



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.



Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe



Original Research Article

AutoCovNet: Unsupervised feature learning using autoencoder and feature merging for detection of COVID-19 from chest X-ray images



Nayeeb Rashid, Md Adnan Faisal Hossain, Mohammad Ali, Mumtahina Islam Sukanya, Tanvir Mahmud, Shaikh Anowarul Fattah *

Department of EEE, BUET, Dhaka, Bangladesh

ARTICLE INFO

Article history:

Received 4 January 2021

Received in revised form

16 September 2021

Accepted 18 September 2021

Available online 20 October 2021

Keywords:

COVID-19 diagnosis

Medical Image Analysis

X-ray

Neural Network

Autoencoder

ABSTRACT

With the onset of the COVID-19 pandemic, the automated diagnosis has become one of the most trending topics of research for faster mass screening. Deep learning-based approaches have been established as the most promising methods in this regard. However, the limitation of the labeled data is the main bottleneck of the data-hungry deep learning methods. In this paper, a two-stage deep CNN based scheme is proposed to detect COVID-19 from chest X-ray images for achieving optimum performance with limited training images. In the first stage, an encoder-decoder based autoencoder network is proposed, trained on chest X-ray images in an unsupervised manner, and the network learns to reconstruct the X-ray images. An encoder-merging network is proposed for the second stage that consists of different layers of the encoder model followed by a merging network. Here the encoder model is initialized with the weights learned on the first stage and the outputs from different layers of the encoder model are used effectively by being connected to a proposed merging network. An intelligent feature merging scheme is introduced in the proposed merging network. Finally, the encoder-merging network is trained for feature extraction of the X-ray images in a supervised manner and resulting features are used in the classification layers of the proposed architecture. Considering the final classification task, an EfficientNet-B4 network is utilized in both stages. An end to end training is performed for datasets containing classes: COVID-19, Normal, Bacterial Pneumonia, Viral Pneumonia. The proposed method offers very satisfactory performances compared to the state of the art methods and achieves an accuracy of 90:13% on the 4-class, 96:45% on a 3-class, and 99:39% on 2-class classification.

© 2021 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

* Corresponding author at: Department of EEE, BUET, ECE Building, West Palashi, Dhaka 1205, Bangladesh

E-mail addresses: tanvirmahmud@eee.buet.ac.bd (T. Mahmud), fattah@eee.buet.ac.bd, s.a.fattah@ieee.org (S.A. Fattah).

<https://doi.org/10.1016/j.bbe.2021.09.004>

0168-8227/© 2021 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

1. Introduction

The novel Coronavirus disease 2019 also known as COVID-19 first appeared in Wuhan, Hubei, China in December 2019 [1], and from then on it turned into a global pandemic affecting millions of lives worldwide. COVID-19 is a novel severe acute respiratory syndrome coronavirus which mostly affects the lungs in the human body [2]. Researchers have found ground-glass opacities, consolidation, and lower zone predominance [3] in chest X-rays of COVID-19 patients. Because of these features in the lung scans, it has been shown that chest X-ray can be used to detect the virus [4] in patients.

Deep learning based methods have seen significant application in chest X-ray image related tasks such as: Nodule classification [5], Tuberculosis detection [6], rib suppression [7], Pneumonia detection [8] and Lung segmentation [9].

Even though CT images can also be used for the task of detecting COVID-19, in comparison to the CT imaging technique, X-ray imaging technique is less expensive and more widely available [10]. X-ray imaging can also be used for mass testing at a faster rate and that is where machine learning technologies can truly contribute. Moreover, X-ray imaging offers the ease of interpretation for various chest related problems. As such, X-ray is used instead of CT images in this study.

One of the main challenges currently with detecting COVID-19 from chest X-ray images using deep learning is the relatively small size of labeled data available. In the deep learning literature, it has been observed that in these cases unsupervised learning can be first used to learn representations which later makes it possible for the supervised learning to converge and generalize even on a small labeled dataset. It was shown by [11] that, using a deep convolutional autoencoder for unsupervised image feature learning made it possible to detect lung nodules with only a small amount of labeled data. It was proposed by [12] that, using a multi-scale representation learning method via sparse autoencoder networks to capture the intrinsic scales in medical images leads to better performance in the classification task. In pathology detection conditional variational autoencoder was used by [13] to learn the reconstruction and encoding distribution of healthy images and the encoder part used these learned features later on for classification task.

Autoencoder-based reconstruction techniques are already being used on chest CT images for COVID-19 detection. Researchers have successfully used U-Net-based architectures [14] to segment multiple COVID-19 infection regions in chest CT images. Some studies [15,16] have shown the use of encoder networks in their system to classify COVID-19 infection from CT images. The method proposed in [17] utilizes contrastive domain invariance enhancement techniques on the output of the feature extractor to further boost their classification performance and to make the system more generalized for detecting COVID-19 in CT images.

Given the success of deep learning-based methods in chest X-ray image-related tasks, it is only natural to use it for classifying COVID-19 from chest X-ray images. A lot of research is being done in this field. COVID-Net [18], a deep convolutional neural network trained for classifying COVID-19 in chest X-

ray images on a dataset containing 3 classes (normal, pneumonia, and COVID) achieved a 93.3% accuracy across the classes. DarkCovidNet another CNN model for this task developed by [19] was trained on both 3 classes and 2 classes (COVID and Non-COVID) and attained an accuracy of 87.02% and 98.08% respectively. Another CNN model based on the Xception [20] architecture named CoroNet [21] was trained on 4 classes (normal, COVID, bacterial pneumonia and viral pneumonia), 3 classes and 2 classes and its accuracy for each of this case was 89.6%, 95% and 99%. A method of segmenting lungs from a chest X-ray image and using random patches from that segmented image to train a pre-trained ResNet-18 [22] to classify COVID-19 was proposed by [23]. Using a small dataset of 50 normal and 50 COVID-19 patients images, [24] trained an InceptionV3, ResNet-50 and Inception-ResnetV2 models and got an accuracy of 97%, 98% and 87% respectively for 2 classes. To get a very satisfactory COVID-19 image detection performance from a relatively small number of the available training dataset is still a difficult and open-ended challenge.

In this study, in order to overcome the problem of getting a very accurate trained model from the given small dataset of patient's chest X-ray images, a two-stage training scheme is developed for detecting COVID-19. Firstly an encoder-decoder based autoencoder network is designed and trained in an unsupervised manner using the X-ray images. Here, the autoencoder network learns to reconstruct the given image and due to the use of an overall optimization scheme, it is expected that the encoder part can preserve detailed information of the image in its different levels and learn relevant features for our dataset. Then the different levels of the encoder part of this autoencoder are connected to the proposed merging network to form an encoder-merging network where the encoder network part is initialized with the weights learned in the first stage. The proposed merging network is developed with unique Merging-blocks (M-blocks) that receive inputs from two different levels of the encoder and merge it in an intelligent way. These M-blocks are arranged in a tree pattern. The encoder-merging network and it is trained in a supervised manner for feature extraction of the X-ray images. The features obtained at the end of this encoder-merging network are passed through densely connected classification layers and these layers make the final prediction. The unique methodology of the proposed method is presented in Fig. 1. The proposed two stage training scheme is different from the traditional approach as instead of initializing our encoder model with random weights or transfer learning from an unrelated dataset, the encoder model first learns about the features of the dataset in the unsupervised training stage and it is later initialized with this learned weight in the second stage. The unsupervised learning using the autoencoder and initializing our encoder-merging classification network with the learned weights enables the model to converge and generalize on a small dataset of labeled chest X-ray images containing the classes: Normal, COVID-19, Bacterial Pneumonia, Viral Pneumonia. An end to end training of the classification network is performed on a balanced dataset of these classes. The addition of this unsupervised learning at the beginning and the use of our uniquely designed encoder-

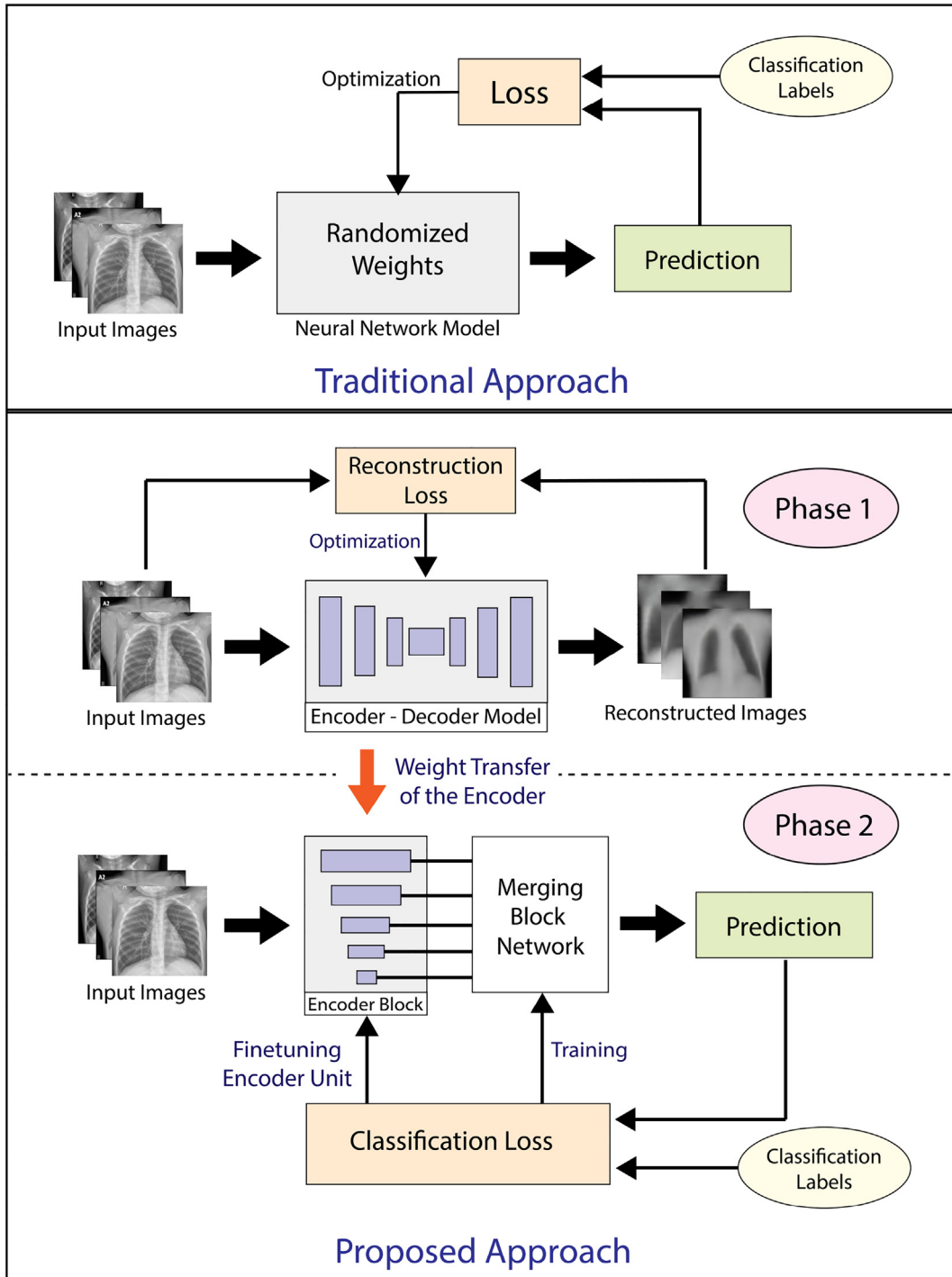


Fig. 1 - The novel approach of the proposed method is presented. In traditional approach images are passed through a randomly initialized neural network model and the model learns to classify the images, in the proposed method there are two phases of training. In the first phase an autoencoder model learns to reconstruct the input X-ray images. In the second phase the encoder portion of the autoencoder is initialized with the weights learned in phase one and connected to a proposed merging block network and this combined model is trained for the classification task.

merging classification blocks result in improved performance across all the traditional metrics.

2. Methodology

In the proposed method, for the purpose of COVID-19 image classification, both unsupervised and supervised deep neural network architectures are utilized in an effective way. The major blocks involved in the proposed scheme are shown in Fig. 2. First, a deep convolutional autoencoder network is designed to perform unsupervised feature extraction from a given chest X-ray image. Next a supervised deep CNN architecture is designed utilizing the extracted first phase features then using these features for supervised learning was performed on the classification network. The classification network is made up of uniquely designed smaller blocks that are arranged in a tree style architecture. Both these networks are trained on chest X-ray images. One major challenge in this work is to handle the classification task in case of limited number of training data, especially for the COVID-19 cases. Hence, in order to obtain a better trained model, prior to the network training stage, an efficient feature extraction stage is incorporated. In view of extracting spatial characteristics of the input image, we propose to utilize an unsupervised feature extraction stage based on the autoencoder-decoder structure. The motivation behind introducing such an additional encoder-decoder step prior to the conventional classification stage is its capability of preserving the detailed information of the given image in its different levels. Since in an autoencoder-decoder structure, a given image needs to be reconstructed at the output

stage by using an overall optimization scheme, it is expected that at the encoder stage, the spatial characteristics of the input image is precisely captured. Hence, if features are extracted from various levels of the encoder, the extracted features can precisely represent a particular class with a better inter-class separation. In the proposed scheme, in order to effectively use the extracted features from various levels of autoencoder, an efficient merging scheme with unique Merging-blocks (M-blocks) is also developed. Use of these merged features in the classification network helps in obtaining better training even with a small dataset of labeled chest X-ray images.

The basic steps of this methodology are shown in Fig. 2. Fig. 2 represents the major blocks used in the proposed method where the first block corresponds to the proposed unsupervised feature extraction stage. In this stage, an architecture with an EfficientNet-B4 model backbone is used to design an encoder-decoder model that performs optimization for each given input image and results in the decoded image. In this process, the encoder extracts different kinds of information from various perspectives which are then encoded in the encoder. These different levels of encoder containing different kinds of information are treated as useful features to be used in the next stage. In Fig. 2, the next stage represents the feature merging block where features taken from different levels of the encoder are efficiently merged by using proposed merging blocks. As a result, features collected from different levels of the encoder are merged in a single feature vector which is then finally used in the classification layer as shown in Fig. 2. These different stages of training are described in the following sections.

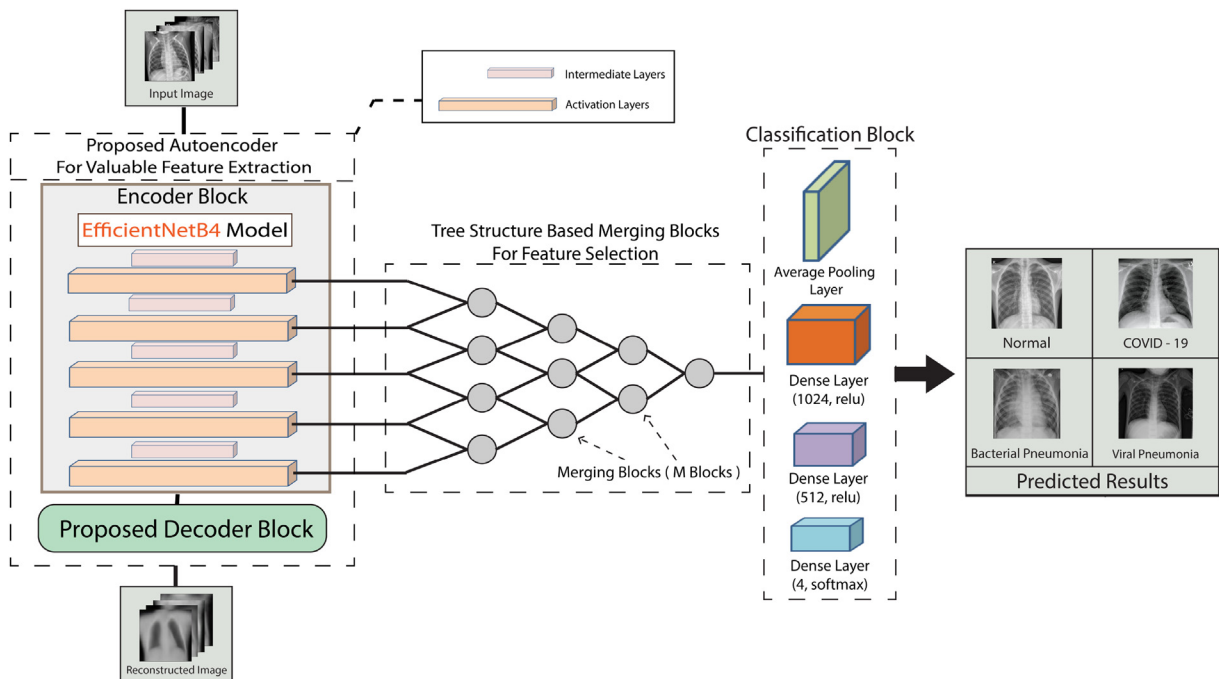


Fig. 2 – The proposed autoencoder framework consists of two-stage training: (1) training the autoencoder network using the input X-ray images and (2) training the intermediate layers of the encoder network (utilizing weights obtained in the first stage) followed by the merging network. Output from the merging network is passed to a classification network that makes the final classification.

2.1. Preprocessing

Prior to using the X-ray images in the deep neural models, a two-stage pre-processing is performed on the images: resizing and normalizing. The input images are resized to 256x256 square images containing three channels. Then a min-max normalization is applied on the resized input images. This makes the training process faster and helps the model converge more easily.

2.2. Proposed unsupervised feature learning architecture

The first phase of the system is an autoencoder that is trained on the unlabeled chest X-ray images and it learns to reconstruct the input images. Autoencoder algorithms are able to use unsupervised learning method to automatically learn features from unlabeled data [25] and they are specially useful in the medical image analysis domain where there is a scarcity of labeled data [11]. An autoencoder consists of two parts: encoder and decoder. The encoder learns the representation of a set of data to efficiently compress and encode it. And the decoder part learns to take that encoded data and reconstruct it as a representation that is as close to the original input data as possible. While selecting a model for this, it is important to note that at the first stage it will be used for an encoder-decoder based feature extraction purpose, and in the last stage, the same architecture will serve as a fundamental classification network. One of the goals is to select the classification architecture in such a way so that two separate architectures won't be needed for these two different stages which will unnecessarily increase the computational burden. This is why in that case the target was to select one classification network which could serve both purposes.

In deep convolutional autoencoders, the encoder part is made by stacking convolution layers which are then followed by pooling layers. As a result, the resolution of the input image is gradually decreased and the channels are increased. This property is similar to the conventional CNN architectures used for classification tasks. Because of this similarity, a conventional classification architecture can be used to implement the encoder block. Among different types of deep convolutional neural networks, the EfficientNet proposed in [26], carefully balances network depth, width, and resolution to obtain a better classification performance. In EfficientNet architecture, a compound scaling scheme is proposed that uniformly scales all dimensions of depth/width/resolution. Such a compound scaling offers the advantage of focusing on more relevant regions with greater object details and can enhance the classification performance by a significant margin in comparison to that obtained by single dimension scaling methods [26]. The compound scaling is defined as:

$$\begin{aligned}
 \text{depth} : d &= \alpha^\phi \\
 \text{width} : w &= \beta^\phi \\
 \text{resolution} : r &= \gamma^\phi \\
 \text{s.t.} : \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\
 \alpha \geq 1, \beta \geq 1, \gamma &\geq 1
 \end{aligned}
 \tag{1}$$

Here, ϕ is a user-specified compound coefficient that controls how many more resources are available for model scaling and α, β, γ are constants that can be determined by a

small grid search [26]. Depending on different scaling operations, there exist various versions of the EfficientNet model, namely B0 to B7. In the proposed study, various EfficientNet models are tested and finally the EfficientNet-B4 architecture is chosen because of its consistently better performance considering the size of input image and the dataset.

For this purpose, different types of available deep convolutional neural network architecture are tested and EfficientNet-B4 [26] model proved to perform the best in terms of accuracy. In the results and simulation section, this study on how the performance would have varied if another state of the art architecture, such as- InceptionV3, Resnet50, VGG11, etc, were used in the proposed scheme. At the first stage, the EfficientNet- B4 model will be used as an encoder-decoder block to get the optimum weights by utilizing different training images. Once the weights of this encoder-decoder block are optimized, these weights will be used as an initial weight at the later stage where the classification task will be performed. And the same encoder-decoder network will be trained at that time in a supervised classification manner. For this reason, using a higher performance accuracy model like the EfficientNet-B4 model reduces computation complexity since it's used as both an encoder block and later as a classification block. EfficientNet-B4 network balances between both these tasks without compromising with performance accuracy. The EfficientNet-B4 model was initialized using the pre-trained weights of Imagenet [27] because the dataset used here is relatively small to be used without Imagenet weights. The fully connected layers at the bottom of the network were omitted and output was taken from the last convolutional block so that it can be used as the encoded data for the autoencoder. For an input image size of (256, 256, 3) the encoder network produces an encoded data of shape (8,8, 1792). The next part of the autoencoder is the decoder. A decoder module was designed to reconstruct the original input image of size (256, 256, 3) from the encoded data of the shape of (8, 8, 1792). This is an opposite operation of the encoder model. A conventional CNN architecture does not perform this kind of operation and as a result a decoder is designed in the proposed scheme to reconstruct the input image from the encoded data produced by the EfficientNet-B4 encoder model. Further analytical details of the decoder can be found in [28].The decoder model consisted of 5 blocks where each block started with a transpose convolutional layer that upsampled the image by a factor of 2. This was followed by a convolutional layer that had the same number of filters as the transposed convolutional layer. The detailed architecture is presented in Fig. 3 and all the layers used in the decoder model are presented in Table 1 with their corresponding output shapes. At the end of the decoded model, there is a convolutional layer with same number of channels as the input image. This layer had a SELU activation function. SELU is the scaled exponential linear unit activation function. It is defined as:

$$\begin{aligned}
 x &= \text{scale} * x; & \text{when } x > 0 \\
 x &= \text{scale} * \alpha * (\exp(x) - 1); & \text{when } x < 0
 \end{aligned}
 \tag{2}$$

Here, scale and alpha are predefined constants with the value: alpha = 1.67326324 and scale = 1.05070098 [29].

Even though the reconstructed X-ray image from this network is not directly used, it is an important by-product of the

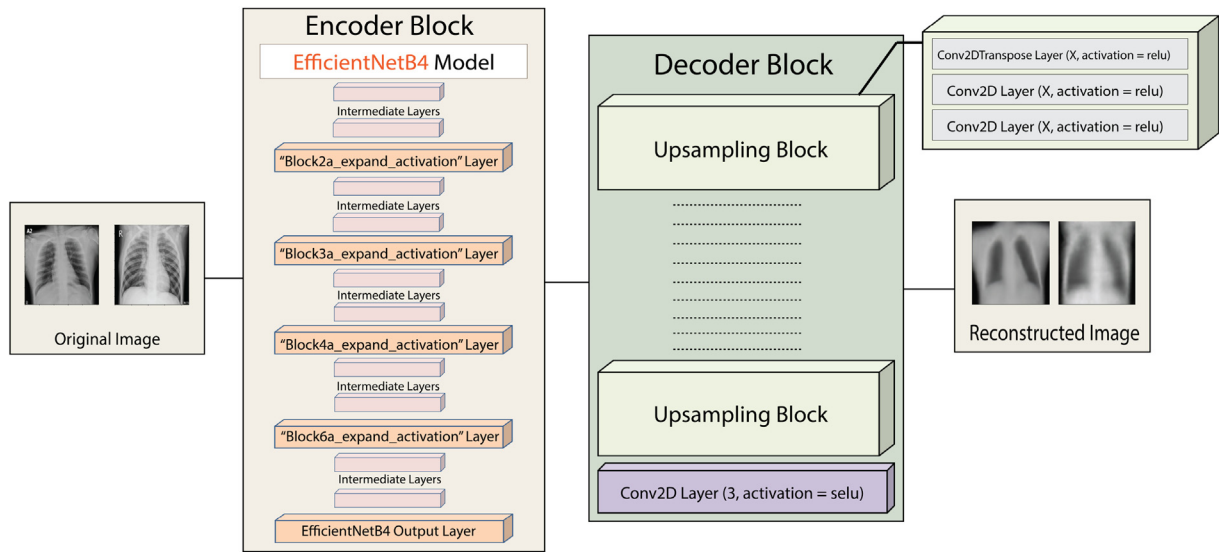


Fig. 3 – Model architecture of the proposed deep convolutional autoencoder.

Table 1 – The layers and their corresponding output shape for the proposed autoencoder model.

Encoder Feature Extraction Layers		Decoder Layers			
Layer (type)	Output Shape	Layer (type)	Output Shape	Layer (type)	Output Shape
"Block2a expand activation" Layer	(128,128,144)	1. Conv2D Transpose Layer	(16,16,512)	8. Conv2D Layer	(64,64,256)
"Block3a expand activation" Layer	(64,64,192)	2. Conv2D Layer	(16,16,512)	9. Conv2D Layer	(64,64,256)
"Block4a expand activation" Layer	(32,32,336)	3. Conv2D Layer	(16,16,512)	10. Conv2D Transpose Layer	(128,128,128)
"Block6a expand activation" Layer	(16,16,960)	4. Conv2D Transpose Layer	(32,32,256)	11. Conv2D Layer	(128,128,128)
EfficientB4 Output Layer	(8,8,1792)	5. Conv2D Layer	(32,32,256)	12. Conv2D Transpose Layer	(256,256,64)
		6. Conv2D Layer	(32,32,256)	13. Conv2D Layer	(256,256,64)
		7. Conv2D Transpose Layer	(64,64,256)	14. Conv2D Layer	(256,256,3)

proposed architecture. It will not be possible to train the encoder network on a small dataset to learn relevant features without this reconstructed X-ray image. Also, the quality of the reconstructed X-ray indicates how well the autoencoder network is converging. If the autoencoder network is trained properly that will help the encoder to preserve detailed information of the images in its different layers that can later be used for the classification task.

As the autoencoder model learns to reconstruct the input image, it does not require a label, and the entire pixel space of the input image works as labels. So even with a small amount of data and also with unlabeled data of other chest X-ray images, this network can be trained and converged. In the process of generating encoded data that is useful for reconstruction, the encoder model manages to perform information preservation from various perspectives by learning unique features of the images in the dataset and these features can then be used for the classification purpose.

2.3. Proposed classification architecture

The next part of the study was to develop a convolutional neural network architecture for the supervised learning scheme of detecting COVID-19 patients from chest X-ray images. At this stage, a classification network is required where the problems to be dealt with are 2-class, 3-class, or 4-class. For this task as mentioned before outputs from the different levels of the encoder part were extracted out from our autoencoder with the weights which were learned in the previous step. Then the features from this encoder network were used and passed through the classification network. The different parts of the classification network are specified below.

2.4. Feature extraction stage

The encoder was the EfficientNet-B4 model that had been trained in the previous step. Since this model had learned

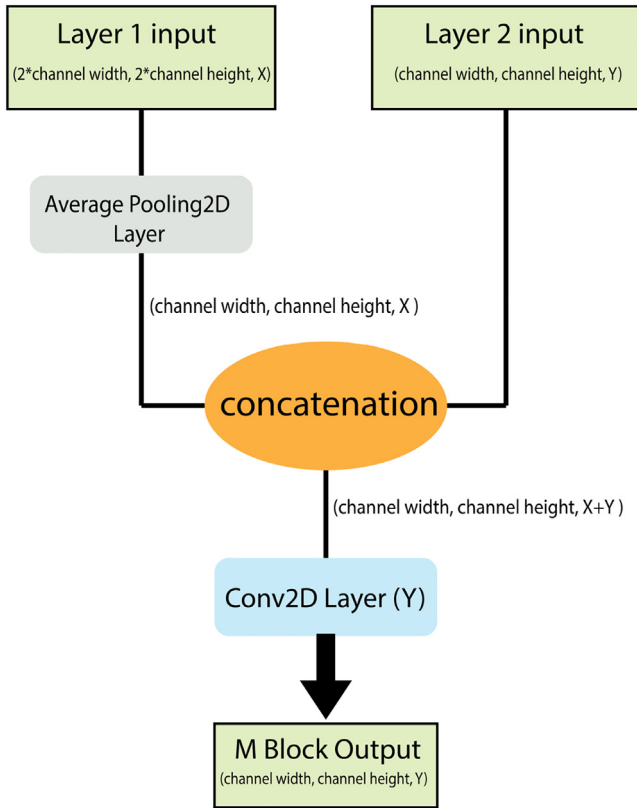


Fig. 4 – Structure of the M-blocks.

to reconstruct chest X-ray images there were valuable features in the intermediate layers of this model that could be used for the purpose of the classification task. In general, CNN models use pooling methods to decrease the input image size which is followed by convolutional operations. This is why outputs of the intermediate layers of each block were extracted after the pooling operation from the encoder model. One may extract features from different number of layers of the encoder model. However, if too many layers are chosen to converge or reduce the information obtained from different layers to a single channel, it may require more stages and if too small number of layers were taken, it may not be capable of preserving information. Hence five (5) num-

ber of layers are chosen in this work considering that to be the most suitable number.

2.5. Merging blocks (M-blocks)

For our classification network, information from different layers of the encoder is taken and reduced to a single channel so that the classification task can be performed. In that case, one major task is to reduce these five layers of information to one layer and to serve this purpose, a unique block called M-blocks is developed that merges features from two layers of the encoder in an intelligent way and later merges features from other M-blocks as well. The block takes two inputs of a 3D tensor with the first one having double the height and width of the second input. The first input is then passed through a pooling layer that uses filter size of (2,2) and averages the value for each window. The output tensor from this step has the same height and width as the second input and these two tensors are then concatenated in the channels axis. Then a convolutional operation is performed on this concatenated tensor with a window size of (1,1) and with the number of filters equal to the second inputs channel number. As a result, the output shape of each M-block is the same as the shape of the second input but it contains features from both the inputs. The structure of this block is presented in detail in Fig. 4. So this block merges features from two other layers and then learns new features on top of them with the convolutional layers. Fig. 5.

2.6. Tree structured feature merging network

The classification network is made with the combination of the encoder model and the M-blocks. The M-blocks in the classification network is connected in a tree network-style architecture as presented in Figure 5. For n number of feature extraction layers the network will have (n-1) stages with each stage having one less M-block than the previous stage. The last stage of this network will have a single M-block. In this study, five feature extraction layers were used and this network has four stages of M-blocks with the first stage taking inputs from the intermediate layers of the encoder and it has five blocks. The second stage of M-blocks takes input features from the first stage and it has four blocks. It continues in this fashion with the last stage having a single M-block with

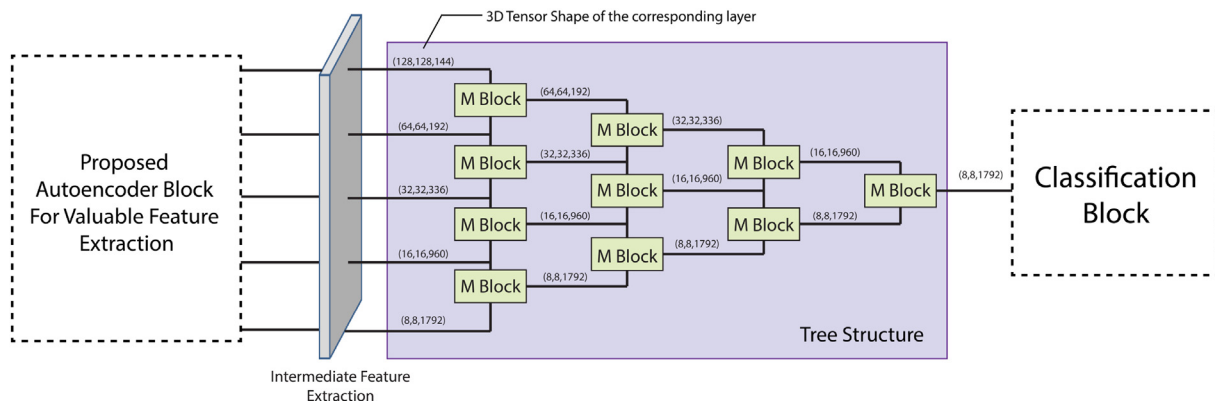


Fig. 5 – Tree Structured Feature Merging Network.

output shape of (8, 8, 1792). Then a global average pooling is performed on this tensor and it outputs a feature vector of size 1792.

These features are then passed through two fully connected dense neural networks with the first one having 1024 neurons and the second one has 512 neurons. After this, softmax activation is performed and classification is completed using that prediction.

3. Result and discussion

In this section, the performance of the proposed method is demonstrated considering different classification cases and various performance measure criteria. Results found by the proposed method are compared with that obtained by some state-of-the-art methods. In what follows, first the dataset and then the results with detailed analysis and comments are presented.

3.1. Dataset

Due to its very recent spread, there is a huge scarcity of publicly available chest X-ray images corresponding to COVID-19 patients. However, datasets for various other types of pneumonia and normal cases are available. Hence, in this research, a combination of two publicly available datasets is used to analyze the performance of the proposed method.

Pneumonia (both viral and bacterial) and normal chest X-ray images were collected from [30], an open-sourced dataset released on the Kaggle platform. The dataset contains 5,863 chest X-ray images with 4273 pneumonia images and 1590 normal images. Out of the 4273 pneumonia class images, there were 2530 images of bacterial pneumonia and 1345 images of viral pneumonia.

For COVID-19 chest X-ray images, the dataset was collected from Dr. Cohen's [31] open-source Github repository. The repository contains an open database of Covid-19 cases with chest X-ray or CT images and is being updated regularly. Chest X-ray images are largely compiled from websites such as Radiopaedia.org, the Italian Society of Medical and Interventional Radiology, and Fig. 1.com [31]. During the time this research was conducted, the repository contained 408 COVID-19 chest X-ray images.

As the dataset was unbalanced, the classes containing a higher number of images were downsampled to make the dataset balanced. Thus, the final dataset consists of 408 bacterial pneumonia, 408 viral pneumonia, 408 normal, and 408 COVID-19 chest X-ray images. After that, these images were randomly distributed into the train and test sub-folders, and five different folds of the dataset were generated for cross-validation. The training set consists of 1306 images of four different classes and the test set contains 326 images, also classified into four different classes. The train and test set images were completely independent. Also, all the images were resized to 256×256 pixels with a resolution of 96 dpi.

In this study two different views of chest X-ray images of this repository are considered in this study: (1) standard frontal PA (Posteroanterior) views and (2) standard AP (Anteroposterior) views. and AP Supine (Anteroposterior laying down)

views (the AP Supine view is avoided due to its confounding image artifacts). The X-ray images of a patient acquired from different views are found to be significantly different. The information that was available in the repository was the total number of people from whom the X-ray images were collected. At the time we conducted the research, X-ray images originate from 408 people of various hospitals across 26 different countries. Unfortunately the person's label was not available and thus we could not split the images on patient-level.

It is to be noted that while some X-rays of different views might originate from a single patient, the number is not that significant to impose a heavy data leakage and cause an over-fitting problem. In order to ensure minimum data leakage and address the over-fitting problem created by the dataset, in each block of the merging network, a Batch Normalization layer that has a regularization effect is used, and even in the feature extractor, there is a batch normalization layer in each block. Hence, in this study, the problem of over-fitting is addressed with the unique learning method proposed in the architecture and the use of the regularization layers.

It is expected that the proposed method can be found suitable for larger datasets as well where splitting the image on patient-level can be addressed. This is left as a future work depending on the availability of required larger datasets.

In Fig. 6 some samples of chest X-ray images from the prepared dataset are shown.

4. Experimental setup

In the first phase of training using the autoencoders, the encoder part was initialized with weights trained on the Imagenet dataset and the decoder part was randomly initialized. This network was optimized on the mean square error loss function with the Adam optimization algorithm. It converged with 50 epochs of training with a batch size of 32. In the second phase the encoder was initialized with the weights learned from the first phase and the M-blocks were randomly initialized. The hyperparameter values used while training the classification model were: learning rate = 0.0001, epoch = 40, batch size = 32. To avoid getting stuck on saddle points in the loss plane learning rate reduction technique is used in both the training phases. The proposed models were implemented with the Keras library using TensorFlow 2.0 backend. The entire training and testing process was performed on Google Colaboratory Server.

5. Performance evaluation

The proposed model is trained and tested on 5 fold cross-validation data containing 3 class. For each of the test set Precision, Sensitivity, F1-score and Accuracy is calculated as the performance metric and it can be seen in the Table 2.

From Table 2 it can be seen that the model got the highest accuracy of 97.97% from fold 1 and the average accuracy for all the 5 folds is 96.41%. The model had accuracy in the range of 95.47% to 97.97% for all of the folds of data. Even the lowest accuracy of 95.47% is still quite high. The same performance metrics are also generated in a class-wise basis for all of the

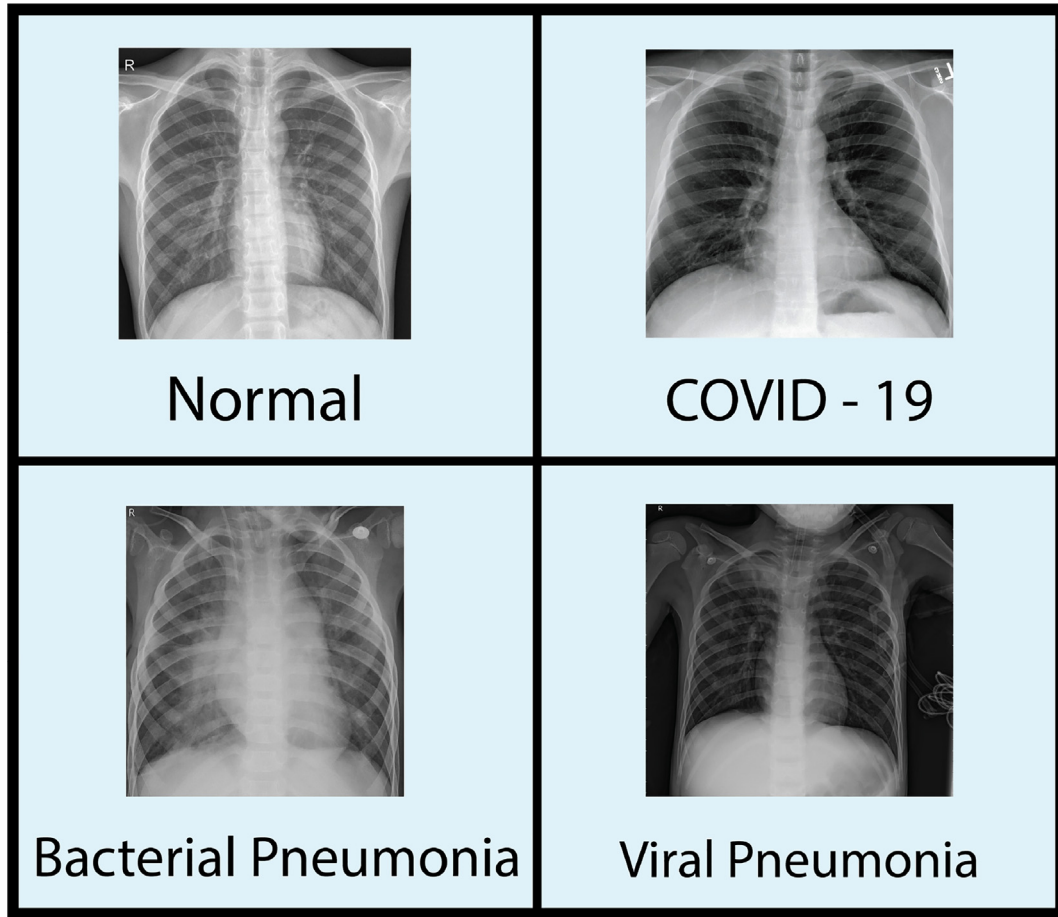


Fig. 6 – Samples of chest X-ray images from prepared dataset.

Table 2 – Precision, Recall, F1-score and Accuracy across all 3 classes for the 5 folds of data.

Folds	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Fold 1	98.05	97.97	97.96	97.97
Fold 2	95.93	95.93	95.93	95.93
Fold 3	95.57	95.53	95.52	95.53
Fold 4	97.13	97.12	97.11	97.12
Fold 5	95.58	95.47	95.47	95.47
Average	96.45	96.41	96.39	96.41

Table 3 – Precision, Sensitivity, F1-score and Accuracy of the 3 classes for Fold 1.

Class	Precision (%)	Sensitivity (%)	F1-score (%)	Accuracy (%)
COVID19	98.79	100	99.39	100
Normal	100	93.90	96.86	93.90
Pneumonia	95.35	100	97.62	100

folds. The class-wise result for fold-1 can be seen in the Table 3.

As evident from Table 3, the model performed exceptionally well in the COVID19 and Pneumonia class getting an accuracy of 100% for both of these classes. While for the Nor-

mal class it gets an accuracy of 93.90%. These claims are further supported by the confusion matrix generated for each of the folds. The confusion matrix for fold-1 and fold-2 are presented in Fig. 7. From Fig. 7 it can be observed that the model accurately predicted all the COVID-19 class images. But some

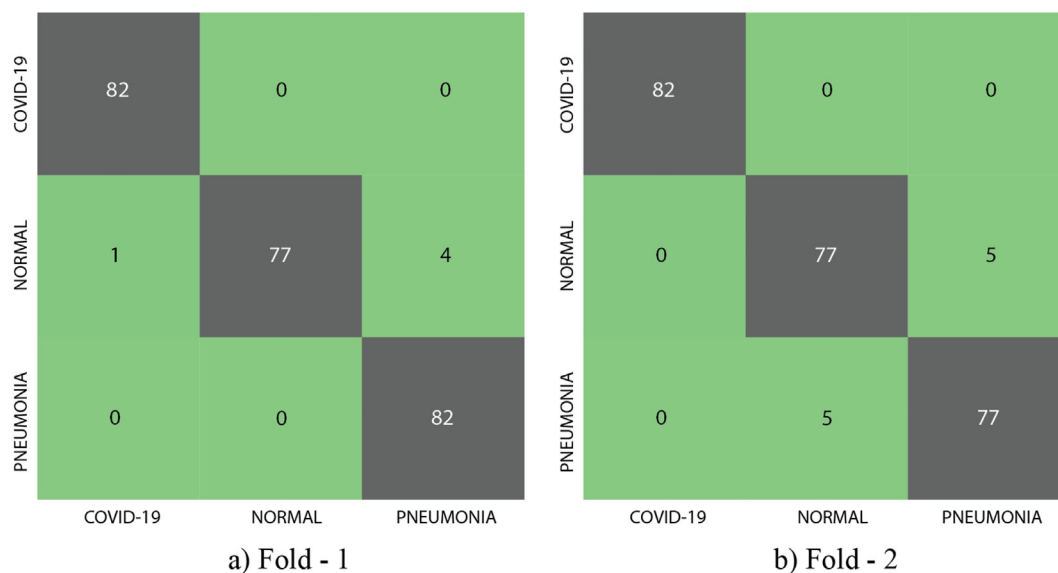


Fig. 7 – Confusion Matrix of the Test Set for 3-class Dataset.

Table 4 – Precision, Recall, F1-score and Accuracy across all 4 classes for the 5 folds of data.

Folds	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Fold 1	91.46	91.46	91.46	91.46
Fold 2	92.07	92.07	92.07	92.07
Fold 3	89.33	89.33	89.33	89.33
Fold 4	90.12	90.12	90.12	90.12
Fold 5	87.65	87.65	87.65	87.65
Average	90.13	90.13	90.13	90.13

Table 5 – Classwise result for 4-class dataset of the best performing Fold.

Class	Precision (%)	Sensitivity (%)	F1-Score (%)	Accuracy (%)
Bacterial Pneumonia	89.74	85.37	87.5	87.5
COVID19	100	100	100	100
Normal	96.25	93.9	95.06	95.06
Viral Pneumonia	84.09	90.24	87.06	87.06

of the Normal class images were classified as the Pneumonia class. While for fold-2, some of the Pneumonia class images were classified as the Normal class.

The model is also trained on a 4 class dataset to separately classify bacterial pneumonia and viral pneumonia. The same performance metrics from the 3-class setup is used in this case as well. The result for cross validation testing is presented in Table 4 and the class-wise result is presented in Table 5.

From these tables, it can be seen that the model gave consistent performance in all of the folds and from the class-wise results it can be seen that the model exceptionally well for the COVID-19 class and reasonably well for the Normal class. The performance dropped a bit when differentiating between bacterial pneumonia and viral pneumonia class. On average, for this 4 class dataset, the model achieved a classification accuracy of 90.13% for the five fold cross validated data. This experimentation was done to see if the model can generalize

for all kinds of low data irrespective of the data source and even if the data are very similar.

Even under these conditions, the model acquired an average accuracy of 90.13% which is a relatively good performance.

The model is trained on a 2-class dataset as well. This dataset was derived from the 3-class dataset where the Normal and Pneumonia classes were labeled as Non-Covid19. The evaluation metric is the same for this task as well. This detailed result is presented in Table 6.

From Table 6 it can be seen that the proposed method performed well on both the classes with an average accuracy of 99.39%. These performances on both the 4 class and 2 class dataset can be further inspected with the confusion matrices presented in Fig. 8. As can be observed from the confusion matrix of Fig. 8, that in the case of the two-class dataset almost all the test images were classified correctly except for two Non-COVID images.

Table 6 – Precision, Sensitivity, F1-score and Accuracy of the 2 classes for Fold 1.

Class	Precision (%)	Sensitivity (%)	F1-score (%)	Accuracy (%)
COVID19	97.62	100	98.8	100
Non-Covid19	100	98.78	99.39	98.78
Average	99.19	99.39	99.19	99.39

As mentioned in the methodology section, the EfficientNet-B4 model was used as the encoder network in this study. But other classification networks, such as Resnet-50, InceptionV3, and the other variants of EfficientNet were also tried as the encoder network and their results on the 4 class and 3 class datasets are compared in Table 7.

From Table 7 it can be observed that even though EfficientNet-B4 performed the best, the other models also provide similar performance which is further proof to the credibility and robustness of the proposed scheme. However, in the results section, in order to report the results in all tables, EfficientNet-B4 is used in the proposed method as the encoder network. To further justify the use of EfficientNet-B4 model as the feature extractor, the Cohen's Kappa score and the Mattheus Correlation Coefficient for the models were evaluated in the 4-class classification scheme using the proposed method. The detailed result of this analysis is presented in Table 8.

To evaluate the effectiveness of the proposed method its results were compared with a simple EfficientNet- B4 classification network pretrained on Imagenet weights. This com-

parison is presented in Table 9. From the results it can be observed that the use of the autoencoder network coupled with the merging block resulted in a performance improvement and this improvement can be specially seen in case of the four class dataset where the classification task becomes much more difficult. To further evaluate the performance of the proposed methodology statistical significance test was performed on the two methods mentioned in Table 9. McNemar's test [32] and Wilcoxon signed ranked test [33] are the two statistical tests that were performed for this purpose. The statistical significance tests are performed on the prediction of the two methods mentioned in Table 9. The prediction of each model on the 326 test set images are compared to the ground label of each of these images and a binary label with correct/incorrect decision is generated based on this comparison. There are two distributions of this binary variable for the two models and the disagreement between the two methods is used as the variable for these statistical significance tests. The test tries to see if it is possible to reject the null hypothesis which states that there is no difference in the disagreement between the two methods. The results of these tests

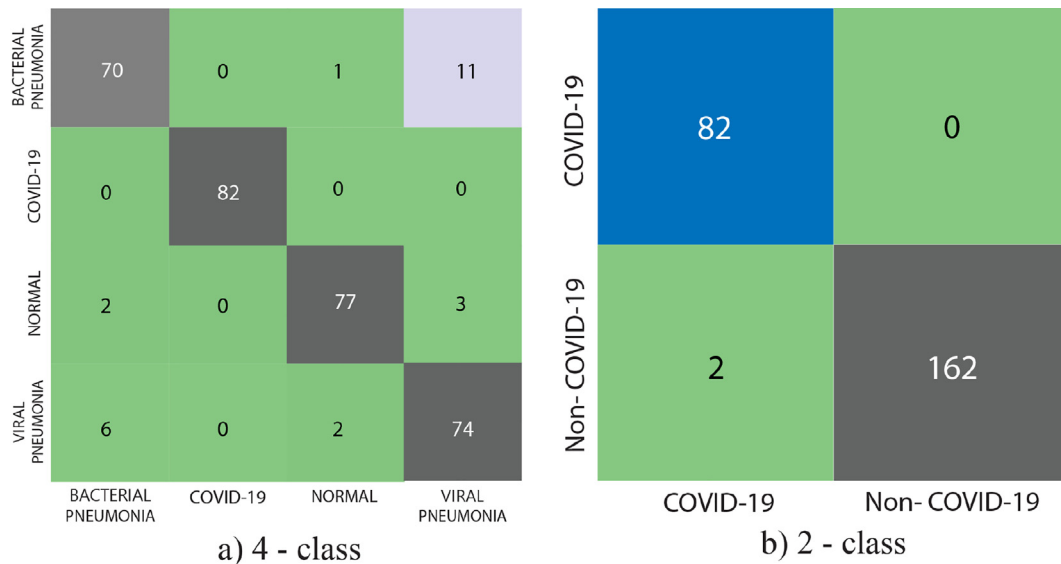


Fig. 8 – Confusion Matrices of the Test Set for 4-class and 2-class Dataset.

Table 7 – Comparisons of different models for 3-class and 4-class classification using our scheme.

Classification Type	Model Accuracy(%)						
	Efficient Net B1	Efficient Net B2	Efficient Net B3	Efficient Net B4	Inception V3	Resnet 50	Vgg-11
3-class	96.75	97.56	97.56	97.97	97.15	96.75	95.53
4-class	90.85	91.31	92.07	92.38	88.11	89.33	86.89

Table 8 – Cohen’s Kappa score and Mattheus Correlation Coefficient of different models for the 4-class classification using the proposed method.

Model	Cohen’s Kappa Score	Mattheus Correlation Coefficient
EfficientNet-B4	0.8861	0.8867
ResNet-50	0.7723	0.7727
InceptionNet-V3	0.8292	0.8321
Vgg-11	0.8252	0.8253

are presented in Table 10. It can be observed from the results that the P-value of McNemar’s test for the 4 class classification scheme was 0.83825 and for the 3 class classification scheme it was 0.68309. The P-value for the Wilcoxon signed ranked test for the 4 class classification scheme was 0.638 and for the 3 class classification scheme, it was 0.084. As for both the test in both classification schemes the P-value was very close to 0.5 it can be inferred that the proposed methodology produced some degree of statistically significant results.

As previously mentioned in section 1, a lot of research work is currently being done on classifying COVID-19 patients from chest X-ray images. These studies are being conducted on both 3-class and 2-class datasets with variation in the number of images in the dataset and the model architecture. A comparison of the proposed system with the existing literature is presented in Table 11. It is to be noted here that in these different methods reported, the number of COVID-19 dataset X-ray images are different in each case.

In order to further prove the superiority of the proposed method, a comprehensive comparative analysis has been performed with the existing literatures that have publicly available implementation of their systems, on the common evaluation protocol of: 408 Covid-19 + 408 Normal + 408 Viral Pneumonia + 408 Bacterial

Pneumonia, 408 Covid-19 + 408 Normal + 408 Pneumonia, 408 Covid-19 + 816 Non-Covid. The results of this analysis are presented in Table 12. From the analysis it can be observed that the proposed method outperforms the other methods in all three classification scheme.

6. Discussion

From the results presented in the previous section on different classification tasks and various performance metrics, the noteworthy observations are presented below.

- In 2-class, 3-class, and as well as in the 4-class setup, the model can always detect COVID-19 classes with very high accuracy.
- The model performance is relatively poor when differentiating between the two classes of bacterial and viral pneumonia, as they look almost identical even to the human eye.

Table 11 compares our proposed method with the existing literature. It can be seen here that the model manages to perform better than the other methods presented in this table. An average accuracy of 96.45% for the 3-class setup and 90.13% for the 4-class setup is found. While for the 2-class setup it is 99.39%. It can be observed that most of the other studies evaluated their system on 3-class and 2-class datasets only. Another point to note here is that more COVID19 class data is used in this study compared to the other studies mentioned. Using a VGG-19 based model [34] acquired an accuracy of 93.48% for the 3-class setup but used 224 COVID, 700 Pneumonia, and 504 normal class images. DarkCovidNet [19] used

Table 9 – Comparison between the Proposed scheme and the traditional transfer learning method.

Classification Scheme	F1-Score(%)	
	3-class	4-class
Traditional Transfer Learning with pre-trained imagenet weights	96.32	88.09
Our proposed scheme with autoencoder + M-block	96.45	90.13

Table 10 – Statistical test results between the Proposed scheme and the traditional transfer learning method.

Classification Scheme	McNemar’s test		Wilcoxon signed ranked test	
	P-Value	Chi-squared Value	P-Value	Statistics
4-class Dataset	0.83825	0.04166	0.638	157.5
3-class Dataset	0.68309	0.16666	0.084	2.500

Table 11 – Comparison of our proposed method with the existing literature.

Work	Amount of chest X-rays	Architecture	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ozturk et al [19]	125 Covid-19 + 500 Normal	DarkCovidNet	98.08	95.13	95.3
	125 Covid-19 + 500 Normal + 500 Pneumonia		87.02	85.35	92.18
Wang and Wong [18]	53 Covid-19 + 5526 Non-Covid	COVID-Net	92.4	93.33	–
Ioannis et al. [34]	224 Covid-19 + 700 Pneumonia + 504 Normal	VGG-19	93.48	92.85	98.75
Sethy and Behra [35]	25 COVID-19 + 25 Non-Covid	ReNet-50/SVM	95.38	95.33	–
Hemdan et al [36]	50 Covid-19 + 50 Non-Covid	VGG-19	90	–	–
Narin et al [24]	50 Covid-19 + 50 Non-Covid	ResNet-50	96.1	91.8	96.6
	305 Covid-19 + 305 Normal		97.4		94.7
	305 Covid-19 + 305 Viral Pneumonia		87.3		85.5
	304 Covid-19 + 305 Bacterial Pneumonia		94.7	–	93.3
Tanvir et al [37]	305 Covid-19 + 305 Viral Pneumonia + 305 Bacterial Pneumonia	CovXNet	89.6		87.6
	305 Covid-19 + 305 Normal + 305 Viral Pneumonia + 305 Bacterial Pneumonia		90.3		89.1
	284 Covid-19 + 310 Normal + 330 Bacterial Pneumonia + 327 Viral Pneumonia		89.6	–	96.4
Khan et al [21]	105 COVID-19 + 11 SARS + 80 Normal	CoroNet	97.35%	98.23%	96.34%
Abbas et al [38]	500COVID-19 + 800 Normal + 400 Pneumonia-Viral + 400 Pneumonia-bacteria	DeTraC	91.2	91.76	93.48
	500COVID-19 + 800 Normal + 800 Pneumonia-bacteria		94.2	92.76	94.56
	500COVID-19 + 800 Normal		99.12	95.36	97.36
Emtiaz et al [39]	371 COVID-19 + 1076 Normal	CoroDet	99.16	97.44	100
	371 COVID-19 + 1076 Normal + 4078 Pneumonia-bacteria		97.40	91.30	84.78
	371 COVID-19 + 1076 Normal + 4078 Pneumonia-bacteria + 4237 Pneumonia-Viral		93.42	89.18	98.92
	408 Covid-19 + 408 Normal + 408 Viral Pneumonia + 408 Bacterial Pneumonia		90.13	91.46	97.15
Proposed Method	408 Covid-19 + 408 Normal + 408 Pneumonia	AutoCovNet	96.45	95.94	97.96
	408 Covid-19 + 816 Non-Covid		99.39	99.39	100

Table 12 – Comparison of the proposed method with the existing literature on common evaluation protocol.

System Architecture	Accuracy on the Common Evaluation Protocols		
	408 Covid-19 + 408 Normal + 408 Viral Pneumonia + 408 Bacterial Pneumonia	408 Covid-19 + 408 Normal + 408 Pneumonia	408 Covid-19 + 816 Non-Covid
CoroNet[21]	82.93	95.12	98.37
DarkCovidNet[19]	89.33	95.93	98.78
Proposed Method	90.13	96.45	99.39

224 COVID, 500 Pneumonia and 500 normal class images and performed five fold cross validation that resulted in an accuracy of 87.02% in the 3-class setup. CoroNet [21] used 310 normal, 330 pneumonia-bacterial and 327 Pneumonia-viral X-ray images with four fold cross validation for their 3-class setup and got an accuracy of 89.6%. While CovxNet [37] used a balanced dataset of 305 images for each of the class. This is why it can be said that using more COVID19 data and improving the performance makes this proposed system a lot more reliable. The studies of [34,35,36,24,40] are based on popular convolutional neural network architectures like VGG-19, ResNet-50 and AlexNet. While [19] used a modified version of the DarkNet architecture and the studies of [37,21,38] have designed convolutional neural networks for this task. None of the literature mentioned here have used two stage training scheme or the use of autoencoder network.

In this study only four classes were considered but if more respiratory diseases are included in the dataset, in fact it will increase the number of classes to be handled by the deep learning network. In that case, the overall accuracy may depend on some well-known factors, such as the intra-class and inter-class feature characteristics of the members belong to the new class, and availability of the training data of the new class. In case of the proposed scheme, it is observed that even when more classes are added to the dataset, the accuracy for the COVID-19 class does not drop significantly. It achieves good generalization and offers very satisfactory COVID-19 detection performance in comparison to some existing methods even when there is more classes included in the dataset.

The proposed classification model is developed with the aim of being used in clinical conditions for detecting

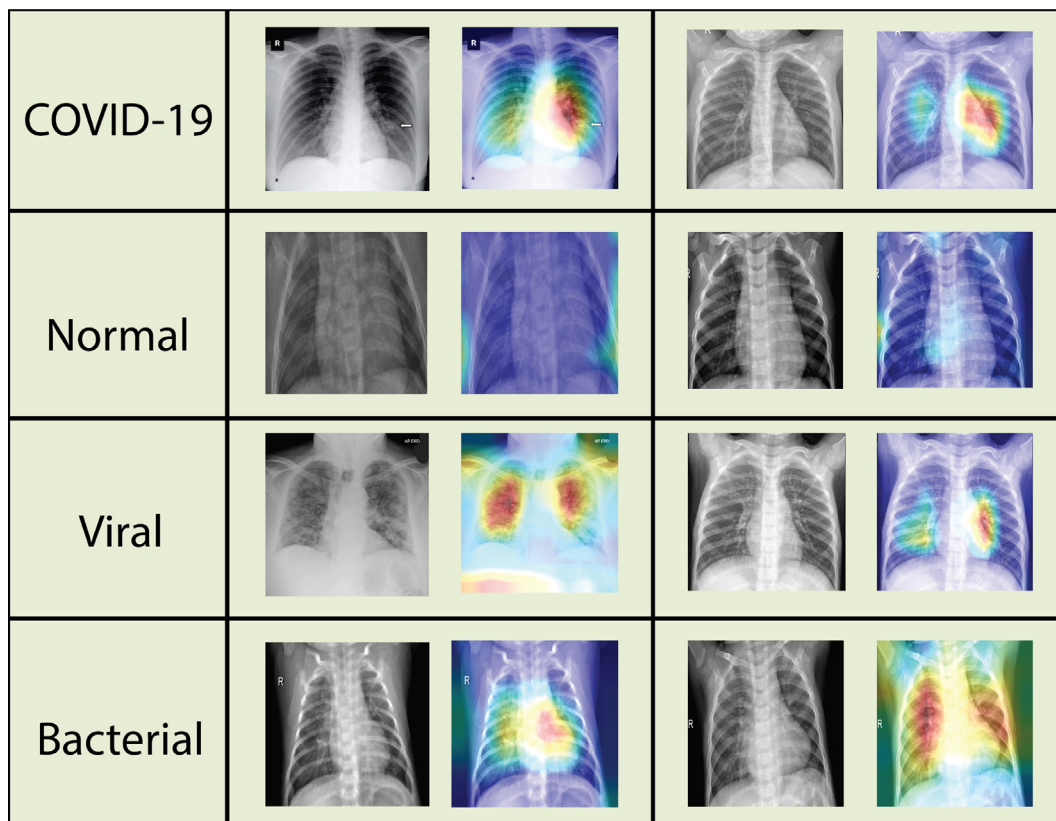


Fig. 9 – Regions of interest in each class of the dataset that is being used in the classification task.

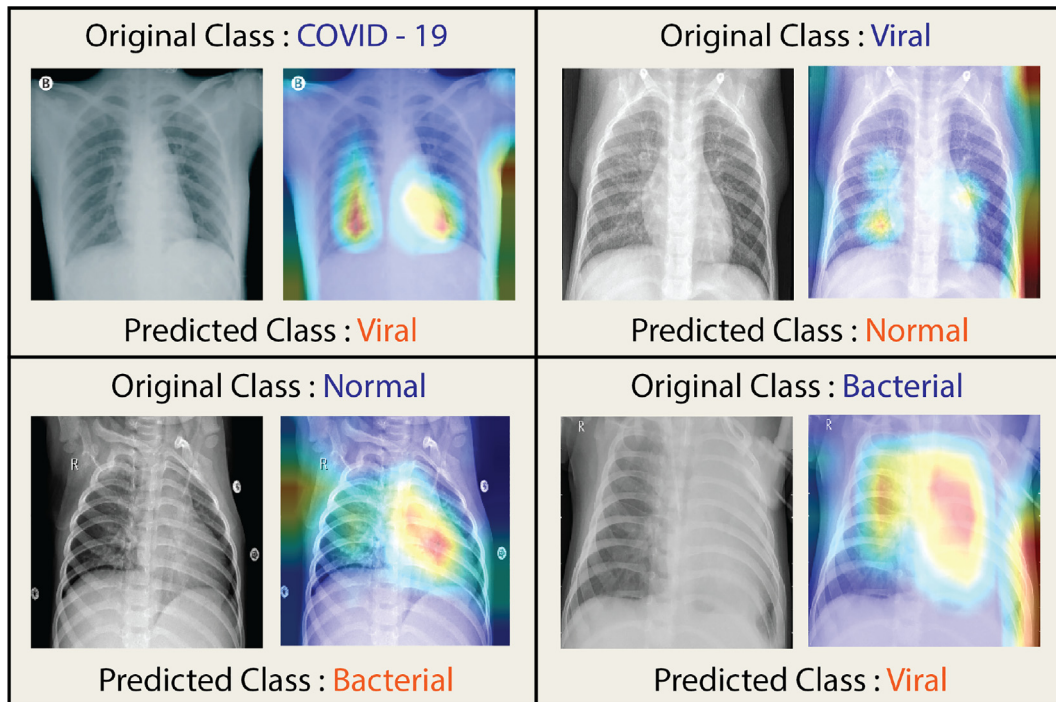


Fig. 10 – Gradient-based activation map for the misclassified instances of the test set.

COVID-19 patients from their chest X-ray images. In such a case only patients showing the known symptoms will undergo this process to verify if they have COVID-19 infection or not. As a result, the purpose of this model is to classify the COVID-19 cases from the other classes with high accuracy. Since pneumonia patients are known to show similar symptoms to that of COVID-19 patients and their chest X-ray images look very similar, the proposed system was trained to differentiate between these two classes and also recognize normal conditioned chest X-ray images similar to the state of the art literature [21,19]. However the proposed system has the potential to be used for detecting multiple respiratory diseases from chest X-ray images similar to this research [41] and this can be further explored in the future research.

It is to be noted that in recent times, several researchers have tested their methods on this dataset as it is publicly available. Apart from the competitive classification performance of the proposed method, it is expected that the users would find the intelligent methodology of this study useful in real-life applications.

7. Gradient based localization

To further inspect the results from our system, the Gradient-weighted Class Activation Mapping (Grad-CAM) [42] algorithm was integrated with our system, and the regions of interest in our X-ray images were identified that are being used for the classification purpose. The result of this integration in our test dataset is presented in Fig. 9.

It can be observed from the heatmap examples of each class that the classification model mostly looked at the

regions of lungs in the chest X-ray images. In parts of the lungs with cloudy regions indicate ground-glass opacity (GGO) and consolidation. For the bacterial and viral pneumonia and as well as the COVID-19 cases the model concentrated on the opacities present in those images with the red and yellow patches indicating the severity of abnormalities present in those regions. While for the normal class images, the heatmaps did not have any red patches in the lungs indicating they did not contain any opacities or abnormalities and as such, they were classified as normal lungs. As mentioned in [3] that, the presence of GGOs in those regions is an indication of there being pneumonia and COVID-19. In [43,44], expert radiologists found GGOs to be the predominant feature in chest X-rays of COVID-19 patients and the regions of abnormality were mostly the lower lobe and bilateral regions. Similar findings, as described above, are obtained in most of the Grad-Cam representations for pneumonia, COVID-19 and normal cases handled by the proposed method.

The misclassified instances of the test set and their corresponding Grad-Cam representations are shown in Fig. 10. It can be observed from the Grad-Cam images that the proposed model makes a wrong classification mostly when the test image of a particular class exhibits close resemblance with the images of another class.

It is to be noted that in many cases, the weights shown in the Grad-cam images cannot be well justified according to the relevant class of the sample image. There was no feedback from the Grad-Cam images that we incorporated in the proposed methodology to improve the classification performance, which could be a potential future work.

8. Conclusion

It has been more than 6 months since the advent of the Covid-19 pandemic, and an automatic detection system of COVID-19 from chest X-ray images is a necessity now. This research was conducted with the aim of developing a deep learning-based system that can generalize even on a small dataset. It is shown that the proposed training scheme utilizing an unsupervised image reconstruction stage for weight initialization of the encoder model and the proposed encoder-merging network that extracts features from different layers of the encoder network and learns to effectively merge them in a supervised training method has the capability to give some very satisfactory consistent results even with a very small dataset. It can handle both binary and multi-class problems in an efficient way. For this reason, it is expected that when a large dataset on this task becomes publicly available, this model will be able to generalize even better. Moreover, the network was designed in such a way that both feature extraction and the classification stage used the same backbone network of EfficientNet-B4. This resulted in more efficient computation and faster convergence.

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] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *JAMA* 2019;323(2020):1239–42.
- [2] Verdecchia P, Cavallini C, Spanevello A, Angeli F. The pivotal link between ace2 deficiency and sars-cov-2 infection. *Eur J Internal Med* 2020.
- [3] Ng M-Y, Lee EY, Yang J, Yang F, Li X, Wang H, et al. Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiol: Cardiothoracic Imaging* 2020;2 e200034.
- [4] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology* 2019;2020 200642.
- [5] Wang C, Elazab A, Wu J, Hu Q. Lung nodule classification using deep feature fusion in chest radiography. *Comput Med Imaging Graph* 2017;57:10–8.
- [6] H. Kim, S. Hwang, Scale-invariant feature learning using deconvolutional neural networks for weakly-supervised semantic segmentation, arXiv preprint arXiv:1602.04984 (2016).
- [7] Soleymanpour E, Pourreza HR, et al. Fully automatic lung segmentation and rib suppression methods to improve nodule detection in chest radiographs. *J Med Signals Sensors* 2011;1:191.
- [8] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T et al., Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning, arXiv preprint arXiv:1711.05225 (2017).
- [9] Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, et al. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans Med Imaging* 2014;33:577–90.
- [10] DeStigter K, Horton S, Atalabi OM, Garcia-Monaco RD, Gharbi HA, Hlabangana LT, et al. Equipment in the global radiology environment: Why we fail, how we could succeed. *J Global Radiol* 2019;5:3.
- [11] Chen M, Shi X, Zhang Y, Wu D, Guizani M. Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Trans Big Data* 2017.
- [12] Tang Q, Liu Y, Liu H. Medical image classification via multiscale representation learning. *Artif Intell Med* 2017;79:71–8.
- [13] Uzunova H, Schultz S, Handels H, Ehrhardt J. Unsupervised pathology detection in medical images using conditional variational autoencoders. *Int J Comput Assist Radiol Surg* 2019;14(3):451–61.
- [14] Chen X, Yao L, Zhang Y, Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images, arXiv preprint arXiv:2004.05645 (2020).
- [15] Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: classification and segmentation. *Comput Biol Med* 2020;126:104037.
- [16] Chen X, Yao L, Zhou T, Dong J, Zhang Y. Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images. *Pattern Recogn* 2021;113:107826.
- [17] Wang Z, Liu Q, Dou Q. Contrastive cross-site learning with redesigned net for covid-19 ct classification. *IEEE J Biomed Health Inf* 2020;24:2806–13.
- [18] Wang L, Wong A, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images, arXiv preprint arXiv:2003.09871 (2020).
- [19] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of covid-19 cases using deep neural networks with X-ray images. *Comput Biol Med* 2020;121:103792.
- [20] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 1251–8.
- [21] Khan AI, Shah JL, Bhat MM. Coronet: A deep neural network for detection and diagnosis of covid-19 from chest X-ray images. *Comput Methods Programs Biomed* 2020;196:105581.
- [22] 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*. p. 770–8.
- [23] Oh Y, Park S, Ye JC. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Trans Med Imaging* 2020.
- [24] 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).
- [25] Ng A et al. Sparse autoencoder. *CS294A Lecture Notes* 2011;72:1–19.
- [26] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.
- [27] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L, et al. *IEEE conference on computer vision and pattern recognition*. Ieee 2009;2009:248–55.
- [28] Lucas A, Iliadis M, Molina R, Katsaggelos AK. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Process Mag* 2018;35:20–36.
- [29] Pedamonti D, Comparison of non-linear activation functions for deep neural networks on mnist classification task, arXiv preprint arXiv:1804.02763 (2018).

- [30] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018;172:1122–1131.e9.
- [31] Cohen JP, Morrison P, Dao L, Covid-19 image data collection, arXiv preprint arXiv:2003.11597 (2020).
- [32] Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput* 1998;10:1895–923.
- [33] Lacoste A, Laviolette F, Marchand M, Bayesian comparison of machine learning algorithms on single and multiple datasets, in: *Artificial Intelligence and Statistics*, PMLR, 2012, pp. 665–675.
- [34] Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 2020;1.
- [35] Sethy PK, Behera SK. Detection of coronavirus disease (covid-19) based on deep features. *Preprints* 2020;2020030300:2020.
- [36] Hemdan EE-D, Shouman MA, Karar ME, Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in X-ray images, arXiv preprint arXiv:2003.11055 (2020).
- [37] Mahmud T, Rahman MA, Fattah SA. Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Comput Biol Med* 2020;122:103869.
- [38] Abbas A, Abdelsamea MM, Gaber MM. Classification of covid-19 in chest X-ray images using detrac deep convolutional neural network. *Appl Intel* 2021;51:854–64.
- [39] Hussain E, Hasan M, Rahman MA, Lee I, Tamanna T, Parvez MZ. Corodet: A deep learning based classification for covid-19 detection using chest X-ray images. *Chaos Solitons Fractals* 2021;142 110495.
- [40] Ibrahim AU, Ozsoz M, Serte S, Al-Turjman F, Yakoi PS. Pneumonia classification using deep learning from chest X-ray images during covid-19. *Cognitive Comput* 2021;1–13.
- [41] Pham HH, Le TT, Tran DQ, Ngo DT, Nguyen HQ. Interpreting chest X-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing* 2021;437:186–94.
- [42] Selvaraju RR, 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*. p. 618–26.
- [43] Weinstock MB, Echenique A, Russell J, Leib A, Miller J, Cohen D, et al. Chest X-ray findings in 636 ambulatory patients with covid-19 presenting to an urgent care center: a normal chest X-ray is no guarantee. *J Urgent Care Med* 2020;14:13–8.
- [44] Wong HYF, Lam HYS, Fong A-T, Leung ST, Chin T-Y, Lo CSY, et al. Frequency and distribution of chest radiographic findings in patients positive for covid-19. *Radiology* 2020;296(2):E72–8.