Received: 8 July 2021 | Revised: 12 September 2021 | Accepted: 13 October 2021

DOI: 10.1111/exsy.12884

ORIGINAL ARTICLE

Expert Systems WILEY

A hybrid learning approach for the stage-wise classification and prediction of COVID-19 X-ray images

Adimoolam M.[1] | Karthi Govindharaju[2] | John A.[3] | Senthilkumar Mohan[4] | Ali Ahmadian[5,6] | Tiziana Ciano[7]

[1]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

[2]Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, India

[3]School of Computer Science and Engineering, Galgotias University, Greater Noida, India

[4]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

[5]Institute of IR 4.0, The National University of Malaysia, UKM, Bangi, Malaysia

[6]Department of Mathematics, Near East University, Nicosia, TRNC, Mersin 10, Turkey

[7]Faculty of Business and Law, University of Portsmouth, Portsmouth, UK

Correspondence

Ali Ahmadian, Institute of IR 4.0, The National University of Malaysia, UKM Bangi, 43600 Selangor, Malaysia.

Email: ali.ahmadian@ukm.edu.my

Abstract

Background: The COVID-19 pandemic has precipitated global apprehensions about increased fatalities and raised concerns about gaps in healthcare infrastructure and accessibility the world over. Consequently, the importance of timely prediction and treatment of the disease to reduce transmission and mortality rates cannot be emphasized enough. Various symptoms of the disease have been identified as it progresses from the time it is contracted. COVID-19 has been found to internally affect the lungs, and the four progressive stages of the infection can be categorized as mild, moderate, severe, and critical. Therefore, an accurate analysis of the current stage of the disease that can help predict its progression has become critical. X-ray imaging has been found to be an effective screening procedure for predicting the various stages of this epidemic. Although many different approaches using machine learning, as well as deep learning were utilized to predict and classify diseases in general, till date, such an approach has not been used to predict the various stages of COVID-19 by using X-ray imaging to identify and classify those stages.

Materials and method: The proposed hybrid method used three public datasets for its implementation. In this work, extensive images were used for the purposes of testing and training. The dataset-1 consists of 1200 COVID-19 as well as 1200 Non-COVID-19 images, while dataset-2 used 700 COVID-19 as well as 700 Non-COVID-19 images, and finally, dataset-III utilized 1900 COVID-19 as well as 1900 Non-COVID-19 images for purposes of testing and training. The proposed work undertook the task of pre-processing using textual and morphological features, while the segmentation and prediction of COVID-19 as well as Non-COVID-19 images were undertaken using VGG-16 with light GBM for better prediction and handing of huge datasets, and finally, the classification of the various stages of COVID-19 images was performed using Deep Belief Network.

Results: The outcomes of the proposed work were subjected to several iterations which were then compared using different parameters such as accuracy, specificity, and sensitivity. In general, the prediction and grouping of the various stages of COVID-19 by using affected images were found to be 99.2%, 99.4% and 99.5%, respectively. The bacterial pneumonia prediction rates were observed to be 98.5%, 99.4% and 98.3%, respectively. The average classification of the stages were found to be 98.1%, 98.6% and 98.3%, while the combined multi-classification prediction rates were observed to be 98.6%, 99.1% and 98.7%, respectively.

KEYWORDS
COVID-19, deep belief network, deep learning, X-ray images, VGG-16

1 | INTRODUCTION

COVID-19 is known to be a transmittable infection, and is found to spread at a rapid rate and affect billions of people around the world. This has ensured that the numbers of affected as well as recovering peoples are increasing day by day. The various symptoms associated with the disease have also undergone changes from its initial days of infecting humans to the present times, and have been seen in many cases as becoming more aggressive. The lung is usually seen to be the most affected organ in COVID-19, as is the case with common bacterial or viral infections. Therefore, it is important to predict the stage of lung disease in a patient once infected by COVID-19 and in this study different technology, especially machine learning techniques, have been introduced in order to predict this, which has demonstrated a fair degree of success and accuracy. The four stages of COVID-19 that were considered in these predictions to determine the degree of progression of lung disease (https://www.sciencedaily.com/releases/2020/04/200423130420.html, n.d.) are (i) Phase-1, where mild cell incursion and viral duplication are observed in the nose and lungs. (ii) Phase-2, where moderate amount replication is seen to affect the lungs and the immune system. (iii) The more severe Phase 3, where the replication is consolidated across all sections of the lungs threatening a collapse of the lungs as they struggle to keep the alveoli open and close. (iv) Phase 4, where the patient is critically affected with an onset of failure of multiple organs. All these various stages of COVID-19 along with the training images used and the related symptoms are shown in Table 1. It is therefore, evident that lung-based initial prediction of the disease and its accurate staging can go a long way in arresting the progression of COVID-19 and facilitating effective treatment.

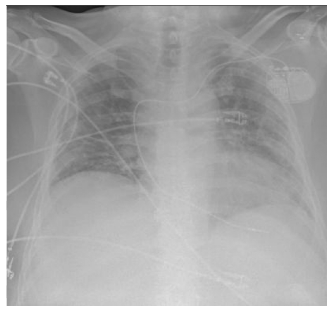
In recent times, several researchers have been seen to use AI-based methodologies, especially machine learning and deep learning, for classifying and predicting images in healthcare where, X-ray, histopathology as well as CT scan images are usually the subject of predictions. Deep Learning has been hailed as a potent tool that can learn cognitive and complex problems. As a result, there has been a surge in the usage, evaluation, and prediction of diseases using deep learning methods. This paper undertook the prediction of COVID-19 stages by utilizing X-ray lung images aided by machine learning as well as deep learning methods. Previous predictions based on deep learning encountered various issues and challenges that this paper has addressed while undertaking to make predictions regarding the progression of COVID-19 by using X-ray lung images are shown below:

i. The previous predictions (Bharati et al., 2020; Jain et al., 2021; Mahdy et al., 2020; Minaee et al., 2020) used fewer images for training, testing and, validation, therefore negatively impacting prediction accuracy rates.
ii. There was, consequently, a concerted effort to improve the efficiency and accuracy of prediction, especially while using larger datasets.
iii. The previous works (Bharati et al., 2020; Jain et al., 2021; Mahdy et al., 2020; Minaee et al., 2020) used fewer features for prediction, whereas significant features are required for improved accuracy.
iv. The previous predictions did not consider the various stages depicted by the COVID-19 images (https://www.sciencedaily.com/releases/2020/04/200423130420.html, n.d.; Singh et al., 2020), even though incorporating the stages or phases in the prediction would help to improve the treatment significantly.

This paper addresses all the above-mentioned issues with its proposed methodology that enabled the contributions.

This work is segregated in the following manner: Section 2 enumerates various works on classification and prediction techniques of images and X-ray images of COVID and other diseases. The materials and methods, including the proposed hybrid methodologies, as well as stage-wise classification of COVID images are introduced in section 3. The implementation of the proposed work, its results, along with performance comparisons and evaluation are presented in Section 4. Finally, a summary as well as the scope of proposed future work has been provided in Section 5.

TABLE 1 Various stages of COVID-19

S. no	Stages	Symptoms	COVID-19 images
1	Stage-1	Invading and replicating in the lungs and nose. (https://www.sciencedaily.com/releases/2020/04/200423130420.html, n.d.; Sharma et al., 2020; Singh et al., 2020)	
2	Stage-2	Replicating moderately and affecting immune system as well as lungs. (https://www.sciencedaily.com/releases/2020/04/200423130420.html, n.d.; Sharma et al., 2020; Singh et al., 2020)	
3	Stage-3	Consolidated replication leading to collapse of parts of the lungs. (https://www.sciencedaily.com/releases/2020/04/200423130420.html, n.d.; Sharma et al., 2020; Singh et al., 2020)	
4	Stage-4	Severe damage to lungs and criticality accompanied by onset of multiple-organ failure. (https://www.sciencedaily.com/releases/2020/04/200423130420.html, n.d.; Sharma et al., 2020; Singh et al., 2020)	

2 | RELATED WORK

Computer-aided diagnosing systems were introduced in 1980, and have been applied to different scenarios of disease prediction and diagnosis. Initially, such image-based diagnosing systems were time-consuming and inefficient. However, with the invention of artificial intelligence and graphic processor units in decision support systems, the prediction rates witnessed a rapid improvement (Bharati et al., 2020). Deep learning models were applied to diseases such as breast cancer, lung cancer, and various other diseases for the purposes of detection and diagnosis. This work is specifically focused on prediction of COVID-19 stages by utilizing X-ray images with a machine learning and deep learning methodology. Various researchers have undertaken similar work with respect to predictions based on the lung X-rays.

Bharati et al. (2020) proposed a hybrid model using VGG, spatial transform, and data argumentation. But the accuracy (73%) of the prediction was relatively lesser since fewer images were utilized for the purposes of training and testing. Cohen et al. (2020) collected various images of COVID-19 and created datasets. But it had only 123 frontal views of images. Mahdy et al. (2020) proposed a model for classifying COVID-19 lung images using multiple levels of thresholds and SVM. Shuja et al. (2021) made a survey of COVID-19 datasets and presented various datasets such as images, text, and speech. Sharma et al. (2020) proposed a model for classifying X-ray images to relevant COVID-19 stages using different augmentations of images. Using this method, the COVID-19 as well as Non-COVID-19 images were classified. Dubey (2020) proposed a model for predicting COVID-19 using X-ray images and deep learning. The ensemble and feature classification techniques such as SVM and bagging classifier were used for prediction. Al-Waisy et al. (2020) proposed a framework based on deep learning for predicting COVID-19 by utilizing X-ray images. Minaee et al. (2020) proposed a deep COVID model for predicting COVID-19 using X-ray images. In this model, testing and training of 184 images were used for prediction. Jain et al. (2021) similarly proposed the detection of COVID-19 by utilizing a deep learning model. The images were predicted using Xceplm, V3 and ResNet. The accuracy rate of the prediction was 97.9%. Cheng et al. (2020) presented images of COVID-19 and introduced chest CT view of scans. Pinter et al. (2020) proposed a machine learning based model utilizing multiple-parameters to predict COVID-19. Hussain et al. (2020) proposed a model for prediction using chest X-rays. The texture and morphological features used to predict utilized the deep learning method. Table 2 presents a summary of the various existing techniques for prediction as well as the number of images used for testing, training and prediction rates.

3 | MATERIALS AND METHODS

In this research work, a hybrid deep learning approach with textual and morphological features has been proposed in order to classify and predict COVID-19 cases by utilizing X-ray images of both affected and normal cases. Sections 3.1 and 3.2 show the dataset details and the proposed work description.

3.1 | Materials: COVID-19 datasets

The datasets consist of COVID-19 affected images as well as normal images. The lung images affected by COVID-19 were gathered from various websites and research organizations. The affected as well as normal datasets have been included in the reference list ((https://www.kaggle.com/tawsifurrahman/covid19-radiography-database, n.d.; https://www.kaggle.com/c/dlai3/data, n.d.; Aboutalebi et al., 2021; Cohen et al., 2020; COVID, C. A. A, 2020; https://github.com/agchung/Actualmed-COVID-chestxray-dataset, 2020)). Table 3 presents the information regarding the training and testing images utilized to predict the images that were COVID-19 positive. The samples COVID-19 affected and normal images are shown in Figure 1.

3.2 | Methodology

The proposed prediction consists of pre-processing, segmentation of affected and non-affected images, training and testing, prediction of COVID-19, and classification of the various stages of prediction. The architecture of the prediction using X-ray lung images has been illustrated in Figure 2. The proposed work is consisted of four steps such as:

TABLE 2 Various methods: prediction rates and number of images

S. no	Techniques	Prediction rate	Images for training and testing
1.	VDSNet (Bharati et al., 2020)	Accuracy (73%)	Limited Number of Images used.
2.	Multi-level thresholding and support vector machine (Mahdy et al., 2020)	Accuracy (95.76%)	25 Normal Images and 25 COVID affected images used.
3	COVID-Net (Wang et al., 2020)	Accuracy (92.4%)	13,975 and 13,870 patient images used.
4.	Deep COVID (Minaee et al., 2020)	Sensitivity (98%) Specificity (92%)	Training- 84 images, Testing- 100 images.
5.	Xception and ResNext (Jain et al., 2021)	Accuracy (97.5)	COVID affected image: 86.
6.	Deep learning and various classification techniques (Hussain et al., 2020)	Multi-classification accuracy (97.5%) prediction accuracy (79.52)	Texture and Morphological features were used to predict. The number of images used was limited.
7.	CNN (Sahlol et al., 2020)	Accuracy (98.7)	Two datasets were used to predict but the number of images were limited.

TABLE 3 Dataset information

S. no	Dataset references	COVID-19 images	Normal images
1	COVID-19 Image Database (Aboutalebi et al., 2021)	1200	1200
2	github.com - COVID Images (https://github.com/GeneralBlockchain/covid-19-chest-xray-lung-bounding-boxes-dataset, n.d.)	700	700 (collected from various sources)
3	Combined Datasets (https://www.kaggle.com/c/dlai3/data, n.d.; https://www.kaggle.com/tawsifurrahman/covid19-radiography-database, n.d.; https://github.com/ieee8023/covid-chestxray-dataset, n.d.; https://github.com/GeneralBlockchain/covid-19-chest-xray-lung-bounding-boxes-dataset, n.d.; Aboutalebi et al., 2021; Cohen et al., 2020; COVID, C. A. A, 2020; https://github.com/agchung/Actualmed-COVID-chestxray-dataset, 2020)	2500	1500

i. Initializing the sample images or visualizing X-ray images from the patient data.
ii. Pre- Processing of datasets.
iii. Predicting COVID-19 (positive) as well as non-COVID (negative) images using VGG-16.
iv. Classification of various stages of COVID-19 images.

3.2.1 | Pre-processing of dataset

The pre-processing procedure converts the image into digital images. The main aim of digital image processing is to improve the features and suppress any distortions in the image features. The images in the datasets have non-uniform shapes and sizes, therefore, before initiation of training, there was a need to resize the images. The pre-processing procedures were performed using texture and morphological features. Energy, Entropy, Correlation, Contrast, Homogeneous are some of the features that were calculated using the textural properties and are represented in the Equations (1)–(5) and these Equations are for Energy, Entropy, Contrast, Correlation and Homogeneity features, respectively. The Grey-level matrix used to calculate the textural structures of the image have been included (Shuai et al., 2017).

$$Energy = \sum_{i,j=1}^{N-1} \left(P_{i,j}\right)^2 \tag{1}$$

$$Entrophy = \sum_{i,j=1}^{N-1} -\ln\left(P_{i,j}\right) P_{i,j} \tag{2}$$

$$Contrast = \sum_{i,j=1}^{N-1} \left(P_{i,j}\right)(i-j)^2 \tag{3}$$

$$Correlation = \sum_{i,j=1}^{N-1} \left(P_{i,j}\right) \frac{(i-\mu)(j-\mu)}{\sigma^2} \tag{4}$$

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{5}$$

where $P_{i,\,j}$ is normalized symmetrical, N is number of grey levels, μ is mean intensity, σ is variance of intensity. The textural features were calculated using various equations used in (Castellano et al., 2004; Esgiar et al., 2002; Guru et al., 2010; Haider et al., 2017; Khalvati et al., 2015; Khuzi et al., 2008; Yu et al., 2017). The prediction size, radius of the prediction area, image equivalence, and dispersion were calculated using morphological features shown in Equations (6) and (7). The morphological features were measured using various equations used in (Shuai et al., 2017). The varying scale morphological analysis (Shuai et al., 2017) was utilized in order to forecast the morphological structures of COVID-19 X-ray lung images (Eren et al., 2019; Fouladi et al., 2021; Gupta et al., 2021; Kundu et al., 2021; Mohan et al., 2021).

$$Difference = \{\omega | \omega \in A, \omega \notin B\} \tag{6}$$

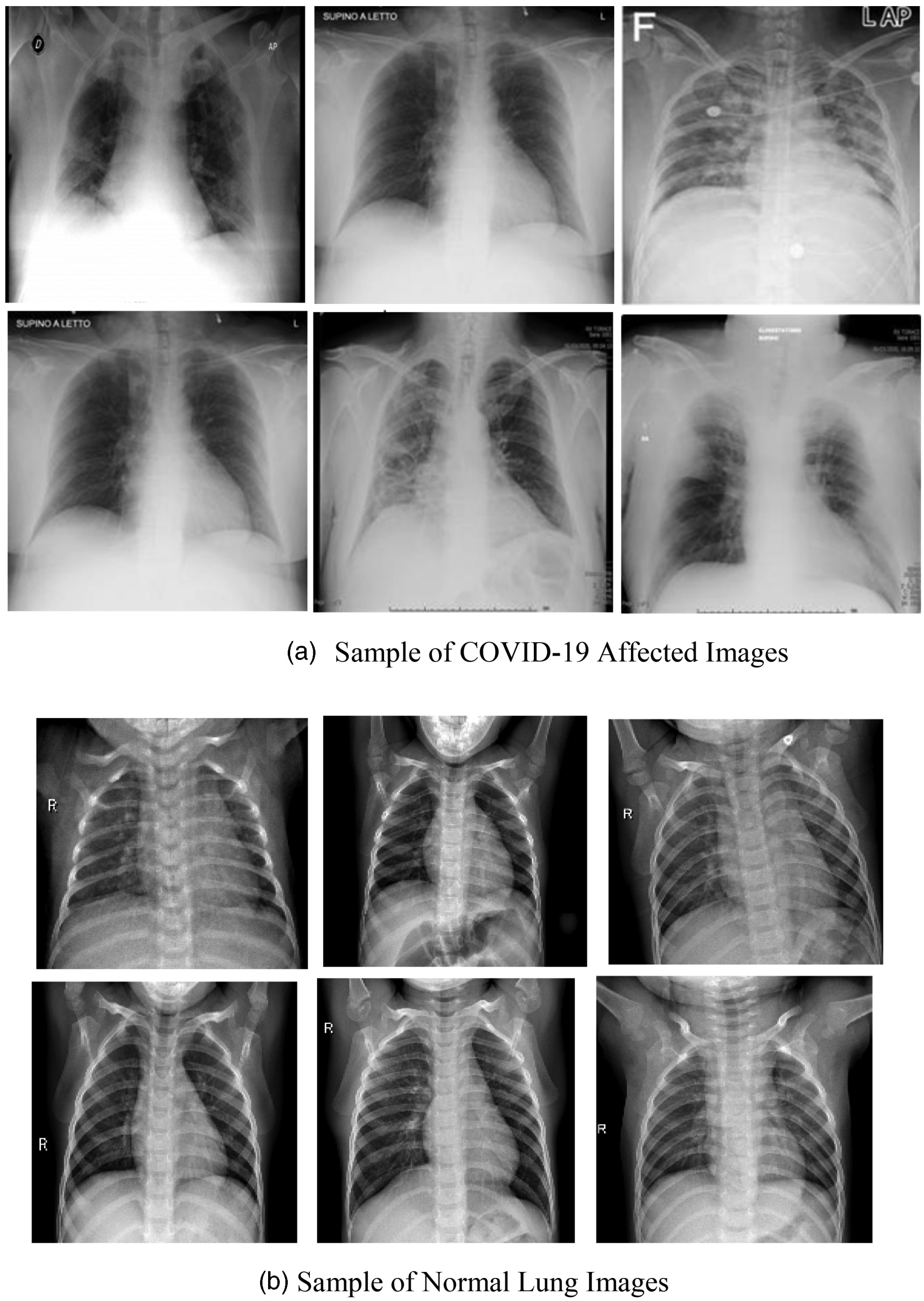


FIGURE 1 Sample of COVID-19 affected images and normal lung images

$$Equalance = \{\omega \,||\, \omega \in A, \omega \in B\} \tag{7}$$

In the proposed work three types of datasets images were resized. In the first dataset (Aboutalebi et al., 2021), 1200 COVID-19 positive as well as 1200 negative images were checked and resized into 256*256 pixels. The rectangular lung images were resized into 256 pixels. For better prediction, the images were augmented in different ways. The various phases of COVID classification from the dataset is described as follows which could be used for training and testing phases. Further various stages of COVID images are also trained and tested. It is represented as five steps as follows.

i. The proposed learning-based hybrid model helped improve the efficiency and prediction accuracy by utilizing X-ray Images.

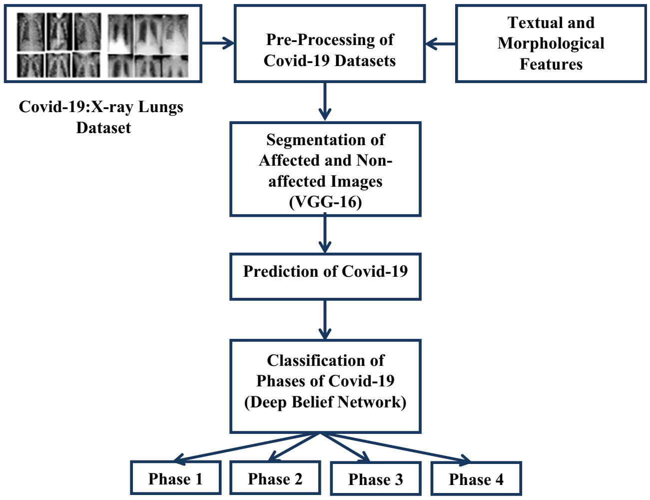


FIGURE 2 Architecture of prediction using X-ray lung images

ii. 3400 COVID affected images and 4400 non-affected images were used during implementation. In total, 7800 images were used for training, testing, and evaluation.
iii. The X-ray image-based predictions incorporated texture and morphological features using hybrid deep learning techniques.
iv. The Prediction and classification of positive as well as negative COVID-19 images were undertaken using VCC-16 and light GBM. The use of Light GBM significantly decreased the complexity and computation time involved.
v. The various classifications and predictions derived from the affected images were carried out by using the layer-wise features of Deep Belief Network (DBN).

3.2.2 | Segmentation of COVID-19:

The VGG-16 was used to segment the COVID-19 (positive) as well as the Non-COVID-19 (negative) images. VGG-16 is a deep conventional neural network technique for classifying or segmenting large-scale datasets or images. The previous works (Ahsan et al., 2021; Luz et al., 2021; Panwar et al., 2020; Sitaula & Hossain, 2021) aimed to only predict, but this work has endeavoured to predict as well as classify the various stages of COVID-19 images. The VGG-16 had an accuracy of 92.7 in the top five tests of images using 138,357,544 parameters, whereas the VGG-19 had an accuracy of 90.2 in the top five tests of images using 143,667,240 parameters. This work used VGG-16 as a means of using fewer parameters and achieving higher accuracy of classification and prediction.

The VGG-16 contained fixed-sized images such as 224*224 as an input and had an output of 1000 values. But in this work, the output of the images were classified as positive and negative. The general representation of the output of VGG-16 is shown in Equation (8).

$$Y = \begin{pmatrix} Y_0 \\ . \\ . \\ Y_{999} \end{pmatrix} \tag{8}$$

In this work, the output prediction of the proposed work is either Negative (Y_0) or Positive (Y_1). The output using VGG-16 representation is shown in Equation (9).

$$Y = \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} \tag{9}$$

The VGG-16 was used to segment the given input images as positive or negative. The diagrammatic representation of the proposed work with VGG-16 and light GBM has been shown in Figure 3.

Light Gradient Boosting was used to boost the outputs as it supported the huge size of datasets. Compared to other ensemble learning methods, it is 10 times faster in the computation of a huge number of datasets. The combination of VGG-16 and light GBM decreased the computation time and increased the accuracy of predictions. VGG-16 and light GBM algorithm are better combination in image-based prediction technique. Even though light GBM is combined with other prediction algorithms, the accuracy and predictions are considerably less.

3.2.3 | Stages of COVID-19 X-ray images

The various stages of the predicted COVID-19 images were performed using DBN. The DBN is a multiplicative graphical representative model and is another class of deep neural networks. It has multiple hidden layers from the observation and tinning layers. The normal DBN performs the classification with each layer having a set of features classifications. One such a belief-based monitoring on real time application and compulsion was discussed for image data driven method (Hirra et al., 2021). The work was that the effective useful features of images were extracted to guide the statistical calculation to monitor further. Yet another work using belief network was for breast cancer prediction from image and it was discussed (Mengistu, 2018) as follows. The unsupervised and supervised algorithm has been utilized for pre-training and fine-tuning processes respectively for automation feature extraction using this belief network model. The images of dataset were patched and logistic regression algorithm was used to support for gaining probability matrix for positive and negative sample prediction. The feature classifications are performed using encoders. The supervised learning data was predicted using gradient functions and unsupervised learning was performed using Restricted Boltzmann machine network or autoencoder. The DBN started from the training layer and extracted the different features in each layer. The main properties of the DBN included training the features of COVID-19 images, pixels, and signals of images directly in each layer. From the second layer onwards, the features increased and every layer's classification features increased in the belief network. The DBN network is shown in Figure 4 and each layer was well interconnected to make the classification of the given dataset.

The supervised model had three parameters for training the nets such as weight (W_{AB}) and related properties of images A_n and B_n, represented in Equations (10) and (11).

$$A_n = [a_1, a_2 \dots a_n] \tag{10}$$

$$Bn = [b_1, b_2 \dots b_n] \tag{11}$$

The energy equation represented for training and computation is represent in Equation (12).

$$E(x,y,\theta) = \sum_{x}^{y} w_{ij} x_i y_j - \sum_{x=1}^{x=n} a_i x_i - \sum_{y=1}^{y=n} b_i y_i \tag{12}$$

The continuous random prediction of COVID X-ray image characteristics and multi-variants of image distribution was performed using Probability Density Function (PDF) and is represented in Equation (13).

$$P(x,y) = \frac{e^{-E(x,y)}}{\sum_{x,y}^{n} e^{-E(x,y)}} \tag{13}$$

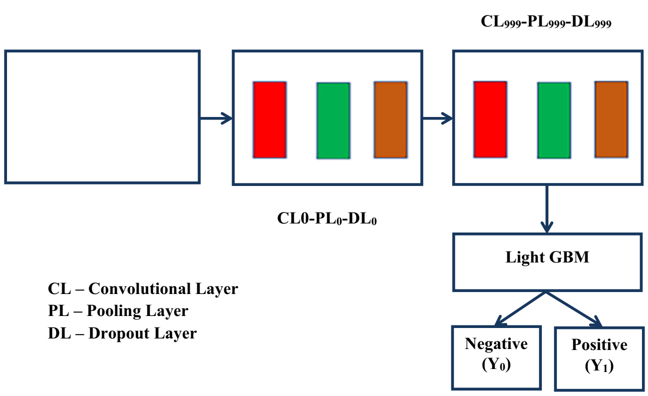


FIGURE 3 Segmentation of images using VGG-16 and light GBM

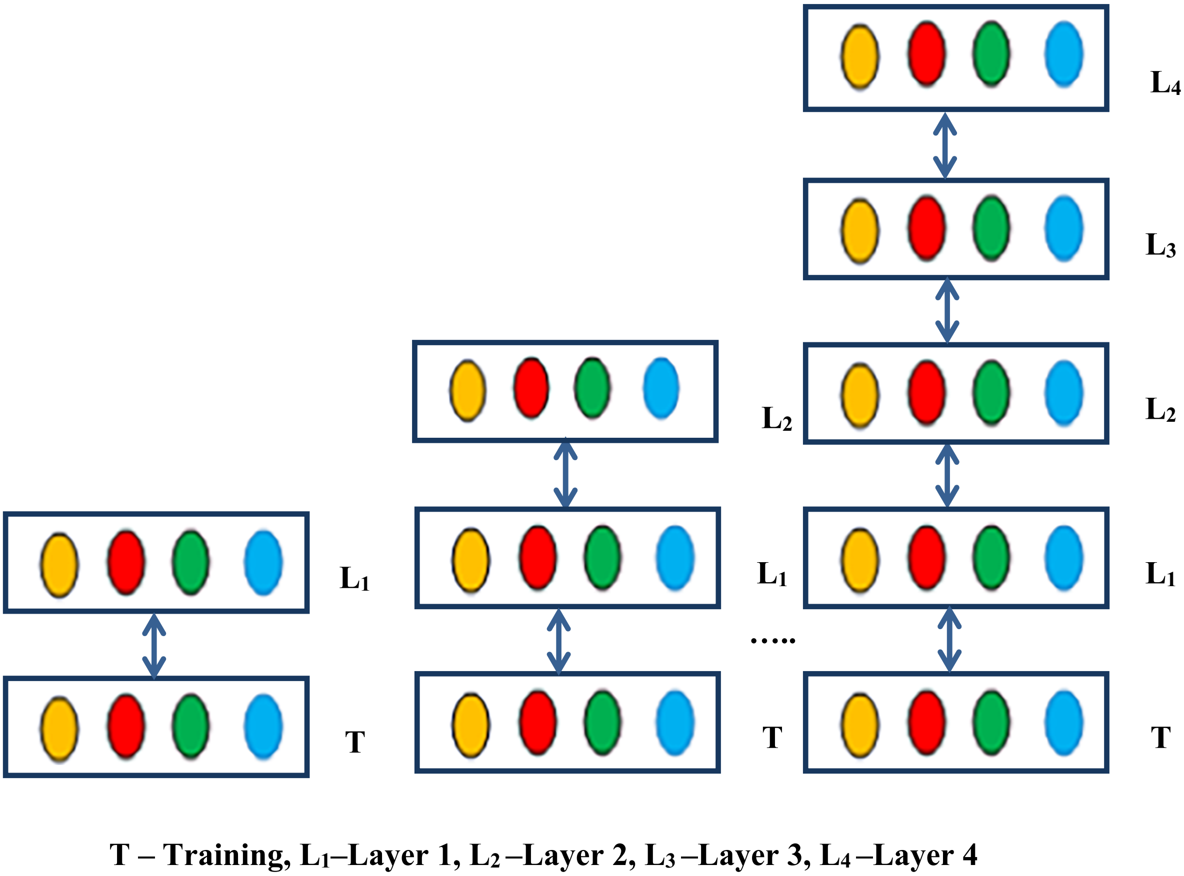


FIGURE 4 Classification of various stages of images using deep belief network

The weight as well as the learning rate of training of X-ray images were performed and is represented in Equation (14).

$$P(wl) = \sum_{i=1}^{n} (w_{ij}x_1 + A_i) + \sum_{j=1}^{n} (w_{ij}y_1 + B_i) \tag{14}$$

Generally, the sample data prediction and classification was performed using supervised and unsupervised learning. The supervised learning was performed using Gradient Descent method and its representation is the shown Equation (15). The supervised learning network was fine tuned from top to bottom. The associated representation of the gradient function with different parameters can be represented as

$$O = W_1A_1B_1 + w_2A_2B_2 + \ldots + W_nA_nB_n \tag{15}$$

Similarly, the unsupervised learning was performed using Boltzmann machine network. It was performed from the bottom to the top layer and matched the features to the maximum possible limit using Equation (16).

$$O = \sum_{i=1}^{n} \left(P\left(\frac{x}{v}\right) \left(\frac{A_nB_n}{w}\right) \right) + \sum_{j=1}^{n} \left(P\left(\frac{y}{v}\right) \left(\frac{A_nB_n}{w}\right) \right) \tag{16}$$

3.2.4 | The working of hybrid approach algorithm

The hybrid approach featured selection was performed using texture and morphological parameters. A huge number of datasets were processed and predicted using VGG-16 and light GBM. The various stages of COVID-19 affected images were classified with the use of DBN. Initially, the collected datasets were noisy and features were predicted with the help of feature selection methods. The texture and morphological features are mentioned in the Equations (1–7). After features selection, the Y_0 to Y_n number of classification and prediction the images were predicted, such as positive (Y_1) or negative (Y_0). The predicted images Y_1 and Y_0 images were classified as per the impact of the COVID infection as per Table 1. The workflow process is presented in Algorithm 1.

For the purpose of continuous predictions, the methods were looped into the input images. As per the looping, the images were examined and the predicted images were sent for classification of various stages of COVID-19 images. These Phases are mild cell incursion and viral duplication of nose and lungs, moderate amount replication which affect the lungs and the immune system, the replication of consolidated sections of the lungs threatening, and critically affected failure of multiple organs respectively.

4 | EXPERIMENTAL RESULTS AND DISCUSSION

4.1 | Datasets and performance parameters

The proposed approach detected and classified the various stages of COVID-19. The implementation of various datasets used has been mentioned in Table 2. All the datasets were classified for training, testing and validations. The 70%, 15% and 15% of the datasets were used for training, testing and validations, respectively. The number of normal and COVID affected images used for training and testing are shown in the Table 4.

In total, 7800 images were utilized in the datasets. The 4400 COVID affected images were used and 3400 normal lung images underwent classifications and predictions. The testing and validation images were not used for training and similarly, the lungs images used for training were not used for testing and validations. The sample predicted images with labels from the huge number of datasets are shown in the Figure 5.

The performance was calculated using various parameters such as accuracy, sensitivity and specificity. These parameters were automatically calculated using other supporting parameters such as true negative, true positive, false negative, as well as false positive, which were used to compute the metrics automatically using the hybrid method. The predicted results were compared with other methods such as the Scratch Model, AlexNet and ResNet50 (Lyu et al., 2018). Compared to the previous models, the proposed model produced better accuracy (Acc), sensitivity (Sen) and specificity (Spc).

Algorithm 1

Input: Various X-ray Images.
Output: i. COVID-19 Predicted images.
ii. Various stages of Predicted Images.
Hybrid Deep Learning Method:
Step 1: Initialize images.
Step 2: Extract each image size (224*224).
Step 3: Extract the texture features [Equations (1)–(5)].
Step 4: Extract the morphological features [Equations (6)–(7)].
Step 5: Initialize the VGG-16 and train the various features.
Step 6: Compute the datasets using GBM.
Step 7: Classify and predict ((Y_1) (Y_0)).
Step 8: Compute the predicted images.
Step 9: Train the various classification phases.
Step 10: Extract the features from the layers.
Step 11: Extract the training and testing feature of phases.
Step 12: Result (O) = Classify (training feature, T).
Step 13: Output: Result (O) = Phases of COVID-19.

TABLE 4 Training, testing and validation ratios of images

Datasets	Normal images	COVID-19 images
Total Images	4400	3400
Training	3080 (70%)	2380 (70%)
Testing	660 (15%)	510 (15%)
Validation	660 (15%)	510 (15%)

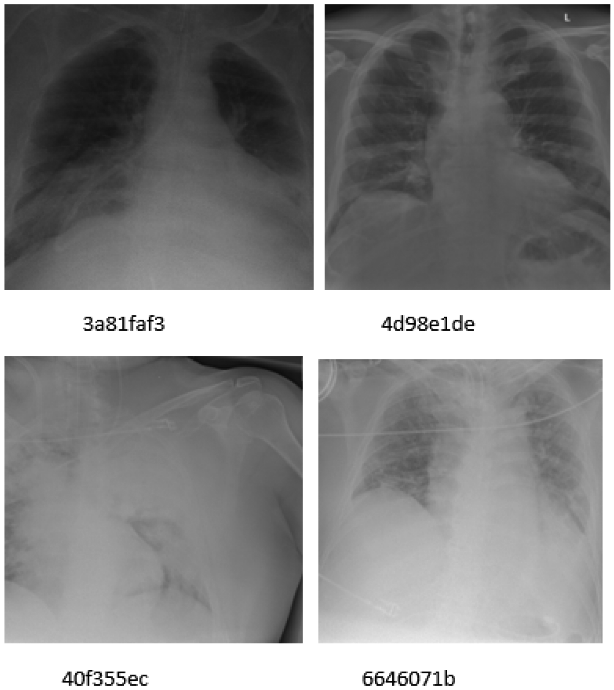


FIGURE 5 COVID predicted images using the hybrid approach

4.2 | Classification and prediction performance

The proposed classification and prediction methods are shown in Figure 6. The proposed hybrid method, the feature extraction, classification and prediction results were received using a combination of VGG-16 and DBN. Each input was trained using the bottom training layer and features were extracted from each layer. Initially, the basic features of the images were extracted from each layer with the help of techniques used for extracting texture and morphological features.

The initial screening and supporting features were seen to produce better accuracy for predictions. Using VGG-16 Net, 0-999 layers were created with each layer having different features for predictions and classifications. This automatically resulted in increased prediction and classification features compared to the other methods. In this proposed hybrid method, a total of 7800 X-ray lung images were utilized for testing, validation as well as training. This automatically resulted in decreased the time complexity of computations. To avoid increased computation time, light GBM was used. The aforementioned methods such as Scratch Model, AlexNet and ResNet50 used lesser percentage of the dataset, and therefore led to lesser computation cost. However, in this proposed hybrid work, light GBM produced overall better classification, accuracy and used less storage capacity in terms of memory and computations.

The presence of Bacterial Pneumonia was used to find the influence of certain bacteria in the affected X-ray images. Figure 7 illustrates the Bacterial Pneumonia prediction rates. Accuracy, sensitivity, as well as specificity parameters were utilized for the purpose of measuring performance. The different labelled and unlabelled features in the images such as smoke, Viral Pneumonia and so on were used to find the influence of the virus. This hybrid method introduced the Bacterial Pneumonia prediction rates, which were not introduced in previous works.

The Figure 8 shows the stage-wise prediction performance and corresponding parameters. The same parameters of accuracy, sensitivity as well as specificity were utilized for the performance calculation and the predicted results were found to be 98.1%, 98.6% and 98.3%, respectively. The VGG-16 layers and BDN provided better classification of the affected images. This classification was used for better decision-making. Table 1 shows the stage-wise identification and sample images of various stages. The DBN layers were used to classify the stages of the COVID images.

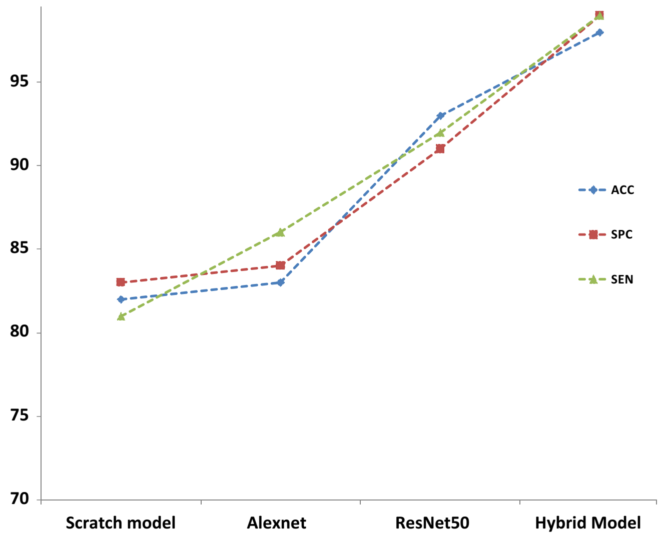


FIGURE 6 Features classification and prediction

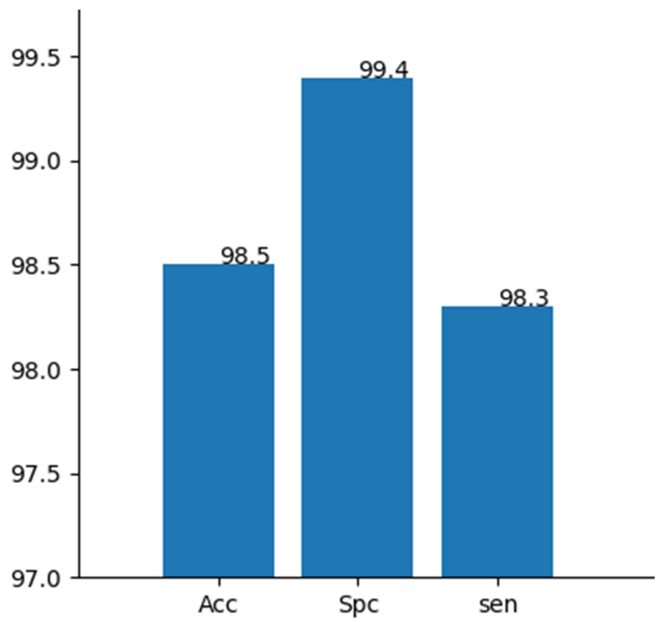


FIGURE 7 Bacterial Pneumonia prediction rates

The combination of 999 layers of VGG-16 and DBN layers produced better classifications of the COVID stages. The number of affected images and corresponding classification of predicted images are shown in Table 5. Similarly, the prediction of total COVID cases, cases of Bacterial Phenomena, the classification of stages and the various overall predictions are shown in Table 6.

5 | CONCLUSION AND FUTURE WORK

COVID-19 has increased fears of infection and mortality rates in families and societies the world over. The healthcare workers are under tremendous pressure to accurately predict the stage of the disease in those affected by it. Therefore, timely predictions play an important role and can save lives. COVID X-ray images are a good resource for determining the status of affected people and can help provide better decision making for treatment protocols. In this work the lung X-ray images were classified into four stages for the purpose of predicting and supporting such decision making. This proposed hybrid method combined various feature extraction methods, prediction and classification methods. The Texture and Morphological methods are used to determine the basic features of the COVID-19 images. The Light GBM and VGG-16 were used for classifying and

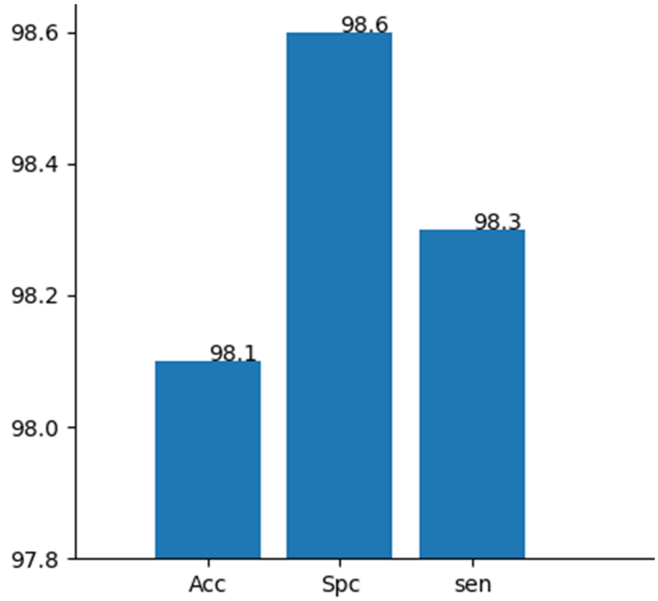


FIGURE 8 Various stages of prediction of COVID-19

TABLE 5 Classification of various stages of COVID images

Total COVID images	Accuracy % of stage wise prediction	Total predicted images	No of Stage-1 COVID images	No of Stage-2 COVID images	No of Stage-3 COVID images	No of Stage-3 COVID images
3400	98.1	3332	1569	1078	613	72

TABLE 6 Various predictions of the proposed method

Predictions	Accuracy	Sensitivity	Specificity
Prediction of COVID-19	99.2	99.4	99.5
Bacterial Pneumonia prediction	98.5	99.4	98.3
Classification of stages	98.1	98.6	98.3
Combined multi-classification prediction	98.6	99.1	98.7

predicting the COVID images. The light GBM helped reduce the computation time and VGG-16 increased the prediction rates. The DBN was used to classify the various predicted images to determine the rate of infection in the lung X-ray images. In this proposed work, a huge number of datasets were used in the implementation. The proposed hybrid method used three parameters for measuring prediction performance such as accuracy, sensitivity and specificity. The COVID-19 images were also predicted in three distinct ways, such as COVID affected and non-affected images, Bacterial influence rates and various stages of the affected images. The overall predicted results were respectively, 98.6%, 99.1% and 98.7% accurate. Therefore, the proposed work was observed to produce improved results in comparison to previous methods, with the stage-wise predictions adding to the novelty of this work. The future direction of this research would benefit from finding ways of further increasing the speed of prediction and reducing the time complexity of the proposed method. Further research can be proposed for the application of multiple prediction models that can lead to increased accuracy of predictions. More fine-tuned prediction-based parameter could be introduced in proposed model in feature to automate the process for prediction. That would be feature span of extension in learning-based research direction. If the more feature extraction specific parameters are included in future, then the proposed algorithm would be modified considerably in all prediction aspects of COVID image. The proposed hybrid algorithm will support future COVID 19 X-ray image predictions even if version changes to be happened. In such a cased some parameters may be included as per future version of COVID 19. The impact of X-ray prediction of COVID 19 is essential in reality since every COVID affected sample images had been appeared different in prediction using proposed algorithm. The proposed algorithm may be extended to automate and predict false positive and false negative predictions in near future.

DATA AVAILABILITY STATEMENT

All the data are fully available in the manuscript.

REFERENCES

Aboutalebi, H., Pavlova, M., Shafiee, M. J., Sabri, A., Alaref, A., & Wong, A. (2021). COVID-net CXR-S: Deep convolutional neural network for severity assessment of COVID-19 cases from chest X-ray images. *arXiv preprint arXiv:2105.00256*.

Ahsan, M., Based, M., Haider, J., & Kowalski, M. (2021). COVID-19 detection from chest X-ray images using feature fusion and deep learning. *Sensors*, *21*(4), 1480.

Al-Waisy, A. S., Al-Fahdawi, S., Mohammed, M. A., Abdulkareem, K. H., Mostafa, S. A., Maashi, M. S., & Garcia-Zapirain, B. (2020). COVID-CheXNet: Hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. *Soft Computing*, 1–16. https://doi.org/10.1007/s00500-020-05424-3

Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked*, *20*, 100391.

Castellano, G., Bonilha, L., Li, L. M., & Cendes, F. (2004). Texture analysis of medical images. *Clinical Radiology*, *59*(12), 1061–1069.

Cheng, S. C., Chang, Y. C., Chiang, Y. L. F., Chien, Y. C., Cheng, M., Yang, C. H., & Hsu, Y. N. (2020). First case of coronavirus disease 2019 (COVID-19) pneumonia in Taiwan. *Journal of the Formosan Medical Association*, *119*(3), 747–751.

Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). COVID-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988*.

COVID, C. A. A. (2020). (19). chest X-ray data initiative.

Dubey, R. K. (2020). Deep learning Based hybrid models for prediction of COVID-19 using chest X-ray.

Eren, L., Ince, T., & Kiranyaz, S. (2019). A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *Journal of Signal Processing Systems*, *91*(2), 179–189.

Esgiar, A. N., Naguib, R. N., Sharif, B. S., Bennett, M. K., & Murray, A. (2002). Fractal analysis in the detection of colonic cancer images. *IEEE Transactions on Information Technology in Biomedicine*, *6*(1), 54–58.

Fouladi, S., Ebadi, M. J., Safaei, A. A., Bajuri, M. Y., & Ahmadian, A. (2021). Efficient deep neural networks for classification of COVID-19 based on CT images: Virtualization via software defined radio. *Computer Communications*, *176*, 234–248.

Gupta, V., Jain, N., Katariya, P., Kumar, A., Mohan, S., Ahmadian, A., & Ferrara, M. (2021). An emotion care model using multimodal textual analysis on COVID-19. *Chaos, Solitons & Fractals*, *144*, 110708.

Guru, D. S., Sharath, Y. H., & Manjunath, S. (2010). Texture features and KNN in classification of flower images. *IJCA*, *1*, 21–29.

Haider, M. A., Vosough, A., Khalvati, F., Kiss, A., Ganeshan, B., & Bjarnason, G. A. (2017). CT texture analysis: A potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer Imaging*, *17*(1), 1–9.

Hirra, I., Ahmad, M., Hussain, A., Ashraf, M. U., Saeed, I. A., Qadri, S. F., & Alfakeeh, A. S. (2021). Breast cancer classification from histopathological images using patch-Based deep learning modeling. *IEEE Access*, *9*, 24273–24287.

https://github.com/agchung/Actualmed-COVID-chestxray-dataset (2020).

https://github.com/GeneralBlockchain/covid-19-chest-xray-lung-bounding-boxes-dataset.

https://github.com/ieee8023/covid-chestxray-dataset

https://www.kaggle.com/c/dlai3/data

https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

https://www.sciencedaily.com/releases/2020/04/200423130420.htm.

Hussain, L., Nguyen, T., Li, H., Abbasi, A. A., Lone, K. J., Zhao, Z., & Duong, T. Q. (2020). Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection. *Biomedical Engineering Online*, *19*(1), 1–18.

Jain, R., Gupta, M., Taneja, S., & Hemanth, D. J. (2021). Deep learning based detection and analysis of COVID-19 on chest X-ray images. *Applied Intelligence*, *51*(3), 1690–1700.

Khalvati, F., Wong, A., & Haider, M. A. (2015). Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC Medical Imaging*, *15*(1), 1–14.

Khuzi, A. M., Besar, R., & Zaki, W. W. (2008). Texture features selection for masses detection in digital mammogram. In *4th Kuala Lumpur international conference on biomedical engineering 2008* (pp. 629–632). Springer, Berlin, Heidelberg.

Kundu, R., Singh, P. K., Ferrara, M., Ahmadian, A., & Sarkar, R. (2021). ET-NET: An ensemble of transfer learning models for prediction of COVID-19 infection through chest CT-scan images. *Multimedia Tools and Applications*, 1–20. https://doi.org/10.1007/s11042-021-11319-8

Luz, E., Silva, P., Silva, R., Silva, L., Guimarães, J., Miozzo, G., & Menotti, D. (2021). Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. *Research on Biomedical Engineering*, 1–14. https://doi.org/10.1007/s42600-021-00151-6

Lyu, Y., Chen, J., & Song, Z. (2018). Image-based process monitoring using deep belief networks. *IFAC-PapersOnLine*, *51*(18), 115–120.

Mahdy, L. N., Ezzat, K. A., Elmousalami, H. H., Ella, H. A., & Hassanien, A. E. (2020). Automatic x-ray COVID-19 lung image classification system based on multi-level thresholding and support vector machine. MedRxiv.

Mengistu, A. D. (2018). The effects of segmentation techniques in digital image based identification of Ethiopian coffee variety. *Telkomnika*, *16*(2), 713–717.

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical Image Analysis*, *65*, 101794.

Mohan, S., Abugabah, A., Kumar Singh, S., Kashif Bashir, A., & Sanzogni, L. (2021). An approach to forecast impact of COVID-19 using supervised machine learning model. *Software: Practice and Experience*, 1–17. https://doi.org/10.1002/spe.2969

Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R., & Singh, V. (2020). Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet. *Chaos, Solitons & Fractals*, *138*, 109944.

Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., & Gloaguen, R. (2020). COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics*, *8*(6), 890.

Sahlol, A. T., Yousri, D., Ewees, A. A., Al-Qaness, M. A., Damasevicius, R., & Abd Elaziz, M. (2020). COVID-19 image classification using deep features and fractional-order marine predators algorithm. *Scientific Reports*, *10*(1), 1–15.

Sharma, A., Rani, S., & Gupta, D. (2020). Artificial intelligence-based classification of chest X-ray images into COVID-19 and other infectious diseases. *International Journal of Biomedical Imaging*, 1–10. https://doi.org/10.1155/2020/8889023

Shuai, J., Shen, C., & Zhu, Z. (2017). Adaptive morphological feature extraction and support vector regressive classification for bearing fault diagnosis. *International Journal of Rotating Machinery*. https://doi.org/10.1155/2017/2384184

Shuja, J., Alanazi, E., Alasmary, W., & Alashaikh, A. (2021). COVID-19 open source data sets: A comprehensive survey. *Applied Intelligence*, *51*(3), 1296–1325.

Singh, N., Suthar, B., Mehta, A., Nema, N., & Pandey, A. (2020). Corona virus: An immunological perspective review. *International Journal of Immunology and Immunotherapy*, *7*, 50. https://doi.org/10.23937/2378-672/1410050

Sitaula, C., & Hossain, M. B. (2021). Attention-based VGG-16 model for COVID-19 chest X-ray image classification. *Applied Intelligence*, *51*(5), 2850–2863.

Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *Scientific Reports*, *10*(1), 1–12.

Yu, H., Scalera, J., Khalid, M., Touret, A. S., Bloch, N., Li, B., & Anderson, S. W. (2017). Texture analysis as a radiomic marker for differentiating renal tumors. *Abdominal Radiology*, *42*(10), 2470–2478.

How to cite this article: M., A., Govindharaju, K., A., J., Mohan, S., Ahmadian, A., & Ciano, T. (2022). A hybrid learning approach for the stage-wise classification and prediction of COVID-19 X-ray images. *Expert Systems*, *39*(4), e12884. https://doi.org/10.1111/exsy.12884