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.

Computers in Biology and Medicine 149 (2022) 105806

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

A dual-stage deep convolutional neural network for automatic diagnosis of COVID-19 and pneumonia from chest CT images☆

Farhan Sadik, Ankan Ghosh Dastider, Mohseu Rashid Subah, Tanvir Mahmud, Shaikh Anowarul Fattah *

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh

ARTICLE INFO

Keywords:
Classification
Clustering
Convolutional neural network
COVID-19
Deep learning
Segmentation

ABSTRACT

In the Coronavirus disease-2019 (COVID-19) pandemic, for fast and accurate diagnosis of a large number of patients, besides traditional methods, automated diagnostic tools are now extremely required. In this paper, a deep convolutional neural network (CNN) based scheme is proposed for automated accurate diagnosis of COVID-19 from lung computed tomography (CT) scan images. First, for the automated segmentation of lung regions in a chest CT scan, a modified CNN architecture, namely SKICU-Net is proposed by incorporating additional skip interconnections in the U-Net model that overcome the loss of information in dimension scaling. Next, an agglomerative hierarchical clustering is deployed to eliminate the CT slices without significant information. Finally, for effective feature extraction and diagnosis of COVID-19 and pneumonia from the segmented lung slices, a modified DenseNet architecture, namely P-DenseCOVNet is designed where parallel convolutional paths are introduced on top of the conventional DenseNet model for getting better performance through overcoming the loss of positional arguments. Outstanding performances have been achieved with an F_1 score of 0.97 in the segmentation task along with an accuracy of 87.5% in diagnosing COVID-19, common pneumonia, and normal cases. Significant experimental results and comparison with other studies show that the proposed scheme provides very satisfactory performances and can serve as an effective diagnostic tool in the current pandemic.

1. Introduction

Early detection and widespread testing are of paramount importance to control the current situation of Coronavirus Disease (COVID-19) [1,2]. Currently, the reverse-transcription polymerase chain reaction (RT-PCR) test is used for the mass screening, which is time-consuming, prone to high false-negative rates and its capacity is greatly limited [3,4]. To mitigate the crisis, radiological imaging-based COVID-19 diagnosis using chest X-ray [5,6] or computed tomography (CT) scan images can be a faster and efficacious complement to the existing testing method [7]. In comparison to the chest X-ray images that are low-resolution by nature, the chest CT scan images provide comparatively better visibility of lesion size with well-performed prediction capability [4,8–10]. It is found that the characteristic signs can be detected with the help of chest CT scans even if the RT-PCR result is negative [4]. Manual screening on CT images is a complicated and tedious job for an individual as lesions in the CT images may reside in a gradual slice by slice manner [11]. Besides, there are lots of similar features between viral pneumonia and novel Coronavirus pneumonia developed by the COVID-19, making it a more difficult job to handle [12]. An artificial intelligence (AI)-based network is therefore earnestly solicited to segment desired regions and classify accordingly [3].

CT screening has long been an essential tool for diagnosing lung diseases including lung cancer and pneumonia [12] and it is now being widely used by the physicians for manually detecting COVID-19. For automatic classification of COVID-19, community-acquired pneumonia (CAP) and non-pneumonia from CT slices, in conventional approaches, CT images are fed directly into the classifier networks [13,14]. In [13], 2D-ResNet-50 deep convolutional neural network (CNN) model is used for handling three classes but in [14], 2D ResNet-152 deep CNN model is used to handle two classes, COVID vs. non-COVID. When the CT images are directly used in a classifier, it may hinder the performance of the network as unnecessary information (non-lung portions) gets passed into it as well. In order to remove the unnecessary portions from the chest CT images, semantic segmentation is generally used, which can be manual or automatic. Methods that utilize semantic segmentation

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.
* Corresponding author.
E-mail address: fattah@eee.buet.ac.bd (S.A. Fattah).

https://doi.org/10.1016/j.compbiomed.2022.105806
Received 7 March 2022; Received in revised form 5 June 2022; Accepted 26 June 2022
Available online 19 July 2022

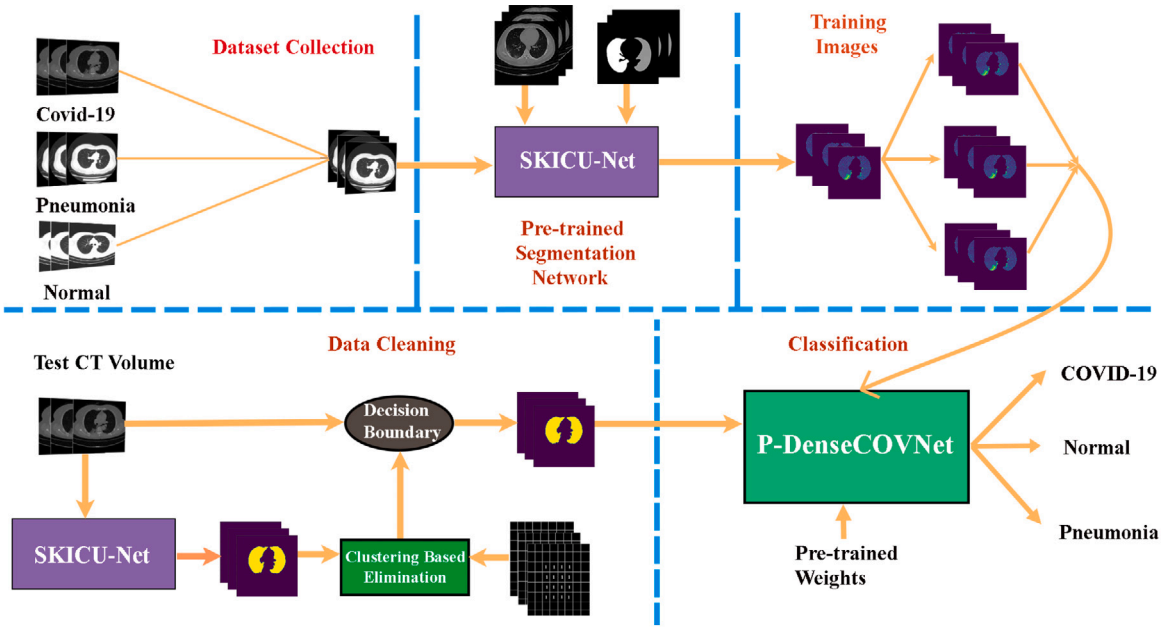

Fig. 1. Pipeline of the proposed method. It includes three fundamental steps: (i) segmenting lung regions from CT images using skip connected version of U-Net (SKICU-Net), (ii) eliminating redundant information by unsupervised hierarchical clustering and (iii) classifying into COVID-19, common pneumonia and normal healthy lungs by the proposed architecture (P-DenseCOVNet).

prior to classification stage, generally at first, employ a segmentation scheme to extract the region of interests from the chest CT images, such as the two lungs and infected regions and then these segmented portions are used in the classifier to detect the COVID-19 cases [15–17]. In [15], the region of interests from the CT slices are identified by two radiologists, which is very time consuming and may not be feasible for a large amount of data. In [16], lung regions are segmented from the 3D chest CT scans by using a pre-trained U-Net deep learning model in an unsupervised manner and then the segmented regions are fed into a 3D ResNet model. Unsupervised segmentation schemes are advantageous when there exists a lack of large, properly annotated train datasets, but they are susceptible to errors [18]. A dual sampling attention network is developed in the scheme proposed in [17] to separate COVID 19 from CAP. In this scheme, after segmenting the lungs and infectious areas using a VB-Net Toolkit [19], a 3D CNN model is used to diagnose the segmented regions. When 3D networks are used for segmentation or classification purposes, it makes the system computationally expensive and more complicated than the 2D networks. Likewise, a multiple instance learning scheme is proposed by [20] where a Pulmonary Toolkit is used to extract the lung field with a bounding box. Moreover, pre-trained Resnet-50 and Xception networks are utilized to capture deep features. Finally, *k*-nearest neighbor search differentiates COVID 19 from CAP based on the features obtained. In both supervised and unsupervised semantic segmentation, U-Net is a widely used network for segmenting biomedical images, with a propitious performance even when the dataset consists insufficient number of images [21]. The segmentation performance of the U-Net is expected to improve further by incorporating various modifications [22–24]. Apart from U-Net in [25], a SegNet-based attention gated (AG) mechanism guided by Dice Loss (DL) and Tversky Loss (TL) is proposed to identify the region of interests. A Linknet architecture is deployed in [26] where the concatenation operation of U-Net is replaced by adding a defying loss of spatial information. With the extracted segmented regions, at the classification stage, various supervised machine learning schemes are commonly utilized [15–17]. During the testing phase with a large number of slices in each CT volume, by proper selection of images with significant importance can reduce the computational burden. A complete scheme for accurately detecting COVID-19 images from normal and pneumonia cases, which offers precise lung segmentation, informative image selection from test CT volume and deep learning-based effective classification utilizing the segmented images, is still in great demand.

In this paper, deep learning-based segmentation followed by classification schemes is proposed to classify COVID-19, normal, and pneumonia patients from chest CT scan images, as shown in Fig. 1. For the purpose of segmenting the lung regions in chest CT images, a skip-connected version of the U-Net model, namely SKICU-Net, and then for the classification, a deep CNN based architecture P-DenseCOVNet are proposed. The proposed SKICU-Net incorporates greater reconstruction of information than the traditional one by introducing extra connections between the encoder and decoder parts. The proposed P-DenseCOVNet upkeeps the positional arguments of the input images in a more vigorous way by utilizing parallel branches along with the DenseNet [27] network and classifies the images into three classes with significant performance. An unsupervised hierarchical clustering scheme is introduced at the testing stage to consider CT images with visible lungs-features for the proposed classification network. Extensive experimentation is carried out on the CT scan image dataset to analyze and compare the performance of the proposed scheme in detecting the COVID-19 patients. All the codes and architectures of this study are publicly available at: https://github.com/fsa125/DualStageCovid.

2. Materials and methods

Three major steps involved in the proposed method are: (a) lung region segmentation from the chest CT scan images utilizing the proposed skip connected U-Net architecture (SKICU-Net), (b) unnecessary information elimination by unsupervised clustering, and (c) prediction of COVID-19 or common pneumonia or normal healthy lungs using the proposed architecture with DenseNet as the backbone network. In this section, datasets used in the study are discussed at first, and the details of the proposed COVID-19 detection scheme are presented in the following subsections.

2.1. Dataset

In this research work, two different publicly available chest CT scan datasets are used: (1) COVID-19 CT Lung and Infection Segmentation dataset from Zenodo [28], and (2) dataset from China

Table 1
Datasets used for the classification task.

Stage	Dataset	Volumes	Images
Train	COVID-19 Lung CT	167	3410
	Pneumonia Lung CT	98	3537
	Normal Lung CT	23	3535
Test	COVID-19 Lung CT	40	3442
	Pneumonia Lung CT	40	6512
	Normal Lung CT	40	3024

Fig. 2. Difference of segmenting performance between traditional U-Net and SKICU-Net.

Consortium of Chest CT Image Investigation (CC-CCII) [12]. The first dataset from [28] contains 20 COVID-19 CT volumes which are fully annotated with left lung, right lung, and infections by two radiologists and verified by an experienced radiologist [28]. Among them, 16 volumes consisting of 3520 2D CT images, are used for the training of the proposed segmentation model (SKICU-Net), and the remaining 4 volumes are used at the test phase. The segmentation model trained on this dataset is later used for segmenting lung regions from the CT images of the second dataset, the CC-CCII dataset one, which is used exclusively for the classification study. To ensure the robustness of our proposed P-DenseCOVNet, 30 volumes of COVID-19 images (30 patients) are tested from "The Cancer Imaging Archive" TCIA[1] dataset.

Table 1 summarizes the different types of datasets used for the classification study. In this study, the terms "volume", "slice" and "image" are used frequently and interchangeably.

2.2. Preprocessing

All the 512 × 512 × 3 CT images from these datasets are resized to 128 × 128 × 3 before feeding them to the skip connected U-Net architecture (SKICU-Net). Resize operation is executed to get rid of time-consuming computations. The resize operation generally affects the images by introducing ripples and degrades the performance by some margin depending on the application. However, different interpolation techniques, such as nearest-neighbor approximation, bicubic interpolation, and bilinear interpolation are introduced in order to preserve the maximum information of an image. In this work, the "area interpolation" technique is used where image shrinking works as bilinear interpolation, and the x-axis and y-axis of the image are shrunk at the same time keeping a proportionate ratio with the original image. Relation of the nearest pixels is also ensured with this technique and therefore, it renders a reliable compression of a large image providing computational efficiency.

2.3. Lung region segmentation

Lung region segmentation is of paramount importance for detection purposes, especially in a two-stage pipeline like the one used in this study [7]. In this work, the goal is to segment lung regions (left lung and right lung) from a fully annotated dataset with masks of the left lung, right lung, and consolidations. Traditional U-Net successfully segments the lung areas but also produces noise as observed in Fig. 2. The main problems of U-Net and FCNs are the dissolving nature of the spatial features due to wide pooling operations. Feature concatenation somewhat tackles the problem but few low-level semantics are hardly detected [29]. To achieve better performance than that is obtained by the traditional one, a skip connected version of U-Net, named as SKICU-Net, is proposed and trained for segmenting lung regions from the CT scan images. The SKICU-Net consists of two parts contracting (encoder) and expansive (decoder). The contracting part is made up of several convolution operations with the same padding which increases the feature maps. From an RGB image of size 128 × 128 × 3, it gets the size of 128 × 128 × 16, which means 16 different feature maps are present at this stage. Max pooling with stride 2 reduces the height and width of the images by half and the size of the images becomes 64 × 64 with feature maps mentioned above. Again convolution operation is applied to the image to create more multi-channel feature maps. These steps are continued serially until reaching the bottom where 256 multi-channel features of 8 × 8 sized images are built. Now upsampling or transposed convolution is introduced to reconstruct the original image which is known as the expansive or decoder path. Upsampling with stride 2 converts 8 × 8 × 256 images to 16 × 16 × 128. But the reconstructed image should have a resemblance to the original one. So, a feature map of size 16 × 16 × 128 from the contracting path is concatenated with the expansive path which preserves the borderline features. Further convolution again weakens the borderline features. In order to solve this, several skip connections are introduced instead of one from contracting path to extensive path to gain maximum reconstruction of information, as shown in Fig. 3. This becomes more vibrant when high-resolution layers are reconstructed. Upsampling and skip connections between two paths (encoder and decoder) continues till it reaches 128 × 128 sized images with 16 feature maps. Later 1 × 1 convolution generates 2 channel images where one channel consists of two lung masks (right lung and left lung) and another one consists of background. Due to serial skip connections noises are prevented to reach maximum performance.

The architecture of the SKICU-Net is shown in Fig. 3. The SKICU-Net is basically a U-shaped staircase network that consists of two parts: contracting (encoder) and expansive (decoder), as mentioned before. For better understanding, a brief discussion on the network is presented here. Each tread in the encoder path is denoted as $C_{row,column}$ format. C_{11} means the first block in the first tread, C_{12} is the second block in the first tread, and so on. Each tread consists of two 3 × 3 convolutions followed by a Rectified linear unit (ReLU) activation unit. Max-pooling operation reduces the dimension of C_{13} and reaches the next tread (C_{21}). The same operations are performed until C_{53} is reached. This is the input to the expansive path where $D_{row,column}$ format is used to describe each tread. Each tread of the expansive or decoder path is constructed by 2 × 2 transposed convolution from the previous tread and followed by a 3 × 3 convolution. It goes from bottom to top performing similar operations on each tread and extracts the desired regions. Additional skip connections between the encoder path and decoder path are used to prevent the loss of border pixels. The relation between the contracting part and the expansive path can be shown by the following equations:

$$D_{i2} = D_{i3} \oplus C_{i3} \tag{1}$$

$$D_{i1} = D_{i2} \oplus C_{i3} \tag{2}$$

where, $i = 1, 2, 3, 4$. For the bottom expansive path, $i = 4$. So we get:

$$D_{42} = D_{43} \oplus C_{43} \tag{3}$$

$$D_{41} = D_{42} \oplus C_{43} \tag{4}$$

And it goes as $i = 3, 2, 1$, where we get 16 channels of size 128 × 128. Then 1 × 1 convolution is performed to get the mask of size

[1] https://wiki.cancerimagingarchive.net/display/Public/COVID-19

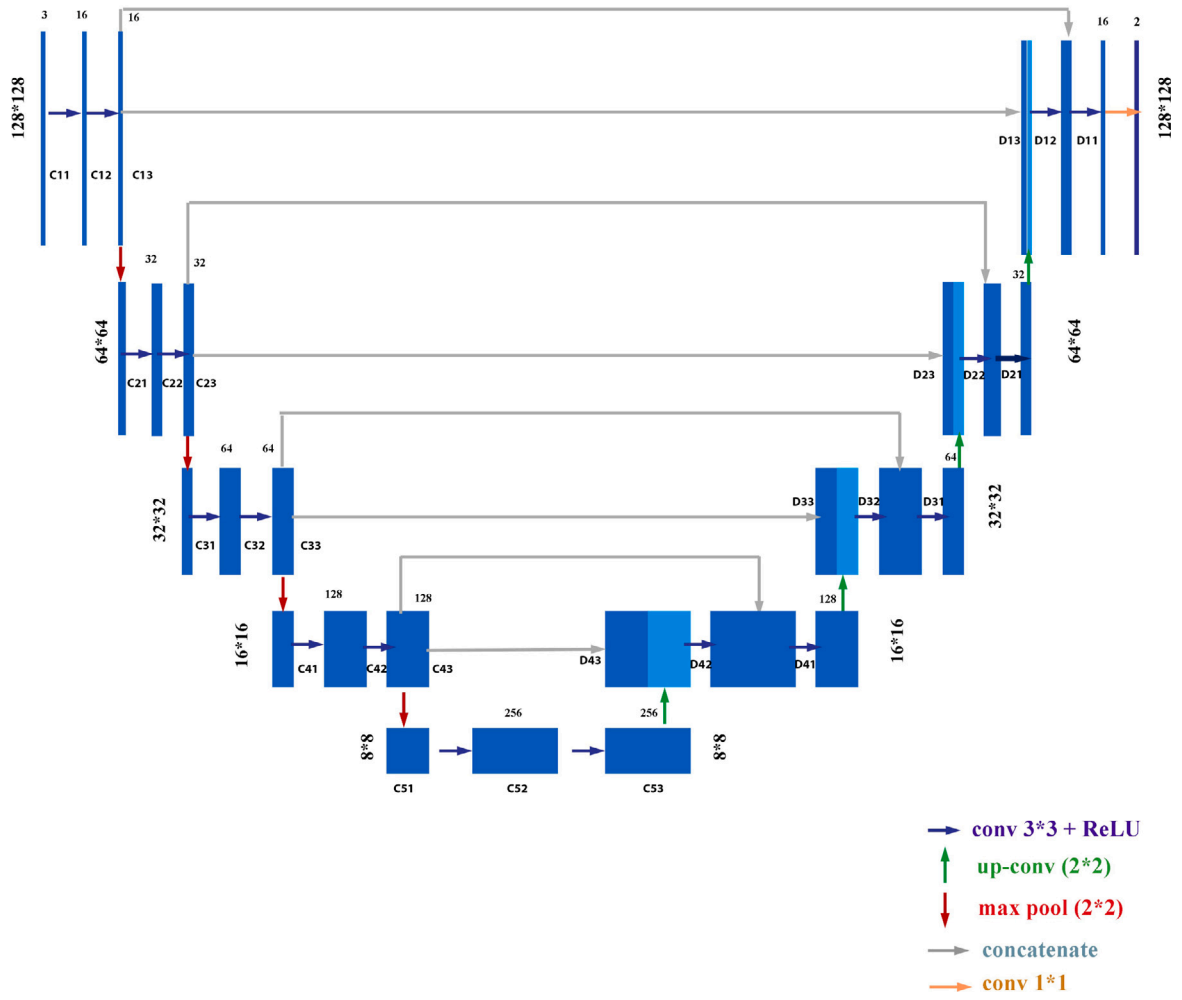

Fig. 3. U-Net architecture with additional skip connections (SKICU-Net) that connects decoded feature maps to encoded ones twice. Blue and light blue boxes correspond to encoded and decoded feature maps, respectively and their width denotes the number of feature maps. The size of each feature map is given on the left (for encoder part) or right (for decoder part) side of the boxes. Multi-colored arrowheads denote different operations, such as convolution, deconvolution, max-pooling, as indicated in the figure.

$128 \times 128 \times 2$, where the first channel contains the background and the other one contains the desired lung regions.

Another important aspect of the proposed model is to define a proper loss function. Considering the right lung and left lung as a single class and background as another class, it becomes a binary class problem. Hence binary cross-entropy loss function, also called sigmoid cross-entropy loss is used as loss function in this type of segmentation problem. It is a sigmoid activation plus a cross-entropy loss. Unlike softmax loss, it is independent for each vector component (class), meaning that the loss computed for every CNN output vector component is not affected by other component values. Equations of cross-entropy and binary cross-entropy are given by:

$$\mathrm{CE} = -\sum_{i}^{N} t_i(\log(\mathrm{f(s)_i})) \tag{5}$$

where N is the total number of classes, t is the respective label, and $\mathrm{f(s)}$ is the softmax function, defined as:

$$\mathrm{f(s)_i} = \frac{\exp(\mathrm{s_i})}{\sum_{j}^{N} \exp(\mathrm{s_j})} \tag{6}$$

For binary cross-entropy, $N = 2$, and Eq. (5) becomes:

$$\mathrm{CE} = -t_1 \log(\mathrm{f(s)_1}) - (1 - t_1)\log(1 - \mathrm{f(s)_1}) \tag{7}$$

The segmented output provides two-channel information: mask and the background. The background channel contains redundant information and therefore eliminated. With the one-channel segmented mask, lung regions from the original CT images are extracted with three-channels. Finally, the trained SKICU-Net model is applied on all the CT images and for each CT image, SKICU-Net generates lung segments for the next step.

2.4. Unsupervised cluster-based elimination

CT screening is 3-dimensional by nature, which provides thorough imaging of lungs in different positions. Among them, lungs are obscure in some images and therefore cannot be considered as useful information, as depicted in Fig. 4. The idea of clustering is to eliminate lung images with negligible information to improve the prediction results of the classification network. One volume of COVID-19 or common pneumonia contains multiple slices, some of them are labeled in the dataset as lesion slices. If the whole volume is trained as COVID-19 or common pneumonia, the non-lesion slices with no clear information of the disease will be wrongly trained as a disease class, producing wrong predictions. Therefore, only the lesion slices are considered and trained for the COVID-19 and common pneumonia cases at the training stage. However, at the test phase, it is needed to predict on the basis of the whole volume, where lesion or non-lesion slices are not marked. That is why the unsupervised clustering-based elimination on the segmented lung slices are performed at the test phase when unseen full CT volumes are provided to the proposed end-to-end architecture.

After the segmentation by the SKICU-Net, segmented lung images are passed to the clustering step. These images are grouped into clusters

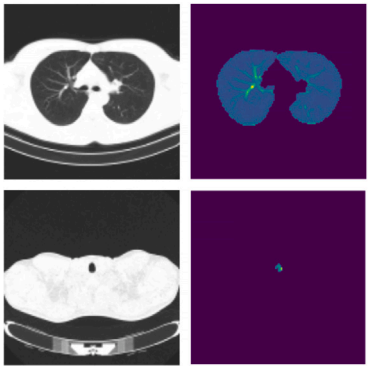

Fig. 4. Normal CT images (left) and segmented lungs by SKICU-Net (right): clearly visible on top; but in the bottom, it is quite obscure and holds negligible information. Unsupervised agglomerative hierarchical clustering method has been used to eliminate them.

based on the visibility of lung portions in the images. Hierarchical clustering is a type of clustering that creates clusters arranging them in a hierarchical tree [30]. Among different kinds of hierarchical clustering methods, agglomerative hierarchical clustering is used in this study, the central operation of which is based on the proximity between the clusters of features. To form the clusters, different types of masks are used to extract the features based on the visibility of lung portions in the segmented images. Depending on the positions of these masks, two types of features are considered: centric features and frontier features. Centric features are extracted from the center region of images, whereas frontier features are extracted from both center and boundary regions. These features are extracted by using gradually extending rectangular masks. Increasing centric and frontier feature values is achieved with the increasing visibility of the lungs. The features are fed into an agglomerative hierarchical clustering algorithm which provides six clusters, found to be the optimum number for this task. Increasing the total number of clusters and eliminating accordingly keeps some of the segmented images with negligible information; while decreasing the total number of clusters erases some segmented images with significant information. In both cases, there is a reduction in the overall accuracy. Therefore in this study, the clustering method is applied to all the segmented test images with lung regions to produce six clusters and eliminate the least significant one. A visualization of the clusters with all the test images is shown in Fig. 5. Some sample CT images corresponding to different clusters are also shown in this figure. The X-axis corresponds to the "Centric features" and the Y-axis corresponds to the "Frontier features". The Z-axis corresponds to a cluster label indicated by a color. It is clearly observed from the figure that the images in "Cluster-1" (the bottom region with red color samples) possess very low values of both features and they have very tiny lung regions. Gradually, the upper regions contain clearer portions of the lungs. Since the red-colored region contains features from the images having negligibly sized lungs, they are eliminated. The objective was to discard the images that carry tiny/dummy lung areas at the time of a single patient testing so that these pieces of information do not take part in the decision-making process and make the scheme reliable. After this cluster-based elimination process, the remaining segmented images are passed to the next prediction step.

There are different kinds of proximity clusters [30], including max, min, group average, and prototype-based clusters or Ward's method, based on which agglomerative hierarchical clustering works. If a and b are two points of two clusters P and Q, min or single link proximity is defined as the shortest distance between them, and max or complete link proximity is defined as the longest distance between them. Group average, as the name suggests, measures the average distance between the points. Ward's method considers proximity between the clusters by merging them in terms of the increase in the SSE (sum of squared error) [30], assuming a cluster's representation with its centroid. Min, max, group average and Ward's method proximity functions are defined as

$$d_{\min(P,Q)} = \min_{a\in P,\ b\in Q}\ d(a,b) \tag{8}$$

$$d_{\max(P,Q)} = \max_{a\in P,\ b\in Q}\ d(a,b) \tag{9}$$

$$d_{\text{avg}(P,Q)} = \sum \frac{d(a,b)}{n_P n_Q} \tag{10}$$

$$d_{\text{ward}(P,Q)} = \sum \frac{(d(a,b))^2}{n_P n_Q} \tag{11}$$

where n_P and n_Q are the points in the two clusters considered to achieve the average result.

2.5. Proposed architecture for the prediction task

In this paper, a new convolutional neural network architecture, named as P-DenseCOVNet, is introduced to classify COVID-19, pneumonia, and normal cases based on lesions and their positions in the images. The DenseNet-121 [27] is used as a backbone network in this proposed classification architecture. Conventional serial convolutional neural network architectures are most likely to lose the positional arguments as serial convolution and max-pooling operations reduce the dimension of feature matrices. So the first layer is split into three parallel convolutional layers with different sizes of kernels: 128, 64, and 32, respectively. Later skip connections are added between them at several points so that positional argument preserves. Several dropout layers are used after convolutional layers to prevent overfitting [31]. The final output is a $32 \times 32 \times 3$ matrix, which is then flattened connecting to a dense layer of size 256. Similarly, the input image is passed into DenseNet-121 with pre-trained weights of ImageNet [32] which gives another flattened layer and connects with another dense layer of the same size of the previous one as shown in Fig. 6. Finally, both the layers are concatenated and passed through three fully connected layers of size 64, 64, and 3, respectively. Dropout layers are introduced among them. Feeding input images ($128 \times 128 \times 3$) directly into DenseNet-121 fails to preserve the positional arguments, whereas the proposed model is capable of keeping those pieces of information and performs well on validation as well as on the new unseen test data. The proposed architecture of the P-DenseCOVNet is shown in Fig. 6. For the three-class prediction, softmax function is used for normalizing the probability vector, the output of which is defined as:

$$P = [p^{\text{COVID}}, p^{\text{pneumonia}}, p^{\text{normal}}] \tag{12}$$

Here p^{COVID}, $p^{\text{pneumonia}}$ and p^{normal} are probabilities of COVID, common pneumonia, and normal cases, respectively.

2.6. Proposed methodology for patient level prediction

The proposed architecture classifies chest CT slices into one of the three classes: COVID-19, common pneumonia, and normal cases. Based on this slice level prediction, next the scheme is developed for patient-level CT volume detection. For this purpose, the full volume of each patient in the test dataset is tested individually and the results are recorded. For each patient, a decision tree approach is taken to determine what class they belong to. The problem is encountered in two steps, considering each step as a binary classification problem. For the first decision, a thresholding approach is adopted to determine whether a particular patient is healthy or not. If the percentage of slices predicted to be normal is more than a certain threshold, the patient is declared to be healthy. If the percentage is less than that value, i.e. the patient is determined to be not healthy, next a decision is taken to determine whether the patient has COVID-19 or common pneumonia.

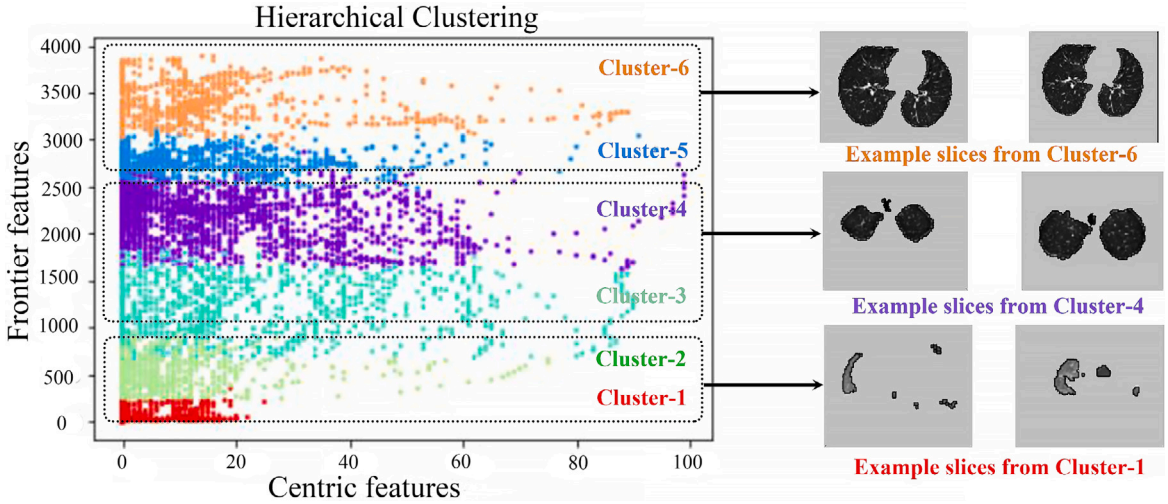

Fig. 5. Clusters based on the centric and frontier features. The bottom portion (in red) portrays the negligible lung segments which were eliminated. Gradually topper portions portray more and more important features. Colorization is performed for visualization only.

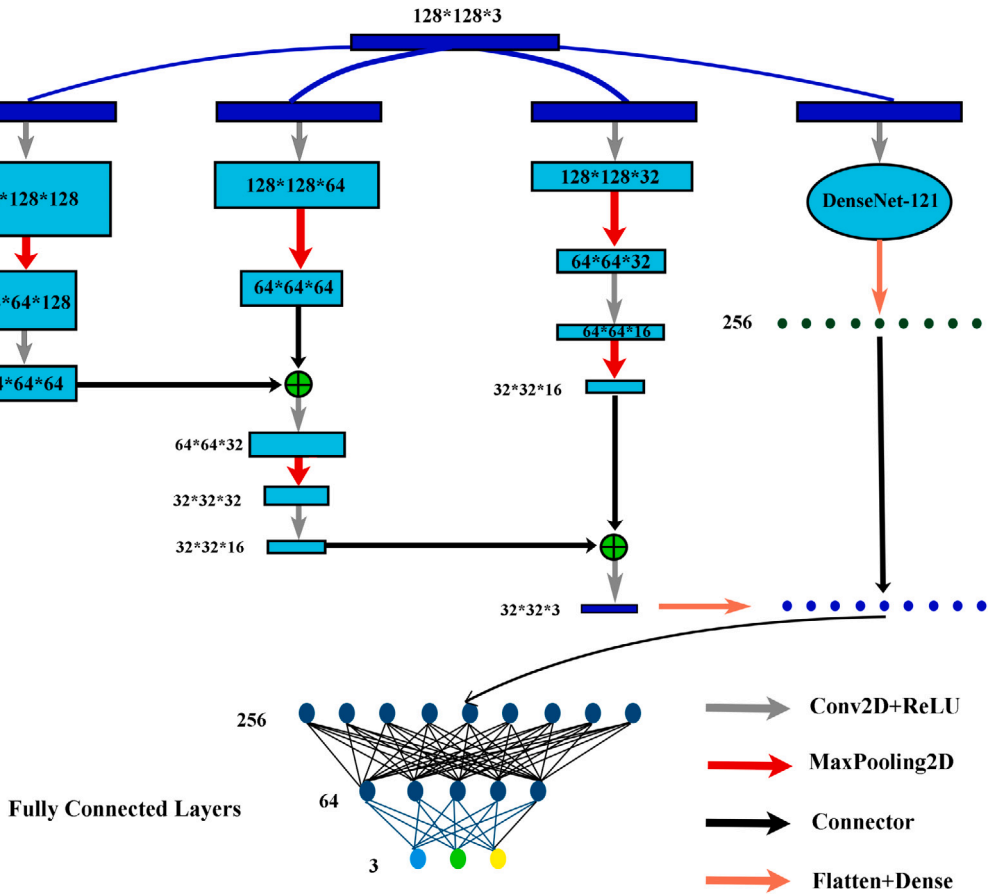

Fig. 6. Proposed architecture for the classification task.

For this second decision, the class that is predicted by more slices is chosen. This procedure is repeated for different threshold values, and the results are examined for all the patients accordingly.

For each patient,

$$\text{Total} = \text{Predicted}_{\text{Normal}} + \text{Predicted}_{\text{COVID}} + \text{Predicted}_{\text{Pneumonia}} \quad (13)$$

A patient is predicted to be healthy if:

$$\frac{\text{Predicted}_{\text{Normal}}}{\text{Total}} > \text{Threshold} \quad (14)$$

If the patient is not healthy, the patient is predicted to have COVID if:

$$\text{Predicted}_{\text{COVID}} > \text{Predicted}_{\text{Pneumonia}} \quad (15)$$

The patient is predicted to have common pneumonia if:

$$\text{Predicted}_{\text{Pneumonia}} > \text{Predicted}_{\text{COVID}} \tag{16}$$

Here,
Total = Total number of slices in a volume
$\text{Predicted}_{\text{Normal}}$ = The number of slices predicted to be normal
$\text{Predicted}_{\text{COVID}}$ = The number of slices predicted to be COVID
$\text{Predicted}_{\text{Pneumonia}}$ = The number of slices predicted to be Pneumonia

3. Experimental results

In this section, the performance of the proposed method is demonstrated considering chest CT scan datasets. Performance of the proposed segmentation architecture, SKICU-Net, in segmenting the lungs portions from the chest CT scans is first presented considering various performance measures. Next, the COVID-19 detection performance of the proposed scheme is discussed considering two major scenarios: COVID-19 slice detection and COVID-19 patient (CT volume) detection. Finally, the performance of the proposed scheme is compared with that of some existing methods.

3.1. Lung segmentation results

Lung segmentation performance is evaluated by comparing the SKICU-Net with some existing architectures, such as conventional U-Net and a fully convolutional network (FCN). F_1 score or dice score, Jaccard index or Intersection over Union (IoU), sensitivity, specificity and balanced average Hausdorff distance are used as evaluation parameters.

F_1 score or dice score is the harmonic mean of precision and recall and it is defined as:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{recall}}{\text{precision+recall}} \tag{17}$$

Jaccard index or Intersection over Union (IoU) is another evaluation metric used to measure similarities between sample sets and is defined as:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{18}$$

Sensitivity or true positive rate and specificity or true negative rate are two measurements used to evaluate the effectiveness of a scheme. They are defined as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{19}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{20}$$

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

The Hausdorff distance measures the similarity to which each point of two sets lies close to each other. These sets can be of images and it can calculate the distance between two points of images in an image set. Thus, this distance is used to determine the degree of resemblance between two images if they are superimposed on one another. Although the average Hausdorff distance has been applied to assess the performance of various medical imaging areas [33–35], it is less suitable for the evaluation of segmentation performance due to its high ranking error. The balanced average Hausdorff distance has a higher median correlation than those by average Hausdorff distance [36] and has been applied in this study. It generates a better ranking of segmentation performance and brings out better comparability between two the predicted image and the ground truth.

The average hausdorff distance between two sets $X = (x_1, x_2, x_3, \ldots, x_n)$ and $Y = (y_1, y_2, y_3, \ldots \ldots y_n)$ is defined as [37]:

$$\mathrm{d}_{AHD}(X,Y) = \left[\frac{1}{2X}\sum_{i=1}^{n}\min_{y\in Y}\ \mathrm{d}(\mathrm{x_i}, \mathrm{y_i}) + \frac{1}{2Y}\sum_{i=1}^{n}\min_{x\in X}\ \mathrm{d}(\mathrm{x_i}, \mathrm{y_i})\right] \tag{21}$$

Table 2
Performance comparison of lung segmentation on the COVID-19 dataset [28] on a patient basis.

Methods	F1 score	IoU	Sensitivity	Specificity	Balanced average Hausdorff distance
FCN	0.739	0.586	0.608	0.996	3.8135
U-Net	0.940	0.886	0.902	**0.998**	2.8312
SKICU-Net	**0.970**	**0.941**	**0.970**	0.997	**2.6455**

Table 3
Performance comparison of lung segmentation on the COVID-19 dataset [28] on a slice basis.

Methods	F1 score	IoU	Sensitivity	Specificity
FCN	0.778	0.675	0.728	0.728
U-Net	0.916	0.844	0.914	0.989
SKICU-Net	**0.966**	**0.935**	**0.942**	**0.991**

In the medical image segmentation, the sets X and Y refer to the voxels or volumes (from a single patient) of the ground truth and the segmentation, respectively, and their elements refer to the pixels of the images. If GtoS is the directed average Hausdorff distance between the ground truth image to predicted image and StoG is the inverse of GtoS, then the balanced average Hausdorff distance for evaluating segmentation performance can be shown as [37]:

$$\text{Balanced average Hausdorff distance} = \left(\frac{\text{GtoS}}{G} + \frac{\text{StoG}}{G}\right)/2 \tag{22}$$

where G is the number of slices/images in a volume. The output gives a rank that determines the segmentation performance. Lower rank value signifies fewer deviations from the ground truth, indicating better performance.

Quantitative results obtained by using three different architectures are presented in Tables 2 and 3. In Table 2, the results obtained on a patient basis are reported. In Table 3, the results obtained on a slice basis are reported. The results shown in these tables are obtained by using the test CT scan images available in dataset [28].

As expected the SKICU-Net offers relatively better performance in terms of various performance measures. On a patient basis, The SKICU-Net surpasses the traditional U-Net by 3.19% in F_1 score and 6.21% in IoU, the two major evaluating parameters. In terms of specificity, the performance of both models is nearly the same. The proposed model offers 7.54% higher sensitivity and 6.56% lower balanced average Hausdorff distance when compared with the traditional U-Net. Compared with the FCN model, F_1 score, IoU, sensitivity, and specificity increase by 31.26%, 60.58%, 59.54%, and 0.1%, respectively. The balanced average Hausdorff distance decrease by 30.63%, in this dataset. It is to noted that the proposed SKICU-Net has the lowest rank among the models, indicating finer segmentation of the lung region. On a slice basis, the SKICU-Net outperforms the traditional U-Net by 5.46% in F_1 score, 10.78% in IoU, 3.06% in sensitivity, and 0.2% in specificity. Compared with the FCN model, F_1 score, IoU, sensitivity, and specificity of the SKICU-Net increase by 24.16%, 38.52%, 29.4%, and 36.13%, respectively.

Apart from the quantitative results, next for qualitative performance evaluation, extracted segments are inspected and for better understanding, few sample cases are shown in Fig. 7. From this figure, it is observed that the performance of the SKICU-Net is better than that of the traditional U-Net and far better than that of the FCN model. Even in cases where there exist sharp changes in the edge or intensity, the SKICU-Net can accurately identify those changes. Hence, it can be concluded that the overall segmentation performance obtained by the SKICU-Net is found satisfactory and the segmented chest CT scans can be effectively utilized without losing necessary information for the purpose of classification.

For the purpose of COVID-19 CT image detection, along with the COVID-19 lung CT dataset available in [28], the CT images of common

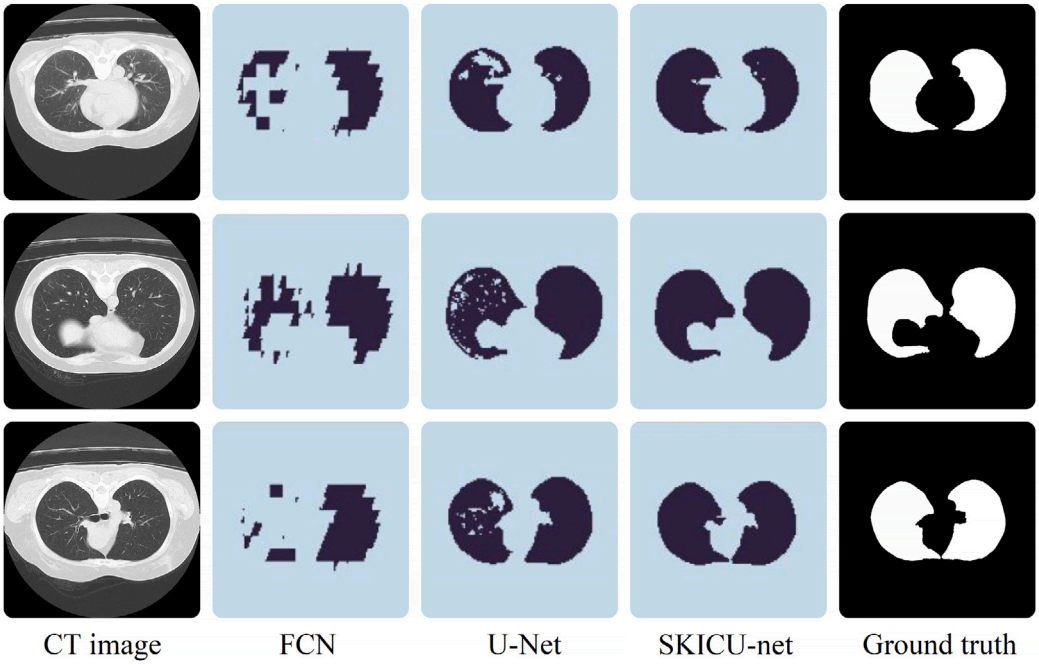

Fig. 7. Visual comparison of Lung segmentation results on COVID-19 dataset. Colorization is performed for visualization only.

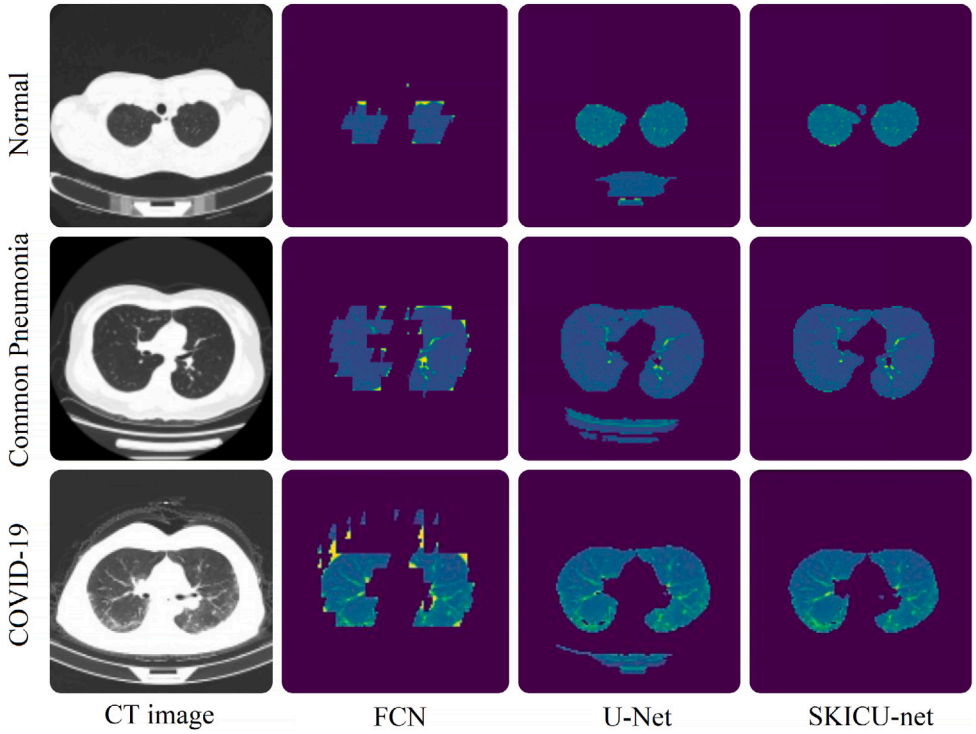

Fig. 8. Visual comparison of Lung segmentation results on COVID-19, common pneumonia, and normal lungs dataset. Colorization is performed for visualization only.

pneumonia patients and healthy cases are collected from the dataset available in [12]. Prior to describing the classification performance, lung segmentation performance obtained in this case is presented in Fig. 8. From visual inspection, one can understand the superiority of the SKICU-Net over the traditional U-Net and FCN models. In Fig. 8, three different chest CT slices are taken from 3 different classes. Here it is observed that the traditional U-Net causes some unnecessary portions included in the segmented images. The FCN model misses major portions in many cases and thus performs poorly compared to both the traditional U-Net and the SKICU-Net. One major concern here is the difference in training and testing datasets for segmentation. It is evident that the SKICU-Net can efficiently handle this factor and perfectly extracts the lung portion only, omitting all unnecessary regions.

3.2. Classification results

3.2.1. Training–testing and optimization

The proposed network is trained with 150 epochs using Adam optimizer [38] and a learning rate of 0.002 with a reduction in learning

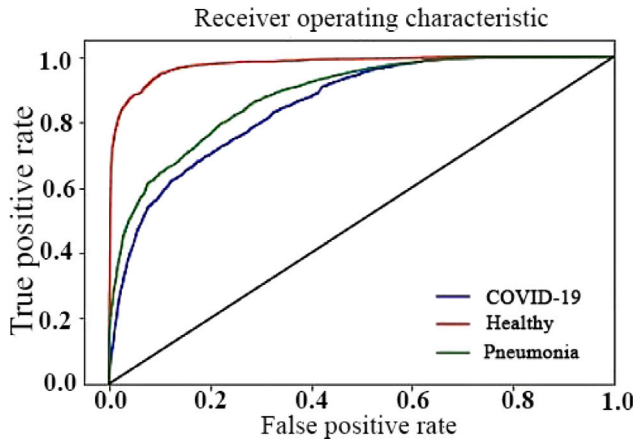

Fig. 9. Classification results evaluating ROC curve. Here, class 1, class 2, and class 3 are COVID-19, healthy and common pneumonia classes, respectively.

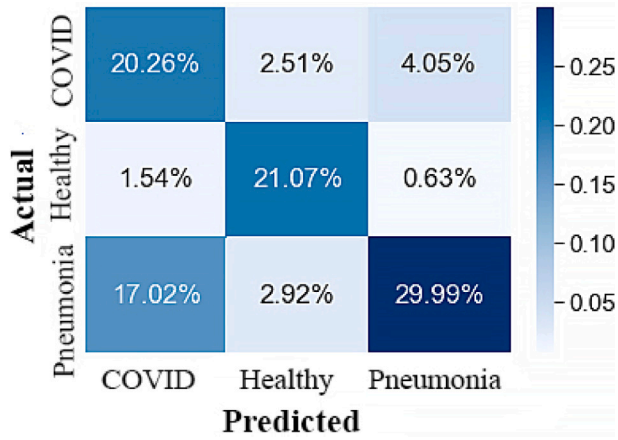

Fig. 10. Confusion matrix for three way classification: COVID-19, common pneumonia and normal cases.

rate enabled when the metric stops learning. Early stopping is used i.e., training is stopped when no improvement in the validation accuracy is seen for a large number of epochs. The batch size used for this case is 64. Data augmentation is performed including rotation (0° ±360°), horizontal and vertical shift (0% ± 20%), scaling (0% ± 20%), horizontal and vertical flips at the training stage. Extensively popular categorical cross-entropy loss function is applied to calculate loss between ground truth labels and predicted results as shown in Eq. (5). Python and Tensorflow[2] frameworks with CUDA and CUDNN support are utilized and the training procedure is done on a NVIDIA Tesla P100 GPU with 16 GB GPU memory.

In [39], the authors discussed the negative effects of augmentation on COVID-19 detection from chest X-ray images. By horizontal and vertical shift or scaling operations, the DL models were found to be struggling while getting the dominant features such as consolidations, GGO from the images. However, in our study, the proposed method consists of three steps, and one of them is segmenting the lung area from the CT scan images. The classification stage is performed following this segmentation stage, therefore images that are used in the classification stage, i.e. included in the augmentation, have no pixel values outside the lung region. In this case, performing augmentation such as scaling, shifting operations make the model more robust; because after the segmentation operation, there are fewer pixels in a CT scan image than the original one, and DL networks are prone to overfitting when it comes to images with fewer pixel values. In the proposed end-to-end architecture, augmentation is therefore adding relevant data.

3.2.2. Slice level results

The performance of P-DenseCOVNet is evaluated on a test set consisting of 12978 CT scan images of 120 patients available in [12]. The trained model is used to predict the probability of three classes: COVID-19, healthy cases and common pneumonia. Predicted results are compared with ground truth labels provided by the dataset, and Receiver operating characteristic (ROC) curve is plotted. The area under ROC curves (ROC AUC) is also calculated. These curves are shown in Fig. 9. From this figure, it is observed that for the three classes, ROC AUC values of 0.85, 0.98, and 0.88, are achieved, respectively.

The confusion matrix of the proposed model is presented in Fig. 10. The confusion matrix is used to determine some performance evaluation metrics for each of the three classes. In summary, accuracy, sensitivity, specificity, and F_1 score is calculated for COVID-19, common pneumonia, and healthy cases, for different classification methods.

[2] https://www.tensorflow.org/.

At first, the CT slices are directly fed into a DenseNet-121 network without going through the preliminary segmentation process. It is observed from Table 4 that this method produces an acceptable output, but still holds scope for improvement. The overall accuracy of this method is 56.4%.

When SKICU-Net is used to segment the lung regions from CT slices before the classification step, accuracy improves by 12.97%, 7.61%, and 6.21%, respectively, in classifying COVID-19, common pneumonia, and healthy patients. Most other parameters also improve by a decent margin, as can be observed from Table 4. The overall accuracy in this case is 65.5%. Segmentation effectively results in the exclusion of all non-lung regions from the images, which holds redundant information, the improvement in performance can be attributed to this fact.

After that, positional characteristics achieved from parallel convolution and concatenation among multi-feature layers are incorporated alongside the basic DenseNet-121 architecture. In this new P-DenseCOVNet, the classification accuracy in detecting COVID-19, common pneumonia and healthy patients increases by 8.03%, 6.93% and 0.88%, respectively. The overall accuracy of this scheme is 71.1%.

When unsupervised clustering is performed before the classification step, 1030 CT slices with the negligible presence of lung regions are removed in total from all the input segmented images, and the remaining 11948 slices are classified, as shown in Fig. 10. This results in a 7.94% reduction in the total number of chest CT slices that need to be classified. It is to be noted that the slices that get eliminated in this procedure are all non-lesion slices, so the overall performance of the P-DenseCOVNet revamps with a reduction in the testing time also. In this scheme, accuracy in detecting COVID-19 patients remains similar while the accuracy in detecting common pneumonia and healthy patients increases by 0.26% and 1.53%, respectively. Most other parameters also improve in this case, corroborating the proposed method. With the help of unsupervised clustering, it is ensured to eliminate images that hold little or no information, before passing it on to the P-DenseCOVNet. This method results in a better performance of the proposed network with an overall accuracy of 72%.

Gradient-weighted Class Activation Mapping which is also known as "Grad-CAM" utilizes the gradients of some pixels flowing into the final convolutional layer to produce a localized heatmap emphasizing the important regions in the image for predicting the class [40]. It is basically a heatmap with portions of an image demonstrating a high probability of a certain class. Comparing with the visual cortex system, human attention maps have high intensity where humans look at the image in order to answer a visual classification problem and both have a correlation of 0.136. Grad-CAM helps to differentiate the visual performance of neural networks. In this paper, the class activation map

Table 4
Classification results for each of the three classes. The proposed method shows better result in the individual classes in a gradual manner.

Method	COVID-19				Common pneumonia				Healthy cases			
	Acc.	Sen.	Spec.	F1	Acc.	Sen.	Spec.	F1	Acc.	Sen.	Spec.	F1
DenseNet-121	0.617	0.705	0.585	0.492	0.657	0.324	**0.992**	0.486	0.853	0.915	0.834	0.747
SKICU-Net + DenseNet-121	0.697	**0.821**	0.653	0.588	0.707	0.462	0.953	0.612	0.906	0.880	0.914	0.816
SKICU-Net + P-DenseCOVNet	**0.753**	0.747	**0.755**	0.619	0.756	0.606	0.905	0.712	0.914	0.896	0.919	0.829
SKICU-Net + Unsupervised clustering + P-DenseCOVNet	**0.753**	0.752	0.753	**0.619**	**0.758**	**0.608**	0.906	**0.714**	**0.928**	**0.916**	**0.932**	**0.858**

Table 5
Patient level classification result for different thresholds.

Threshold	COVID-19				Common pneumonia				Healthy cases			
	Acc.	Sen.	Spec.	F1	Acc.	Sen.	Spec.	F1	Acc.	Sen.	Spec.	F1
0.90	0.775	0.975	0.675	0.743	0.867	**0.700**	0.950	**0.778**	0.858	0.575	**1**	0.730
0.85	0.817	0.975	0.738	0.780	0.867	0.675	0.963	0.771	0.900	0.725	0.988	0.829
0.80	0.858	0.975	0.800	0.821	0.875	0.650	0.988	0.776	0.950	0.900	0.975	0.923
0.75	0.875	0.975	0.825	0.839	0.875	0.650	0.988	0.776	0.967	0.950	0.975	0.950
0.70	**0.892**	**0.975**	**0.850**	**0.857**	**0.875**	0.650	**0.988**	0.776	**0.983**	**1**	0.975	**0.976**

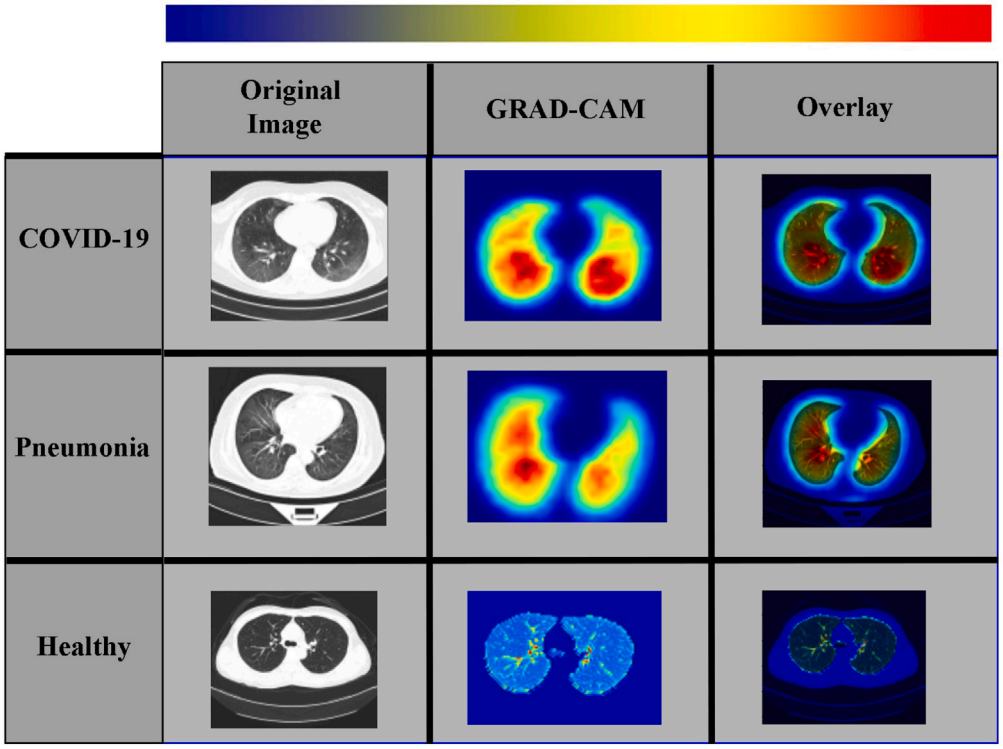

Fig. 11. Attention map generated by the proposed P-DenseCOVNet.

for the three-class images to determine the performance of P-DenseCOVNet is presented in Fig. 11. COVID-19 and pneumonia heatmaps highlight consolidated areas of the lung with warm red colors whereas healthy lungs show a vast amount of blue-colored areas which indicates fewer or zero infectious zones. Infections due to COVID-19 are prone to have more consolidations in the lower lobes of any side of the lung (right or left) or both [41]. Isolated ground-glass opacity (GGO) can also be seen where infections cluster in a region in any portion of the lungs mostly starting from the lower lobes. On the contrary, chest CT images of CAP usually exhibit a more even distribution of consolidations. Similar patterns can also be observed for both COVID 19 and pneumonia for some cases. However, healthy lungs show no sign of consolidations or infections.

From the figure, it is apparent that the COVID-19 images are activated mostly where the consolidation lies. The bottom region of the lung shows consolidation and the attention map also shows these regions along with slight portions from the neighboring area. The same goes for the pneumonia class where consolidated areas are activated (red in the heatmap). In a healthy CT slice, no sign of consolidations is found visually, which is also well-visible from the attention map (no red infectious region). It can be interpreted that the yellow regions indicate the portions that might be infectious and red regions indicate the degree of infections/consolidations. However, the output does not get any feature from the regions outside of the lung due to the segmentation step. So, no dummy features clash with the network's activation, ensuring maximum classification performance. Some COVID cases in the datasets are misclassified as Normal cases while others are misclassified as Pneumonia. The fact is illustrated in Fig. 12, where a COVID slice is predicted as Normal and another COVID slice is predicted as Pneumonia. It is apparent from the figure that the corresponding slice

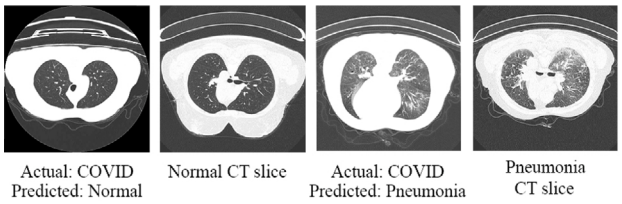

Fig. 12. Misclassified slices from COVID-19 data and standard "Normal" and "Pneumonia" slices for comparison.

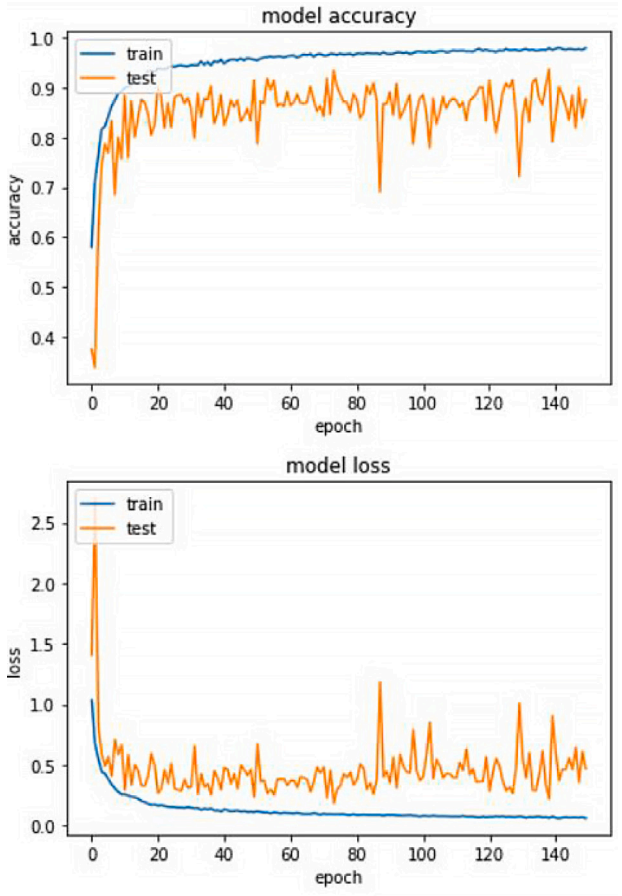

Fig. 13. Convergence analysis of the proposed P-DenseCOVNet.

which is predicted as Normal shows no indication of COVID and the consolidation patterns in some COVID slices resemble Pneumonia. The consolidations in COVID take place in the bottom of the lungs for mild COVID and for severe case, the whole lung is affected and it becomes a complex task for the network to distinguish between the two. However, we propose a thresholding technique to overcome this issue and to correctly classify a patient as COVID, Normal, or Pneumonia.

Convergence analysis is another important aspect to justify the robustness of the model's performance. Steady training performance can be seen in Fig. 13 and P-DenseCOVNet converges quickly with less variability. Stabilization of the validation curve along wth continual decrement of the training curve indicates the stance against overfitting.

3.2.3. Patient level results

Once the performance of the proposed network for COVID-19 slice detection is evaluated, next the scheme is developed to classify patients into one of the three classes based on their full CT volumes. To this end, all the slices of 120 patients are used as the test set. Two consecutive steps are followed to reach a single decision for each patient. The first step involves determining whether a particular patient is healthy or not based on a certain threshold. If the patient is not healthy, a decision is taken on whether it is a COVID-19 case or a common pneumonia case, depending on which class is predicted by more slices. This process is repeated for different thresholds, and the variation in the result is analyzed for all 120 patients. A comprehensive result of the findings is provided in Table 5.

Table 6
Patient level classification result for different thresholds (summary).

Threshold	Overall accuracy	Average sensitivity	Average specificity
0.90	0.750	0.750	0.875
0.85	0.792	0.792	0.896
0.80	0.842	0.842	0.921
0.75	0.858	0.858	0.929
0.70	**0.875**	**0.875**	**0.938**

It is observed from Table 5, that when the threshold is set to 0.70 (i.e., if more than 70% of the slices are predicted to be normal than the volume is of a normal patient), the algorithm reaches a decent performance. At this threshold, there are no false-negative normal cases. As these 12 evaluation metrics change differently when the threshold is varied, the overall accuracy, average sensitivity, and average specificity are calculated in Table 6 to understand the general trend better. From here, the overall result at the threshold value of 0.70 is found to be satisfactory. Total accuracy at this point reaches 87.5%, with an average sensitivity of 87.5% and a high average specificity of 93.8%.

The proposed architecture is evaluated on a two-class COVID-19 detection problem as well. The architecture is trained to classify COVID and Non-COVID slices, where the Non-COVID slices include both common pneumonia and healthy cases. The same test set is used to assess the two-class classification scheme in COVID-19 patient detection. In this case, the volume of each patient is tested separately, and a decision is taken about the overall volume based on the percentage of CT slices that are declared as COVID-19. If the percentage is set to 40% (i.e., if more than 40% of the slices are predicted to have COVID-19 than the patient has COVID-19), the proposed scheme displays satisfactory results. At this value, accuracy, sensitivity, specificity, and F_1 score of the two-class COVID-19 patient detection method is 93.8%, 97.5%, 90%, and 94%, respectively.

The proposed P-DenseCOVNet can cause a comparatively larger amount of network parameters and time consumption at the training stage; however, the additional time is bearable as it is needed only at the training stage and most importantly, it leads to highly efficacious predictions on the test materials within a very short span of time. The proposed end-to-end scheme detects diseases on real time, and the ultimate goal is to integrate the model with a device to implement that. It is to be noted that for a volume with 300 slices, it takes on an average of less than 2 s to predict the result with a reasonable machine configuration.

3.3. Comparison with other methods

In order to evaluate the performance of the COVID-19 detection model, the proposed model is compared with other similar works in the field using unseen test data. The network is compared with seven other state-of-the-art COVID-19 detection models, and promising results are found. It is to be noted that the datasets used in those studies as well as the number of CT scan images used for training and testing purposes are different than the ones used in this study.

The comparison result is presented in Table 7. The overall final accuracy on the patient level is found to be 87.5%. Considering this as a three-way classification, the accuracy is quite promising. For comparison purposes, overall accuracy, average sensitivity, and average specificity of the three classes are chosen as the defining metrics. With

Table 7
Comparison of COVID-19 classification with other related studies. Note that the datasets and the images used in various methods reported in this Table are different.

Literature	Backbone network	Result	Classification type	Quantity
Xu et al. [42]	2D ResNet-18	1. Accuracy: 86.7% 2. F1 Score: I. COVID-19: 83.9% II. Influenza-A: 84.7% III. Healthy: 91.5%	1. COVID-19 2. Influenza-A 3. Healthy	Number of Volumes 1. Train + Validation: 528 2. Test: 90
Shi et al. [43]	Random Forest	1. Accuracy: 87.9% 2. Sensitivity: 90.7% 3. Specificity: 83.3%	1. COVID-19 2. CAP	Number of Volumes 1. Total: 2685 (5-Fold Cross-Validation)
Wang et al. [15]	ModifiedInception Network	1. Accuracy: 82.5% 2. Sensitivity: 75.0% 3. Specificity: 86.0% 4. F1 Score: 72%	1. COVID-19 2. Viral Pneumonia	Number of Images 1. Train: 320 2. Internal Validation: 455 3. External Validation: 290
Jin et al. [44]	ResNet-50	1. Sensitivity: 97.4% 2. Specificity: 92.2%	1. COVID-19 positive 2. COVID-19 negative	Number of Volumes 1. Train: 1136 2. Test: 282
Wu et al. [45]	Res2Net	1. Sensitivity: 95% 2. Specificity: 93%	1. COVID-19 2. Healthy	Number of Volumes (COVID-CS Dataset) 1. Train: 300 2. Test: 400
Wang et al. [16]	DeCovNet	1. Accuracy: 90.1%	1. COVID-19 positive 2. COVID-19 negative	Number of Volumes 1. Train: 499 2. Test: 131
Ouyang et al. [17]	3D AttentionResNet-34	1. Accuracy: 87.5% 2. Sensitivity: 86.9% 3. Specificity: 90.1% 4. F1 Score: 82%	1. COVID-19 2. CAP	Number of Volumes 1. Train + Validation: 2186 2. Test: 2796
Proposed	**P-DenseCOVNet**	**(3-class)** 1. Accuracy: **87.5%** 2. Sensitivity: **87.5%** 3. Specificity: **93.8%** 4. F1 Score: I. COVID-19: **85.7%** II. Pneumonia: **77.6%** III. Healthy: **97.6%**	1. COVID-19 2. Common pneumonia (Viral, bacterial & mycoplasma pneumonia) 3. Healthy	Number of Volumes 1. Train: 288 2. Test: 120
		(2-class) 1. Accuracy: **93.8%** 2. Sensitivity: **97.5%** 3. Specificity: **90.0%** 4. F1 Score: **94%**	1. COVID-19 2. Non-COVID (Common pneumonia & healthy cases)	Number of Volumes 1. Train: 288 2. Test: 120

an average sensitivity score of 87.5% and an average specificity score of 93.8%, the proposed model successfully competes with other similar works in this field. In the two-class COVID-19 patient detection scheme, accuracy, sensitivity, specificity, and F_1 score are also found to be quite satisfactory, with values of 93.8%, 97.5%, 90%, and 94%, respectively. Moreover, the proposed model is also tested with new 30-volume COVID-19 images from the "TCIA" dataset to ensure robustness. For the dataset from TCIA Collections, the accuracy was 73.4% and 85.4% for a slice- and patient-level three-class prediction, respectively. The result is consistent with the aforesaid findings of 75.3% and 87.5% (Tables 4 and 6) for a slice- and patient-level three-class prediction, respectively.

There are some overall concerns presented by the authors in [46], which make a good number of the papers unfit for COVID-19 diagnosis. However, the method presented in this paper is carefully designed by taking into account those concerns. For example, in this paper, completely different datasets are used for segmentation and classification tasks to ensure unbiased results at the segmentation task. For classification, a total of 12,978 images (120 volumes) from a different dataset, other than the previous one, are used to verify our proposed end-to-end scheme. Patient-wise testing has been carried out to avoid biases in the result. A total of 120 patients were tested separately. The final method is selected empirically through the gradual improvement of the architecture. The method of pre-processing (lung segmentation), training parameters, optimizer, learning rate are elaborately mentioned in the paper.

4. Limitations and future work

The study is somewhat limited by the variations of data, as large public datasets from different hospitals are not available. The infections can vary depending on the geographical location and mutation of the virus. Public datasets do not provide additional information like whether the patients are truly COVID-19 positive, or if they have some underlying conditions. Collaboration with hospitals for private data along with public datasets from different domain can further improve the capacity of the proposed model. The SKICU-Net can be integrated with the skip sub-networks proposed in the U-Net++ [22] or the residual attention skip pathways [47] to further enhance the segmentation performance. Moreover, the parallel branches of the P-DenseCOV-Net can be further polished with attention blocks to increase the efficiency of the classification scheme.

5. Conclusion

In this paper, a deep CNN architecture is proposed to detect COVID-19 by classifying the chest CT scan images into COVID-19, common pneumonia, and normal or healthy cases. Instead of using the traditional U-Net, an improved version namely SKICU-Net with skip connections between the encoder and decoder parts is utilized to segment the lung regions from CT images with noteworthy segmentation performance. Unsupervised agglomerative clustering is performed to remove segmented lung regions with insignificant information at the test phase. The segmentation and clustering steps allow the most desired lung

regions to be fed in the succeeding classification architecture, namely P-DenseCOVNet, which employs several parallel branches along with the DenseNet-121 to classify in a proficient manner by preserving the positional arguments. The proposed end-to-end architecture can differentiate COVID-19 from its most competent disease pneumonia, and from healthy cases with promising results in the three-class classification stage. The results get even more efficacious at the two-class classification stage while predicting the COVID-19 cases, segregating them from the Non-COVID class. Performance evaluation and comparison with other works manifest propitious ability to diagnose COVID-19 rapidly, which can play a crucial role in this time. Early detection with this kind of technology allows the clinicians and radiologists to consider the next step for patients quickly, which might be immensely vital to save lives.

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.

Acknowledgments

The authors would like to thank the researchers in Refs. [12,28] for their kind initiative in making the datasets public. The authors would also like to acknowledge the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET) for providing constant support throughout this study.

References

[1] Hongzhou Lu, Charles W. Stratton, Yi-Wei Tang, Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle, J. Med. Virol. 92 (4) (2020) 401–402.

[2] Chen Wang, Peter W Horby, Frederick G Hayden, George F Gao, A novel coronavirus outbreak of global health concern, Lancet 395 (10223) (2020) 470–473.

[3] Xingzhi Xie, Zheng Zhong, Wei Zhao, Chao Zheng, Fei Wang, Jun Liu, Chest CT for typical coronavirus disease 2019 (COVID-19) Pneumonia: Relationship to negative RT-PCR testing, Radiology 296 (2) (2020) E41–E45, PMID: 32049601.

[4] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, Liming Xia, Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, Radiology (2020) 200642.

[5] Lourdes Duran-Lopez, Juan Pedro Dominguez-Morales, Jesús Corral-Jaime, Saturnino Vicente-Diaz, Alejandro Linares-Barranco, COVID-XNet: A custom deep learning system to diagnose and locate COVID-19 in chest X-ray images, Appl. Sci. 10 (16) (2020).

[6] Md Mamunur Rahaman, Chen Li, Yudong Yao, Frank Kulwa, Mohammad Asadur Rahman, Qian Wang, Shouliang Qi, Fanjie Kong, Xuemin Zhu, Xin Zhao, Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches, J. X-Ray Sci. Technol. (Preprint) (2020) 1–19.

[7] Feng Shi, Jun Wang, Jun Shi, Ziyan Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, Dinggang Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19, IEEE Rev. Biomed. Eng. 14 (2021) 4–15.

[8] Heshui Shi, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, Chuansheng Zheng, Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study, Lancet Infect. Dis. (2020).

[9] Li Fan, Dong Li, Huadan Xue, Longjiang Zhang, Zaiyi Liu, Bing Zhang, Lina Zhang, Wenjie Yang, Baojun Xie, Xiaoyi Duan, et al., Progress and prospect on imaging diagnosis of COVID-19, Chin. J. Acad. Radiol. (2020) 1–10.

[10] Michael Chung, Adam Bernheim, Xueyan Mei, Ning Zhang, Mingqian Huang, Xianjun Zeng, Jiufa Cui, Wenjian Xu, Yang Yang, Zahi A. Fayad, Adam Jacobi, Kunwei Li, Shaolin Li, Hong Shan, CT imaging features of 2019 novel Coronavirus (2019-nCoV), Radiology 295 (1) (2020) 202–207, PMID: 32017661.

[11] J. Wang, Y. Bao, Y. Wen, H. Lu, H. Luo, Y. Xiang, X. Li, C. Liu, D. Qian, Prior-attention residual learning for more discriminative COVID-19 screening in CT images, IEEE Trans. Med. Imaging 39 (8) (2020) 2572–2583.

[12] Kang Zhang, Xiaohong Liu, Jun Shen, Zhihuan Li, Ye Sang, Xingwang Wu, Yunfei Zha, Wenhua Liang, Chengdi Wang, Ke Wang, et al., Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography, Cell 181 (6) (2020) 1423 – 1433.e11.

[13] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology 296 (2) (2020) 200905.

[14] Ophir Gozes, Maayan Frid-Adar, Hayit Greenspan, Patrick D. Browning, Huangqi Zhang, Wenbin Ji, Adam Bernheim, Eliot Siegel, Rapid AI development cycle for the Coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis, 2020.

[15] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, Bo Xu, A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19), 2020, http://dx.doi.org/10.1101/2020.02.14.20023028, MedRxiv.

[16] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, C. Zheng, A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT, IEEE Trans. Med. Imaging 39 (8) (2020) 2615–2625.

[17] X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, F. Yan, Z. Ding, Q. Yang, B. Song, F. Shi, H. Yuan, Y. Wei, X. Cao, Y. Gao, D. Wu, Q. Wang, D. Shen, Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia, IEEE Trans. Med. Imaging 39 (8) (2020) 2595–2605.

[18] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, W. Zhang, Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning, IEEE Trans. Med. Imaging 39 (8) (2020) 2584–2594.

[19] Fei Shan, Yaozong Gao, Jun Wang, Weiya Shi, Nannan Shi, Miaofei Han, Zhong Xue, Yuxin Shi, Lung infection quantification of covid-19 in ct images with deep learning, 2020, arXiv preprint arXiv:2003.04655.

[20] Shouliang Qi, Caiwen Xu, Chen Li, Bin Tian, Shuyue Xia, Jigang Ren, Liming Yang, Hanlin Wang, Hui Yu, DR-MIL: deep represented multiple instance learning distinguishes COVID-19 from community-acquired pneumonia in CT images, Comput. Methods Programs Biomed. 211 (2021) 106406.

[21] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[22] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, UNet++: A nested U-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 3–11.

[23] Nabil Ibtehaz, M. Sohel Rahman, MultiResUNet : Rethinking the U-net architecture for multimodal biomedical image segmentation, Neural Netw. 121 (2020) 74–87.

[24] Cheng Chen, Kangneng Zhou, Muxi Zha, Xiangyan Qu, Xiaoyu Guo, Hongyu Chen, Zhiliang Wang, Ruoxiu Xiao, An effective deep neural network for lung lesions segmentation from COVID-19 CT images, IEEE Trans. Ind. Inf. 17 (9) (2021) 6528–6538.

[25] Ümit Budak, Musa Çıbuk, Zafer Cömert, Abdulkadir Şengür, Efficient COVID-19 segmentation from CT slices exploiting semantic segmentation with integrated attention mechanism, J. Digit. Imaging 34 (2) (2021) 263–272.

[26] Qianqian Qi, Shouliang Qi, Yanan Wu, Chen Li, Bin Tian, Shuyue Xia, Jigang Ren, Liming Yang, Hanlin Wang, Hui Yu, Fully automatic pipeline of convolutional neural networks and capsule networks to distinguish COVID-19 from community-acquired pneumonia via CT images, Comput. Biol. Med. 141 (2022) 105182.

[27] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 2261–2269.

[28] Ma Jun, Ge Cheng, Wang Yixin, An Xingle, Gao Jiantao, Yu Ziqi, Zhang Minqing, Liu Xin, Deng Xueyuan, Cao Shucheng, Wei Hao, Mei Sen, Yang Xiaoyu, Nie Ziwei, Li Chen, Tian Lu, Zhu Yuntao, Zhu Qiongjie, Dong Guoqiang, He Jian, COVID-19 CT Lung and Infection Segmentation Dataset, Zenodo, 2020.

[29] Weihao Weng, Xin Zhu, INet: Convolutional networks for biomedical image segmentation, IEEE Access 9 (2021) 16591–16603.

[30] Brian S Everitt, Sabine Landau, Morven Leese, Daniel Stahl, Cluster Analysis 5th ed, John Wiley, 2011.

[31] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[32] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 25, Curran Associates, Inc., 2012, pp. 1097–1105.

[33] Ehab A. AlBadawy, Ashirbani Saha, Maciej A. Mazurowski, Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing, Med. Phys. 45 (3) (2018) 1150–1158.

[34] Michelle Livne, Jana Rieger, Orhun Utku Aydin, Abdel Aziz Taha, Ela Marie Akay, Tabea Kossen, Jan Sobesky, John D Kelleher, Kristian Hildebrand, Dietmar Frey, et al., A U-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease, Front. Neurosci. 13 (2019) 97.

[35] Kimerly A Powell, Tong Liang, Brad Hittle, Don Stredney, Thomas Kerwin, Gregory J Wiet, Atlas-based segmentation of temporal bone anatomy, Int. J. Comput. Assist. Radiol. Surgery 12 (11) (2017) 1937–1944.

[36] Davood Karimi, Septimiu E. Salcudean, Reducing the hausdorff distance in medical image segmentation with convolutional neural networks, IEEE Trans. Med. Imaging 39 (2) (2019) 499–513.

[37] Orhun Utku Aydin, Abdel Aziz Taha, Adam Hilbert, Ahmed A Khalil, Ivana Galinovic, Jochen B Fiebach, Dietmar Frey, Vince Istvan Madai, On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking, Eur. Radiol. Exper. 5 (1) (2021) 1–7.

[38] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, in: Yoshua Bengio, Yann LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[39] Mohamed Elgendi, Muhammad Umer Nasir, Qunfeng Tang, David Smith, John-Paul Grenier, Catherine Batte, Bradley Spieler, William Donald Leslie, Carlo Menon, Richard Ribbon Fletcher, et al., The effectiveness of image augmentation in deep learning networks for detecting COVID-19: A geometric transformation perspective, Front. Med. 8 (2021).

[40] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 618–626.

[41] Yu Guo, Wei Xia, Xuehua Peng, Jianbo Shao, Features discriminating COVID-19 from community-acquired pneumonia in pediatric patients, Front. Pediat. 8 (2020).

[42] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Yanfei Chen, Junwei Su, Guanjing Lang, Yongtao Li, Hong Zhao, Kaijin Xu, Lingxiang Ruan, Wei Wu, Deep learning system to screen coronavirus disease 2019 pneumonia, 2020, arXiv:2002.09334.

[43] Feng Shi, Liming Xia, Fei Shan, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yaozong Gao, He Sui, Dinggang Shen, Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification, 2020, arXiv:2003.09860.

[44] Shuo Jin, Bo Wang, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, Zhuozhao Zheng, et al., AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks, Cold Spring Harbor Laboratory Press, 2020, MedRxiv.

[45] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Chao-Wei Zhao, Ming-Ming Cheng, JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation, 2020.

[46] Michael Roberts, Derek Driggs, Matthew Thorpe, Julian Gilbey, Michael Yeung, Stephan Ursprung, Angelica I Aviles-Rivero, Christian Etmann, Cathal McCague, Lucian Beer, et al., Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, Nat. Mach. Intell. 3 (3) (2021) 199–217.

[47] Qiangguo Jin, Zhaopeng Meng, Changming Sun, Hui Cui, Ran Su, RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, Front. Bioeng. Biotechnol. (2020) 1471.