Mini-COVIDNet : Efficient Light Weight Deep Neural Network for Ultrasound based Point-of-Care Detection of COVID-19

Navchetan Awasthi, *Member, IEEE*, Aveen Dayal, Linga Reddy Cenkeramaddi, *Senior Member, IEEE*, and Phaneendra K. Yalavarthy, *Senior Member, IEEE*

***Abstract*—Lung ultrasound imaging has the potential to be an effective point-of-care test for detection of COVID-19, due to its ease of operation with minimal personal protection equipment along with easy disinfection. The current state-of-the-art deep learning models for detection of COVID-19 are heavy models that may not be easy to deploy in commonly utilized mobile platforms in point-of-care testing. In this work, we develop a light weight mobile friendly efficient deep learning model for detection of COVID-19 using lung ultrasound images. Three different classes including COVID-19, pneumonia, and healthy were included in this task. The developed network, named as Mini-COVIDNet, was bench-marked with other light weight neural network models along with state-of-the-art heavy model. It was shown that the proposed network can achieve the highest accuracy of 83.2% and requires training time of only 24 minutes. The proposed Mini-COVIDNet has 4.39 times less number of parameters in the network compared to its next best performing network and requires a memory of only 51.29 MB, making the point-of-care detection of COVID-19 using lung ultrasound imaging plausible on a mobile platform. Deployment of these light weight networks on embedded platforms shows that the proposed Mini-COVIDNet is highly versatile and provides optimal performance in terms of being accurate as well as having latency in the same order as other light weight networks. The developed light weight models are available at https://github.com/navchetan-awasthi/Mini-COVIDNet.**

***Index Terms*—Point-of-Care Testing, Lung Ultrasound Imaging, COVID-19, Deep Learning, Coronavirus, Detection.**

I. Introduction

CORONAVIRUS disease (2019-nCov) is an illness which is caused by the novel coronavirus, and generally known as COVID-19. It is from the respiratory family of viruses, including middle east respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS). The virus originated from Wuhan City from Hubei Province, China [1]. Human-to-human transmission via direct contact or droplets is a known characteristic of this virus, having a basic reproduction number of 2.24-3.58 and an incubation period of 2-14 days [2]. Patients suffer from rhinorrhoea, chest pain, cough, muscle ache, shortness of breath, confusion, headache, sore throat, diarrhoea, fever, nausea and vomiting according to a study of ninety nine patients with COVID-19 [3]. The standard test included for detection of COVID-19 is molecular diagnostic test i.e. a real-time reverse transcriptase-polymerase chain reaction (RT-PCR) with near perfect specificity and high analytical sensitivity for laboratory based performance [4]. The performance of the same in clinical practice is severely effected by factors such as specimen types, adequacy of specimen, stage of infection and specimen handling, including sample acquisition time from the onset of COVID-19. The rapid spread of COVID-19 has resulted in shortage of reverse RT-PCR test kits for the detection of COVID-19 and has led to exploration of other options, such as chest Computed Tomography (CT), Chest X-ray and lung ultrasound imaging for screening of COVID-19.

X-ray and CT imaging modalities have seen a wider applicability for detection of COVID-19 as RT-PCR tests in the clinical setting have low sensitivity and specificity [4]. Various studies have shown the benefit of using Chest X-ray or CT and proven to improve results for detection of COVID-19 in the clinical scenario [5]–[9]. The low sensitivity of RT-PCR technique requires repeated negative tests resulting in short supply or unavailability of kits at various parts of the globe [5]. Also, CT scans can result in false negatives when the infection is in the early stages and requires time consuming detailed disinfection procedure. Although CT has the required promise for becoming modality of choice for detection of COVID-19 [9], lung ultrasound has got the attention recently due to ultrasound machines being widely available and are relatively cheaper with added advantage of being safe and easy to disinfect.

Ultrasound imaging is a non-invasive technique, it is already replacing X-ray in diagnosis of lung related diseases [10]–[12]. Recently there has been a wide interest in point of care ultrasound based on evidence-based and expert consensus [13], [14]. The benefits of point-of-care ultrasound (POCUS) lies with being cost effective, easy to transport, and bedside accessibility for care of patients. It is not widely used currently as it lacked training pathways and also understanding of evidence behind this modality. It was found that it improves

Submitted on: January 31, 2021. Revised on: March 6, 2021; Accepted on: March 19, 2021. This work was supported in part by WIPRO-GE Collaborative Laboratory on Artificial Intelligence in Healthcare and Medical Imaging and Indo-Norwegian collaboration in Autonomous Cyber-Physical Systems (INCAPS) project: 287918 of International Partnerships for Excellent Education, Research and Innovation (INTPART) program from the Research Council of Norway.

Navchetan Awasthi is with the Massachusetts General Hospital, Boston 02114, MA, USA, and also with the Harvard University, Cambridge 02138, MA, USA (e-mail: navchetanawasthi@gmail.com).

A. Dayal and L. R. Cenkeramaddi are with the Department of Information and Communication Technology, University of Agder, Grimstad, Norway (e-mail: linga.cenkeramaddi@uia.no).

Phaneendra K. Yalavarthy is with the Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India (e-mail: yalavarthy@iisc.ac.in).

traditional examinations in diagnosis and the technology is expanding rapidly. It was also proposed that POCUS should be integrated into the acute and internal medicine curricula for widespread utility [15].

Previous studies proposed that Critical Care Ultrasound (CCUS) can be utilized for monitoring the progression of the COVID-19 pneumonia [16], [17]. Since the disease progression varies and US imaging is cheap, non-invasive and non-radiating, it is the preferred modality of choice for widespread use. The US imaging can be performed daily for patients having severe infection of lung due to COVID-19 for better management [16], [17] . The recent work, Ref. [18] , utilizes POCUS for diagnosis of COVID-19 and was found to be very effective for rapid screening as well as diagnosis of COVID-19 in symptomatic patients. It was also concluded that lung POCUS was not very accurate for patients having history of heart failure, severe emphysema, and interstitial lung disease [18]. The availability and use of POCUS screening helped in identification of symptomatic patients having COVID-19 and does not require performing the RT-PCR test in the resource constrained settings and during peak periods of surge in COVID-19 [18]. In Ref. [19], it was suggested to use positive US signs as markers for identification of persons having COVID-19 unless otherwise proven negative by RT-PCR and suggested the possibility of a classification system based on US images (US-COVID-CORADS). The work presented here is an effort in the direction of building a model for classification and assessment of the lung damage although more analysis/efforts are required by capturing more data with different acquisition schemes. Moreover, the models that are part of this work will allow better triaging of patients in busy centers with lung ultrasound.

This work was aimed at making point-of-care testing of COVID-19 using lung ultrasound imaging a reality. Typical point-of-care testing facilities does not have well trained clinical personnel and even the computing infrastructure in these settings is limited. This work provides a light weight mobile friendly deep learning models for automated detection of COVID-19 based on lung ultrasound images. The aim here is that these models should be trainable in less than half-an-hour and deployable in mobile platforms and provide detection accuracy of COVID-19 on par or better than a human expert. Three different classes including COVID-19, Pneumonia, and healthy were included in this detection task.

The main contributions of this work are:

- Efficient models, both in terms of number of parameters as well as memory, were proposed for the detection of COVID-19 using lung ultrasound images to provide performance on par or better than a human expert.
- These models can be used in mobile or in embedded applications making them universally appealing, especially in the point of care setting.
- The class imbalance problem was managed using the focal loss as the loss function to reduce the bias towards a particular class.
- With parameters being less, these networks are easily trainable on smaller dataset and can provide site/imaging protocol specific models for wider acceptance.
- This work also benchmarks the state-of-the-art light weight networks that were proposed previously for computer vision tasks to show their efficacy in detection of COVID-19 using lung ultrasound images.
- This work also provides a comparison of the proposed as well as available light weight networks in terms of their training as well as inference on low-cost embedded platforms to show the utility in a point of care setting.

II. Related Works

Lung imaging is found to be one of the techniques for capturing the information content and detection of COVID-19. The modalities that are currently being utilized for the diagnosis of the COVID-19 are the following ones and related work utilizing the same has been summarized in the respective subsections. A comparison of lung ultrasound and chest X-ray imaging modalities was also briefly presented for completeness.

A. Chest X-Ray (CXR)

Chest X-ray (CXR) is currently most widely utilized imaging technique for confirming the diagnosis of COVID-19. In Ref. [20], it was shown that the vast majority of patients (566 out of 636) were either having normal or mildly abnormal CXRs (89%) and thus it was concluded that a Chest X-Ray alone may not be effective for detection of COVID-19. In Ref. [21], it was shown that the CXR frequently showed the bilateral peripheral consolidation, but the sensitivity was lower than the RT-PCR testing (69% versus 91%, respectively). In Ref. [22], on a small dataset, it was found that the deep learning based detection model provided an accuracy of 91.24 % and the true positive rate of 0.7879 with 6.88 % false positives. It also highlighted the importance of having good resolution images, lack of large dataset for providing more generalizability. There were various studies earlier to demonstrate the advantages of lung ultrasound imaging for detection of pneumonia and other lung conditions as in Ref. [23], the sensitivity of ultrasound was found to be better than the radiography. In Ref. [24], it was shown that due to higher specificity and sensitivity when compared to chest X-rays, ultrasound was proposed for first-line examinations for acute pneumonia cases. In Ref. [25], lung ultrasound was proven to be highly beneficial for diagnosing the community-acquired pneumonia (CAP) in hospitalized children. In Ref. [26], it was shown that due to high negative prediction value, ultrasound has the potential to replace X-rays for excluding the lung consolidation in children and hence helps in reducing the exposure of radiation in the population. In Ref. [27], it was shown that lung ultrasound performance was superior to CXR for diagnosing pneumonia for those with frailty and advocated to be widely utilized for management and detection of acute respiratory symptoms in older patients.

B. Computed Tomography (CT)

While CXR is one of the important modalities for the detection chest abnormalities, it was also shown that the chest CT

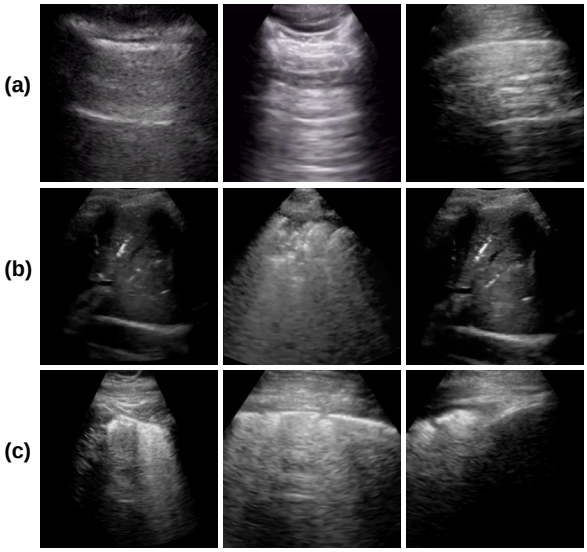


Fig. 1: Example lung ultrasound images utilized in this work representing classes (row-wise) of (a) Healthy lung, (b) Pneumonia infected lung, and (c) Lung infected with COVID-19 exhibiting pleural irregularities and small subpleural consolidation.

is effective for the detection of the abnormalities in the lung. The effectiveness for detection of COVID-19 using chest CT was also investigated earlier [28], [29]. A systematic review was also provided in Ref. [30], in which a literature search of various databases such as Google Scholar, PubMed, Elsevier, and World Health Organization was performed to provide an understanding into the follow-up CT and initial characteristics of COVID-19. Chest CT was found to be showing the greatest severity approximately after 10 days of initial symptoms onset in patients with severe respiratory distress during COVID-19 course [31]. It was found that consolidation at chest imaging or bilateral ground-glass opacities will be the main signatures in assisting the radiologist for possible diagnosis as well as management of COVID-19 [32]. Many deep learning based architectures were proposed for detection of COVID including various pretrained architectures (including ResNet-50, LSTM, DenseNet-201, Location-attention etc. [9], [33]–[38]), architectures based on attention networks, and hybrid architectures. The achieved accuracy was as high as 98 % even with number of samples available for training the network being less. In previous works, it was shown that the pretrained network based architectures performed better as compared to training a network from scratch [34]–[38]. The same strategy of utilization of pretrained networks coupled with utilization of smaller (light weight) models was deployed in this work. The problems of infection control, limited CT availability in parts of the world, CT room decontamination, and couple with high dose makes CT less attractive and portable chest radiography (CXR) was proposed as alternative in identification of lung abnormalities [39].

C. Ultrasound (US)

A timeline of ultrasound findings and comparison with CT can be found in Ref. [40]. The characteristics in chest CT images were highly consistent with the lung ultrasound findings having irregular pleural line, subpleural consolidates, multilobar B-lines and decreased blood flow [41]–[43] and can be expected to follow similar timeline as of CT [40], [43]. Moreover, when the ratio of water, air, tissue is lower in the lung, it does not manifest itself as a complete specular reflector and hence various types of vertical artifacts will be present in the ultrasound images [44]–[46]. COVID-19 was found to be associated with pulmonary embolisms as well as cardiovascular complications and the lung ultrasound can also be effective in the diagnosis, including detection of pulmonary embolisms [47]. It was found that the lung ultrasound is more beneficial because of radiation damage absence, repeatability, safety, low cost, easy disinfection, and point of care use. It was suggested that in case findings of lung ultrasound were inconclusive then the chest CT to be utilized [43]. The lung ultrasound can be effectively utilized to monitor the progress, guide the position, making decisions when to remove the patient from ventilation support and managing extracorporeal membrane therapy [43]. Development of deep learning model for diagnosis of COVID-19 using ultrasound images was achieved by utilization of a VGG-Net architecture [48] and these developments were also highlighted in the recent survey [49]. Transfer learning was also utilized and it was shown that the deeper models are difficult to train and provide inconsistent performance over the different imaging modalities with limited data training [50]. Comparisons of CT with ultrasound imaging showed that they offer complimentary information and can be utilized depending on the need (case to case) basis. While CT scan is more useful in case of severe clinical condition or for an initial assessment, lung ultrasound can be utilized as a first level examination technique in emergency department of low-risk patients and subsequent follow ups [51]. An automatic, unsupervised method was developed using Viterbi algorithm and Hidden Markov Model (HMM) followed by Support Vector Machine (SVM) for localization and detection of pleural lines in ultrasound imaging [52]. This technique achieved accuracy of 84% and 94% for convex and linear probes. Another custom model based on deep learning was proposed for presence and absence of B-lines and gave a sensitivity of 93% and specificity of 96%. This custom model was found to be successful in distinguishing between B-line severity and improved detection of presence and absence of B-lines and was found to be easily integrated in the ultrasound system [53]. Another recent work based on deep learning utilized spatial transformer networks to provide simultaneous prediction of severity score and segmentation of abnormalities related to COVID-19 lung ultrasound images [54] in a weakly supervised manner. The main task in this work [54] was to provide pixel-level segmentation of abnormalities pertaining to COVID-19 in lung ultrasound images. There have been some attempts to make lung ultrasound as a choice of imaging modality for detection of pneumonia and other diseases related to lung utilizing deep learning [55].

TABLE I: Comparison of number of samples (training as well as validation) per class in each of the fold for Covid, Pneumonia and Healthy classes.

Fold	**Covid** (Training/Validation)	**Pneumonia** (Training/Validation)	**Healthy** (Training/Validation)
1	562/116	219/58	142/40
2	557/121	217/60	141/41
3	540/138	206/71	156/26
4	486/192	235/42	149/33
5	567/111	231/46	140/42

TABLE II: Comparison of various light weight deep learning models utilized in this work in terms of parameters, corresponding required memory (in MB), as well as the number of floating-point operations (FLOPs).

Method	**Total no. of Parameters**	**Trainable**	**Non-Trainable**	**Memory (in MB)**	**Giga Flops**
COVID-CAPS	295,616	295,488	128	4.51	8.8
COVID-CAPS(Focal)	295,616	295,488	128	4.51	8.8
COVID-CAPS Scaled	3,094,432	3,093,920	512	37.82	55.4
COVID-CAPS Scaled (Focal)	3,094,432	3,093,920	512	37.82	55.4
POCOVID-Net	14,747,971	2,392,963	12,355,008	225.04	30.7
POCOVID-Net(Focal)	14,747,971	2,392,963	12,355,008	225.04	30.7
Mini-COVIDNet	3,361,091	3,338,947	22,144	51.29	1.15
Mini-COVIDNet(Focal)	3,361,091	3,338,947	22,144	51.29	1.15
MOBILE-Net-V2	2,422,979	2,388,611	34,368	36.97	0.613
MOBILE-Net-V2(Focal)	2,422,979	2,388,611	34,368	36.97	0.613
NASNetMOBILE	4,406,039	4,369,045	36,994	67.23	1.15
NASNetMOBILE(Focal)	4,406,039	4,369,045	36,994	67.23	1.15
ResNet50	23,851,011	23,797,635	53,376	291.30	7.75
ResNet50(Focal)	23,851,011	23,797,635	53,376	291.30	7.75

III. Dataset

The dataset utilized in this work was the same one as given in Ref. [48] and is explained briefly for completeness. The dataset consists of 64 videos of lung ultrasound with 11 of them being belonging to healthy patients, 14 videos of pneumonia patients, and 39 videos of COVID-19 patients. These 64 videos were taken from different sources and hence differ in format and illumination. The videos were assumed to be having a frame rate of 3Hz and a maximum of 30 frames per video to enable the extraction of lung ultrasound images that form the basis of lung ultrasound image dataset utilized in this work. This resulted in a total of 1103 lung ultrasound images (182 healthy, 277 pneumonia, and 678 COVID-19). Since the dataset is small, we utilized a five fold validation for classification of the dataset into the three classes. In five fold cross validation, the dataset was divided into five subsets and the classification/detection task was repeated five times. Each time, one of the subset gets utilized as the test set, while all remaining four subsets become part of training set. The advantage of performing this cross validation is that irrespective of dataset division during the testing and the training, each data sample is present in the test set exactly once and in the training set for exactly four times. The only disadvantage of this technique is that the model needs rerun for five times, which means it takes five times as much computation to make any evaluation [56]. As the developed models here are light weight and geared towards deploying in a limited computing environment (mobile type), the total computational time required for this five fold validation is in the same order as any other traditional deep learning models that gets trained for this purpose. Note that the dataset splitting into these five subsets are at the patient (available video) level and not at the corresponding images level. So the five fold cross validation that was performed on a particular model was with test data that the network has never seen at the patient level. The dataset was further augmented using rotation, horizontal and vertical flip, width and height scaling during training of each network. The number of samples per class in each of the five folds are chosen uniformly and exact sample numbers are provided in the Table-I.

IV. Methods

Various deep learning models have been developed for detection of COVID-19 using CT, CXR, and Ultrasound images. The detection of COVID-19 using lung ultrasound images has been posed as a three class problem, i.e. to accurately classify the lung ultrasound images into healthy, pneumonia, and COVID-19 cases (representative ultrasound images corresponding each of these classes are given in Fig. 1). As the main aim of this work is to propose a COVID-19 detection algorithm based on point-of-care ultrasound imaging, the emphasis has been on smaller (light weight) networks that can run on mobile or embedded systems to provide bed-side and immediate detection without the need for additional computing hardware. The computational complexity of each of the deep learning (CNN) model can be known by the number of floating-point operations (FLOPs) and the same was computed using the TensorFlow built-in profiler. These models were presented in detail below.

A. State of the art models

1) COVID-CAPS: This architecture has been previously utilized for identification of COVID-19 infected cases using chest X-ray (CXR) images and gave high values of specificity and sensitivity. It is an ultra-compact model that utilizes spatial correlation. It consists of four convolutional layers and

three capsule layers [57]. The initial layer is a convolutional layer succeeded by batch normalization (BN). This layer was followed by another convolutional layer superseded by average pooling. Alike, third and fourth layers are convolutional layers, with fourth layer being reshaped to obtain the first capsule layer. Later, these 3 convolutional layers were embedded in the COVID-CAPS for the subdue by agreement process. The ultimate layer has the instantiation parameters of three classes as the task at hand involves classifying lung ultrasound images into three classes. The cross entropy loss function was utilized with Adam [58] as an optimizer with initial learning rate set as 1e-4. The network parameters details as well as memory required was provided in the first row of Table-II. For completeness, a scaled version in terms of matching number of parameters with the proposed Mini-COVIDNet of the model was also included and the details of the same are presented in third row of Table-II.

2) POCOVID-Net: Here a convolutional part of VGG-16 [59], which was known to provide good performance on very large computer vision dataset, was utilized for detection of COVID-19 using lung ultrasound images [48]. It was followed by a hidden layer having ReLU activation with 64 neurons, dropout of 0.5 [60] followed by batch normalization [61]. This was superseded with softmax activated output layer. This model was originally used on ImageNet dataset for extracting the features from images. The last three layers weights were fine tuned during training. The weights of other layers were frozen at the same time resulting in number of parameters as shown in fifth row of Table-II. The loss function in this case was cross entropy and training was performed with Adam [58] optimizer and initial learning rate set as 1e-4.

3) ResNet: Here a convolutional part of ResNet50 [59], which was known to provide good performance on very large computer vision dataset such as ImageNet database, was utilized for detection of COVID-19 using lung ultrasound images [35]. It was followed by a hidden layer having ReLU activation with 64 neurons, dropout of 0.5 [60] followed by batch normalization [61]. This was superseded with softmax activated output layer. This model was originally used on ImageNet dataset for extracting the features from images. The last three layers weights were fine tuned during training. The weights of other layers were frozen at the same time resulting in number of parameters as shown in thirteenth row of Table-II. The loss function in this case was cross entropy and training was performed with Adam [58] optimizer and initial learning rate set as 1e-4.

4) MOBILE-NET-V2: There has been a wide interest in building smaller and efficient neural network models, that can be used in mobile and embedded vision applications. The most widely used one among these light weight and efficient convolutional neural network (CNN) models is MobileNet. It was originally developed for detecting the objects for mobile based and embedded computer vision applications [62]. Following this, many models were developed using the same framework and for providing improved accuracy. One such example is MobileNetV2, which uses inverted residual structure with shortcut connections between bottleneck layers [63]. MOBILE-NET-V2 architecture was shown in previous works

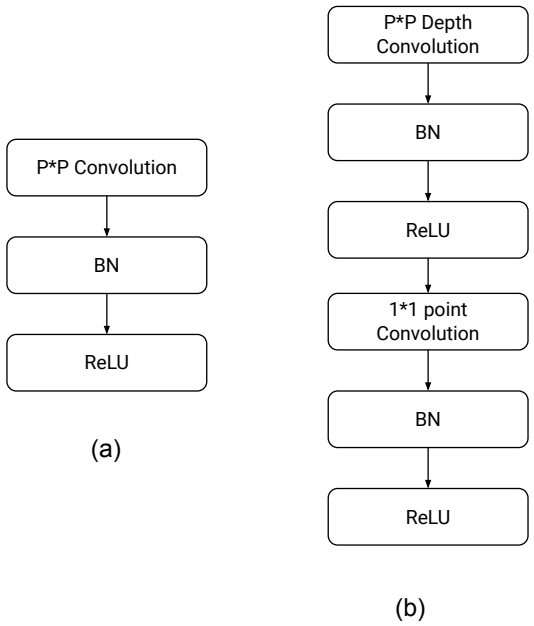


Fig. 2: Steps involved in (a) Traditional Convolution and conversion of the same into (b) Depth-wise Separable Convolution. Here BN refers to batch normalization.

to improve the state of the art performance among light weight deep learning models for benchmarks and other multiple tasks [63]. It is based on an invertible residual structure, where the thin bottleneck layers have shortcut connections between them. The intermediate expansion layer utilizes light weight convolutions to filter features and use those as a source of non-linearity in the model. Also it was found that removing non-linearities which are present in the narrow layers was important to obtain the representational power. This lead to improved performance and provided an intuition leading to design of this architecture. It was also shown that this approach allows decoupling of the input/output domains providing an easy framework for further analysis [63]. The details of this network, such as exact number of parameters and the corresponding memory, were listed in ninth row of Table-II.

5) NASNetMOBILE: NASNetMOBILE architectures utilize a new search space called as "NASNet search space" enabling the transferability with a new regularization technique known as 'ScheduledDropPath' to provide more generalizability of the model [64]. These models have been shown to provide better performance especially for classification tasks and hence have been utilized here and compared with other light weight architectures. In this architecture the search was performed for the best convolutional layer on CIFAR-10 dataset and then applied this layer to the ImageNet dataset by stacking copies of this layer to design a NASNet architecture [64]. The details of this network, such as exact number of parameters and the corresponding memory, were listed in eleventh row of Table-II.

B. Proposed Model

1) Mini-COVIDNet: The MobileNet based models have been shown to be very competent due to fewer number of

TABLE III: The Mini-COVIDNet Architecture utilized in this work detailing each layer type and its corresponding output size. Each convolutional layer was superseded by a batch-normalization layer and a ReLU layer. Notation wise, conv refers to convolution, pw: point-wise, and dw: depth-wise.

Layer (Type)	Output size
input-1 (InputLayer)	(None, 224, 224, 3)
conv1 (Conv2D)	(None, 112, 112, 32)
conv-dw-1 (DepthwiseConv2D)	(None, 112, 112, 32)
conv-pw-1 (Conv2D)	(None, 112, 112, 64)
conv-dw-2 (DepthwiseConv2D)	(None, 56, 56, 64)
conv-pw-2 (Conv2D)	(None, 56, 56, 128)
conv-dw-3 (DepthwiseConv2D)	(None, 56, 56, 128)
conv-pw-3 (Conv2D)	(None, 56, 56, 128)
conv-dw-4 (DepthwiseConv2D)	(None, 28, 28, 128)
conv-pw-4 (Conv2D)	(None, 28, 28, 256)
conv-dw-5 (DepthwiseConv2D)	(None, 28, 28, 256)
conv-pw-5 (Conv2D)	(None, 28, 28, 256)
conv-dw-6 (DepthwiseConv2D)	(None, 14, 14, 256)
conv-pw-6 (Conv2D)	(None, 14, 14, 512)
conv-dw-7 (DepthwiseConv2D)	(None, 14, 14, 512)
conv-pw-7 (Conv2D)	(None, 14, 14, 512)
conv-dw-8 (DepthwiseConv2D)	(None, 14, 14, 512)
conv-pw-8 (Conv2D)	(None, 14, 14, 512)
conv-dw-9 (DepthwiseConv2D)	(None, 14, 14, 512)
conv-pw-9 (Conv2D)	(None, 14, 14, 512)
conv-dw-10 (DepthwiseConv2D)	(None, 14, 14, 512)
conv-pw-10 (Conv2D)	(None, 14, 14, 512)
conv-dw-11 (DepthwiseConv2D)	(None, 14, 14, 512)
conv-pw-11 (Conv2D)	(None, 14, 14, 512)
conv-dw-12 (DepthwiseConv2D)	(None, 7, 7, 512)
conv-pw-12 (Conv2D)	(None, 7, 7, 1024)
conv-dw-13 (DepthwiseConv2D)	(None, 7, 7, 1024)
conv-pw-13 (Conv2D)	(None, 7, 7, 1024)
average-pooling2d (AveragePo	(None, 1, 1, 1024)
flatten (Flatten)	(None, 1024)
dense (Dense)	(None, 64)
batch-normalization (BatchNorm)	(None, 64)
relu (ReLU)	(None, 64)
dropout (Dropout)	(None, 64)
dense-1 (Dense)	(None, 3)

parameters, less model size (memory), and low latency of the model. They gave the benefit of choosing a smaller network which match the latency and size restrictions for a specific application. Generally the smaller models focus on only the size of the model, but do not give much consideration to the speed.

In this work, we propose a modified MobileNet model combined with focal loss, naming it as Mini-COVIDNet for improving the accuracy of the detection of COVID-19. This network utilizes depth-wise separable convolutions and point-wise convolutions for reduction in size [62]. Here, we have utilized the focal loss for MobileNet first time for the ultrasound datasets and compared it with other light weight architectures. The comparison included a study to know the effect of utilization of focal loss on these architectures [65]. Since MobileNet has been trained on the ImageNet dataset, we wanted to utilize the weights of the pre-trained model for the ultrasound dataset. Hence, the original model with the original weights was utilized with final layers being modified. The original model was appended by a hidden layer having ReLU activation with 64 neurons, dropout of 0.5 [60] followed by batch normalization [61]. This was superseded with softmax activated output layer. This model was originally deployed on ImageNet dataset for extracting the features from images. The last three layers weights were fine tuned during training for the problem at hand.

The benefits of using the MobileNet based architectures are as follows:

- Because of its low size, they can be used for mobile and embedded vision applications and can be trained easily with minimal computing hardware requirement.
- Generalize well because of having less number of parameters as compared to using very large networks.

The details of this network/architecture are presented below in detail.

Suppose the input has a size of $D_F \times D_F \times M$ with D_F being the height and width of the input and the number of input channels being represented by M. Assuming the filter has a shape of $D_K \times D_K \times N$ with D_K being the height and width of the filter. The number of filters are represented by N, application of the same on the input results in a size of $D_G \times D_G \times N$, where D_G is the height and width of the output and the number of channels in the output are being represented by N. The computational cost of a regular convolution is

$$D_K \cdot D_K \cdot M \cdot N \cdot D_G \cdot D_G \tag{1}$$

The depth-wise separable convolution for the same input of the same size can be computed in two steps:

- Depth-wise Convolution : Each input channel gets convolved by a filter of size $D_K \times D_K \times 1$ for producing $D_G \times D_G \times M$ sized output. The cost of computation is

$$D_K \cdot D_K \cdot M \cdot D_G \cdot D_G \tag{2}$$

- Point-wise Convolution : A filter of size $1 \times 1 \times N$ was applied to the depth-wise convolution output for producing an output of $D_G \times D_G \times N$. The cost of computation is

$$D_G \cdot D_G \cdot N \cdot M \tag{3}$$

The total cost can be written as the sum of both costs, resulting in

$$M \cdot D_G^2 \cdot D_K^2 \cdot + N \cdot D_G^2 \cdot M \tag{4}$$

The computational cost is reduced for depth-wise separable convolution as compared to regular convolution and is given as

$$\frac{M \cdot D_G^2 \cdot D_K^2 \cdot + N \cdot D_G^2 \cdot M}{D_K^2 \cdot M \cdot N \cdot D_G^2} = \frac{1}{N} + \frac{1}{D_K^2} \tag{5}$$

Thus, if a 3×3 convolution is performed using the depth-wise separable convolution, it has 9 to 8 times less computational complexity as compared to the regular convolution operation owing to a negligible decrease in the accuracy of the model. Fig. 2 shows how a regular convolution can be factorized into a point-wise and a depth-wise convolution. The summary of this model along with layer-wise details were presented in Table-III.

Initially, we trained the network using the cross entropy based metric and realized that since the dataset was imbalanced, we needed scaling of the dataset. Hence, we utilized a focal loss based strategy for training the network to provide

improved performance [66], [67]. The focal loss was explained in detail in the following text. If we denote the actual class by $y \in \{\pm 1\}$ and p_e to denote the estimated probability. The posterior probability p_t then becomes

$$p_t = \begin{cases} p_e & \text{if } y = 1 \\ 1 - p_e & \text{if } y = -1 \end{cases} \tag{1}$$

Here, p_e was computed using $p_e = sigmoid(x)$. The binary cross entropy loss can be written as

$$\epsilon_{BCE}(p_t) = -\log(p_t) \tag{2}$$

$$\frac{d\epsilon_{BCE}(p_t)}{dx} = y(p_t - 1) \tag{3}$$

Note that the gradient will be overshadowed by easily classified negative examples especially when the network was trained using the binary cross entropy loss due to existence of class imbalance. The focal loss can be described as a dynamic scaled version of cross entropy loss with its definition being:

$$\epsilon_{FL}(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \tag{4}$$

$$\frac{d\epsilon_{FL}(p_t)}{dx} = y(1 - p_t)^{\gamma}(\gamma p_t log(p_t) + p_t - 1) \tag{5}$$

The focal loss helps in down weighting the well classified examples and hence improves the accuracy of the classification scheme. The hyperparameter γ helps tuning of weight of different samples. If $\gamma = 0$, it represents the binary cross entropy loss, while with a larger value of γ fewer easily classified samples contribute to the training loss and the samples which are less in number are given more weight. Previous works have introduced hyperparameters to balance the losses from negative and positive samples or by normalizing the negative and positive loss by frequency of corresponding samples. However, it cannot handle the gradient saliency when there are hard negative samples tough to classify. Here, dynamic scaling with the posterior probability p_t is used and a weighted focal loss can be utilized for handling the different classes having different number of samples utilizing the following form of the focal loss.

$$\epsilon_{FL}(p_t) = -\lambda(1 - p_t)^{\gamma} \log(p_t) \tag{6}$$

Here, λ introduces the weight for different classes. The cross entropy loss function was used with Adam [58] optimizer with initial learning rate set as 1e-4. As with the earlier models, the exact number of parameters as well as memory required for the network were given in seventh row of Table-II.

Note that except POCOVID-Net, rest models were never deployed in the detection/classification of COVID-19 lung ultrasound images. Moreover, only COVID-CAPS was utilized earlier for the COVID-19 detection task using chest X-rays. To summarize, we have retrained four new networks in this work on lung ultrasound images and showed a systematic comparison of the same. To handle the class imbalance problem, the utilization of focal loss was implemented in the mini-COVIDNet. Results pertaining to the same were presented as a separate method to show the improvement achieved due to focal loss. Note that all networks except the COVID-CAPS utilized in this work have an addition of a 64 neurons hidden layer with ReLU activation, dropout of 0.5, superseded by batch normalization. This was followed by an output layer for three classes on top of these existing models to perform the task at hand. All networks were trained for 50 epochs to be consistent for comparison and converged before reaching 50 epochs. The time taken for 50 epochs as well as the training time has been shown in Table-VII and Table-IX. All computations in this work were performed on a Linux workstation consisting of a Intel Xeon Silver 4110 CPU with 2.10 GHz clock speed,128 GB RAM and a Nvidia Titan RTX GPU with 24 GB memory.

V. Figures of Merit

For quantitative comparison of performance of the discussed deep learning methods for detection of COVID-19 (along with other two classes), the following figures of merit were utilized.

A. Sensitivity/Recall

This is also known as the true positive rate or recall and can be defined as the proportion of actual positive cases that are correctly classified [68]. It can be written as:

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{6}$$

where TP denotes the true positives, while FN denotes the false negatives.

B. Specificity

It is also called as the true negative rate and defined as the proportion of actual negatives that are correctly identified [68]. It can be computed using:

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

where TN denotes the true negatives, while FP denotes the false positives.

C. Precision

Precision is a measure of the ability of a model not to label a negative sample as positive one [69]. It can be defined as:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

D. F1-Score

It is a measure of a accuracy and provides an easy comparison among different models for the same task [70]. It is given as

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

where Recall is same as Sensitivity defined in Section-V-A.

TABLE IV: Comparison of performance of discussed models (Sec. IV) using five-fold cross validation for the COVID class in terms of figures of merit discussed in Sec. V.

Model	Sensitivity	Specificity	Precision	F1-Score
COVID-CAPS	0.95	0.35	0.71	0.82
COVID-CAPS(Focal)	0.96	0.23	0.66	0.78
COVID-CAPS Scaled	0.92	0.54	0.76	0.83
COVID-CAPS Scaled (Focal)	0.99	0.36	0.67	0.81
POCOVID-Net	0.88	0.76	0.85	0.87
POCOVID-Net(Focal)	0.88	0.79	0.86	0.87
Mini-COVIDNet	0.92	0.68	0.82	0.86
Mini-COVIDNet(Focal)	0.92	0.71	0.83	0.87
MOBILE-Net-V2	0.96	0.61	0.81	0.88
MOBILE-Net-V2(Focal)	0.91	0.64	0.81	0.86
NASNetMOBILE	0.92	0.49	0.73	0.82
NASNetMOBILE(Focal)	0.95	0.52	0.75	0.84
ResNet50	0.84	0.57	0.76	0.80
ResNet50(Focal)	0.95	0.47	0.74	0.83

TABLE V: Comparison of performance of discussed models (Sec. IV) using five-fold cross validation for the pneumonia class in terms of figures of merit discussed in Sec. V.

Model	Sensitivity	Specificity	Precision	F1-Score
COVID-CAPS	0.45	0.92	0.69	0.55
COVID-CAPS(Focal)	0.31	0.99	0.90	0.47
COVID-CAPS Scaled	0.80	0.96	0.88	0.84
COVID-CAPS Scaled (Focal)	0.56	0.99	0.95	0.73
POCOVID-Net	0.91	0.98	0.94	0.93
POCOVID-Net(Focal)	0.94	0.96	0.90	0.92
Mini-COVIDNet	0.82	0.96	0.90	0.86
Mini-COVIDNet(Focal)	0.82	0.97	0.92	0.87
MOBILE-Net-V2	0.64	0.98	0.93	0.76
MOBILE-Net-V2(Focal)	0.68	0.95	0.85	0.75
NASNetMOBILE	0.62	0.98	0.93	0.74
NASNetMOBILE(Focal)	0.70	0.97	0.89	0.78
ResNet50	0.62	0.97	0.89	0.73
ResNet50(Focal)	0.58	0.98	0.93	0.71

TABLE VI: Comparison of performance of discussed models (Sec. IV) using five-fold cross validation for the healthy class in terms of figures of merit discussed in Sec. V.

Model	Sensitivity	Specificity	Precision	F1-Score
COVID-CAPS	0.06	0.94	0.20	0.10
COVID-CAPS(Focal)	0.07	0.94	0.20	0.10
COVID-CAPS Scaled	0.08	0.95	0.25	0.12
COVID-CAPS Scaled (Focal)	0	1	0	-
POCOVID-Net	0.50	0.92	0.55	0.52
POCOVID-Net(Focal)	0.51	0.93	0.58	0.54
Mini-COVIDNet	0.41	0.95	0.60	0.49
Mini-COVIDNet(Focal)	0.51	0.96	0.70	0.59
MOBILE-Net-V2	0.40	0.93	0.52	0.45
MOBILE-Net-V2(Focal)	0.34	0.89	0.39	0.36
NASNetMOBILE	0.28	0.94	0.52	0.36
NASNetMOBILE(Focal)	0.20	0.98	0.65	0.32
ResNet50	0.34	0.84	0.31	0.32
ResNet50(Focal)	0.25	0.95	0.50	0.33

E. Accuracy

Accuracy was defined as the number of images/frames correctly classified and given as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

This figure of merit is for all three classes combined and represent as one number to provide quantitative comparison of discussed models.

Note that the value of above discussed figures of merit will be between 0 to 1 and in all cases, the higher value (close to 1) indicates better performance of a model. These values (except accuracy) will be specific to a particular class (in here the number of classes being three).

VI. Hardware Implementation

All models were deployed on two embedded low-cost hardware devices that are attractive in the point of care settings (low-cost), namely Raspberry Pi 4 Model B and Nvidia Jetson AGX Xavier developer kit, which are also easy to integrate into the existing ultrasound scanner setup.

Raspberry Pi 4 Model B is the latest microprocessor released by Raspberry Pi Foundation. It is a cost effective embedded system with a total cost of only $35. It also provides improved connectivity, memory capacity and processor speed,

TABLE VII: Comparison of various classification models discussed in Sec. IV using five-fold cross validation in terms of accuracy across all classes. The last column of the table shows training time (in minutes) required for each model.

Model	Accuracy	Time for 50 epochs	Training Time (in min.)
COVID-CAPS	0.686	17.74	6.03
COVID-CAPS(Focal)	0.657	16.43	10.19
COVID-CAPS Scaled	0.759	22.37	4.92
COVID-CAPS Scaled(Focal)	0.734	22.42	4.93
POCOVID-Net	0.829	26.03	7.29
POCOVID-Net(Focal)	0.832	18.90	4.05
Mini-COVIDNet	0.811	24.22	1.45
Mini-COVIDNet(Focal)	0.832	24.03	2.88
MOBILE-Net-V2	0.793	17.45	5.93
MOBILE-Net-V2(Focal)	0.763	34.01	20.41
NASNetMOBILE	0.744	28.74	20.69
NASNetMOBILE(Focal)	0.771	17.19	6.88
ResNet50	0.704	16.67	3.67
ResNet50(Focal)	0.748	19.47	3.89

TABLE VIII: Ablation study of various classification models discussed in Sec. IV using five-fold cross validation in terms of accuracy across all classes.

Model	Accuracy (Cross Entropy)	Accuracy (Focal Loss)
COVID-CAPS	0.686	0.657
COVID-CAPS Scaled	0.759	0.734
POCOVID-Net	0.829	0.832
Mini-COVIDNet	0.811	0.832
MOBILE-Net-V2	0.793	0.763
NASNetMOBILE	0.744	0.771
ResNet50	0.704	0.748

when compared with its predecessor Raspberry Pi 3 Model B+. To deploy the models on the embedded system, they must first be converted from Tensorflow to Tensorflow Lite version. Tensorflow Lite is developed by Tensorflow and works as an accelerator to reduce inference time of the deployed models. Tensorflow Lite has been used to deploy several deep learning models on mobile, IoT and embedded systems. All the proposed models were converted into Tensorflow Lite version and were compared based on their inference time and memory size as shown in Table-IX. The entire hardware setup of Raspberry Pi 4 Model B is shown in Fig. 4a. It is important to note that Raspberry Pi can be considered as equal to mobile platform as the hardware specifications are similar to what is available in a typical smart phone (e.g. being 2 GB RAM). The cost specification is provided in Table-X.

Next, the models were also deployed on Nvidia Jetson AGX Xavier developer kit, the latest version in the Jetson family released by Nvidia. Nvidia Jetson AGX Xavier developer kit is a deep learning model accelerator with a 8-core ARM CPU processor and a 512-core Volta GPU with Tensor cores, offering 10 times better energy efficiency and 20 times better performance than its predecessor Nvidia Jetson TX2. The models were deployed in both the Tensorflow version and the Tensorflow Lite version and the comparison based on inference time can be seen in Table-IX. The complete hardware setup of Nvidia Jetson AGX Xavier developer kit is shown in Fig. 4b. As Nvidia Jetson also has the Volta GPU, this was also

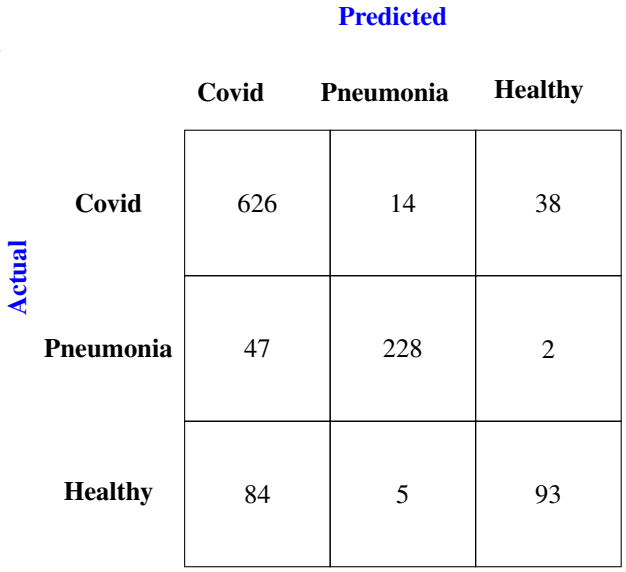


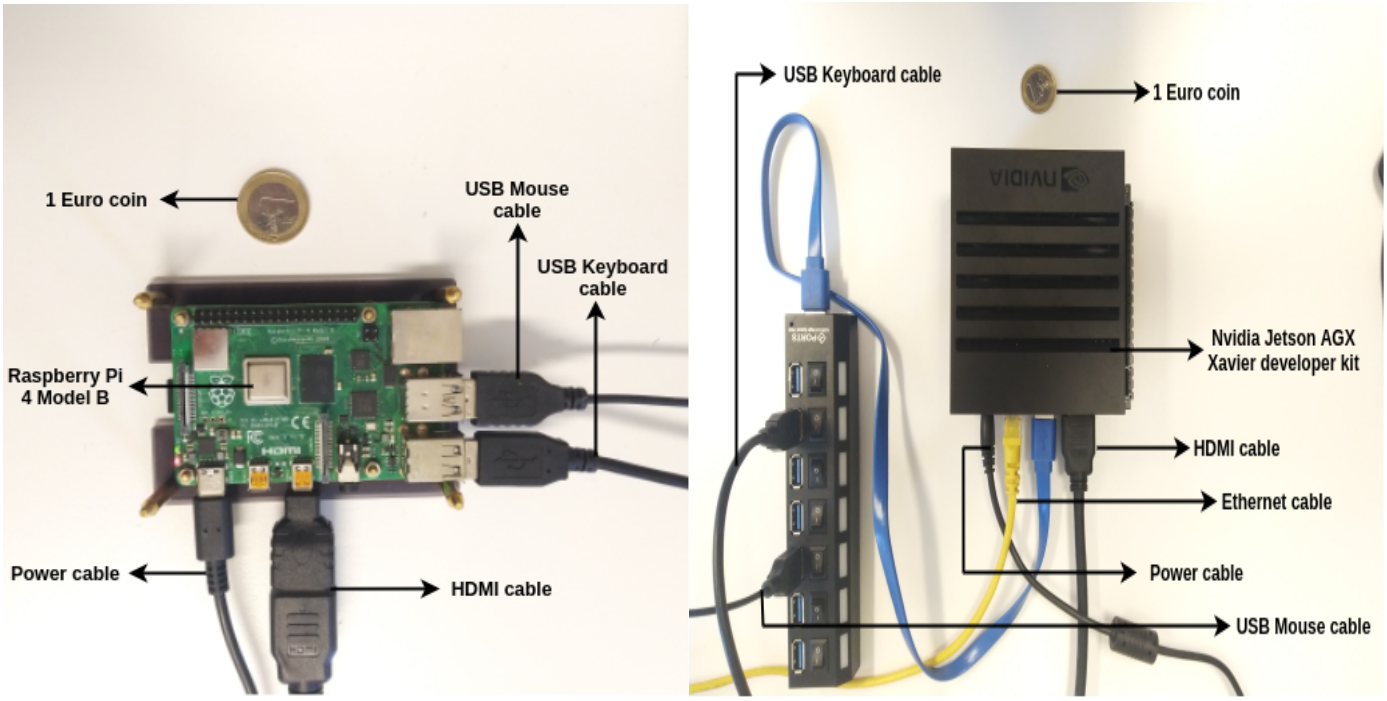
Fig. 3: Confusion Matrix of Mini-COVIDNet with focal loss after the five fold cross validation for the three classes of lung ultrasound images.

deployed for training and the required training time for each model has also been provided in Table-IX.

Results in Table-IX shows that the minimum latency was obtained for the MOBILE-Net-V2 models with the Mini-COVIDNet model being second lowest in the observed latency and hence gives better performance as compared to the other models. The latency was also obtained for the Nvidia Jetson AGX Xavier platform using the Tensorflow as well as Tensorflow Lite models. Again, the minimum latency was obtained for the MOBILE-Net-V2 models with the Mini-COVIDNet models being second lowest in the latency as observed with Raspberry Pi. The memory of the Tensorflow Lite model was also compared and shown in Table-IX. It can be seen that memory requirements for the Mini-COVIDNet models are far less as compared to the POCOVID-Net and ResNet50 models and hence can be easily deployed on a hardware platform. The training time is also shown for the models on the Nvidia Jetson AGX Xavier platform and Mini-COVIDNet has 26.20 minutes as the training time on this system. Given that the proposed Mini-COVIDNet has better accuracy compared to MOBILE-Net-V2 (Table-VII), marginal increase in the latency as well as the training time including memory makes it worthwhile for deployment on hardware platforms.

VII. Results and Discussion

The results obtained in terms of figures of merit utilizing the discussed models in Sec. IV including mini-COVIDNet and POCOVID-Net have been shown in Table-IV, V, and VI for the COVID-19, pneumonia, and healthy classes respectively. The model size with the number of parameters have been summarized in Table-II for all discussed models in this work. The memory size requirement of COVID-CAPS is very less as compared to the other models, but the performance of the same in terms of precision and F1-Score was poor compared



(a) Raspberry Pi 4 Model B

(b) Nvidia Jetson AGX Xavier developer kit

Fig. 4: The hardware setup of Raspberry Pi 4 Model B and Nvidia Jetson AGX Xavier developer kit.

TABLE IX: Comparison of various light-weight deep learning models utilized in this work in terms of latency (in seconds), training time on Nvidia Jetson AGX Xavier as well as corresponding required memory (in MB).

Method	Latency (Raspberry Pi 4, Tensorflow Lite model)	Latency (Nvidia Jetson AGX Xavier, Tensorflow Lite model)	Latency (Nvidia Jetson AGX Xavier, Tensorflow model)	Memory in MB (Tensor-flow Lite model)	Time for 50 epochs (Nvidia Jetson AGX Xavier, in Minutes)	Training Time (Nvidia Jetson AGX Xavier, in Minutes)
COVID-CAPS	0.8493	0.7413	0.7259	1	67.5	22.95
COVID-CAPS(Focal)	0.8595	0.8021	0.8494	1	67.05	41.57
COVID-CAPS Scaled	3.2532	3.7950	1.0085	12	127.20	27.98
COVID-CAPS Scaled (Focal)	3.2892	3.2529	1.0344	12	140.18	30.84
POCOVID-Net	1.8386	2.1073	0.7205	56	72.98	20.43
POCOVID-Net(Focal)	1.8508	2.1120	0.7640	56	72.74	10.18
Mini-COVIDNet	0.1841	0.1150	0.7638	13	26.20	1.57
Mini-COVIDNet(Focal)	0.1849	0.1601	0.9373	13	27.90	3.34
MOBILE-Net-V2	0.1465	0.0660	0.8872	9	23.57	8.01
MOBILE-Net-V2(Focal)	0.1465	0.0600	0.8987	9	24.27	14.56
NASNetMOBILE	0.3018	0.1402	1.7603	17	44.92	32.34
NASNetMOBILE(Focal)	0.3068	0.1488	1.8647	17	46.66	18.66
ResNet50	0.8531	0.5589	0.7964	91	55.05	12.11
ResNet50(Focal)	0.8539	0.5528	0.7574	91	54.10	10.82

to other models. POCOVID-Net performance is similar to Mini-COVIDNet while the number of parameters in the Mini-COVIDNet are lesser by 4.39 times, which results in smaller model and hence can be deployed in a real embedded system (including mobile platforms) for a preliminary detection of COVID-19. This makes Mini-COVIDNet attractive in the point-of-care setting, where the computing power is limited.

The proposed model saves 173.75 MB memory for a single model. Since the amount of data used is less and cross validation is performed using 5 different models. Here, we utilized 5 different models (5 times cross validation) for the classification of the ultrasound into the different classes. Hence, the memory saved is $173.75 \times 5 = 868.75$ MB which will be saved if the compressed models based on MOBILE-Net were utilized for deployment. The results are similar to the POCOVID-Net with an added advantage of proposed model providing the same accuracy with very less memory foot print as well as requiring lesser training time (Table-VII and Table-IX). Here, we have trained the data obtained using the linear transducer, but when the transducer or the ultrasound machine becomes different, the sample images will have different properties, thus requiring the retraining of the network. Thus, having the training time being less is an advantage for these cases and as shown even on embedded platforms the training time is less than 30 minutes.

Generally, the medical images such as PET, CT, MRI or Ultrasound lack standardization in terms of imaging protocols

and vary in terms of quality across different machines at different imaging centres [71]. This is not a problem when routine identification of features was performed in clinical setting by the clinicians. The lung ultrasound imaging protocol in the point-of-care setting is often tailored for a site and adds variation even in terms of image feature set. When these images are analyzed numerically for extracting meaningful features or using deep learning models, these imaging features that distinguish different pathologies may not be consistent across sites. [71]. Thus, having a model which can be trained in less time for extracting these features is more suitable in these settings. Another area where these models can be utilized is in the continual/lifelong learning scenarios [72]. The ability of a model to learn continuously as well as retain the previously learned knowledge is termed as continual/lifelong learning [73], [74]. This also has the disadvantage of catastrophic forgetting/interference where the model learns on new samples while forgetting the previously learned information and hence a decrease in performance of the model is observed [75], [76]. The proposed model is a light-weight one with possibility for training in less time, making it ideal for continual/lifelong learning.

The probes required for acquisition of ultrasound images can be linear or curved. A model trained with one type of probe may not be able to perform well with other. Similarly, when the ultrasound system characteristics or imaging protocol changes, the model needs to be retrained again. To avoid catastrophic forgetting, the model can be easily trained with the whole data since it can be trained easily in less time. This has not been explored here and will be taken up in future work. Also, the recent emphasis is also on designing of various energy efficient processors both for training as well as testing of these deep learning model. This work focused on more efficient and accurate models without worrying about the energy consumption of the models, which is also an important factor. Some of the models, deployed here, have high computational requirements, memory requirements, high computing power as they need to be trained on large datasets. The energy consumption of the MobileNet, Inception-3 and DenseNet model are given in Ref. [77]. The energy consumption is enormous if we include the model tuning, hyper-parameter optimizations into effect as it involves a lot of trail and error. The energy consumption is also an active area of research, has not been discussed here in much detail, the approaches for faster training will definitely help in less energy consumption and hence being very beneficial for the community [77].

Results specific to COVID-19 class were presented in Table-IV. The COVID-CAPS technique gave a higher value of sensitivity, but the specificity, precision and F1-score were less for the detection of COVID-19 class. POCOVID-Net gave better figures of merit, but at the cost of large number of parameters and a higher memory requirement. Mini-COVIDNet has improved sensitivity as compared to the POCOVID-Net and requires less parameters and memory size. Similar trend was observed for the Pneumonia class (Table-V). The results for the healthy class (Table-VI) showed improved performance for the Mini-COVIDNet with focal loss compared to all other models. In Ref. [54], the model produces a output matrix of size 2×3 as the model trains the network by minimizing the loss which consists of Soft ORDinal regression (SORD) [78], consistency loss and prior on the parameters of the transformations. Since, any change in the architecture will change the model, thus, we have not performed the comparisons of this model with the proposed one.

The accuracy of the discussed models were given in Table-VII. From the results the Mini-COVIDNet with focal loss performance is marginally improved compared to the POCOVID-Net in terms of accuracy, while the accuracy of other models are much lower than these. The corresponding confusion matrix was provided in Fig. 3 for the Mini-COVIDNet with focal loss. From this, it is clear that only two samples of Pneumonia class fall into healthy class and only 5 samples of the healthy class fall into pneumonia class giving high values of specificity for both these classes. The Mini-COVIDNet with focal loss model specificity for pneumonia and healthy class were 0.97 and 0.96 respectively as seen from Table-V and VI. The focal loss provides an active way of handling the class imbalance. In some cases, the focal loss did not give better performance as compared to the cross entropy loss [79], [80]. Similarly for the problem at hand in some scenarios, the cross entropy gives a better performance. Other techniques such as oversampling or under-sampling can be utilized along with focal loss for improving the accuracy, but have not been explored in here. Table-VIII shows the ablation study in terms of accuracy across all classes for the cross entropy loss and focal loss for the five-fold cross validation. The focal loss in general provided improved accuracy for the proposed model as well as its immediate competitor POCOVID-Net.

The training time for each of the discussed models is provided in Table-VII. It is possible to train a network and again use pruning [81] and quantization [82] to reduce the size of the model. The accuracy obtained using the ResNet50 architecture was less as compared to the other mobile architectures and hence pruning and quantization are not performed further on this architecture. The ResNet50 model had the largest number of trainable parameters and performs poorly on the test data and with focal loss it clearly shows an improvement over the basic ResNet50 model. Since the original COVID-CAPS model has less parameters as compared to the proposed Mini-COVIDNet, we have also implemented a modified COVID-CAPS Scaled model to increase the number of parameters to make it comparable to Mini-COVIDNet.

Rest of networks training times are in proportion to number of trainable parameters (except for COVID-CAPS, which takes more training time due to depth convolution) and overall the training time is in the order of tens of minutes for the discussed light weight networks. To distangle the performance based on the loss used for the training, all models were trained again utilizing the focal loss as a loss function. The results are shown in Tables-IV,V,VI and VII. Except for COVID-CAPS, COVID-CAPS Scaled and MOBILE-Net-V2, rest models perform better when the focal loss was utilized for training the models in terms of accuracy as seen in Table-VII.

The comparison of various light-weight deep learning models utilized in this work in terms of latency (in seconds), training time on Nvidia Jetson AGX Xavier as well as corre-

TABLE X: Comparison among the different hardware devices used for deploying different models utilized in this work in terms of cost and available memory.

Device	Cost	Available Memory
Raspberry Pi 4	$35	2 GB RAM
Nvidia Jetson AGX	$700	32 GB RAM & 1 GB GPU RAM
Nvidia Titan RTX	$4300	24 GB GPU RAM

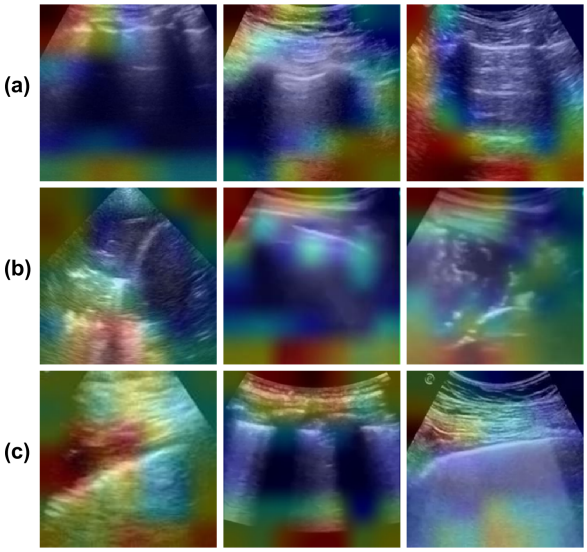


Fig. 5: Example lung ultrasound images after the visualization using the Grad-CAM utilized in this work representing classes (row-wise) of (a) Healthy lung, (b) Pneumonia infected lung, and (c) Lung infected with COVID-19 exhibiting pleural irregularities and small subpleural consolidation.

sponding required memory (in MB) has been shown in Table-IX. The comparison among the different hardware devices used for deploying the different architectures in terms of cost and available memory is shown in Table-X. The Raspberry Pi 4 has the minimum cost and is equal to mobile platform in terms of deployment and hence can be easily deployed along with a clinical ultrasound scanner in a clinical setting.

The visualization of what the model is actually learning is also very important for the purpose of understanding the model. Here, Gradient-weighted Class Activation Mapping (Grad-CAM) [83] was utilized which is class-discriminative for different classes. It highlights the important regions of the image helping in making decision for each of the possible classes. Figure-5 shows the visualizations obtained utilizing the Grad-CAM model for Covid, Pneumonia as well as Healthy lung ultrasounds. As evident, from the figure, the model activates different regions of the input image for different classes for learning the properties and hence classifying using this information. The different regions used for classification are shown and are used inherently in the model. It is also shown that since the activations in the image area are different for different images of the same type of diseased or healthy lung, they can be further improved by utilizing more datasets for the same. The healthy lungs are marked with regular pleural and A-lines [51]. Visualizations given in (Figure-5(a)) capture the pleural and A lines, but failed to capture them for all cases. The pleural consolidations are present in the Pneumonia cases as can be seen from the Figure-5(b), similar to what is reported in Refs. [47], [51]. Recent works [47], [51]–[53] confirms that the characteristics of the lung ultrasound with COVID-19 infection include irregular pleural lines with vertical artifacts. Figure-5(c) shows the lungs infected with COVID-19 and it is clear that the proposed networks is targeting the region near the pleural lines in classifying them into COVID-19 class. This analysis also asserts that the model developed here can rapidly recognize those patients who have significant lung changes (for example, the peripheral distribution of ground-glass opacities (GGOs)) manifesting as B-lines, allowing better triaging of COVID-19 patients.

It is also important to note that the lung ultrasound may not be capable to detect deep lesions within the lung and may not have discerning ability for the diagnosis of COVID19. The current method is sensitive to the damage to the pleural surface of the lung, which has been proven to have prognostic value, commonly observed in intensive care unit–admitted and deceased patients [84]. The results from the developed model has to be interpreted in the clinical and epidemiological context of the patient. The developed model will have better utility in the context of a massive COVID-19 pandemic, where it can better triage patients with pulmonary symptoms (suspected of infection).

The human/expert classification accuracy of lung ultrasound images for the COVID-19 has been only about 0.67 [54]. The results presented in this work (especially in Table-VII) indicate that these light weight models have larger utility due to less number of parameters and less size in memory and can easily perform on par or better than an expert. Even training times being in the order of tens of minutes, makes them highly attractive for a clinical setting. As these deep learning models are light weight, the run times are in the order of seconds (even in mobile platforms) and can be easily executed to give first line of detection of COVID-19 as a rapid point-of-care test using lung ultrasound images.

The final predictions currently come from a single frame of the video. Since there are five different models trained, every test image frame will be able to give us five values of prediction probabilities. These prediction probabilities can be combined differently to get a single value which will decide the class of the frame. There are mainly two basic methods of combining these-

1) Majority Voting : Selecting the classification output on the basis of highest probability among the five models [85].
2) Averaging : Selecting the classification output by averaging the output probabilities from the five models [86].

In this work, we have only performed the averaging for overall accuracy. The other advanced methods will be explored in future for improving overall accuracy.

As lung ultrasound imaging protocols vary significantly across imaging centers [87], developing a universally applicable model will be a challenge across all imaging conditions.

The work presented in here specifically provides a solution to this challenge with utilization of these light weight CNNs requiring very less training time (Table-VII) without compromising the accuracy of the detection, making them attractive as well as easy to deploy in the point-of-care setting, where ultrasound was aimed. These type of methods are essential for making deep learning methods more appealing for point-of-care COVID-19 detection studies. The developed light weight models were made available as an open-source for enthusiastic users at https://github.com/navchetan-awasthi/Mini-COVIDNet.

VIII. Conclusion

Ultrasound imaging has been known to be an effective point-of-care test for diagnosis of pneumonia and is proven to be as competent as chest X-ray. This work showed that lung ultrasound imaging can be utilized as point-of-care test for COVID-19 with deployment of light weight deep neural networks. This work utilized deep learning models that can easily run on mobile platforms for end-to-end detection of COVID-19 using lung ultrasound images, truly making it point-of-care test. As ultrasound is comparatively less infection prone (easy to disinfect), the bed-side detection based on the developed models has been shown to be a reality. The presented light weight networks, including Mini-COVIDNet, have trainable parameters in the order 2-3 million and can be trained in less than thirty minutes to provide an effective detection models, whose performance is on par with a human expert ($> 68\%$). The other advantage of Mini-COVIDNet is that the total number of parameters are 4.39 times lower compared to it's counter part, making it highly versatile to deploy in mobile platforms and is also shown to provide highest accuracy of 83.2 % in detection of COVID-19. The deployment results on the embedded platforms like Raspberry Pi and Nvidia Jetson makes the proposed Mini-COVIDNet very attractive to be utilized in point of care setting allowing better triaging of COVID-19 patients.

Acknowledgments

The authors are thankful to Dr. Jannis Born and other research group members for providing lung ultrasound imaging data as an open source, which made this work possible.

References

[1] Z. A. Memish, S. Perlman, M. D. Van Kerkhove, and A. Zumla, "Middle east respiratory syndrome," *The Lancet*, vol. 395, pp. 1063–1077, 2020.

[2] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) and corona virus disease-2019 (covid-19): the epidemic and the challenges," *International journal of antimicrobial agents*, p. 105924, 2020.

[3] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study," *The Lancet*, vol. 395, no. 10223, pp. 507–513, 2020.

[4] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin *et al.*, "The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society," *Chest*, 2020.

[5] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on covid-19: an update—radiology scientific expert panel," 2020.

[6] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, p. 200432, 2020.

[7] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, p. 200642, 2020.

[8] M. Mossa-Basha, C. C. Meltzer, D. C. Kim, M. J. Tuite, K. P. Kolli, and B. S. Tan, "Radiology department preparedness for covid-19: radiology scientific expert panel," *Radiology*, p. 200988, 2020.

[9] N. Paluru, A. Dayal, H. B. Jenssen, T. Sakinis, L. R. Cenkeramaddi, J. Prakash, and P. K. Yalavarthy, "Anam-net : Anamorphic depth embedding based light-weight cnn for segmentation of anomalies in covid-19 chest ct images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 932–946, 2021.

[10] O. Gehmacher, G. Mathis, A. Kopf, and M. Scheier, "Ultrasound imaging of pneumonia," *Ultrasound in medicine & biology*, vol. 21, no. 9, pp. 1119–1122, 1995.

[11] D. Orso, N. Guglielmo, and R. Copetti, "Lung ultrasound in diagnosing pneumonia in the emergency department: a systematic review and meta-analysis," *European Journal of Emergency Medicine*, vol. 25, no. 5, pp. 312–321, 2018.

[12] 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 *et al.*, "Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis," *Respiratory research*, vol. 15, no. 1, p. 50, 2014.

[13] G. Volpicelli, M. Elbarbary, M. Blaivas, D. A. Lichtenstein, G. Mathis, A. W. Kirkpatrick, L. Melniker, L. Gargani, V. E. Noble, G. Via *et al.*, "International evidence-based recommendations for point-of-care lung ultrasound," *Intensive care medicine*, vol. 38, no. 4, pp. 577–591, 2012.

[14] B. Bouhemad, M. Zhang, Q. Lu, and J.-J. Rouby, "Clinical review: bedside lung ultrasound in critical care practice," *Critical care*, vol. 11, no. 1, p. 205, 2007.

[15] N. Smallwood and M. Dachsel, "Point-of-care ultrasound (pocus): unnecessary gadgetry or evidence-based medicine?" *Clinical Medicine*, vol. 18, no. 3, p. 219, 2018.

[16] T. Zou, W. Yin, and Y. Kang, "Application of critical care ultrasound in patients with covid-19: Our experience and perspective," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2197–2206, 2020.

[17] X. Qian, R. Wodnicki, H. Kang, J. Zhang, H. Tchelepi, and Q. Zhou, "Current ultrasound technologies and instrumentation in the assessment and monitoring of covid-19 positive patients," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 11, pp. 2230–2240, 2020.

[18] D. S. Brenner, G. Y. Liu, R. Omron, O. Tang, B. T. Garibaldi, and T. C. Fong, "Diagnostic accuracy of lung ultrasound for sars-cov-2: a retrospective cohort study," *The ultrasound journal*, vol. 13, no. 1, pp. 1–11, 2021.

[19] S. M. Dudea, "Ultrasonography and sars-cov 2 infection: a review of what we know and do not yet know," *Medical ultrasonography*, vol. 22, no. 2, pp. 129–132, 2020.

[20] M. B. Weinstock, A. Echenique, J. W. R. Dabr, A. Leib, and F. A. Illuzzi, "Chest x-ray findings in 636 ambulatory patients with covid-19 presenting to an urgent care center: A normal chest x-ray is no guarantee," *J Urgent Care Med,(14 (7))*, pp. 13–18, 2020.

[21] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. Chung *et al.*, "Frequency and distribution of chest radiographic findings in covid-19 positive patients," *Radiology*, p. 201160, 2020.

[22] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, "Finding covid-19 from chest x-rays using deep learning on a small dataset," *arXiv preprint arXiv:2004.02060*, 2020.

[23] J.-E. Bourcier, J. Paquet, M. Seinger, E. Gallard, J.-P. Redonnet, F. Cheddadi, D. Garnier, J.-M. Bourgeois, and T. Geeraerts, "Performance comparison of lung ultrasound and chest x-ray for the diagnosis of pneumonia in the ed," *The American journal of emergency medicine*, vol. 32, no. 2, pp. 115–118, 2014.

[24] J.-E. Bourcier, S. Braga, and D. Garnier, "Lung ultrasound will soon replace chest radiography in the diagnosis of acute community-acquired pneumonia," *Current infectious disease reports*, vol. 18, no. 12, p. 43, 2016.

[25] F. Reali, G. F. S. Papa, P. Carlucci, P. Fracasso, F. Di Marco, M. Mandelli, S. Soldi, E. Riva, and S. Centanni, "Can lung ultrasound replace chest radiography for the diagnosis of pneumonia in hospitalized children?" *Respiration*, vol. 88, no. 2, pp. 112–115, 2014.

[26] A.-S. Claes, P. Clapuyt, R. Menten, N. Michoux, and D. Dumitriu, "Performance of chest ultrasound in pediatric pneumonia," *European journal of radiology*, vol. 88, pp. 82–87, 2017.

[27] A. Ticinesi, F. Lauretani, A. Nouvenne, G. Mori, G. Chiussi, M. Maggio, and T. Meschi, "Lung ultrasound and chest x-ray for detecting pneumonia in an acute geriatric ward," *Medicine*, vol. 95, no. 27, 2016.

[28] C. Bao, X. Liu, H. Zhang, Y. Li, and J. Liu, "Coronavirus disease 2019 (covid-19) ct findings: a systematic review and meta-analysis," *Journal of the American College of Radiology*.

[29] E. Y. Lee, M.-Y. Ng, and P.-L. Khong, "Covid-19 pneumonia: what has ct taught us?" *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 384–385, 2020.

[30] S. Salehi, A. Abedi, S. Balakrishnan, and A. Gholamrezanezhad, "Coronavirus disease 2019 (covid-19): a systematic review of imaging findings in 919 patients," *American Journal of Roentgenology*, pp. 1–7, 2020.

[31] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hesketh, L. Yang *et al.*, "Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia," *Radiology*, p. 200370, 2020.

[32] J. P. Kanne, "Chest ct findings in 2019 novel coronavirus (2019-ncov) infections from wuhan, china: key points for the radiologist," 2020.

[33] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang, and D. Shen, "Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2020.

[34] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[36] X. Bai, C. Fang, Y. Zhou, S. Bai, Z. Liu, L. Xia, Q. Chen, Y. Xu, T. Xia, S. Gong *et al.*, "Predicting covid-19 malignant progression with ai techniques," 2020.

[37] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi *et al.*, "Can ai help in screening viral and covid-19 pneumonia?" *arXiv preprint arXiv:2003.13145*, 2020.

[38] C. Butt, J. Gill, D. Chun, and B. A. Babu, "Deep learning system to screen coronavirus disease 2019 pneumonia," *Applied Intelligence*, p. 1, 2020.

[39] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review," *Clinical Imaging*, 2020.

[40] M. Fiala, "Ultrasound in covid-19: a timeline of ultrasound findings in relation to ct," *Clinical Radiology*, 2020.

[41] Y. Huang, S. Wang, Y. Liu, Y. Zhang, C. Zheng, Y. Zheng, C. Zhang, W. Min, H. Zhou, M. Yu *et al.*, "A preliminary study on the ultrasonic manifestations of peripulmonary lesions of non-critical novel coronavirus pneumonia (covid-19)," *Available at SSRN 3544750*, 2020.

[42] E. Poggiali, A. Dacrema, D. Bastoni, V. Tinelli, E. Demichele, P. Mateo Ramos, T. Marcianò, M. Silva, A. Vercelli, and A. Magnacavallo, "Can lung us help critical care clinicians in the early diagnosis of novel coronavirus (covid-19) pneumonia?" *Radiology*, p. 200847, 2020.

[43] Q.-Y. Peng, X.-T. Wang, L.-N. Zhang, C. C. C. U. S. Group *et al.*, "Findings of lung ultrasonography of novel corona virus pneumonia during the 2019-2020 epidemic," *Intensive care medicine*, p. 1, 2020.

[44] M. Demi, R. Prediletto, G. Soldati, and L. Demi, "Physical mechanisms providing clinical information from ultrasound lung images: hypotheses and early confirmations," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 3, pp. 612–623, 2019.

[45] G. Soldati, M. Demi, A. Smargiassi, R. Inchingolo, and L. Demi, "The role of ultrasound lung artifacts in the diagnosis of respiratory diseases," *Expert Review of Respiratory Medicine*, vol. 13, no. 2, pp. 163–172, 2019.

[46] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D. F. Briganti, S. Perlini, E. Torri, A. Mariani, E. E. Mossolani *et al.*, "Is there a role for lung ultrasound during the covid-19 pandemic?" *Journal of Ultrasound in Medicine*.

[47] A. Boccatonda, E. Ianniello, D. D'Ardes, G. Cocco, F. Giostra, C. Borghi, and C. Schiavone, "Can lung ultrasound be used to screen for pulmonary embolism in patients with sars-cov-2 pneumonia?" *European journal of case reports in internal medicine*, vol. 7, no. 7, 2020.

[48] J. Born, G. Brändle, M. Cossio, M. Disdier, J. Goulet, J. Roulin, and N. Wiedemann, "Pocovid-net: Automatic detection of covid-19 from a new lung ultrasound imaging dataset (pocus)," *arXiv preprint arXiv:2004.12084*, 2020.

[49] T. T. Nguyen, "Artificial intelligence in the battle against coronavirus (covid-19): a survey and future research directions," *Preprint, DOI*, vol. 10, 2020.

[50] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, "Covid-19 detection through transfer learning using multimodal imaging data," *IEEE Access*, vol. 8, pp. 149 808–149 824, 2020.

[51] D. Pata, P. Valentini, C. De Rose, R. De Santis, R. Morello, and D. Buonsenso, "Chest computed tomography and lung ultrasound findings in covid-19 pneumonia: A pocket review for non-radiologists," *Frontiers in Medicine*, vol. 7, p. 375, 2020.

[52] L. Carrer, E. Donini, D. Marinelli, M. Zanetti, F. Mento, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi *et al.*, "Automatic pleural line extraction and covid-19 scoring from lung ultrasound data," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[53] B. Cristiana, T. Grzegorz, K. Seungsoo, M. Katelyn, L. Rachel, M. M. Shaw, R. L. McNamara, R. Balasundar, and C. L. Moore, "Automated lung ultrasound b-line assessment using a deep learning algorithm," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[54] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli *et al.*, "Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound," *IEEE Transactions on Medical Imaging*, 2020.

[55] X. Zheng, S. Kulhare, C. Mehanian, Z. Chen, and B. Wilson, "Feature detection and pneumonia diagnosis based on clinical lung ultrasound imagery using deep learning," *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1668–1668, 2018. [Online]. Available: https://doi.org/10.1121/1.5067436

[56] M. H. DeGroot, *Optimal statistical decisions*. John Wiley & Sons, 2005, vol. 82.

[57] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images," *arXiv preprint arXiv:2004.02696*, 2020.

[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[59] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[60] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[61] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[62] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[63] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[64] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[65] A. J. G. Busson, S. Colcher, R. L. Milidiú, B. P. Dias, and A. Bulcão, "Seismic shot gather noise localization using a multi-scale feature-fusion-based neural network," *arXiv preprint arXiv:2005.03626*, 2020.

[66] P. Yun, L. Tai, Y. Wang, C. Liu, and M. Liu, "Focal loss in 3d object detection," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1263–1270, 2019.

[67] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[68] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity." *BMJ: British Medical Journal*, vol. 308, no. 6943, p. 1552, 1994.

[69] M. Buckland and F. Gey, "The relationship between recall and precision," *Journal of the American society for information science*, vol. 45, no. 1, pp. 12–19, 1994.
[70] D. C. Blair, "Information retrieval," *Journal of the American Society for Information Science*, vol. 30, no. 6, pp. 374–375, 1979.
[71] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
[72] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, vol. 113, pp. 54–71, 2019.
[73] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.
[74] S. Thrun and T. M. Mitchell, "Lifelong robot learning," *Robotics and autonomous systems*, vol. 15, no. 1-2, pp. 25–46, 1995.
[75] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," in *Psychology of learning and motivation*. Elsevier, 1989, vol. 24, pp. 109–165.
[76] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory," *Psychological review*, vol. 102, no. 3, p. 419, 1995.
[77] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, "Estimation of energy consumption in machine learning," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 75–88, 2019.
[78] R. Diaz and A. Marathe, "Soft labels for ordinal regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4738–4747.
[79] K. Pasupa, S. Vatathanavaro, and S. Tungjitnob, "Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2020.
[80] B. K. Baloch, S. Kumar, S. Haresh, A. Rehman, and T. Syed, "Focused anchors loss: Cost-sensitive learning of discriminative features for imbalanced classification," in *Asian Conference on Machine Learning*, 2019, pp. 822–835.
[81] R. Reed, "Pruning algorithms-a survey," *IEEE transactions on Neural Networks*, vol. 4, no. 5, pp. 740–747, 1993.
[82] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding," *arXiv preprint arXiv:1510.00149*, 2015.
[83] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
[84] S. M. H. Tabatabaei, H. Talari, F. Moghaddas, and H. Rajebi, "Ct features and short-term prognosis of covid-19 pneumonia: A single-center study from kashan, iran," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200130, 2020. [Online]. Available: https://doi.org/10.1148/ryct.2020200130
[85] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Information fusion*, vol. 6, no. 1, pp. 63–81, 2005.
[86] L. Brewster, J. Dale, T. Guttridge, S. Gruber, A. Hansell, M. Elliott, I. Cowx, N. Whitney, and A. Gleiss, "Development and application of a machine learning algorithm for classification of elasmobranch behaviour from accelerometry data," *Marine biology*, vol. 165, no. 4, p. 62, 2018.
[87] C. J. Patel, H. B. Bhatt, S. N. Parikh, B. N. Jhaveri, and J. H. Puranik, "Bedside lung ultrasound in emergency protocol as a diagnostic tool in patients of acute respiratory distress presenting to emergency department," *Journal of emergencies, trauma, and shock*, vol. 11, no. 2, p. 125, 2018.

Navchetan Awasthi received M.Tech. degree in computational science from Indian Institute of Science (IISc), Bangalore in 2016 and Ph.D. degree in medical imaging from Indian Institute of Science (IISc), Bangalore in 2019. He also received B.Tech. degree in electronics and communication engineering from National Institute of Technology (NIT), Jalandhar in 2011. He is currently a research fellow in Massachusetts General Hospital and Harvard University. His research interests include inverse problems in biomedical optics, medical image analysis, medical image reconstruction, biomedical signal processing, and deep learning.

Aveen Dayal received the bachelor's degree in computer science and engineering from BML Munjal University, Gurgaon, India, in 2020. He is currently a Visiting Research Student with the Department of Information and Communication Technology, University of Agder, Grimstad, Norway. His main research interests are in machine learning and deep learning for autonomous cyber–physical systems.

Linga Reddy Cenkeramaddi (Senior Member, IEEE) received the master's degree in electrical engineering from IIT Delhi, New Delhi, India, in 2004, and the Ph.D. degree in electrical engineering from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2011. He worked for Texas Instruments, Bengaluru, India, in mixed-signal circuit design before joining the Ph.D. program at NTNU. After finishing his Ph.D. degree, he worked in radiation imaging for an atmosphere space interaction monitor (ASIM mission to International Space Station) at the University of Bergen, Bergen, Norway, from 2010 to 2012. He is currently an Associate Professor with the University of Agder, Grimstad, Norway. His main scientific interests are in cyber–physical systems, autonomous systems, and wireless embedded systems.

Phaneendra K. Yalavarthy Senior Member, IEEE) received the M.Sc. degree in engineering from the Indian Institute of Science, Bangalore, India in 2004. He obtained Ph.D. degree in biomedical computation from Dartmouth College, Hanover, NH, USA, in 2007. He is an Associate Professor with the Department of Computational and Data Sciences, Indian Institute of Science, Bangalore. His research interests include medical image computing, medical image analysis, and biomedical optics. He is a senior member of IEEE, SPIE, and OSA. He serves as the associate editor of IEEE Transactions on Medical Imaging.