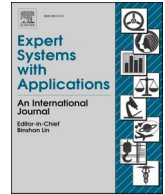




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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



Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble

Tej Bahadur Chandra^{a,*}, Kesari Verma^a, Bikesh Kumar Singh^b, Deepak Jain^c,
Satyabhuwan Singh Netam^c

^a Department of Computer Applications, National Institute of Technology Raipur, Chhattisgarh, India

^b Department of Biomedical Engineering, National Institute of Technology Raipur, Chhattisgarh, India

^c Department of Radiodiagnosis, Pt. Jawahar Lal Nehru Memorial Medical College, Raipur, Chhattisgarh, India

ARTICLE INFO

Keywords:

Coronavirus
Chest X-Ray
nCOVID-19
Pneumonia
SARS-COV-2
Contagious
Pandemic

ABSTRACT

Novel coronavirus disease (nCOVID-19) is the most challenging problem for the world. The disease is caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2), leading to high morbidity and mortality worldwide. The study reveals that infected patients exhibit distinct radiographic visual characteristics along with fever, dry cough, fatigue, dyspnea, etc. Chest X-Ray (CXR) is one of the important, non-invasive clinical adjuncts that play an essential role in the detection of such visual responses associated with SARS-COV-2 infection. However, the limited availability of expert radiologists to interpret the CXR images and subtle appearance of disease radiographic responses remains the biggest bottlenecks in manual diagnosis. In this study, we present an automatic COVID screening (ACoS) system that uses radiomic texture descriptors extracted from CXR images to identify the normal, suspected, and nCOVID-19 infected patients. The proposed system uses two-phase classification approach (normal vs. abnormal and nCOVID-19 vs. pneumonia) using majority vote based classifier ensemble of five benchmark supervised classification algorithms. The training-testing and validation of the ACoS system are performed using 2088 (696 normal, 696 pneumonia and 696 nCOVID-19) and 258 (86 images of each category) CXR images, respectively. The obtained validation results for phase-I (accuracy (ACC) = 98.062%, area under curve (AUC) = 0.956) and phase-II (ACC = 91.329% and AUC = 0.831) show the promising performance of the proposed system. Further, the Friedman post-hoc multiple comparisons and z-test statistics reveals that the results of ACoS system are statistically significant. Finally, the obtained performance is compared with the existing state-of-the-art methods.

1. Introduction

The recent outbreak of the novel coronavirus disease (nCOVID-19) has infected millions of people and killed several individuals across the world ("Coronavirus Disease 2019, 2020"; "Johns Hopkins University, Corona Resource Center, 2020"). The World Health Organization (WHO) has declared this epidemic a global health emergency. nCOVID-19 is caused by a highly contagious virus named severe acute respiratory syndrome coronavirus-2 (SARS-COV-2) in which transmission of infection can even occur from the asymptomatic patients during the incubation period (Huang et al., 2020; Kooraki, Hosseiny, Myers, & Gholamreza-zehad, 2020). As per the expert's opinion, the virus mainly infects the human respiratory tract leading to severe bronchopneumonia with symptoms of fever, dyspnea, dry cough, fatigue, and respiratory failure,

etc. (N.Chen et al., 2020; Cheng et al., 2020). There is no specific vaccine or medication available to cure the disease and prevent further spread. Also, the standard confirmatory clinical test – reverse transcription-polymerase chain reaction (RT-PCR) test for detecting nCOVID-19 is manual, complex, and time-consuming (Chowdhury, Rahman, Khandakar, Mazhar, Kadir, Mahbub, & Reaz, 2020). The limited availability of test-kits and domain experts in the hospitals and rapid increase in the number of infected patients necessitates an automatic screening system, which can act as a second opinion for expert physicians to quickly identify the infected patients, who require immediate isolation and further clinical confirmation.

Chest X-Ray (CXR) is one of the important, non-invasive clinical adjuncts that play an essential role in the preliminary investigation of different pulmonary abnormalities (Chandra & Verma, 2020, 2020a;

* Corresponding author at: Department of Computer Applications, National Institute of Technology Raipur, Chhattisgarh 492010, India.

E-mail addresses: tejbahadur1990@gmail.com (T.B. Chandra), kverma.mca@nitrr.ac.in (K. Verma), bsingh.bme@nitrr.ac.in (B.K. Singh).

<https://doi.org/10.1016/j.eswa.2020.113909>

Received 6 May 2020; Received in revised form 9 August 2020; Accepted 19 August 2020

Available online 26 August 2020

0957-4174/© 2020 Elsevier Ltd. All rights reserved.

Chandra, Verma, Jain, & Netam, 2020; Ke et al., 2019). It can act as an alternative screening modality for the detection of nCOVID-19 or to validate the related diagnosis, where the CXR images are interpreted by expert radiologists to look for infectious lesions associated with nCOVID-19. The earlier studies reveal that the infected patients exhibit distinct visual characteristics in CXR images, as shown in Fig. 1 (Cheng et al., 2020; Chowdhury et al., 2020; Chung et al., 2020; Zhang et al., 2020). These characteristics typically include multi-focal, bilateral ground-glass opacities and patchy reticular (or reticulonodular) opacities in non-ICU patients, while dense pulmonary consolidations in ICU patients (Hosseiny, Kooraki, Gholamrezanezhad, Reddy, & Myers, 2020). However, the manual interpretation of these subtle visual characteristics on CXR images is challenging and require domain expert (Kanne, Little, Chung, Elicker, & Ketai, 2020; L. Wang & Wong, 2020). Moreover, the exponential increase in the number of infected patients makes it difficult for the radiologist to complete the diagnosis in time, leading to high morbidity and mortality (AsnaouiEl Chawki & Idri, 2020).

To fight against nCOVID-19 epidemic, the recent machine learning (ML) techniques can be embedded to develop an automatic computer-aided diagnosis (CAD) system. In this direction, many clinical and

radiological studies have been reported, describing various radio-imaging findings and epidemiology of nCOVID-19 (N. Chen et al., 2020; Huang et al., 2020; Kooraki et al., 2020; Yoon et al., 2020). Further, many deep-learning models like deep convolutional network, recursive network, transfer learning models, etc. have been implemented to automatically analyze the radiological disease characteristics (Chouhan et al., 2020; Jaiswal et al., 2019). Xue et al. (2018) used a convolutional neural network (CNN) to assign a class label to different superpixels extracted from the lungs parenchyma and localize tuberculosis-infected regions in CXR images with average dice index of 0.67. Another work by Pesce et al. (2019), used two novel models, the first model is based on backpropagation neural network that uses weakly labeled CXR images and generates visual attention feedback for accurate localization of pulmonary lesions; the second model used reinforcement learning-based recurrent attention model, which learns the sequence of images to find the nodules. Recently, Purkayastha, Buddi, Nuthakki, Yadav, and Gichoya (2020) introduced CheXNet—deep learning (DL) based model integrated with LibreHealth Radiology Information System, which analyzes uploaded CXR images and assigns one of 14 diagnostic labels.

Motivated by the promising performance of DL models reported in

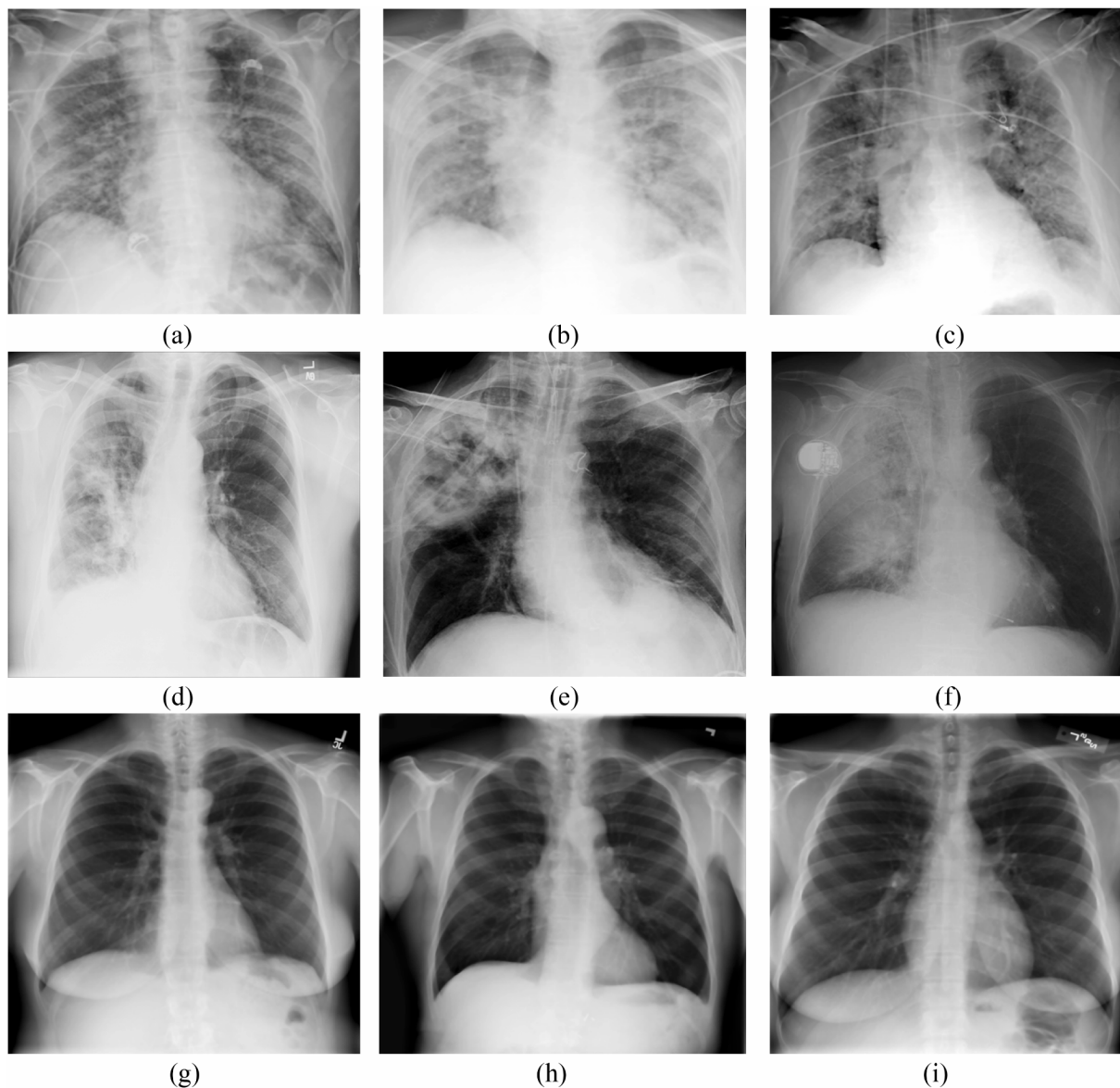


Fig. 1. (a–c) nCOVID-19 infected chest X-Ray images (d–f) Pneumonia infected chest X-Ray (h–i) Normal chest X-Ray images.

the literature and urgent need of an alternate screening tool for early detection of nCOVID-19 infected patients, the research community has applied different DL techniques on chest radiograph images (Abbas, Abdelsamea, & Gaber, 2020; S. Wang et al., 2020; Xu, Jiang, Ma, Du, Li, Lv, & Wu, 2020). The detailed description of different state-of-arts methods, including the imaging modality, dataset size, algorithms, and obtained performance, are recapitulated in Table 1. Initially, the authors have used a mixture of CXR images collected from different hospitals, publications, and older repositories (Abbas, Abdelsamea, & Gaber, 2020; Hemdan, Shouman, & Karar, 2020; Narin et al., 2020). However, the limited availability of annotated CXR images for nCOVID-19 cases to train the data-hungry DL models turned out to be the biggest bottleneck (X. Wang et al., 2017). Latter, to avoid the overfitting of the models, the studies used data augmentation techniques, which generates different variants of the source image by applying random photometric transformations like blurring, sharpening, contrast adjustment, etc. (Chowdhury et al., 2020; S. Wang et al., 2020; Xu et al., 2020). Further, the CT images have also been used to perform in-depth volumetric analysis of subtle disease responses (similar to viral pneumonia or other inflammatory lung diseases) (Maghdid, Asaad, Ghafoor, Sadiq, & Khan, 2020; S. Wang et al., 2020; Xu et al., 2020).

After the retrospective analysis of the above literatures, we found that the existing studies had been performed using a limited number of input CXR or CT images, which may lead to under-fitting of the data-hungry DL models (X. Wang et al., 2017). Moreover, the DL approach requires huge computational resources along with a large number of accurately annotated CXR images to train the model, which restrain its clinical acceptability (Altaf, Islam, Akhtar, & Janjua, 2019; Ho & Gwak, 2019). Conventional ML techniques can be better integrated with CAD systems to overcome these shortcomings. Despite several studies, no one has used conventional ML approaches with ensemble learning using majority voting for the classification of normal and nCOVID-19 infected CXR images.

In this study, we tailored an automatic COVID screening (ACoS) system that employs hierarchical classification using conventional ML algorithms and radiomic texture descriptors to segregate normal, pneumonia, and nCOVID-19 infected patients. The major advantage of the proposed system is that it can be easily modeled using the limited number of annotated images and can be deployed even in a resource-constrained environment.

1.1. Contribution and Organization of paper

The contributions of this study are recapitulated as follows:

- Proposed an ACoS system for detection of nCOVID-19 infected patients using hierarchical classification and augmented images. The proposed model can be used as a retrospective tool or to validate the related diagnosis.
- Applied majority vote based classifier ensemble to aggregate the prediction results of five supervised classification algorithms.
- Review and compare the performance of the proposed ACOS system with the state of the art methods.

The remaining sections of the paper is organized as follows. Section 2 describes the materials and methods used in this study. The obtained results and its detailed analysis are discussed in Section 3. The paper is concluded in Section 4.

2. Materials and methods

2.1. Data acquisition, preprocessing, and augmentation

In this study, we have used dataset from three public repositories – COVID-Chestxray set (Cohen, Morrison, & Dao, 2020), Montgomery set (Candemir et al., 2014; Jaeger et al., 2014), and NIH ChestX-ray14 set

(X. Wang et al., 2017). The detailed statistics of the number of posterior-anterior (PA) view CXR images used from each repository are shown in Table 2. Initially, all input images are preprocessed, which includes image resizing (512×512 pixels), format conversion (Portable Network Graphics), and color space conversion (Gray Scale). Subsequently, the texture preserving guided filter is applied to reduce the inherent quantum noise (Sprawls, 2018). The choice of the de-noising filter is based on our previous study (Chandra & Verma, 2020a).

The preprocessed images are divided into two sub-sets: training-testing set (80%) and validation set (20%). Further, the image augmentation technique is applied to the images of the training-testing set to build a generalized model by incorporating the possible variability in the images, which might occur due to diverse imaging conditions. We applied different random photometric transformations with random parameters between the specified ranges, as described in Table 3.

2.2. Feature extraction and feature selection

The nCOVID-19 infected patients exhibit different radiographic texture patterns such as patchy ground-glass opacities (Fig. 2a), pulmonary consolidations (Fig. 2c), reticulonodular opacities (Fig. 2b), etc. on CXR images (Hosseiny et al., 2020). These subtle visual characteristics can be efficiently represented with the help of radiomic texture descriptors. The study uses eight first order statistical features (FOSF) (Srinivasan & Shobha, 2008), 88 grey level co-occurrence matrix (GLCM) (Gómez, Pereira, & Infantosi, 2012; Haralick, Shanmugam, & Dinstein, 1973) features (in four different orientations) and 8100 histogram of oriented gradients (HOG) (Dalal & Triggs, 2005; Santosh & Antani, 2018) features. The FOSF describes the complete image at a glance by using the mean, variance, roughness, smoothness, kurtosis, energy, and entropy, etc. It can easily quantify the global texture patterns; however, it does not contemplate the local neighborhood information. To overcome this shortcoming, the GLCM and HOG feature descriptor are used to perform the in-depth texture analysis. The GLCM feature describes the spatial correlation among the pixel intensities in radiographic texture patterns along four distinct directions (i.e., 0° , 45° , 90° , 135°) whereas the HOG feature encodes the local shape/texture information. The selection of these statistical texture features is motivated by the fact that it can efficiently encode the natural texture patterns and is widely used in medical image analysis (Chandra & Verma, 2020a; Chandra, Verma, Singh, Jain, & Netam, 2020; Santosh & Antani, 2018; Vajda et al., 2018).

In this study, a total of 8196 features (8 FOSF, 88 GLCM, 8100 HOG) are extracted from each CXR image (described in Appendix-A). However, not all the extracted features are relevant for accurate characterization of visual indicators associated with nCOVID-19. Thus, to select the most informative features, we used a recently developed meta-heuristic approach called—binary grey wolf optimization (BGWO) (Mirjalili, Mirjalili, & Lewis, 2014; Too, Abdullah, Mohd Saad, Mohd Ali, & Tee, 2018). The method imitates the leadership, encircling, and hunting strategy of grey wolves. Unlike the other evolutionary algorithms, the method does not get trapped in local minima, which motivated us to use it in our study (Emary, Zawbaa, & Hassanien, 2016).

Mathematically, the grey wolves are divided into four categories denoted by alpha (α), beta (β), delta (δ), and omega (ω). The α -wolf is the decision-maker and administers the hunting process with the help of beta. The β -wolves are the fittest candidate to replace the alpha when the alpha is very old or dead. The δ -wolves are the next in the hierarchy, which obey the orders from α and β -wolves but command omega wolves. The ω -wolves are the lowest in the hierarchy and report to these leader wolves. The encircling strategy of the wolves is described in Eq. (1).

$$X(t+1) = X_p(t) - A \cdot D \quad (1)$$

$$\text{where } D = |C \cdot X_p(t) - X(t)| \quad (2)$$

Table 1

Existing literature for detection of nCOVID-19 using DL approaches. (Abbreviations: CXR: Chest X-Ray, CT: Computed Tomography, nCOVID: Novel Coronavirus Disease, AUC: Area Under Curve, CNN: Convolutional Neural Network, DT: Decision Tree, SVM: Support Vector Machine, KNN: k-Nearest Neighbor, VGG: Visual Geometry Group).

Articles	Imaging Modality	Dataset Size				Class	Algorithms/Techniques	Performance	
		nCOVID-19	Pneumonia	Normal	Augmented			ACC (%)	AUC
Ozturk et al. (2020)	CXR	127	500	500	–	3	DarkNet	87.02	–
Xu et al. (2020)	CT	219	224	175	2634 patches nCOVID, 2661 patches Pneumonia, 6576 patches Normal	3	ResNet-18	86.70	–
Panwar et al. (2020)	CXR	142	–	142	–	2	nCOVnet	88.10	0.88
S. Wang et al. (2020)	CT	79	180	–	1065 (740 Negative, 325 nCOVID)	2	Deep Learning	89.50	–
Hemdan, Shouman, and Karar (2020)	CXR	25	–	25	–	2	VGG19, DenseNet201, ResNetV2, InceptionV3, InceptionResNetV2, Xception, MobileNetV2	90.00	–
Pathak, Shukla, Tiwari, Stalin, and Singh (2020)	CT	413	439	–	–	2	ResNet-50	93.02	0.93
L. Wang et al. (2020)	CXR	385	5538	8066	–	3	COVID-Net	93.30	–
Maghdid, Asaad, Ghafoor, Sadiq, and Khan (2020)	CXR	85 CXR	–	85 CXR	–	2	AlexNet, Modified CNN	94.00	–
	CT	203 CT	–	158 CT	–			82.0	–
								CT	–
Abbas, Abdelsamea, and Gaber (2020)	CXR	105	11	80	–	3	DeTraC	95.12	–
AsnaouiEl Chawki and Idri (2020)	CXR	–	4273	1583	–	2	CNN, VGG16 VGG19, Inception_V3, Xception, DensNet201, MobileNet_V2, Inception_Resnet_V2, Resnet50	96.61	–
Chowdhury et al. (2020)	CXR	423	1485	1579	6540	3	AlexNet, ResNet18, DenseNet201, SqueezeNet	97.94	–
Narin et al. (2020)	CXR	50	–	50	–	2	ResNet50, ResNetV2, InceptionV3	98.00	–
Ucar and Korkmaz (2020)	CXR	66	3895	1349	4608	3	Deep Bayes-SqueezeNe	98.26	–
Nour, Cömert, and Polat (2020)	CXR	219	1345	1341	765	3	CNN, SVM, DT, KNN	98.97	–
Ardakani, Kanafi, Acharya, Khadem, and Mohammadi (2020)	CT	510	510	–	–		VGG-16, ResNet-18, ResNet-101, AlexNet, VGG-19, Xception, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-50	99.51	0.99
Togaçar et al. (2020)	CXR	295	98	65	–	3	MobileNetV2, SqueezeNet, SVM	99.27	1.00

Table 2

Statistics of the number of CXR images used from different repositories for performance evaluation in training, testing, and validation set.

Dataset property	COVID-Chestxray Set	Montgomery Set	NIH ChestX-ray14 Set	Augmented images	Training -Testing set (80%)	Validation set (20%)
Number of nCOVID-19 CXR images	434	–	–	348	696	86
Number of Pneumonia X-ray images	89	–	345	348	696	86
Number of normal X-ray images	19	80	335	348	696	86
Total Number of X-ray images	542	80	680	1044	2088	258

Table 3

Image augmentation using various photometric transformations.

Transformations	Range	
Sharpening	Automatic	Highlight the fine details by adjusting the contrast between bright and dark pixels.
Gaussian Blur	0.1 to 1.5	Random smoothing of texture information between the specified range of sigma.
Brightness	–20 to 20	Randomly increase or decrease the pixel's intensity between the given range.
Contrast adjustment	Automatic	Adjust the contrast of the image.

where, X_p and X_g denotes the position of the prey and grey wolf in 't' iteration, respectively. The A and C are the coefficient vectors computed using equations Eqs. (3) and (4), respectively.

$$A = 2a.r_1 - a \tag{3}$$

$$C = 2r_2 \tag{4}$$

where, r_1, r_2 denotes the two random numbers between 0 and 1, and 'a' denotes the linearly decreasing encircling coefficient (from 2 to 0) used to balance the tradeoff between searching and exploitation.

Further, the optimal position of the wolfs ($X_\alpha, X_\beta,$ and X_δ) at iteration 't' are updated using Eqs. (5)–(8); $A_1, A_2,$ and A_3 are computed using Eq. (3); and $D_\alpha, D_\beta,$ and D_δ are calculated using Eq. (2).

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{5}$$

$$X_1 = |X_\alpha - A_1.D_\alpha| \tag{6}$$

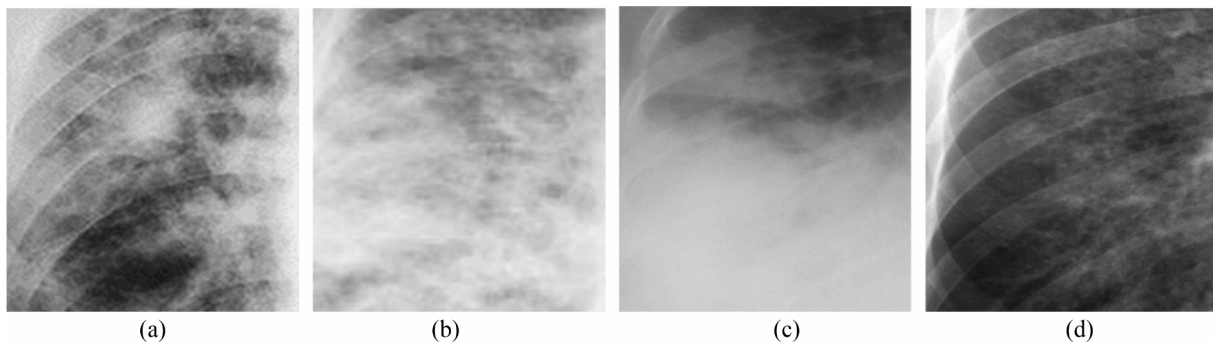


Fig. 2. nCOVID-19 infected Chest X-ray images showing: (a) Ground-glass opacities, (b) Reticular opacities, (c) Pulmonary consolidation, (d) Mild opacities.

$$X_2 = |X_\beta - A_1 \cdot D_\beta| \quad (7)$$

$$X_3 = |X_\delta - A_1 \cdot D_\delta| \quad (8)$$

2.3. Proposed methodology

To develop a robust ACoS system, we hypothesized the following:

Hypothesis 1. The image augmentation technique could improve the robustness of the ACoS system by incorporating variability in input CXR images, which might occur due to diverse imaging modality, exposure time, radiation dose, and varying patient's posture.

Hypothesis 2. The solid mathematical foundation and better generalization capability of support vector machine (SVM) could uncover the subtle radiological characteristics associated with nCOVID-19.

Hypothesis 3. The majority voting based classifier ensemble could act as a multi-expert recommendation and reduce the probable chance of false diagnosis.

To evaluate the hypothetical assumptions, we proposed a prototype (ACoS system) model, as shown in Fig. 3. The proposed system consists of five major steps: pre-processing, image augmentation, feature extraction, classification, and performance evaluation. Initially, to examine Hypothesis 1, the input CXR images are preprocessed (resize, de-noise), and the image augmentation technique is applied (described in Section 2.1). Subsequently, the radiomic textures descriptors are extracted from the complete CXR image and binary gray wolf optimization (BGWO) (Emary et al., 2016; Mirjalili et al., 2014) based feature selection technique is applied to pick the most relevant features. Further to examine Hypothesis 2, the selected features are used to train the model using five supervised classification algorithms, namely – decision tree (DT) (Shalev-Shwartz & Ben-David, 2014), support vector machine (SVM) (Vapnik, 1998), k -nearest neighbor (KNN) (Han, Kamber, & Pei, 2012), naïve Bayes (NB) (Rish et al., 2001) and artificial neural network (ANN) (Artificial Neural Network, 2013). The proposed methodology uses two-phase classification approach. In phase-I, the normal and abnormal (containing nCOVID-19 and Pneumonia) images are segregated. Subsequently, the abnormal images are further classified in phase-II to segregate the nCOVID-19 and pneumonia. Moreover, the fully trained model is validated using a separate validation set. The final prediction of the validation set is the majority vote of seven benchmark classifiers (ANN, KNN, NB, DT, SVM (linear kernel), SVM (radial basis function (RBF) kernel), and SVM (polynomial kernel)), which reduce the probable chance of misclassification (Hypothesis 3). Finally, the performance measures are evaluated for testing and validation sets. All the experiments in this study are implemented using MATLAB R2018a.¹

2.4. Classification

To compute the discriminative performance of the aforementioned features we have used the popular supervised classification algorithms: SVM (linear, radial bias function, polynomial) (Chandra & Verma, 2020a; Vapnik, 1998), ANN (Artificial Neural Network, 2013), KNN (Han, Kamber, & Pei, 2012), NB (Khatami, Khosravi, Nguyen, Lim, & Nahavandi, 2017; Venegas-Barrera & Manjarrez, 2011), and DT (Han, Kamber, & Pei, 2012; Pantazi, Moshou, & Bochtis, 2020). These algorithms are very fast and are widely used in the literature for the classification of pulmonary diseases using CXR images (Chandra & Verma, 2020a; Santosh & Antani, 2018). The selection of these classifiers is motivated by the fact that these algorithms can be efficiently trained using smaller datasets without compromising with the performance. In this study, a discrete set of models were created for phase-I and phase-II, respectively. In phase-I, the models were trained using normal and abnormal images (containing nCOVID-19 and pneumonia) from training –testing set. However, in phase-II, only abnormal images (containing nCOVID-19 and pneumonia) were used to train the models. In both the phases, the performance of the classifiers was evaluated using a 10 fold cross-validation setup. In each fold, all the optimizable learning hyper-parameters were tuned using the Bayesian automatic optimization method (Snoek, Larochelle, & Adams, 2012).

In general, the learning hyper-parameters can be optimized in two ways, called manual and automatic searching. The manual parameter tuning requires expertise. However, when dealing with numerous models and larger datasets, even expertise may not be sufficient (Ucar & Korkmaz, 2020). To overcome this shortcoming, an automatic parameter tuning is used as an alternative. In this study, grid search algorithm is used to select the best hyper-parameters by minimizing the cross-validation loss automatically.²

Moreover, to examine the Hypothesis 3, majority vote based classifier ensemble technique (described in Appendix B, Algorithm 1) is applied (shown in Fig. 3) using a separate validation set (258 CXR images). In order to select the optimal combination of evaluated classifiers for majority vote, we implemented an exhaustive search using recursive elimination method (Chatterjee, Dey, & Munshi, 2019; Q. Chen, Meng, & Su, 2020). Initially, the method starts with all evaluated classifiers, according to the selection criteria, it iteratively eliminates the classifiers until all possible combinations exhausted.

2.5. Performance evaluation metrics

The performance of the proposed ACoS system is assessed using seven performance measures, as shown in Eqs. (9)–(15) (Han, Kamber, & Pei, 2012), where, the number of infected and normal CXR images correctly predicted by the proposed system is denoted by true positive

¹ <https://www.mathworks.com/>.

² <https://www.mathworks.com/help/stats/bayesian-optimization-workflow.html>.

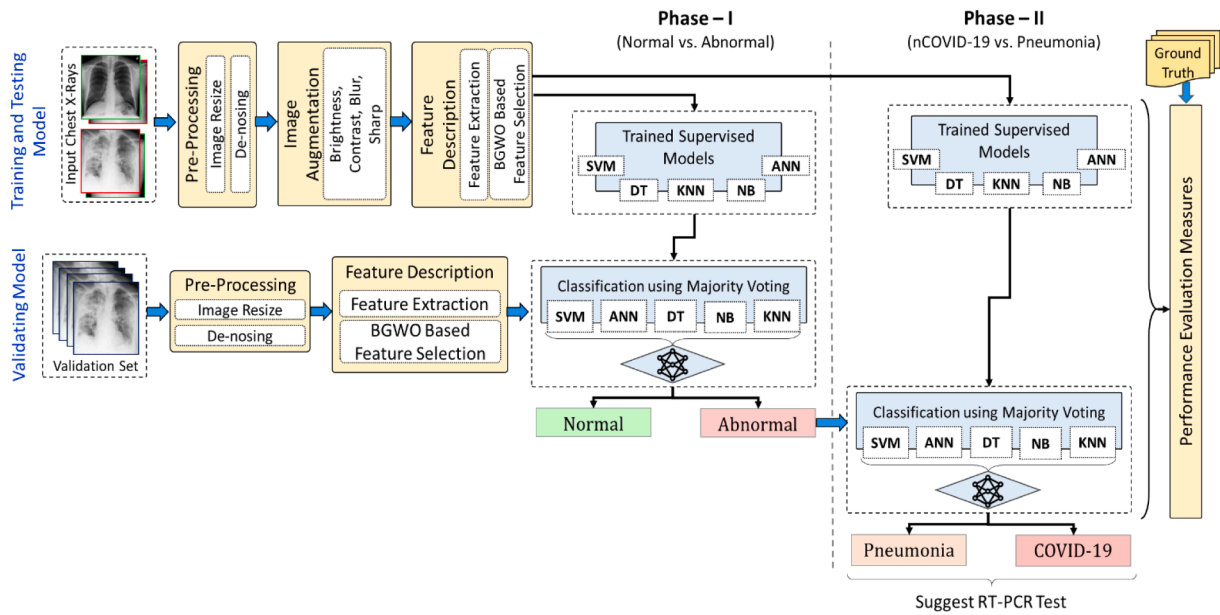


Fig. 3. The prototype of the proposed automatic COVID screening (ACoS) system. (Abbreviations: BGWO: Binary Gray Wolf Optimization, SVM: Support Vector Machine, DT: Decision Tree, KNN: k -Nearest Neighbor, NB: Naïve Bayes, ANN: Artificial Neural Network).

(TP) and true negative (TN), respectively; the false positive (FP) and false-negative (FN) denotes the misclassification of normal and infected images, respectively; $P = TP + FN$ and $N = TN + FP$.

$$\text{Accuracy (ACC)} = \frac{TP + TN}{P + N} \times 100 \quad (9)$$

$$\text{Specificity} = \frac{TN}{N} \times 100 \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (11)$$

$$\text{Recall} = \frac{TP}{P} \times 100 \quad (12)$$

$$F_1 - \text{Measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100 \quad (13)$$

$$\text{Area Under Curve (AUC)} = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right) \quad (14)$$

$$\begin{aligned} \text{Matthews Correlation Coefficient (MCC)} \\ = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times P \times N \times (TN + FN)}} \end{aligned} \quad (15)$$

Finally, the obtained results is statistically validated using z-test and Friedman average ranking and Holm (Holm, 1979) and Shaffer (Shaffer, 1986) post-hoc multiple comparison methods.

3. Experimental results and discussion

This section presents a detailed discussion of the obtained experimental results of the proposed ACoS system. To evaluate the hypothetical assumptions (Hypothesis 1–3), the following experiments were formulated:

Experiment 1—Different photometric transformations were randomly applied to input CXR images, and classification performance was evaluated (Hypothesis 1).

Experiment 2—The classification performance of SVM was assessed and compared with the other benchmark classifiers (Hypothesis 2).

Experiment 3—The classification performance of the majority voting

technique and individual benchmark classifiers are evaluated using a separate validation set (Hypothesis 3).

In this study, we used two-phase classification technique to discriminate the normal, nCOVID-19 and pneumonia X-ray images. Initially, two sets of classification models were created for phase-I and phase-II, respectively using original CXR images from the training-testing set. Subsequently, the image augmentation was performed using different photometric transformations as discussed in Section 2.1. The augmented images along with the original CXR images were used to re-train the models and classification performance was evaluated. From the obtained results shown in Tables 4 and 5, it was observed that the supervised models trained using augmented images performed significantly better compared to the models trained using original CXR images for both the phases (phase-I and phase-II), which confirms the validity of Hypothesis 1. The obtained promising performance using augmented images can be justified by the fact that the augmented images provide sufficient instances to train the model for possible variations in input CXR images, which might occur due to diverse imaging parameters and platforms in different hospitals.

Moreover, the results obtained using different supervised algorithms for phase-I (shown in Table 4) and phase-II (shown in Table 5) demonstrates that the SVM (linear kernel) outperformed the others using a selected feature set (1546 features for phase-I and 2018 features for phase-II). The significant better performance of SVM is due to its generalization capability and ability to learn and infer the intricate natural patterns by efficiently adapting the hyperplane and the soft margins using support vectors. Further, the obtained higher accuracy (ACC) of $99.67 \pm 0.31\%$, area under the curve (AUC) of 1 ± 0.00 , and Matthews Correlation Coefficient (MCC) of 0.99 ± 0.00 for phase-I and ACC of 98.78 ± 0.96 , AUC of 0.99 ± 0.01 , and MCC of 0.98 ± 0.02 for phase-II demonstrates its promising performance and thus justifying the validity of Hypothesis 2.

The nCOVID-19 is highly contagious, and even a single false negative may lead to community spread of the infection. Therefore, to reduce the probable chance of misclassification, we used the majority voting based classifier ensemble of seven benchmark supervised models, as shown in Fig. 3. Further, the classification performance of the majority voting technique and individual benchmark classifiers are evaluated using a separate validation set (which was not used during training of the models). The set consists of 258 CXR images (86 normal, 86 nCOVID-19

Table 4

Phase-I (Normal vs. Abnormal) classification performance of different supervised models using Training-Testing in 10-fold cross-validation setup. (Abbreviations: SVM: Support Vector Machine, DT: Decision Tree, KNN: *k*-Nearest Neighbor, NB: Naïve Bayes, ANN: Artificial Neural Network, STD: Standard Deviation, AUC: Area Under Curve, MCC: Matthews Correlation Coefficient, Note: best performance is highlighted with bold letters).

Classification algorithms	Accuracy (\pm STD)	Specificity (\pm STD)	Precision (\pm STD)	Recall (\pm STD)	F ₁ -Measure (\pm STD)	AUC (\pm STD)	MCC (\pm STD)
<i>Without using Augmented Images</i>							
SVM (RBF Kernel)	66.70 \pm 0.83	3.40 \pm 2.22	66.96 \pm 0.79	98.56 \pm 0.95	79.74 \pm 0.57	0.51 \pm 0.01	0.06 \pm 0.09
DT	90.73 \pm 5.03	85.95 \pm 6.04	92.95 \pm 3.11	93.12 \pm 4.97	93.02 \pm 3.89	0.90 \pm 0.05	0.79 \pm 0.11
NB	97.22 \pm 1.15	97.71 \pm 3.24	98.85 \pm 1.13	96.98 \pm 0.82	97.90 \pm 0.86	0.97 \pm 0.02	0.94 \pm 0.03
KNN	97.41 \pm 2.08	98.28 \pm 1.20	98.14 \pm 1.00	96.99 \pm 3.26	98.02 \pm 0.64	0.98 \pm 0.02	0.94 \pm 0.04
SVM (Poly Kernel)	98.47 \pm 0.81	97.97 \pm 2.39	99.00 \pm 0.86	98.71 \pm 0.81	98.85 \pm 0.61	0.98 \pm 0.01	0.97 \pm 0.02
ANN	98.47 \pm 1.12	98.29 \pm 1.48	99.13 \pm 0.75	98.56 \pm 1.18	98.84 \pm 0.85	0.98 \pm 0.01	0.97 \pm 0.02
SVM (Linear Kernel)	98.85 \pm 1.09	98.57 \pm 1.51	99.21 \pm 0.76	98.99 \pm 0.98	99.13 \pm 0.73	0.99 \pm 0.01	0.97 \pm 0.02
<i>Using Augmented Images</i>							
SVM (RBF Kernel)	82.90 \pm 1.71	48.71 \pm 5.09	79.62 \pm 1.63	98.17 \pm 0.35	88.64 \pm 1.01	0.74 \pm 0.03	0.62 \pm 0.04
DT	95.88 \pm 1.04	93.38 \pm 3.6	96.75 \pm 1.73	97.13 \pm 1.17	96.92 \pm 0.76	0.95 \pm 0.02	0.91 \pm 0.02
NB	96.70 \pm 1.55	97.28 \pm 1.84	98.60 \pm 0.95	96.41 \pm 1.76	97.49 \pm 1.19	0.97 \pm 0.02	0.93 \pm 0.03
ANN	99.33 \pm 0.25	98.13 \pm 1.55	99.18 \pm 0.56	99.43 \pm 0.34	99.50 \pm 0.16	0.99 \pm 0.01	0.98 \pm 0.02
SVM (Poly Kernel)	99.38 \pm 0.55	99.57 \pm 0.17	99.78 \pm 0.15	99.19 \pm 0.17	99.13 \pm 0.52	0.99 \pm 0.01	0.99 \pm 0.01
KNN	99.41 \pm 0.51	99.71 \pm 0.90	99.86 \pm 0.13	99.31 \pm 0.40	99.71 \pm 0.28	1.00 \pm 0.01	0.99 \pm 0.01
SVM (Linear Kernel)	99.67 \pm 0.31	99.57 \pm 0.27	99.79 \pm 0.67	99.86 \pm 0.14	99.82 \pm 0.17	1.00 \pm 0.00	0.99 \pm 0.00

Table 5

Phase-II (nCOVID vs. Pneumonia) classification performance of different supervised models using Training-Testing set in 10-fold cross-validation setup.

Classification algorithms	Accuracy (\pm STD)	Specificity (\pm STD)	Precision (\pm STD)	Recall (\pm STD)	F ₁ -Measure (\pm STD)	AUC (\pm STD)	MCC (\pm STD)
<i>Without using Augmented Images</i>							
DT	72.98 \pm 3.54	70.93 \pm 12.06	73.06 \pm 6.20	74.96 \pm 8.82	73.41 \pm 3.22	0.73 \pm 0.04	0.47 \pm 0.07
ANN	74.29 \pm 5.28	74.29 \pm 32.99	74.29 \pm 8.57	74.29 \pm 30.07	74.29 \pm 17.5	0.74 \pm 0.05	0.49 \pm 0.13
SVM (RBF Kernel)	80.03 \pm 4.40	80.18 \pm 10.61	81.05 \pm 8.12	79.91 \pm 7.44	80.01 \pm 4.30	0.80 \pm 0.04	0.61 \pm 0.09
KNN	80.17 \pm 4.62	83.30 \pm 7.82	82.68 \pm 6.42	77.03 \pm 7.23	79.47 \pm 4.92	0.80 \pm 0.05	0.61 \pm 0.09
NB	80.46 \pm 6.03	79.57 \pm 6.26	80.00 \pm 5.71	81.36 \pm 7.70	80.58 \pm 6.18	0.80 \pm 0.06	0.61 \pm 0.12
SVM (Poly Kernel)	83.34 \pm 3.46	85.89 \pm 5.63	85.42 \pm 4.37	80.77 \pm 5.81	82.87 \pm 3.63	0.83 \pm 0.03	0.67 \pm 0.07
SVM (Linear Kernel)	85.07 \pm 6.94	87.09 \pm 9.64	87.11 \pm 8.73	83.08 \pm 7.39	84.83 \pm 6.74	0.85 \pm 0.07	0.71 \pm 0.14
<i>Using Augmented Images</i>							
NB	84.63 \pm 6.75	85.94 \pm 6.26	85.51 \pm 6.43	83.31 \pm 8.46	84.32 \pm 7.12	0.85 \pm 0.07	0.69 \pm 0.13
DT	84.84 \pm 5.27	84.19 \pm 4.82	84.36 \pm 4.85	85.49 \pm 6	84.91 \pm 5.34	0.85 \pm 0.05	0.7 \pm 0.11
ANN	94.93 \pm 2.54	94.2 \pm 3.52	94.29 \pm 3.1	95.65 \pm 3.5	94.96 \pm 2.55	0.95 \pm 0.03	0.9 \pm 0.05
KNN	97.27 \pm 1.86	95.83 \pm 2.5	95.98 \pm 2.35	98.7 \pm 1.25	97.31 \pm 1.81	0.97 \pm 0.02	0.95 \pm 0.04
SVM (RBF Kernel)	97.7 \pm 1.85	97.41 \pm 2.17	97.49 \pm 2.27	97.99 \pm 1.91	97.71 \pm 1.82	0.98 \pm 0.02	0.95 \pm 0.04
SVM (Poly Kernel)	97.98 \pm 1.99	97.7 \pm 2.26	97.75 \pm 2.18	98.19 \pm 1.58	97.89 \pm 2.05	0.98 \pm 0.02	0.95 \pm 0.04
SVM (Linear Kernel)	98.78 \pm 0.96	98.14 \pm 1.69	98.19 \pm 1.68	99.23 \pm 0.72	98.79 \pm 0.95	0.99 \pm 0.01	0.98 \pm 0.02

and 86 pneumonia). Initially, the radiomic texture features (described in Section 2.2) were extracted from the input CXR images and classified using different supervised models in phase-I. The output of each model acts as an expert suggestion to segregate the input CXR images into normal or abnormal (nCOVID-19 or pneumonia). The classification performance of each model using validation set in phase-I is shown in Table 6. From the obtained results, it can be observed that the performance of majority voting algorithm in phase-I (ACC of 98.062%, AUC of 0.977, and MCC of 0.956) is significantly better compared to the others. Further, all the images which were classified to abnormal category in phase-I were passed to phase-II for differential diagnosis between nCOVID-19 and pneumonia.

In phase-II, the abnormal input images were classified using each

Table 6

Phase-I (Normal vs. Abnormal) classification performance of different supervised models and majority voting algorithm using the validation set. (Abbreviations: SVM: Support Vector Machine, DT: Decision Tree, KNN: *k*-Nearest Neighbor, NB: Naïve Bayes, ANN: Artificial Neural Network, STD: Standard Deviation, AUC: Area Under Curve, MCC: Matthews Correlation Coefficient). (Note: best performance is highlighted with bold letters).

Classification algorithms	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F ₁ -Measure (%)	AUC	MCC
NB	88.372	73.256	87.766	95.930	91.667	0.846	0.734
DT	90.698	80.233	90.659	95.930	93.220	0.881	0.788
SVM (RBF Kernel)	95.349	89.535	94.944	98.256	96.571	0.939	0.895
KNN	95.736	94.186	97.076	96.512	96.793	0.953	0.904
SVM (Linear Kernel)	96.124	90.698	95.506	98.837	97.143	0.948	0.913
SVM (Poly Kernel)	96.124	91.860	96.023	98.256	97.126	0.951	0.912
ANN	96.512	93.023	96.571	98.256	97.406	0.956	0.921
Majority voting	98.062	96.512	98.266	98.837	98.551	0.977	0.956

Table 7

Phase-II (nCOVID vs. Pneumonia) classification performance of different supervised models and majority voting algorithm using the validation set. (Note: best performance is highlighted with bold letters).

Classification algorithms	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F ₁ -Measure (%)	AUC	MCC
KNN	72.093	76.744	74.359	67.442	70.732	0.721	0.444
ANN	73.256	53.488	66.667	93.023	77.670	0.733	0.506
DT	79.070	82.558	81.250	75.581	78.313	0.791	0.583
NB	80.814	72.093	76.238	89.535	82.353	0.808	0.626
SVM (Linear Kernel)	81.977	83.721	83.133	80.233	81.657	0.820	0.640
SVM (Poly Kernel)	86.047	79.070	81.633	93.023	86.957	0.860	0.728
SVM (RBF Kernel)	86.628	83.721	84.615	89.535	87.006	0.866	0.734
Majority voting	91.329	86.207	87.368	96.512	91.713	0.914	0.831

that the majority voting approach outperformed the others achieving fewer Type-I and Type-II errors.

3.1. Statistical analysis

In this section describes the statistical significance of the obtained results from the various experiments performed in this study. Initially, the statistical significance of obtained performance (ACC and F₁-measure) of different supervised models using augmented images and without using augmented images for phase-I and phase-II were validated using z-test statistics. The test consider the null hypothesis as the performance of supervised models before and after applying image augmentation is equal. Alternatively, the models trained using augmented images exhibit higher performance. The test statistics for phase-I and phase-II at 95% confidence interval (or $\alpha = 0.05$) are shown in Table 8. From the statistical results of phase-I, it was observed that the classification performance of ANN, SVM (linear and RBF kernel), DT, and KNN models are significantly higher using augmented images (accepting the alternate hypothesis) compared to the models trained using original CXR images. Similarly, in phase-II, all the models strongly accept the alternate hypothesis (i.e., models trained using augmented images exhibit higher performance or $p < 0.05$).

The statistical significance of the proposed ACoS system was evaluated using Friedman average ranking method and Holm and Shaffer pairwise comparison method for validation set (Chandra & Verma, 2020a; Chandra, Verma, Singh et al., 2020). The Friedman test statistics compare the mean ranks of different classifiers assuming that the performance of all classifiers are equal (null hypothesis). From the average ranks shown in Table 9, we found that the test strongly accepts the alternate hypothesis while rejecting the null, which confirms the substantial difference in the performance of different classification algorithms (at $\alpha = 0.05$) for both the phases. The result can also be verified from the Friedman test (at 7 degrees of freedom) with p -value = 0.0000003993 < 0.05 for phase-I and p -value = 0.000001428 < 0.05 for phase-II. Further, the validity of Hypothesis 3 can be verified from the fact that the majority voting algorithm achieved minimum rank (first rank) in both the phases.

Further, the Friedman average rankings shown in Table 9 demonstrate that the mean ranks of different classification algorithms are

significantly different (p -value < α), therefore it is meaning full to perform the pairwise post-hoc comparisons. In this study, Holm (Holm, 1979) and Shaffer (Shaffer, 1986) post-hoc procedures were used to perform multiple pairwise comparisons. The method considers the null hypothesis as all algorithms performed equally.

In this study, 28 pairs of classification algorithms (denoted by 'i') were compared at $\alpha = 0.05$ level of significance. The Holm and Shaffer method reject those hypotheses that have an unadjusted p -value ≤ 0.002381 and p -value ≤ 0.001786 , respectively for both phase-I and phase-II. The test statistics for phase-I and phase-II are shown in Tables 10 and 11, respectively. From the statistical results, it was observed that the performance of the proposed majority vote based classifier ensemble method is significantly better compared to the other classification algorithms for both the phases confirming the validity of Hypothesis 3.

Finally, the performance of the proposed system is compared with the existing state of the art methods (summarized in Table 1). Initially, the proposed method is compared for two-class (normal vs. abnormal/nCOVID-19) as shown in Table 12. From the table, it was observed that the proposed method performed significantly better compared to Panwar, Gupta, Siddiqui, Morales-Menendez, and Singh (2020), Hemandan, Shouman, and Karar (2020) and Maghdid, Asaad, Ghafoor, Sadiq, and Khan (2020). Further, it achieved comparably equal performance to Narin et al. (2020) and Ozturk et al. (2020). However, one should note that Narin et al. (2020) and Ozturk et al. (2020) used comparably less number of CXR images to train the DL model.

Further, the overall accuracy (for three class: normal vs. nCOVID-19 vs. pneumonia) of the proposed model is evaluated and compared with the existing state of art methods, as shown in Table 13. The table reveals that the proposed method performed significantly better in terms of overall accuracy (ACC = 93.411%) compared to Ozturk et al. (2020) and Wang and Wong (2020). However, it achieved comparably lower performance than Abbas, Abdelsamea, and Gaber (2020), Chowdhury et al. (2020), Ucar and Korkmaz (2020) and Toğaçar, Ergen, and Cömert (2020), which is due to the fact that the author Toğaçar, Ergen, and Cömert (2020) and Abbas, Abdelsamea, and Gaber (2020) used very less number of CXR image to train the DL models. Further, radiological responses of pneumonia and nCOVID-19 are subtle, which confuses the classifier. Overcoming such limitation is still an open-ended research area.

3.2. Discussion

The morbidity and mortality rate due to nCOVID-19 is rapidly increasing, with thousands of reported death worldwide. The WHO has already declared this pandemic as a global health emergency (Coronavirus Disease 2019, 2020). In this study, we presented an ACoS system to detect nCOVID-19 infected patients using CXR image data. We performed two-phase classification to segregate normal, nCOVID-19 and pneumonia infected images. The major challenges we experienced in this study are:

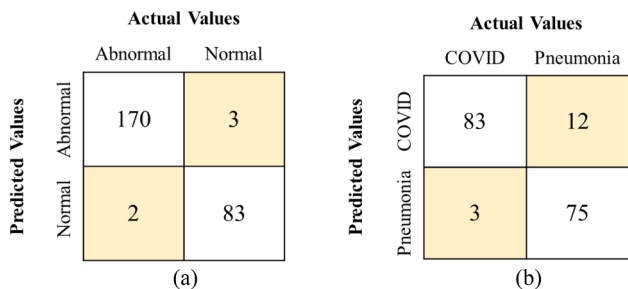


Fig. 4. Confusion matrix using validation set for (a) Majority voting (Phase-I), (b) Majority voting (Phase-II).

Table 8

Computed z-score for comparing the performance (accuracy and F-measure) of different supervised models using augmented images vs. without using augmented images for Training-Testing set in 10-fold cross-validation setup (at 95% significance level or alpha = 0.05). (Abbreviations: SVM: Support Vector Machine, DT: Decision Tree, KNN: k-Nearest Neighbor, NB: Naive Bayes, ANN: Artificial Neural Network, STD: Standard Deviation, AUC: Area Under Curve, MCC: Matthews Correlation Coefficient, Note: Bold value denotes the rejection of alternate hypothesis).

Classifiers	Phase-I				Phase-II			
	Accuracy		F ₁ -Measure		Accuracy		F ₁ -Measure	
	Z-Score	P-Value	Z-Score	P-Value	Z-Score	P-Value	Z-Score	P-Value
ANN	-2.78686	0.00821	-2.84720	0.00693	-16.78630	0.00000	-16.83080	0.00000
SVM (Linear Kernel)	-2.32891	0.02649	-2.04164	0.04963	-13.50650	0.00000	-13.70514	0.00000
SVM (RBF Kernel)	-10.23706	0.00000	-6.70490	0.00000	-18.81539	0.00000	-18.84667	0.00000
SVM (Poly Kernel)	-2.51459	0.01690	-0.77056	0.29647	-15.24470	0.00000	-15.41234	0.00000
DT	-5.79733	0.00000	-5.03525	0.00000	-7.96144	0.00000	-7.74892	0.00000
NB	0.79659	0.29048	0.71026	0.31000	-2.94058	0.00529	-2.62549	0.01271
KNN	-4.74345	0.00001	-4.86842	0.00000	-16.23091	0.00000	-16.75614	0.00000

Table 9

Average ranking of classifiers based on different classification performance metrics using the Friedman test with 7 degrees of freedom. (Note: the minimum value represents the better rank and is highlighted in bold).

Classification algorithms	Average ranking of classification algorithms	
	Phase - I	Phase - II
NB	7.929	5.214
DT	7.071	5.714
SVM (RBF Kernel)	5.714	2.429
KNN	4.000	7.571
SVM (Linear Kernel)	3.714	4.071
SVM (Poly Kernel)	3.929	3.357
ANN	2.571	6.643
Majority Voting	1.071	1.000

- The publicly available nCOVID-19 infected CXR images are limited and lacking standardization.
- The radiological characteristics of nCOVID-19 and pneumonia infections are ambiguous.

Moreover, several studies using DL approaches have been reported in the literature for detection of nCOVID-19 infection in CXR and CT images (as shown in Table 1). Although the DL methods reported promising performance, it suffers from the following shortcomings:

- Resize the input CXR images to lower resolution (like 64 × 64 or 224 × 224, etc.) before processing, which may result in loss of crucial discriminative texture information.
- Demands massive training data to sufficiently train the model.
- Requires expertise to define suitable network architecture and set the many hyper-parameters (like input resolution, number of layers, filters, and filter shape, etc.).

Table 10

p-value and adjusted p-value for pairwise multiple comparisons of different supervised classification algorithms (Phase-I: Normal vs. Abnormal) using the validation set at α = 0.05. (Abbreviations: SVM: Support Vector Machine, DT: Decision Tree, KNN: k-Nearest Neighbor, NB: Naive Bayes, ANN: Artificial Neural Network, STD: Standard Deviation, AUC: Area Under Curve, MCC: Matthews Correlation Coefficient).

i	Algorithms	$z = \frac{R_0 - R_i}{SE}$	p	Holm	Shaffer	Adjusted p-Value	
						P _{Holm}	P _{Shaffer}
28	NB vs. Majority Voting	5.2372	0.0000	0.0018	0.0018	0.0036	0.0036
27	DT vs. Majority Voting	4.5826	0.0000	0.0019	0.0024	0.0037	0.0048
26	NB vs. ANN	4.0916	0.0000	0.0019	0.0024	0.0038	0.0048
25	SVM (RBF Kernel) vs. Majority Voting	3.5460	0.0004	0.0020	0.0024	0.0040	0.0048
24	DT vs. ANN	3.4369	0.0006	0.0021	0.0024	0.0042	0.0048
23	NB vs. SVM (Linear Kernel)	3.2187	0.0013	0.0022	0.0024	0.0043	0.0048
22	NB vs. SVM (Poly Kernel)	3.0551	0.0023	0.0023	0.0024	0.0045	0.0048
21	NB vs. KNN	3.0005	0.0027	0.0024	0.0024	0.0048	0.0048
20	DT vs. SVM (Linear Kernel)	2.5641	0.0103	0.0025	0.0025	0.0050	0.0063
19	DT vs. SVM (Poly Kernel)	2.4004	0.0164	0.0026	0.0026	0.0053	0.0063
18	SVM (RBF Kernel) vs. ANN	2.4004	0.0164	0.0028	0.0028	0.0056	0.0063
17	DT vs. KNN	2.3458	0.0190	0.0029	0.0029	0.0059	0.0063
16	KNN vs. Majority Voting	2.2367	0.0253	0.0031	0.0031	0.0063	0.0063
15	SVM (Poly Kernel) vs. Majority Voting	2.1822	0.0291	0.0033	0.0033	0.0067	0.0067
14	SVM (Linear Kernel) vs. Majority Voting	2.0185	0.0435	0.0036	0.0036	0.0071	0.0071
13	NB vs. SVM (RBF Kernel)	1.6912	0.0908	0.0038	0.0038	0.0077	0.0077
12	SVM (RBF Kernel) vs. SVM	1.5275	0.1266	0.0042	0.0042	0.0083	0.0083
11	SVM (RBF Kernel) vs. SVM	1.3639	0.1726	0.0045	0.0045	0.0091	0.0091
10	SVM (RBF Kernel) vs. KNN	1.3093	0.1904	0.0050	0.0050	0.0100	0.0100
9	ANN vs. Majority Voting	1.1456	0.2519	0.0056	0.0056	0.0111	0.0111
8	KNN vs. ANN	1.0911	0.2752	0.0063	0.0063	0.0125	0.0125
7	DT vs. SVM (RBF Kernel)	1.0365	0.3000	0.0071	0.0071	0.0143	0.0143
6	SVM (Poly Kernel) vs. ANN	1.0365	0.3000	0.0083	0.0083	0.0167	0.0167
5	SVM (Linear Kernel) vs. ANN	0.8729	0.3827	0.0100	0.0100	0.0200	0.0200
4	NB vs. DT	0.6547	0.5127	0.0125	0.0125	0.0250	0.0250
3	KNN vs. SVM (Linear Kernel)	0.2182	0.8273	0.0167	0.0167	0.0333	0.0333
2	SVM (Linear Kernel) vs. SVM	0.1637	0.8700	0.0250	0.0250	0.0500	0.0500
1	KNN vs. SVM (Poly Kernel)	0.0546	0.9565	0.0500	0.0500	0.1000	0.1000

Table 11

p-value and adjusted p-value for pairwise multiple comparisons of different supervised classification algorithms (Phase-II: nCOVID-19 vs. Pneumonia) using the validation set at $\alpha = 0.05$.

i	Algorithms	$z = \frac{R_0 - R_i}{SE}$	p	Holm	Shaffer	Adjusted p-Value	
						P_{Holm}	$P_{Shaffer}$
28	KNN vs. Majority Voting	5.0190	0.0000	0.0018	0.0018	0.0000	0.0000
27	ANN vs. Majority Voting	4.3098	0.0000	0.0019	0.0024	0.0004	0.0003
26	KNN vs. SVM (RBF)	3.9279	0.0001	0.0019	0.0024	0.0022	0.0018
25	DT vs. Majority Voting	3.6006	0.0003	0.0020	0.0024	0.0079	0.0067
24	NB vs. Majority Voting	3.2187	0.0013	0.0021	0.0024	0.0309	0.0270
23	KNN vs. SVM (Poly)	3.2187	0.0013	0.0022	0.0024	0.0309	0.0270
22	ANN vs. SVM (RBF)	3.2187	0.0013	0.0023	0.0024	0.0309	0.0270
21	KNN vs. SVM (Linear)	2.6732	0.0075	0.0024	0.0024	0.1578	0.1578
20	ANN vs. SVM (Poly)	2.5095	0.0121	0.0025	0.0025	0.2418	0.1934
19	DT vs. SVM (RBF)	2.5095	0.0121	0.0026	0.0026	0.2418	0.1934
18	SVM (Linear) vs. Majority Voting	2.3458	0.0190	0.0028	0.0028	0.3417	0.3037
17	NB vs. SVM (RBF)	2.1276	0.0334	0.0029	0.0029	0.5673	0.5339
16	ANN vs. SVM (Linear)	1.9640	0.0495	0.0031	0.0031	0.7926	0.7926
15	DT vs. SVM (Poly)	1.8003	0.0718	0.0033	0.0033	1.0772	1.0772
14	SVM (Poly) vs. Majority Voting	1.8003	0.0718	0.0036	0.0036	1.0772	1.0772
13	KNN vs. NB	1.8003	0.0718	0.0038	0.0038	1.0772	1.0772
12	NB vs. SVM (Poly)	1.4184	0.1561	0.0042	0.0042	1.8728	1.8728
11	KNN vs. DT	1.4184	0.1561	0.0045	0.0045	1.8728	1.8728
10	SVM (Linear) vs. SVM (RBF)	1.2548	0.2096	0.0050	0.0050	2.0957	2.0957
9	DT vs. SVM (Linear)	1.2548	0.2096	0.0056	0.0056	2.0957	2.0957
8	SVM (RBF) vs. Majority Voting	1.0911	0.2752	0.0063	0.0063	2.2019	2.2019
7	ANN vs. NB	1.0911	0.2752	0.0071	0.0071	2.2019	2.2019
6	NB vs. SVM (Linear)	0.8729	0.3827	0.0083	0.0083	2.2964	2.2964
5	ANN vs. DT	0.7092	0.4782	0.0100	0.0100	2.3910	2.3910
4	SVM (Poly) vs. SVM (RBF)	0.7092	0.4782	0.0125	0.0125	2.3910	2.3910
3	KNN vs. ANN	0.7092	0.4782	0.0167	0.0167	2.3910	2.3910
2	SVM (Linear) vs. SVM (Poly)	0.5455	0.5854	0.0250	0.0250	2.3910	2.3910
1	DT vs. NB	0.3819	0.7025	0.0500	0.0500	2.3910	2.3910

Table 12

Two class (normal vs. abnormal) performance comparison of the proposed method with the state of art methods.

Articles	Class	Algorithms/techniques	ACC (%)
Panwar et al. (2020)	2	nCOVnet	88.10
Hemdan, Shouman, and Karar (2020)	2	VGG19, DenseNet201, ResNetV2, InceptionV3, ResNetV2, Xception, MobileNetV2	90.00
Maghddid, Asaad, Ghafoor, Sadiq, and Khan (2020)	2	AlexNet, Modified CNN	94.00
Narin et al. (2020)	2	ResNet50, ResNetV2, InceptionV3	98.00
Ozturk et al. (2020)	2	DarkNet	98.08
Proposed Method (Phase-I)	2	Majority vote based classifier ensemble	98.06

- Requires high computational resources, extensive memory and a significant amount of time to train the network.
- Unlike conventional machine learning, DL approaches are unexplainable in nature.

To overcome the aforementioned limitations, we have used a combination of radiomic texture features with conventional ML algorithms. The following facts can justify the promising performance of the proposed ACoS system:

- The radiomic texture descriptors (FOSF, GLCM, and HOG features) are highly efficient in encoding natural textures and thus can easily quantize the correlation attributes of radiological visual characteristics associated with nCOVID-19 infection.
- The image augmentation technique provides sufficient instances to train the model for possible variable inputs, making the model robust.

Table 13

Three class (normal, nCOVID-19 and pneumonia) performance comparison of the proposed method with the state of art methods.

Articles	Class	Algorithms/techniques	Total number of images	ACC (%)
Ozturk et al. (2020)	3	DarkNet		87.02
L. Wang et al. (2020)	3	COVID-Net		93.30
Abbas et al. (2020)	3	DeTraC	196	95.12
Chowdhury et al. (2020)	3	AlexNet, ResNet18, DenseNet201, SqueezeNet	3487	97.94
Ucar and Korkmaz (2020)	3	Deep Bayes-SqueezeNe	5957	98.26
Nour et al. (2020)	3	CNN, SVM, DT, KNN	3670	98.97
Toğaçar et al. (2020)	3	MobileNetV2, SqueezeNet, SVM	458	99.27
Proposed Method (Overall)	3	Majority vote based classifier ensemble	2346	93.41

- The conventional ML algorithms can be efficiently trained using smaller datasets, fewer resources and minimal hyper-parameter tuning without compromising with the performance.
- The majority vote based classifier ensemble method used in the proposed ACoS system acts as a multi-expert recommendation system and reduces the probable chance of misclassification.

The disadvantages of the proposed system are as follows:

- The subtle radiographic responses of different abnormalities like TB, pneumonia, influenza, etc. confuses the classifier, limiting the diagnostic performance of the system.

In the proposed ACoS system, majority vote based classifier ensemble technique has been exploited to reduce the probable chance of misclassification of nCOVID-19 infected patients. Such method can be easily integrated into mobile radiology van and can work for the welfare of the society.

4. Conclusion

In this study, we have presented an ACoS system for preliminary diagnosis of nCOVID-19 infected patients, so that proper precautionary measures (like isolation and RT-PCR test) can be taken to prevent the further outbreak of the infection. The key findings of the study are summarized as follows:

- The proposed ACoS system demonstrated the promising potential to segregate the normal, pneumonia, and nCOVID-19 infected patients, which can be verified from the significant performance of phase-I (ACC = 98.062%, AUC = 0.977, and MCC = 0.956) and phase-II (ACC = 91.329%, AUC = 0.914 and MCC = 0.831) using the validation set.
- There are significant variations in the input CXR images due to diverse imaging conditions in different hospitals. The proposed system used augmented images, which generate sufficient variability to train the model and improve its robustness.
- The radiomic texture descriptors like FOSF, GLCM, and HOG features are highly efficient in quantizing the correlation attributes of radiological visual characteristics associated with nCOVID-19 infection.
- Unlike the data-hungry DL approaches, the proposed ACoS system used conventional ML algorithms to train the model with limited annotated images and less computational resources. This type of system may have greater clinical acceptability and can be deployed even in a resource-constrained environment.
- The Friedman post-hoc multiple comparison and z-score statistics confirm the statistical significance of the proposed system.

The future work of this study should focus on improving the reliability and clinical acceptability of the system. The integration of the patient’s symptomatology and radiologist’s feedback with the CAD system could be helpful in making a robust screening system. Further, an in-depth analytical comparison of performances between conventional algorithms and deep learning methods could help in establishing its clinical acceptability.

Appendix A. Radiomic texture feature

The ambiguous texture patterns in CXR images due to different infectious diseases is challenging for the radiologist to diagnose and correlate the patterns to a specific disease accurately. nCOVID-19 exhibits various radiological characteristics, as described in Section 2.2. These visual indicators can be efficiently quantized using statistical texture descriptors (FOSF, GLCM, HOG feature). The FOSF encodes the texture according to the statistical distribution of pixel intensities over the entire image deriving a set of histogram statistics (like mean, variance, smoothness, kurtosis, energy and entropy etc.) by waiving the correlation among the pixels (Srinivasan & Shobha, 2008). Further, to encode the spatial correlation of grey-level distributions of disease texture patterns in the local neighborhood, the GLCM texture feature (Gómez et al., 2012; Haralick et al., 1973) is used. It considers the relative position of pixel intensities in a given neighborhood of size $r \times c$ of an input image $I(x,y)$, encoding natural textures (Gómez et al., 2012). It represents the relative frequency of two grey levels ‘i’ and ‘j’ as statistical probability value $P(i,j|d, \theta)$ or $P(i,j|\Delta x, \Delta y)$ that occurs at pair of points separated by distance vector ‘d’ along angle $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. The statistical probability value is described in Eq. (16). The summary of these features are recapitulated in Table 14.

$$P(i,j|\Delta x, \Delta y) = \frac{1}{(r - \Delta x)(c - \Delta y)} \sum_{c=1}^{c-\Delta y} \sum_{r=1}^{r-\Delta x} K \tag{16}$$

$$\text{where, } K = \begin{cases} 0 & \text{if } f(r,c) \neq i \text{ and } f(r + \Delta x, c + \Delta y) \neq j \\ 1 & \text{elsewhere} \end{cases}$$

The HOG features (Dalal & Triggs, 2005; Santosh & Antani, 2018) extract the gradient magnitude and direction of local neighborhood, encoding

Table 14
Radiomic texture features (FOSF, GLCM features) extracted from CXR images.

Category of features	Number of features	Name of features
First Order Statistical Feature (FOSF) (Srinivasan & Shobha, 2008)	8	Mean (m), Variance (μ^2), Standard deviation (σ), Skewness (μ^3), Kurtosis (μ^4), Smoothness (R), Uniformity (U), Entropy (e)
Gray Level Co-occurrence Matrix (GLCM) Texture Feature (Gómez et al., 2012; Haralick et al., 1973)	88 (22×4)	Sum average, Sum variance, Difference variance, Energy, Autocorrelation, Entropy, Sum entropy, Difference entropy, Contrast, Homogeneity I, Homogeneity II, Correlation I, Correlation II, Cluster Prominence, Cluster Shade, Sum of squares, Maximum probability, Dissimilarity, Information measure of correlation I, Information measure of correlation II, Inverse difference normalized, Inverse difference moment normalized
Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005; Santosh & Antani, 2018)	8100	f1, f2, f3, f4, f8100

CRediT authorship contribution statement

Tej Bahadur Chandra: Conceptualization, Methodology, Software, Validation, Investigation, Writing - original draft. **Kesari Verma:** Conceptualization, Formal analysis, Supervision, Writing - review & editing. **Bikesh Kumar Singh:** Formal analysis, Supervision, Writing - review & editing. **Deepak Jain:** Visualization, Validation. **Satyabhanu Singh Netam:** Visualization, Validation.

Declaration of Competing Interest

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

Acknowledgment

The authors would like to acknowledge J.P. Cohen and group for collecting images from various sources and making it publicly available (COVID-Chestxray Database) for the research community.

the local shape and texture information. Initially, the input image is converted into smaller blocks for which gradient magnitude and orientation are computed, as shown in Eqs. (17) and (18), respectively. Further, the histogram bin is created and normalized to create a feature descriptor.

$$g = \sqrt{G_x^2 + G_y^2} \quad (17)$$

$$\theta = \arctan \frac{g_y}{g_x} \quad (18)$$

where, G_x and G_y represents the gradient in x and y direction, respectively. θ represents the gradient direction.

Appendix B. Majority Voting of classifier ensemble

Initially, in step 1 and 2, each image x_i from dataset D is retrieved, and class variables are initialized. The extracted images are tested via all the classification algorithms $m_j \in M$ (i.e., SVM, ANN, KNN, NB, and DT), as shown in step 4. If the classifier classified the sample to a healthy class, the 'Healthy' class label would be incremented by 1 as shown in step 6. Similarly, for the samples classified to unhealthy class, the 'Unhealthy' class label will be incremented (shown step 8). Finally, based on the majority vote, a class label is assigned to each input image, as shown in step 11. Here, the majority vote acts as a multi-expert recommendation and reduce the probable chance of false diagnosis.

Algorithm 1. Majority voting algorithm using ensemble of benchmark classifiers.

Input/Initialize

Validation dataset $D = \{x_1, x_2, \dots, x_n\}$. // n =number of images in the dataset
 Classification model $M = \{m_1, m_2, \dots, m_k\}$ // k = number of classification models
 Healthy = 0; Unhealthy = 0;

Output

Prediction results of each image as healthy or infected.

Algorithm

```

1: for  $x_i = 1$  to  $n$  //where,  $n$ =number of images in the dataset
2:   Healthy = 0; Unhealthy = 0;
3:   for  $m_j = 1$  to  $k$  // where,  $k$ =number of classification models
4:      $x_{i,class}^m \leftarrow \text{predict}(x_i, m_j)$  // extract the class label using model  $m_j$  for input image  $x_i$ 
5:     if  $x_{i,class}^m = \text{'Healthy'}$  // jth model classify the data as 'Healthy'
6:       Healthy = Healthy + 1; // increment the counter by one
7:     else if  $x_{i,class}^m = \text{'Unhealthy'}$  // jth model classifies the data as abnormal
8:       Unhealthy = Unhealthy + 1; // Increment the probability by one
9:     end
10:  end
11:  if Healthy > Unhealthy
12:     $x_{i,class} = \text{'Healthy'}$ ;
13:  else
14:     $x_{i,class} = \text{'Unhealthy'}$ ;
15:  end
16: end

```

References

- Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Retrieved from <http://arxiv.org/abs/2003.13815>.
- Altaf, F., Islam, S. M. S., Akhtar, N., & Janjua, N. K. (2019). Going deep in medical image analysis: Concepts, methods, challenges and future directions. Retrieved from *IEEE Access*, 7, 99540–99572 <http://arxiv.org/abs/1902.05655>.
- Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., & Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine*, 121, Article 103795. <https://doi.org/10.1016/j.cmpbiomed.2020.103795>.
- Asnaoui, K. El Chawki, Y., & Idri, A. (2020). Automated methods for detection and classification pneumonia based on X-Ray images using deep learning. ArXiv Preprint ArXiv:2003.14363. Retrieved from <http://arxiv.org/abs/2003.14363>.
- Artificial Neural Networks. (2013). Retrieved August 8, 2018, from <https://scholastic.technology.com/courses/pattern-classification/lectures/1462278>.
- Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., Singh, R. K., Xue, Z., ... McDonald, C. J. (2014). Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Transactions on Medical Imaging*, 33(2), 577–590. <https://doi.org/10.1109/TMI.2013.2290491>.
- Chandra, T. B., & Verma, K. (2020). Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In B. B. Chaudhuri, M. Nakagawa, P. Khanna, & S. Kumar (Eds.), *Proceedings of third international conference on computer vision & image processing* (pp. 21–33). Singapore: Springer. https://doi.org/10.1007/978-981-32-9088-4_3.
- Chandra, T. B., & Verma, K. (2020a). Analysis of quantum noise-reducing filters on chest X-ray images: A review. *Measurement*, 153, Article 107426. <https://doi.org/10.1016/j.measurement.2019.107426>.
- Chandra, T. B., Verma, K., Jain, D., & Netam, S. S. (2020). Localization of the suspected abnormal region in chest radiograph images. In *2020 First international conference on power, control and computing technologies (ICPC2T)* (pp. 204–209). IEEE. <https://doi.org/10.1109/ICPC2T48082.2020.9071445>.
- Chandra, T. B., Verma, K., Singh, B. K., Jain, D., & Netam, S. S. (2020). Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. *Expert Systems with Applications*, 158, Article 113514. <https://doi.org/10.1016/j.eswa.2020.113514>.
- Chatterjee, S., Dey, D., & Munshi, S. (2019). Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification. *Computer Methods and Programs in Biomedicine*, 178, 201–218. <https://doi.org/10.1016/j.cmpb.2019.06.018>.
- Chen, Q., Meng, Z., & Su, R. (2020). WERFE: A gene selection algorithm based on recursive feature elimination and ensemble strategy. *Frontiers in Bioengineering and Biotechnology*, 8, 496. <https://doi.org/10.3389/fbioe.2020.00496>.
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, 395(10223), 507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
- Cheng, S.-C., Chang, Y.-C., Fan Chiang, Y.-L., 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. <https://doi.org/10.1016/j.jfma.2020.02.007>.
- Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., ... de Albuquerque, V. H. C. (2020). A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Applied Sciences (Switzerland)*, 10(2). <https://doi.org/10.3390/app10020559>.
- Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. Bin, ... Reaz, M. B. I. (2020). Can AI help in screening Viral and COVID-19 pneumonia? (March). Retrieved from <http://arxiv.org/abs/2003.13145>.

- Chung, M., Bernheim, A., Mei, X., Zhang, N., Huang, M., Zeng, X., ... Shan, H. (2020). CT imaging features of 2019 Novel Coronavirus (2019-nCoV). *Radiology*, 295(1), 202–207. <https://doi.org/10.1148/radiol.2020.200230>.
- Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-Chestxray Database. Retrieved June 20, 2020, from <https://github.com/ieee8023/covid-chestxray-dataset>.
- Coronavirus Disease 2019. (2020). Retrieved April 8, 2020, from World Health Organization website: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Vol. 1. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR 05)* (pp. 886–893). <https://doi.org/10.1109/CVPR.2005.177>.
- Emery, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371–381. <https://doi.org/10.1016/j.neucom.2015.06.083>.
- Gómez, W., Pereira, W. C. A., & Infante, A. F. C. (2012). Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE Transactions on Medical Imaging*, 31(10), 1889–1899. <https://doi.org/10.1109/TMI.2012.2206398>.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. San Francisco, CA, itd: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>.
- Ho, T. K., & Gwak, J. (2019). Multiple feature integration for classification of thoracic disease in chest radiography. *Applied Sciences*, 9(19), 4130. <https://doi.org/10.3390/app9194130>.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Hosseiny, M., Kooraki, S., Gholamrezanezhad, A., Reddy, S., & Myers, L. (2020). Radiology perspective of coronavirus disease 2019 (COVID-19): Lessons from severe acute respiratory syndrome and middle east respiratory syndrome. *American Journal of Roentgenology*, (October), 1–5. <https://doi.org/10.2214/AJR.20.22969>.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Jaeger, S., Karagyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., ... McDonald, C. J. (2014). Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging*, 33(2), 233–245. <https://doi.org/10.1109/TMI.2013.2284099>.
- Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. P. C. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement: Journal of the International Measurement Confederation*, 145, 511–518. <https://doi.org/10.1016/j.measurement.2019.05.076>.
- Hemdan, E. E.-D., Shouman, M. A., & Karar, M. E. (2020). COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-Ray images. Retrieved from <http://arxiv.org/abs/2003.11055>.
- Johns Hopkins University, Corona Resource Center. (2020). Retrieved April 8, 2020, from <https://coronavirus.jhu.edu/map.html>.
- Kanne, J. P., Little, B. P., Chung, J. H., Elicker, B. M., & Ketai, L. H. (2020). Essentials for radiologists on COVID-19: An update-radiology scientific expert panel. *Radiology*, 200527. <https://doi.org/10.1148/radiol.2020.200527>.
- Ke, Q., Zhang, J., Wei, W., Polat, D., Woźniak, M., Kośmider, L., & Damaševičius, R. (2019). A neuro-heuristic approach for recognition of lung diseases from X-ray images. *Expert Systems with Applications*, 126, 218–232. <https://doi.org/10.1016/j.eswa.2019.01.060>.
- Khatami, A., Khosravi, A., Nguyen, T., Lim, C. P., & Nahavandi, S. (2017). Medical image analysis using wavelet transform and deep belief networks. *Expert Systems with Applications*, 86, 190–198. <https://doi.org/10.1016/j.eswa.2017.05.073>.
- Kooraki, S., Hosseiny, M., Myers, L., & Gholamrezanezhad, A. (2020). Coronavirus (COVID-19) outbreak: What the department of radiology should know. *Journal of the American College of Radiology*, 17(4), 447–451. <https://doi.org/10.1016/j.jacr.2020.02.008>.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf optimizer. *Advances in Engineering Software*, 69, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
- Maghdid, H. S., Asaad, A. T., Ghafoor, K. Z., Sadiq, A. S., & Khan, M. K. (2020). Diagnosing COVID-19 pneumonia from X-Ray and CT images using deep learning and transfer learning algorithms. Arxiv.Org. Retrieved from <https://arxiv.org/abs/2004.00038>.
- Narin, A., Kaya, C., & Pamuk, Z. (2020). Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. ArXiv Preprint ArXiv:2003.10849.
- Nour, M., Cömert, Z., & Polat, K. (2020). A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Applied Soft Computing*, 106580. <https://doi.org/10.1016/j.asoc.2020.106580>.
- Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 121, Article 103792. <https://doi.org/10.1016/j.combiomed.2020.103792>.
- Pantazi, X. E., Moshou, D., & Bochtis, D. (2020). Artificial intelligence in agriculture. In *Intelligent Data Mining and Fusion Systems in Agriculture* (pp. 17–101). Elsevier. <https://doi.org/10.1016/B978-0-12-814391-9.00002-9>.
- Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Mendez, R., & Singh, V. (2020). Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos, Solitons and Fractals*, 138, Article 109944. <https://doi.org/10.1016/j.chaos.2020.109944>.
- Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., & Singh, S. (2020). Deep transfer learning based classification model for COVID-19 disease. *IRBM*. <https://doi.org/10.1016/j.irbm.2020.05.003>.
- Pesce, E., Joseph Withey, S., Ypsilantis, P. P., Bakewell, R., Goh, V., & Montana, G. (2019). Learning to detect chest radiographs containing pulmonary lesions using visual attention networks. *Medical Image Analysis*, 53, 26–38. <https://doi.org/10.1016/j.media.2018.12.007>.
- Purkayastha, S., Buddi, S. B., Nuthakki, S., Yadav, B., & Gichoya, J. W. (2020). Evaluating the implementation of deep learning in LibreHealth Radiology on Chest X-Rays. *Advances in Intelligent Systems and Computing*, 943, 648–657. https://doi.org/10.1007/978-3-030-17795-9_47.
- Rish, I., et al. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, 3(22), 41–46.
- Santosh, K. C., & Antani, S. (2018). Automated chest x-ray screening: Can lung region symmetry help detect pulmonary abnormalities? *IEEE Transactions on Medical Imaging*, 37(5), 1168–1177. <https://doi.org/10.1109/TMI.2017.2775636>.
- Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395), 826–831.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. Retrieved from *Advances in Neural Information Processing Systems*, 25, 2951–2959 <http://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf>.
- Sprawls, P. (2018). The physical principles of medical imaging, 2nd ed. Retrieved July 25, 2018, from <http://www.sprawls.org/ppmi2/NOISE/>.
- Srinivasan, G., & Shobha, G. (2008). Statistical texture analysis. Retrieved from *Proceedings of World Academy of Science, Engineering and Technology*, 36(December), 1264–1269 <http://staff.fh-hagenberg.at/wbackfri/Teaching/FBA/Uebungen/UE07charRecog/StatTextAnalysisSrinivasan08.pdf>.
- Toğaçar, M., Ergen, B., & Cömert, Z. (2020). COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Computers in Biology and Medicine*, 121, Article 103805. <https://doi.org/10.1016/j.combiomed.2020.103805>.
- Too, J., Abdullah, A., Mohd Saad, N., Mohd Ali, N., & Tee, W. (2018). A new competitive binary Grey Wolf optimizer to solve the feature selection problem in EMG signals classification. *Computers*, 7(4), 58. <https://doi.org/10.3390/computers7040058>.
- Ucar, F., & Korkmaz, D. (2020). COVIDDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical Hypotheses*, 140, Article 109761. <https://doi.org/10.1016/j.mehy.2020.109761>.
- Vajda, S., Karagyris, A., Jaeger, S., Santosh, K. C. C., Candemir, S., Xue, Z., ... Antani, S. (2018). Feature selection for automatic tuberculosis screening in frontal chest radiographs. *Journal of Medical Systems*, 42(8), 146. <https://doi.org/10.1007/s10916-018-0991-9>.
- Vapnik, V. (1998). *Statistical learning theory* (Vol. 3). Wiley, New York.
- Venegas-Barrera, C. S., & Manjarrez, J. (2011). Visual categorization with bags of keypoints. *Revista Mexicana de Biodiversidad*, 82(1), 179–191. <https://doi.org/10.1234/12345678>.
- Wang, L., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. ArXiv Preprint ArXiv:2003.09871. Retrieved from <http://arxiv.org/abs/2003.09871>.
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). 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* (pp. 2097–2106). doi:10.1109/CVPR.2017.369.
- Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... Xu, B. (2020). A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *MedRxiv*. <https://doi.org/10.1101/2020.02.14.20023028>.
- Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., ... Wu, W. (2020). Deep learning system to screen coronavirus disease 2019 pneumonia. Arxiv.Org. Retrieved from <http://arxiv.org/abs/2002.09334>.
- Xue, Z., Jaeger, S., Antani, S., Long, R., Karagyris, A., Siegelman, J., ... Thoma, G. R. (2018). Localizing tuberculosis in chest radiographs with deep learning. In J. Zhang, & P.-H. Chen (Eds.), *Vol. 3. Medical imaging 2018: Imaging informatics for healthcare, research, and applications* (p. 28). <https://doi.org/10.1117/12.2293022>.
- Yoon, S. H., Lee, K. H., Kim, J. Y., Lee, Y. K., Ko, H., Kim, K. H., ... Kim, Y.-H. (2020). Chest radiographic and CT findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of nine patients treated in Korea. *Korean Journal of Radiology*, 21(4), 494. <https://doi.org/10.3348/kjr.2020.0132>.
- Zhang, N., Wang, L., Deng, X., Liang, R., Su, M., He, C., ... Jiang, S. (2020). Recent advances in the detection of respiratory virus infection in humans. *Journal of Medical Virology*, 92, 408–417. <https://doi.org/10.1002/jmv.25674>.