

# Multi-Task Driven Explainable Diagnosis of COVID-19 using Chest X-ray Images

Aakarsh Malhotra<sup>a,1</sup>, Surbhi Mittal<sup>b,2</sup>, Puspita Majumdar<sup>a,1</sup>,  
Saheb Chhabra<sup>a,1</sup>, Kartik Thakral<sup>b,2</sup>, Mayank Vatsa<sup>b,2</sup>, Richa Singh<sup>b,2,\*</sup>,  
Santanu Chaudhury<sup>b,2</sup>, Ashwin Pudrod<sup>c,3</sup>, Anjali Agrawal<sup>d,4</sup>

<sup>a</sup>*IIIT-Delhi, New Delhi, India, 110020*

<sup>b</sup>*IIT Jodhpur, India, 342037*

<sup>c</sup>*Ashwini Hospital and Ramakant Heart Care Centre, India 431602*

<sup>d</sup>*TeleRadiology Solutions, India 560048*

---

## Abstract

With increasing number of COVID-19 cases globally, all the countries are ramping up the testing numbers. While the RT-PCR kits are available in sufficient quantity in several countries, others are facing challenges with limited availability of testing kits and processing centers in remote areas. This has motivated researchers to find alternate methods of testing which are reliable, easily accessible and faster. Chest X-Ray is one of the modalities that is gaining acceptance as a screening modality. Towards this direction, the paper has two primary contributions. Firstly, we present the COVID-19 Multi-Task Network (COMiT-Net) which is an automated end-to-end network for COVID-19 screening. The proposed network not only predicts whether the CXR has COVID-19 features present or not, it also performs semantic segmentation of the regions of interest to make the model explainable. Secondly, with the help of medical professionals, we manually annotate the lung regions and semantic segmentation of COVID19 symptoms in CXRs taken from the ChestXray-14, CheXpert, and a consolidated COVID-19 dataset. These annotations will be released to the research commu-

---

\*Corresponding author

<sup>1</sup>email: {aakarshm, pushpitam, sahebc}@iiitd.ac.in

<sup>2</sup>email: {mittal.5, thakral.1, mvatsa, richa, santanuc}@iitj.ac.in

<sup>3</sup>email: ash147gmc@gmail.com

<sup>4</sup>email: anjali.agrawal@telradsol.com

nity. Experiments performed with more than 2500 frontal CXR images show that at 90% specificity, the proposed COMiT-Net yields 96.80% sensitivity.

*Keywords:* X-ray, COVID-19, Detection, Diagnostics, Deep Learning, Explainable Artificial Intelligence, Multi-task Learning

---

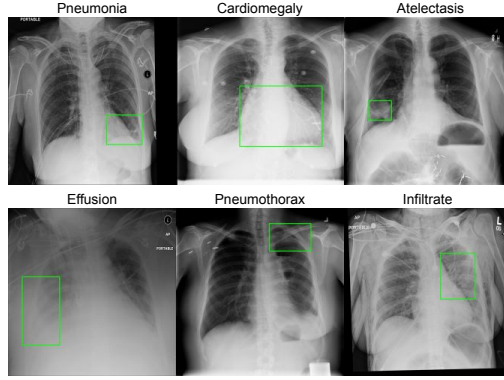
## 1. Introduction

The COVID-19 pandemic has affected the health and well-being of people across the globe and continues its devastating effect on the global population. The total cases have increased at an alarming rate and have crossed 116 million  
5 worldwide [1]. Increasing cases of COVID-19 patients raises the concern for effective screening of infected patients. The current process of testing for COVID-19 is time-consuming and requires availability of testing kits. This necessitates the requirement for alternative methods of screening, which is available to the general population, cost effective, time efficient, and scalable.

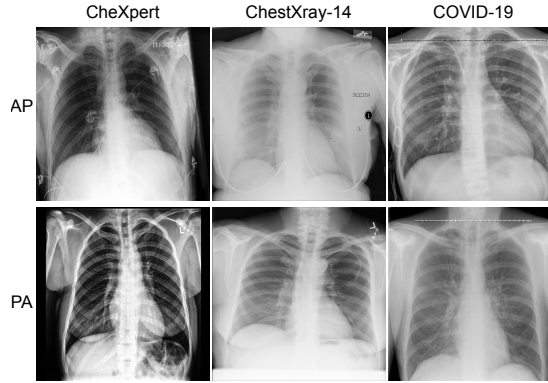
10     Dyspnea is a common symptom for COVID-19. Analyzing the chest X-ray, radiologists have observed that it introduces specific abnormalities in a patient's lungs [2]. For instance, COVID-19 pneumonia has a typical appearance on chest radiographs with bilateral peripheral patchy lung opacities, lower lung distribution, rounded morphology and absence of pleural effusion and lymphadenopathy.  
15     Figure 1 shows samples of chest x-ray images with different lung abnormalities including COVID-19. Motivated by this observation and the fact that x-ray imaging is faster, cheaper, accessible, and has scope for portability, many recent studies have proposed machine learning algorithms to predict COVID-19 using CXRs [3].

### 20     1.1. Literature Review

Researchers have proposed AI-based techniques to detect COVID - 19 using chest CT and x-ray images. Apostolopoulos and Mpesiana [4] explored transfer learning through various CNNs and observed that MobileNet v2 [5] yields the best results. Narin et al. [6] proposed to use three CNN models, namely,



(a) Different kinds of lung abnormalities. The bounding box highlights the diseased region.



(b) AP (anteroposterior) and PA (posteroanterior) views corresponding to CheXpert and ChestXray-14, and COVID-19 datasets.

Figure 1: Samples of chest x-ray images used as a part of this research.

25 ResNet50 [7], InceptionV3 [8], and InceptionResNetV2 [9] for detecting COVID-19 using chest x-ray. The authors fine-tuned these pre-trained deep models for distinguishing COVID-19 from normal x-rays and found that ResNet-50 performed the best. They used 50 chest x-ray images of COVID-19 patients from Github repository [10] and 50 normal chest X-ray images [11]. Nishio et al. [12]  
 30 used a VGG-16 based model for differentiating between COVID-19 pneumonia, non-COVID-19 pneumonia, and healthy CXR images.

Horry et al. [13] proposed a CXR and CT based multi-modal classification for COVID-19 detection. In their work, they propose a data pre-processing

technique and performs transfer learning on various deep learning architec-  
35 tures. Their finding suggests that VGG16 and VGG19 gives promising results  
for COVID-19 vs Pneumonia or normal classification. Further, Oh et al. [14]  
proposed a patch-based CNN approach for COVID-19 diagnosis. During test-  
ing, majority voting from multiple patches at different locations of lungs is  
performed for final decision.

40 For interpretation and explainability, there are limited studies. Mangal et  
al. [15] and Jaiswal et al. [16] utilized DenseNet121 and DenseNet201 [17] for  
classification, respectively. They showed Class Activation Maps (CAM) for  
interpretation. With an emphasis on explainability, the authors showed CAM  
and confusion matrix. On similar lines, Ghoshal and Tucker [18] showed the  
45 application of ResNet50v2 [19] for the above four classes. Authors interpret the  
results using CAM, confusion matrices, Bayesian uncertainty, and Spearman  
correlation. Similarly, Shi et al. [20] and Tsiknakis et al. [21] used CAM for  
better interpretation of COVID detection. Their approach aimed to extract  
relevant features while suppressing inadmissible ones.

50 The problem of small sample size of COVID-19 chest X-ray images was tack-  
led by Loey et al. [22], where they generate new COVID-19 infected images using  
GANs. Wang and Wong [23] introduced COVID-Net for detecting COVID-19  
cases. Further, the authors investigate the predictions made by COVID-Net to  
gain insights on the critical factors associated with COVID-19 cases. In their  
55 work, a three-class classification is performed to distinguish COVID-19 cases  
from regular and Non-COVID cases. In another work, Afshar et al. [24] pro-  
posed a capsule-network based framework referred as COVID-CAPS that uses  
X-ray images for COVID-19 detection. The results from the proposed frame-  
work looks promising as the capsule networks have few parameters to train and  
60 work well on small datasets. Similarly, Shorfuzzaman et al. [25] proposed a  
n-shot meta learning framework. In their work, they use a Siamese neural net-  
work for feature extraction with contrastive loss function in a few-shot learning  
setting for small dataset.

These research demonstrate that AI-driven techniques can diagnose COVID-

19 using chest x-ray images. It could potentially overcome the challenges of limited test kits and speed up the screening process of COVID-19 cases. However, a significant limitation of existing studies is that the algorithms work as a black box. These algorithms predict if the input x-ray is affected by COVID-19 or some related disease. Most studies fail to explain the decisions - for instance, which lung regions are salient for the specific decisions. Secondly, existing studies do not focus on radiological abnormalities such as consolidation, opacities, or pneumothorax. Without a clear emphasis on the lung or the abnormality, it is hard to have the explainability of an algorithm in a crucial application of COVID-19 diagnosis. Further, most of these studies work with a limited number of COVID-19 samples, with around 100 samples under most scenarios. Thirdly, as shown in Figure 1(b), the posteroanterior (PA) and anteroposterior (AP) views of CXR images vary due to the acquisition mechanisms. While training, samples from both classes need to be considered but existing algorithms are generally silent on these details.

## 1.2. Research Contributions

In this research, we propose a deep learning network termed as COVID-19 Multi-Task Network (COMiT-Net), which learns the abnormalities present in the chest x-ray images to differentiate between a COVID-19 affected lung and a Non-COVID affected lung. For medical applications, the explainability of machine learning systems is of paramount importance [26]. In reality, the black-box nature of the deep algorithms refrain us from knowing which regions are getting focused. Furthermore, these algorithms fail to deliver what and where the disease is, which is essential for radiologists and doctors to back their decision. Hence, the proposed network incorporates additional tasks of lung and disease segmentation to provide post-hoc explainability.

The proposed COMiT-Net simultaneously processes the input X-ray for semantic lung segmentation, disease localization, and healthy/unhealthy classification. Incorporating additional tasks while performing the primary task of COVID classification has multiple advantages. While processing for COVID

95 classification, the additional segmentation tasks enforce the network to focus on lung regions and disease-affected areas only. Further, inclusion of healthy/unhealthy classification aids the COMiT-Net to effectively identify a healthy lung. Further, assistance from other tasks reduces dependence on enormous amounts of data required during training. The key research highlights are:

- 100 1. Develop COVID-19 Multi-Task Network (COMiT-Net) for classification and segmentation of the lung and disease<sup>5</sup> regions. The COMiT-Net further predicts if lungs are affected with COVID-19 or Non-COVID-19 disorders and differentiate them from healthy lungs.
2. Inclusion of simultaneous disease segmentation in the COMiT-Net helps  
105 in making the decisions explainable.
3. Extensive evaluation and comparison against the existing deep learning algorithms for COVID-19 prediction, lung, and disease segmentation.
4. Assemble frontal chest x-rays from various sources, that can be used for  
diverse tasks such as classification and semantic segmentation of lungs  
110 and disease. For the assembled dataset, the CXR reports and the CXR were manually verified by a radiologist to affirm the presence/absence of COVID-19 related abnormalities. From different sources, a total of 2513 frontal x-rays of COVID-19 affected patients are collected. Further, these manual annotations for lung and disease semantic segmentation for  
115 healthy, unhealthy, and COVID-19 affected X-ray images will be released to the research community.

## 2. COVID-19 Multi-Task Network (COMiT-Net)

This section provides the details of the proposed COMiT-Net. Multi-task networks are known to learn similar and related tasks together based on the  
120 input data. As shown in Figure 2, multi-task networks have a base network

---

<sup>5</sup>In our context, the terms ‘abnormality’, ‘disease’, and ‘radiological finding’ are used synonymously.

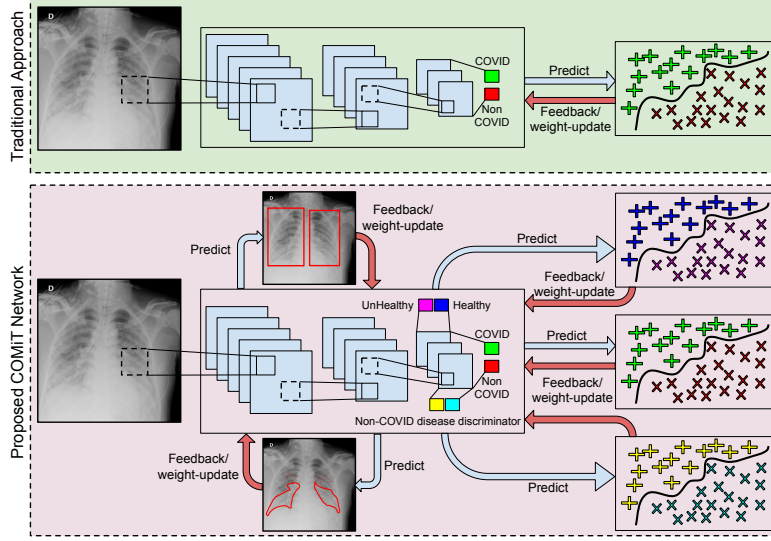


Figure 2: The proposed COMiT-Net to perform multiple related tasks to improve the classification performance for COVID-19 disease diagnostics using frontal x-ray. The figure contrasts the multitask network with single task network.

with multi-objective outputs. Since each task shares the same base network, the weights are learned to be optimal for all functions jointly. The four tasks of COMiT-Net are (i) lung localization, (ii) disease localization, (iii) healthy/unhealthy classification and (iv) multi-label classification for COVID-19 prediction. These tasks are accomplished by using five loss functions: two for segmentation and three for classification. The details of these loss functions are described in the following subsections.

Let  $\mathbf{X}$  be the train set with  $n$  images and  $\mathbf{X}_i$  represent an image.  $\mathbf{X}_i$  is associated with five labels,  $\{\mathbf{L}_i, \mathbf{D}_i, H_i, C_i, O_i\}$  where,  $\mathbf{L}_i$  and  $\mathbf{D}_i$  represent the ground truth binary mask for lung and disease localization, respectively.  $H_i = \{0, 1\}$ ,  $C_i = \{0, 1\}$ , and  $O_i = \{0, 1\}$  represents the healthy/unhealthy, COVID/Non-COVID, and Non-COVID diseases discriminator labels, respectively. Let  $f$  be the proposed COMiT-Net that performs the four different tasks. The task set  $\mathbf{T}$  is defined as  $\mathbf{T} = \{t_1, t_2, t_3, t_4\}$ , where,  $t_1$  and  $t_2$  represent the task of

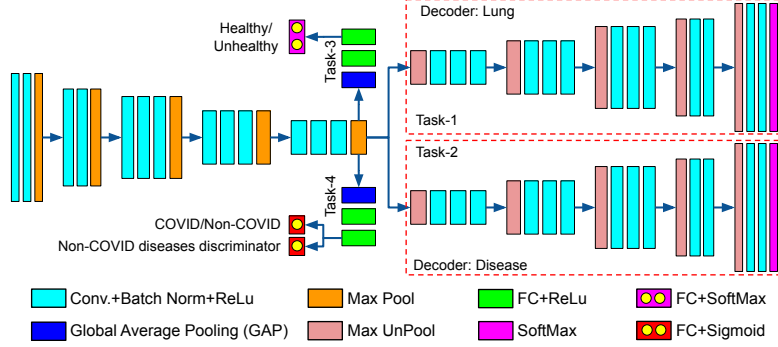


Figure 3: Architecture of the proposed COVID-19 Multi-task Network (COMiT-Net), which is based on a Encoder-Decoder architecture (Best viewed in color).

135 lung and disease localization, respectively.  $t_3$  and  $t_4$  represents the task of  
 140 healthy/unhealthy and COVID/Non-COVID classification, respectively.

### 2.1. Segmentation Loss

Chest x-ray of lungs contain peripheral organs along with lung regions. The  
 primary objective of this research is to differentiate between COVID and Non-  
 140 COVID samples. Since the key information lies in the lungs, the initial task is  
 that of lung segmentation. The second segmentation loss aims to learn semantic  
 segmentation of the diseased regions.

Lung segmentation can be achieved by learning a model that differentiates  
 between the background and foreground lung regions. The COMiT-Net ac-  
 145 complishes this by utilizing a VGG16 Encoder-Decoder architecture [27]. The  
 encoder has VGG16 as a base network. It has five blocks with 2, 2, 3, 3, and 3  
 layers of convolution + batch norm + ReLu layers, respectively. The decoder  
 network builds upon the representation obtained from the encoder network,  
 with a transposed architecture of the encoder network. At the final layers, the  
 150 output is derived from a SoftMax layer. The output dimension equals the input  
 spatial resolution of the X-ray image with the number of channels equaling the  
 number of segmentation classes. Hence, the final layer consists of two channels,



*lung* and *non-lung*.

Similar to lung localization, the disease localization also builds upon the  
 155 encoder representation. However, the disease localization task has a separate  
 decoder branch and is optimized for localizing more than 20+ lung-related dis-  
 orders. For both the lung and disease localization, the gradients are backprop-  
 agated via decoder network into the encoder layers.

Let  $f_{t_1}$  and  $f_{t_2}$  represent the sub-networks for lung and disease localization,  
 160 respectively. For any image  $\mathbf{X}_i$ , the output predicted binary masks for lung and  
 disease localization are represented as:

$$\hat{\mathbf{L}}_i = f_{t_1}(\mathbf{X}_i) \quad \text{and} \quad \hat{\mathbf{D}}_i = f_{t_2}(\mathbf{X}_i) \quad (1)$$

In this research, binary cross entropy loss is used for lung and disease localiza-  
 tion. Mathematically, it is represented as:

$$Z_{1i} = - \sum_{x,y} [\mathbf{L}_i(x,y) \log(\hat{\mathbf{L}}_i(x,y)) + (1 - \mathbf{L}_i) \log(1 - \hat{\mathbf{L}}_i(x,y))] \quad (2)$$

$$Z_{2i} = - \sum_{x,y} [\mathbf{D}_i(x,y) \log(\hat{\mathbf{D}}_i(x,y)) + (1 - \mathbf{D}_i) \log(1 - \hat{\mathbf{D}}_i(x,y))] \quad (3)$$

where,  $Z_{1i}$  and  $Z_{2i}$  are the lung and disease loss, respectively for image  $\mathbf{X}_i$ .  
 165  $\mathbf{L}_i(x,y)$  and  $\mathbf{D}_i(x,y)$  represent the pixel value at location  $(x,y)$  for lung and  
 disease masks, respectively.

## 2.2. Classification Loss

The two classification tasks are  $t_3 = \text{Healthy/Unhealthy}$  classification of the  
 lung X-ray, and  $t_4 = \text{Multi-label classification}$  for the presence of COVID-19 or  
 170 other abnormalities. These tasks are performed using three classification loss  
 functions. The lung and disease localization provides supervision for the three  
 classification tasks. For healthy/unhealthy, COVID/Non-COVID, and Non-  
 COVID diseases discrimination classification, two branches are derived over the

compact encoder representation (after GAP). Each branch has three fully connected layers (FC). For both branches, the first two layers use ReLu activation. The healthy/unhealthy branch uses SoftMax activation at the last FC layer. The multi-label COVID/Non-COVID and Non-COVID diseases discrimination classification branch uses Sigmoid activation at the last FC layer (Figure 3). Let  $f_{t_3}$  and  $f_{t_4}$  represent the sub-networks for healthy/ unhealthy and multi-label classification, respectively. The output of  $f_{t_3}$  for image  $\mathbf{X}_i$  is:

$$P(H_i|\mathbf{X}_i) = f_{t_3}(\mathbf{X}_i) \quad (4)$$

where,  $P(H_i|\mathbf{X}_i)$  is the probability of predicting image  $\mathbf{X}_i$  to  $H_i$ . The loss function for healthy/unhealthy classification is represented as:

$$Z_{3i} = - \sum_{H_i=\{0,1\}} H_i \log(P(H_i|\mathbf{X}_i)) \quad (5)$$

where,  $Z_{3i}$  represents the healthy/unhealthy loss for image  $\mathbf{X}_i$ . For multi-label classification, the output of sub-network  $f_{t_4}$  for an image  $\mathbf{X}_i$  is written as:

$$[\widehat{C}_i, \widehat{O}_i] = f_{t_4}(\mathbf{X}_i) \quad (6)$$

where,  $\widehat{C}_i$  and  $\widehat{O}_i$  represent the output predicted score ( $\in [0, 1]$ ) for COVID/Non-COVID and Non-COVID diseases discriminator, respectively. The radiological findings of COVID-19 pneumonia may overlap those of other viral pneumonia and acute respiratory distress syndrome due to other etiologies. The network needs supervision to segregate COVID-19 pneumonia from Non-COVID lung diseases. Hence, the joint optimization for COVID/Non-COVID along with Non-COVID diseases discrimination helps differentiate COVID-19 affected lungs from lungs affected with diseases other than COVID-19. The joint loss for predicting both COVID/Non-COVID and Non-COVID diseases discrimination is written as:

$$Z_{4i} = -[C_i \log(\widehat{C}_i) + (1 - C_i) \log(1 - \widehat{C}_i)] - [O_i \log(\widehat{O}_i) + (1 - O_i) \log(1 - \widehat{O}_i)] \quad (7)$$

Table 1: Details of the databases used in the experiments.

Database	Healthy	View	Images
Chest X-Ray-14 [28]	Healthy	PA	4088
		AP	2688
	Unhealthy	PA	3469
		AP	3115
CheXpert [29]	Healthy	PA	1331
		AP	2163
	UnHealthy	PA	3279
		AP	11305

195 **Overall Loss Function:** It is possible that the ground truth labels or segmen-  
tation masks are not available for all the images during training. In this case,  
all branches of the networks will not be active during training of COMiT-Net.  
For instance, if the ground truth mask is unavailable for disease segmentation,  
then the sub-network  $f_{t_2}$  will remain inactive and the loss  $Z_{2i}$  for image  $\mathbf{X}_i$   
200 will become zero. In the same manner, other losses can have a 0/1 “switch”.  
Therefore, the total loss  $\mathcal{L}$  is computed as:

$$\mathcal{L} = \sum_i T_{1i}Z_{1i} + T_{2i}Z_{2i} + T_{3i}Z_{3i} + T_{4i}Z_{4i} \quad (8)$$

where,  $T_{1i}$ ,  $T_{2i}$ ,  $T_{3i}$ , and  $T_{4i}$  are the switches pertaining to the tasks  $t_1$ ,  $t_2$ ,  $t_3$ ,  
and  $t_4$ , respectively. The values of these switches are either 0 or 1 depending  
on the availability of ground truth labels/masks of the respective tasks for the  
205  $i^{th}$  image.

### 3. Experimental Details

We next summarize the databases used for training and testing, the lung and  
disease annotations performed as part of this research, and the implementation  
details.

Table 2: Details for the COVID-19 databases used in the experiments.

Source	AP View	PA View	Total Images
GitHub [10]	50	26	76
Italy [30]	30	39	69
Spain [31]	0	110	110
RadioPaedia [32]	9	85	94
BSTI [33]	3	39	42
EuroRad [34]	6	18	24
BIMCV-COVID19+ [35]	NA	NA	2098
<b>Total</b>			<b>2513</b>

### 210 3.1. Database and Protocol

For different tasks of the network, we require a chest X-ray database with multiple annotations and diverse properties. Thus, the database for experiments is created by combining subsets from the ChestXray-14, CheXPert, and COVID-19 infected X-ray databases. We only use frontal X-ray in our experiments from  
215 the following publicly available databases:

- **ChestXray-14 [28]:** The dataset contains healthy and unhealthy x-ray images. It has a total of 112,120 chest x-ray images, out of which 67,310 are PA view images, and remaining 44,810 are AP view. Multiple radiographs of the same patient taken at different times are also present. From  
220 the database, we derive a subset of 13,360 images, spanning both PA and AP views. The unhealthy X-rays are labeled for one or more classes in a total of 14 classes. The 14 classes are: Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, and Pleural Thickening. Ad-  
225 ditionally, the dataset provides localization information of abnormalities for 880 X-rays. The details of the subset drawn from ChestXray-14 is illustrated in Table 1.

Table 3: Details of train-test split across different parameters. Train set for COVID-19 includes augmentation. Cols 1-2 specify number of samples for segmentation tasks. Cols 3-5 specify number of samples for classification task. Cols 6-7 specify total number of AP/PA view samples in the database.

	Mask		Disease-wise			Views	
	Lung	Disease	Normal	Covid	Others	PA	AP
Train	8730	1456	8173	1740	16551	10161	16690
Test	1837	251	2077	2223	4097	2464	3968
Total	10567	1707	10250	3963	20648	12625	20658

- 230
**CheXpert [29]:** The CheXpert dataset contains a total of 223,414 chest x-ray images, out of which 29,420 are PA view, 161,590 are AP view, and the remaining are lateral or single lung view images. Multiple case studies of the same patient are available in the dataset. This dataset contains healthy and unhealthy X-ray images. We selected a subset of 18,078 images. Based on the radiological findings, each X-ray image is labeled positive/negative for 14 pre-defined classes (few overlapping with ChestXray-14). The 14 classes are: No Finding, Enlarged Cardiom, Lung Lesion, Edema, Consolidation, Pneumonia, Atelectasis, Pneumothorax, Cardiomegaly, Pleural Effusion, Pleural Other, Lung Opacity, Fracture, and Support Devices. The details of the x-ray images selected from CheXpert database is shown in Table 1.
- 240
**COVID-19:** For this study, we collected a total of 415 X-rays from various internet sources. The sources have a mixed number of PA and AP view frontal chest x-ray. The number of X-rays collected from each source has been summarized in Table 2. Further, we have also performed additional experiments with 2388 images from the BIMCV+ COVID-19 Database [35]. Details and experimental results of the BIMCV+ database can be found in Section 4.4.

245

Since the above COVID-19 subset has a limited number of images, we per-

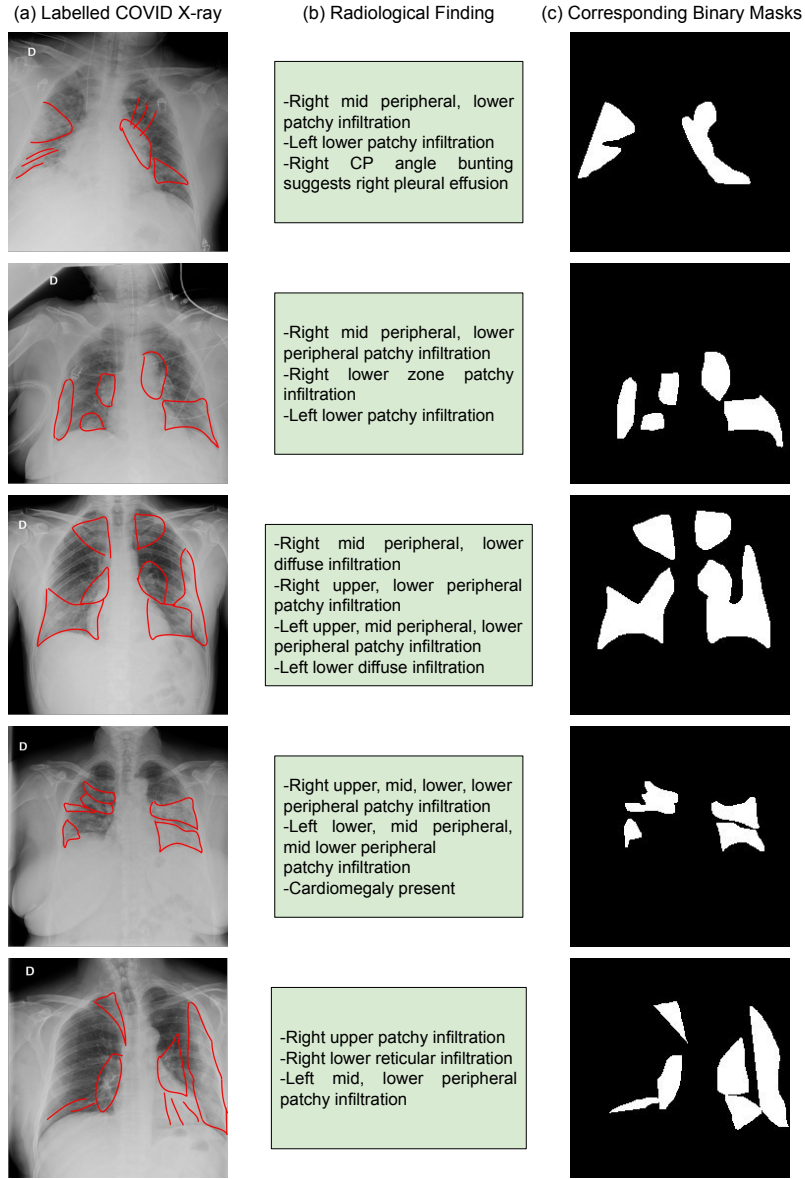


Figure 4: Annotations provided for COVID-19 affected frontal lung x-ray images as a part of this study: (a) Labeled COVID-19 X-ray for locations of radiological finding, (b) Description of the radiological finding, (c) Corresponding binary masks for training deep semantic segmentation algorithms for disease segmentation.

form data augmentation. Each image is augmented five ways - clockwise rotation by  $10^\circ$ , anti-clockwise rotation by  $10^\circ$ , translation by 10 pixels in the X, Y, and XY-directions. Since pneumonia is a closely related pathology to COVID [36], we select all the pneumonia samples of the ChestXray-14 and CheXPert datasets. Further, to accommodate the variations in non-healthy x-ray samples, about 50% more unhealthy samples are selected compared to healthy samples. AP view x-rays are prominent compared to PA views in the CheXpert dataset. Hence, we select more AP view X-ray images.

The data is split into training and testing ensuring that there is no patient overlap in the train and test sets. The details of the train-test data split across different properties are specified in Table 3. The first two columns specify the number of samples for the task of segmentation, i.e. the samples for which disease and lung masks are available. The next three columns specify the number of samples present in different classes (normal, COVID, others) for the task of classification. The last two columns specify the characteristics of the overall database by subdividing the total number of samples into two categories- AP and PA. AP and PA views of chest x-rays are substantially different, and we attempt to balance the two views to provide balanced training. Note that all the train set numbers mentioned in the table are post-augmentation.

### 3.2. Lung and Disease Region Annotation

The datasets mentioned above lack lung localization details. The proposed COMiT-Net requires a ground-truth lung location to identify the lung region from the x-ray. For this purpose, we manually annotated a total of about 9000 lung x-rays. These x-rays include well-balanced healthy/unhealthy, AP/PA subsets taken equally from the CheXpert and ChestXray-14 datasets. All x-ray images available for COVID-19 are also manually annotated for lung segmentation. Mask for each x-ray image has been created by drawing two solid bounding boxes, corresponding to the area covered by each lung. As a part of this study, we also plan to release the ground truth masks for the manually annotated lung regions.

The datasets included as a part of this study have only 880 disease localization annotation images (from ChestXray-14 database). For COVID-19 affected frontal lung x-ray images, we lacked disease segmentation masks. Hence, as a part of this study, the x-ray images are annotated by a radiologist for various radiological findings. The findings radiologists looked for includes: (i) atelectasis, (ii) consolidation, (iii) interstitial shadows (reticular, nodular, ground glass), (iv) pneumothorax, (v) pleural effusion, (vi) pleural thickening, (vii) cardiomegaly, and (viii) lung lesion. The experts annotated a total of 200 COVID-19 affected chest x-rays. A few sample annotations for the same can be seen in Figure 4(a) and the corresponding description in Figure 4(b). While training deep learning algorithms, the model requires binary masks as annotation. Hence, we created these masks based on the annotations (Figure 4(c)). We will release the ground truth binary masks to promote the training of deep semantic segmentation algorithms for abnormality localization.

### 3.3. Implementation Details

The proposed Multi-task network requires input X-ray images of size  $224 \times 224 \times 3$ . The encoder stream is initialized using a pre-trained VGG16 model. With a batch size of 16, the model is optimized over binary cross-entropy loss using Adam optimizer (learning rate =  $5 \times 10^{-5}$ ). Each loss is weighted equally. The model is trained for 30 epochs<sup>6</sup> on NVIDIA GeForce RTX 2080Ti and implemented in PyTorch.

UNet and SegNet models are trained on Nvidia RTX 2080Ti using a PyTorch (v.1.4.0) implementation. The input size is kept the same as  $224 \times 224 \times 3$  with batch size of 4. The model is trained for 25 epochs by minimizing binary Cross-Entropy loss using Adam optimizer with an initial learning rate of 0.0001. Other Python library requirements include torchvision (v.0.5.0), tqdm

---

<sup>6</sup>It is observed that the loss of the model converges around the 30th epoch. To make sure the model is trained and does not overfit, we track the training losses and early stop at 30th epoch.



(v.4.45.0), tensorboardX (v.1.1), and Pillow (7.0.0). Similarly, Mask-RCNN op-  
erates on same sized images with a training batch size of 16. The model is  
trained for 25 epochs by minimizing binary Cross-Entropy loss using SGD op-  
timizer with a learning rate of 0.001 and momentum 0.9. The model is trained  
on Google Colab with Nvidia Tesla T4 as the GPU accelerator using a PyTorch  
(1.6.0+cu101) implementation.

The comparative algorithms include DenseNet121 [15, 16, 17], MobileNetv2 [4,  
5], ResNet18 [6, 37, 7], and VGG19 [13, 38]. For each of these networks, the  
ImageNet pre-trained version is selected. The model is then fine-tuned with  
the dataset and protocol used for the proposed COMiT-Net. The input size,  
batch size, and epochs are kept same as COMiT-Net, i.e.,  $224 \times 224 \times 3$ , 16, and  
30 respectively.

## 4. Results and Analysis

We next evaluate the performance of the proposed COMiT-Net for classifi-  
cation and localization tasks. The performance is compared with existing deep  
learning algorithms for COVID-19 chest radiograph studies. Further, to study  
the effectiveness of the proposed COMiT-Net, we perform experiments by select-  
ing different combinations of sub-networks from the COMiT-Net in subsection  
5.3. Lastly, subsection 5.4 specifically tasks about prediction of COVID-19 af-  
fected CXR. The predictions are presented on a large COVID-19 positive CXRs  
database, validated against predictions from radiologists.

### 4.1. Lung and Disease Localization

In this subsection, the segmentation results of the proposed COMiT-Net  
are compared against region predictions from UNet [39], Mask RCNN [40], and  
SegNet [27]. For lung segmentation, sample predictions of the proposed and  
existing algorithms are shown in Figure 5. Inferring the sample prediction,  
we observe that all four algorithms perform well and give comparable results.  
However, the proposed COMiT-Net yields the most precise bound for lung seg-  
mentation. To support the visual results presented in Figure 5, we additionally

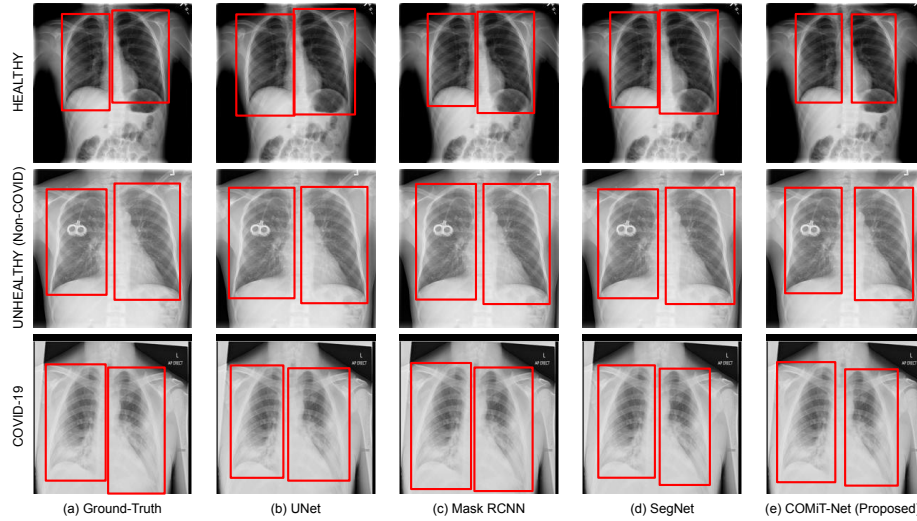


Figure 5: Samples of lung segmentation output for existing algorithms and the proposed COMiT-Net.

report the Intersection over Union (IoU) for lung segmentation. The IoU scores corresponding to the U-Net, Mask-RCNN, SegNet, and COMiT-Net are 0.82, 0.85, 0.83, and 0.85, respectively. The reported IoU for COMiT-Net is same as the state-of-the-art segmentation method Mask-RCNN. We can also visually observe in Figure 5 that COMiT-Net provides the tightest bound for lungs. Since lung and disease localization tasks are performed simultaneously, and diseases are present within the lungs, the lung decoder network learns to focus more on the lung regions rather than the outside the lungs.

The results of disease segmentation are shown in Figure 6. The first two rows of Figure 6 illustrate abnormalities in COVID-19 affected lungs while last two rows have abnormality localization in unhealthy but Non-COVID affected lungs. For disease localization, a more relevant metric is True Positive Rate (TPR). When localizing a disease, we would not want to miss detection of diseased regions (even if some non-diseased regions get predicted as diseased). Hence, the reported TPR corresponding to the U-Net, Mask RCNN, SegNet,

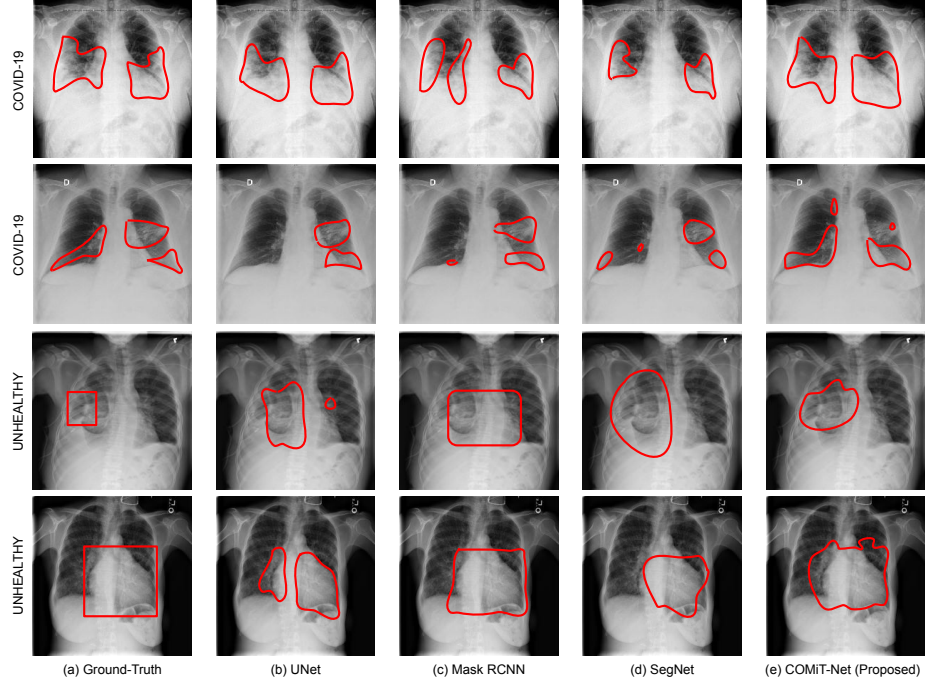


Figure 6: Samples of semantic disease segmentation for existing algorithms and the proposed COMiT-Net. The x-ray images and corresponding abnormality localization for “Unhealthy” are derived from ChestXray-14 database [28].

and COMiT-Net are 0.42, 0.51, 0.31, and 0.87, respectively. The masks for all the four algorithms are predicted at a constant threshold of 0.5 (disease treated as foreground; label=1). Using the predicted mask, we calculate the TPR for the foreground disease classification. From the perspective of shape, Mask-RCNN tends to provide well-defined shape boundaries for Non-COVID unhealthy lungs. SegNet and COMiT-Net provide irregularly shaped predictions, localizing the radiological findings compactly. Overall, we observe that each of the four algorithms predict additional regions for the abnormalities. The detected abnormalities have false positive regions when compared to the ground-truth. As shown in Figure 7(c) and (d), these false positives sometimes arise due to better localization by SegNet and proposed COMiT-Net. The comparative

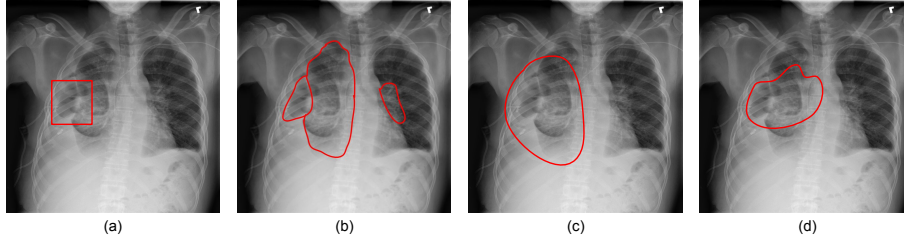


Figure 7: Non-COVID affected unhealthy lung, with (a) ground-truth annotation from ChestXray-14 database, (b) abnormality manually marked by a radiologist (as a part of this study), (c) disease prediction from SegNet, and (d) disease prediction from the proposed COMiT-Net.

ground-truth provided with the database is shown for reference in Figure 7(a).

Further, we observe that for certain abnormalities in ‘Unhealthy’ case, deep  
 360 models fail to localize the abnormality. One of the reasons for this is the limited training data for abnormality localization with large variations in the diseased regions. The unhealthy Non-COVID lung abnormalities are derived from ChestXray-14, which has 700 samples corresponding to 14 labels. As a result of  
 365 a small sample size for each abnormality, the networks cannot localize diseases properly. However, the proposed COMiT-Net has assistance from other tasks. For instance, the lung prediction task would implicitly reinforce COMiT-Net to predict diseases within the lung. Hence, of the four algorithms, the proposed COMiT-Net provides the most overlapping prediction with the ground truth.

370 Compared to 700 samples for 14+ different radiological findings (approx. 50 images per abnormality), there are 290 COVID-19 affected lung x-rays (prior to augmentation). A majority of the COVID-19 affected chest radiographs demonstrate consolidations, which tend to be bilateral and more common in lower zones [41]. Hence, deep models have more samples to learn the localization of COVID-19 specific abnormalities than other diseases (290 vs. 50).  
 375 In retrospect, the first two rows of Figure 6 illustrate that all four models perform relatively better for COVID-19 localization than the last two rows of

“unhealthy” localization. In most cases, each of the four models predict affected regions in the lower lung zones bilaterally. However, the proposed COMiT-Net  
380 outperforms other algorithms. For instance, in the first row of Figure 6, both Mask-RCNN and SegNet tend to leave out the darker region in the right lung, while ground-truth and COMiT-Net have that region marked as diseased. Further, in the low contrast x-ray in row two, the less opaque part of the right lower lung looks darker (though being diseased). Hence, UNet fails to detect  
385 any finding in the right lower lung, while Mask RCNN and SegNet detects a few small region(s). Nevertheless, the proposed COMiT-Net can detect such faint differences in lung density.

#### 4.2. Classification

Next, we evaluate the COMiT-Net’s performance for healthy/unhealthy  
390 (Task 3) classification and multi-label classification of COVID-19 and other diseases (Task 4). For COVID/Non-COVID classification, a branch is derived from the last layer of the encoder network i.e., GAP which outputs the embedding of the input samples. The branch has three fully connected layers (FC) where the first two layers use ReLu activation and the last FC layer uses sigmoid activation  
395 for classification. The results of the COMiT-Net are compared against popular deep networks. These include DenseNet121 [15, 16, 17], MobileNetv2 [4, 5], ResNet18 [6, 37, 7], and VGG19 [13, 38]. Further, we draw a comparison with Random Decision Forest (RDF) [42] and Support Vector Machines (SVM) [43] with three different kernels- sigmoid, gaussian, and radial basis function (RBF).  
400 In the COMiT-Net Embedding + RDF, the embeddings of the input samples are obtained from the GAP layer of the encoder network and are further used to train the RDF classifier. Similarly, for the COMiT-Net Embedding + SVM (Sigmoid), COMiT-Net Embedding + SVM (Gaussian), and COMiT-Net Embedding + SVM (RBF), the output embeddings are used to train SVM classifier  
405 with sigmoid, gaussian, and rbf kernels.

The results for classification performance are presented in Table 4. In the Table, the sensitivity results are reported at two fixed specificities, i.e., 90% and

Table 4: Evaluation and comparison of the proposed COMiT-Net with existing algorithms for COVID-19 prediction (FC = Fully Connected Classification Layers). Y denotes varying specificity values. Col 1 with Y=99% denotes the sensitivity of the network at 99% specificity.

	<b>Sensitivity@Y Specificity</b>		<b>EER (%)</b>
	<b>Y = 99%</b>	<b>Y = 90%</b>	
DenseNet121 + FC	60.80	90.40	9.82
MobileNetv2 + FC	67.20	93.60	8.04
ResNet18 + FC	56.00	81.60	13.78
VGG19 + FC	50.40	82.40	13.70
COMiT-Net Embedding + RDF	79.20	95.20	7.34
COMiT-Net Embedding + SVM (Sigmoid)	6.40	24.00	41.46
COMiT-Net Embedding + SVM (Gaussian)	82.40	88.80	11.38
COMiT-Net Embedding + SVM (RBF)	82.40	88.80	11.38
<b>COMiT-Net (Proposed)</b>	<b>87.20</b>	<b>96.80</b>	<b>7.30</b>

99%. At these fixed specificity values, the observed sensitivity is 96.80% and 87.20%, respectively. Further, for COVID classification, we observe an overall  
410 test classification performance<sup>7</sup> of 98.79%. Lastly, we also computed precision and recall for the proposed method. At 99% specificity, the precision is 64.12% and recall is 87.20%. Additionally, using precision and recall, the F1 score is found as 73.90%.

The proposed COMiT-Net achieves the highest TPR and lowest EER compared to the existing algorithms. With the implicit supervision from lung and  
415 disease localization tasks, the proposed COMiT-Net outperforms all other existing algorithms. To show the stability of different algorithms with different

---

<sup>7</sup>The overall test classification accuracy is =  $\frac{TP+TN}{TP+TN+FP+FN}$

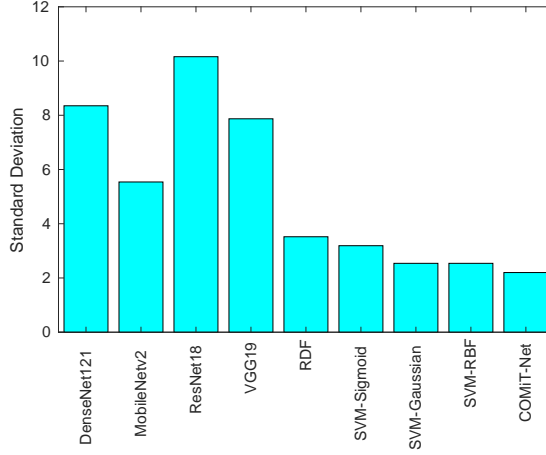
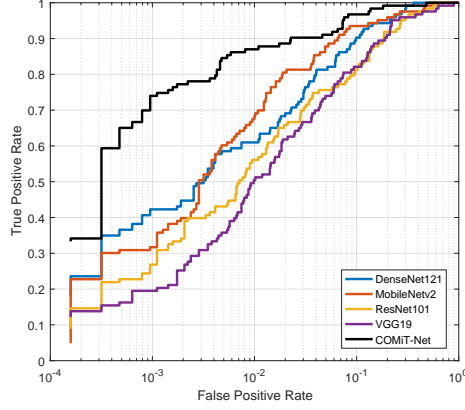


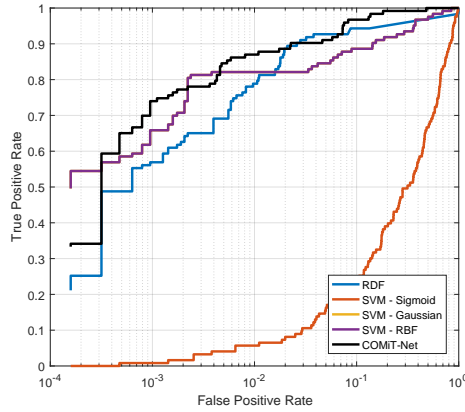
Figure 8: Standard deviation ( $\pm$ ) of Sensitivity (at 1% FAR) for different algorithms. The performance is computed for different initialization of deep networks. The results show the stability in sensitivity for COMiT-Net, delivering consistent results for different initializations.

initialization, the networks are three-times trained with different initialization parameters. Across different training initializations, we report the standard deviation in Sensitivity to evaluate the stability (lower standard deviation implies higher stability). As shown in Figure 8, the proposed COMiT-Net is the most stable algorithm across different initializations. Classifiers that use embeddings from COMiT-Net also report lower standard deviation. Hence, it can be inferred that COMiT-Net consistently provides a discriminative representation, resulting in a stable performance. Figure 9 further shows the comparison using the ROC curves of the proposed COMiT-Net and existing algorithms.

The COMiT-Net’s classification performance for the COVID-19 samples into the healthy and unhealthy class is also analyzed. The proposed network classifies 97.25% of COVID-19 samples into unhealthy class and 2.75% in healthy class. The high TPR of the COVID-19 class and the majority of the COVID-19 samples being classified into unhealthy class showcase the effectiveness of the proposed network for COVID-19 detection. Overall, the classification performance of healthy/unhealthy classification is 75.17% for all the test samples, while for



(a) Comparison of the proposed COMiT-Net with existing deep learning models.



(b) COMiT-Net embedding in combination with different classifiers.

Figure 9: ROC curves summarizing the performance for COVID-19 classification.

Non-COVID disease classification is 73.87%. Based on the proposed COMiT-Net, Figure 10 shows some of the misclassified samples where the network predicts COVID-19 positive instances (as per the RT-PCR test) into healthy (Task 3). Correspondingly, the same samples are also predicted as Non-COVID by Task 4 of the proposed COMiT-Net. In retrospect, we believe that minimal opacities in the lung region could be the probable cause of misclassification. This led us to check the ground truth for the hospitalization day. Of the four



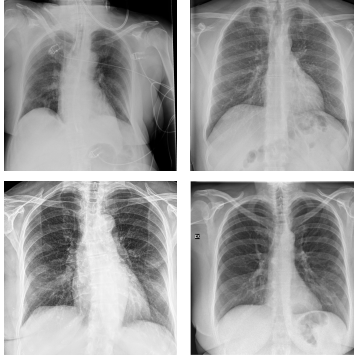


Figure 10: COVID-19 positive case misclassified as both healthy and Non-COVID by the proposed COMiT-Net.

misclassified samples shown in Figure 10, three turned out to be the early days of the patient’s hospitalization (up to day 3). Based on these observations, we believe that the COMiT-Net predicts an x-ray being affected when there is presence of symptoms such as opacities and consolidations.

#### 4.3. Ablation Study

To study the importance of different tasks in the proposed COMiT-Net, we perform an ablation study by choosing different combinations of tasks. The four tasks in the COMiT-Net are Task 1: Semantic lung segmentation, Task 2: Semantic disease segmentation, Task 3: Healthy/Unhealthy classification of the lung X-ray, and Task 4: Multi-label classification for the presence of COVID-19 or other diseases. We perform eight different ablation experiments, presented in the Table 5. It is observed that for COVID-19 prediction, each task (loss function) has an important role. Removing either of the three assisting tasks deteriorates the performance. Of all these three assisting tasks, the lung segmentation task holds a pivotal role. In a COVID-19 affected x-ray, a common trait is that the lungs get affected bilaterally. Hence, a comprehensive view provided by the lung segmentation task provides more weight to lung regions, resulting in better performance with Task 1 than any other task. We perform

Table 5: An ablation study on reducing the number of tasks and observing its effect on COVID-19 prediction.

	COVID-19 (Sensitivity %)
All 4 Tasks	<b>96.80</b>
Task 4 Only	84.40
Task 1 and 4	94.40
Task 2 and 4	57.60
Task 3 and 4	67.80
Task 1, 2 and 4	92.80
Task 1, 3 and 4	87.20
Task 2, 3 and 4	54.40

disease segmentation and healthy/unhealthy classification since their efficacy improves in conjunction with lung segmentation and has a positive impact on the *Non-COVID disease* classification prediction. As validated by the ground-truth t-SNE feature space plot (shown in Figure 11(a)), the predictions of the test COVID-19 samples (Figure 11(b)) are well separated from Non-COVID samples. It shows that the model can distinguish COVID-19 affected samples and can predict unseen test labels correctly. Further, we use Grad-CAM [44] which is a popular tool for producing ‘visual explanations’ for decisions obtained from deep learning architectures. As observed in the Grad-CAM analysis (Figure 12), the proposed COMiT-Net focuses on diseased regions, in both COVID-19 and unhealthy test samples, to make its prediction.

In the real-world scenario, annotating disease and lung masks for chest x-rays is a time consuming and challenging task. The performance of the proposed COMiT-Net is dependent on the available annotated data for the different tasks. It might not be feasible to have an adequate amount of disease and lung masks for the X-ray images in the training set. Moreover, the performance of the proposed model for the task of COVID/Non-COVID classification is dependent on the task of lung and disease segmentation. Removing either of these assisting

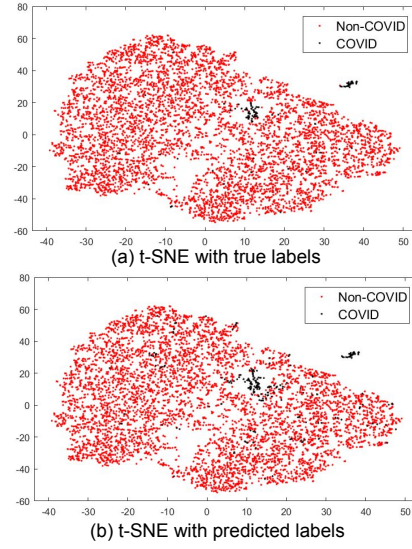


Figure 11: Interpretation of feature representation based on (a) ground-truth and (b) predicted labels using t-SNE plot for COVID/Non-COVID classification.

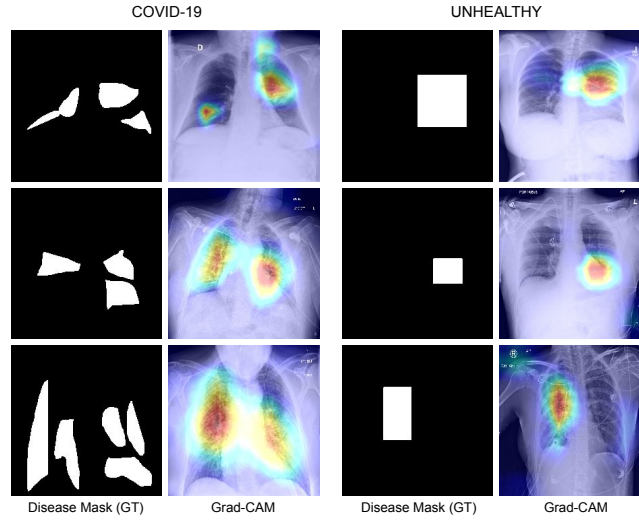


Figure 12: Interpretation of regions focused by COMiT-Net using Grad-CAM. As seen from heat maps, the supervision from disease and lung annotation helps the COMiT-Net to focus on unhealthy regions.

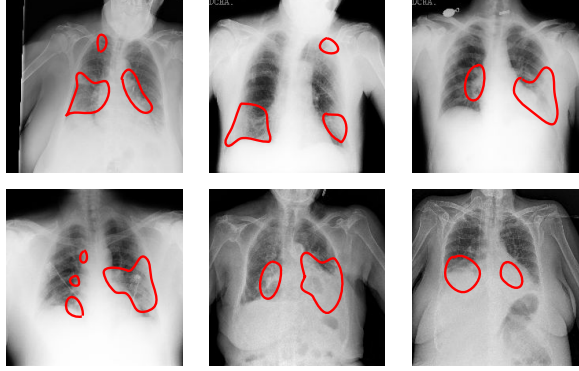


Figure 13: Few instances of semantic disease segmentation from the proposed COMiT-Net for the BIMCV+ COVID-19 database.

tasks deteriorates the performance of the COMiT-Net.

#### 4.4. Prediction on Unseen COVID X-ray Database

Recently, BIMCV-COVID19+ database has been released by the Medical  
 480 Imaging Databank in Valencian Region (BIMCV). To demonstrate the performance of the proposed COMiT-Net in a real-world scenario, we report the results on the BIMCV-COVID19+ database as well. At the time of download, the database had 2388 frontal x-ray images. The x-ray is captured from COVID positive patients during hospitalization. For each patient, the database has one  
 485 or more x-ray along with the report of one or more COVID-19 diagnostics tests (RT-PCR, IGM, IGG) with its timestamps. It is to be noted that the timestamp of the x-ray does not coincide with the timestamp of the diagnostics test. Hence, if an x-ray is taken closer to a negative diagnostics test or is at least 14 days away from the nearest positive diagnostics test, the x-ray is considered  
 490 as COVID-19 negative. Otherwise, we label the x-ray as COVID-19 positive. With this procedure, there are a total of 2,098 COVID-19 positive x-rays which are used for testing purposes only (no training is performed on this dataset). For reproducibility, we will release these labels along with filenames of the radiographs.

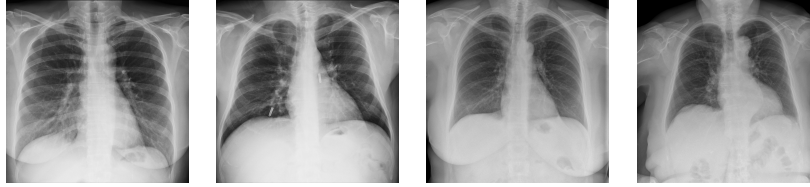


Figure 14: Sample instances that were labeled by BIMCV-COVID19+ as positive but the COMiT-Net and radiologists predicted negative.

495 Out of 2,098 COVID-19 positive x-rays, the trained COMiT-Net correctly classifies 1,793 samples and misclassifies 305 samples (at 10% FAR). Hence, at 90% specificity, the COMiT-Net has a sensitivity of 85.46%. Sample instances of disease segmentation from the correctly classified 1,793 samples are shown in Figure 13. Similar to previous disease segmentation instances, COMiT-Net is able to localize abnormalities bilaterally.

500 To further understand the behavior of the 305 misclassified instances, each of these samples is verified by radiologists to affirm if any radiological abnormality is present or not. Of these 305, there are 115 instances where radiologists confirmed the absence of any abnormality. Few sample instance of these cases where the COMiT-Net and radiologists predicted negative yet BIMCV-COVID19+ labelled positive is shown in Figure 14.

510 Lastly, we show additional results on COVID-19 subsets of the COVID-19 Radiography dataset<sup>8</sup>. Of the six COVID subsets presented in the dataset, our proposed method had already used four (BIMCV, EuroRad, Github, SIRM). For the remaining two subsets, at 90% specificity threshold, the sensitivity is 98.75% for the ARMIRO subset (400 images) and 99.45% for the ML-workgroup subset (183 images).

---

<sup>8</sup><https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>

## 5. Conclusion and Future Work

In the face of the SARS-CoV2 pandemic, it has become essential to perform  
515 mass screening and testing of patients. However, many countries around the  
world are not equipped with enough laboratory testing kits or medical person-  
nel for the same. At the same time, X-rays are amongst the most popular,  
cost-effective and widely available imaging technology across the world. This  
paper presents an “explainable solution” for detecting COVID-19 pneumonia  
520 in patients through chest radiographs. We propose **COMiT-Net** which per-  
forms the tasks of classification and segmentation simultaneously. Experiments  
conducted on different chest radiograph datasets show promising results of the  
proposed algorithm in COVID prediction. The ablation study supports the uti-  
lization of different tasks in the proposed multi-task network. We believe that  
525 the proposed COMit-Net can be used as an attractive alternative solution that  
can assist the doctors and the research community to speed up the screening  
process of COVID cases.

In future, we plan to extend this work and use the proposed framework  
for the task of COVID-19 prediction using modalities other than X-ray such  
530 as CT and ultrasound. Since these modalities provide complementary infor-  
mation such as nature and formation of abnormalities present in the diseased  
region, incorporating them will help in crafting a robust solution. Further, we  
need to understand distinguishable traits between COVID-19 pneumonia and  
non-COVID viral pneumonia. Learning these traits can be useful in detecting  
535 COVID-19 pneumonia.

## Acknowledgments

The authors acknowledge RAKSHAK project under NM-CPS DST and  
iHuB Drishti Jodhpur. Malhotra is partially supported by Visvesvaraya Ph.D.  
Fellowship. Mittal is partially supported by UGC-Net JRF Fellowship. Majum-  
540 dar is partially supported by DST Inspire Ph.D. Fellowship.

## References

- [1] A. Schiffmann, World COVID-19 Stats, <https://ncov2019.live/>, [Accessed: 6-March-2021] (2020).
- [2] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hesketh, L. Yang, et al., Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia, *Radiology* 295 (3) (2020) 715–721.
- [3] T. T. Nguyen, Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions, Preprint, DOI: 10.13140/RG.2.2.36491.23846 10.
- [4] I. D. Apostolopoulos, T. A. Mpesiana, Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, *Physical and Engineering Sciences in Medicine* 43 (2) (2020) 635–640.
- [5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [6] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, *arXiv:2003.10849*.
- [7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
- [9] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, in: *AAAI conference on Artificial Intelligence*, Vol. 31, 2017.

- [10] J. P. Cohen, P. Morrison, L. Dao, Covid-19 image data collection, arXiv 2003.11597[Accessed: 6-Nov-2020].  
 URL <https://github.com/ieee8023/covid-chestxray-dataset>
- [11] P. Mooney, Chest X-Ray Images (Pneumonia)[Accessed: 6-Sept-2020].  
 URL <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- [12] M. Nishio, S. Noguchi, H. Matsuo, T. Murakami, Automatic classification between covid-19 pneumonia, non-covid-19 pneumonia, and the healthy on chest x-ray image: combination of data augmentation methods, Scientific reports 10 (1) (2020) 1–6.
- [13] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, N. Shukla, Covid-19 detection through transfer learning using multimodal imaging data, IEEE Access 8 (2020) 149808–149824.
- [14] Y. Oh, S. Park, J. C. Ye, Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets, IEEE Transactions on Medical Imaging 39 (8) (2020) 2688–2700.
- [15] A. Mangal, S. Kalia, H. Rajgopal, K. Rangarajan, V. Namboodiri, S. Banerjee, C. Arora, CovidAID: COVID-19 Detection Using Chest X-Ray, arXiv preprint arXiv:2004.09803.
- [16] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, M. Kaur, Classification of the covid-19 infected patients using densenet201 based deep transfer learning, Journal of Biomolecular Structure and Dynamics (2020) 1–8.
- [17] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
- [18] B. Ghoshal, A. Tucker, Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection, arXiv preprint arXiv:2003.10769.



- [19] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: European Conference on Computer Vision, Springer, 2016, pp. 630–645.
- 600 [20] W. Shi, L. Tong, Y. Zhuang, Y. Zhu, M. D. Wang, EXAM: An Explainable Attention-based Model for COVID-19 Automatic Diagnosis, in: ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2020, pp. 1–6.
- 605 [21] N. Tsiknakis, E. Trivizakis, E. E. Vassalou, G. Z. Papadakis, D. A. Spandidos, A. Tsatsakis, J. Sánchez-García, R. López-González, N. Papanikolaou, A. H. Karantanas, et al., Interpretable artificial intelligence framework for covid-19 screening on chest x-rays, *Experimental and Therapeutic Medicine* 20 (2) (2020) 727–735.
- 610 [22] M. Loey, F. Smarandache, N. E. M Khalifa, Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning, *Symmetry* 12 (4) (2020) 651.
- [23] L. Wang, Z. Q. Lin, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, *Scientific Reports* 10 (1) (2020) 1–12.
- 615 [24] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, A. Mohammadi, Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images, *Pattern Recognition Letters* 138 (2020) 638–643.
- 620 [25] M. Shorfuzzaman, M. S. Hossain, Metacovid: A siamese neural network framework with contrastive loss for n-shot diagnosis of covid-19 patients, *Pattern Recognition* 113 (2021) 107700.
- [26] C. M. Cutillo, K. R. Sharma, L. Foschini, S. Kundu, M. Mackintosh, K. D. Mandl, T. Beck, E. Collier, C. Colvis, K. Gersing, V. Gordon, R. Jensen, B. Shabestari, N. Southall, M. in Healthcare Workshop Working Group,

- Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency, *Digital Medicine* 3 (1) (2020) 47.
- [27] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12) (2017) 2481–2495.
- [28] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, *arXiv preprint arXiv:1711.05225*.
- [29] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghighi, R. Ball, K. Shpanskaya, et al., Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison, in: *AAAI Conference on Artificial Intelligence*, Vol. 33, 2019, pp. 590–597.
- [30] SIRM, COVID-19 Database[Accessed: 6-Sept-2020].  
URL <https://www.sirm.org/category/senza-categoria/covid-19/>
- [31] C. Imaging, COVID-19 CXR Spain[Accessed: 6-Sept-2020].  
URL <https://threadreaderapp.com/thread/1243928581983670272.html#>
- [32] RadioPaedia, Search results for “covid 19”, [Accessed: 6-Sept-2020].  
URL <https://radiopaedia.org/search?utf8=%E2%9C%93&q=covid+19&scope=all&lang=us>
- [33] BSTI, COVID-19 BSTI IMAGING DATABASE, [Accessed: 6-Sept-2020].  
URL <https://www.bsti.org.uk/training-and-education/covid-19-bsti-imaging-database/>
- [34] EuroRad, EuroRad Search results for COVID-19, [Accessed: 6-Sept-2020].  
URL <https://www.eurorad.org/advanced-search?search=covid-19>

- [35] BIMCV, BIMCV-COVID19, [Accessed: 16-Oct-2020] (2020).  
URL <https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/>
- [36] B. Nazari, Coronavirus and Pneumonia, [Accessed: 10-Nov-2020] (2020).  
URL <https://www.webmd.com/lung/covid-and-pneumonia#1>
- 655 [37] M. Z. Che Azemin, R. Hassan, M. I. Mohd Tamrin, M. A. Md Ali, COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings, International Journal of Biomedical Imaging 2020 (2020) 1988–1996.
- [38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-  
660 scale image recognition, arXiv preprint arXiv:1409.1556.
- [39] O. Ronneberger, P. Fischer, T. 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.
- 665 [40] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
- [41] J. Sawani, How Does COVID-19 Appear in the Lungs? (2020).  
URL <https://labblog.uofmhealth.org/lab-report/how-does-covid-19-appear-lungs>
- 670 [42] H. T. Kam, et al., Random decision forest, in: International Conference on Document Analysis and Recognition, Vol. 1, 1995, pp. 278–282.
- [43] J. A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Processing Letters 9 (3) (1999) 293–300.
- [44] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based  
675 localization, International Journal of Computer Vision 128 (2) (2020) 336–359.