2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW) | 979-8-3503-0261-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSPW59220.2023.10193538

ENSEMBLE METHODS FOR ENHANCED COVID-19 CT SCAN SEVERITY ANALYSIS

Anand Thyagachandran, Hema A Murthy

Department of Computer Science and Engineering,
Indian Institute of Technology Madras, India
{tanand, hema}@cse.iitm.ac.in

ABSTRACT

Computed Tomography (CT) scans provide a high-resolution image of the lungs, allowing clinicians to identify the severity of infections in COVID-19 patients. This paper presents a domain knowledge-based pipeline for extracting infection regions from COVID-19 patients using a combination of image-processing algorithms and a pre-trained UNET model. Then, an infection rate-based feature vector is generated for each CT scan. The infection severity is then classified into four categories using an ensemble of three machine-learning models: Random Forest, Support Vector Machines, and Extremely Randomized Trees. The proposed system is evaluated on the validation and test datasets with a macro F1 score of 58% and 46.31%, respectively. Our proposed model has achieved 3^{rd} place in the severity detection challenge as part of the IEEE ICASSP 2023: AI-enabled Medical Image Analysis Workshop and COVID-19 Diagnosis Competition (AI-MIA-COV19D). The implementation of the proposed system is available at *https://github.com/aanandt/Enhancing-COVID-19-Severity-Analysis-through-Ensemble-Methods.git*

***Index Terms*—** COVID-19, CT-Scans, Infection Segmentation, Machine Learning Methods, Severity Analysis

1. INTRODUCTION

The outbreak of COVID-19 has resulted in an urgent need for prompt and precise diagnostic testing to detect individuals who may have contracted the SARS-CoV-2 virus. Laboratory tests such as Reverse Transcription Polymerase Chain Reaction (RT-PCR) [1] and antigen tests are commonly used for diagnosing COVID-19. While these tests detect the virus's presence from the respiratory samples [2], but fail to provide an accurate analysis of the disease severity. Additional diagnostic tools are needed to assess the extent of lung infections. Radiological imaging of the chest, including chest radiography and CT scans, is essential to determine the severity of COVID-19 infections [3]. At the same time, chest X-rays do not provide sufficient resolution to evaluate the extent of lung damage. CT scans provide a more detailed view of the lungs to identify the distribution and area of infection. Clinicians can comprehend vital information to gauge the severity of the disease and deliver timely and effective treatment to the patients.

Table 1. Summary of the severity detection challenge dataset.

Category	Train	Val	**Description**
Mild	133	31	Few GGOs. Pulmonary parenchymal involvement $\leq 25\%$ or absence
Moderate	124	20	GGOs. Pulmonary parenchymal involvement $25 \leq 50\%$
Severe	166	45	GGOs. Pulmonary parenchymal involvement $50 \leq 75\%$
Critical	39	5	GGOs. Pulmonary parenchymal involvement $\geq 75\%$

Clinicians typically observe ground-glass opacities (GGO) [4], which indicate lung inflammation but do not obstruct the underlying pulmonary vessels. Consolidations are the advanced stage of GGO and hide the underlying vessels [5]. Pleural effusion occurs when fluid accumulates excessively in the pleural space surrounding the lungs and is a highly severe case of COVID-19 [6]. These clinical features are critical in identifying and diagnosing COVID-19 in patients [7]. In addition to GGO and consolidations, features such as the halo sign (central consolidations surrounded by GGO), the reverse halo sign (central ground-glass lucent area with peripheral consolidation), and crazy paving patterns are observed by the clinicians [4, 8]. These features and their distribution provide crucial information to diagnose and manage COVID-19 patients.

Machine learning and deep learning methods have been widely employed for classifying and segmenting infection regions in CT scans. Several studies, including [9, 10, 11, 12], have attempted to classify CT scans as COVID-19 or non-COVID-19. Furthermore, some studies [13, 14, 15, 16] have extended the classification task to include three categories - COVID-19, Community-acquired Pneumonia (CAP), and Normal. Since COVID-19 and CAP have similar clinical features, it is essential to differentiate between these two categories to monitor the progression of COVID-19. Although these methods have shown good results in classification, there is a need to identify the severity of the patients. Many research works have focused on infection region segmentation from CT scans, which can aid in severity analysis. However, generating large medical image datasets for infection segmen-

tation is time-consuming and requires expert annotation. Different strategies, such as semi-supervised [17, 18], weakly supervised [19, 20], and unsupervised methods [21, 22], have been proposed to address the unavailability of a large corpus. Some works, such as [23, 24], have used the infection volume to classify CT scans into healthy, mild, moderate, severe, and critical categories.

The severity of COVID-19 patients is determined using the Chest CT Severity Score (CTSS) method, as described in [25]. This method assigns a score between 0 and 25 based on the distribution and magnitude of abnormalities observed on chest CT scans [26, 27]. In line with this approach, we have developed a method that uses various image processing algorithms and a pre-trained UNET model [28] to extract infected regions from CT scans. We then propose a feature vector based on the infection rate and use various machine-learning models to predict the severity of chest CT scans.

The paper is organized as follows: **Section 2** explains the dataset used in the study. **Section 3** presents the proposed method for segmenting relevant clinical features and classifying COVID-19 patients into different severity classes. **Section 4** discusses results and inferences. Finally, **Section 5** summarizes the present study.

2. DATASET

The COVID-19-CT-DB dataset used in this study was provided as part of the "AI-enabled Medical Image Analysis Workshop, and COVID-19 Diagnosis Competition (AI-MIA-COV19D)" [29, 30, 31, 32, 33, 34]. The CT scans were collected between September 1, 2020, and March 31, 2021, and were annotated by four experienced medical experts, including two radiologists and two pulmonologists. The severity detection challenge uses the subset of the COVID-19-CT-DB dataset, and the summary of the dataset is shown in **Table 1**.

3. PROPOSED METHOD

The proposed method consists of two parts, lesion segmentation from the CT scan and quantifying the severity based on the infection rate. The proposed pipeline uses various image processing algorithms, machine learning, and deep learning models to predict the severity score for COVID-19 patients.

3.1. Infection Segmentation

The infection segmentation pipeline uses domain knowledge to extract the lesion areas. Initially, the lung region is extracted from the chest CT image using a pre-trained UNET model. Then, the resultant image's contrast is enhanced by histogram hyperbolization [35]. Next, the pulmonary blood vessels are enhanced using the top-hat morphological operation [36]. Finally, the resultant image is employed with morphological operations to fine-tune the infection regions.

3.1.1. Lung Mask Generation

The chest CT scan contains the lung region and organs like the trachea, diaphragm, heart, stomach, and tissues. This module extracts the lung region from the chest CT scan and removes the unnecessary areas from further analysis. A pre-trained UNET [28] model is used to extract the lung region from the CT scan images. The UNET model is trained on the HU scale CT images, but the challenge dataset CT scan images are in JPG format. A linear transformation is applied to the JPG images to make them compatible with the input format for the UNET model.

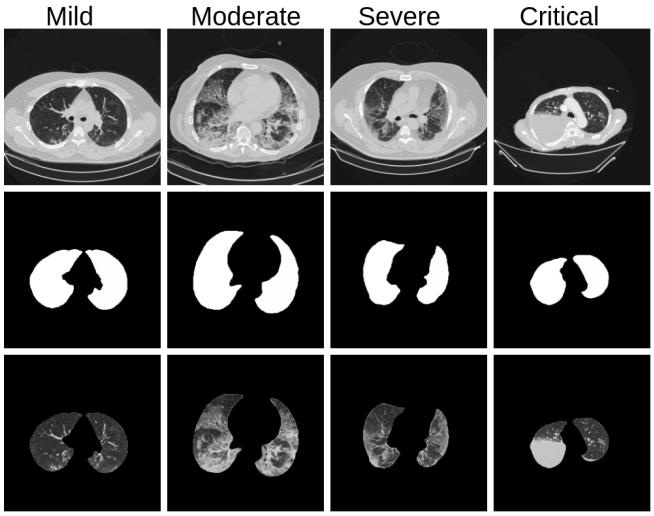

Fig. 1. First row represents CT scan images of mild, moderate, severe, and critical categories provided in the challenge. The second row shows the lung mask generated from the images. The third row depicts the segmented region of interest for further analysis of each category.

The lung mask involvement accounts for variations in the number of CT image slices across patients. An empirical threshold on the lung mask's involvement and a threshold on the number of slices from the middle region (slice number which lies between one-third to two-thirds of the total number of slices) are used to select the CT scan images for further analysis. The segmented CT scan images are applied with a Gaussian filter to smooth the image without significantly reducing the sharpness of the edges. Further, this image is fed into the image enhancement module. The lung masks and extracted regions of interest from different categories of COVID-19 severity are shown in **Fig. 1**.

3.1.2. Image Enhancement using Histogram Hyberbolization

Histogram hyperbolization, introduced in [35], is a non-linear image enhancement method that improves the contrast of an image by adjusting its perceived brightness levels. The primary objective of this approach is that it mimics the process of vision and could correspond to the radiologist's view of the image.

$$J(I) = c(exp[log(1 + \tfrac{1}{c}) * normcm[I]] - 1) \qquad (1)$$

where $J(I)$ is the space-invariant hyperbolization transformation of the image I. c is an empirical threshold.

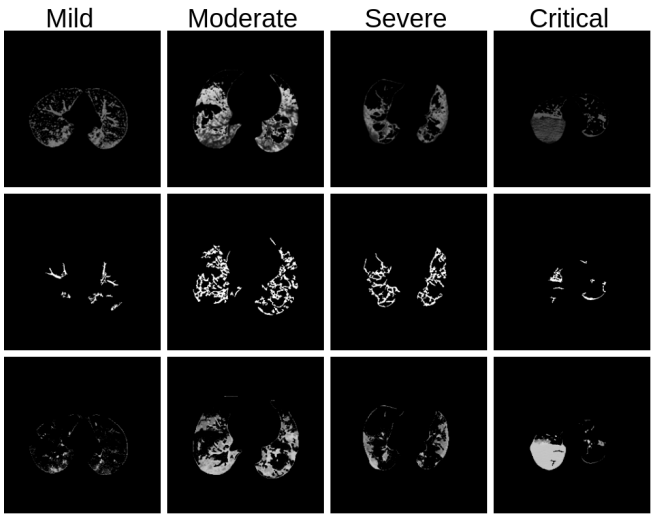

Fig. 2. First row shows the histogram hyperbolized lung region of mild, moderate, severe, and critical classes. The second row depicts enhanced blood vessels using top-hat transform. The third row represents the extracted infection regions.

$normcm$ is the normalized cumulative distribution of the intensity histogram. The contrast-enhanced image contains infection regions and the pulmonary vessels. The contrast-enhanced lung region of the CT scan images using histogram hyperbolization is shown in **Fig. 2**.

3.1.3. Morphological operations

The present work mainly uses three commonly-used morphological operations: dilation, opening, and top-hat [37]. A kernel matrix (a rectangular kernel (3 x 3)) is used to convolve with the input image. The dilation operation chooses the maximum pixel value from each (3 x 3) neighborhood and replaces the center pixel. The erosion operation finds the minimum pixel value from the kernel neighborhood. The opening morphological procedure removes small foreground objects from the binary image by employing the erosion operation followed by the dilation operation. The top-hat transformation enhances blood vessel-like structures by subtracting the image from its opening morphological image. Foreground-connected components are computed from the resultant image, and a criterion on the size or area is applied to retain these connected components.

The OTSU adaptive binarization method [38] is applied to the contrast-enhanced image from the previous module. Morphological operation top-hat transform is applied to enhance and remove the blood vessels. The resultant image contains blob-like structures while removing the blood vessels and are removed by the area-based connected components method. The dilation operation is employed to fine-tune the boundary of the extracted infection region. The enhanced blood vessels and infection masks generated from the different severity classes are shown in **Fig. 2**.

3.2. Lesion To Severity Score

This section describes the various methods explored to identify the correlation between the infection regions in the CT images and their severity class. Inspired by the CTSS method, a weighted average on the percentage of infection in the left and right lungs is estimated. Further, various machine learning algorithms are experimented with to classify the CT scan into different severity classes.

3.2.1. Weighted Average Method (WAM)

The infection rate in the left and right lungs are estimated separately for each slice in the CT scan with the help of the infection mask and the generated UNET lung mask. In CTSS estimation, the right lung is divided into three lobes and the left into two lobes, and the radiologist visually determines each lobe's degree of involvement. The lobe scores are calculated by the percentage of infection involvement in each lobe [39]. The CTSS is estimated by aggregating the average score lobe across the CT scan slices. In the WAM, a score ranging from one to four is assigned based on the rate of infection, one (0- 25%), two (25 - 50%), three (50 - 75%), and four (75 - 100 %). A weighted average method is employed to determine the severity score for each CT scan slice. The right lung region score is multiplied by a weight of three and the left lung region score by two. The severity score is estimated by averaging the scores across the CT scan images.

3.2.2. Non-linear Methods

Each CT scan image is represented by the infection rates of the left and right lungs, resulting in an 80-dimensional feature vector for each patient's scan. Two methods are employed for fixed-length feature representation: for CT scans with more than 40 slices, the slices are uniformly divided into forty regions, and the median slice features are selected from each region; for CT scans with fewer than 40 slices, the feature vector is enhanced by computing the average infection rate from the available slices and appending it to generate an 80-dimensional vector. Machine learning models, including logistic regression (LR), decision tree (DT), gradient boost (Gboost), extreme gradient boosting (XGboost), Ada boost, k-nearest neighbor (kNN), naive Bayes (NB), random forest (RF), extremely randomized tree (ERT) [40], support vector machine (SVM), and voting based ensemble models, are employed to learn from these feature representations to predict the severity classes. The machine learning models are implemented with the help of the sklearn library in Python.

4. RESULTS AND DISCUSSIONS

The proposed method aims to identify the correlation between the lung infection rate and the severity class. The weighted average method (WAM) discovers a linear mapping between the infection rate and the severity classes, which yielded a macro F1 score of 33%. This result led to the realization of

Table 2. The severity analysis of various machine learning classifiers using validation dataset CT scans.

Models	WAM	DT	LR	Adaboost	Gboost	NB	kNN	XGboost	RF	ERT	SVM	Ensemble
Precision	0.42	0.36	0.40	0.43	0.61	0.51	0.57	0.64	0.65	0.69	0.71	0.68
Recall	0.36	0.37	0.45	0.52	0.43	0.61	0.51	0.51	0.52	0.56	0.56	0.55
F1 score	0.33	0.36	0.41	0.45	0.45	0.48	0.52	0.54	0.55	0.59	0.59	0.58

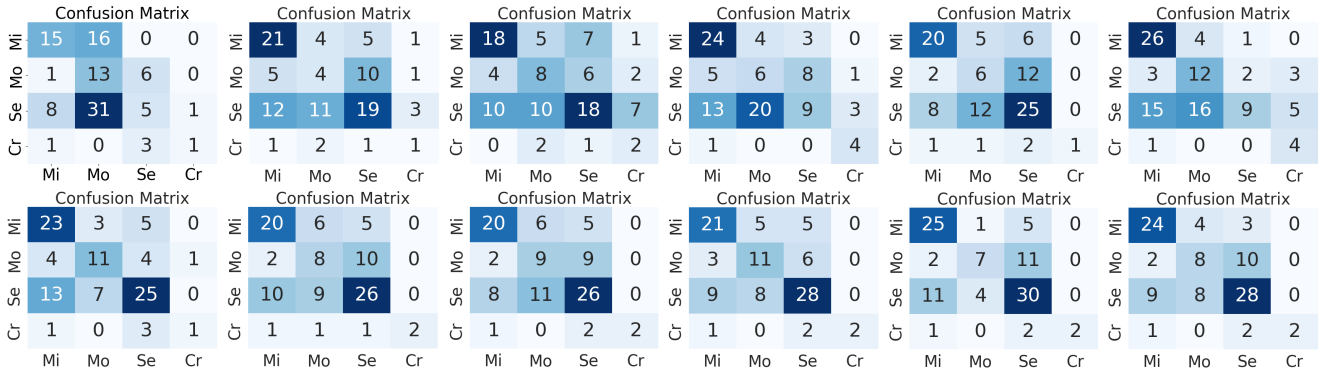

Fig. 3. The first row (from left to right) depicts the confusion matrices of the WAM, DT, LR, Adaboost, Gboost, and NB. The second row represents the kNN, XGboost, RF, ERT, SVM, and Ensemble (RF, ERT, SVM) classifiers. The axis labels such as Mi-mild, Mo-moderate, Se-severe, and Cr-critical.

the non-linear relationship of the infection rate-based severity analysis, prompting the exploration of various machine-learning methods. The results of various classifiers are shown in **Fig. 3** and tabulated in **Table 2**.

The machine learning models achieve better results (regarding macro F1 score) than the WAM. The RF, ERT, and SVM serve top-3 performing models for the severity analysis. The ensemble model can improve the accuracy and robustness of predictions, reduce over-fitting, and improve the model's generalization. The proposed pipeline works well, and it has some class misclassification issues. The confusion matrices show that the mild class achieves higher classification accuracy and less misclassification with other classes. In the case of the moderate category, more misclassifications towards the severe category; the SVM, Ensemble, and RF models provide more misclassifications to the severe category than the mild category. Similarly, the severe class CT scans are misclassified into the mild and moderate classes. We have observed that the boundary detector has issues in some severe cases; appropriate domain information is required to improve lung segmentation accuracy.

The standard evaluation metrics, such as precision, recall, and macro F1 score, are used to evaluate the models. Precision refers to the proportion of true positives among all predicted positive instances. Recall measures the proportion of true positives among all actual positive instances. The Macro F1 score is a harmonic mean of precision and recall, providing an overall measure of model performance across all classes. It considers false positives and negatives and is a valuable metric for imbalanced classes. The performance of linear and non-linear classifiers is shown in **Table 2**. The baseline model is based on a convolutional neural network (CNN)- recurrent neural network (RNN) where the CNN is used as a feature extractor, and RNN is used to model the sequence of features extracted from the CT scan images. The baseline model achieved a macro F1 score of 38% in the validation data. It is observed from the **Table 2** that all non-linear models, except the decision tree, outperformed the baseline model. A voting-based ensemble model (RF, ERT, and SVM) achieves the macro F1 score of 46.31% and performs better than the baseline model (40.06%) on the challenge test dataset.

5. CONCLUSION

The current study proposes a domain knowledge-based preprocessing pipeline to extract relevant lesion regions from chest CT scans. Further, a novel set of infection rate-based features is generated. Machine learning models are trained with these features to predict the patient's COVID-19 infection severity. A voting-based ensemble model comprising Random Forest (RF), Extremely Randomized Trees (ERT), and Support Vector Machine (SVM) achieved a macro F1 score of 58% on the validation data and 46.31% on the test data.

6. REFERENCES

[1] Shannon L Emery et al., "Real-time reverse transcription–polymerase chain reaction assay for sars-associated coronavirus," *Emerging infectious diseases*, vol. 10, no. 2, pp. 311, 2004.

[2] Moustapha Dramé et al., "Should rt-pcr be considered a gold standard in the diagnosis of covid-19?," *Journal of medical virology*, 2020.

[3] Elmehdi Benmalek et al., "Comparing ct scan and chest x-ray imaging for covid-19 diagnosis," *Biomedical Engineering Advances*, vol. 1, pp. 100003, 2021.

[4] Thomas C Kwee and Robert M Kwee, "Chest ct in covid-19: what the radiologist needs to know," *Radiographics*, vol. 40, no. 7, pp. 1848–1865, 2020.

[5] Minhua Yu et al., "Thin-section chest ct imaging of covid-19 pneumonia: a comparison between patients with mild and severe disease," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, pp. e200126, 2020.

[6] Atsushi Nambu et al., "Imaging of community-acquired pneumonia: roles of imaging examinations, imaging diagnosis of specific pathogens and discrimination from noninfectious diseases," *World Journal of Radiology*, vol. 6, no. 10, pp. 779, 2014.

[7] Nan Zhang et al., "Clinical characteristics and chest ct imaging features of critically ill covid-19 patients," *European Radiology*, vol. 30, pp. 6151–6160, 2020.

[8] Liaoyi Lin et al., "Ct manifestations of coronavirus disease (covid-19) pneumonia and influenza virus pneumonia: A comparative study," *American Journal of Roentgenology*, vol. 216, no. 1, pp. 71–79, 2021.

[9] Junlin Hou et al., "Cmc-cov19d: Contrastive mixup classification for covid-19 diagnosis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 454–461.

[10] Dimitrios Kollias et al., "Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 537–544.

[11] Shuang Liang et al., "A hybrid and fast deep learning framework for covid-19 detection via 3d chest ct images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 508–512.

[12] Radu Miron et al., "Evaluating volumetric and slice-based approaches for covid-19 detection in chest cts," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 529–536.

[13] Parnian Afshar et al., "Covid-ct-md, covid-19 computed tomography scan dataset applicable in machine learning and deep learning," *Scientific Data*, vol. 8, no. 1, pp. 1–8, 2021.

[14] Pratyush Garg et al., "Multi-scale residual network for covid-19 diagnosis using ct-scans," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8558–8562.

[15] Shubham Chaudhary et al., "Detecting covid-19 and community acquired pneumonia using chest ct scan images with deep learning," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8583–8587.

[16] Shuohan Xue and Charith Abhayaratne, "Covid-19 diagnostic using 3d deep transfer learning for classification of volumetric computerised tomography chest scans," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8573–8577.

[17] Deng-Ping Fan et al., "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2626–2637, 2020.

[18] Jun Ma et al., "Active contour regularized semi-supervised learning for covid-19 ct infection segmentation with limited annotations," *Physics in Medicine & Biology*, vol. 65, no. 22, pp. 225034, 2020.

[19] Xinggang Wang et al., "A weakly-supervised framework for covid-19 classification and lesion localization from chest ct," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2615–2625, 2020.

[20] Zhongyi Han et al., "Accurate screening of covid-19 using attention-based deep 3d multiple instance learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2584–2594, 2020.

[21] Rui Xu et al., "Unsupervised detection of pulmonary opacities for computer-aided diagnosis of covid-19 on ct images," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 9007–9014.

[22] Han Chen et al., "Unsupervised domain adaptation based covid-19 ct infection segmentation network," *Applied Intelligence*, pp. 1–14, 2022.

[23] Ben Ye et al., "Severity assessment of covid-19 based on feature extraction and v-descriptors," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7456–7467, 2021.

[24] Yazan Qiblawey et al., "Detection and severity classification of covid-19 in ct images using deep learning," *Diagnostics*, vol. 11, no. 5, pp. 893, 2021.

[25] Arthur WE Lieveld et al., "Chest ct in covid-19 at the ed: validation of the covid-19 reporting and data system (co-rads) and ct severity score: a prospective, multicenter, observational study," *Chest*, vol. 159, no. 3, pp. 1126–1135, 2021.

[26] Kunwei Li et al., "Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19)," *European radiology*, vol. 30, pp. 4407–4416, 2020.

[27] Kunhua Li et al., "The clinical and chest ct features associated with severe and critical covid-19 pneumonia," *Investigative radiology*, 2020.

[28] Johannes Hofmanninger et al., "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, no. 1, pp. 1–13, 2020.

[29] Dimitrios Kollias et al., "Ai-mia: Covid-19 detection & severity analysis through medical imaging," *arXiv preprint arXiv:2206.04732*, 2022.

[30] Anastasios Arsenos et al., "A large imaging database and novel deep neural architecture for covid-19 diagnosis," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2022, p. 1–5.

[31] Dimitrios Kollias et al., "Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 537–544.

[32] Dimitrios Kollias et al., "Deep transparent prediction through latent representation analysis," *arXiv preprint arXiv:2009.07044*, 2020.

[33] Dimitris Kollias et al., "Transparent adaptation in deep medical image diagnosis.," in *TAILOR*, 2020, p. 251–267.

[34] Dimitrios Kollias et al., "Deep neural architectures for prediction in healthcare," *Complex & Intelligent Systems*, vol. 4, no. 2, pp. 119–131, 2018.

[35] Werner Frei, "Image enhancement by histogram hyperbolization," *Computer Graphics and Image Processing*, vol. 6, no. 3, pp. 286–294, 1977.

[36] Yanjie Zhu et al., "Automatic segmentation of ground-glass opacities in lung ct images by using markov random field-based algorithms," *Journal of digital imaging*, vol. 25, no. 3, pp. 409–422, 2012.

[37] Petros Maragos and Ronald Schafer, "Morphological filters–part i: Their set-theoretic analysis and relations to linear shift-invariant filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 8, pp. 1153–1169, 1987.

[38] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[39] Ran Yang, Xiang Li, Huan Liu, Yanling Zhen, Xianxiang Zhang, Qiuxia Xiong, Yong Luo, Cailiang Gao, and Wenbing Zeng, "Chest ct severity score: an imaging tool for assessing severe covid-19," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, pp. e200047, 2020.

[40] Pierre Geurts, Damien Ernst, and Louis Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, pp. 3–42, 2006.