be a chance of uncertainty present at pixel level in filtered image [48]. So, the adaptive total variation filter method [49] is used for denoising of uncertainty present at pixel level. In Fig. 1 we have shown raw covid image, histogram Equalised and denoised image, normalized image with distribution of pixel intensities corresponding to the image.

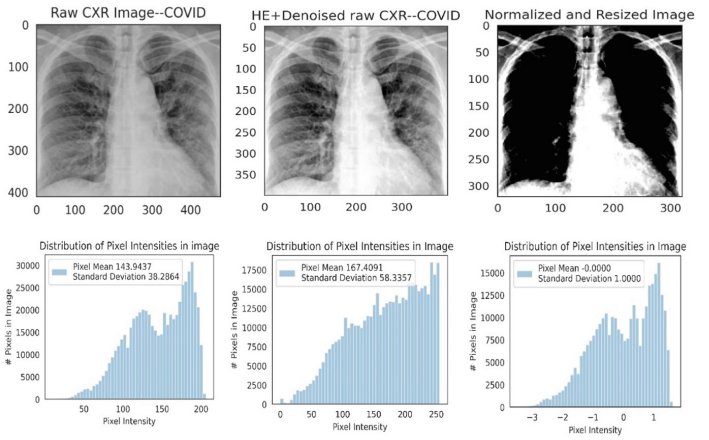


Fig. 1. Corresponding image and distribution of pixel intensities at different stages of preprocessing.

Let I represent grayscale image over given bounded set Ω subset of R^2 then denoised image d approximately matches an observed image y = (y1, y2)∈ Ω, is represented by Eq. (1) as:

$$d = \arg\min_{d}\left(\int_{\Omega}(d - I.\ln d)dy + \left(\int_{\Omega}(w(y)|\nabla d|dy\right)\right) \tag{1}$$

Where $w(y) = \frac{1}{1+K\ mod\ G_\sigma * \nabla d'}$, $GaussianKernel(G_\sigma)$ with variance(σ), contrast parameter (K > 0) and a convolution operator (*). After initial preprocessing, images are converted to grayscale for thresholding process [26]. In order to prevent symbolic and textual noise in images, binary thresholding is applied on initial preprocessed image by thresholding process using Eq. (2).

$$M(x, y) = \begin{cases} Max_{th}, & I(x, y) \geq Min\_th \\ 0, & otherwise \end{cases} \tag{2}$$

Where I (x, y) be input image, *Min_th* and *Max_th* are the minimum and maximum threshold [27] values. Also, morphological analysis of binary threshold image helps to correctly classify some of the dark misclassified [41] areas of target image that are below threshold label. Erosion operation is performed on the threshold image followed by a dilation operation that helps to generate better binary mask [43]. Figure 2 shows the results obtained at various stages of preprocessing related to COVID, normal, virus and bacteria image. First row represents original image, second row shows results after standardization and resizing of image whereas third row represents image resulted after denoising an image and final row shows image obtained by binary thresholding process.

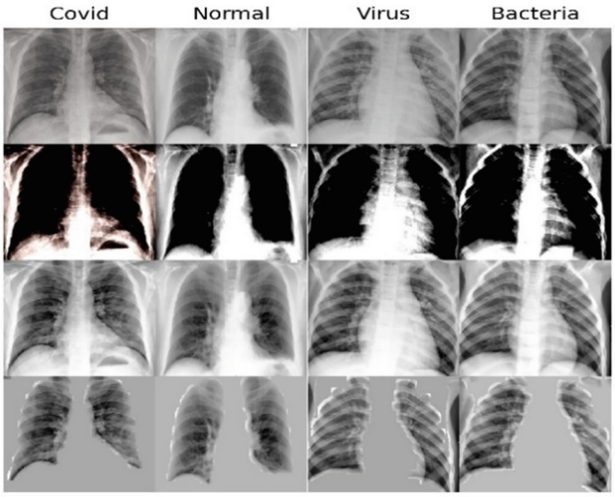


Fig. 2. Results of various stages of CXR preprocessing

The fused CXR datasets COVIDx and COVIDy contain a corpus of 146 training COVID samples with proportion to positive training cases of normal and pneumonia pathology. Similarly, COVIDy virus and bacteria samples are less in number than other

two pathologies. Such imbalanced dataset does not ensure smooth learning of models [42]. So, we used weighted loss-based approach to handle class imbalance challenge. We assign new weights to each class to achieve equal contribution to loss for positive as well as negative cases for each class. For this we multiply each example in training set by a class specific weighted factor (Wpos and Wneg) corresponding to each class computed using Eq. (5) and (6), so that each class contributes equally to the loss computed using Eq. (6) and (7).

$$freq_{pos} = \frac{total\ positive\ examples}{N} \tag{3}$$

$$freq_{neg} = \frac{total\ negative\ examples}{N} \tag{4}$$

$$W_{pos} = freq_{neg}\ and\ W_{neg} = freq_{pos} \tag{5}$$

$$L(X, y_{covid}) = \begin{cases} -w_{pos,covid} logP(Y = 1|X) if Y = 1 \\ -w_{neg,covid} logP(Y = 0|X) if Y = 0 \end{cases} \tag{6}$$

$$L(X, y) = L(X, y_{covid}) + L(X, y_{normal}) + L(X, y_{bacteria}) + L(X, y_{virus}) \tag{7}$$

4.2 Classification Process

Based on the number of classes present in COVIDx and COVIDy CXR image datasets we make binary and multiclass classification strategies for COVID-19 classification as shown in Table 2 with features represented as classes, labels and dataset used in this study. Whereas, viral pneumonia (VP) and bacterial pneumonia (BP) are taken from CXR pneumonia dataset [35] are also combined as other pneumonia class to compare against COVID and normal cases.

Table 2. Summary of classification scenarios and corresponding labels

Classes	Labels	Dataset
COVID-19 vs non- COVID-19	{0, 1}	COVIDx, Binary
COVID-19 vs normal vs pneumonia	{0, 1, 2}	COVIDx, Multi-class
COVID-19 vs normal vs viral pneumonia vs bacterial pneumonia	{0, 1, 2, 3}	COVIDy Multi-class
COVID-19 vs normal vs other pneumonia	{0, 1, 2}	COVIDy Multi-class

All the classification strategies are managed using different one hot encoded data frame. Table 3 depicted clearly that how many CXR images are distributed into each

training, testing and validation for different classification scenarios. Since, covid CXR dataset used for the proposed study does not previously hold any data for training and testing purpose. So, we randomly distribute CXR fused datasets into training, testing and validation set in proportion as shown in Table 3. Data splitting has been performed on patient level to ensure no data leakage between the training, validation and testing dataset.

Table 3. Distribution summary of data into training, testing and validation

Pathology	COVIDx dataset (stage 1)						COVIDy dataset (stage 2)					
	2-class classification			3-class classification			3-class classification			4-class classification		
	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test
COVID -19	146	16	32	146	16	32	137	24	32	137	24	32
Pneumonia	852	76	114	426	37	56	–	–	–	–	–	–
Normal				426	39	58	145	27	41	145	27	41
VP	–	–	–	–	–	–	153	34	48	78	15	23
BP	–	–	–	–	–	–				75	19	25
Total	**998**	**92**	**146**	**998**	**92**	**146**	**435**	**84**	**121**	**435**	**84**	**121**

4.3 Classification Network Architecture

Our classification network aims to classify given image into one of pathology class by extracting features using state-of-the-art deep learning models. We are intended to prevent overfitting by using limited number of data points as CXR images. Deep learning models are data hungry require big volume of data. To work with limited training CXR examinations we adopted deep transfer learning-based approaches using pretrained ImageNet weights. Instead of using categorical cross-entropy loss [44] function we trained SOTA models with improved class weighted loss function to reduce biasness due to class imbalance. All these strategies support in making training process stable even when the size of dataset is small.

Regarding classification we choose five state-of-the-art DL architectures, DenseNet169, InceptionResNetV2, MobileNet, Vgg19 and NASNetMobile [46–48] as baseline models. In Fig. 3 we have shown the classification network architecture that classifies input samples taken from two datasets COVIDx and COVIDy based on different pathological characteristics. These baseline architectures use class weighted loss function to handle class imbalance over different classification scenarios. We added new layers at top of these baseline model. The output of these base models is taken as input followed by flatten layer, Dense layer having 512 neurons, Dropout layer with value of 0.5 to avoid overfitting accompanied with ReLU activation function. Also, L1 regularization is used in dense layer with low learning rate to prevent overfitting. Most popular

ReLU (Rectified linear unit) activation function used by internal layers whereas SoftMax is used by final dense layer as Eq. (12) and Eq. (13). In addition, finally, an output layer consists of 2, 3 or 4 neurons based on the number of classes with SoftMax activation function.

Once models training is over, these models classify input test image into one of pathology class based on maximum probability value given using softmax function. To know the interpretability of models we visualize attention maps by extracting gradients from last convolutional layer to interpret why that sample is classified to particular class. Models takes input processed in 3 channel and resized to (224, 224) before training various SOTA deep learning approaches. Minibatch gradient descent with batch size 32, optimization approach as Adam with initial learning rate 0.00001 and early stopping strategy with patience value 5 is used for training and testing. On early stopping, learning rate is reduced by 10. Also, model finetuning is performed for last 10 epochs with similar parameters by unfreezing top 20 layers. The entire process is accomplished in two stages. In stage 1, the classification network is trained and tested using COVIDx dataset to make one of these possible predictions based on maximum probability value i.e., a) normal, b) non-COVID19 and c) COVID-19 infection or simply a) covid-19 infection or non-covid-19. In stage 2, the classification network is trained and tested using COVIDy dataset. Here pneumonia cases are separated as viral and bacterial pneumonia for more robust and comparative analysis using SOTA approaches mentioned above. This entire work has been carried out using Tensorflow and Keras library via N-Vidia Titan GPU.

$$ReLu = \max(0, x) \tag{8}$$

$$Softmax = \frac{1}{1 + e^{-x}} \tag{9}$$

4.4 Performance Measures

Deep learning models has been evaluated using performance metrices accuracy, sensitivity (Recall), positive predictive value (PPV) and F1-score computed using Eq. (10–13) [16, 45]. COVID-19 sensitivity and COVID-19 PPV are also recorded to find best model for COVID-19 identification. For any input image I_X, class having maximum confidence score as output by softmax function is considered as final prediction for calculating confusion matrix.

$$\text{Accuracy} = \frac{(TN + TP)}{(TN + TP + FN + FP)} \tag{10}$$

$$\text{Sensitivity} = \frac{(TP)}{(TP + FN)} \times 100 \tag{11}$$

$$\text{Postive Prediction(PPV)} = \frac{(TP)}{(TP + FP)} \tag{12}$$

$$F1\,score = 2 \times \frac{PPV \times \text{Sensitivity}}{PPV + \text{Sensitivity}} \tag{13}$$

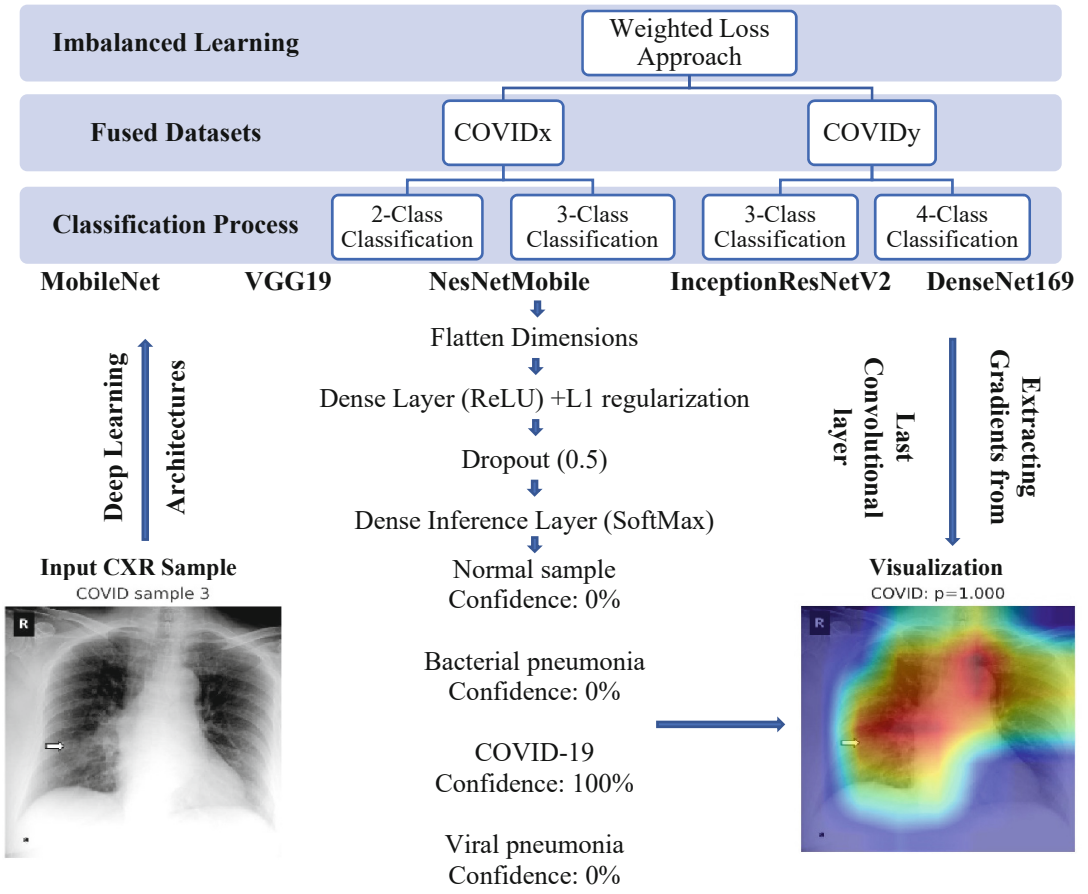


Fig. 3. Classification network architecture

5 Results and Discussion

In this section, we discuss the results achieved after experimentally analyzing CXR images using five transfer learning architectures as baseline models i.e., DenseNet169, InceptionResNetV2, MobileNet,Vgg19 and NASNetMobile. These models are quantitatively evaluated using sensitivity, Positive predictive value, accuracy, F1-score, COVID-19 sensitivity, and COVID-19 positive predictive value. Sample images used for COVID-19 are taken from similar dataset used in majority of studies, So COVID -19 sensitivity and COVID 19 PPV are also taken into consideration. Independent testing set is used for performance analysis whereas model development has been carried out using train and validation sets listed in Table 3. In Table 4 we summarized the best results achieved by CNN architectures on 4 different binary and multiclass classification strategies shown in Table 2. Classification strategy in Table 4 is represented as CX (m, n) and CY (m, n) Where CX and CY refers to COVIDx and COVIDy dataset and "m" is stage number and "n" is the number of classes considered among different classification strategies. All these experiments are performed using improved class weighted loss function. After extensive number of trials, we observed that no particular model performs well on all classification scenarios. But various significant observations have been drawn from these

results. Also, Finetuning is performed for last 10 epochs that significantly improved validation accuracy. We aim to train models until training or validation accuracy ranges more than 95%.

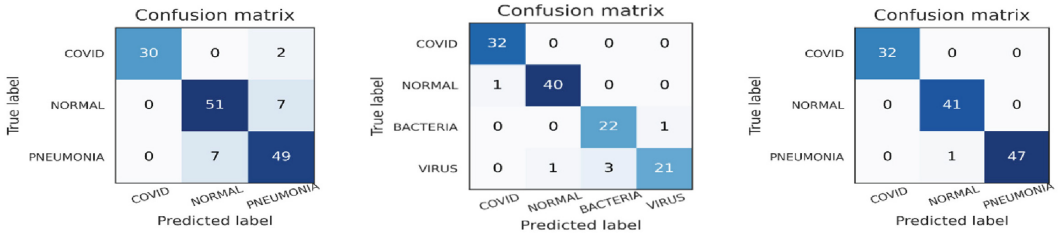
Results have been obtained in two different stages using two separate datasets. In stage 1, we used COVIDx dataset for study with 3 classes resulted in two classification strategies. MobileNet achieves best overall accuracy score but it lacks race in COVID-19 sensitivity and COVID -19 PPV in classifying COVID-19 vs normal vs pneumonia. DenseNet169 claimed to be best for classification of COVID-19 with sensitivity score of 93.75% and PPV of 100%. Similarly, in stage 2 we used COVIDy dataset for more robust analysis of pneumonia into viral and bacterial pneumonia as in other studies. Here, also DenseNet169 achieves best sensitivity value of 100% in both 3 class and 4 class scenarios. Accuracy score is also considerably better than other SOTA models. Also, MobileNet achieves the same but less accurate in distinguishing viral and bacterial pneumonia. DenseNet169 also achieves best PPV value of 96.77% and 100% respectively. However, identification of viral and bacterial pneumonia separately helps more wisely in the identification of COVID-19 cases correctly. Also, VGG19 and NASNetMobile gives very low confidence score as output probability and DenseNet169 gives highest confidence score on given covid-19 sample as input. In Table 4 we have shown the results obtained by five models in stage 1 and stage 2 separately for each classification scenario.

Table 4. Summary of classification results obtained with different scenarios

Stage no.	Classification model	Classification process	Accuracy (%)	sensitivity (%)	Precision (%)	F1-score (%)	COVID19 sensitivity	COVID-19 (PPV) (%)
Results of Stage 1 over (COVIDx Dataset)	MobileNet	CX (1,3)	86.99	86.99	87.11	86.98	90.62	93.55
		CX (1,2)	95.89	95.89	96.10	95.73	81.25	100
	VGG19	CX (1,3)	76.71	76.71	77.58	76.34	56.25	85.71
		CX (1,2)	95.89	95.89	96.10	95.73	81.25	100
	NASNetMobile	CX (1,3)	78.77	78.77	79.89	78.70	65.62	91.30
		CX (1,2)	80.82	80.82	80.82	80.82	56.25	56.25
	InspectionResNetv2	CX (1,3)	78.77	78.77	80.50	78.89	68.75	100
		CX (1,2)	95.21	95.21	95.25	95.05	81.25	96.30
	DenseNet169	CX (1,3)	**89.04**	**89.04**	**89.25**	**89.12**	**93.75**	**100**
		CX (1,2)	**96.58**	**96.58**	**96.79**	**96.63**	**96.88**	**88.57**
Results of Stage 2 over (COVIDy Dataset)	MobileNet	CY (2,4)	90.91	90.91	90.82	90.00	100	96.97
		CY (2,3)	96.69	96.69	96.86	96.70	100	91.43
	VGG19	CY (2,4)	88.43	88.43	88.66	88.25	93.75	93.75
		CY (2,3)	95.04	95.04	95.18	95.01	87.50	96.55
	NASNetMobile	CY (2,4)	76.03	76.03	77.18	76.30	87.50	73.68
		CY (2,3)	89.26	89.26	89.21	89.10	78.12	89.29
	InspectionResNetv2	CY (2,4)	85.95	85.95	86.54	86.24	93.75	90.91
		CY (2,3)	95.04	95.04	95.15	95.09	93.75	88.24
	DenseNet169	CY (2,4)	**95.04**	**95.04**	**95.15**	**94.98**	**100**	**96.97**
		CY (2,3)	**99.17**	**99.17**	**99.19**	**99.17**	**100**	**100**

* *Bold represents the best model*

Furthermore, majority of studies have taken subset of pneumonia and normal samples from benchmark datasets but COVID-19 images are similar in majority of studies. Also, we need to distinguish COVID-19 sample more accurately. Results show that DenseNet169 is best in detecting COVID-19 cases. Outcome values achieved by proposed DenseNet169 shows significant improvements in the identification of COVID-19 cases among others. Likewise, in Fig. 5 we have shown that how training and validation accuracy of DenseNet169 varies with respect to number of epochs. Subplots a) and b) in first row showed graph of training and validation accuracy at y-axis with respect to epoch number at x-axis in 3-class and 2-class respectively in stage 1 using dataset COVIDx. Similarly, subplot c) and d) shows the comparison of accuracy graph in stage 2 using dataset COVIDy in 4-class and 3-class respectively. In initial epochs validation accuracy seems to be more than training accuracy this is because of strong regularization.



(a) 3-class classification (stage 1) (b) 4-class classification (stage 2) (c) 3-class classification (stage 2)

Fig. 4. Confusion matrices corresponding to 3-class and 4-class configurations by DenseNet169.

(a) Visualization of 3-class classification in stage 1 (b) Visualization of 2-class classification in stage 1

(c) Visualization of 4-class classification in stage 2 (d) Visualization of 3-class classification in stage 2

Fig. 5. Visualization of training and validation accuracy vs epoch number for DenseNet169

Based on the principle of class having maximum confidence score as a final predicted class, we have shown resulted confusion matrices for 3-class and 4-class classification scenario on test data in Fig. 4. From results, we can perceive that in stage 1 COVID-19

class gain sensitivity and PPV value of 93.75% and 100% is confused with pneumonia samples because of some overlapping characteristics. But further distinguishing of pneumonia into separate viral and bacterial pneumonia classes in stage 2 resolves the confusion with pneumonia sample. It achieves COVID-19 sensitivity of 100% in stage 2. Moreover, further observations sight that viral class gain sensitivity score of 84% and miserably confused with bacterial patients.

This study is supported by class weighted loss function to handle class imbalance. Further analysis of the proposed class weighted loss function while using proposed model has seen enormous improvements in results. In Table 5 we compared sensitivity values achieved by various classes using categorical cross-entropy loss and proposed weighted loss function on COVIDx dataset. Sensitivity value is observed after every 10 epochs as depicted in Table 5 under similar parameter value of learning rate. Due to a lesser number of COVID-19 samples in this dataset we observed clearly from table that categorical cross-entropy loss prioritizes learning of normal and pneumonia which is against our desired goal. But with weighted loss, improvements can be noticed clearly due to equal importance given to positive and negative samples of each class. We compared the results obtained by proposed DenseNet169 with various state-of-the-art proposed approaches as shown in Table 6, Table 7 and Table 8. In Table 6, we compared our proposed DenseNet169 with very first COVID CXR classification model COVID-Net [28] model that works for three classes as COVID-19, pneumonia and normal. In Table 7 further we compare the proposed model with a popular 4-class COVID-19 classification model named COVID-AID [33] based on DenseNet121 model with ChexNet weights. In Table 8 we compared the results of proposed model with other recently proposed state-of-the-art techniques. Majority of techniques are based on transfer learning and finetuning due to limited availability of COVID-19 CXR images. Table 8 contains features as the reference considered for comparison, best architecture, number of classes, technique used and performance measure in terms of accuracy, COVID-19 sensitivity and COVID-19 positive predictive value as per considerations taken in reference. Table 8 various state-of-the-art approaches are compared with proposed model.

Table 5. Sensitivity comparison with and without weighted loss over COVIDx dataset

Epochs	Classes	1–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80
Categorical Cross Entropy Loss (%)	COVID-19	**6.45**	**12.90**	**16.13**	**19.35**	**58.06**	**61.29**	**64.52**	**74.19**
	Normal	65.52	75.86	81.03	82.76	81.03	82.76	79.31	84.48
	Pneumonia	64.29	69.64	71.43	73.21	75.00	78.57	78.57	75.00
Weighted loss approach (%)	COVID-19	**41.94**	**54.84**	**58.06**	**54.84**	**77.42**	**80.65**	**80.65**	**83.87**
	Normal	53.45	62.07	74.14	77.59	81.03	84.48	81.03	86.21
	Pneumonia	64.29	75.00	76.79	76.79	78.57	78.57	80.36	78.57

* *Bold represents difference in minority class sensitivity before and after data imbalance approach.*

Table 6. Class wise comparison of proposed model with COVID-Net

Pathology	COVID-Net		Proposed	
	PPV	Sensitivity	PPV	Sensitivity
COVID-19	98.9	91	**100**	**93.75**
Normal	90.5	95	87.93	87.93
Pneumonia	91.3	94	84.48	87.50

Table 7. Class wise comparison of proposed model with Covid-AID

Pathology		CovidAID		Proposed	
		PPV	Sensitivity	PPV	Sensitivity
4-class	COVID-19	93.8	100	**96.97**	**100**
	Normal	98.9	76.1	97.56	97.56
	VP	72.1	87.2	95.45	86
	BP	88.1	96.1	88	95.65
3-class	COVID-19	96.8	100	**100**	**100**
	Normal	98.9	74.4	97.62	100
	Pneumonia	86.8	99.5	100	97.92

Table 8. Comparison of proposed model with other state-of-the-art models

Ref.	Best architecture	Classes	Technique	Performance metrics (%)
[17]	VGG16	3	Fine tuning	$Se_{cov} = 100\%$
[19]	Xception	3	Fine tuning	$A_{cov} = 97\%$
[23]	ResNet50	3	Fine tuning	$Se_{cov} = 98\%$, $Sp_{cov} = 93\%$
[24]	NASNetLarge	3	Fine tuning	$A = 96\%$, $Se = 91\%$
[31]	COVID-Net	3	–	$A = 93.3\%$, $Se_{cov} = 91\%$
Proposed DenseNet169		**3**	**Fine tuning**	**$A = 99.17\%$, $Se_{cov} = 100\%$ $Sp_{cov} = 100\%$**
[30]	GoogleNet	4	Fine tuning	$A = 99\%$
[1]	ResNet18	4	Fine tuning	$A = 89\%$, $Se_{cov} = 100\%$
[32]	DenseNet121(COVID-AID)	4	Fine tuning	$Se_{cov} = 100\%$
Proposed DenseNet169		**4**	**Fine tuning**	**$A = 95.04\%$, $Se_{cov} = 100\%$ $PPV_{cov} = 96.97\%$**

* *Bold represents best model, Se represents sensitivity, A as Accuracy and Sp as specificity.*

5.1 Visualizing Learning with GradCAM

To demonstrate output qualitatively, we visualize attention maps focused by proposed model with Gradient weighted Class Activation Mapping (GradCAM). CAM (Class activation maps) aimed at increasing interpretability by looking at where model has focused while classifying given input image to a particular class. In Fig. 6, by taking given COVID-19 test input image we use the Grad-CAM technique to build heatmap showing significant regions of consideration for predicting certain pathological condition. We accomplish this by extracting gradients flowing into last convolutional layer of our proposed DenseNet169 corresponding each predicted class. This is more important to understand whether the interpretation of model focuses on right regions while classifying given input image. From row 1st to 3rd of Fig. 6, with given input positive COVID-19 sample, corresponding saliency map is shown next to the original image with correct predicted class and output probability at top of saliency map image. In saliency map, red signifies region of more importance. Whereas in row 4th and 5th indicate Grad-CAM heatmap on giving correctly classified COVID positive sample as input.

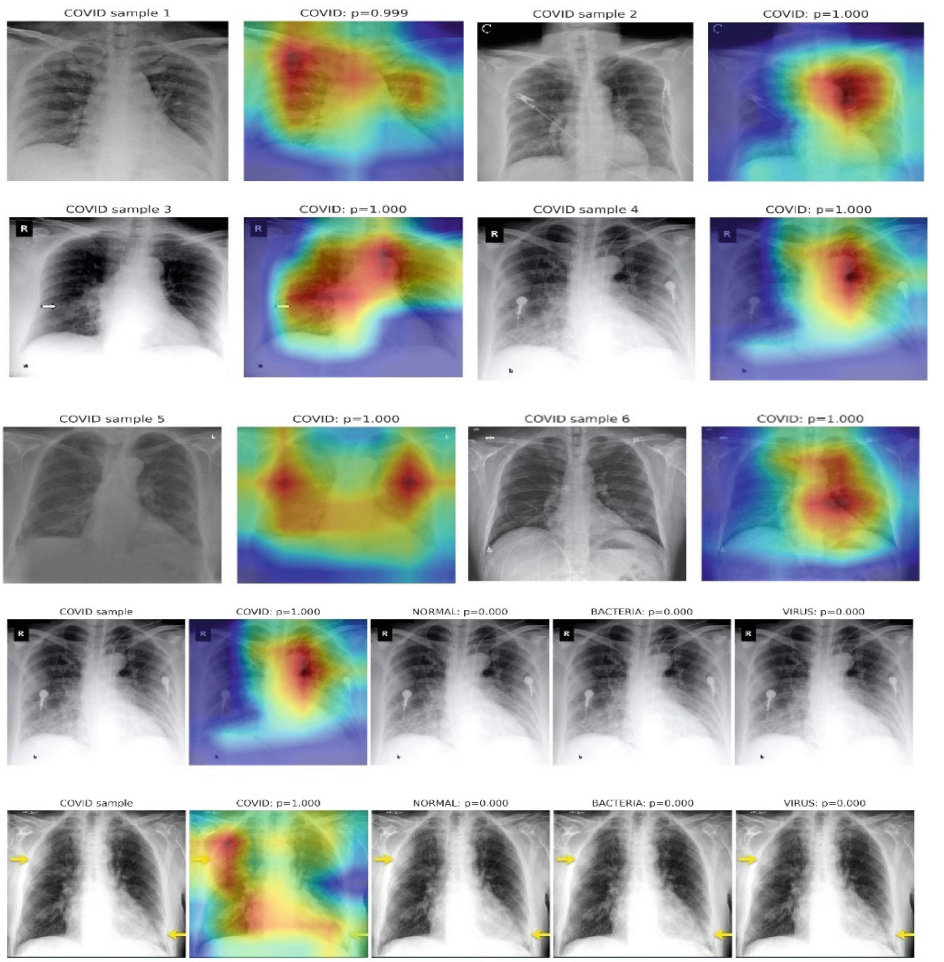


Fig. 6. Visualization of saliency maps over positive COVID-19 CXR sample using Grad-CAM

6 Conclusions

This article presents a comparative study of deep CNN models for automated detection of COVID-19 using limited CXR samples. Five different state-of-the-art transfer learning models are tested to aid the early classification of COVID-19. We observed that the emergency situation of COVID-19 completely shattered the healthcare sector worldwide due to its exponential behaviour and unidentified nature of the disease in its early stages. We believe that timely inference and correct diagnosis of type of disease can save millions of lives every year. Thus, computer-aided diagnosis significantly assists radiologists to capture more better images and real-time identification of COVID-19 or pneumonia just after the acquisition. The idea of early classification will open many application areas of CAD tools. It will be useful for the screening process at airports for early diagnosis of COVID-19, pneumonia and other types of disease. Now, it is the responsibility of artificial intelligence community to develop, explore and test new models to deal with current situation and any such epidemic situation in future. In this study, after number of trials we observed that DenseNet169 outperforms other four comparison models in terms of COVID-19 sensitivity in 3-class and 4-class classification. Also, we are in a view that proposed class weighted loss function performs better than categorical class entropy loss. Although, size of the dataset is small but outcomes look promising. Finally, in global perspectives after critically analyzing the related literature and the results obtained from DenseNet169, we can say that deep learning models has enormous potential in the classification of COVID-19 patients against others. Also, the series of DenseNet models have high sensitivity value and is likely to be more supportive in the early classification of COVID-19 cases among others.

References

1. Oh, Y., Park, S., Ye, J.C.: Deep learning COVID-19 features on CXR using limited training data sets. IEEE Trans. Med. Imaging **39**(8), 2688–2700 (2020)
2. WHO: Coronavirus disease 2019 (COVID-19) Dashboard (2021). https://covid19.who.int/. Accessed 7 Feb 2021
3. Waller, J.V., et al.: Diagnostic tools for coronavirus disease (COVID-19): comparing CT and RT-PCR viral nucleic acid testing. Am. J. Roentgenol. **215**(4), 834–838 (2020)
4. Khatami, F., et al.: A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis. Sci. Rep. **10**(1), 1–12 (2020)
5. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106 (2017)
6. Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K.N., Mohammadi, A.: COVID-CAPS: a capsule network-based framework for identification of COVID-19 cases from X-ray images. Pattern Recogn. Lett. **138**, 638–643 (2020)
7. Zreik, M., Van Hamersvelt, R.W., Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. IEEE Trans. Med. Imaging **38**(7), 1588–1598 (2018)
8. Rajpurkar, P., et al.: Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. **15**(11), e1002686 (2018)

9. Irvin, J., et al.: CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 590–597, July 2019
10. Ketu, S., Mishra, P.K.: Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection. Appl. Intell. **51**(3), 1492–1512 (2021). https://doi.org/10.1007/s10489-020-01889-9
11. Rajpurkar, P., et al.: CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
12. Perumal, V., Narayanan, V., Rajasekar, S.J.S.: Detection of COVID-19 using CXR and CT images using transfer learning and Haralick features. Appl. Intell. **51**(1), 341–358 (2020). https://doi.org/10.1007/s10489-020-01831-z
13. Chakraborty, M., Dhavale, S.V., Ingole, J.: Corona-Nidaan: lightweight deep convolutional neural network for chest X-Ray based COVID-19 infection detection. Appl. Intell. **51**(5), 3026–3043 (2021). https://doi.org/10.1007/s10489-020-01978-9
14. Sedik, A., Hammad, M., Abd El-Samie, F.E., Gupta, B.B., Abd El-Latif, A.A.: Efficient deep learning approach for augmented detection of Coronavirus disease. Neural Comput. Appl., 1–18 (2021)
15. Narin, A., Kaya, C., Pamuk, Z.: Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849 (2020)
16. Sharma, A., Mishra, P.K.: Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. Int. J. Inf. Technol., 1–12 (2021)
17. Civit-Masot, J., Luna-Perejón, F., Domínguez Morales, M., Civit, A.: Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images. Appl. Sci. **10**(13), 4640 (2020)
18. Ibrahim, A.U., Ozsoz, M., Serte, S., Al-Turjman, F., Yakoi, P.S.: Pneumonia classification using deep learning from chest X-ray images during COVID-19. Cognit. Comput., 1–13 (2021)
19. 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)
20. Hemdan, E.E.D., Shouman, M.A., Karar, M.E.: COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in X-ray images. arXiv preprint arXiv:2003.11055 (2020)
21. Ayaz, M., Shaukat, F., Raja, G.: Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. Phys. Eng. Sci. Med. **44**(1), 183–194 (2021). https://doi.org/10.1007/s13246-020-00966-0 AQ2
22. Srivastava, A., Mishra, P.K.: A survey on WSN issues with its heuristics and meta-heuristics solutions. Wireless Pers. Commun. **121**(1), 745–814 (2021). https://doi.org/10.1007/s11277-021-08659-x
23. Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., Soufi, G.J.: Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. Med. Image Anal. **65**, 101794 (2020)
24. 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 (2020). https://doi.org/10.1007/s10489-020-01900-3
25. Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., Pinheiro, P.R.: CovidGAN: data augmentation using auxiliary classifier GAN for improved COVID-19 detection. IEEE Access **8**, 91916–91923 (2020)
26. 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)
27. Mishra, S.: Financial management and forecasting using business intelligence and big data analytic tools. Int. J. Financ. Eng. **5**(02), 1850011 (2018)

28. Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., Dong, Y.: The ensemble deep learning model for novel COVID-19 on CT images. Appl. Soft Comput. **98**, 106885 (2021)
29. Srivastava, A., Mishra, P.K.: State-of-the-art prototypes and future propensity stem on internet of things. Int. J. Recent Technol. Eng. (IJRTE) **8**(4), 2672– 2683 (2019). https://doi.org/10.35940/ijrte.D7291.118419
30. Loey, M., Smarandache, F., M Khalifa, N.E.: Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. Symmetry **12**(4), 651 (2020)
31. Wang, L., Lin, Z.Q., Wong, A.: COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci. Rep. **10**(1), 1–12 (2020)
32. Mangal, A., et al.: CovidAID: COVID-19 detection using chest X-ray. arXiv preprint arXiv:2004.09803 (2020)
33. Abbas, A., Abdelsamea, M.M., Gaber, M.M.: DeTrac: transfer learning of class decomposed medical images in convolutional neural networks. IEEE Access **8**, 74901–74913 (2020)
34. Stein, A.: Pneumonia dataset annotation methods. RSNA pneumonia detection challenge discussion, 2018 (2020). https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/discussion/. Accessed 5 Dec 2020
35. Mooney, P.: Kaggle chest X-ray images (pneumonia) dataset (2018). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed 5 Dec 2020
36. Jaeger, S., Candemir, S., Antani, S., Wáng, Y.X.J., Lu, P.X., Thoma, G.: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med. Surg. **4**(6), 475 (2014)
37. Cohen, J.P., Morrison, P., Dao, L.: COVID-19 image data collection. arXiv preprint arXiv:2003.11597 (2020)
38. Mishra, S., Tripathi, A.R.: IoT platform business model for innovative management systems. Int. J. Financ. Eng. (IJFE) **7**(03), 1–31 (2020)
39. Abdullah-Al-Wadud, M., Kabir, M.H., Dewan, M.A.A., Chae, O.: A dynamic histogram equalization for image contrast enhancement. IEEE Trans. Consum. Electron. **53**(2), 593–600 (2007)
40. Mishra, S., Tripathi, A.R.: AI business model: an integrative business approach. J. Innov. Entrep. **10**(1), 1–21 (2021). https://doi.org/10.1186/s13731-021-00157-5
41. Mishra, S., Triptahi, A.R.: Platforms oriented business and data analytics in digital ecosystem. Int. J. Financ. Eng. **6**(04), 1950036 (2019)
42. Anand, A., Pugalenthi, G., Fogel, G.B., Suganthan, P.N.: An approach for classification of highly imbalanced data using weighting and undersampling. Amino Acids **39**(5), 1385–1391 (2010)
43. Mishra, S., Tripathi, A.R.: Literature review on business prototypes for digital platform. J. Innov. Entrep. **9**(1), 1–19 (2020). https://doi.org/10.1186/s13731-020-00126-4
44. Huan, E.-Y., Wen, G.-H.: Transfer learning with deep convolutional neural network for constitution classification with face image. Multimedia Tools Appl. **79**(17–18), 11905–11919 (2020). https://doi.org/10.1007/s11042-019-08376-5
45. Sharma, A., Mishra, P.K.: State-of-the-art in performance metrics and future directions for data science algorithms. J. Sci. Res. **64**(2) (2020)
46. Rajinikanth, V., Joseph Raj, A.N., Thanaraj, K.P., Naik, G.R.: A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection. Appl. Sci. **10**(10), 3429 (2020)
47. Khan, S., Islam, N., Jan, Z., Din, I.U., Rodrigues, J.J.C.: A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recogn. Lett. **125**, 1–6 (2019)

48. Chaturvedi, S.S., Tembhurne, J.V., Diwan, T.: A multi-class skin cancer classification using deep convolutional neural networks. Multimedia Tools Appl. **79**(39–40), 28477–28498 (2020). https://doi.org/10.1007/s11042-020-09388-2
49. Ketu, S., Mishra, P.K.: A hybrid deep learning model for COVID-19 prediction and current status of clinical trials worldwide. Comput. Mater. Continua **66**(2) (2020)
50. Arias-Londoño, J.D., Gomez-Garcia, J.A., Moro-Velázquez, L., Godino-Llorente, J.I.: Artificial Intelligence applied to chest X-ray images for the automatic detection of COVID-19. A thoughtful evaluation approach. IEEE Access (2020)
51. Chaurasia, B., Verma, A.: A comprehensive study on failure detectors of distributed systems. J. Sci. Res. **64**(2) (2020)