Lung Ultrasound Image Classification using Deep Learning and Histogram of Oriented Gradients Features for COVID-19 Detection

E. A. Nehary, Sreeraman Rajan, and Carlos Rossa
Department of Systems and Computer Engineering
Carleton University, Ottawa, Canada
ebrahimali@cmail.carleton.ca; sreeramanr@sce.carleton.ca; rossa@sce.carleton.ca

2023 IEEE Sensors Applications Symposium (SAS) | 979-8-3503-2307-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/SAS58821.2023.10254002

***Abstract*—Features derived by deep learning models are useful in challenging classification tasks such as the detection of COVID-19 in ultrasound lung images. However, in this process, knowledge-based hand-crafted features engineered by humans are totally neglected. Hand-crafted features do have a significant role in performance enhancement in complicated classification tasks. This paper proposes a fusion of hand-crafted features with abstract features produced by deep learning in the later stages of the classification process for COVID-19 detection. Histogram of Oriented Gradients (HOG) features are used as a hand-crafted feature for fusion with abstract features produced by VGG16 and Vision transformer (ViT). The HOG and abstract features are fused later to improve the classification model's performance. A public COVID-19 dataset is used to demonstrate the improved performance of the proposed classification model. Results show that the proposed fusion technique improves the accuracy achieved by ViT for normal *vs* abnormal classification by 1.75% and for COVID-19 *vs* bacterial pneumonia classification by 1.04% when compared with the traditional ViT classification results. Similarly, when abstract features produced by VGG16 were fused with HOG features, the classification accuracy achieved an improvement of approximately 5.81% and 4.94% for normal *vs* abnormal and COVID-19 *vs* bacterial pneumonia classification, respectively.**

***Index Terms*—Deep learning, Vision transformer, Histogram of Oriented Gradients, COVID-19, Ultrasound images**

I. Introduction

The global COVID-19 pandemic has infected more than half a billion people by spreading either directly or indirectly through droplet transmission, resulting in more than 6 million deaths worldwide [1]. The reduction of mortality and transmission rates of the virus are facilitated by early detection and monitoring of carriers of the COVID-19 virus [2]. The current method employed for COVID-19 detection is the reverse transcription-polymerase chain reaction (RT-PCR), however, it has been reported to have low sensitivity for diagnostic purposes [3]. Computed tomography (CT), although an effective alternative with better sensitivity to detect COVID-19 [4], is expensive and not a viable option as it requires sterilization between scans of different patients [3], [5].

Ultrasound (US) imaging is a better alternative to CT or X-ray as it is portable, easy to disinfect, inexpensive, and non-ionizing [6], [7]. Yet, COVID-19 detection is a challenging task because US images are noisy and have low resolution. Realizing this, several different classification models were proposed in the literature to detect COVID-19 in lung US (LUS) images using machine learning and deep learning algorithms. Several deep learning models, including transfer learning, were proposed for COVID-19 classification in [8]. Deep learning models such as convolutional neural networks (CNN), Linear transformers, Vision transformers, EfficientNet, and gated multilayer perceptron (gMLP) models have been proposed for the identification of COVID-19 infected LUS images [9], [10], [11], [12]. There are several review papers that present the state-of-the-art in COVID-19 detection using LUS images [13], [14].

In all the above-mentioned works, abstract features were automatically extracted by the model and used for COVID-19 classification. Such abstract features often provide an adequate representation of the input image to the classification system. Recently, in the field of radars' application, handcrafted features along with abstract features extracted by deep learning network were used for classifying ships in maritime synthetic aperture radar images [15]. Motivated by [15], in this paper, a fusion of abstract features and handcrafted features for identifying COVID-19 LUS images is proposed in this paper. The premise of this paper is that handcrafted features add user knowledge or expertise into the classification problem and thus help in obtaining better classification results. This is the first work that recognizes the role of handcrafted features in deep learning models for COVID-19 classification and proposes the fusion of abstract features derived from deep learning and handcrafted features.

The paper is organized as follows. In the following section the methodology adopted in this paper is presented. Results are presented and discussed in Section III, and conclusions and future works are presented in Section IV.

II. Methodology

The proposed framework depicted in Figure 1 comprises of the fusion of two streams corresponding to the output of the deep learning model and the HOG features respectively. The deep learning block provides abstract features obtained

by pretrained VGG16 or vision transformers (ViT). The dimensionality of the HOG features is reduced using principal component analysis (PCA) block. Finally, the fusion model is used to fuse the abstract feature $F1$ with the HOG feature $F2$ for the final classification as shown in Figure 2. The following subsections explain briefly the HOG features, the deep learning model, and the fusion block used in this work.

1) HOG features: Histograms of oriented gradients (HOG) features have been proposed in [16] for human detection in image processing applications. HOG features are descriptors that depend on the gradient magnitude and orientations in localized portions of an image. The gradients mainly exist in the peripheral regions of images, making them suitable for characterizing the shape and local appearance of an object. In this work, HOG is utilized to extract features from the A-line, the B-line, the pleural line, and the consolidation from ultrasound images. Lines and consolidations carry useful information that improves classification performance. To the best of our knowledge, this is the first time that HOG features of ultrasound images are employed for classification tasks. For the sake of completeness, the various steps involved in HOG feature extraction are given: (i) The given input image is resized to 128×128 using a bilinear interpolation algorithm [17]; (ii) the amplitude $G(i,j)$ and orientation $\theta(i,j)$ associated with the gradient of the resized image are obtained as follows:

$$G(i,j) = \sqrt{G_x^2(i,j) + G_y^2(i,j)}, \tag{1}$$

where $G_x(i,j)$ and $G_y(i,j)$ are the gradient amplitudes in x and y directions respectively, given as

$$G_x(i,j) = I(i+1,j) - I(i-1,j),$$
$$G_y(i,j) = I(i,j+1) - I(i,j-1),$$

where $I(i,j)$ indicates gray intensity in the i-th row and j-th column of an image. The gradient orientation, ranging from $0 - 360^o$, is given as

$$\theta(i,j) = \arctan\left[\frac{G_y(i,j)}{G_x(i,j)}\right]; \tag{2}$$

(iii) The gradient matrices, namely amplitude and angle, are divided into 8×8 cells, containing 64 pixels each. For each cell, a 9-bin gradient histogram is computed where each bin has an angle range of 20^o. The output histogram from each cell has a size of 9×1; (iv) A block is constructed using four 8×8 cells, and then the gradient histogram of each block has a size of 36×1 and it is normalized to reduce the effect of noise and contrast variation for a specific object in the image; (v) Feature descriptors for each block are produced by concatenating gradient histograms of each cell.

The feature descriptor for the image is integrated into the gradient histogram of each block as shown in Figure 3 for COVID-19 and normal LUS images. The length of the HOG features is 8100. The HOG features for all the training images are collected to form a matrix with 8100 columns. PCA of the matrix of HOG features is done for dimensionality reduction [18]. Dimensionality reduction has been recommended to avoid overfitting issues [15]. PCA is summarized as follows. Features are normalized by subtracting the mean and dividing by the standard deviation. Then, the covariance matrix for each feature pair is calculated. Eigen decomposition of the covariance matrix is obtained to determine the principal component of the data. k eigenvectors have been selected based on the largest k eigenvalues. In our work $k = 64$, thus the HOG features reduce from 8100 to 64.

2) Deep learning: Deep learning models have achieved remarkable results in a variety of applications, including image classification. Deep learning automatically extracts useful information from the input images called abstract features. It feeds these features into fully connected layers with sigmoid activation functions for final classification. In this paper, a pretrained VGG16 model [19] and a ViT [20] are considered for extracting abstract features.

i. **Pretrained VGG16:** Transfer learning method is used due to the limited nature of dataset. Transfer learning is a method in which a deep learning model is trained on a large dataset like ImageNet, and then the trained model is reused as a starting point for the next task. In this paper, the VGG16 model [19] trained using the ImageNet dataset for more than 1000 classes was used. Pre-existing fully connected (FC) layers were removed and replaced by the following layers in succession: (i) global average pooling; (ii) FC (with 512 neurons); (iii) drop-out layer (with ratio of 0.5) and (iv) FC with 2 neurons producing two classes.
   All convolution layers except the ones in the last block are frozen. The last block and the new FC layers are trained using ultrasound image data. Finally, *all* the convolutional layers are frozen, and the new FC layers are removed to provide the abstract features. These features are fed into the fusion block for the final decision.

ii. **Vision transformer (ViT):** The ViT model [20] consists of three primary components; linear embedding, transformer encoding, and final classification. Initially, the input image is divided into non-overlapping patches by the model. Each patch is flattened and fed to a linear embedding layer to compute learnable embedding features concatenated with positional information. The output of the linear embedding layer is then fed to the transformer encoding layer. The transformer encoding layer includes a multi-head self attention layer (MSA), a multi-layer perceptron (MLP), and layer norm (LN). Self-attention is used by MSA to find the connection between different patches in the input image. Finally, the output from the transformer encoder is transferred to the final classification layer. The final classification layer is an FC layer with a softmax activation. More details can be found in [20], [21].

3) Fusion block: Abstract features $F1$ are fused with the HOG features $F2$ using the fusion model shown in Figure 2. First, F1 and F2 are projected using the FC layer with 256

neurons. The projected features are concatenated and fed to three FC layers with 128, 64, and 2 neurons. The two largest FC layers are followed by Gaussian error linear units (Gelu) as activation functions. The final, smallest FC layer is followed by softmax activation function for classification. There are 2 neurons in the last convolutional layer, which correspond to the number of classes.

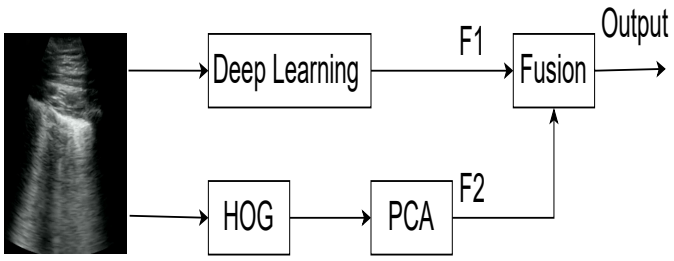

Fig. 1: Overall block diagram of the proposed model.

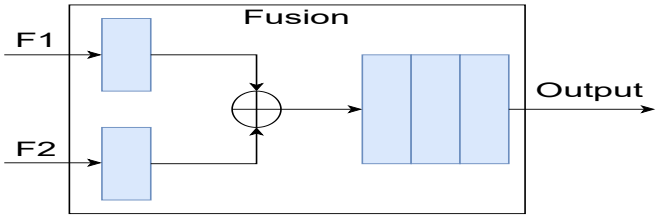

Fig. 2: Block diagram of fusion block for combining abstract features F1 and HOG features F2. Concatenation is indicated as $\oplus$ and the fully connected layers by blue rectangles.

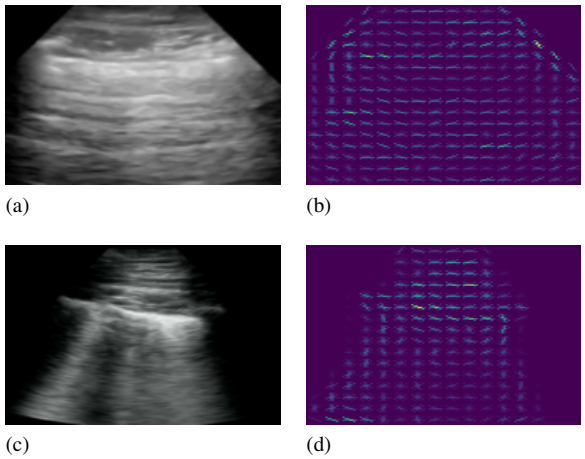

Fig. 3: Illustrations of HOG features descriptors. (a) and (c) are the normal LUS image and the COVID-19 LUS image, respectively, while their HOG features descriptors are shown in (b) and (d).

III. Results

The performance of the proposed model has been evaluated using the lung ultrasound images (LUS) dataset provided publicly by [5], [22]. This data set was gathered from various resources such as Northumbria, Neruppin, publication, GrepMed, Butterfly, education platforms such as the point care of ultraound (POCUS) Altas, and life in fast line (LIFFL) and consequently, there is heterogeneity in the database due to the varying formats, brightness and type of sensors used for data acquisition. Northumbria is a healthcare foundation that serves a population in the northeast region of the United Kingdom. As part of their data collection process, a GE Healthcare convex probe ultrasound machine called the $Venue^{TM}$, which operates within a frequency range of $(2-5)$ MHz was used. The foundation also implemented the BLUE protocol [23] to acquire the data. This protocol is particularly effective in identifying pulmonary pathologies like B-lines and pleural lines, which are crucial in diagnosing COVID-19 [24]. To confirm cases of COVID-19, the foundation relied on RT-PCR tests, while X-ray or CT scans were utilized for detecting bacterial pneumonia. The dataset consists of 70 videos that are collected from 44 males and 26 females. Neuruppin data was acquired at Theodor Fontane Medical School in Neuruppin, Germany. Like Northumbria, Neuruppin utilized a GE Healthcare ultrasound machine and the BLUE protocol. Neuruppin provided 31 videos, of which 28 videos were obtained using a convex probe and 3 videos were obtained using a linear probe. The US data provided by Neuruppin and Northumbria accounts for approximately half of the dataset. GrepMed is a searchable medical image repository that provides 9 videos acquired with a convex probe and 3 videos acquired with a linear probe. On the other hand, Butterfly is a portable ultrasound machine that connects to a smartphone and uses a convex probe to collect ultrasound data. Using this, a total of 18 videos were obtained. The POCUS Atlas, an ultrasound education platform, provided 8 videos collected with a convex probe and 2 videos collected with a linear probe. Additionally, the education platform LITFL provides 5 videos that were acquired using a convex probe. Various publications and websites contributed to the collection as well, providing 45 videos obtained with a convex probe and 10 videos obtained with a linear probe.

As a result, the whole dataset consists of a total of 202 videos. These videos were categorized into COVID-19, bacterial pneumonia, and normal (healthy) lung images. Sample images from each class (COVID-19, bacterial pneumonia, and normal (healthy) lung) are shown in Figure 4, where the first column represents the COVID-19 class, the second column depicts bacterial pneumonia, and the third column shows the healthy lung. It is observed that there is a difference within the same class due to data collection from various sources. Furthermore, only Neuruppin and Northumbria utilized the BLUE protocol to acquire US data while the protocol followed by other resources is unknown. Recently, various protocols have been proposed specifically for the detection of COVID-19, as reported in [3], [25]–[27]. These protocols have shown effective detection of COVID-19. The placement of the probe in the patient, along with the timing of acquisition and frequency of the US probe, play a key role during the collection of

ultrasound videos. These will influence the quality and clarity of the pulmonary artifacts such as A-lines and B-lines. These artifacts are helpful in detecting and classifying COVID-19 [25]–[27]. However, no publicly available data acquired using a single sensor, following an approved protocol is accessible for our work. Therefore, the focus of our work is to develop a model that can improve classification performance using this challenging dataset. Consequently, our proposed model is expected to enhance the classification performance using the collected data that follows the specific protocol for COVID-19 when it becomes available.

In order to train the model, individual frames were extracted from each video and used as input for training the model. The frames extracted from the videos were based on the assumption that all videos have a 3 Hz frame rate with a maximum of 30 frames per video. Of the 5460 US frames extracted from the videos, 2061 were attributed to COVID-19, 1944 to bacterial pneumonia, and 1455 to the healthy lung images. Due to the small dataset, five-fold cross-validation was used to train and test the proposed model. It may be noted that the set of extracted frames from each video appears either in the training fold or in the test folds (and not in both), thus avoiding any possible correlation between the extracted frames from the same video. The US images from the datasets were resized to 128×128 pixels using bilinear interpolation. All models are trained using a variant of Adam optimizer (AdamW) with a weight decay of 0.0001 and a learning rate of 0.001 for 100 epochs. The binary cross entropy is used as a loss function for VGG16 while sparse categorical cross entropy was used for ViT. Softmax activation function is used in the final layer for classification. Details of other hyperparameters used for ViT are given in Table I. The class imbalance in the dataset was handled using class weights.

The performance of the proposed fusion model for COVID-19 classification was evaluated using accuracy, sensitivity, precision, and F1_ score. The classification results shown in Table II are for normal (healthy) *vs* abnormal (COVID-19 and bacterial pneumonia) classification and the classification results shown in Table III are for COVID-19 *vs* bacterial pneumonia classification. It is clear that when HOG features are fused with abstract features from ViT or VGG16 the results are improved. When the features from ViT and HOG features are fused, the accuracy for normal and abnormal classification is enhanced approximately by 2% while when the features from VGG16 and HOG features are fused, the accuracy improved approximately by 6%. COVID-19 and bacterial pneumonia classifications also improved by fusing HOG features with abstract features as shown in Table III. Accuracy improved by nearly 5% and 1% for VGG16 and ViT, respectively. This result demonstrates that HOG features complement abstract features from deep learning algorithms and therefore provide better classification. Moreover, HOG features yield competitive results to deep learning models as demonstrated in Tables IV and V for normal versus abnormal classification and COVID-19 versus bacterial pneumonia classification using traditional classifiers. Out of all implemented traditional classifiers, support vector machine (SVM) with radial basis function (RBF) kernel provided the best accuracy of 88.43% with HOG features as input for normal and abnormal classifications. SVM with RBF also provided the best accuracy of 85.34% for distinguishing COVID-19 from bacterial pneumonia.

TABLE I: Parameters for training ViT.

Parameter	Value
Image size	128
Patch size	32
Projection dimension	32
Heads numbers	4
Transformer layers	8
MLP head units	512, 256

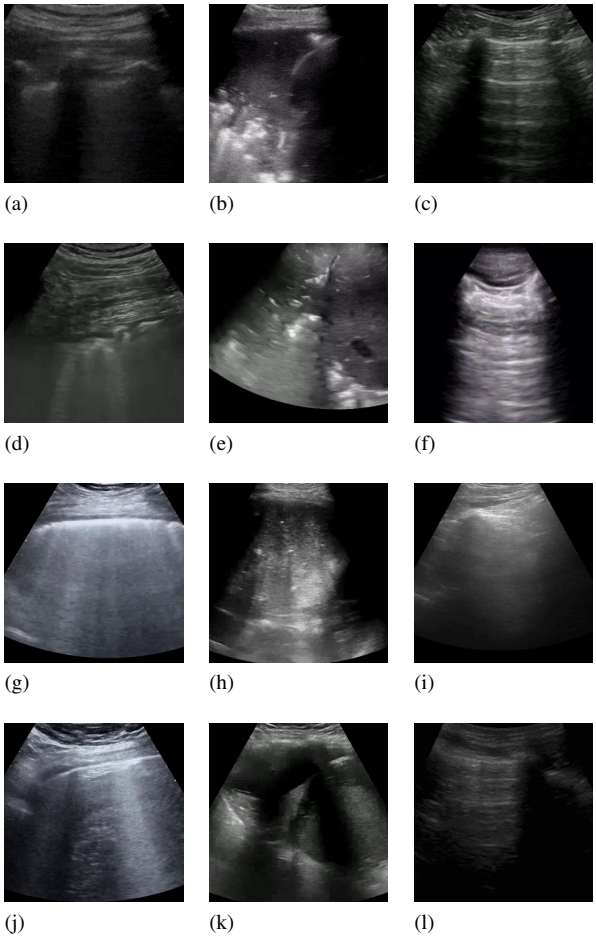

Fig. 4: Samples of LUS images of COVID-19 are shown in the first column, bacterial pneumonia in the second column, and normal (healthy) lungs in the third column.

IV. Conclusions and Future Works

Deep learning fails to give good classification accuracy for low-resolution, noise-like ultrasound images [8]. Hand-crafted

TABLE II: Normal *vs* abnormal classification ($\times 100\%$) of ViT and VGG16 with and without HOG features.

Models	Sensitivity	Specificity	Precision	F1_score	Accuracy
ViT	0.8364	0.8364	0.8434	0.8365	0.8528
ViT and HOG	0.8450	0.8450	0.8689	0.8529	0.8703
VGG16	0.7416	0.7416	0.7225	0.7252	0.7939
VGG16 and HOG	0.8205	0.8205	0.8666	0.8292	0.8520

TABLE III: COVID-19 *vs* bacterial pneumonia classification ($\times 100\%$) of ViT and VGG16 with and without HOG features.

Models	Sensitivity	Specificity	Precision	F1_score	Accuracy
ViT	0.9024	0.9024	0.8975	0.8973	0.8992
ViT and HOG	0.9076	0.9076	0.9070	0.9068	0.9096
VGG16	0.8139	0.8139	0.8143	0.8114	0.8169
VGG16 and HOG	0.8653	0.8653	0.8810	0.8617	0.8663

features complement abstract features obtained from deep-learning algorithms to achieve better classification results. In this work, the fusion of abstract features from a deep-learning-based algorithm with the hand-crafted feature was proposed. HOG features were used as hand-crafted features. The results demonstrate the superiority of such feature fusion.

This paper has provided only preliminary results for the fusion of hand-crafted features and abstract features from deep learning methods. Systematic experiments need to be carried out to further understand the advantages and disadvantages of the proposed method. Future work will investigate the fusion of spatial features like Haar, Canny edge detector, SIFT, KAZE, and AKAZE with abstract features derived from different deep learning models. Different fusion methods may also be considered.

V. Compliance with ethical standards

This study was conducted retrospectively using human subject data made available in open access by the LUS dataset.

VI. Acknowledgments

This work was financially supported by Natural Sciences and Engineering Research Council of Canada (NSERC). The authors thank Mr. Alec Cotton and Ms. Ankita Dey for proofreading the document.

TABLE IV: Normal *vs* abnormal classification ($\times 100\%$) of the traditional classifiers with HOG features.

Models	Sensitivity	Specificity	Precision	F1_score	Accuracy
Random Forest	0.6615	0.6615	0.8610	0.6651	0.7586
SVM (Linear kernel)	0.5000	0.5000	0.3220	0.3917	0.6439
SVM (RBF kernel)	0.8573	0.8573	0.8979	0.8677	0.8843
Decision Tree	0.7636	0.7636	0.7784	0.7656	0.7885
Extra Tree	0.7676	0.7676	0.8764	0.7828	0.8276
KNN	0.8043	0.8043	0.8311	0.8081	0.8303
MLP	0.8606	0.8606	0.8887	0.8663	0.8811
LogisticRegressionCV	0.8552	0.8552	0.8864	0.8620	0.8789
PassiveAggressive	0.7273	0.7273	0.8107	0.7112	0.7640
LDA	0.8505	0.8505	0.8879	0.8588	0.8763
QDA	0.8111	0.8111	0.8838	0.8272	0.8556
Adaboost	0.7925	0.7925	0.8287	0.8021	0.8290
Bagging	0.7610	0.7610	0.7984	0.7693	0.7991
GradientBoosting	0.7824	0.7824	0.8483	0.7954	0.8296
HistGradientBoosting	0.7862	0.7862	0.8462	0.7982	0.8304
Voting	0.7941	0.7941	0.8936	0.8122	0.8482
Stacking Classifier	0.8369	0.8369	0.8964	0.8502	0.8720

TABLE V: COVID-19 *vs* bacterial pneumonia classification ($\times 100\%$) of the traditional classifiers with HOG features.

Models	Sensitivity	Specificity	Precision	F1_score	Accuracy
Random Forest	0.7304	0.7304	0.8387	0.7349	0.7732
SVM (Linear kernel)	0.8260	0.8260	0.8794	0.8323	0.8455
SVM (RBF kernel)	0.8474	0.8474	0.8712	0.8471	0.8534
Decision Tree	0.7711	0.7711	0.7946	0.7768	0.7885
Extra Tree	0.8182	0.8182	0.8703	0.8251	0.8389
KNN	0.8174	0.8174	0.8533	0.8209	0.8322
MLP	0.7990	0.7990	0.8251	0.8018	0.8117
PassiveAggressive	0.5365	0.5365	0.4508	0.4221	0.5735
LDA	0.8136	0.8136	0.8386	0.8168	0.8253
QDA	0.8244	0.8244	0.8469	0.8269	0.8359
Adaboost	0.7992	0.7992	0.8257	0.8039	0.8149
BaggingClassifier	0.7683	0.7683	0.8092	0.7764	0.7926
GradientBoosting	0.8077	0.8077	0.8453	0.8125	0.8248
HistGradientBoosting	0.8184	0.8184	0.8481	0.8240	0.8343
Voting	0.7877	0.7877	0.8553	0.7965	0.8155
Stacking Classifier	0.8386	0.8386	0.8759	0.8415	0.8509

References

[1] "Centers for disease control and prevention (CDC)." Available at https://coronavirus.jhu.edu/map.html. Accessed: 2023-05-01.

[2] W. Xue *et al.*, "Modality alignment contrastive learning for severity assessment of COVID-19 from lung ultrasound and clinical information," *Medical Image Analysis*, vol. 69, p. 101975, 2021.

[3] C. McDermott, M. Łącki, B. Sainsbury, J. Henry, M. Filippov, and C. Rossa, "Sonographic diagnosis of COVID-19: A review of image processing for lung ultrasound," *Frontiers in Big Data*, vol. 4, p. 612561, 2021.

[4] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, 2020.

[5] J. Born *et al.*, "Accelerating detection of lung pathologies with explainable ultrasound image analysis," *Applied Sciences*, vol. 11, no. 2, p. 672, 2021.

[6] Diaz-Escobar *et al.*, "Deep-learning based detection of COVID-19 using lung ultrasound imagery," *Plos One*, vol. 16, no. 8, p. e0255886, 2021.

[7] S. Desai, A. Pareek, and M. Lungren, "Deep learning and its role in COVID-19 medical imaging," *Intelligence-Based Medicine*, vol. 3, p. 100013, 2020.

[8] N. Awasthi, A. Dayal, L. Cenkeramaddi, and P. 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.

[9] S. Perera, S. Adhikari, and A. Yilmaz, "Pocformer: A lightweight transformer architecture for detection of COVID-19 using point of care ultrasound," in *IEEE International Conference on Image Processing (ICIP)*, pp. 195–199, 2021.

[10] M. Rahhal, Y. Bazi, R. Jomaa, M. Zuair, and F. Melgani, "Contrasting EfficientNet, ViT, and gMLP for COVID-19 detection in ultrasound imagery," *Journal of Personalized Medicine*, vol. 12, no. 10, p. 1707, 2022.

[11] G. Muhammad and M. Hossain, "COVID-19 and non-COVID-19 classification using multi-layers fusion from lung ultrasound images," *Information Fusion*, vol. 72, pp. 80–88, 2021.

[12] C. Chakraborty and A. Abougreen, "Intelligent internet of things and advanced machine learning techniques for COVID-19," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 7, no. 26, 2021.

[13] J. Wang *et al.*, "Review of machine learning in lung ultrasound in COVID-19 pandemic," *Journal of Imaging*, vol. 8, no. 3, p. 65, 2022.

[14] L. Zhao and L. Bell, "A review of deep learning applications in lung ultrasound imaging of COVID-19 patients," *BME Frontiers*, vol. 2022, 2022.

[15] T. Zhang *et al.*, "HOG-ShipCLSNet: A novel deep learning network with

HOG feature fusion for SAR ship classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–22, 2021.
[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, 2005.
[17] M. Mastyło, "Bilinear interpolation theorems and applications," *Journal of Functional Analysis*, vol. 265, no. 2, pp. 185–207, 2013.
[18] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 17–32, 1981.
[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[20] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[22] J. Born *et al.*, "L2 accelerating COVID-19 differential diagnosis with explainable ultrasound image analysis: An AI tool," *Thorax*, vol. 76, no. Suppl 1, pp. A230–A231, 2021.
[23] D. Lichtenstein, "BLUE-protocol and FALLS-protocol: Two applications of lung ultrasound in the critically ill," *Chest*, vol. 147, no. 6, pp. 1659–1670, 2015.
[24] K. Jackson, R. Butler, and A. Aujayeb, "Lung ultrasound in the COVID-19 pandemic," *Postgraduate Medical Journal*, vol. 97, no. 1143, pp. 34–39, 2021.
[25] G. Soldati *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.
[26] T. Perrone *et al.*, "A new lung ultrasound protocol able to predict worsening in patients affected by severe acute respiratory syndrome coronavirus 2 pneumonia," *Journal of Ultrasound in Medicine*, vol. 40, no. 8, pp. 1627–1635, 2021.
[27] L. Demi *et al.*, "New international guidelines and consensus on the use of lung ultrasound," *Journal of Ultrasound in Medicine*, vol. 42, no. 2, pp. 309–344, 2023.