

RESEARCH



SpecMEn-DL: spectral mask enhancement with deep learning models to predict COVID-19 from lung ultrasound videos

Farhan Sadik, Ankan Ghosh Dastider and Shaikh Anowarul Fattah*

Abstract

Lung Ultrasound (LUS) images are considered to be effective for detecting Coronavirus Disease (COVID-19) as an alternative to the existing reverse transcription-polymerase chain reaction (RT-PCR)-based detection scheme. However, the recent literature exhibits a shortage of works dealing with LUS image-based COVID-19 detection. In this paper, a spectral mask enhancement (SpecMEn) scheme is introduced along with a histogram equalization pre-processing stage to reduce the noise effect in LUS images prior to utilizing them for feature extraction. In order to detect the COVID-19 cases, we propose to utilize the SpecMEn pre-processed LUS images in the deep learning (DL) models (namely the SpecMEn-DL method), which offers a better representation of some characteristics features in LUS images and results in very satisfactory classification performance. The performance of the proposed SpecMEn-DL technique is appraised by implementing some state-of-the-art DL models and comparing the results with related studies. It is found that the use of the SpecMEn scheme in DL techniques offers an average increase in accuracy and F_1 score of 11% and 11.75%, respectively, at the video-level. Comprehensive analysis and visualization of the intermediate steps manifest a very satisfactory detection performance creating a flexible and safe alternative option for the clinicians to get assistance while obtaining the immediate evaluation of the patients.

Keywords: COVID-19, Lung ultrasound, Image processing, Spectral mask, Disease classification

Introduction

In December 2019, the world discovered a novel type of coronavirus causing viral pneumonia outbreak which quickly reached the global stage, and the World Health Organization (WHO) declared this Coronavirus Disease (COVID-19) as a global pandemic [1]. The rapid spread of this disease has created a worldwide emptiness in the medical capacity, demanding an efficient complementary scheme to detect COVID-19 at the earliest period and thereby curtailing its spread. The reverse transcription-polymerase chain reaction (RT-PCR) test, the gold standard for detecting COVID-19, is of limited capacity, time-consuming, and strictly dependent on swab-collection techniques [2]. Complementary attempts are aimed at using computed tomography (CT) scan, X-Ray,

and Lung Ultrasound (LUS) images [3–5]. Considering the radiation hazards, cost, and flexibility, LUS is better than CT scans or X-Rays with even better performance in some cases than the others [6]. Therefore, introducing LUS imaging techniques in COVID-19 diagnosis by accurately separating it from pneumonia or regular healthy cases can be a vital step to fight the current pandemic by ensuring rapid care for the patients.

Most of the machine learning (ML)-based works on COVID-19 detection are devoted to analyze the LUS images through the classification into certain categories [7, 8], sometimes followed by a supervised or unsupervised segmentation step. Supervised segmentation needs properly annotated data, which is a mammoth task; and a publicly available annotated LUS dataset related to COVID-19 is quite inadequate. Video-based grading and frame-based disease severity score prediction are other ways to deal with the LUS images [9]. However, investigations on COVID-19 detection through LUS

*Correspondence: fattah@eee.buet.ac.bd

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh

images are somewhat limited compared to the studies on other relevant imaging-based diagnostic fields. In [7], a VGG16-based simple classification network namely POCOVID-Net is implemented with moderate classification performance. Here, the authors presented a collection of the open-source LUS dataset on COVID-19, namely POCUS dataset, which is getting richer day by day. In [10], the same research group demonstrated explainable LUS image analysis on raw recordings for accelerating COVID-19 differential diagnosis. However, these attempts are quite straightforward and leave scope for improvement in this field. In [8], a comparative analysis of COVID-19 detection performance is presented on CT scan, X-Ray, and ultrasound datasets, and it is concluded that the LUS images exhibit comparatively the best result in terms of disease detection accuracy. Here, the authors utilized the same dataset for LUS that was used in [7, 10]. It is to be noted that these studies were conducted considering only a selected portion of LUS images from the large dataset. An adroit classification scheme with efficacious detection performance under source-independent conditions is therefore much coveted to predict the disease class accurately and rapidly.

In this paper, an automatic scheme is proposed for classifying the LUS images into COVID-19, pneumonia, and regular/healthy categories. The main idea here is to develop an efficient LUS image enhancement scheme and

utilize the resulting enhanced LUS images in the deep learning (DL)-based classification networks for achieving better classification performance. The pipeline of the proposed method is presented in Fig. 1. For the purpose of LUS image enhancement, first, the contrast-limited adaptive histogram equalization (CLAHE) pre-processing is performed. A spectral mask enhancement (namely SpecMEn) scheme is proposed that generates a mask by utilizing the CLAHE enhanced images and the mask is then employed on each image plane to further reduce the effect of noise. The 3-channel pre-processed LUS images are headed towards the DL classification network. A thorough evaluation of the proposed scheme by both frame-level and video-level results, and comparison with related studies manifest its capability of enhancing the prediction performance of the classification networks.

Materials and methods

Dataset

In this paper, the point-of-care ultrasound (POCUS) dataset is utilized [7], which is publicly available and an open-sourced dataset. It comprises various types of videos from the COVID-19, pneumonia, and regular/healthy cases. The COVID-19 class includes some subclasses, such as pregnant cases, dialytic cases, and some unlabelled cases. Similarly, the pneumonia class includes viral pneumonia and unlabelled pneumonia cases. In

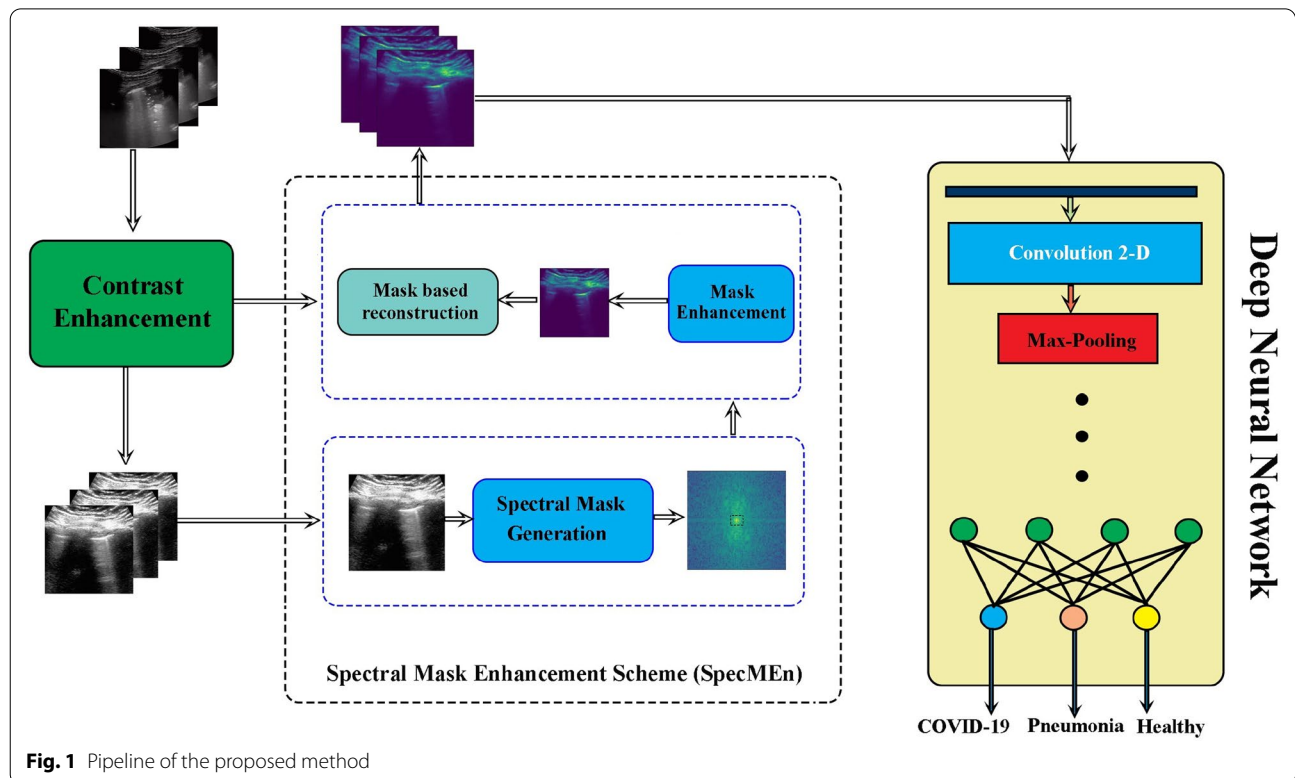


Fig. 1 Pipeline of the proposed method

this paper, all these three classes, namely COVID-19, pneumonia, and healthy cases are considered (excluding the unlabeled data). The dataset used in this study consists of 123 LUS videos. After extracting the frames from each video, a total of 41,529 images are obtained. Among them, 74 videos (60%) with 27,920 frames are placed in the training set, and the other 49 videos (40%) with 13,609 frames are placed in the testing set. The detailed distribution of these training and testing sets for each of the three classes is presented in Table 1.

Preprocessing

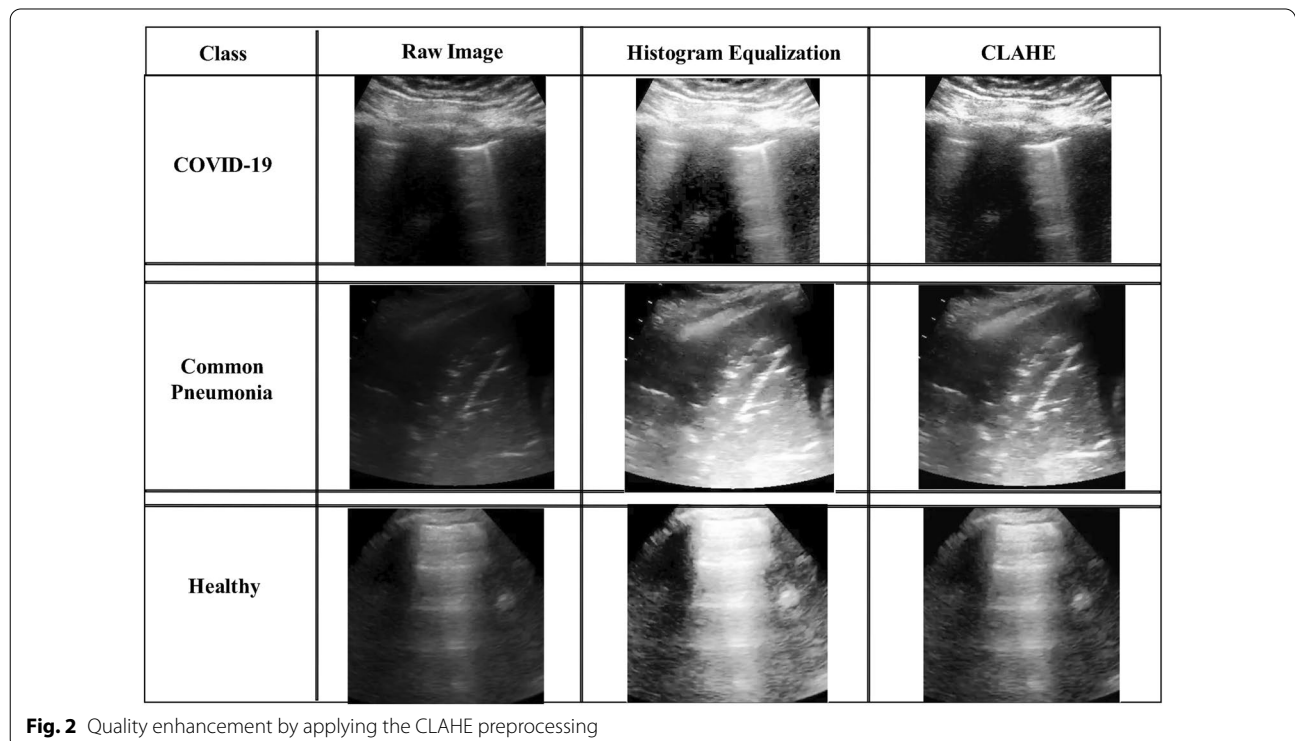
It is well known that noises can be introduced in ultrasound images during the data acquisition, transmission, storage, and retrieval processes. The presence of noise generally tends to reduce the image resolution and

contrast, thereby reducing its diagnostic capability. Especially, the low-intensity regions of ultrasound images with very low contrast may create an obstacle to extract resolvable details for differentiating various classes.

In order to reduce the effect of noise in ultrasound images, the contrast-limited adaptive histogram equalization (CLAHE) is employed, which is found to be very effective in enhancing the ultrasound images [8, 11, 12]. In the CLAHE method, better equalization in terms of maximum entropy is obtained and it limits the contrast of an image [13]. Here, the neighboring block boundaries are eliminated using a bilinear interpolation. The traditional adaptive histogram equalization (AHE) method over-amplifies the contrast in comparatively homogeneous regions of the image, resulting in an increased amount of noise as well [14]. Although the CLAHE method offers better performance in comparison to the performance of the AHE method, in many cases, over-enhancement is observed. Due to the over-enhancement, noises may get boosted in some cases. In order to demonstrate the noise-reduction performance of the CLAHE pre-processing technique, in Fig. 2, the LUS images of three different classes (COVID-19, pneumonia, and normal) are shown considering three cases: without using any pre-processing (raw images), using histogram equalization and using the CLAHE. It is observed from the figure that after applying the CLAHE pre-processing technique, exposure and contrast of the images are

Table 1 Dataset used in this study

Stage	Class	Videos	Frames	Total frames
Train	COVID-19	25	10,559	27,920
	Pneumonia	16	4088	
	Regular/healthy	33	13,961	
Test	COVID-19	16	3662	13,609
	Pneumonia	11	1485	
	Regular/healthy	22	8036	



increased, which makes the darker portions of the images more visible. The CLAHE method shows a better performance in this regard, but a boost to noise still exists to some extent. In order to overcome this problem, in the proposed method, a spectral domain noise enhancement scheme is introduced.

Proposed spectral mask enhancement scheme

In the proposed method, a spectral-domain LUS image enhancement approach is introduced prior to the classification stage, so that better detection performance is achieved. As discussed before, after the CLAHE-based pre-processing stage, both the significant features and noises may also get enhanced. However, the effect of noises should be eliminated while preserving the major features.

The strength of the Fourier transform in analyzing ultrasound image characteristics is widely known [15]. For this purpose, 2D spectral analysis is performed using the discrete Fourier transform (DFT) which is found to be very effective in analyzing ultrasound image characteristics [15]. For a 2D signal $g(x, y)$, the DFT is defined as [16]:

$$G(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x, y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})}, \quad (1)$$

where $x = 0, 1, \dots, (M - 1)$; $y = 0, 1, \dots, (N - 1)$, and u, v are the frequency coordinates.

From a recent study on COVID-19 and pneumonia affected ultrasound images, it is experimentally found that COVID-19 affected ultrasound images exhibit some distinguishable sonographic features, such as thickening, blurring, and discontinuities in the pleural lines of the ultrasound images [17, 18]. The spectral domain representation of an image obtained by using the DFT follows the geometric structure in the spatial domain. For example, the high-frequency components arise due to the sharp intensity changes in the border regions. The Fourier spectrum of an ultrasound image is expected to exhibit bright rays emitting from the central frequency based on the information related to directions of dominant discontinuities in edges and other geometric textures [15]. Hence, the spectral analysis of ultrasound images under consideration can extract some key information and present that through some frequency components [16, 19]. By investigating the 2D spectral masks of several ultrasound images of three different classes: healthy, pneumonia, and COVID-19, it is found that the central region of the spectral mask exhibits significant differences among the three classes, as expected. More energy concentration is also observed in the central regions of the spectrum.

The pre-processed LUS images, following the implementation of CLAHE-based image enhancement, are resized into $128 \times 128 \times 3$ RGB channels and converted into grayscale images. Considering the efficient implementation, 2D fast Fourier transform (FFT) is applied to the pre-processed image. In the resulting magnitude spectrum, it is observed that the low-frequency components exhibit a higher magnitude than that of the high-frequency components. More energy concentration is also observed in the central regions of the spectrum. As a result, a brighter area can be found near the central region. A rectangular window covering that low-frequency region is used to adjust the magnitudes near the central region, as shown in Fig. 3. Such a magnitude scaling operation helps to further control the contrast. This scaled magnitude spectrum is used along with the phase spectrum to construct the spectral mask. The enhanced image can be reconstructed from the spectral mask through inverse fast Fourier transform (IFFT), and the resulting single-channel grayscale image is found, the normalized version of which is then multiplied with the 3-channel CLAHE pre-processed LUS images. The resulting spectral mask enhanced 3-channel images are then employed in the DL models to perform the classification task. The proposed spectral mask enhancement stage followed by the DL models to classify the LUS images is termed as SpecMEn-DL method.

Classification architecture

From the LUS images obtained after the SpecMEn stage, the target is now to extract effective features for classifying the images into three classes: COVID-19, pneumonia, and normal/regular. Instead of using the conventional ML-based techniques, in the proposed scheme, a deep convolutional neural network (CNN) architecture is employed. The CNN is capable of automatically pulling out multi-variant features and learning the spatial hierarchies of features using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers. Back-propagation calibrates the weights of a neural network based on the error rate and helps to minimize the cost function in each iteration [20].

There are various efficient deep CNN architectures available in the literature. The objective of this study is not to design a new deep CNN architecture, rather demonstrate the effectiveness of the proposed spectral mask enhancement (SpecMEn) scheme in classifying the LUS images into three classes (i.e. COVID-19, pneumonia, and normal) using state-of-the-art efficient deep CNN models (overall SpecMEn-DL method). For this purpose, DenseNet-201 [21], ResNet-152V2 [22], Xception [23], VGG19 [24], and NasNetMobile [25] architectures pre-trained with ImageNet [26] are


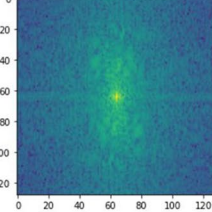

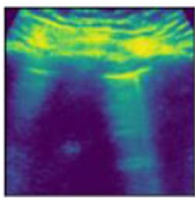

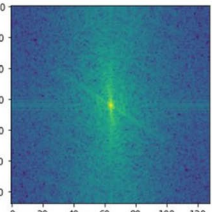
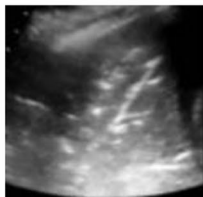
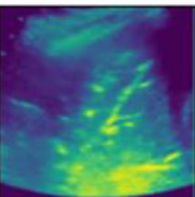
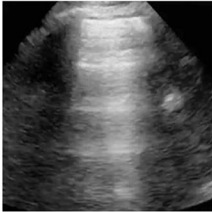
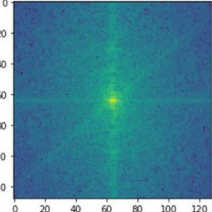

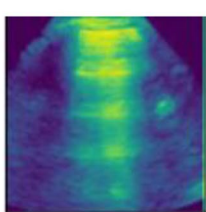
Pre-processed Image	Spectral Mask	SpecMEn Output		
		Gray-Scale	JET Colour Map	
				COVID-19
				Common Pneumonia
				Healthy

Fig. 3 Noise reduction and quality enhancement by the proposed SpecMEn method

considered. The ResNet and DenseNet architectures are designed in such a way that these deep networks are capable of overcoming the vanishing gradient problem. The sigmoid activation function is used that squeezes the input so that larger values can be easily computed. With the increase in number of layers, these values may tend to zero as every layer consists of activation functions. This problem is known as the vanishing gradient problem. As a result, as the network goes deeper, its performance does not get saturated. In order to solve this problem, the DenseNet allows layers to obtain additional information from all previous layers and pass on its feature-maps to all subsequent layers by

concatenating them and the ResNet tends to pass features skipping some layers in between. The Xception architecture inspired from the Inception network is a linear stack of depth-wise separable convolution layers with residual connections (skip connections like ResNet). Depth-wise separable convolutions are faster to compute as it splits the multiplication operation into two parts. In depthwise separable convolution, three separate kernels of size $3 \times 3 \times 1$ are used to replace a single filter of size $3 \times 3 \times 3$. $(M - 2) \times (N - 2) \times 1$ feature maps are generated from an input of size $M \times N \times 3$ and combining these maps, images of size $(M - 2) \times (N - 2) \times 3$ are found. The VGG19

architecture is a very large neural network with three fully connected layers at the closure path along with a softmax function for classification. Due to its depth, it is very troublesome to train. Images are passed through a stack of convolutional layers, where the kernels are used with a very small receptive area which takes a lot of time to calculate. The NasNetMobile or NasNet-A architecture, a family of NasNet that utilizes the NAS framework, is used for both frame-based and video-level detection due to its high performance. The Neural Architecture Search (NAS) is a data-driven intelligent approach that allows the network block to learn from data through reinforcement learning instead of experiments.

Decision tree scheme for video-level detection

The proposed method is implemented to classify the LUS images into three classes for both frame-level and video-level results. Following the frame-level stage, analysis is carried out at the video-level as well. Each LUS video contains a large number of frames but one particular class label among the three possible classes and predicting that particular video class is the ultimate target of a classification network.

In the LUS video dataset, only the labeling for each video is available. As a result, all the images of a particular video are assigned to the same class according to the class label of the video. For example, if a COVID-19 labeled video contains 300 image frames, all the 300 frames are considered as COVID-19. But in real life LUS-imaging, it is not necessary that all the frames of a COVID-19 labeled video must exhibit the characteristics of COVID-19. There may be some frames that contain normal or pneumonia characteristics. A COVID-19 labeled LUS video may contain frames that depict a healthy condition, whereas the rest of the frames are infectious. In this study, all the frames of a certain class of video are considered as members of that class; as individual frames are not annotated in the dataset, rather the videos are labeled as a whole. During the testing phase, individual video-level cases are considered where for each video, a decision-tree approach is followed to predict the class label. The analysis is performed in two steps, where a thresholding approach is applied at the first step to detect whether the frames of a certain video are healthy or not. If the number of frames in a video predicted as healthy crosses beyond a threshold, it is termed as normal or healthy case. If it does not cross the threshold, the decision on whether it is a COVID-19 case or pneumonia case is made by analyzing the predictions made on the other two types. The

process is repeated for various threshold values and the results are presented for each of the thresholds.

Experimental results

Training-testing and optimization

The deep learning models used in the proposed scheme are trained with a learning rate of 0.002, batch size 64, and the number of epochs 30. The Adam Optimizer [27] is used in each of the stages as an optimizing function. Various types of data augmentation are utilized at the training phase including rotation, horizontal and vertical shifts, scaling, and flips. The categorical cross-entropy loss function [28] is applied to calculate loss between ground truth labels and predicted results.

Frame-level results

The performance of the proposed method is evaluated on a test set consisting of 13,609 frames acquired from 49 LUS videos available in [7]. The models trained on 27,920 frames from 74 LUS videos are used to predict the frames into one of the three classes: COVID-19, pneumonia, and regular or healthy cases. Some standard statistical measures, such as the accuracy, sensitivity, specificity, and F_1 score are considered as the parameters for evaluating the performance of the proposed method. Five deep learning architectures, namely the DenseNet-201, VGG16, Xception, ResNet152V2, and NasNetMobile are employed to train through the proposed strategy and applied on the test set. Detailed results obtained for each of these five models are presented in Table 2 considering the two cases: with and without using the proposed spectral mask enhancement scheme (SpecME). The average increase in COVID-19 detection accuracy ranges up to 4% as noticeable from the Table. For example, in the case of the Xception model, for all three classes, each performance measure exhibits higher values when the proposed SpecME is applied, except a slightly lower specificity value in the regular class (0.952 and 0.949). A similar scenario is observed for the case of NasNetMobile, with relatively low accuracy in comparison to that achieved with the Xception model. It can be concluded that by using the proposed SpecME scheme, a significant increase is obtained in most of the performance evaluating parameters.

The models are trained to perform two-class classification by categorizing the images into healthy and diseased (COVID-19 and pneumonia) classes. The overall accuracy, weighted sensitivity, specificity, and F_1 score are presented in Table 3. For all the five models, the use of the proposed technique achieves a propitious performance by improving all the evaluating parameters in a congruous manner, as distinct by the results.

Table 2 Detailed frame-based result showing the improvement by the proposed technique. Five well-known deep CNN models are considered for manifestation purpose

Class	Model	DenseNet-201		VGG19		Xception		ResNet152V2		NasNetMobile	
		Without SpecME _n	With SpecME _n	Without SpecME _n	With SpecME _n	Without SpecME _n	With SpecME _n	Without SpecME _n	With SpecME _n	Without SpecME _n	With SpecME _n
COVID-19	Accuracy	0.895	0.904	0.857	0.882	0.863	0.906	0.850	0.850	0.869	0.886
	Sensitivity	0.910	0.892	0.905	0.906	0.882	0.909	0.888	0.845	0.869	0.835
	Specificity	0.879	0.917	0.806	0.857	0.843	0.902	0.811	0.854	0.870	0.939
Pneumonia	F1 score	0.899	0.905	0.867	0.887	0.868	0.904	0.859	0.852	0.871	0.882
	Accuracy	0.905	0.929	0.868	0.930	0.869	0.908	0.892	0.860	0.899	0.916
	Sensitivity	0.967	0.903	0.705	0.723	0.544	0.784	0.664	0.924	0.520	0.685
Regular	Specificity	0.898	0.932	0.967	0.955	0.910	0.924	0.920	0.852	0.945	0.944
	F1 score	0.687	0.728	0.701	0.801	0.478	0.665	0.574	0.595	0.529	0.637
	Accuracy	0.891	0.883	0.743	0.929	0.894	0.891	0.841	0.859	0.856	0.887
	Sensitivity	0.724	0.850	0.872	0.883	0.800	0.801	0.698	0.661	0.820	0.902
	Specificity	0.993	0.903	0.664	0.935	0.952	0.949	0.928	0.979	0.878	0.878
	F1 score	0.834	0.846	0.720	0.730	0.852	0.852	0.768	0.780	0.812	0.859

Table 3 Two-class i.e. healthy vs. diseased cases (COVID-19 and pneumonia) classification results by using five different deep learning models

Model	Accuracy	Sensitivity	Specificity	F1 score
DenseNet	0.836	0.836	0.751	0.827
DenseNet+SpecMEn	0.849	0.849	0.781	0.843
VGG19	0.837	0.837	0.745	0.826
VGG19+SpecMEn	0.848	0.848	0.765	0.840
Xception	0.835	0.835	0.735	0.822
Xception+SpecMEn	0.861	0.861	0.792	0.856
ResNet152V2	0.872	0.872	0.811	0.868
ResNet152V2+SpecMEn	0.904	0.904	0.865	0.902
NasNetMobile	0.816	0.816	0.762	0.812
NasNetMobile+SpecMEn	0.827	0.827	0.745	0.818

Video-level results

The ultimate goal of a classification network is to predict the class of an individual video through a certain decision, based on the frame-based results. In order to achieve this goal, the proposed classification scheme is implemented on the videos separately to check its efficacy at the individual video-level as well. For this purpose, the same 49 videos used for the frame-based results are tested by espousing the methodology presented in “Decision tree scheme for video-level detection” section. It is observed from the results in both Table 2 and 3 that NasNetMobile provides the closest result to the proposed technique, with slightly improved performance in a few parameters. Thence, the NasNetMobile model is considered for conveying the picture of improvement by the proposed method at the video-level.

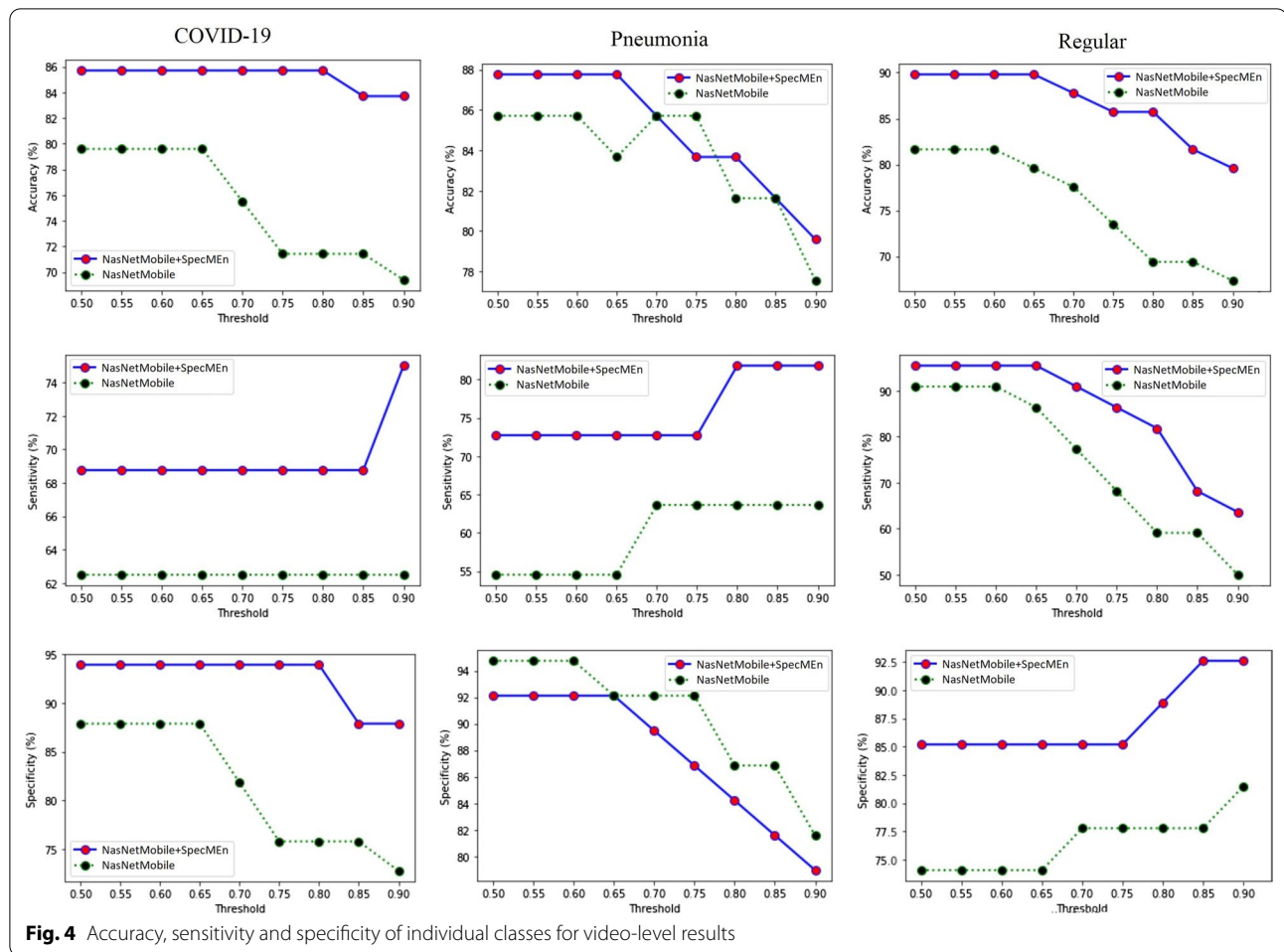
The results for each of the thresholds are shown in Table 4. It is evident from the table that for the threshold of 60%, i.e., when the individual video is predicted as normal if > 60% frames of that individual video are predicted

as normal, the best results are achieved. Examining the results, an unvarying hike is conspicuous with an increase of 8.1%, 6.4%, and 8.6% in accuracy, specificity, and F_1 score, respectively. The improvement in accuracy, sensitivity, and specificity for the three cases separately are shown in Fig. 4. The COVID-19 cases are predicted with an accuracy of $85.26 \pm 0.9\%$ by the proposed technique, whereas it becomes $75.28 \pm 1.46\%$ for NasNetMobile only. The average increase in sensitivity and specificity for COVID-19 prediction is 7.64% and 11%, respectively, compared with the traditional NasNetMobile model alone. Similarly, pneumonia and regular cases are predicted with the accuracies of 87.76% and 89.80%, which are respectively 2% and 8% greater than that achieved by the NasNetMobile. After the threshold of 0.6, the model converges and the same result is achieved for both 0.55 and 0.50 thresholds. The consistent improvement is noticeable in the visualization of the performance evaluating parameters in Fig. 4. For the various thresholds, the average classification accuracy, specificity, and F_1 score increases by an average of 11%, 7.96%, and 11.75%, respectively, than the NasNetMobile model alone.

It is to be noted that the proposed technique is implemented on a large scale of data with all the 49 videos, comprised of a total of 13,609 frames. This is the very first study following the training and testing on this massive amount of LUS frames to the best of our knowledge. Among the videos, 3 COVID-19 videos are falsely predicted by both NasNetMobile and NasNetMobile+SpecMEn as pneumonia and 2 videos as healthy. Result deviates from the usual format in these unique cases where most of the frames are predicted wrongly. Reshaping the test set with the elimination of certain frames from these unique sources tremendously increases both individual and overall accuracy. However, the proposed technique is employed regardless of these to convey the true picture of the efficacy of this model.

Table 4 Video-level classification results for varying thresholds

Threshold	Overall accuracy		Average sensitivity		Average specificity		Average F1 score	
	NasNetMobile	NasNetMobile+SpecMEn	NasNetMobile	NasNetMobile+SpecMEn	NasNetMobile	NasNetMobile+SpecMEn	NasNetMobile	NasNetMobile+SpecMEn
0.90	0.572	0.714	0.572	0.714	0.786	0.880	0.572	0.720
0.85	0.612	0.735	0.612	0.735	0.792	0.886	0.613	0.740
0.80	0.612	0.776	0.612	0.776	0.792	0.895	0.613	0.779
0.75	0.653	0.776	0.653	0.776	0.803	0.884	0.655	0.777
0.70	0.694	0.796	0.694	0.796	0.823	0.890	0.693	0.794
0.65	0.714	0.816	0.714	0.816	0.826	0.896	0.708	0.812
0.60	0.735	0.816	0.735	0.816	0.832	0.896	0.726	0.812
0.55	0.735	0.816	0.735	0.816	0.832	0.896	0.726	0.812



Comparison with related studies

The studies related to automatic prediction based on LUS datasets are limited until now. In [8], a relatively small dataset is used to consider the LUS cases through two different experiments: (1) 226 normal vs. 235 COVID-19 and 220 pneumonia cases, and (2) 235 COVID-19 vs. 220 pneumonia cases. VGG19 model is trained to classify the images into two classes each time, three-class classification is not performed there. The train-test split in that study is a bit unclear. Although the dataset used in this study can hardly be compared with them, a relative presentation of classifying the LUS images into healthy and unhealthy (COVID-19 and pneumonia) is provided in Table 5. In this study, 3662 COVID-19, 1485 pneumonia, and 8036 normal images are utilized in the testing set which is unseen at the training phase. For the same task, in [8], the amount of data was 5% to ours, with 235 COVID-19, 220 pneumonia, and 225 normal images.

In both [7] and [10], they utilized a selected portion of the POCUS dataset. In [7], they utilized 654 COVID-19,

Table 5 Comparison for two-class classification

Model	Class	Precision	Recall	F1 score	Quantity
VGG19	Healthy	0.96	0.60	0.74	3662 COVID-19, 1485 Pneumonia, 8036 Normal
	Unhealthy	0.80	0.99	0.88	
VGG19+ SpecMEen	Healthy	0.95	0.63	0.76	235 COVID-19, 220 Pneumonia, 226 Normal
	Unhealthy	0.81	0.98	0.89	
[8]	Healthy	0.94	0.98	0.96	235 COVID-19, 220 Pneumonia, 226 Normal
	Unhealthy	0.99	0.97	0.98	

277 bacterial pneumonia, 172 healthy images from 64 videos; whereas in [10], they utilized 693 COVID-19, 377 bacterial pneumonia, and 295 healthy images from 86 videos and 28 images. In both works, they gathered the images through manual processing with 30 frames per video as the maximum rate. It is apparent from our analysis in “Video-level results” section, that neglecting a portion of the dataset holds the capability of magnifying the overall performance tremendously.

Conclusion

In this paper, a spectral-domain enhancement scheme along with a histogram equalization pre-processing technique is implemented to extract the noise-reduced LUS images, which are used in the DL-based classification networks. Instead of directly using the given LUS images, the proposed SpecMEN-DL scheme utilizes noise reduced LUS images which helps in extracting better features for the classification networks and enhancing the classification performance to a significant margin. For example, at the frame-level evaluation, the proposed SpecMEN-DL scheme can enhance the COVID-19 and pneumonia detection accuracy by up to 4–6% in both 3-class and 2-class problems. At the video-level, where a single prediction is done on a particular patient's video, the detection accuracy, specificity, and F_1 score improve drastically by an average of 11%, 7.96%, and 11.75%, respectively, in comparison to the results obtained by the traditional DL model. Rigorous analysis with five established DL models in the source-independent conditions is presented to appraise the skill of the proposed technique. Consistently promising performance in both frame-level and video-level results demonstrate the superior ability of the proposed scheme in automatic COVID-19 detection from the LUS data, which can be a vital tool in this ongoing pandemic.

Acknowledgements

The authors would like to acknowledge the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET) for providing constant support throughout this study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest to disclose.

Received: 2 April 2021 Accepted: 24 May 2021

Published online: 9 July 2021

References

- Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The Lancet*. 2020;395(10223):470–3.
- Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020;296:200642.
- Ulhaq A, Born J, Khan A, Gomes DPS, Chakraborty S, Paul M. COVID-19 control by computer vision approaches: a survey. *IEEE Access*. 2020;8:179437–56.
- Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*. 2020;296(2)
- Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, Perlini S, Torri E, Mariani A, Mossolani EE, Tursi F, et al., Is there a role for lung ultrasound during the COVID-19 pandemic? *J Ultrasound Med*. 2020
- Amatya Y, Rupp J, Russell FM, Saunders J, Bales B, House DR. Diagnostic use of lung ultrasound compared to chest radiograph for suspected pneumonia in a resource-limited setting. *Int J Emerg Med*. 2018;11(1):8.
- Born J, Brandle G, Cossio M, Disdier M, Goulet J, Roulin J, Wiedemann N, POCVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). 2020.
- Horry MJ, Chakraborty S, Paul M, Ulhaq A, Pradhan B, Saha M, Shukla N. COVID-19 detection through transfer learning using multimodal imaging data. *IEEE Access*. 2020;8:149808–24.
- Roy S, Menapace W, Oei S, Luijten B, Fini E, Saltori C, Huijben I, Chen-nakeshava N, Mento F, Sentelli A, Peschiera E, Trevisan R, Maschietto G, Torri E, Inchingolo R, Smargiassi A, Soldati G, Rota P, Passerini A, van Sloun RJG, Ricci E, Demi L. Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. *IEEE Trans Med Imaging*. 2020;39(8):2676–87.
- Born J, Wiedemann N, Brandle G, Buhre C, Rieck B, Borgwardt K, Accelerating COVID-19 differential diagnosis with explainable ultrasound image analysis. 2020.
- Reza AM. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J VLSI Signal Process Syst Signal Image Video Technol*. 2004;38(1):35–44.
- Singh P, Mukundan R, De Ryke R, Feature enhancement in medical ultrasound videos using contrast-limited adaptive histogram equalization. *J Digital Imaging*. 2019;1–13
- Kouame D, Gregoire JM, Pourcelot L, Girault JM, Lethiecq M, Ossant F, Ultrasound imaging: signal acquisition, new advanced processing for biomedical and industrial applications. In: *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 (Vol. 5, pp. v/993–v/996 Vol. 5)*. IEEE
- Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ter Haar Romeny B, Zimmerman JB, Zuiderveld K. Adaptive histogram equalization and its variations. *Comput Vis Graphics Image Process*. 1987;39(3):355–68.
- Martinez-Mas J, Bueno-Crespo A, Khazendar S, Remezal-Solano M, Martinez-Cendan JP, Jassim S, Du H, Al Assam H, Bourne T, Timmerman D. Evaluation of machine learning methods with Fourier Transform features for classifying ovarian tumors based on ultrasound images. *PLoS ONE*. 2019;14:1–14.
- Najarjan K, Splinter R. *Biomedical signal and image processing*. New York: Taylor & Francis; 2012.
- Cid X, Wang A, Heiberg J, Canty D, Roysce C, Li X, El-Ansary D, Yang Y, Haji K, Haji D, et al. Point-of-care lung ultrasound in the assessment of patients with COVID-19: a tutorial. *Australas J Ultrasound Med*. 2020;23(4):271–81.
- Tan G, Lian X, Zhu Z, Wang Z, Huang F, Zhang Y, Zhao Y, He S, Wang X, Shen H, et al. Use of lung ultrasound to differentiate coronavirus disease 2019 (COVID-19) pneumonia from community-acquired pneumonia. *Ultrasound Med Biol*. 2020;46(10):2651–8.
- Smith SW, et al. *The scientist and engineer's guide to digital signal processing*. San Diego: California Technical Pub; 1997.
- Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning*. New York: MIT Press; 2016.
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700–4708)*.
- He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *European conference on computer vision 2016 (pp. 630–645)*. Springer, Cham.
- Chollet F. Xception: deep learning with depthwise separable convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017;2017:1800–7.
- Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition. 2015.
- Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 8697–8710)*.

26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in neural information processing systems*, vol. 25. New York: Curran Associates Inc; 2012. p. 1097–105.
27. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014.
28. Murphy KP. *Machine learning: a probabilistic perspective*. New York: MIT press; 2012.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.