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

Narinder Singh Punn · Sonali Agarwal

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 multiclass 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, NASNet-Large 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.

Narinder Singh Punn
IIIT Allahabad, Prayagraj, India, 211015
Tel.: +91-7018466740
E-mail: pse2017002@iiita.ac.in

Sonali Agarwal
IIIT Allahabad, Prayagraj, India, 211015
Tel.: +91-9415647042
E-mail: sonali@iiita.ac.in

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 syn-
drome (SARS) occurred in 2003, whereas the second outbreak began in 2012 in
Saudi Arabia with the middle east respiratory syndrome (MERS). The novel coro-
navirus disease 2019 (COVID-19), began as an outbreak from epicentre Wuhan,
Peoples Republic of China in late December 2019, and till April 15, 2020, it caused
1,996,681 infections and 127,590 deaths worldwide [52]. The coronavirus (COVID-
19) outbreak was declared a public health emergency of international concern by
WHO on January 30, 2020 [53]. 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 [33]. 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 iden-
tified infected person 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 test-
ing 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 [12]. 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 diag-
nose pneumonia, lung inflammation, abscesses, and/or enlarged lymph nodes [2].
Due to its easy transmission, developing techniques to accurately and easily iden-
tify 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 [10]. 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 [34]. 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 popu-

lar choice among researchers [22]. CNNs have been successfully deployed in the analysis of video endoscopy [11] and CT images, and also used for the diagnosis of pediatric pneumonia via chest Xray images [6, 19]. Chouhan et al. [7] proposed a transfer learning based deep network approach pre-trained on ImageNet for pneumonia detection. Wang et al. [51] proposed a customized VGG16 model for lung regions identification to classify different types of pneumonia. Later, Ronneburger et al. [37] 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. [35] proposed a dense CNN with 121layers to detect several pathologies including pneumonia using chest Xrays. Lakhani et al. [23] obtained an area under the curve (AUC) of 0.95 in pneumonia detection using AlexNet and GoogLeNet along with image augmentation. Recently, several deeplearning based COVID19 detection techniques have been proposed [17, 49, 50]. Linda et al. [49] introduced a deep CNN, named COVIDNet for the detection of COVID19 cases from the chest Xray images. Shuai et al. [50] achieved accuracy, specificity and sensitivity of 89.5%, 88% and 87% respectively for COVID19 identification using CT images.

There are many datasets available about chest X-rays for the detection of pneumonia [8, 16, 28, 44]; but in present research work, COVID-19 X-ray chest images [8] and Radiological Society of North America (RSNA) images [44] 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 paper 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 II, and section III briefs the dataset. Section IV contains the proposed methodology followed by the evaluation metrics in section V. Results are discussed in section VI 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, 16, 28, 44]. 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. [19] 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. [45] 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. [38] 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 [24] 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. [54] 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. [27] 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 paper 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. [20] 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, Squeeznet 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. [40] proposed a deep learning based model to identify coronavirus infections using CXR images. Deep features from CXR images have been extracted and support vect or 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 [9], where conventional ResNet-50 model
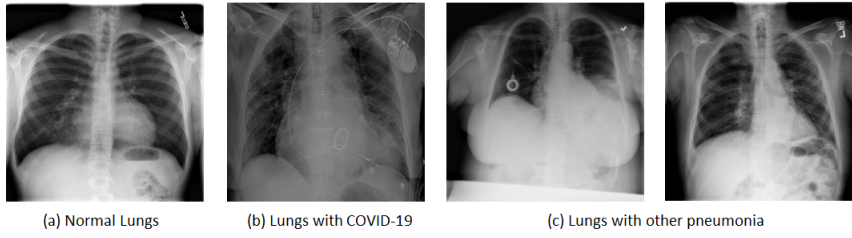
(a) Normal Lungs   (b) Lungs with COVID-19   (c) Lungs with other pneumonia

Fig. 1 Sample chest radiographs.

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 128x128x3, 224x224x3 and 229x229x3 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. [56] 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 [31].

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 two datasets are utilized for experiments: COVID-19 image [8] and Radiological Society of North America (RSNA) [44]. 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 patients confidentiality (Fig.1(b)). 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 SARSr-CoV-2 or COVID-19, SARSr-CoV-1 or SARS, Streptococcus spp., 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 [44]. Out of 30,000 selected images, 15,000 examinations had positive cases of pneumonia and from the remaining 15000 cases, 7500 cases had no findings and other 7500 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. 1(c). This dataset has been published in two stages. In Stage one, 25,684 training
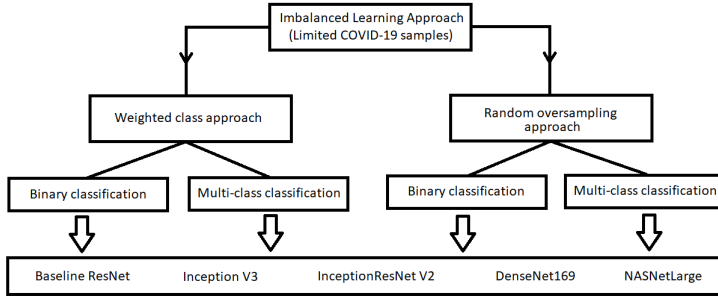
Fig. 2 Schematic representation of the proposed components for COVID-19 identification.

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.

4 Proposed Contribution

The era of artificial intelligence has brought significant improvements in the living society [26]. 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 [25]. 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, [41] and fused modalities [32] 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 and RSNA 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 vs other types of pneumonia vs 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 NASNet-Large [15, 46, 48, 57]. 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 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 [18].

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.

$$w(c) = C_c . \frac{\sum\limits_{c=0}^{N} n_c}{N.n_c} \tag{1}$$

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. chlamydophila, SARS, streptococcus, legionella, 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 normal case findings. The three classes are provided with labels as "0" being a normal case, "1" being a COVID-19 case, and "2" being other pneumonia cases.

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 3 [21]:

$$f(s)_i = \frac{e^{s_i}}{\sum\limits_{C}^{j} e^{s_i}} \tag{2}$$

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

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 models 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 another dataset, RSNA [44], which consists of posteroanterior CXR images covering sample cases labelled as pneumonia and normal. Table 1 describes the distribution of training, testing, and validation sets using the above discussed datasets for different class imbalance strategies i.e. class weighted loss function that penalizes the model for any false negative prediction and random oversampling [55] of minority classes which in this case is COVID-19.

Table 1 Posteroanterior CXR images distribution into training, validation, and test sets with datasets fusion for different problem definitions.

Dataset	Finding	Total PCXR Images	Class imbalance learning strategy											
			class weighted loss function						Random oversampling					
			Binary class classification			Multi-class classification			Binary class classification			Multi-class classification		
			Tr	Val	Tst	Tr	Val	Tst	Tr	Val	Tst	Tr	Val	Tst
COVID-19 Image Dataset	COVID-19	108	88	9	11	88	9	11	793	9	11	422	9	11
	Other pneumonia	45				422	42	51				422	42	51
RSNA	Other pneumonia	470	793	78	97				793	78	97			
	Normal	453				371	36	46				422	36	46

The CXR images in the aggregated dataset also consist 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
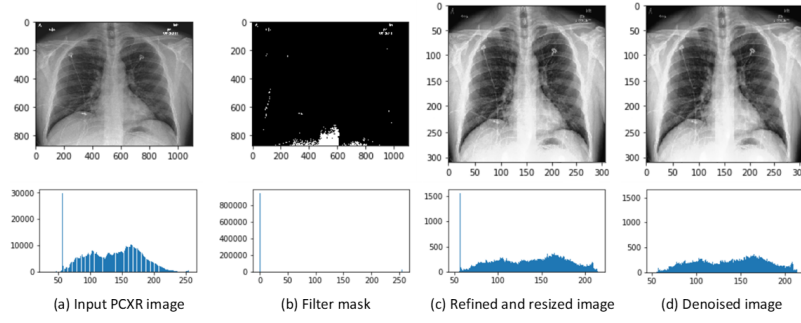
(a) Input PCXR image　　(b) Filter mask　　(c) Refined and resized image　　(d) Denoised image

Fig. 3 Data preprocessing stages of raw posteroanterior CXR image.

binary thresholding [30] as given by equation 4, followed by resizing the images to a fixed size resolution of 331x331x3.

$$M(x, y) = \begin{cases} max\_th, & i(x, y) \geq min\_th. \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

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 [13]. The denoising or removal of such uncertainty is carried through the adaptive total variation method [47] 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 $x = (x_1, x_2) \; \epsilon \; \Omega$ - pixels, given as

$$u = \underset{u}{\arg\min} \left(\int_{\Omega} (u - f.\ln u) dx + \int_{\Omega} (\omega(x) | \bigtriangledown u | dx) \right) \tag{5}$$

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

Fig. 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 2 along with their respective contribution, parameters, and performance on the standard benchmark datasets. The inception deep convolutional architectures proposed by GoogLeNet are considered as the
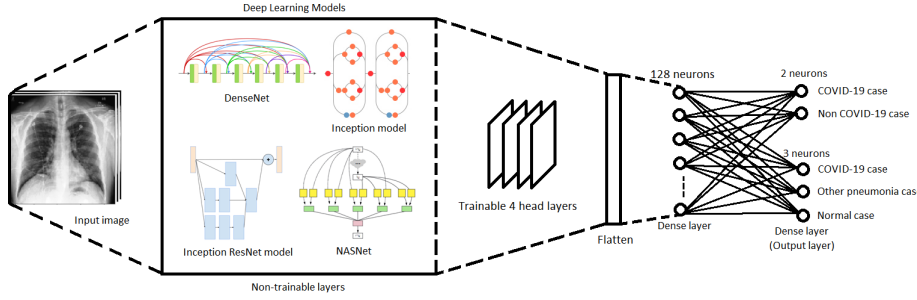
Fig. 4 Schematic representation of training framework for deep learning architectures via transfer learning.

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 [48]. 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. [46]. 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. [15]. 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. Google Brain research team proposed NAS-Net model based on reinforcement learning search methods [57]. 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. 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 [14], 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 2 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, accompanied with the softmax activation function at the output or decision layer.

Table 2 Recent deep learning architectures that are reported with top 5 error rate per year.

Architecture	Year	Contribution	Param.	Dataset	E. rate
Inception-v3 [48]	2015	Avoided the bottleneck representations, Dimension reduction promotes faster learning.	23.6M	ILSVRC	5.6
Inception ResNet-v2 [46]	2016	Focused on residual connection rather than filter connection via split and merge approach.	56M	ILSVRC	4.9
DenseNet169 [15]	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 [57]	2018	Search and transfer the architecture block from small dataset to large dataset, Scheduled drop path regularization.	22.6M	CIFAR-10	2.4

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 Results and discussion

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. Fig. 6 shows the performance curves of the NASNetLarge model evaluated and monitored on the test set for each iteration during the training phase, generated using the tensorboard. The DenseNet169 displayed similar performance to NASNetLarge in contrast to other networks. The best scores of the models corresponding to each scenario are highlighted in Table 3 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, CM, RB, and RM where C and R indicates class
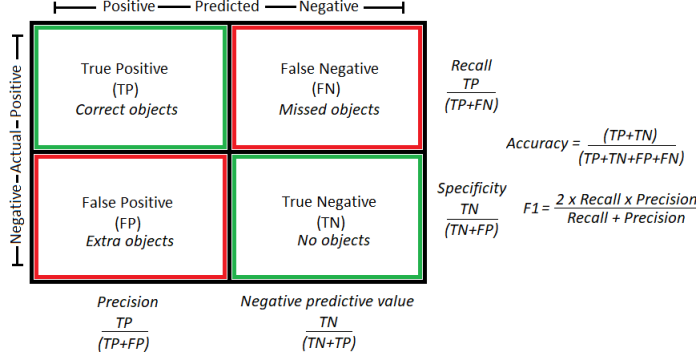
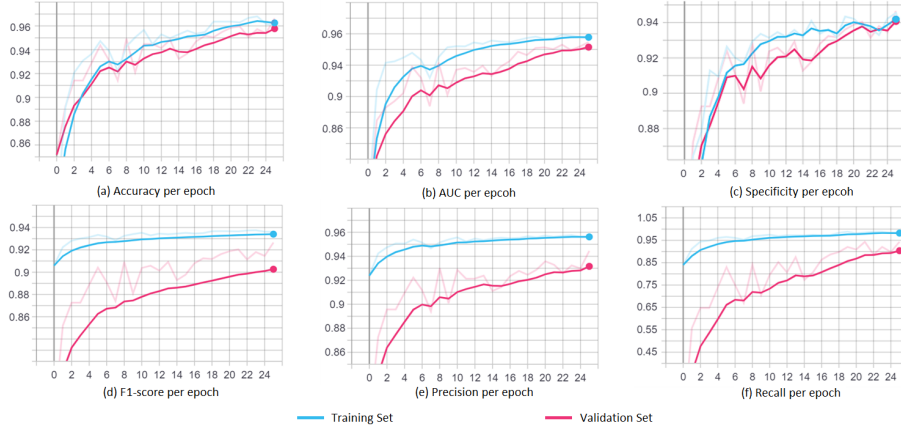**Fig. 5** Confusion matrix and performance evaluation metrics.

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

imbalanced learning approaches, namely weighted class loss functions and random oversampling of the minority class, whereas B and M indicates the binary and multi-class classification schemes, respectively. Furthermore, the results obtained are compared with the other recently proposed approaches to detect COVID-19 positive symptoms via X-ray images as shown in Table 4. It is observed that the proposed approach outperforms other approaches, whereas Apostolopoulos et al. [3] achieved similar results with the help of VGG19 model [43] but with approximately double trainable parameters (144M) as compared to the proposed approach (83M).

7 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

Table 3 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	0.76	0.62	0.89	0.90	0.77	0.73
	RB	0.89	0.56	0.56	0.89	0.60	0.56
	RM	0.76	0.65	0.43	0.83	0.77	0.52
Inception v3	CB	0.89	0.77	0.80	0.93	0.89	0.78
	CM	0.87	0.74	0.88	0.90	0.88	0.80
	RB	0.88	0.78	0.70	0.90	0.91	0.74
	RM	0.87	0.90	0.82	0.91	0.90	0.86
Inception ResNet v2	CB	0.95	**0.86**	0.67	0.98	0.87	0.75
	CM	0.93	0.86	**0.92**	**0.99**	**0.89**	**0.89**
	RB	0.90	0.85	0.75	0.94	0.90	0.80
	RM	0.92	**0.97**	**0.97**	**0.99**	0.94	**0.97**
DenseNet 169	CB	0.90	0.81	0.77	0.92	0.91	0.79
	CM	0.88	0.87	0.91	0.96	**0.89**	**0.89**
	RB	0.95	**0.94**	**0.96**	0.97	**0.96**	**0.95**
	RM	0.92	0.93	0.95	0.94	**0.95**	0.94
NASNet Large	CB	**0.97**	0.83	**0.91**	**0.99**	**0.98**	**0.87**
	CM	**0.94**	**0.88**	0.91	0.95	**0.89**	**0.89**
	RB	**0.98**	0.88	0.90	**0.99**	0.95	0.89
	RM	**0.96**	0.93	0.90	0.95	0.94	0.91

Table 4 Comparative analysis of the proposed approach with recently proposed strategies.

Authors	Technique	Network	Classes	Accuracy	Precision	Recall	AUC	Specificity
Yujin et al. [29]	Fine tuning	Res-Net18	4	0.88	0.83	0.86	-	0.96
Afshar et al. [1]	Fine tuning	Capsule	4	0.95	-	0.9		0.95
Wang et al. [49]	Full training	COVID-Net	3	0.92	0.91	0.88	-	-
Apostolopoulos et al. [3]	Fine tuning	VGG19	3	0.87	-	0.92	-	0.98
			2	0.98	-		-	
Proposed	**Fine tuning**	**NASNet Large**	**3**	**0.96**	**0.93**	**0.91**	**0.95**	**0.94**
			2	**0.98**	**0.88**	**0.91**	**0.99**	**0.98**

are many visualization techniques for example class activation maps (CAM) [39], saliency maps (SM) [42], local interpretable model-agnostic explanations (LIME) [36], 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 architecture is considered to generate LIME explanations for some samples taken from the test set, whereas class activation maps tend to present the patterns learned by the model for classification of samples as shown in Fig. 7. Fig. 8 presents the LIME technique applied to three samples belonging to three distinct classes as COVID-19, other types of pneumonia, and normal cases. The red and green areas in the LIME generated explanation (Fig. 8) correspond to the regions that contributed against the predicted class and towards the predicted class respectively.
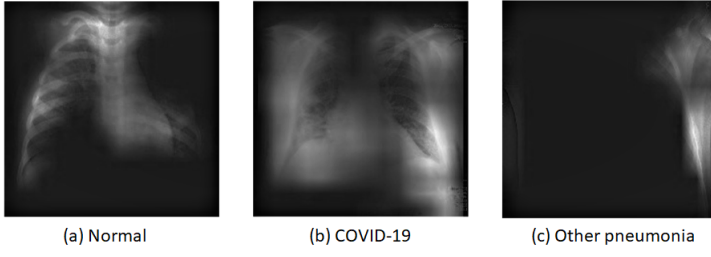
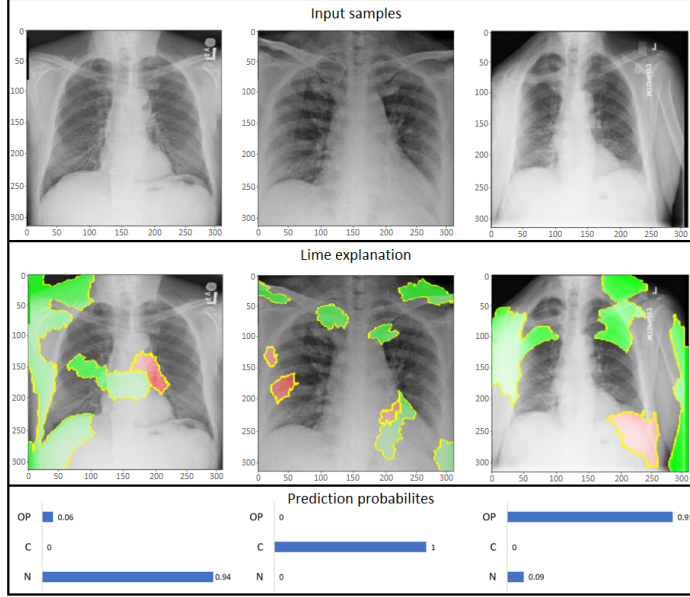**Fig. 7** CAM of NASNetLarge model for (a) Normal, (b) COVID-19, and (c) Other pneumonia.

Fig. 8 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).

8 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. 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.

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, K.N., Mohammadi, A.: Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. arXiv preprint arXiv:2004.02696 (2020)
2. Allen, J.N., Davis, W.B.: Eosinophilic lung diseases. American journal of respiratory and critical care medicine **150**(5), 1423–1438 (1994)
3. Apostolopoulos, I.D., Mpesiana, T.A.: Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine p. 1 (2020)
4. Bukhari, S.U.K., Bukhari, S.S.K., Syed, A., SHAH, S.S.H.: The diagnostic evaluation of convolutional neural network (cnn) for the assessment of chest x-ray of patients infected with covid-19. medRxiv (2020)
5. Chen, Q., Montesinos, P., Sun, Q.S., Heng, P.A., et al.: Adaptive total variation denoising based on difference curvature. Image and vision computing **28**(3), 298–306 (2010)
6. Choe, J., Lee, S.M., Do, K.H., Lee, G., Lee, J.G., Lee, S.M., Seo, J.B.: Deep learning–based image conversion of ct reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. Radiology **292**(2), 365–373 (2019)
7. Chouhan, V., Singh, S.K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., de Albuquerque, V.H.C.: A novel transfer learning based approach for pneumonia detection in chest x-ray images. Applied Sciences **10**(2), 559 (2020)
8. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv preprint arXiv:2003.11597 (2020)
9. Farooq, M., Hafeez, A.: Covid-resnet: A deep learning framework for screening of covid19 from radiographs. arXiv preprint arXiv:2003.14395 (2020)
10. Fourcade, A., Khonsari, R.: Deep learning in medical image analysis: A third eye for doctors. Journal of stomatology, oral and maxillofacial surgery **120**(4), 279–288 (2019)
11. Gómez, P., Semmler, M., Schützenberger, A., Bohr, C., Döllinger, M.: Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. Medical & biological engineering & computing **57**(7), 1451–1463 (2019)
12. Guo, Y.R., Cao, Q.D., Hong, Z.S., Tan, Y.Y., Chen, S.D., Jin, H.J., Tan, K.S., Wang, D.Y., Yan, Y.: The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak–an update on the status. Military Medical Research **7**(1), 1–10 (2020)
13. Hasinoff, S.W.: Photon, poisson noise. Computer Vision: A Reference Guide,Boston, MA pp. 608–610 (2014)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)

15. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708 (2017)
16. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., 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, pp. 590–597 (2019)
17. Joaquin, A.S.: 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 (2020). [Online; accessed April 12, 2020]
18. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. Journal of Big Data 6(1), 27 (2019)
19. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5), 1122–1131 (2018)
20. Khalifa, N.E.M., Taha, M.H.N., Hassanien, A.E., Elghamrawy, S.: 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 preprint arXiv:2004.01184 (2020)
21. Koidl, K.: Loss functions in classification tasks. School of Computer Science and Statistic Trinity College, Dublin (2013)
22. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
23. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284(2), 574–582 (2017)
24. Liang, G., Zheng, L.: A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Computer methods and programs in biomedicine p. 104964 (2019)
25. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26 (2017)
26. Makridakis, S.: The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms. Futures 90, 46–60 (2017)
27. 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)
28. Oakden-Rayner, L.: Chexnet: an in-depth review. https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/ (2020). [Online; accessed April 12, 2020]
29. Oh, Y., Park, S., Ye, J.C.: Deep learning covid-19 features on cxr using limited training data sets. IEEE Transactions on Medical Imaging (2020)
30. OpenCV: Image thresholding. https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html. (2020). [Online; accessed April 12, 2020]
31. Pang, G., Shen, C., van den Hengel, A.: Deep anomaly detection with deviation networks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 353–362 (2019)
32. Punn, N.S., Agarwal, S.: Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16(1), 1–15 (2020)
33. Punn, N.S., Sonbhadra, S.K., Agarwal, S.: Covid-19 epidemic analysis using machine learning and deep learning algorithms. medRxiv (2020)
34. Rajinikanth, V., Dey, N., Raj, A.N.J., Hassanien, A.E., Santosh, K., Raja, N.: Harmony-search and otsu based system for coronavirus disease (covid-19) detection using lung ct scan images. arXiv preprint arXiv:2004.03431 (2020)
35. Rajpurkar, P., Irvin, J., Ball, R.L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C.P., et al.: Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. PLoS medicine 15(11), e1002686 (2018)
36. Ribeiro, M.T., Singh, S., Guestrin, C.: " 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 (2016)

37. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer (2015)

38. Saraiva, A., Ferreira, N., Sousa, L., Carvalho da Costa, N., Sousa, J., Santos, D., Soares, S.: Classification of images of childhood pneumonia using convolutional neural networks. In: 6th International Conference on Bioimaging, pp. 112–119 (2019)

39. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp. 618–626 (2017)

40. Sethy, P.K., Behera, S.K.: Detection of coronavirus disease (covid-19) based on deep features. Preprints **2020030300**, 2020 (2020)

41. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual review of biomedical engineering **19**, 221–248 (2017)

42. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013)

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

44. Stein, A.: Pneumonia dataset annotation methods. rsna pneumonia detection challenge discussion, 2018. `https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/discussion/` (2020). [Online; accessed April 12, 2020]

45. Stephen, O., Sain, M., Maduh, U.J., Jeong, D.U.: An efficient deep learning approach to pneumonia classification in healthcare. Journal of healthcare engineering **2019** (2019)

46. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence (2017)

47. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9 (2015)

48. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826 (2016)

49. Wang, L., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. arXiv preprint arXiv:2003.09871 (2020)

50. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., et al.: A deep learning algorithm using ct images to screen for corona virus disease (covid-19). MedRxiv (2020)

51. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: IEEE CVPR (2017)

52. WHO: Coronavirus disease 2019 (covid-19) situation report 81. `https://www.who.int/docs/default-source/coronaviruse/situation-reports/` (2020). [Online; accessed April 12, 2020]

53. WHO: Who timeline - covid-19. `https://www.who.int/news-room/detail/08-04-2020-who-timeline---covid-19` (2020). [Online; accessed April 12, 2020]

54. Wu, H., Xie, P., Zhang, H., Li, D., Cheng, M.: Predict pneumonia with chest x-ray images based on convolutional deep neural learning networks. Journal of Intelligent & Fuzzy Systems (Preprint), 1–15 (2020)

55. Yap, B.W., Rani, K.A., Rahman, H.A.A., Fong, S., Khairudin, Z., Abdullah, N.N.: 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), pp. 13–22. Springer (2014)

56. Zhang, J., Xie, Y., Li, Y., Shen, C., Xia, Y.: Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12338 (2020)

57. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697–8710 (2018)