IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 July 2022, accepted 13 September 2022, date of publication 30 September 2022, date of current version 22 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3211080

RESEARCH ARTICLE

Reliability of Machine Learning in Eliminating Data Redundancy of Radiomics and Reflecting Pathophysiology in COVID-19 Pneumonia: Impact of CT Reconstruction Kernels on Accuracy

YAUHEN STATSENKO[1,2], TETIANA HABUZA[3,4], TATSIANA TALAKO[1,2], TETIANA KURBATOVA[1], GILLIAN LYLIAN SIMIYU[1], DARYA SMETANINA[1], JUANA SIDO[5], DANA SHARIF QANDIL[6], SARAH MERIBOUT[7], JURI G. GELOVANI[8,9], KLAUS NEIDL-VAN GORKOM[1], TALEB M. ALMANSOORI[1], FATMAH AL ZAHMI[10,11], TOM LONEY[11], ANTHONY BEDSON[12], NERISSA NAIDOO[11], ALIREZA DEHDASHTIAN[13], MILOS LJUBISAVLJEVIC[1,2], JAMAL AL KOTEESH[1,13], AND KARUNA M. DAS[1]

[1]College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates
[2]Abu Dhabi Precision Medicine Virtual Research Institute (AD PM VRI), United Arab Emirates University, Al Ain, United Arab Emirates
[3]College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates
[4]Big Data Analytics Center (BIDAC), United Arab Emirates University, Al Ain, United Arab Emirates
[5]Eye Microsurgery Center "Voka", 220089 Minsk, Belarus
[6]Ras Al Khaimah Medical and Health Sciences University, Al Qusaidat, Ras al Khaimah, United Arab Emirates
[7]Department of Medicine, University of Constantine 3, Constantine 25000, Algeria
[8]Siriraj Hospital, Mahidol University, Salaya 73170, Thailand
[9]Biomedical Engineering Department, College of Engineering, Wayne State University, Detroit, MI 48202, USA
[10]Mediclinic Middle East Parkview Hospital, Dubai, United Arab Emirates
[11]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates
[12]Sheikh Shakhbout Medical City, Abu Dhabi, United Arab Emirates
[13]Radiology Department, Tawam Hospital, Abu Dhabi, United Arab Emirates

Corresponding authors: Yauhen Statsenko (e.a.statsenko@gmail.com), Tetiana Habuza (gabuzat@gmail.com), Taleb M. Almansoori (taleb.almansoor@uaeu.ac.ae), and Jamal Al Koteesh (jkoteesh@seha.ae)

This work was supported in part by the Mohammed Bin Rashid University Collaborative Research under Award 31M418-21M153, and in part by the United Arab Emirates University College of Medicine and Health Sciences Research Grant 12M119.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Department of Health Abu Dhabi under Reference No. DOH/CVDC/2020/887.

ABSTRACT **Background:** Radiomical data are redundant but they might serve as a tool for lung quantitative assessment reflecting disease severity and actual physiological status of COVID-19 patients. **Objective:** Test the effectiveness of machine learning in eliminating data redundancy of radiomics and reflecting pathophysiologic changes in patients with COVID-19 pneumonia. **Methods:** We analyzed 605 cases admitted to Al Ain Hospital from 24 February to 1 July, 2020. They met the following inclusion criteria: age≥18 years; inpatient admission; PCR positive for SARS-CoV-2; lung CT available at PACS. We categorized cases into 4 classes: mild <5% of pulmonary parenchymal involvement, moderate - 5-24%, severe - 25-49%, and critical ≥ 50%. We used CT scans to build regression models predicting the oxygenation level, respiratory and cardiovascular functioning. **Results:** Radiomical findings are a reliable source of information to assess the functional status of patients with COVID-19. Machine learning models can predict the oxygenation level, respiratory and

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

cardiovascular functioning from a set of demographics and radiomics data regardless of the settings of reconstruction kernels. The regression models can be used for scoring lung impairment and comparing disease severity in follow up studies. The most accurate prediction we achieved was 6.454±3.715% of mean absolute error/range for all the features and 7.069±4.17% for radiomics. **Conclusion:** The models may contribute to the proper risk evaluation and disease management especially when the oxygen therapy impacts the actual values of the functional findings. Still, the structural assessment of an acute lung injury reflects the severity of the disease.

INDEX TERMS Blended machine learning model, COVID-19, functional outcomes, lung structural changes, pneumonia, radiomics, SARC-CoV-2, structure-function association.

ABBREVIATIONS AND ACRONYMS

AUC -	area under the curve
CAP -	community-acquired pneumonia
CI -	confidence interval
con_ -	consolidation
COVID-19 -	coronavirus disease 2019
CRP -	C-reactive protein
CT -	computed tomography
eff_ -	pleural effusion
ggo_ -	ground glass opacity
G6PD -	glucose-6-phosphate dehydrogenase
HCT -	hematocrit
HGB -	hemoglobin
IRF -	immature reticulocyte fraction
LDH -	lactate dehydrogenase
LLL -	left lower lobe
LUL -	left upper lobe
MCH -	mean corpuscular hemoglobin level
MCHC -	mean corpuscular hemoglobin concentration
MCV -	mean corpuscular volume
ML -	machine learning
MPV -	mean platelet volume
PCR -	polymerase chain reaction
PTT -	activated partial thromboplastin
_rate -	percentage occupied with lesions of either type in the total lung volume
RBC -	red blood cells
RDW-CV -	red blood cell distribution width
RLL -	right lower lobe
RML -	right middle lobe
RUL -	right upper lobe
SAP -	SARS-CoV-2 associated pneumonia
SARS-CoV-2 -	severe acute respiratory syndrome-related coronavirus 2
_std -	standard deviation
WBC -	white blood cells

I. INTRODUCTION

Atypical viral pneumonia associated with COVID-19 has been a significant burden of morbidity and mortality across the world. The clinical variants of the disease range from mild forms to severe respiratory failure [1]. Most patients are diagnosed with mild and moderate forms of COVID-19 and have a favorable prognosis. The severe course of the infection can present with secondary hypoxemia and result in acute respiratory distress syndrome [2], [3], [4], [5]. From the beginning of the pandemic, clinicians started researching the distinctive features of the disease pathogenesis [1] and its course to stratify risks for the optimal case management [2], [6]. The idea under this was to study pathophysiology which may have the following application. *First*, it will reveal the pathologic processes underlying non-typical devastation by the disease, i.e. what accounts for the higher rates of morbidity and mortality compared to those of community acquired pneumonia (CAP). *Second*, COVID-19 specific mechanisms of acute respiratory failure could provide a clue to optimal therapy.

Physicians found a dissociation between the degree of lung injury and the severity of the coronavirus disease emerged in 2019 [7]. To explain this, they proposed to look at COVID-19 associated pneumonia as an interplay of two major pathophysiologic mechanisms: an impairment of the lung parenchyma and disarrangement of metabolism at the system level. None of them can attribute solely for the progression specific to the disease. Logically, medical doctors should consider both the factors while assessing the severity of a clinical case or reporting the radiologic findings. Few models provide information on the lung impairment and pathophysiologic changes during the disease progression and most of the models are classification algorithms (see Table 1). The results of correlation analysis and classification may reflect the disease course inadequately [2], [8], [9], [10], [11], [12]. Still these methods can be used to model lifelong dynamics of the studied data.

Bioengineering and data science suggests using the following technical approaches to increase the accuracy of diagnostics and risk stratification. *First*, *radiomics* provides a quantification of the structural changes from diagnostic imaging. One may use it for assessing pulmonary compromise on computed tomography (CT) in pneumonia patients. *Second*, *machine learning* (ML) allows us to build multivariate models with a set of predictors (e.g., metabolites) put into a model. With ML a physician may solve a very challenging task of assessing the examinee's metabolome. The approach enables us to quantify the hypermetabolic state in acute respiratory failure which is a contributing factor to the extraordinary ventilatory and oxygenation demands in the infected patients. *Finally*, applying artificial intelligence

TABLE 1. Performance of machine learning models in which structural data were used to predict pathologic findings, pneumonia aetiology and outcomes.

Reference	Predictors	Ground truth	Accuracy, %	Sensitivity, %	Specificity, %	AUC	Odds ratio
DISEASE PATHOPHYSIOLOGY AND PATHOMORPHOLOGY							
[13]	CT images on admission	SpO_2 < 90% on admission	69	70	69	0.75	
[14]	Low-dose CT images	Shortness of breath					2.15
[15]	CT images	Clinical severity				0.95	
[16]	Radiomics		74	83	68	0.84	
[17]	Radiomics					0.87	
[17]	Radiomics + age, comorbidity		87	100	82	0.92	
[18]	Radiomics + CT images		81	88	78	0.86	
[19]	Radiomics	Opacification type				0.99	
[20]	Radiomics	Residual lung lesions 3 months after discharge		92	83	0.88	
OUTCOMES							
[19]	Radiomics	Recovery vs death				0.85	
[21]	Radiomics					0.81	
[22]	Radiomics					0.84-0.87	
[23]	Radiomics and clinical features		88	88	89	0.95	
[24]	Radiomics + CT images, age, CRP	14 day outcomes		88.8	73.0	0.88	
[25]	Radiomics	PCR negative status during treatment		60.0	63.0	0.81	
[21]	Radiomics	ICU admission				0.84	
[22]	Radiomics					0.81	
DIFFERENTIATION BETWEEN COVID-19 AND NON-COVID-19 PNEUMONIA							
[26]	Radiomics + lesion charactetistics, lymph nodes enlargement, pleural effusion	COVID-19 vs another type of viral pneumonia	94.4	92.9	97,1	0.96	
[27]	Radiomics, volumentric and clinical features	Positive vs negative PCR test for COVID-19		81.6	92.3	0.93	
[28]	Radiomics derived from radiograms			85	67	0.87	
[29]	Radiomics		98			1.00	
[30]			85.2	69.5	91.6	0.88	
[31]						1.00	
[32]		COVID-19 vs Influenza A pneumonia				0.87	
[33]	Radiomics + CT images		90.3	89.9	90.7	0.96	

* All the papers reported findings on radiomics retrieved from lung CT except for [28] where radiograms were used

to the consecutive studies of the laboratory findings and medical imaging dwells promises to combine the evidence coming from the diagnostic modalities based on conceptually different methodologies [34], [35], [36]. This will help to cover all the known pathophysiologic mechanisms of atypical pneumonia caused by SARS-CoV-2.

The motivation for the current study was as follows. *First*, we tried to fill the gap between the predictive models built by data scientists and the real needs of clinicians. Most of the existing ML models for risk stratification in COVID-19 are not widely used in clinical practice as they predict the disease outcome but not its progression. The clinical utility of such a prediction is limited because of the relatively low mortality, however the number of severe cases still remains high. *Our second motivation* was to contribute to the optimal patient management which is a demanding task for admission, pulmonology and ICU departments. The essential tool for such management is a quantitative assessment of individual risks on admission. The same quantitative score can be later used on the follow-up examination to detect disease worsening. Today's practice is to quantify structural changes through radiomics: the radiomics data reflect the lung involvement as well as background diseases. However, medical doctors have to deal with a big number of numeric values.

For this reason, *the third task* was to reduce the number of analyzed features by providing a summary score for case assessment. With the ML algorithm based on a combination of radiomic data and functional parameters of respiratory and cardiovascular systems one can calculate a single measure of disease severity and progression, e.g., the oxygen saturation level.

A. STRUCTURAL CORRELATES OF THE LUNG IN ACUTE RESPIRATORY FAILURE

Machine learning prognostication in patients with COVID-19 can be improved by combining the clinical variables with

the initial radiographs [37]. This is because the level of lung impairment correlates with the outcomes (e.g. risk of mortality in infected patients, etc.) [38]. A typical approach to the analysis of radiologic findings is to quantify them. There are several ways to do this. The easiest one is human-driven scoring the lesions. When applied into clinical monitoring, such a system improved prediction of in-hospital mortality in patients with COVID-19 [38]. However, the most promising way to quantify radiologic findings is to extract features for analysis of diagnostic images with radiomics. It provides a summary of both the studied disease and background conditions (e.g., the level of pulmonary fibrosis, emphysema, the volume of pulmonary effusion, etc.). Logically, confounders may change the disease course thus impacting the supposed outcomes.

The advantages of applying radiomics into practice are as follows. *First*, it is a useful tool for quantifying both normal anatomic structures and lesions. In pulmonology it assesses the pulmonary functional volume and the level of lung involvement. Quantifying chest X-ray and lung CT could be very important in the clinical management of pneumonia [39], [40] as radiological imaging plays a crucial role in evaluating the course of the disease and in choosing the proper therapeutical tactics [41]. *Second*, radiomics attracts the attention of radiologists because of its ability to uncover the characteristics that may have otherwise been misreported [42]. The typical human-driven radiological assessment is subject to bias and interobserver reliability [41]. In contrast to it, AI-assisted tools might be powerful but they do not substitute the judgment of a clinician [6], [43]. In this context, radiomics is an objective tool which might serve to standardise the results of radiological description and turn them into a measurable outcome.

To use the approach efficiently, researchers have to address *the following issues*. The first one is *the repeatability and reproducibility of the radiomics features*. A lot of factors may affect the quality of diagnostic images. The issue becomes an obstacle because of a variance in the technical parameters (e.g., tube currents, reconstruction filters, etc.) across clinics. Presumably, the radiomics approach should also change with regard to the image acquisition settings [44]. For instance, there is evidence that the reconstruction kernel, also referred as the filter, affects the image quality [45], [46]. This is analogous to the effect of the reconstruction settings on the human observer's ability to detect small lesions [47].

The second issue associated with the utility of radiomics is *the necessity to use exclusively valuable features and to avoid information noise from the useless ones* [48].

The last issue to resolve is *standardisation of the imaging features for radiomics analysis*. Researchers suggest that generalization of the prognostic impact of radiomics should be done cautiously [36], [46]. There are certain considerations on how to treat the features that are database-dependent. In a multicenter study on radiomics, the performance was improved by standardising data from different clinics in three distinct manners: min-max normalization, z-score normalization, and whitening from the principle component analysis. The authors suggested that a conversion of CT numbers to electron densities (electron density transformation) is the crucial feature to standardise as it does not depend on scanning protocols [42].

B. HYPERMETABOLIC STATE IN ACUTE RESPIRATORY FAILURE

1) METABOLIC CHANGES IN COMMUNITY ACQUIRED PNEUMONIA

Studies on community-acquired pneumonia (CAP) revealed the pathophysiological mechanisms of pulmonary inflammatory disease. It manifests with metabolic dysregulation caused by the systemic inflammatory response to the infection and detected in the peripheral blood of patients with CAP [49], [50]. This allows predicting the outcomes of the pneumonia from ICU scoring systems such as simplified acute physiology score II and sequential organ failure assessment score [50]. The predictors that comprise the scoring systems are functional variables [e.g., the heart rate (HR), the systolic blood pressure (SBP), blood gas tension, etc.] along with the biochemical findings (BUN, urine output, creatinine, sodium, potassium, bicarbonate, bilirubin) and the hematological estimates (white blood cells, platelets).

By modeling the outcomes, researchers want to improve early accurate diagnosis and provide timely treatment to prevent devastating complications. As age is an important risk factor, these studies have investigated the outcomes of severe pneumonia predominantly in the elderly, while few studies concentrate on the altered metabolic profiles that clearly differentiate the survivors and non-survivors among young patients with CAP. This is why identification of both diagnostics markers and prognostics markers of CAP remains actual with many questions remaining unanswered [49]. For instance, some cases of CAP are not documented microbiologically. Designing a study on specificity of the markers to etiology of CAP is a challenging task.

2) ML-BASED MODELS OF METABOLIC CHANGES IN COVID-19

Although considerable progress has been made in understanding the molecular mechanisms underlying pulmonary infection, a satisfactory prognosis remains limited [43]. The laboratory findings may reflect the disease course and mortality. In a previous study we justified threshold values of biochemical and hematological parameters for proper in-hospital management of patients with COVID-19 [2]. Some researchers also built *univariate predictive models* to forecast the disease severity from the level of D-dimer [51], lymphocyte count and lactate dehydrogenase (LDH) at presentation [52]. Other scientists programmed *multivariate models* based on the level of serum LDH activity, C-reactive protein (CRP), the coefficient of variation of red blood cell (RBC) distribution width (RDW-CV), blood urea nitrogen (BUN), direct bilirubin, lower albumin and age [53]. As the

performance of those models remained low (maximum sensitivity 77.5%, specificity 78.4%), our research group trained a neural network to predict the severe form of COVID-19 which requires a transfer of the patient to the ICU. The resulting accuracy was admissible in the model with the top valuable tests (APTT, CRP, and Fibrinogen: AUC 0.86; CI 0.486 - 0.884) and it improved when all significant laboratory findings were considered (total bilirubin, ALT, AST, D-dimer, APTT, CK, CRP, LDH, troponin, ferritin, fibrinogen; AUC 0.90) [2]. Other authors mentioned a combination of IL-6 and D-dimer (sensitivity 96.4%, specificity 93.3%) [54] and age (hazard ratio=1.67, p=0.024) [55] as the predictors of the disease severity. Several studies revealed the possibility to forecast the COVID-19 mortality from LDG (odds ratio=6.53), CRP (sensitivity 51%, specificity 88%) and markers of coagulation system dysfunction: D-dimer (AUC 0.74) and prothrombine time (AUC 0.64) [51], [56], [57]. The laboratory and morphological findings proved to be more predictive if used in combination [37]. To make the predictive models more accurate, in this study we decided to perform data blending using both radiomics and laboratory findings as predictors.

II. OBJECTIVES

Here we aimed to test the effectiveness of machine learning in eliminating data redundancy of radiomics and reflecting pathophysiologic changes in patients with COVID-19 pneumonia.

Hypothetically, injury to the lung in pneumonia patients should correlate with the individual risk factors (particularly age) and it should accurately depict the disease severity which is commonly assessed with the laboratory findings (e.g., system inflammatory markers, etc.). An efficient way to quantify radiologic findings and measure disease severity is by extracting radiomics features. Radiomic analysis with advanced statistical methods could help to compare follow-up studies, detect diseases worsening and stratify risks thus improving patient management. As reconstruction kernels affect the image quality, the settings of the filters may impact the diagnostic value of CT-scans.

For addressing the research questions, we formulated the following objectives:

1. To quantify the structural changes in computed tomography (CT) images of patients with COVID-19 associated pneumonia and to study the relationship between the radiomics features extracted from lung CT images, demographics and the laboratory markers of disease severity.
2. To find the radiomics and biochemical features reflective of the functional changes in patients with COVID-19.
3. To compare the reconstruction kernels of lung CT images with regard to the predictive potential for identifying the values of the clinical severity markers.

III. MATERIALS AND METHODS

A. STUDY COHORT

We did a retrospective analysis of the data obtained as a standard of primary and secondary care. The study cohort included 605 consecutive patients who were treated from 24 February to 1 July, 2020 in Al Ain Hospital which is an acute care hospital serving Al Ain city (N 634,000) in the United Arab Emirates. The inclusion criteria were as follows: age 18 years or older; inpatient admission; SARS-CoV-2 positive real-time reverse-transcriptase polymerase chain reaction (PCR) from nasopharyngeal swabs only; computed tomography (CT) images. At the beginning of the pandemic, 'National Guidelines for Clinical Management and Treatment of COVID-19' [58] compelled to hospitalize everybody tested positive for SARS-CoV-2, disregarding the disease severity (e.g., the symptoms). This enabled us to explore a unique dataset of early phase examinations. The patients presented with all possible disease forms, from mild to critical (for details see subsection III-B4). As per the Guidelines, the patients underwent a thorough examination with a set of functional (cardiorespiratory data), hematological (total blood count), biochemical (e.g., inflammatory biomarkers of inflammation, oxydative status, disease severity, etc.) and radiological (lung CT) data being collected.

B. METHODS USED

1) ACQUIRING HEMATOLOGICAL, BIOCHEMICAL AND FUNCTIONAL DATA

From the hospital dataset we retrieved the demographic data (age, sex) and the laboratory findings that were determined with automatic hematologic and biochemical analyzers. Oxygen saturation (SpO_2) was measured with pulse oximetry. Physiological parameters of cardiovascular function [e.g., breath (BR) and heart rate (HR), systolic (SBP) and diastolic blood pressure (DBP)] were acquired in a typical way at the time of physical examination of the patient on admission.

2) CT SCANNING

The high-resolution CT scan protocol was as follows: the tube voltage 120kV, the electric current 195mA, the exposure time 0.5s and the slice thickness of 1mm. The scanning range was from lung apex to diaphragm in the axial plane taken at the end inspiration. Before acquiring images for the current study, we checked the Hounsfield unit with a standard water phantom which is used for the quality assurance of the computed tomography scanner. The raw data obtained from each scan were reconstructed with three reconstruction filters (kernels) that vary in smoothness (B30f, B60f and B80f).

3) IMAGE PREPROCESSING

The first part of the solution is to derive a large number of features from the medical images. For this we utilized the U-Net architecture proposed in [59] for segmentation of lung lobes and lesions (the ground-glass opacity, consolidation, and pleural effusion). The batch normalization was added

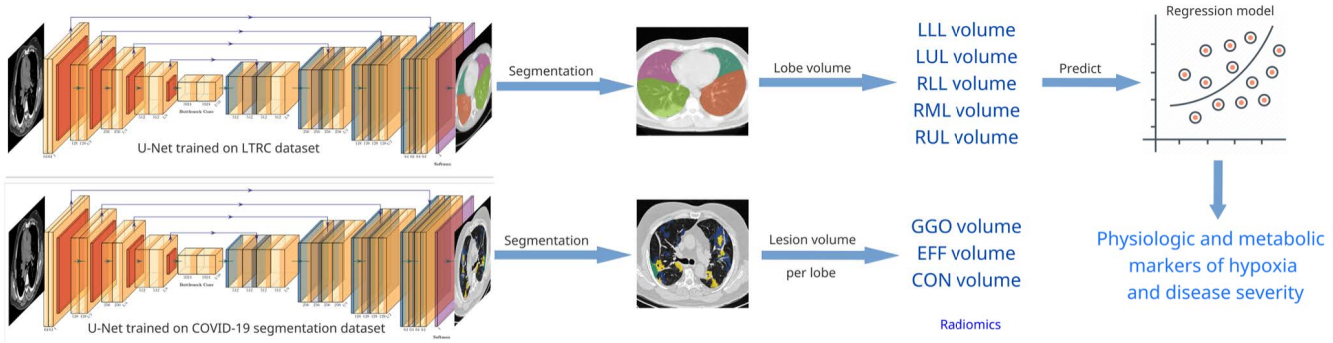


FIGURE 1. **Computational analysis and architecture of U-Net model used to segment anatomic structures and lesions.**

after each convolutional layer of the model [60]. For further details we refer the reader to the original paper of Ronneberger et. al. because we did not change the architecture of the model. The U-Net model is widely used for image segmentation in medicine [61], [62]. It was also trained to delineate the lung lobe boarders on the LTRC, lung lesions - on the COVID-19 CT dataset [63], [64]. We used the models to produce lobe and lesion segmentation masks for the study dataset (see Fig.1).

To collect radiomics data for the entire lung and their lobes, we applied lung masks segmented with the deep learning U-net model trained on large and diverse dataset as described in [63]. Ground glass opacity (ggo_), consolidation (con_), and pleural effusion (eff_) are the most common types of lung lesions in COVID-19. These lesions were segmented with CT Thorax Covid-19 model from MedSeg tool [64]. By multiplying the number of voxels in the mask by the voxel size we received the total lung volume as well as the volumes of the lung lobes and lesions. Then, we utilized fslstats tool from FSL framework to report certain summary statistics for an input 3D/4D image [65]. Specifically, we calculated the mean density of the lungs and their lobes, the standard deviation (_std) and entropy of the density (lungs_entropy). With the same tool we calculated the center of gravity (center_of_gravity), i.e., the point at which the density is evenly dispersed and all sides are in balance. The gravity center has coordinates along x-, y-, and z-axes that correspond to coronal, sagittal and axial reconstruction planes. The characteristics of the density were given in Hounsfield units (HU). Finally, all the volume variables were normalized, or expressed as percentages of the total lung volume (ggo_rate, con_rate, eff_rate). To calculate total lung involvement (pathology_rate) we summed up the percentages of either type of the lesions. The total number of 186 volumetric variables and results achieved with the fslstats tool formed the radiomical findings taken into further analysis. See the illustration of the proposed framework in Fig. 2.

4) METHODOLOGY OF DATA ANALYSIS

Fig. 3 shows the general idea of the proposed structure-function association model.

Working on the first objective, we looked for the radiological markers of COVID-19-associated severity. The total CT involvement score was calculated by summing up the score for involvement of each lung lobe (1 for $< 5\%$, 2 for 5-25%, 3 for 26-49%, 4 for 50-74%, and 5 for $\geq 75\%$). Scoring the severity in this way is convenient in case of visual assessment. As we did the automatized assessment, we calculated the proportion of the total involvement to the entire lung volume. By applying the following criteria, we categorized the cases into 4 classes: mild <5% of pulmonary parenchymal involvement, moderate [5%;25%), severe [25%;50%), and critical $\geq$ 50%. Then we studied the separability of the demographics, laboratory findings and radiomics values on admission to hospital concerning the class of severity. As the variables of the datasets had the non-normal distribution, we utilized non-parametric tests for the analysis.

To address the second objective, we categorized the features into the robust and redundant ones. To filter the features, we resorted to several methods. First, we compared all the variables in the classes with diverse lung impairment. For this we employed a Kruskal-Wallis test as the features were non-normally distributed. Second, we assessed the association between the features and the disease severity level with Pearson's correlation coefficients. Third, we performed correlation feature selection. For this we ranked all the features associated pronouncedly with the oxygen saturation and anion gap concerning the strength of the association, i.e. by the value of correlation coefficient r. These steps allowed us to compute the relevance and redundancy of the features. Data retrieval (see Section III-B3) followed by reduction of redundancy resemble a standard two-fold solution which is typically applied to radiomics [42].

For the third objective, we utilized conventional ML regression models to predict the values of markers of oxygenation, respiratory and cardiovascular function (SpO_2, HCO_3, BR, HR, SBP, DBP, AG, serum potassium and sodium). We used radiomics data as predictors for building the conventional model with the following list of regressors: AdaBoost, Extra Trees, Gradient Boosting, K nearest neighbours, Lasso, Random Forest. Regression models were trained with the 10-fold cross-validation technique.

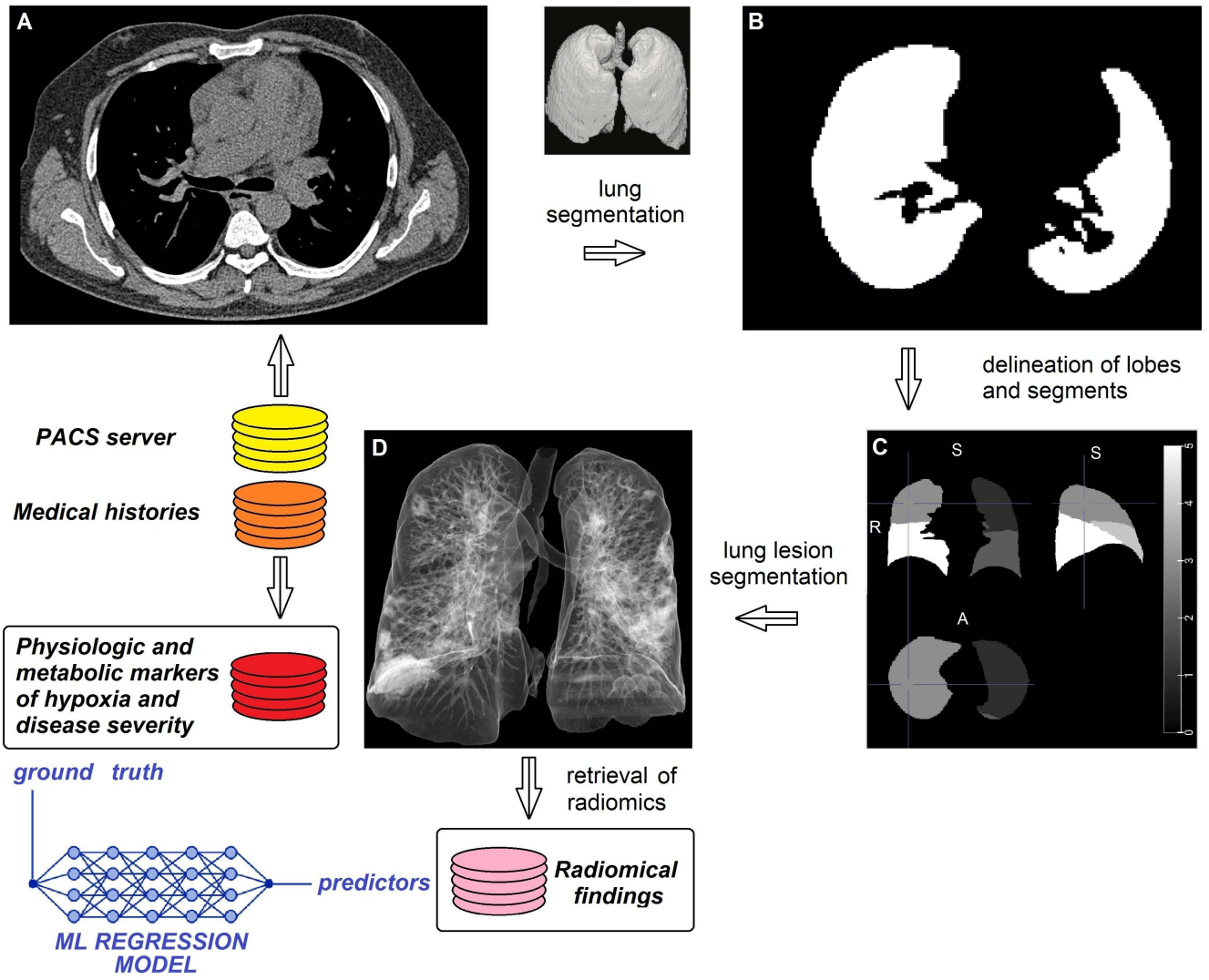


FIGURE 2. Research pipeline: steps of image pre-processing, radiomics retrieval and machine learning: A. Lung CT; B. Lung mask; C. Lobe masks; D. 3D volume rendering of lungs with peripheral lesions.

The default settings were applied to the models. We used the described pipeline while working with all the samples of CT images acquired with distinct reconstruction kernels. To compare the predictive value of the imaging findings, we evaluated the performance of the models for distinct reconstruction kernels. The ratio of mean average error to the range of values was the final performance metric.

C. HARDWARE AND SOFTWARE USED

We used the computational power of the Linux Ubuntu 18.04 Nvidia DGX-1 machine learning server with 40 CPU cores and 8x NVIDIA Tesla V100 GPU with 32 GB memory each, accessed with a web-based multi-user concurrent job scheduling system [66]. The experimental work was conducted using programming language Python and its libraries for DL, Data Processing, and Data visualization, such as tensorflow-gpu v.2.3.1, keras v.2.4.3, SciPy v.1.16.4, NumPy, Pandas, Matplotlib, Seaborn. To automate the deployment of the applications within the software containers, we installed Neurodocker, which wraps up the aforementioned software in a complete file system [67].

IV. RESULTS

A. ASSOCIATION BETWEEN RADIOMICS FEATURES, DEMOGRAPHICS AND MARKERS OF DISEASE SEVERITY

The data presented in Tables 2-3 allow comparison of injury to the lung with demographics, the clinical and laboratory findings in the patients with COVID-19.

Table 2 shows the results of the physical examination and the laboratory findings on admission in patients stratified by lung involvement from the mild to critical level. Table 4 depicts basic radiomics for the entire lung in the same groups of patients. Correlation coefficients reflect the strength of association between the severity of the disease and

TABLE 2. Comparison of patients with regard to severity of COVID-19-associated lung pneumonia: demographics, result of physical examination and laboratory findings on admission.

	Total n=605	Disease severity					Correlation with severity level	
		mild n_1=357(59.01%)	moderate n_2=215(35.54%)	severe n_3=31(5.12%)	critical n_4=2(0.33%)	p_{1-4}	r	p-value
DEMOGRAPHICS								
Age	39.78 [31.26-47.38]	38.28±10.73*	40.91±11.85	47.9±12.23*	61.82±0.54*	$\mathbf{1.03\times10^{-5}}$	0.146	$\mathbf{3.07\times10^{-4}}$
18-39 years	341(56.36%)	218(61.06%)*	113(52.56%)	10(32.26%)*	0(0.0%)			
40-64 years	253(41.82%)	136(38.1%)*	97(45.12%)	18(58.06%)	2(100.0%)			
65+ years	11(1.82%)	3(0.84%)	5(2.33%)	3(9.68%)*	0(0.0%)			
Sex						0.83301	0.027	0.51458
Male	519(85.79%)	303(84.87%)	187(86.98%)	27(87.1%)	2(100.0%)			
Female	86(14.21%)	54(15.13%)	28(13.02%)	4(12.9%)	0(0.0%)			
MARKERS OF OXYGENATION, RESPIRATORY AND CARDIOVASCULAR FUNCTION								
SpO_2	0.99 [0.98-1.0]	0.99±0.01*	0.99±0.01	0.95±0.05*	0.79±0.01*	$\mathbf{3.84\times10^{-6}}$	-0.141	**0.00049**
HCO_3	25.04 [23.7-26.3]	25.24±2.97*	24.91±3.02	23.76±2.46*	22.45±1.15	**0.00672**	-0.098	**0.01550**
Breath rate	18.26 [18.0-18.0]	17.87±1.01*	18.3±2.04	21.81±6.41*	30.5±5.5*	$\mathbf{7.12\times10^{-6}}$	0.118	**0.00361**
AG	17.31 [16.5-18.3]	17.23±2.69	17.29±2.61	18.19±2.02	17.25±0.05	0.14781	0.081	**0.04513**
Potassium	4.03 [3.8-4.3]	4.02±0.36	4.03±0.39	4.13±0.36	3.85±0.15	0.42716	0.054	0.18226
Sodium	139.36 [138.0-141.0]	139.69±2.24*	139.01±2.75*	138.16±3.67	137.5±5.5	**0.02901**	-0.121	**0.02941**
Heart rate	85.47 [76.0-92.0]	83.98±13.79*	87.0±13.95*	91.24±16.45	99.0±19.0	**0.01147**	0.125	**0.00211**
Systolic blood pressure	135.83 [124.0-150.0]	137.26±18.9	136.06±25.77	121.67±40.19*	75.95±0.85*	**0.00139**	-0.062	0.12518
Diastolic blood pressure	81.17 [73.0-92.0]	83.74±13.8*	80.29±19.25	60.41±27.01*	37.05±11.65*	$\mathbf{1.28\times10^{-5}}$	-0.116	$\mathbf{4.26\times10^{-3}}$
DISEASE SEVERITY MARKERS								
CRP	9.89 [0.8-7.7]	5.6±15.23*	13.2±24.03*	36.68±58.39*	5.25±4.65	$\mathbf{9.24\times10^{-8}}$	0.239	$\mathbf{2.51\times10^{-9}}$
D-dimer	0.37 [0.19-0.37]	0.35±0.43*	0.37±0.28	0.63±0.55*	0.37±0.0	$\mathbf{2.51\times10^{-6}}$	0.175	$\mathbf{1.57\times10^{-5}}$
Ferritin	360.79 [136.0-398.0]	280.66±296.27*	441.58±559.44*	728.25±669.03*	284.0±208.0	$\mathbf{1.52\times10^{-6}}$	0.214	$\mathbf{1.07\times10^{-7}}$
ENZYMES								
LDH	227.02 [185.0-236.0]	211.18±48.92*	244.62±102.65*	285.04±97.44*	263.51±36.49	$\mathbf{7.86\times10^{-9}}$	0.228	$\mathbf{1.48\times10^{-8}}$
Alkaline phosphatase	80.64 [67.0-83.0]	79.49±31.49	82.32±55.12	82.2±33.12	81.82±1.18	0.77191	0.011	0.78927
Amylase	79.83 [56.0-87.0]	81.32±60.4	75.9±33.49	88.8±53.5	96.42±16.58	0.45871	-0.023	0.56666
G6PD	10.9 [10.9-11.2]	10.9±2.5	10.93±2.12	10.66±1.8	10.9±0.0	0.63044	-0.041	0.31217
Lipase	41.26 [26.0-43.0]	36.29±15.8*	44.76±52.95	70.81±86.86*	94.63±53.37	**0.00036**	0.137	$\mathbf{7.38\times10^{-4}}$
BIOCHEMICAL SUBSTRATES AND ELECTROLYTES								
Total protein	78.08 [76.0-80.0]	78.46±4.24*	77.5±4.55*	77.79±2.36	78.08±0.0	0.15004	-0.081	**0.04534**
Albumin	39.94 [39.0-42.0]	40.99±3.26*	38.84±4.8*	35.6±5.64*	37.97±1.97	4.13×10^{-12}	-0.286	$\mathbf{8.08\times10^{-13}}$
Creatinine	81.11 [66.0-87.0]	77.14±22.11	81.78±58.52	123.42±139.52	61.0±17.0	0.18894	0.066	0.10570
Urea	4.09 [3.0-4.6]	3.91±1.67	4.16±2.21	5.6±3.45*	4.2±0.4	**0.00577**	0.084	**0.03808**
Uric acid	306.38 [260.0-346.0]	306.74±78.61	307.78±82.72	295.85±70.29	255.5±13.5	0.41808	-0.021	0.61373
Total bilirubin	8.99 [6.2-9.7]	9.08±5.91	8.78±4.11	9.34±4.77	8.99±0.0	0.85357	0.015	0.70453
Direct bilirubin	3.54 [2.5-3.6]	3.61±4.19	3.4±1.45	3.79±1.53	3.54±0.0	0.50329	0.037	0.36959
Glucose random	6.41 [5.13-6.41]	6.23±2.18*	6.54±2.35*	7.65±3.21*	5.69±0.72	**0.00047**	0.158	$\mathbf{9.83\times10^{-5}}$
Calcium	2.35 [2.29-2.41]	2.36±0.09*	2.34±0.11	2.31±0.1*	2.29±0.05	**0.00482**	-0.126	**0.00196**
Magnesium	0.84 [0.8-0.89]	0.84±0.07	0.84±0.08	0.84±0.06	0.85±0.0	0.99048	-0.000	0.99590
Phosphorus	1.12 [0.97-1.24]	1.13±0.22	1.11±0.23	1.08±0.22	1.15±0.04	0.73054	-0.041	0.30984
HEMATOLOGIC FINDINGS								
HGB	144.04 [135.0-154.0]	145.88±14.49*	143.07±17.92	131.19±20.82*	118.5±3.5*	$\mathbf{9.47\times10^{-5}}$	-0.103	**0.01138**
HCT	0.43 [0.41-0.45]	0.43±0.04*	0.42±0.05	0.39±0.06*	0.36±0.01*	$\mathbf{1.70\times10^{-5}}$	-0.124	**0.02185**
RBC	4.5 [4.5-5.39]	4.49±1.78	4.65±1.57	3.91±1.8*	0.16±0.05*	**0.00288**	-0.027	0.50686
RDW CV	1.85 [0.12-0.14]	1.78±4.61	1.71±4.86	2.5±5.53*	18.4±1.5*	**0.00566**	0.025	0.54592
Reticulocyte count	61.2 [41.9-72.1]	62.98±27.0*	58.23±25.35*	57.15±24.72	125.7±64.5	0.07416	-0.097	**0.01656**
Reticulocyte %	0.13 [0.01-0.02]	0.13±0.52	0.11±0.48	0.15±0.45	2.47±2.33*	0.07468	-0.008	0.84975
IRF	9.44 [6.0-11.0]	9.3±5.23	9.4±4.84	10.7±4.21	18.22±8.78	0.10187	0.053	0.19248
MCH	338.14 [331.0-346.0]	337.91±11.76	338.53±13.42	338.45±9.44	332.5±4.5	0.52274	0.046	0.25972
MCHC	338.14 [331.0-346.0]	337.91±11.76	338.53±13.42	338.45±9.44	332.5±4.5	0.52274	0.046	0.25972
MCV	83.46 [81.1-86.6]	83.73±5.12	83.04±5.97	83.17±6.14	84.65±5.35	0.91202	-0.027	0.51216
MPV	10.46 [9.8-11.0]	10.43±0.91	10.46±0.89	10.72±1.14	10.55±0.15	0.44909	0.044	0.28024
Platelet	260.17 [210.0-299.0]	259.96±71.73	261.84±80.11	256.74±92.05	171.0±23.0	0.26810	-0.003	0.93571
WBC	6.87 [5.6-7.9]	6.89±2.03	6.83±1.95	6.81±1.61	6.87±0.0	0.99653	0.003	0.93569
Lymphocyte count	2.17 [1.59-2.66]	2.26±0.8*	2.05±0.77*	1.88±0.82	1.8±0.2	**0.00555**	-0.137	**0.00076**
Lymphocyte %	32.42 [24.6-39.2]	33.8±10.36*	30.85±9.95*	28.09±11.62*	21.1±2.7	**0.00061**	-0.144	**0.00037**
Basophil count	0.03 [0.02-0.04]	0.03±0.02	0.03±0.02	0.03±0.02	0.04±0.0	0.14678	-0.083	**0.04158**
Basophil %	0.43 [0.3-0.6]	0.44±0.24*	0.41±0.25*	0.4±0.21	0.5±0.0	0.08453	-0.098	**0.01541**
Eosinophil count	0.17 [0.05-0.21]	0.17±0.22	0.18±0.25	0.15±0.17	0.36±0.11	0.16144	-0.049	0.22418
Eosinophil %	2.43 [0.8-3.0]	2.42±2.95	2.49±3.07	2.01±2.13	4.25±1.35	0.19844	-0.051	0.21386
Monocyte count	0.59 [0.44-0.7]	0.58±0.22	0.6±0.23	0.58±0.2	0.7±0.05	0.50062	0.016	0.69938
Monocyte %	8.77 [6.7-10.2]	8.66±2.68	9.0±3.13	8.53±2.54	8.1±0.5	0.75119	0.031	0.45034
Neutrophil count	3.99 [2.79-4.74]	3.92±1.82	4.06±2.07	4.28±1.63	5.64±0.4	0.15882	0.043	0.29655
Neutrophil %	55.94 [48.1-64.5]	54.66±11.58*	57.25±11.49	60.97±13.5*	66.05±3.55	**0.00503**	0.120	**0.00313**

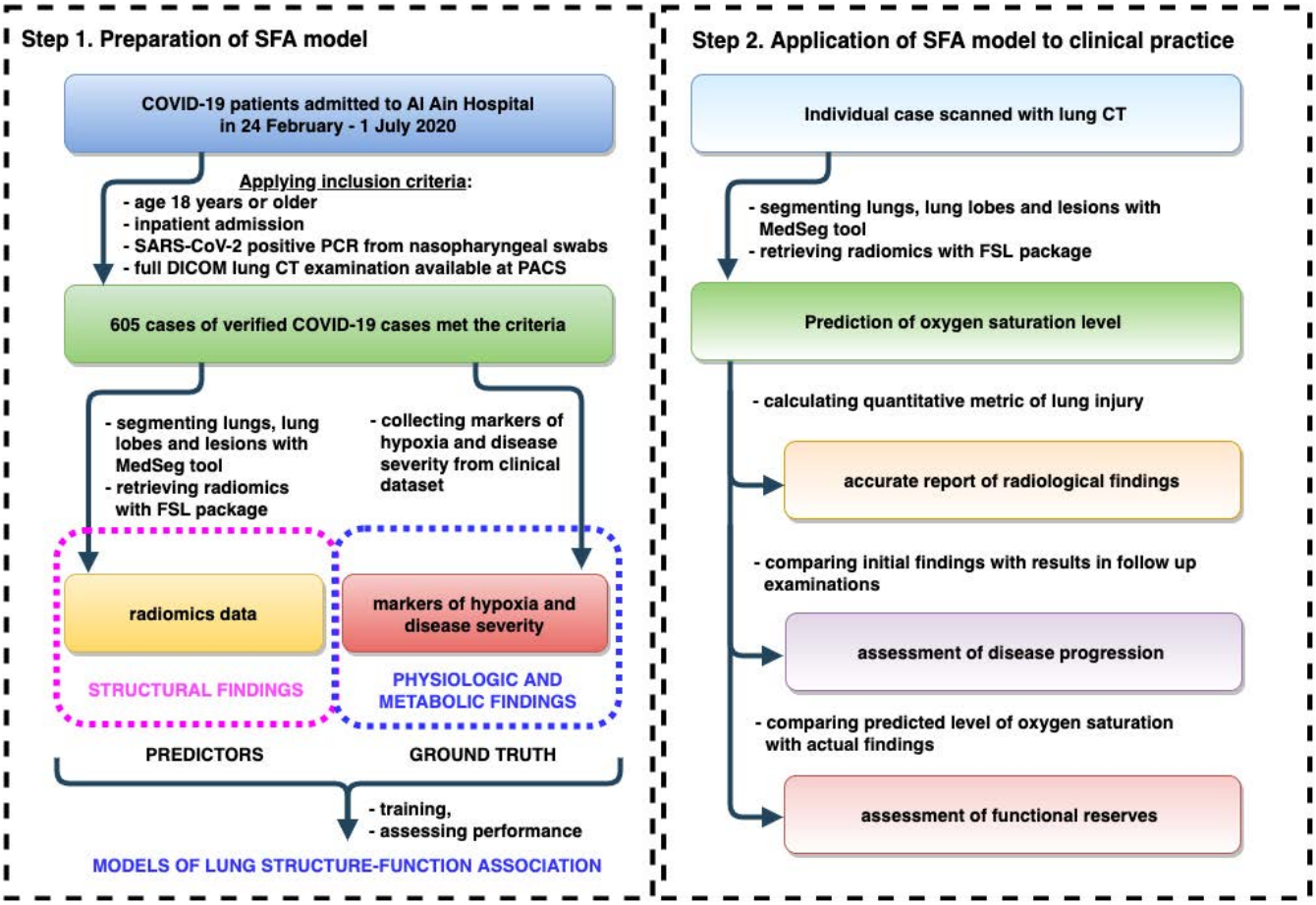


FIGURE 3. Preparation and application of the proposed structure-function association model to clinical practice.

TABLE 3. Comparison of patients with regard to radiomics features calculated for lobes: LLL, LUL, RLL, RML and RUL.

	Mean	CI	Lung lobes					
			LLL	LUL	RLL	RML	RUL	p-value
volume liter	0.67	[0.45-0.84]	0.8±0.25*	0.76±0.3*	0.8±0.32*	0.31±0.1*	0.67±0.19*	$\mathbf{7.20396\times10^{-291}}$
volume %	19.61	[17.18-23.87]	23.79±2.51*	21.92±3.01*	23.15±3.68*	9.26±1.76*	19.94±2.72*	**0**
entropy	0.1	[0.0-0.17]	0.14±0.15*	0.13±0.16*	0.15±0.17*	0.01±0.0*	0.06±0.1*	$\mathbf{2.26539\times10^{-170}}$
mean	-13.93	[-17.78–8.82]	-17.19±5.63*	-15.25±6.91*	-16.18±7.48*	-6.83±2.36*	-14.21±4.24*	$\mathbf{2.76344\times10^{-252}}$
std	101.08	[81.08-119.33]	115.03±21.37*	104.45±27.37*	107.52±29.12*	73.64±14.33*	104.74±17.65*	$\mathbf{1.21905\times10^{-205}}$
PATHOLOGICAL FINDINGS								
pathology_rate	5.92	[0.08-5.51]	2.3±6.25*	9.75±14.32*	13.17±16.61*	1.66±5.08*	2.7±6.48*	$\mathbf{4.92246\times10^{-162}}$
ggo_lobe_rate	5.33	[0.06-4.76]	1.93±5.28*	8.81±12.82*	12.19±15.18*	1.34±4.14*	2.36±5.58*	$\mathbf{1.03566\times10^{-170}}$
ggo_rate	1.06	[0.01-0.97]	0.45±1.15*	1.73±2.4*	2.48±2.97*	0.11±0.32*	0.51±1.25*	$\mathbf{1.57244\times10^{-212}}$
con_lobe_rate	0.59	[0.01-0.23]	0.37±1.39	0.94±2.57*	0.98±3.06*	0.32±1.48*	0.34±1.45*	$\mathbf{3.04467\times10^{-79}}$
con_rate	0.1	[0.0-0.05]	0.09±0.31	0.17±0.42*	0.17±0.5*	0.02±0.1*	0.07±0.3*	$\mathbf{6.44126\times10^{-115}}$

the aforementioned parameters. Table 3 compares radiomics for distinct lung lobes (LLL, LUL, RLL, RML and RUL) with regard to the severity level.

Fig. 4 illustrates an evident association between physiological (e.g. breath and heart rate, systolic and diastolic pressure, etc.) and biochemical markers of oxygenation status and disease severity (SpO_2, HCO_3, potassium, AG). Age as an individual risk factor of worsening correlates positively with the level of CRP ($r = 0.22$; $p < 0.05$) and the blood saturation level ($r = 0.27$; $p < 0.05$).

A negative correlation ($r = -0.23$; $p < 0.05$) between the levels of oxygen saturation and CRP justifies each of the

TABLE 4. Comparison of patients with regard to severity of COVID-19-associated lung pneumonia: radiomics features.

	Total n=605		Disease severity					Correlation with severity level	
			mild n_1=357(59.01%)	**moderate** n_2=215(35.54%)	**severe** n_3=31(5.12%)	**critical** n_4=2(0.33%)	p_{1-4}	r	p-value
LUNG INVOLVEMENT									
Pathology_rate	6.71	[0.81-9.24]	1.55±1.39*	10.93±4.95*	33.71±6.46*	56.66±5.63*	$\mathbf{8.43\times10^{-98}}$	0.843	$\mathbf{1.8\times10^{-164}}$
ggo_rate	5.43	[0.45-7.87]	1.14±1.21*	9.23±4.52*	26.48±7.21*	35.48±1.96*	$\mathbf{2.04\times10^{-94}}$	0.828	$\mathbf{9.1\times10^{-154}}$
ggo_std	32.07	[13.38-47.59]	17.88±11.14*	49.05±10.49*	73.95±10.29*	89.95±8.9*	$\mathbf{3.09\times10^{-90}}$	0.809	$\mathbf{2.0\times10^{-141}}$
con_rate	1.28	[0.14-1.08]	0.41±0.56*	1.69±1.98*	7.23±7.62*	21.18±3.67*	$\mathbf{4.87\times10^{-47}}$	0.577	$\mathbf{6.0\times10^{-55}}$
con_std	4.89	[1.9-5.82]	3.01±2.29*	6.31±4.07*	14.85±9.53*	33.97±4.56*	$\mathbf{4.77\times10^{-40}}$	0.530	$\mathbf{4.9\times10^{-45}}$
eff_rate	0.0	[0.0-0.0]	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.975783	-0.001	**0.98744**
CT score	6.77	[5.0-7.0]	5.33±0.6*	8.03±1.54*	13.81±1.75*	18.5±0.5*	$\mathbf{1.40\times10^{-97}}$	0.841	$\mathbf{9.7\times10^{-163}}$
lungs_entropy	0.71	[0.69-0.74]	0.71±0.03	0.71±0.04	0.71±0.07	0.77±0.06	0.18023	0.061	0.13512
centre_of_gravity_x	255.71	[255.24-256.09]	255.77±0.89	255.62±0.63	255.64±0.75	255.43±0.39	0.47918	-0.053	0.19238
centre_of_gravity_y	256.68	[255.49-257.7]	256.92±2.11*	256.38±1.57*	255.94±1.01*	256.06±1.04	**0.00151**	-0.150	**0.00021**
centre_of_gravity_z	154.62	[143.66-165.03]	157.69±14.38*	150.51±16.43*	148.18±14.32*	149.28±18.2	$\mathbf{4.84\times10^{-7}}$	-0.227	$\mathbf{1.64\times10^{-8}}$
VOLUMES OF LUNGS AND THEIR LOBES									
Lung volume, liters	3.4	[2.62-3.96]	3.83±1.04*	2.81±0.7*	2.63±0.68*	2.77±0.17	$\mathbf{4.86\times10^{-36}}$	-0.515	$\mathbf{3.36\times10^{-42}}$
LLL_volume	0.8	[0.63-0.94]	0.89±0.24*	0.69±0.2*	0.62±0.21*	0.59±0.16	$\mathbf{1.21\times10^{-26}}$	-0.440	$\mathbf{4.72\times10^{-30}}$
LUL_volume	0.76	[0.54-0.92]	0.89±0.3*	0.58±0.17*	0.46±0.11*	0.47±0.02	$\mathbf{1.73\times10^{-45}}$	-0.575	$\mathbf{1.66\times10^{-54}}$
RUL_volume	0.67	[0.53-0.78]	0.72±0.18*	0.58±0.16*	0.62±0.18	0.6±0.08	$\mathbf{2.60\times10^{-19}}$	-0.362	$\mathbf{3.19\times10^{-20}}$
RML_volume	0.31	[0.24-0.37]	0.34±0.1*	0.27±0.08*	0.22±0.08*	0.16±0.04*	$\mathbf{5.90\times10^{-20}}$	-0.369	$\mathbf{6.32\times10^{-21}}$
RLL_volume	0.8	[0.56-0.98]	0.95±0.31*	0.61±0.18*	0.45±0.14*	0.56±0.04	$\mathbf{1.82\times10^{-51}}$	-0.617	$\mathbf{1.23\times10^{-64}}$
LOBE INVOLVEMENT									
LLL_ggo_lobe_rate	1.93	[0.05-1.01]	0.21±0.39*	2.36±3.41*	16.26±10.08*	42.08±15.21*	$\mathbf{3.94\times10^{-69}}$	0.703	$\mathbf{2.07\times10^{-91}}$
LLL_con_lobe_rate	0.37	[0.01-0.14]	0.06±0.16*	0.41±0.93*	2.96±3.98*	11.43±0.15*	$\mathbf{1.32\times10^{-43}}$	0.551	$\mathbf{3.12\times10^{-49}}$
LUL_ggo_lobe_rate	8.81	[0.61-12.45]	1.78±2.48*	14.87±10.0*	45.41±13.98*	45.44±1.8*	$\mathbf{2.74\times10^{-84}}$	0.782	$\mathbf{5.22\times10^{-126}}$
LUL_con_lobe_rate	0.94	[0.04-0.52]	0.19±0.36*	1.18±1.93*	6.96±6.74*	14.75±1.79*	$\mathbf{4.27\times10^{-46}}$	0.570	$\mathbf{2.09\times10^{-53}}$
RLL_ggo_lobe_rate	12.19	[0.55-19.69]	2.57±3.42*	22.7±12.14*	48.21±12.61*	40.45±6.34*	$\mathbf{5.11\times10^{-89}}$	0.807	$\mathbf{7.3\times10^{-140}}$
RLL_con_lobe_rate	0.98	[0.01-0.45]	0.15±0.41*	1.28±2.48*	7.26±8.05*	20.57±2.43*	$\mathbf{2.14\times10^{-50}}$	0.599	$\mathbf{4.60\times10^{-60}}$
RML_ggo_lobe_rate	1.34	[0.0-0.46]	0.13±0.4*	1.53±2.75*	12.67±10.17*	22.21±7.46*	$\mathbf{2.93\times10^{-50}}$	0.591	$\mathbf{3.79\times10^{-58}}$
RML_con_lobe_rate	0.32	[0.0-0.07]	0.04±0.21*	0.33±0.89*	2.54±4.28*	14.1±2.93*	$\mathbf{3.83\times10^{-39}}$	0.521	$\mathbf{2.48\times10^{-43}}$
RUL_ggo_lobe_rate	2.36	[0.05-1.91]	0.38±0.72*	3.17±3.78*	17.14±11.84*	40.42±1.58*	$\mathbf{1.92\times10^{-68}}$	0.700	$\mathbf{3.10\times10^{-90}}$
RUL_con_lobe_rate	0.34	[0.01-0.1]	0.05±0.21*	0.38±0.98*	2.54±4.35*	12.06±0.85*	$\mathbf{1.30\times10^{-40}}$	0.532	$\mathbf{1.58\times10^{-45}}$

variables as functional and biochemical indicators of disease severity. The physiological markers of COVID-19 severity (heart and breath rate) were also positively associated with the CRP level ($r = 0.21$ and 0.42; $p < 0.05$).

The percentage of the lung involvement strongly correlated with the lung CT score ($r = 0.97$; $p < 0.001$). Both metrics of the lung structural changes are intimately associated with most of the physiological and biochemical markers of oxygen deprivation. However, the association of the oxygen saturation level with the percentage of the lung involvement was slightly stronger than with the CT score ($r = -0.53$, $p < 0.001$ vs. $r = -0.52$, $p < 0.001$).

The total CT score is a semi-quantitative score of pulmonary involvement. It rates the percentage of each of the five lobes that is injured: < 5%, 5-25%, 26-49%, 50-75% and > 75% involvement. Its validity for the assessment of COVID-19 severity has been already shown in previous studies [68]. Meanwhile the parenchymal involvement percentage is a purely quantitative; it provides a more accurate assessment of cases. Still, it is less available in real clinical settings where human-driven visual scoring remains in use due to its simplicity. The percentage of the pulmonary parenchymal involvement outperforms the preexisting biomarker of the lung lesions – the lung CT score – in reflecting pathophysiologic changes.

B. RADIOMICS AND BIOCHEMICAL FEATURES REFLECTIVE OF FUNCTIONAL CHANGES IN PATIENTS WITH COVID-19

To build predictive models of functional parameters, we started from feature selection from the list of physiological, biochemical and radiomics findings. For this we looked at the separability of classes with regard to the values of the supposed predictors and analyzed correlations (Tables 2-3). Fig. 5-6 show correlation feature selection for predicting AG and blood oxygen saturation from the dataset of demographics data, laboratory and radiomics findings. There we ranked all features significantly correlated with the oxygen saturation and anion gap concerning the strength of the association, i.e. by the value of correlation coefficient r.

Fig. 5 shows noteworthy (p<0.05) features ranked with regard to the strength of their association with AG. The features that are positively correlated with AG attribute to protein and heme metabolism (total protein, albumin, urea, creatinine, total and direct bilirubin). Another set of the informative predictors is reflective of the total count of leukocytes and subtypes (neutrophil, monocytes, lymphocytes). A single radiomical finding of the lung lobe involvement (RMO_ggo_lobe_rate) stays in the list of valuable features.

Fig. 6 depicts correlation feature selection for the model predicting the SpO_2 level. The information gain of radiomics

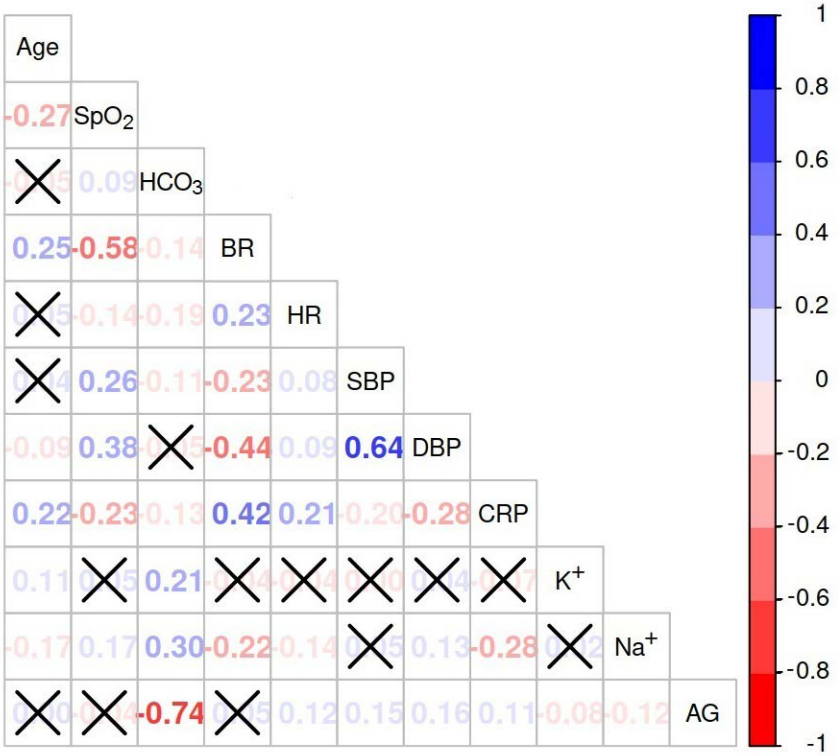


FIGURE 4. Associations between age and features used as outcomes of machine learning model. Significant associations ($p < 0.05$) are presented with values of Spearman's rank correlation coefficient. Coefficients printed in blue stand for direct correlation, in red – for inverse one. Heatmap encodes p-value in color.

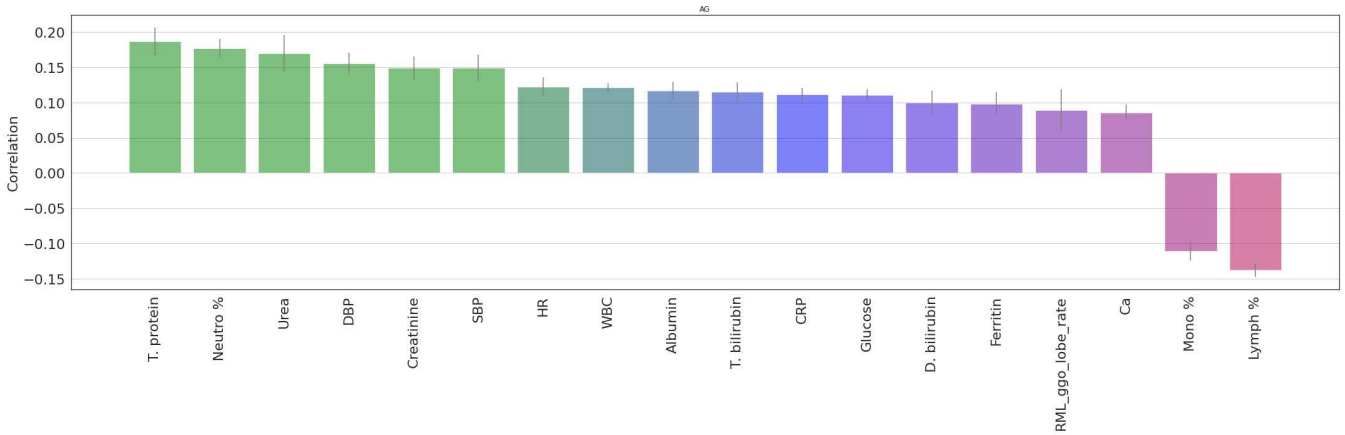


FIGURE 5. Predictors ranked by correlation feature selection method employing 10-fold cross-validation technique to predict AG. Only significant correlations displayed.

variables into the final prediction is considerably higher in this model compared to the model for AG. The involvement of the total lung and its specific lobes correlated negatively with the oxygen saturation. Contrarily, there is a pronounced positive correlation between the lung lobes volumes and SpO_2. As in the previous model, the markers of protein metabolism showed a notable positive association with SpO_2. The same is true for the RBC count, HCT, the level of hemoglobin and percentage of lymphocytes. The age, the concentration of CRP and LDH activity exert negative correlation with the level of oxygenation.

C. COMPARISON OF RECONSTRUCTION KERNELS BY PREDICTIVE POTENTIAL TO IDENTIFY CLINICAL SEVERITY FROM DIAGNOSTIC IMAGES

We studied whether the settings of the reconstruction kernels influence the radiomical findings in terms of their possibility

to reflect the clinical status. For each reconstruction kernel we calculated radiomics and used three sets of radiomical data to predict the physiological (e.g., SpO_2) and biochemical markers of disease severity. Fig. 7 presents the accuracy of models built on radiomics for each reconstruction kernel. For most of the models the least accurate prediction is observed once images are acquired with B60f kernel. The data suggests employment of B30f kernel for identification of the clinical status from CT images. However, the accuracy differs insignificantly among disparate reconstruction kernels.

Table 5 shows the performance of distinct regressors used in AI models predicting the markers of oxygenation and disease severity from radiomics taken separately and in conjunction with the clinical features. To calculate radiomics, we used CT scans reconstructed in B30f, B60f, and B80f kernels. The most accurate is the prediction made by AdaBoost regressor from all the features with radiomics based on CT images acquired with B30f reconstruction kernels (the ratio of MAE to range of values is $6.454 \pm 3.715\%$). Random Forest regressor demonstrates the top prediction from radiomics features ($7.069 \pm 4.17\%$).

V. DISCUSSION

A. ASSOCIATIONS OF RADIOMICS FEATURES WITH DEMOGRAPHICS, BIOCHEMICAL AND HEMATOLOGIC MARKERS OF DISEASE SEVERITY

The CT imaging features of COVID-19 pneumonia resemble various conditions such as organizing pneumonia or inflammatory lung processes [69]. It results in respiratory failure because of the organized buds of granulation tissue. The tissue obstructs the alveolar lumen and bronchioles thus causing respiratory failure [70]. Moreover, these pathological findings can be steady and form the so-called secondary organizing pneumonia which accounts for persistent symptoms even after an acute phase [71].

1) RADIOMICS

Acute respiratory distress syndrome is a form of hypoxemic respiratory failure characterized by lung tissue edema and injury, inflammatory responses and compromised gas exchange [72]. As seen from our data, blood oxygenation correlates with the level of lung injury. This justifies that radiomics approach used in the study assesses patient's worsening reliably (see Table 4 which reflects an evident association between major radiomical findings and the disease severity). From Table 3, the injury to the right lower lobe (both of ground glass opacity and consolidation types) is prevalent among other parts of the lung. This is compliant with an early study on the chest computed tomography findings in COVID-19 pneumonia. The study marked the following common CT features of this type of pneumonia: the involvement of the lower lobe and multiple types of the opacity (ground-glass, ground-glass and consolidation, and consolidation alone) [73]. Interestingly, radiologic features of SARS-Cov-2-associated pneumonia resemble settings of non-viral pneumonia better than the ones for non–SARS-CoV-2 viruses [74].

2) EPITHELIAL DESTRUCTION AND PROTEIN METABOLISM

In our study, there was a marked association between the lung injury and the markers of protein metabolism. This happens because an increase in protein permeability across the endothelial and epithelial barriers of the lung is the most fundamental early physiologic characteristic of acute lung injury [75]. Macrophage activation, surfactant dysfunction, and epithelial destruction follow compromised gas exchange. They result in injury to both the vascular endothelium and the alveolar epithelium [72].

3) DEMOGRAPHICS, PHYSIOLOGICAL AND BIOCHEMICAL MARKERS OF DISEASE SEVERITY

From the comparison of patients grouped in four categories according to the disease severity, there is a pronounced variability in age, biochemical markers of inflammation and coagulation (CRP, D-dimer), oxygenation (SpO_2, LDH), respiratory and cardiovascular function (breath and heart rate, blood pressure) among the groups (see Table 2).

4) WHITE BLOOD CELLS

In pneumonia, inflammatory cells could be activated to produce a large number of mediators in the early stage of the disease [72]. In our study, the percentage of neutrophils in the total WBC count increases with advancing severity score. Parallel to this, a low lymphocyte count and percentage reflect weakening of the immune response with a rise in the severity level. These findings are compliant with the data of other authors who have addressed lymphocytopenia as a marker of the disease severity in COVID-19 [76]. Apart from marked lymphopenia at admission to hospital, nonsurvivors developed more severe lymphopenia over time [77]. So, the hematologic findings worsen parallel to the results of the radiologic assessment.

5) RED BLOOD CELLS

From our data, a steady decrease in the level of hemoglobin, hematocrit and RBC count is accompanied by an increase in the concentration of ferritin while disease worsening. The prevalence of anemia in a non-severe case of pneumonia is not as well-studied as in severe ones. Research on community-acquired pneumonia reported anemia increasing with illness severity and being more common in females, patients with comorbidities and poor outcomes [78]. Our findings proved that this is also true for atypical pneumonia caused by SARS-CoV-2. Alternatively, preexisting anemia can be a risk factor of high incidence or severity of pneumonia. A study that tested the hypothesis did not show association of the likelihood of developing pneumococcal pneumonia either with the frequency or with the severity of anemia [79].

TABLE 5. Different regressors trained on radiomics and clinical features: model performance in terms of MAE.

Data		SpO_2	BR	HCO_3	CRP	K	Na	AG	HR	SBP	DBP	Mean±SD
Clinical features												
AdaBoost	MAE	0.0118	1.6744	11.1386	0.2768	1.3102	1.3340	1.0784	10.3752	16.7978	12.3278	
	MAE/range, %	5.3600	3.2400	4.8300	11.5300	7.2800	2.7300	3.5900	12.9700	7.8100	9.7800	**6.593±3.654**
Extra Trees	MAE	0.0124	1.6804	8.7902	0.2766	1.3130	1.3384	1.1480	10.7192	16.9672	12.2706	
	MAE/range, %	5.6400	3.2500	3.8200	11.5200	7.2900	2.7400	3.8300	13.4000	7.8900	9.7400	6.598±3.792
Gradient Boosting	MAE	0.0128	1.8202	8.5810	0.2852	1.3518	1.4546	1.2720	10.7980	17.2286	12.6708	
	MAE/range, %	5.8200	3.5200	3.7200	11.8800	7.5100	2.9800	4.2400	13.5000	8.0100	10.0600	6.798±3.792
K nearest neighbours	MAE	0.0126	1.8398	9.1512	0.3088	1.8432	1.4842	1.2022	11.5154	18.5290	13.8730	
	MAE/range, %	5.7300	3.5600	3.9700	12.8700	10.2400	3.0400	4.0100	14.3900	8.6200	11.0100	7.381±4.303
Lasso	MAE	0.0120	1.7200	10.3056	0.2762	1.2650	1.3960	1.1116	10.6014	17.0726	12.8734	
	MAE/range, %	5.4500	3.3300	4.4700	11.5100	7.0300	2.8600	3.7100	13.2500	7.9400	10.2200	6.617±3.695
Random Forest	MAE	0.0122	1.8668	9.0690	0.2758	1.3132	1.4840	1.1598	10.4192	16.4254	12.4160	
	MAE/range, %	5.5500	3.6100	3.9400	11.4900	7.3000	3.0400	3.8700	13.0200	7.6400	9.8500	6.607±3.605
Radiomics from B80f kernel												
AdaBoost	MAE	0.0104	1.8154	11.7398	0.2972	1.9506	1.4354	1.0096	11.1048	16.5130	12.6728	
	MAE/range, %	4.7300	3.5100	5.1000	12.3800	10.8400	2.9400	3.3700	13.8800	7.6800	10.0600	7.159± 4.211
Extra Trees	MAE	0.0108	1.8996	10.3136	0.3034	2.0404	1.5098	1.0656	11.1240	17.1888	12.5274	
	MAE/range, %	4.9100	3.6700	4.4800	12.6400	11.3400	3.0900	3.5500	13.9000	7.9900	9.9400	7.286±4.294
Gradient Boosting	MAE	0.0110	1.9316	10.6936	0.3172	2.0826	1.5094	1.1440	11.4104	16.7202	13.2366	
	MAE/range, %	5.0000	3.7400	4.6400	13.2200	11.5700	3.0900	3.8100	14.2600	7.7800	10.5100	7.457±4.425
K nearest neighbours	MAE	0.0114	2.0114	9.8888	0.3218	2.0548	1.5974	1.1330	11.7048	17.9420	13.7742	
	MAE/range, %	5.1800	3.8900	4.2900	13.4100	11.4200	3.2700	3.7800	14.6300	8.3500	10.9300	7.58±4.503
Lasso	MAE	0.0116	1.7900	11.1626	0.2880	1.9328	1.3798	1.0232	10.9810	17.3352	13.0746	
	MAE/range, %	5.2700	3.4600	4.8400	12.0000	10.7400	2.8300	3.4100	13.7300	8.0600	10.3800	7.149±4.118
Random Forest	MAE	0.0104	1.8734	10.2646	0.2918	1.9284	1.4554	1.0224	11.0346	16.6922	12.7518	
	MAE/range, %	4.7300	3.6200	4.4600	12.1600	10.7100	2.9800	3.4100	13.7900	7.7600	10.1200	**7.069±4.17**
Clinical features and radiomics from B30f kernel												
AdaBoost	MAE	0.0102	1.6556	10.8264	0.2790	1.3008	1.3216	1.0296	10.3198	16.4668	12.0938	
	MAE/range, %	4.6400	3.2000	4.7000	11.6200	7.2300	2.7100	3.4300	12.9000	7.6600	9.6000	**6.454±3.715**
Extra Trees	MAE	0.0106	1.7096	8.9806	0.2846	1.3400	1.3776	1.0424	10.6966	16.6514	11.9498	
	MAE/range, %	4.8200	3.3100	3.9000	11.8600	7.4400	2.8200	3.4700	13.3700	7.7400	9.4800	6.526±3.886
Gradient Boosting	MAE	0.0110	1.7794	8.7856	0.2852	1.3778	1.4538	1.1086	11.0134	16.4332	12.6130	
	MAE/range, %	5.0000	3.4400	3.8100	11.8800	7.6500	2.9800	3.7000	13.7700	7.6400	10.0100	6.652±3.926
K nearest neighbours	MAE	0.0116	1.9962	9.6696	0.3248	2.0858	1.6412	1.1222	12.0806	18.3518	14.1906	
	MAE/range, %	5.2700	3.8600	4.2000	13.5300	11.5900	3.3600	3.7400	15.1000	8.5400	11.2600	7.688±4.637
Lasso	MAE	0.0112	1.8090	10.8504	0.2822	1.3098	1.3840	1.0376	11.0426	17.2472	13.0260	
	MAE/range, %	5.0900	3.5000	4.7100	11.7600	7.2800	2.8400	3.4600	13.8000	8.0200	10.3400	6.718±3.877
Random Forest	MAE	0.0104	1.8878	8.9728	0.2770	1.3448	1.5350	1.0222	10.6512	16.4722	12.3736	
	MAE/range, %	4.7300	3.6500	3.8900	11.5400	7.4700	3.1500	3.4100	13.3100	7.6600	9.8200	6.534±3.757
Clinical features and radiomics from B60f kernel												
AdaBoost	MAE	0.0108	1.6860	10.7232	0.2838	1.3216	1.3368	1.0626	10.3178	16.4498	12.6708	
	MAE/range, %	4.9100	3.2600	4.6500	11.8200	7.3400	2.7400	3.5400	12.9000	7.6500	10.0600	**6.534±3.719**
Extra Trees	MAE	0.0108	1.7232	8.7964	0.2850	1.3182	1.3818	1.0606	10.8708	16.6854	12.4354	
	MAE/range, %	4.9100	3.3300	3.8200	11.8800	7.3200	2.8300	3.5400	13.5900	7.7600	9.8700	6.553±3.928
Gradient Boosting	MAE	0.0110	1.7298	8.4972	0.2924	1.3708	1.4316	1.1400	10.9854	16.2228	12.9434	
	MAE/range, %	5.0000	3.3500	3.6900	12.1800	7.6200	2.9300	3.8000	13.7300	7.5500	10.2700	6.65±3.981
K nearest neighbours	MAE	0.0114	2.0454	9.4680	0.3236	2.0162	1.5818	1.0914	11.6578	18.0940	13.8334	
	MAE/range, %	5.1800	3.9600	4.1100	13.4800	11.2000	3.2400	3.6400	14.5700	8.4200	10.9800	7.533±4.51
Lasso	MAE	0.0114	1.8192	10.8782	0.2786	1.3136	1.3828	1.0528	10.8006	17.0136	13.0166	
	MAE/range, %	5.18	3.52	4.72	11.6100	7.3000	2.83	3.5100	13.5	7.9100	10.33	6.676±3.769
Random Forest	MAE	0.0104	1.8864	9.1400	0.2756	1.3392	1.5518	1.0666	10.6008	16.3742	12.5738	
	MAE/range, %	4.73	3.6500	3.97	11.48	7.4400	3.18	3.56	13.25	7.62	9.98	6.542±3.705
Clinical features and radiomics from B80f kernel												
AdaBoost	MAE	0.0104	1.6760	10.5908	0.2780	1.3170	1.3284	1.0430	10.3076	16.8334	12.5976	
	MAE/range, %	4.7300	3.2400	4.6000	11.5800	7.3200	2.7200	3.4800	12.8800	7.8300	10.0000	**6.487±3.704**
Extra Trees	MAE	0.0108	1.7392	9.4302	0.2814	1.3342	1.3752	1.0812	10.8268	16.7208	12.3108	
	MAE/range, %	4.9100	3.3600	4.0900	11.7200	7.4100	2.8200	3.6000	13.5300	7.7800	9.7700	6.58±3.86
Gradient Boosting	MAE	0.0106	1.7698	9.4862	0.2938	1.3654	1.4730	1.2206	10.6808	17.0070	12.8044	
	MAE/range, %	4.8200	3.4200	4.1200	12.2400	7.5900	3.0200	4.0700	13.3500	7.9100	10.1600	6.727±3.85
K nearest neighbours	MAE	0.0114	1.9902	9.7950	0.3254	2.0274	1.5434	1.1688	11.8466	18.5750	13.5656	
	MAE/range, %	5.1800	3.8500	4.2500	13.5600	11.2600	3.1600	3.9000	14.8100	8.6400	10.7700	7.623±4.562
Lasso	MAE	0.0108	1.8184	10.8042	0.2818	1.3222	1.3994	1.0620	10.8742	17.1848	12.8760	
	MAE/range, %	4.9100	3.5200	4.6900	11.7400	7.3500	2.8700	3.5400	13.5900	7.9900	10.2200	6.689±3.824
Random Forest	MAE	0.0102	1.8842	9.2316	0.2788	1.3472	1.5284	1.0748	10.5472	16.4000	12.4992	
	MAE/range, %	4.6400	3.6400	4.0100	11.6200	7.4800	3.1300	3.5800	13.1800	7.6300	9.9200	6.546±3.721

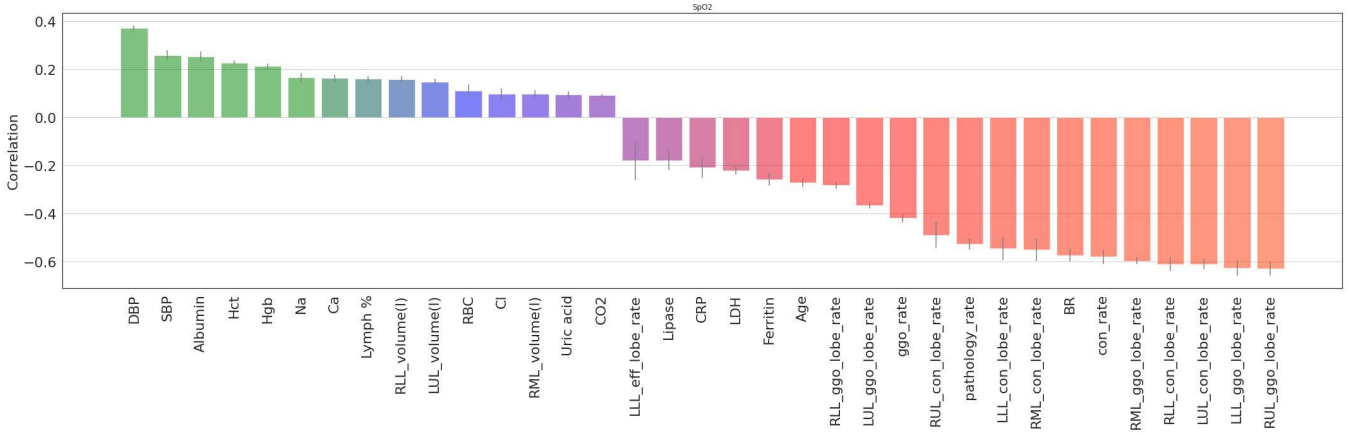


FIGURE 6. Predictors ranked by correlation feature selection method employing 10-fold cross-validation technique to predict level of SpO_2. Only significant correlations displayed.

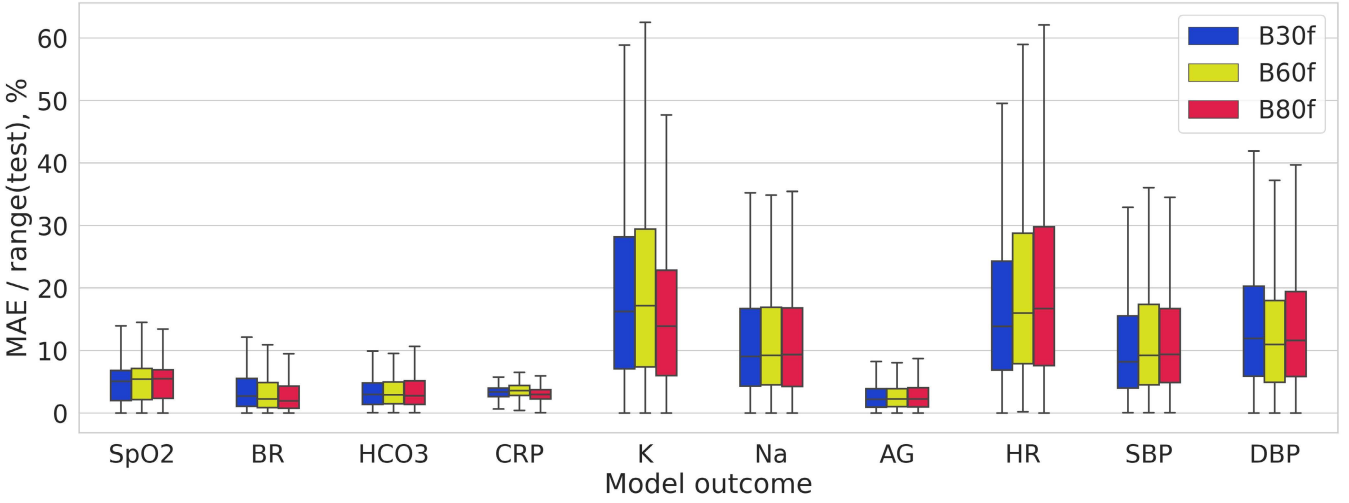


FIGURE 7. Distribution of MAE/range of values on different kernels for AdaBoost ML model trained on clinical features and radiomics jointly.

B. INFORMATION GAIN OF RADIOMICS AND BIOCHEMICAL FEATURES AS PREDICTORS OF FUNCTIONAL CHANGES IN COVID-19 ASSOCIATED PNEUMONIA

Researchers resort to various predictors for risk assessment and management in patients with SARS-CoV-2 pneumonia. In a study on a predictive model based on laboratory findings, potassium, chlorine and sodium were markedly higher in non-survivors group versus discharge group. In that study the groups differed in many laboratory features. However, none of the features provided the adequate accuracy in predicting the outcome of SARS-CoV-2 pneumonia [48]. This does not justify a low predictive value of laboratory findings. When analyzed together, they provide a reliable forecast as shown in our previous study for three top valuable predictors (the level of CRP, fibrinogen and activated partial thromboplastin time) [2]. Other authors showed that the level of severity correlates with plasma levels of IL-2, IL-7, IL-10, G-CSF, IP10, MCP1, MIP1A, and TNF-α [80]. Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 as disease severity predictors [81]. IL-8 is among the top valuable predictors for prognostication of acute lung injury [82]. Furthermore, avoiding or mitigating the cytokine storm may be a key treatment for SARS-CoV-2 [83].

In the current study we combined radiomics with humoral and cellular factors as potential biological markers of acute lung injury. The idea of blending the data from distinct visual and laboratory modalities comes from bioinformatics. To elucidate the pathogenesis of lung injury through an unbiased 'big-picture' approach, it uses advances in genomics, proteomics, metabolomics, etc. [75].

We put age as a predictor into the blending models as the informative value of laboratory findings differs among age groups. Other studies evidence the value of age in the clinical assessment of the patients infected with SARS-CoV-2. For instance, the most important laboratory reports

are normal or temporary elevated CRP, conflicting WBC count results and procalcitonin in children, lymphopenia, elevated CRP in adults along with elevated LDH in the elderly [84]. Supposedly, for this reason, age stays among strong predictors in the model forecasting oxygen saturation (see Fig. 6).

While some researchers work with molecular biology, we resorted to radiomics as it can be retrieved from the examinations conducted as part of standard care. COVID-19 pneumonia has typical imaging features which can help screening for highly suspected cases and evaluating the extent of the disease and its severity. The features are ground glass opacity or mixed ground glass opacity with consolidation, vascular enlargement in the lesions, their peripheral distribution, multifocal and bilateral involvement predominantly in the lower parts of the lung [85].

To stratify cases by the extent of the disease, we used the percentage of the total lung involvement as the precise automatic quantification of the lung lesions. Other researchers employ human-driven assessment of the extent of the disease. To calculate the CT involvement score, they rate the percentages of each of the five lobes that are involved. The CT involvement score reflects the severity and extent of the disease [85]. There is an ongoing discussion about an optimal scoring method for scaling the lung involvement by diagnosticians. Some argue towards including of additional qualitative features of lung involvement: GGO, consolidations, crazy-paving pattern, character other than enlisted [41]. To meet this requirement, we utilized a way of automatic segmentation that assessed common types of lesions separately (GGO, consolidations, pleural effusion).

It remains unclear if the radiological findings reflect the clinical status more or less reliably than the laboratory findings. Concentrating on distinguishing COVID-19 pneumonia from other viral pneumonia researchers found out that the clinical, laboratory, and especially radiological findings may aid in the differential diagnosis of non–SARS-CoV-2 pathogens from COVID-19 [74]. Unlike other authors, we put much effort to elucidate the clinical value of radiomics retrieved from chest CT. Our study gives an insight into an additive value of the distinct diagnostic modalities while managing cases of pneumonia. For instance, we ranked all features significantly correlated with the oxygen saturation and anion gap concerning the strength of the association, i.e. by the value of correlation coefficient r (Fig. 5-6) A single radiomical finding of the lung lobe involvement (RMO_ggo_lobe_rate) stays in the list of the valuable features. Interestingly, the correlation of AG with the laboratory estimates of the protein and heme metabolism (the level of total protein, urea, creatinine, albumin, total and direct bilirubin) is stronger than with other biochemical findings (e.g., glucose level). Some hematologic parameters (the percentage of neurophils, lymphocytes and monocytes) are also strongly correlated with AG.

If applied to community-acquired pneumonia, the analysis we did is supposed to provide similar findings. This is because chest CT in COVID-19 is more prone to resemble nonviral cases of pneumonia than viral pneumonia [74].

C. COMPARISON OF RECONSTRUCTION KERNELS BY PREDICTIVE POTENTIAL TO IDENTIFY THE CLINICAL SEVERITY FROM DIAGNOSTIC IMAGES

Our study shows that the reconstruction kernel does not affect the informative value of CT. Notched boxplots in Fig. 7 present the accuracy of regression models predicting markers of hypoxia and COVID-19 severity. On average, the accuracy of prediction from either the clinical features or radiomics is almost equal. Training the models on the full set of data does not enhance the accuracy. The models for AG and CRP have the best prediction metrics. This justifies the utility of AG and CRP for reflecting the clinical status of patients with COVID-19.

The effects of the reconstruction kernel (also referred to as algorithms or filters) on image quality is a common issue in radiology studies [44], [46], [86], [87]. This is because visual diagnostics depends heavily on the image quality (e.g., noise and spatial resolution) and the kernel may impact the quality settings. A kernel should be selected carefully for an examination. A smoother (lower resolution) kernel gives a more accurate representation. A sharper (higher resolution, edge-enhancing) kernel generates images with higher spatial resolution, but increases the image noise [86]. The image acquisition settings should correspond to the size and appearance of the targeted structure and the general background. For instance, the evaluation of small *low-contrast* structures should advance from the application of sharper high-resolution kernels [88]. In contrast, the ability to detect small *high-contrast* lesions improves as the reconstruction kernel becomes smoother [47]. Sharp image reconstruction kernels result in higher CT measurements of emphysema than smooth kernels [89]. Supposedly, the tendencies are true for the human-driven visual diagnostics. Our study does not justify this for computer-aided diagnostics. When done in an automatic way, the assessment of the disease severity and the predictive value of the radiomical findings retrieved from the images do not depend heavily on the kernel settings.

VI. LIMITATIONS OF THE STUDY

The study has some weaknesses as well as strengths. *A known limitation* of our study is that we worked specifically with the CT scans acquired right on admission to hospital when patients were hospitalized within a day or two after the disease emerged and they tested positive. As the radiological findings vary across the disease phases, the models we built were not trained to work with the data typical for the intermediate and late phases of the illness. Future studies are required to extend the utility of the ML algorithms by applying them to the results of the follow-up studies. *One more limitation* of the current study is that we tested patients exceptionally for SARS-CoV-2. However, coinfections may occur, and this should be considered [74].

Studies on distinguishing COVID-19 pneumonia from other viral pneumonia with chest CT showed that, surprisingly, COVID-19 is more prone to resemble nonviral cases of pneumonia than viral pneumonia [74]. *The positive side* of the statement is that, supposedly, the models we built can be transferable from COVID-19 cases exceptionally to CAP patients. *The negative side* is that a non–SARS-CoV-2 pathogen does not exclude accompanying COVID-19. This may account for false predictions. Such a limitation is typical for the majority of COVID-19 studies.

The future research will overcome the limitations of the current study by training the models on the follow-up examinations of the patients tested for the most common respiratory viruses including SARS-CoV-2. This will validate such a method for the quantitative assessment of the disease progression, which can aid clinical practice in the following ways. *First*, it will help physicians to adjust risk from admission to discharge thus optimizing individual case management. *Second*, an accurate marker of pneumonia severity will help the practitioners to perform clinical trials of the effectiveness of therapy.

A strength of the study is that the constructed models allow us to reduce the number of radiomical features which should be analyzed for the comprehensive assessment of COVID-19 cases considerably. This can simplify the routine patient management and decision-making by physicians in real clinical settings. *Another strength* of the study is that the comparison of the predicted and actual levels of oxygen saturation may reflect functional reserves in COVID-19 patients. To justify this, new studies should be conducted.

VII. CONCLUSION

- Pulmonology needs a reliable tool for quantitative assessment of the lung involvement in COVID-19 on admission and follow-up examination. An optimal biomarker should reflect disease severity and the actual physiological status of the patient. Radiomical findings might serve as the marker, but they are redundant and hard to deal with. We eliminated data redundancy with a machine learning model that accurately predicts the functional and biochemical markers of hypoxia. The model outcomes can serve as a single measure of the structural changes in the lungs due to COVID-19 thus substituting a big number of radiomical data. Physicians can compare the values calculated with the model from radiomical data to assess the disease severity and progression. The computations can be automatized and implemented into real clinical settings.
- The study proposed an efficient way of automatizing lung injury assessment in pneumonia and measuring disease severity. To model pathophysiologic change in the lung of patients with COVID-19, we built regression models predicting the oxygenation level, respiratory and cardiovascular functioning from the extracted radiomics features. The analysis of radiomics with advanced statistical methods helps to compare follow-up studies, detects diseases worsening and stratifies risks thus improving patient management. We also compared the reconstruction kernels of lung CT images with regard to the predictive potential for reflecting the clinical severity markers.
- Radiomics aids in prediction of the oxygenation level, respiratory and cardiovascular functioning from a set of demographics data, biochemical and hematologic findings. The average accuracy (MAE/range, %) of the models based on radiomics is 7.069±4.17, on the clinical features - 6.593±3.654 and on their combination - 6.454±3.715.
- The features which are positively correlated with anion gap (AG) attribute to protein and heme metabolism (total protein, albumin, urea, creatinine, total and direct bilirubin), total count of leukocytes and subtypes (neutrophils, monocytes, lymphocytes) and the involvement of the right middle lobe.
- The information gain of radiomics is significantly higher in the model predicting SpO_2. The oxygen saturation level correlates negatively with the involvement of the total lung and its specific lobes, CRP and LDH activity and positively - with lung lobes volumes, markers of protein metabolism, RBC count, HCT, the level of hemoglobin and percentage of lymphocytes.
- The settings of the reconstruction kernels do not impact the capacity of radiomics data to reflect the clinical status. The least accurate prediction is observed once images are acquired with B60f kernel, the most accurate - with B30f kernel. However, the accuracy differs insignificantly among disparate reconstruction kernels.

COMPLIANCE WITH ETHICAL STANDARDS

The study underwent ethical review by Department of Health Abu Dhabi (reference Number: DOH/CVDC/2020/887) and got approval for the retrospective analysis of the data obtained as a standard of care. No potentially identifiable personal information is presented in the study.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

All authors contributed to the creation of the article as follows: Yauhen Statsenko and Jamal Al Koteesh contributed to the conceptual idea of the paper; Yauhen Statsenko and Tetiana Habuza formulated the objectives, wrote the manuscript; Tetiana Habuza performed the statistical analysis, prepared the figures and tables for data presentation and illustration, Tatsiana Talako, Tetiana Kurbatova, Juana Sido, Gillian Lylian Simiyu, Darya Smetanina, Dana Sharif Qandil, Sarah Meribout, Juri G. Gelovani, Klaus Neidl-Van Gorkom, with Taleb M. Almansoori, Milos Ljubisavljevic, Fatmah Al Zahmi, Tom Loney, Anthony Bedson, Nerissa

Naidoo, Alireza Dehdashtian, and Karuna M. Das contributed to the literature review and data analysis.

DATA AVAILABILITY STATEMENT

Generated Statement: The datasets generated for this study are available on request at the site of *Big Data Analytics Center* at https://bi-dac.com

REFERENCES

[1] T. S. Simonson, T. L. Baker, R. B. Banzett, T. Bishop, J. A. Dempsey, J. L. Feldman, P. G. Guyenet, E. J. Hodson, G. S. Mitchell, E. A. Moya, B. T. Nokes, J. E. Orr, R. L. Owens, M. Poulin, J. M. Rawling, C. N. Schmickl, J. J. Watters, M. Younes, and A. Malhotra, "Silent hypoxaemia in COVID-19 patients," *J. Physiol.*, vol. 599, no. 4, pp. 1057–1065, Feb. 2021, doi: 10.1113/JP280769.

[2] Y. Statsenko, F. Al Zahmi, T. Habuza, K. N.-V. Gorkom, and N. Zaki, "Prediction of COVID-19 severity using laboratory findings on admission: Informative values, thresholds, ML model performance," *BMJ Open*, vol. 11, no. 2, Feb. 2021, Art. no. e044500, doi: 10.1136/bmjopen-2020-044500.

[3] Y. Statsenko, F. Al Zahmi, T. Habuza, T. M. Almansoori, D. Smetanina, G. L. Simiyu, K. N.-V. Gorkom, M. Ljubisavljevic, R. Awawdeh, H. Elshekhali, M. Lee, N. Salamin, R. Sajid, D. Kiran, S. Nihalani, T. Loney, A. Bedson, A. Dehdashtian, and J. Al Koteesh, "Impact of age and sex on COVID-19 severity assessed from radiologic and clinical findings," *Frontiers Cellular Infection Microbiology*, vol. 11, Feb. 2022, doi: 10.3389/fcimb.2021.777070.

[4] F. Al Zahmi, T. Habuza, R. Awawdeh, H. Elshekhali, M. Lee, N. Salamin, R. Sajid, D. Kiran, S. Nihalani, D. Smetanina, T. Talako, K. N.-V. Gorkom, N. Zaki, T. Loney, and Y. Statsenko, "Ethnicity-specific features of COVID-19 among arabs, africans, south asians, east asians, and Caucasians in the United Arab Emirates," *Frontiers Cellular Infection Microbiology*, vol. 11, p. 1241, Mar. 2022, doi: 10.3389/fcimb.2021.773141.

[5] Y. Statsenko, "Deep learning-based automatic assessment of lung impairment in COVID-19 pneumonia: Predicting markers of hypoxia with computer vision," *Front. Med.*, vol. 1996, May 2022, Art. no. 882190, doi: 10.3389/fmed.2022.882190.

[6] J. P. Cohen, L. Dao, K. Roth, P. Morrison, Y. Bengio, A. F. Abbasi, B. Shen, H. K. Mahsa, M. Ghassemi, H. Li, and T. Duong, "Predicting COVID-19 pneumonia severity on chest X-ray with deep learning," *Cureus*, vol. 12, no. 7, pp. 1–14, Jul. 2020, doi: 10.7759/cureus.9448.

[7] P.-J. Yu, H. Cassiere, K. Bocchieri, S. DeRosa, S. Yar, and A. Hartman, "Hypermetabolism in critically ill patients with COVID-19 and the effects of hypothermia: A case series," *Metabolism Open*, vol. 7, Sep. 2020, Art. no. 100046, doi: 10.1016/j.metop.2020.100046.

[8] Y. Statsenko, E. Fursa, V. Laver, N. Altakarli, T. Almansoori, F. A. Zahmi, K. Gorkom, M. Sheek-Hussein, A. Ponomareva, G. Simiyu, D. Smetanina, M. Ljubisavljevic, M. Szolics, and J. Al Koteesh, "Risk stratification and prediction of severity of hemorrhagic stroke in dry desert climate–A retrospective cohort study in eastern region of abu dhabi emirate," *J. Neurological Sci.*, vol. 429, Oct. 2021, Art. no. 117760, doi: 10.1016/j.jns.2021.117760.

[9] Y. Statsenko, T. Habuza, I. Charykova, K. Gorkom, N. Zaki, T. Almansoori, M. Ljubisavljevic, M. Szolics, J. Al Koteesh, A. Ponomareva, G. Simiyu, D. Smetanina, and M. Belghali, "Ai models of age-associated changes in CNS composition identified by MRI," *J. Neurological Sci.*, vol. 429, Oct. 2021, Art. no. 118303, doi: 10.1016/j.jns.2021.118303.

[10] Y. Statsenko, T. Habuza, I. Charykova, K. N.-V. Gorkom, N. Zaki, T. M. Almansoori, G. Baylis, M. Ljubisavljevic, and M. Belghali, "Predicting age from behavioral test performance for screening early onset of cognitive decline," *Frontiers Aging Neurosci.*, vol. 13, p. 282, Jul. 2021, doi: 10.3389/fnagi.2021.661514.

[11] Y. Statsenko, "Correlation between lifelong dynamics of psychophysiological performance and brain morphology. ESNR 2021," *Neuroradiology*, vol. 63, pp. 41–42, May 2021, Art. no. 1-P11, doi: 10.1007/s00234-021-02791-y.

[12] K. Gorkom, "Comparison of brain volumetric changes with functional outcomes in physiologic brain aging. ESNR 2021," *Neuroradiology*, vol. 63, pp. 43–44, Jun. 2021, Art. no. 1-P15, doi: 10.1007/s00234-021-02791-y.

[13] A. Esposito, A. Palmisano, R. Cao, P. Rancoita, G. Landoni, D. Grippaldi, E. Boccia, M. Cosenza, A. Messina, S. La Marca, D. Palumbo, C. D. Serio, M. Spessot, M. Tresoldi, P. Scarpellini, F. Ciceri, A. Zangrillo, and F. De Cobelli, "Quantitative assessment of lung involvement on chest CT at admission: Impact on hypoxia and outcome in COVID-19 patients," *Clin. Imag.*, vol. 77, pp. 194–201, Sep. 2021, doi: 10.1016/j.clinimag.2021.04.033.

[14] P. Brouqui, S. Amrane, M. Million, S. Cortaredona, P. Parola, J.-C. Lagier, and D. Raoult, "Asymptomatic hypoxia in COVID-19 is associated with poor outcome," *Int. J. Infectious Diseases*, vol. 102, pp. 233–238, Jan. 2021, doi: 10.1016/j.ijid.2020.10.067.

[15] H.-B. Tan, F. Xiong, Y.-L. Jiang, W.-C. Huang, Y. Wang, H.-H. Li, T. You, T.-T. Fu, R. Lu, and B.-W. Peng, "The study of automatic machine learning base on radiomics of non-focus area in the first chest CT of different clinical types of COVID-19 pneumonia," *Sci. Rep.*, vol. 10, no. 1, Nov. 2020, doi: 10.1038/s41598-020-76141-y.

[16] M. Zhang, X. Zeng, C. Huang, J. Liu, X. Liu, X. Xie, and R. Wang, "An AI-based radiomics nomogram for disease prognosis in patients with COVID-19 pneumonia using initial CT images and clinical indicators," *Int. J. Med. Informat.*, vol. 154, Oct. 2021, Art. no. 104545, doi: 10.1016/j.ijmedinf.2021.104545.

[17] Z. Xie, H. Sun, J. Wang, H. Xu, S. Li, C. Zhao, Y. Gao, X. Wang, T. Zhao, S. Duan, C. Hu, and W. Ao, "A novel CT-based radiomics in the distinction of severity of coronavirus disease 2019 (COVID-19) pneumonia," *BMC Infectious Diseases*, vol. 21, no. 1, pp. 1–11, Jun. 2021, doi: 10.1186/s12879-021-06331-0.

[18] C. Li, D. Dong, L. Li, W. Gong, X. Li, Y. Bai, M. Wang, Z. Hu, Y. Zha, and J. Tian, "Classification of severe and critical COVID-19 using deep learning and radiomics," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3585–3594, Dec. 2020, doi: 10.1109/JBHI.2020.3036722.

[19] F. Homayounieh, R. Babaei, H. Karimi Mobin, C. D. Arru, M. Sharifian, I. Mohseni, E. Zhang, S. R. Digumarthy, and M. K. Kalra, "Computed tomography radiomics can predict disease severity and outcome in coronavirus disease 2019 pneumonia," *J. Comput. Assist. Tomogr.*, vol. 44, no. 5, pp. 640–646, Sep. 2020, doi: 10.1097/RCT.0000000000001094.

[20] J. Huang, F. Wu, L. Chen, J. Yu, W. Sun, Z. Nie, H. Liu, F. Yang, and C. Zheng, "CT-based radiomics helps to predict residual lung lesions in COVID-19 patients at three months after discharge," *Diagnostics*, vol. 11, no. 10, p. 1814, Sep. 2021, doi: 10.3390/diagnostics11101814.

[21] F. Homayounieh, S. Ebrahimian, R. Babaei, H. K. Mobin, E. Zhang, B. C. Bizzo, I. Mohseni, S. R. Digumarthy, and M. K. Kalra, "CT radiomics, radiologists, and clinical information in predicting outcome of patients with COVID-19 pneumonia," *Radiol., Cardiothoracic Imag.*, vol. 2, no. 4, Aug. 2020, Art. no. e200322, doi: 10.1148/ryct.2020200322.

[22] C. Arru, S. Ebrahimian, Z. Falaschi, J. V. Hansen, A. Pasche, M. D. Lyhne, M. Zimmermann, F. Durlak, M. Mitschke, A. Carriero, J. E. Nielsen-Kudsk, M. K. Kalra, and L. Saba, "Comparison of deep learning, radiomics and subjective assessment of chest CT findings in SARS-CoV-2 pneumonia," *Clin. Imag.*, vol. 80, pp. 58–66, Dec. 2021, doi: 10.1016/j.clinimag.2021.06.036.

[23] I. Shiri, M. Sorouri, P. Geramifar, M. Nazari, M. Abdollahi, Y. Salimi, B. Khosravi, D. Askari, L. Aghaghazvini, G. Hajianfar, A. Kasaeian, H. Abdollahi, H. Arabi, A. Rahmim, A. R. Radmard, and H. Zaidi, "Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104304, doi: 10.1016/j.compbiomed.2021.104304.

[24] J. Xu, M. Zhou, Z. Wang, T. Liao, Y. Ma, G. Hu, S. Wang, J. Gu, Z. Yin, and Y. Jin, "Development and validation of a predictive model based on radiomics to predict the short-term outcomes of patients with COVID-19," Res. Square, Sep. 2020, doi: 10.21203/rs.3.rs-74208/v1.

[25] Q. Cai, S.-Y. Du, S. Gao, G.-L. Huang, Z. Zhang, S. Li, X. Wang, P.-L. Li, P. Lv, G. Hou, and L.-N. Zhang, "A model based on CT radiomic features for predicting RT-PCR becoming negative in coronavirus disease 2019 (COVID-19) patients," *BMC Med. Imag.*, vol. 20, no. 1, pp. 1–10, Oct. 2020, doi: 10.1186/s12880-020-00521-z.

[26] X. Fang, X. Li, Y. Bian, X. Ji, and J. Lu, "Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2," *Eur. Radiol.*, vol. 30, no. 12, pp. 6888–6901, Jul. 2020, doi: 10.1007/s00330-020-07032-z.

[27] H. J. Chen, L. Mao, Y. Chen, L. Yuan, F. Wang, X. Li, Q. Cai, J. Qiu, and F. Chen, "Machine learning-based CT radiomics model distinguishes COVID-19 from non-COVID-19 pneumonia," *BMC Infectious Diseases*, vol. 21, no. 1, pp. 1–13, Sep. 2021, doi: 10.1186/s12879-021-06614-6.

[28] J. R. Ferreira Junior, D. A. Cardona Cardenas, R. A. Moreno, M. D. F. de Sá Rebelo , J. E. Krieger, and M. A. Gutierrez, "Novel chest radiographic biomarkers for COVID-19 using radiomic features associated with diagnostics and outcomes," *J. Digit. Imag.*, vol. 34, no. 2, pp. 297–307, Feb. 2021, doi: 10.1007/s10278-021-00421-w.

[29] S. Masoud Rezaeijo, M. Ghorvei, and M. Alaei, "A machine learning method based on lesion segmentation for quantitative analysis of CT radiomics to detect COVID-19," in *Proc. 6th Iranian Conf. Signal Process. Intell. Syst. (ICSPIS)*, Dec. 2020, pp. 1–5, doi: 10.1109/ICSPIS51611.2020.9349605.

[30] J. Guiot, "Development and validation of an automated radiomic ct signature for detecting COVID-19," *Diagnostics*, vol. 11, p. 41, Dec. 2020, doi: 10.3390/diagnostics11010041.

[31] S. M. Rezaeijo, R. Abedi-Firouzjah, M. Ghorvei, and S. Sarnameh, "Screening of COVID-19 based on the extracted radiomics features from chest CT images," *J. X-Ray Sci. Technol.*, vol. 29, no. 2, pp. 229–243, Mar. 2021, doi: 10.3233/XST-200831.

[32] Q. Zeng, K. I. Zheng, J. Chen, Z. Jiang, T. Tian, X. Wang, H. Ma, K. Pan, Y. Yang, Y. Chen, and M. Zheng, "Radiomics-based model for accurately distinguishing between severe acute respiratory syndrome associated coronavirus 2 (SARS-CoV-2) and influenza a infected pneumonia," *MedComm*, vol. 1, no. 2, pp. 240–248, Sep. 2020, doi: 10.1002/mco2.14.

[33] Y. Huang, Z. Zhang, S. Liu, X. Li, Y. Yang, J. Ma, Z. Li, J. Zhou, Y. Jiang, and B. He, "CT-based radiomics combined with signs: A valuable tool to help radiologist discriminate COVID-19 and influenza pneumonia," *BMC Med. Imag.*, vol. 21, no. 1, p. 31, Feb. 2021, doi: 10.1186/s12880-021-00564-w.

[34] T. Habuza, N. Zaki, Y. Statsenko, F. Alnajjar, and S. Elyassami, "Deep learning for predicting cognitive gap as a reliable biomarker of dementia," *MedRxiv*, Jan. 2021.

[35] T. Habuza, N. Zaki, E. A. Mohamed, and Y. Statsenko, "Deviation from model of normal aging in alzheimer's disease: Application of deep learning to structural MRI data and cognitive tests," *IEEE Access*, vol. 10, pp. 53234–53249, 2022, doi: 10.1109/ACCESS.2022.3174601.

[36] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, and J. R. Weir-McCall, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Mach. Intell.*, vol. 3, no. 3, pp. 199–217, 2021.

[37] Y. J. Kwon, D. Toussie, M. Finkelstein, M. A. Cedillo, S. Z. Maron, S. Manna, N. Voutsinas, C. Eber, A. Jacobi, A. Bernheim, Y. S. Gupta, M. S. Chung, Z. A. Fayad, B. S. Glicksberg, E. K. Oermann, and A. B. Costa, "Combining initial radiographs and clinical variables improves deep learning prognostication in patients with COVID-19 from the emergency department," *Radiology: Artif. Intell.*, vol. 3, no. 2, Mar. 2021, Art. no. e200098, doi: 10.1148/ryai.2020200098.

[38] A. Borghesi, A. Zigliani, S. Golemi, N. Carapella, P. Maculotti, D. Farina, and R. Maroldi, "Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy," *Int. J. Infectious Diseases*, vol. 96, pp. 291–293, Jul. 2020, doi: 10.1016/j.ijid.2020.05.021.

[39] M. A. Orsi, G. Oliva, T. Toluian, C. V. Pittino, M. Panzeri, and M. Cellina, "Feasibility, reproducibility, and clinical validity of a quantitative chest X-ray assessment for COVID-19," *Amer. J. Tropical Med. Hygiene*, vol. 103, no. 2, pp. 822–827, Aug. 2020, doi: 10.4269/ajtmh.20-0535.

[40] K. M. Das, J. A. Alkoteesh, M. Sheek-Hussein, S. A. Alzadjali, M. T. Alafeefi, R. Singh, Y. Statsenko, E. S. Soteriades, V. Singh, and K. Van Gorkom, "Role of chest radiograph in MERS-cov pneumonia: A single tertiary referral center experience in the United Arab Emirates," *Egyptian J. Radiol. Nucl. Med.*, vol. 52, no. 1, pp. 1–7, May 2021, doi: 10.1186/s43055-021-00517-x.

[41] P. Wasilewski, B. Mruk, S. Mazur, G. Półtorak-Szymczak, K. Sklinda, and J. Walecki, "COVID-19 severity scoring systems in radiological imaging—A review," *Polish J. Radiol.*, vol. 85, no. 1, pp. 361–368, 2020, doi: 10.5114/pjr.2020.98009.

[42] A. Haga, W. Takahashi, S. Aoki, K. Nawa, H. Yamashita, O. Abe, and K. Nakagawa, "Standardization of imaging features for radiomics analysis," *J. Med. Invest.*, vol. 66, nos. 1–2, pp. 35–37, Feb. 2019, doi: 10.2152/jmi.66.35.

[43] T. Habuza, A. N. Navaz, F. Hashim, F. Alnajjar, N. Zaki, M. A. Serhani, and Y. Statsenko, "Ai applications in robotics, diagnostic image analysis and precision medicine: Current limitations, future trends, guidelines on cad systems for medicine," *Inform. Med. Unlocked*, vol. 24, Jun. 2021, Art. no. 100596, doi: 10.1016/j.imu.2021.100596.

[44] J. Choe, S. M. Lee, K.-H. Do, G. Lee, J.-G. Lee, S. M. Lee, and J. B. Seo, "Deep Learning–based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses," *Radiology*, vol. 292, no. 2, pp. 365–373, Aug. 2019, doi: 10.1148/radiol.2019181960.

[45] A. Traverso, L. Wee, A. Dekker, and R. Gillies, "Repeatability and reproducibility of radiomic features: A systematic review," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 102, no. 4, pp. 1143–1158, Nov. 2018, doi: 10.1016/j.ijrobp.2018.05.053.

[46] M. Meyer, J. Ronald, F. Vernuccio, R. C. Nelson, J. C. Ramirez-Giraldo, J. Solomon, B. N. Patel, E. Samei, and D. Marin, "Reproducibility of CT radiomic features within the same patient: Influence of radiation dose and CT reconstruction settings," *Radiology*, vol. 293, no. 3, pp. 583–591, Dec. 2019, doi: 10.1148/radiol.2019190928.

[47] P. F. Judy and R. G. Swensson, "Detection of small focal lesions in CT images: Effects of reconstruction filters and visual display Windows," *Brit. J. Radiol.*, vol. 58, no. 686, pp. 137–145, Feb. 1985, doi: 10.1259/0007-1285-58-686-137.

[48] G. Wu, S. Zhou, Y. Wang, W. Lv, S. Wang, T. Wang, and X. Li, "A prediction model of outcome of SARS-CoV-2 pneumonia based on laboratory findings," *Sci. Rep.*, vol. 10, no. 1, Aug. 2020, doi: 10.1038/s41598-020-71114-7.

[49] B. Zhou, B. Lou, J. Liu, and J. She, "Serum metabolite profiles as potential biochemical markers in young adults with community-acquired pneumonia cured by moxifloxacin therapy," *Sci. Rep.*, vol. 10, no. 1, Mar. 2020, doi: 10.1038/s41598-020-61290-x.

[50] J. Pereira, J. Paiva, J. Baptista, F. Froes, and J. Gonçalves-Pereira, "Severe community-acquired pneumonia: Risk factors for in-hospital mortality," *Crit. Care*, vol. 16, no. S1, Mar. 2012, doi: 10.1186/cc10645.

[51] I. Huang, R. Pranata, M. A. Lim, A. Oehadian, and B. Alisjahbana, "C-reactive protein, procalcitonin, d-dimer, and ferritin in severe coronavirus disease-2019: A meta-analysis," *Ther. Adv. Respir. Dis.*, vol. 14, pp. 1–14, Jul. 2020, doi: 10.1177/1753466620937175.

[52] D. Ji, D. Zhang, J. Xu, Z. Chen, T. Yang, P. Zhao, G. Chen, G. Cheng, Y. Wang, J. Bi, L. Tan, G. Lau, and E. Qin, "Prediction for progression risk in patients with COVID-19 pneumonia: The CALL score," *Clin. Infectious Diseases*, vol. 71, no. 6, pp. 1393–1399, Sep. 2020, doi: 10.1093/cid/ciaa414.

[53] J. Gong, "A tool to early predict severe 2019-novel coronavirus pneumonia (COVID-19): A multicenter study using the risk nomogram in Wuhan and Guangdong, China," *MedRxiv*, May 2021.

[54] Y. Gao, T. Li, M. Han, X. Li, D. Wu, Y. Xu, Y. Zhu, Y. Liu, X. Wang, and L. Wang, "Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19," *J. Med. Virology*, vol. 92, no. 7, pp. 791–796, Jul. 2020, doi: 10.1002/jmv.25770.

[55] E. V. Robilotti, "Determinants of covid-19 disease severity in patients with cancer," *Nat. Med.*, vol. 26, no. 8, pp. 1218–1223, Jun. 2020, doi: 10.1038/s41591-020-0979-0.

[56] B. M. Henry, G. Aggarwal, J. Wong, S. Benoit, J. Vikse, M. Plebani, and G. Lippi, "Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: A pooled analysis," *Amer. J. Emergency Med.*, vol. 38, no. 9, pp. 1722–1726, Sep. 2020, doi: 10.1016/j.ajem.2020.05.073.

[57] H. Long, L. Nie, X. Xiang, H. Li, X. Zhang, X. Fu, H. Ren, W. Liu, Q. Wang, and Q. Wu, "D-dimer and prothrombin time are the significant indicators of severe COVID-19 and poor prognosis," *BioMed Res. Int.*, vol. 2020, pp. 1–10, Jun. 2020, doi: 10.1155/2020/6159720.

[58] (Jun. 2020). *National Guidelines for Clinical Management and Treatment of COVID-19—Version 4.1*. Accessed: Aug. 1, 2020. [Online]. Available: https://www.dha.gov.ae/en/HealthRegulation/Documents/National_Guidelines_of_COVID_19_1st_June_2020.pdf

[59] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 9315, Munich, Germany, Nov. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[60] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2015, pp. 448–456.

[61] X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method," *Med. Phys.*, vol. 44, no. 10, pp. 5221–5233, Aug. 2017, doi: 10.1002/mp.12480.
[62] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, "NnU-Net: Self-adapting framework for U-net-Based medical image segmentation," 2018, *arXiv:1809.10486*.
[63] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur. Radiol. Experim.*, vol. 4, no. 1, pp. 1–11, Aug. 2020, doi: 10.1186/s41747-020-00173-2.
[64] MedSeg. *COVID-19 CT Segmentation Dataset*. Accessed: Jun. 12, 2022. [Online]. Available: http://medicalsegmentation.com/covid19
[65] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782–790, Aug. 2012, doi: 10.1016/j.neuroimage.2011.09.015.
[66] T. Habuza, K. Khalil, N. Zaki, F. Alnajjar, and M. Gochoo, "Web-based multi-user concurrent job scheduling system on the shared computing resource objects," in *Proc. 14th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2020, pp. 221–226, doi: 10.1109/IIT50501.2020.9299110.
[67] Nipype Tutorial. *Docker Image for the Interactive Nipype Tutorial*. Accessed: Apr. 12, 2021. [Online]. Available: https://miykael.github.io/nipype_tutorial/notebooks/introduction_docker.html
[68] M. Francone, F. Iafrate, G. M. Masci, S. Coco, F. Cilia, L. Manganaro, V. Panebianco, C. Andreoli, M. C. Colaiacomo, M. A. Zingaropoli, M. R. Ciardi, C. M. Mastroianni, F. Pugliese, F. Alessandri, O. Turriziani, P. Ricci, and C. Catalano, "Chest CT score in COVID-19 patients: Correlation with disease severity and short-term prognosis," *Eur. Radiol.*, vol. 30, no. 12, pp. 6808–6817, Jul. 2020, doi: 10.1007/s00330-020-07033-y.
[69] F. Pan, "Time course of lung changes at chest ct during recovery from coronavirus disease 2019 (COVID-19)," *Radiology*, vol. 295, no. 3, pp. 715–721, Feb. 2020, doi: 10.1148/radiol.2020200370.
[70] D. Chandra, R. Maini, and D. M. Hershberger, "Cryptogenic organizing pneumonia," in *StatPearls [Internet]*. Treasure Island, FL, USA: StatPearls, 2021.
[71] K. Kanaoka, S. Minami, S. Ihara, T. Tanaka, H. Yasuoka, and K. Komuta, "Secondary organizing pneumonia after coronavirus disease 2019: Two cases," *Respiratory Med. Case Rep.*, vol. 32, 2021, Art. no. 101356, doi: 10.1016/j.rmcr.2021.101356.
[72] X. Fang, C. Bai, and X. Wang, "Bioinformatics insights into acute lung injury/acute respiratory distress syndrome," *Clin. Trans. Med.*, vol. 1, p. 9, Jun. 2012, doi: 10.1186/2001-1326-1-9.
[73] F. Fu, J. Lou, D. Xi, Y. Bai, G. Ma, B. Zhao, D. Liu, G. Bao, Z. Lei, and M. Wang, "Chest computed tomography findings of coronavirus disease 2019 (COVID-19) pneumonia," *Eur. Radiol.*, vol. 30, no. 10, pp. 5489–5498, May 2020, doi: 10.1007/s00330-020-06920-8.
[74] A. O. Alpaydin, N. S. Gezer, G. O. Simsek, K. C. Tertemiz, O. O. E. Kutsoylu, A. N. Zeka, I. Guzel, M. Soyturk, A. A. Sayiner, and V. A. Oguz, "Clinical and radiological diagnosis of non–SARS-CoV-2 viruses in the era of COVID-19 pandemic," *J. Med. Virology*, vol. 93, no. 2, pp. 1119–1125, Feb. 2021, doi: 10.1002/jmv.26410.
[75] J. F. Pittet, R. C. Mackersie, T. R. Martin, and M. A. Matthay, "Biological markers of acute lung injury: Prognostic and pathogenetic significance.," *Amer. J. Respiratory Crit. Care Med.*, vol. 155, no. 4, pp. 1187–1205, Apr. 1997, doi: 10.1164/ajrccm.155.4.9105054.
[76] L. Tan, Q. Wang, D. Zhang, J. Ding, Q. Huang, Y.-Q. Tang, Q. Wang, and H. Miao, "Lymphopenia predicts disease severity of COVID-19: A descriptive and predictive study," *Signal Transduction Targeted Therapy*, vol. 5, no. 1, pp. 1–3, Mar. 2020, doi: 10.1038/s41392-020-0148-4.
[77] D. Wang, "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China," *Jama*, vol. 323, no. 11, pp. 1061–1069, Feb. 2020, doi: 10.1001/jama.2020.1585.
[78] M. C. Reade, L. Weissfeld, D. C. Angus, J. A. Kellum, and E. B. Milbrandt, "The prevalence of anemia and its association with 90-day mortality in hospitalized community-acquired pneumonia," *BMC Pulmonary Med.*, vol. 10, no. 1, pp. 123–145, Mar. 2010, doi: 10.1186/1471-2466-10-15.
[79] S. M. Doshi, A. M. Rueda, V. F. Corrales-Medina, and D. M. Musher, "Anemia and community-acquired pneumococcal pneumonia," *Infection*, vol. 39, no. 4, pp. 379–383, May 2011, doi: 10.1007/s15010-011-0122-8.
[80] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, and L. Zhang, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, Jun. 2020, doi: 10.1016/S0140-6736(20)30183-5.
[81] H. Han, Q. Ma, C. Li, R. Liu, L. Zhao, W. Wang, P. Zhang, X. Liu, G. Gao, F. Liu, Y. Jiang, X. Cheng, C. Zhu, and Y. Xia, "Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors," *Emerg. Microbes Infections*, vol. 9, no. 1, pp. 1123–1130, May 2020, doi: 10.1080/22221751.2020.1770129.
[82] T. C. Allen and A. Kurdowska, "Interleukin 8 and acute lung injury," *Arch. Pathol. Lab. Med.*, vol. 138, no. 2, pp. 266–269, Feb. 2014, doi: 10.5858/arpa.2013-0182-RA.
[83] M. Rokni, M. R. Hamblin, and N. Rezaei, "Cytokines and COVID-19: Friends or foes?" *Hum. Vaccines Immunotherapeutics*, vol. 16, no. 10, pp. 2363–2365, Aug. 2020, doi: 10.1080/21645515.2020.1799669.
[84] S. Vakili, A. Savardashtaki, S. Jamalnia, R. Tabrizi, M. H. Nematollahi, M. Jafarinia, and H. Akbari, "Laboratory findings of COVID-19 infection are conflicting in different age groups and pregnant women: A literature review," *Arch. Med. Res.*, vol. 51, no. 7, pp. 603–607, Oct. 2020, doi: 10.1016/j.arcmed.2020.06.007.
[85] W. Zhao, Z. Zhong, X. Xie, Q. Yu, and J. Liu, "Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: A multicenter study," *Amer. J. Roentgenology*, vol. 214, no. 5, pp. 1072–1077, May 2020, doi: 10.2214/AJR.20.22976.
[86] J. Paul, B. Krauss, R. Banckwitz, W. Maentele, R. W. Bauer, and T. J. Vogl, "Relationships of clinical protocols and reconstruction kernels with image quality and radiation dose in a 128-slice CT scanner: Study with an anthropomorphic and water phantom," *Eur. J. Radiol.*, vol. 81, no. 5, pp. e699–e703, May 2012, doi: 10.1016/j.ejrad.2011.01.078.
[87] D. S. Gierada, A. J. Bierhals, C. K. Choong, S. T. Bartel, J. H. Ritter, N. A. Das, C. Hong, T. K. Pilgram, K. T. Bae, B. R. Whiting, J. C. Woods, J. C. Hogg, B. A. Lutey, R. J. Battafarano, J. D. Cooper, B. F. Meyers, and G. A. Patterson, "Effects of CT section thickness and reconstruction kernel on emphysema quantification," *Academic Radiol.*, vol. 17, no. 2, pp. 146–156, Feb. 2010, doi: 10.1016/j.acra.2009.08.007.
[88] S. Prevrhal, K. Engelke, and W. A. Kalender, "Accuracy limits for the determination of cortical width and density: The influence of object size and ct imaging parameters," *Phys. Med. Biol.*, vol. 44, no. 3, p. 751, Mar. 1999, doi: 10.1088/0031-9155/44/3/017.
[89] K. L. Boedeker, M. F. McNitt-Gray, S. R. Rogers, D. A. Truong, M. S. Brown, D. W. Gjertson, and J. G. Goldin, "Emphysema: Effect of reconstruction algorithm on CT imaging measures," *Radiology*, vol. 232, no. 1, pp. 295–301, Jul. 2004, doi: 10.1148/radiol.2321030383.

YAUHEN STATSENKO is currently an Associate Professor of radiology with the College of Medicine and Health Sciences, United Arab Emirates University. His technical background as an IT Developer helps him to implement data analysis into medical studies and to supervise studies in biophysics and bioinformatics. He uses machine learning to analyze clinical data and focuses on mathematical modeling of physiological processes and diseases.

TETIANA HABUZA received the bachelor's degree in applied mathematics and the master's degree in informatics from Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, and the Ph.D. degree in computer science from United Arab Emirates University. Her research interests include bioinformatics and control theory. She focuses on advanced machine learning techniques for automatic analysis of medical images and clinical data.

TATSIANA TALAKO received the Ph.D. and M.D. degrees. She is currently an Assistant Professor and a consultant in hematology and rheumatology. She has many-year experience in teaching and research on thrombosis and thrombolysis, hemorrhage, and hemostasis pathologies. She is a patent holder and the author of over 50 research articles.

TETIANA KURBATOVA is currently a Research Assistant with the College of Medicine and Health Sciences, United Arab Emirates University, Al Ain. After finishing Kyiv National Taras Shevchenko University, she participated in research projects. Some of them were related to infectious diseases, including COVID-19.

GILLIAN LYLIAN SIMIYU is currently a Research Assistant with the College of Medicine and Health Sciences, United Arab Emirates University, Al Ain. She has hands-on experience at governmental institutions. She is an investigator with a particular interest in radiology, radiomics, and communicable diseases.

DARYA SMETANINA is currently a Research Assistant with the College of Medicine and Health Sciences, United Arab Emirates University, Al Ain. She is engaged in projects related to public health. Her research interests include epidemiology of COVID-19, neurodegenerative diseases, and obesity.

JUANA SIDO is currently an Administrator with the Eye Microsurgery Center "Voka," Minsk, Belarus. She is also a Research Volunteer and a Project Manager. Her research interests include computer vision and diagnostic imaging in conjunction with diagnostic machine learning algorithms.

DANA SHARIF QANDIL is currently a Medical Student with Ras Al Khaimah Medical Health and Sciences University, United Arab Emirates. She has a strong adherence to research work. Her current research interest includes quantitative assessment of lung and brain impairment with radiologic findings.

SARAH MERIBOUT graduated from the University of Constantine 3, Algeria. She is currently a Research Volunteer with the University of Constantine 3, Algeria. Her research interest includes automatic assessment of the internal organ structure from medical images. Due to her high motivation, she has knowledge in interpreting imagery presentations of a wide range of pathologies.

JURI G. GELOVANI is currently a Professor of surgery and radiology with the College of Medicine and Health Sciences, United Arab Emirates University, and an Adjunct Professor with the Department of Biomedical Engineering, the Department of Oncology, and the Department of Neurosurgery, Wayne State University, and the Siriraj Hospital, Mahidol University, Bangkok, Thailand. He specializes in neuroimaging, neurology, neurosurgery, and neurotraumatology. He is the pioneer of molecular-genetic in vivo imaging who holds more than 15 patents. He has published more than 160 articles and book chapters, and edited a major book in molecular imaging.

KLAUS NEIDL-VAN GORKOM is currently a Chairperson of the Radiology Department, College of Medicine and Health Sciences, United Arab Emirates University. His research interests include neuroradiology and computer vision. He is an established radiologist and nuclear medicine specialist, a SEHA-approved consultant with many-year experience who shares knowledge with colleagues and students.

TALEB M. ALMANSOORI is currently an Assistant Dean for the Clinical Affairs, an Assistant Professor of radiology with the College of Medicine and Health Sciences, United Arab Emirates University. He is a clinical practitioner and a researcher in general radiology who specializes in stroke treatment and different predictors of the surveillance.

FATMAH AL ZAHMI is currently an American Certified Emirati Neurologist who has dual training in stroke and neuromuscular medicine. She is particularly interested in studies on stroke in children and young adults, women, inflammatory myopathies, nerve conduction, and electromyography studies. At the time of the COVID-19 pandemic, she initiated and conducted studies dedicated to disease epidemiology and pathophysiology.

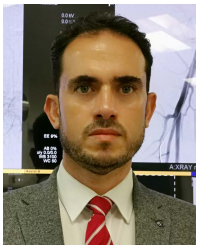
TOM LONEY is currently an Associate Professor of public health and epidemiology with broad interests and expertise in clinical and population-based research. He worked at the United Arab Emirates for the past 12 years conducting research and providing public, occupational, and environmental health advisory services to the defense, industrial and healthcare sectors. He is a member of the WHO/ILO Occupational Systematic Review Methodology Working Group.

ANTHONY BEDSON is currently a Radiology Manager with over 30 years experience in radiography with a strong interest in cross sectional imaging. He has previously been involved in several research projects for the Cognitive Neuroscience Department, University of Bangor, North Wales, U.K.

NERISSA NAIDOO is currently an Assistant Professor of anatomy at the Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates. Most recently, she was selected as a Next Generation Women Leader 2020 by McKinsey Company. Her research contributions pertain to the degenerative and biomechanical aspects of three-dimensional musculoskeletal models, anomalies of arterial trees, and blended instructional models in lesson plan dissemination.

ALIREZA DEHDASHTIAN received the M.D. Diploma degree from the University of Szeged, Hungary. He is currently a SEHA-Certified Neuroradiologist with Tawam Hospital, Al Ain, United Arab Emirates. He specializes in diagnosing neurological pathologies.

MILOS LJUBISAVLJEVIC is currently a Professor of physiology with the College of Medicine and Health Sciences, United Arab Emirates University. His research interests include physiologic and pathophysiologic mechanisms of cortical plasticity and motor control with special interests in movement disorders. He serves for various national and international grant committees. He serves as a member for the editorial board and an ad-hoc reviewer for several international journals.



JAMAL AL KOTEESH is currently an Interventional Radiologist and a Chair of the Radiology Department, Tawam Hospital, Al Ain, and an Adjunct Professor of radiology with the CMHS, UAEU. He is also one of the founders of interventional radiology service in the Abu Dhabi. He brings expertise on acute stroke interventions, erectile dysfunction treatment, and all types of biopsies and drainages under radiology guidance.

KARUNA M. DAS received the M.D. degree in diagnostic radiology from the Postgraduate Institute of Medical Education and Research, Chandigarh, India. He had a Fellowship in Interventional Radiology at Saint James's University Hospital, Leeds, U.K. He is a SEHA-approved consultant in radiology. He is currently an Associate Professor with the Radiology Department, College of Medicine and Health Sciences, United Arab Emirates University. His research interest includes assessment of lung impairment caused by viral infections, in particular, MERS-CoV2 and COVID-19.