Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2022)
IEEE Xplore Part Number: CFP22N67-ART; ISBN: 978-1-6654-9707-7

University Indoor Scene Classification using Transfer Learning

Dipak Ramoliya
Computer Science & Engineering Department
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Changa, Anand, India
dipakramoliya.ce@charusat.ac.in

Parth Goel
Computer Science & Engineering Department
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Changa, Anand, India
parthgoel.ce@charusat.ac.in

Jinish Vaidya
Computer Science & Engineering Department
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Changa, Anand, India
jinishvaidya@gmail.com

Poojan Vadaliya
Computer Science & Engineering Department
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Changa, Anand, India
poojanvadaliya@gmail.com

2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA) | 978-1-6654-9707-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIRCA54612.2022.9985516

***Abstract*— Image classification is used to classify images using a machine without any human intervention. This will help obtain all possible information of that particular capture just by feeding the image or by just clicking pictures from a device. The proposed study has created an extensive dataset of labs and its amenity by comprising 4502 images. Here there are 4 classes in the dataset: Apple Lab, Sophos Lab, Sophos Rack, and Virtual Reality Lab. The proposed work is based on transfer learning. Data augmentation was performed after the proposed model was passed through VGG 16, ResNet50, Inception V3, and Xception pre-trained models from which features were extracted automatically and categorized into 4 distinct classes. Size of the model, inference time, training accuracy, testing accuracy, recall, precision, F1 score, and FLOPs were evaluated while analyzing through pre-trained models. Best experimental results were obtained using Xception pre-trained model among all 4 pre-trained network models with 99.90% as training accuracy, 99.75% as testing accuracy, 99.35% as recall, 99.85% as precision, and 99.60% as F1 score; having a model size of 83 MB.**

Keywords*— *Image classification, Self-created dataset, Transfer learning, Pre-trained models.

I. INTRODUCTION

Image classification is used to get information and predict the outcomes using the machine. People visit various places in but are not sure about every detail about the site. This leads to taking help from someone to know more information about that site. This problem can be solved with help of a machine learning model, which has the ability to classify the image. It is just required to input the image into the model by clicking a picture or taking an image from any of the available resources. This developed model can be integrated with an application so that it can be also used by an organization like companies, universities, schools, colleges, etc. Organizations can use this model for visitors, employers, students, etc. in the sense if someone doesn't know about the place then that person can click a photo from the app and the app itself will classify the image using the pre-trained model and will deliver all the information about that facility or accessory without anyone's help. Task will get easier with such a system as it will be useful for getting detailed information about the location, and its external and internal structure in just on few clicks. This will be helpful for visitors, learners, and many other peoples.

In our designed imaged classification model, we have applied the transfer learning approach, where we used the pre-trained models for lab and its appliances classification. Using a pre-trained model is beneficiary as compared to building a model from scratch because they are already trained on an ImageNet dataset [1] having 1000 classes and a huge number of images, so retraining it for a customized dataset having lesser images will be sufficient due to which, it will take lesser training time and image prediction will have better accuracy. In this model, the lab and its amenities dataset were manually collected and augmentation was performed on it to make a more robust dataset than before. All the images were passed through VGG 16 [2], ResNet50 [3], Inception V3 [4], and Xception [5] pre-trained models, and the last layer of pre-trained models were connected to flatten layer followed by a dense layer having 4 neurons for 4 different class and SoftMax activation function. After that size of the model, inference time, training & testing accuracy, precision, recall, F1-score, and FLOPs were compared for all the four models.

Highlights of this paper are summarized below:

- Methodology for image classification from the manually collected dataset.
- Analyzing and concluding the best pre-trained model for image classification for this dataset
- Comparison of evaluation metrics such as the size of the model, inference time, training & testing accuracy,

precision, recall, f1-score, & FLOPs among the four pre-trained models (VGG16, ResNet50, InceptionV3, and Xception)

The paper is structured as follows:

An overview of some research work related to image classification is in section 2. Details of the dataset used for the proposed work are in section 3. Section 4 presents a methodology of the architecture of the model used for the classification. Outcomes of work are analyzed and compared in section 5. In final section 6, the summary and scope of enhancement have been stated.

II. Related Work

Image classification has numerous applications in different domains, and various research has already been done on it. For instance, image classification can be used for vehicle classification which can be helpful to form an Intelligent Transportation System(ITS) [6], and handwriting recognition for various languages to automate the process which will reduce time and would be solving the contemporary problem of digit recognition in various corporation sectors [7], classifying crop type mapping for prediction of crop yield which would help us for inspecting and examining land cover dynamics for better field management and utilization of land [8], fruit image classification for collecting various useful information about the ultimate product [9]. Similar to this image classification helped in various areas for advancing from the manual methodologies to automation to perform the task with optimal output.

Image classification can be a complex procedure that can get affected by numerous amounts of factors. Using ANN and SVM with some other models were among some of the most recommended models for machine learning used for image classification which got assorted in a number of ways [10][11]. Multiple approaches have been experimented on diverse datasets by configuring the pre-trained convolutional neural network models, especially for image classification [12][13][14]. Utilizing and developing a well-established CNN model for image classification on a custom dataset with the introduction of transfer learning have eased the task by employing the pre-calculated weights of the pre-established models like VGG16, Resnet, InceptionV3, Xception, etc. to obtain the classified outputs.

The process of how transfer learning has helped us out over the pace of time has resulted in an intriguing solution in the field of computer vision and image classification. Transfer learning for automating brain image classification was used by implementing Deep Convolutional Neural Network (DCNN) [15] in which pre-trained DCNNs like Inception, Resnet, and many more were implemented from which it revealed Alex-net to result in the best among all. Bansal et al. used VGG16 accompanied by various handcrafted feature extraction methods like SIFT, SURB, and ORB for corner detector algorithm classified through Random Forest with the result of 93.78% accuracy [16]. Goel et al. demonstrated the efficiency of transfer learning on the classification of Gujarati handwritten CIFAR dataset with an accuracy of 94.98% of the EfficientNet model [17]. Narvekar et al. analyzed the transfer learning implementation for creating small-sized lightweight models [18]. Noor et al. applied DCNN with layer freezing and data augmentation to achieve higher accuracy in animal facial recognition [19][22] with achieving a 100% test accuracy. Huang et al. proposed and compared based on three pre-trained CNN models to find out the most optimal output for waste classification in [20]. Also, Kumar proposed a technique that helped to simplify to control of large voluminous data which he showed in [21] by using some CNN and transfer learning techniques. The approach of transfer learning indeed helped out in a number of ways to develop a lightweight model with a significant accuracy ratio.

III. Dataset Description

As this model will be useful for the exploration of our university, we have used our dataset for image classification. Here dataset has a total of 4502 color images of 256 x 256 resolution and all these images were divided for testing and training. Images of three different labs present in our university are captured for the dataset in which labs like Apple Lab, Sophos Lab, and Virtual Reality Lab are present. Also, for further level classification of lab amenities, we have introduced one more category that is Sophos rack which is one of the amenities present in one of the labs. All the images were gathered manually where photographs were captured in such a way that every portion of labs and the amenity gets covered for the dataset; additionally, for every portion, there are images with different angles, so that if a person captures a photo while visiting the lab which could be from any angle, then the model can predict the category of the image class and deliver the perfect classified output. Each category has a different number of images in the dataset as per the size of the lab or the lab amenity. The larger the size of the lab, the more the number of images present in the dataset. Below table shows the dataset description which has the count of images of each category.

TABLE I. Dataset Descriptor

Collection	Training Image set	Testing Image set	Total Images
Apple Lab	853	146	999
Sophos Lab	1983	172	2155
Sophos Rack	422	38	460
Virtual Reality Lab	838	50	888
Total	4096	406	4502

Here training and testing dataset was divided in 91% and 9% respectively.

IV. Methodology

In this section, we proposed data preprocessing and the framework which is built using the pre-trained convolutional neural network. Since the dataset was completely collected manually, therefore the first step for the dataset was to preprocess the input data that is the image captured is passed through the acquisition process where the image is converted into a numerical array and then the images are resized from any shape to a standardized shape which we have kept as

224x224x3. Different augmentation techniques were implemented for getting better results in the classification of an image where augmentation techniques from the Image Data Generator library like Horizontal and Vertical Shift Augmentation of 0.2, Horizontal and Vertical Flips Augmentation was set to true, Random Zoom Augmentation of 0.2, Random Rotation Augmentation of 40, Shear Range of 0.2 and Fill mode with value nearest were applied to increase the training dataset image frequency and also help in generalizing the input images which helped in increase the data size and train the models with more images accumulating multiple scenarios

overcome this problem, transfer learning has gained tremendous popularity in recent years. Through the use of knowledge from the appropriate source domain, transfer learning aims to improve performance in the target domain.

Transfer learning reduces the amount of data required to execute the intended job. To apply transfer learning on any given dataset for a particular task necessitates the knowledge of what pre-trained model to select, the analogy between the source and the target domain, the size of the selected problem, and the amount of augmentation and fine-tuning needed to apply for the

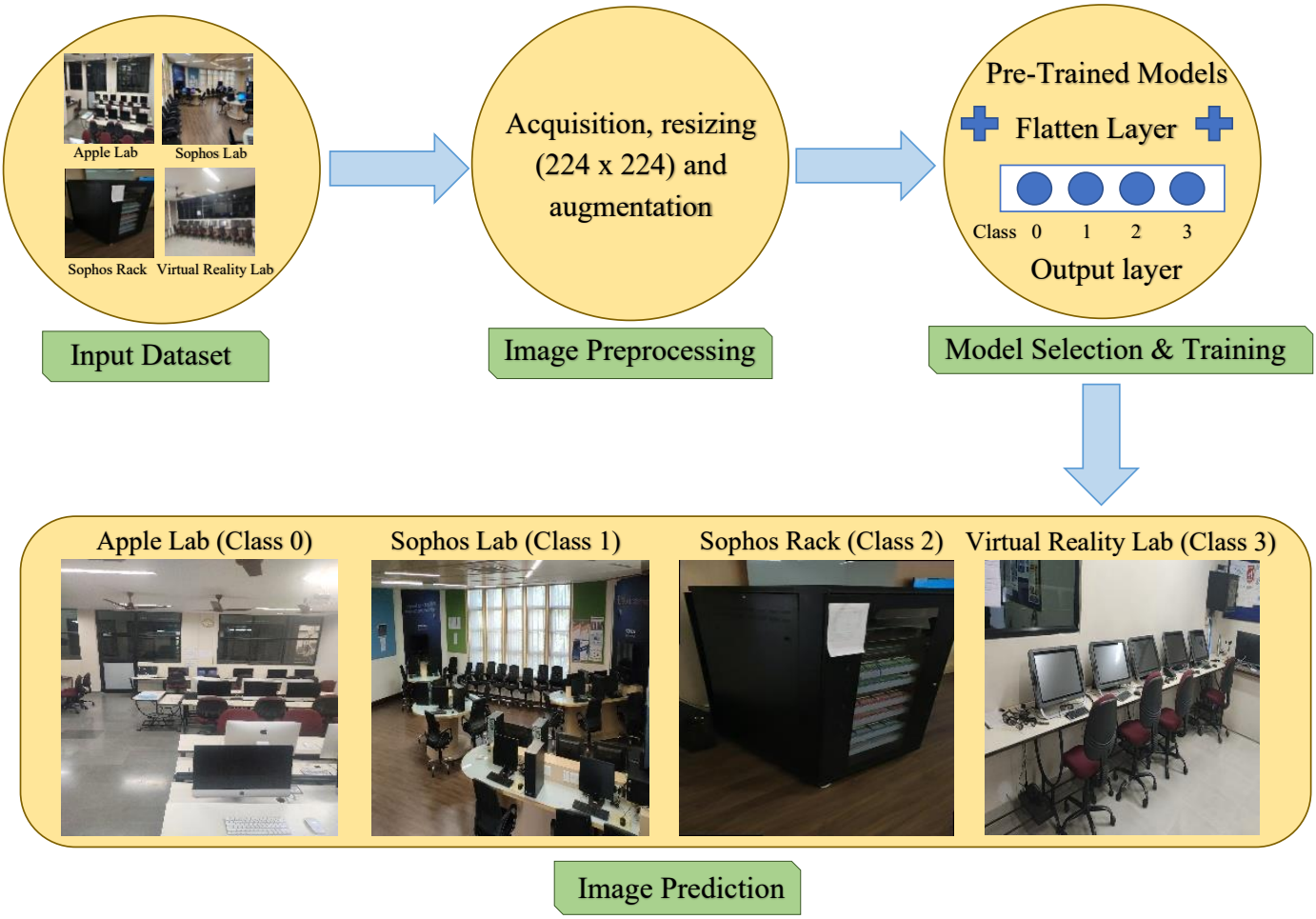

Fig. 1. Workflow of Proposed work displaying how the input data is then preprocessed and then passed through the pretrained model with the addition of the final layers which then are classified into the relevant classes as per the training.

in the image.

Four popular CNN architectures are trained on the handmade dataset which are VGG-16, InceptionV3, Resnet50, and Xception followed by a flattening and the output layer. These four models are utilized for the extraction of the features of the images by using the transfer learning approach. CNN is considered one of the deep learning techniques to solve the computer vision problems without extracting the manually created features. It can be difficult to get the massive amounts of data needed for classification and huge volumes of data are frequently required to train a CNN. As a modern approach to

best results if required. For fine-tuning we have set various hyperparameters for all the four pre-trained models. First of all, we have kept batch size as 32, activation function as softmax in the last layer, loss function as categorical_crossentropy, and epochs as 10 for all pre-trained models. Besides this, we have used adam optimizer with a learning rate of 0.001 for VGG16, adam optimizer with a learning rate of 0.0001 for ResNet50, and the RMSprop optimizer with a learning rate of 0.0001 for InceptionV3, and the SGD optimizer with a learning rate of 0.001 for Xception pre-trained model.

The pre-trained model VGG-16, InceptionV3, Resnet50, and Xception is in the proposed framework system from which Xception was accounted for the study since it delivered the best results among others. The ImageNet dataset which is used for the training of the pre-trained CNN networks is regarded as the source domain. Whereas the target domain consists of recognizing the input images by telling us which of the four classes which are Apple lab, Sophos lab, Sophos Rack, and VR lab the image belongs. Here as the target domain is quite analogous to the source task hence, we decided to use the freezing and training methodology. The implemented technique accelerated the neural network by progressively freezing the hidden layers of the pre-trained model and giving the end layers open for modification as per the required target domain results. After passing the dataset images from the pre-trained model it has been passed through the flatten layer where the resultant output images of the last layer of the pre-trained CNN models were flattened into a single dimensional long feature array for input to the next layer which consists of a fully connected layer of 4 neurons as there consist 4 target classes in our target task which are Apple lab, Sophos lab, Sophos rack, and Virtual Reality lab. The classification of the four classes is carried out by the softmax classifier in the output layer which would provide the result of delivering the final classified output of the image to which of the four classes the image belongs to.

V. Result Analysis

In this section, the result obtained from the research work is presented. The Hardware configuration of the system used for this research work is Intel 2-core Xeon 2.2 GHz CPU, Nvidia Tesla K80 GPU, 13GB Ram, and 108 GB HDD. Python 3.7.13 and deep learning frameworks like Keras, and TensorFlow has been used to build the project. The training dataset contains 4096 images while the test dataset has 406 images constituting the former dataset to be 91% of total images while the latter dataset consists of 9% of total images. A detailed description of the dataset used for classification is in section 3. Here images were reshaped to 224 X 224 pixels and passed through the pre-trained model. Batch size of 32 and 10 epochs was used for model training. The output of the pre-trained model was passed to flatten layer to convert into a linear vector after which the output of the flattening layer was passed to the dense layer having 4 neurons having SoftMax as activation function which will classify images into 4 different classes (Apple lab, Sophos lab, Sophos rack, and VR lab). Various evaluation parameters were assessed for the obtained result which includes training and testing accuracy, precision, recall, F1 score, size of the model, Inference Time, and FLOPs. Analysis and comparison of results obtained from 4 different pre-trained models which were used for the proposed work enumerated in Table 2.

In this research work, out of all four pre-trained model testing accuracy, recall, precision, and F1 score are minimum for VGG16 with 97.22% as training accuracy, 94.79% as testing accuracy, 93.58% as recall, 96.87% as precision, 94.92% as F1 Score; having a model size of 54.7 MB, whereas the maximum training & testing accuracy, recall, precision, and F1 score is obtained by using Xception pre-trained model with 99.90% as training accuracy, 99.75% as testing accuracy, 99.35% as recall, 99.85% as precision, and 99.60% as F1 score; having the model size of 83 MB. The time required to predict any random image (Inference time) is similar and low for all the pre-trained is around 50-60 milliseconds. VGG 16 is having the minimum number of Floating Points Operations (FLOPs) which is 14810834 as compared to other three pre-trained model: ResNet50, InceptionV3, and Xception having 23856378, 2195635, and 22153911 Floating Point Operations (FLOPs) respectively.

TABLE II. Analysis and Results of Various Evaluation Parameters Obtained from Pretrained Models.

Model Name	VGG 16	ResNet50	InceptionV3	Xception
Model Size	54.7MB	95.1MB	85.6MB	83MB
Inference Time	0.049 s	0.050 s	0.057 s	0.056 s
Training Accuracy	97.22%	96.79%	99.83%	99.90%
Testing Accuracy	94.79%	96.77%	99.50%	99.75%
Recall	93.58%	97.44%	98.75%	99.35%
Precision	96.87%	97.35%	99.70%	99.85%
F1 Score	94.92%	97.34%	98.21%	99.60%
FLOPs	14810834	23856378	21956035	22153911

VI. Conclusion And Future Work

In this paper, the image classification of the lab and its resources is addressed using the Transfer Learning approach. At first, we had created our dataset of 3 labs and 1 amenity present in one of the labs. Preprocessing with data augmentation was performed on images of the dataset. After that images were passed through 4 pre-trained network models: VGG 16, ResNet50, Inception V3, and Xception with some modification in the last final layers. Finally, the proposed model was evaluated based on 8 criteria which include the size of the model, inference time, training accuracy, testing accuracy, recall, precision, F1 score, and B-Flops on different pre-trained network models. The best outcome was obtained by using Xception pre-trained model among all 4 pre-trained neural network models with training accuracy of 99.90%, testing accuracy of 99.75%, recall of 99.35%, and precision of 99.85%, F1 score of 99.60%, and model size of 83 MB.

In the future, the dataset can be changed and extended as per the needs. Also, different approaches and different pre-trained models can be taken into consideration for obtaining better results outputs.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556, 2014

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[5] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," arXiv preprint, p. 1610.02357, 2017

[6] M. M. Hasan, Z. Wang, M. A. I. Hussain, and K. Fatima, "Bangladeshi Native Vehicle Classification Based on Transfer Learning with Deep Convolutional Neural Network," *Sensors*, vol. 21, no. 22, p. 7545, Nov. 2021, doi: 10.3390/s21227545.

[7] Z. Le, "A transfer learning approach for handwritten numeral digit recognition," in *ACM International Conference Proceeding Series*, Jan. 2020, pp. 140–145. doi: 10.1145/3378936.3378970.

[8] A. Moumni and A. Lahrouni, "Machine Learning-Based Classification for Crop-Type Mapping Using the Fusion of High-Resolution Satellite Imagery in a Semiarid Area," *Scientifica (Cairo)*, vol. 2021, 2021, doi: 10.1155/2021/8810279.

[9] Q. Xiang, G. Zhang, X. Wang, J. Lai, R. Li, and Q. Hu, "Fruit image classification based on Mobilenetv2 with transfer learning technique," Oct. 2019. doi: 10.1145/3331453.3361658.

[10] Chandra, M.A., Bedi, S.S. Survey on SVM and their application in image classification. Int. j. inf. tecnol. 13, 1–11 (2021). https://doi.org/10.1007/s41870-017-0080-1

[11] Zheng Fang, Gong Zhang, Qijun Dai, Biao Xue, Peng Wang. (2022) A hybrid network for PolSAR image classification based on polarization orientation angles. Remote Sensing Letters 13:4, pages 383-393.

[12] Gadri, S., Neuhold, E. (2020). Building Best Predictive Models Using ML and DL Approaches to Categorize Fashion Clothes. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds) Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science(), vol 12415. Springer, Cham. https://doi.org/10.1007/978-3-030-61401-0_9

[13] Patel, H., Upla, K.P. A shallow network for hyperspectral image classification using an autoencoder with convolutional neural network. Multimed Tools Appl 81, 695–714 (2022). https://doi.org/10.1007/s11042-021-11422-w

[14] Andrea Loddo, Mauro Loddo, Cecilia Di Ruberto,A novel deep learning based approach for seed image classification and retrieval,Computers and Electronics in Agriculture,Volume 187,2021,106269,ISSN 0168-1699,https://doi.org/10.1016/j.compag.2021.106269.

[15] Kaur, T., Gandhi, T.K. Deep convolutional neural networks with transfer learning for automated brain image classification. Machine Vision and Applications 31, 20 (2020). https://doi.org/10.1007/s00138-020-01069-2

[16] Bansal, M., Kumar, M., Sachdeva, M. et al. Transfer learning for image classification using VGG19: Caltech-101 image data set. J Ambient Intell Human Comput (2021). https://doi.org/10.1007/s12652-021-03488-z

[17] P. Goel and A. Ganatra, "A Pre-Trained CNN based framework for Handwritten Gujarati Digit Classification using Transfer Learning Approach," *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022, pp. 1655-1658, doi: 10.1109/ICSSIT53264.2022.9716483.

[18] C. Narvekar and M. Rao, "Flower classification using CNN and transfer learning in CNN- Agriculture Perspective," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 660-664, doi: 10.1109/ICISS49785.2020.9316030.

[19] Alam Noor, Yaqin Zhao, Anis Koubaa, Longwen Wu, Rahim Khan, Fakheraldin Y.O. Abdalla, Automated sheep facial expression classification using deep transfer learning, Computers and Electronics in Agriculture, Volume 175, 2020, 105528, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2020.105528.

[20] Huang, G-L, He, J, Xu, Z, Huang, G. A combination model based on transfer learning for waste classification. Concurrency Computat Pract Exper. 2020; 32:e5751. https://doi.org/10.1002/cpe.5751

[21] Kumar, Dr. (2020). Video based Traffic Forecasting using Convolution Neural Network Model and Transfer Learning Techniques. Journal of Innovative Image Processing. 2. 128-134. 10.36548/jiip.2020.3.002.

[22] Patel, Priyang, et al. "Advancements in Cloud-Based Solution for Medical Imaging: A Survey." 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2021.