Multi-modal lung ultrasound image classification by fusing image-based features and probe information

Gabriel Iluebe Okolo
School of Computing, Engineering and Physical Sciences
University of the West of Scotland
Paisley, United Kingdom
gabriel.okolo@uws.ac.uk

Stamos Katsigiannis
Department of Computer Science
Durham University
Durham, United Kingdom
stamos.katsigiannis@durham.ac.uk

Naeem Ramzan
School of Computing, Engineering and Physical Sciences
University of the West of Scotland
Paisley, United Kingdom
naeem.ramzan@uws.ac.uk

***Abstract*—Lung ultrasound is a widely used portable, cheap, and non-invasive medical imaging technology that can be used to identify various lung pathologies. In this work, we propose a multi-modal approach for lung ultrasound image classification that combines image-based features with information about the type of ultrasound probe used to acquire the input image. Experiments on a large lung ultrasound image dataset that contains images acquired with a linear or a convex ultrasound probe demonstrated the superiority of the proposed approach for the task of classifying lung ultrasound images as "COVID-19", "Normal", "Pneumonia", or "Other", when compared to simply using image-based features. Classification accuracy reached 99.98% using the proposed combination of the Xception pre-trained CNN model with the ultrasound probe information, as opposed to 96.81% when only the pre-trained EfficientNetB4 CNN model was used. Furthermore, the experimental results demonstrated a consistent improvement in classification performance when combining the examined base CNN models with probe information, indicating the efficiency of the proposed approach.**

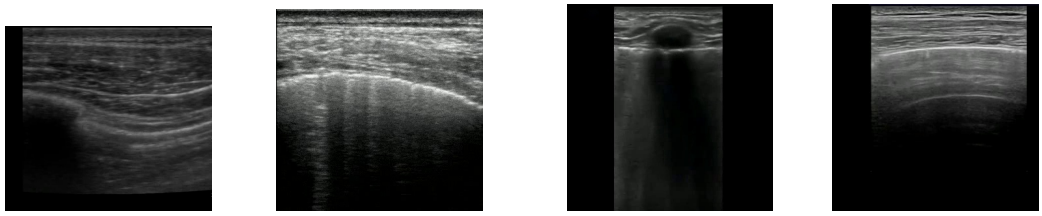
***Index Terms*—lung ultrasound images, COVID-19, image classification, multi-modal, CNN**

I. Introduction

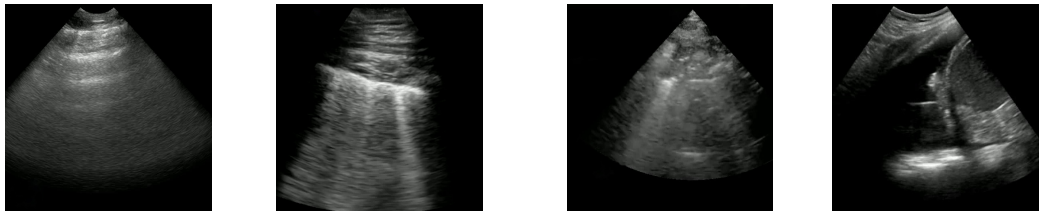
Lung ultrasound is a well known medical imaging technique for the detection of pneumonia and related lung sicknesses [1]. As an alternative to CT and X-ray, lung ultrasound (LUS) is a portable, cheap, non-invasive, and easy-to-use medical imaging technology that can be used to identify lung illnesses [2]. In addition, LUS imaging is an effective technique for early diagnosis and follow-up of COVID-19 patients, according to recent medical literature [3]–[5]. Fiala et al. [4] provide a succinct summary of LUS results in COVID-19 patients. Some of these abnormalities, such as numerous fused bilateral B lines, subpleural pulmonary consolidations, uneven pleural lines, and poor blood flow, are consistent with the results of CT scans. A significant difference between subpleural lesions in COVID-19 patients and other pulmonary disorders, such as bacterial pneumonia, tuberculosis, and cardiogenic pulmonary oedema, is one of the key findings Fiala et al. [4] highlighted.

The use of biomedical imaging methods (e.g. US, X-ray, CT) has recently shown potential in the detection of COVID-19, allowing in most cases for faster diagnosis than with the widely used RT-PCR approach [6], [7]. Furthermore, deep learning methods and more specifically convolutional neural networks (CNNs), have demonstrated high efficiency in many computer vision-oriented tasks, including the classification and segmentation of images, as well as object detection [8]. Consequently, research on AI/machine learning systems has extensively focused on automating image analysis in the clinical field [9], and deep learning-based automated detection of COVID-19 using LUS imaging has been shown to achieve high performance [10]–[12].

Diaz-Escobar et al. [13] classified the POCUS LUS dataset [10] using binary (COVID-19 vs. pneumonia and COVID-19 vs. healthy) and three-class (COVID-19, pneumonia, and healthy) pre-trained deep learning (DL) models, such as VGG19 [14], InceptionV3 [15], Xception [16], and ResNet50 [17]. Their findings demonstrated that InceptionV3 achieved the highest AUC of 0.97 for distinguishing COVID-19 cases from healthy controls and patients with pneumonia. A lightweight DL design for COVID-19 LUS diagnosis was proposed by Awasthi et al. [18]. The newly developed technique, known as Mini-CovidNet, alters MobileNet with focus loss, achieving an accuracy of 0.83 on the POCUS dataset. Che et al. [19] evaluated a dataset made up of both the POCUS [10] and ICLUS-DB [20] datasets, using a multi-scale residual CNN with a feature fusion approach, achieving an average accuracy of 0.95. Horry et al. [6] conducted an experiment on US images of COVID-19, normal and bacterial pneumonia, by applying transfer learning for the classification approach using eight CNN-based models, achieving a recall and precision of 1.0. Similarly, Roberts et al. [21] carried out a study on classifying COVID-19, bacterial pneumonia, and normal cases on the POCUS dataset [10] by applying DL techniques using the VGG16 [14] and ResNet18 [17] architectures, achieving an accuracy of 0.86 and an AUC of 0.90. Sadik et al. [22] also experimented on the POCUS dataset using four DL models (DenseNet201 [23], ResNet152V2 [17], Xception [16], and VGG19 [14]), achieving an accuracy of 0.91 and an F1-score of 0.90. Zheng et al. [24] proposed a multi-modal approach that combines imaging and text data using a neural network that can classify COVID-19 vs. non-COVID-19 cases, achieving an accuracy of 0.98 and a F1-score of 0.99. It must be noted that in many research works that



(a) Normal (linear) (b) COVID-19 (linear) (c) Pneumonia (linear) (d) Other (linear)



(e) Normal (convex) (f) COVID-19 (convex) (g) Pneumonia (convex) (h) Other (convex)

Fig. 1: Example US images from the COVIDx-US dataset for each class and probe type. (a-d) linear probe, (e-h) convex probe.

focused on COVID-19 detection using DL approaches, multi-modal approaches have proven to be more effective, achieving better classification performance than methods relying on a single modality [6], [24], [25].

Despite numerous research works on automated LUS image classification, the performance achieved is not yet on par with the level required for real-world clinical practice. To this end, in this work, we employed a large lung ultrasound image dataset (COVIDx-US [26]) that contains 22,776 LUS images, obtained using a linear or convex US probe, and characterised as referring to COVID-19, pneumonia, other pathologies, or normal images. We propose a multi-modal approach for LUS image classification that fuses image-based features, extracted using common CNN models pre-trained on ImageNet [27], with information about the probe used to obtain the input LUS image. The proposed approach was evaluated on the examined dataset against commonly used pre-trained CNN models, demonstrating a consistent improvement in performance against all the examined models.

The contribution of this work can be summarised as follows: (i) We evaluate the performance of several well-proven CNN architectures that have been shown to perform well on general image classification tasks, and examine them on the task of classifying LUS images as belonging to COVID-19, normal, pneumonia, and others. (ii) We propose a multi-modal approach where image-based features extracted by the CNN models are combined with information about the type of probe used in order to acquire the ultrasound image, in order to enhance the final classification performance. (iii) We provide a detailed performance evaluation of the proposed multi-modal architecture on the examined LUS image dataset.

II. Methodology

In this work, we attempt to improve the performance of common pre-trained CNN models on the task of classifying lung ultrasound images as COVID-19, pneumonia, other lung diseases/conditions, or normal. To this end, we propose and evaluate a multi-modal approach where image-based features extracted by the CNN models are combined with information about the type of probe used in order to acquire the ultrasound image, in order to enhance the final classification performance. The proposed models were trained and evaluated on a dataset containing 22,776 lung ultrasound images.

A. Dataset

The COVIDx-US [26] (version 1.5) collection is heterogeneous in nature and includes lung ultrasonic imaging data from numerous sources with different properties, such as different US probe types (convex or linear), symptoms exhibited by the patients, demographic information, and others. It contains 220 lung ultrasound videos created using linear or convex probes, from which 22,776 processed ultrasound images are extracted. The ultrasound videos in the dataset were acquired from nine different sources: 1) ButterflyNetwork, 2) GrepMed, 3) LITFL, 4) The PocusAtlas, 5) Radiopaedia, 6) CoreUltrasound, 7) University of Florida (UF), 8) Scientific Publications, and 9) Clarius. In the current COVIDx-US version, US images are divided into four categories: “COVID-19”, “normal”, “pneumonia”, and “other”. The dataset contains 9,227 US images categorised as “COVID-19”, 2,245 categorised as “normal”, 4,300 as “pneumonia”, and 7,004 as “other”. In the absence of an official training/test split, we opted to split the dataset into 90% training and 10% test, using random stratified sampling in order to preserve the classes distribution. The images in the training set were further divided into 80% for training and 20% for validation in order to facilitate the training of the machine learning models. Apart from the class that it belongs to, each image was also annotated with the type of US probe (linear or convex) used for its acquisition. Linear probes have a flat array and look, producing pictures with a better resolution but less tissue penetration, whereas convex probes (or curved linear probes) have a curved array with a

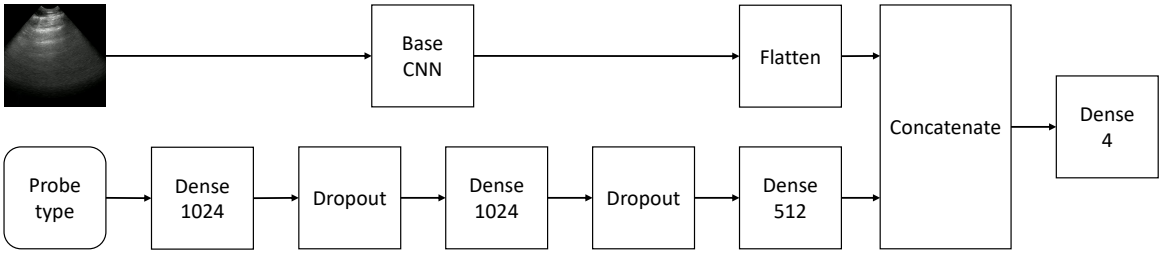


Fig. 2: Outline of the proposed architecture.

wider field of view at a lower frequency, offering wider depth and deeper penetration [28]. Examples of LUS images from the COVIDx-US dataset for each class and probe type are shown in Fig. 1.

B. Proposed architecture

Preliminary experimentation on the examined dataset using various established CNN models, pre-trained on ImageNet and fine-tuned on the training set of the dataset, demonstrated that such an approach can achieve impressive performance, with accuracies and F1-scores up to 96.5%. Nevertheless, as explained in the description of the dataset, the COVIDx-US dataset contains LUS images acquired using a linear or a convex probe. LUS images acquired with different types of probes exhibit some different characteristics, as can be seen in Fig. 1. As a result, and despite the high performance achieved during our preliminary experimentation, we hypothesise that the presence in the dataset of US images obtained using various types of probes (linear, convex) may hinder the ability of the machine learning model to precisely classify a US image. Dividing the dataset into LUS images acquired using a linear probe and images acquired using a convex probe could be a potential solution to this issue. However, this would lead to: (a) a reduction in the number of training samples, (b) the requirement for creating and training two separate machine learning models, one for images acquired using a linear probe, and one for images acquired using a convex probe, resulting in reduced practicality and increased computational cost, and (c) potential reduction in the generalisation ability of the trained models, since the high accuracy and F1-score achieved during the preliminary experimentation demonstrated that despite their differences, images acquired with these two types of probes exhibit similar features that can be successfully extracted using a single machine learning model.

To address this issue and avoid dividing the dataset according to probe type, thus having to train two separate models, we propose a multi-modal approach where features extracted using a CNN model, pre-trained on ImageNet, are fused with information about the type of probe used to acquire the input image. An outline of the proposed approach is provided in Fig. 2. The proposed deep learning approach consists of a neural network with two branches. The first branch takes a LUS image as an input and passes it through the convolutional base of a pre-trained CNN, i.e. the convolutional layers without the final fully connected layers used for classification. The output of the convolutional layers is then flattened before being fused with the output of the second branch of the neural network. In this work, we examined the performance of the following nine established CNN models for the first branch of the proposed model: MobileNetV2 [29], InceptionV3 [15], Inception-ResNetV2 [30], Xception [16], ResNet50V2 [17], EfficientNetB4 [31], DenseNet121, DenseNet169, DenseNet201 [23].

The second branch of the proposed neural network takes as an input the type of probe used (linear or convex) and maps it to a 1-dimensional embedding of size 512. To compute the probe embedding, the input is first passed through a fully connected (dense) layer of size 1024, followed by another fully connected layer of size 1024, and finally a fully connected layer of size 512. A ReLU activation function is used for all the fully connected layers. To reduce overfitting and improve the generalisation ability of the model, a dropout layer [32] with a dropout rate of 0.4 is added after each of the first two fully connected layers, as shown in Fig. 2. The output of both branches is then concatenated and passed through a fully connected layer of size 4 (equal to the number of classes in the dataset) that uses a softmax activation function for the final classification of the input image.

C. Training and classification

To evaluate the efficiency of the proposed multi-modal approach, we compared its performance (see Section III) to the performance of the base CNNs used, fine-tuned on the examined dataset. All base CNN models used were pre-trained on ImageNet and the Keras library was used for both the pre-trained models, as well as for implementing the proposed architecture. Both the proposed and the baseline models were trained using the Adam optimiser, a learning rate of 0.0001, a batch size of 16, and sparse categorical cross-entropy as the loss function. In addition, the training process stopped after 20 epochs with no improvement in validation accuracy, and the learning rate was multiplied by a factor of 0.2 after 2 epochs with no improvement, with a lower bound on the learning rate of 10^{-6}.

III. Results & Discussion

The proposed approach using the nine base CNN models, as well as the nine base CNN models individually, were trained and evaluated on the COVIDx-US dataset for the task of classifying LUS images into “COVID-19”, “Normal”, “Pneumonia”,

TABLE I: Classification performance (%) of the baseline and the proposed methods for various base CNN models. Results in bold indicate the best performance for each metric and approach. Underlined results indicate the overall best performance.

Base CNN	**Base CNN**					**Base CNN + Probe info (proposed)**				
	Params.	**Accuracy**	**Precision**	**Recall**	**F1**	**Params.**	**Accuracy**	**Precision**	**Recall**	**F1**
MobileNetV2	2,508,868	96.73	**96.51**	96.46	96.49	4,087,364	99.32	99.14	99.42	99.27
InceptionV3	22,007,588	96.49	96.45	96.67	96.55	23,586,084	99.34	99.18	99.39	99.28
InceptionResNetV2	54,490,340	96.44	96.28	95.80	96.03	56,068,836	99.39	99.23	99.46	99.36
Xception	21,262,892	94.38	95.03	95.24	95.11	22,841,388	**99.98**	**99.99**	**99.94**	**99.97**
ResNet50V2	23,966,212	96.57	96.32	96.35	96.33	25,544,708	99.71	99.70	99.70	99.70
EfficientNetB4	18,025,052	**96.81**	96.47	**96.70**	**96.58**	19,603,548	99.41	99.24	99.22	99.37
DenseNet121	7,238,212	96.53	96.30	96.18	96.24	8,816,708	99.34	99.18	99.39	99.28
DenseNet169	12,969,028	96.75	96.48	96.58	96.53	14,547,524	99.36	99.17	99.49	99.30
DenseNet201	18,698,308	96.49	96.35	96.10	96.22	20,276,804	99.36	99.14	99.41	99.30
Mean		96.35	96.24	96.23	96.23		99.47	99.33	99.49	99.43
St. Dev.		0.75	0.46	0.47	0.46		0.23	0.30	0.21	0.24

Xception + Probe info (proposed)

Actual \ Predicted	Normal	COVID-19	Pneumonia	Other
Normal	448	1	0	0
COVID-19	0	1845	0	0
Pneumonia	0	0	860	0
Other	0	0	0	1400

EfficientNetB4

Actual \ Predicted	Normal	COVID-19	Pneumonia	Other
Normal	448	1	0	0
COVID-19	2	1843	0	0
Pneumonia	0	0	799	61
Other	0	0	81	1319

Fig. 3: Confusion matrices for the best performing proposed model (Xception + Probe information) and the best performing base CNN model (EfficientNetB4).

and "Other" (4-class problem). Classification performance was measured using the following metrics: classification accuracy, precision, recall (sensitivity), and F1-score. Furthermore, since precision, recall, and F1-score depend on the class that is considered as positive, they were computed for all classes separately and the average across the four classes was reported as the final metric value. It must be noted that all experiments were carried out using the TensorFlow library, the Keras API, and the Python programming language on a Tesla K40c and a GeForce GTX Titan GPU. Additionally, all LUS images were resized to 224×224 pixels before being fed as input to the examined models.

A. Experimental results

The classification performance achieved using the baseline and the proposed approaches is reported in TABLE I, in terms of the classification accuracy, precision, recall, and F1-score metrics. From this table, it is evident that the fusion of the probe information and the image-based CNN features consistently led to an improvement in all classification metrics over the baseline CNN approach. The overall best performing model was the proposed "Xception + Probe information" that achieved an F1-score of 99.97%, an accuracy of 99.98%, a precision of 99.99%, and a recall of 99.94%, miss-classifying only one LUS image, as shown in Fig. 3. The proposed approach outperformed the baseline approach regardless of the base CNN model used, with the worst performing model, i.e. "MobileNetV2 + Probe information", achieving a higher F1-score compared to the best performing baseline model, i.e. EfficientNetB4, (99.27% vs. 96.58%). Among the baseline models, EfficientNetB4 achieved the best classification performance, with a classification accuracy of 96.81%, F1-score of 96.58%, precision of 96.47%, and recall of 96.70%. The confusion matrix for the EfficientNetB4 baseline model is depicted in Fig. 3. It is worth mentioning that although EfficientNetB4 achieved the best performance among the baseline models in terms of accuracy, F1-score, and recall, MobileNetV2 achieved a marginally higher precision (96.51% vs. 96.47%), although such a difference can be considered as negligible. The difference in classification metrics between the baseline and the proposed models is presented in TABLE II, whereas the confusion matrices for the best performing proposed and baseline models are depicted in Fig. 3.

B. Comparison to state of the art

In a recent work that used the examined COVIDx-US dataset, Adedigba et al. [33] reported a classification accuracy of 99.74% using MobileNetV2 [29] and 99.75% using SqueezeNet [34]. However, this work was conducted on a previous version of the dataset that contained a total of 174 LUS videos, compared to 220 contained in the current version (Version 1.5) that is used in this work. Furthermore and most importantly, in their work, Adedigba et al. [33] opted to simplify the examined 4-class problem and focus on the COVID-19 vs. Non-COVID-19 problem by grouping together all images that did not belong to the COVID-19 class and annotating them as Non-COVID-19. Consequently, the reported accuracy refers to the aforementioned binary classification problem. To evaluate our proposed approach and provide a

TABLE II: Difference (Δ) between the metrics' values achieved for the proposed method and the base CNN method.

Base CNN	Δ Accuracy	Δ Precision	Δ Recall	Δ F1
MobileNetV2	2.59	2.63	2.96	2.78
InceptionV3	2.85	2.73	2.72	2.73
InceptionResNetV2	2.95	2.95	3.66	3.33
Xception	5.60	4.96	4.70	4.86
ResNet50V2	3.14	3.38	3.35	3.37
EfficientNetB4	2.60	2.77	2.52	2.79
DenseNet121	2.81	2.88	3.21	3.04
DenseNet169	2.61	2.69	2.91	2.77
DenseNet201	2.87	2.79	3.31	3.08
Mean	3.11	3.09	3.26	3.19
St. Dev.	0.90	0.69	0.61	0.63

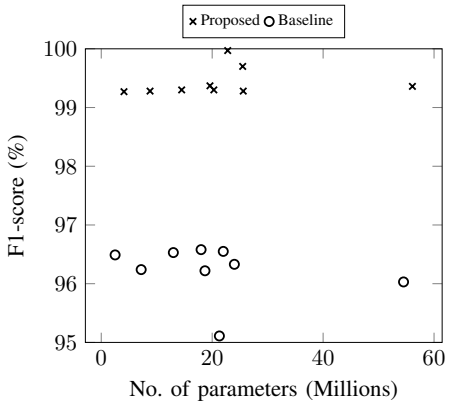


Fig. 4: F1-score (%) achieved for the proposed and the baseline models in relation to the number of parameters of each model.

fair evaluation against the Adedigba et al. [33] approach, we first replicated their experiment using the current version of the COVIDx-US dataset. Classification performance for the COVID-19 vs. Non-COVID-19 binary problem was indeed consistent with the one reported by Adedigba et al. [33]. However, when examined on the 4-class problem, the accuracy of the MobileNetV2 model dropped to 96.73%, as shown in TABLE I, whereas the proposed approach using MobileNetV2 as its base CNN achieved an enhanced classification accuracy of 99.32%.

C. Additional Discussion

It is clear from TABLE I and TABLE II that for each of the base models under consideration, the proposed technique outperformed the baseline approach in terms of all performance metrics. In addition, regardless of the base CNN model utilised, the proposed approach consistently offered better classification performance in terms of all the metrics. As shown in TABLE II, the introduction of the US probe information led to an average increase of $+3.11\%$ (± 0.90) in accuracy, $+3.09\%$ (± 0.69) in precision, $+3.26\%$ (± 0.61) in recall, and $+3.19\%$ (± 0.63) in F1-score for all the baseline CNN models. The results achieved for both the proposed and the baseline approach are quite stable, with the proposed models achieving an F1-score between 99.27%, using MobileNetV2 as the base CNN, and 99.97% using Xception as the base CNN, compared to the baseline models that achieved an F1-score between 95.11% for Xception and 96.58% for EfficientNetB4. The stability of the examined models is also evidenced by the low standard deviations computed across all models for all the examined performance metrics, as shown in the last row of TABLE I.

Considering the above, as well as the presented results, it is evident that the proposed multi-modal method of combining LUS image-based CNN features with information about the US probe type used to acquire the input LUS image can outperform models that rely solely on pre-trained CNN models for classifying LUS images into "COVID-19", "Normal", "Pneumonia", and "Other" (4-class problem). Furthermore, despite the proposed model based on the Xception CNN model achieving the best classification performance, it is evident from TABLE I that all models performed considerably well, thus providing the flexibility to the user to decide on the model to be used based on its computational complexity, by selecting the model with the lowest amount of parameters that achieves an acceptable performance, as shown in Fig. 4.

IV. Conclusion

In this work, we proposed a multi-modal approach for the classification of lung ultrasound images for the task of distinguishing them between "COVID-19", "Normal", "Pneumonia", and "Other (lung diseases/conditions)". The proposed approach relies on the fusion of image-based CNN features with information regarding the type of ultrasound probe (linear or convex) used to acquire the input image. Experimental results on a large lung ultrasound image dataset containing 22,776 lung ultrasound images demonstrated the superiority of the proposed approach compared to solely using the base CNN model for the classification of the images. The highest classification accuracy of 99.98% and highest F1-score of 99.97% were achieved using the proposed combination of probe information with the pre-trained Xception CNN model, compared to the highest accuracy of 96.81% and highest F1-score of 96.58% achieved using solely the EfficientNetB4 pre-trained CNN model. For future work, we plan to examine the incorporation to the model of more information related to the patients, such as various symptoms, as well as evaluate our models on additional lung ultrasound datasets that contain information about the type of probe used to acquire the images.

References

[1] A. Pagano, F. G. Numis, G. Visone, C. Pirozzi, M. Masarone, M. Olibet, R. Nasti, F. Schiraldi, and F. Paladino, "Lung ultrasound for diagnosis of pneumonia in emergency department," *Internal and emergency medicine*, vol. 10, no. 7, pp. 851–854, 2015.

[2] F. Mojoli, B. Bouhemad, S. Mongodi, and D. Lichtenstein, "Lung ultrasound for critically ill patients," *American journal of respiratory and critical care medicine*, vol. 199, no. 6, pp. 701–714, 2019.

[3] Y. Tung-Chen, "Lung ultrasound in the monitoring of covid-19 infection," *Clinical Medicine*, vol. 20, no. 4, p. e62, 2020.

[4] M. J. Fiala, "A brief review of lung ultrasonography in covid-19: is it useful?" *Annals of emergency medicine*, vol. 75, no. 6, pp. 784–785, 2020.

[5] Q.-Y. Peng, X.-T. Wang, and L.-N. Zhang, "Findings of lung ultrasonography of novel corona virus pneumonia during the 2019–2020 epidemic," *Intensive care medicine*, vol. 46, no. 5, pp. 849–850, 2020.

[6] 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.

[7] G. I. Okolo, S. Katsigiannis, T. Althobaiti, and N. Ramzan, "On the use of deep learning for imaging-based COVID-19 detection using chest x-rays," *Sensors*, vol. 21, no. 17, 2021.

[8] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, 2021, (Early Access).

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[10] 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.

[11] 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*, vol. 39, no. 8, pp. 2676–2687, 2020.

[12] J. Born, N. Wiedemann, M. Cossio, C. Buhre, G. Brändle, K. Leidermann, J. Goulet, A. Aujayeb, M. Moor, B. Rieck *et al.*, "Accelerating detection of lung pathologies with explainable ultrasound image analysis," *Applied Sciences*, vol. 11, no. 2, p. 672, 2021.

[13] J. Diaz-Escobar, N. E. Ordóñez-Guillén, S. Villarreal-Reyes, A. Galaviz-Mosqueda, V. Kober, R. Rivera-Rodriguez, and J. E. L. Rizk, "Deep-learning based detection of COVID-19 using lung ultrasound imagery," *Plos one*, vol. 16, no. 8, p. e0255886, 2021.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. on Learning Representations (ICLR)*, 2015.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[16] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[18] N. Awasthi, A. Dayal, L. R. Cenkeramaddi, and P. K. Yalavarthy, "Mini-COVIDNet: efficient lightweight deep neural network for ultrasound based point-of-care detection of COVID-19," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 6, pp. 2023–2037, 2021.

[19] H. Che, J. Radbel, J. Sunderram, J. L. Nosher, V. M. Patel, and I. Hacihaliloglu, "Multi-feature multi-scale CNN-derived COVID-19 classification from lung ultrasound data," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 2618–2621.

[20] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D. F. Briganti, S. Perlini, E. Torri, A. Mariani, E. E. Mossolani *et al.*, "Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method," *Journal of Ultrasound in Medicine*, vol. 39, no. 7, pp. 1413–1419, 2020.

[21] J. Roberts and T. Tsiligkaridis, "Ultrasound diagnosis of COVID-19: robustness and explainability," *arXiv preprint arXiv:2012.01145*, 2020.

[22] F. Sadik, A. G. Dastider, and S. A. Fattah, "SpecMEn-DL: spectral mask enhancement with deep learning models to predict COVID-19 from lung ultrasound videos," *Health Information Science and Systems*, vol. 9, no. 1, pp. 1–11, 2021.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[24] W. Zheng, L. Yan, C. Gou, Z.-C. Zhang, J. J. Zhang, M. Hu, and F.-Y. Wang, "Pay attention to doctor–patient dialogues: multi-modal knowledge graph attention image-text embedding for covid-19 diagnosis," *Information Fusion*, vol. 75, pp. 168–185, 2021.

[25] Z. Hu, Z. Liu, Y. Dong, J. Liu, B. Huang, A. Liu, J. Huang, X. Pu, X. Shi, J. Yu *et al.*, "Evaluation of lung involvement in COVID-19 pneumonia based on ultrasound images," *BioMedical Engineering OnLine*, vol. 20, no. 1, pp. 1–15, 2021.

[26] A. Ebadi, P. Xi, A. MacLean, S. Tremblay, S. Kohli, and A. Wong, "COVIDx-US - an open-access benchmark dataset of ultrasound imaging data for AI-driven COVID-19 analytics," *arXiv:2103.10003*, 2021.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[28] S. Muny, "What probe do I need for my ultrasound system?" 2019, Accessed 11/08/2022. [Online]. Available: https://www.probomedical.com/blog/what-probe-do-i-need-for-my-ultrasound-system/

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[30] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. on Artificial Intelligence*, 2017, p. 4278–4284.

[31] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th International Conference on Machine Learning (ICML)*, 2019.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.

[33] A. P. Adedigba and S. A. Adeshina, "Deep learning-based classification of COVID-19 lung ultrasound for tele-operative robot-assisted diagnosis," in *2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)*. IEEE, 2021, pp. 1–6.

[34] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1mb model size," *arXiv preprint, arXiv:1602.07360*, 2016.