

Article

Contrasting EfficientNet, ViT, and gMLP for COVID-19 Detection in Ultrasound Imagery

Mohamad Mahmoud Al Rahhal ¹, Yakoub Bazi ^{2,*}, Rami M. Jomaa ³, Mansour Zuair ² and Farid Melgani ⁴

¹ Applied Computer Science Department, College of Applied Computer Science, King Saud University, Riyadh 11543, Saudi Arabia

² Computer Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

³ Computer Science Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

⁴ Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy

* Correspondence: ybazi@ksu.edu.sa

Abstract: A timely diagnosis of coronavirus is critical in order to control the spread of the virus. To aid in this, we propose in this paper a deep learning-based approach for detecting coronavirus patients using ultrasound imagery. We propose to exploit the transfer learning of a EfficientNet model pre-trained on the ImageNet dataset for the classification of ultrasound images of suspected patients. In particular, we contrast the results of EfficientNet-B2 with the results of ViT and gMLP. Then, we show the results of the three models by learning from scratch, i.e., without transfer learning. We view the detection problem from a multiclass classification perspective by classifying images as COVID-19, pneumonia, and normal. In the experiments, we evaluated the models on a publically available ultrasound dataset. This dataset consists of 261 recordings (202 videos + 59 images) belonging to 216 distinct patients. The best results were obtained using EfficientNet-B2 with transfer learning. In particular, we obtained precision, recall, and F1 scores of 95.84%, 99.88%, and 97.41%, respectively, for detecting the COVID-19 class. EfficientNet-B2 with transfer learning presented an overall accuracy of 96.79%, outperforming gMLP and ViT, which achieved accuracies of 93.03% and 92.82%, respectively.

Keywords: coronavirus; classification; deep learning; transfer learning; ultrasound images



Citation: Rahhal, M.M.A.; Bazi, Y.; Jomaa, R.M.; Zuair, M.; Melgani, F. Contrasting EfficientNet, ViT, and gMLP for COVID-19 Detection in Ultrasound Imagery. *J. Pers. Med.* **2022**, *12*, 1707. <https://doi.org/10.3390/jpm12101707>

Academic Editor: Amir Hajjam El Hassani

Received: 20 July 2022

Accepted: 10 October 2022

Published: 12 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

COVID-19 is an infectious disease caused by a virus from the coronavirus strain. Coronavirus is a phrase derived from the Latin word corona, which means crown. The World Health Organization (WHO) reports that the number of those infected with this virus is rising quickly. According to statistics released on 12 March 2021, more than 116 million instances and around 2.5 million deaths had already been confirmed. [1]. The most well-known symptoms of COVID-19 are fever, fatigue, and dry cough; other less-common symptoms can include pain, nasal congestion, and loss of taste and smell. The risk of serious complications is more significant among the elderly and people with health problems.

In addition to the ordinary real-time polymerase chain reaction (RT-PCR) test, medical images are progressively used for screening and monitoring the disease. Indeed, medical imaging such as ultrasound, computed tomography (CT), and X-ray are important elements in medical practice as they allow scientists to learn more about the disease in a noninvasive manner. Furthermore, the automatic analysis of these images using machine learning methods can greatly assist in monitoring the effectiveness of the treatment and adjusting protocols based on its severity. More recently, ultrasound imaging has also been used for disease screening as it has many advantages in terms of being relatively cost effective

compared to other imaging techniques such as CT, being safe, and providing efficient and immediate healthcare information [2].

Deep learning techniques have recently demonstrated their dominance over traditional approaches in numerous domains, such as computer vision and image classification [3,4]. These techniques have achieved encouraging results in classifying medical images [5,6]. Several research works have applied deep learning networks to detect COVID-19 using lung ultrasound (LUS) [7–13], CT scan [14–22], or CXR images [23–31]. The development of an efficient and accurate system for detecting COVID-19 is still challenging, and the need for detecting COVID-19 cases as soon as possible with high accuracy could save the rest of the community from this pandemic. Furthermore, the appearance of new variants of SARS-CoV-2 is encouraging researchers to develop and improve new systems that are able to detect patients infected by new variants.

The objective of this work is to design an automatic system for detecting COVID-19 using LUS. To this end, we employ several deep learning models. In the proposed models, we explore two learning approaches: transfer learning and learning from scratch. The former approach, transfer learning, is used in the case of a relatively limited number of training samples; therefore, we fine-tune a pre-trained model instead of training a new model from scratch as in the later approach. More specifically, we develop three well-known models—EfficientNet-B2 [32], gMLP [33], and ViT [34]—for classifying LUS images. We evaluate the efficiency of these models within a multiclass problem where the goal is to classify LUS images into COVID-19, pneumonia, or normal classes.

The main contributions of this work are as follows:

- We propose a deep learning-based model for the automatic detection of COVID-19 using LUS images, to increase the accuracy and the speed of detecting COVID-19 compared with the routine rRT-PCR test.
- Three deep learning models are proposed and evaluated: EfficientNet-B2, gMLP, and ViT.
- We explore the effectiveness of the proposed models in two learning approaches: transfer learning and learning from scratch.

The remainder of this article is organized as follows: Section 2 lists the main related works. In Section 3, we present a detailed description of the proposed deep learning models. Section 4 is dedicated to describing the dataset and presenting the experimental results obtained with the proposed approach and discussing our findings. Finally, Section 5 presents conclusions and future works.

2. Related Works

The literature on COVID-19 reports several methods for the analysis of medical images such as of CT images [35–41], X-ray [42–47], and LUS [7–10]. For instance, Silva et al. [35] proposed a voting-based approach, where the images from a given patient are classified using a group in a voting system. In [36], the authors implemented a bidirectional classification system using differential evolution algorithms. In [37], the authors proposed a contrastive learning method for jointly learning on heterogenous datasets. In [38], the authors proposed a multiscale feature fusion method for enhancing the detection accuracy. In [39], Zhou et al. proposed a method that allows the segmentation and identification of the infected region images from different sources. In another work [40], a method based on an adaptive feature selection deep guided forest method was proposed.

Similarly other approaches have been developed for disease detection using X-ray imagery [42–47]. For example, in [42], the authors developed a model for classifying images into three different classes: non-COVID, pneumonia, and COVID. The authors in [43] used different pre-trained Convolutional Neural Network (CNN) architectures for feature generation and investigated different classifiers to classify the extracted features. They found that the best results were obtained using MobileNet as a pre-trained CNN combined with a Support Vector Machine (SVM) classifier. The authors in [44] proposed a transfer learning approach based on decomposition techniques to detect the class boundaries.

In [45], the authors proposed a capsule network composed of four convolutional layers and three capsule layers for handling class imbalance problems. In [46], the authors proposed a COVID data-based network that combines segmentation and data augmentation in order to improve the detection accuracy. In [47], the authors suggested employing a bilateral low-pass filter and a histogram equalization technique to pre-process the images. Then, a pseudo-color image created from the original and filtered images is given progressively to a CNN model for classification.

In the context of detecting COVID-19 using LUS, the authors in [9] suggested a spatial transformer network that simultaneously predicts the severity of the disease and provides weakly supervised localization of pathological artifacts. They also presented a technique for frame score aggregation at the video-level based on uninorms. Their proposed model achieved an *F1* score of 71.4% in frame-based classification. The authors of [13] applied pre-trained residual CNN models (ResNet18/ResNet50). They evaluated their proposed models using a dataset with four to seven classes; their proposed models achieved good results, with an *F1* score of 98%. The author of [10] proposed a light deep learning model for detecting COVID-19 using ultrasound images. Their model achieved very good results in terms of the training time, but with a low overall accuracy of 83%.

The literature shows that it is indeed possible to develop a deep learning system for the automatic detection of COVID-19 using LUS, to improve the performance of the systems in terms of classification accuracy, which encourages us to present this study.

3. Materials and Methods

3.1. Dataset Description

In the experiments, we used the ultrasound dataset proposed in [8,11]. This dataset consists of 255 LUS recordings (196 videos and 59 images) belonging to 216 distinct patients. This dataset includes samples from COVID-19, bacterial pneumonia, and healthy patients, as indicated in Table 1 [8].

Table 1. Number of videos and images in the dataset per class and probes.

	Convex		Linear		Total
	#Video	#Image	#Video	#Image	
COVID-19	64	18	6	4	92
Bacterial pneumonia	49	20	2	2	73
Viral pneumonia	3	–	3	–	6
Normal	66	15	9	–	90
Total	182	53	20	6	261

The dataset was collected from different sources, including clinical information provided voluntarily by hospitals or ultrasound instructors at academic institutions, LUS recordings published in other scholarly works, community platforms, open medical repositories, and health-tech firms. The data were acquired using a variety of ultrasound devices with linear and convex probes. Because of their higher frequency, linear probes have a higher resolution, which makes it easier to analyze problems along the pleural line [48]. The linear probe does not analyze deeper lung tissue because it penetrates the tissue less than the convex probe, which can interfere with the distinction of B-lines [49].

Figure 1 shows examples of ultrasound images obtained from different probes and records, where the left two columns (a) represent images from several records captured using convex sensor, and the right two columns (b) represent images from different records captured using a linear sensor.

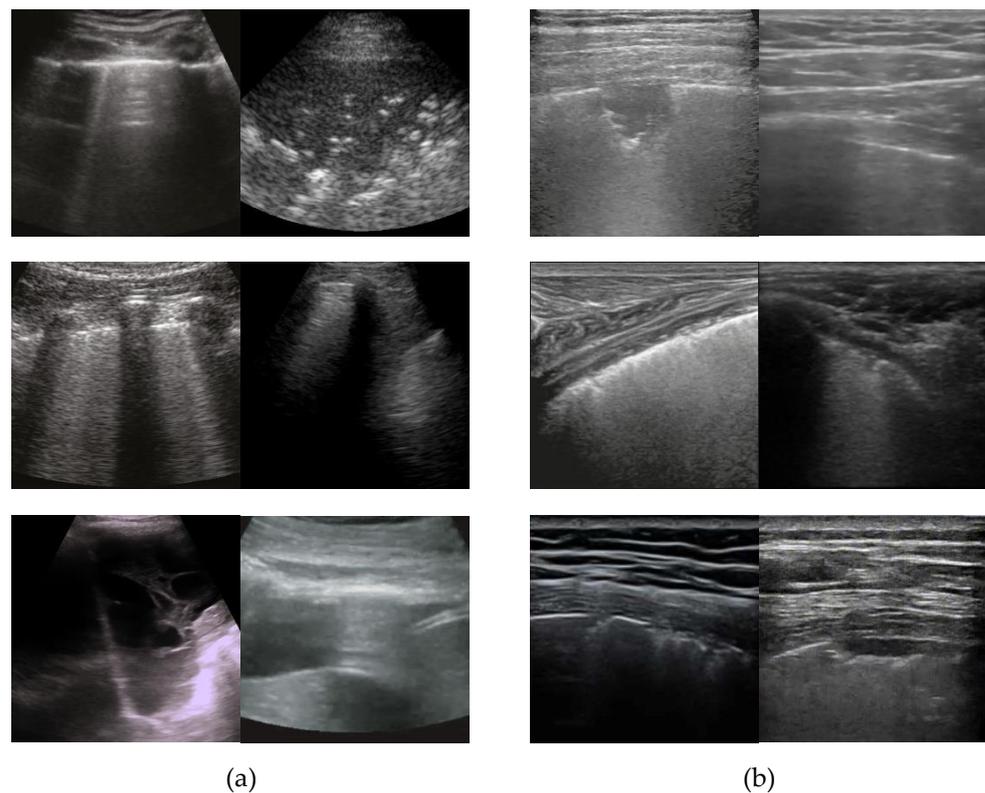


Figure 1. Examples of ultrasound images obtained from different probes and records using (a) a convex sensor and (b) a linear sensor.

3.2. Compound Scaling Network

The recent trend of image classification problems is training deep neural network architectures, mainly CNN, to predict the labels of test images. A CNN network is a cascade of several convolution layers that have trainable weights and biases. Each layer performs a convolution operation on the input data followed by optional nonlinear operation. Following the success of AlexNet [50] on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), CNNs have been successfully applied in various research domains, including medical image analysis. The performance of ConvNets has been significantly improved by training deeper architectures (network architectures with several convolution layers arranged in various ways) such as GoogleNet [51], ResNet [52], and more recently, GPipe [53]. These architectures perform scaling of the ConvNets by increasing the number of layers (depth increase), number of channels in each layer (width increase), or image resolution. EfficientNet [32] is currently the only model that performs scaling in all of the three dimensions in a principled way.

The authors of EfficientNet show that although scaling ConvNets in one of the three dimensions (depth, width, and image resolution) improves performance, the gain saturates quickly as the network becomes bigger. In order to overcome this, they proposed a compound scaling method that uniformly scales network depth, width, and resolution using fixed scaling coefficients. Moreover, they validated the importance of balancing a network in all dimensions first by developing a mobile-size baseline network called EfficientNet-B0. Then, starting from this baseline network, they applied the proposed scaling method to obtain eight variants of the EfficientNet model. Figure 2 shows the architecture of EfficientNet-B2. The proposed models significantly outperform other ConvNet architectures on the ImageNet classification problem while having fewer parameters and running much faster during inference, an important property for real-time applications such as the one considered in this work. In addition, the features learned by the networks are transferable and achieve impressive results on a wide range of datasets.

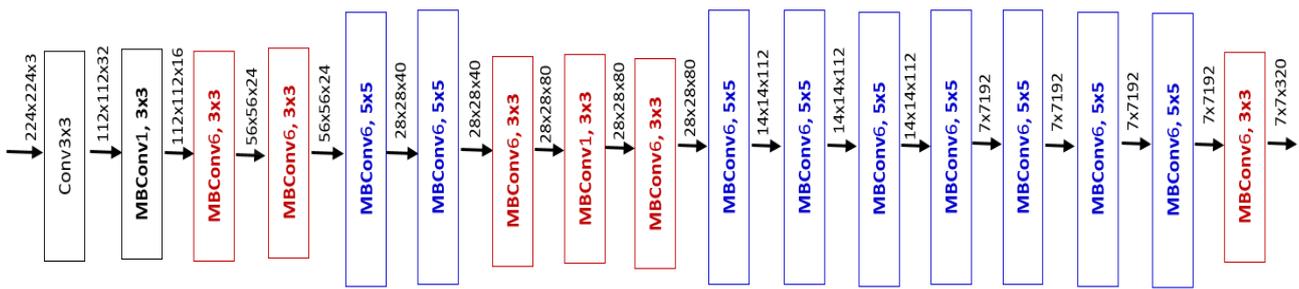


Figure 2. The EfficientNet-B2 Architecture. The numeric value next to MBCConv indicates a multiplication factor for the input channels. For example, MBCConv6 means the output channel size is 6 times the input channel size. Labels in the form of $a \times b$ represent kernel size [32].

The baseline network (EfficientNet-B0) uses the mobile inverted bottleneck layer as the main building block as shown in Figure 3, which is an inverted residual block combined with squeeze-and-excitation (SE) blocks. An inverted residual block first projects an input feature map into a higher dimensional space and then applies depth-wise convolution operation in the new space. The new feature map is projected back to a low-dimensional space using point-wise convolution (1×1 convolution) with linear activation. Finally, a residual connection is added from the input to the output of the point-wise convolution operation, resulting in an output feature map. SE blocks, on the other hand, learn to weight channels of an input feature map adaptively. First, they convert the input to a feature vector of size equal to the number of channels (c) and then feed it to a two-layer neural network. The output of this network, which is a vector of size c , is used to scale each channel based on its importance. Additionally, the baseline network is developed by leveraging a multi-objective neural architecture search technique that takes into account accuracy and real-world latency on mobile devices. Starting from this baseline network, the authors applied the compound scaling method to obtain seven different EfficientNet models (EfficientNet-B1 to B7).

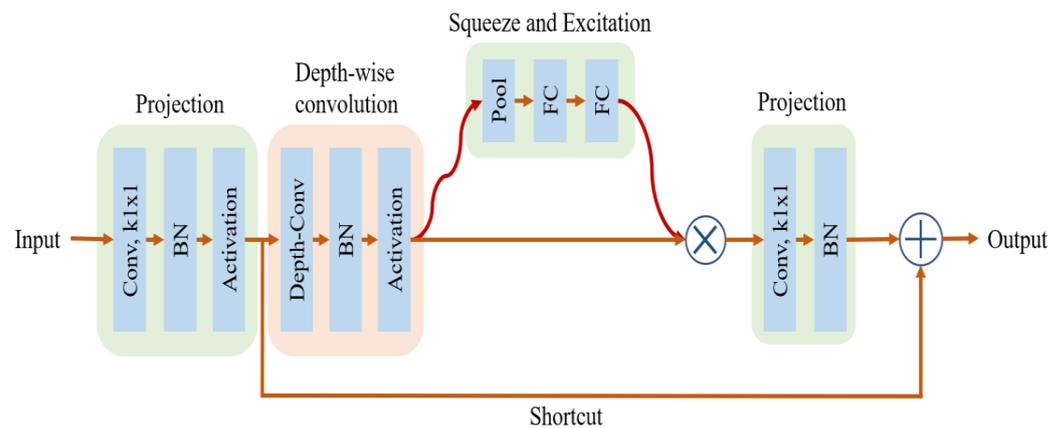


Figure 3. A mobile inverted bottleneck layer used for the EfficientNet family of models. (BN—Batch normalization layer; FC—Fully connected layer).

3.3. Network Optimization on Ultrasound Images

It is a known fact that training deep neural network models such as EfficientNet requires having large labeled training examples, and collecting such data is time consuming and costly. An alternative remedy is to either use pre-trained models as off-the-shelf feature extractors and train a generic classifier (such as SVMs) or fine-tune the model for the classification problem at hand. Since we have a limited number of samples for the problem we are trying to address, we have choose to fine-tune the EfficientNet-B2 model for classifying LUS images.

3.4. Vision Transformers

Consider the collection of n chest medical images $S = \{X_i, y_i\}_{i=1}^n$, where X_i and y_i are sample images and their associated class label, $y_i \in \{1, 2, \dots, m\}$, and m is the number of identified classes for a set of images. The method's goal is to learn how to translate the ultrasound image input to the appropriate class label.

The prototype was inspired by a Vision Transformer (ViT). The vanilla Transformer [34], which has attracted a lot of attention recently for its capacity to deliver state-of-the-art (SOTA) performance in machine translation and other applications involving natural language processing [54], serves as the sole architectural foundation for ViT.

Encoder–decoder blocks of the Transformer design enable the concurrent processing of sequential data without the need for recurrent networks. The self-attention mechanism, which is suggested to capture long-range links between the sequence's pieces, is substantially responsible for the success of Transformer models. In an effort to apply the standard Transformer to image categorization, Vision Transformer was proposed. Without incorporating any architecture tailored to certain types of data, the major objective was to generalize image categorization to modalities other than text. ViT performs classification by mapping a series of image patches to the semantic label using the encoder module of the Transformer in particular. Contrary to traditional CNN architectures, which frequently employ filters with local receptive fields, the Vision Transformer's attention mechanism enables it to be applied across various regions of the image and to integrate data from throughout the entire image.

Three key building components make up the proposed ViT model: an embedding layer, an encoder, and a final classifier. The input image is separated into non-overlapping patches in the first stage, which is then supplied into the embedding layer, encoder, and final classifier. We go into great detail about the model's elements in the subsections that follow. The proposed ViT model's general structure is shown in Figure 4.

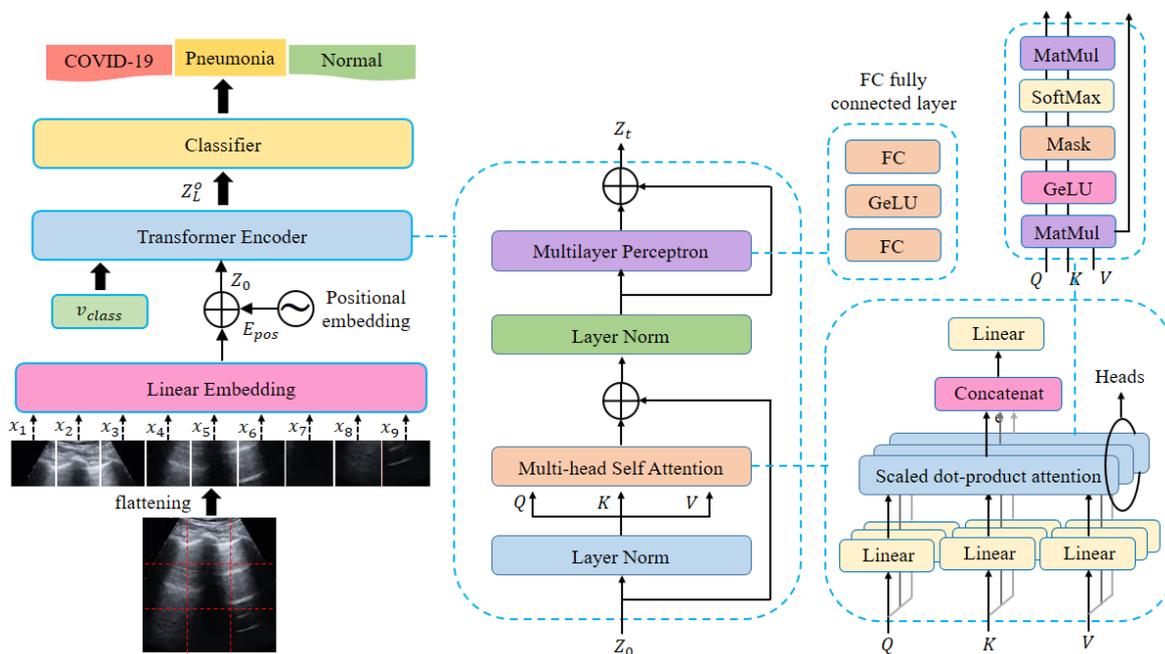


Figure 4. Vision Transformer Architecture.

3.4.1. Linear Embedding Layer

First, a sequence of separate, non-overlapping patches are created from the input image. The input image x which has the dimensions $h \times w \times c$ (where h , w , and c are the height, width, and number of channels, respectively) is then split into a series of lengths m by dividing it into small patches $x = \{x_p^1, x_p^2, \dots, x_p^m\}$, which have the dimensions $p \times p$,

which is fixed, and m is equal to $h \times w/p^2$. A typical patch size is 16×16 or 32×32 ; a smaller patch size yields a longer sequence, and vice versa.

The word tokens from the first Transformer are comparable to these patches. Using a learnt embedding matrix E , the sequence of patches is linearly projected onto a vector of the model dimension d before being fed into the encoder. The learnable classification word x_{class} that is necessary to complete the classification job is concatenated with the embedding representations after that. The flattened image patches are fed into a linear embedding layer E to match their dimension to the model dimension d , and afterwards they are transformed into embeddings.

Each patch embedding is added to its appropriate positional information to avoid the flattening operation from erasing the positional information. The learned class token x_{class} is attached to the resulting position-aware embeddings. Through a self-attention process, the categorization token and patch embeddings communicate with one another.

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^m E] + E_{pos} \text{ While } E \in R^{(p^2 \cdot c) \times d}, \wedge E_{pos} \in R^{(m+1) \times d} \quad (1)$$

3.4.2. ViT's Encoder Module

The transformer encoder receives the resulting embedded patch sequence z_0 . The encoder is made up of a stack of L identical layers, each of which is made up of two primary building blocks: a feed-forward network (FFN) block and a multi-head self-attention (MSA) block. The Transformer encoder's main component, the MSA, uses the self-attention (SA) technique to identify dependencies between various patches of the input image. (2) and (3) both display specifics of the calculations that occur in the SA block. The input sequence is first used to create three distinct matrices: the key K , the query Q , and the value V . An attention map is produced by using an inner product to match the query matrix to the key matrix. After being scaled by the key's dimension d_K , the output is then obtained using the SoftMax function. In order to concentrate on more crucial parameters, the outcome is finally multiplied by the value V .

$$[Q, K, V] = z U_{QKV}; U_{QKV} \in R^{d \times 3d_K} \quad (2)$$

$$SA(z) = \text{softmax}\left(QK^T / \sqrt{d_K}\right) \cdot V \quad (3)$$

Using multiple self-attention heads ($SA_1, SA_2 \dots SA_h$), where h is the number of heads, the multi-head self-attention is an extension of SA that performs the SA procedure concurrently. The purpose of using h heads is to let each head to concentrate on a unique relationship between the image's patches. Following that, a linear layer projects the outputs of all heads to the final dimension, as shown in Equation (4):

$$MSA(z) = \text{Concat}(SA_1(z); SA_2(z); \dots SA_h(z)) W^O, W^O \in R^{h \cdot d_K \times d} \quad (4)$$

where W^O stands for the final projection matrix's learned parameters.

The MSA block is followed by the second block in the encoder layer, called FNN. It has a *GeLU* activation function [55] sandwiched between two completely linked layers. Each block of the two encoder layers is followed by a layer of normalization (LN). The outputs are calculated utilizing residual connections and the following Equations (5) and (6):

$$z'_l = MSA(\text{ln}(z_{l-1})) + z_{l-1}, l = 1 \dots L \quad (5)$$

$$z_l = FNN(\text{ln}(z'_l)) + z'_l, l = 1 \dots L \quad (6)$$

The classification layer, which consists of a fully connected layer (FC) with a SoftMax activation function to generate the class labels, receives the output of the ViT encoder. We

instruct the classifier to predict the class label using the classification token represented by the first element of the encoder output z_L^0 .

$$y = \text{Softmax}\left(\text{FC}\left(z_L^0\right)\right) \quad (7)$$

3.5. gMLPs

Transformers' multi-head self-attention layers are simplified by gMLP, which is suggested as evidence that self-attention is not essential for ViT [33]. gMLP is a simple variation of MLP with gating that includes static parameterized channel projections and spatial projections. As depicted in Figure 5, it consists of a stack of L identical blocks. The element-wise multiplication (linear gating) procedures known as linear projection operations \odot are used. BERT for NLP and ViT for vision are the input and output protocols. Positional encodings are not necessary for gMLPs, unlike Transformers, nor is there any need to mask out the paddings throughout the fine-tuning of NLP.

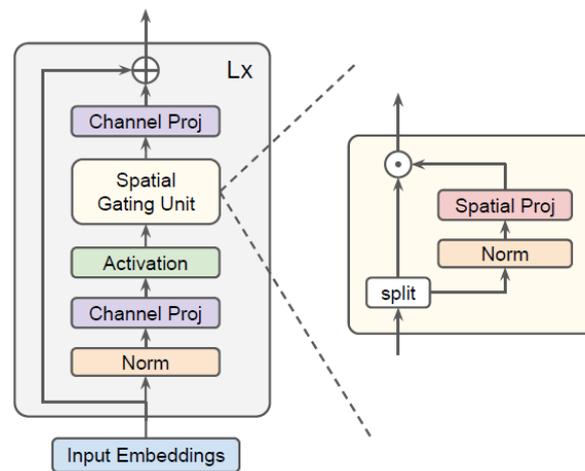


Figure 5. gMLP architecture with Spatial Gating Unit (SGU) [33].

4. Results

Given an ultrasound image of a patient, a multiclass model tells whether the patient has COVID-19 (has the coronavirus), is normal (does not have the coronavirus), or has pneumonia. As mentioned in the previous section, we applied transfer learning for this problem. To achieve this task, we fine-tuned three well-known architectures: EfficientNet-B2, ViT, and gMLP networks. Furthermore, we repeated the experiments without the transfer learning approach, i.e., learning from scratch.

4.1. Dataset Preparation

In order to conduct the experiments, we followed the same procedure followed by the author of the dataset [8]. They considered the recordings of convex probes and discarded the recordings of linear probes (20 videos and 6 images) and the recordings of viral pneumonia patients (6 videos). The convex videos vary in length and type and are composed of 160 ± 144 frames at a frame rate of 25 ± 10 Hz. These recordings (179 videos and 56 images) were manually processed and split into frames at a rate of 3 Hz (with 30 frames per video at maximum). At the end, the constructed ultrasound database contained images from three classes (i.e., 1204 COVID-19, 704 bacterial pneumonia, and 1326 healthy images). The irrelevant data (i.e., measure bars, texts, and artifacts on the borders) were cropped from the images before they were resized to 224×224 pixels.

Similar to [8], all experiments in this study were repeated five times according to a 5-fold cross-validation procedure. In the 5-fold database split, a patient-based split was used, in which the images belonging to a single video are included in the same fold, and all three classes must appear in each fold. Applying the cross-validation method for

database splitting is one of the known methods to evaluate the generalization capabilities and prevent overfitting in the predictive models [56,57].

4.2. Performance Evaluation

As performance metrics, we report the accuracy (8), precision (9), recall (10), and F1 scores (11).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

where TP , TN , FP , and FN are the true positive, true negative, false positive, and false negative values, respectively. These evaluation metrics are used by several similar works in the literature [8,13,34,41]; thus, using them gives the ability to compare our results with those of the other studies and to be consistent with the assessment procedures of medical diagnostic systems [58]. These metrics are calculated from a confusion matrix.

Accuracy is one of the main measures that could be used to analyze the performance of the classification systems. Moreover, precision and recall are two other performance measures that aid in discrimination between the classification systems. How many of the predictions are accurate is determined by precision, also known as positive prediction. On the other hand, recall (also known as sensitivity) demonstrates a system's ability to identify a patient's ailment. [58]. Recall shows how well the system is performing; false-negative predictions, or the inaccurate evaluation of infected patients as non-infected patients, come at a significant cost. Consequently, recall can be regarded as one of the most crucial criteria in the event of pandemics such as COVID-19.

F-measure or F1 score is a well-known measure in the classification systems. It is calculated as the weighted harmonic mean of both precision and recall. F1 score is considered as an extra evaluation metric to differentiate between the systems that could generate a comparable value of precision and recall.

These metrics' equations are used for multiclass classification problems as in our problem by simply considering TP as the class we want to calculate metrics for and TN as the rest of the classes.

4.3. Transfer Learning

As described earlier, we developed three models to detect COVID-19 through ultrasound imagery by applying a deep transfer learning approach. In this subsection, we report the results of the three proposed architectures.

4.3.1. EfficientNet-B2

We fine-tuned EfficientNet-B2 for detecting COVID-19 using ultrasound images. We fine-tuned only the last 100 layers of this network and froze the remaining ones. We used the stochastic gradient descent (ADAM) for optimization with number of training epochs set to 20, mini-batch size set to 20, and learning rate set to 0.0001. Table 2 shows the detailed values of the different parameters of EfficientNet-B2, gMLP, and ViT networks.

Figure 6 shows the confusion matrices of the different five trails of our experiments of the EfficientNet-B2 model. We noticed that in five trials, COVID-19 class images were correctly classified at a percentage higher than 99.41%, leading to a high recall of 99.81%, which shows the ability of the proposed model to recognize COVID-19 images in most cases, even in the fifth trial, which shows low performance in classifying healthy and pneumonia classes.

Starting from the confusion matrices in Figure 6, we calculated the accuracy of the EfficientNet-B2 model on the three classes (normal, pneumonia, and COVID-19) and present the results in Table 3.

Table 2. Parameter optimization for EfficientNet-B2, gMLP, and ViT.

Parameter	EfficientNet-B2	gMLP	ViT
Number of iterations	25	25	25
Batch size	20	20	16
No. of classes	3	3	3
Input size	224 × 224	224 × 224	224 × 224
Optimizer	Adam	AdamW	Adam
No. of epochs	20	50	50
Learning rate	1 × 10 ⁻⁴	1 × 10 ⁻⁴	1 × 10 ⁻⁴

Table 3. Accuracy per class (%) of the EfficientNet-B2 model with a transfer learning approach. We present the results of the five trials, where each trial is an experiment of the 5-fold cross-validation procedure. The last column represents the overall accuracy of the model.

	Normal	Pneumonia	COVID-19	Overall Accuracy
Trial no. 1	100	99.20	100	99.83
Trial no. 2	100	100	100	100
Trial no. 3	100	100	100	100
Trial no. 4	100	100	100	100
Trial no. 5	61.44	87.30	99.41	84.41
Average ± std.	99.28 ± 19.27	97.30 ± 6.22	99.88 ± 0.29	96.79 ± 7.07

The overall accuracy of the proposed architecture was 100% in three of the five trials and 99.83% in the first trial. The only exception is in the fifth trial, with an overall accuracy of 84.41%. It is worth noting that the accuracy of the system for detecting COVID-19 is excellent at 99.88% on average and 100% in four of the five trials.

The high accuracy of the architecture in detecting COVID-19 leads us to report the other evaluation metrics of detecting COVID-19 in Table 4.

Table 4. Performance metrics (%) of the five trials of the EfficientNet-B2 model with a transfer learning approach; we report the results of COVID-19 class.

	Precision	Recall	F1
Trial no. 1	99.69	100	99.84
Trial no. 2	100	100	100
Trial no. 3	100	100	100
Trial no. 4	100	100	100
Trial no. 5	77.72	99.41	87.24
Average ± std.	95.84 ± 1.95	99.88 ± 0.29	97.41 ± 5.68

Table 4 demonstrates that the proposed architecture exhibits high performance, with an average recall of 99.88%, precision of 95.84%, and F1 score of 97.41%. The excellent value for recall informs us that the system correctly classifies the images with COVID-19 markers.

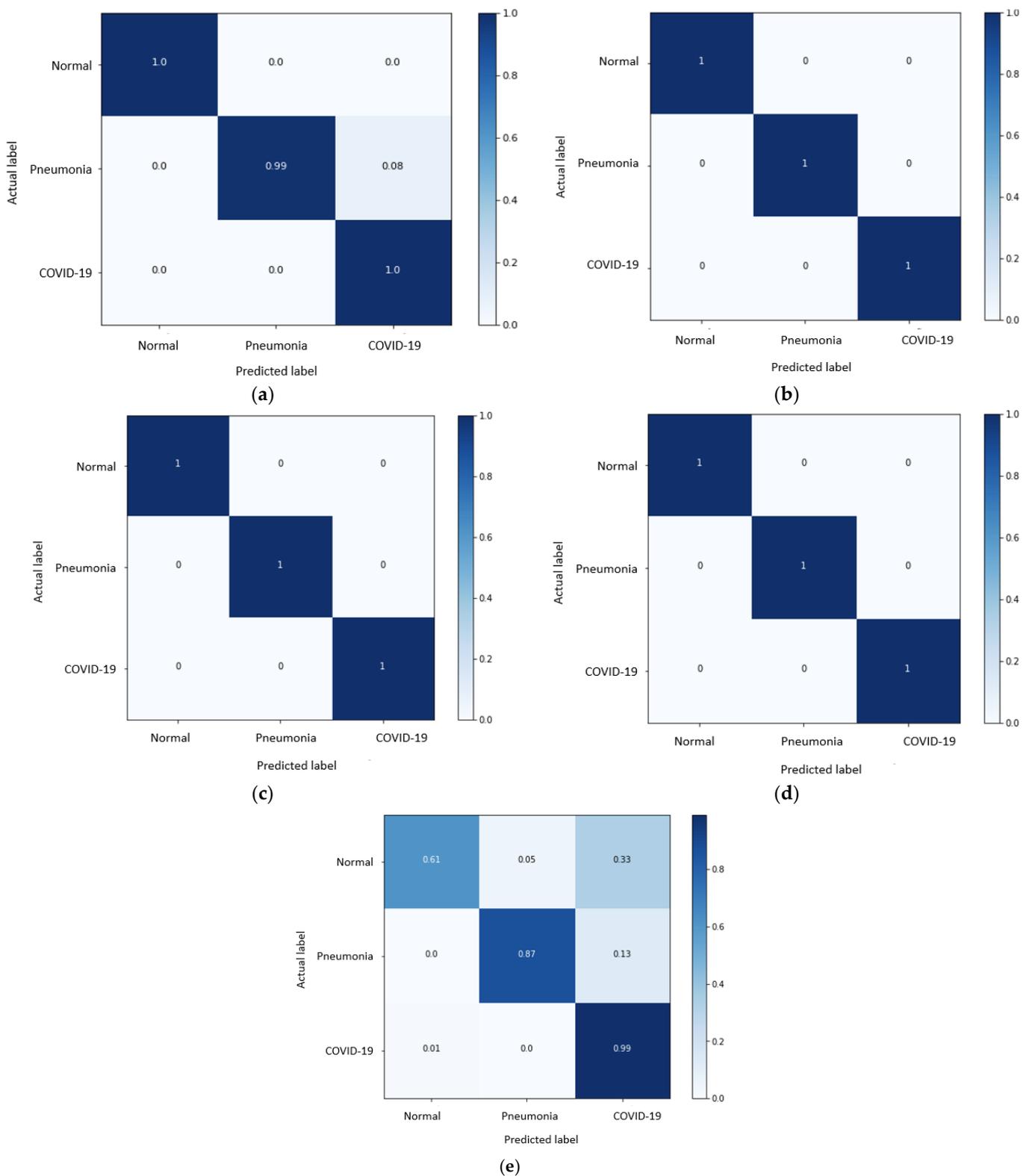


Figure 6. Confusion matrices of different 5-fold trials (a–e) of EfficientNet-B2. The rows represent the actual true values of the different classes; the columns represent the predicted values.

4.3.2. Contrasting EfficientNet-B2 with ViT and gMLP

In order to show the high performance of the proposed EfficientNet-B2 in classifying the ultrasound imagery, we now compare its accuracy in detecting COVID-19 with the

results of two well-known deep learning techniques: gMLP and ViT16. The results are reported in Table 5.

Table 5. Comparing the accuracy of EfficientNet-B2 in classifying COVID-19 class images with the accuracies of gMLP and ViT16.

	EfficientNet-B2	gMLP	ViT16
Trial no.1	99.83	91.51	90.7
Trial no.2	100	92.36	99.06
Trial no.3	100	98.62	97.64
Trial no.4	100	95.34	97.41
Trial no.5	84.41	83.58	80.38
Average \pm std.	96.79 \pm 7.07	92.82 \pm 5.60	93.03 \pm 7.78

In Figure 7, we show the receiver operating characteristic (ROC) curve with the average scores' area under the curve (ROC-AUC) of the different classes by applying the three models. The results show that EfficientNets yields better results compared to ViT and gMLP.

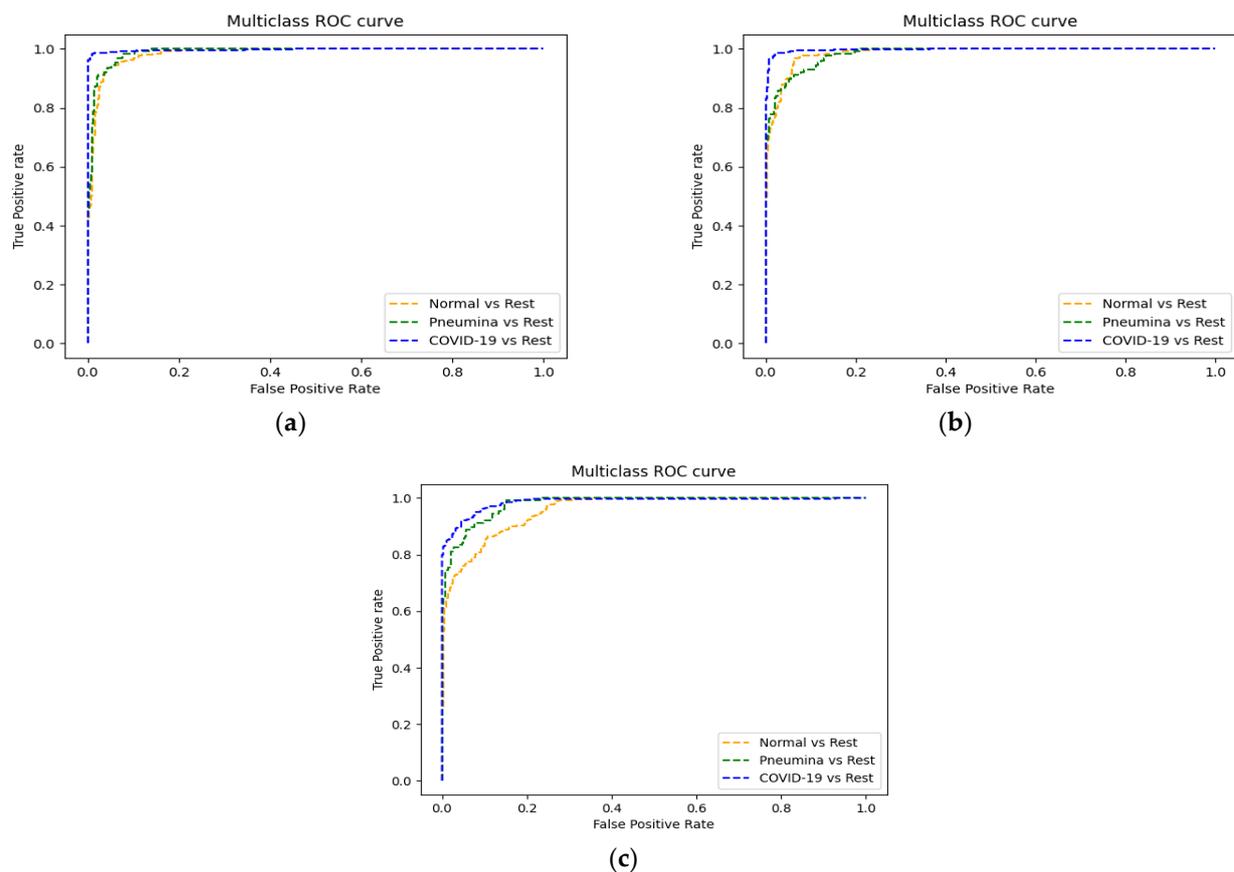


Figure 7. ROC curves of the three models: (a) EfficientNet-B2, (b) ViT, and (c) gMLP. We show the one-vs-rest ROC curve for the three classes.

4.4. Learning from Scratch

In the previous sections, we explored the efficiency of the EfficientNet-B2 through transfer learning in classifying ultrasound images. Furthermore, we found that it outperforms gMLP and ViT. Therefore, in this section, we utilize EfficientNet-B2 to solve our problem with learning from scratch, i.e., without a transfer learning approach. Likewise,

we compare the performance of EfficientNet-B2 with the performance of gMLP and ViT16 in terms of the accuracy of classifying COVID-19 class, as depicted in Table 6.

Table 6. Accuracy of the proposed architecture by applying learning from scratch. The accuracy is compared with the accuracies of gMLP and ViT16.

	EfficientNet-B2	gMLP	ViT16
Trial no. 1	84.8	88.1	82.7
Trial no. 2	89.9	71.8	62.2
Trial no. 3	89.8	77.2	67.8
Trial no. 4	89.8	70.9	68.0
Trial no. 5	76.5	79.4	74.1
Average \pm std.	86.2 \pm 5.82	77.48 \pm 6.93	70.96 \pm 7.79

From the results in Table 6, we can also see that EfficientNet-B2 beats the other models with an average accuracy of 86.2% compared with accuracies of 77.48% and 70.96% for gMLP and ViT16, respectively.

4.5. Comparing with the State of the Art

In order to prove the superiority of our proposed models, we compare the results of our best model EfficientNet-B2 with the models of the authors of the dataset as presented in Table 7.

Table 7. Performance of the proposed architecture compared with the accuracies of models proposed by the authors of the dataset.

	Overall Accuracy %	Precision %	Recall %	F1 %
Born et al. [8]	87.8	90	88	89
Born et al. [12]	89	88	96	92
EfficientNetB2 (Transfer Learning) Proposed	96.79	95.84	99.88	97.41

From the reported results, we find that our proposed model achieves very good performance in terms of the evaluation metrics, with unbeatable improvement in all the reported evaluation metrics.

4.6. Discussion

The reported results in the previous section show that the proposed approaches could clearly improve the performance of the COVID-19 detection systems using LUS. The results presented in Table 3 show that employing the transfer learning approach is a good choice to increase the accuracy of detecting COVID-19. Compared to the results of the learning from scratch approach shown in Table 6, transfer learning shows superiority and represents great improvement.

Furthermore, the proposed models based on transfer learning outperform the state-of-the-art methods on the same LUS dataset as shown in Table 7, where we can see that EfficientNet-B2 and even the other two models (gMLP and ViT) produce better results than the other works.

In order to deploy the proposed model at the point of care, we can simply develop a web-based application that is connected to a remote server where the model operates. The technician at the point of care captures the LUS record and uploads it to the web-based application to send it for analysis. The model then receives the record and the frames are

extracted, preprocessed, and classified. Finally, the result (the class of the LUS record) is sent back to the web-based application as shown in Figure 8.

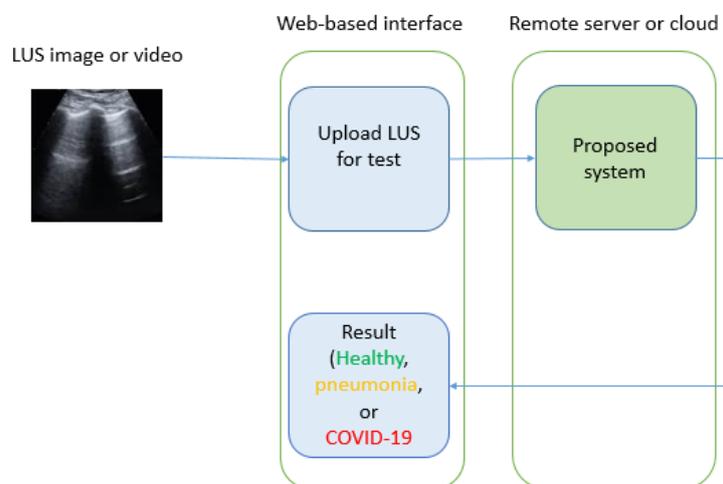


Figure 8. The deployment of the proposed model at the point of care.

The deployment of such a system has no bad effect on the patients or the medical teams and it is safe to be implemented and operated at the points of care. This is because ultrasound is considered a safe medical imagery technique and is cost effective. Furthermore, the only tools required are one ultrasound device and a computer at each point of care and one central computer for the model.

5. Conclusions

In this work, we propose a deep learning approach to detect COVID-19 patients from ultrasound images. More specifically, we applied the transfer learning approach by fine-tuning a model of the well-known family of EfficientNet models, i.e., EfficientNet-B2. Moreover, we explored the performance of the model without using the transfer learning approach, i.e., learning from scratch. Furthermore, we contrasted the performance of EfficientNet-B2 with other well-known deep learning models (gMPL and ViT16) to classify LUS images. The experimental results for EfficientNet-B2 with transfer learning show exceptional performance, outperforming both gMPL and ViT16 models. EfficientNet-B2 shows acceptable performance even when applying it with a learning from scratch approach. Furthermore, our models outperform the models of the authors of the database, which demonstrates the high performance of our model. One of the limitations of our model (EfficientNet-B2) is the relative large number of parameters to learn. As for future work, we are planning to develop an automatic system for detecting COVID-19 from multimodal imagery, i.e., CT, CXR, and LUS.

Author Contributions: Data curation, R.M.J.; funding acquisition, M.Z. and Y.B.; investigation, M.M.A.R. and F.M.; methodology, Y.B. and M.M.A.R.; project administration, Y.B.; writing—original draft, Y.B., R.M.J. and M.M.A.R.; writing—review and editing, M.M.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship for Research and Innovation, Ministry of Education, Saudi Arabia, the project number IFKSUDR_H157.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to the Deanship for Research and Innovation, Ministry of Education, Saudi Arabia, for funding this research work through the project number IFKSUDR_H157.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO. Director-General's Opening Remarks at the Media Briefing on COVID-19. 10 April 2020. Available online: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---10-april-2020> (accessed on 10 April 2020).
2. Raheja, R.; Brahmavar, M.; Joshi, D.; Raman, D. Application of Lung Ultrasound in Critical Care Setting: A Review. *Cureus* **2019**, *11*, e5233. [[CrossRef](#)] [[PubMed](#)]
3. Chatfield, K.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Return of the Devil in the Details: Delving Deep into Convolutional Nets. *arXiv* **2014**, arXiv:1405.3531.
4. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
5. Shen, D.; Wu, G.; Suk, H.-I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
6. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
7. Diaz-Escobar, J.; Ordóñez-Guillén, N.E.; Villarreal-Reyes, S.; Galaviz-Mosqueda, A.; Kober, V.; Rivera-Rodriguez, R.; Rizk, J.E.L. Deep-Learning Based Detection of COVID-19 Using Lung Ultrasound Imagery. *PLoS ONE* **2021**, *16*, e0255886. [[CrossRef](#)]
8. Born, J.; Wiedemann, N.; Cossio, M.; Buhre, C.; Brändle, G.; Leidermann, K.; Aujayeb, A.; Moor, M.; Rieck, B.; Borgwardt, K. Accelerating Detection of Lung Pathologies with Explainable Ultrasound Image Analysis. *Appl. Sci.* **2021**, *11*, 672. [[CrossRef](#)]
9. Roy, S.; Menapace, W.; Oei, S.; Luijten, B.; Fini, E.; Saltori, C.; Huijben, I.; Chennakeshava, N.; Mento, F.; Sentelli, A.; et al. Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound. *IEEE Trans. Med. Imaging* **2020**, *39*, 2676–2687. [[CrossRef](#)]
10. Awasthi, N.; Dayal, A.; Cenkeramaddi, L.R.; Yalavarthy, P.K. Mini-COVIDNet: Efficient Lightweight Deep Neural Network for Ultrasound Based Point-of-Care Detection of COVID-19. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **2021**, *68*, 2023–2037. [[CrossRef](#)]
11. Born, J.; Wiedemann, N.; Cossio, M.; Buhre, C.; Brändle, G.; Leidermann, K.; Aujayeb, A. L2 Accelerating COVID-19 Differential Diagnosis with Explainable Ultrasound Image Analysis: An AI Tool. *Thorax* **2021**, *76*, A230–A231. [[CrossRef](#)]
12. Born, J.; Brändle, G.; Cossio, M.; Disdier, M.; Goulet, J.; Roulin, J.; Wiedemann, N. POCOVID-Net: Automatic Detection of COVID-19 From a New Lung Ultrasound Imaging Dataset (POCUS). *arXiv* **2020**, arXiv:2004.12084.
13. La Salvia, M.; Secco, G.; Torti, E.; Florimbi, G.; Guido, L.; Lago, P.; Salinaro, F.; Perlini, S.; Leporati, F. Deep Learning and Lung Ultrasound for Covid-19 Pneumonia Detection and Severity Classification. *Comput. Biol. Med.* **2021**, *136*, 104742. [[CrossRef](#)] [[PubMed](#)]
14. Song, Y.; Zheng, S.; Li, L.; Zhang, X.; Zhang, X.; Huang, Z.; Chen, J.; Zhao, H.; Jie, Y.; Wang, R.; et al. Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT Images. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 2775–2780. [[CrossRef](#)] [[PubMed](#)]
15. Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M.; Guo, J.; Cai, M.; Yang, J.; Li, Y.; Meng, X.; et al. A Deep Learning Algorithm Using CT Images to Screen for Corona Virus Disease (COVID-19). *Eur Radiol* **2021**, *31*, 6096–6104. [[CrossRef](#)] [[PubMed](#)]
16. Wang, X.; Deng, X.; Fu, Q.; Zhou, Q.; Feng, J.; Ma, H.; Liu, W.; Zheng, C. A Weakly-Supervised Framework for COVID-19 Classification and Lesion Localization From Chest CT. *IEEE Transactions on Medical Imaging* **2020**, *39*, 2615–2625. [[CrossRef](#)]
17. Panwar, H.; Gupta, P.K.; Siddiqui, M.K.; Morales-Menendez, R.; Bhardwaj, P.; Singh, V. A Deep Learning and Grad-CAM Based Color Visualization Approach for Fast Detection of COVID-19 Cases Using Chest X-Ray and CT-Scan Images. *Chaos Solitons Fractals* **2020**, *140*, 110190. [[CrossRef](#)]
18. Li, L.; Qin, L.; Xu, Z.; Yin, Y.; Wang, X.; Kong, B.; Bai, J.; Lu, Y.; Fang, Z.; Song, Q. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. *Radiology* **2020**, 200905.
19. Chen, J.; Wu, L.; Zhang, J.; Zhang, L.; Gong, D.; Zhao, Y.; Hu, S.; Wang, Y.; Hu, X.; Zheng, B.; et al. Deep Learning-Based Model for Detecting 2019 Novel Coronavirus Pneumonia on High-Resolution Computed Tomography: A Prospective Study. *medRxiv* **2020**. [[CrossRef](#)]
20. Chen, X.; Yao, L.; Zhang, Y. Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images. *arXiv* **2020**, arXiv:2004.05645.
21. Soares, E.; Angelov, P.; Biaso, S.; Froes, M.H.; Abe, D.K. SARS-CoV-2 CT-Scan Dataset: A Large Dataset of Real Patients CT Scans for SARS-CoV-2 Identification. *medRxiv* **2020**. [[CrossRef](#)]
22. Liang, L.; Ma, L.; Qian, L.; Chen, J. An Algorithm to Attack Neural Network Encoder-Based Out-Of-Distribution Sample Detector. *arXiv* **2020**, arXiv:2009.08016.
23. Li, X.; Li, C.; Zhu, D. COVID-MobileXpert: On-Device COVID-19 Screening Using Snapshots of Chest X-ray. *arXiv* **2020**, arXiv:2004.03042.

24. Wang, L.; Wong, A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-ray Images. *Sci. Rep.* **2020**, *10*, 19549. [[CrossRef](#)]
25. Hemdan, E.E.-D.; Shouman, M.A.; Karar, M.E. Covidx-Net: A Framework of Deep Learning Classifiers to Diagnose Covid-19 in x-Ray Images. *arXiv* **2020**, arXiv:2003.11055.
26. Apostolopoulos, I.D.; Mpesiana, T.A. COVID-19: Automatic Detection from x-Ray Images Utilizing Transfer Learning with Convolutional Neural Networks. *Phys. Eng. Sci. Med.* **2020**, *43*, 635–640. [[CrossRef](#)] [[PubMed](#)]
27. Sethy, P.K.; Behera, S.K.; Ratha, P.K.; Biswas, P. Detection of Coronavirus Disease (COVID-19) Based on Deep Features and Support Vector Machine. *Preprints* **2020**, 2020030300. [[CrossRef](#)]
28. Narin, A.; Kaya, C.; Pamuk, Z. Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks. *arXiv* **2020**, arXiv:2003.10849. [[CrossRef](#)]
29. Farooq, M.; Hafeez, A. Covid-Resnet: A Deep Learning Framework for Screening of COVID19 from Radiographs. *arXiv* **2020**, arXiv:2003.14395.
30. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al-Emadi, N.; et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? *arXiv* **2020**, arXiv:2003.13145. [[CrossRef](#)]
31. Ucar, F.; Korkmaz, D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet Based Diagnosis of the Coronavirus Disease 2019 (COVID-19) from X-Ray Images. *Med. Hypotheses* **2020**, *140*, 109761. [[CrossRef](#)]
32. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR, 2019. Volume 97, pp. 6105–6114.
33. Liu, H.; Dai, Z.; So, D.R.; Le, Q.V. Pay attention to mlps. *Adv. Neural Inf. Process Syst.* **2021**, *34*, 9204–9215.
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Processing Syst.* **2017**, *30*, 5998–6008.
35. Silva, P.; Luz, E.; Silva, G.; Moreira, G.; Silva, R.; Lucio, D.; Menotti, D. COVID-19 Detection in CT Images with Deep Learning: A Voting-Based Scheme and Cross-Datasets Analysis. *Inform. Med. Unlocked* **2020**, *20*, 100427. [[CrossRef](#)] [[PubMed](#)]
36. Pathak, Y.; Shukla, P.K.; Arya, K.V. Deep Bidirectional Classification Model for COVID-19 Disease Infected Patients. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *18*, 1234–1241. [[CrossRef](#)]
37. Wang, Z.; Liu, Q.; Dou, Q. Contrastive Cross-Site Learning With Redesigned Net for COVID-19 CT Classification. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2806–2813. [[CrossRef](#)] [[PubMed](#)]
38. Rahhal, M.M.A.; Bazi, Y.; Jomaa, R.M.; Zuair, M.; Ajlan, N.A. Deep Learning Approach for COVID-19 Detection in Computed Tomography Images. *Comput. Mater. Contin.* **2021**, *67*, 2093–2110. [[CrossRef](#)]
39. Zhou, L.; Li, Z.; Zhou, J.; Li, H.; Chen, Y.; Huang, Y.; Xie, D.; Zhao, L.; Fan, M.; Hashmi, S.; et al. A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis. *IEEE Trans. Med. Imaging* **2020**, *39*, 2638–2652. [[CrossRef](#)]
40. Sun, L.; Mo, Z.; Yan, F.; Xia, L.; Shan, F.; Ding, Z.; Song, B.; Gao, W.; Shao, W.; Shi, F.; et al. Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification With Chest CT. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2798–2805. [[CrossRef](#)]
41. Al Rahhal, M.M.; Bazi, Y.; Jomaa, R.M.; AlShibli, A.; Alajlan, N.; Mekhalif, M.L.; Melgani, F. COVID-19 Detection in CT/X-ray Imagery Using Vision Transformers. *J. Pers. Med.* **2022**, *12*, 310. [[CrossRef](#)]
42. Arias-Londoño, J.D.; Gómez-García, J.A.; Moro-Velázquez, L.; Godino-Llorente, J.I. Artificial Intelligence Applied to Chest X-Ray Images for the Automatic Detection of COVID-19. A Thoughtful Evaluation Approach. *IEEE Access* **2020**, *8*, 226811–226827. [[CrossRef](#)]
43. Ohata, E.F.; Bezerra, G.M.; Chagas, J.V.S.d.; Neto, A.V.L.; Albuquerque, A.B.; Albuquerque, V.H.C.d.; Filho, P.P.R. Automatic Detection of COVID-19 Infection Using Chest X-Ray Images through Transfer Learning. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 239–248. [[CrossRef](#)]
44. Abbas, A.; Abdelsamea, M.M.; Gaber, M.M. Classification of COVID-19 in Chest X-Ray Images Using DeTraC Deep Convolutional Neural Network. *Appl. Intell.* **2021**, *51*, 854–864. [[CrossRef](#)] [[PubMed](#)]
45. Afshar, P.; Heidarian, S.; Naderkhani, F.; Oikonomou, A.; Plataniotis, K.N.; Mohammadi, A. COVID-CAPS: A Capsule Network-Based Framework for Identification of COVID-19 Cases from X-Ray Images. *Pattern Recognit. Lett.* **2020**, *138*, 638–643. [[CrossRef](#)] [[PubMed](#)]
46. Tabik, S.; Gómez-Ríos, A.; Martín-Rodríguez, J.L.; Sevillano-García, I.; Rey-Area, M.; Charte, D.; Guirado, E.; Suárez, J.L.; Luengo, J.; Valero-González, M.A.; et al. COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3595–3605. [[CrossRef](#)]
47. Heidari, M.; Mirniaharikandehi, S.; Khuzani, A.Z.; Danala, G.; Qiu, Y.; Zheng, B. Improving the Performance of CNN to Predict the Likelihood of COVID-19 Using Chest X-Ray Images with Preprocessing Algorithms. *Int. J. Med. Inform.* **2020**, *144*, 104284. [[CrossRef](#)]
48. Soldati, G.; Smargiassi, A.; Inchingolo, R.; Buonsenso, D.; Perrone, T.; Briganti, D.F.; Perlini, S.; Torri, E.; Mariani, A.; Mossolani, E.E.; et al. Is There a Role for Lung Ultrasound During the COVID-19 Pandemic? *J. Ultrasound Med.* **2020**, *39*, 1459–1462. [[CrossRef](#)]
49. Lichtenstein, D.A. Lung Ultrasound in the Critically Ill. *Ann. Intensive Care* **2014**, *4*, 1. [[CrossRef](#)]

50. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA, 3–6 December 2012; Bartlett, P.L., Pereira, F.C.N., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; pp. 1106–1114.
51. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
52. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
53. Huang, Y.; Cheng, Y.; Bapna, A.; Firat, O.; Chen, D.; Chen, M.X.; Lee, H.; Ngiam, J.; Le, Q.V.; Wu, Y.; et al. GPipe: Efficient Training of Giant Neural Networks Using Pipeline Parallelism. In Proceedings of the Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019; Wallach, H.M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E.B., Garnett, R., Eds.; pp. 103–112.
54. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186.
55. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). *arXiv* **2020**, arXiv:1606.08415.
56. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
57. Berrar, D. Cross-Validation. In *Encyclopedia of Bioinformatics and Computational Biology*; Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds.; Academic Press: Oxford, UK, 2019; pp. 542–545. ISBN 978-0-12-811432-2.
58. Šimundić, A.-M. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC* **2009**, *19*, 203–211.