Biomedical Signal Processing and Control 81 (2023) 104340

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

An efficient lung disease classification from X-ray images using hybrid Mask-RCNN and BiDLSTM

Varadharajan Indumathi [a,*], Rathinavelayutham Siva [b]

[a] Research Scholar School of Computing - Department of Computational Intelligence, SRM Institute of Science and Technology, Chengalpattu, India
[b] School of Computing - Department of Computational Intelligence, SRM Institute of Science and Technology, Chengalpattu, India

ARTICLE INFO

Keywords:
COVID-19 prediction
Deep learning
Chest X-ray
Bidirectional long-short-term-memory
Mask region-based convolutional neural network
Crystal algorithm

ABSTRACT

Lung diseases mainly affect the inner lining of the lungs causing complications in breathing, airway obstruction, and exhalation. Identifying lung diseases such as COVID-19, pneumonia, fibrosis, and tuberculosis at the earlier stage is a great challenge due to the availability of insufficient laboratory kits and image modalities. The rapid progression of the lung disease can be easily identified via Chest X-rays and this serves as a major boon for the terminally ill patients admitted to Intensive Care Units (ICU). To enhance the decision-making capability of the clinicians, a novel lung disease prediction framework is proposed using a hybrid bidirectional Long-Short-Term-Memory (BiDLSTM)-Mask Region-Based Convolutional Neural Network (Mask-RCNN) model. The Crystal algorithm is used to optimize the scalability and convergence issues in the Mask-RCNN model by hyperparameter tuning. The long-range dependencies for lung disease prediction are done using the BiDLSTM architecture which is connected to the fully connected layer of the Mask RCNN model. The efficiency of the proposed methodology is evaluated using three publicly accessible lung disease datasets namely the COVID-19 radiography dataset, Tuberculosis (TB) Chest X-ray Database, and National Institute of Health Chest X-ray Dataset which consists of the images of infected lung disease patients. The efficiency of the proposed technique is evaluated using different performance metrics such as Accuracy, Precision, Recall, F-measure, Specificity, confusion matrix, and sensitivity. The high accuracy obtained when comparing the proposed methodology with conventional techniques shows its efficiency of it in improving lung disease diagnosis.

1. Introduction

Lung disease is a disorder that prevents the lungs from proper functioning and is mainly caused due to smoking, pollution, and abnormal cell growth. Lung disease affects the airways, air sacs, interstitium, blood vessels, pleura, and chest wall thereby causing many respiratory problems such as chronic obstructive pulmonary disease, pneumonia, asthma, tuberculosis, fibrosis, etc [1]. As per the WHO report, chronic obstructive pulmonary disease (COPD) becomes 3rd leading cause of death in 2019 and is increasing day by day. COPD, a chronic inflammatory condition, impairs breathing by obstructing airflow to the lungs. Therefore, an early and accurate detection process helps to control the worsening condition of these diseases. The medical industry now has access to several research methods, including machine learning and deep learning approaches, to effectively diagnose lung disease. Lung disorders are effectively recognized at an early stage by using X-ray image databases [2].

Because of differences in lifestyle, environment, climate change, and other variables, diseases have a greater impact on conditions. The likelihood of becoming unhealthy will then rise. Because of chronic obstructive pulmonary disease (COPD), around 3.4 million people died in 2016 [3]. Generally, the lung diseases risk is rigorous, particularly in middle-low and developing incomes countries, wherever millions of people were facing air pollution and poverty. According to estimates from the WHO, more than 4 million people die prematurely each year due to air pollution-related diseases, including pneumonia and asthma [4]. Therefore, it must be important to take action to reduce air pollution and carbon emissions. Consequently, it is crucial to put in place an efficient diagnostics system that may aid in lung disease identification. Tuberculosis is an infectious disease that is caused by bacteria known as mycobacterium tuberculosis. Microorganisms enter our bodies mostly by inhalation through the lungs. Pneumonia is an inflammatory disease caused by either a bacteria or a virus [5]. In addition, pneumonia can be induced by another foreign object or by inhaling vomit. Asthma is a

* Corresponding author.
E-mail address: vindhumathi.mail@gmail.com (V. Indumathi).

https://doi.org/10.1016/j.bspc.2022.104340
Received 29 June 2022; Received in revised form 4 October 2022; Accepted 23 October 2022
Available online 24 November 2022

chronic condition characterized by frequent episodes of wheezing and shortness of breath. COPD is a curable and preventive condition characterized by airflow restriction that is not completely reversible [6]. The disease process can be anticipated by predictors such as uneven interfaces, parenchymal band manifest, coarse reticular patterns, and interstitial thickening.

COVID-19 appeared as a large epidemic in over 150 countries beginning in December 2019, threatening the health of many individuals. Because there was no medicine or vaccination for COVID at the time, the suspected individual was promptly isolated, limiting the possibility of virus propagation [7]. Virus-infected individuals are identified using screening procedures such as an RT-PCR (reverse transcription-polymerase chain reaction) test and a blood test. This virus will easily spread to healthy persons. COVID-19 first affects the lungs and produces respiratory problems [8]. A timely diagnosis is necessary to save the COVID patient's life. Lung disorders such as TB, pneumonia, and asthma are identified by chest radiography imaging, such as computed tomography (CT) or X-ray imaging [9]. However, for the identification of lung disorders, numerous machine learning and image processing approaches are used. The visual index associated with COVID is represented in X-ray imaging, and chest CT has higher sensitivity for COVID-19 detection. The ground glass pattern was noticed in a COVID-infected person at an early stage in the location where the pulmonary arteries are confined. These flaws are rapidly identified using AI and machine learning methods [10]. Lung disease prediction can take advantage of the advancements and efficacy of deep learning algorithms in disease diagnosis and treatment [26,27]. The literature [11–15] has presented a variety of deep learning algorithms for COVID-19 detection from CT scans and X-rays, including Convolutional Neural Networks, Deep Neural Networks (DNN), etc.

Motivated by the increasing number of COVID19 and lung disease patients in the world, this study presents a hybrid Mask Region-Based Convolutional Neural Network (Mask-RCNN)- bidirectional Long-Short-Term-Memory (BiDLSTM) model to enhance the prediction accuracy of Lung disease prediction. In this paper, the Mask RCNN algorithm is deployed to segment the lung disease-infected areas, and the Crystal algorithm is utilized to improve the segmentation accuracy by selecting the optimal region for segmentation. The main novelty of this study lies in the hybrid Mask RCNN-BiDLSTM model for both segmentation and classification in an end-to-end manner. The proposed model not only offers accurate lung disease prediction but also assists in efficient decision-making. The abnormalities present in the frontal lung projection (upper, middle, and lower zones) are identified with a higher accuracy using the Crystal optimized Mask RCNN network via an edge detection branch and path augmentation pyramid network (PAFPN). Using the BiDLSTM layer, the model predicts the lung disease types such as COVID19, pneumonia, and tuberculosis. Extensive experiments conducted on different datasets such as proves the efficiency of our proposed Hybrid Mask RCNN-BiDLSTM framework in terms of both binary and multiclass classification scenarios.

The rest of the paper is structured in the following manner. The prior research employing pulmonary disease detection is presented in Section 2. The proposed approach is covered in Section 3 concerning various stages of lung disease image processing. The results and discussions are presented in Section 4. In Section 5, the conclusion along with the future scope is discussed.

2. Literature survey

CMT-CNN (Contrastive Multi-Task Conventional Neural Network) learning for automated CT and X-ray diagnosis was reported by Li et al. [11], and it comprises two main tasks. Identifying COVID-19 in healthy subjects and pneumonia patients is a crucial task. The additional job is to use a contrastive loss to enhance local aggregation. To begin, each image undergoes a sequence of augmentations. Then the model is transformed into a comparable embedded presentation of the same images in the latent space while the individual images are distinct. Therefore, while maintaining the spread-out properties of the data, this CMT-CNN may generate predictions that are transformation-invariant. Experiments are done with CT datasets and X-ray datasets obtained in the hospital. As a result, contrastive learning improves deep learning model accuracy on X-ray (0.96 percent to −2.42 percent) and CT (5.49 percent to 6.45 percent) without the requirement of extra annotations according to experimental data.

Dong et al. [12] presented a Robust COVID-19 Deep Network (RCoNet) for COVID19 prediction from X-ray images. The Deformable Mutual Information maximization (DeIM) allows the Mutual Information (MI) among latent representation and input data to be accurately calculated and increased to capture disentangled and dense representational properties. In addition, Mixed High-order Moment Feature (MHMF) may extract the distinguishing traits of complex distributions while comprehensively examining the benefits of using high-order statistics in medical imaging. Finally, Multi-expert Uncertainty-aware Learning (MUL) constructs many parallel dropout networks for each Chest X-ray (CXR) image to evaluate uncertainty and hence minimize performance degradation brought on by data noise. This method attains state-of-the-art performance across numerous measures on open-source COVIDx datasets. The result shows better accuracy than the old methods in the presence of noise.

CXR is the most effective and cost-benefit way to detect Tuberculosis (TB), COVID-19, and Pneumonia. Mahbub et al. [13] used deep features to detect pulmonary abnormalities in CXR related to the above-mentioned diseases. The non-healthy CXR screening is considered with a lightweight Deep neural network (DNN) which is reported in the state-of-the-art literature with the minimum number of parameters and epochs. The developed DNN gives accuracies: 99.87 % on covid-19, 99.76 %on TB, and 99.55 % on pneumonia versus healthy datasets. Similarly, the accuracies for non-healthy CXR screening are 98.89 % for covid-19 vs pneumonia, 100 % for pneumonia vs TB, and 98.99 % for COVID-19 vs TB. The results show that the developed CNN can utilize in resource-constraint regions for mass screening. However, this approach improves performance only when the entire dataset is used; otherwise, the accuracy is quite poor. The authors also failed to address the overfitting, exploding gradient, and class imbalance problems associated with the CNN architecture.

Dey et al. [14] developed an optimal fuzzy ensemble model using CNN for identifying TB from CXR images. The challenge of manual tuning is overcome by the *meta*-heuristic optimization algorithm. This algorithm is used to set the measures of fuzzy optimally during the models' training process. The ensemble technique's performance on the validation set is used to determine the best measures of fuzzy. The characteristics are retrieved from photos using three state-of-the-art methods before using the ensemble approach. The datasets used for this evaluation are the publicly available TB dataset of CXR images.

Medical approaches and computer vision provide considerable obstacles in medical image segmentation. Due to lung disease, difficulties occur in the lung and the varied dimensions of the lungs, as well as the structure and different types of medical imaging instruments. CT (Computed tomography) is a clinical diagnostic technology that has numerous systems based on CAD (computer-aided detection) that have been enhanced or established in conjunction with a computational approach derived from CT images. Xu, Y., et al. [15] presented lung segmentation models based on CT generalizations. Mask R-CNN (Mask Regions that employ Convolutional Neural Network) is used in combination with image processing techniques such as K-means region growth, fine-tuning of the Parzen window, and clustering. The outcome met expectations with 97 % accuracy, 99 % sensitivity, 98 % dice coefficient, and 97 % Positive Predictive Value (PPV) for the best MPK model. A second data set is utilized to improve the findings, and the outcomes demonstrate the system's effectiveness.

Lung disease is common all over the world. These contain asthma, pneumonia, fibrosis, chronic obstructive pulmonary disease, and

tuberculosis. Bharati, S. et al [2] utilized X-ray images to diagnose lung diseases using the hybrid deep learning (HDL) technique. Some tools are used for the implementation of this technique like Keras, Notebook, Jupyter, and Tensorflow. A dataset of NIH chest X-ray images was used to train the new model, which was downloaded from the Kaggle site. VDSNet has a 73 % verification accuracy throughout the whole database, whereas Vanilla RGB, Vanilla Gray, the hybrid VGG and CNN, and the modified capsule network have accuracy values of 69.5 %, 69 %, 63.8 %, and 67.8 %, respectively. The verification accuracy is slightly reduced when the sample database is used instead of the entire database, but VDSNet needs relatively little training time. Therefore, the proposed VDSNet infrastructure will facilitate the diagnosis of lung disease by experts and medical professionals.

Chetia, R. et al. [16] presented an ISQEENSAT (improved Sobel quantum edge extraction with non-maximum suppression and adaptive threshold) approach to suppress low noises in the chest and extract clinical information to study the COVID-19 pattern and presence in the lungs. The developed technique can detect very sharp and precise clinical edges of peripheral ground-glass opacity that appear in the early stages of Covit-19 patients, compared with conventional classical edge extraction operators. The findings of the edge extraction ensure the differentiation and detection of COVID -19 infection. ISQEENSAT can be a beneficial technique for helping COVID-19 examination and can assist the physician in determining how much of the region has been contaminated.

Kundu, R. et al [17] proposed an Ensemble of transfer learning with a conventional neural network (ET-NET). Faster COVID-19 screening is possible with the use of deep learning technology that makes use of chest CT-scan images. By integrating bootstrap or packed ensemble transfer learning models into three categories, the supervised image classification problem is addressed. ResNet34, DenseNet201, and Inception v3 were used to enhance certain models. A publicly accessible database is used to evaluate ET-NET, their developed framework.

Zhang et al.[18] developed a convolutional Siamese neural network with content-based image retrieval (CBIR-CSNN) to identify tuberculosis and lung cancer using CT images. The combination of content-based image retrieval (CBIR) and deep learning is used to identify lung cancer (LC) by analyzing nodular/mass atypical tuberculosis (NMTB) in CT images. The CBIR-CSNN has achieved great performance in the patch levels with 0.947 accuracies, 0.953 Mean Average Precision(MAP), and 0.970 of the area under the curve (AUC) values.

Deep learning has been widely used in the analysis of Chronic obstructive Pulmonary Disease (COPD) [32]. For multichannel lung sound analysis, various techniques such as Hibert Huang transform [32] and second order difference plot (SODP) [33]. Altan et al. [31] presented a deep Extreme Learning Machine (ELM) architecture using ELM autoencoder kernels such as lower upper triangulation ELM and Hessenberg decomposition ELM for Chronic Obstructive Pulmonary disease prediction. The main aim of the authors is to incorporate the benefits of the ELM autoencoder kernels such as rapid training and generalization capability. The study is mainly conducted using the 12-channel lung sound assessment.

Narh et al. [34] identified the impact of deep learning architectures in COVID19 prediction using the Local Histogram Equalization (LHE) technique. The experimental analysis is conducted using the dataset formed with images concerned with different disease categories such as Covid19 and pneumonia. The right and left lobes of the lung was segmented and the classification performance is evaluated using different disk value for the LHE. Narli et al. [35] also predicted COVID19 using high-performance transfer learning models and Contrast Limited Adaptive Histogram Equalization (CLAHE). The different transfer learning architectures utilized are AlexNet, MobileNet, VGG16, and DarkNet19. The crucial regions in the lungs such as airways and pathologies are preprocessed using the CLAHE technique. The CLLLAHE parameters utilized are a disk value of 56 and a clip limit of 0.2. The CLAHE-based transfer learning model offered a maximum accuracy of 95.878 %.

3. Materials and methods

3.1. Dataset description

COVID-19 dataset-1(COVID-19 radiography database): This dataset was created by a team of researchers from different locations such as Bangladesh, Pakistan, Qatar, Doha, etc. It consists of a total of 3616 COVID-19 images and 10,192 normal classes [24,25]. However, the database is periodically updated with new COVID-19 X-ray image samples. Fig. 1(a)-(c) presents the sample images present in the COVID-19 dataset-1 along with the different masks in Fig. 2.

Tuberculosis-dataset-2 (Tuberculosis (TB) Chest X-ray Database): A research team from Dhaka University, Bangladesh, and the University of Qatar, Doha, with help from a Malaysian collaborator from the doctors at Hamad Medical Corporation, Bangladesh, has developed a chest X-ray dataset for tuberculosis (TB) cases that are tested as positive along with the normal images. This dataset [29] contains 700 publically accessible TB images, 2800 images of TB that can be downloaded from the TB portal of NIAID by contract signing approach, and 3500 disease-free images.

NIH Chest X-ray Dataset-3 (National Institute of Health Chest X-ray Dataset): In this dataset [2], there are 112,120 X-ray images from 30,805 unique patients with disease labels. To create these labels from the accompanying radiological reports, the experts text-mine disease categories using natural language processing. This dataset contains 62 Pneumonia images and 84 Fibrosis images. Fig. 3 presents the sample images from Dataset-3 and 4.

3.2. Evaluation metrics

In order to evaluate the performance of the proposed method, the performance metrics such as accuracy, recall, specificity, precision, and f-score are utilized. These metrics are measured to illustrate the effectiveness of the proposed method. The detailed decryption of these metrics is given as follows,

- ***Accuracy*** (Ac)

It is the measurement of correctly predicted disease from the total number of disease records.

$$Ac = \frac{\rho_T + N_T}{\rho_T + N_T + \rho_F + N_F} \tag{1}$$

- ***Recall*** (Re)

Recall indicates the correctly predicted diseased classes from the total number of abnormal classes.

$$Re = \frac{\rho_T}{\rho_T + N_F} \tag{2}$$

- ***Specificity*** (Sp)

It is the ratio of correctly predicted non-diseased classes to the sum of true negatives and false positives.

$$Sp = \frac{N_T}{\rho_F + N_T} \tag{3}$$

- ***Precision*** (Pr)

The ratio of correctly predicted diseased classes to the total diseased records in the dataset.

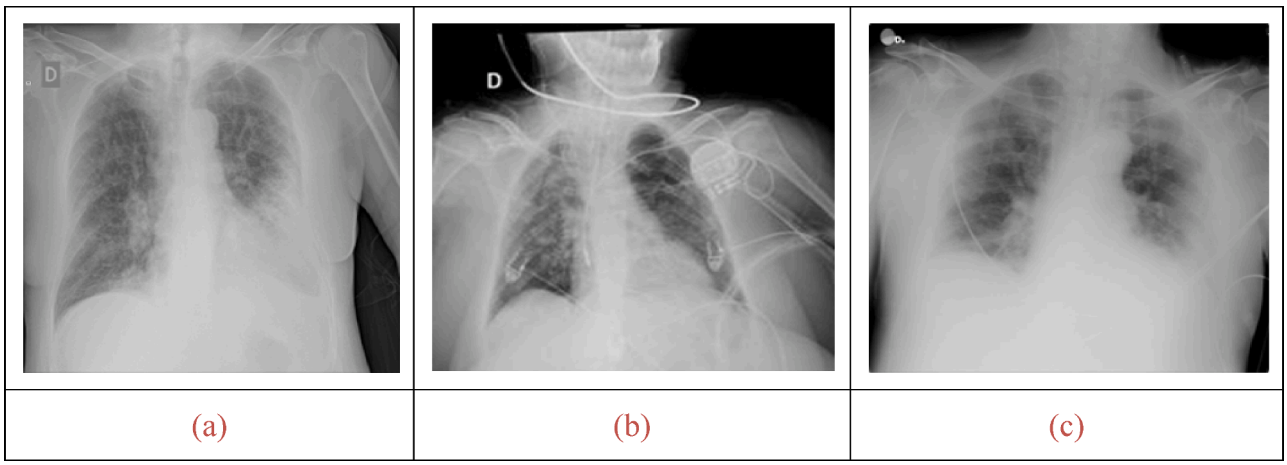

Fig. 1. Images present in the COVID-19 radiography database. (a)-(c): Sample images 1–3.

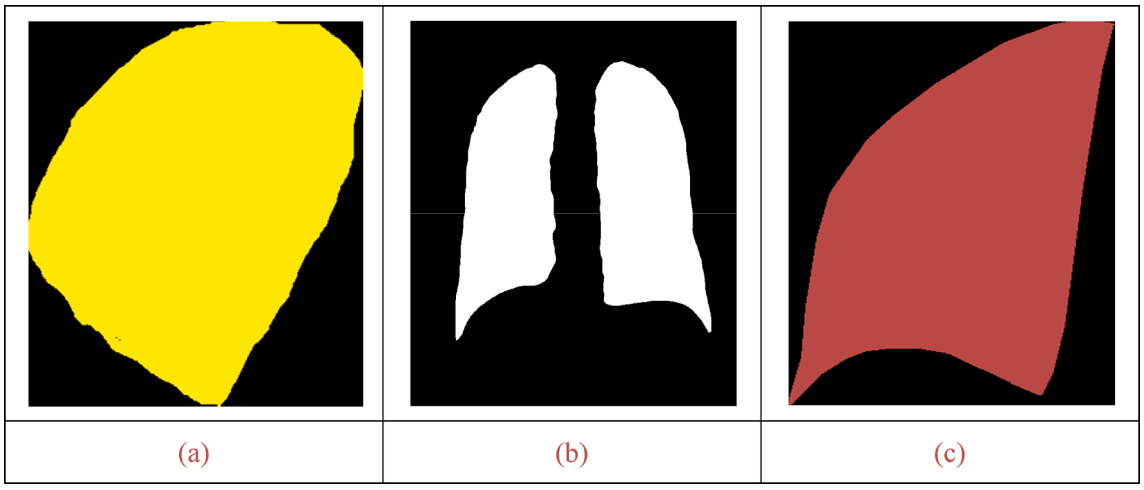

Fig. 2. Masks generated for the images present in COVID-19 dataset-2. (a)-(c): Masks 1–3.

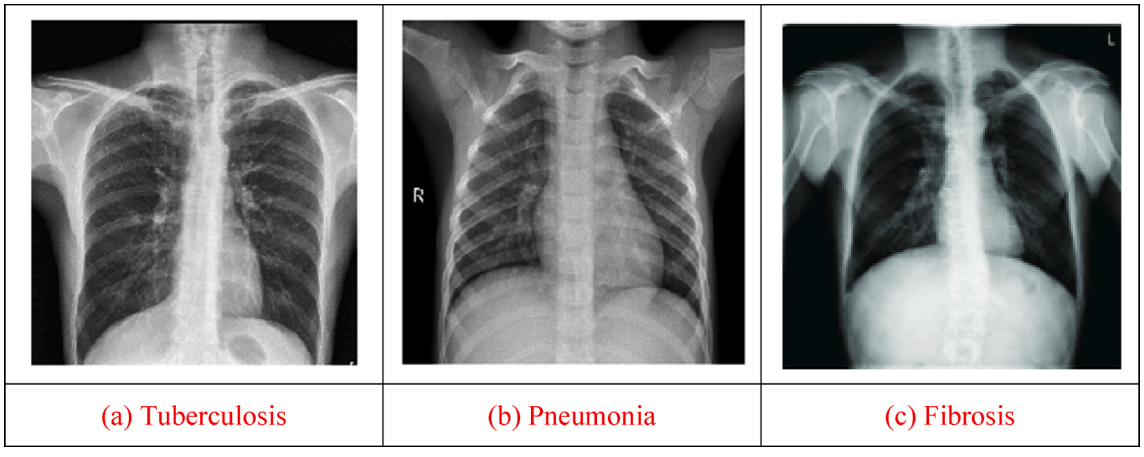

Fig. 3. Sample images present in dataset-3 and dataset-4.

$$Pr = \frac{\rho_T}{\rho_T + \rho_F} \tag{4}$$

- ***F-score*** (Fs)

F-score also known as F-measure achieves a tradeoff between precision and recall.

$$Fs = 2 * \left(\frac{Re \times Pr}{Re + Pr}\right) \tag{5}$$

From the above equations, the terms such as ρ_T,N_T,ρ_F and N_F

represents the true positives, true negatives, false positives, and false negatives respectively.

3.3. Proposed model design

Both the Mask RCNN and BiDLSTM have their own special features and the hybrid MR-CNN integrates them for predicting lung disease. The detailed descriptions of both the layers are presented in this paper and the proposed architecture is presented in Fig. 4. The Crystal algorithm is used to optimize the scalability and convergence issues in the Mask-RCNN model by hyperparameter tuning. The output layer of the MR-CNN is interconnected with the BiDLSTM cell features and several hidden layers. The BiDLSTM layer classifies whether the given image is COVID19 positive, tuberculosis positive, pneumonia positive, fibroid positive, or normal. A hybrid architecture is presented to learn the deep feature hierarchies from the Chest X-ray images. In this regard, we have added the crystal-optimized Mask R-CNN architecture for segmentation and fine-tuned the BiDLSTM layer for lung disease prediction. The final layers of the BiDLSTM architecture are customized based on the different target disease categories (COVID19, pneumonia, and tuberculosis).

The Crystal algorithm optimizes the Mask RCNN architecture and is modified to segment the appropriate lung disease regions from the Chest X-ray images. Mask RCNN is one of the recent additions to the Region-based CNN family and it was introduced in the year 2018. Due to the highest accuracy, it offered in the coco segmentation challenge, we have selected this model in our paper for segmentation. It mainly creates the mask at a pixel level for each object detected. To tune the Mask RCNN parameters based on the lung disease prediction task, we have utilized the crystal algorithm to tune the different hyperparameters associated with Mask R-CNN. It also overcomes the different problems associated with the ResNet backbone of Mask RCNN such as high memory usage, and poor segmentation accuracy, and offers improved processing speed. These problems are minimized using the Crystal algorithm in this work.

3.3.1. Preprocessing

The images are from several datasets of varying sizes and circumstances, and preprocessing is used to eliminate or minimize the effects on performance. Different types of X-rays are used with different technologies and resolutions. Eq. (6) represents the dataset normalization.

$$M_i = \frac{y_i - \min(y)}{\max(y) - \min(y)} \tag{6}$$

where y is for original images and M is a normalized image. According to Eq. (7), X is the standardized image. The mean and standard deviation (S.D) are used from the training set to unite the data distribution by applying standardization to the test set and validation [19].

$$X_i = \frac{M_i - mean(M)}{S.D(M)} \tag{7}$$

The histogram equalization is applied to extend the pixel intensity from its actual range (0–255). A better image was produced with somewhat higher contrast and intensity after histogram equalization. To minimize the noise in the input image a gaussian blur with a filter size of 5×5 size is applied and the edge information is preserved by applying a bilateral filter with a diameter, sigma color, and space values of 5,75,

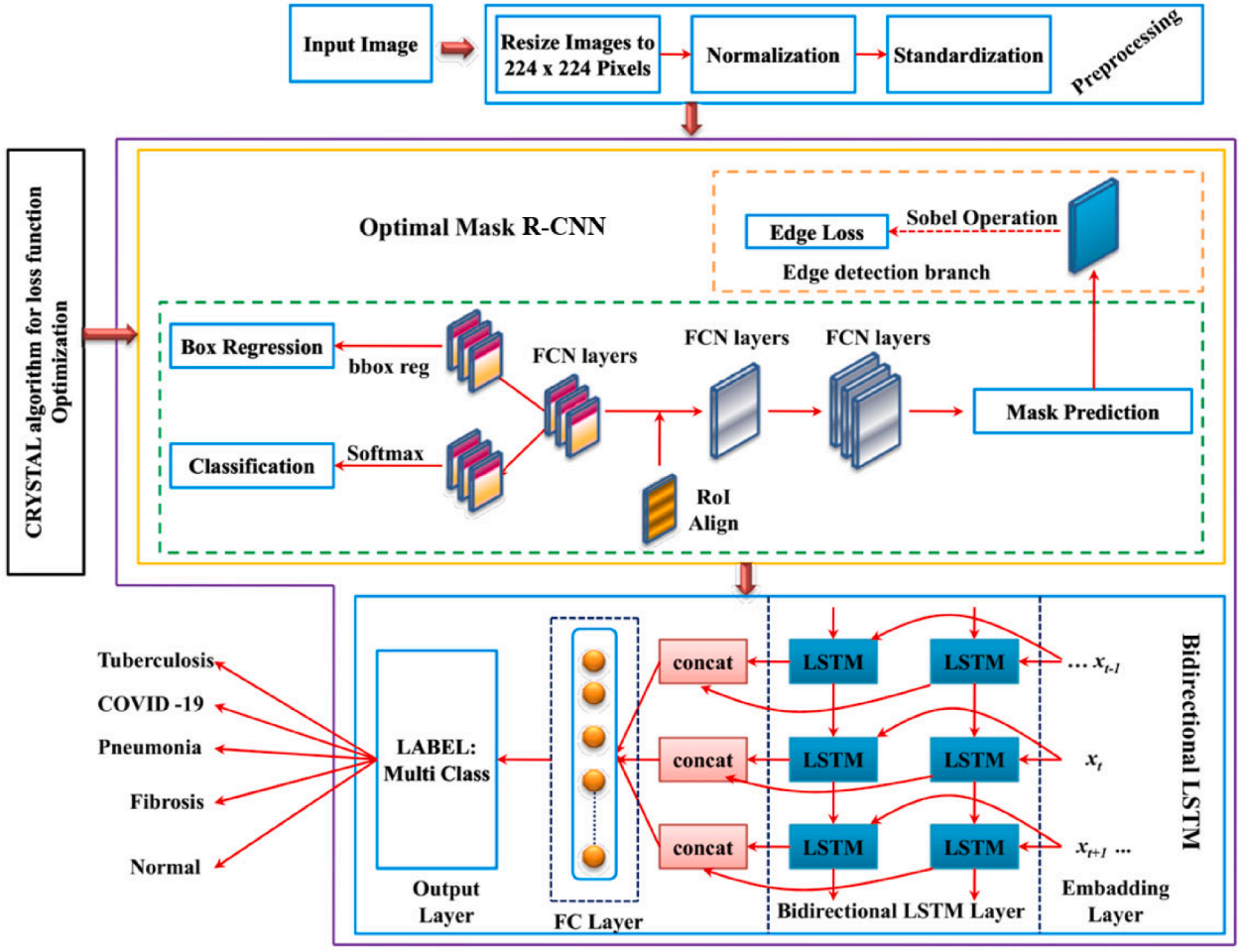

Fig. 4. Overall architecture.

and 75. To enhance the classification accuracy even further, the diaphragm is removed via an adaptive masking technique [30] by computing the maximum and minimum intensity of the pixel and applying binary thresholding. The binary threshold value is (100,255) and a blur value of (15,0). At last, the morphological closing operation is applied which results in an adaptive mask that eliminates the diaphragm from the input image. The different preprocessing operations applied to an actual image in the dataset are presented in Fig. 5.

3.3.2. Mask R-CNN

The Mask R-CNN constitutes two different phases (i.e. initial stage and final stage). In the initial stage, the ResNet FPN framework is employed in the extraction of hierarchical features. Following the extraction process, the region proposal framework generates the candidate object bounding box. The last phase includes class prediction as well as identifying the bounding box. In addition to this, the mask R-CNN is capable of generating binary masks for every individual ROI by employing a fully convolutional network. Also, the mask R-CNN employs ROI alignment (i.e. ROI alignment) for replacing the pooling process in faster R-CNN thereby performing coarse quantization which results in ROI misaligning as well as extracted features [21].

3.3.2.1. Path augmentation feature pyramid network. The high-level features are strongly interconnected with the overall objects in the CNN detectors and the low-level features mainly represent the local information. The high-level features are mainly integrated by introducing a top-down path in FPN for lung disease classification. The information loss due to pooling and deconvolution can be overcome by using the Path augmentation-based feature pyramid network (PAFPN) that adds a bottom-up path augmentation component to improve the low-level feature hierarchy. In this work, a short path with a total of 8 layers is inserted from the low layers to the topmost one. The hierarchical feature maps are computed by the feature extraction layer in a bottom-up manner. The feature maps generated in each layer are of the size $1/2, 1/4, 1/18, 1/16,$ *and* $1/32$ due to the presence of the max-pooling operations in between the convolutional layer. The feature maps are upsampled with a factor of 2.

The FPN created the PAFPN-generated features map with equal spatial sizes. In the addition of bottom-up paths augmentation in the feature, maps were represented by K_2, K_3, K_4, K_5, K_6 and related H_2 to H_6. Initially, K_i was processed into 3×3 convolutional layers of stride 2 at two times it was down-sampled. Therefore the down-sampled features map was connected by H_{i+1} . To generate P_{i-1}/P_1, the P_i value is upsampled two times and integrated with the convolutional part C_{i-1} of the final ResNet101 convolutional layer.

To generate the new feature map K_{i+1} with other 3×3 convolutional layers in the fused future maps. The feature fusions operation was carried out by higher features to lower levels for the contextual features of fusion. In the up-bottom feature fusions path were generated for the feature maps to be denoted as G_2, G_3, G_4, G_5, G_6 .

As a result, bounding boxes encircled the lung disease-affected areas, which were often elongated rectangles in shape.

3.3.2.2. Edge detection strategy. The input of a particular network is considered as a prediction outcome of the mask branch present in mask R-CNN. In order to attain an optimal output, an edge detection filter named Sobel edge detection filter is employed in performing the convolution operation with respect to the inputs. Finally, the differences among the edge detection outcome are computed as edge losses which are then finally added to network loss.

Here, the term edge detection losses are referred to as root mean square error (RMSE) among the true ones and the predicted outcome. The mathematical expression involved in determining the edge detection loss is stated in the following equation.

$$\mu_{(Edge)} = \frac{1}{n} \sum_{1 \leqslant j \leqslant n} \left[\left| Z_j - \hat{Z}_j \right| \right]^2 \tag{8}$$

From the above equation, the total number of images is n. The term Z_j and $\hat{Z}_j$ indicates the edge detection outcome and equivalent ground truth

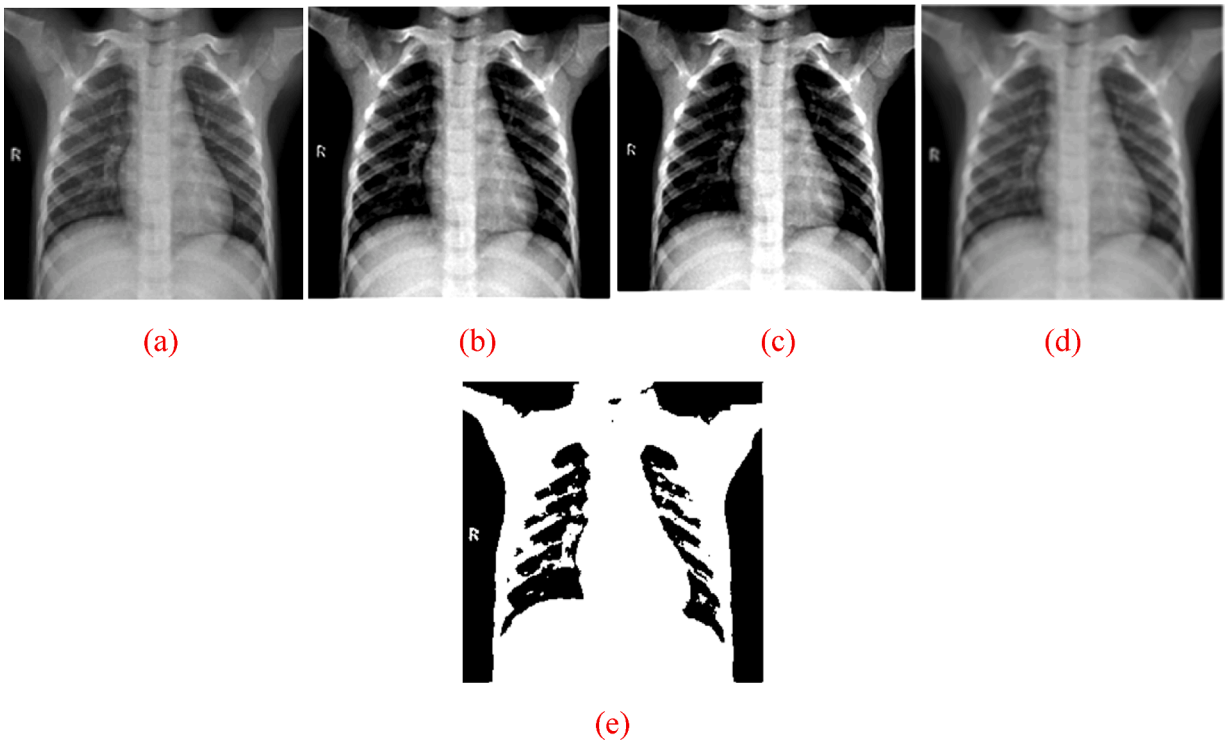

Fig. 5. Image preprocessing. (a) Actual image, (b) Input image with histogram equalization, (c) Application of bilateral filter, (d) Application of gaussian blur, and (e) application of binary thresholding.

of a *jth* image.

3.3.2.3. Loss function. The loss function for mask generation, box regression, and classification are the same as mask R-CNN (regions with the convolutional neural network) while the loss function for edge detection differs [22]. For every region of interest (ROI) sample, the four-loss function tasks are trained together to obtain multi-task loss (l).

$$l = l_M + l_B + l_C + l_E \tag{9}$$

The term l_M represents the mask generation loss, l_B depicts the box regression loss, l_C signifies the classification loss, and l_E implies the edge detection loss. In a faster R-CNN approach, the loss for classification and box regression are considered to be similar which is expressed as follows,

$$l_C + l_B = \frac{1}{n_C}\sum_x l_C\left(R_x, R_x^*\right) + \beta\frac{1}{n_B}\sum_x R_x^* l_B\left(T_x - T_x^*\right) \tag{10}$$

From the above equation, R_x depicts the forecasted probability of anchor x, R_x^* portrays the ground truth label of anchor, and x implies an anchor index in mini-batch. The value of R_x^* is one when the anchor x is positive and R_x^* is zero when anchor x is negative. Moreover, T_x is a vector coordinate of forecasted four parameters, T_x^* denotes the co-ordinates of ground truth, and β depicts the balancing parameter of two terms. The classification and regression are normalized using n_C and n_B terms.

The log loss function is applied for classification loss and the numerical expression is provided as,

$$l_C\left(\left\{R_x, R_x^*\right\}\right) = -R_x^*\log R_x^* - \left(1 - R_x^*\right)\log\left(1 - R_x^*\right) \tag{11}$$

In a similar way, the smooth l_1 loss is utilized for box regression,

$$l_B\left(\left\{T_x, T_x^*\right\}\right) = l_1^S\left(T_x - T_x^*\right) \tag{12}$$

The mask generation loss is defined using mean binary cross-entropy loss and it is given by,

$$l_M = -\frac{1}{M^2}\sum_{1\leqslant x,y\leqslant M} q_{xy}\log^{\circ} q_{xy}^d + \left(1 - q_{xy}\right)\log\left(1 -^{\circ} q_{xy}^d\right) \tag{13}$$

3.3.3. Abnormality merging

In the process of database construction, the full abnormal region is considered to be cropped into different image sets. Therefore, all of the identified images that include abnormal fluid accumulation are initially combined in order to merge the abnormal areas utilising the image location information that was earlier gathered during image acquisition. Assume that the predicted images are represented as $Z = \{Z_1, Z_2, .., Z_n\}$. The predicted abnormal regions are depicted as $P_d = \left\{P_x^d, x = 1, ..., n; d = 1, ..., D_x\right\}$. The terms P_x^d and D_x signifies the d^{th} predicted abnormal region in x^{th} image and total number of abnormal predicted regions in x^{th} image. The shortest distance between two neighboring abnormal regions is computed using the below equation,

$$G\left(S_x, S_y\right) = \sqrt{\left(S_{fp} - S_{yp}\right)^2 + \left(S_{fq} - S_{yq}\right)^2} \tag{14}$$

$$D = Min(Min(G)) \tag{15}$$

The term G indicates the matrix in which it defines the distance between S_f and S_y point sets, the coordinates of point sets S_f and S_y is represented as S_{fp},S_{yp},S_{fq} and S_{yq}. The proposed method combines nearby abnormal areas when they meet two requirements: (1) the shortest distance between the abnormal regions must be smaller than the threshold value, and (2) the same abnormal classes must be determined by comparing the mask values of the abnormal regions. This technique determines interconnected abnormal areas by analyzing surrounding abnormal regions for regional membership [23]. The hyperparameters of the MaskRCNN architecture is presented in Table 1.

Table 1
Mask RCNN hyperparameters.

Parameters	Values
Input	1
Outputs	5
Number of hidden units	100
Mini batch size	512
Learning rate drop factor	0.5
Initial learning rate	0.01
Number of iterations	50

3.3.4. Crystal structure algorithm (CryStAl)

The minerals in the solid are mainly formed of molecules, atoms, and ions which are mainly in a crystallographic order. The word crystal is a Greek word that means frozen by cold and the crystals possess both anisotropic and isotropic properties. There was also a popular belief that when water is stored at a very low temperature for more time it will become stable with a temperature increase. The lattice is the underlying component of the crystal and the basis mainly identifies the position of the atoms in the crystal structure associated with each lattice point (periodic array of points). Based on the combinations of the basis and lattice, the crystals are identified. The crystal configurations are mainly represented using the Bravais model in the CryStAl algorithm. By taking the infinite lattice shape, the periodic crystal structure is formed, and based on the position of the lattice point, the lattice pint value is acquired as shown below:

$$a = \sum m_i r_i \tag{16}$$

where i is the number of corners present in the crystal. The shortest distance vector between the principal crystallographic directions is represented as r_i and the integer value is represented as a_i. The crystal is statistically modeled as follows where each individual of the CryStaAl algorithm is an individual crystal in the search space. During initialization, the number of crystals is randomly identified for iterative purposes and the crystals (X) are identified as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{2i} \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} a_1^1 & a_1^2 & \cdots & a_1^j & a_1^{\partial} \\ a_2^1 & a_2^2 & \cdots & a_2^j & a_2^{\partial} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_i^1 & a_i^2 & \cdots & a_i^j & a_i^{\partial} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_m^1 & a_m^2 & \cdots & a_m^j & a_m^{\partial} \end{bmatrix} \tag{17}$$

The number of crystals along with their dimensionality is represented as m and ∂. In the search space, the crystal positions are randomly initialized as follows:

$$a_i^j(0) = a_{i,\min}^j + \kappa\left(a_{i,\max}^j - a_{i,\min}^j\right) \tag{18}$$

In the above equation, κ is a random number present in the interval [0,1]. The initial position of the crystals is taken as $a_i^j(0)$ and the maximum and minimum value that can be substituted is represented as $a_{i,\max}^j$ and $a_{i,\min}^j$. The crystals in the corners are taken as main crystals (X_{main}) which are identified based on the concept of crystallography. The values are mainly determined by taking the crystals generated initially and the current crystal X is omitted for the random selection. The best solution is represented as X_{best} and the mean value of randomly generated crystals is represented as μ. The position of the candidate solutions in the search space is updated using the basic lattice principles via four different types of updating processes as shown in Table 2.

In table-1, X_{new} and X_{old} mainly represent the old and new positions. The random numbers are specified as *rand1-rand3*. The different position update equations shown in Table 1 are responsible for the

Table 2
Different types of position updates in the crystal algorithm.

Name of the updating process	Equation
Normal cubicle	$X_{new} = X_{old} + X_{main}$
Cubicle with X_{best}	$X_{new} = X_{old} + rand1X_{main} + rand2X_{best}$
Cubicle with X_{mean}	$X_{new} = X_{old} + rand1X_{main} + rand2\mu$
Cubicle with both X_{best} and X_{mean}	$X_{new} = X_{old} + rand1X_{main} + rand2X_{best} + rand3\mu$

exploration and exploitation phases of the Crystal algorithm [28] from which the local and global searches are conducted. To handle the solution variables violating the boundary, a mathematical flag is used to define the variables outside the range. The boundary change of the violating variables is identified by the flag. Based on the maximum number of iterations, the termination condition is determined and the algorithm terminates after the predefined iteration value. The pseudo-code of the crystal algorithm is given in Algorithm-1.

Algorithm-1: Crystal algorithm.

Algorithm Crystal
Randomly generate the initial positions for the crystal
Compute the fitness value of each crystalWhile
(iter < Max_iteration)
For i = 1 to number of crystals
Generate X_{main}
Generate new crystals using $X_{new} = X_{old} + X_{main}$
Generate X_{best}
Generate new crystals using $X_{new} = X_{old} + rand1X_{main} + rand2X_{best}$
Generate μ
Generate new crystals using $X_{new} = X_{old} + rand1X_{main} + rand2\mu$
Generate new crystals using $X_{new} = X_{old} + rand1X_{main} + rand2X_{best} + rand3\mu$
If the new crystals generated violates the boundary condition then
Manage the position of the crystal and amend them
End If
Compute the fitness value of the new crystal
Update the value of the best solution found as the global best
End For
iter=*iter*+1
End While
Return Global_best
End **algorithm**

The Mask R-CNN has several issues associated with it and the main one is the convergence issue it can be overcome by using hyperparameter tuning instead of default parameters. The convergence issue mainly arises when the size of the global batch size increases up to 128. The crystal algorithm is mainly used to schedule the learning rate of the Mask-RCNN and also for hyperparameter tuning. The crystal algorithm decreases the initial learning rate and increases the number of iterations when there is a large number of gradients present in the initial iteration. The crystal algorithm mainly decays single times in the learning rate decay stage instead of three times. The Crystal algorithm also offers improved scalability with increasing batch size. The hyperparameters used to improve the Mask R-CNN performance by the Crystal algorithm are shown in Table 3. The Mask-RCNN model convergence is enhanced by the parameter tuning offered by the Crystal algorithm.

The Crystal algorithm objective function is mainly dependent on the image features taken for segmentation. This is established using a match operation that identifies the match that exists in the actual feature data and the solution vector feature. If a match occurs, then the feature is allocated a score. This mainly emphasizes that the feature selected is correct for segmentation. If a feature is an aid to have frequent occurrence then it is allocated a higher score.

Table 3
Hyperparameter optimization of the Mask R-CNN algorithm by the Crystal algorithm.

Parameters	Value
Learning rate decay stage	Only once in 10 epoch
Base learning Rate	0.01
Batch size	8
Linearly scaled batch size	0.24

$$Fitness = \frac{1}{M}\sum_{j=1}^{M}\frac{\left\{\sum_{i=1}^{F} S\left(match(X_j(i), A_l(i))\right)\right\}}{F} \tag{19}$$

where M is the number of features extracted from the input dataset, j is the solution space vector, i is the position, $A_l(i)$ is the actual feature in the l^{th} vector of the dataset, and the features corresponding to each vector are represented as F. The quality of each solution is identified by evaluating the solution vector $X_j(i)$. If a match is obtained between the $X_j(i)$ and $A_l(i)$ values then the score value S is incremented. The overall score is computed for each feature present in the vector and this score gives a matching value of a single vector in the dataset. The value of the optimal feature selected by the Crystal algorithm is checked with the feature values of the entire dataset. The solution to our optimization problem mainly has a maximum value since it is a maximization problem.

3.3.5. Bidirectional LSTM

The control flow of the conventional LSTM is the same as RNNs, and it analyzes the data that propagates information forward. The variant functions allow the LSTM to recall or forget information. The cell stage and various gates are the fundamental principles of LSTM. The cell state serves as a transit channel for pertinent data throughout the data processing process. It may be thought of as a memory of the network. The gates are a collection of neural networks that determine which information about a cell state is permissible. Throughout the training, the gates will explore which information is important to retain or forget. Three gates govern the information flows. They are, the input gate, output gate, and forgot gate. The input gate decides which data from the current state must be added. The output gate determines what the next concealed state should be. Finally, the forget state decides what information from the previous step must be retained.

The drawback in LSTM classification was improved by implementing bidirectional LSTM which strengthens the features of LSTM. In this case, two types of LSTM input data were trained: original input data and a reversed replica of input data. However, BiDLSTM [20] is considered a straightforward technique. In this strategy, the first recurrent layer is initially replicated, and then the original form of input data is presented as input to the first layer, while a reversed replica input data is provided to the duplicated layer. This approach is used in ordinary RNNs to solve the problem of vanishing gradients.

With the forward and backward hidden layer, the BiDLSTM has operated its input data in two directions i.e. from right to left and left to right. The mathematical expression for forwarding hidden layer (→Gr), backward hidden layer (←Gr), and the output (Z) is given below:

$$\overrightarrow{G}_r = \mathrm{B}\left(y_{q\overrightarrow{G}} q_r + y_{\overrightarrow{G}\overrightarrow{G}}\overrightarrow{G}_{r-1} + u_{\overrightarrow{G}}\right) \tag{20}$$

$$\overleftarrow{G}_r = \mathrm{B}\left(y_{q\overleftarrow{G}} q_r + y_{\overleftarrow{G}\overleftarrow{G}}\overleftarrow{G}_{r-1} + u_{\overleftarrow{G}}\right) \tag{21}$$

$$Z_r = y_{\overrightarrow{G}Z}\overrightarrow{G}_r + y_{\overleftarrow{G}Z}\overrightarrow{G}_r + u_Z \tag{22}$$

From the above equations, the hidden layer is depicted as B, the weight matrices are represented as y whereas $y_{q\overrightarrow{G}}$ represents the forward input hidden weight matrix and the backward input hidden weight matrix was denoted as $y_{q\overleftarrow{G}}$. The variable u represents the bias vector and $u_{\overrightarrow{G}}$,$u_{\overleftarrow{G}}$ represents the bias vector in the forward and backward directions. The BiLSTM outputs the probability of the image into five classes namely normal, COVID19, tuberculosis, fibrosis, and pneumonia. The class with the highest probability is chosen as the final outcome. The input taken by the bidirectional LSTM is a total of 224, 648 images

which were divided into three sets namely training (70 %), testing (20 %), and validation (10 %). The experiments have been conducted a total of three times to prove their effectiveness and we didn't perform data augmentation due to the balanced number of training samples. The hyperparameters of the BiLSTM algorithm are presented in Table 4.

4. Results and discussions

This section compares the efficiency of the proposed methodology with different state-of-art techniques such as CMT-CNN, RcoNet, DNN, CNN, Mask R-CNN, HDL, ISQEENSAT, ET-NET, and CBIR-CSNN. The dataset description along with the different evaluation metrics used is also presented in this section. The proposed methodology is implemented in python using an 11th Gen Intel Core i5-11400 processor - equipped with a 1 TB 7200RPM 3.5″ SATA HDD hard drive. The different models taken for comparison along with their parameter settings are presented in Table 5.

4.1. Comparative analysis results

The graphical representation of model accuracy with respect to different epochs is illustrated in Fig. 7 (a). The training and testing accuracy of the proposed method is evaluated to determine its efficiency. The experimental investigation proves that both the training and testing accuracy of the proposed model goes beyond 90 % after the 10th epoch. This evaluation demonstrates clearly that the proposed model has a higher tendency to detect and diagnose lung diseases effectively in the early stage.

Fig. 6 (b) demonstrates the training and testing loss analysis of the proposed model with respect to different epochs. The experimental analysis is conducted to validate the effectiveness of the proposed model. The result portrays that the training and testing losses of the proposed model are minimized extensively after the 25th epoch. Also shows that the training loss of the proposed model decreases more than the testing loss. The minimized training and testing losses ultimately reduce the model's complexities like over-fitting effectively. The performance metrics such as precision, recall, specificity, F-score, and accuracy are computed. The performance metrics of Mask R-CNN and optimized Mask R-CNN are evaluated under different lung disease datasets. Table 6 depicts the Mask R-CNN and Optimized Mask R-CNN performance for multiclass classification. The proposed method has a higher precision rate of 98.78 % and 98.70 % on the COVID19 and tuberculosis datasets. Also, the testing accuracy is maximum for optimized Mask R-CNN compared to the Mask R-CNN on different classes. The optimized Mask R-CNN is an efficient method for determining lung diseases on the basis of precision, recall, specificity, F-score, and accuracy.

The cross-validation score for the multiclass lung disease classification problem is shown in Fig. 7. The mean score value for the categorization of multiclass lung disease is shown by the dotted line in the graph. The performance of the model was determined by implementing the stratified K-fold cross-validation with 10 folds of the K set. Initially, the dataset is shuffled in this process and then divided into K groups. From the figure, the mean validation score of multiclass lung disease classification is 0.997. The graph is plotted with a stratified score against a validation score. The validation score for 10 folds was computed by repeating the process until each K-folds operates as test data.

Table 4
BiLSTM hyperparameters.

Parameters	Values
Input	1
Outputs	5
Number of hidden units	100
Mini batch size	512
Learning rate drop factor	0.5
Initial learning rate	0.01
Number of iterations	50

Table 5
Parameter setting of state-of-art architectures.

Model	Parameters
CMT-CNN [11]	Momentum = 0.9 Weight decay = 0.0005 Number of epochs = 20 Batch size = 4 Learning rate = 0.01
RCoNet [12]	Learning rate of 2×10^{-4} Weight decay factor = 1×10^{-4} Batch size = 8 drop rate of each dropout layer ={0.1, 0.3, 0.5}
DNN [13]	Number of epochs = 10 Optimizer = Adam Number of hidden layers = 3 Batch size = 100 Learning rate = 0.001 Activation function = Linear, ReLU Regularization strength = 0.001
CNN [14]	Mini Batch size = 500 Maximum number of epochs = 15 Momentum = 0.9 Learning rate drop factor = 0.2 Initial learning rate = 0.05
Mask R-CNN [15]	Iteration = 100 Batch size-4 Learning rate = 1e-6 Optimizer = Adam Anchor aspect ratio = [0.5,1.2] Anchor scale = [8,16,32,64,128]
HDL [2]	Number of convolutional layers = 10 Max-Pooling after layers = 2, 4, 7, and 10 Output layer = Softmax Localization network learning rate = 0.001 Batch size = 64
ISQEENSAT [16]	Threshold T1 = 0.1 and T2 = 0.9 Batch size = 4 Learning rate = 1.10^{-4} Drop Length of the detection region = 7 Edge length = 7
ET-NET [17]	Loss function = Binary Cross-Entropy Loss Batch size of InceptionV3, ResNet34, and ResNet201 is 16,32, and 32 Learning Rate0.001 Momentum = 0.99 Input Size = 299x299x3 (Inception v3), 224x224x3 (Others)
CBIR-CNN [18]	Activation Function = ReLU Pooling = Max Pooling Dropout value = 0.5 First Dense LayerLinear(30976, 1024) , Second Dense LayerLinear(1024, 128) , Final Dense LayerLinear (128,2)

The accuracy of the optimized Mask R-CNN-BiDLSTM evaluation with the existing techniques is shown in Fig. 8. The evaluation was carried out under multiple lung disease classes such as fibrosis, COVID19, tuberculosis, and pneumonia. In comparison to previous approaches, the optimized Mask R-CNN has maximum accuracy of 95.71 % for COVID19 and 89.56 % for TB. The proposed method was compared with CMT-CNN, DNN, Mask R-CNN, ET-NET, and CBIR-CSNN. The optimized Mask R-CNN-BiDLSTM has better detection accuracy and enhances the performance of conventional LSTM.

Table 7 illustrates the results obtained by comparing the proposed technique with the existing literary works in terms of accuracy, precision, sensitivity, specificity, and F1 score. The better results are obtained for this model mainly by employing the crystal algorithm and the BiDLSTM architecture along with the Mask R-CNN network. The accuracy and robustness of the proposed methodology mainly lie in evaluating the patient images obtained from different countries. Even though

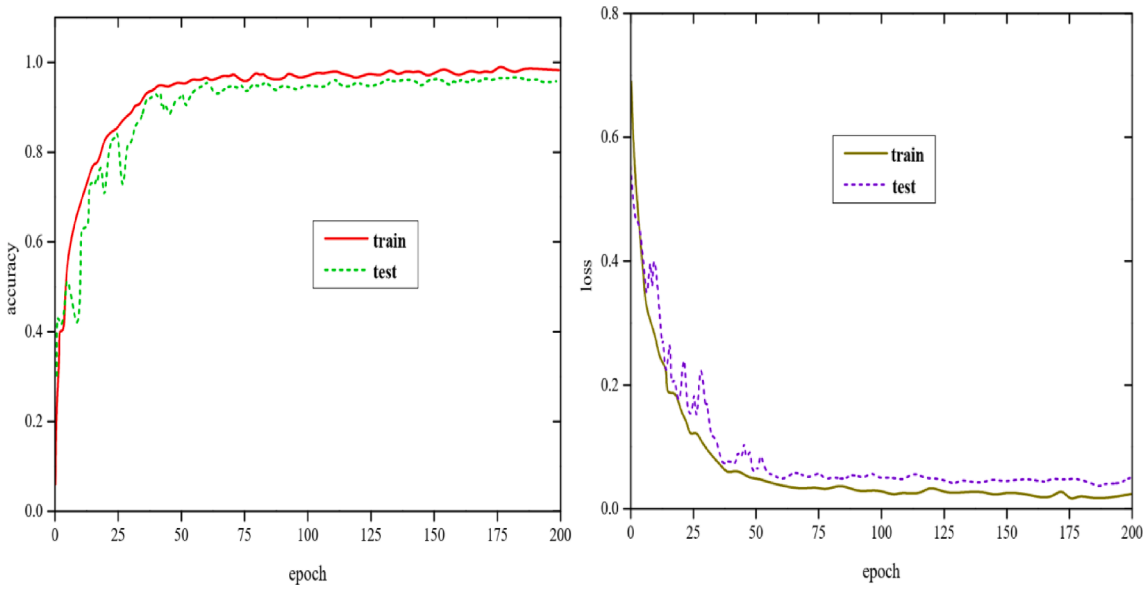

Fig. 6. Training and testing analysis. (a) Accuracy and (b) Loss.

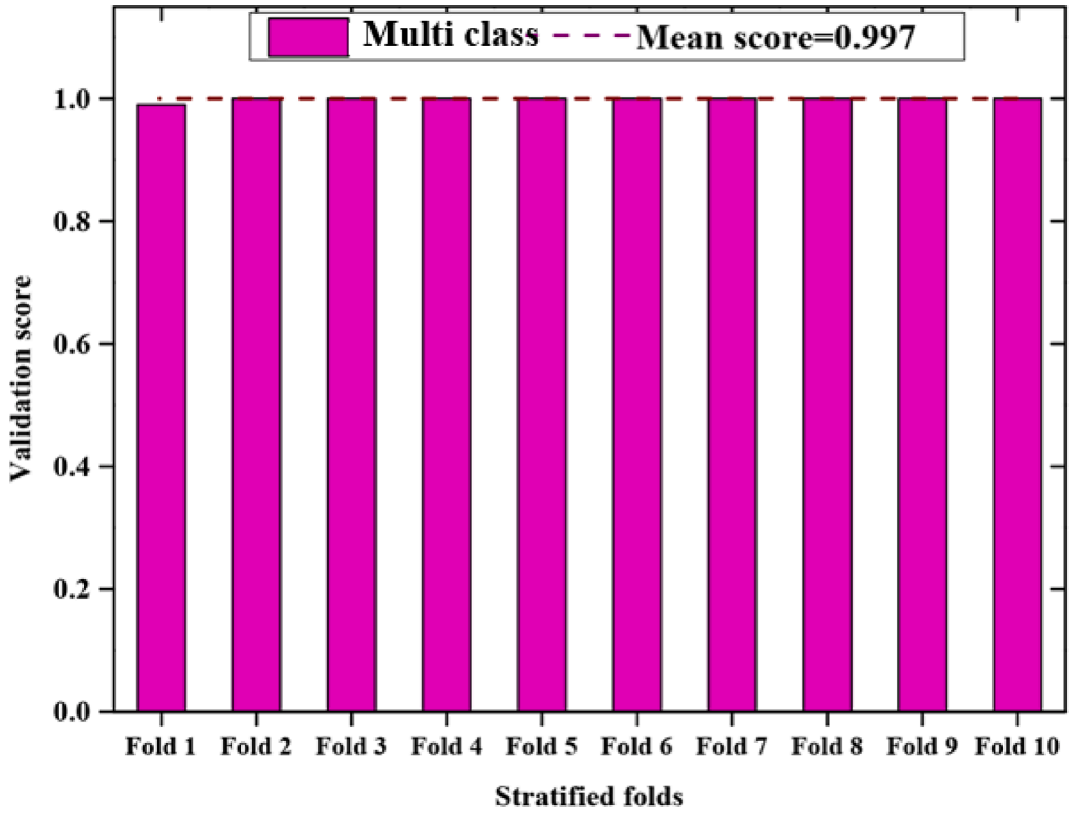

Fig. 7. The cross-validation score of multiclass lung disease classification.

the existing techniques offer improved accuracy it often fails to minimize the time consumption and complexity. The accuracy drop in different techniques such as DNN and CNN is due to the absence of efficient preprocessing techniques.

Table 8. evaluates the Performance of Optimized Mask R-CNN and Bidirectional LSTM in terms of different disease classes. The results show that the proposed Optimized Mask R-CNN and Bidirectional LSTM network provide better results based on the parameters such as precision, F1-score, Recall, and Accuracy. The accuracy of the proposed methodology for identifying different diseases such as COVID-19, Tuberculosis, Fibrosis, and Pneumonia are 99 %, 97.86 %, 95.93 %, and 95.90 % respectively.

4.2. *Ablation study*

To evaluate the efficiency of the proposed model with the addition of the crystal algorithm optimized MaskRCNN architecture we have conducted the ablation experiments and it is presented in Table 9. The

Table 6
Mask R-CNN and Optimized Mask R-CNN performance for multiclass classification.

Performance metrics	Mask R-CN- BiDLSTM				Optimized Mask R-CNN- BiDLSTM			
	COVID19 (%)	Tuberculosis (%)	Pneumonia (%)	Fibrosis (%)	COVID19 dataset-1 (%)	Tuberculosis (%)	Pneumonia (%)	Fibrosis (%)
Precision	98.61	96.81	94	93	98.78	98.70	97	96
Recall	87.72	82.59	93.90	92	97.82	97.86	98.3	96.70
Specificity	96.42	89.62	95.45	91	99.16	98.85	98	96.71
F-score	92.69	85.71	94.77	91.3	98.38	97.95	98.2	96.35
Accuracy	90.36	81.64	93	91	98.71	97	98	96

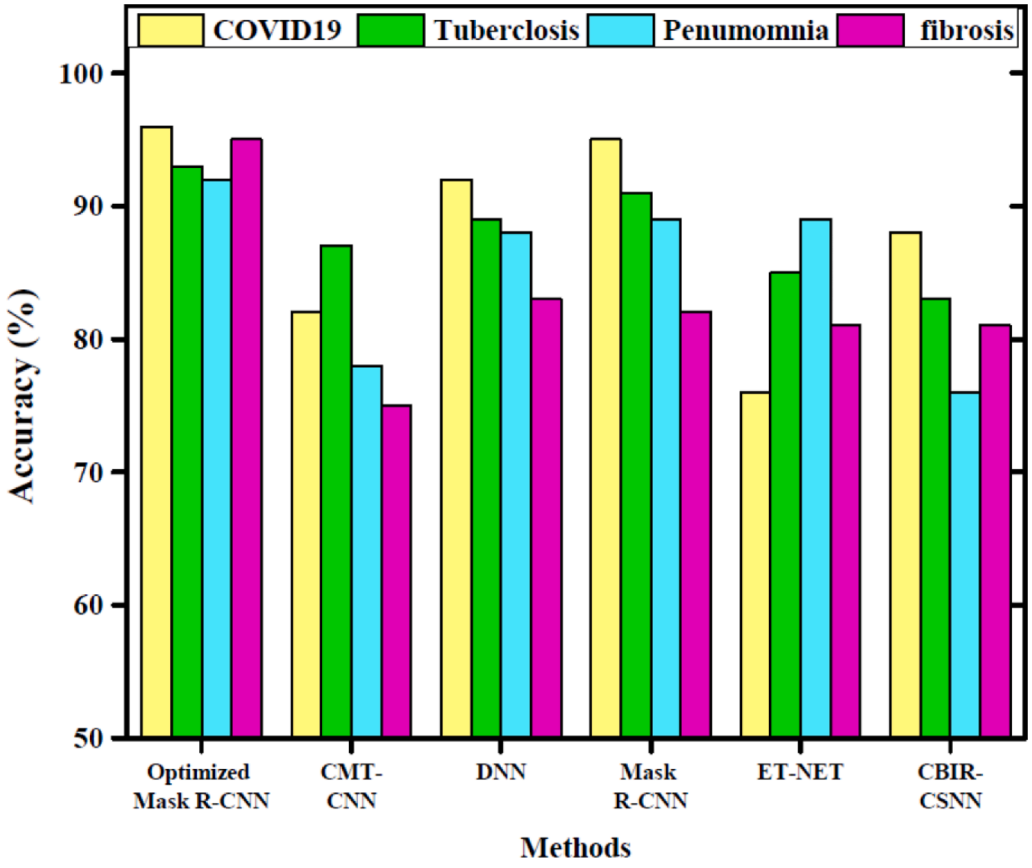

Fig. 8. Accuracy calculation of Optimized Mask R-CNN with the existing techniques.

Table 7
Performance evaluation with different state-of-art techniques in terms of different modalities.

References	Technique title	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
[11]	CMT-CNN	95.2	94.8	94.3	95.6	96.5
[12]	RCoNet	94.4	92.2	97.8	96.5	94.0
[13]	DNN	92.2	92.0	91.1	92.1	91.2
[14]	CNN	92.0	91.0	92.1	91.5	–
[15]	Mask R-CNN	93.2	–	95.4	94.3	–
[2]	HDL	96.0	–	–	92.1	97.3
[16]	ISQEENSAT	94.3	94.6	93.1	91.4	94.3
[17]	ET-NET	92.2	98.0	95.4	93.6	92.5
[18]	CBIR-CSNN	94.6	–	98.4	94.4	94.5
Proposed	Optimized Mask R-CNN and Bidirectional LSTM	99.16	99.18	99.16	98.61	99.14

proposed model is mainly composed of image preprocessing, segmentation, and classification. Initially, the MaskRCNN is removed and the BiLSTM architecture is adopted for classification. The proposed model is evaluated in terms of accuracy, precision, recall, and F1 score. The experiments are conducted for the COVID19 radiography datasets. The accuracy index mainly decreased when the model was evaluated using Mask RCNN alone. The accuracy, precision, recall, and F1-score of the Mask RCNN + BiLSTM are 93.54 %, 94.54 %, 93.36 %, and 93.56 %. A 7.11 % improvement can be noticed when the crystal algorithm and Mask RCNN are integrated with the BiLSTM classifier. After the

Table 8
Performance evaluation of optimized Mask R-CNN and bidirectional LSTM.

Disease classes	Performance Evaluation Metrics			
	Precision (%)	Accuracy (%)	F1-Score (%)	Recall (%)
Covid-19	99	99	99	99
Tuberculosis	98.70	97.86	98.85	97.95
Fibrosis	96.71	95.93	96.66	96.75
Pneumonia	95.35	95.90	96.45	96.77
Normal	97	97	97	97

preprocessing techniques are applied, the proposed model offers an improvement nearly equal to 1 % in accuracy, precision, recall, and F1 scores when compared to the Crystal algorithm + MaskRCNN + BiLSTM model without preprocessing.

4.3. Discussions

In this study, we present an AI platform for different lung disease predictions such as COVID19, tuberculosis, and pneumonia. By employing a Hybrid architecture our model offered a competitive performance in the baseline datasets. The results derived in the previous section show that the proposed model offers optimal performance in lung disease prediction. With the help of the Crystal optimized Mask RCNN architecture and the BiDLSTM architecture, the proposed model outperforms different state-of-art techniques such as CMT-CNN [11], RCoNet [12], DNN [13], CNN [14], Mask R-CNN [15], HDL [2], ISQEENSAT [16], ET-NET [17], and CBIR-CNN [18]. The efficiency of the proposed model is determined in terms of different performance metrics such as Precision, Accuracy, F1-Score, and Recall. The AUC-ROC curves shown in Fig. 9 also support the performance. When compared to the existing methods in terms of Accuracy, F1-Score, and Recall, the proposed model offers optimal performance. The essential factor in determining lung disease via X-ray images is to observe the white spots also known as infiltrates and fluid surrounding the lungs.

The proposed work mainly aims to capture the different lung disease patient conditions and classify them into fibrosis, COVID19, tuberculosis, and pneumonia. For accomplishing this task, we have utilized a total of 224,648 images in the dataset for a total of 1,726 iterations. The Crystal algorithm is implemented for selecting the crucial features in the image for segmentation by identifying the final hyperparameters for the Mask RCNN architecture. For comparison, we have taken the Particle Swarm Optimization (PSO) [38] and Genetic Algorithm (GA) [39] for analysis, and the results of the convergence performance are presented in Figure. The number of solutions, upper and lower limit of segmentation features, number of iterations, and dimensionality are set as 20, (1,0), 10, and 22 respectively. The parameters of the GA and PSO are also set in the same manner. The convergence performance of the proposed model is shown in Fig. 10. The results show that the proposed model converges when it reaches the 100th iteration with an overall fitness value nearly equal to 0.7. The results show that the Crystal algorithm achieved the optimal feature set for the segmentation tas k of the MaskRCNN.

The precision of a prediction outcome can be measured via the 95 % Confidence Interval (CI) by placing a sequence of values consistent with the data. The prediction outcome is said to be precise when the confidence interval is very narrow [36,37]. The sensitivity and specificity are reported here with a 95 % confidence interval for more than 100 randomized trials. The sensitivities and specificities of the proposed model are verified in terms of a Receiver Operating Characteristics (ROC) curve. The diagnostic performance shows that the likelihood ratios of the proposed model are statistically similar with a 95 % confidence interval with the human experts. The proposed model was evaluated using the testing dataset and the ROC curve plots the area between the Sensitivity (true positive rate) and 1-specificity (False positive rate). The proposed hybrid architecture was evaluated in terms of ROC curves using both training (70 %) and validation data (30 %) with the BiLSTM architecture. The results in Fig. 11 (a) and (b) show that the hyperparameter tuning improves the overall performance of the model. During training, the AUC of the proposed Hybrid BiLSTM is 0.98 whereas the AUC value of the BiLSTM architecture is 0.91. The AUC value for the validation dataset is 0.94 which is relatively high when evaluated using the BiLSTM model alone.

The Dice Similarity Coefficient (DSC) mainly measures the similarity between the target and predicted masks. Here we mainly measure the

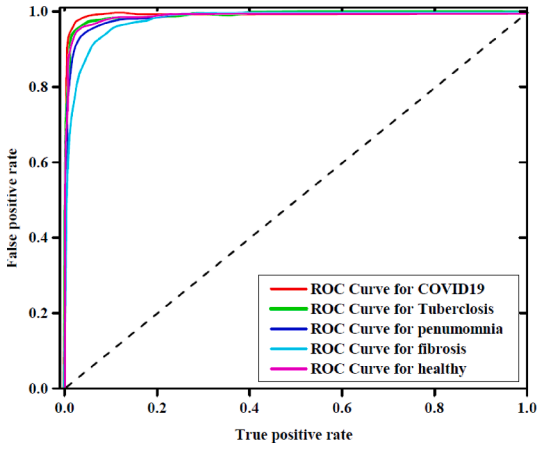

Fig. 9. ROC curve for multiclass lung disease prediction.

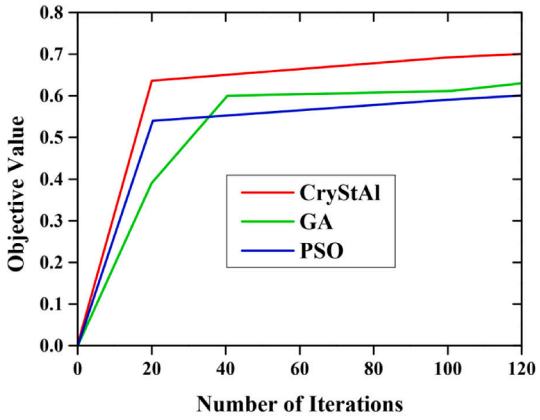

Fig. 10. Convergence analysis of the Crystal algorithm.

Table 9
Results of the ablation experiments.

Model	Accuracy	Precision	Recall	F1-score
BiLSTM	90.54	90.21	89.54	90.14
MaskRCNN	89.56	84.45	87.65	87.65
MaskRCNN + BiLSTM	93.54	94.54	93.36	93.56
Crystal algorithm + MaskRCNN + BiLSTM	97.65	97.58	96.65	97.65
Proposed model	98.71	98.78	97.82	98.38

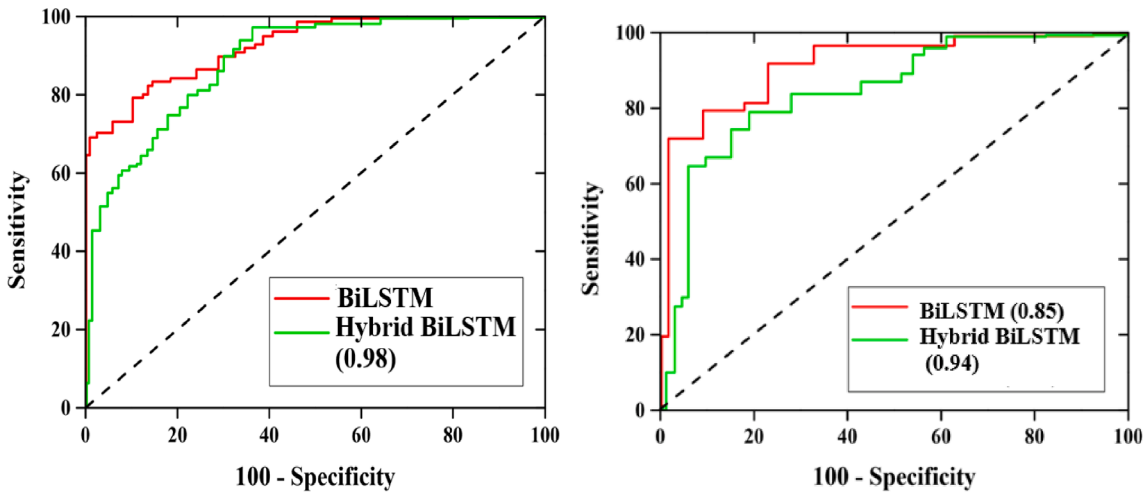

Fig. 11. ROC curves. (a) Training and (b) Validation.

spatial overlap between two target segmentation target regions X and Y as shown follows:

$$DSC(X,Y) = \frac{2(A \cap B)}{(A+B)} \tag{23}$$

The Jaccard Index (JI) is a performance metric that is closely interrelated with the DSC metric and it mainly computes the Intersection over Union (IoU) tactic to identify the proportion of overlap between the predicted and target mask. It is mainly used as a loss function during training.

$$JI(T,P) = \frac{T \cap P}{T \cup P} \tag{24}$$

where T and P represent the target and prediction scores.

The segmentation accuracy of the proposed Crystal algorithm optimized MaskRCNN model is compared with different state-of-art techniques such as DeIM [12], ISQEENSAT [16], Mask RCNN [15], LHE [34], and CLAHE [35]. Table 10 presents the Accuracy, JI, and DSC scores obtained from different models. The statistical analysis is conducted using different metrics such as Dice Similarity Coefficient (DSC) and JI. The results show that the proposed model offers minimal non-overlapped patches and it is less affected by different abnormalities. The proposed model segments the infiltrate region accurately when compared to the existing techniques such as CLAHE and LHE. The proposed model misclassifies certain instances in patients affected with Ground Glass Opacity, cardiomegaly, and consolidation. The misclassification is mainly due to the patches such as mid axillary, midclavicular, and axilla in these regions. The results in Table 10 demonstrate that the proposed model outperforms other state-of-art techniques in terms of segmentation.

The results of the proposed model are also visualized in terms of the Grad-CAM technique and the affected regions of the lung are presented in terms of different colors. In Grand-CAM visualization, the importance decreases from red to blue areas. Fig. 12 presents the Grand-CAM output of the proposed model for different disease classes. The results show that the proposed model offers optimal performance in determining lung disease features.

Table 10
Comparative analysis using segmentation techniques.

Models	Accuracy	JI	DSC
DeIM [12]	0.9121	0.8958	0.9145
ISQEENSAT [16]	0.9145	0.9014	0.9104
Mask RCNN [15]	0.8975	0.8965	0.9187
LHE [34]	0.8975	0.8947	0.8954
CLAHE [35]	0.8356	0.8236	0.8547
Proposed Crystal algorithm optimized MaskRCNN architecture	0.9654	0.97487	0.9856

The matrix representation of the overall performance of the proposed model is referred to as the confusion matrix. Fig. 13 describes the confusion matrix of five cases: COVID-19, tuberculosis, pneumonia, fibrosis, and normal classes. The letters in bold represent the number of classes predicted correctly. For the fibrosis class, the proposed model predicts a total of 99 % of images correctly as fibrosis, and 1 % of the image is predicted as tuberculosis.

4.4. Limitations

The proposed model is reliable when it comes to identifying different lung diseases such as COVID19, Tuberculosis, and pneumonia. The proposed model faces different issues when implemented in the real-time medical decision support system. The main issue is accepting the responsibility for the hybrid Crystal optimized BiDLSTM- Mask-RCNN model responsibility. In this scenario, we have to notice one point our proposed model is mainly designed to improve the decision-making capability of the physician and not to replace them. To assist the physician one step further, we have incorporated the details such as how the proposed model works and classifies the disease of the patient. We have mainly incorporated these details to help the physician analyze the purpose of the module and to provide the reason why the image was classified as belonging to a certain category. The other main issue faced by our model is the quality of the images used during training. We obtained the images from different datasets such as the COVID-19 radiography database, the Tuberculosis (TB) Chest X-ray Database, and the National Institute of Health Chest X-ray Datasets. These datasets are mainly selected based on the fact that they are correctly labeled by an expert. In a chest X-ray analysis system, image labeling is the most laborious, time-consuming, and complex task. Since COVID19 is a recently identified disease, the number of samples acquired for this disease is limited. Even though some commercial companies are working on COVID19 datasets based on chest X-rays, their ultimate objective is to keep them private.

5. Conclusion

A hybrid Crystal optimized BiDLSTM- Mask-RCNN model is developed in this paper to detect lung disease leveraging X-ray images. The COVID19 radiography dataset, the Tuberculosis (TB) Chest X-ray Database, and the National Institute of Health Chest X-ray Dataset are the four lung disease datasets used in this study. The precision rate and

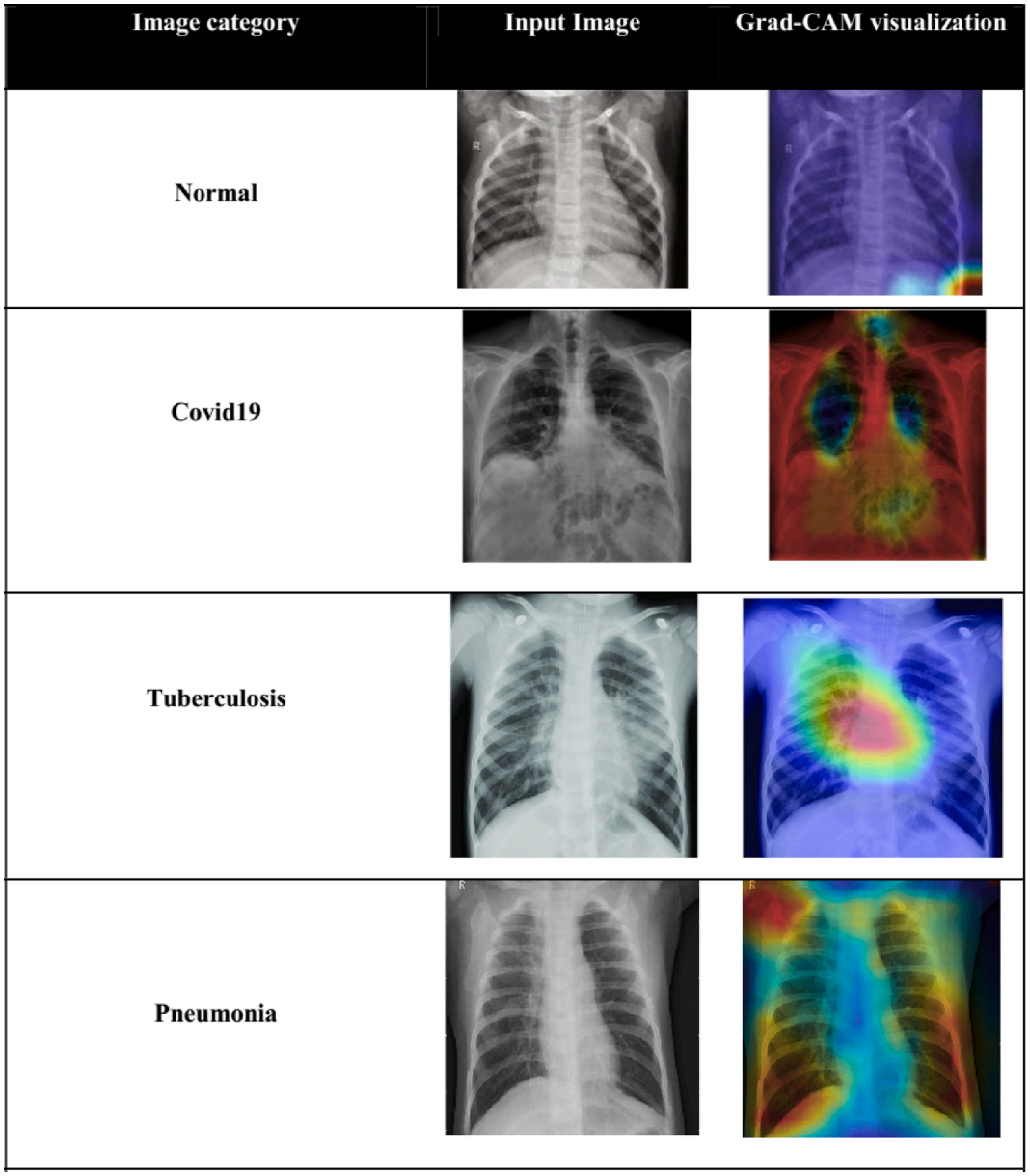

Fig. 12. Grand CAM visualization for different lung disease classes.

	COVID19	Tuberculosis	Pneumonia	Fibrosis	Normal
COVID19	**99%**	0%	1%	0%	0%
Tuberculosis	0%	**99%**	0%	1%	0%
Pneumonia	1%	0%	**98%**	1%	0%
Fibrosis	0%	1%	0%	**99%**	0%
Normal	0%	0%	0%	0%	**100%**

Fig. 13. Confusion matrix.

the testing accuracy are higher for optimized Mask R-CNN compared to other methods. The logistic value for Mask R-CNN, LSTM, and optimized Mask R-CNN is measured as 0.93, 0.95, and 0.99 respectively. The cross-validation score and mean score of the classification problem of multi-class classification were determined. The performance metrics namely accuracy, recall, precision, f-score and specificity are determined to examine the capability of the proposed method. Moreover, the proposed model's efficiency is checked by analyzing the training and testing accuracy as well as testing and training loss. The results showed that the proposed method achieves a higher percentage of testing and training accuracy and less testing and training loss. In the future, we plan to identify the severity of the COVID19 disease in patients using images obtained from more than one modality.

Ethical statements:

Funding: Not applicable

Ethics approval:

Compliance with Ethical Standards

Human and Animal Rights

This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material:

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Code availability: Not applicable.

CRediT authorship contribution statement

Varadharajan Indumathi: Conceptualization, Methodology, Software, Visualization, Investigation. **Rathinavelayutham Siva:** Data curation, Writing – original draft, 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.

Data availability

No data was used for the research described in the article.

References

[1] R. Kundu, R. Das, Z.W. Geem, G.T. Han, R. Sarkar, Pneumonia detection in chest X-ray images using an ensemble of deep learning models, PLoS ONE 16 (9) (2021) e0256630.

[2] S. Bharati, P. Podder, M.R.H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, Inf. Med. Unlocked 20 (2020), 100391.

[3] D. Keidar, D. Yaron, E. Goldstein, Y. Shachar, A. Blass, L. Charbinsky, I. Aharony, L. Lifshitz, D. Lumelsky, Z. Neeman, M. Mizrachi, COVID-19 classification of X-ray images using deep neural networks, Eur. Radiol. 31 (12) (2021) 9654–9663.

[4] W. Ausawalaithong, A. Thirach, Marukatat, S. Wilaiprasitporn, T. Automatic lung cancer prediction from chest X-ray images using the deep learning approach. In 2018 11th Biomedical Engineering International Conference (BMEiCON) (pp. 1-5). IEEE (2018).

[5] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays, Comput. Methods Programs Biomed. 196 (2020), 105608.

[6] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning, Med. Image Anal. 65 (2020), 101794.

[7] M.D. Li, N.T. Arun, M. Gidwani, K. Chang, F. Deng, B.P. Little, D.P. Mendoza, M. Lang, S.I. Lee, A. O'Shea, A. Parakh, Automated assessment and tracking of COVID-19 pulmonary disease severity on chest radiographs using convolutional siamese neural networks. Radiology: Artificial Intelligence, 2(4)(2020) p.e200079.

[8] J. Zhang, Y. Xie, Pang, Liao, Z. Verjans, J. Li, W. Sun, Z. He, J. Li, Y. Shen, C. Xia, Y.. Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection. arXiv preprint arXiv:2003.12338 (2020).

[9] I.A. van den Berk, M.M. Kanglie, T.S. van Engelen, S. Bipat, M.G. Dijkgraaf, P. M. Bossuyt, W. de Monyé, J.M. Prins, J. Stoker, OPTimalIMAging strategy in patients suspected of non-traumatic pulmonary disease at the emergency department: chest X-ray or ultra-low-dose CT (OPTIMACT)—a randomised controlled trial chest X-ray or ultra-low-dose CT at the ED: design and rationale, Diagnostic and Prognostic Research 2 (1) (2018) 1–10.

[10] A.D. Gunasinghe, A.C. Aponso, H. Thirimanna, Early prediction of lung diseases. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-4). IEEE. (2019).

[11] J. Li, G. Zhao, Y. Tao, P. Zhai, H., Chen, H. He, T., Cai,. Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19. Pattern Recognition, 114 (2021) p.107848.

[12] S. Dong, Q. Yang, Y. Fu, M. Tian, C. Zhuo, Rconet: Deformable mutual information maximization and high-order uncertainty-aware learning for robust covid-19 detection, IEEE Trans. Neural Networks Learn. Syst. 32 (8) (2021) 3401–3411.

[13] M.K. Mahbub, M. Biswas, L. Gaur, F. Alenezi, K.C. Santosh, Deep features to detect pulmonary abnormalities in chest X-rays due to infectious DiseaseX: covid-19, pneumonia, and tuberculosis, Inf. Sci. (2022).

[14] S. Dey, R. Roychoudhury, R. Malakarand Sarkar, An optimized fuzzy ensemble of convolutional neural networks for detecting tuberculosis from Chest X-ray images, Appl. Soft Comput., 114 (2022) 108094.

[15] Y. Xu, L.F. Souza, I.C. Silva, A.G. Marques, F.H. Silva, V.X. Nunes, T. Han, C. Jia, V. H.C. de Albuquerque, P.P. Rebouças Filho, A soft computing automatic based in deep learning with use of fine-tuning for pulmonary segmentation in computed tomography images, Appl. Soft Comput., 112 (2021) 107810.

[16] R. Chetia, P.P. Sahu, Quantum edge extraction of chest CT image for the detection and differentiation of infected region of COVID-19 patient, Arabian J. Sci. Eng. (2022) 1–12.

[17] R. Kundu, P.K. Singh, M. Ferrara, A. Ahmadian, R. Sarkar, ET-NET: an ensemble of transfer learning models for prediction of COVID-19 infection through chest CT-scan images, Multimedia Tools and Appl. (2021) 1–20.

[18] K. Zhang, S. Qi, J. Cai, D. Zhao, T. Yu, Y. Yue, Y. Yao, W. Qian, Content-based image retrieval with a convolutional siamese neural network: distinguishing lung cancer and tuberculosis in CT images, Comput. Biol. Med., 140 (2022) 105096.

[19] D. Arias-Garzón, J.A. Alzate-Grisales, S. Orozco-Arias, H.B. Arteaga-Arteaga, M.A. Bravo-Ortiz, A. Mora-Rubio, J.M. Saborit-Torres, J.Á.M. Serrano, M. de la IglesiaVayá, O. Cardona-Morales, R. Tabares-Soto, COVID-19 detection in X-ray images using convolutional neural networks, Mach. Learning Appl., 6 (2021) 100138.

[20] Y. Imrana, Y. Xiang, L. Ali, Z. Abdul-Rauf, A bidirectional LSTM deep learning approach for intrusion detection, Expert Syst. Appl. 185 (2021), 115524.

[21] Y. Xu, D. Li, Q. Xie, Q. Wu, J. Wang, Measurement 178 (2021), 109316.

[22] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Adv. Neural Inform. Process. Syst. (2015) 91–99.

[23] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[24] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z. B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M.T. Islam, Can AI help in screening Viral and COVID-19 pneumonia? IEEE Access 8 (2020) 132665–132676.

[25] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S.B.A. Kashem, M.T. Islam, S.A. Maadeed, S.M. Zughaier, M.S. Khan, M.E. Chowdhury, Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images, 2020.

[26] G. Maguolo, L. Nanni, A critic evaluation of methods for COVID-19 automatic detection from X-ray images, Information Fusion 76 (2021) 1–7.

[27] E. Tartaglione, C.A. Barbano, C. Berzovini, M. Calandri, M. Grangetto, Unveiling covid-19 from chest x-ray with deep learning: a hurdles race with small data, Int. J. Environ. Res. Public Health, 17(18) (2020) 6933.

[28] S. Talatahari, M. Azizi, M. Tolouei, B. Talatahari, P. Sareh, Crystal structure algorithm (CryStAl): a metaheuristic optimization method, IEEE Access 9 (2021) 71244–71261.

[29] T. Rahman, A. Khandakar, M.A. Kadir, K.R. Islam, K.F. Islam, R. Mazhar, T. Hamid, M.T. Islam, S. Kashem, Z.B. Mahbub, M.A. Ayari, Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization, IEEE Access 8 (2020) 191586–191601.

[30] A. Giełczyk, A. Marciniak, M. Tarczewska, Z. Lutowski, Pre-processing methods in chest X-ray image classification, PLoS ONE 17 (4) (2022) e0265949.

[31] G. Altan, Y. Kutlu, A. Gökçen, Chronic obstructive pulmonary disease severity analysis using deep learning onmulti-channel lung sounds, Turkish J. Electr. Eng. Comput. Sci. 28 (5) (2020) 2979–2996.

[32] G. Altan, Y. Kutlu, N. Allahverdi, Deep learning on computerized analysis of chronic obstructive pulmonary disease, IEEE J. Biomed. Health. Inf. 24 (5) (2019) 1344–1350.

[33] G. Altan, Y. Kutlu, A.Ö. Pekmezci, S. Nural, Deep learning with 3D-second order difference plot on respiratory sounds, Biomed. Signal Process. Control 45 (2018) 58–69.

[34] S.S. Narlı, G., Altan, Impact of Local Histogram Equalization on Deep Learning Architectures for Diagnosis of COVID-19 on Chest X-rays.

[35] S.S. Narli, G., Altan, Clahe based Enhancement to Transfer Learning in COVID-19 Detection. Gazi Mühendislik Bilimleri Dergisi, pp.1-11.

[36] S. Leena Nesamani, S. Nirmala Sugirtha Rajini, M.S. Josphine, J. Jacinth Salome,. Deep Learning-Based Mammogram Classification for Breast Cancer Diagnosis Using Multi-level Support Vector Machine. Advances in Automation, Signal Processing, Instrumentation, and Control, (2021) 371-383.

[37] G. Altan, Y. Kutlu, N. Allahverdi, Deep learning on computerized analysis of chronic obstructive pulmonary disease, IEEE J. Biomed. Health Inform., 24(5) (2019) 1344-1350.

[38] F. Marini, B.Walczak, Particle swarm optimization (PSO). A tutorial, Chemometrics and Intelligent Laboratory Systems, 149 (2015) 153-165.

[39] S. Mirjalili, Genetic algorithm, in: Evolutionary Algorithms and Neural Networks, Springer, Cham, 2019, pp. 43–55.