



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.



## Automated lung ultrasound scoring for evaluation of coronavirus disease 2019 pneumonia using two-stage cascaded deep learning model

Wenyu Xing<sup>a,b,1</sup>, Chao He<sup>c,1</sup>, Jiawei Li<sup>d,e</sup>, Wei Qin<sup>f</sup>, Minglei Yang<sup>g</sup>, Guannan Li<sup>h</sup>, Qingli Li<sup>h</sup>, Dean Ta<sup>a,b,j</sup>, Gaofeng Wei<sup>i,\*</sup>, Wenfang Li<sup>c,\*</sup>, Jiangang Chen<sup>f,h,\*</sup>

<sup>a</sup> Center for Biomedical Engineering, School of Information Science and Technology, Fudan University, Shanghai 200438, China

<sup>b</sup> Human Phenome Institute, Fudan University, Shanghai 201203, China

<sup>c</sup> Department of Emergency and Critical Care, Changzheng Hospital, Naval Medical University, Shanghai 200003, China

<sup>d</sup> Department of Medical Ultrasound, Fudan University Shanghai Cancer Center, Shanghai 200032, China

<sup>e</sup> Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China

<sup>f</sup> Engineering Research Center of Traditional Chinese Medicine Intelligent Rehabilitation, Ministry of Education, Shanghai 201203, China

<sup>g</sup> Artificial Intelligence and Clinical Innovation Research, Neusoft Medical Systems Co., Ltd, Shenyang 110167, China

<sup>h</sup> Shanghai Key Laboratory of Multidimensional Information Processing, School of Communication & Electronic Engineering, East China Normal University, Shanghai 200241, China

<sup>i</sup> Naval Medical Department, Naval Medical University, Shanghai, 200433, China

<sup>j</sup> Academy for Engineering & Technology, Fudan University, Shanghai 200433, China

### ARTICLE INFO

#### Keywords:

LUS  
Automated scoring  
COVID-19 pneumonia  
Cascaded model  
Deep learning

### ABSTRACT

Coronavirus disease 2019 (COVID-19) pneumonia has erupted worldwide, causing massive population deaths and huge economic losses. In clinic, lung ultrasound (LUS) plays an important role in the auxiliary diagnosis of COVID-19 pneumonia. However, the lack of medical resources leads to the low using efficiency of the LUS, to address this problem, a novel automated LUS scoring system for evaluating COVID-19 pneumonia based on the two-stage cascaded deep learning model was proposed in this paper. 18,330 LUS images collected from 26 COVID-19 pneumonia patients were successfully assigned scores by two experienced doctors according to the designed four-level scoring standard for training the model. At the first stage, we made a secondary selection of these scored images through five ResNet-50 models and five-fold cross validation to obtain the available 12,949 LUS images which were highly relevant to the initial scoring results. At the second stage, three deep learning models including ResNet-50, Vgg-19, and GoogLeNet were formed the cascaded scored model and trained using the new dataset, whose predictive result was obtained by the voting mechanism. In addition, 1000 LUS images collected another 5 COVID-19 pneumonia patients were employed to test the model. Experiments results showed that the automated LUS scoring model was evaluated in terms of accuracy, sensitivity, specificity, and F1-score, being 96.1%, 96.3%, 98.8%, and 96.1%, respectively. They proved the proposed two-stage cascaded deep learning model could automatically score an LUS image, which has great potential for application to the clinics on various occasions.

### 1. Introduction

The coronavirus disease caused by SARS-CoV-2 that broke out in 2019 is highly contagious and has a serious impact on the economic development and the people's lives around the world [1]. The World

Health Organization has adjusted the risk assessment level for the prevention and control of coronavirus disease 2019 (COVID-19) pneumonia to the highest level. Currently, influenced by variant strains such as delta and Omicron, COVID-19's prevention, control, and treatment are still one of the serious challenges in the world. As of January 19, 2022,

\* Corresponding authors.

E-mail addresses: [wxying20@fudan.edu.cn](mailto:wxying20@fudan.edu.cn) (W. Xing), [hechao19820110@163.com](mailto:hechao19820110@163.com) (C. He), [jiaweili2006@163.com](mailto:jiaweili2006@163.com) (J. Li), [wqin6891@hotmail.com](mailto:wqin6891@hotmail.com) (W. Qin), [yangml60@midea.com](mailto:yangml60@midea.com) (M. Yang), [gqli@cs.ecnu.edu.cn](mailto:gqli@cs.ecnu.edu.cn) (G. Li), [qili@cs.ecnu.edu.cn](mailto:qili@cs.ecnu.edu.cn) (Q. Li), [tda@fudan.edu.cn](mailto:tda@fudan.edu.cn) (D. Ta), [highpeak8848@163.com](mailto:highpeak8848@163.com) (G. Wei), [chzhedlwf@163.com](mailto:chzhedlwf@163.com) (W. Li), [jgchen@cee.ecnu.edu.cn](mailto:jgchen@cee.ecnu.edu.cn) (J. Chen).

<sup>1</sup> Contributed equally to the work.

<https://doi.org/10.1016/j.bspc.2022.103561>

Received 17 November 2021; Received in revised form 20 January 2022; Accepted 2 February 2022

Available online 7 February 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.

334966436 cases were diagnosed worldwide, 5575544 cases died, and the number of deaths exceeds 1.66%. Therefore, how to effectively realize the timely diagnosis and treatment of patients with COVID-19 pneumonia is the top priority of the current response to the epidemic.

Ultrasound is a fast, convenient, radiation-free, low-cost, and targeted examination technology, which has been widely employed in clinical diagnosis [2,3]. In recent years, lung ultrasound (LUS) has been playing an increasingly important role in the diagnosis of various lung abnormalities, including pleural effusion, pneumothorax, chronic obstructive pulmonary disease (COPD), especially the COVID-19 pneumonia [4,5]. As reported, the specificity and sensitivity of LUS in diagnosing the COVID-19 pneumonia are better than chest X-ray, and close to CT [6]. When the outbreak of COVID-19 pneumonia occurred in 2019, characterized as “small and fast”, ultrasound examinations significantly contributed to solve a variety of real difficulties, such as a huge quantity of patients, limited medical resources, and contaminated environment. It also has become the main imaging equipment to enter the intensive care unit (ICU) in the epidemic area, providing effective bedside supports for real-time evaluation of COVID-19 pneumonia [7–14].

The LUS imaging findings of COVID-19 pneumonia is like common pneumonia, showing as a single or multiple A-lines, B-lines, irregular pleural lines, pulmonary consolidation, and other morphological changes to normal LUS images [12]. Researchers have used lung ultrasound scores (LUSS) to evaluate pneumonia with good clinical achievements. Yusuf et al. [15] introduced the use of contrast-enhanced ultrasound in COVID-19 lung imaging and its diagnosis of three cases. Gutsche et al. [16] adopted a specific LUS score to classify the patients as LUS-positive and -negative cases. Mento et al. [17] combined different number of scanning areas and corresponding scoring system to evaluate the severity of patients. Anderson et al. [18] divided the chest wall into 8 areas, including the two front areas and the two side areas of each hemithorax, and evaluated the B-line in each area space to obtain the corresponding score. Soldati et al. [19] proposed a four-level scoring system of Scores 0–3 based on the presence/absence and spatial extent of A-lines, B-lines, sub-pleural consolidations, and white lung. Zhu et al. [20] employed the point of care LUS and statistical analysis to achieve the classification of moderate, severe, and critically ill for evaluating COVID-19 pneumonia patients. Zhao et al. [21] assigned different LUSS for the COVID-19 pneumonia according to the most severe finding from the LUS images collected from ten zones on every patient’s chest wall. Li et al. [22] employed the LUSS of Scores 0–3 to provide valuable semi-quantitative information for evaluating the COVID-19 pneumonia in neonates. Moreover, other LUS scoring methods can be found in references [23–25]. Although such methods provided semi-objective evaluation of the pneumonia, the LUSS still relied on the visual judgment of experienced clinicians. This limited the application value of LUSS in the clinical diagnosis of pneumonia, especially in those areas with limited medical resources.

Nowadays, there have been many studies applying computer techniques such as image processing or machine learning for the automatic B-line detection and scoring of LUS images, with an aim to scale the image to difficult severities. Brusasco et al. [26] applied the k-means algorithm to divide pixels into two subsets including B-lines and no B-lines, thereby detecting the B-lines in an automated manner. Brattain et al. [27] detected the B-lines via classification models based on five characters extracted from the angular slices. van et al. [28] proposed a weakly supervised deep learning method for automatically detecting and localizing B-lines in an ultrasound scan. Roy et al. [11] and Carrer et al. [9] also focused on application of different classifiers for assigning the LUS scoring, including machine learning and deep learning.

However, the traditional B-line score only focused on the change of B-lines, which could reflect limited aspects of the lung disease. There indeed exists some LUS scoring systems that comprehensively considers changes in structural characteristics including A-line, B-line, pleural line, and lung consolidation, including those for evaluating the COVID-

19 pneumonia. For example, Carrer et al. [9] employed the eight features of pleural line and its underlying area into the machine learning for the automatic LUS scoring. Chen et al. [29] proposed a quantitative feature extraction method, and used the neural networks, support vector machines, and decision trees for automatic scoring of LUS images. Wang et al. [30] designed four features related to pleural line and four features related to B-line to analyze the lung ultrasound images, and realized the binary severe/non-severe classification by support vector machine. Most of the studies are based on self-designed feature extraction algorithms combined with machine learning models to achieve the severity assessment of COVID-19 pneumonia. This kind of methods is inefficient, and cannot fully consider the characteristics of various indicators of the entire LUS image and the correlation between them.

With such an awareness, this paper realized the design and verification of the automatic scoring system for lung ultrasound based on the four-level scoring standard of Scores 0–3 [17,19] and the two-stage cascaded deep learning model, which provides reliable technical support for the rapid, convenient, and accurate assessment of the severity of COVID-19 pneumonia in clinic.

The main contributions of this paper are summarized as follows.

Through the cross-validation and multiple ResNet-50 models, the effective selection of data sets was realized, which reduced the errors when clinicians performed manual scoring, and provided reliable data support for the accuracy of model training.

Through the design of the cascaded deep learning model, a LUS automated scoring system with high-accuracy for evaluating COVID-19 pneumonia was obtained.

The method proposed in this paper improved the clinical significance of LUS in the diagnosis of COVID-19 pneumonia, and facilitated real-time detection of patient conditions in areas with insufficient medical resources.

## 2. Materials and methods

### 2.1. Patients

The experiment involved 31 patients who were admitted to the Wuhan Huoshenshan Hospital from February 23, 2020 to April 2, 2020, and confirmed as COVID-19 pneumonia with computed tomography (CT, United Imaging, uCT760) and positive RT-PCR test. Their clinical indicators are shown in Table 1.

All patients underwent LUS examinations for 12 standard fields (Fig. 1), with six chest areas per hemithorax based on a division of each hemithorax into anterior/lateral/posterior/and upper/lower zones [31–33], according to the BLUE (bedside LUS in emergency) protocol (Lichtenstein and Meziere7). In each field, with the probe kept stationary, three ultrasound cine loops with a duration of 5 s (roughly more than one respiration cycle) were collected and stored in the format of Digital Imaging and Communications in Medicine (DICOM). The ultrasound equipment LOGIQ e (GE Healthcare, Wauwatosa, WI, USA) was utilized and equipped with a curved array low frequency transducer (1–5 MHz) (image depth: 15 cm, focal depth: 7.5 cm, mechanical index (MI): 1.2, thermal index (TI): 0.7, operation mode: penetration). No more than 50 LUS images were selected in each cine loop, in a way of

**Table 1**  
Patient statistics data.

Indicators	Value
Number of patients tested by CT	31
Number of patients tested by RT-PCR	31
Age	55 ± 21
Height/cm	168 ± 13
Weight/kg	70 ± 18
Pulse/bpm	84 ± 26
Blood pressure/mmHg	128/72 ± 23/19
Oxygen saturation/%	96 ± 4

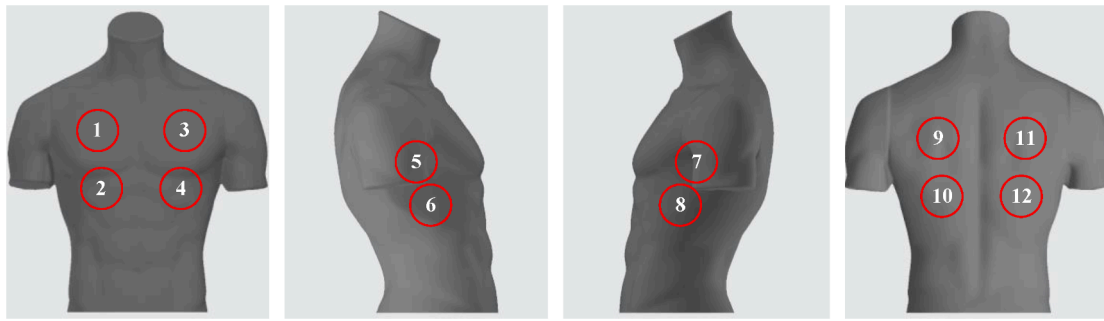


Fig. 1. Schematic representation of the twelve acquisition areas on chest.

that the number of frames between adjacent images was not less than 10 frames. Such an operation was under the consideration that the adjacent frames of the ultrasound cine loops may be of high similarity with each other and make the model overfitting. A total of 20,163 LUS images were extracted from the afore collected cine loops. Clinically, the worst case in multiple regions is usually regarded as the clinical manifestation of patients.

This study was also approved by the Ethics Committee of Huoshenshan Hospital, Wuhan, China (Approval number: HSSL030). All patients provided written informed consents by themselves or family members. The data will be provided on the website in the future (<https://bio-hsi.ecnu.edu.cn/>).

2.2. System design

In this paper, the design of the automated LUS scoring system for COVID-19 pneumonia is shown in Fig. 2. The system was divided into three parts: i) establishing scoring standard and performing the initial scoring by the two experienced clinicians, ii) verifying the initial scoring

based on multiple ResNet-50 models for image secondary selection, and iii) building the final accurate scoring model based on the cascaded deep learning models with re-selected new LUS data set.

2.3. Scoring standard

As determined by experienced clinicians, referring to clinical indicators and the previously proposed scoring system specifically for evaluating the COVID-19 pneumonia [19,20], the degree of lesions and scoring standard are divided into Scores 0, 1, 2, and 3, as shown in Fig. 3. This was also a widely accepted scoring standard for evaluating the clinical severities of patients with COVID-19 pneumonia using LUS.

Score 0: The pleural line is continuous and regular with A-lines present.

Score 1: The pleural line is indented and discontinued, with multiple spaced B-lines spreading at an interval of approximately 7 mm in between.

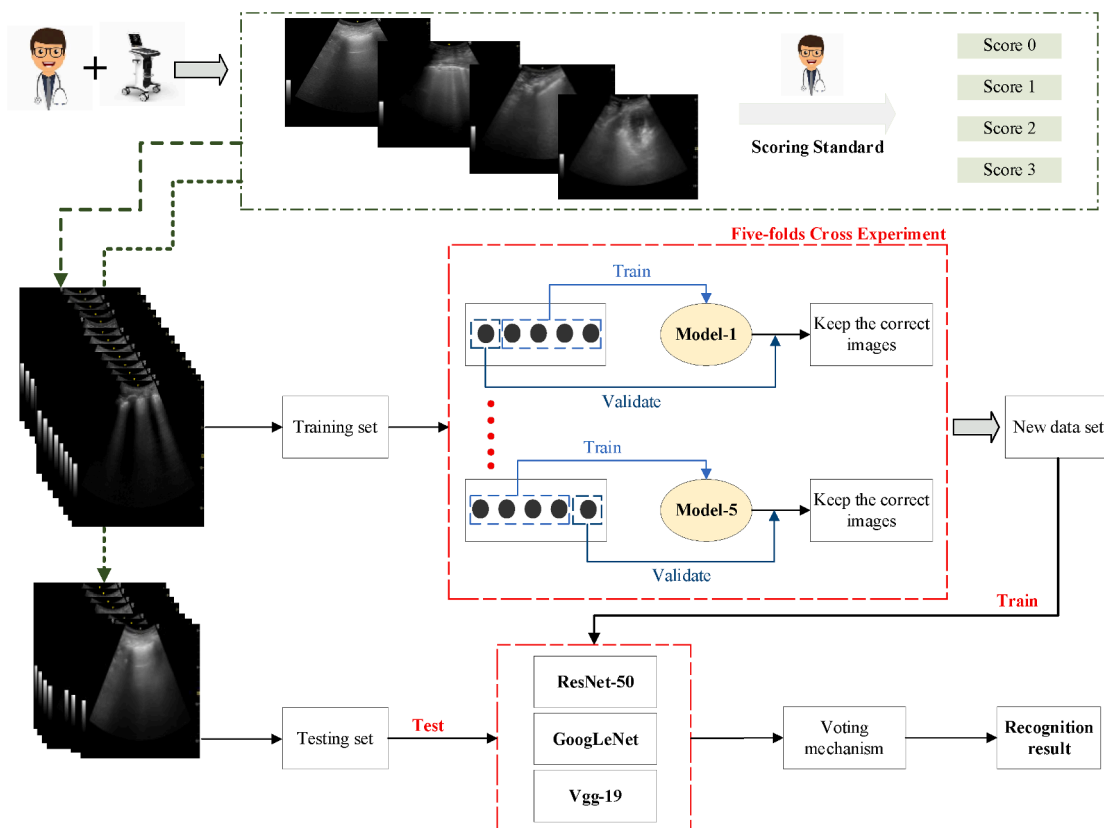


Fig. 2. Schematic diagram of automated LUS scoring system.

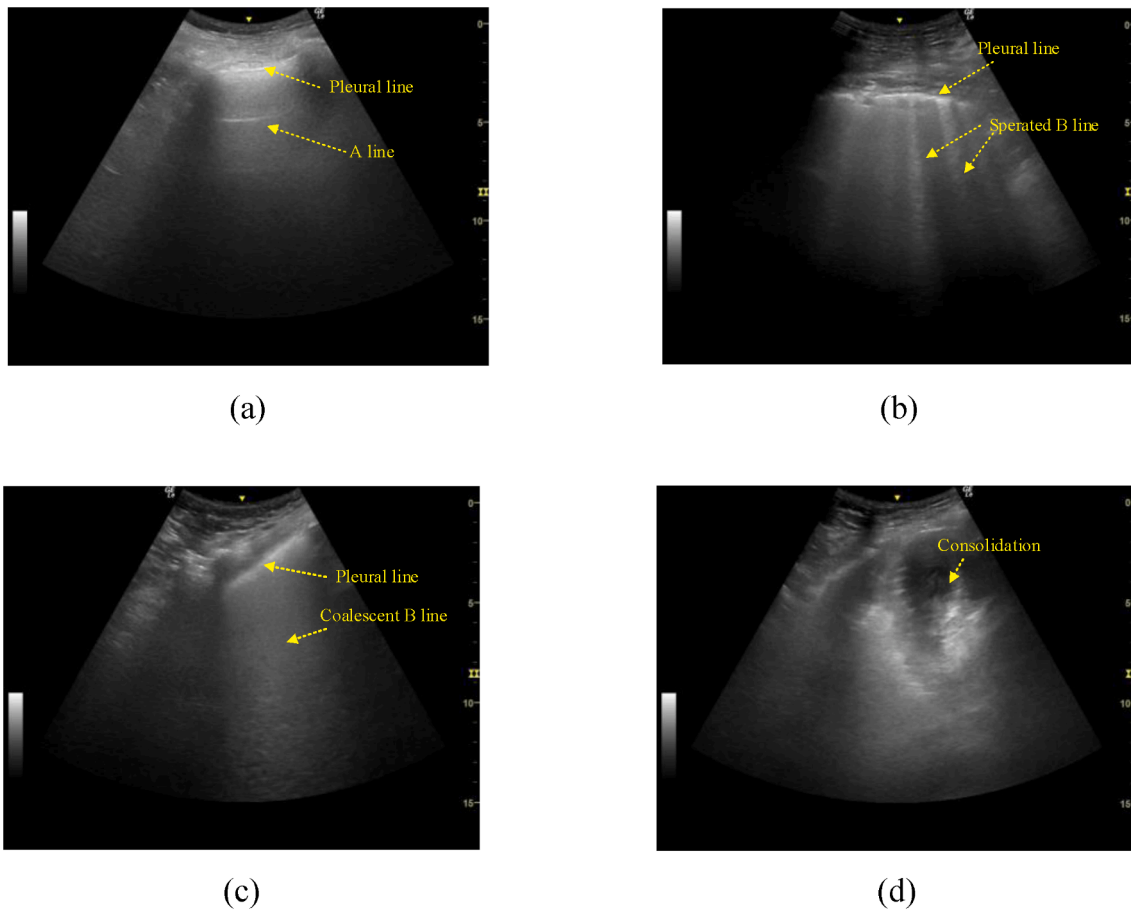


Fig. 3. Four lung ultrasound patterns according to the scoring standard. (a) Score 0, (b) Score 1, (c) Score 2, (d) Score 3.

Score 2: The pleural line is severely broken with coalescent B-lines at an interval of  $\leq 3$  mm in between.

Score 3: There are dense and largely extended white lung with or without larger consolidations.

#### 2.4. Automated scoring system

COVID-19 pneumonia has different indicators in LUS images, like consolidation, B-line, A-line, etc., which has no accurate gold standard. They can only rely on clinician’s identification as a semi-standard to ensure that the data used in training are accurate. Therefore, every collected LUS images was scored by two experienced clinicians (more than 6 years in using LUS) according to the above scoring standard and procedure (Figs. 3 and 4), respective. If the two scores were equal, we then assigned the result to the LUS image as the initial score. Otherwise, we discarded this image. As a result, we can increase the credibility of the labels. In addition, those images that the clinicians were not able to recognize also could not be recognized by our model, since the model was the learner of the clinicians.

However, the poor quality of some ultrasound images may cause the scores of some images of low confidence, possibly deteriorating the performance of the developed automated scoring model. In our study, we can regard these LUS images with low scoring confidence as possible scoring interference items in the image. Therefore, to remove these images and address this problem, the two-stage cascaded model was designed with the first stage to exclude those images of low confidence, and the second stage to build the classification model for LUS scoring. Meanwhile, there are some other similar designs published in literatures. Ren et al. [34] used the two-stage cascaded model to select the most promising candidate sets to achieve the image set classification

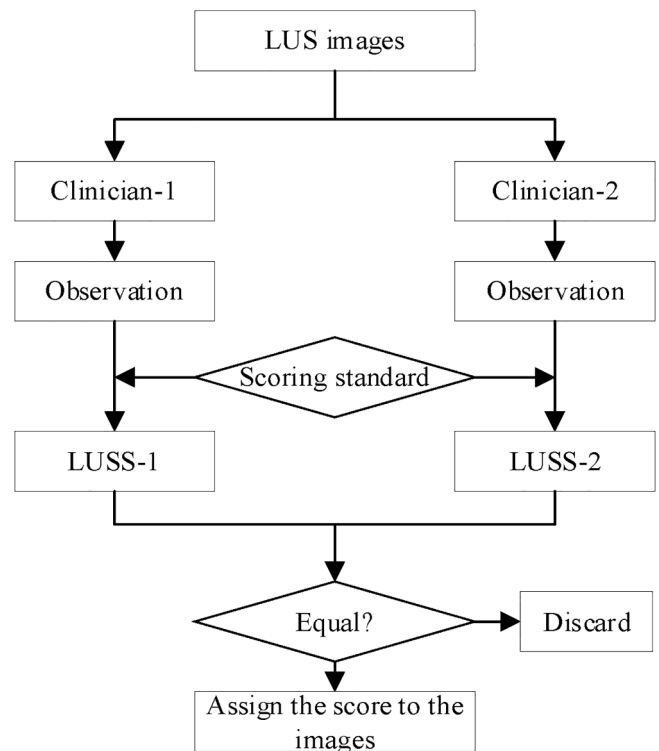


Fig. 4. Procedure of LUSS assignment.

task. Nam et al. [35] removed noise signal in the first stage and realize target signal classification in the second stage.

#### 2.4.1. Secondary selection of LUS images

According to Fig. 1, this paper adopted a secondary selection model based on ResNet-50 model and 5-fold cross validation for traversing the entire dataset. By dividing the data set into five parts, each experiment put four input network models to train. The well-trained neural network model was then used to identify the remaining data, with the correct samples remained and the wrong samples abandoned. The above operation repeated five times, as a result, we have the data set of good quality and ensured score, based on which we can carry out the subsequent automatic scoring model training.

In the first stage, five deep learning models based on ResNet-50 [36], as shown in Fig. 5, were applied for the secondary selection of LUS images using the stochastic gradient descent with momentum (SGDM) optimizer [37] and the SoftMax classifier [38].

The model was mainly composed of the convolution block and the identity block. The former is mainly used to change the dimension of the network, while the latter is mostly used to deepen the depth of the network. In the shallow network, by normalizing the data in the middle layer, it can ensure that the network adopts the stochastic gradient descent (SGD) algorithm in the back propagation. The network can thereby reach convergence. However, when we need to extract the deeper features of the complex images, the traditional network design and linking methods may make the SGD algorithm lose efficacy in deep layers. The model was thereby difficult to get converge. As proposed in the study, the residual module was developed based on the convolutional neural network model, which can send parts of the input data directly to the output layer without passing through the convolution layer, retaining parts of the original information, effectively solving the problem of gradient dispersion in back propagation. Such an operation could realize the training of deeper network and the extraction of deeper features. In addition, the Zero-Pad layer was added before the convolution layer, with the purpose of making the input image and the feature image after the convolution layer have the same dimension. Assuming the input image size is  $W_1 \times Z_1 \times D_1$ , the size calculation method of the input feature maps is shown in (1).

$$\begin{aligned} W_2 &= (W_1 - F + 2P)/S + 1 \\ H_2 &= (H_1 - F + 2P)/S + 1 \\ D_2 &= K \end{aligned} \quad (1)$$

where  $K$  is the number of filters,  $F$  is the size,  $S$  is the step size, and  $P$  is the boundary filling.

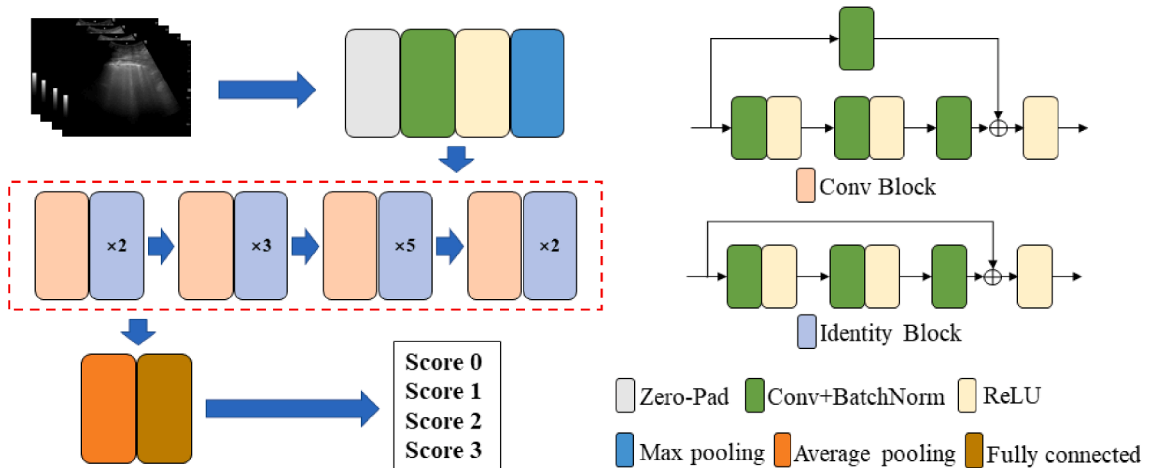


Fig. 5. Schematic diagram of ResNet-50 CNN model.

#### 2.4.2. Establishment of final scoring model

After the secondary selection of LUS images integrated with ResNet-50 model in the first stage, a new data set with more accurate LUSS than the original data was obtained. This process reduced the errors of subjective and objective factors in the initial scores, which improved the accuracy of the final scoring model.

The second stage of the final accurate scoring system was designed based on the transfer learning of multiple deep learning network models (Fig. 6), they all used the new data set to re-train the model. At the end, the predictive results of the testing images were obtained by the voting mechanism of three different deep learning models including ResNet-50, Vgg-19, and GoogLeNet.

- ResNet-50 has been introduced in Part 2.4.1.
- Vgg-19 [39] contains 19 layers (16 convolutional and 3 fully connected), the whole network uses the same convolution kernel size ( $3 \times 3$ ) and maximum pool size ( $2 \times 2$ ), which can handle a large number of parameters and better than a large filter ( $5 \times 5$  or  $7 \times 7$ ) convolution layer.
- GoogLeNet [40] contains 22 layers with the inception module, which maintains the sparsity of neural network structure and makes full use of the high computational performance of dense matrix.

#### 2.4.3. Loss function

To improve the automated LUS scoring model, the cross-entropy loss function  $L_{CE}$ , as shown in (2), was used in the network to minimize the loss of the model.

$$L_{CE} = - \sum_{i=1}^N t_i \log(p_i) \quad (2)$$

where  $N$  is the number of categories,  $t_i$  and  $p_i$  is the real-value and prediction probability of the category, respectively.

#### 2.5. Statistical analysis

Statistical analysis was performed using SPSS 22.0 for Windows system (SPSS Inc., Chicago, IL, USA). The data, which consist of age, height, weight, pulse, blood pressure, and oxygen saturation, were summarized as median  $\pm$  range. The results analysis indicators, which consist accuracy, sensitivity, specificity, and F1 score, were analyzed in this experiment.



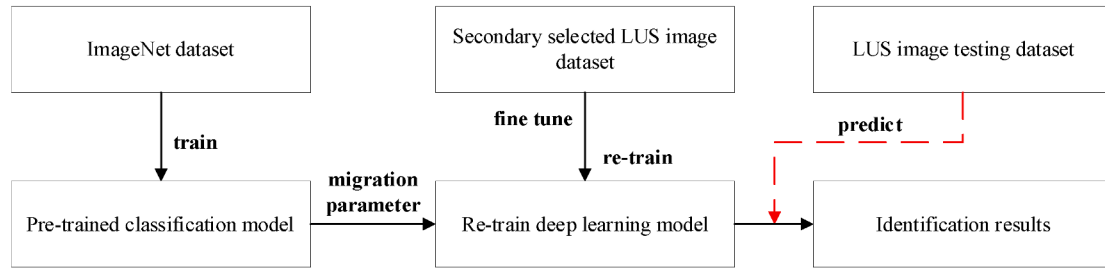


Fig. 6. Schematic diagram of transfer learning.

### 3. Experimental results and analysis

A total of 20,163 LUS images were collected from 31 patients who were admitted to the Wuhan Huoshenshan Hospital from February 23, 2020 to April 2, 2020, and confirmed as COVID-19 pneumonia. After double-blind scoring by two experienced clinicians, 19,330 LUS images assigned with the LUSS were employed in the experiment accounting for 95.87%, most of the data removed were of poor acquisition quality (i.e., rib blocking or poor contact). Among them 18,330 LUS images with the composition of Score 0: 4930 (26.90%), Score 1: 5710 (31.15%), Score 2: 4070 (22.20%), and Score 3: 3620 (19.75%) collected from 26 patients were used for training. 1000 LUS images with each of the four scores being 250 collected from another 5 patients were used for testing, which did not participate in any training process. This allocation method of training set and testing set can ensure that there was no cross between each other, and ensured the reliability of the experimental results.

The deep learning models were performed using Pytorch framework, running on a computer with a CPU: Intel Xeon Gold 6248R, RAM: 256G, and GPU: Tesla V100.

#### 3.1. Secondary selection result of initial data

The original training set (18330 LUS images) were trained using ResNet-50 for five times with the method of 5-fold cross validation. During the training process, it was found that the model performance no longer changed after the 2000th iterations. Therefore, the number of iterations of the experiment was fixed as 2000. The batch size was set to 4 according to our data quantity and hardware equipment conditions. The other three parameters including learning rate, regularization parameter, and momentum parameter were set according to the empirical values [41,42]. The training parameters of the CNN models are shown in Table 2. The training process and loss value of five ResNet-50 models is shown in Fig. 7. We used the testing set (1000 LUS images) to test the five well-trained models, the accuracy was 82.51%, 88.25%, 87.80%, 84.45%, and 86.98%, respectively (Table 3). It shows that the proposed cascaded model can be applied for the secondary selection. As a result, the one-fold data set (validation set) in each experiment can be predicted by the well-trained model based on other four-fold data set. Thus, each LUS images in the training set can be verified through the five experiments. A new LUS image dataset with highly relevance to the original scoring results was obtained for the second stage experiment.

Table 2  
Training parameters of classification network.

Parameter	Value
Iterations	2000
Batch size	4
Learning rate	0.001
L2 regularization	0.0001
Momentum	0.9
Loss function	$L_{CE}$
Frame	Matlab 2019b

The experimental results of model testing and LUS image selection are shown in Table 4 with the new dataset containing 12,949 (Score 0: 27.14%, Score 1: 32.76%, Score 2: 14.20%, Score 3: 25.90%) LUS images.

#### 3.2. Accuracy of automated scoring model

The new data set containing 12,949 LUS images were put into the ResNet-50, Vgg-19, and GoogLeNet models, respectively. The training parameters were same as Table 2. The training process was shown in Fig. 8. After 2000 iterations, the training accuracy of three models were 94.09%, 99.12%, and 98.85%, respectively. The loss value of three models were 0.18, 0.10, and 0.02, respectively.

Thus, the score of 1000 LUS images in testing set were predicted using the three well-trained deep learning models. Table 5 shows the testing accuracy of these three models. In addition, we also used the idea of cascade prediction, for one testing image. If two or three models have the same prediction results, we would assign this score to the testing image. If each model had different prediction results, we would assign the results with the highest confidence and higher than 80% to the testing image. The comparison was shown in Fig. 9, which shows that the prediction accuracy of the cascade model was significantly improved in most scores, only 0.2% lower than that of Vgg-19 model on Score 1. The final accuracy of Scores 0–3 based on overall cascaded model is 95.6%, 99.2%, 89.6%, and 100%, respectively, with the average being 96.1%.

#### 3.3. Evaluation of automated scoring model

Fig. 10 was the confusion matrix of different scores of the cascaded LUS scoring model, which shows that the proposed scoring model has a strong correlation with Scores 0–3. According to the confusion matrixes, four evaluation indicators of accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1 score (F1), as shown in (3)–(6), were adopted to evaluate the effectiveness of the automated scoring model for COVID-19 pneumonia.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$SEN = \frac{TP}{TP + FN} \quad (4)$$

$$SPE = \frac{TN}{TN + FP} \quad (5)$$

$$F1 = \frac{2TP}{2TP + FN + FP} \quad (6)$$

Here,  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  are true positive, true negative, false positive, and false negative, respectively.

The experimental results of four evaluation indicators are shown in Table 6. The mean value of four indexes can reach 96.1%, 96.3%, 98.8%, and 96.1%, respectively. It shows the proposed methods has high accuracy, precision, sensitivity, and specificity, and indicates the promise

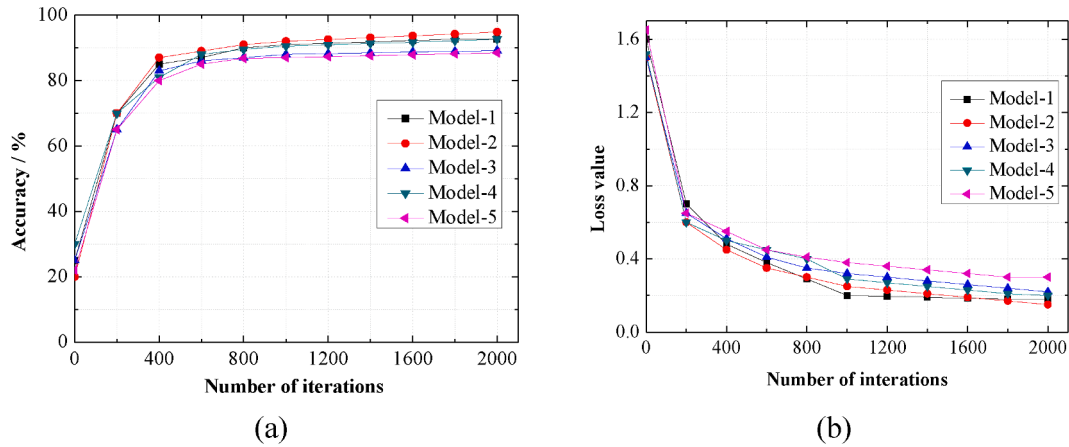


Fig. 7. (a) Training process and (b) loss value of the first five ResNet-50 models.

**Table 3**  
Experimental results of the first stage.

Score	Accuracy of testing set/%				
	1	2	3	4	5
Score 0	92.90	84.79	91.08	65.72	95.74
Score 1	83.19	86.34	96.53	92.99	94.75
Score 2	55.03	87.47	64.37	83.78	57.74
Score 3	98.90	94.41	99.21	95.30	99.69
Average	82.51	88.25	87.80	84.45	86.98
	85.99				

**Table 4**  
Experimental results of the first stage.

Score	Number of correct samples in each experiment				
	1	2	3	4	5
Score 0	952	547	779	358	878
Score 1	711	850	694	1002	985
Score 2	316	469	566	386	102
Score 3	586	717	715	679	657
Sum	2565	2583	2754	2425	2622
	12,949				

for clinical applications.

3.4. Comparison with other methods

Automated LUS scoring has been applied to scale the severities of COVID-19 pneumonia. Previously, Carrer et al. [9] proposed a pleural line detection and localization method which employed the eight

features of pleural line and its underlying area into the machine learning for the automated scoring. Chen et al. [29] also designed a similar scoring model including five steps, Step 1 transferred image format, Step 2 performed the pleural line detection, Step 3 selected the ROI, Step 4 extracted 28 different features, and Step 5 achieved the automatic scoring (Scores 0–3) based on deep learning. This model analyzed more features than Carrer’s method [9] and obtained a great performance for scoring LUS images.

In this study, we compared our method with the Chen’s (Fig. 11: method-1) [29] using the data set described in this manuscript, the accuracy of the proposed two-stage cascaded deep learning model was improved by about 9%. What’s more, we used the same testing set to evaluate these models used in the first and second stages, the accuracy gains of 10.11% is obtained compared with the average accuracy of five well-trained ResNet-50 models in the first stage (85.99%, Fig. 11: method-2). The average increments are 5.73% and 4.26% compared with the three independent deep learning models used in the second stage including ResNet-50, Vgg-19, and GoogLeNet, which were trained using the initial data set (18330 LUS images, Fig. 11: methods-3,4,5) and the secondary selection dataset (12949 LUS images, Fig. 11: methods-7,8,9), respectively. In addition, when we used the voting mechanism

**Table 5**  
Experiment results of the final scoring model.

Model	Training accuracy/%	Testing accuracy/%				
		Score 0	Score 1	Score 2	Score 3	Average
ResNet-50	94.09	82.4	93.2	76.4	99.2	87.8
Vgg-19	99.12	97.6	96.0	86.4	99.6	94.9
GoogLeNet	98.85	88.4	98.8	85.2	98.8	92.8

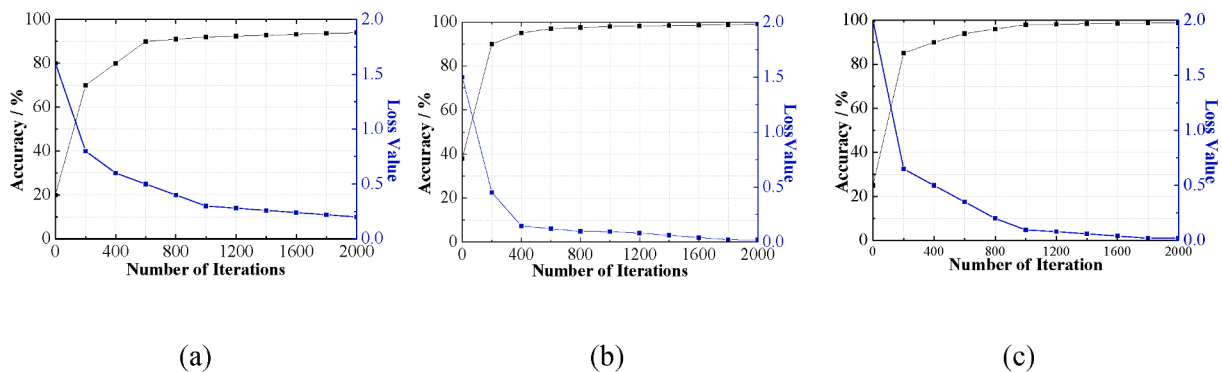


Fig. 8. Training process of (a) ResNet-50, (b) Vgg-19, and (c) GoogLeNet model.



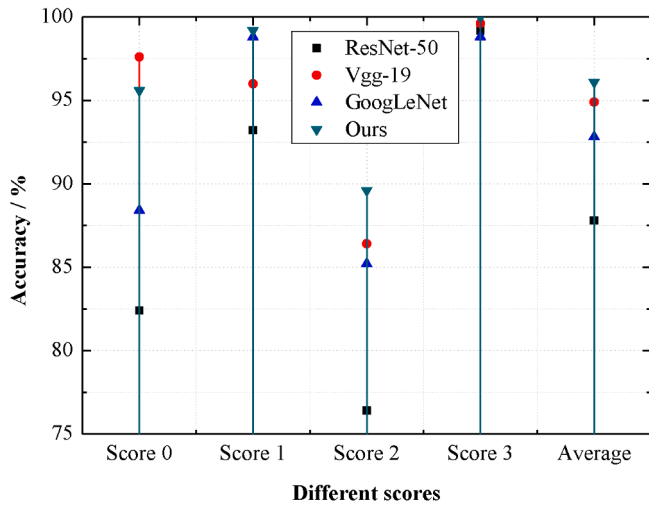


Fig. 9. Comparison of different models.

#### 4. Discussion

In this paper, we proposed a two-stage cascaded deep learning model for automated scoring of the COVID-19 pneumonia with a data set of 20,163 LUS images collected from 31 patients in Wuhan Huoshenshan Hospital. Via the two-stage experimental process including image secondary selection and accurate scoring model establishment, we obtained the LUS automated scoring model which has a high prediction accuracy of 96.1%.

In previous studies, various scoring methods based on LUS images were proposed to evaluate the severities of COVID-19 pneumonia, which are based on feature extraction and machine learning [29,30]. Through quantitative analysis of pleural line [9] or B-line [43], they realized the clinical auxiliary diagnosis. But these methods mainly focused on the analysis of single or two indicators and ignored the correlation of multiple indicators and the overall analysis of the LUS images. This may cause the loss of some information and make the auxiliary diagnosis results inaccurate. As a contrast, the proposed method based on deep learning models can analyze the image from a deeper level. Through

Table 6

Evaluation of the final automated scoring model.

Index	ACC/%	SEN/%	SPE/%	F1/%
Score 0	95.6	95.2	98.6	95.4
Score 1	99.2	96.5	99.7	97.8
Score 2	89.6	97.4	96.8	93.3
Score 3	100.0	96.2	100.0	98.0
Average	96.1	96.3	98.8	96.1

to evaluate the performance of ResNet-50, Vgg-19, and GoogLeNet trained with the initial data set (without five-fold cross validation filtering, Fig. 11: method-6), the results show that the scoring accuracy is 93%, which is lower than 3.1% of the proposed method in this paper.

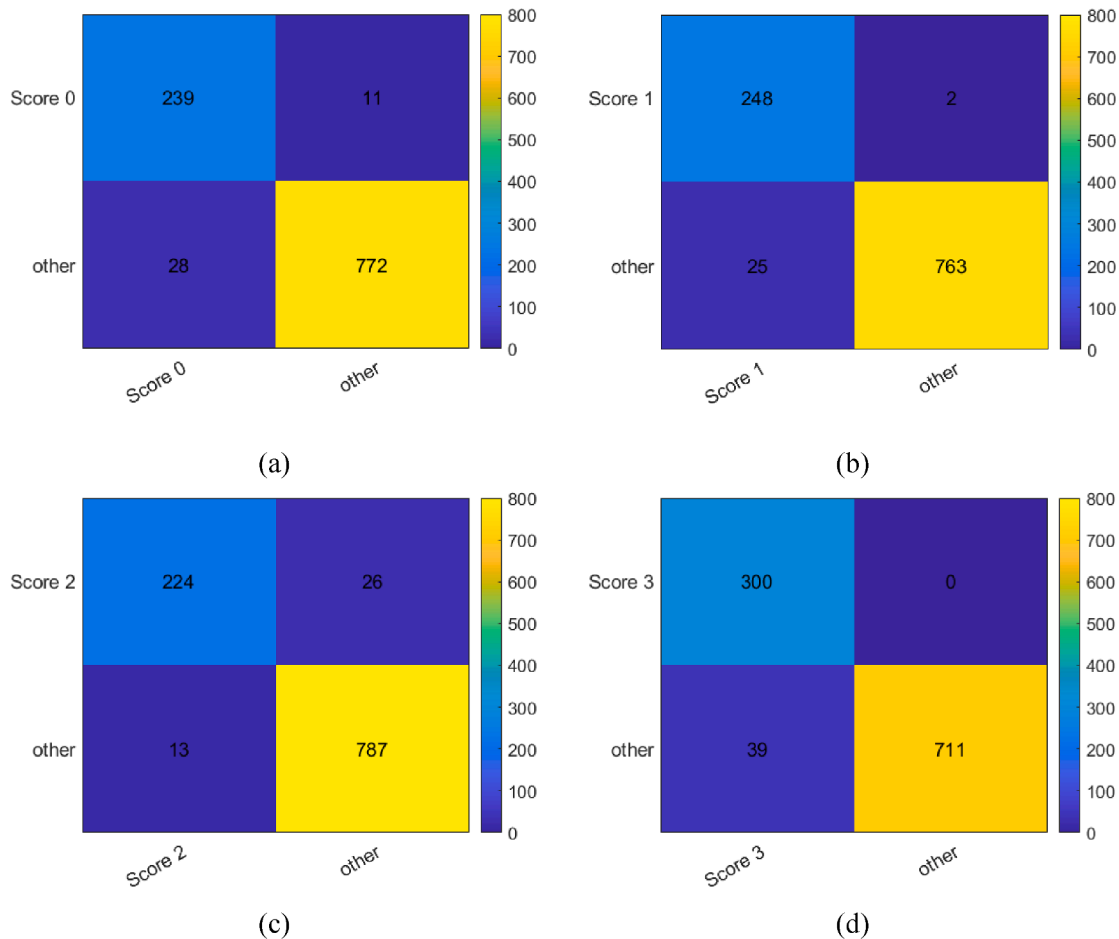
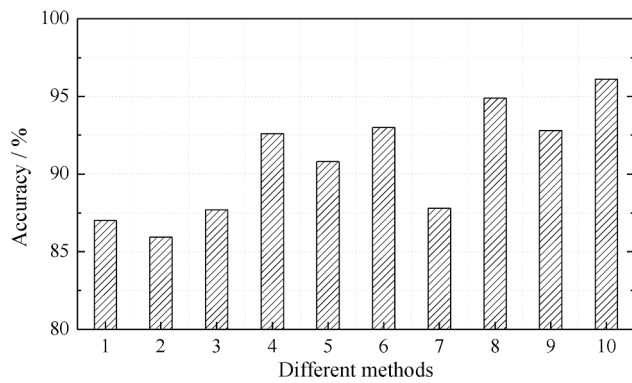


Fig. 10. Confusion matrix of (a) Score 0, (b) Score 1, (c) Score 2, and (d) Score 3.



**Fig. 11.** Comparison with different methods. 1-Chen et al., 2-Average of first stage, 3-ResNet-50 (initial dataset), 4-Vgg-19 (initial dataset), 5-GoogLeNet (initial dataset), 6-Voting mechanism (initial dataset), 7-ResNet-50 (re-selected dataset), 8-Vgg-19 (re-selected dataset), 9-GoogLeNet (re-selected dataset), 10-Ours.

comparing with Chen's experimental results [29], it is also confirmed that the proposed method has a significant improvement in the accuracy of automated LUS scoring. Meanwhile, the effective designs of the scoring model using multiple strategies including double blind score, cascade models, five-fold cross validation, and voting mechanism ensure the feasibility of the model as much as possible and have an improvement compared with the single deep learning model.

However, although promising results were achieved, there are several limitations of the study. First, this paper integrated the images collected from multiple standard fields of patients with COVID-19 pneumonia into one data set to train the scoring model, with the highest score was used as the diagnostic result [44]. According to the references [44] and [45], the highest scores were mainly focused on the lateral and posterior areas, whereas the lowest scores were focused on the anterior area. However, this paper lacked this research limited by the data set. Secondly, the proposed model was mainly based on the deep learning networks, whose solvability is poor and may have an unknown impact on clinical application. Thirdly, the time was urgent and the equipment resources were limited at that time, the data applied in this paper were collected from the same site using the same ultrasound scanner, and its practical clinical application needs to be verified in multi center data.

In future work, we will study the automatic detection methods of the pleural, A-lines and B-lines to help the manual annotation of the images, achieve the quantitative analysis of image, and improve the clinical solvability. Meanwhile, we will collect more LUS data with explicit field to study how the fields affect the performance, improve our model, and attempt to validate the effect on multiple communities. Therefore, the proposed method will be better served the auxiliary diagnosis for clinicians.

## 5. Conclusion

In this paper, we proposed an automated LUS scoring technique for evaluating COVID-19 pneumonia based on the two-stage cascaded deep learning model. Through the setting of LUS scoring standard, secondary selection of LUS images, and establishment of automated scoring model, this method achieved accurate and automatic LUS scoring of COVID-19 pneumonia. The results proved this method has a better performance than previous methods and a great clinical application potential.

## 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.

## Acknowledgements

This work was supported by the Program of Shanghai Academic Research Leader [grant number 19XD1400500], the National Natural Science Foundation of China [grant number 61975056], the Shanghai Natural Science Foundation [grant number 19ZR1416000], the Science and Technology Commission of Shanghai Municipality [grant number 14DZ2260800], the Key Research Fund of Logistics of PLA [grant number BWS14C018], and the Shanghai Health and Family Planning Commission [grant number 4802016ZB0201].

## References

- [1] W.-J. Guan, Z.-y. Ni, Y.u. Hu, W.-H. Liang, C.-Q. Ou, J.-X. He, L. Liu, H. Shan, C.-L. Lei, D.S.C. Hui, B. Du, L.-J. Li, G. Zeng, K.-Y. Yuen, R.-C. Chen, C.-I. Tang, T. Wang, P.-Y. Chen, J. Xiang, S.-Y. Li, J.-L. Wang, Z.-J. Liang, Y.-X. Peng, L.i. Wei, Y. Liu, Y.-H. Hu, P. Peng, J.-M. Wang, J.-Y. Liu, Z. Chen, G. Li, Z.-J. Zheng, S.-Q. Qiu, J. Luo, C.-J. Ye, S.-Y. Zhu, N.-S. Zhong, Clinical characteristics of coronavirus disease 2019 in China, *N. Engl. J. Med.* 382 (18) (2020) 1708–1720.
- [2] C.C. Liu, B.Y. Li, Y. Li, W.Y. Mao, C. Chen, R. Zhang, D.A. Ta, Ultrasonic backscatter difference measurement of bone health in preterm and term newborns, *Ultrasound Med. Biol.* 46 (2020) 305–314.
- [3] P.H. Mayo, R. Copetti, D. Feller-Kopman, G. Mathis, E. Maury, S. Mongodi, F. Mojoli, G. Volpicelli, M. Zanobetti, Thoracic ultrasonography: a narrative review, *Intensive Care Med.* 45 (9) (2019) 1200–1211.
- [4] B. Bouhemad, H. Brisson, M. Le-Guen, C. Arbelot, Q. Lu, J.J. Rouby, Bedside ultrasound assessment of positive end-expiratory pressure-induced lung recruitment, *Am. J. Respir. Crit. Care Med.* 183 (2011) 341–347.
- [5] G. Soldati, M. Demi, A. Smargiassi, R. Inchingolo, L. Demi, The role of ultrasound lung artifacts in the diagnosis of respiratory diseases, *Expert Rev. Respir. Med.* 13 (2019) 163–172.
- [6] Q.Y. Peng, X.T. Wang, L.N. Zhang, C.C. Chinese, Findings of lung ultrasonography of novel corona virus pneumonia during the 2019–2020 epidemic, *Intensive Care Med.* 46 (2020) 849–850.
- [7] J. Chen, J. Li, C. He, W. Li, Q. Li, Automated pleural line detection based on radon transform using ultrasound, *Ultrasound Imaging* 43 (1) (2021) 19–28.
- [8] D. Buonsenso, F. Raffaelli, E. Tamburrini, D.G. Biasucci, S. Salvi, A. Smargiassi, R. Inchingolo, G. Scambia, A. Lanzone, A.C. Testa, F. Moro, Clinical role of lung ultrasound for diagnosis and monitoring of COVID-19 pneumonia in pregnant women, *Ultrasound Obstet. Gynecol.* 56 (1) (2020) 106–109.
- [9] L. Carrer, E. Donini, D. Marinelli, M. Zanetti, F. Mento, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, F. Bovolo, L. Bruzzone, Automatic pleural line extraction and COVID-19 scoring from lung ultrasound data, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 67 (11) (2020) 2207–2217.
- [10] W.Z. Lu, S.S. Zhang, B.H. Chen, J.X. Chen, J.Z. Xian, Y.H. Lin, H. Shan, Z.Z. Su, A clinical study of noninvasive assessment of lung lesions in patients with coronavirus disease-19 (COVID-19) by bedside ultrasound, *Ultraschall Med.* 41 (2020) 300–307.
- [11] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli, E. Peschiera, R. Trevisan, G. Maschietto, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, P. Rota, A. Passerini, R.J.G. van Sloun, E. Ricci, L. Demi, Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound, *IEEE Trans. Med. Imaging* 39 (8) (2020) 2676–2687.
- [12] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D.F. Briganti, S. Perlini, E. Torri, A. Mariani, E.E. Mossolani, F. Tursi, F. Mento, L. Demi, Is there a role for lung ultrasound during the COVID-19 pandemic? *J. Ultrasound Med.* 39 (2020) 1459–1462.
- [13] A.C.H. Yu, L. Demi, M. Muller, Q. Zhou, Ultrasound imaging: a silent hero in COVID-19 and lung diagnostics, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 67 (11) (2020) 2194–2196.
- [14] B. Zhou, X. Yang, X. Zhang, W.J. Curran, T. Liu, Ultrasound elastography for lung disease assessment, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 67 (11) (2020) 2249–2257.
- [15] G.T. Yusuf, A. Wong, D. Rao, A. Tee, C. Fang, P.S. Sidhu, The use of contrast-enhanced ultrasound in COVID-19 lung imaging, *J. Ultrasound* (2020) 1–5.
- [16] H. Gutsche, T.G. Lesser, F. Wolfram, T. Doenst, Significance of lung ultrasound in patients with suspected COVID-19 infection at hospital admission, *Diagnostics* 11 (6) (2021) 921.
- [17] F. Mento, T. Perrone, V.N. Macioce, F. Tursi, D. Buonsenso, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, On the impact of different lung ultrasound imaging protocols in the evaluation of patients affected by coronavirus disease 2019 how many acquisitions are needed? *J. Ultrasound Med.* 40 (10) (2021) 2235–2238.
- [18] K.L. Anderson, J.M. Fields, N.L. Panebianco, K.Y. Jenq, J. Marin, A.J. Dean, Inter-rater reliability of quantifying pleural B-lines using multiple counting methods, *J. Ultrasound Med.* 32 (2013) 115–120.

- [19] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D.F. Briganti, S. Perlini, E. Torri, A. Mariani, E.E. Mossolani, F. Tursi, F. Mento, L. Demi, Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, Reproducible Method, *J. Ultrasound Med.* 39 (2020) 1413–1419.
- [20] S.-T. Zhu, F.-Y. Tao, J.-H. Xu, S.-S. Liao, C.-L. Shen, Z.-H. Liang, B.-B. Shi, Q. Li, Utility of point-of-care lung ultrasound for clinical classification of COVID-19, *Ultrasound Med. Biol.* 47 (2) (2021) 214–221.
- [21] L. Zhao, K. Yu, Q.i. Zhao, R. Tian, H. Xie, L. Xie, P. Deng, G. Xie, A. Bao, J. Du, Lung ultrasound score in evaluating the severity of coronavirus disease 2019 (COVID-19) pneumonia, *Ultrasound Med. Biol.* 46 (11) (2020) 2938–2944.
- [22] W. Li, M. Fu, C. Qian, X. Liu, L. Zeng, X. Peng, Y. Hong, H. Zhou, L.i. Yuan, Quantitative assessment of COVID-19 pneumonia in neonates using lung ultrasound score, *Pediatr. Pulmonol.* 56 (6) (2021) 1419–1426.
- [23] L. Gargani, M. Doveri, L. D'Errico, F. Frassi, M.L. Bazzichi, A.D. Sedie, M.C. Scali, S. Monti, S. Mondillo, S. Bombardieri, D. Caramella, E. Picano, Ultrasound lung comets in systemic sclerosis: a chest sonography hallmark of pulmonary interstitial fibrosis, *Rheumatology* 48 (11) (2009) 1382–1387.
- [24] H. Li, Y.D. Li, W.W. Zhu, L.Y. Kong, X.G. Ye, Q.Z. Cai, L.L. Sun, X.Z. Lu, A simplified ultrasound comet tail grading scoring to assess pulmonary congestion in patients with heart failure, *Biomed Res. Int.* 2018 (2018) 1–10.
- [25] H.-F. Zong, G. Guo, J. Liu, L.-L. Bao, C.-Z. Yang, Using lung ultrasound to quantitatively evaluate pulmonary water content, *Pediatr. Pulmonol.* 55 (3) (2020) 729–739.
- [26] C. Brusasco, G. Santori, E. Bruzzo, R. Tro, C. Robba, G. Tavazzi, F. Guarracino, F. Forfori, P. Boccacci, F. Corradi, Quantitative lung ultrasonography: a putative new algorithm for automatic detection and quantification of B-lines, *Crit. Care* 23 (2019) 288.
- [27] L.J. Brattain, B.A. Telfer, A.S. Liteplo, V.E. Noble, Automated B-line scoring on thoracic sonography, *J. Ultrasound Med.* 32 (2013) 2185–2190.
- [28] R.J.G. van Sloun, L. Demi, Localizing B-lines in lung ultrasonography by weakly supervised deep learning, in-vivo results, *IEEE J Biomed. Health Inform.* 24 (4) (2020) 957–964.
- [29] J. Chen, C. He, J. Yin, J. Li, X. Duan, Y. Cao, L.i. Sun, M. Hu, W. Li, Q. Li, Quantitative analysis and automated lung ultrasound scoring for evaluating COVID-19 pneumonia with neural networks, *IEEE Trans Ultrason Ferroelectr Freq Control* 68 (7) (2021) 2507–2515.
- [30] Y. Wang, Y. Zhang, Q. He, H. Liao, J. Luo, Quantitative analysis of pleural line and B-lines in lung ultrasound images for severity assessment of COVID-19 pneumonia, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 69 (1) (2022) 73–83.
- [31] W. Xue, C. Cao, J. Liu, Y. Duan, H. Cao, J. Wang, X. Tao, Z. Chen, M. Wu, J. Zhang, H. Sun, Y. Jin, X. Yang, R. Huang, F. Xiang, Y. Song, M. You, W. Zhang, L. Jiang, Z. Zhang, S. Kong, Y. Tian, L.i. Zhang, D. Ni, M. Xie, Modality alignment contrastive learning for severity assessment of COVID-19 from lung ultrasound and clinical information, *Med. Image Anal.* 69 (2021) 101975, <https://doi.org/10.1016/j.media.2021.101975>.
- [32] A. Soummer, S. Perbet, C. Arbelot, J.E. Bazin, J.M. Constantin, Q. Lu, J.J. Rouby, Ultrasound assessment of lung aeration loss during a successful weaning trial predicts postextubation distress, *Intensive Care Med.* 37 (2011) S109–S209.
- [33] R. Kruisselbrink, V. Chan, G.A. Cibinel, S. Abrahamson, A. Goffi, I-AIM (Indication, Acquisition, interpretation, medical decision-making) framework for point of care lung ultrasound, *Anesthesiology* 127 (3) (2017) 568–582.
- [34] Z. Ren, B. Wu, X. Zhang, Q. Sun, Image set classification using candidate sets selection and improved reverse training, *Neurocomputing* 341 (2019) 60–69.
- [35] Y. Nam, C. Lee, Cascaded convolutional neural network architecture for speech emotion recognition in noisy conditions, *Sensors* 21 (13) (2021) 4399.
- [36] He KM, Zhang XY, Ren SQ, Sun J, Ieee. Deep Residual Learning for Image Recognition. 2016 *Ieee Conference on Computer Vision and Pattern Recognition*. New York: Ieee, 2016. 770-78.
- [37] S. Jayalakshmy, G.F. Sudha, Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks, *Artif. Intell. Med.* 103 (2020) 10.
- [38] J. Luo, W.W. Shi, N. Lu, J. Wang, H. Chen, Y.J. Wang, X.F. Lu, X.F. Wang, X.H. Hei, Improving the performance of multisubject motor imagery-based BCIs using twin cascaded softmax CNNs, *J. Neural Eng.* 18 (2021) 13.
- [39] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint 2014; arXiv: 1409.1556.
- [40] Szegegy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015; 1-9.
- [41] S. Gehlot, A. Gupta, R. Gupta, SDCT-AuxNet $\theta$ : DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis, *Med. Image Anal.* 61 (2020), 101661.
- [42] Jianbo Mao, Yuhao Luo, Lei Liu, Jimeng Lao, Yirun Shao, Min Zhang, Caiyun Zhang, Mingzhai Sun, Lijun Shen, Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks, *Acta Ophthalmol.* 98 (3) (2020), <https://doi.org/10.1111/aos.v98.3.10.1111/aos.14264>.
- [43] Qing Deng, Yao Zhang, Hao Wang, Liao Chen, Zhaohui Yang, Zhoufeng Peng, Ya Liu, Chuangli Feng, Xin Huang, Nan Jiang, Yijia Wang, Juan Guo, Bin Sun, Qing Zhou, Semiquantitative lung ultrasound scores in the evaluation and follow-up of critically ill patients with COVID-19: a single-center study, *Acad. Radiol.* 27 (10) (2020) 1363–1372.
- [44] L. Demi, F. Mento, A.D. Sabatino, A. Fiengo, U. Sabatini, V.N. Macioce, M. Robol, F. Tursi, C. Sofia, C.D. Cienzo, A. Smargiassi, R. Inchingolo, T. Perrone, Lung ultrasound in COVID-19 and post-COVID-19 patients, an evidence-based approach, *J. Ultrasound Med.* (2021).
- [45] Federico Mento, Tiziano Perrone, Veronica Narvena Macioce, Francesco Tursi, Danilo Buonsenso, Elena Torri, Andrea Smargiassi, Riccardo Inchingolo, Gino Soldati, Libertario Demi, On the impact of different lung ultrasound imaging protocols in the evaluation of patients affected by coronavirus disease 2019 how many acquisitions are needed? *J. Ultrasound Med.* 40 (10) (2021) 2235–2238.