Performance of a computer aided diagnosis system for SARS-CoV-2 pneumonia based on ultrasound images

Shiyao Shang [a,1], Chunwang Huang [a,1], Wenxiao Yan [a,1], Rumin Chen [a], Jinglin Cao [a], Yukun Zhang [a], Yanhui Guo [b,*], Guoqing Du [a,*]

[a] Department of Ultrasound, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China
[b] Department of Computer Science, University of Illinois Springfield, Springfield, IL USA

ARTICLE INFO

ABSTRACT

Purpose: In this study we aimed to leverage deep learning to develop a computer aided diagnosis (CAD) system toward helping radiologists in the diagnosis of SARS-CoV-2 virus syndrome on Lung ultrasonography (LUS).
Method: A CAD system is developed based on a transfer learning of a residual network (ResNet) to extract features on LUS and help radiologists to distinguish SARS-CoV-2 virus syndrome from healthy and non-SARS-CoV-2 pneumonia. A publicly available LUS dataset for SARS-CoV-2 virus syndrome consisting of 3909 images has been employed. Six radiologists with different experiences participated in the experiment. A comprehensive LUS data set was constructed and employed to train and verify the proposed method. Several metrics such as accuracy, recall, precision, and F1-score, are used to evaluate the performance of the proposed CAD approach. The performances of the radiologists with and without the help of CAD are also evaluated quantitively. The p-values of the t-test shows that with the help of the CAD system, both junior and senior radiologists significantly improve their diagnosis performance on both balanced and unbalanced datasets.
Results: Experimental results indicate the proposed CAD approach and the machine features from it can significantly improve the radiologists' performance in the SARS-CoV-2 virus syndrome diagnosis. With the help of the proposed CAD system, the junior and senior radiologists achieved F1-score values of 91.33% and 95.79% on balanced dataset and 94.20% and 96.43% on unbalanced dataset. The proposed approach is verified on an independent test dataset and reports promising performance.
Conclusions: The proposed CAD system reports promising performance in facilitating radiologists' diagnosis SARS-CoV-2 virus syndrome and might assist the development of a fast, accessible screening method for pulmonary diseases.

1. Introduction

SARS-CoV-2 virus syndrome pandemic has struck about 255 million people and is responsible for more than five million deaths [1]. It has led to unprecedented pressure on global healthcare services. A rapid and reliable diagnostic method for SARS-CoV-2 virus syndrome is a prerequisite for infectious disease control including isolation and treatment.

In the early stage of SARS-CoV-2 virus syndrome, reverse transcriptase polymerase chain reaction (RT-PCR) probably shows negative results. The false-negative rate decreases from 100% on day 1 to 20% on day 8 since exposure and symptom onset [2]. Imaging is an important supplementary method to indicate further RT-PCR testing among suspected SARS-CoV-2 virus syndrome patients. It may be particularly useful in Emergency Department for advanced patients triage and faster decision making while waiting for the RT-PCR results.

High-resolution computed tomography (HR-CT) in thoracic imaging was accepted as the gold standard imaging method of SARS-CoV-2 virus syndrome evaluation, on account of its high reliability [3]. However, for those critical patients in intensive care units (ICU) who are difficult to be

transported, or for patients in low- and middle-income countries where a CT exam is unavailable, an alternative imaging technique is required. Furthermore, the high-dose radiation exposure of CT is also an important consideration for patients who need imaging follow-up repeatedly, children and pregnant women. Chest X-ray can be obtained easily with minimal dose radiation exposure. However, studies showed that the chest X-ray exam has a relatively low sensitivity and a considerable proportion of the SARS-CoV-2 virus syndrome patients have been misdiagnosed by chest X-ray [4,5]. HR-CT and chest X-ray have limitations for SARS-CoV-2 virus syndrome evaluation, and supplementary method is a pressing demand in clinical practice.

Lung ultrasonography (LUS) has been used in different lung disease diagnoses for decades, such as interstitial lung disease, subpleural consolidations, and acute respiratory distress syndrome. A meta-analysis reported that the sensitivity and specificity for pneumonia diagnosis using LUS were as high as 94% (95% CI, 92–96%) and 96% (94–97%) [6]. During the progression of SARS-CoV-2 virus syndrome, the distal regions of the lung tended to be more frequently involved where can be detected easily and accurately by LUS [7]. The sonographic appearances were characterized with the pleural line irregularities and thickening, increment of B-line artifacts with severity and less frequently pleural effusion, and subpleural consolidations [8,9].

Numerous studies investigated the diagnostic value of LUS in SARS-CoV-2 virus syndrome. Researchers have found a strong correlation between CT and LUS in detecting lung lesions of SARS-CoV-2 virus syndrome [10,11]. In those studies, LUS adequately detected lung lesions in patients with SARS-CoV-2 virus syndrome. Compared with the chest X-ray, LUS provided a higher sensitivity in the diagnosis of SARS-CoV-2 virus syndrome (88.9% vs. 51.9%) [5]. A recent systematic review of 51 studies with 10,155 patients with SARS-CoV-2 infection evaluated the diagnostic value of thoracic CT, chest X-ray and ultrasound. The pooled sensitivities of thoracic CT, chest X-ray and LUS were 87.9%,80.6% and 86.4%, while the pooled specificities were 80.0%,71.5% and 54.6%, respectively. These findings demonstrate that chest CT, chest X-ray and ultrasound all have moderate sensitivity in diagnosing SARS-CoV-2 infection, with lower specificity for lung ultrasound and lower sensitivity for chest X-ray [12].

Since the ACR did not recommend the routine use of evaluating patients with suspected SARS-CoV-2 infection by CT [13], LUS is a promising supplementary method for SARS-CoV-2 virus syndrome evaluation with a lot of advantages: portable, nonradiative, ability on beside examination, minimizing the risk of occupational exposure and cross infection. It is particularly useful in ICU and Emergency Department for monitoring and triage of SARS-CoV-2 infection. Nevertheless, ultrasound quality is operator-dependent, and limited experience may lead to difficulties in utilizing this technique. Due to the massive increase in the patient population, healthcare professionals are thrown into a very high workload. Many physicians from different specialities are redeployed to treat patients with suspected or confirmed SARS-CoV-2 infection. Identifying and diagnosing SARS-CoV-2 infection with ultrasound data are challenging and time-consuming for novice utilizers. Computer aided diagnosis (CAD) system based on artificial intelligence (AI) provides an efficient way to assist physicians in disease diagnosis. CAD system is a promising tool to assist physicians in clinical decisions with higher accuracy, less complexity, and less time consumption.

Our study on CAD of SARS-CoV-2 virus syndrome is different from previous works. We not only simply propose a CAD system for LUS images and justify its performance on different datasets, but also verifies its efficiency on facilitating and improving the performance of physicians with different level of experiences. The new contribution of this study includes two folders: (1) a novel CAD system for SARS-CoV-2 virus syndrome diagnosis is proposed using a transfer learning on a deep residual convolutional neural network (CNN); (2) clinical experiments are studied to justify the performance improvement of radiologists with CAD's help.

2. Materials and methods

2.1. Materials

Images are from a public dataset whose detailed information can be found at https://github.com/jannisborn/covid19_ultrasound/blob/master/data/README.md. There are 262 patients include 92 SARS-CoV-2 virus syndrome, 79 non-SARS-CoV-2 pneumonia and 90 healthy cases. All ultrasound videos are split into images at a frame rate of 3 Hz. A total of 3909 images were obtained from GitHub and the public dataset [14]. All samples have been categorized into three classes using the final diagnosis results, including 1664 SARS-CoV-2 virus syndrome images, 876 non-SARS-CoV-2 pneumonia images, and 1369 health images. The final diagnosis results were confirmed by nasopharyngeal RT-PCR for SARS-Cov2.

2.2. Proposed methods

In our study, a transfer learning based residual network was developed to extract features on LUS and help radiologists with SARS-CoV-2 virus syndrome diagnosis. The proposed methods are introduced as follows.

2.2.1. Proposed transfer learning network

Transfer learning is able to reduce the time in training and improves the network's generalization ability. In this stage, rather than building a model from scratch, a pre-trained ResNet50 network, which was trained using ImageNet, is selected as a backbone to extract the features from LUS images. Inside it, a fully connected layer is modified to match the output numbers of classified categories, and a binary cross-entropy function is used as the loss function which computes the binary cross-entropy (BCE) between predictions and targets [15]. Compared with the mean square error (MSE) loss function, BCE can slow down the gradient dispersion and speed up the training process. The architecture of the pre-trained ResNet is modified by adding a fully connected layer for feature extraction and adding a global average pooling to interpret these features in the classification task. The idea is to generate one feature map for each corresponding category of the classification task in the last convolutional layer. The global average pooling is more native to the convolution structure by enforcing correspondences between feature maps and categories. Thus, the feature maps can be easily interpreted as categories confidence maps. Also, there is no parameter to optimize in the global average pooling thus overfitting is avoided.

2.2.2. Gradient-weighted class activation mapping

The proposed CAD system aims to provide both the classification results and the salient features on LUS images to help radiologists. The gradient-weighted class activation mapping (Grad-CAM) method creates an activation map that highlights the crucial areas [16]. In the Grad-CAM method, the gradients of the layers flowing into the final convolutional layer produce a rough localization map in which the important areas are highlighted. Grad-CAM uses the gradient information flowing into the last convolutional layer to assign significance values to each neuron which respondents to class-specific information in the image.

In the proposed CAD system, the Grad-CAM approach was employed to enhance the visuality by focusing attention on the critical lesion regions on US images. These regions where the CNN model was activated were colored from the lowest activation to the highest activation. The top three high activations are marked using three circles. The red circle represents the highest activation and the pink one is the second highest activation.

2.3. Platform settings

The modified ResNet50 is trained on a server with a 2 × Six-Core

Intel Xeon processor and 128 GB of memory. The server is equipped with an NVIDIA Tesla K40 GPU with 12 GB of memory.

To verify the efficiency to help radiologists, a web-based graphic user interface (WGUI) is designed to provide a platform for radiologists to perform diagnoses on LUS images. This WGUI randomly selects images from the dataset, displays them in front of radiologists, and provides diagnosis options for radiologists to select. It also helps radiologists by providing the diagnosis results of CAD, displaying color Grad-CAM maps, and marking salient feature reference regions. An example is shown in Fig. 1 in demonstrating the diagnosis procedure with the WGUI.

2.4. Image dataset settings

In the LUS dataset, to evaluate the classification performance on ratios of different categories, images have been selected twice to construct a balanced dataset and an unbalanced dataset, respectively.

In the balanced dataset where each category was represented by the same number of images, 2628 images were selected randomly from the whole dataset. In the unbalanced dataset, all 3909 images were selected from the original dataset. The dataset was split into the training, validation, and testing sets (Table 1).
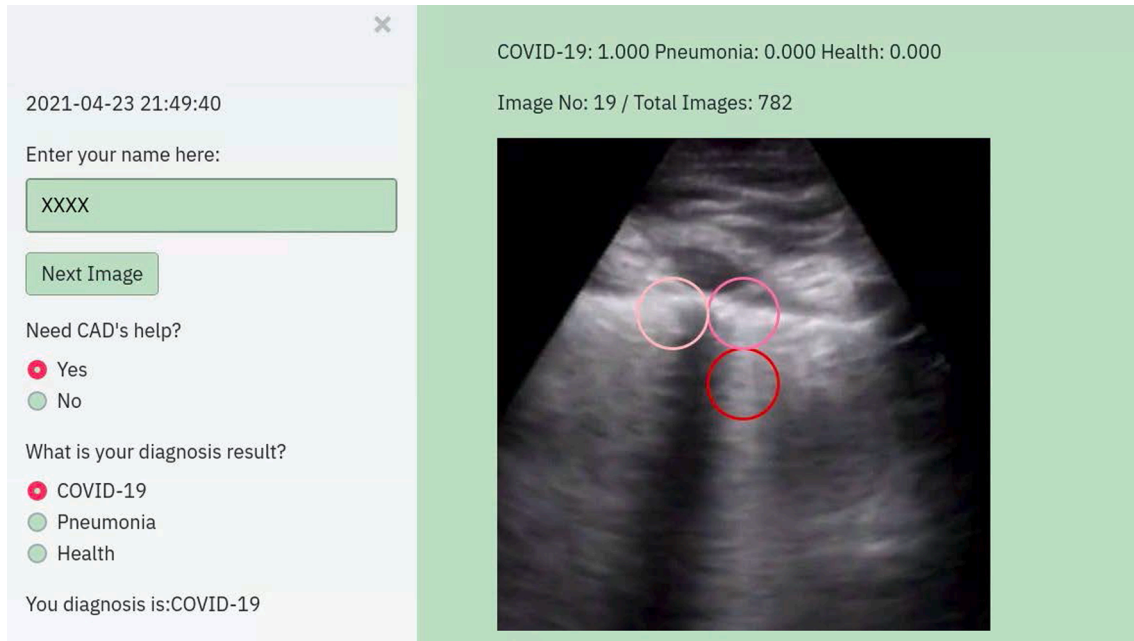
(a)
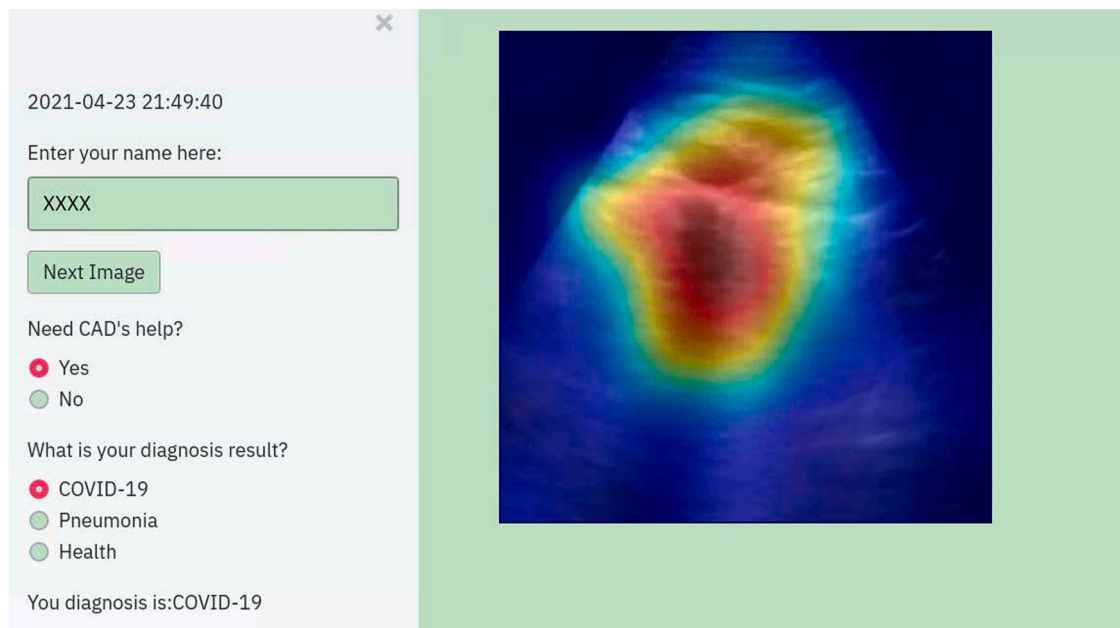
(b)

Fig. 1. An example of the diagnosis procedure with the WGUI. (a) Diagnosis interface with the diagnosis results from the CAD. (b) Diagnosis interface with Gad-CAM map.

Table 1
Data-split table.

Group	Balanced dataset			Unbalanced dataset		
Categories	S	NS	H	S	NS	H
	876	876	876	1664	876	1369
Training	1840			2736		
Validation	263			391		
Testing	525			782		
Total	2628			3909		

2.5. Evaluation metrics

To verify the performance of the transfer learning model, two commonly used machine learning algorithms have been employed to compare with the proposed model on the same dataset. In machine learning, the support vector machine (SVMs) [17] is one of the most robust supervised learning models for classification and regression analysis. An SVM maps training examples to points in space to maximize the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. The K-nearest neighbors (KNN) algorithm is a type of instance-based classification method where an unknown object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its K nearest neighbors. In the parameters of KNN, 5 neighbors are selected, Euclidean distance is the distance metric, and all features are standardized in the range of [0, 1].

A confusion matrix is a table used to evaluate classification performance. Each row of the confusion matrix represents the instances of an actual class and each column represents the instances of a predicted class. In the figures of the confusion matrix, the rows correspond to the predicted class (Output Class) and the columns correspond to the true class (Target Class). The diagonal elements correspond to correctly classified observations, and the off-diagonal cells correspond to

Table 2
Comparison of confusion matrices between different machine learning methods and radiologists on the balanced dataset.

Method	Total accuracy	Average precision	Average recall	Average F1-score	Confusion Matrix				
					Predicted				
					S	NS	H		
Proposed CAD	99.05%	99.05%	99.07%	99.05%	175	0	0	S	
					0	170	0	NS	
					0	5	175	H	
SVM	98.48%	98.48%	98.54%	98.48%	172	0	0	S	
					0	170	0	NS	
					3	5	175	H	
KNN	98.86%	98.86%	98.88%	98.86%	174	0	0	S	
					0	170	0	NS	
					1	5	175	H	
JR	73.59%	73.59%	74.69%	72.60%	415	34	184	S	Actual
					55	481	78	NS	
					55	10	263	H	
JR with CAD	93.14%	93.14%	93.24%	93.17%	490	14	44	S	
					7	500	4	NS	
					28	11	477	H	
SR	77.84%	77.84%	79.32%	77.87%	453	79	136	S	
					23	423	39	NS	
					49	23	350	H	
SR with CAD	95.49%	95.49%	95.59%	95.51%	500	2	17	S	
					6	498	2	NS	
					19	25	506	H	

Abbreviations: S: SARS-CoV-2 virus syndrome, NS: Non-SARS-CoV-2 pneumonia, H: Healthy, JR: Junior radiologists, SR: Senior radiologists.

incorrectly classified observations. The numbers at the rightest column are the percentages of all the examples predicted to belong to each class that is correctly and incorrectly classified. The number at the bottom row are the percentages of all the examples belonging to each class that is correctly and incorrectly classified. Four parameters namely accuracy, precision, recall and F1-score were calculated and compared.

In addition, a statistic testing, a pair *t*-test is used to evaluate the difference between radiologists' performances without CAD help and with CAD help. The p-values of the *t*-test show if there are significant improvements on diagnosis performances.

3. Results

We employed two groups of radiologists with different levels of experience to perform diagnoses with and without the help of the proposed CAD system on the platform. Among them, three junior radiologists had an average age of 26 years with less than three years' experience in LUS diagnosis and three senior radiologists had an average age of 35 years with more than eight years' experience in LUS diagnosis.

3.1. Evaluation results on the balanced dataset

In the balanced dataset where each category was represented by the same number of images, 2628 images were selected randomly from the whole dataset. In the balanced dataset, 70% of the patch images were used for training, 10% for validation, and 20% for testing. 1840 images were in the training set, 263 in the validation set, and 525 in the testing set.

Table 2 and Fig. 2 show the confusion matrices on the proposed CAD approach, SVM, and KNN methods. In the confusion matrix figures, the grey scales in each cell visualize the values in the cells. The higher, the larger the percentage values are. In the classification results of the proposed CAD, all samples are classified correctly and more accurately than other methods. The values of Accuracy, Recall, Precision, and F1-score are compared in Table 2. These results also indicate improvement of our proposed model over the SVM and KNN models in the evaluation metrics.

The evaluation results of the radiologists' performance on balanced dataset are summarized in Table 2 and Fig. 3.Without the help of our proposal CAD system, only 79.0% and 86.3% SARS-CoV-2 virus syndrome were identified correctly by junior and senior radiologists, and this proportion increased to 93.3% and 95.2%, respectively with the help of our proposal CAD system. After using our proposal CAD system, both junior and senior radiologists achieved higher accuracy, precision, recall and F1-scores values.

The details about the evaluation results of different models and radiologists on balanced dataset are listed in Table E1, E2 (supplementary material). The proposed CAD system achieves precision, recall, and F1-score values of all 100% for diagnosing SARS-CoV-2 virus syndrome, 97.14%, 100%, and 98.55% for non-SARS-CoV-2 pneumonia, and 100%, 97.22%, and 98.59% for healthy cases. With the assistance of the proposed CAD system, the precision, recall and F1 scores were higher for diagnosing SARS-CoV-2 virus syndrome, non-SARS-CoV-2 pneumonia, and healthy cases in both junior and senior radiologists' groups. Line chart for different models and radiologist in diagnosing with LUS on balanced dataset is presented in Fig. 6 a. Our CAD system outperforms other classifiers and increases the performance of both junior and radiologists.

3.2. Evaluation with the unbalanced dataset

In the unbalanced dataset, all 3909 images were selected from the original dataset. The same ratio (70%, 10%, 20%) was used to split the dataset into the training, validation, and testing sets. These splits were applied to each category. From the 3909 images, 2736 were selected for training, 391 for validation, and 782 for testing.
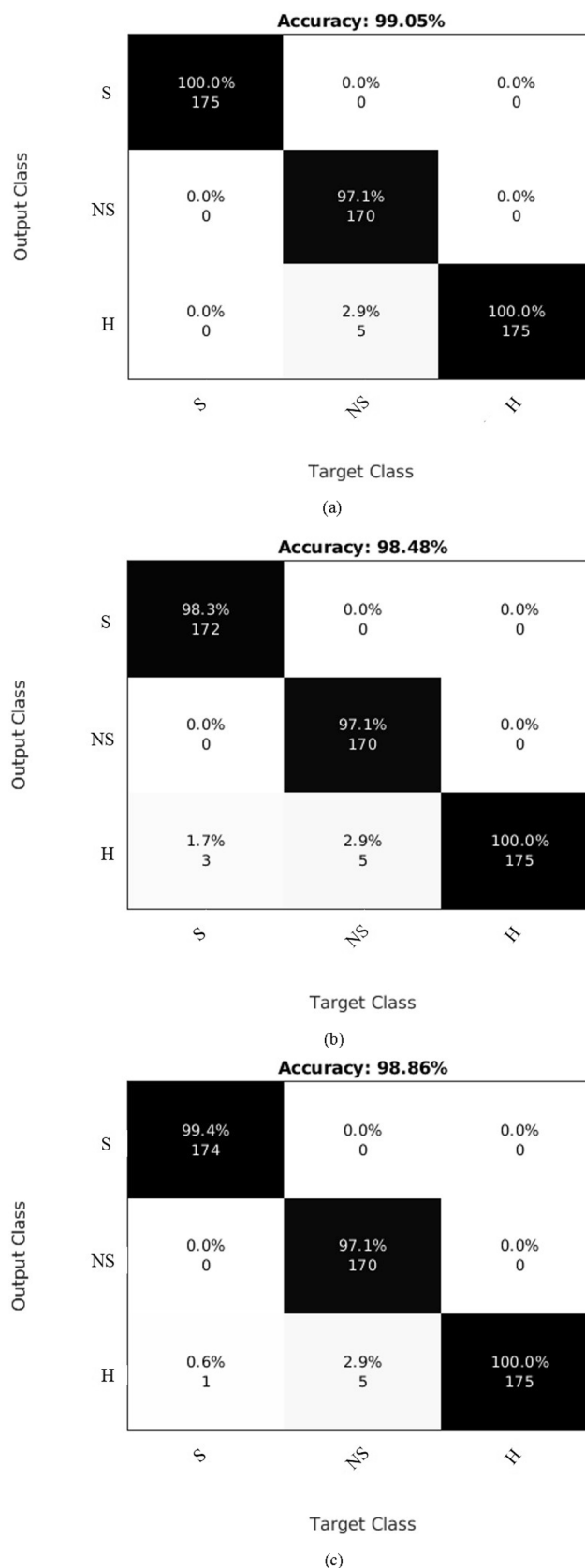

Fig. 2. Confusion matrix of different algorithms on balanced data set. (a) CAD. (b) SVM. (c) KNN. Abbreviations:S:SARS-CoV-2 virus syndrome, NS:Non-SARS-CoV-2 pneumonia, H:Healthy.
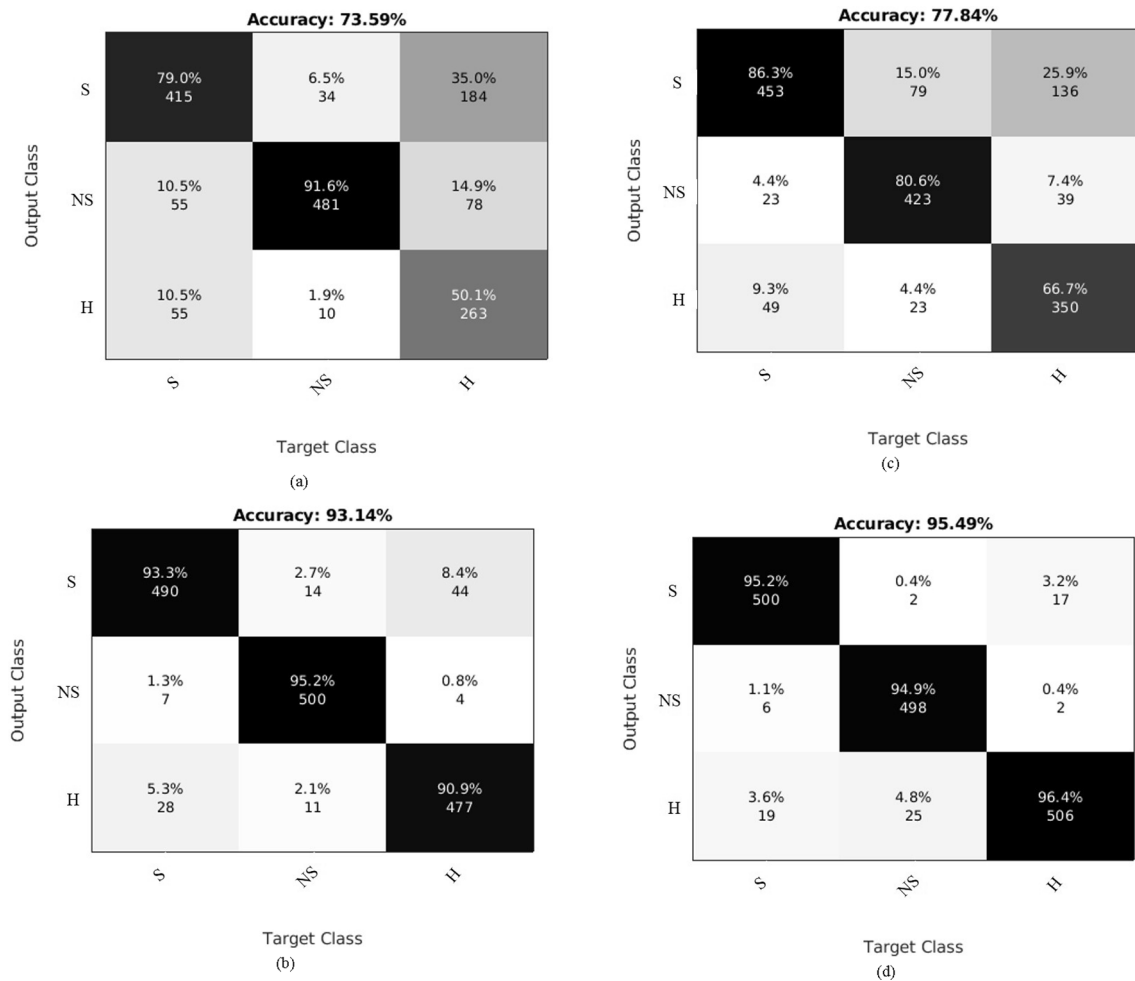
Fig. 3. Confusion matrix of different radiologists on balanced data set. (a) Junior radiologists. (b) Junior radiologists with CAD. (c) Senior radiologists. (b) Senior radiologists with CAD. Abbreviations:S:SARS-CoV-2 virus syndrome, NS:Non-SARS-CoV-2 pneumonia, H:Healthy.

Table 3 and Fig. 4 show the confusion matrices and results of the evaluation metrics between our proposed model and the SVM and KNN methods on the unbalanced dataset. Our model and other classifiers all achieve high performance metrics.

Table 3 and Fig. 5 show the confusion matrix and results of different radiologists' performance with and without the help of the proposed CAD system on unbalanced data. The details about the evaluation results of different models and radiologists on unbalanced dataset are listed in Table E3 and E4 (supplementary material). The proposed CAD system achieves precision, recall, and F1-score values of 98.20%,99.70% and 98.94 for diagnosing SARS-CoV-2 virus syndrome, 100%, 96.15%, and 98.04% for non-SARS-CoV-2 pneumonia, and 98.54%, 99.26%, and 98.90% for healthy cases. Again, the CAD system improved radiologists' performance on unbalanced dataset. Line chart for different models and radiologist in diagnosing with LUS on unbalanced dataset is presented in Fig. 6 b. Both groups of radiologists achieved higher precision, recall, and F1-score values with the help of the CAD system.

Statistical testing experiments were taken to evaluate the performance differences of radiologists with and without the help of the CAD system and test if there are significant improvements between them. Four comparisons were taken on junior and senior radiologists on different datasets. The p-values of the *t*-test in Table 4 are all less than 0.05 and show that with the help of the CAD system, both junior and senior radiologists significantly improve their diagnosis performance on both balanced and unbalanced datasets.

4. Discussion

Compared to the exploded deep learning studies of SARS-CoV-2 virus syndrome with thoracic CT or chest X-ray, relatively few researchers of the LUS CAD system can be found in the literature. Some work has been done on exploiting LUS image analysis and deep learning to detect and quantify B-line, extract pleural line, and subpleural lesions [18–20]. Born *et al.* trained a deep convolutional neural network on a lung ultrasound dataset consisted of 1103 images for SARS-CoV-2 virus syndrome detection, with a sensitivity of 0.96, a specificity of 0.79, and an F1-score of 0.92[21]. Roy *et al.* used a deep learning model to detect LUS imaging patterns associated with SARS-CoV-2 virus syndrome where the disease severity score was predicted, and the pathological artifacts of SARS-CoV-2 virus syndrome were located [22].

Our proposed CAD system using the transfer learning of ResNet is employed to identify SARS-CoV-2 virus syndrome on LUS images. The results demonstrate that the proposed system accomplished high classification precisions on both balanced and unbalanced dataset. The evaluation metrics results also justify the better performance of the proposed method than the existing classifiers. The experimental results on the ability to assist radiologists demonstrate that the CAD system can significantly improve the radiologists' performance on the SARS-CoV-2 virus syndrome diagnosis.

Six radiologists participated in this study and were divided into junior and senior groups according to their experience on LUS diagnosis. Radiologists inside each group have similar prior knowledge levels, which can prevent bias due to knowledge inequalities. The performance

Table 3

Comparison of confusion matrices between different machine learning methods and radiologists on the unbalanced dataset.

Method	Total accuracy	Average precision	Average recall	Average F1-score	Confusion Matrix				
					Predicted				
					S	NS	H		
Proposed CAD	98.72%	98.91%	98.37%	98.63%	327	0	1	S	
					4	175	3	NS	
					2	0	270	H	
SVM	98.72%	98.87%	98.55%	98.70%	326	0	1	S	
					3	174	1	NS	
					4	1	272	H	
KNN	98.98%	99.07%	98.72%	98.89%	328	0	1	S	
					4	174	1	NS	
					1	1	272	H	
JR	69.35%	71.22%	72.35%	68.39%	784	50	312	S	Actual
					165	474	141	NS	
					50	1	369	H	
JR with CAD	94.80%	95.04%	94.68%	94.85%	934	21	29	S	
					26	504	7	NS	
					39	0	786	H	
SR	74.42%	75.53%	75.10%	74.40%	788	59	241	S	
					125	453	76	NS	
					86	13	505	H	
SR with CAD	96.38%	96.92%	95.63%	96.17%	932	0	2	S	
					44	520	11	NS	
					23	5	809	H	

Abbreviations: S: SARS-CoV-2 virus syndrome, NS: Non-SARS-CoV-2 pneumonia, H: Healthy, JR: Junior radiologists, SR: Senior radiologists.

in the identification of lung lesions of senior radiologists was better than junior ones, however, the accuracy of both groups was not high. With the assistance of the proposed CAD system, junior radiologists achieved an accuracy similar to seniors, and both accuracies are higher than 90%. This finding provides evidence to help more physicians without much experience in LUS, especially those from the emergency room and ICU to efficiently utilize LUS for SARS-CoV-2 virus syndrome evaluation and follow-up. During the pandemic of SARS-CoV-2 virus syndrome, this finding has much practical significance.

In our experiment, it was interesting that the performance of radiologists with the assistance of the CAD system on the diagnosis of SARS-CoV-2 virus syndrome was still inferior to the proposed CAD system, which was contrary to the experiment designer's assumption. It might be due to the radiologists' lack of clinical experience in the new disease, and less confidence in CAD. In some cases, it was difficult for radiologists to distinguish mild lesions in SARS-CoV-2 virus syndrome and healthy lungs. Our CAD system can detect the subtle differences between them.

With the assistance of our proposed CAD system, physicians are released from complicated training and repetitive image reading tasks, and spare time for clinical decision making and treatment implementation.

Buonsenso et al. [23] introduced a specific LUS evaluation procedure for children with suspected SARS-CoV-2 virus syndrome. The clinical examination and lung imaging can be performed concomitantly by just two operators: one pediatrician and another assistant. It minimized the use of healthcare staff and medical image equipment, which reduce the risk of exposure of both clinicians and patients. Although further studies are required, we believe that with the help of our proposed CAD system, clinicians will make better use of LUS for SARS-CoV-2 virus syndrome evaluation and the novice clinicians on LUS will obtain high accuracy in identifying SARS-CoV-2 virus syndrome patients from healthy people.

A large quantity of research work on AI for SARS-CoV-2 virus syndrome has been published. Conversely, few of those systems were subsequently applied to the clinic. It is insufficient for clinical decision-making with the computer alone. Humans are the main subject of
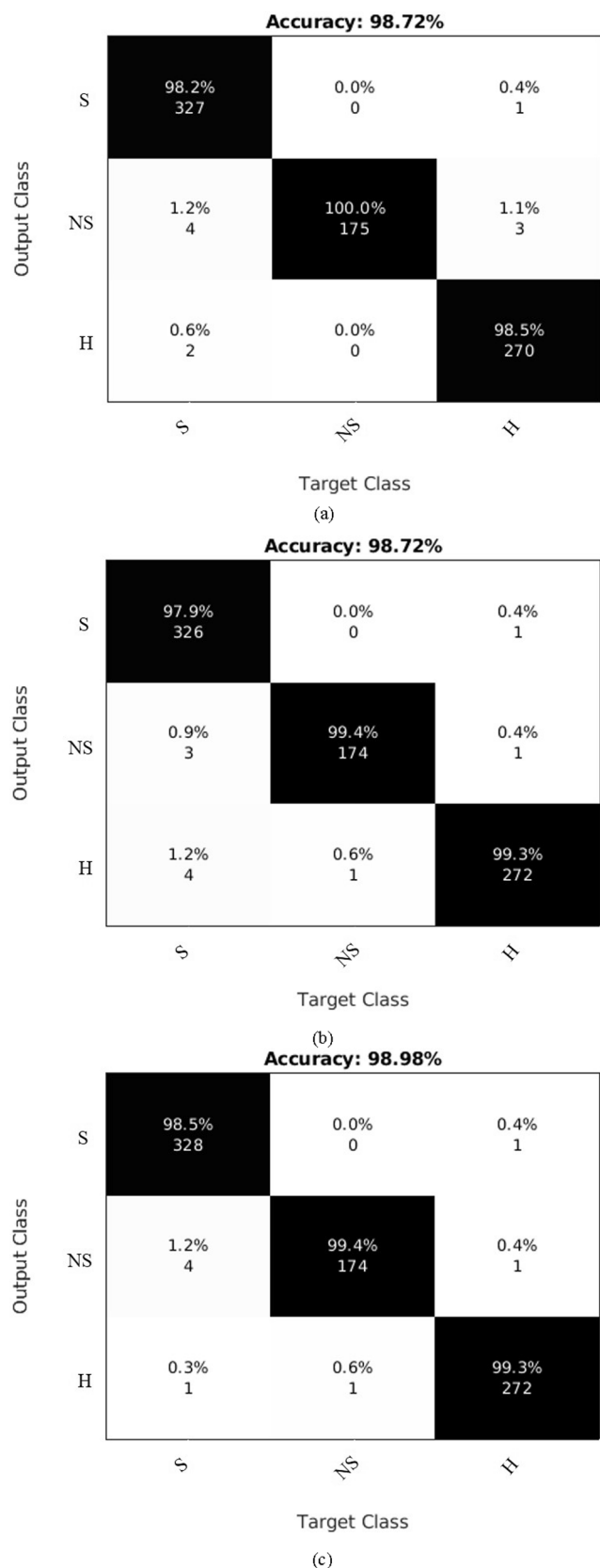
Fig. 4. Confusion matrix on the unbalanced data set. (a) Proposed CAD. (b) SVM. (c) KNN. Abbreviations:S:SARS-CoV-2 virus syndrome, NS:Non-SARS-CoV-2 pneumonia, H:Healthy.

medical practice. Explanatory techniques were used in our proposed CAD system to facilitate the utilization of it. The output of our CAD system was presented as diagnosing results with probability, highlighted critical lesion regions, and Grad-CAM maps. Clinicians can enter the field easily and give their verdicts according to the visualized diagnostic basis by the CAD system as well as their own experience. In our study, images have been selected twice to construct a balanced dataset and an unbalanced dataset to evaluate the classification performance on ratios of different categories. The training, validation and testing dataset were disjoint. Radiologists who participated in the tests were with different experience on LUS diagnosis, and there were weeks between two tests to washout radiologists' memory. The proposed system accomplished high classification precisions on both balanced and unbalanced dataset, and improved the performance of both senior and junior radiologists on different dataset. It demonstrates that our proposed CAD system has good reproducibility and generalization.

Another point worth considering is whether the CAD system is too sensitive for SARS-CoV-2 virus syndrome detection. Since most of the patients infected showed mild symptoms and can be monitored at home [24], it seems unnecessary to detect every minor lesion in SARS-CoV-2 virus syndrome. However, we still recommend the use of the CAD system. Firstly, because of the infectious nature of SARS-CoV-2 virus syndrome, it is very important to identify and isolate the infected patients. Moreover, some of the patients with moderate symptoms also have the possibility of progression to the severe situation. A fast and accurate method is essential for disease evaluation.

Our study has some limitations. Firstly, it was technically impossible to blind the radiologists to the method because they had to know whether to make a diagnosis with or without the CAD system. The level of trust in the CAD system may influence the diagnosis. Even so, the performance of all the radiologists were significantly improved after using the CAD system. In this study, the images obtained from GitHub and the public dataset are inevitably heterogeneous. The ultrasound devices, gain, and imaging depths were diverse. Clinical information such as age, gender, and symptoms is unknown. Ultrasonography is operator-dependent, placing a barrier for the objective analysis. Because our CAD system demonstrated its accuracy in evaluating LUS data of SARS-CoV-2 virus syndrome with both balanced and unbalanced datasets, it's reasonable to presume that it is easily generalized to analyze other LUS datasets, and eventually provides clinicians actionable clues for therapy strategy.

Our proposed CAD system on LUS images improved the SARS-CoV-2 virus syndrome diagnosis by using the transfer learning of ResNet and help improving radiologists' performance. In the future, we plan to apply our proposed method to a large dataset with a greater diversity of classes and a more varied imaging environment. With the high generalization ability of our method, we expect that it can be easily extended to similar applications in different diseases with few adaptations.

Funding

CRediT authorship contribution statement

Shiyao Shang: Writing – original draft. **Chunwang Huang:** Resources, Visualization. **Wenxiao Yan:** Project administration. **Rumin Chen:** Validation. **Jinglin Cao:** Investigation. **Yukun Zhang:** Formal analysis. **Yanhui Guo:** Software, Data curation. **Guoqing Du:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence
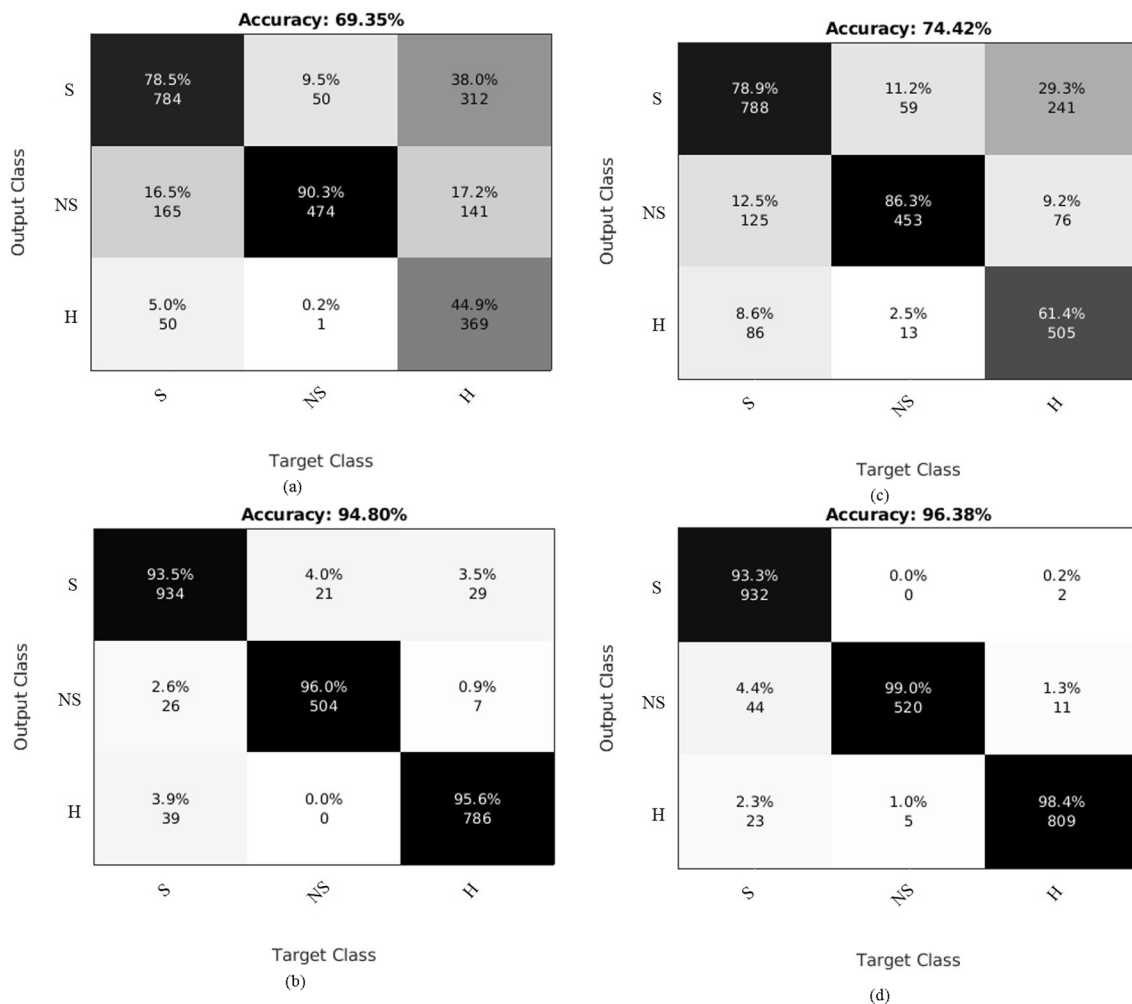
Fig. 5. Confusion matrix of different radiologists on the unbalanced data set. (a) Junior radiologists. (b) Junior radiologists with CAD. (c) Senior radiologists. (b) Senior radiologists with CAD. Abbreviations:S:SARS-CoV-2 virus syndrome, NS:Non-SARS-CoV-2 pneumonia, H:Healthy.

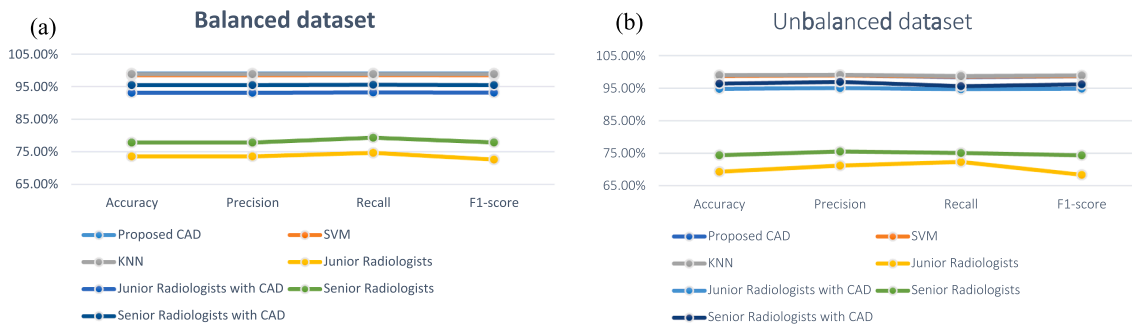**Fig. 6.** Evaluation results for different classifiers and radiologists (a) Balanced dataset. (b) Unbalanced dataset.

Table 4
t-test between performance of radiologists with and without CAD system.

	Comparison	p-value
Balanced dataset	JR vs JR + CAD	4.64×10^{-2}
	SR vs SR + CAD	1.63×10^{-12}
Unbalanced dataset	JR vs JR + CAD	2.0×10^{-2}
	SR vs SR + CAD	2.01×10^{-2}

Abbreviations; R: Junior radiologists, SR: Senior radiologists.

the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2021.110066.

References

[1] Available via https://coronavirus.jhu.edu/map.html (accessed on 18 Nov 2021).

[2] L.M. Kucirka, S.A. Lauer, O. Laeyendecker, D. Boon, J. Lessler, Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2

tests by time since exposure, Ann. Intern. Med. 173 (4) (2020) 262–267, https://doi.org/10.7326/M20-1495.

[3] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, Radiology (2020), https://doi.org/10.1148/radiol.2020200642.

[4] E.S. Kim, B.S. Chin, C.K. Kang, N.J. Kim, Y.M. Kang, J.P. Choi, D.H. Oh, J.H. Kim, B. Koh, S.E. Kim, N.R. Yun, J.H. Lee, J.Y. Kim, Y. Kim, J.H. Bang, K.H. Song, H. B. Kim, K.H. Chung, M.D.C. Oh, Korea national committee for clinical management of, clinical course and outcomes of patients with severe acute respiratory syndrome coronavirus 2 infection: a preliminary report of the first 28 patients from the Korean cohort study on COVID-19, J. Korean Med. Sci. 35 (13) (2020), https://doi.org/10.3346/jkms.2020.35.e142.

[5] J.R. Pare, I. Camelo, K.C. Mayo, M.M. Leo, E.M. Schechter-Perkins, Point-of-care lung ultrasound is more sensitive than chest radiograph for evaluation of COVID-19, West. J. Emerg. Med. 21 (4) (2020) 771–778, https://doi.org/10.5811/westjem.2020.5.47743.

[6] M.A. Chavez, N. Shams, L.E. Ellington, N. Naithani, R.H. Gilman, M.C. Steinhoff, M. Santosham, R.E. Black, C. Price, M. Gross, W. Checkley, Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis, Respir. Res. 15 (1) (2014), https://doi.org/10.1186/1465-9921-15-50.

[7] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, L. Demi, Is there a role for lung ultrasound during the COVID-19 pandemic? Lung Ultrasound (LUS) findings in Covid-19, J. Ultrasound Med.: Off. J. Am. Inst. Ultras. Med. 39 (7) (2020) 1459–1462, https://doi.org/10.1002/jum.15284.

[8] A.D. Kumar, S.C. Hung, Y. Duanmu, S. Graglia, J. Kugler, Lung ultrasound findings in patients hospitalized with Covid-19 [published online ahead of print], J. Ultras. Med. (2021), https://doi.org/10.1002/jum.15683.

[9] C. Xing, Q. Li, H. Du, W. Kang, J. Lian, L. Yuan, Lung ultrasound findings in patients with COVID-19 pneumonia, Critical Care 24 (1) (2020) 174, https://doi.org/10.1186/s13054-020-02876-9.

[10] E. Poggiali, A. Dacrema, D. Bastoni, V. Tinelli, E. Demichele, P.M. Ramos, M. Silva, A. Vercelli, A. Magnacavallo, Can lung US help critical care clinicians in the early diagnosis of novel coronavirus (COVID-19) pneumonia? Radiology 295 (3) (2020) https://doi.org/10.1148/radiol.2020200847.

[11] K. Ökmen, D.K. Yıldız, E. Soyaslan, Comparison of lung ultrasonography findings with chest computed tomography results in coronavirus (COVID-19) pneumonia, J. Med. Ultrason. 48 (2) (2021) 245–252, https://doi.org/10.1007/s10396-021-01081-7.

[12] N. Islam, S. Ebrahimzadeh, J.P. Salameh, S. Kazi, N. Fabiano, L. Treanor, M. Absi, Z. Hallgrimson, M.M. Leeflang, L. Hooft, C.B. van der Pol, R. Prager, S.S. Hare, C. Dennie, R. Spijker, J.J. Deeks, J. Dinnes, K. Jenniskens, D.A. Korevaar, J. F. Cohen, A. Van den Bruel, Y. Takwoingi, J. van de Wijgert, J.A. Damen, J. Wang, M.D. McInnes, C.-D.T.A.G. Cochrane, Thoracic imaging tests for the diagnosis of

COVID-19, Cochrane Database Syst. Rev. 3 (2021) CD013639, https://doi.org/10.1002/14651858.CD013639.pub4.

[13] American College or Radiology. ACR Recommendations for the Use of Chest Radiography and Computed Tomography (ct) for Suspected Covid-19 Infection. Available via https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection.

[14] Available via https://github.com/jannisborn/covid19_ultrasound.

[15] A. Buja, W. Stuetzle, Y. Shen, Loss functions for binary class probability estimation and classification: structure and application, 2005, pp. 119. https://doi.org/10.1.1.184.5203.

[16] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, IEEE Int. Conf. Comput. Vision (2017) 618–626, https://doi.org/10.1109/ICCV.2017.74.

[17] C. Cortes, C. Cortes, V. Vapnik, C. Llorens, V.N. Vapnik, C. Cortes, M. Côrtes, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297, https://doi.org/10.1023/A:1022627411411.

[18] R. 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, https://doi.org/10.1109/JBHI.2019.2936151.

[19] 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, https://doi.org/10.1109/TUFFC.2020.3005512.

[20] Y. Xu, Y. Zhang, K. Bi, Z. Ning, L. Xu, M. Shen, G. Deng, Y. Wang, Boundary restored network for subpleural pulmonary lesion segmentation on ultrasound images at local and global scales, J. Digit. Imaging 33 (5) (2020) 1155–1166, https://doi.org/10.1007/s10278-020-00356-8.

[21] J. Born, G. Brndle, M. Cossio, M. Disdier, N. Wiedemann, POCOVID-Net: Automatic Detection of COVID-19 From a New Lung Ultrasound Imaging Dataset (POCUS), arXiv:2004.12084, 2021. (latest version).

[22] S. Roy, W. Menapace, S. Oei, B. Luijten, L. Demi, Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound, IEEE Trans. Med. Imaging 1 (99) (2020) 2676–2687, https://doi.org/10.1109/TMI.2020.2994459.

[23] D. Buonsenso, D. Pata, A. Chiaretti, COVID-19 outbreak: less stethoscope, more ultrasound, Lancet Respir. Med. 8 (5) (2020) e27, https://doi.org/10.1016/S2213-2600(20)30120-X.

[24] P.K. Samudrala, P. Kumar, K. Choudhary, N. Thakur, G.S. Wadekar, R. Dayaramani, M. Agrawal, A. Alexander, Virology, pathogenesis, diagnosis and in-line treatment of COVID-19, Eur. J. Pharmacol. 883 (2020) 173375, https://doi.org/10.1016/j.ejphar.2020.173375.