TITLE PAGE

Title:

Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: A multicenter study

Authors and affiliations:

Xiaolong Qi[1]#*, Zicheng Jiang[2]#, Qian Yu[3]#, Chuxiao Shao[4], Hongguang Zhang[5], Hongmei Yue[1], Baoyi Ma[6], Yuancheng Wang[3], Chuan Liu[1], Xiangpan Meng[3], Shan Huang[3], Jitao Wang[1], Dan Xu[1], Junqiang Lei[1], Guanghang Xie[1], Huihong Huang[2], Jie Yang[4], Jiansong Ji[4], Hongqiu Pan[5], Shengqiang Zou[5], Shenghong Ju[3]*

1. CHESS-COVID-19 center, The First Hospital of Lanzhou University, Lanzhou, China
2. CHESS-COVID-19 center, Ankang Central Hospital, Ankang, China
3. Department of Radiology, Jiangsu Key Laboratory of Molecular and Functional Imaging, Zhongda Hospital, Medical School, Southeast University, Nanjing, China
4. CHESS-COVID-19 center, Zhejiang University Lishui Hospital & Lishui Central Hospital, Lishui, China
5. Department of Infectious Diseases and Critical Care Medicine, The Affiliated Third Hospital of Jiangsu University, Zhenjiang, China
6. Department of Respiratory Medicine, The People's Hospital of LinXia Hui Prefecture, Linxia, China

# These authors contributed equally to this work.

Correspondence to:

Xiaolong Qi, MD, Professor of Medicine

CHESS-COVID-19 center, The First Hospital of Lanzhou University, Lanzhou, China

Phone: 86-18588602600

E-mail: qixiaolong@vip.163.com

OR

Shenghong Ju, MD, Professor of Medicine

Department of Radiology, Zhongda Hospital, Medical School of Southeast University, Nanjing, China

E-mail: jsh0836@126.com

Author contributions:

Xiaolong Qi, Zicheng Jiang and Qian Yu contributed equally.

Concept and design: Xiaolong Qi, Shenghong Ju. Acquisition of data: Zicheng Jiang, Chuxiao Shao, Hongguang Zhang, Hongmei Yue, Baoyi Ma, Jitao Wang, Junqiang Lei, Huihong Huang, Jie Yang, Jiansong Ji, Hongqiu Pan, Shengqiang Zou. Analysis and interpretation of data: Qian Yu, Chuan Liu, Yuancheng Wang, Xiangpan Meng, Shan Huang, Dan Xu, Guanghang Xie, Xiaolong Qi, Shenghong Ju. Drafting of the manuscript: Qian Yu, Xiaolong Qi. Critical revision of the manuscript for important intellectual content: Xiaolong Qi, Shenghong Ju. Administrative, technical, or material support: Xiaolong Qi, Zicheng Jiang, Chuxiao Shao, Hongguang Zhang, Hongmei Yue, Hongqiu Pan. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Financial support:

The study was supported by Gansu Provincial COVID-19 Science and Technology Major Project; Shanxi Provincial Emergency Research Project for Chinese Medicine for Prevention and Treatment of COVID-19 (2020-YJ005); Zhenjiang Key Research and Development Plan for COVID-19 Emergency Project (SH2020001)。

Competing interests:

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the

submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing:

The imaging or algorithm data used in this study are available upon request.

Transparency statement:

Lead author and the manuscript's guarantor affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained.

Word count:

1356 words

Abstract

Objectives

To develop and test machine learning-based CT radiomics models for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection.

Design

Cross-sectional

Setting

Multicenter

Participants

A total of 52 patients with laboratory-confirmed SARS-CoV-2 infection and their initial CT images were enrolled from 5 designated hospitals in Ankang, Lishui, Zhenjiang, Lanzhou, and Linxia between January 23, 2020 and February 8, 2020. As of February 20, patients remained in hospital or with non-findings in CT were excluded. Therefore, 31 patients with 72 lesion segments were included in the final analysis.

Intervention

CT radiomics models based on logistic regression (LR) and random forest (RF) were developed on features extracted from pneumonia lesions in training and inter-validation

datasets. The predictive performance was further evaluated in test dataset on lung lobe- and patients-level.

Main outcomes

Short-term hospital stay (≤10 days) and long-term hospital stay (>10 days).

Results

The CT radiomics models based on 6 second-order features were effective in discriminating short- and long-term hospital stay in patients with pneumonia associated with SARS-CoV-2 infection, with areas under the curves of 0.97 (95%CI 0.83-1.0) and 0.92 (95%CI 0.67-1.0) by LR and RF, respectively, in the test dataset. The LR model showed a sensitivity and specificity of 1.0 and 0.89, and the RF model showed similar performance with sensitivity and specificity of 0.75 and 1.0 in test dataset.

Conclusions

The machine learning-based CT radiomics models showed feasibility and accuracy for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection.

Introduction

The coronavirus disease 2019 (COVID-19) from Wuhan, China has become a global challenge since the December 2019.[1-3] Clinical characteristics of patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection have been reported, and the median hospital stay of forty-seven discharged patients was 10 days.[1] Studies showed that patients' condition in Wuhan worsens on the 10th day after illness onset.[2] However, patients with symptoms longer than 10 days outside Wuhan were less severe than those in Wuhan.[3] Therefore, the hospital stay in patients with SARS-CoV-2 infection is one of the prognostic indicators, and its non-invasive predicting tool is important for assessing the patients' clinical outcome.

Chest CT is recommended as a routine test in the diagnoses and monitoring of COVID-19 since ground-glass opacities and consolidation are the most relative imaging features in pneumonia associated with SARS-CoV-2 infection.[4-6] On the basis of our previous work utilizing quantitative CT for COVID-19,[7] we hypothesized that high-throughout information hidden behind CT images[8] had potential in discriminating the hospital stay. The study aimed to develop and test machine learning-based CT radiomics models for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection.

Methods

Study population

The multicenter study was conducted according to principles of the Declaration of Helsinki and approved by all institutional review board. The need for written informed consent from the participants was waived. Patients with laboratory-confirmed SARS-CoV-2 infection and their initial CT images were enrolled from 5 designated hospitals between January 23, 2020 and February 8, 2020, with final follow-up on February 20, 2020 (**Figure 1**). Most patients received antiviral treatment with interferon inhalation, lopinavir and ritonavir, combined with probiotics. Patients were discharged once the results of two real-time fluorescence polymerase-chain-reaction tests taken 24 hours apart were negative for SARS-CoV-2 antigens. Patients without pneumonia findings or those remained in hospital were excluded. Sample size consideration was shown in supplementary file. In this study, the optimal cut-off value of hospital stay was determined to be 10 days based on previous studies[1-3], by which patients were classified into short-term hospital stay (≤10 days) and long-term hospital stay (>10 days).

CT radiomics features and model building

The pipeline of radiomics model was shown (**Figure 1**), and features extraction and model building were performed on lung lobe-level. Images preprocessing was shown in supplementary file. Images containing lesions were segmented using Python (3.6, https://www.python.org) and 3Dslicer (version 4.10.0; https://www.slicer.org/) with two

steps. First, lung lobes of each patient were segmented automatically using algorithms based on U-net[9], and results were checked and modified by one radiologist (Q.Y). Next, lesions in each lung lobe were labeled semi-automatically using serval seeds placed within lesion region to generate the contours. Three radiologists (S.H, Q.Y and X.P.M) evaluated segments of each lesion and reached a consensus. All imaging processes were blinded to clinical data.

In total, 1218 features were calculated per lesion patch. First-order, shape and second-order features were extracted from original images and wavelet filter applied images using pyradiomics.[10] Two supervised learning algorithms, logistic regression (LR) and random forest (RF), were used to build the model and verify the robustness of features (supplementary file).[11] We applied 5-fold cross-validation on the training dataset to prove model performance.

Model performance

The cutoff point was defined on receiver operating characteristic (ROC) curves of training data by maximizing the sum of sensitivity and specificity. The model performance was evaluated using test dataset on lung lobe-level. Areas under the ROC curve (AUC), sensitivities, specificities, positive predictive value (PPV), and negative predictive value (NPV) were recorded. On patient-level, one was defined as long-term hospital stay once more than one lesion of lung lobe was labeled as long-term stay lesion, if not, as short-term hospital stay.

Statistical analysis

Categorical data were expressed in numbers (percentages), and continuous variables as median (interquartile range). Demographic and clinical characteristics of patients were assessed using Chi square test (Fisher's exact test as appropriate) for categorical variables, and Mann-Whitney test for continuous variables in SPSS (version 22.0. IBM Crop. Armonk, NY, USA). Feature selection and model building were implemented with FeAture Explorer (FAE, v0.2.5, https://github.com/salan668/FAE) on Python (3.6.8, https://www.python.org/)). Test values like areas under the receiver operating characteristic curves (95% confidence interval), sensitivity, specificity was calculated in SPSS and Python. A P-value < 0.05 was considered statistically significant.

Patient and Public Involvement

This cross-sectional study did not involve patients in study design, outcome measures, or writing or editing of this study.

Results

Patient characteristics

A total of 52 patients with laboratory-confirmed SARS-CoV-2 infection and initial CT images were enrolled from 5 designated hospitals in Ankang, Lishui, Zhenjiang, Lanzhou, and Linxia, China. As of February 20, 14 patients were still hospitalized, and 7 patients had non-findings in CT images. Therefore, 31 patients with 72 lesion segments were included in the final analysis. The training and inter-validation cohort comprised 26 patients (12 from Ankang, 8 from Lishui, 4 from Lanzhou, and 2 from Linxia) with 59 lesion segments, and test cohort comprised 5 patients from Zhenjiang with 13 lesion segments. The median age was 38.00 (interquartile range, 26.00-47.00) years and 17 (57%) were male. Comorbidities, symptoms and laboratory findings at admission were summarized in **Table 1**.

Performance of CT radiomics model

The CT radiomics model, based on 6 features (**supplementary Table1**), showed the highest AUC on the training and inter-validation dataset. The performance of modeling using LR and RF methods was shown in **Figure 2**. On lung lobe-level, models using LR method significantly distinguished short- and long-term hospital stay (In training and inter-validation datasets, cut-off value 0.31, AUC 0.94 (95%CI 0.92-0.97), sensitivity 1.0, specificity 0.87, NPV 1.0, and PPV 0.88; In test dataset, AUC 0.97 (95%CI 0.83-1.0), sensitivity 1.0, specificity 0.89, NPV 1.0, and PPV 0.8). Besides, models using RF method obtained satisfied results (In training and inter-validation datasets, cut-off value 0.68, AUC 1.0 (95%CI 1.0-1.0),

sensitivity 1.0, specificity 1.0, NPV 1.0, and PPV 1.0; In test dataset, AUC 0.92 (95%CI 0.67-1.0), sensitivity 0.75, specificity 1.0, NPV 0.9, and PPV 1.0).

On patients-level, in training and inter-validation datasets, 6 of 6 patients were correctly classified as short-term stay by both models, and 20 of 20, 16 of 20 patients were correctly identified as long-term stay by RF and LR models, respectively. In test dataset, 1 of 2 patients was correctly classified as short-term stay, and 3 of 3 were correctly identified as long-term stay by RF and LR models.

As of February 28, we followed up for a prospective dataset of 6 newly discharged patients with 24 lesions from designated hospitals (Comorbidities, symptoms and laboratory findings were described in supplementary Table 2). All patients were correctly recognized as long-term stay using both RF and LR models developed on raw 52 patients.

Discussion

In this study, machine learning-based CT radiomics models were developed and tested for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection. CT radiomics features hidden within lesions, ground-glass opacities and consolidation, were extracted, and machine-learning models demonstrated robust performance using multicenter cohorts for training and inter-validation, and an independent dataset and a prospective dataset for test.

Though there were slightly differences in CT scan parameters among centers, key features included in models were second-order, and focused on distribution, correlation and variance in gray level intensities, which described the relationship between voxels and hold quantitative information on the spatial heterogeneity of pneumonia lesions.[11,12] Compared with first-order features, second-order features were not sensitive to absolute value and thus more robust. Moreover, the models showed satisfied AUCs more than 90% on both training and test process, which indicated that the models could be applied in a general situation. Similarity in AUCs, sensitivity and specificity for RF and LR models also demonstrated the robustness, according to prior study that classification method showed most dominant in variability of model.[11]

The study was limited by small sample size. The percentage of short-term hospital stay is low in our multicenter cohorts, and semi-automated lesion segmentation might result in selection bias. A larger prospective multicenter cohort is needed to tune and test the machine learning-based CT radiomics models.

In summary, the machine learning-based CT radiomics models showed feasibility and accuracy for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection.

Summary boxes

What is already known on this topic

The coronavirus disease 2019 (COVID-19) from Wuhan, China has become a global challenge since the December 2019. The median hospital stay of discharged patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection was reported as 10 days.

The condition of patients in Wuhan worsens on the 10th day after illness onset. However, patients with symptoms longer than 10 days outside Wuhan were less severe than those in Wuhan.

The hospital stay is one of the prognostic indicators, and its non-invasive predicting model based on CT radiomics features is important for assessing the patients' clinical outcome.

What this study adds

In the multicenter study with patients from 5 designated hospitals in China, the CT radiomics models using logistic regression and random forest method showed satisfied diagnostic performance with sensitivity of 1.0 and 0.75, specificity of 0.89 and 1.0, respectively, in independent test dataset.

The machine learning-based CT radiomics models showed feasibility and accuracy for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection.

References

1. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. [published February 7, 2020]. *JAMA*. doi:10.1001/jama.2020.1585.

2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. [published January 24, 2020]. *Lancet*. doi: 10.1016/S0140-6736(20)30183-5.

3. Xu XW, Wu XX, Jiang XG, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ*. 2020;368:m606.

4. Zu ZY, Jiang MD, Xu PP, et al. Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology*. 2020:200490.

5. Pan F, Ye T, Sun P, et al. Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia. *Radiology*. 2020:200370.

6. Lei J, Li J, Li X, et al. CT Imaging of the 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology*. 2020:200236.

7. Qi X, Lei J, Yu Q, Xi Y, Wang Y, Ju S. CT imaging of coronavirus disease 2019 (COVID-19): from the qualitative to quantitative. [published January 20, 2020]. *Ann Transl Med.* doi:10.21037/atm.2020.02.91

8. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577.

9. Hofmanninger J, Prayer F, Pan J, Rohrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem. *ArXiv200111767 Phys Stat*. January 2020. http://arxiv.org/abs/2001.11767.

10. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77(21):e104-e107.

11. Parmar C, Grossmann P, Bussink J, et al. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep*. 2015;5(1):13087.

12. Elshafeey N, Kotrotsou A, Hassan A, et al. Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma. *Nat Commun*. 2019;10(1):3170.

Table 1 Baseline characteristics

Characteristics	All	Short-term stay (≤10days)	Long-term stay (>10days)	P value
Age, median (interquartile range [IQR]), years	38.00	32.00 (26.00-46.50)	41.00 (26.00-47.00)	0.72
Gender, No. (%), male	17 (55%)	2 (25%)	15 (65%)	0.10
Comorbidities, No. (%)	3 (10%)	1 (13%)	2 (9%)	1
Hypertension, No. (%)	2 (6%)	1 (13%)	1 (4%)	
History of cancer, No. (%)	2 (6%)	0 (0%)	2 (9%)	
Symptoms				
Fever, No. (%)	22 (71%)	5 (63%)	17 (74%)	
Myalgia, No. (%)	6 (19%)	2 (25%)	4 (17%)	
Diarrhea, No. (%)	2 (6%)	1 (13%)	1 (4%)	
Headache, No. (%)	11 (35%)	3 (38%)	8 (35%)	
Cough, No. (%)	25 (81%)	7 (88%)	18 (78%)	
Laboratory findings				
Blood leukocyte count, median (IQR), $\times 10^9$/L	4.42	5.09 (3.70-6.89)	4.30 (3.56-4.66)	0.14
Neutrophil count, median (IQR), $\times 10^9$/L	2.83	3.25 (2.45-4.19)	2.80 (1.90-3.30)	0.10
Neutrophil percentage, median (IQR), %	62.30	63.15 (61.63-71.80)	61.30 (52.90-65.50)	0.32

Lymphocyte count, median (IQR), $\times 10^9$/L	1.30	1.38 (1.16-1.93)	1.22 (0.91-1.40)	0.17
Lymphocyte percentage, median (IQR), %	29.00	26.70 (21.63-29.45)	31.90 (23.2-36.10)	0.30

Figure legends

Figure 1 Flow chart of study population and model building.

PCC, Pearson correlation coefficient; LR, logistic regression; RF, random forest.

Figure 2 The predictive performance of CT radiomics models in training, inter-validation and test datasets.

ROC, receiver operating characteristic; AUC, areas under the ROC curves; LR, logistic regression; RF, random forest; PPV, positive predictive value; NPV, negative predictive value.

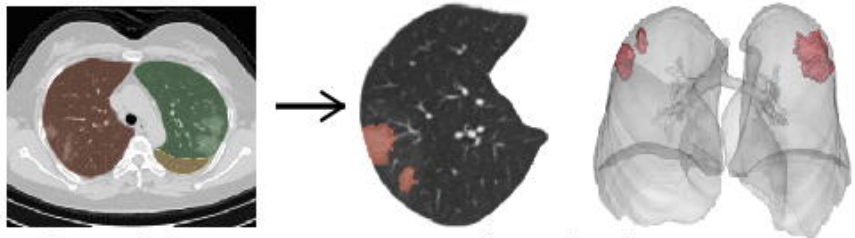
Patients

SARS-Cov-2 infection & initial CT
January 23 2020 - February 8 2020
Patients (N=52) Lesions (N=111)

Excluded:
14 hospitalized patients
7 patients non-findings in initial CT

Enrolled: Patients (N=31) Lesions (N=72)

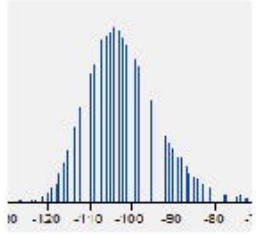
I. Segmentation



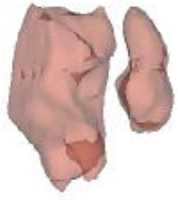
Lung lobes

Lung lesions

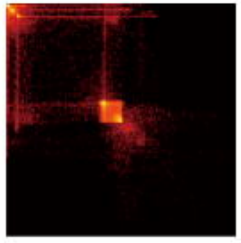
II. Feature extraction



First-order

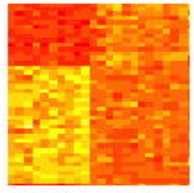


Shape

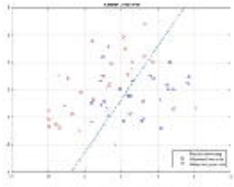


Second-order

III. Model Building

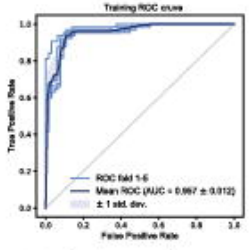


Feature selection
PCC & Relief

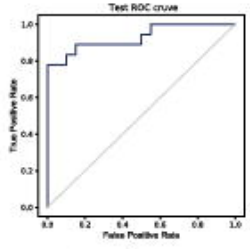


Classifier
LR & RF

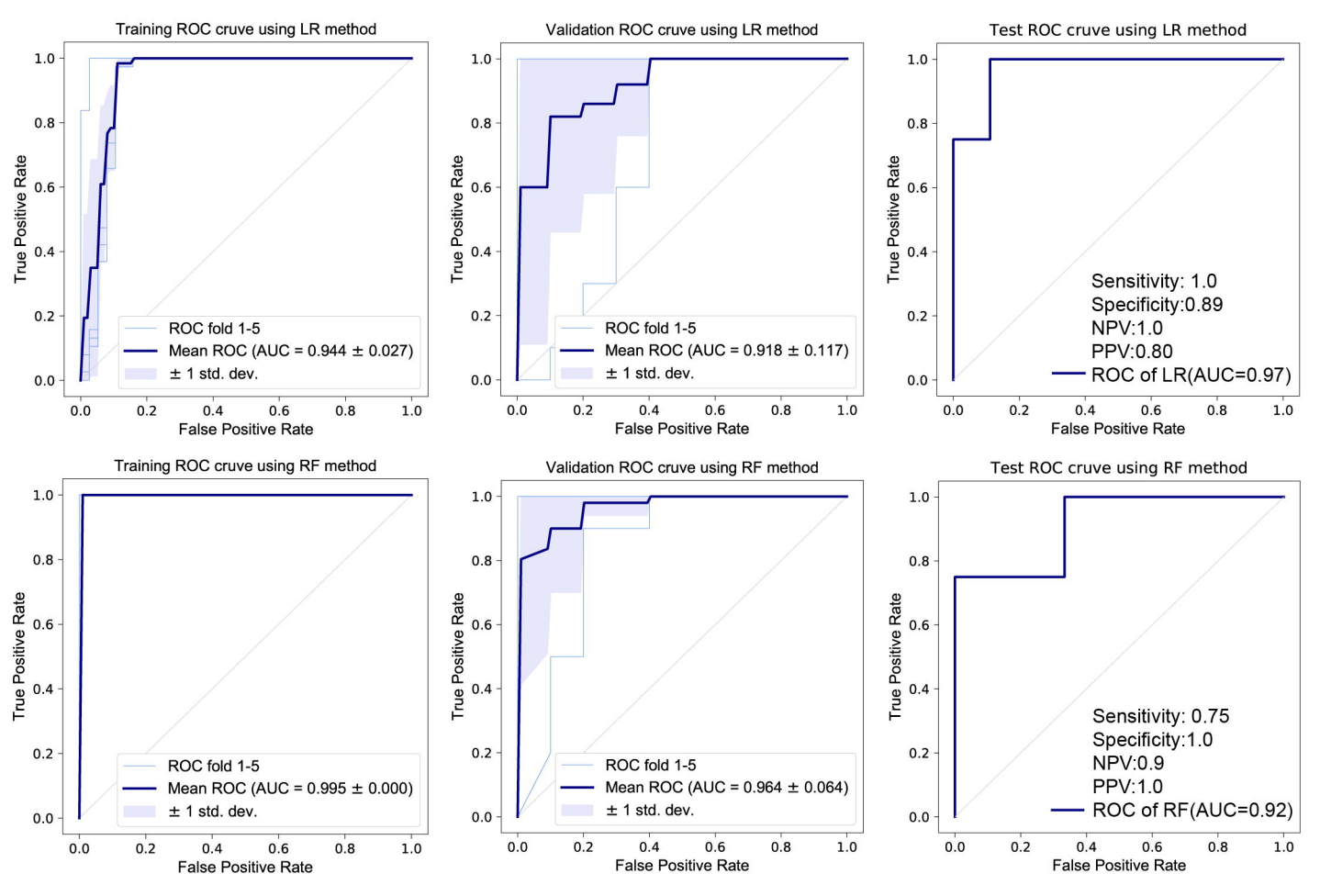
IV. Data Analysis



Training dataset



Test dataset


Training ROC cruve using LR method
True Positive Rate
False Positive Rate
ROC fold 1-5
Mean ROC (AUC = 0.944 ± 0.027)
± 1 std. dev.
Validation ROC cruve using LR method
True Positive Rate
False Positive Rate
ROC fold 1-5
Mean ROC (AUC = 0.918 ± 0.117)
± 1 std. dev.
Test ROC cruve using LR method
True Positive Rate
False Positive Rate
Sensitivity: 1.0
Specificity:0.89
NPV:1.0
PPV:0.80
ROC of LR(AUC=0.97)
Training ROC cruve using RF method
True Positive Rate
False Positive Rate
ROC fold 1-5
Mean ROC (AUC = 0.995 ± 0.000)
± 1 std. dev.
Validation ROC cruve using RF method
True Positive Rate
False Positive Rate
ROC fold 1-5
Mean ROC (AUC = 0.964 ± 0.064)
± 1 std. dev.
Test ROC cruve using RF method
True Positive Rate
False Positive Rate
Sensitivity: 0.75
Specificity:1.0
NPV:0.9
PPV:1.0
ROC of RF(AUC=0.92)