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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Automatic pulmonary auscultation grading diagnosis of Coronavirus Disease 2019 in China with artificial intelligence algorithms: a cohort study

Hongling Zhu, Jinsheng Lai, Bingqiang Liu, Ziyuan Wen, Yulong Xiong, Honglin Li, Yuhua Zhou, Quyun Fu, Guoyi Yu, Xiaoxiang Yan, Xiaoyun Yang, Jianmin Zhang, Chao Wang, Hesong Zeng

PII: S0169-2607(21)00574-5
DOI: https://doi.org/10.1016/j.cmpb.2021.106500
Reference: COMM 106500

To appear in: Computer Methods and Programs in Biomedicine

Received date: 5 April 2021
Accepted date: 20 October 2021

Please cite this article as: Hongling Zhu, Jinsheng Lai, Bingqiang Liu, Ziyuan Wen, Yulong Xiong, Honglin Li, Yuhua Zhou, Quyun Fu, Guoyi Yu, Xiaoxiang Yan, Xiaoyun Yang, Jianmin Zhang, Chao Wang, Hesong Zeng, Automatic pulmonary auscultation grading diagnosis of Coronavirus Disease 2019 in China with artificial intelligence algorithms: a cohort study, Computer Methods and Programs in Biomedicine (2021), doi: https://doi.org/10.1016/j.cmpb.2021.106500

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.
Highlights

- To the best of our knowledge, it is the first study to investigate the auscultation characteristics of covid-19 with different severities and by analyzing auscultation data with deep learning algorithms.
- Three new CNN models in classifying Covid-19 patients based on pulmonary auscultation for two-classification and four-classification tasks achieved superior performance.
- The unidimensional raw data of lung sound is used to perform classification, which preventing complex pre-processing and manual feature extraction.
Automatic pulmonary auscultation grading diagnosis of Coronavirus Disease 2019 in China with artificial intelligence algorithms: a cohort study

Hear 2019-nCov

Hongling Zhu*, Jinsheng Lai*, Bingqiang Liu, Ziyuan Wen, Yulong Xiong, Honglin Li, Yuhua Zhou, Qiuyun Fu, Guoyi Yu, Xiaoxiang Yan, Xiaoyun Yang, Jianmin Zhang, Chao Wang, Hesong Zeng

*Contributed equally

Division of Cardiology, Department of Internal Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, P.R. China
(H Zhu MD PhD, J Lai MD PhD, H Li BS, X Yang MD PhD, Prof H Zeng MD PhD)

School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China; Wuhan National Laboratory of Optoelectronics, Wuhan, Hubei 430074, China
(B Liu, Z Wen BE, Q Fu PhD, C Yu PhD, Prof C Wang PhD)

School of Artificial Intelligence, Jianghan University, Wuhan, Hubei 430056, China
Prof J Zhang PhD

Department of Cardiology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, 197 Rui Jin Road II, Shanghai, 200025, China
Prof X Yan MD PhD, Y Zhou MD PhD

Correspondence to:
Prof. Hesong Zeng, Division of Cardiology, Department of Internal Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, P.R. China
zenghs@tjh.tjmu.edu.cn

or

Prof. Chao Wang, 1. School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China 2. Wuhan National Laboratory of Optoelectronics, Wuhan, Hubei 430074, China
chao_wang_me@hust.edu.cn
Abstracts

Background and Objective

Research on automatic auscultation diagnosis of COVID-19 has not yet been developed. We therefore aimed to engineer a deep learning approach for the automated grading diagnosis of COVID-19 by pulmonary auscultation analysis.

Methods

172 confirmed cases of COVID-19 in Tongji Hospital were divided into moderate, severe and critical group. Pulmonary auscultation were recorded in 6-10 sites per patient through 3M littmann stethoscope and the data were transferred to computer to construct the dataset. Convolutional neural network (CNN) were designed to generate classifications of the auscultation. F1 score, the area under the curve (AUC) of the receiver operating characteristic curve, sensitivity and specificity were quantified. Another 45 normal patients were served as control group.

Results

There are about 56.52%, 59.46% and 78.85% abnormal auscultation in the moderate, severe and critical groups respectively. The model showed promising performance with an averaged F1 scores (0.9938 95% CI 0.9923–0.9952), AUC ROC score (0.9999 95% CI 0.9998–1.0000), sensitivity (0.9938 95% CI 0.9910–0.9965) and specificity (0.9979 95% CI 0.9970–0.9988) in identifying the COVID-19 patients among normal, moderate, severe and critical group. It is capable in identifying crackles, wheezes, phlegm sounds with an averaged F1 scores (0.9475 95% CI 0.9440–0.9508), AUC ROC score (0.9762 95% CI 0.9848–0.9865), sensitivity (0.9482 95% CI 0.9393–0.9578) and specificity (0.9835 95% CI 0.9806–0.9863).

Conclusions

Our model is accurate and efficient in automatically diagnosing COVID-19 according to different categories, laying a promising foundation for AI-enabled auscultation diagnosing systems for lung diseases in clinical applications.

Keywords

Auscultation, Coronavirus Disease 2019 (COVID-19), convolutional neural
Introduction

The 2019 novel coronavirus (2019-nCov) emerged December 2019, and caused a cluster of acute respiratory illness called Coronavirus Disease 2019 (COVID-19). As of Dec 3, 2020, more than 89,906 cases were confirmed and 4,642 cases died of it in China. While in the world, the confirmed cases increased sharply to more than 47007194 and death cases reached to 1208224, which induced huge damage to the people worldwide.

The 2019-nCov is a kind of coronavirus infecting humans through the angiotension converting enzyme (ACE) receptor, inducing organ dysfunction including lungs, heart and kidneys. The main manifestation of the patient varied from fever, cough, dyspnea and vomiting and so on, accompanied with leukocytopenia and lymphocytopenia. The virus can be transmitted and spread in humans at all ages and genders through close contact and droplets, even high concentration of aerosol. Thus, it is a highly infectious and dangerous disease. According to the severity of COVID-19, the Chinese Center for Disease Control and Prevention (CDC) divided the patients into four types called light, moderate, severe and critical. As there is no effective drug or vaccine available thus far, the severity level of the patients is essential information for the treatment of COVID-19. However, the main method to decide the severity level of COVID-19 patients are syndromes, computed tomography (CT) and blood testing of inflammation markers. These testing methods are radioactive, expensive or invasively. There is no efficient, cost effective and reproducible method yet.

Mosby’s Medical and Nursing Dictionary defines the physical examination as ‘An investigation of the body to determine its state of health using any or all of the techniques of inspection, palpation, percussion, auscultation, and olfaction’. Specifically, auscultation is the process of listening to the internal sounds of human body through a stethoscope and it is an effective and widely-used tool to diagnose
especially the lung diseases and abnormalities. Through the stethoscope, physicians may hear various abnormal lung sounds including wheezes, crackles, squawk, rhonchus and phlegm sounds as well as the normal lung sounds according to the patients’ diseases. Thus the auscultation is an essential, but simple and patient-friendly method for testing the pneumonia conditions. While in COVID-19, in order to prevent from infecting with this virus, we have updated protective equipment that make it difficult for the doctors to do auscultation in the patients. What’s more, as the quality of auscultation relies much on the environment around and the diagnostic value relies on the physicians’ experience which are prone to inherent subjectivity, physicians’ manual auscultation usually leads to a low value of auscultation in lung disease diagnosing and testing.

This study is the first to investigate the auscultation characteristics of COVID-19 with different severities and by analyzing auscultation data with deep learning algorithms to offer a practical, high accuracy, cost effective, and comprehensive automatic auscultation diagnosis framework, which can be used as a reliable tool for the diagnosis and prediction for not only COVID-19, but also various pathological respiratory conditions.

Methods

Study design and participants

172 confirmed cases of COVID-19 treated in Tongji Hospital of Huazhong University of Science and Technology, Wuhan, China from Mar 31 to Apr 5, 2020 were included in this study. Cases were confirmed by next generation sequencing or real-time PCR methods or according to the clinical diagnosis criteria including CT scans, symptoms and laboratory results. And the patients were divided into moderate, severe and critical group according to Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial Version 7) in China (Supplementary Table 1). Outcomes were followed up until Apr 10, 2020. As mild case means the clinical symptoms were mild, and there was no sign of pneumonia on imaging, we did not
include this type in our study. Another 45 normal patients with normal chest CT scanning, Echocardiography and Electrocardiography were included as control group.

This study complied with the edicts of the 1975 Declaration of Helsinki\textsuperscript{14} and was approved by the Ethical Committee of Tongji Hospital with Institutional Review Board Approval (IRB) number of TJ-C20201202. Written informed consent was obtained from patients or patients’ next of kin.

**Clinical Data Collection**

Epidemiological, clinical, laboratory and radiological features were extracted from electronic medical records in Tongji Hospital. Throat-swab specimens were obtained after clinical remission of symptoms such as fever, cough, and dyspnea, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA was detected. Routine laboratory examinations consisted of complete blood count (Sysmex XN-2000 and its original reagent, Kobe, Japan), liver and renal function tests (Roche cobas 8000 and its original reagent, Basel, Switzerland), myocardial enzymes, and inflammatory cytokines like hsCRP (\(< 1 \text{ mg/L low risk}, 1-3 \text{ mg/L moderate risk}, > 3 \text{ mg/L high risk}, > 10 \text{ mg/L infection and inflammation}\) (Roche cobas e602 and its original reagent, Basel, Switzerland). Non-contrast CT scanning (GE Healthcare, Philips, or Toshiba Medical Systems) of the thorax were performed in almost all patients in the supine position during end-inspiration.

**Pulmonary Auscultation**
The pulmonary auscultation was recorded in 10 sites per patient with 30 seconds for each site according to the Diagnosing guidelines through 3M littmann stethoscope (Model 3200, 3M Health Care, Saint Paul Minnesota, America). The 10 sites are showed as following: 1. The left middle pulmonary, 2. The left bottom pulmonary, 3. The right middle pulmonary, 4. The right bottom pulmonary, 5. The left Midaxillary pulmonary, 6. The right Midaxillary pulmonary, 7. The left middle pulmonary on the back, 8. The left bottom pulmonary on the back, 9. The right middle pulmonary on the back, 10. The right bottom pulmonary on the back (Figure 1). To the patients who were endotracheal intubated and could not sit up or turn over, the first six sites were recorded. The 10 sites were employed in the normal group as well, and all 10 sites per normal patients were served as the normal control dataset with mixed auscultation.

**Figure 1: Auscultation site distribution model.**

**Annotation procedure/Dataset preparation and usage**

The auscultation data were transferred into digital WAVE files to the Littmann StethAssist™ software (1.3.230) via the 3M littman stethoscope built-in Bluetooth transceiver to construct the dataset. The auscultation was diagnosed by a committee consisting of two independent doctors, who are board-certified actively-practicing doctors majored in cardiology and respirology in Tongji Hospital. Doctor committees firstly annotated auscultation records independently, then discussed the records that they did not reach an agreement. After discussing comprehensively, all the auscultation records were annotated by consensus, providing an expert standard for artificial intelligence (AI) model evaluation.

**Overview of the deep learning AI models**
Three convolutional neural network (CNN) models were developed for lung sound classification, as three classifications of lung sounds were realized to study the automatic pulmonary auscultation hierarchical diagnosis of COVID-19 patients by artificial intelligence methods. The raw records of lung sounds were segmented into audio clips with a duration of 4 seconds. The input of the three CNN models is a 16000×1 vector as the sampling rate of lung sound is 4 kHz/s. One of the CNN models has a 1×2 vector output for two-category classification, while the other two models have a 1×4 vector output for four-category classification. More specifically, the CNN model with an output of 1×2 vector was developed to classify normal and abnormal lung sounds. For the other two CNN models with an output of 1×4 vector, each element of the output vector corresponds to one of four specific lung sound classes or one of four COVID-19 patient conditions, depending on the training and classification purpose. In our study, the two four-category classification models were trained to classify four COVID-19 patient conditions including normal, moderate, severe, and critical cases, and four major lung sounds including normal, crackles, wheezes, and phlegm sounds, respectively.

To classify lung sounds collected from COVID-19 patients and normal subjects into two categories or four categories, lightweight CNN model and residual CNN models were developed respectively. Actually, for the lung sound signals, the lightweight CNN models and residual networks are used for classification, and achieved good results. For the two-classification of normal and abnormal lung sounds, the using of depthwise separable convolution not only makes the network more lightweight in terms of multiply-add operations and weight parameters, but also achieves better performance compared with the model without depthwise separable convolution. For the four major lung sounds, the residual structure can effectively prevent gradient exploding or vanishing by reducing the backpropagated error, and therefore the performance of the four-classification CNN model can be significantly improved by increasing the depth of the network. Experiments were also carried out on neural network models with and without depthwise separable convolution/residual structure. It is confirmed that lightweight CNN models and residual networks perform
well for lung sound signals.

For these CNN models, rectified linear unit was used to prevent vanishing gradient and dropout was used to prevent overfitting. We chose Adam for stochastic optimization and used cross-entropy as the loss function. In the output layers, the softmax function was used to calculate the probabilities of input test sample corresponding to two/four categories respectively, and the one with the largest probability is chosen as the final classification result. The structures of the deep CNN models and related information including the layer type, kernel stride, filter shape, and input size are shown in appendix 1 (Supplementary Figure 1, 2 and Table2, 3).

**Statistical analysis**

We evaluated the performance of the CNN models by some criteria, such as the prediction accuracy, the area under the curve (AUC) of the receiver operating characteristic (ROC), F1 score (harmonic mean of the precision and sensitivity), sensitivity, and specificity, with two-sided 95% CI. Confusion matrices were also used to evaluate whether the predictions of the CNN models were consistent with the labelled results from committee consensus. The CNN models can be assessed comprehensively through these different criteria.

For the prediction accuracy, we evaluated the ratio between the predicted result $\hat{Y}_s$ and the labelled result $Y_s$, according to the following expression:

$$\text{Accuracy} = \frac{1}{S} \sum_{s=1}^{S} 1[\hat{Y}_s = Y_s]$$  \hspace{1cm} (1)

where $S$ is the number of the lung sound segments to be assessed.

We also calculated the macro sensitivity, specificity, precision, F1 score, by using the classification results including true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates. The formulas are defined by the following equations\textsuperscript{15,19}:

$$\text{sensitivity} = \frac{TP}{TP + FN} \hspace{1cm} (2)$$
specificity = \frac{TN}{TN+FP} \quad (3)
precision = \frac{TP}{TP+FP} \quad (4)
F1 \text{ score} = \frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (5)

We used the macro F1 score to evaluate the performance with all classes according to the expression:

\text{mean F1 score} = \frac{1}{N} \sum_{n=1}^{N} F1_n \quad (6)

where N is the number of the classes.

In terms of clinical characteristics, all statistical analyses were performed using SPSS (Statistical Package for the Social Sciences) version 21.0 software (SPSS Inc). Categorical variables were described as frequency rates and percentages, and continuous variables were described with mean± standard deviation (SD) when the data were normally distributed, otherwise median and inter quartile range (IQR) values. Means for continuous variables were compared using independent group t tests when the data were normally distributed, otherwise, the Mann-Whitney test was used. Proportions for categorical variables were compared using the χ² test. For unadjusted comparisons, a 2-sided α of less than 0.05 was considered statistically significant.

Results

Presenting Characteristics of Clinical features and laboratory parameters in moderate, severe and critical groups of COVID-19

The mean age of the 172 COVID-19 patients (50% male) is 64.9±14.9 and the BMI is 23.2 (20.2, 25.1) kg/m². There was no significant difference in age and BMI among the three groups, but a significantly increased male ratio in severe (n=74, 48.6% male) and critical group (n=52, 63.5% male) compared to that in moderate group (n=46, 37% male). 2019-nCov was detected in 138 (79.7%) of the totally 172 patients
by real-time PCR, of which 40 (87%), 57 (77%) and 40 (76.9%) cases confirmed in moderate,

severe and critical groups respectively. The top three comorbidities included hypertension (73, 42.4%), diabetes (39, 22.7%) and coronary heart diseases (CHD, 21, 12.2%), and patients treated in the critical groups (51.9%, 30.8%, 15.4%), compared with patients in the moderate group (39.1%, 17.4%, 6.5%) and severe group (37.8%, 20.3%, 13.5%), were more likely to have underlying comorbidities. The most common symptoms at onset of illness were fever (133, 77.3%), cough (120, 69.8%), and dyspnea (115, 66.9%). And there were significantly increased cough and dyspnea in severe group, and increased dyspnea in critical group, compared to that in the moderate group. The 2019-nCov patients showed a degree of heart and kidney dysfunctions and inflammation reactions and increased IgM and IgG of 2019-nCov. Specifically, the critical group showed a significantly increased White blood cell (11.2±7.3), Neutrophil (11.2±13.2), hsCRP (70.6±66.6), IL-6 (50.5±56.8), Myoglobin (236.1±322.7), NT-proBNP (2660.0±4354.1), Aspartate aminotransferase (45.3±43.3), Blood urea nitrogen (8.2±5.3) and a decreased Lymphocyte (0.9±0.5) compared to that of the moderate group. As of Apr 10, 135 (78.5%) patients were discharged and 4 died (overall mortality, 2.3%), and the remaining patients were still hospitalized. (Please see detailed information in Supplementary Table 4)

| Total  | Moderate | Severe | Critical |
|--------|----------|--------|----------|
| (N=172)| (N=46)   | (N=74) | (N=52)   |
| Demographic and clinical characteristics | | | |
| Age (mean±SD) | 64.9±14.9 | 61.4±15.5 | 67.7±14.5 | 66.9±14.5 |
| Sex, male(%) | 88 (50.0) | 17 (37.0) | 36 (48.6) * | 33 (63.5) * |
| BMI (kg/m2) | 23.2 | 21.5 | 23.8 | 23.5 |
| median (IQR) | (20.2, 25.1) | (19.7, 23.6) | (21.5, 26.0) | (20.3, 26.8) |
| Comorbidity | | | |
| Hypertension(%) | 73 (42.4) | 18 (39.1) | 28 (37.8) | 27 (51.9) |
| Diabetes(%) | 39 (22.7) | 8 (17.4) | 15 (20.3) | 16 (30.8) |
| Coronary heart diseases (%) | 21 (12.2) | 3 (6.5) | 10 (13.5) | 8 (15.4) |
| Cerebrovascular | 20 (11.6) | 2 (4.4) | 8 (10.8) * | 10 (19.2) * |
| Symptom                  | Fever(%) | Cough(%) | Dyspnee(%) | Outcomes |
|--------------------------|----------|----------|------------|----------|
|                          | 133 (77.3) | 120 (68.8) | 115 (66.9) | discharges (%) |
|                          | 33 (71.7)  | 27 (58.7)  | 18 (39.1)  | 135 (78.5) |
|                          | 59 (79.7)  | 59 (79.7)  | 61 (82.4)  | 43 (93.5)  |
|                          | 41 (78.9)  | 34 (65.4)  | 36 (69.2)  | 68 (91.9)  |
|                          |           |           |            | 24 (46.2)  |
|                          |           |           |            |           |

| Outcomes                | Discharged (%) | Still hospitalized (%) | died(%) |
|-------------------------|----------------|------------------------|--------|
|                         | 135 (78.5)     | 32 (18.6)              | 4 (2.3) |
|                         | 43 (93.5)      | 3 (6.5)                | 0 (0)  |
|                         | 68 (91.9)      | 4 (5.4)                | 2 (2.7) |
|                         | 24 (46.2)      | 25 (48.1)              | 2 (3.9) |

| Laboratory findings     | IgM of 2019-nCov(≤10AU/ml) | IgG of 2019-nCov(≤10AU/ml) | White blood cell count(3.5-9.5*10⁹/L) | Neutrophil count(1.0-6.3*10⁹/L) | Lymphocyte count(1.1-3.2*10⁹/L) | hsCRP (mg/l) | IL-6(<7pg/ml) | cTnI (<34.2pg/ml) | Myoglobin (<154.9ng/ml) | NT-proBNP (<241pg/ml) | HbA1C (4-8%) | Alanine aminotransferase (≤41IU/L) | Aspartate aminotransferase (≤40U/L) | Blood urea nitrogen (3.6-9.5mmol/l) | Creatinine() |
|-------------------------|---------------------------|---------------------------|--------------------------------------|---------------------------------|---------------------------------|-------------|----------------|------------------------|--------------------------|--------------------------|--------------|-------------------------------|---------------------------|-------------------------------|--------------|
|                         | 59.9±107.2                | 144.2±116.5               | 8.9±5.2                              | 7.6±8.2                        | 1.4±4.0                         | 57.1±65.7   | 32.8±44.4       | 276.9±1711.5          | 187.8±256.0             | 1908.5±6334.6          | 6.8±1.8      | 37.3±56.9                    | 39.5±49.0                | 6.5±4.9                      | 59-104 umol/l male, |
|                         | 75.8±148.6                | 142.9±94.7                | 6.9±3.3                              | 5.0±3.1                        | 1.4±0.9                         | 40.1±58.2   | 19.6±31.1       | 518.8±1280.3           | 78.2±195.2              | 518.8±1280.3            | 7.0±1.7      | 26.0±19.4                    | 29.8±17.8                | 5.2±3.1                      | 69.7±28.5     |
|                         | 52.2±70.7                 | 144.0±126.9               | 6.4±3.7                              | 6.8±3.7                        | 1.7±6.1                         | 58.2±67.8   | 27.8±38.6       | 2132.1±8703.4          | 117.5±204.7             | 2132.1±8703.4           | 6.8±2.1      | 40.8±67.9                    | 41.3±63.5                | 6.2±5.2                      | 67.3±290.3   |
|                         | 57.4±110.5                | 145.6±120.0               | 11.2±7.3                             | 11.2±13.2#                     | 0.9±0.5*                        | 70.6±66.6   | 50.5±58.6       | 2660.0±4354.1          | 236.1±322.7             | 2660.0±4354.1           | 6.5±1.5      | 42.2±61.2                    | 45.3±43.3*               | 8.2±5.3#                     | 84.2±56.1    |

**Table 1.** Presenting Characteristics of Clinical features and laboratory parameters in moderate, severe and critical groups of COVID-19 onset of hospital admission. Data are
showed as mean± SD, n (%), or n/N (%), where N is the total number of patients with available data. (* p<0.05 moderate vs severe/critical group, #p<0.05 severe vs critical group)

CT and pulmonary auscultation in moderate, severe and critical groups of COVID-19

There are totally 96.2% patients (n=50/52) in critical group that showed pathological changes in both right and left lung, higher than that in the moderate (n=39/45, 1 patient lack of CT) with 86.7% and the severe group (n=67/73, 1 patient lack of CT) with 91.8%. Specifically, the pathological changes are mainly divided into ground glass change, consolidation, fibre stripe shadow, patch shadow, combination, and Pleural effusion in combination. The critical group showed 5 (9.6%) ground glass change, lower than that in the moderate group with 17 (23.8%) or severe group with 16 (21.9%), however, it showed 23 (44.2%) as patch shadow, higher than that in the moderate group with 9 (20.0%) or severe group with 16 (21.9%). The three groups showed a high frequency in combination changes of CT scans, in which the critical group exhibited the most pleural effusion occurrence with a ratio of 63.6%, higher than that of 31.3% in moderate group and 41.7% in the severe group. Similar tendency was observed onset of the auscultation time. (Please see detailed information in Supplementary Table 5). There are about 56.52%, 59.46% and 78.85% abnormal auscultation in moderate, severe and critical group respectively. The acoustic characteristics of different types of lung sounds varied from each other in time duration, frequency range and sound description20 (Please see detailed information in Supplementary Table 6). In our study, the abnormal auscultation includes crackles, wheezes and phlegm sounds in every group (Please see detailed information in Supplementary Table 7). We also showed the representative CT scanning pictures compared with the auscultation spectrogram in different groups and different abnormal sounds of Covid-19. Thus for each auscultation site, we have related pulmonary site in the CT scanning to be corresponded, and the corresponded CT
scanning is showed on the left. (Please see detailed information in Supplementary Figure 3). We found that there were positive correlations between the sounds signals and CT scanning in diagnosing 2019-nCov diseases.

**Deep learning models in three classifications**

The original lung sound signals were segmented with a duration of 4 seconds. To handle the imbalance problem in the data set, we used data augmentation methods including noise addition, time shifting, etc., and segmented the signals with a duration of 4 seconds. For the two-classification of normal and abnormal lung sounds, there are 18,848 segments of normal lung sound and 14,745 segments of abnormal lung sound. For the classification of COVID-19 severity levels, the number of normal, moderate, severe and critical segments are 9890, 13,126, 20 and 8996, respectively. For the four major lung sounds, the number of normal, crackles, wheezes and phlegm sound segments are 6446, 5379, 2732 and 2714, respectively. We randomly scrambled the lung sound segments, and then divided the segment data into training, validation, and testing datasets with the ratio of 6:2:2. We used the testing dataset to calculate the accuracies of the CNN models. ROC curves and AUC were plotted to assess the model discrimination of each class in the three CNN models (Table 3 and Figure 2a, b). The confusion matrices were also used to illustrate the discordance between the CNN’s predictions and the labelled results from committee consensus. (Figure 3). Experiments were carried out on neural network models with and without depthwise separable convolution/residual structure. It is found that lightweight CNN models and residual networks perform well for audio signals. Supplementary Figure 1, 2 and Table 2, 3 showed detailed information of classification performance of the three CNN models.

For the classification of normal and abnormal lung sounds, the accuracy of the two-category CNN model was 98.73%. The mean AUC ROC score of the algorithm was 0.9994 (95% CI 0.9992–0.9995), with sensitivity of 0.9880 (95% CI 0.9840–0.9919) and specificity of 0.9899 (95% CI 0.9872–0.9926). Besides, the mean F1 score of the model was 0.9875. We can see that the CNN model can distinguish the
abnormal lung sound signals from the normal signals with a very high prediction accuracy. As for the classification of COVID-19 severity levels, the first four-category CNN model performed very well in general, which achieved a prediction accuracy of 99.18%. The mean AUC ROC score was 0.9999 (95% CI 0.9998-1.0000). The mean sensitivity was 0.9938 (95% CI 0.9910–0.9965), and the mean specificity was 0.9979 (95% CI 0.9970–0.9988). The mean F1 score was also very high which reached 0.9938. As for the classification for normal, crackles, wheezes and phlegm sounds, the accuracy of the second four-category CNN model was 95.17%. The mean AUC ROC score of the CNN model was 0.9762 (95% CI 0.9848–0.9865), with sensitivity 0.9482 (95% CI 0.9393–0.9578) and specificity 0.9835 (95% CI 0.9806–0.9863). Besides, the mean F1 score of the model was 0.9475.

| Table 2. Performance summary of the deep learning model in COVID-19 classification of normal and abnormal, four severities and four major lung sounds. AUC represents for area under the curve, ROC represents for receiver operating characteristic. | Model AUC ROC (95% CI) | Model sensitivity (95% CI) | Model specificity (95% CI) | Model F1 score (95% CI) |
| COVID-19 classification of normal and abnormal | 0.9994 (0.9992-0.9995) | 0.9880 (0.9840-0.9919) | 0.9899 (0.9872-0.9926) | 0.9875 (0.9859-0.9891) |
| Normal | 0.9968 (0.9965-0.9970) | 0.9630 (0.9623-0.9666) | 0.9876 (0.9832-0.9920) | 0.9441 (0.9399-0.9482) |
| Moderate | 0.9999 (0.9998-1.0000) | 0.9919 (0.9844-0.9944) | 0.9973 (0.9965-0.9981) | 0.9928 (0.9914-0.9943) |
| Severe | 0.9999 (0.9998-0.9999) | 0.9919 (0.9866-0.9972) | 0.9973 (0.9955-0.9990) | 0.9909 (0.9890-0.9928) |
| Critical | 0.9999 (0.9998-1.0000) | 0.9932 (0.9913-0.9952) | 0.9977 (0.9971-0.9984) | 0.9937 (0.9919-0.9956) |
| Mean | 0.9999 (0.9998-1.0000) | 0.9938 (0.9910-0.9965) | 0.9979 (0.9970-0.9988) | 0.9938 (0.9923-0.9952) |
| COVID-19 classification of four major lung sounds | Mean | 0.9762 (0.9848-0.9865) | 0.9482 (0.9393-0.9578) | 0.9835 (0.9806-0.9863) | 0.9475 (0.9440-0.9508) |
| Normal | 0.9968 (0.9965-0.9970) | 0.9630 (0.9493-0.9766) | 0.9876 (0.9832-0.9920) | 0.9441 (0.9399-0.9482) |
| Crackles | 0.9569 (0.9543-0.9594) | 0.8311 (0.8113-0.8548) | 0.9468 (0.9404-0.9532) | 0.8904 (0.8835-0.8973) |
| Wheezes | 0.9933 (0.9929-0.9937) | 0.9998 (0.9995-1.0000) | 0.9999 (0.9998-1.0000) | 0.9719 (0.9708-0.9729) |
| Phlegm sounds | 0.9957 (0.9955-0.9959) | 0.9990 (0.9973-1.0000) | 0.9997 (0.9991-1.0000) | 0.9835 (0.9822-0.9847) |
| Mean | 0.9762 (0.9848-0.9865) | 0.9482 (0.9393-0.9578) | 0.9835 (0.9806-0.9863) | 0.9475 (0.9440-0.9508) |
Specifically, the confusion matrix shows that it is more difficult to classify between normal class and crackle class. The main reason why the performance index is worse than the first two CNN models is that some of the crackle signals were wrongly classified into the normal signal category. Actually, crackles may sometimes occur in healthy subjects during a deep inspiration\textsuperscript{21}. This may explain why the prediction accuracy of the third CNN model is lower than that of the first two CNN models.
Figure 3. Confusion Matrix of deep learning model in identifying abnormal auscultation from normal ones (up), classification of COVID-19 with normal, moderate, severe and critical (middle) and classification of COVID-19 with normal, crackle, wheezing, and phlegm sounds (bottom).
To further evaluate the proposed model, we also used the International Conference on Biomedical Engineering and Health Informatics (ICBHI) 2017 scientific challenge respiratory sound database\textsuperscript{22} as an independent dataset to evaluate our models. For the classification of normal and abnormal respiratory sounds, the accuracy, sensitivity and specificity were 82.59\%, 97.10\% and 80.59\% respectively, while for classification of four kinds of different respiratory sounds, the accuracy, sensitivity and specificity were 91.59\%, 97.10\% and 90.83\% respectively. The test results are comparable with the state-of-the-art research\textsuperscript{19}.

Discussion

In our study, we found that the auscultation paradigm represented by end-to-end automatic deep learning showed high potential as a new approach with high efficiency to help auxiliary diagnosis of 2019-nCov with exact severity classifications, also revealed its pathophysiological condition in a new way compared with the radiological characteristics of traditional CT.

The stethoscope was one of the first medical diagnostic instrument in clinical ever used and the auscultation of the respiratory system is broadly utilized since it is costless, noninvasive and safe. However, correct distinguishing auscultation of lung sounds related to different lung diseases is an art that requires rigid practice and experience. Nowadays several investigations have aimed to identify respiratory sound signal using machine learning or deep learning which make an automatic analysis of lung sound is available\textsuperscript{10,15,16,19,21,23,24}. However, these studies are limited to common pulmonary diseases without any medical or physiological parameters of the patients\textsuperscript{25,26}, thus weakened the clinical influence and application of their AI models.

In our study, we have thoroughly distinguished the pathological and physiological parameters of the patients onset of hospital admission and auscultation, applying more information to research the lung sounds of the patients. However, more precise diagnosis of COVID-19 presents various challenges. Most crucially, 2019-nCov is a kind of coronavirus inducing serious infection and danger to people,
resulting in wide spreading and deaths in the world. Thus, in order to decrease the 
virus exposure and healthcare-associated infection, strict and updated protection is 
urgently needed for the doctors and nurses, which make it difficult to do stethoscope 
testing manually. Using a stethoscope for auscultation offers an opportunity for the 
clinical physicians and patients to establish a support relationship that goes beyond 
what can be established by conversation.\textsuperscript{27} It is a process that clinicians examine the 
patients in a caring but professional manner, and help patients to relieve a sense of 
xiety and even fear.\textsuperscript{28} This is crucially important in the COVID-19 treatment as all 
the COVID-19 patients lived in quarantine room of quarantine hospitals\textsuperscript{29}. In our 
study, the auscultation was recorded in 10 sites per patient with 30 seconds for each site, 
supplying a comprehensive information for the lung auscultation. The 
auscultation of the 172 COVID-19 patients and 45 normal patients were analyzed by 
three CNN models, that’s the two-category CNN model classified normal and 
abnormal lung sounds, the other two four-category CNN models classified four 
COVID-19 patient conditions including normal, moderate, severe, and critical cases, 
and four major lung sounds including normal, crackles, wheezes, and phlegm sounds, 
respectively. It exhibited that we make the most use of the auscultation data. The three 
different models showed high accuracy and precision in hierarchical diagnosing of 
COVID-19, laying a foundation of the clinical application of our models.

Specifically, based on the data collected and analysis conducted, no evidence 
was found to show there are significant differences in acoustic characteristics between 
COVID-19 pneumonia and other pneumonia. However, in our manuscript, we found 
that the auscultation is different in normal, moderate, severe and critical groups, 
which indicate that pulmonary auscultation can be valuable for classifying normal 
patients and 2019-nCov patients, and classifying the severity of 2019-nCov into 
moderate, severe and critical groups.

Furthermore, we have systematically analyzed the CT abnormalities in our study 
with the classifications of ground glass change, consolidation, patch shadow and 
pleural effusion, which is closely related with pathophysiological condition of the
lungs and thus, leading to different kinds of auscultation abnormalities, including crackles, wheezes and phlegm sounds. We also showed representative CT scanning pictures compared with the auscultation spectrogram in different groups and different abnormal sounds of COVID-19 in order to detect the relationship of CT and auscultation in a more intuitive way. We found that there were positive correlations between the sounds signals and CT scanning in diagnosing 2019-nCov diseases. However, CT showed limited resolution in identifying the density of the shadow, leading to limited words as ground glass, patch, shadow, and it is hardly to distinguish the exact difference among them and subtle changes among several CTs tested in different times of one patient. Besides, CT showed limited efficacy in classifying some lung diseases such as bronchospasm. Auscultation can be divided into dozens of abnormalities, and the data can be obtained frequently, serving as a more detailed and real-time marker compared with CTs. The auscultation by stethoscope is easy to be accomplished in clinical work. Being a completely safe and noninvasive technique, it has advantages in terms of availability, costless, comfort, and diagnostic potential. Unlike CT and X-rays emitting ionizing radiation, it is safe for all patients; unlike MRI, it is available for those with metal implant and pacemakers. It is also accessible, available and portal equipment whenever and wherever the patients need medical check for lung condition, without taking off the ventilator and going to the specific CT rooms. In our study, we found that the patients in the critical group were too sick that they were not tolerable to be tested with CT in the specific CT rooms but were surrendered to X rays testing instead in about every 3 days or even 1 day as their conditions became worse. This will definitely cause much fees and radiation harm to the patients, as well as less accurate result by X rays. The diversity contrasts of auscultation and different radiological tests implies that the auscultation showed desired sensitivity and superior features in clinical usage, indicating a necessity of auscultation in diseases diagnosis and treatment.

This study had some limitations. We have limited number of patients included in each kind of classifications. Our dataset is consisted of the lung sounds acquired from patients in isolation ward because of the characteristics of 2019-nCov, thus the
auscultation was inevitably recorded with several kinds of noises, such as talking, cell phone ringing and air conditioner running, which made the lung sounds impure and could induce some interference for the model. However, this is also an essential way to keep the auscultation prone to the real clinical condition. Fallibility is inevitable with a large sample size even gold standard achieved at the expert-level by the committee members.

To the best of our knowledge, our study is the first to systematically identify the lung sound on COVID-19 with deep learning methods. Our study showed that deep learning algorithm performs with high precision in distinguish abnormal lung sound, what's more, it performs essentially valuable precision in identifying different kinds of abnormal lung sounds. This is the first time that a deep learning approach has been used to systematically diagnose almost all classifications of COVID-19, resulting in an end-to-end computerized, AI-based diagnosis model. As for the special characteristics of COVID-19, which is a serious viral pneumonia with widely spreading and without targeted medicine or vaccines, we foresee this deep learning approach showing a potentially promising use in infectious diseases, and newly appeared diseases fields. At the meantime, this AI-based diagnosis system also can be used in telemedicine, especially for rural areas and hospitals where experienced doctors are scarce. Our models with the advantages of cost-effectiveness and efficiency in classifying COVID-19 lung sounds, could potentially be used for helping real-time diagnosis and observing through wearable devices that monitor pulmonary and cardiovascular conditions through lung and cardiac auscultation.

In principle, our deep learning approach shows a superior capability with accuracy, efficiency and precision to classify COVID-19 in different severity level, and different kinds of abnormal lung sounds, indicating the deployment of automatic, computerized and AI-based decision-support systems in clinical environments. It provides clinicians with useful early prognostic information to facilitate pretreatment risk stratification for COVID-19, and guides the medical staff to conduct more intensive surveillance and treatment to patients at high risk of severe illness to
improve outcomes.

Competing of interests
All authors declare no competing interests.

Funding
This work is partially supported by National Natural Science Foundation of China (Grant 81873523, 61974053, 82100531), Fundamental Research Funds of the Central Universities of China (Grant 2019KFYXJJS049), and Wuhan Science and Technology Foundation of Application Frontier (Grant 2020020601012320).
References

1. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* Feb 15 2020;395(10223):507-513.

2. Guo T, Fan Y, Chen M, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA cardiology.* Mar 27 2020.

3. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front. Med.* Mar 12 2020.

4. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama.* Feb 7 2020.

5. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* Feb 15 2020;395(10223):497-506.

6. Commission CNH. Diagnosis and treatment of novel coronavirus pneumonia in China (trial version 7) https://www.who.int/docs/default-source/wpro---documents/countries/china/covid-19-briefing-nhc/1-clinical-protocols-for-thediagnosis-and-treatment-of-covid-19-v7.pdf?sfvrsn=c6cbfba4_2 accessed April 14, 2020 2020.

7. Yan CX, Li J, Shen X, Luo L, Li Y, Li MY. Biological Product Development Strategies for Prevention and Treatment of Coronavirus Disease 2019. *Sichuan Da Xue Xue Bao Yi Xue Ban.* Mar 2020;51(2):139-145.
8. Luan RS, Wang X, Sun X, et al. Epidemiology, Treatment, and Epidemic Prevention and Control of the Coronavirus Disease 2019: a Review. *Sichuan Da Xue Xue Bao Yi Xue Ban.* Mar 2020;51(2):131-138.

9. Marshall HR. Mosby's medical and nursing dictionary. *Nurse Educ. Today.* 1983/01/01/ 1983;3(4):86.

10. Sreejyothi S, Renjini A, Raj V, Swapna M, Sankararaman SI. Unwrapping the phase portrait features of adventitious crackle for auscultation and classification: a machine learning approach. *J. Biol. Phys.* 2021;47(2):1-13.

11. Mosby's Dictionary of Medicine, Nursing & Health Professions - Seventh edition
Mosby Mosby's Dictionary of Medicine, Nursing & Health Professions - Seventh edition 2272 Mosby 9780723433934 0723433933. *Nurs. Stand.* Feb 8 2006;20(22):36.

12. World Health O. *Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance, 28 January 2020.* Geneva: World Health Organization; 2020 2020.

13. Lei J, Li J, Li X, Qi X. CT Imaging of the 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology.* 2020;295(1):18-18.

14. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama.* 2013;310:2191-2194.

15. Shuvo SB, Ali SN, Swapnil SI, Hasan T, Bhuiyan MIH. A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram. *IEEE Journal of Biomedical and Health*
16. Ma Y, Xu X, Yu Q, Zhang Y, Wang G. LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using bi-ResNet Deep Learning Algorithm. Paper presented at: BioCAS2019.

17. Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017.

18. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016, 2016.

19. Acharya J, Basu A. Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning. *IEEE Transactions on Biomedical Circuits and Systems.* 2020;PP(99):1-1.

20. Sovijärvi ARA, Malmberg LP, Charbonneau G, Vanderschoot J, Earis JE. Characteristic of breath sounds and adventitious respiratory sounds. *European Respiratory Review.* 2000;10(77):591-596.

21. Sovijärvi ARA, Vanderschoot J, Earis JE. Standardization of computerized respiratory sound analysis. *European Respiratory Review.* 2000;10(77).

22. “ICBHI 2017 Challenge,” 2017, [Online]. Available: https://bhichallenge.med.auth.gr/.

23. Abbas A, Fahim A. An automated computerized auscultation and diagnostic system for pulmonary diseases. *J. Med. Syst.* Dec 2010;34(6):1149-1155.

24. Aykanat M, Ozkan K, Kurt B, Saryal S. Classification of lung sounds using convolutional neural networks. *Eurasip Journal on Image & Video Processing.*
25. Chen Q, Zhang W, Xiang T, Zhang X, Lei W. Automatic heart and lung sounds classification using convolutional neural networks. Paper presented at: 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)2016.

26. Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks. Artif. Intell. Med. 2018:S0933365717302051.

27. Fakoya FA, du Plessis M, Gbenimacho IB. Ultrasound and stethoscope as tools in medical education and practice: considerations for the archives. Advances in medical education and practice. 2016;7:381-387.

28. Nardone DA, Lucas LM, Palac DM. Physical examination: a revered skill under scrutiny. South. Med. J. Jun 1988;81(6):770-773.

29. Zhu J, Tan Y, Huang B, Zhu Y, Gao XH. Don't throw the stethoscope away! European heart journal. May 4 2020.

30. Dick EA, Varma D, Kashef E, Curtis J. Use of advanced imaging techniques during visits to emergency departments—implications, costs, patient benefits/risks. The British Journal of Radiology. 2016;89(1061):20150819.