Comparison of COVID-19 Classification via Imagenet-based and RadImagenet-based Transfer Learning Models with Random Frame Selection

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/SAS58821.2023.10254111

***Abstract*—Transfer learning methods are used in lung ultrasound (US) video classification as the training data for learning is generally limited. However, transfer learning models trained on the ImageNet dataset have been considered in the literature. Recently, the RadImageNet dataset, which contains medical images, was made available with four pre-trained models. In this work, pretrained models from both of these datasets are used as the backbone network, and models are developed for COVID-19 classification based on frames and videos. Random frame selection method is proposed and compared with the commonly used constant and non-adjacent frame selections. Video classification based on selected frames (VCBSF), and video classification based on the whole video (VCBWV) using voting are studied as part of video classification approach. A publicly available COVID-19 dataset is used to do this comparative study. The pre-trained models using the ImageNet dataset outperformed those using the RadImageNet dataset for frame classification. The ResNet50 and DenseNet121 models using ImageNet outperformed their counterparts using RadImageNet for VCBSF classification, while the Inception_ResNet_V2 and Inception_V3 models using RadImageNet performed better for VCBWV classification. These findings provide insight into improving the accuracy of COVID-19 detection using deep learning techniques and random frame selection.**

***Index Terms*—COVID-19 , Random frame selection, RadImageNet, ImageNet, Transfer learning, Frame classification, Video classification**

I. Introduction

In 2019, the World Health Organization (WHO) declared COVID-19 a pandemic, which has impacted over half a billion people worldwide and led to more than 6 million deaths [1]. Early detection of COVID-19 can significantly reduce its spread among individuals and consequently lower mortality rates [2]. Although the reverse transcription-polymerase chain reaction (RT-PCR) test is considered the gold standard for detecting COVID-19 , it has been found to have low sensitivity [3]. Medical imaging modalities are an alternative viable option, as they are considered to be accurate and highly sensitive diagnostic tools. Computed tomography (CT) [4], [5] and ultrasound (US) imaging [6], [7], are particularly found effective for detection of COVID-19 . However, CT is expensive, ionizing, and requires sterilization between scans of different patients [3], [8], when compared to US imaging.

US imaging is portable, affordable, radiation-free, and easy to disinfect when compared to CT [9], [10]. However, the classification of COVID-19 using US images is challenging due to noise and low resolution. Several deep-learning models have been proposed in the literature for the classification of COVID-19 . Due to the limited availability of US images, transfer learning methods are widely used to classify COVID-19 into one of three classes: a lung infected with COVID-19, a lung infected with bacterial pneumonia, or a healthy lung. These models were usually trained using the ImageNet dataset. For instance, pre-trained deep learning models such as POCOVID-Net, based on pre-trained VGG16, were proposed in [8], [11], and pre-trained models based on VGG19, InceptionV3, Xception, and ResNet50 were introduced in [9]. Pre-trained VGG16 and ResNet18 were proposed in [12], ResNet was employed in [13], and additional pre-trained models were discussed in [6], [14], [15], [16], [17]. As mentioned previously, all pre-trained models were trained using the large generic ImageNet dataset [18]. Recently, a large dataset of medical images named RadImageNet was made available, as reported in [19]. Four models, namely ResNet50 [20], DenseNet121 [21], Inception_ResNet_V2 [22], and Inception_V3 [23], were trained using this dataset and have been made publicly available. In this work, these pre-trained models based on RadImageNet are utilized and their classification performance against the same models trained on the ImageNet dataset is studied.

The majority of models proposed in the literature extract frames from US video clips using a constant frame rate, except for the approach presented in [24], which extracts frames utilizing the non-adjacent frame selection method. The constant frame selection method is based on the assumption that all videos have a constant frame rate, while the non-adjacent method selects frames that are not adjacent. To address these limitations, a random frame selection method is proposed that does not rely on any assumptions or constraints. Once the frames are extracted, they are utilized for COVID-19 classification. It is worth noting that most studies split the extracted frames into train and test sets, which may lead to inaccurate results, While other papers address this issue in

their work, [7], [8]. This is because the frames from the same video clip may end up in both the training and testing sets simultaneously, resulting in data leakage. To ensure accurate performance and prevent data leakage, this study divides the data into training and testing based on the videos themselves.

The main contributions of this study can be summarized as follows.

1) This is the first ever study to use RadImageNet trained models for transfer learning for COVID-19 classification.
2) Random frame selection method is proposed and compared with previous frame selection techniques (constant and non-adjacent methods) that are the current state-of-the-art.
3) Four pre-trained models (ResNet50 , DenseNet121 , Inception_ResNet_V2 , and Inception_V3) are utilized to classify lung images into three categories: lungs infected with COVID-19 , lungs infected with bacterial pneumonia, and healthy lungs.
4) The performance comparison of pre-trained models trained on ImageNet with those trained with RadImageNet is provided.
5) Two approaches for video classification that utilize majority voting, one based on selected frames (VCBSF) and the other based on the entire video (VCBWV), are presented.

The remaining paper is organized as follows: Section II presents the methodology, Section III provides the results, and Section IV discusses the conclusions and future works.

II. Methodology

Our proposed method comprises two parts:

1) Frame selection, and
2) Utilization of pre-trained models with RadImageNet and ImageNet weights for comparison purposes.

Four pre-trained models, namely ResNet50 , DenseNet121 , Inception_ResNet_V2 , and Inception_V3 , which are trained separately using ImageNet and RadImageNet data, are employed. Thus two sets of pre-trained models are considered in this paper. These models are used to classify the COVID-19 US images into three classes namely, COVID-19 , bacterial pneumonia, and healthy. Examples for each class from lung ultrasound (LUS) images are shown in Figure 2 where the first row represents the COVID-19 class, the second row is bacterial pneumonia, and the last row is the healthy class. Moreover, majority voting is used to provide video classification either based on selected frames (VCBSF) or based on the whole video (VCBWV).

A. Random frame selection

A random selection method is proposed to extract a frame from each video and compare it with two state-of-the-art methods: non-adjacent and constant. The constant method assumes that all videos have a consistent frame rate. The frame extraction rate is determined by dividing the video frame rate by the assumed constant frame rate. For instance, as shown in Figure 2 (a), if the assumed frame selection is 3 frames in a video with a frame rate of 30 frames per second, then there will be 10 consecutive frames skipped with every frame selected for classification. It is important to note that the frame extraction rate depends on the video frame rate and that there should be no overlap in the output frames as shown in Figure 2 (a). The frame extraction process continues until the end of the video or the maximum of an extracted frame is met. In this work, the maximum is set to 30.

The non-adjacent method was proposed in [24], where nonadjacent frames were used for classification. Figure 2 (b) shows the non-adjacent method with a constraint of 5 frames per second. Non-adjacent is similar to constant with the exception that the number of frames may not be dependent on the frame rate. From frame to frame, the position of the chosen frames may be different but non-adjacent. Random selection, on the other hand, eases all the constraints. The number of chosen frames and the position of the frames within a video can be different. From frame to frame within a video, there would be no pattern. Thus the random selection of frames does not depend on the video frame rate. Figure 2 (c) presents an example of this method where some extracted frames are closer to each other while others are farther apart. This enhances the likelihood of obtaining useful frames, as demonstrated in the results section.

B. Transfer learning

Transfer learning is a method that involves training a deep learning model on a large dataset and then using the learned knowledge to tackle a new task with a smaller dataset, such as the COVID-19 dataset. This approach effectively addresses the issue of overfitting and can lead to improved performance on the task at hand. In this work, ResNet50 [20], DenseNet121 [21], Inception_ResNet_V2 [22], and Inception_V3 [23] are employed for the purpose of the classification. Two different datasets are used to train these models and retrieve two distinct pre-trained weights: ImageNet and RadImageNet. ImageNet is a generic dataset containing approximately 1.3 million images [18]. On the other hand, RadImageNet is a medical images dataset [19] with 1.35 million radiologic images obtained from computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) modalities. RadImageNet covers 11 anatomical regions, such as brain, hip, knee, and liver, and was annotated by radiologists with 165 classes.

The original fully connected layers are removed from all pre-trained models and replaced with a sequence of layers including an average pooling layer, a dense layer, a batch normalization layer, a PRELU activation function, a dropout layer, and a final dense layer with a softmax activation function, in that order. The first and second dense layers have 128 and 3 neurons, respectively, and the dropout ratio is 0.5. All convolution layers are frozen except for the last five. The models are trained using an Adam optimizer (Adam) with a weight decay of 0.0001 and a learning rate of 0.001 for 100 epochs and with a categorical cross-entropy loss function. To

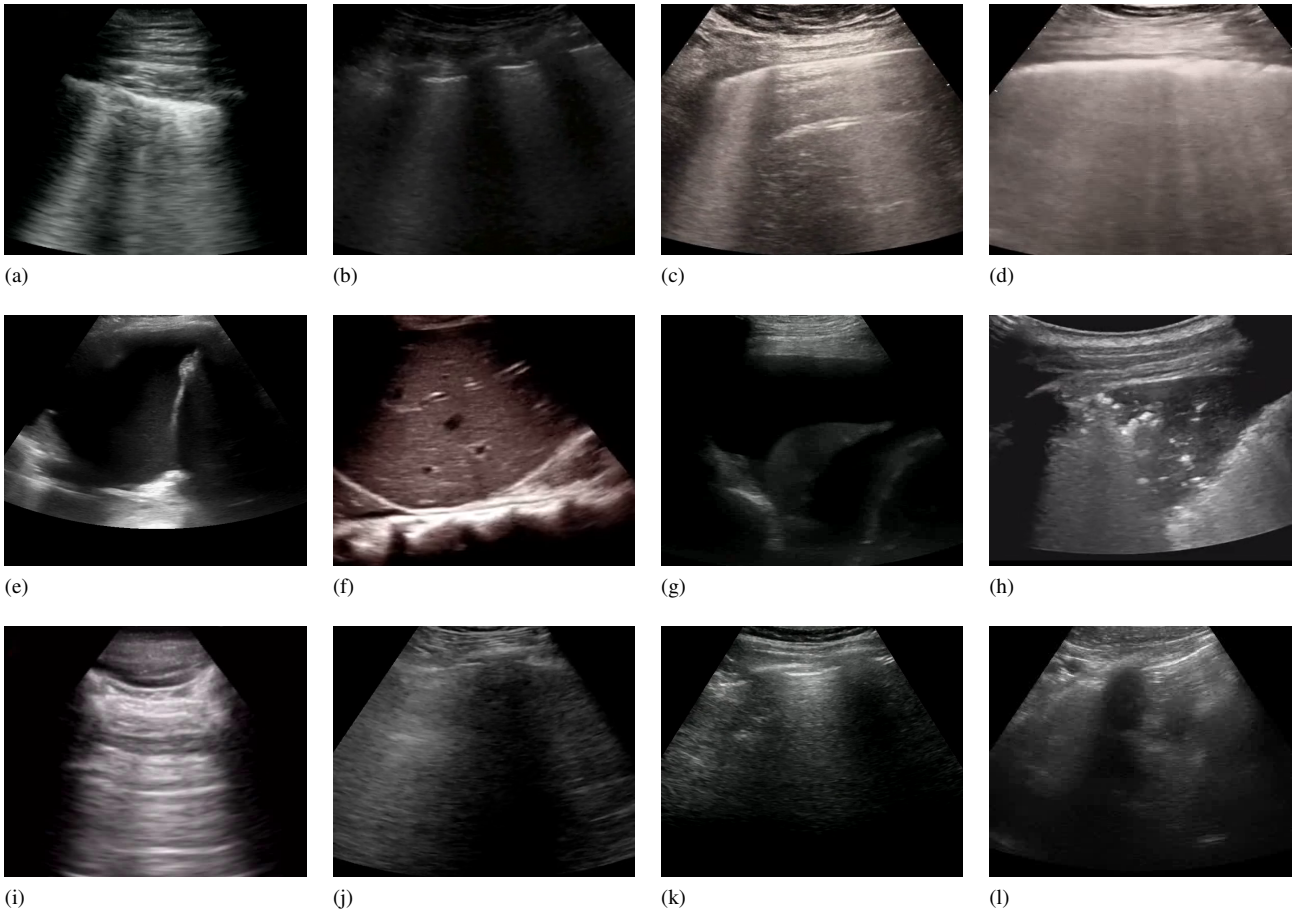


Fig. 1: Samples of US images of COVID-19, bacterial pneumonia, and healthy lungs. Images (a) to (d) are COVID-19, and (e) to (h) are bacterial pneumonia while (i) to (l) are healthy lungs.

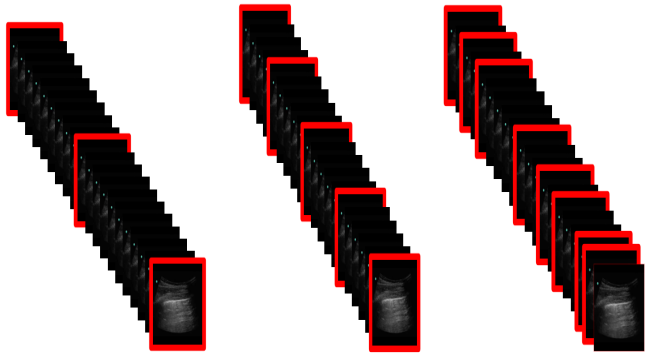


Fig. 2: Illustrations of frames selections from the video clip using (a) constant, (b) non-adjacent, (c) random.

prevent overfitting, early stopping techniques are implemented with a patience of 20, whereby the training process is halted if there is no decrease in the validation loss. A comparison is provided to evaluate the performance of the pre-training models, ImageNet or RadImageNet, and to determine which one is better suited for our classification task.

III. Results

The US dataset provided by [8], [25], was used to evaluate the pre-trained models and frame selection methods. The dataset consists of 202 videos and 59 images, which were recorded using either convex or linear US probes and distributed across four classes: COVID-19 , bacterial pneumonia, viral pneumonia, and healthy lungs. The data was gathered from 41 distinct resources, with around half of the dataset obtained from Northumbria data and Neuruppin. The presence of COVID-19 was confirmed using RT-PCR by Northumbria data and Neuruppin, while bacterial pneumonia was confirmed using X-ray and CT. Further, the entire dataset annotations have been reviewed and validated by two medical experts, one of whom is a pediatric physician with more than 10 years of clinical US experience, and the other is an academic US course instructor. For the evaluation, only the data gathered by the convex probe was used, and the viral pneumonia class was excluded due to its low sample size (only 3 video clips). To prevent data leakage, a five-fold cross-validation method was employed based on the video, which ensured that the frames

extracted from the video appeared only in the train or test sets. The results are provided as the mean of five repetitions.

Table I shows the result of frame classification, video classification based on selected frames from the test video clips (VCBSF), and video classification based on the whole test video clip (VCBWV). The two video classifications are achieved using majority voting of the frame classifications. It shows the results obtained when the four pre-trained models (ResNet50 , DenseNet121 , Inception_ResNet_V2 , and Inception_V3), utilize ImageNet weights. The classification results provided by ResNet50 and Inception_ResNet_V2 demonstrate that the random frame selection outperforms the other frame selections for all classifications (frame and video classifications VCBSFand VCBWV). The non-adjacent method yields the best results when the pre-trained model is DenseNet121 for all classification types. Again, the non-adjacent method provides the best frame classification only when Inception_V3 is employed. However, the best results of video classification are provided using random (VCBSF) and non-adjacent (VCBWV) frame selection methods when the model Inception_V3 is used. The important thing to note is that selecting frames randomly or in a non-adjacent manner consistently outperforms the most commonly used constant frame technique used in previous research. The best overall results are achieved when random frame selection is utilized for the frame and VCBWVclassifications, while the best overall results for the VCBSFclassification are yielded when non-adjacent frame selection is utilized, as illustrated by the yellow color in Table I. The Inception_ResNet_V2 model achieved the best performance for both frame classification and video classification based on selected frames. On the other hand, for video classification based on the whole video (VCBWV), the ResNet50 model produced the optimal performance.

Table II is similar to Table I, except that it uses RadImageNet weights instead of ImageNet weights. The best results were achieved when ResNet50 , DenseNet121 , and Inception_V3 were employed with non-adjacent frame selection. On the other hand, the Inception_ResNet_V2 model provided the best results when using random frame selection. The yellow color in Table II indicates the best overall performance. Specifically, the overall best results for frame classification are achieved through non-adjacent frame selection and the use of the Inception_V3 model. Meanwhile, for video classification VCBSFand VCBWV, the best outcomes are obtained by utilizing random frame selection and the Inception_ResNet_V2 model. Once again, random and non-adjacent frame selections outperform the traditional constant frame selection.

Figure 3 depicts the comparison between four models using either ImageNet weights or RadImageNet weights (based on accuracy) with all three types of frame selections considered in this paper. When using random frame selection, the model utilizing ImageNet outperforms the RadImageNet model in all classification types (frame and both video classifications). However, the performance varies across different classification types when using non-adjacent frame selection. ImageNet outperforms RadImageNet for frame classification in all models, whereas ResNet50 and DenseNet121 perform better with ImageNet than with RadImageNet for both video classifications. On the other hand, Inception_ResNet_V2 and Inception_V3 perform better with RadImageNet than with ImageNet for both video classifications. In conclusion, most models perform better with ImageNet than with RadImageNet, likely because the ImageNet dataset captures more useful information than RadImageNet. Further experiments are necessary to explore the efficacy of RadImageNet. It may be worthwhile to download RadImageNet and train additional models, such as Vision Transformer and VGG16, using the pre-trained models for our classification task.

IV. Conclusions and Future Works

The methods used for extracting frames from US video clips include constant and non-adjacent frame selection. The constant method assumes that all videos in the dataset have the same frame rate, while the non-adjacent method requires that the selected frames are not adjacent. A random frame selection method that does not rely on any assumptions or constraints was proposed in this paper and the classification performance obtained with the three frame selection methods were compared in this paper. Since the US dataset is limited, transfer learning was considered for COVID-19 classification. Traditionally, all transfer learning methods employ only pre-trained models using ImageNet, despite the availability of a pre-trained model using the medical images dataset RadImageNet. Therefore, in this paper, performance comparison was carried out using pre-trained models obtained using each of these datasets. In order to carry out the classification comparison in LUS classification, pre-trained models using ImageNet and RadImageNet were employed for classification. The results show that random and non-adjacent frame selections outperform constant selections. Random selection is favored because it does not require any assumptions or conditions. Additionally, pre-trained models using ImageNet provided the best frame classification results, with ResNet50 and DenseNet121 for VCBSF and VCBWV. Conversely, Inception_ResNet_V2 and Inception_V3 with RadImageNet yielded the best performance for VCBSF and VCBWV classification. In the future, ensemble deep learning trained on both ImageNet and RadImageNet may be considered to exploit the useful information provided by each of the datasets and improve classification performance. Furthermore, the RadImageNet dataset may also be utilized to train alternative models such as VGG16 and vision transformers, which can be subsequently utilized in LUS classification task. An advanced voting technique such as the differential evolution algorithm can also be employed to enhance video classification performance.

V. ACKNOWLEDGMENTS

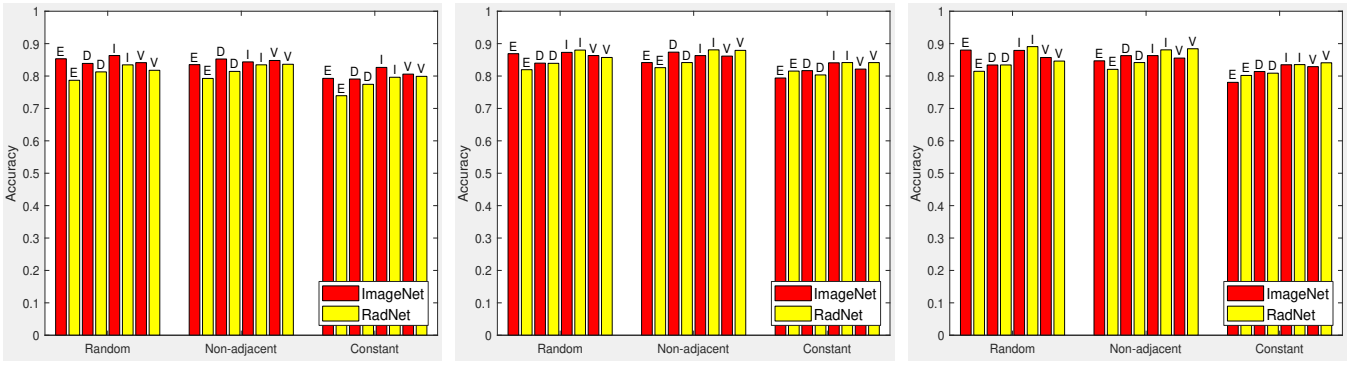
This work was financially supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

TABLE I: The results of 4 ImageNet pre-trained models evaluated based on different frame selection methods, including random (R), non-adjacent (N), and constant (C). The optimal results are highlighted in bold for each frame selection approach, while the overall best results are highlighted in yellow. F. S. indicates frame selection.

		Frame classification					VCBSF					VCBWV				
Models	F. S.	Sensitivity	Specificity	Precision	F1-Score	Accuracy	Sensitivity	Specificity	Precision	F1-Score	Accuracy	Sensitivity	Specificity	Precision	F1-Score	Accuracy
ResNet50	R	**0.8577**	**0.9277**	**0.8529**	**0.8504**	**0.8533**	**0.8770**	**0.9361**	**0.8734**	**0.8667**	**0.8690**	**0.8872**	**0.9416**	**0.8839**	**0.8779**	**0.8801**
	N	0.8401	0.9191	0.8386	0.8316	0.8352	0.8496	0.9231	0.8461	0.8383	0.8416	0.8540	0.9257	0.8519	0.8434	0.8469
	C	0.7610	0.8898	0.7796	0.7462	0.7928	0.7758	0.8936	0.8050	0.7567	0.7939	0.7965	0.8940	0.7896	0.7629	0.7805
DenseNet121	R	0.8437	0.9207	0.8456	0.8363	0.8389	0.8473	0.9214	0.8500	0.8370	0.8398	0.8421	0.9190	0.8461	0.8319	0.8341
	N	**0.8558**	**0.9271**	**0.8548**	**0.8494**	**0.8525**	**0.8779**	**0.9374**	**0.8781**	**0.8715**	**0.8737**	**0.8674**	**0.9321**	**0.8649**	**0.8605**	**0.8630**
	C	0.7686	0.8954	0.7622	0.7540	0.7907	0.8026	0.9076	0.8239	0.7951	0.8167	0.8183	0.9097	0.8149	0.8072	0.8139
Inception_ResNet_V2	R	**0.8674**	**0.9324**	**0.8659**	**0.8610**	**0.8633**	**0.8777**	**0.9372**	**0.8778**	**0.8716**	**0.8730**	**0.8829**	**0.9401**	**0.8857**	**0.8776**	**0.8791**
	N	0.8506	0.9224	0.8451	0.8426	0.8436	0.8725	0.9316	0.8669	0.8640	0.8631	0.8725	0.9315	0.8660	0.8640	0.8631
	C	0.8034	0.9064	0.8181	0.7968	0.8266	0.8324	0.9168	0.8565	0.8273	0.8408	0.8452	0.9186	0.8438	0.8319	0.8348
Inception_V3	R	0.8467	0.9220	0.8409	0.8385	0.8413	0.8703	**0.9330**	0.8651	**0.8612**	**0.8633**	0.8636	**0.9304**	0.8595	**0.8544**	**0.8572**
	N	**0.8544**	**0.9255**	**0.8507**	**0.8454**	**0.8480**	**0.8710**	0.9326	**0.8661**	0.8600	0.8615	**0.8643**	0.9297	**0.8615**	0.8538	0.8554
	C	0.7908	0.8996	0.7940	0.7818	0.8059	0.8125	0.9091	0.8227	0.8061	0.8214	0.8389	0.9162	0.8357	0.8269	0.8290

TABLE II: The results of four RadImageNet pre-trained models were evaluated based on different frame selection methods, including random (R), non-adjacent (N), and constant (C). The optimal results are highlighted in bold values for each frame selection approach, while the overall best results are highlighted in yellow. F. S. indicates frame selection.

		Frame classification					VCBSF					VCBWV				
Models	F. S.	Sensitivity	Specificity	Precision	F1-Score	Accuracy	Sensitivity	Specificity	Precision	F1-Score	Accuracy	Sensitivity	Specificity	Precision	F1-Score	Accuracy
ResNet50	R	0.7934	0.8942	0.7938	0.7844	0.7869	0.8263	0.9103	**0.8307**	0.8175	0.8196	0.8222	0.9078	**0.8269**	0.8127	0.8146
	N	**0.8002**	**0.8970**	**0.7980**	**0.7923**	**0.7926**	**0.8364**	**0.9128**	0.8305	**0.8280**	**0.8259**	**0.8293**	**0.9103**	0.8256	**0.8227**	**0.8206**
	C	0.7182	0.8627	0.7176	0.7081	0.7393	0.8060	0.9045	0.8272	0.8048	0.8153	0.8103	0.9023	0.8166	0.8003	0.8018
DenseNet121	R	0.8182	0.9065	**0.8231**	0.8105	0.8128	0.8460	0.9192	**0.8570**	0.8383	0.8392	0.8418	0.9163	**0.8538**	0.8345	0.8342
	N	**0.8221**	**0.9081**	0.8168	**0.8138**	**0.8143**	**0.8494**	**0.9228**	0.8486	**0.8398**	**0.8416**	**0.8494**	**0.9228**	0.8486	**0.8398**	**0.8416**
	C	0.7592	0.8839	0.7474	0.7421	0.7744	0.7952	0.9006	0.8021	0.7856	0.8034	0.8179	0.9059	0.8151	0.8069	0.8088
Inception_ResNet_V2	R	**0.8399**	0.9177	**0.8365**	**0.8332**	0.8345	**0.8872**	0.9398	**0.8862**	**0.8804**	0.8801	**0.8981**	**0.9456**	**0.8943**	**0.8911**	**0.8908**
	N	0.8395	**0.9182**	0.8364	0.8326	**0.8346**	0.8869	**0.9414**	0.8850	0.8796	**0.8808**	0.8869	0.9414	0.8850	0.8796	0.8808
	C	0.7840	0.8941	0.7804	0.7741	0.7962	0.8313	0.9185	0.8490	0.8309	0.8416	0.8396	0.9171	0.8443	0.8354	0.8353
Inception_V3	R	0.8257	0.9111	0.8221	0.8138	0.8177	0.8665	0.9300	0.8628	0.8561	0.8573	0.8566	0.9251	0.8540	0.8437	0.8462
	N	**0.8426**	**0.9191**	**0.8381**	**0.8350**	**0.8363**	**0.8870**	**0.9408**	**0.8846**	**0.8787**	**0.8792**	**0.8911**	**0.9431**	**0.8901**	**0.8837**	**0.8842**
	C	0.7847	0.8967	0.7731	0.7721	0.7990	0.8329	0.9198	0.8412	0.8283	0.8415	0.8490	0.9221	0.8439	0.8379	0.8409



(a) (b) (c)

Fig. 3: Accuracy of 4 pre-trained models using ImageNet or RadImageNet weights with random, non-adjacent, and constant frame selections for (a) frame classification, (b) VCBSF, (c) VCBWV. ResNet50, DenseNet121, Inception_ResNet_V2, and Inception_V3 are indicated by E, D, I, and V, respectively.

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] L. Sun *et al.*, "Adaptive feature selection guided deep forest for COVID-19 classification with chest CT," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2798–2805, 2020.

[5] Z. Wang, Q. Liu, and Q. Dou, "Contrastive cross-site learning with redesigned net for COVID-19 CT classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2806–2813, 2020.

[6] 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.

[7] 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.

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

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

[10] 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.
[11] J. Born *et al.*, "POCOVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS)," *arXiv preprint arXiv:2004.12084*, 2020.
[12] J. Roberts and T. Tsiligkaridis, "Ultrasound diagnosis of COVID-19: Robustness and explainability," *arXiv preprint arXiv:2012.01145*, 2020.
[13] H. Che *et al.*, "Multi-feature multi-scale CNN-derived COVID-19 classification from lung ultrasound data," in *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2618–2621, 2021.
[14] 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.
[15] 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.
[16] 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.
[17] L. Zhao and M. Lediju, "A review of deep learning applications in lung ultrasound imaging of COVID-19 patients," *BME Frontiers*, vol. 2022, 2022.
[18] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015.
[19] X. Mei *et al.*, "RadImageNet: An open radiologic deep learning research dataset for effective transfer learning," *Radiology: Artificial Intelligence*, vol. 4, no. 5, p. e210315, 2022.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
[21] G. Huang, Z. Liu, V. D. Maaten, and K. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.
[22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proceedings of Association for the Advancement of Artificial Intelligence(AAAI)*, vol. 31, 2017.
[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.
[24] B. Barros, P. Lacerda, C. Albuquerque, and A. Conci, "Pulmonary COVID-19: Learning spatiotemporal features combining CNN and LSTM networks for lung ultrasound video classification," *Sensors*, vol. 21, no. 16, p. 5486, 2021.
[25] 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.