Joseph Edradan
CSC621-01

Quantification Research: Methods and Procedures

YELLOW: What you should read and is probably good information
CYAN: Information/Understanding (definitions and context)
ORANGE: Stuff you should look up or is questionable
GREEN: Might be good information?
RED: Adversity/Problems (VERY IMPORTANT STUFF)
PURPLE: Stuff that Joseph cares about, don't bother

Severity and Consolidation Quantification of COVID-19 From CT Images Using Deep Learning Based on Hybrid Weak Labels

Abstract

"Both the extent and type of pulmonary opacities help assess disease severity. However, manually pixel-level multi-class labelling is time-consuming, subjective, and non-quantitative. In this article, we proposed a hybrid weak label-based deep learning method that utilize both the manually annotated pulmonary opacities from COVID-19 pneumonia and the patient-level disease-type information available from the clinical report. A UNet was firstly trained with semantic labels to segment the total infected region. It was used to initialize another UNet, which was trained to segment the consolidations with patient-level information using the Expectation-Maximization (EM) algorithm. "

Paper

"The proposed deep learning method1 consisted of the following two steps: Step 1: training a semantic segmentation network for infected lung regions based on strong label. Step 2: training a semantic segmentation network for consolidation based on patient-level weak label. In step 1, a 2D UNet [33] was employed to segment the infection regions from the CT images (UNet-1). The training labels were CT images with pixel-level annotation of being infected or not. In step 2, UNet-1 was further finetuned to segment consolidation from the infected regions (UNet-2). A subset of the training images was annotated regarding the existence of consolidation for each patient. The consolidation network was trained in the framework of EM"

E. Severity and Consolidation Quantification

The severity and consolidation quantifications are given as:

$$\text{severity score} = \frac{\text{area of infected region}}{\text{area of lung}}, \tag{14}$$

and

$$\text{consolidation score} = \frac{\text{area of consolidation}}{\text{area of lung}}. \tag{15}$$

"UNet-1 was trained on batches of 16 by the Adam algorithm for 200 epochs in total. The learning rate is 10^{-2}, 10^{-3} and 10^{-4} for epoch 1-50, 50-100, and 100-200. UNet-2 was initialized from UNet-1 and trained with batch size of 16. Adam algorithm was used in the M-step. 50 epochs were trained with learning rate of 0.0005. We implemented both φ_1 and φ_2 as in (12) and (13) as the prior function and tried various hyperparameters b_1, b_2 and k_2. For φ_1, $b_1 = 1$ achieved the best Dice coefficient on the MedSeg dataset, which is equivalent to setting all the pixels in images with consolidation to consolidation. For φ_2, $b_2 = 9$, $k_2 = 0.5$ achieved the best Dice. $b_2 = 9$ is equivalent to -34 HU before the gray value normalization. We also implemented thresholding as the baseline method [25], where pixels larger than -200 HU inside the predicted infected region were considered as consolidation."

"Although UNet-1 performs a standard segmentation task and is trained in a conventional supervised manner, a good performance is still crucial because UNet-1 is the basis for the following consolidation segmentation. UNet-1 also provides the total infected region to calculate the severity score (14), which is clinically important for the estimation of the lung function. According to Fig. 2, the predicted scores were underestimated compared to the label. One of the main causes is that the model tends to miss regions with very mild GGO infections, which are not very different from normal lung tissue. Furthermore, the annotations also tend to dilate from the visible boundaries of GGO. Meanwhile, the model predictions are closer to the boundaries, leading to smaller regions compared to the annotation."

"Currently the thresholding method is considered as a reliable method to separate GGO and consolidation [25]. We also observed a good Dice coefficient using thresholding only compared to the proposed weakly labeled learning. However, further investigation found that thresholding tends to misclassify vessels inside the infected regions to consolidation, as they have higher HU compared to GGO and other lung tissues. Most of the FPs in thresholding results came from these vessels, which are far from real consolidation regions. Although the proposed approach had similar FPR compared to thresholding, these FPs are mostly from difference between the boundaries of the labeled and predicted consolidation.

We believe that deep learning-based quantification can help address the need in patients with worsening respiratory status and moderate or severe infection where chest CT scan is recommended, and often performed [8]. The developed deep learning-based CT segmentation can serve as an important tool to help assess disease severity and progression as well as to predict prognosis. The predicted severity score of COVID-19 pneumonia along with other clinical and laboratory markers such as patient age, comorbidities, and oxygen saturation can help caretaking physicians determine patients in need of intubation or ICU admission. Although there are no known and approved treatment for COVID-19 pneumonia, multiple, ongoing clinical trials involving antiviral agents and antibodies can benefit from the proposed method which helps quantify the disease burden and thus, assess disease response or progression in an objective manner. Consistent scoring facilitated by the developed automatic tool can empower both cross-sectional and longitudinal comparisons, enable us better to understand the populational characteristics and the temporal evolution of the COVID-19 disease.
”

Code

https://github.com/wudufan/lung_seg_em

Notes:

- 12 October 2020
- Look at the equations in the image above (Severity and Consolidation Quantification)
- This paper was mostly about their model

Quantifying prognosis severity of COVID-19 patients from deep learning based analysis of CT chest images

Abstract

“...A very high specificity Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) test is the principal technique in use for diagnosing the COVID-19 patients. Also, CT scans have helped medical professionals in patient severity estimation & progression tracking of COVID-19 virus. In study we present our own extensible COVID-19 viral infection tracking prognosis technique. It uses annotated dataset of CT chest scan slice images created with the help of medical professionals. The annotated dataset contains bounding box coordinates of different features for COVID-19 detection like ground glass opacities, crazy paving pattern, consolidations, lesions etc. We qualitatively identify the severity of the patient for later prognosis stages in our study to assist medical staff for patient prioritization. First we detected COVID-19 positive patients with pre-trained Siamese Neural Network (SNN) which obtained 87.6% accuracy, 87.1% F1-Score & 95.1% AUC scores. These metrics were achieved after removal of 40% quantitatively highly similar images from the COVID-CT dataset. This reduced dataset was further medically annotated with COVID-19 features for bounding box detection. After this we assigned severity scores to detected COVID-19 features and calculated the cumulative severity score for COVID-19 patients. For qualitative patient prioritization with prognosis clinical assistance information, we finally converted this score into a multi-classification problem which obtained 47% weighted-average F1-score.

Paper

“These COVID-19 lung disorder features definitely serve as an indicator for analyzing severity of the patient. With ground glass opacity, a type of opacity that appears in initial stages of contraction of COVID-19 by a patient. Later, it mostly transforms into chunks of consolidation patches or crazy paving patterns even accompanied by lesions causing more serious respiration issues and probable requirement for breathing aid. But, these imaging methods do not have reliable specificity as the features of COVID-19 do overlap with a large number of respiratory diseases like influenza A & B, pneumonia & HSV pneumonia, atypical bacterial pneumonia, interstitial lung disease, pulmonary edema etc. This indicates that diagnosis with reliable certainty may be lacking for COVID-19 from radiological findings only. With this argument we focus our efforts on evaluating the severity of a COVID-19 patient after the diagnosis of the disease is done with RT-PCR.

This paper is aimed at expounding on a severity estimation approach of COVID-19 patients which leverages the feature detection and localization in CT chest scan slice images. Based on the proposed severity metric we can assist medical professionals to prioritize the patients for the required vital resources. Which would in turn help to allocate resources with maximum utilization and cater to the maximum number of patients. We specifically focus on detection and localization of ground glass opacities, crazy paving pattern, reverse halo effect, vascular dilatation, subpleural band, bronchiectasis, traction bronchiectasis, consolidations and lesions. Based on these features we classify the severity of the patient. An earlier clinical study also uses a severity estimation approach [38]. It is done by manually accumulating scores over the different segment regions of lungs in a CT scan image slice instead of directly focusing on detecting COVID-19 features. But, this manual analysis from medical experts consumes vital human resources which is not good for medical facilities under stress while handling large numbers of patients.

We also tackle the issue of availability of quality dataset annotations for both CT chest scan images and provide medically validated annotations for CT chest scan dataset [39]. The annotations provided in the dataset are reviewed by medical professionals for creating good quality annotation dataset. Our data is biased towards few COVID-19 features because of their frequent recurrence with other COVID-19 features in the dataset. Hence, our model performance is also skewed towards predicting better results for those features. Also, we acknowledge the imbalanced & biased nature of data and the negative effects on the predictions caused by it [6]. We try to limit these effects by using few shot learning techniques with hyperparameter optimizations and data augmentation that deliver best results out of smaller datasets…
"

“For calculating patient extremity estimation, we develop a two step deep learning pipeline where the first step is for detecting positive COVID-19 cases with the N-shot learning approach of SNNs. This step filters out negative COVID-19 images and feeds only COVID-19 positive images for our second feature localization and detection model, as shown in Fig. 1. Also, COVID-19 negative images constitute multiple lung diseases except COVID-19 having similar lung disorders and regular lung CT scan slice images that are disease free. First step, in this figure demonstrates the preprocessing pipeline for selecting images with high variance features eliminating quantitatively similar images. Second step, defines our deep learning pipeline to detect COVID-19 positive images and provide prognosis severity analysis accordingly. In this figure, CVi & NCVi represents high variance subset of image datasets Ci & NCi for COVID-19 positive and negative cases respectively. First detection step is also complementary to our annotation effort for designing a better quality dataset where we are specifically focused on labelling false positives and negative slice images from our detection step. Hence, we reorganized our labelling efforts more around the miss-classified images from the first step for creating a better severity analysis model and annotated dataset."

"We utilized the data reduction step to minimize our annotation efforts without compromising much on prediction performances of both models in our deep learning pipeline. This preprocessing step reduces the volume of quantitatively similar data by comparing the images with cosine similarity metric instead of qualitative features like blur, brightness, sharpness etc. We deployed a VGG16 [34] like similarity measurement neural network that measures variance of an image with respect to all dataset images for both COVID-19 positive and negative dataset split respectively. The high variance metric qualitatively serves as a measure for capturing maximum numbers peculiar features of an image with respect to other images in the dataset. These features can be COVID-19 features in COVID-19 positive dataset and some other distinctive features in case of COVID-19 negative images dataset. We were able to maintain our accuracy for our SNN model with minimum training images. As even after removal of 41.35 % highly similar images from the training dataset the performance of the model is not compromised beyond 1.5 %. Because their contribution is limited to model accuracy improvement and is majorly compensated with data augmentation techniques. Also, this variance maximization approach allowed us to capture a significant number of different instances per class for our localization model training.

Our contributions in this paper are summarized in this paragraph. In this research paper we provide a newly annotated dataset that captures features of COVID-19 in CT chest scans image slices. The annotations are presented in Pascal VOC and csv format in two different image resolutions of 300x300 and 500x500 pixels. Second, we introduce image similarity based data reduction methodology that omits unnecessary images that are not valuable to model accuracy improvement. Third, we propose a two step deep learning pipeline that is tuned to work well even on smaller datasets. Both steps are independent of one another and can be utilized individually with first being employed for diagnosis and second one for prognosis. Here, we have utilized the diagnosis and data reduction steps to minimize our labelling effort for the prognosis dataset. But, these efforts can be skipped for calculating severity metric if bounding box annotations are available. The clinical motivation behind this study is to propose a prognosis deep learning pipeline that is robust & extensible even on a small volume of data. This approach would also assist medical professionals to prioritize resources for more severe patients and would give assisting metric data to informatively supplement their decision."

"The base dataset under consideration in this paper is biased in nature with most CT chest scans belonging to mid-30s & mid-60s age group patients and approximately 60% of patients under consideration being biologically male. With no specifically detailed metadata, prior comorbidities information and quality control for given CT scan images slices in the dataset, the results would be tampered and unreliable in nature…"

"For localization and detection of an instance, SSDs are often used considering their simplicity and high frames-per-second (FPS) processing speed [24]. We limit our study to 2D CT chest scan sliced representation of images in this paper. But, also while scanning across 3D CT chest scan volumes in video like formats, these architectures would yield very high performance as well. Hence, it would have been even better model architecture to localize and detect dynamic bounding boxes per frame. SSDs fundamental idea is to use bounding boxes of different scales and aspect ratios per feature map for localization and remove region proposal network steps like Fast RCNN [13]. At prediction time, the network combines predictions from multiple feature maps to handle different size objects and generates a presence score based on intersection over union (IoU) method to create adjustments for matching the box better to instance shape. In our study we utilized the SSD approach for detecting COVID-19 lung disorders in a more granular manner for better medical interpretability."

"The COVID-19 dataset images collected are resized into standard sizes of 300x300 and 500x500 pixels [39]. This dataset is originally created from medRxiv and bioRxiv available preprint images on COVID-19 patients. Metadata information like gender, age, location, offset time period, severity of COVID-19 and radiology is also provided. Whereas negative COVID-19 images constituting of healthy & non-COVID-19 lung disorders CT scan slices are added from sources like MedPix, LUNA, Radiopaedia and PubMed Central. The patient data used from these sources is completely anonymized as per the privacy regulations and compliance procedures. Multiple images of a single patient are available that are visually very similar in nature are also present in this dataset. This dataset is already split into 349 images of 216 COVID-19 positive patients and 397 images of negative COVID-19 samples. Also, the COVID-19 patient dataset is skewed in nature with relatively more cases of male gender and older age groups. Specifically, 62.77% of the COVID-19 positive patients are biologically male in this dataset. And the patient data contains a relatively large number of samples for middle-aged (32 to 51 years) & older age groups (62 to 81 years). The Fig. 3 highlights some sample COVID-19 positive and negative images for a facile review of COVID-19 opacities and lung disorders. This figure demonstrates the similarity of features in both COVID-19 positive and negative samples. Demonstrating that without specific COVID-19 related domain knowledge of Chest CT scan it is a very hard task to classify these images and provide diagnosis information.1"

"We used the training set images from the detection experiment as input images for training the severity estimation SSD networks. From these images COVID-19 features were detected and localized in the first step of the network. We pipelined multiple data augmentation techniques for performance improvements in our SSD models, as shown in Fig. 8. Red color represents the background class whereas blue and green color represents COVID-19 feature classes in the given image demonstrating three sample data augmentation transformations. The predictions for different feature classes and bounding box coordinates on COVID-19 positive images were made by serving batches of these data augmented images as input to our SSD networks. From which effective area of each predicted attribute class was calculated and we selected the top 36 predicted classes excluding the background class. Which was then fed into (3) for severity estimate value calculation. Finally, we divided the derived severity estimates from test images into three classes known as initial, intermediate & severe. From multiclass categories derived on these three severity types, we measure the prognosis capabilities of our best SSD network.

"For severity analysis estimates, we used (3) to determine COVID-19 patient's condition. In our test dataset, we had 30 images of COVID-19 positive patients and an almost equivalent number of negative patients. Our diagnostic model eliminated the COVID-19 negative images containing background class only and the prognosis model didn't provide any bounding box results for those images. Hence, we determined severity estimates only for the 30 COVID-19 positive images under consideration in our test dataset. The results from SSD models are fed into the severity estimation (3) and the final numerical values from this equation were normalized into the range of [0, 1] values only.

The Table 7 provides the complete result report of severity analysis estimates for the COVID-19 patients predicted by our model. The estimated severity numeric value instead of being presented as standard [0, 1] values were transformed into categorical initial, intermediate and severe states based on the equal division of [0, 1] numeric space. The actual patient severity categories were determined manually from the analysis of metadata associated with the test set images. Hence, the problem was transformed into a multi-class classification analysis for determining quantitative model performance. The confusion matrix corresponding to this problem is presented in Fig. 11 and results are summarized in result Table 7. Also, the total accuracy of the prognosis model is 47 % as derived from the confusion matrix."

"The results in Table 7 shows that our model performs best when the feature type in the images are Initial ones in nature. This can be justified with an abundance of annotations being available in the annotation dataset for Vascular Dilation and Ground Glass Opacity classes. These classes have the lowest severity amongst all COVID-19 features. Also, our model is having good specificity value for Severe class as well. This is especially helpful in order to prioritize severe patients more accurately. Also, our test results demonstrate 47% accuracy while determining the exact category of the COVID-19 positive patient severity. And reports even higher specificity of 60% to avoid wrong prognosis analysis.

Also in Fig. 12, we qualitatively observe our model predictions for both correctly and incorrectly predicted images. From the first incorrectly predicted image in the second row, we observe that sometimes the model might consider consolidation patches appearing near to background features like lung walls and might miss-identify them. From Fig. 12, we also observe that our SSD models are very good at capturing vascular dilation and ground glass opacity features as these are the most abundant annotations present in the dataset. Also, we observe that the model does identify the lesions and bronchiectasis present in the images quite correctly even if their annotations are not present in large amounts for this class in particular. This can be attributed to their simpler circle-like shape and almost the same size in every instance."

"Our deep learning pipeline demonstrates that it can quantify severity of COVID-19 patients from annotated CT chest scan slice images which would further assist in the prognosis process. This novel approach instead of using segmentation or regression based modelling techniques for diagnosis adopts a bounding box estimation approach for COVID-19 feature detection and affected area estimation. It provides the prognosis information by evaluating the sum of areas covered by different COVID-19 features for severity estimations. This analysis quantifies the COVID-19 feature affected areas in lungs which would serve as deescalation or escalation assisting factors for admitting patients in intensive care units (ICU) based on the calculated severity score. The analytical information from this system provides valuable information for resource management in medical facilities."

Notes:

- 10 January 2022
- They manually decided what is initial, intermediate and severe
- They used bounding boxes instead of segmentation which is a novel solution, though not exactly what we are using (we are using segmentation).

Severity Quantification and Lesion Localization of COVID-19 on CXR using Vision Transformer

Abstract:

"Under the global pandemic of COVID-19, building an automated framework that quantifies the severity of COVID-19 and localizes the relevant lesion on chest X-ray images has become increasingly important. Although pixel-level lesion severity labels, e.g. lesion segmentation, can be the most excellent target to build a robust model, collecting enough data with such labels is difficult due to time and labor-intensive annotation tasks. Instead, array-based severity labeling that assigns integer scores on six subdivisions of lungs can be an alternative choice enabling the quick labeling. Several groups proposed deep learning algorithms that quantify the severity of COVID-19 using the array-based COVID-19 labels and localize the lesions with explainability maps. To further improve the accuracy and interpretability, here we propose a novel Vision Transformer tailored for both quantification of the severity and clinically applicable localization of the COVID-19 related lesions. Our model is trained in a weakly-supervised manner to generate the full probability maps from weak array-based labels. Furthermore, a novel progressive self-training method enables us to build a model with a small labeled dataset. The quantitative and qualitative analysis on the external testset demonstrates that our method shows comparable performance with radiologists for both tasks with stability in a real-world application."

Paper:

"Although pixel-level segmentation labels have the most abundant information toward this goal, it is hard to collect a large dataset due to its time-consuming annotation. To mitigate this issue, simple array-based severity labeling methods are introduced, where integer-valued severity scores are assigned on the six or eight subdivisions of CXR images [1, 20]. With the array labels, several algorithms [2, 19] quantify the severity of COVID-19 and generate explainability maps using convolutional neural networks and visualization methods such as GradCAM [16], LIME [15], etc. However, the probability values on the explainability maps, usually based on the normalized activation [16], are not directly related to the real probability value of the lesion existence. Accordingly, comparisons of the saliency maps with the true lesion annotation from the radiologists are rarely made.

To provide clinically meaningful quantification of severity and localization of COVID-19 lesion, here we propose a novel Vision Transformer (ViT) trained in a weakly-supervised manner using severity array labels. Recently, Vision Transformer (ViT) [7] was shown to attain state-of-the-art

(SOTA) performance on the image classification tasks by learning long-range dependency among pixels using a self-attention mechanism [22]. Training a vanilla ViT requires a vast dataset to learn inductive biases, so that the authors of [7] suggest using a hybrid ViT that uses a convolutional neural network (CNN) as a feature embedding network on the small-sized dataset. By extending the idea, our vision Transformer is trained using the low-level CXR feature corpus that are generated using a feature extraction network pretrained on a large CXR dataset. Additionally, we use ROI max-pooling layer that can bridge between pixel-level supervision and the severity array label in a weakly-supervised manner [14].

One of the important advantages of our novel ViT scheme for severity quantification and lesion localization is that the global attention maps from Transformer can lead to full lesion maps where each pixel value directly means the probability of the abnormality of COVID-19. Moreover, our novel progressive self-training, which was inspired by [23], enables to utilize the large unlabeled dataset in addition to the small severity-labeled dataset. By performing both quantitative and qualitative evaluation using the external test data set, we validate the model's performance and its generalization capability for different institute data set."

"For external testing of the quantification and the localization performance of the model, CXR images from another independent domestic institution, Chungnam National Univerity Hospital [CNUH], are used for the quantitative evaluation. The severity labels are annotated from the same two radiologists who labeled CXR images from the other domestic training dataset. In the publicly available BIMCV dataset [4], the COVID-19 lesion segmentation label is annotated on 12 frontal images. We use these images for the qualitative analysis of our model. The numbers of images for quantifying severity and the localization of COVID-19 lesion are summarized in Table 1."

Notes:

- 12 Mar 2021
- Nothing important, paper is too specialized and does not list quantification process

[Automated quantification of COVID-19 pneumonia severity in chest CT using histogram-based multi-level thresholding segmentation]

Abstract

"Chest computed tomography (CT) has proven its critical importance in detection, grading, and follow up of lung affection in COVID-19 pneumonia. There is a close relationship between clinical severity and the extent of lung CT findings in this potentially fatal disease. The extent of lung lesions in CT is an important indicator of risk stratification in COVID-19 pneumonia patients. This study aims to explore automated histogram-based quantification of lung infection in COVID-19 pneumonia in volumetric computed tomography (CT) images in comparison to conventional semi-quantitative severity scoring. This retrospective study enrolled 153 patients with proven COVID-19 pneumonia. Based on the severity of clinical presentation, the patients were divided into three groups: mild, moderate and severe. Based upon the need for oxygenation support, two groups were identified as

follows: common group that incorporated mild and moderate severity patients who did not need intubation, and severe illness group that included patients who were intubated. An automated multi-level thresholding histogram-based quantitative analysis technique was used for evaluation of lung affection in CT scans together with the conventional semi-quantitative severity scoring performed by two expert radiologists. The quantitative assessment included volumes, percentages and densities of ground-glass opacities (GGOs) and consolidation in both lungs. The results of the two evaluation methods were compared, and the quantification metrics were correlated.

Paper

"Typical CT findings in COVID-19 pneumonia include initially, bilateral ground-glass opacities (GGOs) located peripherally, followed later by consolidation, with multi lobe affection of both lungs in most cases. Crazy paving, reticulation and reversed halo sign are encountered with progression of the disease [13–17]. COVID-19 tends to cause severe changes on CT that reflect the widespread aggressive lung injury characteristic of this disease [18].

Lung affection patterns including the number, size, and density of lesions, as well as the overall extent of diseased lung, are all indicators of lung damage and remaining lung reserve. Identifying objective prognostic parameters in CT images of COVID-19 patients may allow better management decisions that hopefully lead to improved clinical outcome [19].

The great sensitivity of chest CT for diagnosing COVID-19 pneumonia poses a significantly increasing workload on radiologists to visually identify and evaluate the extent of COVID-19 affection from thin-thickness CT images, which becomes more challenging if follow up CT imaging is needed [20, 21]."

Previous studies have shown a close relationship between clinical and lung CT severity in COVID-19 pneumonia [14, 22, 23]. The extent of lung lesions in CT is an important indicator of risk stratification in COVID19 patients [19].

Quantitative CT which is already widely used for assessment of diffuse lung diseases, especially in interstitial disease, is an objective tool for detection of image features that are not visually recognizable by the radiologists [24, 25].

The internationally adopted COVID-19 Reporting and Data System (CO-RADS), recommended by the Radiological Society of North America and other radiological societies [26, 27], uses a scoring system from 0 to 5 to classify lung involvement in CT images from very unlikely to very likely, respectively. The CO-RADS has shown a very good performance for predicting the likelihood of COVID-19 infection with substantial interobserver agreement [28].

Most of the imaging studies since the outbreak of COVID-19 have focused on lung CT findings, with only few studies concerning the quantitative analysis of these findings [29].

Unfortunately, this role has been reluctantly integrated into the routine radiological practice because the radiologist's semi-quantitative visual assessment is subjective, time consuming and lacking inter-observer consistency. A semi-quantitative visual severity scoring (SS) system for lung affection has been proposed [30] for assessment of lung affection in SARS, and for evaluation of ARDS [31].

The scoring of disease severity is performed in each of the five lung lobes on a scale from 0 to 5, with 0 indicating no involvement, 1 indicating less than 5% involvement, 2 indicating 5–25% involvement, 3 indicating 26–49% involvement, 4 indicating 50–75% involvement, and 5 indicating more than 75% involvement. The sum of individual lobar scores represents the total SS that ranges from 0 (no involvement) to 25 (maximum involvement). Tis semi-quantitative lobar-based visual scoring system has been adopted in the assessment of the severity of COVID-19 lung affection [15, 32, 33].

Quantitative CT analysis is superior to semi-quantitative visual SS in assessment of the severity of COVID19 infection. Computerized quantitative segmentation methods could provide objective assessment of the percentage of the diseased part of the lung containing GGOs and consolidation to determine the disease burden [34, 35].

The lung volumes measured by CT are well correlated with pulmonary function test results such as total lung capacities and forced vital capacities [36]. Quantitative indices have higher reproducibility than visual scoring and are significantly correlated with lung function and clinical parameters [37].

Still there is no consensus regarding the grading of severity of lung affection in CT images of COVID-19 pneumonia. Simple subjective descriptive terms are cordially used by most radiologists for determination of the severity of lung affection such as mild, moderate, severe or critical."

"Semi-quantitative visual severity scoring assessment Te high-resolution computed tomography (HRCT) images were independently reviewed by two expert radiologists on a picture archiving and communication system (PACS) computer workstation at window setting for lung parenchyma (center, −600 HU; width, 1600 HU). A lobar-based visual SS was independently identified by two expert radiologists, and the scores were then averaged to determine the mean total SS of COVID-19 lung affection. The scoring system considered the overall extent of parenchymal abnormalities, including the GGOs and consolidation (Co), using the definitions of the Fleischner Society glossary of terms for thoracic imaging [39]. Any coexisting reticular pattern (interlobular, intra-lobular and/or peribronchial thickening) or other types of opacities (crazy paving or reversed halo), or pleural effusion were also documented alongside the severity scoring."

- From −1024 to −950 HU (red), representing emphysema (low-attenuation areas, LAA).
- From −949 to −750 HU (yellow), corresponds to healthy lung tissue.
- From −749 to −300 HU (blue), it represents the lung parts which are more dense than healthy lung (high-attenuation areas, HAA) and can be used to quantify ground-glass opacities
- From −299 to +40 HU (violet), this group corresponds to areas with further increase in density, including the semi-consolidation and consolidation.

"For density-based quantitative analysis, 1.5 mm high resolution slices were reconstructed at sharp kernel settings. The lung analysis software of the workstation automatically generates the histogram of distribution of the density of each voxel within the lung and calculates the mean of distribution. In quantitative lung analysis, the following metrics were automatically extracted from the lung density histogram:

- The volume of each lung and the total volume of both lungs (TLV) in cubic centimeter (cc)
- The mean density of each lung and the mean density of both lungs (MLD) in HU
- The volume and percentage of GGOs (volGGO and %GGO, respectively) and the volume and percentage of consolidation (volCo and %Co, respectively) in both lungs.
- The total volume of diseased lung was manually calculated as the sum of volGGO and volCo, and the total percentage of the diseased lung or total lesion load (TLL), representing the disease burden, was calculated as the sum of %GGO and %Co.
- The volumes and percentages of normal and hyperinflated parts of both lungs

“The key for containment of COVID-19 pandemic is early detection and early isolation [40]. CT plays an important role in COVID-19 diagnosis, monitoring, severity stratification, and evaluation of treatment response [6, 41, 42]. Te overall CT picture of COVID-19 pneumonia is based on the severity of lung abnormalities and its distribution. The severity of COVID-19 pneumonia should be objectively stratified on the basis of quantitative data. The severity of lung affection is a critical metric in treatment and prognosis of COVID-19 patients [19, 25]. Severe abnormalities in lung CT at an early stage are suggestive of poor prognosis [40]. Fast, accurate, and reproducible quantitative analytical tools are especially needed for assessment of COVID-19 pneumonia in CT images because, in addition to being multifocal, lung lesions often show rapid progression and change of its pattern [33]. Semi-quantitative visual assessment of COVID-19 lesions is impractical in clinical routine because it is time-consuming, lacks reproducibility and suffers inter-observer and even intra-observer variations. The objective assessment of the disease burden expressed as the percentage of the affected lung relative to the total lung volume is a sensitive and specific metric for estimation of disease progression and treatment response [15, 34].

Computer-aided diagnosis (CAD) has become an important auxiliary diagnostic tool. The automated segmentation of lung lesions from volumetric 3D images allows calculation of the total burden of COVID-19 pneumonia as a percentage subvolume of the total lung volume. Quantitative methods for determination of the severity of lung affection in CT images of covid-19 patients could improve the diagnostic efficiency and mitigate the workload of radiologists, allowing more timely and appropriate treatment decisions for COVID19 patients. Quantitative CT analysis represents a reproducible assessment that allows fast, reliable and potentially predictive tool for assessing disease progression and response to treatment in COVID-19 pneumonia [43].

“In this study, the density-based multi-level thresholding technique was utilized for quantification of COVID19 lung affection in high-resolution thin-cuts volumetric whole-lung CT images. The findings of this study show that the extent of COVID-19 lesions visually scored by radiologists from HRCT images does significantly correlate with the volumetric measurements obtained by the quantitative computer-assisted histogram-based automatic method.

The data obtained in this study show that both semiquantitative visual severity scoring and quantitative CT performed nearly equally, and their parameters correlated well with the clinical data of patients. Our findings indicate that the severity of COVID-19 pneumonia could be accurately stratified on the basis of objective CAD

This is consistent with the results of Li et al. [48] who described a high consistency between COVID-19 severity scoring assessed by visual semi-quantitative CT analysis with the clinical classification of COVID-19 pneumonia. On the other hand, Colombi et al. [24] reported a good correlation between the well-aerated lung volume and the patient’s clinical outcome in COVID-19 pneumonia.

Lanza et al. [49] used density-based quantitative lung analysis to predict clinical outcome in COVID-19 pneumonia regarding the need for respiratory support and the risk of in-hospital mortality. They reported that the compromised lung volume was the most accurate metric in this regard.

Bressem et al. [50] used density-based classification to detect correlation between high-density lung volume (diseased lung) and the severity of COVID-19 pneumonia requiring intensive care unit (ICU) admission and assisted ventilation.

In a recent study, Salvatore et al. [51] used a computer-assisted density-based quantitative technique for automatic segmentation of CT images to calculate the volumes of GGOs, consolidation and the residual healthy lung in COVID-19 pneumonia. Their findings have shown that the results of these quantitative methods are good predictors of COVID-19 patient outcome.

In another recent study, Romanov et al. [52] used histogram-based analysis and HU thresholding to automatically extract CT imaging biomarkers in atypical pneumonia caused by COVID-19 and influenza viruses. The authors reported that the derived imaging biomarkers correlate with the clinical severity scale and the inflammatory laboratory markers."

"In ARDS, a ratio less than 40% between well-aerated lung volume and the total lung capacity was reported to be associated with a higher mortality risk [43].

Colombi et al. [24] used lung attenuation thresholds to quantify well-aerated lung volume in admittance CT images and stated that it can be used to predict the risk of adverse outcome in COVID-19 patients. CT score of the diseased lung has been reported as a risk factor for mortality in ARDS [53].

Patients with COVID-19 lung affection frequently develop ARDS [54, 55] which is the primary cause of death in COVID-19 pneumonia, especially in old aged patients with comorbidities [25].

In all cases of this study regardless its clinical or radiological severity, in presence of GGOs, the automatically extracted whole lung attenuation histograms show blunted peak that is shifted to the right compared with the normal lung, while, in presence of alveolar consolidation, the histograms show a high sharp peak that is shifted more to the right compared to ground-glass opacities and normal lung. These findings are consistent with those described by Sumikawa et al. [44].

Conclusions

The results of this study show that the automated histogram-based quantification of COVID-19 disease burden is a rapid, reliable and reproducible method for objective estimation of lung affection in CT images. There is good correlation between the conventional semi-quantitative CT severity scoring and automated quantitative analysis methods. The automated quantitative methods are especially important in situations like the current COVID-19 pandemic in which radiology departments are overloaded with cases. Some limitations do exist in this study. First and foremost, the lack of longitudinal assessment of disease progression of patients enrolled in the study. The study also included a relatively small number of patients.

Notes:

- 09 December 2021
- Arbitrary assignment of disease severity from 0 to 5 based on "involvement"
    - 0 indicating no involvement,
    - 1 indicating less than 5% involvement,
    - 2 indicating 5–25% involvement,
    - 3 indicating 26–49% involvement,
    - 4 indicating 50–75% involvement,
    - 5 indicating more than 75% involvement
- Cannot use 0 to 5 scales for COVID-19 pneumonia. Must use mild, moderate, severe, or critical.
    - From −1024 to −950 HU (red), representing emphysema (low-attenuation areas, LAA).

    - From −949 to −750 HU (yellow), corresponds to healthy lung tissue.
    - From −749 to −300 HU (blue), it represents the lung parts which are more dense than healthy lung (high-attenuation areas, HAA) and can be used to quantify ground-glass opacities
    - From −299 to +40 HU (violet), this group corresponds to areas with further increase in density, including the semi-consolidation and consolidation.
- Semi-quantitative visual assessment is time consuming, inaccurate, and lacks reproducibility
- "... percentage of the affected lung relative to the total lung volume is a sensitive and specific metric for estimation of disease progression and treatment response"
- If you have acute respiratory distress syndrome (ARDS), you are more likely to die

Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi

Abstract

"...There was no difference in mean DSC between the 2D and 3D U-Net systems. The newly-devised 2D and 3D U-Net approaches were found to be more effective than a commercial auto-segmentation tool. Even the relatively shallow 2D U-Net which does not require high-performance computational resources was effective enough for the lung segmentation. Semantic segmentation using deep learning was useful in radiation treatment planning for lung cancers."

Paper

"In conclusion, we validated the efficacy of semantic segmentation of the lung excluding the trachea and main bronchi using 2D and 3D Net systems. The results revealed that both U-Net systems yielded higher accuracy than the conventional method and had the same mean DSC (0.990). Our newly devised approach was useful for increasing the accuracy of lung contour delineation.

2D and 3D U-Nets did not differ in terms of the accuracy of contour delineation of the lung. Considering that several human organs have longitudinally-oriented structures, like the lung, even relatively shallow 2D delineation may provide useful information allowing the lung contour to be delineated with accuracy sufficient for clinical purposes. Thus, we believe that our novel 2D U-Net system has practical medical applications."

Notes:

- 11 February 2020
- About comparing models to ground truth and not covid infected lung area to total lung area

Other Notes:

MIDL 2020, Keynote by Nikos Paragios: AI-Driven Quantification, Staging and Prognosis of COVID-19

- Extracted from CT: disease extent, number of diseased areas, fat ratio

- Include attributes like: age, sex, lymphocyte count, crp, D-dimers, DLP, CTDivol, preexisting medical conditions
- Do cross validation on different models and compare them to ground truth (radiologist)

Conclusion

Based on the papers that our group has read, quantifying the severity of a covid-19 infected lung is very subjective and therefore any conclusions based on those subjective results does not guarantee that those conclusions are true. The results of Quantification alone can only tell you information about how much a patient's lung is infected, no other information can be drawn from that. However, in conjunction with other information from an infected patient such as a patient's age, comorbidities, and oxygen saturation, as stated by *Severity and Consolidation Quantification of COVID-19 From CT Images Using Deep Learning Based on Hybrid Weak Labels*, a medical expert can make a more accurate prognosis. In addition to having information about the patient along with the covid-19 severity quantification, the accumulation of this type of information along with other similar types of information from multiple patients allows medical experts to make even more accurate prognoses based on the correlations in information, diagnoses, and outcomes among similar patients as mentioned in *Automated quantification of COVID-19 pneumonia severity in chest CT using histogram-based multi-level thresholding segmentation*.

The method of quantifying the severity of covid-19 in a patient's lungs are based on the equations listed in *Severity and Consolidation Quantification of COVID-19 From CT Images Using Deep Learning Based on Hybrid Weak Labels*. The quantification equations are (severity score) = (area of infected region) / (area of lung) and (consolidation score) = (area of consolidation) / (area of lung). It is important to note that the equations listed are based on 2D slices on a patient's lungs which do not fully represent the entire severity of infection of a patient's lungs; therefore, reconstruction of a patients' lungs in 3D is preferable in determining a better quantification of severity..For this case, Joseph has taken the responsibility of figuring out this method of quantification; though, it's not guaranteed that he will figure out that process when this project is due.

The reason why our group decided to go with the equations mentioned above is because it's a very easy method of determining how much of a patient's lungs is infected by covid-19. The reasons why we didn't choose any other methods of quantifying the severity of covid-19 is that:

- The amount of infection in the lungs does not necessarily tell you everything about the patient's health and future health
- We do not have a lot of additional information about the patient's medical records nor is the information we got along with their scans useful.
- We don't have the computational resources and appropriate datasets for proper analysis

If we had access to proper datasets, such as the ones listed in the research papers, along with sufficient computation, we could have built a machine learning classification model that would have the results of our covid-19 severity quantification along with patient data to predict the future outcome of an infected patient.