RESEARCH ARTICLE

A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images

Zongsheng Hu[1] | Zhenyu Yang[2] | Kyle J. Lafata[2,3,4] | Fang-Fang Yin[1,2] | Chunhao Wang[2]

[1]Medical Physics Graduate Program, Duke Kunshan University, Kunshan, Jiangsu, China

[2]Department of Radiation Oncology, Duke University, Durham, North Carolina, USA

[3]Department of Radiology, Duke University, Durham, North Carolina, USA

[4]Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, USA

Correspondence
Chunhao Wang, Department of Radiation Oncology, Duke University Medical Center, Durham, NC 27710, USA.
Email: chunhao.wang@duke.edu

Abstract

Purpose: To develop a deep learning model design that integrates radiomics analysis for enhanced performance of COVID-19 and non-COVID-19 pneumonia detection using chest x-ray images.

Methods: As a novel radiomics approach, a 2D sliding kernel was implemented to map the impulse response of radiomic features throughout the entire chest x-ray image; thus, each feature is rendered as a 2D map in the same dimension as the x-ray image. Based on each of the three investigated deep neural network architectures, including VGG-16, VGG-19, and DenseNet-121, a pilot model was trained using x-ray images only. Subsequently, two radiomic feature maps (RFMs) were selected based on cross-correlation analysis in reference to the pilot model saliency map results. The radiomics-boosted model was then trained based on the same deep neural network architecture using x-ray images plus the selected RFMs as input.

The proposed radiomics-boosted design was developed using 812 chest x-ray images with 262/288/262 COVID-19/non-COVID-19 pneumonia/healthy cases, and 649/163 cases were assigned as training-validation/independent test sets. For each model, 50 runs were trained with random assignments of training/validation cases following the 7:1 ratio in the training-validation set. Sensitivity, specificity, accuracy, and ROC curves together with area-under-the-curve (AUC) from all three deep neural network architectures were evaluated.

Results: After radiomics-boosted implementation, all three investigated deep neural network architectures demonstrated improved sensitivity, specificity, accuracy, and ROC AUC results in COVID-19 and healthy individual classifications. VGG-16 showed the largest improvement in COVID-19 classification ROC (AUC from 0.963 to 0.993), and DenseNet-121 showed the largest improvement in healthy individual classification ROC (AUC from 0.962 to 0.989). The reduced variations suggested improved robustness of the model to data partition. For the challenging non-COVID-19 pneumonia classification task, radiomics-boosted implementation of VGG-16 (AUC from 0.918 to 0.969) and VGG-19 (AUC from 0.964 to 0.970) improved ROC results, while DenseNet-121 showed a slight yet insignificant ROC performance reduction (AUC from 0.963 to 0.949). The achieved highest accuracy of COVID-19/non-COVID-19 pneumonia/healthy individual classifications were 0.973 (VGG-19)/0.936 (VGG-19)/ 0.933 (VGG-16), respectively.

Conclusions: The inclusion of radiomic analysis in deep learning model design improved the performance and robustness of COVID-19/non-COVID-19

pneumonia/healthy individual classification, which holds great potential for clinical applications in the COVID-19 pandemic.

1 | INTRODUCTION

Since its first discovery in 2019, the coronavirus disease (COVID-19) has affected more than 100 million people globally, and more than 5 million deaths related to COVID-19 were reported by the end of 2021.[1] Accurate and efficient diagnosis of COVID-19 is crucial to interrupt disease transmission and to start treatments of affected individuals. Currently, reverse transcription-polymerase chain reaction (RT-PCR) has been recognized as the gold standard for COVID-19 diagnosis for its high specificity.[2] While RT-PCR test may have limited sensitivity and long time of processing (a few hours to 2 days),[3] radiographic procedures, including chest x-ray and CT exams, have been adopted clinically as alternative diagnosis tools.[4] While COVID-19 related abnormalities could be more easily found in volumetric CT images,[5] planar chest x-ray has its unique advantages in COVID-19 diagnosis. Specifically, the short imaging time on a more accessible X-ray unit enables rapid COVID-19 exams, which can be critical in areas with high-volume patients and/or limited-resource medical facilities. To date, pilot studies have revealed that certain x-ray image features, including peripheral consolidations and ground-glass opacities, have been widely observed in COVID-19 infected patients.[4–9] However, the prevalent application of chest x-ray imaging in COVID-19 diagnosis is challenged by relatively limited sensitivity and specificity.[5] Additionally, radiographic exams including chest x-rays may not be optimal for radiologists' reading in the differentiation of non-COVID-19 pneumonia from COVID-19, which is important for early patient stratification that can lower COVID-19 mortality rate with more targeted treatments.[10,11]

Computed aided diagnosis systems (CAD) may have the potential to solve the aforementioned problem with high throughput quantitative analysis. In the last several months under the COVID-19 pandemic, studies revealed that CAD systems outperformed radiologists in radiographic-based COVID-19 diagnosis; with CAD information as reference information, radiologist reading results could be significantly improved.[11–13] One approach for such a CAD system is radiomics-based image analysis, which first extracts radiomics features as computation image biomarkers and then uses the extracted features in hand-made or machine learning classifier tasks. Although handcrafted radiomics features are commonly used in medical image analysis with possible qualitative image interpretability, the reported accuracy (75%–80%) of COVID-19 diagnosis is still limited in the representative radiomics-based CAD works.[14–16] Driven by recent theoretical developments and access to massive computation power, deep learning has demonstrated its great potential in CAD developments. It has been reported that deep learning solutions based on artificial neural network deployment could achieve high (>90%) specificities in COVID-19 diagnosis against healthy individuals[17,18]; moreover, decent specificities (>85%) of differentiating COVID-19 from non-COVID-19 pneumonia have been achieved.[11,19–21] Nevertheless, like all other deep learning applications in medical image analysis, the hyperparameters in the neural network are generated without explicit human knowledge intervention. Thus, the "black box" nature of deep learning-based CAD inhibits their interpretability, and potential clinical applications of these CADs could be impaired by limited interpretability by clinicians.

In this work, we aim to develop a radiomics-boosted deep learning CAD design for chest x-ray based COVID-19 diagnosis. Because hand-crafted radiomics and deep learning are complementary approaches to image representation, their integration may facilitate better model performance and interpretation.[22] Innovative implementation of radiomics analysis was included to analyze deep features from 3 custom-trained neural networks for COVID-19, non-COVID-19 pneumonia, and healthy individuals classification, and such radiomics analysis results were then incorporated as additional image-based input sources for 3 improved neural network designs. The proposed methodology may enhance deep learning interpretability for COVID-19 diagnosis from a current radiomics knowledge perspective.

2 | MATERIALS AND METHODS

2.1 | Image dataset

In this IRB-waived retrospective study, a total of 812 chest x-ray images were collected from three public databases,[23–25] including 262/288/262 images of COVID-19/non-COVID-19 pneumonia/healthy individuals, respectively. The image numbers from the three categories were approximately the same for eliminating categorical bias during deep learning training. All collected images were verified by experienced medical physicists with proper lung x-ray display settings and with absences of the overlaid image reading
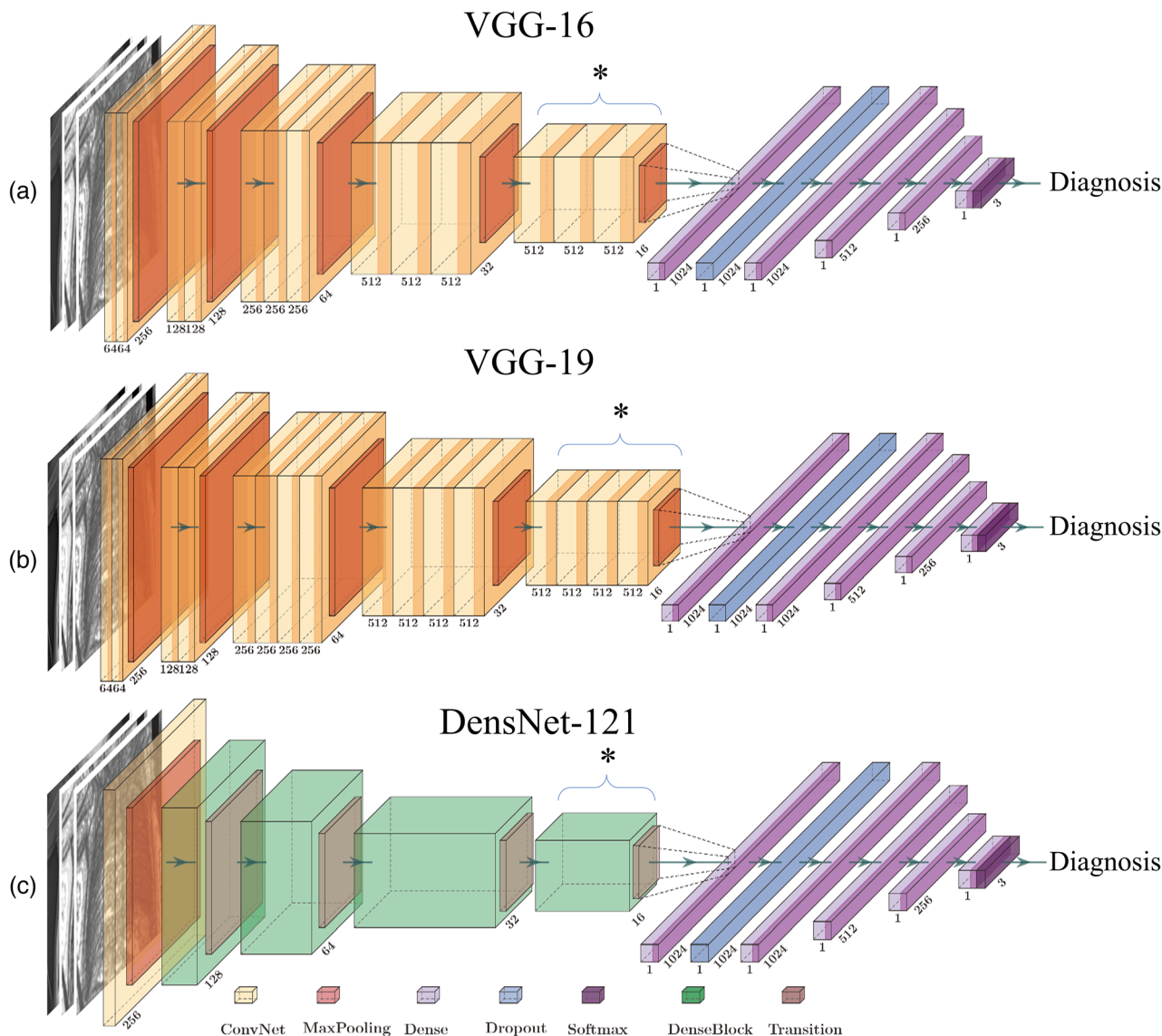
FIGURE 1 Diagrams of the three studied deep neural networks. (a) VGG-16, (b) VGG-19, and (c) DenseNet-121

annotations. To unify image data size, all images were resized to a 256 × 256 matrix grid size using b-spine interpolation and were normalized to 256 gray levels. 649 and 163 images (8:2) were assigned for the model training set and the independent test set, respectively.

2.2 | Neural network architecture

To investigate radiomics-boosted deep learning as a general methodology, we studied three deep neural networks, VGG-16 (Figure 1a), VGG-19 (Figure 1b), and DenseNest-121 (Figure 1c), for COVID-19/non-COVID-19 pneumonia/healthy individual classification. Based on pre-trained deep learning schemes,[26,27] the investigated three deep neural networks share the same two-part design: the 1st part is the convolutional

base: In VGG-16 and VGG-19, the convolutional bases consist of five convolutional blocks. Each convolutional block is stacked by two to four convolutional layers and a max-pooling layer. In each convolutional layer, the filter size is 3 × 3 with padding and stride of 1. Max-pooling is performed over a 2 × 2-pixel window with a stride of 2. In DenseNet-121, the convolutional base consists of one convolutional block and four dense blocks. The convolutional block consists of 7 × 7 convolutional layers with a stride of 2 and a 3 × 3 Max-pooling with a stride of 2. The dense blocks consist of 6, 12, 24, and 16 convolutional units, respectively. Each unit is stacked by a 1 × 1 and a 3 × 3 convolutional layer. The first three dense blocks are followed by three transition layers, which consist of a 1 × 1 convolutional layer with stride 1 and a 2 × 2 average pool with stride 2; the 2nd part is the Dense part, which is the stack of dense layers.
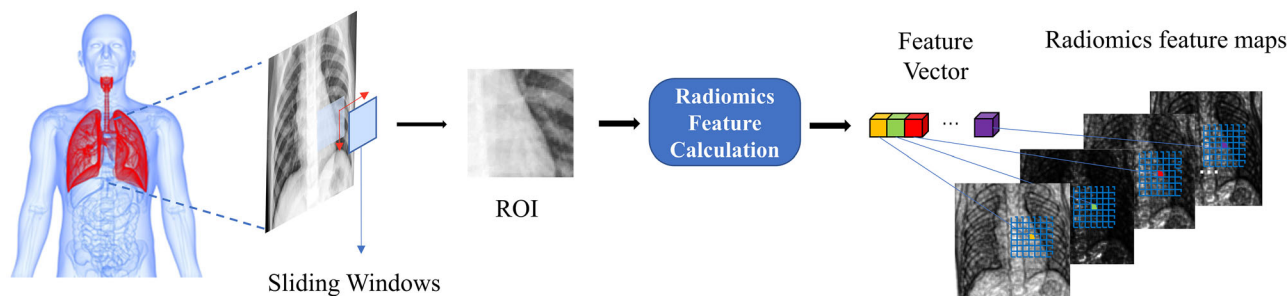
FIGURE 2   A workflow summary of radiomic feature map (RFM) calculation in this work

Depending on specific classification tasks, the number and size of dense layers can be customized. In all three deep neural network architectures, the self-defined Dense classifier connects the convolutional base and consists of five Dense layers with the size of 1024, 1024, 512, 256, and 3, respectively. The input of the neural network is a three-channel image with a $256 \times 256 \times 3$ shape size, while the output is one of the three categorical binary label vectors, i.e., [1,0,0], [0,1,0], and [0,0,1], which correspond to COVID-19, non-COVID-19 pneumonia, and healthy results, respectively. To deal with relatively small data size in this work, the convolutional bases loaded the weights that were pre-trained on ImageNet as a transfer learning scheme.[28] In addition, in order to make the models more relevant for the problem at hand, fine-tune technique was used with the last few convolutional layers (marked * in Figure 1: the last convolutional blocks in VGG-16 and VGG-19, as well as the last dense block in DenseNet121) being set as free parameters for task-specific training. To avoid the occurrence of overfitting, a dropout layer was added between the first two dense layers with a dropout possibility of 0.5, and soft-max activation was used in the output layer.

2.3 | Radiomic feature map extraction

Classic radiomics analysis calculates radiomic features as scalar values from a pre-defined region-of-interest (ROI) in image space. While this approach has been widely adopted to capture the overall textures in ROI, it cannot capture the anatomy-driven subtle texture variations within the ROI. As such, we implement a RFM calculation workflow,[29] which is summarized in Figure 2. For RFM generation, a 2D kernel (13×13 matrix size) was adopted to form a ROI, and 37 radiomic features were extracted as a 1×37 vector within this ROI following classic GLCOM[30] (21 features) and GLRLM[31] (16 features) feature extraction methods using 32 grey levels. For each feature, the calculated feature value was assigned as the pixel value centered at the ROI. By moving this 2D kernel across the x-ray image as a sliding window operation, 37 feature maps were formed in the

same dimension as original x-ray images. All radiomic analysis was done using custom code that was bench-marked with digital phantoms[32] and complies with the imaging biomarker standardization initiative.[33]

2.4 | Model training and evaluation

For each of the three investigated deep neural network architecture, two model versions were trained: in the 1st pilot model, the x-ray image is the sole model input. To accommodate the input shape of pre-trained neural networks, the grayscale x-ray images were broadcast to three channels as a network input variable. This model serves as the benchmarking deep learning model in this work; in the 2nd version as a radiomics-boosted model, the grayscale x-ray image and two derived RFMs were stacked as the 3-channel neural network input variable. These two RFMs were selected based on the analysis of the 1st model's saliency map (SM), which indicates how important each pixel is with respect to the final classification results of the neural network in the benchmarking model. It is calculated as the absolute gradient of class activation which is defined as the dot product of prediction output and target divided by the input image[34]: a pixel with a higher intensity value in SM indicates higher importance of that pixel in neural network's attention for diagnosis. The two RFMs with the highest average cross-correlation (CC) values against the SM results in training data were selected. This action amplifies certain pixels (and regions) with potentially high importance of disease diagnosis in the image space, which could improve the overall diagnostic accuracy of the proposed radiomics-boosted model in comparison with the 1st pilot model.

To investigate model robustness, 50 runs of each model version were trained using the training data set (649 images). In each run, the training and validation samples were randomly selected following a 7:1 ratio. During deep learning training within the TensorFlow environment using a Nvidia™ (Santa Clara, CA) Tesla V100 graphic card, the loss function was categorical cross-entropy, and Adam optimizer was selected. For model evaluation, sensitivity, specificity, accuracy, and
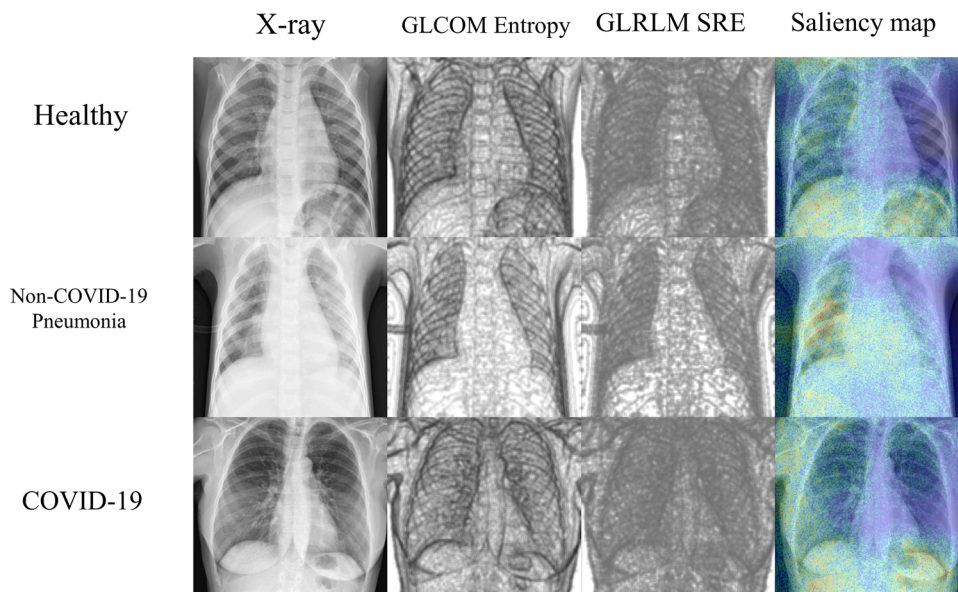
FIGURE 3 Image comparisons from three example cases for the VGG-16 pilot model. The GLRLM SRE RFMs and saliency map (overlaid with X-ray image) are illustrated in 0.3 power scale

ROC area under the curve (AUC) results from both model versions were analyzed. Statistical significance of comparison was determined by Wilcoxon signed rank tests at level 0.05.

3 | RESULTS

Figure 3 shows an example of image comparison in the pilot model of VGG-16, that is, the one using x-ray only as input. As illustrated, the identified RFMs, GLCOM entropy (CC = 0.33) RFM, and GLRLM short run emphasis (SRE) (CC = 0.31) render more tissue textural variations in both lung and other soft tissue regions than original x-ray images. Similarly, GLCOM entropy (CC = 0.32) and GLCOM sum entropy (CC = 0.31) RFMs were selected from VGG-19 pilot model, while the GLCOM sum average (CC = 0.27) and GLRLM short run high gray level emphasis (CC = 0.28) RFMs were selected from DenseNet-121 pilot model.

Table 1 summarizes the quantitative comparisons of sensitivity, specificity, accuracy, and ROC AUC between pilot models and radiomics-boosted models. For the VGG-16 architecture, the radiomics-boosted deep learning model achieved statistical improvements in all parameters with $p < 0.05$. The largest improvements were observed in non-COVID-19 pneumonia diagnosis. Additionally, the reduced standard deviations of the reported statistics indicated the enhanced robustness of the radiomics-boosted deep learning design. These quantitative results highlight the superiority of the proposed radiomics-boosted deep learning model in the context of VGG-16 architecture. The standard deviation

of 50 runs is less than 3%, reaching a high level of robustness. For VGG-19 architecture, the mean values of all parameters were higher in the radiomics-boosted model than the pilot model; however, the observed numerical improvements were small, and only a few improvements were found with statistical significance in COVID-19 and healthy class results. The standard deviations were also reduced as indicators of improved model robustness. Results from DenseNet-121 architecture were similar to VGG-16 and VGG-19 results except for non-COVID-19 pneumonia classification, in which mixed impacts in the radiomics-boosted model were presented; nevertheless, the radiomics-boosted model did improve COVID-19 and healthy individual classification performance and increased model robustness in the healthy individual classification. As a summary, the radiomics-boosted design achieved best performance in COVID-19 diagnosis in both VGG-16 and VGG-19 architecture applications, while it achieved the best performance in healthy individual classification in DenseNet-121 architecture application.

Figure 4 summarizes the ROC analysis results of the three studied architectures. The blue and red solid lines represent the average ROC results of 50 runs of two deep model versions (x-ray only vs. x-ray + RFM), and the colored bands represent the model performance variation as ± 1 standard deviation. For VGG-16 architecture (Figure 4a), radiomics-boosted design improved ROC results of all three classification tasks, and the largest performance improvement was observed in non-COVID-19 pneumonia diagnosis. Additionally, the radiomics-boosted deep learning model has narrower ROC bandwidth, which suggests

TABLE 1 Sensitivity, specificity, accuracy, and ROC AUC results summaries of top: VGG-16 architecture, middle: VGG-19 architecture, bottom: DenseNet-121 architecture. The mean values and standard deviation of 50 trained runs are reported

VGG-16	Healthy		Non-COVID-19 pneumonia		COVID-19	
	X-ray	X-ray + RFM	X-ray	X-ray + RFM	X-ray	X-ray + RFM
Sensitivity	0.854 ± 0.065	$0.922 \pm 0.059^*$	0.780 ± 0.092	$0.857 \pm 0.0361^*$	0.903 ± 0.071	$0.949 \pm 0.036^*$
Specificity	0.918 ± 0.044	$0.938 \pm 0.022^*$	0.941 ± 0.041	$0.963 \pm 0.023^*$	0.940 ± 0.037	$0.973 \pm 0.020^*$
Accuracy	0.895 ± 0.029	$0.933 \pm 0.023^*$	0.892 ± 0.029	$0.931 \pm 0.016^*$	0.927 ± 0.028	$0.965 \pm 0.016^*$
AUC	0.948 ± 0.027	$0.979 \pm 0.012^*$	0.918 ± 0.043	$0.969 \pm 0.017^*$	0.963 ± 0.023	$0.993 \pm 0.006^*$
VGG-19	**Healthy**		**Non-COVID-19 pneumonia**		**COVID-19**	
	X-ray	X-ray + RFM	X-ray	X-ray + RFM	X-ray	X-ray + RFM
	0.902 ± 0.097	0.921 ± 0.053	0.824 ± 0.059	0.854 ± 0.041	0.933 ± 0.047	$0.969 \pm 0.039^*$
Specificity	0.920 ± 0.030	$0.940 \pm 0.013^*$	0.969 ± 0.017	0.972 ± 0.020	0.969 ± 0.028	0.975 ± 0.023
Accuracy	0.914 ± 0.032	0.933 ± 0.019	0.925 ± 0.025	0.936 ± 0.013	0.956 ± 0.023	$0.973 \pm 0.017^*$
AUC	0.970 ± 0.015	0.978 ± 0.007	0.964 ± 0.028	0.970 ± 0.013	0.987 ± 0.013	$0.994 \pm 0.006^*$
DenseNet-121	**Healthy**		**Non-COVID-19 pneumonia**		**COVID-19**	
	X-ray	X-ray + RFM	X-ray	X-ray + RFM	X-ray	X-ray + RFM
Sensitivity	0.810 ± 0.210	$0.948 \pm 0.048^*$	0.902 ± 0.070	0.872 ± 0.042	0.747 ± 0.150	0.800 ± 0.061
Specificity	0.920 ± 0.049	$0.924 \pm 0.045^*$	0.850 ± 0.109	$0.922 \pm 0.036^*$	0.971 ± 0.037	$0.994 \pm 0.014^*$
Accuracy	0.881 ± 0.056	$0.920 \pm 0.020^*$	0.866 ± 0.060	$0.907 \pm 0.017^*$	0.896 ± 0.038	$0.928 \pm 0.024^*$
AUC	0.962 ± 0.022	$0.989 \pm 0.008^*$	0.963 ± 0.017	0.949 ± 0.022	0.964 ± 0.022	0.972 ± 0.027

the enhanced robustness of its design under different data sample uses. The same improvements were also observed in VGG-19 results (Figure 4b), but the improvements' magnitudes were smaller than ones in VGG-16 results, which was mainly contributed by the higher performance of VGG-19 prior to radiomics-boosted implementation. In Figure 4c of DenseNet-121 results, while the ROC improvement by radiomics-boosted design was prominent in healthy individual classification, the improvement in COVID-19 diagnosis was limited. As reported in Table 1, the ROC result in non-COVID-19-pneumonia showed a slightly decreased performance after radiomics-boosted design, though such decrease has found no statistical significance.

As an example of saliency map visualization, the SM results of the VGG-16 architecture are illustrated in the last column of Figure 3. The pixel values in SM can be interpreted as the attention of the deep learning model. As seen, the attention patterns, that is, colored hot regions distribution in SMs, demonstrated prominent spatial heterogeneity across the image field-of-view. In addition, the current SM illustration suggests potential class-specific spatial patterns of deep network attention: more attention might be drawn to lateral lung regions for COVID-19 detection, while mediastinum regions might be attention focus for non-COVID-19 pneumonia detection. In order to quantitatively analyze the attention patterns across different patient cohorts, we calculated the CC matrix for SMs of all three radiomics-boosted architectures in the test set in Figure 5, which includes all the CCs between paired SM results from the test set. For VGG-16 and VGG-19 results, CCs within each cohort

were relatively higher than those calculated across different cohorts. This result suggests that the developed deep learning model captured cohort-specific features for the classification task. Additionally, the mean CC result of COVID-19 versus non-COVID-19 pneumonia (VGG-16:0.12; VGG-19: 0.10) cohort was slightly higher than the result of COVID-19 versus healthy cohorts (VGG-16:0.07; VGG-19: 0.08) and non-COVID-19 pneumonia versus healthy cohorts (VGG-16:0.09; VGG-19: 0.09). This observed COVID-19/non-COVID-19 pneumonia similarity supports the clinical reports of challenges in COVID-19/non-COVID-19 pneumonia differentiation.[10,11] For the DenseNet-121, however, all reported CCs were very small and were an order smaller than those in VGG-16 and VGG-19 results. These results suggest that SM from DenseNet-121 architecture did not capture meaningful class-specific spatial patterns.

4 | DISCUSSION

To our best knowledge, this work is the first of its kind for combining radiomic analysis and deep neural network implementation. The results of this work demonstrated that the inclusion of RFMs, as a new form of handcrafted imaging biomarker rendering, can improve deep learning-based COVID-19 detection. With the aid of RFMs, we achieve higher model performances for COVID-19/non-COVID-19 pneumonia/healthy classification with a smaller (812 patients in total) than reported work. For example, Zhang et al.[11]
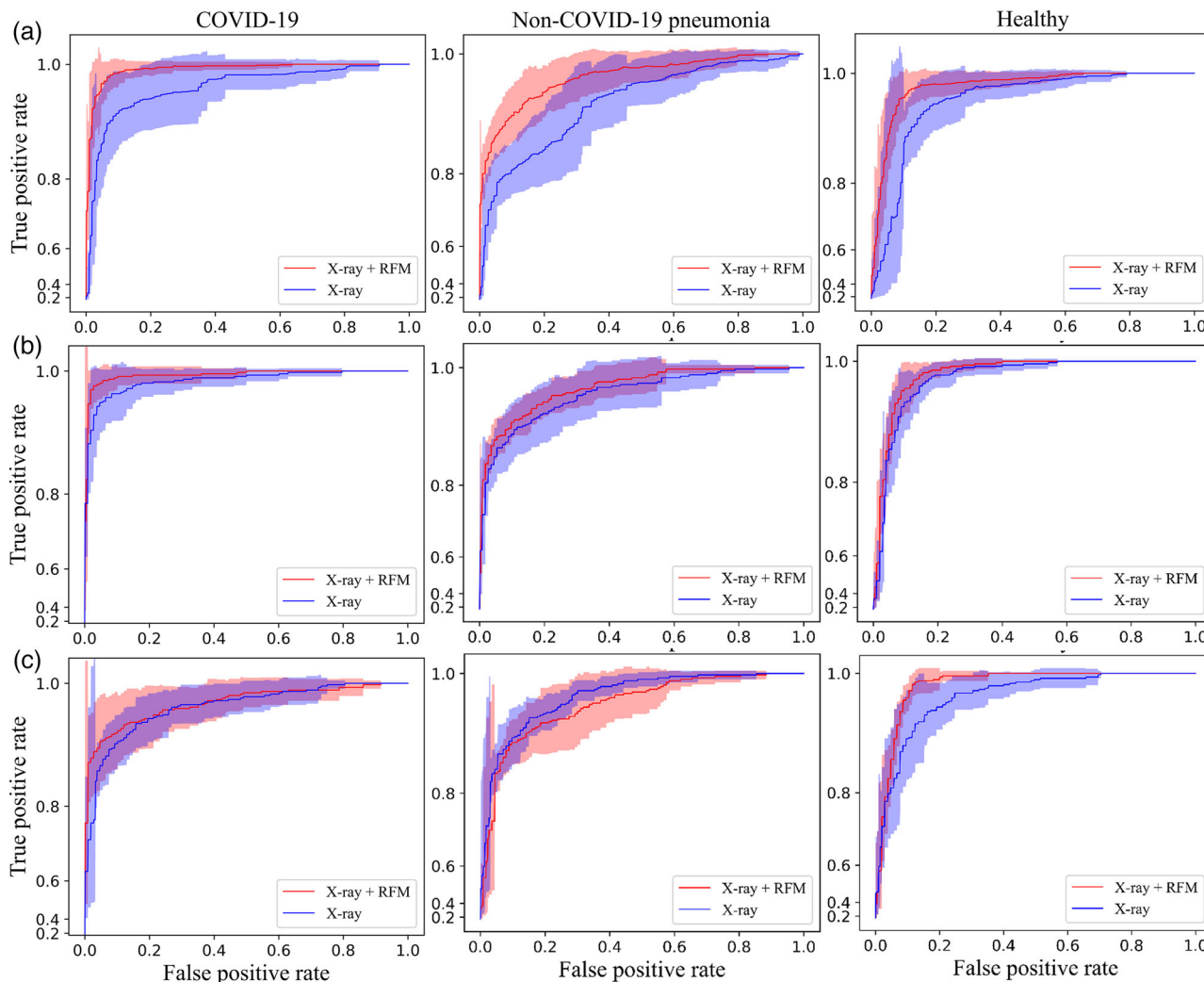
FIGURE 4 The ROC results of pilot model versus radiomics-boosted model using (a) VGG-16, (b) VGG-19, and (c) DenseNet-121 deep neural network architecture. 0.3 power scale was used in the y axis to highlight the difference
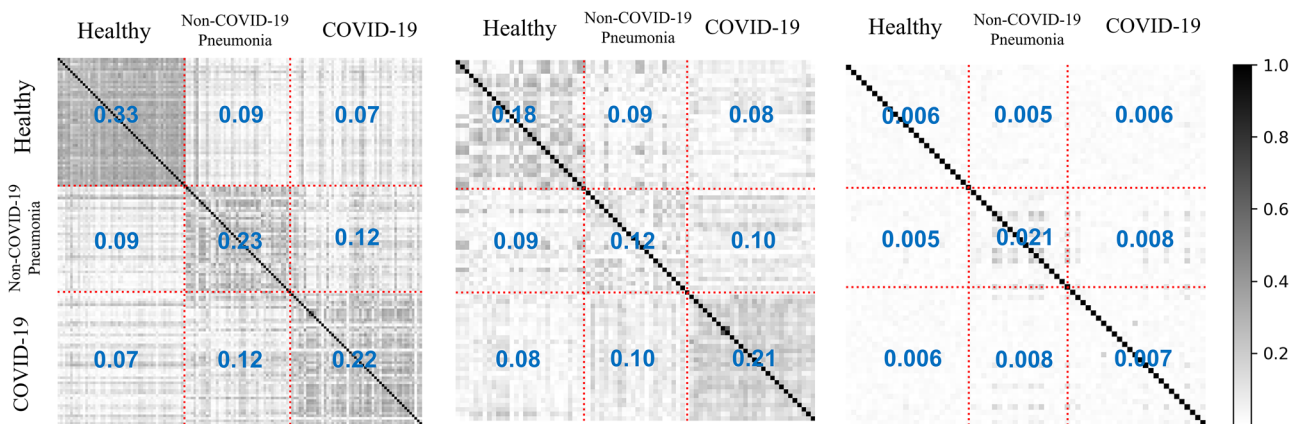

FIGURE 5 The SM cross-correlation matrix of radiomics-boosted model on test set for left: VGG-16; middle: VGG-19; right: DenseNet121 architectures. The x and y axes represent the sample ID in the test set, sorting with the order of healthy/non-COVID-19 pneumonia/COVID-19 cohorts

achieved a sensitivity of 88% and a specificity of 79% in COVID-19/non-COVID-19 pneumonia diagnosis with a dataset of 2060 patients. Nishio et al.[19] achieved an accuracy of 83.7% for three categories of classification (healthy/non-COVID-19 pneumonia/COVID-19) using the VGG-16 model trained on 1248 images. Tulin et al.[21] achieved an accuracy of 87.0% for tri-class classification (healthy/non-COVID-19 pneumonia/COVID-19) using the Darknet-19. In this work, we studied a total of 812 chest x-ray images from three public datasets,[23–25] which were not curated by the typical medical image study protocols. As a result, proper image processing is necessary for streamlined deep learning implementation. In particular, we resized all images to 256×256 grid size and normalized all images to 256 gray levels as uint8 format. These operations are standard in digital image processing which will facilitate the data reproducibility of this work. In addition, the robustness of the developed model was systematically analyzed. For each model design, we trained 50 runs of models using randomly selected training and validation samples following a ratio of 7:1. The small standard deviation (< 0.06) of selected metrics and ROC results revealed the enhanced robustness of the developed model, which further demonstrates the potential of the radiomics-boosted deep learning design in clinical situations using different x-ray image data sources.

The deep learning implementations in this work adopted three commonly used deep neural network architectures based on a transfer learning scheme. VGG-16 was first selected for the following reason: (1) it has been widely studied for medical image analysis tasks as a transfer learning scheme; (2) in comparison with other prevalent candidates, VGG-16 possesses a smaller number of trainable parameters under the transfer learning scheme and thus leads to reduced calculation workload for network training; and (3) previous studies[19] reported that VGG-16 achieved the highest accuracy in COVID-19 diagnosis tasks in comparison with several other pre-trained deep neural network architectures. In addition to VGG-16, we studied VGG-19 and DenseNet-121 for the proposed radiomics-boosted design. As reported in Figures 3 and 4, results of VGG-19 after the radiomics-boosted design were similar to the ones in VGG-16 implementation, which can be attributed to the high similarity of network architecture shown in Figure 1.[26] On the other hand, the performance improvement after the radiomics-boosted design was higher in VGG-16 application than in VGG-19 application: this can be explained by the fact that classification performance from VGG-19 using x-ray image only was higher than ones from VGG-16. Although the total trainable parameter numbers under the transfer learning scheme were approximately the same in VGG-16 and VGG-19, VGG-19 has a larger dimension with more total parameters due to the three additional convolutional layers. Overall, the proposed radiomics-boosted design still improved VGG-19 performance in terms of the 4 classification evaluators as well as model robustness. In the DenseNet-121 study, the improvements after radiomics-boosted design were rather limited. While radiomics-boosted design improved healthy individual classification results with statistical significance, ROC results in non-COVID-19 pneumonia classification showed a slight performance decrease. A plausible explanation is the fact that DenseNet-121 has a larger dimension with more trainable parameters (>70 m) than VGG-16 and VGG-19 (~40 m) under the transfer learning scheme. In addition, DenseNet-121 has more parameters in total (>75 m) than VGG-19 (~55 m). As such, a larger x-ray image set might be necessary to exploit the full potential of the proposed radiomics-boosted design in the DenseNet-121 application. Nevertheless, the current results, especially the prominent performance improvement in healthy individual classification, are sufficient to support the benefit of radiomics-boosted design in the DenseNet-121 application.

The inclusion of RFMs is a key technical innovation. Instead of calculating radiomic features as scalar values from selected volumes in image space, RFMs capture the anatomy-driven subtle texture variations within ROIs. It has been demonstrated that radiomics are associated with pulmonary function[35] and lung ventilation measurements[29]; as such, the potential functional information in RFMs contributes to the enhanced COVID-19 diagnosis accuracy. The selection of two RFMs from 37 RFMs is more than a trivial task. While direct comparisons of all possible RFM combinations are feasible, it requires a high computational cost without the potential of transferring this technique to other clinical applications. Driven by the hypothesis that certain RFMs can be related to neural network hyperparameters, we selected RFMs based on similarity metrics between RFMs and neural network saliency maps which measure the attention pattern of the network implementation. Such a pattern can be used as an auxiliary tool for radiologists in image reading, that is, highlighting specific regions as visual clues for human reading. This potential human-aid tool can be important in accurate COVID-19/non-COVID-19 pneumonia differentiation, which can be a challenging task for radiologists using chest x-ray images without volumetric information.[11] RFMs with higher similarities to SMs could emphasize regional information to enhance neural network attention, which increases synchronously with SM results. It is worth mentioning that both VGG-16 and VGG-19 results identified GLCOM Entropy based on SM results, which emphasizes the similarity of the two deep learning architectures. In addition to the current design, it would be of interest to investigate other RFM selection mechanisms that can be complementary to the SM results (i.e., based on "dissimilarity") for potential performance improvement. DenseNet-121 SM

results in Figure 5, however, did not show recognized class-specific SM spatial patterns in comparison with VGG-16 and VGG-19. The absence of such patterns may stem from the potentially limited sample size due to large variable numbers of DenseNet-121. Additionally, ImageNet-based transfer learning may affect SM calculation results as well. Future works of DenseNet-121 analysis, based on full training from scratch or transfer learning from an x-ray-specific dataset, are of our interest to continue SM-based deep learning interpretability studies.

The presented design of combining radiomics analysis and deep neural network implementation may create a new paradigm of the CAD system. The implemented RFM calculation workflow may also enhance neural network performance in other tasks, particularly those where multi-channel imaging data are required as input.[36] Additionally, the proposed method provides a radiomics perspective of deep learning interpretability. The hyperparameters in the neural network are trained without explicit human knowledge intervention and thus are hard to interpret by empirical knowledge. For deep learning-based CAD systems, the "black box" nature impaired the clinical deployments of such systems without clinicians' confidence. As a step towards deep learning interpretability, we investigated neural network attention information using a radiomics-based analysis. Radiomics has been widely studied as computational imaging biomarkers for disease detection and outcome monitoring,[37] and it has been demonstrated that radiomic feature spaces can be mathematically decomposed to provide interpration.[38] Following the saliency map analysis approach in this work, additional parameters can be used to enhance deep learning interpretability, such as histology sample images from biopsy and anatomy contours from radiation therapy. These directions will be studied in future works when appropriate datasets become available.

5 | CONCLUSION

In this study, we proposed a radiomics-boosted deep learning design for x-ray based COVID-19 diagnosis and non-COVID-19 pneumonia diagnosis. An innovative RFM calculation workflow was implemented to generate additional input sources for deep neural networks, and such design was tested using three deep neural network architectures. Results showed that the proposed radiomics-boosted deep learning design improved the performance and robustness of COVID-19/non-COVID-19 pneumonia/healthy individual classification in concurrence with a radiomics viewpoint of deep learning interpretation. It holds great potential for clinical applications for COVID-19 diagnosis and generalization in other diagnosis tasks.

CONFLICT OF INTEREST
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.

DATA AVAILABILITY STATEMENT
Image data in this work come from publically available datasets; the developed radiomics feature map calculation tool and the presented deep neural network implementations can be shared upon request to the corresponding author.

DISCLOSURE
The authors have no conflicts to disclose.

REFERENCES

1. Worldometer. *COVID-19 Coronavirus Pandemic*. https://www.worldometers.info/coronavirus/
2. Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA*. 2020;323(18):1843-1844.
3. Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 2020;296(2):E115-E117.
4. Ai T, Yang ZL, Hou HY, et al. Correlation of chest CT and RT-PCR testing for Coronavirus Disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020;296(2):E32-E40.
5. Yoon SH, Lee KH, Kim JY, et al. Chest radiographic and CT findings of the 2019 novel Coronavirus Disease (COVID-19): analysis of nine patients treated in Korea. *Korean J Radiol*. 2020;21(4):494-500.
6. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing. *Radiology*. 2020;296(2):E41-E45.
7. Wong HYF, Lam HYS, Fong AHT, et al. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology*. 2020;296(2):E72-E78.
8. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382(18):1708-1720.
9. Shi HS, Han XY, Jiang NC, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis*. 2020;20(4):425-434.
10. Bai HX, Hsieh B, Xiong Z, et al. Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. *Radiology*. 2020;296(2):E46-E54.
11. Zhang R, Tie X, Qi Z, et al. Diagnosis of coronavirus disease 2019 pneumonia by using chest radiography: value of artificial intelligence. *Radiology*. 2021;298(2):E88-E97.
12. Bai HX, Wang R, Xiong Z, et al. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology*. 2020;296(3):E156-E165.
13. Wang H, Wang L, Lee EH, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. *Eur J Nucl Med Mol Imaging*. 2021;48(5):1478-1486. https://doi.org/10.1007/s00259-020-05075-4
14. Fang X, Li X, Bian Y, Ji X, Lu J. Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2. *Eur Radiol*. 2020;30(12):6888-6901.

15. Santone A, Belfiore MP, Mercaldo F, Varriano G, Brunese L. On the adoption of radiomics and formal methods for COVID-19 coronavirus diagnosis. *Diagnostics (Basel)*. 2021;11(2): 293.

16. Wang L, Kelly B, Lee EH, et al. Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable radiomics features. *Eur J Radiol*. 2021;136:109552.

17. Minaee S, Kafieh R, Sonka M, Yazdani S, Jamalipour SG. Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. *Med Image Anal*. 2020;65:101794.

18. Khan AI, Shah JL, Bhat MM. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed*. 2020;196:105581.

19. Nishio M, Noguchi S, Matsuo H, Murakami T. Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Sci Rep*. 2020;10(1):17532.

20. Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. *Comput Biol Med*. 2020;126:104037.

21. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med*. 2020;121:103792.

22. Lafata KJ, Wang Y, Konkel B, Yin F-F, Bashir MR. Radiomics: a primer on high-throughput image phenotyping. *Abdominal Radiology*. 2021;1-17. https://doi.org/10.1007/s00261-021-03254-x

23. Google. *COVID-19 X-ray dataset*. https://drive.google.com/file/d/1AfC8emc3mGCfIYo1jK-R6suK7a2Kv2m2

24. Cohen J, Morrison P, Dao L. COVID-19 image data collection. *arXiv*. 2020. https://doi.org/10.48550/arXiv.2003.11597.

25. Paul M, *Chest X-Ray Images (Pneumonia)*. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

26. Simonyan K, Zisserman A Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556*. 2014.

27. Huang G, Liu Z, Maaten L, Weinberger KQ Densely Connected Convolutional Networks. *arXiv*. 2018.

28. Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2009:248-255. https://doi.org/10.1109/cvpr.2009.5206848.

29. Yang Z, Lafata K, Chen X, et al. Quantification of lung function on CT images based on pulmonary radiomic filtering. *arXiv*. 2021. arXiv:2105.11171.

30. Haralick RM, Shanmugam K, Dinstein IH. Textural feature for image classification. *IEEE Trans Syst Man Cybern*. 1973;SMC-3(6):610-621.

31. Galloway MM. Textural analysis using gray level run lengths. *Computer Graphics and Imaging Processing*. 1975;4(2):172-179.

32. Chang Y, Lafata K, Wang C, et al. Digital phantoms for characterizing inconsistencies among radiomics extraction toolboxes. *Biomedical Physics & Engineering Express*. 2020;6(2):025016. https://doi.org/10.1088/2057-1976/ab779c.

33. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338.

34. Karen S, Andrea V, Andrew Z. Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv*. 2014. arXiv:1312.6034v2.

35. Lafata KJ, Zhou Z, Liu JG, Hong J, Kelsey CR, Yin FF. An exploratory radiomics approach to quantifying pulmonary function in CT images. *Sci Rep*. 2019;9(1):11509.

36. Wang C, Liu C, Chang Y, et al. Dose-distribution-driven PET image-based outcome prediction (DDD-PIOP): a deep learning study for oropharyngeal cancer IMRT application. *Front Oncol*. 2020;10:1592.

37. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762.

38. Lafata KJ, Chang Y, Wang C, et al. Intrinsic radiomic expression patterns after 20 Gy demonstrate early metabolic response of oropharyngeal cancers. *Med Phys*. 2021;48(7):3767-3777.