A human-computer collaboration for COVID-19 differentiation: combining a radiomics model with deep learning and human auditing

Xiaoyan Xin[1#], Ran Mo[2#], Mingran Shao[1], Wen Yang[1], Daixin Li[3], Yanqiu Zhang[1], Han Wang[1], Baiyun Liu[4], Song Tian[4], Weidao Chen[4], Jiangfen Wu[4], Bin Zhu[1], Kefeng Zhou[1], Chao Du[3], Bing Zhang[1]

[1]Department of Radiology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China; [2]Department of Burns and Plastic Surgery, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, China; [3]Department of Radiology, the Second Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China; [4]Beijing Infervision Technology Co. Ltd., Beijing, China

Contributions: (I) Conception and design: B Zhang; (II) Administrative support: X Xin, R Mo, B Liu, J Wu; (III) Provision of study materials or patients: M Shao, W Yang; (IV) Collection and assembly of data: Y Zhang, H Wang; (V) Data analysis and interpretation: M Shao, Y Zhang, S Tian, W Chen, Y Chen, D Li, C Du; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.
[#]These authors contributed equally to this work.

Correspondence to: Chao Du. Department of Radiology, the Second Affiliated Hospital of Nanjing University of Chinese Medicine, No. 121 Jiangjia Yuan, Nanjing, China. Email: duchao2009du@163.com; Bing Zhang. Department of Radiology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, No. 321 Zhongshan Road, Nanjing, China. Email: zhangbing_nanjing@nju.edu.cn.

Background: This study aimed to build a radiomics model with deep learning (DL) and human auditing and examine its diagnostic value in differentiating between coronavirus disease 2019 (COVID-19) and community-acquired pneumonia (CAP).

Methods: Forty-three COVID-19 patients, whose diagnoses had been confirmed with reverse-transcriptase polymerase-chain-reaction (RT-PCR) tests, and 60 CAP patients, whose diagnoses had been confirmed with sputum cultures, were enrolled in this retrospective study. The candidate regions of interest (ROIs) on the computed tomography (CT) images of the 103 patients were determined using a DL-based segmentation model powered by transfer learning. These ROIs were manually audited and corrected by 3 radiologists (with an average of 12 years of experience; range 6–17 years) to check the segmentation acceptance for the radiomics analysis. ROI-derived radiomics features were subsequently extracted to build the classification model and processed using 4 different algorithms (L1 regularization, Lasso, Ridge, and Z test) and 4 classifiers, including the logistic regression (LR), multi-layer perceptron (MLP), support vector machine (SVM), and extreme Gradient Boosting (XGboost). A receiver operating characteristic curve (ROC) analysis was conducted to evaluate the performance of the model.

Results: Quantitative CT measurements derived from human-audited segmentation results showed that COVID-19 patients had significantly decreased numbers of infected lobes compared to patients in the CAP group {median [interquartile range (IQR)]: 4 [3, 4] and 4 [4, 5]; P=0.031}. The infected percentage (%) of the whole lung was significantly more elevated in the CAP group [6.40 (2.77, 11.11)] than the COVID-19 group [1.83 (0.65, 4.42); P<0.001], and the same trend applied to each lobe, except for the superior lobe of the right lung [1.81 (0.09, 5.28) for COVID-19 *vs.* 1.32 (0.14, 7.02) for CAP; P=0.649]. Additionally, the highest proportion of infected lesions were observed in the CT value range of (–470, –370) Hounsfield units (HU) in the COVID-19 group. Conversely, the CAP group had a value range of (30, 60) HU. Radiomic model using corrected ROIs exhibited the highest area under ROC (AUC) of 0.990 [95% confidence interval (CI): 0.962 –1.000] using Lasso for feature selection and MLP for classification.

Conclusions: The proposed radiomics model based on human-audited segmentation made accurate differential diagnoses of COVID-19 and CAP. The quantification of CT measurements derived from DL could potentially be used as effective biomarkers in current clinical practice.

Keywords: coronavirus disease 2019 (COVID-19); deep learning (DL); radiomics; classification; human audit

Introduction

Since the coronavirus disease 2019 (COVID-19) outbreak started in late December 2019, researchers from all over the world have devoted great efforts to recognizing, characterizing, and treating the disease. COVID-19 was reported as a novel respiratory pandemic disease caused by a coronavirus, and to date, viral nucleic acid detection using real-time reverse-transcriptase polymerase-chain-reaction (RT-PCR) remains the known standard of reference. However, chest computed tomography (CT) has proven its capacity to outperform RT-PCR in the diagnosis and course monitoring of COVID-19 (1,2). Abundant evidence has suggested that the characteristic features displayed on CT images reveal early signs of COVID-19. Typical CT signatures, including ground-glass opacity (GGO), bilateral or peripheral distributed lesions, septal thickening, and consolidation, also referred to as "crazy-pavings" in the advanced stage (3,4), have been widely reported in the literature (5).

Advanced artificial intelligence (AI) techniques, such as deep learning (DL) and machine learning, have been actively involved in attempts to speed up clinical tasks for patients' benefits (6,7). During this period, many AI studies have successfully performed COVID-19 detection and classification tasks (8), typically to differentiate between COVID-19 and other popular causes of pneumonia [including community-acquired pneumonia [CAP]] (7,9). However, the performance of these models has often suffered from variations and a lack of explanations due to the black-box nature of the DL.

Thus, in this study, we aimed to develop a conventional CT-based radiomics model that implemented DL and human auditing. By manually reviewing and editing the DL-based segmentation results, we extracted radiomics features that represented a better reflection of lesion characteristics, which in turn facilitated the differentiation between COVID-19 from CAP. We present the following article in accordance with the STROBE reporting checklist (available at https://dx.doi.org/10.21037/apm-20-2625).

Methods

The retrospective study was approved by the Institutional Review Board of the Affiliated Hospital of Nanjing University Medical School (No. 2019-100-01). Informed consent was waived due to the retrospective nature. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Patients and datasets

We obtained CT images of 50 COVID-19 patients, comprising 14 patients from the Affiliated Drum Tower Hospital of Nanjing University Medical School and 36 patients from the Second Affiliated Hospital of the Nanjing University of Chinese Medicine between January 25, 2020, and March 1, 2020. Repeated swab tests confirmed positive diagnoses of COVID-19. 60 CAP patients were randomly selected from the participating hospitals for the same period. To be eligible to participate in the study, the patients had to meet the following inclusion criteria: (I) have confirmation of a pneumonia etiology based on swab tests for COVID-19 patients and sputum cultures for CAP patients; (II) have decent quality CT scans; and (III) have positive CT findings of pneumonia. Conversely, patients were excluded from the study if they met any of the following exclusion criteria: (I) had an unconfirmed etiology of pneumonia; and/or (II) had no CT scans available or had negative CT findings of pneumonia. Detailed clinical information was available from patients at 1 participating hospital (n=34), including data on onset symptoms and laboratory results, such as white blood cell and lymphocyte counts.

CT protocols

The chest CT examinations were performed using scanners from several manufacturers with standard imaging protocols. Each volumetric chest CT was scanned at the end of inhalation. The scan ranged from the apex of the lung to the diaphragm. The scan parameters were as follows: 120 kVp, rotation time of 0.5 s, the pitch of 0.75, a slice-thickness range from the mediastinal window of 3 mm, and a lung window of 1.25 mm.

DL-based segmentation and human auditing

The DL-based segmentation algorithm was built in-house

on the InferScholar research platform by InferVision (https://www.infervision.com/, Beijing, China) to segment the infected pneumonia areas and generate quantitative measurements of segmented masks. The algorithm development and validation processes have been described previously in detail (10). Briefly, this algorithm was implemented with a U-net-like deep convolution neural network model and employed to segment regions of interest (ROIs), including lungs, lobes, and detected opacities. These automatic segmented ROIs were subsequently reviewed by 3 experienced radiologists who were blinded to the patients' clinical status. Radiologists were asked to audit the DL-derived results independently, and adjust the segmented ROIs manually if necessary. This manual adjustment included enlarging or shrinking the ROIs based on the subjective evaluations by radiologists who had reached a consensus, removing obvious false positives, such as vascular artifacts, subpleural interstitial changes, and adding missed lesions (false negatives) by manual annotation with the hand tool implemented in the platform. Eventually, after human auditing, the delineated ROIs were processed into the algorithm analysis module to generate quantitative measurements, including the infected lobe numbers, the percentage of involved lesions' volumes (of the whole lung and each lobe), and lesion percentages based on different CT attenuation values. *Figure 1* depicts our proposed methodological framework for human-centered segmentation auditing derived from DL and radiomic model building.

Feature extraction and selection

For each CT series, 1,454 features, which could be subdivided into 7 classes by definition, were extracted from adjusted ROIs based on DL-based segmentation and human auditing. After dimension reduction, 7-dimensional features remained, including shape, texture, and first-order statistics. The shape features calculated the largest three-dimensional diameter and surface area, and the texture features were defined as regular texture features, such as gray-level dependence matrix and (gray level size zone matrix. First-order statistics reflect the distribution of voxel intensities within a region. To reduce overfitting or solve multicollinearity issues, we considered 4 feature selection approaches, including L1 regularization, the least absolute shrinkage, and selection operator (LASSO), ridge regression, and the Z test (11). The feature extraction and selection process were implemented using Python

3.6. Subsequently, the radiological features most closely associated with the determination of the 2 disease groups were obtained.

The development of a personalized classification model

Four different classifiers, including the logistic regression (LR), multi-layer perceptron (MLP), support vector machine (SVM), and extreme Gradient Boosting (XGboost), were used to predict the COVID-19 response. The combination of 4-feature selection methods and 4 classifiers was investigated by conducting a 5-fold cross-validation, a standard validation technique (12). The feature selection methods were included in the cross-validation algorithm to contribute to the final model fit reflected in the performance metrics (13). The classification performances were thus evaluated using the receiver operating characteristic (ROC) and the area under the ROC curve (AUC).

Statistical analysis

The statistical analysis was performed using SPSS (version 24.0, IBM Crop, NY, USA). Distribution normality was assessed using the Shapiro-Walk test. Continuous variables were expressed as mean [standard deviation [SD]] for normal distributed data and median [interquartile range [IQR]] for non-normal distributed data. Categorical variables were presented as frequencies (percentages, %). Patients' demographic and clinical characteristics were assessed using the Chi-squared test (or Fisher's exact test as appropriate). The quantitative measurements based on the segmentation results between the COVID-19 and CAP group were carried out using the Mann-Whitney test. A P value <0.05 was considered statistically significant.

Results

Patients population

A total of 50 COVID-19 patients were initially selected for this study; however, 4 patients were excluded due to ultra-thin CT slice thickness (≤0.50 mm) with a super-resolution beyond the algorithm prediction range, and 3 patients were excluded due to poor segmentation quality as determined by the human-audited segmentation results. Ultimately, our dataset comprised 43 COVID-19 patients (mean age: 41±15 years old, of whom 58.1% were male) and 60 CAP patients (mean age: 55±18 years old, of whom 63.3% were
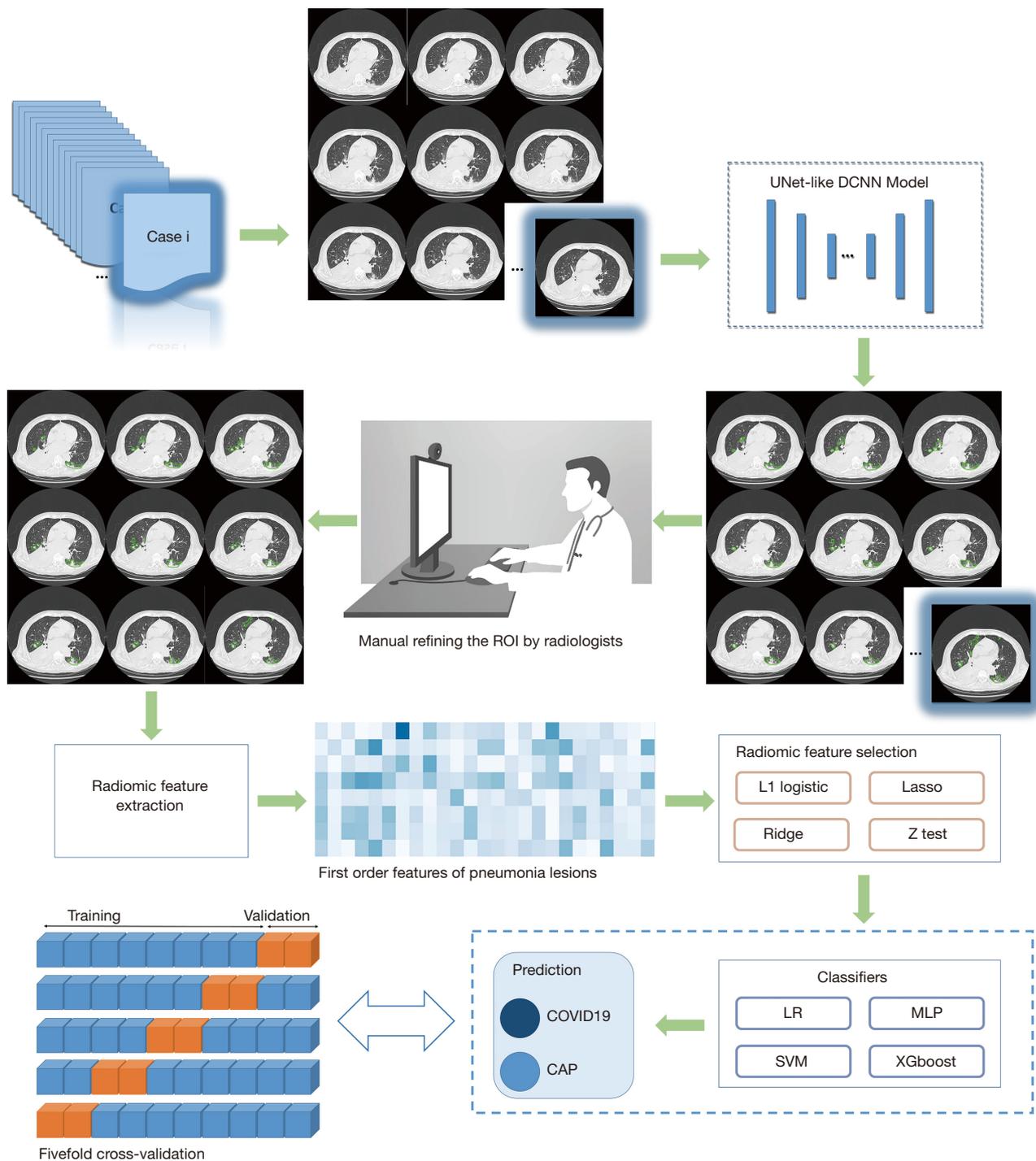
Figure 1 Our proposed methodological framework for human-centered segmentation auditing derived from DL and radiomic model building. DCNN, deep convolutional neural network; ROI, regions of interest; COVID-19, coronavirus disease 2019; CAP, community acquired pneumonia; LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine; XGboost, extreme gradient boosting.

Table 1 Demographic and clinical information of COVID-19 and CAP groups

Characteristics	COVID-19 group (n=43)	CAP group (n=60)	P value
Gender, n (%)	n=4	n=60	0.59
Female	18 (41.9)	22 (36.7)	
Male	25 (58.1)	38 (63.3)	
Age (years), mean (SD) (n=103)	41.0 (15.5)	54.7 (17.8)	<0.001*
Symptoms, n (%)	n=14	n=30	
Fever	7 (87.5)	12 (48.0)	0.098
Cough	4 (44.4)	13 (52.0)	1.000
Muscle ache	3 (33.3)	5 (20.0)	0.649
Lab test: WBC, n (%)	n=8	n=25	0.01*
Increased (>9.5×10^9/L)	0 (0.0)	12 (48.0)	
Normal (3.5–9.5×10^9/L)	6 (66.7)	12 (48.0)	
Decreased (<3.5×10^9/L)	2 (22.2)	1 (4.0)	
Lab test: lymphocytes, n (%)	n=8	n=25	0.007*
Increased (>3.2×10^9/L)	1 (11.1)	1 (4.0)	
Normal (1.1–3.2×10^9/L)	6 (66.7)	7 (28.0)	
Decreased (<1.1×10^9/L)	1 (11.1)	17 (68.0)	

Data were expressed as the number (percentage) of patients. Ages were reported as means (standard deviations). *, a statistically significant difference between groups. COVID-19, coronavirus disease 2019; CAP, community acquired pneumonia; WBC, white blood cell count.

male). The most prevalent onset symptoms, such as fever and coughing, were found in both the COVID-19 and CAP groups (*Table 1*). All the enrolled COVID-19 patients were classified as moderate cases according to the Diagnosis and Treatment of Novel Coronavirus Pneumonia (trial version seven) published by the National Health Commission of the People's Republic of China (14).

Quantitative CT measurements generated by DL-based segmentation and human auditing

After a deep examination of false positives and negatives of the AI-enabled segmentation results and the insufficiently segmented regions, final segmentation masks were confirmed by 3 experienced radiologists and human auditing (*Figure 2*). The quantitative CT measurements were obtained and are set out in *Table 2*. The results showed that the numbers of infected lobes were significantly lower in the COVID-19 group [median (IQR): 4 (3 to 4)] than the CAP group [4 (4 to 5)] (P=0.031) Concerning the percentage of

lung involvement in the whole lung, that of the CAP group was significantly more elevated than that of the COVID-19 group [median (IQR): 1.83% (0.65%, 4.42%) *vs.* 6.40% (2.77%, 11.11%); P<0.001]. Similarly, the percentage of lung involvement per lobe was significantly higher in the CAP group than the COVID-19 group, except for that of the right upper lobe [1.81 (0.09, 5.28) for the COVID-19 group *vs.* 1.32 (0.14, 7.02) for the CAP group; P=0.649].

We also investigated the percentage of lung involvement in varying CT attenuation value ranges and observed that the highest proportion of lesions in the COVID-19 group were in the CT value of (–470, –370) HU, and of (30 to 60) HU in the CAP group. Significant differences were observed between the 2 groups for all CT value ranges (P<0.05) except for the range of (–370, –270). As *Table 3* shows, we also investigated the intra-class correlation coefficient (ICC) statistics between the segmentation results (infected volume fractions) derived from DL alone and human auditing. The results showed that there was good consistency in the volume proportion of the total lung
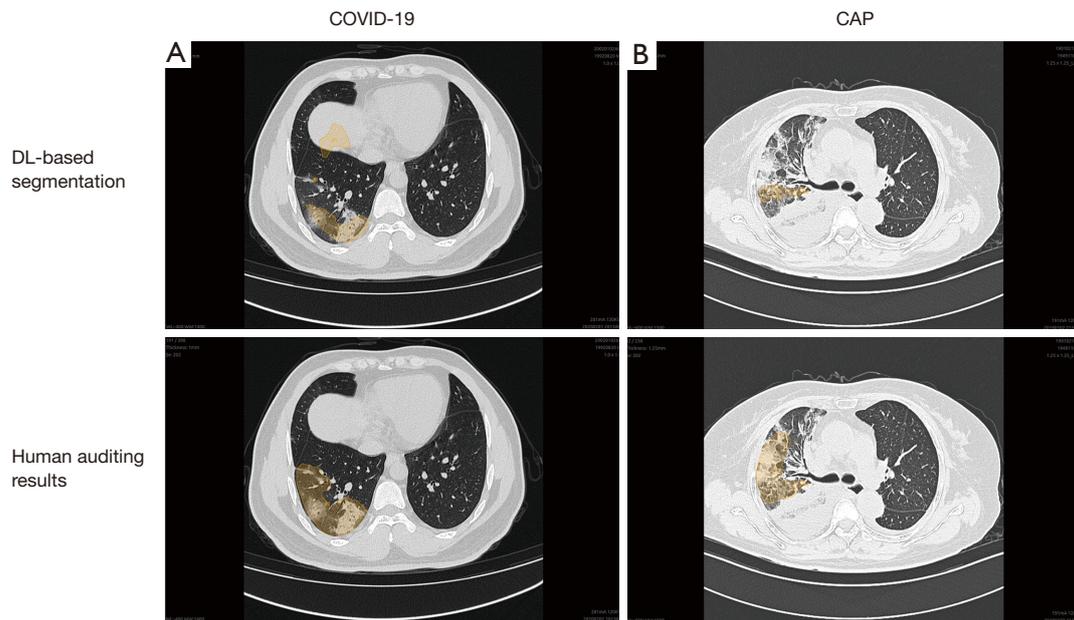
Figure 2 A comparison of the segmentation results derived from DL and human auditing. A CT slice from a representative COVID-19 (A) and CAP (B) case overlaid with original segmented ROIs (yellow masks) derived from DL segmentation results (top), and manually corrected segmented masks (bottom). DL, deep learning; COVID-19, coronavirus disease 2019; CAP, community acquired pneumonia.

Table 2 The infected lobes and distributions of the COVID-19 group and the CAP group

Descriptors	COVID-19 group (n=43), median (IQR), %	CAP group (n=60), median (IQR), %	P value
No. of infected lobes	4 (3 to 4)	4 (4 to 5)	0.031*
Locations			
The whole lung	1.83 (0.65, 4.42)	6.40 (2.77, 11.11)	<0.001*
Superior lobe of right lung	1.81 (0.09, 5.28)	1.32 (0.14, 7.02)	0.649
Middle lobe of right lung	0.01 (0.00, 0.67)	0.16 (0.00, 6.44)	0.030*
Inferior lobe of right lung	1.77 (0.32, 7.10)	12.72 (1.77, 25.78)	<0.001*
Superior lobe of left lung	0.25 (0.00, 1.65)	3.20 (0.25, 6.62)	<0.001*
Superior lobe of left lung	1.17 (0.05, 7.39)	10.90 (1.94, 23.26)	<0.001 *
CT value ranges			
(–1,000, –570) HU	1.70 (0.40, 4.20)	2.18 (0.56, 5.71)	0.418
(–570, –470) HU	7.74 (1.61, 23.01)	3.63 (1.16, 10.05)	0.040*
(–470, –370) HU	9.35 (1.39, 22.72)	4.24 (1.30, 10.62)	0.040*
(–370, –270) HU	6.32 (1.05, 17.55)	4.64 (1.87, 11.63)	0.561
(–270, –170) HU	2.81 (0.86, 9.13)	6.22 (2.40, 11.26)	0.017*
(–170, –70) HU	1.57 (0.33, 3.81)	6.37 (2.52, 11.39)	<0.001*
(–70, 30) HU	1.22 (0.15, 2.45)	7.16 (3.03, 12.86)	<0.001*
(30, 60) HU	1.10 (0.10, 2.21)	8.10 (3.26, 13.36)	<0.001*
(60, 1,000) HU	0.83 (0.09, 1.73)	7.61 (3.17, 13.91)	<0.001*

Data were displayed as median [interquartile range (IQR)]. *, P<0.05 indicates statistical significance based on a Mann-Whitney test. COVID-19, coronavirus disease 2019; CAP, community acquired pneumonia; IQR, inter-quartile range; CT, computed tomography; HU, Hounsfield units.

Table 3 ICC statistics between segmentation results (infected volumes fractions) derived from DL and human auditing

Locations (DL *vs.* human auditing)	COVID-19 group (n=43), ICC (95% CI)	CAP group (n=60), ICC (95% CI)
The whole lung	0.908 (0.758, 0.958)	0.931 (0.887, 0.958)
Superior lobe of right lung	0.352 (0.054, 0.591)	0.266 (0.015, 0.486)
Middle lobe of right lung	0.491 (0.230, 0.687)	0.896 (0.832, 0.937)
Inferior lobe of right lung	0.610 (0.381, 0.769)	0.779 (0.656, 0.861)
Superior lobe of left lung	0.902 (0.823, 0.946)	0.858 (0.772, 0.913)
Superior lobe of left lung	0.935 (0.882, 0.964)	0.698 (0.541, 0.808)

COVID-19, coronavirus disease 2019; CAP, community acquired pneumonia; DL, deep learning; ICC, intra-class correlation coefficient; CI, confidence interval.

Table 4 Summary of AUC values for classifiers and feature selection methods with average predictive performance taken over the 5-fold validation

| Feature selection	Classifiers	AUC	AU-PRC	SEN	SPEC	F1_Score	ACC
L1 regularization	LR	0.928	0.820	0.935	0.885	0.896	0.897
	MLP	0.948	0.804	0.935	0.869	0.887	0.888
	SVM	0.948	0.788	0.848	0.902	0.866	0.869
	XGboost	0.938	0.772	0.957	0.820	0.869	0.869
LASSO	LR	0.975	0.921	0.957	0.967	0.952	0.953
	MLP	0.990	0.942	0.957	0.984	0.962	0.963
	SVM	0.981	0.904	0.891	0.984	0.932	0.935
	XGboost	0.946	0.772	0.957	0.820	0.869	0.869
RIDGE	LR	0.949	0.836	0.935	0.902	0.905	0.907
	MLP	0.968	0.880	0.848	0.984	0.912	0.916
	SVM	0.962	0.817	0.870	0.918	0.885	0.888
	XGboost	0.938	0.808	0.913	0.885	0.886	0.888
Z test	LR	0.946	0.793	0.913	0.869	0.877	0.879
	MLP	0.961	0.769	0.870	0.869	0.857	0.860
	SVM	0.928	0.752	0.891	0.836	0.849	0.850
	XGboost	0.928	0.784	0.870	0.885	0.867	0.869

LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine; Xgboost, eXtreme gradient boosting; AUC, area under the receiver characteristic curve; AU-PRC, area under the precision-recall curve; SEN, sensitivity; SPEC, specificity; ACC, accuracy.

infection between the groups (*Table 3*).

Performance of CT radiomic model in differentiating between COVID-19 and CAP

Sixteen models were established in this study. For each model, the evaluation metrics were AUC, the area under the precision-recall curve (AU-PRC), sensitivity (SEN),

specificity (SPEC), F1-score, and accuracy (ACC). *Table 4* summarizes the varying performance of each model for each classifier concerning the different feature selection methods. Among all the models, the Lasso regression yielded higher AUC values for all used classifiers. Notably, the MLP classifier obtained the highest AUC of 0.990 [95% confidence interval (CI): 0.962–1.000]. The results indicated that combining the LASSO with the MLP
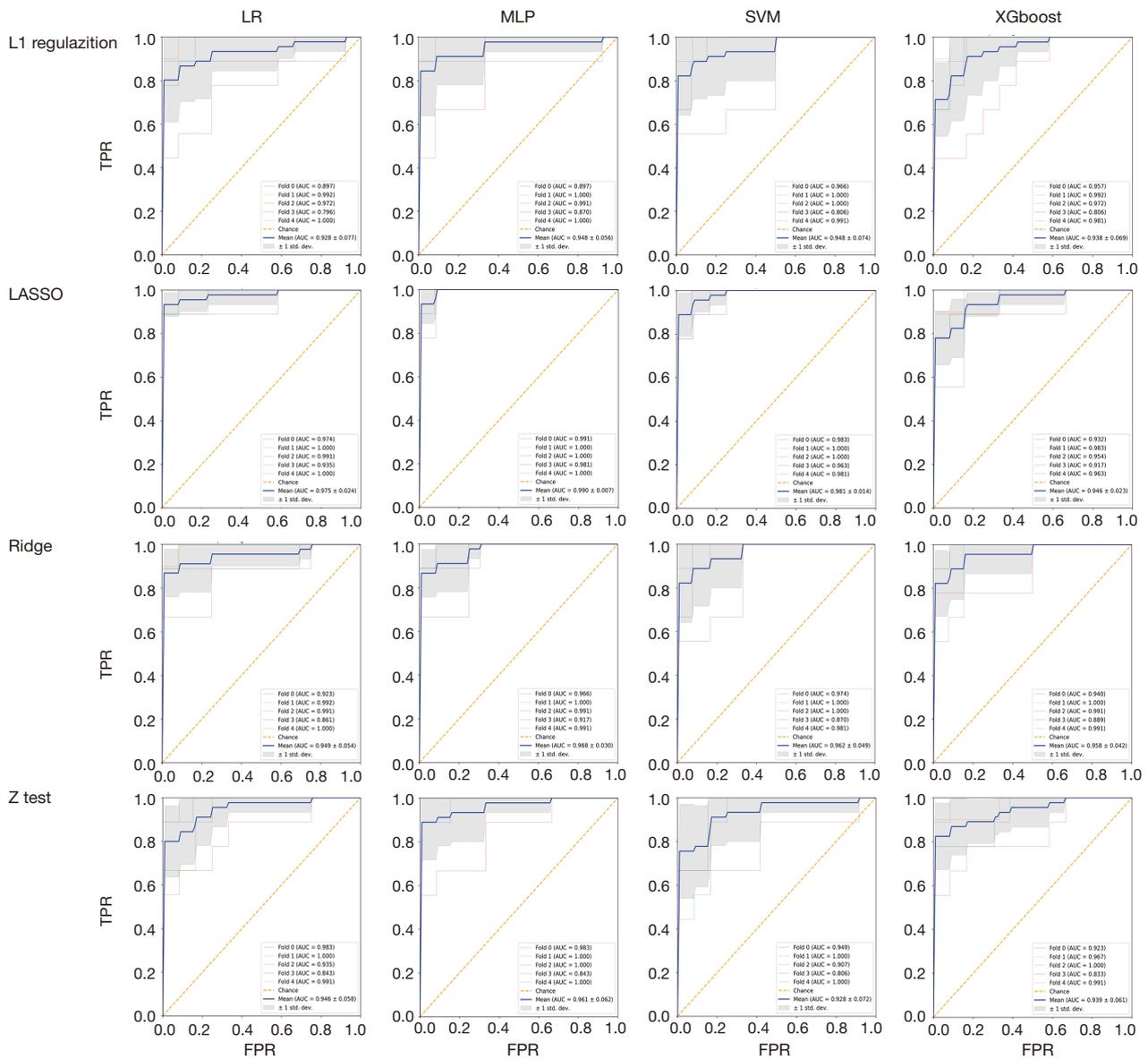
Figure 3 ROC curves of CT Radiomic models (over the 5-fold validations) for each classifier across the various feature selection methods in differentiating COVID-19 from CAP. COVID-19, coronavirus disease 2019; CAP, community acquired pneumonia; LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine; Xgboost, eXtreme gradient boosting; TPR, true positive rate; FPR, false positive rate; AUC, area under the receiver operating characteristic curve; std.dev, standard deviation.

classifiers resulted in the best-performing model with the highest ACC (96.3%), SEN (95.7%), SPEC (98.4%), and AU-PRC (0.942) (*Figure 3*).

Discussion

In this study, we explored a methodological framework

to differentiate between COVID-19 and CAP using conventional CT radiomics models that were implemented with DL and human auditing. The segmentation results derived from the DL-based segmentation algorithms were manually corrected by experienced radiologists and eventually used to extract radiomics features for classification purposes. The quantitative CT

measurements provided the quantitative volume fractions of lung involvement for the whole lung and each lobe and different CT attenuation ranges. Using this information, radiologists were able to evaluate the involvement of infected regions in both lungs. This information could also be used as effective biomarkers for monitoring illness progress and curative effects.

Previous studies have employed subjective evaluations using a scoring system rated by radiologists to assess disease severity (15-17). However, this semi-quantitative approach often suffers from inaccuracy and inconsistency among different readers. Additionally, it is difficult to make findings based on visual interpretations of images alone due to the disease's lack of specificity (18,19). As our results showed, patients in the CAP group had a significantly higher lung involvement percentage in both the whole lung and individual lobes (except for the right upper lobe) than those in the COVID-19 group. Conversely, their corresponding distributions across the HU spectrum differed significantly. A higher proportion of infected lesions were found in the low-density range (–570, –270) in the COVID-19 group than the CAP group, and lung involvement peaked at higher CT value ranges (–270, 60). As different CT values represented different types of lesions (20), the results suggest that a lower density of GGO was more common in the COVID group than the CAP group, which presented with lesions of higher density.

Conventional radiomics models were developed and evaluated concerning different feature selection algorithms and classifiers to identify COVID-19. Among the 16 different combinations, Lasso regression with MLP was the most predictive classifier with an AUC of 0.990 (95% CI: 0.962–1.000). This predictive accuracy was comparable to the DL models proposed by Li *et al*. (AUC =0.96 for COVID-19) and Song *et al*. (AUC =0.98 at the image level and 0.99 at the patient level) (7,21). Unlike the DL CAP algorithm, which requires a significant amount of labeled data for training and validation, our proposed model enabled binary classification using features derived from DL-based segmentation masks reviewed by humans. The involvement of a human factor maximizes the success of developing algorithms powered by DL in radiology.

This study had several limitations. First, as it was a multi-center study, we could not obtain clinical information from all patients. Some laboratory data were missing, and thus, comparisons between the groups could not be made. Second, the study had several common limitations, such

as small sample size and the lack of an external validation dataset. However, a methodological foundation was established for further analysis to achieve a larger-scale differential diagnosis of COVID-19 pneumonia. Finally, a prospective study using the proposed model is needed to address the clinical diagnostic value of the model.

Conclusions

In conclusion, we developed CT radiomics models implemented with human-centered DL-based algorithms to differentiate between COVID-19 and CAP. This framework could provide a methodological foundation for differential diagnoses and potentially reduce the clinical burden caused by the pandemic. Future work will extend this approach to a larger dataset to further refine this technology for diffuse pulmonary diseases, such as pulmonary alveolar proteinosis and interstitial pneumonia. Additionally, the fusion of radiomics features and local binary pattern-based edge-texture features may be able to undertake the classification task with a limited dataset in medical imaging.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at https://dx.doi.org/10.21037/apm-20-2625

Data Sharing Statement: Available at https://dx.doi.org/10.21037/apm-20-2625

Peer Review File: Available at https://dx.doi.org/10.21037/apm-20-2625

Conflicts of Interest: All authors have completed the ICMJE

uniform disclosure form (available at https://dx.doi.org/10.21037/apm-20-2625). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The retrospective study was approved by the Institutional Review Board of the Affiliated Hospital of Nanjing University Medical School (No. 2019-100-01). Informed consent was waived due to the retrospective nature. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Pan F, Ye T, Sun P, et al. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology 2020;295:715-21.
2. Wang Y, Dong C, Hu Y, et al. Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. Radiology 2020;296:E55-64.
3. Chung M, Bernheim A, Mei X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 2020;295:202-7.
4. Ng MY, Lee EYP, Yang J, et al. Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. Radiol Cardiothorac Imaging 2020;2:e200034.
5. Bao C, Liu X, Zhang H, et al. Coronavirus Disease 2019 (COVID-19) CT Findings: A Systematic Review and Meta-analysis. J Am Coll Radiol 2020;17:701-9.
6. Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020;26:1224-8.
7. Li L, Qin L, Xu Z, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020;296:E65-71.
8. Xu X, Jiang X, Ma C, et al. A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. Engineering (Beijing) 2020;6:1122-9.
9. Wang S, Kang B, Ma J, et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 2021;31:6096-104.
10. Huang L, Han R, Ai T, et al. Serial Quantitative Chest CT Assessment of COVID-19: A Deep Learning Approach. Radiol Cardiothorac Imaging 2020;2:e200075.
11. Lee J, Jeong J, Jun C. Markov blanket-based universal feature selection for classification and regression of mixed-type data. Expert Syst Appl 2020;158:113398.
12. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018;2:36.
13. Delzell DAP, Magnuson S, Peter T, et al. Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data. Front Oncol 2019;9:1393.
14. National Health Commission of the Peoples Republic of China. Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment. 2020. Available online: http://kjfy.meeting.so/msite/news/show/cn/3337.html
15. Yang R, Li X, Liu H, et al. Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. Radiol Cardiothorac Imaging 2020;2:e200047.
16. Liang T, Liu Z, Wu CC, et al. Evolution of CT findings in patients with mild COVID-19 pneumonia. Eur Radiol 2020;30:4865-73.
17. Jin YH, Cai L, Cheng ZS, et al. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Mil Med Res 2020;7:4.
18. Simpson S, Kay FU, Abbara S, et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA - Secondary Publication. J Thorac Imaging 2020;35:219-27.
19. Shi H, Han X, Zheng C. Evolution of CT Manifestations in a Patient Recovered from 2019

Novel Coronavirus (2019-nCoV) Pneumonia in Wuhan, China. Radiology 2020;295:20.

20. Du S, Gao S, Huang G, et al. Chest lesion CT radiological features and quantitative analysis in RT-PCR turned negative and clinical symptoms resolved COVID-19 patients. Quant Imaging Med Surg 2020;10:1307-17.

21. Song Y, Zheng S, Li L, et al. Deep learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT images. IEEE/ACM Trans Comput Biol Bioinform 2021. doi: 10.1109/TCBB.2021.3065361. [Epub ahead of print].

Cite this article as: Xin X, Mo R, Shao M, Yang W, Li D, Zhang Y, Wang H, Liu B, Tian S, Chen W, Wu J, Zhu B, Zhou K, Du C, Zhang B. A human-computer collaboration for COVID-19 differentiation: combining a radiomics model with deep learning and human auditing. Ann Palliat Med 2021;10(7):7329-7339. doi: 10.21037/apm-20-2625