



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



SVD-CLAHE boosting and balanced loss function for Covid-19 detection from an imbalanced Chest X-Ray dataset

Santanu Roy^{a,*}, Mrinal Tyagi^b, Vibhuti Bansal^b, Vikas Jain^c

^a School of Engineering and Technology, Christ (Deemed to be University), Bangalore 560074, India

^b Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India

^c School of Computer Science Engineering and Technology, Bennett University, Greater Noida, UP 201310, India

ARTICLE INFO

Keywords:

Class imbalance problem
Covid-19 detection
Chest X-Ray (CXR) images
Data augmentation
Categorical Cross Entropy (CCE)
Contrast Limited Adaptive Histogram
Equalization (CLAHE)
Singular Value Decomposition (SVD)

ABSTRACT

Covid-19 disease has had a disastrous effect on the health of the global population, for the last two years. Automatic early detection of Covid-19 disease from Chest X-Ray (CXR) images is a very crucial step for human survival against Covid-19. In this paper, we propose a novel data-augmentation technique, called SVD-CLAHE Boosting and a novel loss function Balanced Weighted Categorical Cross Entropy (BWCCE), in order to detect Covid 19 disease efficiently from a highly class-imbalanced Chest X-Ray image dataset. Our proposed SVD-CLAHE Boosting method is comprised of both oversampling and under-sampling methods. First, a novel Singular Value Decomposition (SVD) based contrast enhancement and Contrast Limited Adaptive Histogram Equalization (CLAHE) methods are employed for oversampling the data in minor classes. Simultaneously, a Random Under Sampling (RUS) method is incorporated in major classes, so that the number of images per class will be more balanced. Thereafter, Balanced Weighted Categorical Cross Entropy (BWCCE) loss function is proposed in order to further reduce small class imbalance after SVD-CLAHE Boosting. Experimental results reveal that ResNet-50 model on the augmented dataset (by SVD-CLAHE Boosting), along with BWCCE loss function, achieved 95% F1 score, 94% accuracy, 95% recall, 96% precision and 96% AUC, which is far better than the results by other conventional Convolutional Neural Network (CNN) models like InceptionV3, DenseNet-121, Xception etc. as well as other existing models like Covid-Lite and Covid-Net. Hence, our proposed framework outperforms other existing methods for Covid-19 detection. Furthermore, the same experiment is conducted on VGG-19 model in order to check the validity of our proposed framework. Both ResNet-50 and VGG-19 model are pre-trained on the ImageNet dataset. We publicly shared our proposed augmented dataset on Kaggle website (<https://www.kaggle.com/tr1gg3rtrash/balanced-augmented-covid-cxr-dataset>), so that any research community can widely utilize this dataset. Our code is available on GitHub website online (<https://github.com/MrinalTyagi/SVD-CLAHE-and-BWCCE>).

1. Introduction

Severe Acute Respiratory Syndrome Corona Virus2 (SARS-CoV2) [1] was declared as 'pandemic' by World Health Organization (WHO) in March 2020. Still now, over the last two years, more than 230 million people have been affected by this novel Corona Virus and more than 4.8 million people were declared dead [2] due to this virus. One of the features of this SARS-CoV2 is that it directly attacks human respiratory system and causes different forms of lung opacity, pneumonia and chest infections in human body [3]. Moreover, it can penalize human life, by destroying their immune systems. Covid-19 virus has become dangerous because it is contagious, which means it can be transmitted from person to person through their physical contact

or breath contact [4]. One of the crucial steps into the battle against COVID-19 disease is to incorporate accurate, automated and easily available screening methods of infected patients, like radiology examination using chest radiography, Computed Tomography (CT) scanning and Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) [5] etc. Although RT-PCR is the most common golden standard screening method for Covid-19 detection, it is also a very tedious and manual process, and several researchers have already identified that RT-PCR can not detect Covid-19 with higher accuracy and sensitivity [6]. Thus, RT-PCR has not been considered in the field of machine learning or computer vision for automatic Covid-19 detection. Several researchers ([7],[8]) have proposed automatic Covid-19 detection model using

* Corresponding author.

E-mail addresses: santanuroy35@gmail.com (S. Roy), mrinaltyagi24@gmail.com (M. Tyagi), bansal.vibhuti25@gmail.com (V. Bansal), vikasjain.cse@gmail.com (V. Jain).

<https://doi.org/10.1016/j.complbiomed.2022.106092>

Received 6 March 2022; Received in revised form 15 August 2022; Accepted 3 September 2022

Available online 28 September 2022

0010-4825/© 2022 Elsevier Ltd. All rights reserved.

computer vision and deep learning model from CXR or CT scan images. In this paper, we have considered CXR images' dataset because this is the most easily available screening method and not very costly [9] like CT scanning. The main objective of this research is to classify the CXR images into four classes (I) Covid 19, (II) Lung Opacity, (III) Normal and (IV) Viral Pneumonia.

Class imbalance is a very common problem in medical image diagnosis, since in hospitals the number of patients for different diseases varies considerably. The problem of class imbalance appears when the number of images in one class outnumbered the other classes [10] and consequently it may affect the final results of the classification. That means images from minor classes may be misclassified into major class which is undesirable. One of the ways to tackle this problem is to deploy Under-Sampling, in order to balance the dataset. Random Under-Sampling (RUS) [11] is the most frequently employed under-sampling method by researchers and it is known for its simplicity. It simply randomly chooses some images from major class in order to exclude them. However, RUS does not work efficiently for a highly-imbalanced dataset, since large number of randomly excluded images may contain significant features for the classification task. Thus, many researchers preferred over-sampling over under-sampling for resolving this class imbalance problem. In the recent advent, the most frequently employed over-sampling method for CNN is Data Augmentation technique [12] by zooming, cropping, rotation and flipping. SMOTE [13], KNN [14] and GAN-Oversampling ([15],[16]) etc. have also been widely utilized by several researchers. However, there is no guarantee that these data-augmentation methods will be feature invariant for the classification task. Indeed, this is dependent on the statistics of the dataset. SMOTE [13] generates synthetic images, based on random interpolation between their neighbor samples in a minor class. According to Z. Chen et al. [17], SMOTE can generate a very different statistical distribution of dataset than original dataset which is not desirable. GAN-Oversampling [16] is a comparatively efficient method, where synthetic images are produced by Generative Adversarial Network (GAN). However, GAN-oversampling is computationally very costly than other data-augmentation techniques, moreover, larger number of images are required to perform GAN-Oversampling [16]. This is not feasible for small datasets. Another significant limitation of these data-augmentation methods is that they may produce very similar images (almost replicas), in large amounts in the augmented dataset. This does not improve the model performance, moreover, this may further induce overfitting in the model [18]. Hence, any data-augmentation method has a trade-off that, it should not produce exactly same images, as well as it should not produce very dis-similar images (otherwise that will be feature variant for final classification task). Another limitation of over-sampling is that it further increases the training times of CNN model. Several researchers ([11], [19]) have come up with a hybrid method of employing both oversampling and under-sampling in recent times, but it does not resolve the class imbalance problem in a generalized way.

Boosting is another technique which generally boosts the performance of a weak classifier [20]. Boosting can be any method like oversampling, under-sampling [11], ensemble method [17] etc. RUS Boosting [11], Ada Boosting [20], SMOTE Boosting [21] etc. have been widely utilized by numerous researchers. Ada-Boosting is an ensemble technique which uses variants of Nearest Neighbor Classifiers to enhance the performance of this weak classifier. However, this kind of ensemble method may increase computational complexity of the model and consequently, it may induce huge overfitting. RUS Boosting and SMOTE Boosting are further modifying Ada Boost by combining data sampling (oversampling and under-sampling) and ensemble method. J. Sun et al. [22] recently proposed an AdaBoost SVM Ensemble model along with SMOTE and time weighting in order to resolve class imbalance problem for financial distress prediction. All of these aforementioned Boosting methods are very time consuming and more feasible for time series data (or in NLP) than digital images. In

this paper, we propose a novel SVD-CLAHE Boosting which is based on only over-sampling and under-sampling methods. We call this method as Boosting, because we have observed there is a huge boosting performance of standard CNN models, after employing proposed augmented dataset. Further it is explained in depth in Section 2.

Another way to resolve this issue, is to incorporate a cost sensitive [23] Deep Learning model in which different kind of loss function is used in order to alleviate class imbalance problem. For example, many researchers employed Weighted Categorical Cross Entropy (WCCE) ([24], [25]) for CNN model. The idea is to give little bit more weightage to the minor class and a little bit less weightage to the major class. However, WCCE may not work if the dataset is highly imbalanced. This is further explained in depth in the Section 3.3. There are many different loss functions like focal loss [26], center loss [27], distribution balanced loss [28], Anchor loss [29] proposed by several scientists who have worked in the same direction of class imbalance problem. Y. Cui et al. [30] found that the problem of class imbalance can be resolved just by modifying weights in loss function for different classes, rather than employing entirely different loss function. They have come up with a novel methodology of assigning weights, which are inversely proportional to the number of images to the corresponding class. However, we found that their methodology of weights assignment is not dependent on the number of classes, thus, it is not adaptive with multi-class classification task. Focal loss [26], Tversky loss [31], Unified focal loss [32] etc. are used very frequently to resolve class imbalance problem for image segmentation. However, for classification task, we didn't find any specific loss function which efficiently works with any kind of class imbalance problem, in a generalized way.

The Covid 19 dataset is taken from publicly available Kaggle website ([33], [34]) This dataset consists of four different classes. (I) Covid, (II) Lung Opacity (LO), (III) Normal, (IV) Viral Pneumonia (VP). A team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh have created this database of CXR images. In the first update, they released 219 COVID-19, 1341 normal and 1345 viral pneumonia CXR images. In the 2nd update, they have increased the COVID-19 class to 1200 CXR images. In the 3rd update, they have modified the database into total 3616 COVID-19 positive cases, along with 10,192 Normal, 6012 Lung Opacity and 1345 Viral Pneumonia (VP) images.

There are some challenges in this dataset which are as follows.

- Clearly this can be observed that the number of images in each class, varies significantly in this dataset. Thus, conventional CNN model may not work on this dataset efficiently. This is a class imbalance problem.
- Intra-class variance in Covid class, is comparatively higher than that of other classes, which is further discussed in section-4.2. This kind of intra-class statistical variability in the dataset further makes the classification task a lot more complicated.
- Many of the images have poor contrast and poor background luminance, thus, it may lead to poor feature extraction in CNN model.

In order to resolve these aforementioned challenges, we have proposed a novel framework in this paper. The main contributions of this paper are explained as follows:

1. A novel data-augmentation technique, "SVD-CLAHE Boosting", is proposed for resolving class imbalance problem from a highly imbalanced Covid 19 Chest X-Ray (CXR) dataset.
2. A novel SVD-based contrast enhancement method, CLAHE 0.5 and CLAHE 1.0, are employed for over-sampling, whereas under-sampling is done by RUS in major classes.
3. A novel loss function, "Balanced Weighted Categorical Cross-Entropy (BWCCE)", is proposed for eliminating little class imbalance present after employing SVD-CLAHE Boosting.

4. For training the CNN models, a transfer learning approach is deployed in which pre-trained weights are taken from a large ImageNet dataset.
5. A unique framework (i.e. "SVD-CLAHE Boosting" data augmentation along with BWCCE loss function) is proposed for ResNet-50 model, which performs more efficiently than individual ResNet-50 model and other existing models.
6. For the validity of the aforementioned framework, the same experiment is also conducted on the VGG-19 model.
7. We have also shared the augmented dataset (of 30,033 images), generated by proposed SVD-CLAHE Boosting method, on the Kaggle site. To the best of our knowledge, this kind of augmented and balanced Covid CXR dataset was not available so far in public.

The rest of the paper is organized in the following way: Section 2 presents a brief explanation of existing methods for Covid-19 detection from CXR images. Section 3 describes the entire proposed methodology, that is SVD-CLAHE Boosting and BWCCE Loss function, performed on ResNet-50 CNN model. Moreover, in Section 4, quantitative and qualitative results of several CNN models are compared with the proposed framework. In Section 5, we present our concluding remarks.

2. Existing methods of Covid detection

M. Shiddhartha and Avik Santra [35] proposed a very light weighted CNN model, called Covid Lite model, which is based on Depth-wise Separable Convolutional Neural Network (DSCNN). The advantage of deploying DSCNN over DCNN is that, DSCNN reduces computational complexity of the model (during training) considerably, since sequential point wise convolution is incorporated. Moreover, White Balance and CLAHE image processing methods are utilized as pre-processing methods before feeding the images into DSCNN. The authors have employed a very small dataset of CXR images, having only 1823 images. Thus, their proposed Covid-Lite model is feasible for training this small dataset. L. Wang et al. [36] first time designed a novel CNN architecture, especially dedicated to Covid-19 detection from Chest X-Ray (CXR) images. The authors deployed a CNN architecture, called Projection Expansion Projection Extension (PEPX), based on human-machine collaborative design strategy. They incorporated a very light weighted CNN so that it can efficiently train the Covid dataset from the scratch. They called their architecture as Covid-Net. Moreover, they made a new CXR dataset Covid-x publicly available, which consists of 13,975 CXR images. Their dataset had three classes: Covid, Pneumonia and Normal. Y. Xu et al [37] recently employed a novel Mask Attention (MANet) based model for covid-19 detection from CXR images for five classes: Covid, Normal, Tuberculosis (TB), Bacterial Pneumonia (BP) and Viral Pneumonia (VP). Their model is comprised of two stage network. In the first stage, they segmented only the lung portions from the CXR images using a ResUnet (ResNet Backbone Unet) model. Thereafter, in the second stage, they employed several standard CNN models like ResNet-34, ResNet-50, VGG-16 and Inception-V3 etc., in order to accomplish the final classification task. They employed ResUnet model as an attention-based model [38], which gives more attention to the important lung portions of the image. They have shown experimentally that accuracy of those CNN models is improved 0.5-1%, after employing MANet.

M. Togacar et al. [39] proposed a novel framework in which they have created two different augmented datasets, by using a fuzzy color-based image processing method and original dataset stacking technique. The main purpose of utilizing fuzzy color based pre-processing technique is to convert original images into the images with lesser noise and clear the foreground information. Moreover, a dataset stacking technique is incorporated to fuse the original images onto these processed images, further providing better clarity and high contrast images. They have combined two datasets from two different sources,

in order to make a small dataset of 458 images. Moreover, they have employed two very lightweight CNNs, which are Mobile-Net V2 [40] and Squeeze-Net [41] for feature extraction from very small dataset. This is followed by an SVM classifier for the final classification task. Although, their method used data-augmentation, it did not work in the direction of reducing class imbalance from the dataset. Instead, they tried to improve the contrast of the original images, for better feature extraction. Recently, M. Mamelakis et al. [42] proposed a deep transfer learning pipeline DenResCov-19, which is based on ensemble of ResNet-50 [43] and DenseNet-121 models [44]. These models are pre-trained on ImageNet dataset. According to the author, combining both of these models further improved the model's overall performance significantly. D. Das et al. [45] proposed a truncated Inception-Net model in which they modified the architecture of the traditional Inception-V3 model [46] a little bit, to reduce its complexity. Moreover, they deployed both Max-pooling and Global Average Pooling operation in order to reduce the dimension of the image considerably. Although they reduced the complexity of Inception-V3 considerably, their preparation of datasets was entirely a manual process. Other than these works, a lot of comparative analysis have been done in the direction of deep learning-based CNN model for Covid 19 detection [47–54] from CXR images and CT images. S. Nayak et al. [7] presented a survey paper for Covid 19 detection from CXR images, in which they implemented many standard CNN models AlexNet [55], VGG-16 [56], Google Net [57], ResNet-34 [43], ResNet-50 [43], InceptionV3 [46] etc., for CXR Covid dataset.

All of the methods mentioned above did not work toward alleviating class imbalance problems from a particular dataset. Instead, we have found that most of the researchers [36,39,42,45,49] combined different Covid CXR datasets from different sources and manually discarded many images, in order to produce a balanced dataset. This is a wrong approach, in our perspective, because they have manually reduced the challenges of the dataset. Moreover, different datasets can have different kinds of statistics; thus, combining all the datasets into one is not feasible. A.M.Khan et al. [49] claimed that still now there is no balanced Covid dataset available online. To the best of our knowledge, our proposed method is the first attempt to provide an automatic deep learning framework that can efficiently work on a highly class imbalanced Covid-19 dataset. Unlike other existing research work, our proposed augmentation technique (of preparing a dataset) is purely automatic.

In this research, we intentionally chose this dataset with the challenges mentioned above. Many researchers already found 97–98 percentage accuracy, F1-score, precision and recall for covid-19 detection from other existing CXR datasets [36,58]. However, the same is not true for the employed dataset. We found conventional CNN model does not work efficiently for this employed dataset [33] due to the challenges as mentioned earlier. Hence, we believe that there is a huge research scope to further improve the existing models' performance for this challengeable dataset.

Our research mainly focuses on resolving the class imbalance problem in a highly imbalanced and challengeable dataset. We have not found an equally challenged and imbalanced dataset for Covid-19 detection from CXR images; thus, other existing datasets are not feasible for our proposed methodology. Instead, we have prepared three more augmented datasets along with this original dataset. We tested the performance of two different CNN models on those datasets for the validity of the proposed framework.

3. Methodology

The methodology for our proposed method can be divided into three parts: I. Choosing a suitable CNN model, II. SVD-CLAHE Boosting for class imbalance problem, III. Balanced Weighted Categorical Cross-Entropy (BWCCE).

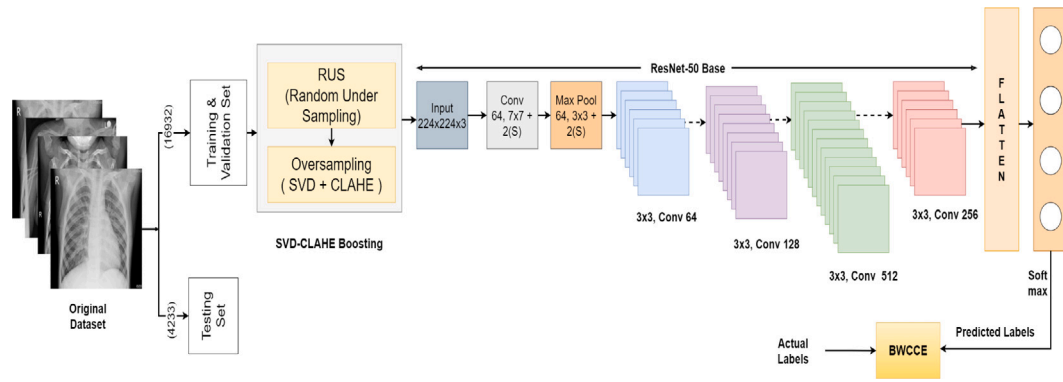


Fig. 1. Block Diagram of entire proposed model (SVD-CLAHE Boosting + ResNet-50 + BWCCE).

3.1. Choosing suitable CNN model

Since the employed dataset is very challenging, which is mentioned in the previous section, our first task is to find a suitable CNN model for this dataset. We have tested various CNN models like InceptionV3 [46], Xception [59], DenseNet-121 [44], VGG 16, VGG 19 [56], ResNet-50 [43] etc. on this dataset, which is further presented in the results and analysis section. We observed that VGG 16, VGG19, and ResNet-50 have slightly better results than Xception, DenseNet-121, InceptionV3, etc. VGG models are also known for their simplicity which is a direct modification of Alex-Net. Because of this simplicity, their model converges faster, and consequently, they have better performances than other complicated models (Inception-V3, DenseNet-121, Xception). ResNet-50 model also performs efficiently for this dataset, despite its complicated structure. To the best of our knowledge, due to the skip connection present in the ResNet-50 model [43] it can alleviate the problem of vanishing gradients, which generally appear during weights updating by the back-propagation algorithm (especially for a large and complicated network like InceptionV3, Xception). Moreover, the ResNet-50 model has a higher (50) number of layers, enabling the network to make very complicated decisions. Thus, we have chosen ResNet-50 as our proposed model. Additionally, we have also employed VGG-19 to check the validity of the proposed framework (i.e., SVD-CLAHE boosting + BWCCE loss function). The entire proposed model is presented in Fig. 1.

3.2. SVD-CLAHE boosting for class imbalance problem

Any class imbalance problem can be defined as follows: Assume the data samples in an image dataset is represented by $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m)\}$ where m is the total number of samples in the dataset, X_i is i th original image and Y_i is its corresponding label. The total number of classes in the dataset is K . Thus, labeled data $y \in [1, 2, \dots, K]$.

The estimated probabilities for each class are represented by $p = [p_{i,c1}, \dots, p_{i,cL}, \dots, p_{i,cK}]^T$, where $p_{i,cL} \in [0, 1], \forall L = 1, 2, \dots, K$; here $p_{i,cL}$ means that the probability of a sample ' i ', will be correctly classified into the class ' cL '.

A dataset is said to have class imbalance problem if and only if:

$$\frac{n_{cl}}{m} \ll \frac{n_{cH}}{m} \quad (1)$$

where n_{cl} is the number of samples in a minor class ' cl ', and n_{cH} is the number of samples in a major class ' cH '. The name ' cl ' is chosen for the lower class, the name ' cH ' is chosen to indicate the higher class, $m = \sum_c n_c$ is the total number of samples in the dataset.

In Eq. (1), $\frac{n_{cl}}{m}$ can also be interpreted as the probability that one sample will be correctly classified into minor class ' cl '. Thus, the purpose of any over-sampling method is to significantly increase this probability of classification for minor classes so that it will be comparable to the probability of classification for major classes.

Hence, the purpose of any over-sampling method is,

$$p_{i,cl} \approx p_{j,cH} \quad (2)$$

Here, $p_{i,cl}$ is the probability of i th sample, correctly classified into minor class ' cl '. $p_{j,cH}$ is the probability of j th sample, correctly classified into major class ' cH '.

The performance of ResNet-50 is fraught with the problem of a highly class imbalance of dataset, although it is working better than other models. This can be further observed in Section 5. We have noticed a slight fluctuation in the precision and recall results for different classes; moreover, we think 90% accuracy (by ResNet-50) can be further improved for the multi-class classification task for this imbalanced dataset.

This paper develops a novel data-augmentation technique, based on contrast enhancement by SVD based method and CLAHE method. Our proposed data-augmentation method neither generates exactly same images, nor they generate very dis-similar images, in terms of statistics. Thus, it overcomes the trade-off of conventional data augmentation mentioned earlier in Section 1. The idea is to generate a synthetic, pre-processed, and balanced dataset so that the model has enough images per class to learn how to distinguish various classes of images. By SVD based contrast enhancement, we produce synthetic images with slightly different statistics in terms of luminance and contrast. Whereas CLAHE 0.5 and CLAHE 1.0 further improved the clarity and contrast of the images, which can strengthen the edges during feature extraction by the CNN model. Some of the pre-processed images by CLAHE 0.5 and SVD+CLAHE 0.5 are shown in Fig. 2. We call our proposed oversampling and RUS under-sampling method SVD-CLAHE Boosting because it can significantly boost the performance of a standard CNN model by alleviating the class imbalance problem from the dataset.

The exact method of the proposed SVD-CLAHE Boosting method is further explained below in-depth in three parts: (a) SVD based contrast enhancement method, (b) CLAHE Contrast enhancement method, (c) Proposed SVD-CLAHE boosting according to intra-class variance.

3.2.1. SVD based contrast enhancement method

The first step of this method is to decompose the images into two orthogonal matrices and one singular value (diagonal) matrix by Singular Value Decomposition (SVD) [60], which is given in Eq. (3). After that, we modify only the singular matrix by multiplying it with a real constant, named as 'ratio'. In Eq. (3), these two decomposed orthogonal matrices have all the essential information of the images; thus, only multiplying a real constant value with the singular value matrix does not lose any critical information, to the best of our knowledge.

There was already a similar method of SVD based contrast enhancement, which was proposed for satellite images [61]; however, by their method, the value of ratio does not give a higher value for the employed Covid image dataset. Thus, we proposed a novel SVD based contrast enhancement method that will be suitable for the CXR

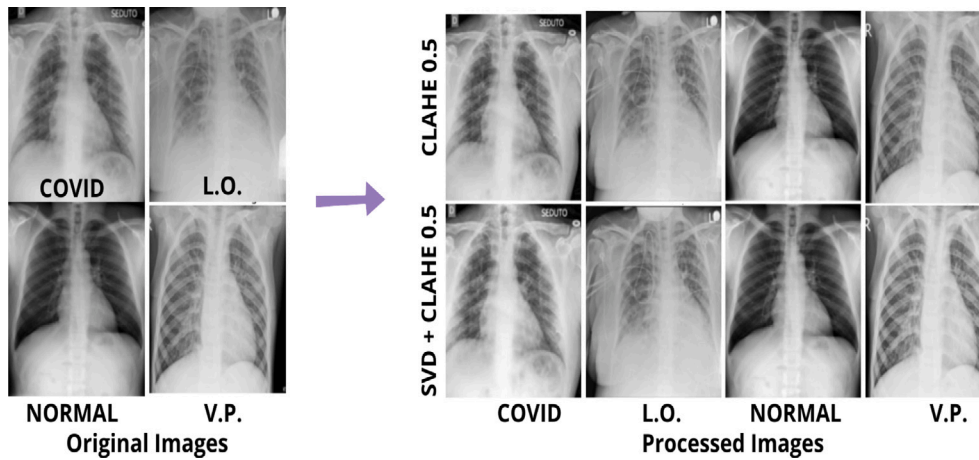


Fig. 2. Example of proposed Augmented Dataset (by SVD-CLAHE Boosting).

dataset. In our proposed method, we choose a reference image with very good contrast. We choose the value of ‘ratio’ according to the contrast difference between the reference image and the source image. The SVD contrast enhancement method is further explained below in Eqs. (3)–(8).

Step1: Decompose I into U , S and V , whereas U , V are orthogonal matrices, S is the singular value matrix, which is a diagonal matrix.

$$I = USV \quad (3)$$

Step2: The real constant ‘ratio’ is determined in the following way, which is presented from Eq. (4)–(7). Here, σ_g means global standard deviation, and μ_g means the global mean of the image. C_{tar} is the global contrast of the target image, C_{So} is the global contrast of the source image.

$$C_{tar} = \frac{\sigma_g(tar)}{\mu_g(tar)} \quad (4)$$

$$C_{So} = \frac{\sigma_g(Source)}{\mu_g(Source)} \quad (5)$$

if $C_{tar} > C_{So}$

$$ratio = 1.05 + \frac{C_{tar} - C_{So}}{C_{tar} + C_{So}} \quad (6)$$

else

$$ratio = 1.05 \quad (7)$$

Step3: Multiply ‘ S ’ (the singular matrix) with the real constant ‘ratio’ given in the following Eq. (8).

$$S' = S * ratio \quad (8)$$

Step4: Concatenate $US'V$ into I_p , whereas, I_p is the processed image.

Explanation: The idea is to generate some synthetic images by modifying the singular matrix in SVD space. In order to fix the constant ‘ratio’ (by which the singular matrix is to be multiplied), we have chosen one target image from each class that has very good contrast and luminance property. After that, we compute the global contrast, according to S. Roy et al. [62], which is nothing but the ratio of global standard deviation to the global mean value for both reference image and source image. These are in Eqs. (4) and (5), respectively. We fix the ratio by Eq. (6), which should be proportional to the contrast between the reference and source images. A similar kind of adaptive transformation is recently proposed by S. Roy et al. [63] for color normalization of histopathology images. However, we observe that some CXR images ($C_{tar} - C_{So}$) can be negative. Because there is no guarantee that reference image contrast is always greater than that

of any other source images in the dataset. In that case, no further contrast enhancement is required because the source image contrast is already greater than that of the reference image. Thus, the value of ‘ratio’ should be chosen 1. However, we choose the by-default value of ‘ratio’ to 1.05 instead of 1, which are given in Eqs. (6) and (7). Because it will ensure that there will always be around 5 percent contrast enhancement for every image. Otherwise, the same images could be produced by the proposed SVD based contrast enhancement, which is not desirable for data augmentation, mentioned in Section 1. Therefore, it was necessary to produce slightly different images (in terms of contrast and luminance statistics) during oversampling in a minor class, thereby choosing the default value of ratio to 1.05. Some of the images from the proposed augmented dataset are shown in Fig. 2.

3.2.2. CLAHE Contrast enhancement method

Contrast Limited Adaptive Histogram Equalization (CLAHE) [64] is a modified version of Global Histogram Equalization (GHE) [65] in which histogram stretching operation is limited by a maximum clipping value (e.g., 0.5, 1.0). Here HE method is employed locally, which means it is performed in every local region (or window) of the image. We have chosen this window size 8×8 empirically for this operation. This is to clarify that, CLAHE 0.5 means it specifies a scale of contrast clipping 0.5 (this is not a version). For example, CLAHE 0.5 will do less contrast enhancement than CLAHE 1.0. However, we have also noticed that the conventional CLAHE 2.0 method (according to open cv library it is, CLAHE 2.0) is not inevitable of data loss. It can do over-contrast enhancement for some CXR images. Consequently, it might not work for CNN models. Therefore, we have chosen CLAHE 0.5 and CLAHE 1.0 so that there will not be excess contrast enhancement and, consequently, there will be comparatively little data loss. In order to ensure that fact, we have computed the correlation coefficient between the processed image and the original image in the overall CXR dataset for different scales of CLAHE. We have found that the mean correlation co-efficient value does not go less than 0.93 for CLAHE 0.5 and does not go less than 0.90 for CLAHE 1.0. Therefore, we have empirically chosen the CLAHE 0.5 and CLAHE 1.0 contrast enhancement methods with an 8×8 window for data augmentation.

3.2.3. Proposed SVD-CLAHE Boosting according to intra-class variance

We have done a series of experiments on SVD-CLAHE boosting. Table 1 represents a different number of images per class we have chosen for various experiments. The results of their performances can be found in more depth in Section 5. First, we have performed conventional data-augmentation techniques [12] like rotation, flipping (horizontal and vertical), zooming, and cropping in minor classes (e.g., Viral Pneumonia and the Covid class) in order to do oversampling. However, we

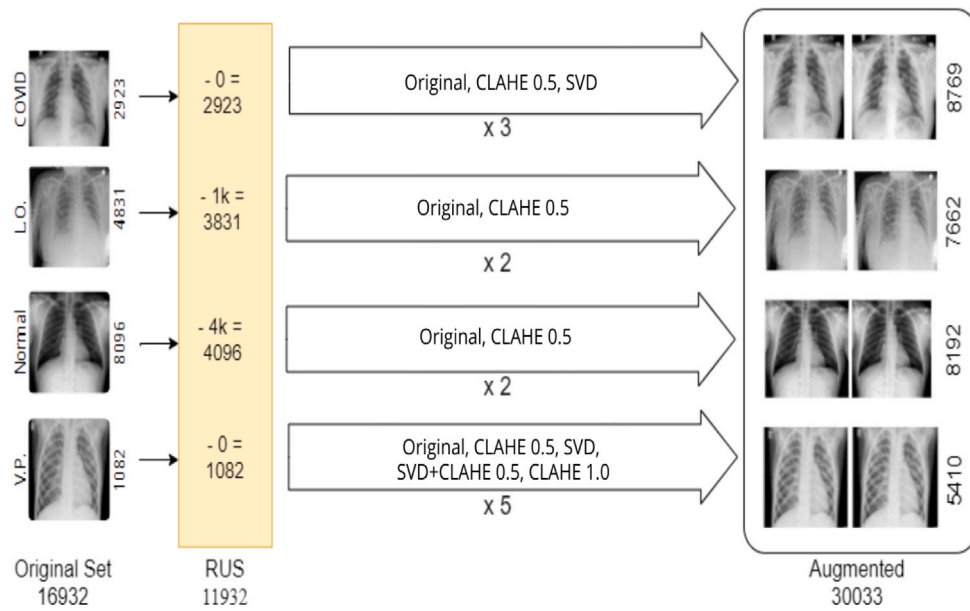


Fig. 3. The entire scheme of proposed SVD-CLAHE Boosting.

Table 1
Comparisons of various experiments with different number of images per classes.

| Dataset (training) | Covid | Normal | LO | VP |
|--|-------|--------|------|------|
| Original | 2923 | 8096 | 4831 | 1082 |
| SVD-CLAHE Boosting (Equal images per class) | 5355 | 5355 | 5355 | 5355 |
| Proposed SVD- CLAHE Boosting | 8769 | 8192 | 7662 | 5410 |

observed that incorporating such data augmentation produces worse results.

Because we have found that these data augmentation techniques are not exactly featured invariant for these CXR images, for example, if we flip those images horizontally, we notice a very different image is produced. This will differ the feature extraction process for the classification task. Consequently, it further increases the complexity during the training of the CNN model. Later, we generate an augmented dataset by SVD-CLAHE oversampling and RUS, where each class has an equal number of images (5355) to resolve the datasets class imbalance problem. This is shown in Table 1. However, we have found that producing a precisely same (or very similar) number of images per class did not resolve this class imbalance problem entirely from this dataset. However, it improved the model performance a little bit.

In the final experiment, a dataset of 30,033 images is prepared by the proposed SVD-CLAHE Boosting method. In this method, the number of images per class is chosen based on an intra-class variance in each class. To the best of our knowledge, if in one class the intra-class variance among the images is huge (that means different images have different kinds of statistics in a class), in that case, the number of images trained to the CNN model must also be higher for the corresponding class. Because of higher statistical variance, neural networks may need a higher number of such images to converge faster. Hence, intra-class variance significantly impacts this kind of dataset for the final classification task. We have chosen ten images per class with the most similarity among all images in that class to compute the intra-class variance in the dataset. We have employed five different (expert) people in order to choose the most similar images from each class, and thereafter, we computed intra-similarity by estimating correlation co-efficient [66] between the chosen image and all other images in that class, which is given in Eq. (9). This is to clarify that, here

Table 2
Mean Values of Intra-class similarity and Intra-class variance for different classes of Covid 19 dataset.

| Classes | Covid | Normal | LO | VP |
|------------------------|---------------|--------|--------|---------------|
| Intra-class similarity | 0.5283 | 0.5572 | 0.5852 | 0.7052 |
| Intra-class variance | 0.4717 | 0.4428 | 0.4147 | 0.2958 |

expert people means they are experts in computer vision task and thus, understand well statistical similarity inside images. Subsequently, we take an average of all such values and subtract it from one to compute the final intra-class variance.

$$\sigma(\text{intra}) = 1 - \frac{1}{M(N-1)} \sum_{m=1}^M \sum_{n=1}^{N-1} \text{Corr}(x_m, y_n) \quad (9)$$

where $M = 10$ per class, which is the chosen number of images having the most similarity per class, $N = a$ total number of images in one class, x_m is the selected image (from those 10 images) in one class which are the most similar, y_n is any other image in that class, Corr is representing correlation coefficient between two random variables x and y . Having more intra-class similarity (structural similarity) means that it has a lesser intra-class variance. Therefore, we subtracted the computed correlation coefficient from 1 in Eq. (9) to have a probabilistic kind of intuition to compute the final intra-class variance in each class.

We have found mean intra-class variance for different classes are 0.4717, 0.4428, 0.4147, and 0.2948 for Covid, Normal, Lung Opacity, and Viral Pneumonia classes, respectively, according to the Eq. (9). This can be further observed from Table 2, that the Covid class has the highest intra-class variance, which means that in the covid class the statistics among images are comparatively dis-similar than other classes.

Hence, after incorporating the notion of intra-class variance, the purpose of proposed SVD-CLAHE Boosting will be slightly different than Eq. (2). This is given by Eq. (10).

$$p_{i,c_l} * [1 - \sigma_{c_l}(\text{intra})] \approx p_{j,c_H} * [1 - \sigma_{c_H}(\text{intra})] \quad (10)$$

Here p_{i,c_l} is the probability that a sample 'i' will be correctly classified taken from minor class 'c_l'. p_{j,c_H} is the probability of a sample 'j' will be correctly classified taken from major class 'c_H'. Indeed, we found for this SVD-CLAHE Boosting, for minor class probability p_{i,c_l}

multiplied by its intra-class similarity (that is, nothing but $[1 - \sigma_c]$), should be similar to that of major class. Hence, for the Viral Pneumonia (VP) class, the probability of classification multiplied by 0.7052 (from Table 2), should be similar to the probability of classification for the Normal class multiplied by 0.5572. This perhaps enables us to choose fewer images from the VP class than other classes. Because in the case of the VP class, this intra-class similarity among images is higher than of other classes.

Viral Pneumonia (VP), which has the least number (1082) of images for training, is oversampled by SVD-based contrast enhancement, CLAHE 0.5 and CLAHE 1.0. Moreover, the proposed SVD+CLAHE 0.5 and the original image sets are also included to oversample VP (minor) class 5 times larger ($1082 \times 5 = 5410$) than previous. Now, keeping this number of image combination constant for the VP class, we try to compute what should be the number of images for other classes according to the formula of intra-class variance. We have found those numbers should be 8657, 7610, and 8125 for covid, lung opacity, and normal classes, respectively. Therefore, for covid class, we have generated thrice the size of the dataset ($2923 \times 3 = 8769$), by incorporating SVD contrast enhancement method and CLAHE 0.5 method, along with original images. Moreover, for lung opacity, we employed RUS in order to exclude 1k images from that class ($4831-1000 = 3831$). This is followed by CLAHE 0.5 method (along with original images), to get twice of rest of the data size ($3831 \times 2 = 7662$). For normal class, we have excluded 4k images by RUS, followed by CLAHE 0.5 (along with original) to twice the remaining dataset ($4096 \times 2 = 8192$). This can be observed that the new augmented dataset (of 30,033 images) has a very similar number of images in each class, compared to the required number of images, according to its intra-class variance. The entire scheme of generating this augmented dataset is presented in Fig. 3.

3.3. Balanced Weighted Categorical Cross Entropy (BWCCE)

By employing SVD-CLAHE Boosting, we generated a new augmented dataset, which is more balanced than the original dataset. However, we observed that the performance of the ResNet-50 model on this augmented dataset is still a little bit imbalanced. This can be observed from its classification report for different classes and its confusion matrix. Therefore, a new loss function is proposed to alleviate a little bit of class imbalance from the model after employing SVD-CLAHE Boosting.

Categorical Cross-Entropy (CCE) [67] loss function can be expressed as the following mathematical formula.

$$L_{CCE}(y, p) = -\frac{1}{N} * \sum_{i=1}^N \sum_{c=1}^K y_{i,c} * \log(p_{i,c}) \quad (11)$$

where, $p_{i,c}$ is the probability that the i th sample belongs to class c , $y_{i,c}$ is the class weight for i th sample for total K no of classes and for each class $y_{i,c}$ is same, N is the total number of samples in a particular class.

In case of conventional CCE, $y_{i,c}$ is by-default chosen as

$$y_{i,c} = 1 \quad (12)$$

For class imbalance problem, Weighted Categorical Cross-Entropy (WCCE) is widely employed by researchers [24,25]. The expression of the WCCE loss function is given in the following equation.

$$L_{WCCE}(y, p) = -\frac{1}{N} * \sum_{i=1}^N \sum_{c=1}^K \beta_c * \log(p_{i,c}) \quad (13)$$

Whereas β_c is the weight in each class. β of the minor classes must be inversely proportional to the number of images per class and β of the major classes keep remain the same, that is 1. However, the number of images in minor class may differ a lot from the same of major class. In that case, β of the minor class can be a very higher value. Consequently, the CNN model will be more biased to a particular

(minor) class, and it will further push the accuracy and precision of the minor class closer to 1. However, the accuracy and precision of other classes will fluctuate considerably in that case, which is not desirable. This is further demonstrated in Section 4.

In order to resolve the problem mentioned above, we propose a novel BWCCE loss function in which bias weights β are assigned inversely proportional to the number of images in class. However, unlike WCCE, bias weights of BWCCE are assigned based on the probability notion. The following equation can express the proposed BWCCE loss function.

$$L_{BWCCE}(y, p) = -\frac{1}{N} * \sum_{i=1}^N \sum_{c=1}^K \beta_c * \log(p_{i,c}) \quad (14)$$

where, the bias weight in each class, β_c is given by the following equation.

$$\beta_c = \frac{1}{K-1} * (1 - \frac{n_c}{\sum_c n_c}) \quad (15)$$

where, ($K > 1$)

Here, in Eq. (15), n_c is the total number of samples in class c , $\sum_c n_c$ is the total number of samples in the entire dataset, ($K - 1$) is just a normalization factor which ensures that sum of all β 's will not go beyond the value 1, K is a total number of classes.

Lemma 1. If weight of each class in a Weighted Cross Entropy, is chosen as $\beta_c = \frac{1}{K-1} * (1 - \frac{n_c}{\sum_c n_c})$, where $\forall c = 1, 2, \dots, K$ and ($K > 1$), then sum of weights of all class will be equal to 1.

That is,

$$\sum_{c=1}^K \beta_c = 1 \quad (16)$$

Proof. The number of classes = K . We have to prove that $\sum_{c=1}^K \beta_c = 1$ if $\beta_c = \frac{1}{K-1} * (1 - \frac{n_c}{\sum_c n_c})$. The total number of images in entire dataset $\sum_c n_c$ will be a same constant number, let us assume this 'm'.

Thus, substituting the value from Eq. (15) into Eq. (16) we get,

$$\sum_{c=1}^K \beta_c = \frac{1}{K-1} [(1 - \frac{n_1}{m}) + (1 - \frac{n_2}{m}) + \dots + (1 - \frac{n_K}{m})] \quad (17)$$

$$\text{or, } \sum_{c=1}^K \beta_c = \frac{1}{K-1} [K - \frac{(n_1 + n_2 + \dots + n_K)}{m}] \quad (18)$$

$$\text{Now, } (n_1 + n_2 + \dots + n_K) = \sum_{c=1}^K n_c = m \quad (19)$$

Hence, substituting the value from Eq. (19) into Eq. (18), we get,

$$\sum_{c=1}^K \beta_c = \frac{1}{K-1} (K-1) = 1 (\text{Proved}) \quad (20)$$

As $K \neq 1$, from Eq. (20), this is proved that sum of β 's for BWCCE will be equal to 1. Hence, the proposed loss function BWCCE supports the notion of probability for assigning weights. Therefore, we have observed that the proposed loss function BWCCE produces much more balanced classification results than conventional CCE and WCCE. In every class, the results of accuracy, precision, and recall are a little more balanced (do not fluctuate too much) by the proposed BWCCE.

Explanation: First, let us understand why did we choose $\beta_c = \frac{1}{K-1} * (1 - \frac{n_c}{\sum_c n_c})$. Here, we assigned weights β with the notion of probability. The ratio $\frac{n_c}{\sum_c n_c}$ perhaps shows a probabilistic perspective, and thus, it is subtracted from 1, which is the total probability. For example, the Viral Pneumonia has 5410 images after the proposed SVD-CLAHE boosting out of 30,033 images. Hence, the probability that one sample from viral pneumonia will be classified correctly is $5410/30,033 = 0.1801$. This is comparatively lesser than the average probability for four classes (i.e., 0.25). Thus, subtracting this probability value from 1

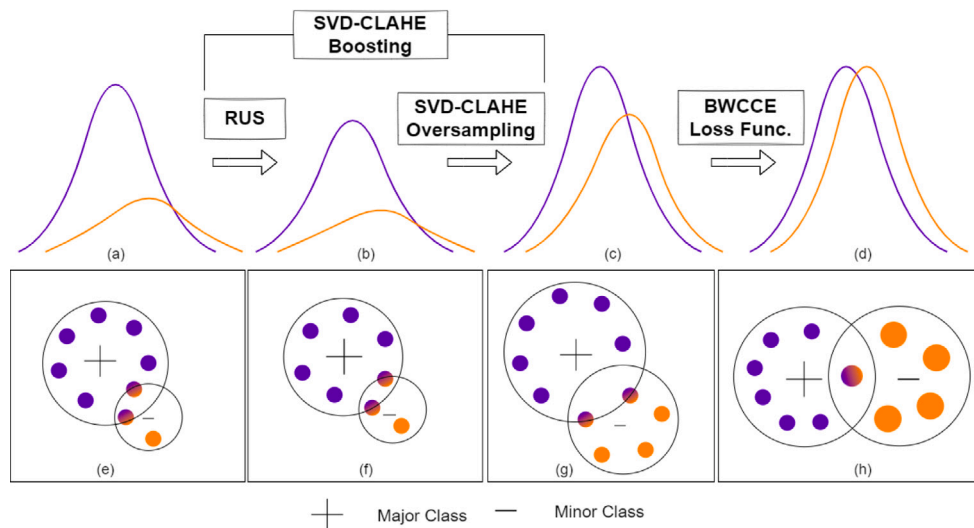


Fig. 4. Visualization of entire proposed methodology Fig. 4(a) presents the distribution of major and minor classes in the original dataset, based on a number of images. Figs. 4(b) to 4(d) present the changes in distribution after employing the proposed methodology. Fig. 4(e) indicates the cluster representation of major and minor classes for the original dataset. Fig. 4(f) to Fig. 4(h) indicate the changes in cluster representation in major and minor classes after employing RUS, SVD-CLAHE Over-sampling, and BWCCE loss function, respectively. All these diagrams of distribution and cluster representation are completely imaginary and have not been taken from any statistical plot of the dataset.

(which is the sum of total probability) and then normalizing the value gives you the intuition of how much bias weight should be chosen for viral pneumonia. That is, $\frac{1}{3} * (1 - 0.1801) = 0.2733$ which is now a slightly higher than average weight per class i.e., 0.25. Here $(4 - 1 = 3)$ is the normalization factor, because number of classes is 4. In Lemma 1, it is already proved that sum of β 's will be equal to one, if the normalization factor is $(K - 1)$. Similarly, if you compute the β for the major class, we will get a lesser weight, i.e., less than the average weight 0.25. Unlike WCCE, the weights of different classes for BWCCE are not exactly chosen inversely proportional to number of images in that class, rather it is chosen based on the notion of probability of classes. Thus, weights for different classes do not deviate too much from the average weight 0.25. Consequently, deep neural networks will not be biased too much towards one particular class for BWCCE. In the results and analysis section, this can be further observed that ResNet-50 with BWCCE loss function not only improves classification results a little bit, but also, it enables the neural network to provide a very stable validation graph, which is a significant improvement. Furthermore, we have observed that ResNet-50 model along with BWCCE loss function, converges a little bit faster, during training. This justifies the necessity of employing the proposed loss function.

3.4. Physical interpretation of proposed methodology

From Fig. 4, the necessity of the proposed methodology can be easily visualized. Figs. 4(a)–4(d) represent the distribution of major (+) and minor (-) class based on several images. Clearly, Fig. 4(a) shows that there is a huge class imbalance in the original dataset, which means the number of images in major class and minor class is very different. Fig. 4(b) indicates the distribution of major class peak is considerably reduced after employing RUS, whereas the distribution of minor class is same. Because images are only excluded from major classes, but not from minor classes. Fig. 4(c) represents major and minor class distribution after employing SVD-CLAHE oversampling. Clearly peak of both major and minor classes is enhanced after incorporating SVD-CLAHE oversampling, so that the major class distribution is now comparable to the distribution of minor class. Fig. 4(d) represents major and minor class distribution after employing the BWCCE loss function. Clearly, the BWCCE function was deployed to alleviate a little bit of class imbalance present after SVD-CLAHE Boosting (i.e., RUS + SVD-CLAHE Over-sampling). Fig. 4(d) shows that both classes' distribution becomes exactly equivalent because by the BWCCE loss function, the

minor class is given more weightage (or, bias) than the major class. Thus, effectively the distribution of both classes becomes equivalent. Figs. 4(e)–4(h) represents major and minor classes cluster visualization. Increasing the number of inner circles in a class is equivalent to increasing the number of samples in that class. Increasing the size of the inner circle in a class is equivalent to increasing the weight bias in that class. In Fig. 4(f), this can be visualized that the number of inner circles in major class decreases after RUS, which means the number of samples is decreased from the major class by RUS. Fig. 4(g) indicates that the number of inner circles in both major and minor classes is considerably increased after incorporating SVD-CLAHE over-sampling. Now the size of the bigger circle for major and minor classes is kind of comparable. However, still, there was little class imbalance present after employing SVD-CLAHE Boosting. Fig. 4(h) represents the final cluster after employing SVD-CLAHE Boosting and BWCCE loss function. Clearly, the size of the inner circles of the minor class is increased here, which means in the BWCCE loss function, the minor class is getting more biased than the major class. Hence, finally, the bigger circle of both major and minor classes in Fig. 4(h) is now equivalent. This reveals that our proposed methodology makes the distribution of major class and minor class similar; hence it alleviates the class imbalance problem completely. All these diagrams of distribution and cluster are entirely imaginary. They have not been taken from any statistical plot of the dataset, and they are just employed for better visualization of the proposed methodology.

4. Experimental results and analysis

The following points can summarize the entire experimental results and analysis section:

1. First, experimental results of several CNN models like InceptionV3, DenseNet-121, VGG 16, VGG 19, ResNet-50 etc. are compared for this Covid CXR dataset and shown in Table 3. Moreover, two existing models (only for Covid detection), Covid-Lite by M.Shiddhartha et al. [35], and Covid-Net by Wang et al. [36], are implemented on this CXR dataset. This can be observed from Table 3 that ResNet-50 and VGG-19 have better results than other models. The reason was already mentioned in Section 4.1.

2. We have chosen the ResNet-50 model as the proposed model. Moreover, different data-augmentation techniques are employed to prepare three more augmented datasets and the original dataset. These augmented datasets are: (I) an augmented dataset by traditional data-augmentation technique (i.e., by rotation, flipping, zooming, etc.), (II) an augmented dataset prepared by SVD-CLAHE Boosting in which a number of images per class is the same, and finally (III) the proposed augmented dataset by SVD-CLAHE Boosting, in which a number of images per class is chosen based on intra-class variance. After that, the ResNet-50 model is performed on these four datasets. Furthermore, different loss functions are incorporated in the ResNet-50 model for the proposed augmented dataset. All the results of the ResNet-50 model experiments are presented in Table 4.
3. We have chosen another CNN model, VGG-19, to check the proposed framework's validity. Thus, the same experiments are also conducted on the VGG-19 model. The results of VGG-19 are presented in Table 5.
4. A comparison of training and validation graphs for both ResNet-50 and VGG-19 models are shown in Fig. 5 and Fig. 6, respectively.

4.1. Training specification

- All of the experiments have been performed using Tensorflow and Keras. Tesla P100 GPU, provided by Google Colab Pro service, is used to train the models. 25 GB RAM was also available from the service to prevent RAM crashes during the experiments.
- Adams-optimizer is employed as the preferred choice of optimizer for all the experiments with a learning rate of $1e^{-7}$, beta 1 value of 0.9, beta 2 value of 0.999, and epsilon value of $1e^{-7}$.
- A batch size of 32 was used while training CNN models. All the experiments are performed with 80% training data and 20% testing data of the total dataset. This training-testing data splitting is done randomly. Out of entire training dataset, 20 percent are further chosen as a validation set randomly, to check the validity of model performance.
- All the images are resized into 224×224 before feeding them into the CNN model.
- Transfer learning approach is deployed for training all the standard CNN models. All these models are pretrained on the ImageNet dataset [68] (by Keras) and finetuned on CXR dataset. However, existing Covid-Lite and Covid-Net models are trained from the scratch since their pre-trained weights are not available on Keras.
- Early stopping call-back to stop the model training (if the model starts to overfit) is employed based on the validation loss. Patience of 5 epochs is incorporated in the Early Stopping call-back, which minimizes the validation loss. That means, if the validation loss keeps increasing for 5 consecutive epochs, then the call-back will automatically stop the model training in order to avoid overfitting. Weights are restored to the last checkpoints where the model was not overfitting.
- All the experiments are performed with the same hyperparameters and with the same aforementioned methodology during training.

4.2. Evaluation metrics

Standard 'accuracy' metric is used as evaluation metric to check how accurately the model is working. Besides that precision, recall as well as F1 Score are also evaluated in order to make sure the model is not affected by any class imbalance problem. The mathematical formulae for the aforementioned metrics are given below:

$$Precision = \frac{TP}{TP + FP} \quad (21)$$

Table 3

Comparisons of Mean values of evaluation metrics for various existing models with the proposed framework on the test set.

| Methodology | F1 score | Accuracy | Precision | Recall | AUC |
|--|-------------|-------------|-------------|-------------|-------------|
| Covid Lite by [35] | 0.85 | 0.85 | 0.84 | 0.87 | 0.88 |
| Covid-Net by [36] | 0.87 | 0.87 | 0.88 | 0.86 | 0.88 |
| Inception-V3 | 0.78 | 0.8 | 0.77 | 0.79 | 0.80 |
| Xception | 0.79 | 0.8 | 0.81 | 0.77 | 0.78 |
| DenseNet-121 | 0.84 | 0.82 | 0.82 | 0.86 | 0.89 |
| VGG-16 | 0.88 | 0.88 | 0.89 | 0.87 | 0.90 |
| VGG-19 | 0.91 | 0.90 | 0.92 | 0.90 | 0.91 |
| ResNet-50 | 0.90 | 0.89 | 0.91 | 0.91 | 0.92 |
| VGG-19 with proposed framework | 0.95 | 0.94 | 0.96 | 0.95 | 0.96 |
| ResNet-50 with proposed framework | 0.95 | 0.94 | 0.96 | 0.95 | 0.96 |

$$Recall = \frac{TP}{TP + FN} \quad (22)$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (23)$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (24)$$

TP stands for True Positives, i.e., instances labeled as positive and classified correctly as a positive instance. TN stands for True negatives, which represent instances where the actual label is negative and is predicted likewise. FP stands for False positives where the instance belongs to the negative label but is predicted to be positive. FN stands for False negatives, where instances are positive but are predicted as negative. Besides these metrics, we have also computed AUC metric which is an important metric for measuring class imbalance.

4.3. Results analysis and discussions

First, various deep convolutional neural networks, namely, ReNet50, InceptionV3, DenseNet-121, VGG-16, VGG-19, Covid-Lite, proposed by M. Siddhartha et al. [35], and Covid-Net by Wang et al. [36] are performed to the imbalance dataset. This provides us an insight into the problem by observing how this dataset responds to naive methods of Deep Learning. This can be further observed from Table 3. The exact training specification mentioned in Section 4.1 is followed for all the experiments, except the Covid-Net and Covid-Lite models.

For the Covid-Net model, a deep learning framework with PEPX blocks proposed by Wang et al. [36], has been trained on the original dataset from scratch with 183,695,108 parameters with all of them as trainable. A batch size of 8, for 30 epochs is employed and call-back of early stopping is taken with the patience of 5. The model ran for 25 epochs to give an accuracy of 0.87 on the testing data. Although this model is trained from scratch on the original dataset, it provides good results in terms of accuracy, F1 score, precision, and recall which can be observed from Table 3. However, we have found that this model is still a slightly complicated model (having many no. of layers) and thus, inducing overfitting in this CXR dataset.

For Covid-Lite model, original dataset is pre-processed by CLAHE and white balance image processing techniques, as mentioned by M.Shiddhartha et al. [35] in their paper. The dataset is then split into training, validation and testing sets with a ratio of 0.7, 0.1 and 0.2. Thereafter, they are fed into the Covid-Lite model to train it from scratch for 50 epochs with a batch size of 32. It was observed that training and validation parameters are highly fluctuating, and

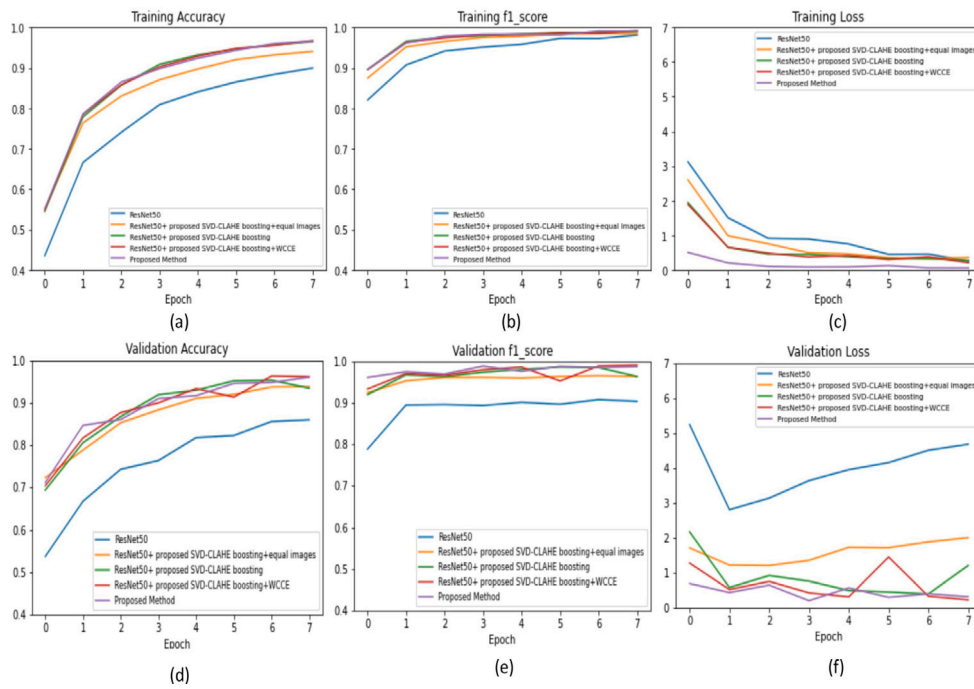


Fig. 5. Comparisons of performances of several experiments on ResNet-50 Model (a) training accuracy, (b) training F1 score, (c) training loss, (d) validation accuracy, (e) validation F1 score, (f) validation loss. The experiments are employed are already labeled in the diagram, those are ResNet-50 on original dataset, ResNet-50 on augmented dataset (by SVD-CLAHE boosting) with equal no of images per class, ResNet-50 on augmented dataset (by proposed SVD-CLAHE boosting), ResNet-50+ SVD-CLAHE Boosting +WCCE, proposed method (ResNet-50+ SVD-CLAHE Boosting +BWCCE).

the number of trainable parameters is 1,019,396. The performance by their model is little bit poor. Due to the class imbalance problem present in the original dataset, there is huge fluctuation of results in their classification report. Their proposed model is only suitable for a balanced dataset.

Standard CNN models like Inception-V3, Xception, and DenseNet-121 have little bit poor results than other models due to the huge complexity of their models. ResNet50, VGG 16, and VGG 19 performed better than other models in terms of accuracy, precision, recall, and F1 score, which can be observed in Table 3. VGG 16 and VGG 19 are comparatively simpler models than Inceptionv3, Xception, and DenseNet-121. Therefore, those VGG models have no overfitting in their model performances. ResNet-50 is another model that shows promising results, shown in Table 3, despite its complicated structure. We believe that the skip connection present in the model can alleviate the problem of vanishing gradients, and it also reduces the complexity of the model a little bit.

We have chosen ResNet-50 as proposed model over VGG-19 and VGG-16, because we have observed the graphs of validation loss and validation accuracy (which can be observed in Fig. 5 and Fig. 6) fluctuates more for VGG-19. Moreover, according to our perspective, due to higher number of layers (50) present in ResNet-50, it can accomplish very complicated task and consequently its performance graph of validation is more stable than that of VGG-19. However, we have chosen VGG-19 for checking the validity of the proposed framework, because it is having the second best results after ResNet-50.

The following experiments are conducted with the ResNet-50 model. Important observations from all these experiments are presented below.

- First, conventional data-augmentation techniques like rotation, flipping (horizontal and vertical), zooming, etc., are employed for making an augmented dataset using the ImageDataGenerator, available in Tensorflow. After that, the ResNet-50 model is

Table 4
Ablation Study of of various experiments conducted on ResNet-50 model on the testing set.

| Methodology | F1-score | Accuracy | Precision | Recall | AUC |
|---|-------------|-------------|-------------|-------------|--------------|
| ResNet-50 with original dataset | 0.90 | 0.89 | 0.91 | 0.91 | 0.943 |
| ResNet-50 with conventional data augmentation (flipping,rotation) | 0.89 | 0.88 | 0.91 | 0.87 | 0.907 |
| ResNet-50 with SVD-CLAHE Boosting with equal no of images | 0.92 | 0.91 | 0.91 | 0.93 | 0.937 |
| ResNet-50 with proposed SVD-CLAHE Boosting | 0.94 | 0.93 | 0.96 | 0.93 | 0.966 |
| ResNet-50 with proposed SVD-CLAHE Boosting with WCCE | 0.94 | 0.94 | 0.96 | 0.93 | 0.965 |
| Proposed method (ResNet-50+SVD-CLAHE Boosting+BWCCE) | 0.95 | 0.94 | 0.96 | 0.95 | 0.967 |

performed on that augmented dataset. The ResNet-50 model with this conventional data augmentation produces poor results than the ResNet-50 model on the original dataset. This can be further noticed in Table 4. The reason we already discussed in Section 4. According to our perspective, these data-augmentation techniques are not feature invariant for the final classification task.

- SVD-CLAHE Boosting is performed to produce a balanced augmented dataset in which each class has an equal number of images. After that, the ResNet-50 model is performed on this augmented dataset. The performance of this data augmentation is

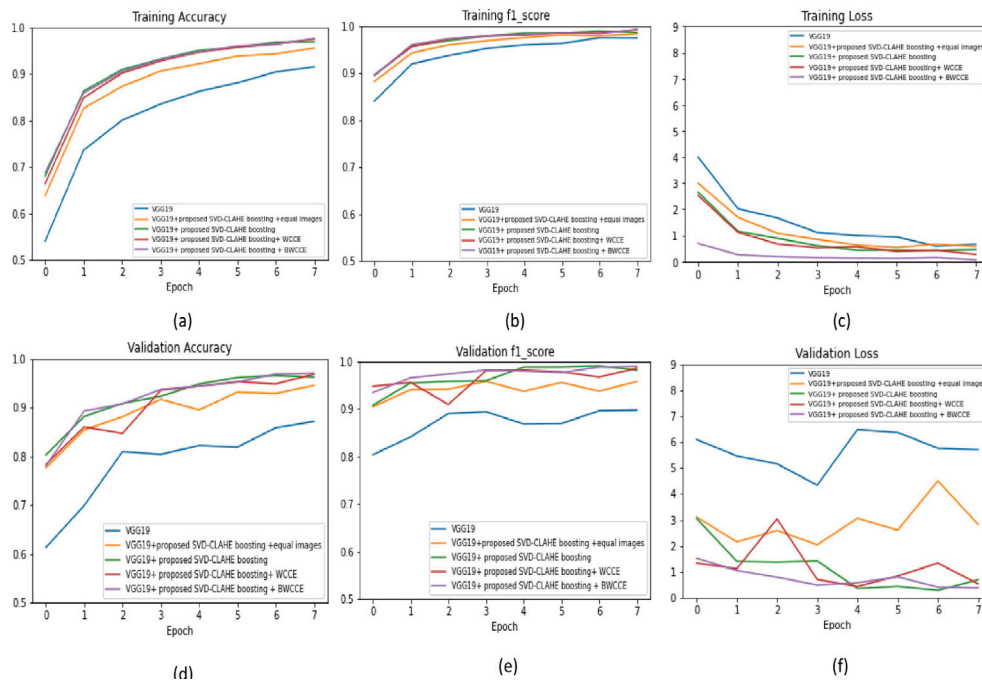


Fig. 6. Comparisons of performances of several experiments on VGG-19 Model (a) training accuracy, (b) training F1 score, (c) training loss, (d) validation accuracy, (e) validation F1 score, (f) validation loss. The experiments are employed are already labeled in the diagram, those are VGG-19 on original dataset, VGG-19 on augmented dataset (by SVD-CLAHE boosting) with equal no of images per class, VGG-19 on augmented dataset (by proposed SVD-CLAHE boosting), VGG-19+ SVD-CLAHE Boosting +WCCE, proposed method (VGG-19+ SVD-CLAHE Boosting +BWCCE).

slightly improved than the performance of ResNet-50 on the original dataset. F1 score, accuracy, and recall factor are improved by 1-2%, which can be observed in Table 4. However, by observing its confusion matrix, we found that little class imbalance exists.

- Proposed SVD-CLAHE Boosting (data-augmentation) method is performed to produce a third augmented dataset. The number of images per class is chosen according to the intra-class variance, which is already explained in Section 4. ResNet-50 model on this augmented dataset (by proposed SVD-CLAHE Boosting) has performed very effectively, as it produced accuracy, precision, recall, and F1 scores 93%, 96%, 93%, and 94% respectively, given in Table 4. This is an approx 4%–5% improvement from ResNet-50 with the original dataset. This is a significant improvement or boosting performance compared to other existing methods mentioned in Table 3. Thus, this justifies the proposed data-augmentation method and its name ‘SVD-CLAHE Boosting’.
- Although the proposed SVD-CLAHE boosting works very effectively, we still have found a little bit of class imbalance, which is noticed by its confusion matrix and from the classification report. Therefore, the Weighted Categorical Cross-Entropy (WCCE) loss function is incorporated in the ResNet-50 model, performed on the proposed augmented dataset. We have observed a little bit of improvement (1%) in the accuracy, compared to the performance of ResNet-50 on the proposed augmented dataset, after employing WCCE. However, from Fig. 5, this can be visualized that WCCE further incorporates a little bit of fluctuation of validation graphs for accuracy, F1 score, and loss. This is undesirable. Overall, we have observed that WCCE pushes the accuracy and precision of the model slightly more for minor classes, however, this induces a little bit of fluctuations in the performances among other classes.
- Therefore, we develop a novel Balanced Weighted Categorical Cross-Entropy loss function (BWCCE) to alleviate the limitation mentioned earlier. This can be observed from Table 4 that BWCCE performs slightly better than conventional CCE and WCCE for this proposed augmented dataset. ResNet-50 model (along with

BWCCE loss function) on the proposed augmented (30k) dataset has achieved the accuracy, precision, recall, and F1 scores of 94%, 96%, 95%, and 95%, respectively. Overall, there is a 1% improvement in accuracy, 1% improvement in F1 score, and 2% improvement in recall than the ResNet-50 model after SVD-CLAHE Boosting. Additionally, from Fig. 5 this can be observed that the proposed framework (that is, ResNet-50+SVD-CLAHE Boosting+BWCCE) provides the least validation loss; moreover, it provides validation accuracy and F1 score graph (shown in Fig. 5), more stable than that of WCCE and CCE. This justifies the effectiveness of the BWCCE loss function.

Furthermore, this has also been observed from Fig. 7(a) that the ResNet-50 model with BWCCE loss function converges much faster (by only 9 epochs) than that of CCE (20 epochs) and WCCE (13 epochs). Fig. 7(b) also shows that the average time taken per epoch during training by the proposed BWCCE loss function is significantly lesser than that of other loss functions. Hence, this can be concluded that the proposed BWCCE loss function simplifies the optimization problem of a complicated CNN model; thus, it enables the CNN model to converge much faster during training.

The confusion matrix in Fig. 8 helps us to visualize and analyze the prediction performances of different experiments in a better way. This gives a definite comparison of the evaluation metrics in a matrix form. For example, the values 705, 1012, 1804, and 270 are the True Positive values for the ResNet-50 model on an original dataset for the classes Covid, Lung Opacity, Normal and Viral Pneumonia, respectively. By comparing Fig. 8(a) and Fig. 8(b), we can visualize that true-positives for Lung Opacity (LO) is significantly increased, whereas Normal and VP classes are a little bit improved while utilizing SVD-CLAHE Boosting. However, it fails to improve performances for overall classes (not for Covid class). Because still, little class imbalance was present after SVD-CLAHE Boosting. We incorporated the BWCCE loss function on the ResNet-50 model to resolve this problem. Fig. 8(c) indicates that BWCCE enables us to improve overall performances in

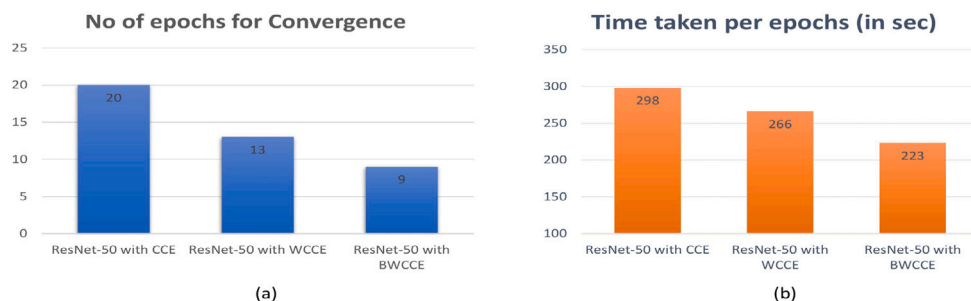


Fig. 7. (a) No of Epochs of Convergence for ResNet-50 with different loss functions on proposed augmented dataset, (b) Average time taken per epochs in sec, for ResNet-50 with different loss functions on proposed augmented dataset.

| CM1 | Covid | LO | Normal | VP | CM2 | Covid | LO | Normal | VP | CM3 | Covid | LO | Normal | VP |
|--------|-------|------|--------|-----|--------|-------|------|--------|-----|--------|-------|------|--------|-----|
| Covid | 705 | 61 | 51 | 2 | Covid | 690 | 23 | 52 | 0 | Covid | 717 | 15 | 37 | 2 |
| LO | 10 | 1012 | 148 | 0 | LO | 1 | 1120 | 153 | 0 | LO | 5 | 1084 | 100 | 1 |
| Normal | 13 | 100 | 1804 | 0 | Normal | 2 | 48 | 1840 | 0 | Normal | 2 | 82 | 1926 | 2 |
| VP | 2 | 2 | 53 | 270 | VP | 0 | 0 | 11 | 293 | VP | 0 | 0 | 3 | 257 |

Fig. 8. Confusion matrix for different experiments on ResNet-50 model, (a) Confusion Matrix (CM1) for ResNet-50 on original dataset, (b) Confusion Matrix (CM2) for ResNet-50+SVD-CLAHE Boosting, (c) Confusion Matrix (CM3) for Proposed methodology (ResNet-50+SVD-CLAHE Boosting+ BWCCE).

Table 5
Ablation Study of various experiments conducted on VGG-19 model on the testing set.

| Methodology | F1-score | Accuracy | Precision | Recall | AUC |
|---|-------------|-------------|-------------|-------------|--------------|
| VGG-19 with original dataset | 0.91 | 0.90 | 0.92 | 0.90 | 0.928 |
| VGG-19 with conventional data augmentation (flipping, rotation) | 0.85 | 0.86 | 0.89 | 0.82 | 0.883 |
| VGG-19 with SVD-CLAHE Boosting with equal no of images | 0.93 | 0.93 | 0.95 | 0.94 | 0.950 |
| VGG-19 with proposed SVD-CLAHE Boosting | 0.94 | 0.93 | 0.95 | 0.92 | 0.961 |
| VGG-19 with proposed SVD-CLAHE Boosting with WCCE | 0.95 | 0.94 | 0.95 | 0.94 | 0.964 |
| Proposed method (VGG-19+SVD-CLAHE Boosting+BWCCE) | 0.95 | 0.94 | 0.96 | 0.95 | 0.967 |

all classes (except little degradation in the VP class). Thus, it gives a balanced result compared to other experiments. The number of true positives in normal and Covid classes is significantly increased. Due to the number of normal class images being higher, it enables the model to achieve its best performance in terms of accuracy, precision, recall, and F1 score, which we already observed in Table 4.

In order to check the validity of the proposed framework, we have conducted the same experiments on the VGG-19 model as well. Because we wanted to check whether the proposed framework works well in a generalized way or not. An ablation study of various experiments on VGG-19 (on the testing dataset), is presented in Table 5. First, VGG-19 model is experimented with conventional data augmentation with rotation, flipping, etc. However, those data-augmentation techniques are not featured invariant; thus, it produces worse results. This data-augmentation has degraded the overall performance by 3%–8%, shown

in Table 5. The performance on the second augmented dataset, however, has improved a bit. Overall, there is a 2%–4% improvement from this augmented dataset to the original dataset. Furthermore, proposed SVD-CLAHE Boosting based on intra-class variance works very efficiently on the VGG-19 model as well. From Table 5, this can be observed that SVD-CLAHE Boosting produces an overall 2%–3% boosting of performance, over VGG 19 on the original dataset. Hence, SVD-CLAHE Boosting works in a generalized way, to the best of our knowledge. In order to reduce a little bit of class imbalance still present after SVD-CLAHE Boosting, we employed the proposed loss function BWCCE, incorporated in the VGG-19 model. Experimental results in Table 5 reveal that BWCCE provides a little bit better and more balanced results than the previous results. Table 5 shows that the proposed framework (i.e., VGG-19+ SVD-CLAHE Boosting + BWCCE) provides 1%–3% further improvement in performances than VGG-19 on the proposed augmented dataset (by SVD-CLAHE Boosting).

Moreover, this can be observed from Fig. 5 that by WCCE loss functions, the VGG-19 model provides a little bit spike (or oscillation) in all validation results (accuracy, F1 score, and loss graph). This is undesirable. However, experiments with the BWCCE loss function in Fig. 5 provides more stable results than WCCE. Overall, the BWCCE loss function provides the best accuracy, F1 score, and least loss metrics than other experiments for both models, shown in Figs. 5 and 6. This is a significant improvement. Therefore, the proposed BWCCE justified its usefulness for the VGG-19 model. Hence, the proposed framework (i.e., data-augmentation by SVD-CLAHE boosting and proposed loss function BWCCE) works very efficiently for both models.

5. Conclusion

A novel data-augmentation method SVD-CLAHE Boosting was proposed to solve a multi-class classification task for a highly imbalanced Covid-19 CXR dataset. First, the ResNet-50 model was proposed for the classification task; moreover, ResNet-50 was employed on the augmented dataset (by proposed SVD-CLAHE Boosting), which provided better results than other models. The proposed augmented dataset (of

30,033 images) was able to distinguish between different classes more efficiently, thereby generalizing more and avoiding the issue of class imbalance problem significantly. This augmented dataset was shared online publicly on the Kaggle website. The boosting performance of evaluation metrics further justified its name 'SVD-CLAHE Boosting'. In order to check the validity of the proposed data-augmentation method, same experiment was conducted on the VGG-19 model as well. The experimental result suggested that the proposed SVD-CLAHE Boosting worked well in a generalized way. However, a little bit of a class imbalance problem was still observed in the model (in their confusion matrix); thus, a novel BWCCE loss function was employed, which assigned the bias weightage of several classes based on the notion of probability. This novel loss function provided improved performance metrics for both ResNet-50 and VGG-19 models and provided a stable validation graph for loss function and accuracy compared to WCCE. Another attractive characteristic of the proposed loss function was that for complicated models (ResNet-50), it converged a little bit faster than conventional CCE and WCCE. Hence, this can be concluded that the proposed data-augmentation technique, 'SVD-CLAHE boosting' along with BWCCE loss function, worked efficiently for both the ResNet-50 and VGG-19 models for this imbalanced imbalance Covid-19 dataset. The mean of evaluation metrics indicated that the proposed framework outperformed any other methods both qualitatively and quantitatively.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] C.S.G. of the International, et al., The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, *Nat. Microbiol.* 5 (4) (2020) 536.
- [2] H.K. Koh, A.C. Geller, T.J. VanderWeele, Deaths from COVID-19, *JAMA* 325 (2) (2021) 133–134.
- [3] F. Shirani, A. Shayanfar, S. Hajiahmadi, COVID-19 pneumonia: a pictorial review of CT findings and differential diagnosis, *Egypt. J. Radiol. Nucl. Med.* 52 (1) (2021) 1–8.
- [4] X. Peng, X. Xu, Y. Li, L. Cheng, X. Zhou, B. Ren, Transmission routes of 2019-nCoV and controls in dental practice, *Int. J. Oral Sci.* 12 (1) (2020) 1–6.
- [5] I. Smyrlaki, M. Ekman, A. Lentini, N. Rufino de Sousa, N. Papanicolaou, M. Vondracek, J. Aarum, H. Safari, S. Muradrasoli, A.G. Rothfuchs, et al., Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-PCR, *Nature Commun.* 11 (1) (2020) 1–12.
- [6] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest CT for COVID-19: comparison to RT-PCR, *Radiology* 296 (2) (2020) E115–E117.
- [7] S.R. Nayak, D.R. Nayak, U. Sinha, V. Arora, R.B. Pachori, Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study, *Biomed. Signal Process. Control* 64 (2021) 102365.
- [8] A. Ulhaq, A. Khan, D. Gomes, M. Paul, Computer vision for COVID-19 control: a survey, 2020, arXiv preprint arXiv:2004.09420.
- [9] A.A. Soltan, S. Kouchaki, T. Zhu, D. Kiyasseh, T. Taylor, Z.B. Hussain, T. Peto, A.J. Brent, D.W. Eyre, D. Clifton, Artificial intelligence driven assessment of routinely collected healthcare data is an effective screening test for COVID-19 in patients presenting to hospital, *MedRxiv* (2020).
- [10] M. Buda, A. Maki, M.A. Mazurkowski, A systematic study of the class imbalance problem in convolutional neural networks, *Neural Netw.* 106 (2018) 249–259.
- [11] C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse, A. Napolitano, RUSBoost: A hybrid approach to alleviating class imbalance, *IEEE Trans. Syst. Man Cybern.* 40 (1) (2009) 185–197.
- [12] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *Int. Conf. Med. Imag. Comput. Comput.-Assisted Interv.*, Springer, 2015, pp. 234–241.
- [13] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artificial Intelligence Res.* 16 (2002) 321–357.
- [14] P. Chaudhari, H. Agarwal, V. Bhatija, Data augmentation for cancer classification in oncogenomics: an improved KNN based approach, *Evol. Intell.* 14 (2) (2021) 489–498.
- [15] K. Deepshikha, A. Naman, Removing Class Imbalance using Polarity-GAN: An Uncertainty Sampling Approach, 2020, arXiv preprint arXiv:2012.04937.
- [16] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, P.R. Pinheiro, Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection, *IEEE Access* 8 (2020) 91916–91923.
- [17] Z. Chen, J. Duan, L. Kang, G. Qiu, Class-imbalanced deep learning via a class-balanced ensemble, *IEEE Trans. Neural Netw. Learn. Syst.* (2021).
- [18] C. Drummond, R.C. Holte, et al., C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling, in: *Workshop on Learning from Imbalanced Datasets II*, 11, Citeseer, 2003, pp. 1–8.
- [19] Q. Kang, X. Chen, S. Li, M. Zhou, A noise-filtered under-sampling scheme for imbalanced classification, *IEEE Trans. Cybern.* 47 (12) (2016) 4263–4274.
- [20] Y. Freund, R.E. Schapire, et al., Experiments with a new boosting algorithm, in: *ICML*, vol. 96, Citeseer, 1996, pp. 148–156.
- [21] N.V. Chawla, A. Lazarevic, L.O. Hall, K.W. Bowyer, Smoteboost: Improving prediction of the minority class in boosting, in: *European Conference on Principles of Data Mining and Knowledge Discovery*, Springer, 2003, pp. 107–119.
- [22] J. Sun, H. Li, H. Fujita, B. Fu, W. Ai, Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting, *Inf. Fusion* 54 (2020) 128–144.
- [23] S.H. Khan, M. Hayat, M. Bennamoun, F.A. Sohel, R. Togneri, Cost-sensitive learning of deep feature representations from imbalanced data, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (8) (2017) 3573–3587.
- [24] Ö. Özdemir, E.B. Sönmez, Weighted cross-entropy for unbalanced data with application on COVID X-ray images, in: *2020 Innovations in Intelligent Systems and Applications Conference, ASYU*, IEEE, 2020, pp. 1–6.
- [25] Y. Ho, S. Wookey, The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling, *IEEE Access* 8 (2019) 4806–4813.
- [26] K. Pasupa, S. Vatathanavaro, S. Tungjitnob, Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification, *J. Ambient Intell. Humaniz. Comput.* (2020) 1–17.
- [27] R. Zhang, Q. Wang, Y. Lu, Combination of ResNet and center loss based metric learning for handwritten Chinese character recognition, in: *2017 14th IAPR International Conference on Document Analysis and Recognition, ICDAR, 5*, IEEE, 2017, pp. 25–29.
- [28] T. Wu, Q. Huang, Z. Liu, Y. Wang, D. Lin, Distribution-balanced loss for multi-label classification in long-tailed datasets, in: *European Conference on Computer Vision*, Springer, 2020, pp. 162–178.
- [29] S. Ryou, S.-G. Jeong, P. Perona, Anchor loss: Modulating loss scale based on prediction difficulty, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5992–6001.
- [30] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, S. Belongie, Class-balanced loss based on effective number of samples, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9268–9277.
- [31] S.S.M. Salehi, D. Erdogmus, A. Gholipour, Tversky loss function for image segmentation using 3D fully convolutional deep networks, in: *International Workshop on Machine Learning in Medical Imaging*, Springer, 2017, pp. 379–387.
- [32] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation, *Comput. Med. Imaging Graph.* 95 (2022) 102026.
- [33] M.E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al Emadi, et al., Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* 8 (2020) 132665–132676.
- [34] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S.B.A. Kashem, M.T. Islam, S. Al Maadeed, S.M. Zughair, M.S. Khan, et al., Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images, *Comput. Biol. Med.* 132 (2021) 104319.
- [35] M. Siddhartha, A. Santra, COVIDLite: A depth-wise separable deep neural network with white balance and CLAHE for detection of COVID-19, 2020, arXiv preprint arXiv:2006.13873.
- [36] L. Wang, Z.Q. Lin, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images, *Sci. Rep.* 10 (1) (2020) 1–12.
- [37] Y. Xu, H.-K. Lam, G. Jia, MANet: A two-stage deep learning method for classification of COVID-19 from Chest X-ray images, *Neurocomputing* 443 (2021) 96–105.
- [38] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (11) (1998) 1254–1259.
- [39] M. Toğaçar, B. Ergen, Z. Cömert, COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches, *Comput. Biol. Med.* 121 (2020) 103805.
- [40] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv preprint arXiv:1704.04861.
- [41] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, 2016, arXiv preprint arXiv:1602.07360.
- [42] M. Mamalakis, A.J. Swift, B. Vorselaars, S. Ray, S. Weeks, W. Ding, R.H. Clayton, L.S. Mackenzie, A. Banerjee, DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays, 2021, arXiv preprint arXiv:2104.04006.

- [43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [44] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
- [45] D. Das, K. Santosh, U. Pal, Truncated inception net: COVID-19 outbreak screening using chest X-rays, *Australas. Phys. Eng. Sci. Med.* 43 (3) (2020) 915–925.
- [46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
- [47] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-covid: Predicting covid-19 from chest X-ray images using deep transfer learning, *Med. Image Anal.* 65 (2020) 101794.
- [48] A.K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, J.J. Rodrigues, Identifying pneumonia in chest X-rays: a deep learning approach, *Measurement* 145 (2019) 511–518.
- [49] A.I. Khan, J.L. Shah, M.M. Bhat, CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images, *Comput. Methods Programs Biomed.* 196 (2020) 105581.
- [50] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks, *Pattern Anal. Appl.* (2021) 1–14.
- [51] T.B. Chandra, K. Verma, B.K. Singh, D. Jain, S.S. Netam, Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble, *Expert Syst. Appl.* 165 (2021) 113909.
- [52] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, W. Zhang, Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning, *IEEE Trans. Med. Imaging* 39 (8) (2020) 2584–2594.
- [53] S.A. Harmon, T.H. Sanford, S. Xu, E.B. Turkbey, H. Roth, Z. Xu, D. Yang, A. Myronenko, V. Anderson, A. Amalou, et al., Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets, *Nature Commun.* 11 (1) (2020) 1–7.
- [54] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Comput. Biol. Med.* 121 (2020) 103792.
- [55] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012).
- [56] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
- [57] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
- [58] J.P. Cohen, P. Morrison, L. Dao, K. Roth, T.Q. Duong, M. Ghassemi, Covid-19 image data collection: Prospective predictions are the future, 2020, arXiv preprint arXiv:2006.11988.
- [59] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
- [60] M.E. Wall, A. Rechtsteiner, L.M. Rocha, Singular value decomposition and principal component analysis, in: A Practical Approach to Microarray Data Analysis, Springer, 2003, pp. 91–109.
- [61] H. Demirel, C. Ozcinar, G. Anbarjafari, Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition, *IEEE Geosci. Remote Sens. Lett.* 7 (2) (2009) 333–337.
- [62] S. Roy, S. Lal, J.R. Kini, Novel color normalization method for Hematoxylin & Eosin stained histopathology images, *IEEE Access* 7 (2019) 28982–28998.
- [63] S. Roy, S. Panda, M. Jangid, Modified reinhard algorithm for color normalization of colorectal cancer histopathology images, in: 2021 29th European Signal Processing Conference, EUSIPCO, IEEE, 2021, pp. 1231–1235.
- [64] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J.B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, *Comput. Vis. Graph. Image Process.* 39 (3) (1987) 355–368.
- [65] R.C. Gonzales, R.E. Woods, *Digital image processing*, second ed., Prentice Hall, 2001.
- [66] Z. Wang, A.C. Bovik, A universal image quality index, *IEEE Signal Process. Lett.* 9 (3) (2002) 81–84.
- [67] S. Jadon, A survey of loss functions for semantic segmentation, in: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB, IEEE, 2020, pp. 1–7.
- [68] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.