Multi-feature Multi-Scale CNN-Derived COVID-19 Classification from Lung Ultrasound Data

Hui Che, Jared Radbel, Jag Sunderram, John L. Nosher, Vishal M. Patel, and Ilker Hacihaliloglu

Abstract—The global pandemic of the novel coronavirus disease 2019 (COVID-19) has put tremendous pressure on the medical system. Imaging plays a complementary role in the management of patients with COVID-19. Computed tomography (CT) and chest X-ray (CXR) are the two dominant screening tools. However, difficulty in eliminating the risk of disease transmission, radiation exposure and not being cost-effective are some of the challenges for CT and CXR imaging. This fact induces the implementation of lung ultrasound (LUS) for evaluating COVID-19 due to its practical advantages of noninvasiveness, repeatability, and sensitive bedside property. In this paper, we utilize a deep learning model to perform the classification of COVID-19 from LUS data, which could produce objective diagnostic information for clinicians. Specifically, all LUS images are processed to obtain their corresponding local phase filtered images and radial symmetry transformed images before fed into the multi-scale residual convolutional neural network (CNN). Secondly, image combination as the input of the network is used to explore rich and reliable features. Feature fusion strategy at different levels is adopted to investigate the relationship between the depth of feature aggregation and the classification accuracy. Our proposed method is evaluated on the point-of-care US (POCUS) dataset together with the Italian COVID-19 Lung US database (ICLUS-DB) and shows promising performance for COVID-19 prediction.

I. INTRODUCTION

The COVID-19 pandemic has increased the burden of excess morbidity and mortality worldwide. The high transmissibility and long incubation time of the SARS CoV-2 virus increase the difficulty in containing viral spread. A rapid diagnosis and severity classification in the early stage of the disease can significantly reduce the risk of further infections and help mitigate the excess morbidity and mortality of critically ill patients. At present, the main detection methods for COVID-19 infection are the genetic test (reverse transcription-polymerase chain reaction (RT-PCR)) [1] and serology test. Radiological assessment, based on CT and CXR, has been incorporated to improve the management of COVID-19 disease. However, difficulty in eliminating the risk of disease transmission, radiation exposure, not being cost-effective are some of the challenges for CT and CXR imaging [2]. CT scan can also be not performed bedside limiting its use in the intensive care unit (ICU) settings.

Lung ultrasound (LUS) is non-invasive, rapid, repeatable, and provides bedside imaging providing a safer alternative to CXR and CT. As such, LUS use for rapid assessment of the severity of COVID-19 pneumonia has been reported [1]–[3]. However, early lesions or less obvious tissue changes are difficult to distinguish by the human eyes. Furthermore, differences in medical pathology around various regions and the varied LUS experience of clinicians can result in misdiagnosis. Thus, developing standardized systems to report and interpret disease findings is a challenge with LUS [2].

Artificial intelligence (AI)-based solutions in medical imaging have demonstrated the potential to establish objective and unified interpretation standards. In [1] a new convolutional neural network (CNN) architecture, termed POCOVID-Net, was proposed. A VGG-16 architecture was used as the backbone and was fine-tuned during network training [1]. The reported average 3-class classification accuracy was 89% [1]. In [4], a multi-task CNN architecture was proposed. The network achieved an $F_1$ score of 61%, a precision of 70%, and a recall of 60% for risk prediction. Although promising early results, CNN-based methods for processing B-mode US data are affected by the image acquisition settings and quality of the collected data [5]. Finally, the limited availability of COVID-19 LUS data is also another bottleneck.

To address the above problems, we propose using a multi-feature multi-scale CNN-based approach to achieve a more accurate COVID-19 classification. Given that incorporating local-phase image tissue features can improve the accuracy of CNNs [5] for processing B-mode US data, local phase US image-based COVID-19 signatures are extracted for diverse and robust representations. Then we adopted the feature-fusion strategy to realize the effect of feature complement. To enlarge the network perception dimensions for more discriminative features of the input images, extra convolutional layers with different-size kernels are used in our CNN architecture. Our proposed approach is evaluated on 1752 scans obtained from 76 subjects.

II. METHODS

Our method mainly consists of two parts, local phase features extraction and binary classification based on multi-feature multi-scale CNNs. In this work, the use of local phase information aims to enhance the appearance of lung tissues and recover the pertinent tissue structure from LUS data. The
extraction of local phase image features also increases the
dimensionality for training. The model applied for the clas-
sification task is based on the multi-scale two-dimensional
(2D) residual neural network (ResNet) architecture similar to
the one reported in [6]. Three different fusion architectures
are investigated during this work.

A. Local Phase Image Features

Image phase information is a key component in the inter-
pretation of a scene and has been used in various applications
for processing US data [5], [7]. In this part, we first obtain
the local phase energy feature image, denoted as $LPE(x,y)$,
which is extracted using:

$$LPE(x,y) = \sum_{s} [USM1] - \sqrt{USM2^2 + USM3^2}$$  \hspace{1cm} (1)

In the above formula, $s$ represents the number of filter scales
set to 2 throughout the experimental evaluation, and $USM$ is
the group of monogenic signal images computed using the
vector-valued odd filter (Riesz filter) [7] on band-pass filtered
LUS image, denoted as $USB(x,y)$, as follows:

$$USM(x,y) = [USM1(x,y), USM2(x,y), USM3(x,y)]$$
$$= [USB(x,y), USB(x,y) \otimes h1(x,y),$$
$$USB(x,y) \otimes h2(x,y)]$$  \hspace{1cm} (2)

where $\otimes$ represents the convolution operation and $h1(x,y)$,
$h2(x,y)$ are components in Riesz filter. For bandpass filtering
$\alpha$-scale space derivative quadrature filters (ASSD) [7] are
used to output $USB(x,y)$.

Then US signal transmission map is modelled with scat-
tering and attenuation information to get enhanced im-
age $USE(x,y)$, with maximized visibility of high intensity
$LPE(x,y)$ features inside a local region. A linear interpola-
tion model is selected to combine the two interactions:

$$LPE(x,y) = USE(x,y)US(x,y) + (1 - USE(x,y))\beta$$  \hspace{1cm} (3)

Here, $LPE(x,y)$ is the local phase energy image, $USE(x,y)$
is the signal transmission map and $USE(x,y)$ is the enhanced
image. $\beta$ is a constant value representative of the tissue
echogenicity in the local region. Our aim is $USE(x,y)$ and
we hope to get two different enhancement results with
different $\beta$ value settings (60% and 90% of the maximum
intensity value of $LPE(x,y)$). Once the signal transmission
map $USE(x,y)$ is obtained using the well-established Beer
Lambert Law, $USE(x,y)$ can be calculated according to
Equation (4):

$$USE(x,y) = \frac{LPE(x,y) - \beta}{[\max(USE(x,y), \epsilon) + \beta]}$$  \hspace{1cm} (4)

$\delta$ is related to tissue attenuation coefficient, $\eta$, and $\epsilon$ is a
small constant to avoid division by zero. Throughout the
experimental evaluation $\eta = 0.85$, $\epsilon = 0.0001$. ASSD filter
parameters were kept same as reported in [7].

In Fig. 1, the two enhanced local phase images $USE(x,y)$
denoted as $USEE1(x,y)$ and $USEE2(x,y)$ are shown. These two
enhanced images are used as the input for radial symmetry
tissue extraction. Fast radial symmetry transform algorithm is
applied on the local phase images, aiming to detect points of
interest [7]. Fig. 1 shows radial symmetry images $S1(x,y)$ and
$S2(x,y)$ corresponding to the local phase enhanced images,
$USEE1(x,y)$ and $USEE2(x,y)$, it can be seen that the transforma-
tion highlights the points of interest that are characterized
by radial symmetry as well as high contrast.

B. Network Architecture

The multi-scale 2D ResNet is a light-weighted classifica-
tion network even though it simultaneously captures features
from multiple receptive fields. This network is composed of
three functional parts: 1) one convolutional layer for primary
feature map extraction, 2) multiple residual blocks with the
multi-scale convolutional layer, 3) a fully connected layer
with a softmax activation function to act as a classifier. All
input images were resized to $256 \times 256$. We investigate three
different fusion architectures with the $USE(x,y)$, $USEE1(x,y),
USEE2(x,y), S1(x,y)$, and $S2(x,y)$ images as an input (Fig.2).

Figure 2 illustrates various network architectures. During
this work, we adopt multi-scale receptive fields to focus on
different scale information. Three ResCNN blocks are basic
components in all designs to extract multi-scale features. They
have different receptive kernels, the sizes set to $3 \times 3, \ 5 \times 5$
and $7 \times 7$. In every ResCNN block, there are three
sub-blocks and each sub-block contains two convolutional
layers. The skip connection is added in each sub-block to
avoid the degradation problem [5]. An average pooling layer
follows after the convolution operation to output the final
feature map. At the end of the network, a fully connected
layer with an activation function is used to act as a classifier,
with the input of the concatenation of the final feature maps.

Feature-fusion function is utilized in different levels of
the CNN model to construct early-, mid-, and late-fusion
structures [4]. To achieve early fusion, all the images are
concatenated at the pixel level to form input with more channels. In the mid-fusion model, multiple input images are input to the network separately, processed by the initial convolutional layer to obtain corresponding primary feature maps. Concatenation is performed to finish feature aggregation for the deeper extraction. Late fusion is operated before the fully connected layer processing to fuse final feature maps from each input image.

C. Data

The dataset used in this work was obtained from [1] and [3] and consisted of 1276 COVID-19 LUS scans from 51 subjects, 254 bacterial pneumonia LUS scans from 13 subjects, and 222 LUS scans from 12 healthy subjects. Bacterial pneumonia and healthy LUS scans are joined as a non-COVID-19 class. Therefore, our focus for this work, was on binary classification.

D. Experiment Implementation

We performed 5-fold cross-validation to evaluate the performance of our proposed method. During the evaluation, the same patient data was not included in the training and testing data. The reported final results show the mean of the 5-fold cross-validation. All datasets maintain the same data distribution, including the $US(x, y), US_E_1(x, y), S_1(x, y)$ and $S_2(x, y)$ datasets.

All CNN models are trained by using the cross-entropy loss function and ADAM optimizer with a learning rate of $1e^{-5}$. Classification performance is measured by four metrics: accuracy, precision, recall, and $F_1$ score. To evaluate the effectiveness of image processing methods and feature-fusion strategies, we compare the results of just using $US(x, y)$ image as the input and the combination between two groups of processed images ($US_E_1(x, y)$ and $US_E_2(x, y)$, $S_1(x, y)$ and $S_2(x, y)$). Furthermore, we investigate the accuracy of the
model by using 5 kinds of images as input.

Experiments are implemented in the PyTorch framework with an Intel Core GPU at 3.70 GHz and a NVIDIA GeForce GTX 1080Ti GPU.

### III. RESULTS

Quantitative results of our proposed method are presented in Table 1. Our proposed multi-scale network achieves an average classification accuracy of 89.94% when using only LUS data $(US(x, y))$. The average accuracy increases to 91.93% when using enhanced images $USE(x, y)$, and 90.68% when using radial symmetry images $S(x, y)$ (Table 1). The best performance was obtained when combining the LUS $US(x, y)$ images with the enhanced images $USE(x, y)$, where an average accuracy of 95.11% was obtained. The comparison among results of the first three sets of experiments demonstrates that the local phase feature is beneficial to enhance tissue characteristics for network learning, especially, feature fusion performed in the early stage. As seen in Table 1 the results present that late-fusion design obtains the highest accuracy (95.11%), $F1_{score}$ (96.70%) significantly outperforming the other fusion operations in these two metrics. When all the image features were combined early fusion architecture obtained the best results compared to other fusion networks investigated. We further observe that by using local phase image features the performance of the network for classifying non-COVID-19 data is also improved ($F1_{score}$ of 90.48% vs 80.46%).

### IV. CONCLUSION

In this work, we have explored the use of local phase lung US image features for improved classification of COVID-19 using a multi-scale CNN architecture. Quantitative and qualitative results confirmed that the use of local phase information and multi-feature multi-scale CNN contributes to improved COVID-19 classification performance in LUS data. Fusing LUS features and local phase features at a late stage gives the highest accuracy reaching 95.11%, at the same time, other metrics prove a balanced classification capability of the model. In most cases, the early-fusion strategy shows a better classification performance. One of the drawbacks of our work is the limited amount of dataset. Compared to the publicly available CT and CXR image datasets for COVID-19, the number of available LUS is scarce. However, we are in the process of collecting new data which we believe will strengthen the evaluation of our proposed CNN model.

Our future work will include the evaluation of the proposed method on a larger scale dataset. We also would like to extend our network for multi-class classification for differentiating regular pneumonia from COVID-19 and evaluation against other CNN-based classification methods [1], [4]. Finally, optimization of the local phase image filter parameters based on CNN performance will be investigated as well.

### REFERENCES


