



Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks

Narinder Singh Punn¹ · Sonali Agarwal¹

Published online: 17 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

The novel coronavirus 2019 (COVID-19) is a respiratory syndrome that resembles pneumonia. The current diagnostic procedure of COVID-19 follows reverse-transcriptase polymerase chain reaction (RT-PCR) based approach which however is less sensitive to identify the virus at the initial stage. Hence, a more robust and alternate diagnosis technique is desirable. Recently, with the release of publicly available datasets of corona positive patients comprising of computed tomography (CT) and chest X-ray (CXR) imaging; scientists, researchers and healthcare experts are contributing for faster and automated diagnosis of COVID-19 by identifying pulmonary infections using deep learning approaches to achieve better cure and treatment. These datasets have limited samples concerned with the positive COVID-19 cases, which raise the challenge for unbiased learning. Following from this context, this article presents the random oversampling and weighted class loss function approach for unbiased fine-tuned learning (transfer learning) in various state-of-the-art deep learning approaches such as baseline ResNet, Inception-v3, Inception ResNet-v2, DenseNet169, and NASNetLarge to perform binary classification (as normal and COVID-19 cases) and also multi-class classification (as COVID-19, pneumonia, and normal case) of posteroanterior CXR images. Accuracy, precision, recall, loss, and area under the curve (AUC) are utilized to evaluate the performance of the models. Considering the experimental results, the performance of each model is scenario dependent; however, NASNetLarge displayed better scores in contrast to other architectures, which is further compared with other recently proposed approaches. This article also added the visual explanation to illustrate the basis of model classification and perception of COVID-19 in CXR images.

Keywords COVID-19 · Classification · Deep learning · Transfer learning · Pneumonia · Chest X-ray (CXR) · Imbalanced learning

1 Introduction

Coronaviruses are a large family of viruses that can cause severe illness to the human being. The first known severe epidemic is severe acute respiratory syndrome (SARS) occurred in 2003, whereas the second outbreak began in 2012 in Saudi Arabia with the middle east respiratory

syndrome (MERS). The novel coronavirus disease 2019 (COVID-19), began as an outbreak from epicentre Wuhan, People's Republic of China in late December 2019, and till April 15, 2020, it caused 1,996,681 infections and 127,590 deaths worldwide [58]. The coronavirus (COVID-19) outbreak was declared a public health emergency of international concern by WHO on January 30, 2020 [59]. On March 11, as the number of COVID-19 cases has increased thirteen times apart from China with more than 118,000 cases in 114 countries and over 4,000 deaths, WHO declared this a pandemic [38]. Globally, many researchers of medicine, clinical and artificial intelligence areas are trying hard to mobilize preventive action plans for COVID-19 with identified research priorities. Since this disease is highly contagious, the most desirable preventive measure is to identify the infected people to control the spread. Unfortunately, there is no well-known treatment available to cure COVID-19, therefore the identified infected person

This article belongs to the Topical Collection: *Artificial Intelligence Applications for COVID-19, Detection, Control, Prediction, and Diagnosis*

✉ Narinder Singh Punn
pse2017002@iiita.ac.in

Sonali Agarwal
sonali@iiita.ac.in

¹ IIT Allahabad, Prayagraj 211015, India

must be kept in isolation to break the transmission chain as this patient may become the source of community transfer. Till now, the testing kit is the only available option for diagnosis of COVID-19. Unavailability of testing kits due to excessive demand all over the world is a severe problem in the mission against this pandemic. Though several healthcare organizations are claiming for successful development of testing kits, there is a huge gap in demand and supply. The healthcare agencies have accelerated the rate of development of low-cost testing kits, but the inability to diagnose at early-stage and due to exponential growth of COVID-19 cases, medical professionals are bound to rely on other diagnostic measures.

Clinical studies have shown that most COVID-19 patients suffer from lung infection [13]. Although chest CT is a more effective imaging technique for lung-related disease diagnosis; CXR is preferred because it is widely available, faster and cheaper than CT. Since COVID-19 infection attacks the epithelial cells that line our respiratory tract, hence X-rays images can be used to analyse the lungs to diagnose pneumonia, lung inflammation, abscesses, and/or enlarged lymph nodes [2]. Due to its easy transmission, developing techniques to accurately and easily identify the presence of COVID-19 and distinguish it from other forms of flu and pneumonia is crucial.

Biomedical image analysis (segmentation and classification) is an admired area of research to make the healthcare system more promising [11]. In this area, advancement in computing infrastructure makes it possible to deploy the deep learning techniques for complex medical image analysis tasks. Recent works have shown that the chest X-rays of patients suffering from COVID-19 depicts certain abnormalities in the radiography [39]. For medical image analysis; deep learning techniques, specifically, convolutional neural networks (CNN) are very effective and efficient in feature extraction and learning, hence becoming the most popular choice among researchers [25]. CNNs have been successfully deployed in the analysis of video endoscopy [12] and CT images, and also used for the diagnosis of pediatric pneumonia via chest X-ray images [6, 22]. Chouhan et al. [7] proposed a transfer learning based deep network approach pre-trained on ImageNet [9] for pneumonia detection. Wang et al. [57] proposed a customized VGG16 model for lung regions identification to classify different types of pneumonia. Later, Ronneburger et al. [42] demonstrated the effectiveness of image augmentation with CNN in the presence of a small set of images. In the area of biomedical image classification, Rajpurkar et al. [40] proposed a dense CNN with 121-layers to detect several pathologies including pneumonia using chest X-rays. Lakhani et al. [26] obtained an area under the curve (AUC) of 0.95 in pneumonia detection using AlexNet and GoogLeNet along with image augmentation. Recently,

several deep-learning based COVID-19 detection techniques have been proposed [20, 55, 56]. Linda et al. [55] introduced a deep CNN, named COVID-Net for the detection of COVID-19 cases from the chest X-ray images. Shuai et al. [56] achieved accuracy, specificity and sensitivity of 89.5%, 88% and 87% respectively for COVID-19 identification using CT images.

There are many datasets available about chest X-rays for the detection of pneumonia [8, 18, 32, 50]; but in present research work, COVID-19 X-ray chest images, [8] and Radiological Society of North America (RSNA) images [50] are utilized to generate all possible samples of chest infection and also to make the study comparable with other research works. The number of COVID-19 infected samples present in this dataset is very limited that may lead to biased outcome, hence the objective of this research is to maximize the learning ability in presence of a small set of positive class samples. For early diagnosis of COVID-19, this article presents the effectiveness of random oversampling and weighted class loss function approaches for unbiased fine-tuned learning (transfer learning) in various state-of-the-art deep learning techniques.

Rest of the manuscript is organized as follows: recent research articles are discussed in Section 2, and Section 3 briefs the dataset. Section 4 contains the proposed methodology followed by the evaluation metrics in Section 5. Results are discussed in Section 6 whereas the last section contains concluding remarks.

2 Related work

Due to the ample availability of X-ray machines, disease diagnosis using CXR images are widely used by healthcare experts. In case of any suspect of COVID-19; instead of using test kits, an alternate way to detect pneumonia from the CXR images is required, so that further investigation can be narrowed down for COVID-19 identification. Many studies have been performed on similar ground with several CXR datasets for diagnosis of pneumonia and other complications [8, 18, 32, 50]. These studies also advocate the need of an automated system for quick diagnosis, because the manual methods of X-ray analysis are time consuming and unable to serve the purpose due to limited availability of X-ray machine operators or radiologists.

Amid the COVID-19 outbreak, many companies at the global level around the world embraced a flurry of Artificial Intelligence (AI) based solutions to detect COVID-19 on chest X-ray scans. It is evident that deep learning tools are effectively used to screen mild cases, triage new infections, and monitor disease advancements. This way of diagnosis can reduce the growing burden on radiologists, and also supplant standard nucleic acid tests as the primary

diagnostic tool for coronavirus infection. It is also reported that a swab test needs isolation for testing procedure, whereas chest X-ray based detection can be easily manageable. Kermany et al. [22] proposed CXR image-based deep learning model to detect pneumonia and classify other diseases using different medical datasets with testing accuracy of 92.80%. In another similar research, Stephen et al. [51] illustrated an efficient deep learning approach for pneumonia classification, by using four convolutional layers and two dense layers in addition to classical image augmentation and achieved 93.73% testing accuracy. Later, Saraiva et al. [44] experimented convolutional neural networks to classify images of childhood pneumonia by using a deep learning model with seven convolutional layers along with three dense layers while achieving 95.30% testing accuracy. Liang and Zheng [27] demonstrated a transfer learning method with a deep residual network for pediatric pneumonia diagnosis with 49 convolutional layers and two dense layers and achieved 96.70% testing accuracy. In similar research, Wu et al. [60] focused on convolutional deep neural learning networks and random forest to propose a pneumonia prediction using CXR images and achieved 97% testing accuracy.

Afterwards, Narin et al. [31] proposed a deep convolutional neural network based automatic prediction model of COVID-19 with the help of pre-trained transfer models using CXR images. In this research, authors used ResNet50, InceptionV3 and Inception-ResNetV2 pre-trained models to obtain a higher prediction accuracy for a subset of X-ray dataset. Apostolopoulos et al. [3] in their study, utilised state-of-the-art convolutional neural network architectures for classifying the CXR images. Transfer Learning was adopted to handle various abnormalities present in the dataset. Two datasets from different repositories have been used to study images of three classes: COVID-19, bacterial/viral pneumonia and normal condition. The article establishes the suitability of the deep learning model with the help of accuracy, sensitivity, and specificity parameters.

In another research, generative adversarial networks (GAN) are used by Khalifa et al. [23] to detect pneumonia from CXR images. The authors addressed the overfitting problem and claimed its robustness by generating more images through GAN. The dataset containing 5863 CXR images of two categories: normal and pneumonia, has been used with typical deep learning models such as AlexNet, GoogLeNet, SqueezeNet and Resnet18 to detect pneumonia. This research highlights that the Resnet18 outperformed among other deep transfer models in combination with GAN. Further, Sethy et al. [46] proposed a deep learning based model to identify coronavirus infections using CXR images. Deep features from CXR images have been extracted and support vector machine (SVM) classifier

is used to measure accuracy, false positive rate, F1 score, Matthew's correlation coefficient (MCC) and kappa. It is found that ResNet50 in combination with SVM is statistically superior when compared to other models. Later, Bukhari et al. [4] also used ResNet-50 CNN architectures on 278 CXR images, partitioned under 3 groups as normal, pneumonia and COVID-19. This approach gave promising results and indicated substantial differentiation of pulmonary changes caused by COVID-19 from the other types of pneumonia.

Recently, in another research work, an improved ResNet-50 CNN architecture named COVIDResNet has been proposed [10], where conventional ResNet-50 model is applied with different training techniques including progressive resizing, cyclical learning rate finding, and discriminative learning rates to gain fast and accurate training. The experiment is performed through progressively re-sizing of input images to $128 \times 128 \times 3$, $224 \times 224 \times 3$ and $229 \times 229 \times 3$ pixels, and automatic learning rate selection for fine-tuning the network at each stage. This work claimed to be computationally efficient and highly accurate for multi-class classification.

A new deep anomaly detection model is developed by Zhang et al. [62] for fast and more reliable screening. To evaluate the model performance, CXR image data of COVID-19 cases and other pneumonia has been collected from two different sources. To eliminate the data imbalance problem in the collected samples, authors proposed a CXR based COVID-19 screening model through anomaly detection task [35].

Following this context, this article proposes to contribute for early diagnosis of COVID-19 using the state-of-the-art deep learning architectures, assisted with transfer learning and class imbalance learning approaches.

3 Dataset description

In this research three datasets are utilized for experiments: COVID-19 image [8], Radiological Society of North America (RSNA) [50] and U.S. national library of medicine (USNLM) collected Montgomery country - NLM(MC) [19]. COVID-19 image dataset is a public database of pneumonia cases with CXR images related to COVID-19, MERS, SARS, and ARDS collected by Cohen et al. [8] from multiple resources available at public domains without infringing patient's confidentiality (Fig. 1b). It is claimed that this dataset can help to identify characteristics of COVID-19 in contrast to other types of pneumonia; therefore it can play a major role in predicting survival rate. The dataset includes the statistics up to March 25, 2020 consisting of 5 types of pneumonia such as SARS-CoV-2 or COVID-19, SARS-CoV-1 or SARS, Streptococcus spp.,

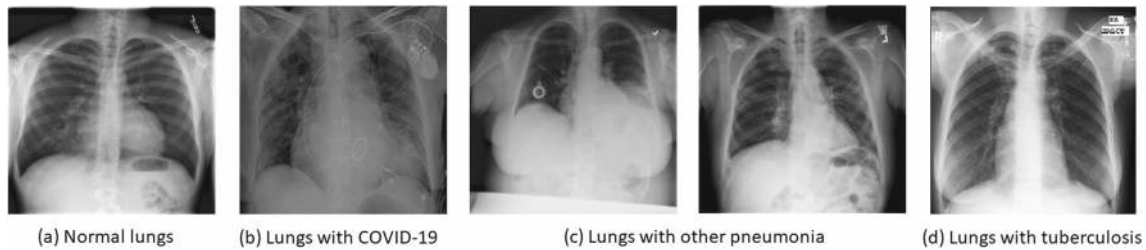


Fig. 1 Sample chest radiographs

Pneumocystis spp. and ARDS with following attributes: patient ID, offset, sex, age, finding, survival, view, modality, date, location, filename, doi, url, license, clinical notes, and other notes.

Another dataset utilized in this study is published under RSNA pneumonia detection challenge is a subset of 30,000 examinations taken from the NIH CXR14 dataset [50]. Out of 30,000 selected images, 15,000 examinations had positive cases of pneumonia and from the remaining 15,000 cases, 7,500 cases had no findings and other 7,500 cases had symptoms other than pneumonia. All these images are annotated by a group of experts including radiologists in two stages. A sample image is shown in Fig. 1c. This dataset has been published in two stages. In Stage one, 25,684 training images were considered to test 1,000 images. Later in stage two 1000 testing samples were added to the training set to form the dataset of 26,684 training images and a new set of 3,000 radiographs were introduced for the test. For robust testing and comprehensive coverage of the comparative analysis, NLM(MC) [19] dataset is also utilized that consists of 138 chest posterior-anterior x-rays samples of tuberculosis and normal cases. A sample image is represented in Fig. 1d. Table 1 presents the class summary details of the fused dataset resulting from the above discussed datasets which is utilized for training, testing and validation of the proposed approach. The fused dataset is composed of 1214 posteroanterior chest x-ray samples with classes labeled as COVID-19 (108), other pneumonia (515), tuberculosis (58) and normal (533). The generated fused dataset is publicly available [36].

Table 1 Fused dataset disease class summary details

Dataset	Findings	PCXR images	Total samples
COVID-19	COVID-19	108	153
	Other pneumonia	45	
RSNA	Normal	453	923
	Other pneumonia	470	
NLM(MC)	Tuberculosis	58	138
	Normal	80	

4 Proposed contribution

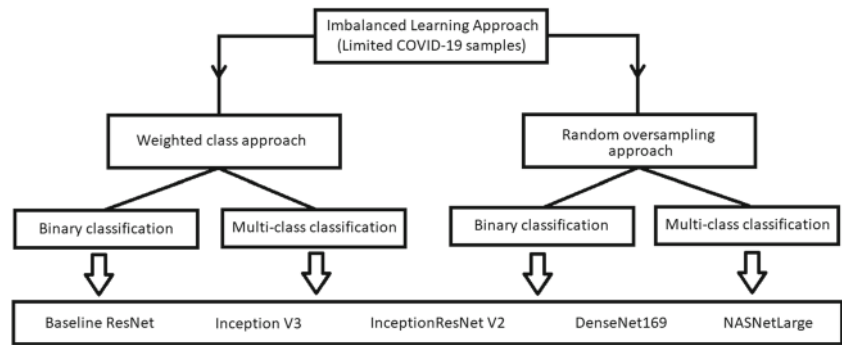
The era of artificial intelligence has brought significant improvements in the living society [30]. The recent advancements in deep learning have extended its domain in various applications such as healthcare, pixel restoration, visual recognition, signal processing, and a lot more [28]. In healthcare domain, the deep learning based image processing approaches for classification and segmentation are applied for faster, efficient, and early diagnosis of the deadly diseases e.g. breast cancer, brain tumor, etc. by using different imaging modalities such as X-ray, CT, MRI, [47] and fused modalities [37] along with its future possibilities. The success of these approaches is dependent on the large amount of data availability, which however is not in the case of automated COVID-19 detection.

The main contribution of the work is divided into the components as shown in Fig. 2. It has two concrete components: data preprocessing and classification. COVID-19 image, RSNA and NLM(MC) datasets are used to generate the final working set. The newly generated dataset contains CXR images of the following classes: coronavirus caused diseases, pneumonia, other diseases and normal cases. Further, binary classification (COVID-19 vs others) and multi-class classification (COVID-19, other types of pneumonia, tuberculosis and normal) are achieved using random oversampling and weighted class loss function approaches for unbiased fine-tuned learning (transfer learning) in various state-of-the-art deep learning approaches such as baseline ResNet, Inception-v3, Inception ResNet-v2, DenseNet169, and NASNetLarge [17, 52, 54, 63]. The trained models are utilized for identification and classification of COVID-19 in novel samples. Later, visualization techniques are utilized to understand and elaborate the basis of the classification results.

4.1 Imbalanced learning approach

Class balancing techniques are necessary when minority classes are more important. The dataset used in this research is highly imbalanced which may lead to biased learning of the model. Number of coronavirus infected CXR images are

Fig. 2 Schematic representation of the proposed components for COVID-19 identification



very less compared to other classes, hence class balancing techniques must be insured to smoothen the learning process. This section discusses two approaches to handle the class imbalance problem: weight class approach and random oversampling [21].

4.1.1 Weighted class approach

In this approach, the intention is to balance the data by altering the weights that each training sample class carries when computing the loss. Normally, each class carries equal weights, but sometimes certain classes with minority samples are required to hold more weights if they are more important because training examples within that class should have a significant effect on the loss function. In the used dataset the coronavirus infected image class samples must be given more weights as they are more significant. In this article, the weights for each class is generated based on the Eq. (1).

$$w(c) = C_c \cdot \frac{\sum_{c=0}^N n_c}{N \cdot n_c} \tag{1}$$

where C_c is the class constant for a class c , N is the number of classes, and n_c is the number of samples in a class c . The computed class weights are later fused with the objective function (loss function) of the deep learning model in order to heavily penalize the false predictions concerned with the minority samples, which in this case is coronavirus.

4.1.2 Random oversampling approach

In this approach, the objective is to increase the number of minority samples by utilizing the existing samples belonging to the minority class. The minority samples are increased until the samples associated with every class become equal. Hence the procedure follows by identifying the difference between the number of samples in majority and minority class. To fill this void of difference, the samples are generated from the randomly selected sample belonging to the minority class by applying certain

statistical operations. In this work, the samples of CXR image of COVID-19 positive cases are less as compared to other classes, therefore, these minority class images are randomly oversampled by means of rotation, scaling, and displacement with the objective to achieve equal distribution of classes and accommodate unbiased learning among the deep learning models.

4.2 Classification Strategy

Based on the type of data samples availability of CXR images the COVID-19 classification is divided into two following schemes:

- Binary Classification - In this classification scheme, the coronavirus positive samples labelled as “1” (COVID-19) are identified against the rest of the samples labelled as “0” (non COVID-19 case) which involves other cases e.g. chlamyidophila, SARS, streptococcus, tuberculosis, etc., along with the normal cases.
- Multi-class Classification—In this classification scheme, the aim is to distinguish and identify the COVID-19 samples from the other pneumonia cases along with the presence of tuberculosis and normal case findings. The multi-class classification is performed with three and four classes. The three classes are provided with labels as “0” being a normal case, “1” being a COVID-19 case, and “2” being other pneumonia and tuberculosis cases, whereas four classes are labeled as “0” being a normal case, “1” being a COVID-19 case, and “2” being other pneumonia case and “3” as tuberculosis case.

In both the classification strategies, the deep learning models are trained with the above discussed imbalanced learning approaches using the weighted categorical cross entropy (WCE) loss function as given by Eq. (2) and Eq. 3 [24]:

$$f(s)_i = \frac{e^{s_i}}{\sum_C e^{s_i}} \tag{2}$$

$$WCE = -\sum_i^C w(i).t_i \cdot \log(f(s)_i) \tag{3}$$

In categorical cross entropy, the distribution of the predictions (the activations in the output layer, one for each class) is compared with the true distribution only, to ensure the clear representation of the true class as one-hot encoded vector; here, closer the model’s outputs are to that vector, the lower the loss.

4.3 Data Preprocessing

In this article, due to the limited samples of posteroanterior chest X-ray images concerned with positive COVID-19 [8] cases, the data samples are mixed with the other randomly selected CXR images selected from other datasets-, RSNA [50] and NLM(MC) [19]. The RSNA and NLM(MC) datasets consists of posteroanterior CXR images covering sample cases labelled as pneumonia and tuberculosis respectively along with normal samples. Table 2 describes the distribution of training, testing, and validation sets using the fused dataset for binary and multi-class classification along with different class imbalance strategies i.e. class weighted loss function that penalizes the model for any false negative prediction and random oversampling [61] of minority classes which in this case is COVID-19.

The CXR images in the aggregated dataset also consists of unwanted artifacts such as bright texts, symbols, varying resolutions and pixel level noise, which necessitates its preprocessing. In order to suppress the highlighted textual and symbolic noise, the images are inpainted with the image mask generated using binary thresholding [34] as given by Eq. (4), followed by resizing the images to a fixed size resolution of 331 × 331 × 3.

$$M(x, y) = \begin{cases} max_th, & i(x, y) \geq min_th. \\ 0, & otherwise. \end{cases} \tag{4}$$

where $i(x,y)$ is an input image, max_{th} and min_{th} are max and min thresholds to design the mask. Despite filtering the unwanted information, there is still the possibility of uncertainty at the deep pixel level representation [15]. The denoising or removal of such uncertainty is carried through the adaptive total variation method [53] while preserving the original distribution of pixel values.

Let for a given grayscale image f , on a bounded set Ω over \mathbb{R}^2 , where $\Omega \subset \mathbb{R}^2$, denoising image u that closely matches to observed image $x = (x_1, x_2) \in \Omega$ - pixels, given as

$$u = \arg \min_u \left(\int_{\Omega} (u - f \cdot \ln u) dx + \int_{\Omega} (\omega(x) | \nabla u | dx) \right) \tag{5}$$

where $\omega(x) = \frac{1}{1+k \text{ mod } G_{\sigma} * \nabla u}$, G_{ρ} - the Gaussian kernel for smoothing with σ variance, $k > 0$ is contrast parameter and $*$ is convolution operator.

Figure 3 illustrates the data preprocessing stages by considering an instance of COVID-19 case consisting of textual and symbolic artifacts from the generated dataset. The resulting distributed pixels histograms at each stage of preprocessing shown in Fig. 3, illustrates that the preprocessing approach tends to preserve the original nature of distribution of the pixels while removing the irregular intensities. The preprocessed images are then divided into training, testing, and validation set for training and evaluation of the state-of-the-art deep learning classification models.

4.4 Deep learning models

This section incorporates the state-of-the-art deep learning models utilized in the present research work as shown in Table 3 along with their respective contribution, parameters, and performance on the standard benchmark datasets. The inception deep convolutional architectures

Table 2 Posteroanterior CXR images distribution into training, validation, and test sets from the fused datasets for different problem definitions

Findings	Class imbalanced learning strategy																	
	Classification with class weighted loss function									Classification with random oversampling								
	Binary (2 classes)			Multi-class (3 classes)			Multi-class (4 classes)			Binary (2 classes)			Multi-class (3 classes)			Multi-class (4 classes)		
	Tr	Val	Tst	Tr	Val	Tst	Tr	Val	Tst	Tr	Val	Tst	Tr	Val	Tst	Tr	Val	Tst
Normal	906	90	110	437	44	52	437	44	52	960	90	110	469	44	52	437	44	52
Tuberculosis				469	46	58	47	5	6				469	46	58	437	5	6
Other pneumonia							422	41	52							437	41	52
COVID-19	88	9	11	88	9	11	88	9	11	960	9	11	469	9	11	437	9	11
Total	994	99	121	994	99	121	994	99	121	1920	99	121	1407	99	121	1748	99	121

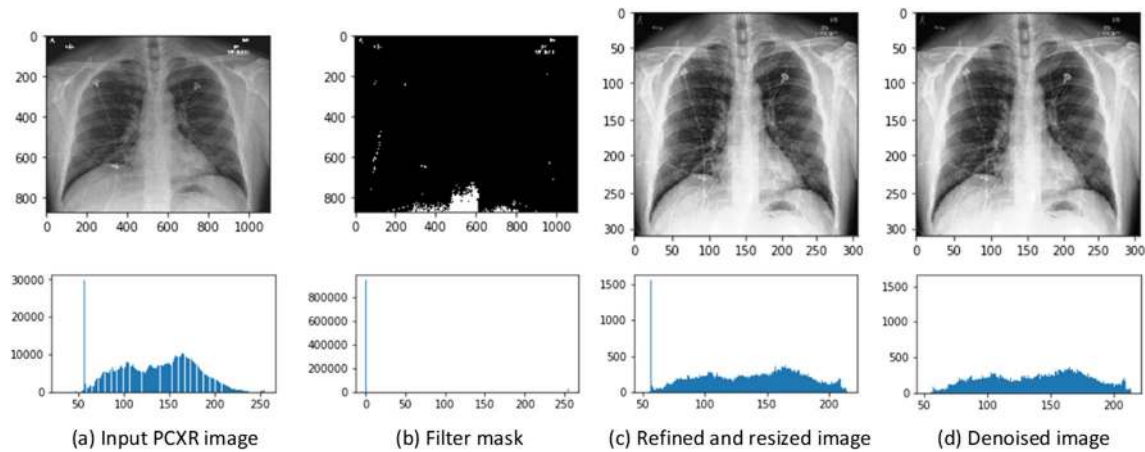


Fig. 3 Data preprocessing stages of raw posteroanterior CXR image

proposed by GoogLeNet are considered as the state-of-the-art deep learning architectures for image analysis and object identification with the basic model as inception-v1 [5]. Later, this base model was refined by introducing the batch normalization and established as the inception-v2 [54]. In further iterations, additional factorisation was introduced and released as the inception-v3. It is one of the pre-trained models to perform two types of specific tasks: dimensionality reduction using CNN and classification using fully-connected and softmax layers. Since it is originally trained on over a million images consisting of 1,000 classes of ImageNet, its head layers can be retrained for the generated dataset using historical knowledge to reduce the extensive training and computational power.

Later, Inception-ResNet-v2 was proposed by Szegedy et al. [52]. This hybrid model is a combination of residual connections and a recent version of Inception architecture. It is intended to train very deep convolutional models by the additive merging of signals, both for image recognition and object detection. This network is more robust and learns rich feature representations. Afterwards, DenseNet was proposed by Huang et al. [17]. It works on the concept of reuse, in which each layer receives inputs from all previous layers and yields a condensed model to pass its own feature-maps to all subsequent layers. This makes the network thinner and compact with the fewer number of channels, while improving variation in the input of subsequent layers, and becomes easy to train and highly parameter efficient.

Table 3 Recent deep learning architectures that are reported with top 5 error rate per year

Architecture	Year	Contribution	Param.	Dataset	E. rate
Inception-v3 [54]	2015	Avoided the bottleneck representations, Dimension reduction promotes faster learning.	23.6M	ILSVRC	5.6
Inception ResNet-v2 [52]	2016	Focused on residual connection rather than filter connection approach.	56M	ILSVRC	4.9
DenseNet169 [17]	2017	Dense connection blocks for activation flow across the layers.	17.8M	CIFAR-10 CIFAR-100 CIFAR-10+ CIFAR-100+	5.19 19.64 3.46 17.18
NASNetLarge [63]	2018	Search and transfer the architecture block from small dataset to large dataset, Scheduled drop path regularization.	22.6M	CIFAR-10	2.4

Google Brain research team proposed NASNet model based on reinforcement learning search methods [63]. It creates search space by factoring the network into cells and further dividing it into a number of blocks. Each block is supported by a set of popular operations in CNN models with various kernel size e.g: convolutions, max pooling, average pooling, dilated convolution, depth-wise separable convolutions etc.

4.5 Fine tuning

Training deep learning models require (Inception-v3, InceptionResNet-v2, etc.) exhaustive amount of resources and time. While these networks, attain relatively excellent performance on ImageNet [9], training them on a CPU is an exercise in futility. These CNNs are often trained for a couple of weeks or more using arrays of GPUs to get good results on the complex and complicated datasets. In most deep CNNs the first few convolution layers learn low-level features (edges, curves, blobs) and with progress, through the network, it learns more mid/high-level features or patterns associated with the on-going task. In fine tuning, the aim is to keep or freeze these trained low-level features, and only train the high-level features needed for our new image classification problem.

In this article, the trials of training the deep learning classification models initiates with the baseline residual network that is composed of five residual blocks [16], defined by two convolutions whose rectified linear unit (ReLU) activation is concatenated with the input of the first convolution layer, followed by max-pooling and instance normalization, except the final output layer which uses softmax activation function. Later, transfer learning is utilized on the state-of-the-art architectures discussed in Table 3 to utilize the pre-trained models while fine tuning the head layers to serve the purpose of classifying the COVID-19 samples, as illustrated in Fig. 4. This follows by enabling the four head layers of the network to adjust

its trainable parameters along with the addition of fully connected layers with 128 neurons and 2 or 3 neurons in the output layer depending on binary or multi-class classification, accompanied with the softmax activation function.

5 Evaluation metrics

The deep learning models are trained using the training and validation set under consideration of each possible combination of the discussed approaches of class imbalance learning and classification strategy, thereby making four possible scenarios for training a model. Later, the trained models are evaluated on the test set using the standard benchmark performance metrics such as accuracy, precision, recall (selectivity), area under curve (AUC), specificity and F1 score (dice coefficient) as shown in Fig. 5.

6 Training and testing

The proposed approach is trained and tested on the above discussed fused dataset using mini-batch gradient descent with learning rate optimizer as adam [43] via N-Vidia Titan GPU. The training, testing, and validation sets are highlighted in Table 2. The training phase is assisted with the batch size of 10, 4-fold cross validation, switch normalization [29] to adapt to instance or batch or layer normalization, and earlystopping technique that aims to stop the training process if the validation error stops decreasing. For each epoch the above discussed evaluation metrics are computed on the training and validation set to analyse the model's performance and improve the classification results. Later, the test set is utilized to evaluate the results of the proposed approach.

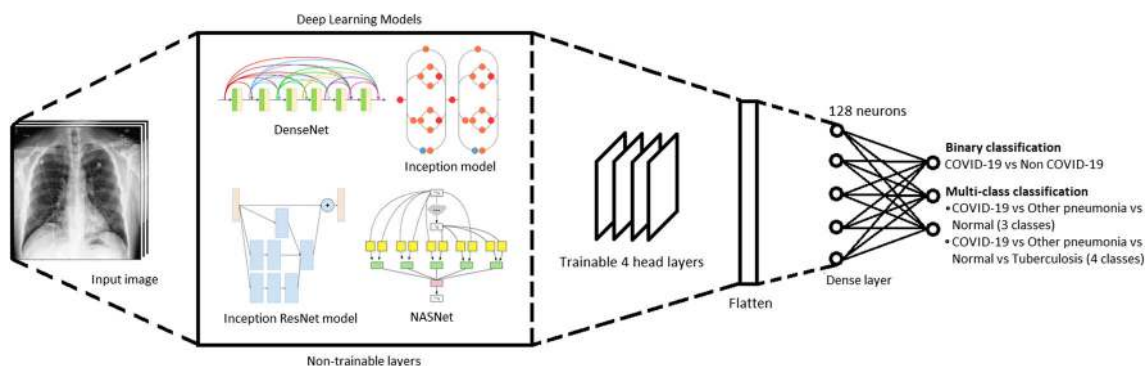
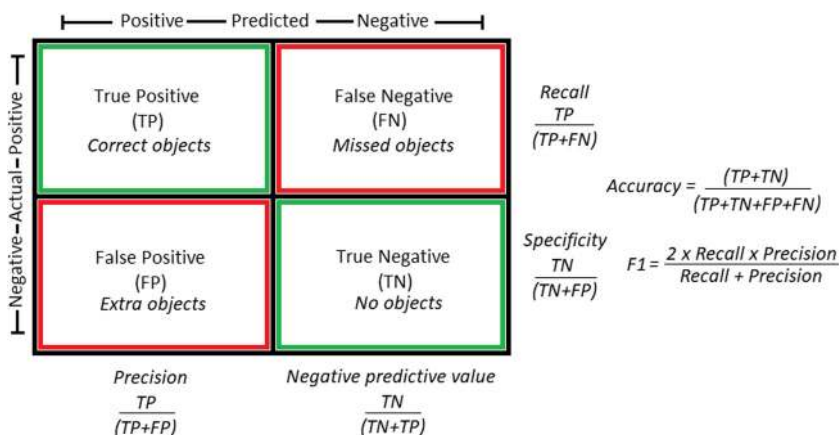


Fig. 4 Schematic representation of training framework for deep learning architectures via transfer learning

Fig. 5 Confusion matrix and performance evaluation metrics



7 Results and discussion

s With extensive trials, it is observed that there is no individual model that displayed best performance for all the scenarios in terms of accuracy, precision, recall, specificity, AUC, and F1-score. However, the NASNetLarge mostly claimed the best scores, thus making it best fit for the classification of COVID-19 samples. It is also observed that the results of binary classification (COVID-19 vs non-COVID-19) are better than the multi-class classification (COVID-19 vs other classes). With this it is evident that by grouping other diseases together as non-COVID-19 samples, models can efficiently learn features and patterns belonging to the COVID-19. Figure 6 shows the average performance curves of the NASNetLarge model evaluated and monitored on the test set for each iteration during

the training phase for classification of COVID-19 samples, generated using the tensorboard. Since the models are initialized with the trained weights on the ImageNet and head layers are fine-tuned for classification, the training of the models initiates with descent metrics values. The best scores of the models corresponding to each scenario are highlighted in Table 4 along with the detailed comparative analysis of the results for classifying the COVID-19 samples in terms of discussed evaluation metrics for different scenarios CB, CM₃, CM₄ RB, RM₃ and RM₄ where C and R indicates class imbalanced learning approaches, namely weighted class loss functions and random oversampling of the minority class, whereas B indicates the binary classification and M₃ (3 classes), M₄ (4 classes) indicate multi-class classification schemes. Furthermore, the results obtained are compared with the other recently proposed

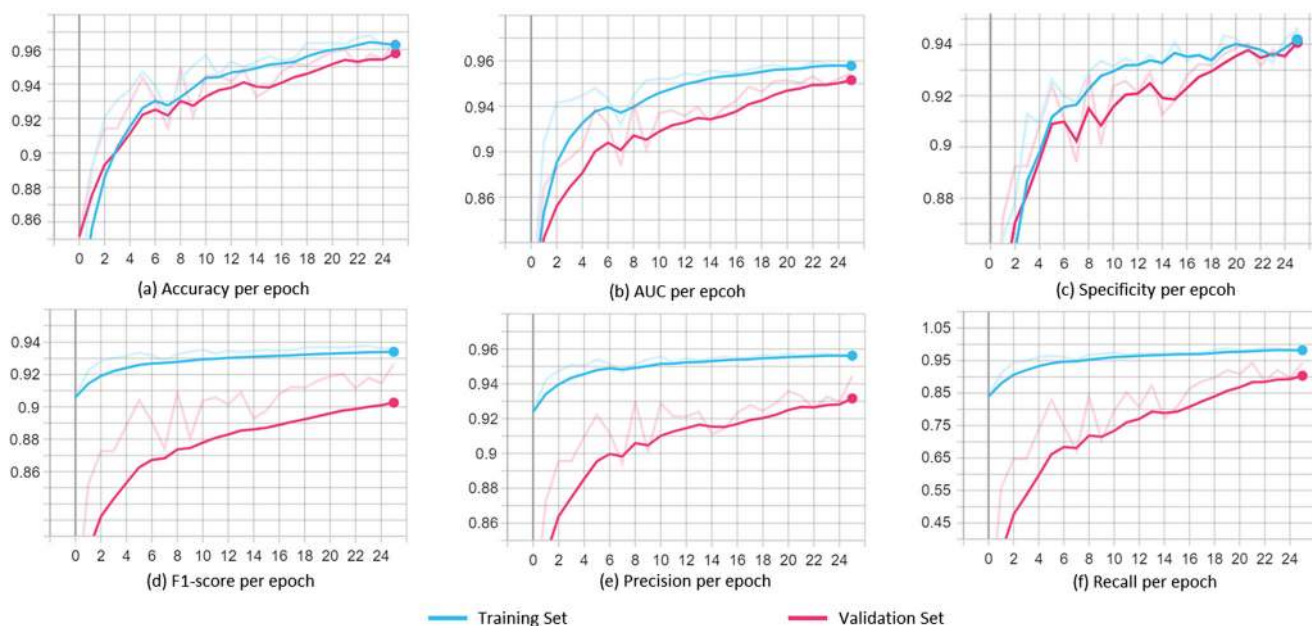


Fig. 6 Monitoring every epoch of training with performance curves for the NASNetLarge model with average **a** accuracy, **b** AUC, **c** specificity, **d** F1-score, **e** precision, and **f** recall. The curves are smoothed using moving exponential average

Table 4 Comparative analysis of the classification results of the deep learning models for each scenario

Model	Scenario	Accuracy	Precision	Recall	AUC	Specificity	F1-score
Baseline ResNet	CB	0.84	0.67	0.89	0.90	0.85	0.76
	CM ₃	0.77	0.62	0.89	0.90	0.77	0.73
	CM ₄	0.77	0.61	0.83	0.90	0.77	0.70
	RB	0.89	0.56	0.56	0.89	0.60	0.56
	RM ₃	0.76	0.63	0.42	0.83	0.77	0.50
	RM ₄	0.76	0.65	0.43	0.82	0.73	0.52
Inception v2	CB	0.88	0.77	0.79	0.93	0.89	0.78
	CM ₃	0.87	0.73	0.88	0.90	0.88	0.80
	CM ₄	0.87	0.73	0.85	0.89	0.85	0.79
	RB	0.88	0.78	0.70	0.88	0.91	0.74
	RM ₃	0.87	0.90	0.81	0.91	0.90	0.85
	RM ₄	0.84	0.89	0.81	0.91	0.88	0.85
Inception ResNet v2	CB	0.95	0.86	0.66	0.98	0.87	0.75
	CM ₃	0.92	0.85	0.92	0.99	0.89	0.88
	CM ₄	0.91	0.83	0.91	0.98	0.89	0.87
	RB	0.90	0.85	0.75	0.93	0.88	0.80
	RM ₃	0.92	0.97	0.96	0.99	0.94	0.96
	RM ₄	0.91	0.95	0.93	0.98	0.93	0.94
DenseNet 169	CB	0.90	0.82	0.78	0.91	0.91	0.80
	CM ₃	0.87	0.87	0.91	0.95	0.89	0.89
	CM ₄	0.86	0.87	0.89	0.93	0.89	0.88
	RB	0.95	0.94	0.96	0.97	0.95	0.95
	RM ₃	0.91	0.93	0.95	0.93	0.95	0.94
	RM ₄	0.92	0.94	0.95	0.94	0.95	0.94
NASNet Large	CB	0.97	0.82	0.91	0.99	0.98	0.86
	CM ₃	0.94	0.89	0.91	0.94	0.89	0.90
	CM ₄	0.95	0.88	0.89	0.92	0.89	0.88
	RB	0.98	0.87	0.90	0.99	0.94	0.98
	RM ₃	0.96	0.93	0.91	0.96	0.94	0.92
	RM ₄	0.95	0.95	0.90	0.95	0.94	0.92

Bold numbers highlight the best scores of the models corresponding to each scenario

approaches to detect COVID-19 positive symptoms via X-ray images as shown in Table 5. It is observed that the

proposed approach outperforms other approaches, whereas Apostolopoulos et al. [3] achieved similar results with the

Table 5 Comparative analysis of the proposed approach with recently proposed strategies

Authors	Technique	Network	Classes	Accuracy	Precision	Recall	AUC	Specificity
Yujin et al. [33]	Fine tuning	Res-Net18	4	0.88	0.83	0.86	–	0.96
Afshar et al. [1]	Fine tuning	Capsule		0.95	–	0.90		0.95
Proposed	Fine	NASNet Large		0.95	0.95	0.90	0.94	0.92
Wang et al. [55]	Full training	COVID-Net	3	0.92	0.91	0.88	–	–
Apostolopoulos et al. [3]	Fine tuning	VGG19		0.87	–	0.92	–	0.98
Proposed	Fine tuning	NASNet Large		0.96	0.93	0.91	0.96	0.94
Hall et al. [14]	Fine tuning	Ensemble	2	0.91	–	0.78	0.94	0.93
Apostolopoulos et al. [3]	Fine tuning	VGG19		0.98	–	0.92	–	0.98
Proposed	Fine tuning	NASNet Large		0.98	0.87	0.90	0.99	0.98

Bold numbers highlight the results obtained using proposed approach

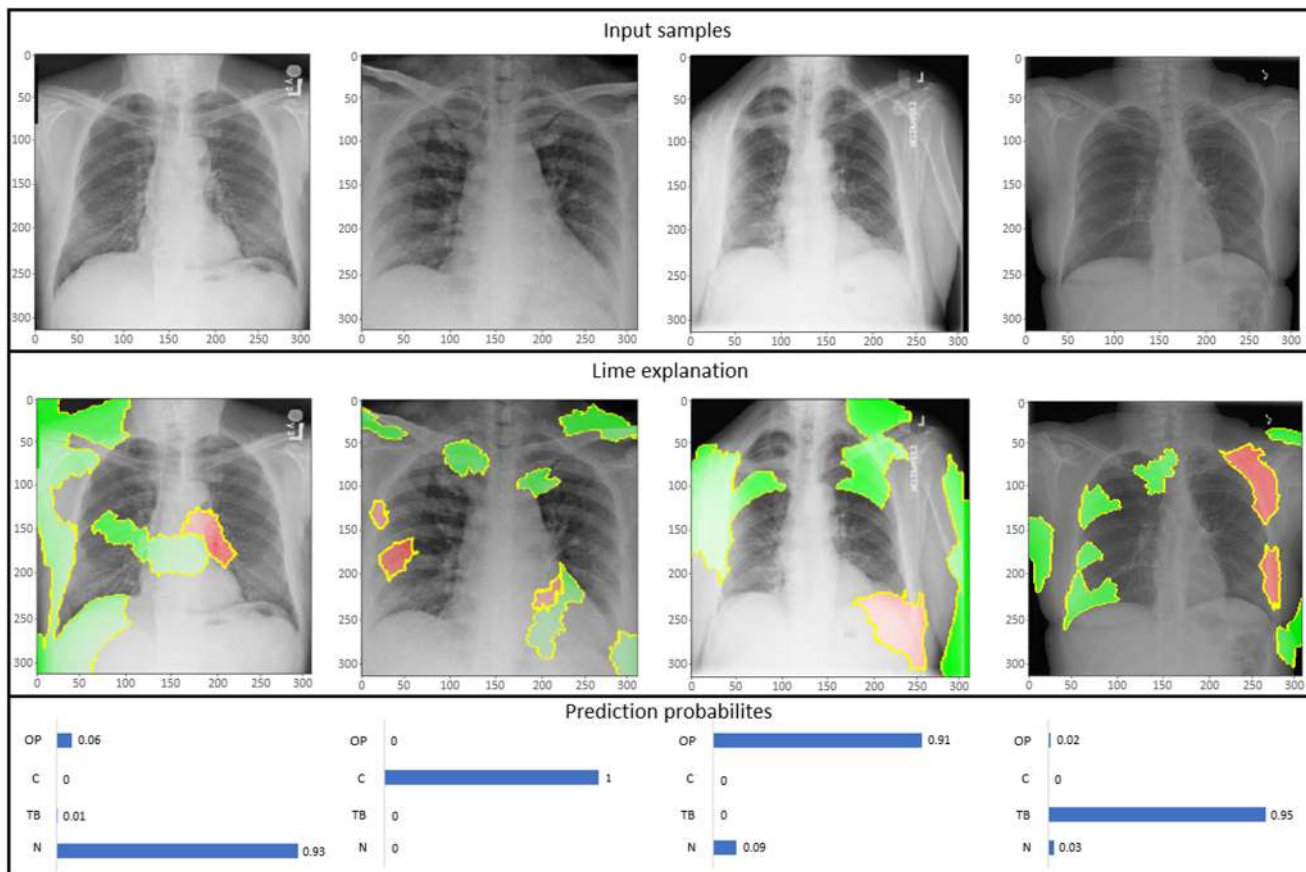


Fig. 7 LIME explanation of three distinct class samples using NASNetLarge along with the prediction probabilities of sample being normal (N), COVID-19 (C), and other pneumonia (OP)

help of VGG19 model [49] but with approximately double trainable parameters (144M) as compared to the proposed approach (83M).

8 Visualization

Training the model and getting the results is not sufficient unless it is understood that what is triggering the concerned output. To handle this blackbox, visualization techniques assist in illustrating the basis of prediction of the model.

There are many visualization techniques for example class activation maps (CAM) [45], saliency maps (SM) [48], local interpretable model-agnostic explanations (LIME) [41], and a lot more. In this article, activation maps and LIME techniques are utilized to present the model perception of identifying and classifying the COVID-19 samples from CXR images. CAM aims at understanding the feature space of an input image that influences the prediction, whereas LIME is an innovative explanation technique to represent the model prediction with local fidelity, interpretability and model agnostic. For instance, fine-tuned NASNetLarge

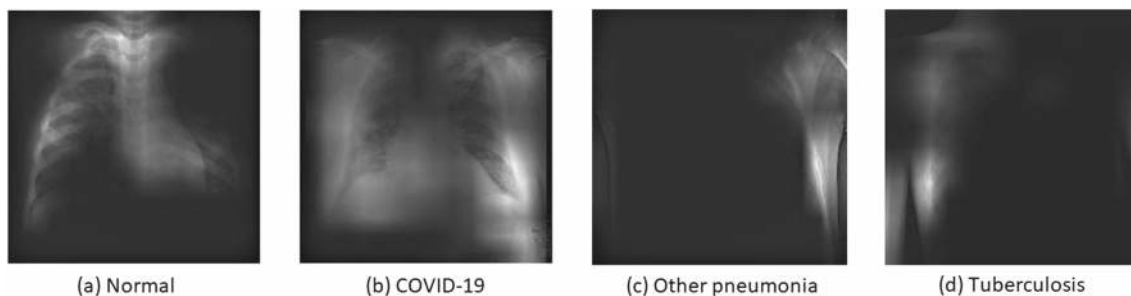


Fig. 8 CAM of NASNetLarge model for a Normal, b COVID-19, and c Other pneumonia

architecture is considered to generate LIME explanations for some samples taken from the test set (see Fig. 7), whereas class activation maps tend to present the patterns learned by the model for classification of samples as shown in Fig. 8. Figure 7 presents the LIME technique applied to four samples belonging to four distinct classes as COVID-19, other types of pneumonia, tuberculosis and normal cases. The red and green areas in the LIME generated explanation correspond to the regions that contributed against the predicted class and towards the predicted class respectively.

9 Conclusion

This article proposes to leverage the state-of-the-art deep learning models to aid in early identification and diagnosis of COVID-19 virus by using the limited posteroanterior chest X-ray images. Each trained model was evaluated using benchmark performance metrics e.g. accuracy, precision, recall, area under curve, specificity, and F1 score under four different scenarios concerned with imbalanced learning and classification strategy. With extensive trials, it was observed that models achieve different scores in different scenarios, among which NASNetLarge displayed better performance specially in binary classification of COVID-19 samples. The visual representation based on local interpretable model agnostic explanations is utilized to understand the basis of prediction of the model. As an extension to this work more deep learning models and preprocessing techniques can be explored to achieve better results.

Acknowledgements We are very much obliged to our institute, Indian Institute of Information Technology Allahabad (IIITA), India and Big Data Analytics (BDA) lab for ensuring the needed resources and support. We also would like to extend our thanks to the colleagues for their valuable guidance and suggestions. The manuscript is also available in the pre-print arXiv with the eprint-2004.11676 under the license Attribution 4.0 International (CC BY 4.0).

Compliance with Ethical Standards

Conflict of interest The authors have no conflict of interest to declare.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A (2020) Covid-caps: a capsule network-based framework for identification of covid-19 cases from x-ray images. arXiv:2004.02696
2. Allen JN, Davis WB (1994) Eosinophilic lung diseases. *Am J Respir Crit Care Med* 150(5):1423–1438
3. Apostolopoulos ID, Mpesian TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 1:635–640. <https://doi.org/10.1007/s13246-020-00865-4>
4. Bukhari SUK, Bukhari SSK, Syed A, SHAH SSH (2020) The diagnostic evaluation of convolutional neural network (cnn) for the assessment of chest x-ray of patients infected with covid-19. medRxiv
5. Chen Q, Montesinos P, Sun QS, Heng PA et al (2010) Adaptive total variation denoising based on difference curvature. *Image Vis Comput* 28(3):298–306
6. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, Seo JB (2019) Deep learning-based image conversion of ct reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology* 292(2):365–373
7. Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, de Albuquerque VHC (2020) A novel transfer learning based approach for pneumonia detection in chest x-ray images. *Appl Sci* 10(2):559
8. Cohen JP, Morrison P, Dao L (2020) Covid-19 image data collection. arXiv:2003.11597
9. Deng J, Dong W, Socher R, Li L, Kai L, Li F-F (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp 248–255
10. Farooq M, Hafeez A (2020) Covid-resnet: a deep learning framework for screening of covid19 from radiographs. arXiv:2003.14395
11. Fourcade A, Khonsari R (2019) Deep learning in medical image analysis: a third eye for doctors. *J Stomatol Oral Maxillofac Surg* 120(4):279–288
12. Gómez P, Semmler M, Schützenberger A, Bohr C, Döllinger M (2019) Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Med Biol Eng Comput* 57(7):1451–1463
13. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, Tan KS, Wang DY, Yan Y (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak—an update on the status. *Mil Med Res* 7(1):1–10
14. Hall LO, Paul R, Goldgof DB, Goldgof GM (2020) Finding covid-19 from chest x-rays using deep learning on a small dataset. arXiv:2004.02060
15. Hasinoff SW (2014) Photon, poisson noise. *Computer Vision: A Reference Guide*, Boston, pp 608–610
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
18. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghighi B, Ball R, Shpanskaya K et al (2019) Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 590–597
19. Jaeger S, Candemir S, Antani S, Wang YXJ, Lu PX, Thoma G (2014) Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg* 4(6):475
20. Joaquin AS (2020) Using deep learning to detect pneumonia caused by ncov-19 from x-ray images. <https://towardsdatascience.com/using-deep-learning-to-detect-ncov-19-from-x-ray-images-1a89701d1acd> (Online; Accessed 12 April 2020)
21. Johnson JM, Khoshgoftaar TM (2019) Survey on deep learning with class imbalance. *J Big Data* 6(1):27
22. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F et al (2018) Identifying

- medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172(5):1122–1131
23. Khalifa NEM, Taha MHN, Hassanien AE, Elghamrawy S (2020) Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset. arXiv:2004.01184
 24. Koidl K (2013) Loss functions in classification tasks. School of Computer Science and Statistic Trinity College, Dublin
 25. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
 26. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284(2):574–582
 27. Liang G, Zheng L (2019) A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer Methods and Programs in Biomedicine*, pp 104964
 28. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. *Neurocomputing* 234:11–26
 29. Luo P, Ren J, Peng Z, Zhang R, Li J (2018) Differentiable learning-to-normalize via switchable normalization. arXiv:1806.10779
 30. Makridakis S (2017) The forthcoming artificial intelligence (ai) revolution: its impact on society and firms. *Futures* 90:46–60
 31. Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv:2003.10849
 32. Oakden-Rayner L (2020) Chexnet: an in-depth review. <https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/> (Online; Accessed 12 April 2020)
 33. Oh Y, Park S, Ye JC (2020) Deep learning covid-19 features on cxr using limited training data sets. *IEEE Trans Med Imaging* 39:2688–2700. <https://doi.org/10.1109/TMI.2020.2993291>
 34. OpenCV (2020) Image thresholding. https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html (Online; Accessed 12 April 2020)
 35. Pang G, Shen C, van den Hengel A (2019) Deep anomaly detection with deviation networks. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp 353–362
 36. Pun N (2020) Covid-19 posteroanterior chest x-ray fused (cpcxr) dataset. <https://github.com/nspunn1993/COVID-19-PA-CXR-fused-dataset> (Online; Accessed 21 Jul 2020)
 37. Pun NS, Agarwal S (2020) Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. *ACM Trans Multimed Comput Commun Appl (TOMM)* 16(1):1–15
 38. Pun NS, Sonbhadra SK, Agarwal S (2020) Covid-19 epidemic analysis using machine learning and deep learning algorithms medRxiv
 39. Rajinikanth V, Dey N, Raj ANJ, Hassanien AE, Santosh K, Raja N (2020) Harmony-search and otsu based system for coronavirus disease (covid-19) detection using lung ct scan images. arXiv:2004.03431
 40. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz CP et al (2018) Deep learning for chest radiograph diagnosis: a retrospective comparison of the chexnext algorithm to practicing radiologists. *PLoS Med* 15(11): e1002686
 41. Ribeiro MT, Singh S, Guestrin C (2016) “Why should I trust you?” Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp 1135–1144
 42. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp 234–241
 43. Ruder S (2016) An overview of gradient descent optimization algorithms. arXiv:1609.04747
 44. Saraiva A, Ferreira N, Sousa L, Carvalho da Costa N, Sousa J, Santos D, Soares S (2019) Classification of images of childhood pneumonia using convolutional neural networks. In: *6th International conference on bioimaging*, pp 112–119
 45. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*, pp 618–626
 46. Sethy PK, Behera SK (2020) Detection of coronavirus disease (covid-19) based on deep features. Preprints, 2020030300, 2020
 47. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. *Ann Rev Biomed Eng* 19:221–248
 48. Simonyan K, Vedaldi A, Zisserman A (2013) Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv:1312.6034
 49. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
 50. Stein A (2020) Pneumonia dataset annotation methods. rsn pneumonia detection challenge discussion, 2018. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/discussion/> (Online; Accessed 12 April 2020)
 51. Stephen O, Sain M, Maduh UJ, Jeong DU (2019) An efficient deep learning approach to pneumonia classification in healthcare. *J Healthcare Eng* 2019. <https://doi.org/10.1155/2019/4180949>
 52. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI conference on artificial intelligence*
 53. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1–9
 54. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2818–2826
 55. Wang L, Wong A (2020) Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. arXiv:2003.09871
 56. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X et al (2020) A deep learning algorithm using ct images to screen for corona virus disease (covid-19). medRxiv
 57. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers R (2017) Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *IEEE CVPR*
 58. WHO (2020) Coronavirus disease 2019 (covid-19) situation report–81. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/>. Online; Accessed 12 April 2020
 59. WHO (2020) Who timeline—covid-19. <https://www.who.int/news-room/detail/08-04-2020-who-timeline---covid-19> (Online; Accessed 12 April 2020)
 60. Wu H, Xie P, Zhang H, Li D, Cheng M (2020) Predict pneumonia with chest x-ray images based on convolutional deep neural learning networks. *J Intell Fuzzy Syst (Preprint)* 1–15
 61. Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah NN (2014) An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In:

- Proceedings of the first international conference on advanced data and information engineering (DaEng-2013). Springer, pp 13–22
62. Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv:2003.12338
 63. Zoph B, Vasudevan V, Shlens J, Le QV (2018) Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8697–8710

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Narinder Singh Punn is presently pursuing his PhD from Indian Institute of Information Technology Allahabad, Prayagraj, U. P., India 211015. He is working as a senior research fellow in the “Big Data Analytics” laboratory of IITA. His research areas are: Machine Learning, Deep Learning and Biomedical Image Analysis.



Dr. Sonali Agarwal is presently working as Associate Professor in department of information technology at IIT Allahabad, Prayagraj, U. P., India 211015. She is having hands on experience on stream computing and complex processing platforms such as Apache Spark, Apache Flink and ESPER. Her research interests are in the areas of Stream Analytics, Big Data, Stream Data Mining, Complex Event Processing System, Support

Vector Machines and Software Engineering.