

1 High-Dimensional Multinomial Multiclass Severity 2 Scoring of COVID-19 Pneumonia Using CT Radiomics 3 Features and Machine Learning Algorithms

4
5 Isaac Shiri¹, Shayan Mostafaei², Atlas Haddadi Avval³, Yazdan Salimi¹, Amirhossein Sanaat¹, Azadeh
6 Akhavanallaf¹, Hossein Arabi¹, Arman Rahmim^{4,5} and Habib Zaidi^{1,6,7,8}

7
8 1 Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211
9 Geneva, Switzerland

10 2 Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska
11 Institutet, Stockholm, Sweden

12 3 School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

13 4 Departments of Radiology and Physics, University of British Columbia, Vancouver BC, Canada

14 5 Department of Integrative Oncology, BC Cancer Research Institute, Vancouver BC, Canada

15 6 Geneva University Neurocenter, Geneva University, Geneva, Switzerland

16 7 Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University
17 Medical Center Groningen, Groningen, Netherlands

18 8 Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

19
20
21
22 **Corresponding author:**

23 Habib Zaidi, Ph.D
24 Geneva University Hospital
25 Division of Nuclear Medicine and Molecular Imaging
26 CH-1211 Geneva, Switzerland
27 **Tel:** +41 22 372 7258
28 **Fax:** +41 22 372 7169
29 **email:** habib.zaidi@hcuge.ch
30

1 Abstract

2 We aimed to construct a prediction model based on computed tomography (CT) radiomics
3 features to classify COVID-19 patients into severe-, moderate-, mild-, and non-pneumonic. A
4 total of 1110 patients were studied from a publicly available dataset with 4-class severity
5 scoring performed by a radiologist (based on CT images and clinical features). CT scans were
6 preprocessed with bin discretization and resized, followed by segmentation of the entire lung
7 and extraction of radiomics features. We utilized two feature selection algorithms, namely
8 Bagging Random Forest (BRF) and Multivariate Adaptive Regression Splines (MARS), each
9 coupled to a classifier, namely multinomial logistic regression (MLR), to construct multiclass
10 classification models. Subsequently, 10-fold cross-validation with bootstrapping (n=1000)
11 was performed to validate the classification results. The performance of multi-class models
12 was assessed using precision, recall, F1-score, and accuracy based on the 4×4 confusion
13 matrices. In addition, the areas under the receiver operating characteristic (ROC) curve
14 (AUCs) for multi-class classifications were calculated and compared for both models using
15 “multiROC” and “pROC” R packages. Using BRF, 19 radiomics features were selected, 9
16 from first-order, 6 from GLCM, 1 from GLDM, 1 from shape, 1 from NGTDM, and 1 from
17 GLSZM radiomics features. Ten features were selected using the MARS algorithm, namely 2
18 from first-order, 1 from GLDM, 2 from GLRLM, 2 from GLSZM, and 3 from GLCM
19 features. The Mean Absolute Deviation and Median from first-order, Small Area Emphasis
20 from GLSZM, and Correlation from GLCM features were selected by both BRF and MARS
21 algorithms. Except for the Inverse Variance feature from GLCM, all selected features by BRF
22 or MARS were significantly associated with four-class outcomes as assessed within MLR
23 (All p-values<0.05). BRF+MLR and MARS+MLR resulted in pseudo-R² prediction
24 performances of 0.295 and 0.256, respectively. Meanwhile, there were no significant
25 differences between the feature selection models when using a likelihood ratio test (p-value
26 =0.319). Based on confusion matrices for BRF+MLR and MARS+MLR algorithms, the
27 precision was 0.861 and 0.825, the recall was 0.844 and 0.793, whereas the accuracy was
28 0.933 and 0.922, respectively. AUCs (95% CI) for multi-class classification were 0.823
29 (0.795-0.852) and 0.816 (0.788-0.844) for BRF+MLR and MARS+MLR algorithms,
30 respectively. Our models based on the utilization of radiomics features, coupled with machine
31 learning, were able to accurately classify patients according to the severity of pneumonia, thus
32 highlighting the potential of this emerging paradigm in the prognostication and management
33 of COVID-19 patients.

34 **Keywords:** COVID-19, Pneumonia, Radiomics, Machine Learning, Computed Tomography,
35 High-Dimensional Data.

36

37

1 INTRODUCTION

2 The highly contagious SARS-CoV-2 virus has led to significant morbidity and mortality
3 worldwide ¹. Pneumonia is regarded as one of the main complications of COVID-19 disease,
4 which can lead to lethal conditions while escalating the cost of healthcare ². The most popular
5 diagnostic test considered as the gold standard for coronavirus disease is the reverse
6 transcription polymerase chain reaction (RT-PCR) assay ³. While highly specific, RT-PCR
7 has shown low sensitivity, as studies have reported significant false-negatives in patients who
8 had abnormalities in their chest CT images confirmed with secondary follow-up RT-PCR to
9 be positive for COVID-19 ⁴.

10 CT aids in the diagnosis and management of COVID-19 patients and could be potentially
11 used as an outcome/survival prediction tool, towards enhanced treatment planning ⁵⁻⁷. CT
12 scanning has been utilized as a highly sensitive tool for COVID-19 diagnosis ⁸ since it is fast
13 and generates quantifiable features (e.g., the extent to which lung lobes are involved) and non-
14 quantifiable features (e.g., ground-glass opacities and their laterality) to assess COVID-19
15 pneumonia, besides the enhanced sensitivity compared to RT-PCR ⁹.

16 Severity can be defined as an index that depicts the effects of a disease on mortality,
17 morbidity, and comorbidities ¹⁰ and has the potential to help physicians manage the patients
18 more decently whether in patients with cancer or with non-cancer diseases ^{11,12}. A number of
19 severity scoring systems have been proposed to quantify disease advancement in patients,
20 including general assessments (e.g., APACHE score) and disease-specific ones (e.g., Child-
21 Pugh score) ¹³. Several conventional scoring systems have been proposed for COVID-19
22 severity assessment ¹⁴. These include the usage of patient clinical, comorbidity, and laboratory
23 data, which are all helpful in constructing predictive models for severity assessment in
24 COVID-19 ¹⁵.

25 There has also been a growing interest in using imaging data of patients, such as thoracic
26 CT images. For example, a study by Sanders et al. ¹⁶ computed the score of CT images in
27 patients with cystic fibrosis and evaluated the prognostic ability. A promising line of research
28 that emerged recently reported on the CT severity index and its correlation with acute
29 pancreatitis severity ¹⁷⁻¹⁹. The COVID-19 Reporting and Data System (CO-RADS) was
30 suggested for standardized visual assessment of COVID-19 pneumonia to enhance agreement
31 between radiologists ²⁰. This system includes features for the diagnosis of COVID-19 and
32 consists of a 5-point scale for categorizing patient CT images. In addition, other guidelines
33 aiming to reach consensus when interpreting COVID-19 suspected chest CT images were
34 proposed ²¹. These guidelines are mostly based on visual assessment of images; e.g. the

1 amount to which lung lobes are involved, the volume of which is infected, and anatomical
2 assessments.

3 Francone et al.²² reported a study on the correlation between CT score and the severity of
4 coronavirus disease. Zhao et al.²³ also conducted research on the measurement of the extent
5 to which lung lobes are infected and evaluation in COVID-19 patients' prognosis. Li et al.²⁴
6 also confirmed the association between chest CT score and COVID-19 pneumonia severity.
7 At the same time, most scoring systems involve visual assessment and hence are time-
8 consuming^{23,24}. In this regard, medical image analysis using machine learning and radiomics
9 has been applied to quantify features to tackle these main challenges²⁵⁻³⁵.

10 The field of radiomics opens pathways for the study of normal tissues, cancer, and many
11 other diseases, including potentially the newly emerging COVID-19 disease^{6,7,29,36-40}.
12 Specifically, Xie et al.⁴¹ evaluated the potential of a radiomics framework to diagnose
13 COVID-19 from CT images. Di et al.⁴² also studied whether radiomics features can help to
14 distinguish between pneumonia of COVID-19 and that of other viral/bacterial causes. A
15 number of studies reported on the application of radiomics analysis to CT images towards
16 COVID-19 classification and prognostication⁴³. Homayounieh et al.⁴⁴ assessed the prognostic
17 power of CT-based radiomics features to determine severe and non-severe cases. In another
18 study, Li et al.⁴⁵ proposed a radiomics model based on CT images and classified patients
19 based on the criticality of their disease. A recent study by Yip et al.⁴⁶ applied a robust
20 radiomics model to CT images to predict the severity of COVID-19 disease in patients. All
21 above models pursued binary task performance, which reduced multiclass classification to
22 two class approaches. However, in the real clinical triage situation, scoring systems consist of
23 multi-class datasets. In the present study, involving a large cohort of patients, we aimed to
24 construct a CT radiomics-based multi-class classification model to predict the severity of
25 COVID-19 pneumonia.

26

27

28

29

30

31

32

1 MATERIALS AND METHODS

2 Data Description

3 Figure 1 presents the different steps performed in this study. All experiments were performed
4 in accordance with relevant guidelines and regulations.

5 Datasets and Segmentation

6 This study is based on the MosMed Dataset (mosmed.ai) consisting of 1110 patient CT scans,
7 also utilized in other efforts^{46,47}. Ethics approval and consent to participate **were** not needed
8 since the study was performed on open access online dataset. The patients were referred to the
9 Municipal Hospital in Moscow, Russia, and were classified based on clinical and visual CT
10 findings as follows.

11 In the zero class, the patient has neither clinical symptoms (e.g. fever) nor CT findings in
12 favor of any kind of pneumonia (Class 0, non-pneumonic). The 1st class contains patients who
13 have a low-temperature fever ($t < 38\text{ }^{\circ}\text{C}$) in addition to a mild increase in respiratory rate (RR
14 < 20) while showing none or $< 25\%$ ground-glass opacity (GGO) involvement (Class 1,
15 COVID-19 with mild severity). Patients in the 2nd class have a higher body temperature ($t >$
16 $38.5\text{ }^{\circ}\text{C}$) with a RR of 20-30, while CT scan shows 25-50% involvement of lung parenchyma
17 (Class 2, COVID-19 with moderate severity). Patients in the 3rd class have high body
18 temperature and RR of 30 or more, with CT findings of 50% to diffuse involvement in
19 addition to organ failure and shock signs (Class 3, severe COVID-19). Each of the classes,
20 namely 0, 1, 2, and 3, included 254, 684, 125, and 47 patients, respectively. The median age
21 was 47 (ranging from 18 to 97), and 42% of patients were female. Figure 2 shows an example
22 of representative CT images for each class.

23 All CT images were automatically segmented using a deep learning-based algorithm for
24 whole lung segmentation^{48,49}. After whole-lung 3D segmentation, all images were reviewed
25 and modified to ensure correct 3D-volume lung segmentation.

26

27 Image Preprocessing and Feature Extraction

28 All images were resized to isotropic voxel size $1 \times 1 \times 1\text{ mm}^3$ and image intensity was
29 discretized by 64-gray level binning, followed by feature extraction. The extracted features
30 from the whole-lung segmented regions, totalling 110, included shape ($n=16$), intensity
31 ($n=19$), and texture features, namely second-order texture of gray-level co-occurrence matrix
32 (GLCM, $n=24$), and high-order features, namely gray-level size-zone matrix (GLSZM, $n=16$),

1 neighbouring gray tone difference matrix (NGTDM, n=5), gray-level run-length matrix
2 (GLRLM, n=16) and gray-level dependence matrix (GLDM, n=14). Radiomics feature
3 extraction was performed using the Pyradiomics Python library ⁵⁰, which is compliant with
4 the Image Biomarker Standardization Initiative (IBSI) ⁵¹.

5 **Feature Selection and Classification and Evaluation**

6 In this study, we used two different feature selection algorithms, including Bagging Random
7 Forests (BRF) ⁵² and Multivariate Adaptive Regression Splines (MARS) ⁵³. BRF and MARS
8 algorithms were implemented in "VSURF" and "earth" R packages, respectively. For
9 multiclass classification, we implemented multinomial logistic regression using the "mnlogit"
10 R package. The MLR model fitness indices included p-value of the Wald test (corrected for
11 false-discovery rate via Benjamini and Hochberg method), pseudo R² (goodness of fit criteria
12 in a logistic regression model), as well as coefficient and Standard of Error (SE). In the MLR
13 model, class 0 served as a reference class whereas statistical comparison between two models
14 (the two feature selectors) was performed by the Likelihood Ratio Test. Ten-fold cross-
15 validation with bootstrapping (n=1000) was used to validate model performance. We report
16 precision, recall, F1-score, and accuracy for different class for each model. In addition, the
17 areas under the receiver operating characteristic (ROC) curve (AUCs) for multi-class
18 classification models were calculated and compared for both models using "multiROC" and
19 "pROC" R packages, respectively.

20
21
22
23
24
25
26
27
28
29

1 RESULTS

2 Table 1 summarizes the selected features and their importance value (IV) by BRF and MARS
3 for multiclass classification. Nineteen radiomics features were selected by BRF, including 9
4 from first-order, 6 from GLCM, one from GLDM, one from shape, one from NGTDM, and
5 one from GLSZM. Among these features, Mean Absolute Deviation (IV: 80%), Robust Mean
6 Absolute Deviation (IV: 72%) and kurtosis (IV: 70%) features from first-order, and
7 Correlation (IV: 75%), and Cluster Tendency (IV: 73%) features from GLCM were selected
8 as the most important ones. In the MARS algorithm, 10 features were selected with high IVs,
9 including 2 from first-order, 1 from GLDM, 2 from GLRLM, 2 from GLSZM, and 3 from
10 GLCM. The highest IV was achieved by Gray Level Variance from GLDM (IV: 94%), Zone
11 Entropy from GLSZM (IV: 93%), and Small Area Emphasis from GLSZM (IV: 83%). Mean
12 Absolute Deviation and Median from first-order features, Small Area Emphasis from
13 GLSZM, and Correlation from GLCM, were selected by both BRF and MARS algorithms.
14 Figure 3 depicts the feature map of different radiomics features in different classes. Figure 4
15 represents the feature selection process for multi-class classification by BRF and MARS.

16 Table 2 summarizes the adjusted p-value (by Benjamini and Hochberg method) of the
17 Wald test and coefficient (standard of error) for selected features by BRF and MARS
18 algorithms. Except for Inverse Variance from GLCM, all selected features yielded a
19 significant p-value (<0.05). BRF+MLR and MARS+MLR resulted in pseudo R^2 values of
20 0.295 and 0.256, respectively. However, there were no significant differences between both
21 models when using a likelihood ratio test (p-value =0.319).

22 Table 3 summarizes classification power indices (SD), including Precision, Recall, F1-
23 score, Accuracy, and AUC via multinomial logistic regression with 1000 bootstrapping
24 samples for each model. In terms of F1-score, classes 2 and 3 resulted in the lowest precision
25 (mean (sd)) in BRF+MLR (0.798 (0.106)) and MARS+MLR (0.752 (0.099)), whereas four-
26 class mean F1-scores were 0.847 and 0.805 for BRF+MLR and MARS+MLR algorithms,
27 respectively. The mean precision was 0.861 and 0.825, whereas the mean recall was 0.844
28 and 0.793 for BRF+MLR and MARS+MLR algorithms, respectively. BRF+MLR and
29 MARS+MLR algorithms achieved an accuracy of 0.933 and 0.922, respectively, in four-class
30 classification. AUCs (95% CI) for multi-class classification were 0.823 (0.795-0.852) and
31 0.816 (0.788-0.844) for BRF+MLR and MARS+MLR algorithms, respectively. Figure 5
32 depicts the ROC curves for our four-class classification method.

33

1 **DISCUSSION**

2 In the current study, we constructed a CT radiomics-based model to predict the severity of
3 COVID-19 patients in a large cohort of patients. To this end, we extracted radiomics features
4 from whole lung segmentations and selected high-importance features utilizing two different
5 algorithms, namely BRF and MARS. The selected features were then fed to a multinomial
6 logistic regression classifier for multiclass severity scoring. We achieved 0.823 (95% CI:
7 0.795-0.852) and 0.816 (95% CI: 0.788-0.844) for AUC, and 0.933 and 0.922 for accuracy in
8 BRF- and MARS-selected features, respectively.

9 We used an automatic model⁴⁸ to segment chest CT images for two reasons. First, most
10 CT scans performed in the COVID-19 pandemic era are low-dose. In addition, these scans are
11 acquired with a high pitch. Hence, it is difficult for radiologists to find and follow lung
12 fissures to manually detect or segment the anatomical lobes. As such, we used our previously
13 constructed deep learning model to fully segment the entire lung of each patient.

14 Yip et al.⁴⁶ conducted a study on the same dataset utilized in this work, aiming to evaluate
15 some radiomics features towards severity class prediction in patients. They included all 1110
16 patient CT scans and extracted 107 radiomics features. The maximum relevance minimum
17 redundancy (MRMR) and recursive feature elimination (RFE) algorithms were exploited for
18 feature selection and analysis of the selected features using univariate and multivariate
19 approaches using a logistic regression model to classify as accurately as possible. In their
20 study, the patients were categorized into three severity categories, namely mild, moderate, and
21 severe, to perform two-class classification tasks (mild vs. severe and moderate vs. severe) by
22 splitting the data into training (60%) and test (40%) sets. The authors obtained an AUC of
23 0.65 in differentiating between moderate and severe cases, while their model performed better
24 (AUC = 0.85) in distinguishing mild vs. severe forms of COVID-19 disease. In this work, we
25 reached an overall AUC of 0.823. In our study and the one by Yip et al.⁴⁶, feature extractions
26 were performed using Pyradiomics⁵⁰ as applied to the entire lung. Interestingly, there were
27 some commonly selected features arrived at via feature selection in both studies, including
28 Small Area Emphasis from GLSZM, Correlation and Informational Measure of Correlation
29 from GLCM, and Median and Mean Absolute Deviation from first-order features. These
30 selected features in both studies could potentially be used as predictors as they provide
31 information about the intensity and heterogeneity of the lung in COVID-19 patients.

32 A noticeable advantage of the study by Yip *et al.*⁴⁶ was the use of a second radiologist
33 observer who classified patients' images into mild, moderate, and severe classes without

1 paying attention to the default classification of the dataset provider. This method helped to
2 observe the prediction power of the models in both “provider” and “radiologist” datasets. In
3 addition, they split the dataset into training and test sets. In contrast, we applied the
4 bootstrapping technique to estimate and ensure the reproducibility of our results. In addition,
5 the study by Yip *et al.*⁴⁶ may have reduced generalizability as it only predicts mild versus
6 severe, and moderate versus severe disease, having reduced multiclass classification into two-
7 class approaches. In the real clinical triage situation, the radiologist may benefit from a
8 multiclass classification scheme for enhanced patient management, as provided by our study.

9 Homayounieh *et al.*⁵⁴ included 315 patients in their study and extracted CT-based
10 radiomics features from the lung to show that radiomics can predict patients’ outcome
11 (inpatient vs. outpatient management) with an AUC of 0.84 while the radiologist assessment
12 alone achieved an AUC of 0.69. Feature extraction was performed by applying the different
13 preprocessing algorithms on images, with classification performed using logistic regression.
14 They reported that adding clinical variables to the radiomics model can notably improve the
15 predictability of a model for patient outcome prediction (AUC improved from 0.75 to 0.84).
16 Another study conducted by Wei *et al.*⁵⁵ evaluated the predictive ability of two models (one
17 CT texture-based and one clinical) for determining the severity of each of the 81 COVID-19
18 patients. They showed that CT texture features could modestly predict whether the patient has
19 common COVID-19 pneumonia or a severe one with an AUC of 0.93, which is comparable to
20 that of the clinical-only model (AUC = 0.95). They also observed that several texture features
21 had a moderate correlation with the clinical variables of patients.

22 Chaganti *et al.*⁵⁶ studied Ground Glass Opacity (GGO) and consolidations that appear on a
23 CT image of COVID-19 patients in an attempt to propose an automated method for
24 segmenting and quantifying COVID-19 lesions. Their proposed method calculated the
25 percentage of opacity and lung severity score using deep learning algorithms and was able to
26 predict the severity with a decent performance. However, Chaganti *et al.*⁵⁶ proposed a
27 method trained only on the mentioned abnormalities and had a limited performance in other
28 abnormalities quantification. Even with improving segmentation algorithms, this method
29 would be limited because of the highly heterogeneous nature of COVID-19 pneumonia in
30 addition to ignoring the shape and texture of segmented lesions. Moreover, providing accurate
31 lobe segmentation of COVID-19 patients would be challenging from typical low-dose and
32 high pitch chest CT scans. In the current and previous studies^{44,46,55}, radiomics features, as
33 extracted from the entire lung (less challenging segmentation task for deep learning
34 algorithms), were evaluated to provide fast and robust severity scoring in COVID-19 patients.

1 In this work, chest CT was used for assessment. At the same time, there are few studies on
2 other modalities such as chest X-ray radiography in prognostication and outcome prediction
3 evaluation of COVID-19 patients. For example, Bae and colleagues⁵⁷ utilized radiomics
4 features and modeled them on chest X-rays of 514 patients and found out that their radiomics-
5 and deep learning-based model can accurately predict mortality and the need for mechanical
6 ventilation in patients (AUCs = 0.93 and 0.90, respectively). Providing a severity score using
7 chest X-rays is a valuable venue to explore. Yet, such work requires extensive comparisons
8 with CT-based frameworks to assess the relative value of each modality for different tasks.

9 This study suffered from a few limitations, including the fact that our model was trained on
10 single-center data. At the same time, we evaluated our models using a 10-fold cross-validation
11 and bootstrapping technique to evaluate the repeatability and robustness of our results. In any
12 case, further research should be conducted on multicentric data and patient images with
13 multiple observers for improved training of the models and enhanced generalizability.

14

15 **Conclusion**

16 We evaluated high-dimensional multinomial multiclass severity scoring of pneumonia using
17 CT radiomics features and machine learning algorithms. We applied two feature selectors
18 (BRF and MARS) coupled to one classifier (multiclass logistic regression model) on a large
19 cohort of COVID-19 patients. Our radiomics model was validated to depict accurate
20 classification of patients according to multi-class pneumonia severity assessment criteria,
21 highlighting the potential of this emerging paradigm in the assessment and management of
22 COVID-19 patients.

23

24 **Acknowledgments**

25 This work was supported by the Swiss National Science Foundation under grant SNRF
26 320030_176052.

27

28 **Conflict of Interest statement**

29 The authors declare that they have no conflict of interest.

30

31

1

2 **Abbreviations**

3 CT: Computed Tomography

4 COVID-19: Coronavirus disease 2019

5 AUC: Area under the receiver operating characteristic curve

6 BRF: Bagging Random Forest

7 FS: Feature Selection

8 GGO: Ground Glass Opacity

9 IBSI: The Image Biomarker Standardization Initiative

10 MARS: Multivariate Adaptive Regression Splines

11 MLR: Multinomial Logistic Regression

12 RT-PCR: Reverse transcription polymerase chain reaction

13 GLCM: Gray-Level Co-Occurrence Matrix

14 GLSZM: Gray-Level Size-Zone Matrix

15 NGTDM: Neighbouring Gray Tone Difference Matrix

16 GLRLM: Gray-Level Run-Length Matrix

17 GLDM: Gray-Level Dependence Matrix

18

19

20

21

22

23

24

25

26

27

28

29

Table 1. Selected features by Bagging Random Forests (BRF) (“VSURF” R package) and multivariate adaptive regression splines (MARS) (“earth” R package) for multi-class classification

Feature Selection Algorithm	Selected Variables	Feature type	Importance Value
Bagging Random Forests	First Order	Mean Absolute Deviation	80%
	GLCM	Correlation	75%
	GLCM	Cluster Tendency	73%
	First Order	Robust Mean Absolute Deviation	72%
	First Order	Variance	70%
	First Order	Interquartile Range	71%
	First Order	Kurtosis	72%
	First Order	Skewness	71%
	GLDM	Dependence Entropy	68%
	First Order	Median	48%
	First Order	Entropy	69%
	GLCM	Sum Entropy	67%
	GLCM	Inverse Variance	42%
	GLCM	Joint Entropy	66%
	First Order	Uniformity	63%
	GLCM	Contrast	62%
	Shape	Major Axis Length	30%
	NGTDM	Contrast	60%
GLSZM	Small Area Emphasis	50%	
Multivariate Adaptive Regression Splines	First Order	Mean Absolute Deviation	22%
	GLDM	Gray Level Variance	94%
	GLRLM	Gray Level Non Uniformity	57%
	GLRLM	Short Run Emphasis	70%
	GLSZM	Small Area Emphasis	83%
	First Order	Median	24%
	GLCM	Imc l	46%
	GLSZM	Zone Entropy	86%
	GLCM	Correlation	38%
	GLCM	Difference Entropy	40%

Table 2. Model fitness indices for application of multinomial logistic regression (MLR) (“mnlogit” R package) to selected features.

Feature Selection Algorithm	Feature type		Adj. P-value	Coefficient (SE)	Pseudo R ² for logistic regression
Bagging Random Forests	First order	Mean Absolute Deviation	0.001	0.066 (0.007)	0.295
	GLCM	Correlation	0.001	14.3 (1.85)	
	GLCM	Cluster Tendency	0.002	0.20 (0.024)	
	First Order	Robust Mean Absolute Deviation	0.001	0.093 (0.10)	
	First Order	Variance	0.001	17×10 ⁻⁴ (2×10 ⁻⁵)	
	First Order	Interquartile Range	0.001	0.042 (0.0047)	
	First Order	Kurtosis	0.002	-0.35 (0.042)	
	First Order	Skewness	0.001	-2.44 (0.269)	
	GLDM	Dependence Entropy	0.001	4.58 (0.585)	
	First order	Median	0.002	0.022 (0.0031)	
	First order	Entropy	0.001	4.43 (0.495)	
	GLCM	Sum Entropy	0.001	4.14 (0.461)	
	GLCM	Inverse Variance	0.790	-2.17 (8.15)	
	GLCM	Joint Entropy	0.001	2.42 (0.270)	
	First Order	Uniformity	0.001	-21.8 (2.60)	
	GLCM	Contrast	0.002	1.18 (0.136)	
	Shape	Major Axis Length	0.005	-0.012 (0.016)	
	NGTDM	Contrast	0.028	171.4 (20.60)	
GLSZM	Small Area Emphasis	0.031	-50.2 (7.71)		
Multivariate Adaptive Regression Splines	First Order	Mean Absolute Deviation	0.001	0.066 (0.007)	0.256
	GLDM	Gray Level Variance	<0.001	0.705 (0.082)	
	GLRLM	Gray Level Non Uniformity	<0.001	-0.006 (8×10 ⁻⁶)	
	GLRLM	Short Run Emphasis	<0.001	42.2 (4.90)	
	GLSZM	Small Area Emphasis	0.001	-50.2 (7.0)	
	First Order	Median	0.001	0.022 (0.003)	
	GLCM	Imc1	0.001	-44.8 (6.82)	
	GLSZM	Zone Entropy	0.001	6.3 (0.76)	
	GLCM	Correlation	0.001	14.3 (1.85)	
	GLCM	Difference Entropy	0.001	8.7 (0.98)	

P-value by Wald chi-square test; Adj. P-value: P-value adjusted by Benjamini and Hochberg false discovery (FDR) correction method; class 1 as a reference class; SE= Standard of Error; statistical comparison between two models showed non-significant difference by Likelihood Ratio Test: P-value=0.319.

Table 3. Classification performance indices by multinomial logistic regression with 1000 bootstrapping samples based feature selection. SD is shown in brackets.

Feature Selection Algorithm	Class	Precision	Recall	F1-score	Accuracy	AUC (95% CI)
Bagging Random Forests	Class 1	0.743 (0.078)	0.948 (0.093)	0.833 (0.099)	0.912 (0.098)	0.823 (0.795-0.852)
	Class 2	0.926 (0.022)	0.861 (0.017)	0.892 (0.017)	0.872 (0.015)	
	Class 3	0.866 (0.101)	0.739 (0.099)	0.798 (0.106)	0.958 (0.110)	
	Class 4	0.906 (0.10)	0.829 (0.097)	0.866 (0.096)	0.989 (0.104)	
	Average/total	0.861	0.844	0.847	0.933	
Multivariate Adaptive Regression Splines	Class 1	0.732 (0.095)	0.862 (0.101)	0.792 (0.081)	0.895 (0.126)	0.816 (0.788-0.844)
	Class 2	0.901 (0.019)	0.865 (0.014)	0.883 (0.026)	0.859 (0.017)	
	Class 3	0.824 (0.079)	0.764 (0.111)	0.793 (0.091)	0.955 (0.123)	
	Class 4	0.842 (0.112)	0.680 (0.108)	0.752 (0.099)	0.980 (0.106)	
	Average/total	0.825	0.793	0.805	0.922	

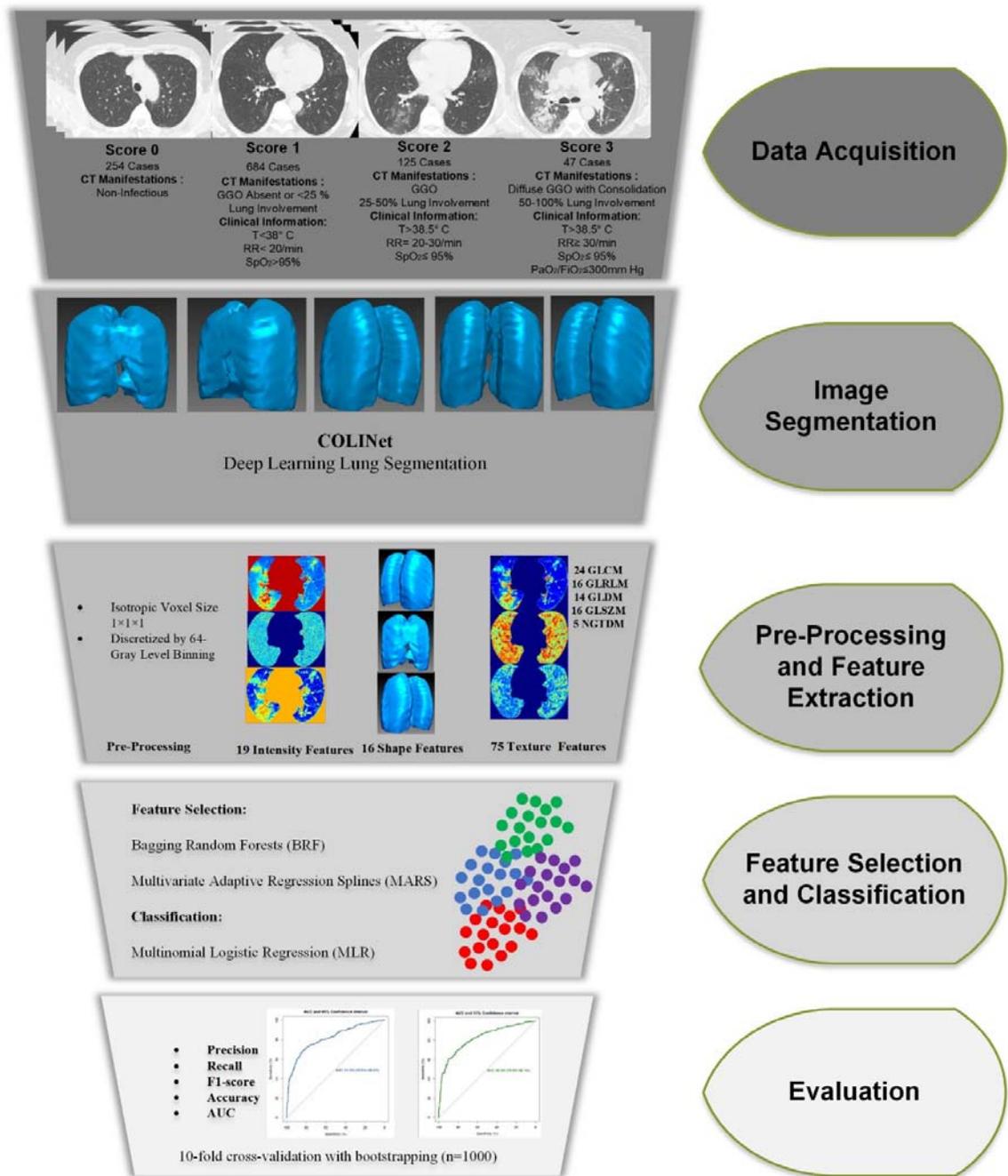


Figure 1: Different steps of current study. GGO : ground glass opacities, T: Temperature, RR: Respiratory Rate, SpO₂: Peripheral Capillary Oxygen Saturation, PaO₂: Partial Pressure of Oxygen. FiO₂=Fraction of Inspired Oxygen.

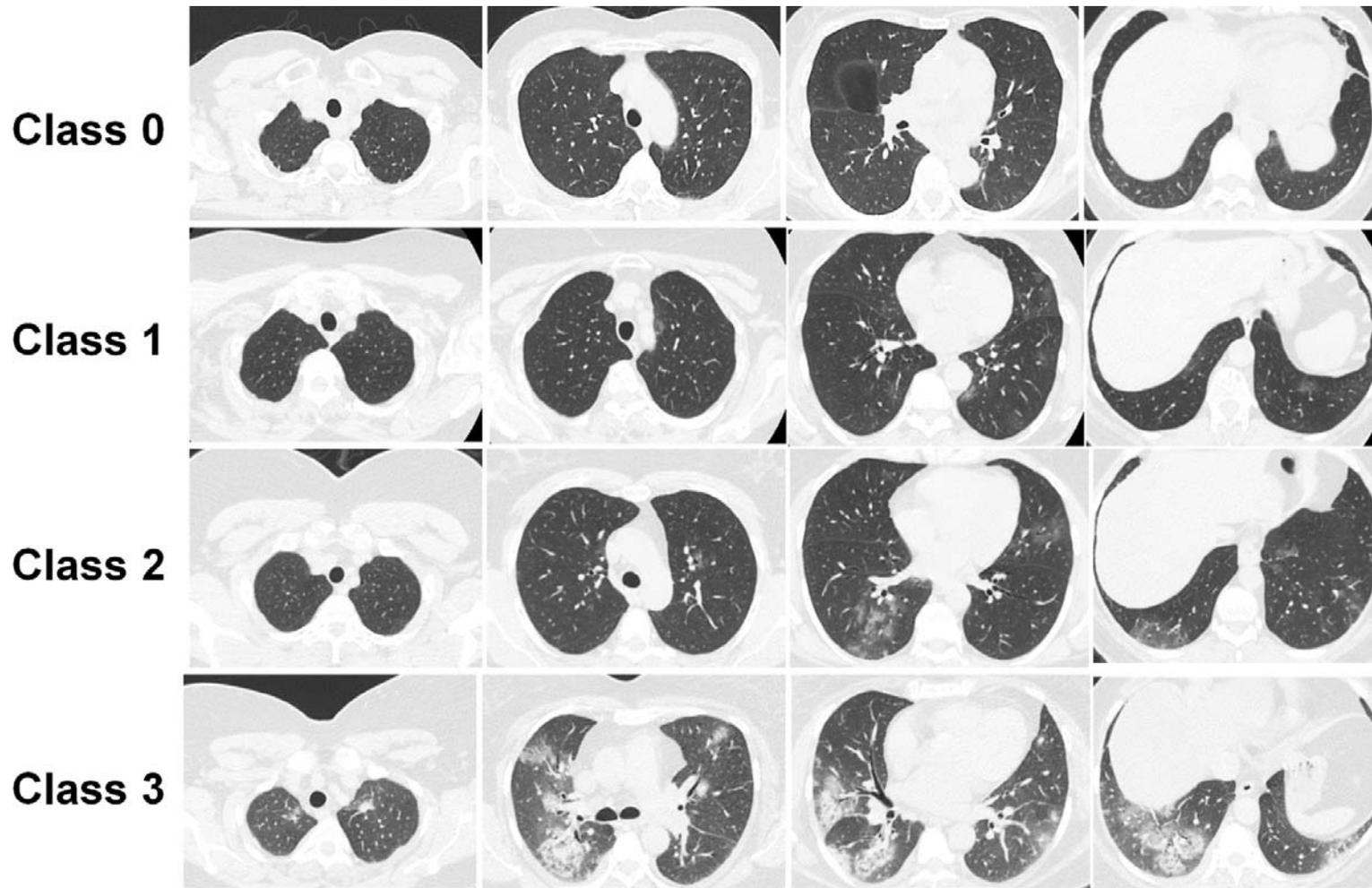


Figure 2: Example of patient CT images for different class.

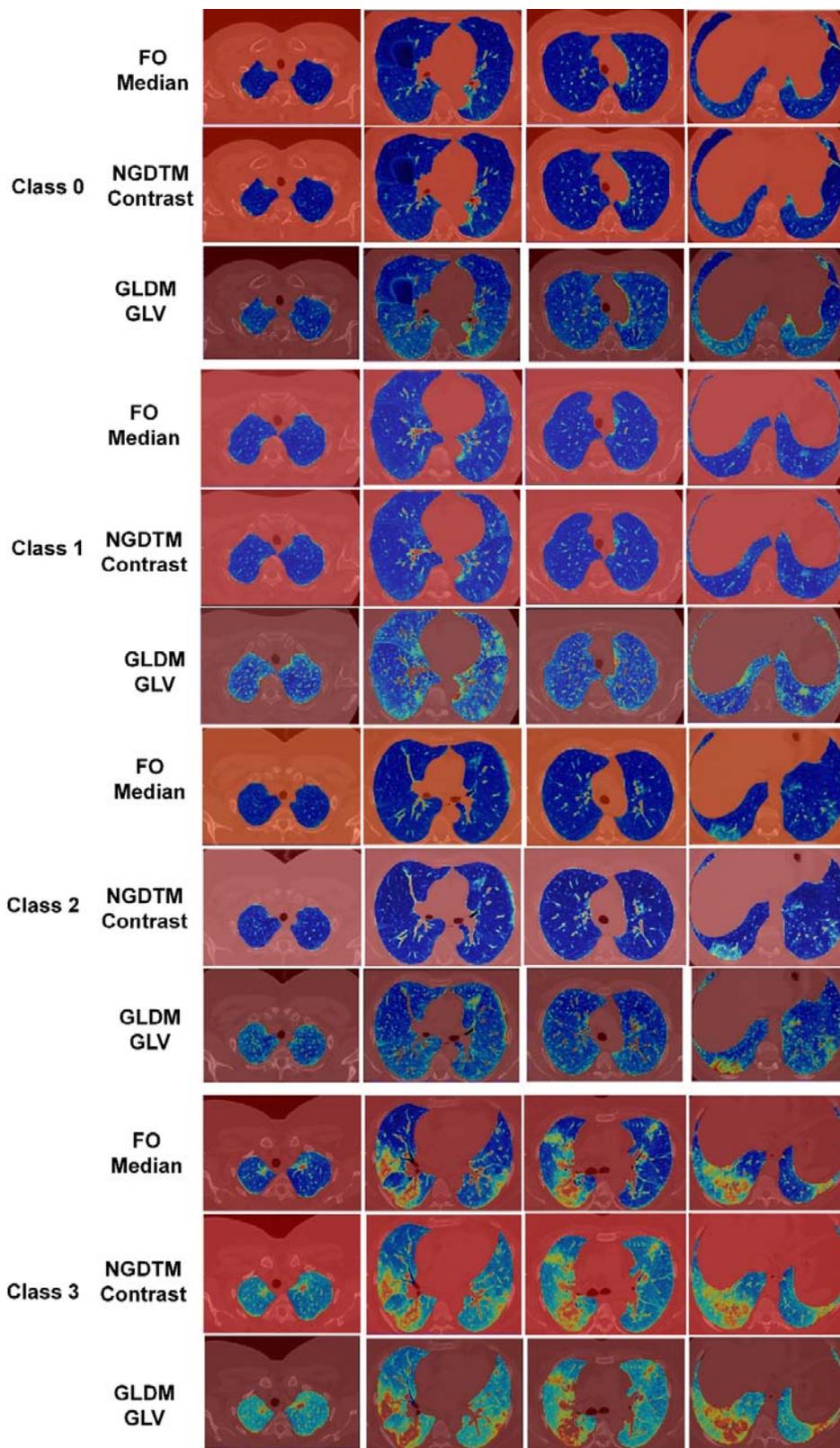


Figure 3: Example of selected features (Median feature from First Order, Contrast feature from NGDTM and GLV features from GLDM) in different class cases.

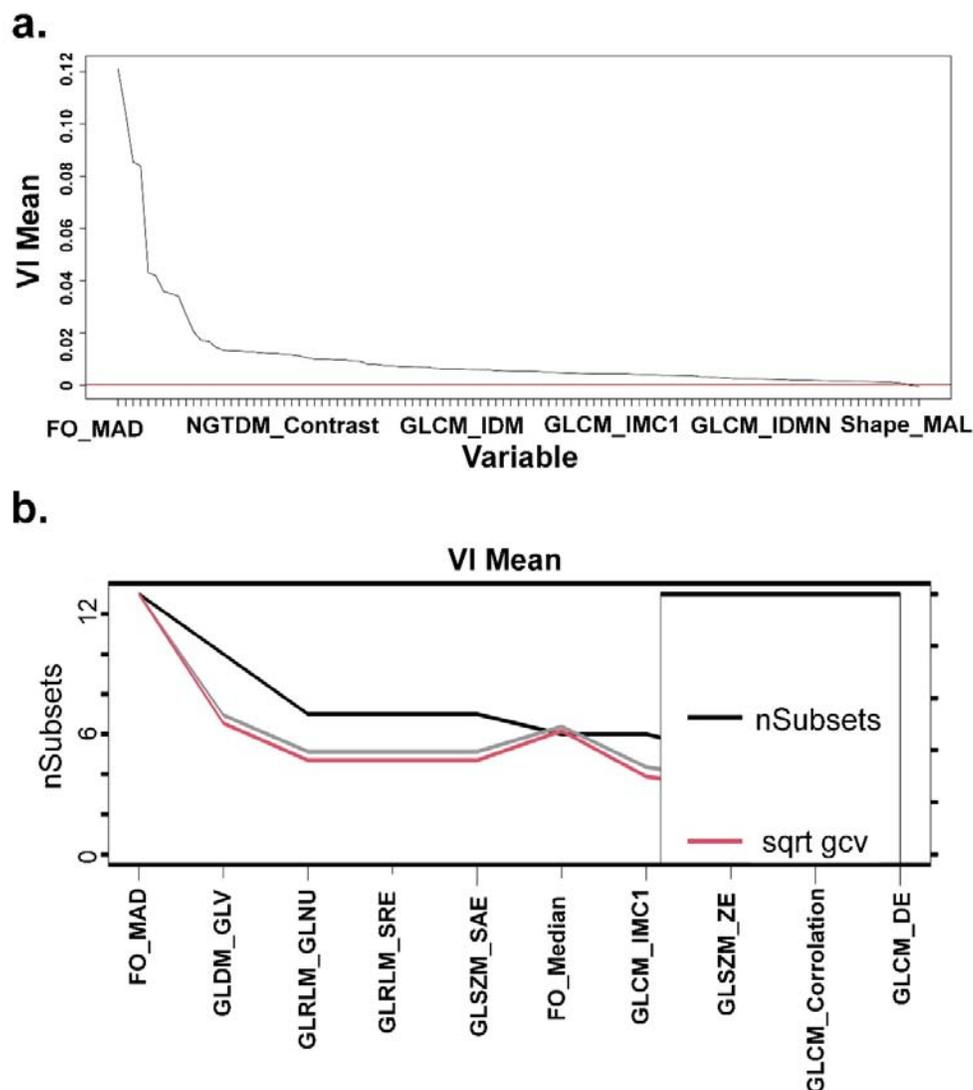


Figure 4: Feature selection process for multi-class classification by (a) Bagging Random Forests (number of selected features=19), and (b) Multivariate Adaptive Regression Splines (number of selected features=10).

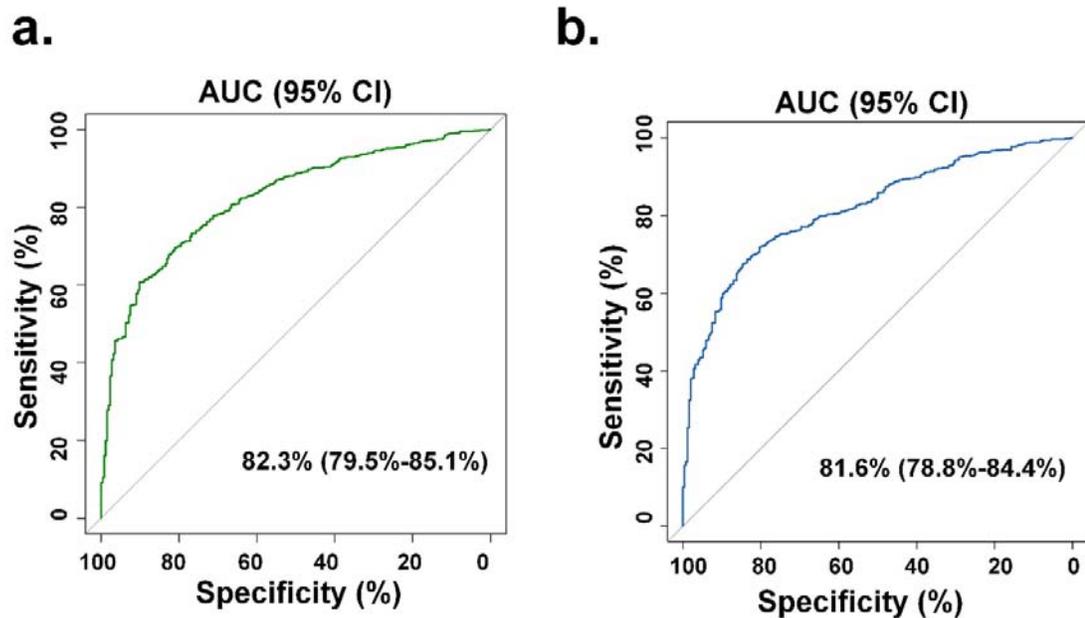


Figure 5: (a) ROC curve for assessing power of multi-class classification of the selected features in Bagging Random Forests (AUC=0.823), and (b) Multivariate Adaptive Regression Splines (AUC=0.816). Statistical comparison of ROC curves by “pROC” R package indicated non-significant difference ($Z=-1.164$, $P\text{-value}=0.244$).

REFERENCES

1. Meyerowitz-Katz, G. & Merone, L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int J Infect Dis* **101**, 138-148 (2020).
2. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C. & Di Napoli, R. Features, Evaluation, and Treatment of Coronavirus. in *StatPearls* (StatPearls Publishing Copyright © 2020, StatPearls Publishing LLC., Treasure Island (FL), 2020).
3. Corman, V.M., *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **25**(2020).
4. La Marca, A., *et al.* Testing for SARS-CoV-2 (COVID-19): a systematic review and clinical guide to molecular and serological in-vitro diagnostic assays. *Reprod Biomed Online* **41**, 483-499 (2020).
5. Schmidt, C.W. CT scans: balancing health risks and medical benefits. *Environ Health Perspect* **120**, A118-A121 (2012).
6. Shiri, I., *et al.* COVID-19 prognostic modeling using CT radiomic features and machine learning algorithms: Analysis of a multi-institutional dataset of 14,339 patients. *Comput Biol Med* **145**, 105467 (2022).
7. Shiri, I., *et al.* Diagnosis of COVID-19 Using CT image Radiomics Features: A Comprehensive Machine Learning Study Involving 26,307 Patients. *medRxiv* (2021).
8. Li, Y. & Xia, L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. *AJR. American journal of roentgenology* **214**, 1280-1286 (2020).
9. Long, C., *et al.* Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *European journal of radiology* **126**, 108961 (2020).
10. Gambert, S. Disease Severity. in *Encyclopedia of Behavioral Medicine* (eds. Gellman, M.D. & Turner, J.R.) 606-606 (Springer New York, New York, NY, 2013).
11. Tai, S.Y., *et al.* Symptom severity of patients with advanced cancer in palliative care unit: longitudinal assessments of symptoms improvement. *BMC palliative care* **15**, 32 (2016).
12. Fjerstad, M., Trussell, J., Lichtenberg, E.S., Sivin, I. & Cullins, V. Severity of infection following the introduction of new infection control measures for medical abortion. *Contraception* **83**, 330-335 (2011).
13. Bouch, D.C. & Thompson, J.P. Severity scoring systems in the critically ill. *Continuing Education in Anaesthesia Critical Care & Pain* **8**, 181-185 (2008).
14. Robilotti, E.V., *et al.* Determinants of COVID-19 disease severity in patients with cancer. *Nature Medicine* **26**, 1218-1223 (2020).
15. Li, X., *et al.* Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *The Journal of allergy and clinical immunology* **146**, 110-118 (2020).
16. Sanders, D.B., Li, Z., Brody, A.S. & Farrell, P.M. Chest computed tomography scores of severity are associated with future lung disease progression in children with cystic fibrosis. *Am J Respir Crit Care Med* **184**, 816-821 (2011).
17. Sahu, B., *et al.* Severity assessment of acute pancreatitis using CT severity index and modified CT severity index: Correlation with clinical outcomes and severity grading as per the Revised Atlanta Classification. *Indian J Radiol Imaging* **27**, 152-160 (2017).
18. Raghuwanshi, S., Gupta, R., Vyas, M.M. & Sharma, R. CT Evaluation of Acute Pancreatitis and its Prognostic Correlation with CT Severity Index. *J Clin Diagn Res* **10**, TC06-TC11 (2016).

19. Alhajeri, A. & Erwin, S. Acute pancreatitis: value and impact of CT severity index. *Abdominal imaging* **33**, 18-20 (2008).
20. Prokop, M., *et al.* CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19-Definition and Evaluation. *Radiology* **296**, E97-e104 (2020).
21. Neri, E., *et al.* Structured reporting of chest CT in COVID-19 pneumonia: a consensus proposal. *Insights into Imaging* **11**, 92 (2020).
22. Francone, M., *et al.* Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. *Eur Radiol* **30**, 6808-6817 (2020).
23. Zhao, W., Zhong, Z., Xie, X., Yu, Q. & Liu, J. Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study. *AJR. American journal of roentgenology* **214**, 1072-1077 (2020).
24. Li, K., *et al.* The Clinical and Chest CT Features Associated With Severe and Critical COVID-19 Pneumonia. *Investigative radiology* **55**, 327-331 (2020).
25. Yip, S.S.F. & Aerts, H.J.W.L. Applications and limitations of radiomics. *Physics in medicine and biology* **61**, R150-R166 (2016).
26. Cunliffe, A., *et al.* Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *International journal of radiation oncology, biology, physics* **91**, 1048-1056 (2015).
27. Nazari, M., Shiri, I. & Zaidi, H. Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients. *Comput Biol Med* **129**, 104135 (2020).
28. Mostafaei, S., *et al.* CT imaging markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking regression algorithm. *La radiologia medica* **125**, 87-97 (2020).
29. Shiri, I., *et al.* Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients. *Comput Biol Med* **132**, 104304 (2021).
30. Shayesteh, S., *et al.* Treatment response prediction using MRI-based pre-, post-, and delta-radiomic features and machine learning algorithms in colorectal cancer. *Med Phys* **48**, 3691-3701 (2021).
31. Amini, M., *et al.* Multi-level multi-modality (PET and CT) fusion radiomics: prognostic modeling for non-small cell lung carcinoma. *Phys Med Biol* **66**(2021).
32. Khodabakhshi, Z., *et al.* Overall Survival Prediction in Renal Cell Carcinoma Patients Using Computed Tomography Radiomic and Clinical Information. *J Digit Imaging* **34**, 1086-1098 (2021).
33. Khodabakhshi, Z., *et al.* Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature. *Comput Biol Med* **136**, 104752 (2021).
34. Shiri, I., *et al.* Impact of feature harmonization on radiogenomics analysis: Prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images. *Comput Biol Med* **142**, 105230 (2022).
35. Rahmim, A., *et al.* Tensor Radiomics: Paradigm for Systematic Incorporation of Multi-Flavoured Radiomics Features. *arXiv preprint arXiv:2203.06314* (2022).
36. Liu, Z., *et al.* The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* **9**, 1303-1322 (2019).
37. Edalat-Javid, M., *et al.* Cardiac SPECT radiomic features repeatability and reproducibility: A multi-scanner phantom study. *Journal of Nuclear Cardiology* (2020).

38. Abdollahi, H., Shiri, I. & Heydari, M. Medical Imaging Technologists in Radiomics Era: An Alice in Wonderland Problem. *Iran J Public Health* **48**, 184-186 (2019).
39. Amini, M., *et al.* Overall Survival Prognostic Modelling of Non-small Cell Lung Cancer Patients Using Positron Emission Tomography/Computed Tomography Harmonised Radiomics Features: The Quest for the Optimal Machine Learning Algorithm. *Clin Oncol (R Coll Radiol)* **34**, 114-127 (2022).
40. Avard, E., *et al.* Non-contrast Cine Cardiac Magnetic Resonance image radiomics features and machine learning algorithms for myocardial infarction detection. *Comput Biol Med* **141**, 105145 (2022).
41. Xie, C., *et al.* Discrimination of pulmonary ground-glass opacity changes in COVID-19 and non-COVID-19 patients using CT radiomics analysis. *European journal of radiology open* **7**, 100271 (2020).
42. Di, D., *et al.* Hypergraph learning for identification of COVID-19 with CT imaging. *Medical image analysis* **68**, 101910 (2020).
43. Bouchareb, Y., *et al.* Artificial intelligence-driven assessment of radiological images for COVID-19. *Computers in biology and medicine*, 104665 (2021).
44. Homayounieh, F., *et al.* Computed Tomography Radiomics Can Predict Disease Severity and Outcome in Coronavirus Disease 2019 Pneumonia. *J Comput Assist Tomogr* **44**, 640-646 (2020).
45. Wang, Y., *et al.* Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. *Radiology* **296**, E55-e64 (2020).
46. Yip, S.S.F., *et al.* Performance and Robustness of Machine Learning-based Radiomic COVID-19 Severity Prediction. *medRxiv : the preprint server for health sciences*, 2020.2009.2007.20189977 (2020).
47. Jin, C., *et al.* Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nature communications* **11**, 5088 (2020).
48. Shiri, I., *et al.* COLI-NET: Fully Automated COVID-19 Lung and Infection Pneumonia Lesion Detection and Segmentation from Chest CT Images. *medRxiv*, 2021.2004.2008.21255163 (2021).
49. Shiri, I., *et al.* COLI-Net: Deep learning-assisted fully automated COVID-19 lung and infection pneumonia lesion detection and segmentation from chest computed tomography images. *Int J Imaging Syst Technol* (2021).
50. van Griethuysen, J.J.M., *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* **77**, e104-e107 (2017).
51. Zwanenburg, A., *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328-338 (2020).
52. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. VSURF: an R package for variable selection using random forests. (2015).
53. Zhang, W., Wu, C., Li, Y., Wang, L. & Samui, P. Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 1-14 (2019).
54. Homayounieh, F., *et al.* CT Radiomics, Radiologists, and Clinical Information in Predicting Outcome of Patients with COVID-19 Pneumonia. *Radiology: Cardiothoracic Imaging* **2**, e200322 (2020).
55. Wei, W., Hu, X.W., Cheng, Q., Zhao, Y.M. & Ge, Y.Q. Identification of common and severe COVID-19: the value of CT texture analysis and correlation with clinical characteristics. *Eur Radiol* **30**, 6788-6796 (2020).

56. Chaganti, S., *et al.* Automated Quantification of CT Patterns Associated with COVID-19 from Chest CT. *Radiology: Artificial Intelligence* **2**, e200048 (2020).
57. Bae, J., *et al.* Predicting Mechanical Ventilation Requirement and Mortality in COVID-19 using Radiomics and Deep Learning on Chest Radiographs: A Multi-Institutional Study. *ArXiv* (2020).