Noname manuscript No.
(will be inserted by the editor)

AI-based Diagnosis of COVID-19 Patients Using X-ray Scans with Stochastic Ensemble of CNNs

Ridhi Arora* · Vipul Bansal* · Himanshu Buckchash* · Rahul Kumar* · Vinodh J Sahayasheela · Narayanan Narayanan* · Ganesh N Pandian · Balasubramanian Raman

*equal contribution.
Corresponding Authors: Ganesh N Pandian (ganesh@kuchem.kyoto-u.ac.jp) and Rahul Kumar (rkumar9@cs.iitr.ac.in).

Ridhi Arora
Department of Computer Science & Engineering, Indian Institute of Technology Roorkee, India
E-mail: rarora@cs.iitr.ac.in

Vipul Bansal
Department of Mechanical & Industrial Engineering, Indian Institute of Technology Roorkee, India
E-mail: vbansal@me.iitr.ac.in

Himanshu Buckchash, Rahul Kumar
Department of Computer Science & Engineering, Indian Institute of Technology Roorkee, India
E-mail: {hbuckchash, rkumar9}@cs.iitr.ac.in

Vinodh J Sahayasheela
Institute of Integrated Cell Material Sciences (WPI-iCeMS), Kyoto University of Advanced Study, Kyoto, Japan
E-mail: vinodh@chemb.kuchem.kyoto-u.ac.jp

Narayanan Narayanan
Centre for Research and Graduate Studies, University of CyberJaya, Malaysia
E-mail: nn@annalakshmi.net

Ganesh N Pandian
Institute of Integrated Cell Material Sciences (WPI-iCeMS), Kyoto University of Advanced Study, Kyoto, Japan
Centre for Research and Graduate Studies, University of CyberJaya, Malaysia
E-mail: ganesh@kuchem.kyoto-u.ac.jp

Balasubramanian Raman
Department of Computer Science & Engineering, Indian Institute of Technology Roorkee, India
Centre for Research and Graduate Studies, University of CyberJaya, Malaysia
E-mail: bala@cs.iitr.ac.in

Abstract According to WHO, COVID-19 is an infectious disease and has a significant social and economic impact. The main challenge in fighting against this disease is its scale. Due to the imminent outbreak, the medical facilities are over-exhausted and unable to accommodate the piling cases. A quick diagnosis system is required to address these challenges. To this end, a stochastic deep learning model is proposed. The main idea is to constrain the deep-representations over a gaussian prior to reinforce the discriminability in feature space. The model can work on chest X-ray or CT-scan images. It provides a fast diagnosis of COVID-19 and can scale seamlessly. This work presents a comprehensive evaluation of previously proposed approaches for X-ray based disease diagnosis. Our approach works by learning a latent space over X-ray image distribution from the ensemble of state-of-the-art convolutional-nets, and then linearly regressing the predictions from an ensemble of classifiers which take the latent vector as input. We experimented with publicly available datasets having three classes – COVID-19, normal, Pneumonia. Moreover, for robust evaluation, experiments were performed on a large chest X-ray dataset with five different very similar diseases. Extensive empirical evaluation shows how the proposed approach advances the state-of-the-art.

Keywords:COVID-19, X-ray images, Feature Extraction, Deep Learning, Machine Learning

1 Introduction

Gerry, a former United Nations employee, unexpectedly finds himself in a race against time as he investigates a threatening virus that turns humans into zombies. The situation that world is facing today resembles that of fictional movie though today the coronavirus is wreaking havoc. Novel corona virus (SARS-COV-2) has brought the whole world to its knees. Coronavirus officially known as the Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) causes an infectious disease that has spread to over 215 countries [1]. The deadly disease has no cure till now and is contagious which makes it even a bigger threat to the human beings. It has crossed the figure of 7 million of confirmed cases and took more than $402,793$ lives [2]. As per the statistics by WHO [3] and presented in Figure 1, an exponential increase in the number of cases has been reported [4]. People with weak immune systems and already suffering from other diseases like Sugar, Blood pressure, etc. are more prone to the infection [5]. The virus is known to be at the most contagious stage during the first 3 days and the symptoms take typically 5 days to appear but it can vary from 2 to 14 days, hence making this as the biggest problem of the situation [6]. The virus is most contagious at the time when the symptoms have not yet appeared. The person feels cough, fever, and pain in throat and in severe cases difficulty in breathing. It spreads through close contact or the droplets when the infected coughs, sneezes or talks [7]. For a vaccine to develop for COVID-19, diagnosis through real-time reverse transcription-polymerase chain reaction (RT-PCR) is confirmed [8]. However, RT-PCR still

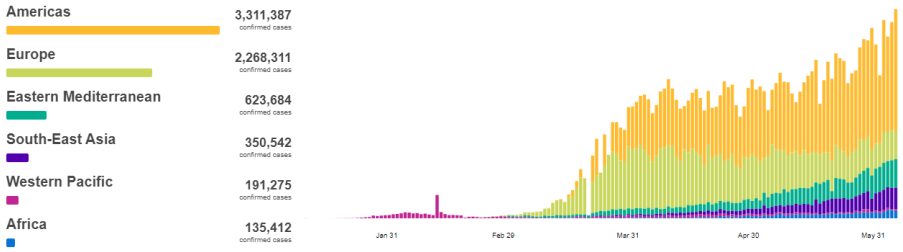


Fig. 1: Statistics as presented by World Health Organization (WHO) for the COVID-19 cases across globe (Last Updated: June 8, 2020, 8:36am CEST.) [3]

lacks through the issue of time management and there still lies a long chain of trials from animal to humans and moreover the virus is mutating regularly. Due to this, it may arise the condition that the infected person may not be recognized and may not get the suitable treatment and may spread the virus to healthy population. This cannot be acceptable in this pandemic situation.

The pandemic has parted many families apart, slayed the world economy, aggravated global poverty, and molested the mental peace of all of us. Various governments are devising different strategies to win the battle against this contagion and save their people. Even after around 60 days of complete lockdown in India, we are still under partial lockdown to implement social distancing. Every nation's government is trying it's best to fight against this virus by using different strategies. Almost all the countries are in lockdown *i.e.* the population is under self-quarantine as the only weapon we have in this battle is the social distancing [9]. The governments are trying to maintain a smooth supply of the essential items and maintain law and order in the countries. Doctors are working day and night to make and test the vaccines. The estimated time to finalize a trusted vaccine is still 14 to 18 months which leads us to how the world will handle the situation till then?

Research expertise in diagnosing chest X-rays and computed tomography (CT) images have come to the front and handled this adverse situation [10]. Because of the low sensitivity of the RT-PCR test, even with the negative results, the symptoms can be elevated with the radiological screening of the images [11; 12]. It has been observed that X-rays and CT images are sensitive to the screening of the COVID-19 among the patients. A chest X-ray is performed in suspected or confirmed patients through specific circuits. It is a discriminating element wherein the patient's further diagnosis depends on the clinical situation and the chest X-ray film results. Many recent works depending on the image features with clinical results may help in early detection of COVID-19 [13; 14; 15].

Technological advancement using the applications of deep learning and machine learning for automatic prediction of COVID-19 traces in people have taken over the minds of people [16; 17; 18; 19]. Inspired by the dire need to develop means to fight COVID-19, and intrigued by the works of open

access support of the research community, this work leverages the variational aspects of encoding the radiological image features into classification scores for predicting the COVID-19 related lung opacities. Here, a deep learning based stacked pipeline incorporating deep CNNs is encouraged. DenseNet and GoogLeNet in particular, are used for feature extraction with the involvement of a variational environment to further extract the latent space of the extracted features. This model is trained on the publically available COVID-19 datasets for performing diagnostic tests. The main contributions of this paper are as follows:

- We propose a framework for COVID-19 detection based on X-ray images, which outperforms recent state-of-the-art CNN models.
- We further propose a two-step ensemble approach under a variational setting for learning the joint representation of representations generated by fine tuned CNNs.
- We perform extensive experiments on COVID-19 dataset along with a challenging multi-class benchmark on X-ray images.

The paper is organized as follows: Section 2 discusses the current diagnostic methods in the interest of COVID-19 pandemic with the dataset specification in Section 3. The proposed model pipeline is explained in detail in Section 4. Section 5 describe the experimentation involving the implementation details, the evaluation criteria with in-depth results, and analysis. Finally, conclusions and future aspects are drawn in Section 6.

2 Related Work

This situation has undoubtedly marked its place in the history of the world that a virus has made even the big countries to bow down to knees. Considering the damage that the virus has brought to the world, many researchers and scientists are trying to put their best foot forward by implementing and testing their knowledge against all odds. Therefore global health systems are dealing with the crisis with the available technology and resources. However we, the scientific community have tried their hands on various technological advancements in the area of machine learning (ML) and deep learning (DL) algorithms. Recently, several machine and deep learning methods have been used to detect and classify the disease with radiologist medical image dataset. Various type of medical images have been widely used for COVID-19 classification and detection. Hemdan *et al.* [20] used deep learning models over X-ray images to detect and diagnose COVID-19 and proposed a COVIDX-Net model consisting of several CNN models [17; 18; 19]. Li *et al.* have used CNNs over Chest CT images to classify COVID-19 and healthy patients.

Wang *et al.* [21] has presented a classification technique for COVID-19 prediction by using the 453 chest CT images and achieved an accuracy of 82.9%. In [22], the author has presented a multi-class classification using a modified

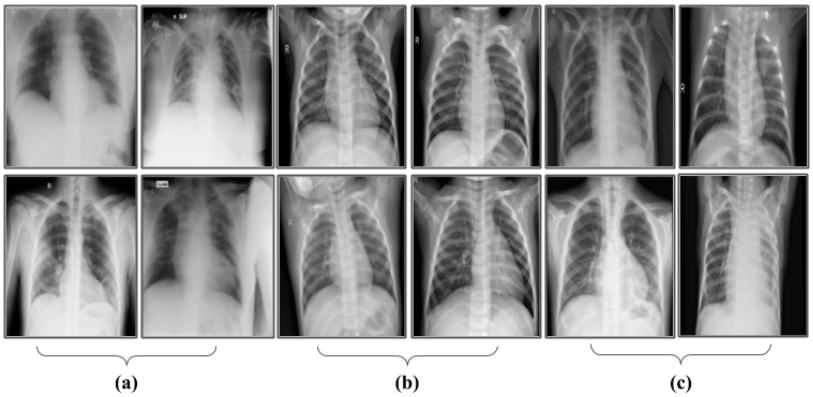


Fig. 2: Sample images from $\mathcal{D}_1$ and $\mathcal{D}_2$ dataset used for the experimentation purpose in the course of this study; (a) COVID-19 images, (b) Normal images, and (c) Pneumonia images.

version of pre-trained ResNet-50 named as DRE-Net over CT images (100 patients with bacterial pneumonia, 86 non-COVID patients and 88 patients with COVID-19 infected) to distinguish among bacterial pneumonia, non-COVID, and COVID-19 viral infection. In [23], Sethy *et al.* presented a COVID viral infection disease detection using the chest X-ray images. First of all, by using the CNNs, the authors extracted the deep features using the pre-trained ImageNet, and SVM [24] were used at the last layer to classify it. Santosh [16] has introduced a modern AI-driven framework to identify COVID-19 outbreaks better and forecast the scale of their worldwide spread by utilizing active learning-based cross-validation models. In another study, Wang *et al.* [25] implemented a CNN based COVID-Net, a multi-class classification technique to distinguish COVID-19 patients with non-COVID as healthy people and bacterial infected patients by using $16,756$ chest x-ray images that were scanned from radiography X-ray with 13645 subjects. Das *et al.* [26] introduced Truncated Inception Net to classify positive COVID-19 chest X-ray images w.r.t pneumonia and stable subject. They reported 99.96% accuracy with 100% AUC whereas an accuracy of 99.92% with 99% AUC to distinguish between positive COVID-19 v/s other subjects (mixed pneumonia, tuberculosis and stable chest X-ray CXRs). In another study [27], Mukharjee *et al* has represented a shallow CNN for classification of COVID-19 vs non-COIVD-19 using chest X-ray images and reported an $\langle Ac, Sen,\ and\ AUC\rangle \equiv \langle 96.92\%, 94.2\%, and\ 98.69\%\rangle$.

Gozes *et al.* [28] designed a CT analysis method for the identification and quantitative analysis of COVID-19 based on AI. Rahmatizadeh *et al.* [29] introduced a three-step AI-based model to improve the critical care of patients is introduced. Application in intensive care unit (ICU) is performed where the input evidence include surgical, paraclinical, customized medicine (OMICS) and epidemiology data. The performance included assessment, therapy; the stratification of threats, prognosis, and direction, and finally, it concludes that

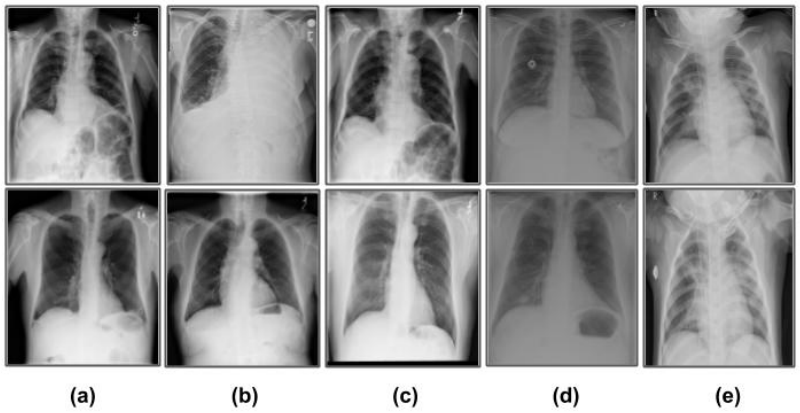


Fig. 3: Sample images from $\mathcal{D}_3$ dataset used for the experimentation purpose in the course of this study; (a) Atelectasis (b) Effusion, (c) Infiltration, (d) Nodule and (e) Pneumonia.

the health care system's efforts to overcome the problem of COVID-19 detection in patients with the help of today's AI-based technology can easily and effectively tackle the patients, particularly for the patients in ICU suffering from COVID-19 [30].

3 Dataset Specification

The 2019 novel coronavirus has shown numerous unique symptoms to detect the viral disease COVID-19 from the patients. It has been deduced that the COVID-19 patterns can be perceived from either CT or chest X-ray images. Considering the seriousness of the situation worldwide, the publically available datasets have been collected and worked upon.

Initially, we have taken Chest X-ray images (Pneumonia) [31], which is represented as $\mathcal{D}_2$, containing a total of 5840 images with 5216 for training (1341 for Normal class and 3875 for Pneumonia class) and remaining 624 (234 for Normal class and 390 for Pneumonia class) for testing.

Another dataset $\mathcal{D}_1$ for COVID-19 images is collected which has been taken from 3 COVID-19 datasets which includes COVID-19 Radiography Database [32] containing of approximately 219 images with another dataset from [33] containing a total of 69 images with 60 in training and 9 in testing. The third dataset has been taken from an open source Github repository [34] which contained a total of 35 COVID-19 images. All the datasets are combined together to form a collection of 308 images which are further used for the experimentation purpose on the proposed framework. In all, the 308 images have been segregated as 196 for training and the rest 112 for testing.

Table 1: DATASET-1 ($\mathcal{D}_1$): Number of images for each class in COVID-19 dataset.

Dataset	**Classes**			
Bifurcation	**COVID-19**	**Normal**	**Pneumonia**	**Total**
Train set	196	210	210	616
Test set	112	113	113	338
Total	308	323	323	954

Table 2: DATASET-2 ($\mathcal{D}_2$): Number of images for each class in Pneumonia dataset.

Dataset	**Classes**		
Bifurcation	**Normal**	**Pneumonia**	**Total**
Train set	1341	3875	4216
Test set	234	390	624
Total	1575	4265	5840

Table 3: DATASET-3 ($\mathcal{D}_3$): Number of images for each class in NIH Chest X-ray dataset.

Dataset	**Classes**					
Bifurcation	**Atelectasis**	**Effusion**	**Infiltration**	**Nodule**	**Pneumonia**	**Total**
Train set	3414	2788	7327	2248	3875	19625
Test set	801	1167	2220	457	390	5035
Total	4215	3955	9547	2705	4265	24660

We have comprehensively performed experiments on the NIH Chest X-ray image (dataset $\mathcal{D}_3$) [1] containing 14 categories, out of which we have selected 5 categories namely, Atelectasis, Effusion, Infiltration, Nodule, and Pneumonia which have higher correlation among them (tabulated in Table 3). Considering the dataset specifications, we found that the Pneumonia subset of images was limited in number which easily got mixed with other classes (upon initial experimental trials). Therefore, to remove this apprehension, we amalgamated the Pneumonia images from $\mathcal{D}_2$ dataset in the $\mathcal{D}_3$ dataset. The snag about the dataset is that the images have been extracted based on the query utilizing the natural language processing (NLP) which has confused the reader with

[1] https://nihcc.app.box.com/v/ChestXray-NIHCC/file/220660789610

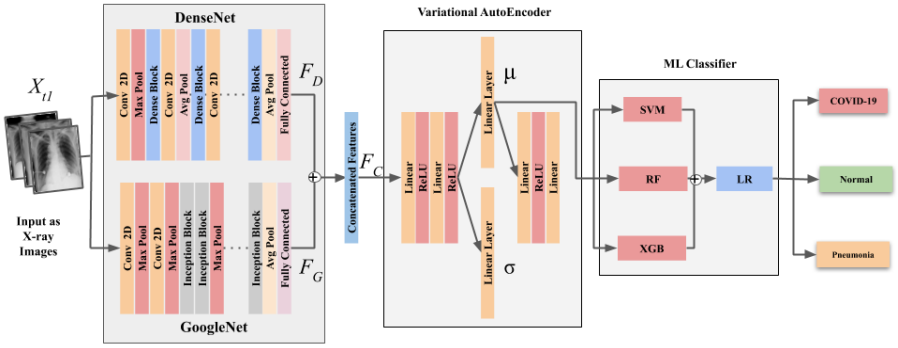


Fig. 4: Block diagram of the proposed model framework depicting different modules for the effective classification of COVID-19 v/s Normal v/s Pneumonia.

multiple classes for a single image. Therefore, to avoid confusion we have taken the diseases with minimal overlapping and maximum images in the concerned image category. Considering the consequences in these complex times, there are no clinical implications associated with the acquired datasets.

Some of the sample images from the COVID-19, Normal and Pneumonia and the NIH Chest X-ray image dataset are shown in Figure 2 with the images from the NIH chest X-ray dataset are shown in Figure 3.

4 Proposed Methodology

In this section, we propose the detailed involvement of deep learning modules for estimating the feasibility of using the chest X-rays to diagnose COVID-19. Here, the concept of the stacked architecture of DenseNet and GoogleNet along with the settings of variational autoencoder is proposed. The deep modules are trained on $\mathcal{D}_2$ dataset which includes the $\{X_t, Y_t\}$ set of images and their labels for Normal and Pneumonia patients. It contains a total of 4265 Pneumonia with 1575 Normal images.

DenseNet [35], in its basic architecture is a deep CNN that helps in eradicating the issue of vanishing-gradient efficiently depicts cross-connection and allows the use of previously extracted features by the network. This advantage is captured to achieve better performance with lesser computation. For the binary class classification of the network, two fully connected (FC) layers with 1024 hidden features are added which resulted in a final output of two classes. To fully utilize the network gradients, ReLU activation function is used in between the two FC layers which has the advantage of not activating all the neurons at once and enhancing the network for elevating the performance. At the approaching end of the model, *logsoftmax* function instead of the *softmax* function is encouraged to provides better numerical computation and gradient optimization.

For the DenseNet neural network, the key task of the convolution layer is to extract the features from the image, preventing the error of manually extracting features. It consists of a series of feature maps. A convolution kernel holds these feature maps and consists of weights as parameters, and is often called convolution filters. This convolution kernel consists of learnable parameters and is transformed with feature maps of the layers lying before joined using cross-connections allowing information flow, and the resulting elements are stored followed by offset, and finally passing it through a nonlinear ReLU activation to acquire an extracted feature map. A multiplicity of different convolution kernels allow the extraction of more complex features. The formula for calculation shall be as shown in Equation 1.

$$x_{\beta}^{L} = g\left\{\sum_{\alpha}^{Z^{L-1}} x_{\alpha}^{L-1} k_{\alpha\beta}^{L-1} + \gamma_{\beta}^{L}\right\} \tag{1}$$

Whereas $g(\cdot)$ denotes the activation function with non-linearity. Z^{L-1} denotes the feature maps input to the convolution kernel selected by the $L-1$ layer, and denotes the convolution operation performed using kernel $k_{\alpha\beta}$, γ denotes the offset added to extracted kernel features. L denotes the number of layers in our convolution neural network, and $k_{\alpha\beta}$ denotes the convolution kernel used over input x_{α}^{L-1} of the feature map β between the L th layer and the feature map α of the $(L-1)^{th}$ layer.

After training, the output layer, last FC layer, and ReLU activation function are removed from the *DenseNet* model and the remaining part is used as a feature extractor. The model $DenseNet_{1024}$ gives 1024 features as an output.

GoogleNet [36], is a state of the art deep learning architecture that allows better and more computationally efficient results. This network consists of inception modules where every penultimate layer and the next layer is joined with four connections. First, a 1×1 convolution is performed on the output of the previous layer. This is done because the output of the previous layer consists of some useful extracted features which further need to be used, then a 3×3 and a 5×5 convolutions are performed on output from the previous layer.

For dimensionality reduction, a 1×1 filter is used before performing a 3×3 and 5×5 convolution. Also, a 3×3 max-pooling layer is followed by a 1×1 filter convolution which helps in achieving an optimal sparse structure. The model with pre-trained weights is trained on ImageNet [37] dataset, followed by replacing the last FC layer of the network with the addition of two FC layers having 1024 hidden features and a final output of class probabilities. The activation function is amalgamated between the added layers. The output layer of the network is of the *logsoftmax* function with the 'Adam' optimizer to finally train the network.

The individually trained DenseNet and GoogLeNet are now fused together to form an ensemble acting as a feature extractor. The X-ray images of COVID-19, Normal, and Pneumonia are taken from the dataset D_1 which consists of

images and their corresponding labels as $\{X_c, Y_c\}$, respectively. The images are further divided into train and test sets where every image is resized to a dimension of 224×224.

These images are then given as input to the trained model of GoogLeNet and DenseNet extracting 1024 features, represented as $GoogleNet_{1024}$ and $DenseNet_{1024}$,respectively, at the output, giving us $G(c_i)$ and $D(c_i)$ which contains the features F_G and F_D, respectively as depicted in Equation 2.

$$E(c_i) = D(c_i) \oplus G(c_i), i = 1, 2, ...X_c \tag{2}$$

where $E(c_i)$ are the total extracted output features with $D(c_i)$ and $G(c_i)$. $\oplus$ denotes the concatenation operation.

F_G and F_D are then concatenated to form features F_C with a dimension of 2048, which is used for training the variational autoencoder followed by ML-based classification features.

Prior to sending the feature vector from the aforementioned ensemble into the variational architecture, normalization is performed on F_C to make it compatible with the autoencoder. For this, mean μ_C and standard deviation σ_C is calculated which is used to normalize F_C to provide features F_N.

The concatenated features F_C have a very high dimension which brings out a requirement to use a dimensionality reduction. For learning better distributions and sparse features from F_C, we integrated variational autoencoders. The variational aspect of the pipeline consists of an encoder-decoder assembly which rather than giving the latent space features gives us the distribution of the latent features in the form of its mean and standard deviation as:

$$\{F_\mu, F_\sigma\} \leftarrow VAE_{Encoder}\{F_N\} \tag{3}$$

The reparametrization of F_μ and F_σ is then performed by associating a constant ϵ with F_σ and adding the resultant to F_μ to give features $F_{reparam}$.

$$F_{reparam} = F_\mu + \epsilon \times F_\sigma \tag{4}$$

The latent space encoder is then used to give the output feature space of 100 feature.

A specialized loss function, L_{VAE}, consists of two factors, one which penalizes the reconstruction error and another which allows the learned distribution to be similar to our predefined distribution which is assumed to be a Gaussian distribution. The loss function is a sum of binary cross-entropy loss *i.e.* L_{BCE} and KL divergence loss *i.e.* L_{KL} as given:

$$L_{KL} = 0.5 \times \Sigma(1 + log(\sigma^2) - \mu^2 - \sigma^2) \tag{5}$$

where, μ and σ denote the mean and standard deviation. However, in the proposed model (as shown in Figure 4), F_μ and F_σ are used in equation 5 for μ and σ thereby giving us L_{KL}.

Algorithm 1: Proposed model framework

Require: Input chest X-ray images
Ensure: Three-class classification among COVID-19 v/s Pneumonia v/s Normal patients
1: Initialization:
$\{X_t, Y_t\} \longleftarrow D_1$
$\{X_c, Y_c\} \longleftarrow D_2$
2: Training CNN Models on Dataset D_1 $Y_t \leftarrow DenseNet\{X_t\}$
$Y_t \leftarrow GoogleNet\{X_t\}$
3: Making the chosen CNN modules compatible to the proposed model settings for feature extraction:
$DenseNet_{1024}\{X_t\} \leftarrow DenseNet\{X_t\}$
$GoogleNet_{1024}\{X_t\} \leftarrow GoogleNet\{X_t\}$
4: Extracting features from D_2 dataset:
$F_D \leftarrow DenseNet_{1024}\{X_{c1}\}$
$F_G \leftarrow GoogleNet_{1024}\{X_{c1}\}$
$F_C \leftarrow \{F_G, F_D\}$
5: Normalizing by calculating the mean and standard deviation of the extracted features:
$F_N \leftarrow \frac{F_C - \mu_C}{\sigma_C}$
6: Training of the features is performed in the variational environment (VAE):
$F_N \leftarrow VAE\{F_N\}$
$L_{VAE} \leftarrow L_{BCE} + L_{KL}$
7: Feature Extraction from the VAE module:
$\{VAE_{encoder}, VAE_{decoder}\} \leftarrow VAE\ \{F_\mu, F_\sigma\} \leftarrow VAE_{encoder}\{F_N\}$
8: ML-based predictive classification:
$Y_{SVM} \leftarrow SVM\{F_\mu\}$;
$Y_{RF} \leftarrow RF\{F_\mu\}$;
$Y_{XGB} \leftarrow XGB\{F_\mu\}$
$\hat{Y} \leftarrow LR\{Y_{SVM}, Y_{RF}, Y_{XGB}\}$

In nutshell, a VAE with $VAE_{encoder}$ and $VAE_{decoder}$ have regularised encoding distribution during training which gives mean and the standard deviation to understand the distribution of the latent space. Therefore, for increasing overall model robustness, F_N are passed through $VAE_{encoder}$ to extract F_μ and F_σ which properly represent the latent space distribution.

Machine Learning based classification: At the last stage, we employed the ML predictive classifiers where in F_μ is used to train the stacked arrangement of three classifiers, namely Support Vector Machine (SVM) [24], Random Forest (RF) [38], and XGBoost [39]. The output of this ensembling is later predicted by logistic regression (LR).

The framework gives the output, $\hat{Y}$, which is the final label to the X-ray image which either belongs to COVID-19, Normal or Pneumonia patient. The above proposed network architecture is as described in Algorithm 1.

5 Experiments

Implementation details: The proposed classification system for COVID-19 diagnosis is implemented using Python 3.8 programming language with a processor of Intel® Xeon(R) Gold 5120 CPU @ 2.20GHz×56 with 93.1 GiB RAM on Ubuntu 18.04.2 LTS with NVIDIA Quadro P5000 with 16GB Graphics.

5.1 Evaluation Criteria

To evaluate the efficiency of the proposed framework, the confusion matrix $\langle TP, TN, FP, FN \rangle$ also exploited in [40] along with area under curve (AUC) [41; 42] are estimated. They provide an understanding about the proposed methodology and its potential for detailed classification. AUC of ROC curve helps to check how well a classifier is able to distinguish among various classes.

The classification model's usefulness and productivity are measured using the traditional metrics of Accuracy (Ac), Sensitivity (Sen), and Specificity (Spe) as in Equations 6 7 and 8, respectively.

$F_1 - score$ as given in equation 9 is a measure that reports the balance between precision and recall.

ROC-AUC: ROC curve is the curve between true positive rate (TPR) and false positive rate (FPR) for a classifier. Area under the curve for ROC [41; 42] is an effective measure to check the efficacy of ML classifiers.

$$Accuracy\ (Ac) = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

$$Sensitivity\ (Sen) = \frac{TP}{TP + FN} \tag{7}$$

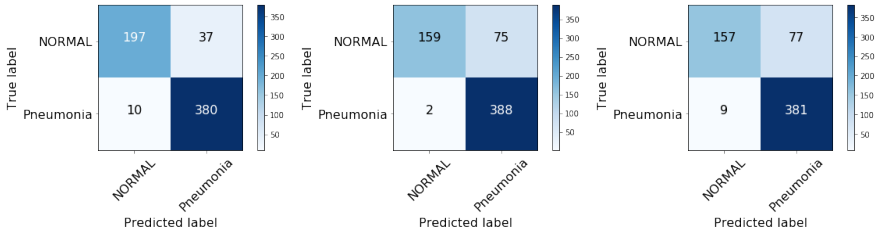
$$Specificity\ (Spe) = \frac{TN}{TN + FP} \tag{8}$$

$$F_1 - score = \frac{2 * TP}{2 * TP + FP + FN} \tag{9}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

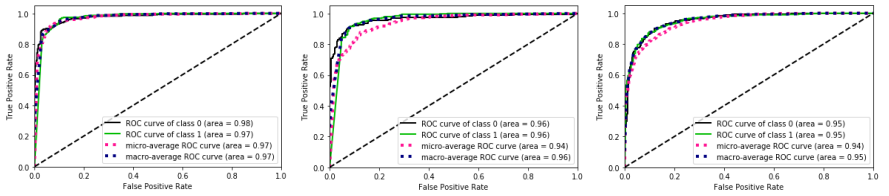
5.2 Results and Analysis

To ethically make a comparative analysis and selection of best CNN model for feature extraction, we initially show the baseline results on the deep modules, DenseNet, ResNet, and GoogleNet on the dataset $\mathcal{D}_2$ in Figure 5. The three chosen models are compared against the area under the curve (AOC) property



(a) (b) (c)

Fig. 5: Confusion matrix of the baseline results on DenseNet, ResNet and GoogLeNet.



(a) (b) (c)

Fig. 6: ROC curves of the baseline results on DenseNet, ResNet and GoogLeNet.

in receiver operating curves (ROC) curves (as shown in Figure 6) from which the best of two, DenseNet and GoogLeNet, are selected and are worked upon for further analysis. We observed that the baseline approach of three deep modules didn't perform satisfactorily at the training stage (whose evaluated metrics are depicted in Table 4) of which the training results on the ResNet are worse. Therefore, to tackle the issue of appropriate and efficient feature extraction, only the two modules, DenseNet and GoogLeNet are taken for further assessment and are operated upon.

Keeping the above conditions in mind, we extended our experiments on $\mathcal{D}_1$ dataset using the stacked architecture of DenseNet and GoogLeNet. The extracted features are then tested using SVM whose results in the form of confusion matrix and ROC curves are depicted in Figure 7 and 8, respectively. We observe that the chosen deep models with such high dimensions are not adequate for the feature extraction process. Therefore, to upright the process, a more robust, with fewer dimensions and better feature recognition capability segment is integrated which can have better generalization ability.

Variational Autoencoder (VAE): To further upgrade the accuracy with fewer dimensions, we need to extract more sparsity among the features which will help us to generate more meaningful latent space. This purpose is solved using the aspect of VAE. Post training, a non-linear dimensionality re-

Table 4: Baseline training-based performance on the GoogLeNet, ResNet and DenseNet.

Module	Performance Metrics				
	Ac	**Sen**	**Spe**	**F1-score**	**AUC**
GoogleNet	0.876	0.837	0.837	0.857	0.957
ResNet	0.862	0.823	0.823	0.841	0.952
DenseNet	0.924	0.908	0.908	0.917	0.969

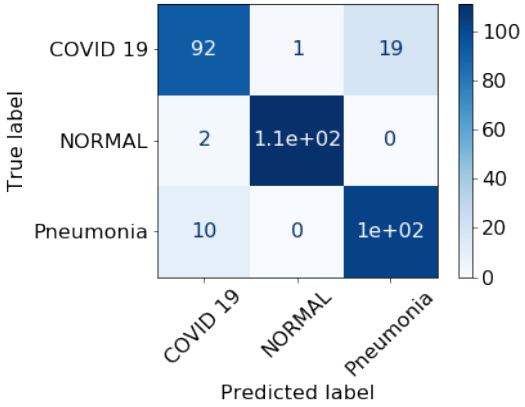


Fig. 7: Confusion Matrix of the DenseNet and GoogleNet feature Extraction modules with SVM-based classification.

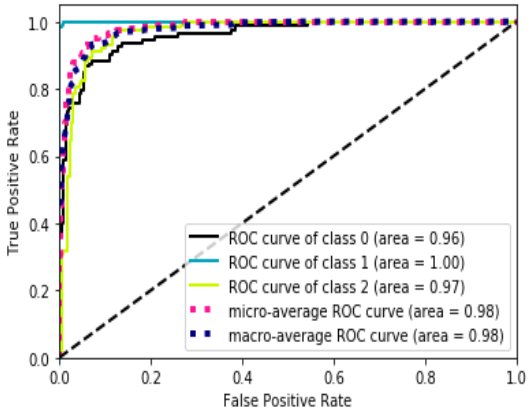
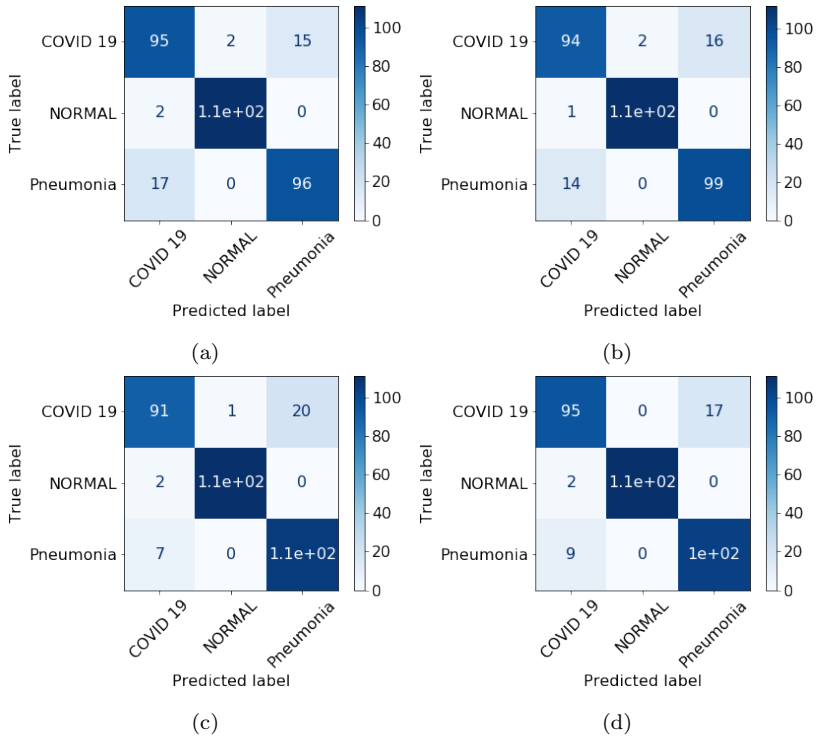


Fig. 8: ROC curves of the DenseNet and GoogleNet feature Extraction modules with SVM-based classifiers.

Table 5: Performance evaluation of ML-based classifiers for classification between COVID-19, Normal, Pneumonia.

Module	Performance Metrics				
	Ac	**Sen**	**Spe**	**F1-score**	**AUC**
SVM	0.911	0.910	0.955	0.910	0.971
RF	0.902	0.902	0.951	0.901	0.974
XGB	0.893	0.893	0.946	0.893	0.972
Ensemble	**0.917**	**0.916**	**0.958**	**0.917**	**0.976**



(a) (b) (c) (d)

Fig. 9: Confusion matrix displaying the final classification among the three classes of COVID-19, Normal and Pneumonia.

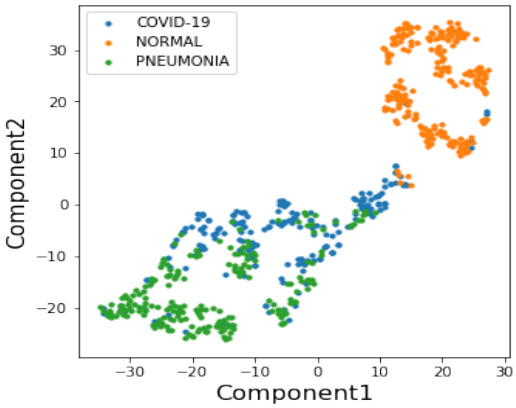


Fig. 10: Visualization of latent space of VAE with 2 components using t-SNE.

duction technique, t-SNE [43], is applied on the concatenated features. t-SNE representations calculates a similarity match between the data instances in high and low dimensions and later optimizing them. Thereafter, a latent space of 100 features is generated whose visualization is shown in Figure 10.

5.3 Quantitative Analysis

Now, we are in a state of understanding the latent space distribution of the extracted features which are further classified using the ML-based predictive classifiers. Considering the tactfulness of ML classifiers, we chose three basic classifiers, viz. SVM, RF, and XGBoost. These classifiers have successfully and easily classified the latent features into their basic classes. Finally, the probabilities predicted from these classifiers are used as meta-features for Logistic Regression to give the final label to our input images.

The entries are tabulated in Table 5. Elaborated results in the form of confusion matrix are shown in Figure 9, where in the Figure 9 (a), (b) and (c) depicts the classification for the SVM, RF and XGB with the Figure 9 (d) depicting the ensemble of above mentioned classifier with a final estimator of Logistic Regression for the final classification. The combined ROC curves for the three ML-based classifiers and the combined final estimator of LR is also shown in Figure 12.

5.4 Multi-class classification results on a challenging dataset

To further analyze the viability of our proposed framework, we implement the above mentioned framework on $\mathcal{D}_3$ dataset (as mentioned in Section 3). For this, we have reported the results for multi-class classification of the diseases

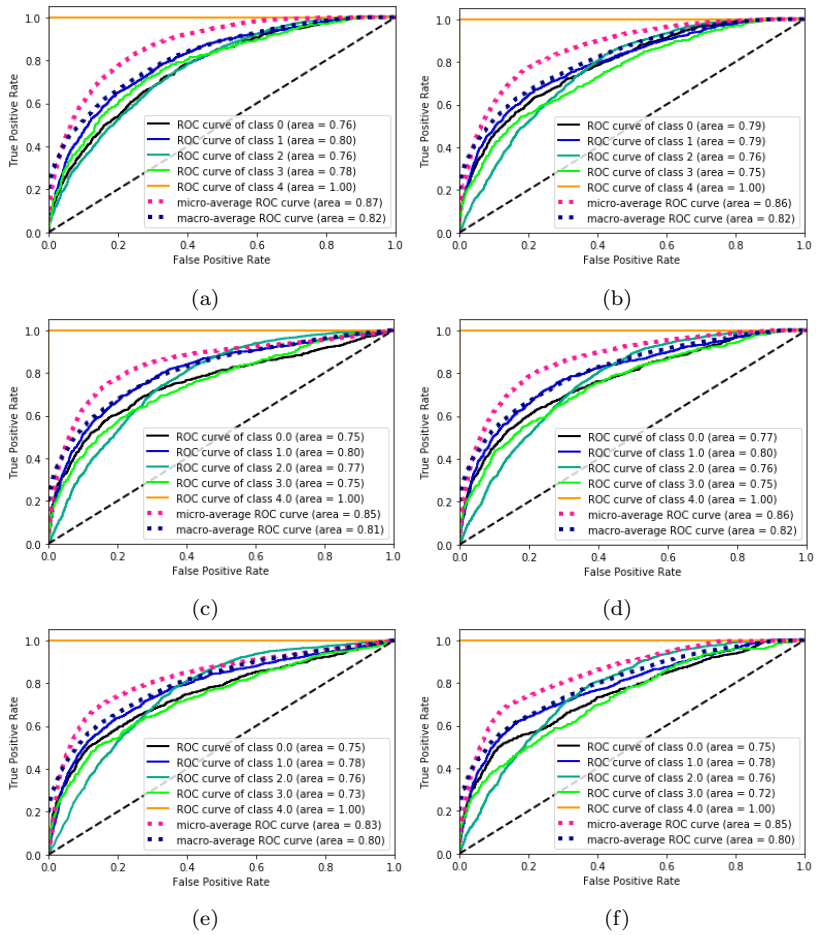


Fig. 11: ROC curves displaying the multi-class classification results on the NIH chest X-ray images.

which include Atelectasis, Effusion, Infiltration, Nodule, and Pneumonia. Figure 11 shows the corresponding ROC curves for the different stage results on the said dataset. Figure 11(a) and (b) show the ROC curves for the initial training of the DenseNet and GoogLeNet for the above mentioned five classes with Figure 11(c), (d) and (e) showing the corresponding final 5-label classification by SVM, RF, and XGBoost classifiers, respectively.

The extracted latent space from the VAE are converted into components using t-SNE which is visualized in Figure 13. It can be seen from the figure that many data points are distorted from the cluster of their labeled classes which in turn affects the performance of our classifiers which is clearly visible in table 6. From the table it can be observed that the quantitative performance varies greatly, in which the proposed pipeline has performed superior to the works

Table 6: Performance evaluation of ML-based classifiers for multi-class classification on ($\mathcal{D}_3$) dataset.

Module	Performance Metrics				
	Ac	**Sen**	**Spe**	**F1-score**	**AUC**
GoogleNet	0.603	0.586	0.882	0.591	0.818
DenseNet	0.614	0.597	0.886	0.605	0.813
SVM	0.614	0.597	0.886	0.605	0.813
RF	0.611	0.589	0.884	0.598	0.815
XGB	0.609	0.594	0.884	0.600	0.803
Ensemble	**0.616**	**0.597**	**0.886**	**0.605**	**0.802**

Table 7: Comparative analysis of the AUC for multi-class classification for the proposed model setting.

Study	Classes			
	Atelectasis	**Effusion**	**Infiltration**	**Nodule**
Wang *et al.* [44]	0.70	0.73	0.61	0.71
Our Approach	**0.75**	**0.78**	**0.76**	**0.72**

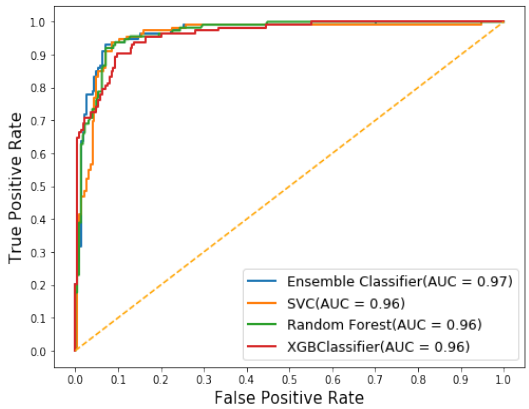


Fig. 12: ROC for Machine Learning based classification among the three classes of COVID-19, Normal and Pneumonia.

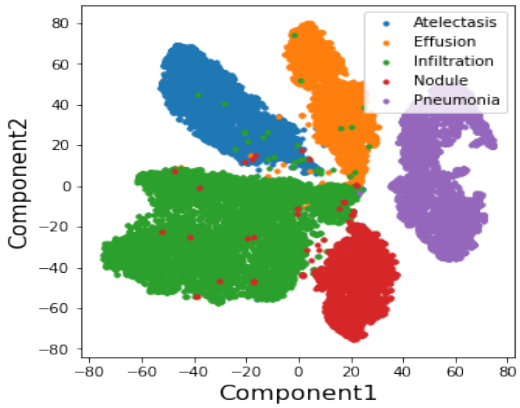


Fig. 13: Visualization of latent space of VAE with 2 components using t-SNE for Dataset 3

done by Xiaosong *et al.* [44]. The comparative study for the AUC for multi-class classification for the 4 classes (leaving out Pneumonia) is as depicted in Table 7 where the proposed method has outperformed the state-of-the-art results.

6 Conclusions

In this article, we proposed a novel framework to classify people using their chest X rays between COVID-19, Normal, and Pneumonia. This study utilizes state-of-the-art deep learning architectures for feature extraction and image processing. The input images are passed through the CNN models, DenseNet and GoogleNet, to generate features that are concatenated and sent to a variational autoencoder (VAE). VAE later learns a meaningful latent space from the features and further inputs the extracted features to ML-based classifiers. Final predictions are performed by ML-based predictive classifiers which help to improve the accuracy and robustness of the model. This proposed study thus helps to achieve an overall accuracy and AUC of 0.91 and 0.97, respectively. Also, we have tested the scalability and efficacy of the proposed framework on a challenging dataset to classify between Atelectasis, Effusion, Infiltration, Nodule, and Pneumonia and to check for viability and the diverse nature of the framework in applicant fields of medical image classification. The results of this dataset showed significant improvement over the state-of-the-art.

Acknowledgment

We acknowledge Consulate General of India, Osaka-Kobe for constant support and encouragement for this Bilateral India-Japan Artificial Intelligence (INJA-

AI) module between the researchers of Indian Institute of Technology Roorkee (IITR), India and Kyoto University, Japan to better manage COVID-19 by non-invasive prediction.

Declarations

Funding Information

There is no funding provided in the course of this study.

Compliance with Ethical Standards

All authors declare that they have no conflict of interest, financial or otherwise.

Availability of data and material

This article does not contain any studies with human participants or animals performed by any of the authors. All the database is acquired from the public logging system (Internet source) whose appropriate references are added in the sections above.

Code Availability

Any public available code is cited in the text at appropriate places and the novel code for the work can taken from the authors upon request.

References

1. World Health Organization; Naming the coronavirus disease (COVID-19) and the virus that causes it. Accessed on: June 6, 2020, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.
2. Worldometers: COVID-19 CORONAVIRUS PANDEMIC. Accessed on: June 6, 2020, https://www.worldometers.info/coronavirus/,.
3. World Health Organization; WHO Coronavirus Disease (COVID-19) Dashboard. Accessed on : June 3, 2020, https://covid19.who.int/.
4. World Health Organization; Coronavirus disease (COVID-19) Situation Report – 126. Accessed on: June 6, 2020, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200525-covid-19-sitrep-126.pdf?sfvrsn=887dbd66$_2$.

5. World Health Organization; WHO releases guidelines to help countries maintain essential health services during the COVID-19 pandemic. Accessed on: June 6, 2020, https://www.who.int/news-room/detail/30-03-2020-who-releases-guidelines-to-help-countries-maintain-essential-health-services-during-the-covid-19-pandemic.
6. World Health Organization; WHO Coronavirus Disease (COVID-19) Dashboard. Accessed on: June 6, 2020, https://covid19.who.int/?gclid=EAIaIQobChMI8aKK-PLR6QIVgdeWCh2t-gO-EAAYASAAEgLTv$_{DB}$wE.
7. Centers for Disease Control and Prevention; Symptoms of Coronavirus. Accessed on: June 6, 2020, cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.
8. Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*, page 200642, 2020.
9. World Health Organization; Rolling updates on coronavirus disease (COVID-19). Accessed on: June 6, 2020, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.
10. Zi Yue Zu, Meng Di Jiang, Peng Peng Xu, Wen Chen, Qian Qian Ni, Guang Ming Lu, and Long Jiang Zhang. Coronavirus disease 2019 (covid-19): a perspective from china. *Radiology*, page 200490, 2020.
11. Jeffrey P Kanne, Brent P Little, Jonathan H Chung, Brett M Elicker, and Loren H Ketai. Essentials for radiologists on covid-19: an update—radiology scientific expert panel, 2020.
12. Xingzhi Xie, Zheng Zhong, Wei Zhao, Chao Zheng, Fei Wang, and Jun Liu. Chest ct for typical 2019-ncov pneumonia: relationship to negative rt-pcr testing. *Radiology*, page 200343, 2020.
13. Weifang Kong and Prachi P Agarwal. Chest imaging appearance of covid-19 infection. *Radiology: Cardiothoracic Imaging*, 2(1):e200028, 2020.
14. Heshui Shi, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, and Chuansheng Zheng. Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study. *The Lancet Infectious Diseases*, 2020.
15. Jasper Fuk-Woo Chan, Shuofeng Yuan, Kin-Hang Kok, Kelvin Kai-Wang To, Hin Chu, Jin Yang, Fanfan Xing, Jieling Liu, Cyril Chik-Yan Yip, Rosana Wing-Shan Poon, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223):514–523, 2020.
16. KC Santosh. Ai-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. *Journal of Medical Systems*, 44(5):1–5, 2020.
17. Ridhi Arora, Prateek Kumar Rai, and Balasubramanian Raman. Deep feature–based automatic classification of mammograms. *Medical & Biological Engineering & Computing*, pages 1–13, 2020.

18. Taufik Rahmat, Azlan Ismail, and Sharifah Aliman. Chest x-rays image classification in medical image analysis. *Applied Medical Informatics.*, 40(3-4):63–73, 2018.
19. Ivo M Baltruschat, Hannes Nickisch, Michael Grass, Tobias Knopp, and Axel Saalbach. Comparison of deep learning approaches for multi-label chest x-ray classification. *Scientific Reports*, 9(1):1–10, 2019.
20. Ezz El-Din Hemdan, Marwa A Shouman, and Mohamed Esmail Karar. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*, 2020.
21. Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *medRxiv*, 2020.
22. Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *medRxiv*, 2020.
23. Prabira Kumar Sethy and Santi Kumari Behera. Detection of coronavirus disease (covid-19) based on deep features. 2020.
24. Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
25. Linda Wang and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. *arXiv preprint arXiv:2003.09871*, 2020.
26. D Das, KC Santosh, and U Pal. Truncated inception net: Covid-19 outbreak screening using chest x-rays. *PREPRINT (Version 1) available at Research Square*, 3, 2020.
27. Himadri Mukherjee, Subhankar Ghosh, Ankita Dhar, Sk Obaidullah, KC Santosh, Kaushik Roy, et al. Shallow convolutional neural network for covid-19 outbreak screening using chest x-rays. 2020.
28. Ophir Gozes, Maayan Frid-Adar, Hayit Greenspan, Patrick D Browning, Huangqi Zhang, Wenbin Ji, Adam Bernheim, and Eliot Siegel. Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. *arXiv preprint arXiv:2003.05037*, 2020.
29. Shahabedin Rahmatizadeh, Saeideh Valizadeh-Haghi, and Ali Dabbagh. The role of artificial intelligence in management of critical covid-19 patients. *Journal of Cellular & Molecular Anesthesia*, 5(1):16–22, 2020.
30. Ratnabali Pal, Arif Ahmed Sekh, Samarjit Kar, and Dilip K Prasad. Neural network based country wise risk prediction of covid-19. *arXiv preprint arXiv:2004.00959*, 2020.
31. Chest X-Ray Images (Pneumonia). Accessed on : June 6, 2020, https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia.
32. COVID-19 Radiography Database. Accessed on : June 6, 2020, https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

33. COVID Dataset. Accessed on : June 6, 2020, https://drive.google.com/a/cs.iitr.ac.in/uc?id=1coM7x3378f-Ou2l6Pg2wldaOI7Dntu1a.
34. Figure1-COVID-chestxray-dataset. Accessed on : June 6, 2020, https://github.com/agchung/Figure1-COVID-chestxray-dataset.
35. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
36. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
37. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
38. Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
39. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
40. A. Gupta, R. Kumar, H. Singh Arora, and B. Raman. MIFH: A machine intelligence framework for heart disease diagnosis. *IEEE Access*, 8:14659–14674, 2020.
41. Mark H Zweig and Gregory Campbell. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. chem.*, 39(4):561–577, 1993.
42. James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.
43. Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
44. Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

RIDHI ARORA received her B.Tech degree in Computer Science & Engineering from MM University, Ambala in 2014 and M.Tech degree in Computer Science & Engineering from DCRUST, Sonipat, INDIA in 2016. She is currently enrolled in the PhD program in the Department of Computer Science & Engineering at Indian Institute of Technology Roorkee, Uttarakhand, INDIA. Her research interests include Biomedical engineering, image processing,

machine learning and feature extraction methods in the domain of biomedical imaging.

VIPUL BANSAL is currently pursuing the bachelor's degree with the Department of Mechanical and Industrial Engineering, IIT Roorkee, Roorkee, India. His research interests include using the machine and deep learning techniques in the field of medical imaging and focused on improving healthcare with the help of these techniques.

HIMANSHU BUCKCHASH received the B.Tech. degree in computer science and engineering from the Harcourt Butler Technological Institute, in 2011, and the M.Tech. degree in computer engineering from the National Institute of Technology, Kurukshetra, in 2014. He is currently pursuing the Ph.D. degree in computer science and engineering with the Indian Institute of Technology, Roorkee, where he is advised by Prof. Balasubramanian Raman. Before his masters, he worked with Newgen Software Tech. Ltd., Delhi, on BPM and ECM related products. His research interests include unsupervised learning, video analysis, emotion recognition, machine learning, and object detection.

RAHUL KUMAR received the B.Tech. degree in information technology from the Muzaffarpur Institute of Technology Muzaffarpur, India in 2012, and the M.Tech. degree in computer science and engineering from Dr. B. R. Ambedkar National Institute of Technology Jalandhar, Jalandhar, India, in 2015. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering from IIT Roorkee, Roorkee, India. His research interests include image processing, and application of machine and deep learning in biomedical imaging.

VINODH J SAHAYASHEELA received his B.Tech and M.Tech degree in Biotechnology from Anna University, India in 2013 and 2015 respectively. Later he joined CSIR-National Chemical Laboratory as Senior Research Fellow and his research focus was on Natural product and Cancer Biology. He has now joined Professor Sugiyama's Chemical Biology group at Kyoto University as a Doctoral student with Japanese government MEXT fellowship. His research interests are the synthesis and biological evaluation of small molecules for sequence specific DNA binding, Nanopore sequencing and G-quadruplex. He has qualified CSIR, DBT and ICMR fellowship and has

five publications to his credits.

NARAYANAN NARAYANAN is a member in the advisory board of International strategies, University of Cyberjaya, Malaysia. He is a consultant with more than 33 years of experience with proven participation in the early stage investment and in structuring the capital of one of the most prestigious Pharma companies in India He has recently been awarded "Movers and shakers" award by Asia HRD congress based in Malaysia for having trained school dropouts in running venture business. He has been mentoring several start-ups including CFS Healthcare (which has developed a B2B portal for sourcing pharma products by the pharmacies from distributors). He is also advising GenMi, a company which is into revolutionary pathway development in Genetic Engineering

NAMASIVAYAM GANESH PANDIAN started his research in Indian Institute of Technology Madras and moved to Niigata University with Japanese Government scholarship to complete his PhD (Biosciences) in 2009. He relocated as a Research Associate at Kyoto University in 2010 after serving as a visiting scientific advisor in Ushiki patent office. He is now a tenured Associate Professor (Jr.) and Principal Investigator in World Premier International Research Center – Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Japan.

BALASUBRAMANIAN RAMAN received his Ph.D. degree from IIT Madras, in 2001. He was a Post-Doctoral Fellow with the University of Missouri, Columbia, MO, USA, from 2001 to 2002, and a Post-Doctoral Associate with Rutgers, The State University of New Jersey, USA, from 2002 to 2003. He was a Visiting Professor and a member with the Computer Vision and Sensing Systems Laboratory, Department of Electrical and Computer Engineering, University of Windsor, Canada, in 2009. He is currently a Professor with the Department of Computer Science and Engineering, IIT Roorkee, Roorkee, India. His research interests include fractional transform theory, wavelet analysis, biometrics, content-based video retrieval, video skimming and summarization, medical imaging, long-range imaging, and hyperspectral imaging. He has more than 150 research publications in reputed journals and conference proceedings. He is a member of the IEEE Society, Uttar Pradesh section and acted as a Joint Secretary of the Executive Committee in the IEEE Uttarakhand Sub-Section from 2011 to 2013. He was a recipient of the BOYSCAST Fellowship from DST, India.