Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

European Journal of Radiology 137 (2021) 109602

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Texture feature-based machine learning classifier could assist in the diagnosis of COVID-19

Zhiyuan Wu [a,b,1,2], Li Li [c,1,2], Ronghua Jin [c], Lianchun Liang [c], Zhongjie Hu [c], Lixin Tao [a,b], Yong Han [a,b], Wei Feng [a,b], Di Zhou [a], Weiming Li [a,b], Qinbin Lu [d], Wei Liu [e], Liqun Fang [e], Jian Huang [f], Yu Gu [g,h], Hongjun Li [c,**], Xiuhua Guo [a,b,*]

[a] Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, China
[b] Beijing Municipal Key Laboratory of Clinical Epidemiology, Capital Medical University, Beijing, China
[c] Beijing Youan Hospital, Capital Medical University, Beijing, China
[d] Department of Laboratorial Science and Technology, School of Public Health, Peking University, Beijing, China
[e] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China
[f] School of Mathematical Sciences, University College Cork, Cork, Ireland
[g] Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing, 100029, China
[h] Department of Chemistry, Institute of Inorganic and Analytical Chemisty, Goethe-University, Frankfurt, 60438, Germany

ARTICLE INFO

Keywords:
Coronavirus disease 2019
Computed tomography
Texture analysis
Machine learning

ABSTRACT

Purpose: Differentiating COVID-19 from other acute infectious pneumonias rapidly is challenging at present. This study aims to improve the diagnosis of COVID-19 using computed tomography (CT).
Method: COVID-19 was confirmed mainly by virus nucleic acid testing and epidemiological history according to WHO interim guidance, while other infectious pneumonias were diagnosed by antigen testing. The texture features were extracted from CT images by two radiologists with 5 years of work experience using modified wavelet transform and matrix computation analyses. The random forest (RF) classifier was applied to identify COVID-19 patients and images.
Results: We retrospectively analysed the data of 95 individuals (291 images) with COVID-19 and 96 individuals (279 images) with other acute infectious pneumonias, including 50 individuals (160 images) with influenza A/B. In total, 6 texture features showed a positive association with COVID-19, while 4 features were negatively associated. The mean AUROC, accuracy, sensitivity, and specificity values of the 5-fold test sets were 0.800, 0.722, 0.770, and 0.680 for image classification and 0.858, 0.826, 0.809, and 0.842 for individual classification, respectively. The feature 'Correlation' contributed most both at the image level and individual level, even compared with the clinical factors. In addition, the texture features could discriminate COVID-19 from influenza A/B, with an AUROC of 0.883 for images and 0.957 for individuals.
Conclusions: The developed texture feature-based RF classifier could assist in the diagnosis of COVID-19, which could be a rapid screening tool in the era of pandemic.

Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; ACE2, angiotensin-converting enzyme 2; COVID-19, coronavirus disease 2019; CT, computed tomography; RT-PCR, real-time polymerase chain reaction; GGO, ground glass opacity; RF, random forest; AUROC, area under the receiver operating characteristic curve; NSDTCT, non-subsampled dual-tree complex contourlet transform; GLCM, grey level co-occurrence matrix; IQR, interquartile range.

* Corresponding author at: Department of Epidemiology and Health Statistics, School of the Public Health, Capital Medical University, No. 10 Xitoutiao, Youanmen Street, Beijing, 100069, China.

** Corresponding author at: Department of Radiology, Beijing Youan Hospital, Capital Medical University, No. 8 Xitoutiao, Youanmen Street, Beijing, 100069, China.

E-mail addresses: wuxiaozhi@ccmu.edu.cn (Z. Wu), 15001017169@139.com (L. Li), 93353503@qq.com (R. Jin), llc671215@sohu.com (L. Liang), yfcyt@139.com (Z. Hu), 13426176692@163.com (L. Tao), hy_vip@126.com (Y. Han), sharkip@qq.com (W. Feng), 18810675096@163.com (D. Zhou), Lucien_Lee727@163.com (W. Li), qingbinlu@bjmu.edu.cn (Q. Lu), lwbime@163.com (W. Liu), fang_lq@163.com (L. Fang), j.huang@ucc.ie (J. Huang), guyu@mail.buct.edu.cn (Y. Gu), lihongjun00113@126.com (H. Li), statguo@ccmu.edu.cn (X. Guo).

[1] First co-author.
[2] Zhiyuan Wu and Li Li analyzed the data and drafted the manuscript together.

https://doi.org/10.1016/j.ejrad.2021.109602
Received 6 August 2020; Received in revised form 11 January 2021; Accepted 9 February 2021
Available online 15 February 2021

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel positive-sense single-stranded RNA virus, is transported into host cells by an interaction between the spike (S) protein and the peptidase domain of angiotensin-converting enzyme 2 (ACE2) [1]. Similar to severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), SARS-CoV-2 can infect and replicate in multiple human cell types and cause coronavirus disease 2019 (COVID-19), which mainly affects lung tissue [2]. COVID-19 was first reported in December 2019 in Wuhan, China, spreads through acute human-to-human transmission, and has become a public health threat worldwide [3]. The WHO characterized COVID-19 as a pandemic on 11 March 2020 [4], and the cumulative number of confirmed COVID-19 cases has increased rapidly to 52,991,639 as of 12 November 2020; 1,294,483 patients have died worldwide, which indicates a high transmissibility of the novel coronavirus [5].

COVID-19 may progress to acute severe pneumonia, along with intestinal, liver, nervous and multiple system damage and other corresponding symptoms [6]. Most infected patients develop fever and dyspnoea, with pulmonary inflammatory lesions, as revealed by chest radiography, such as X-ray examination and computed tomography (CT) imaging [7]. Currently, the genome sequencing information of SARS-CoV-2 has been identified, and the diagnosis of COVID-19 largely depends on laboratory tests of virus-specific nucleic acids by real-time polymerase chain reaction (RT-PCR). CT examination has been used to evaluate the pulmonary condition of those suspected SARS-CoV-2 infections [8]. CT supplies evidence for diagnosing COVID-19, especially when RT-PCR is negative in subjects who are highly suspected to have COVID-19 according to the latest WHO guidance [9]. In the early stages of COVID-19, CT imaging characteristics reveal multiple small plaques and interstitial changes, which may develop into ground glass opacity (GGO) in severe cases [10]. However, these CT findings of COVID-19 always overlap with other acute respiratory infections, such as influenza A/B pneumonia. It is difficult and challenging to accurately and rapidly differentiate COVID-19 from other acute respiratory infections, especially influenza A/B pneumonia, which is also of high incidence at present.

Radiomics is an emerging technique that extracts high-dimensional quantitative image features. In radiomics, texture features describe the spatial relationship between image intensities and are of great interest. An increasing number of studies have suggested that CT texture features, which reflect biological heterogeneity, could provide extensive diagnostic information in clinical practice [11]. Combined with machine learning classifiers, CT texture analysis has been successfully applied in the identification of many diseases, such as cancers [12] and Alzheimer's disease [13]. Huseyin et al. [14] compared the detection accuracy of COVID-19 images using texture features, but they only focused on the image level. Zeng et al. [15] composed a nomogram tool combining CT and clinical information to differentiate COVID-19 patients from influenza A patients. Whether a CT texture feature-based classifier could assist in differentiating between COVID-19 and other infectious pneumonias including influenza A and influenza B, remains unclear. The importance of CT texture features and clinical factors has not been performed.

In this study, we proposed that our classification tool can be used to distinguish COVID-19 images and individuals from other acute respiratory infections using CT texture features and the random forest (RF) algorithm. We evaluated the importance of texture features and their potential applicability for rapid COVID-19 diagnosis in pandemic situations.

2. Material and methods

2.1. Study sample

Individual information was collected from Youan Hospital, one of the affiliated hospitals of Capital Medical University, a designated centre for COVID-19 patients in Beijing. The inclusion criteria were as follows: 1) available confirmed diagnosis of COVID-19 or other infectious pneumonia; and 2) available CT images on admission. The patients were scanned before medical treatment. The exclusion criteria were as follows: 1) no related lesion on CT image; and 2) immunosuppressed individuals, such as those with HIV, and post chemotherapy patients.

COVID-19 was confirmed by virus nucleic acid testing and epidemiological history according to WHO interim guidance [16], while other acute respiratory infections were diagnosed by antigen testing. Chest CT images of all of the individuals were collected in DICOM format, and overall, there were 570 images involving pulmonary inflammatory lesions. In addition, sociodemographic information, clinical symptoms, and biochemistry examination results were obtained from the hospital information system. The CT images were uniformly collected from the PHILIPS Brilliance iCT 256 system. Data were analysed with EpiData 3.1 software (EpiData Association, Denmark). This study followed the guidelines of the Helsinki Declaration and was approved by the Ethics Committees of Capital Medical University. The individuals gave informed consent before taking part in the study.

2.2. Image preprocessing

All patients were scanned using a 5 mm CT slice thickness. Two radiologists with at least 5 years of work experience in the radiology department were committed to manually delineating the regions of interest, namely, the maximal regions involving inflammatory lesions. In general, 2–5 slices were extracted from each patient as the sizes of lesions on CT differed between individuals. The slices containing clear lesions were used in the analysis. ImageJ software (NIH, USA) was used to perform the image processing.

2.3. Texture feature extraction

In our study, we calculated the image texture features by combining the nonsubsampled dual-tree complex contourlet transform (NSDTCT) and the grey level cooccurrence matrix (GLCM). The ROI image was decomposed into 96 sub-bands with NSDTCT by following two steps. First, a dual-tree complex wavelet transform was used to decompose the original image into 2 low-frequency bands and 6 high-frequency bands of six directions ($\pm15^\circ$, $\pm45^\circ$, $\pm75^\circ$); then, 2^n band coefficients were set on the high-frequency level based on a nonsubsampled directional filter bank. After the NSDTCT process, GLCM was performed, and 14 texture features were extracted from each of the 96 sub-bands, including energy, inertia, inverse difference of moment, entropy, correlation, cluster of tendency, contrast, homogeneity, variance, maximum of probability, sum of mean, difference of mean, sum of entropy, and difference of entropy. The computational formulas of the texture features are provided in Table A.1. The mean value of each feature from 96 sub-bands was calculated as the ultimate determinant for an image. Then, the mean value of each feature from all of the images for one person was calculated as the ultimate determinant for an individual. MATLAB 2018b software (MathWorks, USA) was used to perform the feature extraction.

2.4. RF classifier

Machine learning is used to build a model from labelled training data and then to predict outcomes in new data [17]. In this study, the RF classifier was adopted due to its high accuracy and ability to handle large features with small samples. RF is also able to reveal the

importance and contribution of each input feature in the model [18]. In the RF algorithm, we constructed 500 decision trees as the model became robust at this point. RF randomly selected m variables as candidate variables from all the M features at each node, and the parameter "mtry" was set as the square root of M. In addition, the contribution of each feature was displayed based on the accuracy index.

2.5. Statistical analysis

Continuous variables are expressed as medians and interquartile ranges (IQRs), they and were compared with the Mann-Whitney *U* test Categorical variables are expressed as numbers and percentages (%) and were compared with the chi-squared test or Fisher's exact test. For texture features, the Bonferroni-adjusted P value was calculated to reduce class I statistical errors. The data were divided into a training set (80 %) and a test set (20 %) using a 5-fold cross validation strategy at the individual level, and the image data were divided correspondingly, such that all of the images from one individual were divided as a whole. The RF algorithm was implemented using the 'randomForest' package in R 3.5.3 software (R Foundation for Statistical Computing, Austria). The model performance was evaluated by accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC) analyses with the 'pROC' package in R 3.5.3 software. Statistical analyses were performed with SPSS 24.0 software (IBM, USA). Results were statistically significant with a two-sided $P < 0.05$.

3. Results

3.1. Characteristics of the individuals

From Dec 2019 to Mar 2020, we retrospectively enrolled 95 confirmed COVID-19 individuals. Subsequently, 96 individuals with other acute respiratory infections, including 50 influenza A/B pneumonia cases and 46 other infectious pneumonia cases, who presented to Youan Hospital during the same period were included in this study, as shown in Fig. 1. A total of 191 individuals were finally enrolled in this study. A total of 63.2 % of the COVID-19 individuals had cough symptoms, which was lower than that of individuals with other infectious pneumonias (63.2 % vs 83.3 %, $P = 0.003$). The white blood cell count, neutrophil percentage, and C-reactive protein were lower in the COVID-19 group, and the lymphocyte percentage was higher. There were no significant differences in age, sex, or fever. The details of the demographic and clinical characteristics are shown in Table 1.

3.2. Texture analysis

A total of 570 CT images containing the maximal pulmonary inflammatory lesions were manually segmented as shown in Fig. A.1. As described in texture feature extraction, we produced 14 features for each image. Then, the mean value of each feature from all of the images of one person was calculated as the ultimate determinant for that

Table 1
Characteristics of individuals with COVID-19 and other infectious pneumonias.

Clinical factors	Overall (n = 191)	Comparison between individuals		
		other infectious pneumonia (n = 96)	COVID-19 (n = 95)	P value
Age (years) [a]	54.50 [42.50,60.00]	55.00 [50.55,59.08]	50.00 [37.50,65.00]	0.698
Sex (male/female) [b]	94/97	51/45	43/52	0.346
Fever (yes/no) [b]	164/27	87/9	77/18	0.091
Cough (yes/no) [b]	140/51	80/16	60/35	0.003
White blood cell (109/L) [a]	6.10 [4.03,6.84]	6.60 [6.12,7.01]	4.26 [3.50,5.82]	<0.001
Neutrophil (%) [a]	71.30 [59.90,77.30]	75.60 [71.10,78.80]	63.50 [51.50,72.00]	<0.001
Lymphocyte (%) [a]	17.50 [15.10,28.75]	15.45 [14.93,16.00]	26.10 [18.80,34.55]	<0.001
Procalcitonin (mg/L) [a]	0.12 [0.10,0.15]	0.13 [0.10,0.15]	0.12 [0.10,0.15]	0.652
C-reactive protein (mg/L) [a]	39.45 [12.50,49.88]	44.60 [39.45,49.95]	16.30 [3.79,39.95]	<0.001

[a] Median (IQR), Mann-Whitney *U* test.
[b] Numbers of each category, chi-square test.

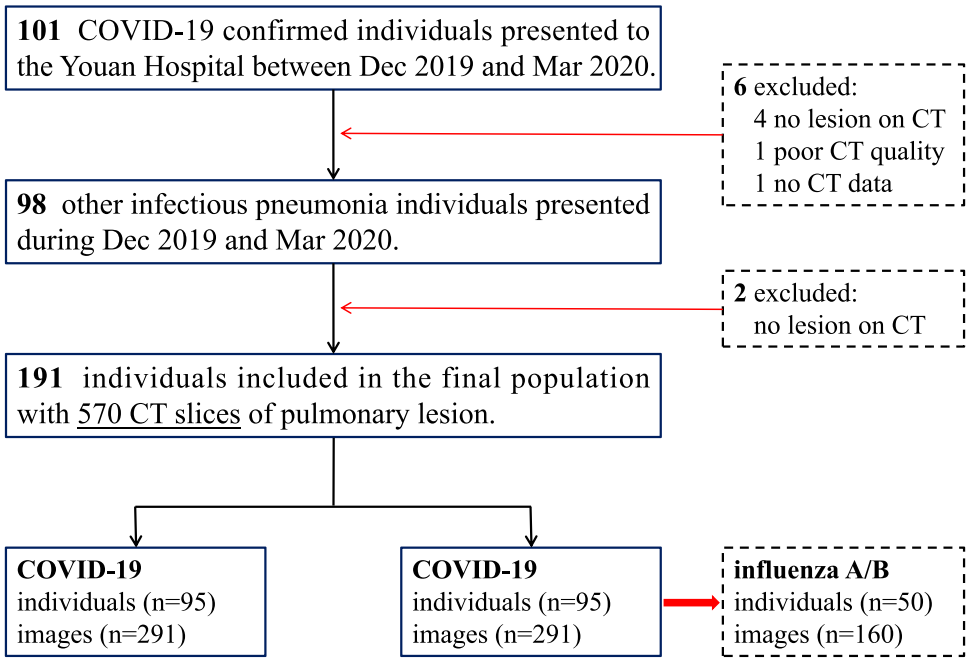

Fig. 1. Flow diagram of individuals included in the study.

individual. In addition, the distributions of these texture features at the image level and individual level between COVID-19 and other infectious pneumonias were analysed. As Table 2 shows, 6 texture features increased and 4 decreased both at the image level and at the individual level for COVID-19 subjects after Bonferroni adjustment. The box plots of the 10 texture features at the image level and the individual level are presented in Fig. 2.

Table 2
The distribution of texture features in the images and individuals between COVID-19 and other infectious pneumonias.

Comparison between images			
Texture feature	other infectious pneumonia (n = 279)	COVID-19 (n = 291)	P value
Cluster of Tendency	17.02[16.86,17.15]	16.99 [16.86,17.12]	0.526
Contrast	10.56[9.63,11.84]	11.21 [10.27,12.09]	<0.001*
Correlation	−0.08[−0.09, 0.07]	−0.06[−0.07, 0.06]	<0.001*
Difference of Entropy	3.54[3.47,3.62]	3.58[3.52,3.64]	<0.001*
Difference of Mean	2.37[2.22,2.54]	2.48[2.34,2.61]	<0.001*
Energy	0.03[0.02,0.03]	0.02[0.02,0.03]	<0.001*
Entropy	3.42[3.18,3.65]	3.65[3.45,3.84]	<0.001*
Homogeneity	0.44[0.42,0.46]	0.42[0.41,0.44]	<0.001*
Inverse Difference of Moment	0.36[0.34,0.39]	0.35[0.33,0.36]	<0.001*
Maximum of Probability	0.06[0.05,0.08]	0.05[0.05,0.06]	<0.001*
Inertia	8.52[8.44,8.58]	8.50[8.44,8.56]	0.586
Sum of Entropy	3.84[3.75,3.94]	3.94[3.85,4.02]	<0.001*
Sum of Mean	15.03[14.87,15.15]	15.00 [14.87,15.13]	0.526
Variance	78.05[76.61,79.28]	78.19 [76.98,79.31]	0.255
Comparison between individuals			
Texture feature	other infectious pneumonia (n = 96)	COVID-19 (n = 95)	P value
Cluster of Tendency	16.99[16.91,17.12]	17.00 [16.91,17.08]	0.619
Contrast	10.44[9.73,11.51]	11.31 [10.39,12.11]	0.002
Correlation	−0.08[−0.09, 0.07]	−0.07[−0.07, 0.06]	<0.001*
Difference of Entropy	3.53[3.46,3.58]	3.58[3.53,3.63]	<0.001*
Difference of Mean	2.36[2.23,2.51]	2.49[2.35,2.60]	<0.001*
Energy	0.03[0.02,0.03]	0.02[0.02,0.03]	<0.001*
Entropy	3.40[3.21,3.60]	3.61[3.43,3.76]	<0.001*
Homogeneity	0.44[0.42,0.46]	0.42[0.41,0.44]	<0.001*
Inverse Difference of Moment	0.36[0.35,0.39]	0.35[0.33,0.36]	<0.001*
Maximum of Probability	0.06[0.06,0.08]	0.06[0.05,0.06]	<0.001*
Inertia	8.50[8.46,8.56]	8.50[8.46,8.55]	0.73
Sum of Entropy	3.84[3.75,3.92]	3.92[3.86,3.99]	<0.001*
Sum of Mean	15.00[14.92,15.13]	15.01 [14.92,15.09]	0.619
Variance	77.90[77.03,78.89]	78.20 [77.26,79.01]	0.205

* $P < 0.05$ after Bonferroni adjust.

3.3. Differentiating COVID-19 from other infectious pneumonia

The data were divided into 5 folds with 4 folds for model training and 1-fold for testing alternately. To identify COVID-19 images and individuals from other infectious pneumonias, the RF classifier was built using texture features. The classification performance was validated in 5 test sets successively, and the results are presented in Table 3. The out-of-bag error was used to evaluate the RF model in the training set. The mean values of accuracy, sensitivity, and specificity of the 5-fold test sets were 0.72, 0.77, and 0.68 at the image level and 0.83, 0.81, and 0.84 at the individual level, respectively. The AUROC values were all above 0.75 for image classification and above 0.80 for individual identification in the test sets, as shown in Fig. 3. The RF model was also applied to the whole dataset to illustrate the relative importance of each texture feature. As shown in Fig. 4, the texture feature 'Correlation' ranked first for individual recognition compared to the clinical factors (age, sex, fever, cough, white blood cell count, neutrophil percentage, lymphocyte percentage, procalcitonin, and C-reactive protein). In addition, 'Correlation' was poorly correlated with these clinical factors, as shown in Table A.2.

3.4. Differentiating COVID-19 from influenza A/B pneumonia

Differential texture features were also used to differentiate COVID-19 from influenza A/B pneumonia at the image level and individual level. As with the first analysis with other infectious pneumonias, the mean AUROC, accuracy, sensitivity, and specificity values of all 5-fold test sets were 0.883, 0.796, 0.883, and 0.657 for image classification and 0.957, 0.841, 0.863, and 0.800 for individual identification, as shown in Table A.3 and Fig. A.2.

4. Discussion

In this study, we investigated the CT texture features of COVID-19 and proposed a radiomic feature-based RF classifier to assist in making differential diagnoses at both the image and individual levels. We found that COVID-19 patients showed similar clinical manifestations as other acute pneumonia patients, while a specific pattern of CT texture features was attributed to COVID-19. Our data show that texture features may override clinical signs regarding the recognition of COVID-19 and could differentiate it from not only influenza A/B but also other acute infectious pneumonias.

The novel coronavirus pandemic is circulating rapidly worldwide and has caused serious worldwide transmission and deaths. Control of the current situation largely depends on the early and accurate diagnosis of infected patients. In this study, we found that white blood cell count, neutrophil percentage, C-reactive protein, and lymphocyte percentage were significantly different between the COVID-19 group and the other infectious pneumonia groups. The clinical symptoms are similar, apart from the cough incidence being relatively lower in COVID-19. These results are consistent with previous studies [15], and they indicate the difficulty in making differential diagnoses only based on this information. In addition, real-time RT-PCR testing does not completely rule out SARS-CoV-2 infection, and CT could supplement diagnostic information, especially when RT-PCR testing is negative in subjects who are highly suspected to have COVID-19 [19].

Most COVID-19 patients present pulmonary pathological changes at an early stage [20]. However, it is difficult to differentiate COVID-19 from other infectious pneumonias, as the CT findings of these infectious pneumonias always overlap with each other, and the common radiographic features on chest CT are not sensitive enough to be captured by the naked eye [21,22]. In contrast, mathematical texture features can be easily utilized by machine learning algorithms. Radiomics is an emerging field of research attempting to explore the practical and clinical value of medical images from a quantitative perspective. To date, it has been applied to discover potential imaging biomarkers

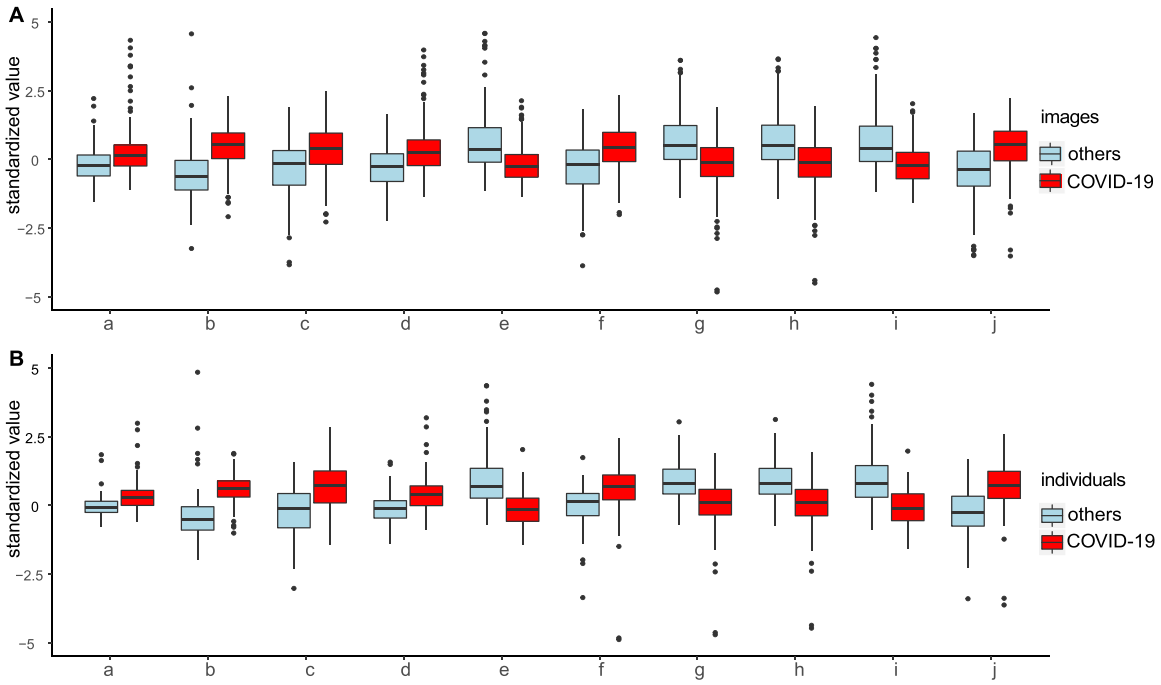

Fig. 2. Distribution of texture features in the images and individuals between COVID-19 and other infectious pneumonias.
Legend:
A: Standardized values of the significantly different features in the images between COVID-19 cases and others.
B: Standardized values of the significantly different features in the individuals between COVID-19 cases and others.
Abbreviations:
a: Contrast; b: Correlation; c: Difference of Entropy; d: Difference of Mean; e: Energy; f: Entropy; g: Homogeneity.
h: Inverse Difference of Moment; i: Maximum of Probability; j: Sum of Entropy.

Table 3
Classification results at the image level and the individual level using 5-fold cross validation.

	Training set	Testing set						
	OBB error (%)	TP	FP	TN	FN	Accuracy	sensitivity	specificity
Images								
fold 1	24.34	40	19	45	10	0.746	0.800	0.703
fold 2	22.08	46	17	43	17	0.724	0.730	0.717
fold 3	22.81	46	18	38	12	0.737	0.793	0.679
fold 4	19.96	39	14	27	16	0.688	0.709	0.659
fold 5	25.50	40	25	45	9	0.714	0.816	0.643
mean	23.18	–	–	–	–	0.722	0.770	0.680
Individuals								
fold 1	21.05	13	2	21	3	0.872	0.813	0.913
fold 2	19.08	16	4	15	4	0.795	0.800	0.789
fold 3	21.74	16	3	16	4	0.821	0.800	0.842
fold 4	18.42	19	2	13	5	0.821	0.792	0.867
fold 5	19.74	16	4	16	3	0.821	0.842	0.800
mean	20.14	–	–	–	–	0.826	0.809	0.842

OBB: out of bag; TP: number of true positives; FP: number of false positives; TN: number of true negatives; FN: number of false negatives.

associated with different diseases, including progression and prognosis [23–25]. In this study, we proposed using the NSDTCT-based GLCM algorithm to analyse the features of CT images from 96 different directions, which could capture more information than only one original level, as discussed previously [26]. Similarly, we found that some texture features are positively or negatively associated with COVID-19, which implies that the specific texture pattern could be captured and used to recognize COVID-19 images and individuals. In this study, we found that ten of 14 CT texture features were significantly associated with COVID-19, and we established an RF classifier with these radiomic features to differentiate COVID-19 from other infectious pneumonias. The RF classifier achieved optimal performance at both the image level (AUROC: 0.800; accuracy: 0.722; sensitivity: 0.770; specificity: 0.680) and the individual level (AUROC: 0.858; accuracy: 0.826; sensitivity: 0.809; specificity: 0.842). The performance was better when used to differentiate COVID-19 from influenza A/B.

In addition, our data show that the feature ‘Correlation’ contributes most to the classification of COVID-19, even when compared with clinical factors (age, sex, fever, cough, white blood cell count, neutrophil percentage, lymphocyte percentage, procalcitonin, and C-reactive protein). ‘Correlation’ is poorly correlated with these clinical factors, which means that texture features could be a potential and irreplaceable radiomic biomarker for COVID-19. From a statistical perspective, ‘Correlation’ mainly reflects the mathematical distance of the image pixels, and it increases in the CT image of COVID-19. This finding is probably because the pulmonary lesions of COVID-19 individuals are more

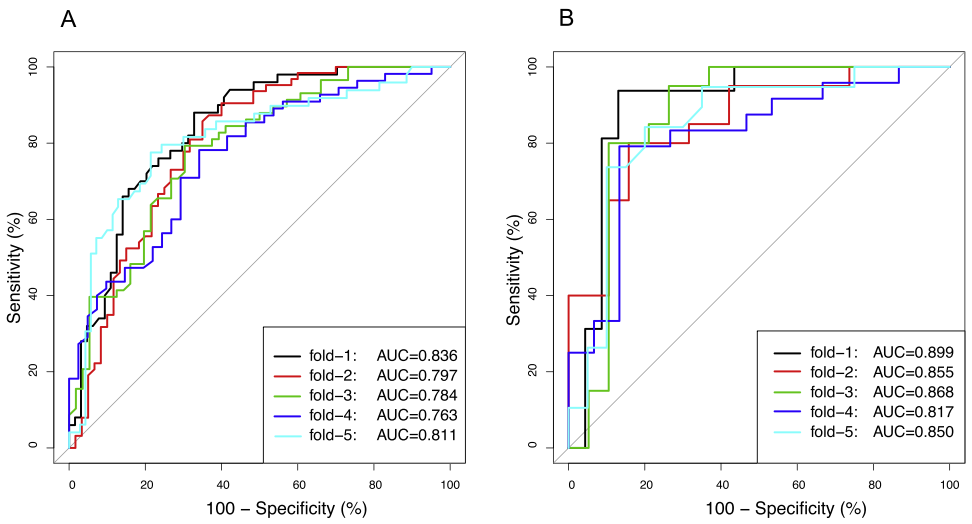

Fig. 3. The ROC plots of texture features using 5-fold cross validation.
A: The ROC plots in the images of COVID-19 and other infectious pneumonias.
B: The ROC plots in the individuals with COVID-19 and other infectious pneumonias.

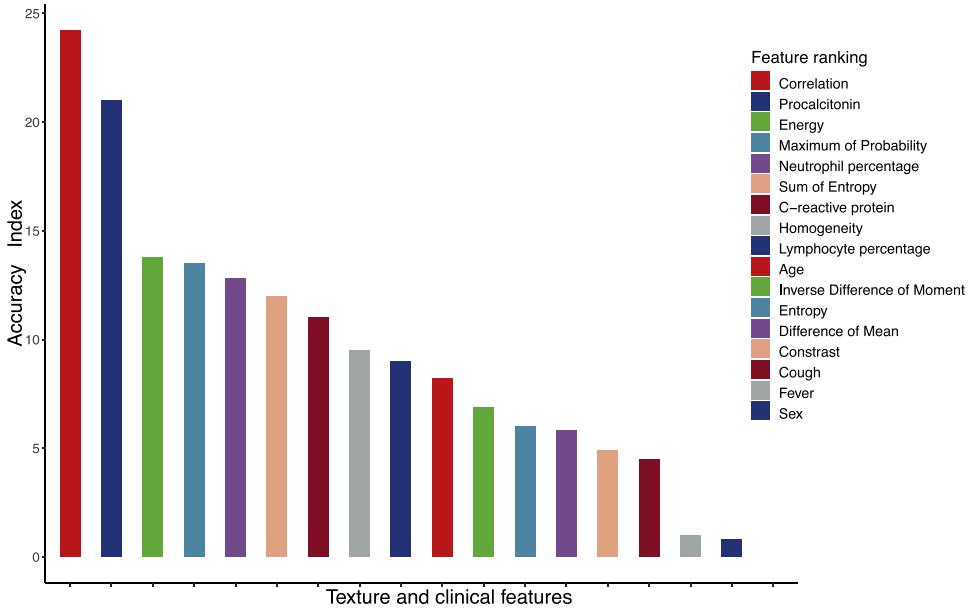

Fig. 4. Contribution of radiomic and clinical features in the classification task assessed by the accuracy criterion.

extensive and severe. Additionally, areas of GGO are found at a relatively early stage in COVID-19 individuals compared with individuals with other infectious pneumonias.

A strength of our study was the developed classification tool for differentiating COVID-19 from other infectious pneumonia and influenza A/B both on the image level and individual level. Additionally, we discovered a radiomic feature 'Correlation' strongly associated with COVID-19. There were several limitations of our study. First, the sample size was relatively small. The machine learning model and AI algorithm should be data driven, and the accuracy largely depends on the sample size. The proposed assistive diagnosis model needs further validation in a larger population. Second, this research was a retrospective cross-sectional study, and we could not clarify the influence of specific texture patterns on the subsequent outcome or disease prognosis.

Our study shows that a specific pattern of CT texture features contributes to COVID-19. The proposed texture feature-based RF classifier, as an assistive classification tool, could be used to rapidly differentiate COVID-19 from other infectious pneumonia and influenza A/B in the present pandemic.

Funding

This work was supported by the National Natural Science Fund of China (No. 81530087 and No. 81773542) and the Natural Science Fund of Beijing Municipal Education Commission (No. KZ201810025031). The study sponsors had no involvement in this research.

CRediT authorship contribution statement

Zhiyuan Wu: Writing - original draft, Methodology. **Li Li:** Writing - original draft. **Ronghua Jin:** Investigation. **Lianchun Liang:** Data curation. **Zhongjie Hu:** Data curation. **Lixin Tao:** Validation. **Yong Han:** Validation. **Wei Feng:** Software. **Di Zhou:** Software. **Weiming Li:** Software. **Qinbin Lu:** Writing - review & editing. **Wei Liu:** Writing -

review & editing. **Liqun Fang:** Formal analysis. **Jian Huang:** Formal analysis. **Yu Gu:** Writing - review & editing. **Hongjun Li:** Supervision. **Xiuhua Guo:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2021.109602.

References

[1] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. Mcguire, D. Veesler, Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2) (2020).

[2] J. Machhi, J. Herskovitz, A.M. Senan, D. Dutta, B. Nath, M.D. Oleynikov, W. R. Blomberg, D.D. Meigs, M. Hasan, M. Patel, P. Kline, R.C.-C. Chang, L. Chang, H. E. Gendelman, B.D. Kevadiya, The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections, J. Neuroimmune Pharmacol. 15 (3) (2020) 359–386.

[3] J.W. Tang, P.A. Tambyah, D.S.C. Hui, Emergence of a novel coronavirus causing respiratory illness from Wuhan, China, J. Infect. 80 (3) (2020) 350–371.

[4] WHO (World Health Organization), WHO Characterizes COVID-19 Outbreak as Pandemic, 2020 (Accessed 13 November 2020), https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020.

[5] J. Chen, Pathogenicity and transmissibility of 2019-nCoV—A quick overview and comparison with other emerging viruses, Microbes Infect. 22 (2) (2020) 69–71.

[6] J.F.-W. Chan, S. Yuan, K.-H. Kok, K.K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C.-Y. Yip, R.W.-S. Poon, H.-W. Tsoi, S.K.-F. Lo, K.-H. Chan, V.K.-M. Poon, W.-M. Chan, J.D. Ip, J.-P. Cai, V.C.-C. Cheng, H. Chen, C.K.-M. Hui, K.-Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet 395 (10223) (2020) 514–523.

[7] C. Huang, Y. Wang, X. Li, L. Ren, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (10223) (2020).

[8] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing, Radiology 296 (2) (2020), 200343.

[9] World Health Organization, Use of Chest Imaging in COVID-19, 2021 (Accessed 13 November 2020), https://www.who.int/publications/i/item/use-of-chest-imaging-in-covid-19.

[10] M. Chung, A. Bernheim, X. Mei, N. Zhang, H. Shan, CT imaging features of 2019 novel coronavirus (2019-nCoV), Radiology 295 (1) (2020), 200230.

[11] R. Yuan, S. Shi, J. Chen, G. Cheng, Radiomics in RayPlus: a web-based tool for texture analysis in medical images, J. Digit. Imaging 32 (2) (2019) 269–275.

[12] S. Srisajjakul, P. Prapaisilp, S. Bangchokdee, CT and MR features that can help to differentiate between focal chronic pancreatitis and pancreatic cancer, Radiol. Med. 125 (12) (2020).

[13] F. Zhang, S. Tian, S. Chen, Y. Ma, X. Guo, Voxel-based morphometry: improving the diagnosis of Alzheimer's disease based on an extreme learning machine method from the ADNI cohort, Neuroence 414 (2019).

[14] H. Yasar, M. Ceylan, A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods, Multimed. Tools Appl. (2020) 1–25.

[15] Q.i. Zeng, K.I. Zheng, J. Chen, Z. Jiang, M. Zheng, Radiomics-based Model for Accurately Distinguishing Between Severe Acute Respiratory Syndrome Associated Coronavirus 2 (SARS✕oV and Influenza a Infected Pneumonia, MedComm, 2020.

[16] World Health O, Clinical Management of Severe Acute Respiratory Infection When Novel Coronavirus (2019-nCoV) Infection is Suspected: Interim Guidance, 28 January 2020, World Health Organization, Geneva, 2020, 2020.

[17] B. Zhang, X. He, F. Ouyang, D. Gu, Y. Dong, L. Zhang, X. Mo, W. Huang, J. Tian, S. Zhang, Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma, Cancer Lett. (2017) 21.

[18] H. Albarakati, H. Saigo, R. Newman, B.K.C. Dukka, RF-GlutarySite: a random forest based predictor for glutarylation sites, Mol. Omics 15 (3) (2019) 189–204.

[19] W. Wu, H. Shi, B. Liang, F. Wu, D. Xiang, Z. Qin, H. Zhou, C. Zheng, L. Wang, Radiological and Clinical Characteristics of 22 Suspected COVID-19 Patients With Multiple Negative RT-PCR Testing, 2020.

[20] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, Ja. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 395 (10223) (2020) 507–513.

[21] J.P. Kanne, Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist, Radiology 295 (1) (2020), 200241.

[22] J. Chen, Pathogenicity and transmissibility of 2019-nCoV—A quick overview and comparison with other emerging viruses, Microbes Infect. 22 (2) (2020) 69–71.

[23] Y. Kawashima, A. Fujita, K. Buch, B. Li, O. Sakai, Using texture analysis of head CT images to differentiate osteoporosis from normal bone density, Eur. J. Radiol. 116 (2019).

[24] Liu Zhenyu, Wang Shuo, Dong Di, Wei Jingwei, Fang Cheng, The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges, Theranostics 9 (5) (2019) 1303–1322.

[25] L. Abdel-Hamid, Glaucoma detection from retinal images using statistical and textural wavelet features, J. Digit. Imaging 33 (2019) 151–158.

[26] Y. Han, Y. Ma, Z. Wu, F. Zhang, X. Guo, Histologic subtype classification of non-small cell lung cancer using PET/CT images, Eur. J. Nucl. Med. Mol. Imaging (2) (2020).