COVID-19 Severity Prediction from Chest X-ray Images Using an Anatomy-Aware Deep Learning Model

Nusrat Binta Nizam[1] · Sadi Mohammad Siddiquee[1] · Mahbuba Shirin[2] · Mohammed Imamul Hassan Bhuiyan[3] · Taufiq Hasan[1,4]

Abstract

The COVID-19 pandemic has been adversely affecting the patient management systems in hospitals around the world. Radiological imaging, especially chest x-ray and lung Computed Tomography (CT) scans, plays a vital role in the severity analysis of hospitalized COVID-19 patients. However, with an increasing number of patients and a lack of skilled radiologists, automated assessment of COVID-19 severity using medical image analysis has become increasingly important. Chest x-ray (CXR) imaging plays a significant role in assessing the severity of pneumonia, especially in low-resource hospitals, and is the most frequently used diagnostic imaging in the world. Previous methods that automatically predict the severity of COVID-19 pneumonia mainly focus on feature pooling from pre-trained CXR models without explicitly considering the underlying human anatomical attributes. This paper proposes an anatomy-aware (AA) deep learning model that learns the generic features from x-ray images considering the underlying anatomical information. Utilizing a pre-trained model and lung segmentation masks, the model generates a feature vector including disease-level features and lung involvement scores. We have used four different open-source datasets, along with an in-house annotated test set for training and evaluation of the proposed method. The proposed method improves the geographical extent score by 11% in terms of mean squared error (MSE) while preserving the benchmark result in lung opacity score. The results demonstrate the effectiveness of the proposed AA model in COVID-19 severity prediction from chest X-ray images. The algorithm can be used in low-resource setting hospitals for COVID-19 severity prediction, especially where there is a lack of skilled radiologists.

Keywords Chest x-ray analysis · Anatomy-aware modeling · COVID-19 severity prediction

Introduction

The novel Coronavirus (COVID-19) pandemic has become a grave public health concern worldwide. According to the World Health Organization (WHO) [1], there were more than 683 million confirmed COVID-19 cases worldwide and 6.8 million deaths as of now in 2023. With the increasing number of patients, the need for streamlined patient management has become extremely important. Disease severity monitoring can be of significant importance for the timely allocation of hospital resources and predicting which patients may need intensive care unit (ICU) support. Biosensors, Artificial intelligence (AI), and the Internet of Medical Things (IoMT) have paved the pathway of modern COVID-19 diagnostic methods [2]. Chest x-ray imaging is a widely available non-invasive tool used for frequent assessment of COVID-19 severity. Although the High-Resolution and Computer-Aided Computed Tomography methods are

✉ Taufiq Hasan
taufiq@bme.buet.ac.bd

1 mHealth Research Group, Department of Biomedical Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1205, Bangladesh

2 Department of Radiology and Imaging, Bangabandhu Sheikh Mujib Medical University, Shahbagh, Dhaka 1000, Bangladesh

3 Department of Electrical and Electronics Engineering (EEE), Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh

4 Center for Bioengineering Innovation and Design (CBID), Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

best suited for COVID-19 severity assessment [3, 4], CXRs can also be useful for regular monitoring due to their relative speed, low cost, availability, and portability, especially in low-resource settings. With the advent of portable equipment, CXR imaging is more reasonable for critically ill patients who cannot be transported for a CT scan.

During the peak of the pandemic, numerous admitted patients endured additional constraints in terms of medical image interpretation in the hospitals. In high-income countries, high patient volumes may result in a higher workload on the existing radiologists, whereas in low-income countries, there may not be a sufficient number of skilled radiologists to perform the necessary image interpretation. In these circumstances, automated analysis of CXRs for the severity prediction of COVID-19 can be of significant value, as the resulting analysis could be potentially used for making decisions regarding escalation or de-escalation of care and monitoring of treatment efficacy.

In recent times, Artificial intelligence (AI) algorithms, particularly deep learning, have demonstrated remarkable progress in the automated interpretation of medical images. During high workloads and in locations where the number of radiologists is limited, automated CXR analysis can play a vital role in efficient and effective patient management. Previous research already revealed that well-trained AI algorithms could perform at or above the level of third-year radiology residents when it comes to identifying various characteristics on radiographs [5]. Thus, effectively trained AI-based algorithms for automated CXR analysis tools can significantly benefit low-resource hospital systems.

Automatic assessment of COVID-19 using CXR images has been an active area of research since the inception of this pandemic [6–12]. Most previous work focuses on the detection of COVID-19 pneumonia from CXR images. However, in hospital settings, automatic detection of the presence of COVID-19 is not of significant value as most admitted patients are confirmed through RT-PCR tests. Methods for automatic assessment of COVID-19 severity can be more valuable as they can be used for monitoring treatment efficacy, especially in the ICU. Therefore, in this work, we focus on the assessment of COVID-19 severity using an existing dataset [13]. In Cohen et al. [13], a COVID-19 pneumonia severity prediction dataset was released including two different disease severity metrics, (i) the Geographic Extent Score, and (ii) the lung Opacity Score. The *geographic extent score* represents the extent of lung involvement, whereas *opacity score* signifies the degree of lung opacity.

The main contributions of this paper are as follows:

1. We propose a deep learning-based unified model that integrates anatomy information with the feature extraction scheme to increase the performance of the architecture for COVID-19 severity scoring

2. A regression model that uses the extracted features for COVID-19 severity scoring assesses lung involvement in COVID-19 patients.
3. Experimental evaluation on an available COVID-19 CXR dataset and validation on an in-house annotated CXR dataset proves the efficiency of this method.
4. This model can be incorporated into teleradiology platforms to reduce the number of severe COVID-19 patients by providing early predictive severity scores.

Motivation for Anatomy-Aware Analysis

COVID-19 pneumonia mainly increases lung density which is seen as whiteness in the radiography images depending on the severity of the pneumonia. When hazy gray areas partially obscure dark lung markings in CXR, the observation is referred to as the *ground-glass opacity*. A ground-glass opacity is a hazy increase in attenuation in interstitial and alveolar processes. Linear opacities (Peripheral, coarse, horizontal white lines, bands, or reticular changes) may also be seen along with ground-glass opacity. In severe cases, lung markings are completely lost because of whiteness, which is referred to as consolidation. These changes are more likely to be peripheral and lower zones, but the whole lung can also be affected. Bilateral lung involvement is most commonly observed. The appearance of nodules, pneumothorax, or pleural effusion might be incidental for COVID-19 cases [14]. Since the infected regions are always associated with the lungs, a trained radiologist naturally interprets a COVID-19 CXR image by first outlining the anatomical structure of the lung. Similarly, an anatomy-informed algorithm can analyze the CXR images for a superior analysis and subsequent severity prediction. For example, if the lung regions are pre-identified, the deep learning models can be trained using higher resolution images only of the relevant portion of the images. We refer to this type of method as being *anatomy-aware*.

This paper proposes a novel COVID-19 severity detection method for chest x-ray images using a lung-annotated anatomy-aware deep learning model. The proposed model integrates the generic x-ray image features with pre-identified anatomical information extracted using an auxiliary segmentation model. The proposed model generates both disease-level features and lung involvement percentage to estimate the severity of pneumonia in COVID-19 patients.

Related Work

Previous work on the automated assessment of CXR images for COVID-19 has focused on disease classification, segmentation, and severity prediction. Transfer learning techniques have proven to be effective ways to classify

COVID-19 from CXR images. In [15], ResNet50, InceptionV3, and InceptionResNetV2 architectures are used to classify COVID-19 disease from CXR images, achieving accuracies of about 98.0%, 97.0%, and 87%, respectively. A number of deep learning architectures have been explored to classify COVID-19 disease from CXR images [6–11, 16–22]. Generative Adversarial Network (GAN)-based data augmentation to detect COVID-19 from CXR images has also been studied, and the result surpasses other augmentation methods [23, 24]. A densely attention-based deep network with high accuracy was proposed for automatic COVID-19 recognition [25]. In [26], a majority voting-based classifier ensemble of five benchmark algorithms was shown to provide promising results in detecting COVID-19 from CXR images. In [27], a wavelet-based depthwise convolution network with Grad-CAM visualization is used for the diagnosis of COVID-19. DMFL-Net, a deep learning framework is proposed ensuring latency, energy usage, and privacy of sharing data across hospitals [28]. A reliable COVID-19 detection architecture (ReCovNet) was developed to detect COVID-19 out of 14 different thoracic diseases using CXRs compiling QaTa-COVID-19 dataset of their previous study [29, 30]. DarkCovidNet, a deep learning model was proposed for both multi and binary classification of COVID-19 with an accuracy of 87.02% and 98.08% for multi and binary class, respectively [31]. Several approaches have been explored to reduce redundant input information using different feature extraction approaches [32–39]. In [40], pre-trained BiT models were analyzed, where DenseNet (using an additional Dense layer with 512 perceptrons) performed best with 92% classification efficiency for automatic COVID-19 prediction. However, none of these previous methods consider the analysis of lung anatomical structure for disease localization or classification. On a separate note, although the COVID-19 detection methods using CXR have achieved promising results, the detection method is not suitable for replacing a traditional RT-PCR diagnostic test. Several approaches have been explored to develop an architecture for both CXRs and CT scan images [41, 42]. In [43], a lightweight CNN-tailored deep neural network (DNN) is proposed to detect COVID-19 from both CT scans and CXRs using a small number of images. Transfer learning-based feature extraction with a capsule network has also been used on both CXRs and CT images to improve performance and avoid overfitting [44]. In our experience, radiologists and general physicians are capable of detecting COVID-19 markers from CXR images, so automatic detection of its presence may not provide a significant advantage. However, disease severity analysis may be more valuable in resource-constrained settings.

There has also been a considerable amount of work on COVID-19 diseased region segmentation in recent years. In [45], a deep neural network-based method is used for the detection of COVID-19 symptoms from CXR images, and gradient-guided class activation maps (Grad-CAM++) and layer-wise relevance propagation (LRP) are used to highlight class-discriminating regions. A segmentation-based deep fusion network is used to classify 14 thoracic diseases using a CXR image, where local lung region images are extracted with discriminative features for improved performance [46]. Anatomy-XNet, an anatomy-aware attention-based network incorporating the spatial features guided by the pre-identified anatomy regions proves the efficacy of utilizing the segmentation knowledge to improve the classification of 14 thoracic diseases [47]. Unsupervised clustering techniques are also used to localize the suspected abnormal regions in the lung field of the CXR images [48]. Additionally, integrating an attention gate (AG) into a standard convolutional neural network (CNN) increases the precision and predictive accuracy of automatic segmentation of COVID-19 regions in CT images using minimum computation [49].

Previous work on COVID-19 severity prediction from CXR images has mainly used variants of deep learning architectures for transfer learning [13, 50–53]. However, these methods also do not explicitly use anatomical information to interpret similar to a radiologist.

Methodology

This work proposes a unified framework in which an anatomy-aware (AA) deep learning model performs feature extraction while a regression model performs severity prediction. The overview of the entire architecture is illustrated in Fig. 1.

The proposed AA model consists of a pre-processing model and a backbone model. The pre-processing model, named the *Anatomy Annotation* (ANT) model, infuses the anatomy structure information of CXRs, and the backbone model is used for dense pooling. Then, the feature map is passed through a gating system consisting of two paths. One path utilizes the Feature Pyramid Attention (FPA) module [54, 55] for improved pixel-level attention and extraction of features. This feature map is then pooled to obtain the disease-level features (feat 1). The other path uses Probabilistic Class Activation Map (P-CAM) pooling [55], which is known to have excellent localization ability. We generate disease-wise heatmaps with P-CAM pooling and calculate the lung involvement of each disease in decimal form (feat 2). The disease-level features and the percentage of lungs covered by the heatmaps are used as feature vectors and concatenated to generate the final feature vector. Finally, a linear regression model is applied to this feature vector to obtain the COVID-19 severity scores.
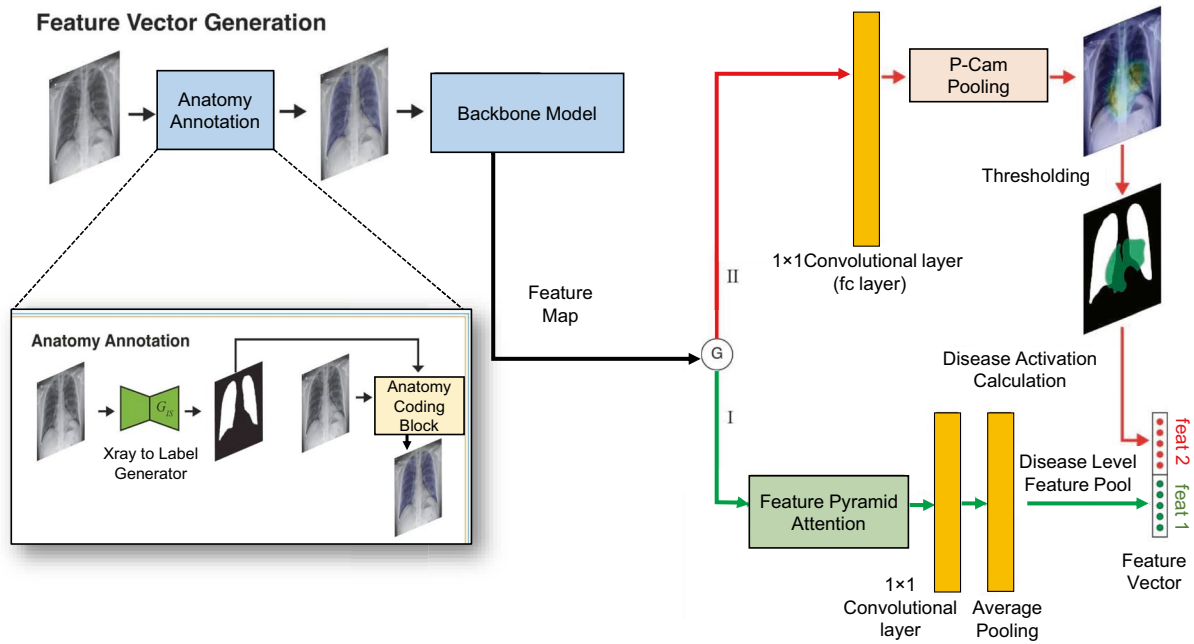
Fig. 1 Overview of the proposed Anatomically Aware Network used for COVID-19 severity prediction. We use a Cycle-GAN trained image to the label generator for lung segmentation. The lung information is then incorporated in the original radiographs through the anatomy coding block which performs Eq. 2. This is followed by the Densenet-121 feature extraction and the gating function. In path I, we use Feature Pyramid Attention for improved pixel-level attention and in consequence, generating five diseases (Atelectasis, Consolidation, Edema, Pleural Effusion, and consolidation) level features. On the other hand, path II is used for class-wise heatmap generation with PCAM pooling. The lung involvement score is calculated using the generated heatmaps and lung masks. Finally, disease-level features and lung involvement scores are used for regression

Anatomy Annotation Block

To incorporate the anatomical features in the AA model it is important to emphasize on different anatomical structures, e.g., lungs, heart, ribs, clavicles, diaphragm, etc. In this work, we present a novel pre-processing method termed as *Anatomy Informed Annotation*.

To perform lung segmentation, we employ a Cycle-GAN-based semi-supervised method that has shown superior performance compared to current methods for this task [56, 57]. The entire architecture is illustrated in Fig. 2. The first generator (G_{IS}), which corresponds to the segmentation network that we want to obtain, learns a mapping from an image to its segmentation labels. The first discriminator (D_S) attempts to differentiate these generated labels from real segmentation masks. Conversely, the second generator (G_{SI}) learns to map a segmentation mask to its original x-ray image. In our semi-supervised segmentation setting, this generator is only used to improve the training procedure. Similarly, the second discriminator (D_I) receives an image as input and predicts whether this image is real or generated. To enforce cycle consistency, the generators are trained such that feeding the labels generated by G_{IS} for an image into G_{SI} provides the same image, and passing back to G_{IS} the image generated by G_{SI} for a segmentation mask gives the same mask. We use the SCR dataset segmentation masks and 200 synthetic segmentation masks obtained from the Stanford Chexpert dataset (generated by a pre-trained U-Net [58]) as labeled chest X-rays. In this way, we generate segmentation masks for the entire Chexpert dataset in a semi-supervised way, using the anatomy-informed annotation.

Let **x** be a chest X-ray image and **y** be the generated segmentation mask. The output of the neural network, which performs anatomy-informed segmentation, can be expressed as a function:

$$\mathbf{y}_{i,j} = \prod_{i,j} \mathbf{f}_{ks}((\mathbf{x}_{(s)i+(\delta)i,\,(s)j+(\delta)j})_{0\,\leq\,(\delta)i,\,(\delta)j\,\leq\,k}) \qquad (1)$$

where k is the kernel size, s is the stride or subsampling factor, and $\mathbf{f}_{ks}$ is the layers of the neural network which is determined by the layer type. The output **y** is used for merging anatomy information in the original CXR image, **x**. We

Fig. 2 The proposed Cycle-GAN-based architecture for semi-supervised lung segmentation. G_{IS} and G_{SI} are the two generators, where D_I and D_S work as discriminators. These four networks are trained simultaneously
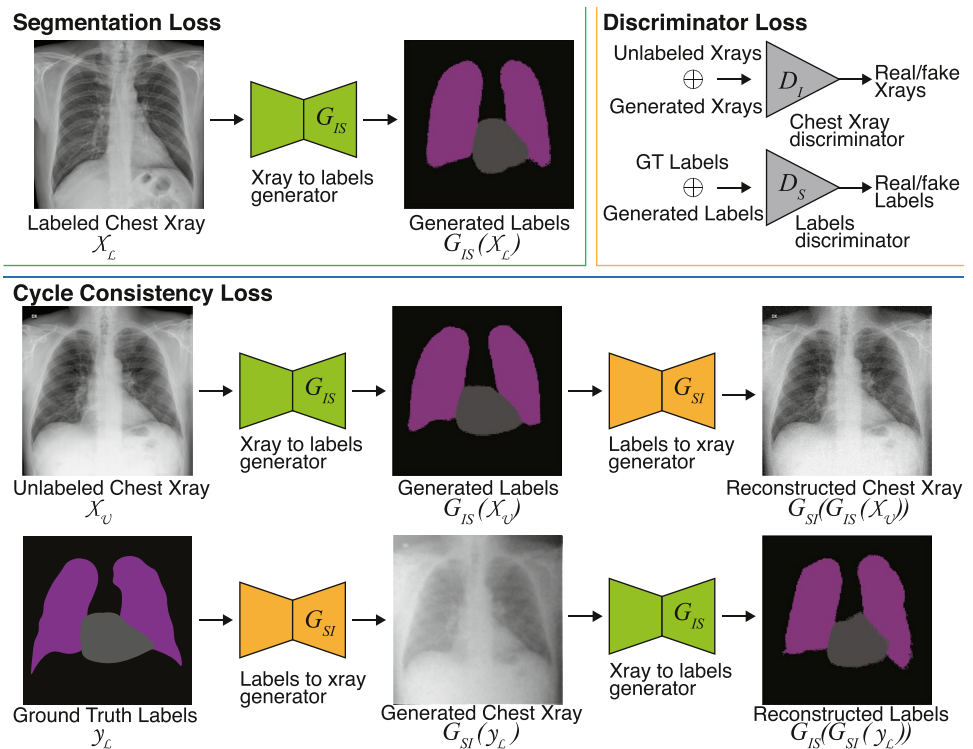
can denote the RGB vector image as three separate column vectors, like, $\mathbf{x} = [\mathbf{x}^0\ \mathbf{x}^1\ \mathbf{x}^2]$. These column vectors can be denoted as three RGB channel column matrices. We can write as, $[\mathbf{x}^0\ \mathbf{x}^1\ \mathbf{x}^2] = [\mathbf{X}^R\ \mathbf{X}^G\ \mathbf{X}^B]$. Then this data vector is infiltrated by a gray-scale segmentation mask, $\mathbf{Y} = [\mathbf{y}_{ij}]$, which is composed of column vectors generated by the neural network shown in Eq. 1.

$$\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{I} \cdot \mathbf{P}^T \cdot \mathbf{Y} \qquad (2)$$

We can find the anatomy-informed image using Eq. 2, where $\mathbf{I}$ is the identity matrix, $\mathbf{P}$ is a hyper-parameter matrix for controlling the measures of infiltration in a specific RGB channel. In order to keep the actual information of the chest radiographs as much as possible, only the Blue channel has been infiltrated with the anatomical information in a small amount, so the hyper-parameter matrix, $\mathbf{P}$ can be written as $\mathbf{P} = [1\ 1\ p]$, $0 < p < 1$. This anatomy-informed image, $\tilde{\mathbf{x}} = [\tilde{\mathbf{X}}^R\ \tilde{\mathbf{X}}^G\ \tilde{\mathbf{X}}^B]$, is then fed into the backbone model.

Backbone Model

The backbone model is a traditional deep learning classification network used in transfer learning. In this case, we use Densenet-121 [59] as its dense block has been well-known for its *feature-reuse* capability during feature extraction. Training this model using the Anatomy-annotated images, $\tilde{\mathbf{x}}$, ensures that the model is aware of the chest radiograph anatomy.

Feature Pyramid Attention

To produce improved pixel-level attention, we incorporate a Feature Pyramid Attention (FPA) module [54] into the system. The Pyramid Attention module first fuses features from two different branches: three $n \times n$ pyramid scales convolution ($n = 3,5,7$) and origin features from CNNs going through a 1×1 convolution. Then, a global average pooling branch feature is added with the output features to select the discriminative multi-resolution feature representation. When these features overlap significantly, final pooling is applied to extract the disease-level features (pre-softmax or pre-sigmoid output).

P-CAM Pooling

We use P-CAM pooling for improved heatmap generation. P-CAM pooling explicitly leverages the excellent localization ability of CAM [60] during training in a probabilistic fashion. The backbone network first processes the input CXR image and generates a feature map. Then, for a particular disease label, such as "Consolidation," each feature embedding within the feature map goes through a fully connected (FC) layer implemented as a 1×1 convolutional layer and generates the class activation score that monotonically measures the disease likelihood of each embedding. It is further bounded with the sigmoid function and interpreted as the disease probability of each embedding. Finally, the output probability map is normalized to attention weights of each embedding, following the

multiple-instance learning (MIL) framework, which is used to pool the original feature map by weighted average pooling. The pooled embedding goes through the same FC layer introduced above and generates the image-level disease probability for training. During inference time, the probability map is directly used for localization. Then, we apply simple hyperparameter thresholding to obtain disease regions. Finally, for the disease d, we define its activation score as follows:

$$A_d = \frac{\sum_{0,0}^{h,w}[L_{i,j} \bigcap R_{i,j}^d]}{\sum_{0,0}^{h,w} L_{i,j}} \quad (3)$$

where $L_{i,j}$ is lung segment (right and left both stacked), $R_{i,j}^d$ is the d region of the disease created from heatmap. The summation is across the segment and region's height and width. According to Eq. 3, disease activation will be between 0 to 1.

Data Sources

We use four different datasets for the training and evaluation of the proposed COVID-19 severity detection framework. These include chest x-ray image datasets including anatomy segmentation masks and disease labels. The datasets are as follows:

JSRT Dataset

The JSRT dataset [61] consists of 247 images (154 nodule and 93 non-nodule images), with a resolution of 2048×2048. This dataset also includes patients information such as: age, gender, diagnosis (malignant or benign), X and Y coordinates of nodule, simple diagram of nodule location.

SCR Dataset

SCR dataset [62] is a database of posterior-anterior chest radiographs where the manual segmentation of lungs, heart, and clavicles is provided. This dataset includes chest radiographs of 247 subjects where annotations of the anatomical structures of the images of the JSRT database, e.g., left lung, right lung, heart, left clavicle, and right clavicle, are included.

Stanford Chexpert Dataset

The Chexpert dataset [63] is a large public dataset for chest x-ray interpretation. This dataset contains 224, 316 radiographic images of 65, 240 patients labeled with 14 observations such as the following: Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion, Pneumonia, Pneumothorax, Enlarged Cardiom., Lung Lesion, Lung Opacity, Pleural Other, Fracture, Support Devices, No Finding.

COVID-19 Pneumonia Severity

The COVID-19 Pneumonia Severity Dataset [13] is a small dataset with 94 images, where each corresponding image has two severity scores: the geographic extent score and the opacity score. The geographic extent score denotes the extent of lung involvement by ground-glass opacity or consolidation for each lung. Here, 0 = no involvement, 1 = <25% involvement, 2 = 25–50% involvement, 3 = 50–75% involvement, and 4 = >75% involvement. The total extent score ranges from 0 to 8, including both lungs. On the other hand, the opacity score signifies the degree of opacity for each lung. Here, 0 = no opacity, 1 = ground-glass opacity, 2 = consolidation, and 3 = white-out. This score ranges from 0 to 6, accumulating both lungs. The labeling of the geographical extent and opacity scores is shown in Fig. 3.

Dataset Organization

In this work, we only use lung annotations for our experiments. First, a U-Net model is trained on the JSRT dataset using SCR labels for lung segmentation. This U-Net is then applied for initially generating 200 lung masks from the Chexpert dataset. Due to dataset mismatch, the U-Net performance on CheXpert is inferior compared to the JSRT data. However, these particular 200 images and their corresponding lung annotations are selected through manual examination to provide an adequate lung mask. These 200 synthetically labeled images together with SCR clinical ground truth annotations are later used for the lung annotation of the entire Chexpert dataset using the Cycle-GAN-based semi-supervised segmentation approached discussed in the "Anatomy Annotation Block" section. Finally, these lung segments are used for appending anatomical structural information based on Eq. (2) on the
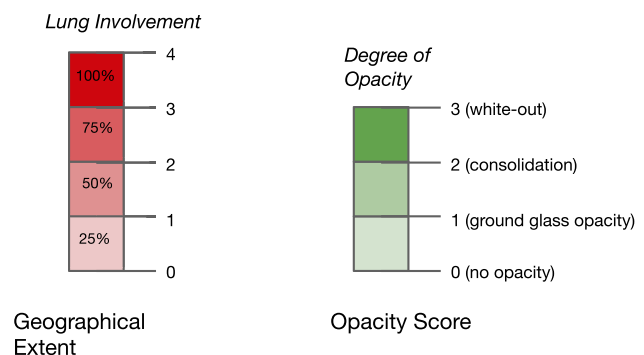

Fig. 3 Labeling of geographical extent and opacity score

Table 1 Datasets required for the training and evaluation of the proposed framework

Dataset	Number of CXRs
JSRT Dataset [61]	247 CXRs with patients information
SCR Dataset [62]	247 CXRs with annotated anatomical information
Stanford Chexpert Dataset [63]	224,316 CXRs with 14 observations
COVID-19 Pneumonia Severity Dataset [13]	94 images with geographic extent score and opacity score
In-house Dataset	12 independent CXRs annotated by experienced radiologist

original CheXpert images. These anatomy-annotated x-ray images are then used for training the Densnet121-FPA and Densnet121-PCAM models. The datasets utilized in this study are summarized in Table 1.

To evaluate the performance of our framework, we use the COVID-19 Pneumonia severity dataset [13]. We perform prediction tasks on a held-out test set using a 5-fold cross-validation scheme. An overall flow diagram describing the data organization and experimental design of this paper is illustrated in Fig. 4.

Experimental Evaluation

Training Phase

In the training phase, we train a U-net model on the SCR dataset. We infer this U-net model on the 200 selected Stanford Chexpert images and generate lung

segmentation masks. These 200 synthetic labels together with SCR ground truth labels are used to train the Cycle-GAN model. Next, we generate the lung segments for the entire Chexpert dataset and incorporate them within the x-ray images using Eq. (2). Afterward, we train our feature vector generation module using these lung-annotated radiographs. The Densenet121-FPA/PCAM model is trained using the following training protocols. Densenet initialization is done by ImageNet weights [64], the batch size is set to 24, images are resized to 256×256, and the binary cross-entropy (BCE) loss is used. The training is done for 5 epochs on 3 RTX 2080*ti* graphics processing units (GPUs). We use this model for feature vector extraction.

Prediction Phase

In the prediction stage, we use two different types of features for the regression analysis. The first type consists of
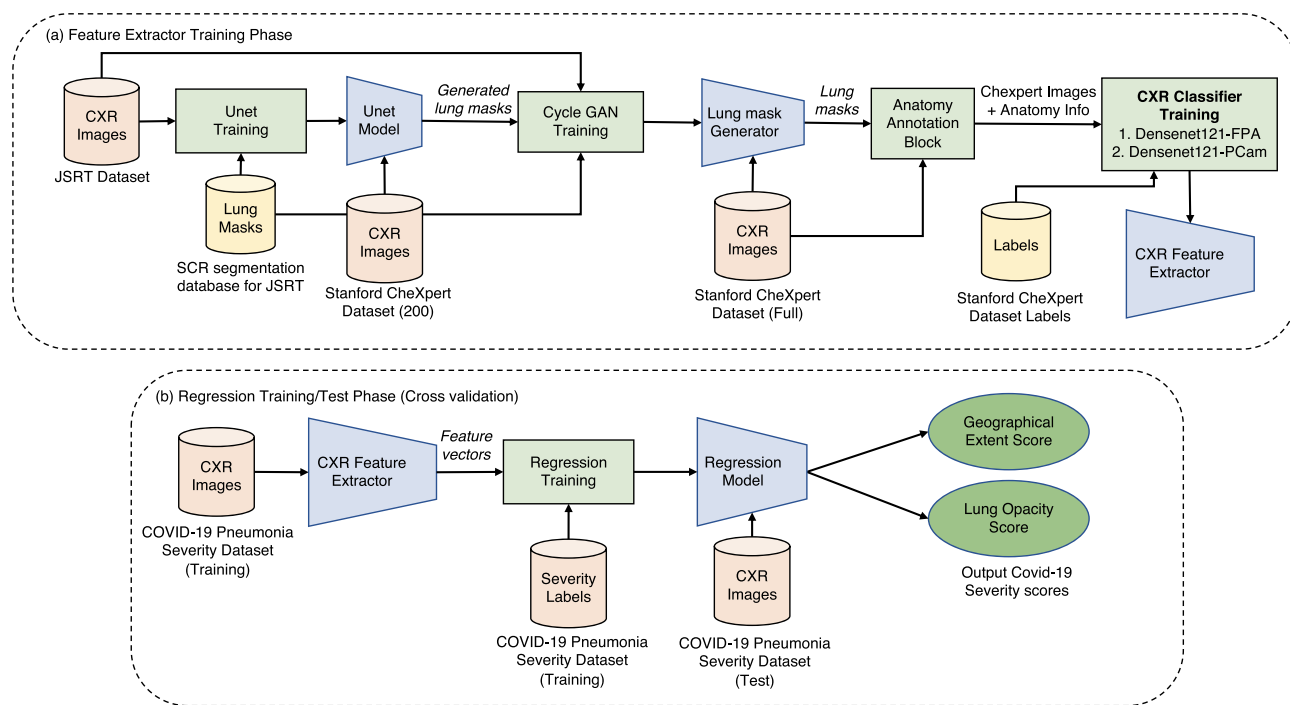
Fig. 4 Dataset organization and overall flow diagram in the feature extractor training and regression training/testing phases

Table 2 Performance evaluation of the proposed COVID-19 severity prediction architecture compared to existing models. In the table: 3 diseases include: Atelectasis, Edema and Consolidation related features. In 4 diseases features Pleural Effusion related features are added. Also results have been generated with the individual disease and the best one is shown as single class (like in case of Densenet-FPA with AA, Edema has the best correlation with the geographic extent. These results also compared with the baseline scores from Cohen et al. [13])

| Task	Method	Features	MSE	MAE	R^2
Geographical Extent Score	Densenet-FPA without AA	3 diseases	2.25 ± 0.62	1.22 ± 0.12	0.59 ± 0.10
		4 diseases	1.93 ± 0.63	1.16 ± 0.15	0.63 ± 0.11
		single disease	3.16 ± 0.72	1.45 ± 0.18	0.41 ± 0.04
	Densenet-FPA with AA	3 diseases	1.90 ± 0.45	1.14 ± 0.12	0.64 ± 0.05
		4 diseases	1.85 ± 0.29	1.21 ± 0.16	0.63 ± 0.10
		single disease	3.45 ± 1.56	1.34 ± 0.20	0.38 ± 0.15
	Densenet-FPA with AA and Disease Activation features	3 diseases	1.87 ± 0.51	1.15 ± 0.09	0.63 ± 0.10
		4 diseases	1.90 ± 0.39	1.12 ± 0.11	0.63 ± 0.05
		single disease	3.52 ± 1.24	1.34 ± 0.20	0.35 ± 0.10
	Cohen et al. [13]	lung opacity	2.06 ± 0.34	1.14 ± 0.10	0.60 ± 0.09
Opacity Score	Densenet-FPA without AA	3 diseases	1.20 ± 0.14	0.76 ± 0.04	0.57 ± 0.09
		4 diseases	1.08 ± 0.22	0.80 ± 0.10	0.61 ± 0.07
		single disease	1.37 ± 0.47	0.94 ± 0.16	0.39 ± 0.10
	Densenet-FPA with AA	3 diseases	0.94 ± 0.20	0.82 ± 0.04	0.55 ± 0.09
		4 diseases	0.97 ± 0.23	0.85 ± 0.11	0.56 ± 0.11
		single disease	1.54 ± 0.47	1.10 ± 0.12	0.32 ± 0.08
	Densenet-FPA with AA and Disease Activation features	3 diseases	0.94 ± 0.21	0.81 ± 0.05	0.46 ± 0.20
		4 diseases	0.96 ± 0.24	1.15 ± 0.31	0.45 ± 0.20
		single disease	1.45 ± 0.50	1.52 ± 0.21	0.20 ± 0.36
	Cohen et al. [13]	lung opacity	0.86 ± 0.11	0.78 ± 0.05	0.58 ± 0.09
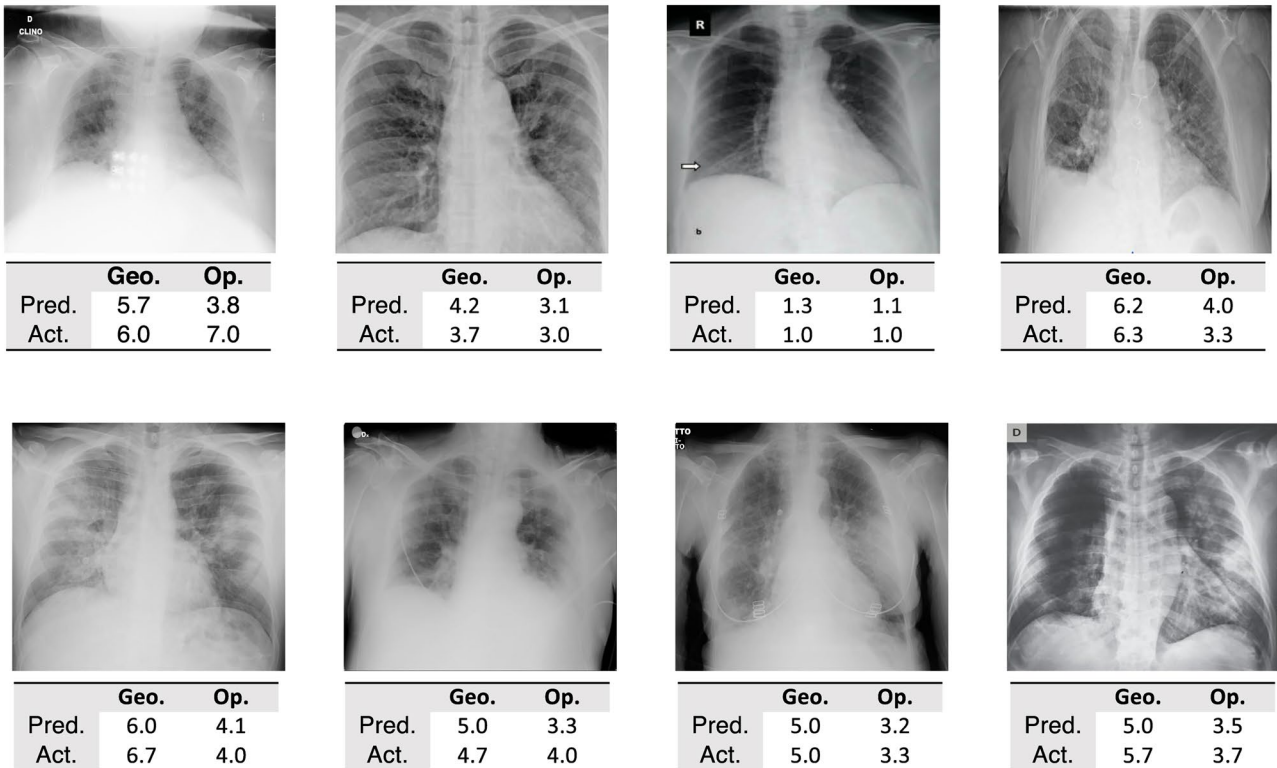
Fig. 5 Comparison of predicted (Pred.) and ground truth/actual (Act.) scores of selected x-ray images for geographical extent score (Geo.) and lung opacity score (Op.)

the outputs from the Densenet121-FPA classifier pipeline considering specific lung-related diseases. These diseases include Edema, Consolidation, Atelectasis, and Pleural Effusion. The second type of feature includes the activation values calculated from the disease activation as defined in Eq. (3) considering the same lung-related diseases. In total, we have a maximum of 8 feature parameters that are available for disease severity prediction via regression. During our experiments, we have selected different combinations of these features for linear regression. We report the mean squared error (MSE), mean absolute error (MAE), and correlation coefficient as performance metrics comparing the predicted COVID-19 severity and the ground truth for the geographic extent and opacity scores. The results are summarized in Table 2.

Overall Results

Our baseline model using Densenet121-FPA without the proposed anatomy-aware (AA) block provides very similar results as reported in [13]. In terms of geographical
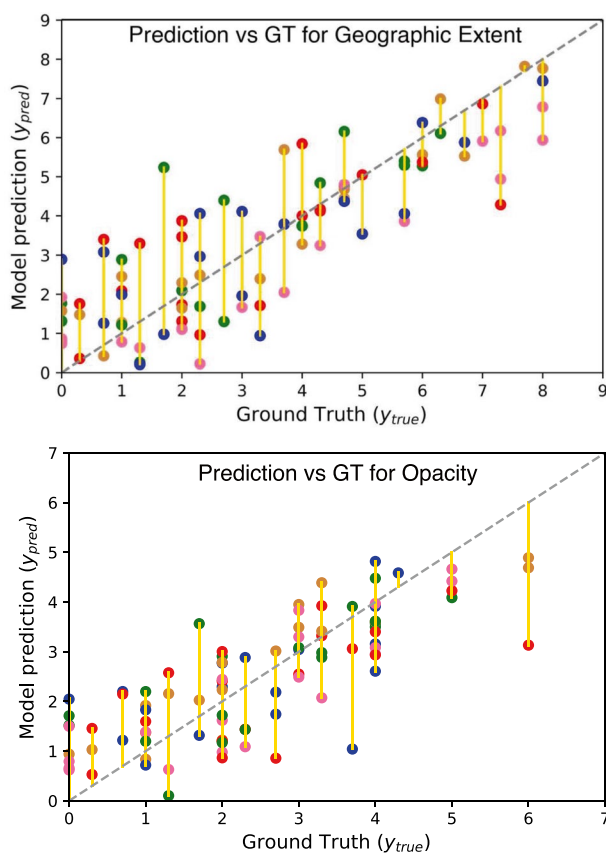
Fig. 6 Model's prediction comparison with ground truth (GT). Here different colored Cartesian coordinates represent different folds in 5-fold cross-validation settings

extent score, Densenet121-FPA and [13] provide MSE scores of 1.93 ± 0.63 and 2.06 ± 0.34, respectively (p-value $= 0.001 < 0.01$). In the case of lung opacity score, Densenet121-FPA and [13] provide MSE scores of 1.08 ± 0.22 and 0.86 ± 0.11 (p-value $= 0.0 < 0.01$), respectively. Thus, we can conclude that the baseline model consisting of the Densenet121-FPA pipeline is equivalent to [13] in terms of performance and statistical significance. Table 2 also shows that the overall performance improves after including the anatomy-aware block.

The best-performing model is Densenet121-FPA, with the AA block included in geographic extent and opacity scores. In terms of geographical extent score, the MSE improves from 1.93 ± 0.63 to 1.85 ± 0.29 after including the AA block over the baseline architecture. On the other hand, for lung opacity score, the MSE improves from 1.08 ± 0.22 to 0.97 ± 0.23 after including the AA block along with the baseline model. However, Table 2 also shows that including the disease activation score does not provide the best result for the prediction of lung opacity score. Overall, analyzing the results of Table 2 implies that including additional anatomical information to the competitive baseline model further increases the ability of the system for disease severity prediction. In Fig. 5, we show a few example images along with predicted and ground truth COVID-19 severity scores. Figure 6 shows the comparison between our model's performance with human annotation in a 5-fold cross-validation setup for geographic extent and lung opacity scores, respectively. The plots again indicate a high correlation between the ground truth disease severity and the predicted value.

Evaluation on In-house Annotated Data

In order to validate the performance of the proposed method on an independent test set, we selected a small set of chest x-ray images and performed annotation by an experienced radiologist. In this study, 12 chest x-ray images were selected from a separate COVID-19 dataset. The same severity scoring method used in [13] was followed during the in-house annotation by the experienced radiologist. The geographic extent and lung opacity scores provided by our radiologist and the proposed anatomy-aware model are summarized in Table 3 and Fig. 7. We found significant improvement in the geographic extent score and opacity score (p-value < 0.01)

Table 3 Evaluation of the proposed COVID-19 severity prediction model on Selected Chest X-ray images annotated by an In-house experienced radiologist

Scoring Method	MAE	MSE
Geographical Extent Score	1.55±0.98	3.35±3.51
Opacity Score	0.62±0.48	0.59±0.89
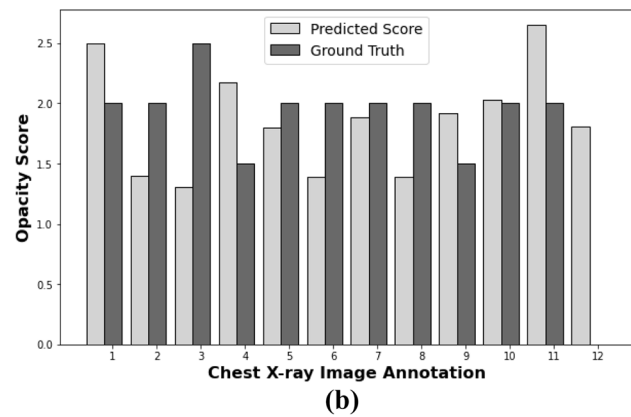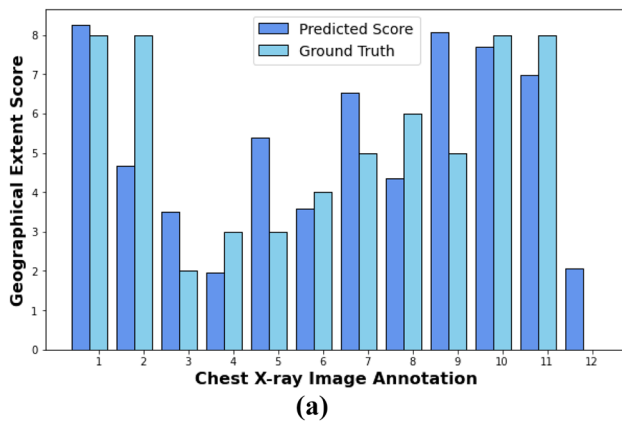
Fig. 7 a Predicted value and ground truth of geographical extent score of selected chest X-ray images annotated by an In-house experienced radiologist, and **b** Predicted value and ground truth of opacity score of selected chest X-ray images annotated by an In-house experienced radiologist

compared with [13] using the selected CXRs. As evident from the results, our model shows low MAE values for both geographic extent and lung opacity scores. However, the MSE is found to be lower for lung opacity score compared to that of the geographic extent scores. These results further demonstrate the effectiveness of the proposed method. However, we acknowledge that this independent validation set is small in size. Further evaluation on a larger independent dataset would provide more convincing evidence of the effectiveness of the proposed method.

Discussion

Most of the previous research on chest x-ray image analysis involves traditional deep learning architectures typically used for image classification. There are few studies that are involved with COVID-19 severity prediction. Basically, for severity prediction, these studies are involved with pretrained models without anatomical information [13]. Additionally, there is a need for a case-by-case comparison to find out the possible causes of error to improve the image analysis architecture [53]. Also, there is a lack of CXRs for performing the segmentation tasks, which sometimes makes it difficult, and there may be anomalies in the findings. However, the existing models are not specifically aware of the anatomical structure present in typical chest x-ray images. In contrast, an experienced radiologist always first identifies the thoracic organs before looking for markers of the disease. Therefore, we hypothesized that adding anatomical information to existing models can help to improve the model's performance in predicting disease severity. Also, anatomical information can reduce the computational complexity by introducing disease-specific features, which

helps the model to learn precisely. The experimental results presented in this work demonstrate that including simple anatomical information, e.g., coloring the x-ray images by lung segmentation masks, can improve the performance of the COVID-19 severity prediction model. Our baseline architecture performs similarly compared to existing models on the COVID-19 pneumonia severity prediction task. We have also included a case-by-case comparison study with the annotated CXRs to quantify the performance of the proposed algorithm. However, the proposed anatomy-informed method yields about 4.1% and 11% relative improvement in MSE for the geographic extent and lung opacity scores, respectively. For the future perspective of this study, there is a need for further clinical validation on a larger-scale independent dataset and validation in a clinical workflow for effective deployment in low-resource hospital settings.

Conclusion

This paper has proposed a novel anatomy-aware deep learning framework for COVID-19 disease severity prediction from chest X-ray images. While traditional methods generally do not specifically consider anatomical information for medical image analysis, expert radiologists tend to always consider their human anatomy knowledge before making a diagnostic decision. In this work, we have utilized a semi-supervised model for automatically generating lung segmentation masks that are subsequently fused within the chest X-ray images. Our best-performing model has provided a relative improvement of 11% in MSE compared to existing methods when evaluated on a COVID-19 pneumonia severity prediction dataset. Experimental comparisons between systems with and without anatomy

information integrated clearly show the effectiveness of the proposed method. Our model also shows promising results on an unseen in-house clinical evaluation dataset that an experienced radiologist has annotated. The experimental evaluations demonstrate the effectiveness of the proposed anatomy-aware architecture for COVID-19 disease severity prediction. Though the COVID-19 transmission rate has significantly surged down, the proposed approach can be incorporated into other AI-based platforms for the severity prediction of other respiratory diseases. In different low-income settings and underprivileged areas, people with comorbid conditions are at risk of severe infections and there is a lack of medical facilities. Chest radiography is among the major radiological diagnostic methods in these regions. Hence, there is a need for computer-aided diagnostic methods to provide healthcare facilities in underserved communities where the proposed method can provide assistance to healthcare professionals. Thus, the extensions of the proposed method can be used for COVID-19 or other respiratory disease severity prediction in low-resource settings.

Author Contribution Nusrat Binta Nizam: literature review, software/coding, experiments, writing manuscript, editing manuscript, preparing response to review document. Sadi Mohammad Siddiquee: conceptualization, methodology development, software/coding, experiments, writing first draft. Mahbuba Shirin: data collection and annotation. Mohammed Imamul Hassan Bhuiyan: conceptualization, supervision, writing — reviewing and editing. Taufiq Hasan: conceptualization, supervision, writing — reviewing and editing.

Declarations

Ethics Approval No human subjects were involved in this research. Existing open-access medical image datasets were used for this work.

Consent to Participate Not applicable, since no human subjects were involved in this research.

Consent for Publication Not applicable, since no human subjects were involved in this research.

Competing Interests The authors declare no competing interests.

References

1. W.H.O. (2023) WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, https://covid19.who.int/, (Accessed on 03/16/2021)

2. Irkham I, A. U. Ibrahim, C. W. Nwekwo, F. Al-Turjman, Y. W. Hartati (2022) Current technologies for detection of COVID-19: Biosensors, artificial intelligence and internet of medical things (IOMT). Sensors 23(1):426.

3. M. Francone, F. Iafrate, G. M. Masci, S. Coco, F. Cilia, L. Manganaro, V. Panebianco, C. Andreoli, M. C. Colaiacomo, M. A. Zingaropoli, et al., Chest ct score in COVID-19 patients: correlation with disease severity and short-term prognosis, European Radiology 30 (12) (2020) 6808–6817.

4. S. Punitha, F. Al-Turjman, T. Stephan, Genetically optimized computer-aided diagnosis for detection and classification of COVID-19, AI-powered IoT for COVID-19 (2020) 105–122.

5. J. T. Wu, K. C. Wong, Y. Gur, N. Ansari, A. Karargyris, A. Sharma, M. Morris, B. Saboury, H. Ahmad, O. Boyko, et al., Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents, JAMA network open 3 (10) (2020) e2022779–e2022779.

6. E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, M. Z. Parvez, Corodet: A deep learning based classification for COVID-19 detection using chest x-ray images, Chaos, Solitons & Fractals 142 (2021) 110495.

7. R. A. Al-Falluji, Z. D. Katheeth, B. Alathari, Automatic detection of COVID-19 using chest x-ray images and modified resnet18-based convolution neural networks, Comput. Mater. Contin. (2021) 1301–1313.

8. A. Abbas, M. M. Abdelsamea, M. M. Gaber, Classification of COVID-19 in chest x-ray images using detrac deep convolutional neural network, Appl. Intell. 51 (2) (2021) 854–864.

9. A. Afifi, N. E. Hafsa, M. A. Ali, A. Alhumam, S. Alsalman, An ensemble of global and local-attention based convolutional neural networks for COVID-19 diagnosis on chest x-ray images, Symmetry 13 (1) (2021) 113.

10. M. Umer, I. Ashraf, S. Ullah, A. Mehmood, G. S. Choi, Covinet: a convolutional neural network approach for predicting COVID-19 from chest x-ray images, J. Ambient Intell. Humaniz. Comput. (2021) 1–13.

11. G. Wang, X. Liu, J. Shen, C. Wang, Z. Li, L. Ye, X. Wu, T. Chen, K. Wang, X. Zhang, et al., A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest x-ray images, Nature Biomedical Engineering (2021) 1–13.

12. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proc. IEEE CVPR, 2017, pp. 2097–2106.

13. J. P. Cohen, L. Dao, K. Roth, P. Morrison, Y. Bengio, A. F. Abbasi, B. Shen, H. K. Mahsa, M. Ghassemi, H. Li, et al., Predicting COVID-19 pneumonia severity on chest x-ray with deep learning, Cureus 12 (7) (2020).

14. H. J. Koo, S. Lim, J. Choe, S.-H. Choi, H. Sung, K.-H. Do, Radiographic and ct features of viral pneumonia, Radiographics 38 (3) (2018) 719–739.

15. A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks, arXiv preprint arXiv:2003.10849 (2020).

16. A. M. Ismael, A. Şengür, Deep learning approaches for COVID-19 detection based on chest x-ray images, Expert Systems with Applications 164 (2021) 114054.

17. A. U. Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, P. S. Yakoi, Pneumonia classification using deep learning from chest x-ray images during COVID-19, Cognitive Computation (2021) 1–13.

18. A. S. Al-Waisy, S. Al-Fahdawi, M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. S. Maashi, M. Arif, B. Garcia-Zapirain, COVID-chexnet: hybrid deep learning framework for identifying COVID-19 virus in chest x-rays images, Soft computing 27 (5) (2023) 2657–2672.

19. A. Hamza, M. A. Khan, M. Alhaisoni, A. Al Hejaili, K. A. Shaban, S. Alsubai, A. Alasiry, M. Marzougui, D2bof-covidnet: A framework of deep bayesian optimization and fusion-assisted optimal

deep features for COVID-19 classification using chest x-ray and MRI scans, Diagnostics 13 (1) (2023) 101.

20. Y. Kaya, E. Gürsoy, A mobilenet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection, Soft Computing (2023) 1–15.

21. K. Jyoti, S. Sushma, S. Yadav, P. Kumar, R. B. Pachori, S. Mukherjee, Automatic diagnosis of COVID-19 with mca-inspired tqwt-based classification of chest x-ray images, Computers in Biology and Medicine 152 (2023) 106331.

22. M. M. Islam, M. Z. Islam, A. Asraf, M. S. Al-Rakhami, W. Ding, A. H. Sodhro, Diagnosis of COVID-19 from x-rays using combined CNN-RNN architecture with transfer learning, BenchCouncil Transactions on Benchmarks, Standards and Evaluations (2023) 100088.

23. S. Motamed, P. Rogalla, F. Khalvati, Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and COVID-19 in chest x-ray images, arXiv preprint arXiv:2006.03622 (2021).

24. T. Zebin, S. Rezvy, Covid-19 detection and disease progression visualization: Deep learning on chest x-rays for classification and coarse localization, Appl. Intell. 51 (2) (2021) 1010–1021.

25. Z. Ullah, M. Usman, S. Latif, J. Gwak, Densely attention mechanism based network for COVID-19 detection in chest x-rays, Scientific Reports 13 (1) (2023) 261.

26. T. B. Chandra, K. Verma, B. K. Singh, D. Jain, S. S. Netam, Coronavirus disease (COVID-19) detection in chest x-ray images using majority voting based classifier ensemble, Expert Syst. Appl. 165 (2021) 113909.

27. K. K. Singh, A. Singh, Diagnosis of COVID-19 from chest x-ray images using wavelets-based depthwise convolution network, Big Data Mining and Analytics 4 (2) (2021) 84–93.

28. H. Malik, A. Naeem, R. A. Naqvi, W.-K. Loh, DMFL_net: A federated learning-based framework for the classification of COVID-19 from multiple chest diseases using x-rays, Sensors 23 (2) (2023) 743.

29. A. Degerli, M. Ahishali, S. Kiranyaz, M. E. Chowdhury, M. Gabbouj, Reliable COVID-19 detection using chest x-ray images, in: 2021 IEEE International Conference on Image Processing (ICIP), IEEE, 2021, pp. 185–189.

30. A. Haghanifar, M. M. Majdabadi, Y. Choi, S. Deivalakshmi, S. Ko, COVID-CXNET: Detecting COVID-19 in frontal chest x-ray images using deep learning, Multimedia Tools and Applications 81 (21) (2022) 30615–30645.

31. T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, U. R. Acharya, Automated detection of COVID-19 cases using deep neural networks with x-ray images, Computers in biology and medicine 121 (2020) 103792.

32. F. Khozeimeh, D. Sharifrazi, N. H. Izadi, J. H. Joloudari, A. Shoeibi, R. Alizadehsani, J. M. Gorriz, S. Hussain, Z. A. Sani, H. Moosaei, et al., Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients, Scientific Reports 11 (1) (2021) 15343.

33. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need. in advances in neural information processing systems, in: Proc. NIPS, 2007, pp. 5998–6008.

34. K. Li, Z. Wu, K.-C. Peng, J. Ernst, Y. Fu, Tell me where to look: Guided attention inference network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9215–9223.

35. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

36. N. Sarafianos, X. Xu, I. A. Kakadiaris, Deep imbalanced attribute classification using visual attention aggregation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 680–697.

37. M. Jaderberg, K. Simonyan, A. Zisserman, et al., Spatial transformer networks, Advances in neural information processing systems 28 (2015).

38. S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.

39. J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 3146–3154.

40. S. Albahli, N. Ayub, M. Shiraz, Coronavirus disease (COVID-19) detection using x-ray images and enhanced densenet, Applied Soft Computing 110 (2021) 107645.

41. G. K. Yenurkar, S. Mal, Effective detection of COVID-19 outbreak in chest x-rays using fusionnet model, The Imaging Science Journal (2023) 1–21.

42. L. T. Duong, P. T. Nguyen, L. Iovino, M. Flammini, Automatic detection of COVID-19 from chest x-ray and lung computed tomography images using deep neural networks and transfer learning, Applied Soft Computing 132 (2023) 109851.

43. H. Mukherjee, S. Ghosh, A. Dhar, S. M. Obaidullah, K. Santosh, K. Roy, Deep neural network to detect COVID-19: one architecture for both CT scans and chest x-rays, Applied Intelligence 51 (2021) 2777–2789.

44. S. Roy, A. K. Das, Deep-cov: An integrated deep learning model to detect COVID-19 using chest x-ray and CT images, Computational Intelligence (2023).

45. M. Karim, T. Döhmen, D. Rebholz-Schuhmann, S. Decker, M. Cochez, O. Beyan, et al., DeepCOVIDexplainer: Explainable COVID-19 predictions based on chest x-ray images, arXiv preprint arXiv:2004.04582 (2020).

46. H. Liu, L. Wang, Y. Nan, F. Jin, Q. Wang, J. Pu, Sdfn: Segmentation-based deep fusion network for thoracic disease classification in chest x-ray images, Comput. Med. Imaging Graph. 75 (2019) 66–73.

47. U. Kamal, M. Zunaed, N. B. Nizam, T. Hasan, Anatomy-xnet: An anatomy aware convolutional neural network for thoracic disease classification in chest x-rays, IEEE Journal of Biomedical and Health Informatics 26 (11) (2022) 5518–5528.

48. T. B. Chandra, K. Verma, D. Jain, S. S. Netam, Localization of the suspected abnormal region in chest radiograph images, in: Proc. IEEE ICPC2T, IEEE, 2020, pp. 204–209.

49. Ü. Budak, M. Çıbuk, Z. Cömert, A. Şengür, Efficient COVID-19 segmentation from CT slices exploiting semantic segmentation with integrated attention mechanism, Journal of Digital Imaging 34 (2021) 263–272.

50. J. Zhu, B. Shen, A. Abbasi, M. Hoshmand-Kochi, H. Li, T. Q. Duong, Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs, PloS one 15 (7) (2020) e0236621.

51. Z. Jiao, J. W. Choi, K. Halsey, T. M. L. Tran, B. Hsieh, D. Wang, F. Eweje, R. Wang, K. Chang, J. Wu, et al., Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study, The Lancet Digital Health 3 (5) (2021) e286–e294.

52. S. Park, G. Kim, Y. Oh, J. B. Seo, S. M. Lee, J. H. Kim, S. Moon, J.-K. Lim, J. C. Ye, Vision transformer using low-level chest x-ray feature corpus for COVID-19 diagnosis and severity quantification, arXiv preprint arXiv:2104.07235 (2021).

53. A. Signoroni, M. Savardi, S. Benini, N. Adami, R. Leonardi, P. Gibellini, F. Vaccher, M. Ravanelli, A. Borghesi, R. Maroldi, et al., BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset, Med. Image Anal. 71 (2021) 102046.

54. H. Li, P. Xiong, J. An, L. Wang, Pyramid attention network for semantic segmentation. arXiv preprint arXiv:1805.10180 (2018).

55. W. Ye, J. Yao, H. Xue, Y. Li, Weakly supervised lesion localization with probabilistic-cam pooling, arXiv preprint arXiv:2005.14480 (2020).

56. A. K. Mondal, A. Agarwal, J. Dolz, C. Desrosiers, Revisiting cyclegan for semi-supervised segmentation, arXiv preprint arXiv:1908.11569 (2019).

57. J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proc. IEEE CVPR, 2017, pp. 2223–2232.

58. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Proc. MICCAI, Springer, 2015, pp. 234–241.

59. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proc. IEEE CVPR, 2017, pp. 4700–4708.

60. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proc. IEEE CVPR, 2016, pp. 2921–2929.

61. J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, K. Doi, Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules, Am. J. Roentgenol. 174 (1) (2000) 71–74.

62. B. Van Ginneken, M. B. Stegmann, M. Loog, Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database, Med. Image Anal. 10 (1) (2006) 19–40.

63. J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al., Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison, in: Proc. AAAI Conf. on Artificial Intelligence, Vol. 33, 2019, pp. 590–597.

64. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proc. IEEE CVPR, Ieee, 2009, pp. 248–255.