Science and Technology Publishing (SCI & TECH)
ISSN: 2632-1017
Vol. 8 Issue 12, December - 2024

Comparison Of Pre-Trained Machine Learning Models For Hot Pepper Disease Detection And Classification

[1]Ezema D.C.

Department of Electrical and Electronic Engineering
[3]State University of Medical and Applied Sciences(SUMAS),Igbo-Eno,
Enugu State, Nigeria
donatus.ezema@sumas.edu.ng

[3]Geku Diton

Department of Electrical and Electronic Engineering
Federal University Otuoke, Bayelsa State.
Email: gekudd@fuotuoke.edu.ng

[3]Okafor, Joyce Odu

Department of Electrical and Electronic Engineering,
Federal University Otuoke, Bayelsa State
ORCID: 0009-0001-3983-6994
Odujp@fuotuoke.edu.ng

***Abstract*— In this paper comparison of pre-trained machine learning models for Hot Pepper disease detection and classification is presented. Specifically, the following four different machine learning models are considered: Residual Neural Network 50, Inception V3, Residual Neural Network 152 V2 and Mobile Net V2. The image dataset used in this work were obtained from the Hot Pepper disease images documented by the Syiah Kuala University, Indonesia. Auto-orientation of pixel data (with EXIF-orientation stripping) was done to ensure the diseased areas are effectively captured as accurate as possible; each image was cropped to 640 x 640 pixels from each original image. In order to train and validate the models, dataset splitting was done as follows: 70% of training dataset, 10% of validation dataset, and 20% of testing dataset. The models were trained and validated over 20 epochs. The results showed that MobileNetV2 has the highest training accuracy of 99.74 % but the validation accuracy of 62 % whereas the InceptionV3 has the training accuracy of 94.39 % and validation accuracy of 78 %. It can be concluded that based on accuracy performance, the InceptionV3 model is the best model. Again, the InceptionV3 has the highest precision value of 80.0% and the highest recall value of 79.0%. In addition, the InceptionV3 has the highest F1 score value of 78.0%. in all, the InceptionV3 model is the best model among the four models considered.**

Keywords — Residual Neural Network 50, Pre-trained Model, Inception V3, Machine Learning Model, Residual Neural Network 152 V2 , transferred Learning , Mobile Net V2

1. INTRODUCTION

In recent years, the growing population and worsening global climatic conditions are putting mounting pressure on farmers to meet the global food demand [1,2,3]. The problem has attracted many researchers and solution approaches. Some researchers have advocated for genetically modified food while some have opted for precision farming and smart agricultural techniques [4,5,6].

Again, one of the perennial challenges face by farmers is disease; plant diseases can greatly affect the plant yield, cause great losses and spread to other farms which extends the damages due to the plant disease [7,8,9]. Early detection of the plant disease can greatly avert these losses [10,11]. As such, farmers are employing technologies to ensure early detection and classification of plant diseases to minimize the damages that can be caused by plant diseases. In this study, machine learning models are used to carryout Hot pepper disease detection and classification. The performance of the selected machine learning models are then compared [12,13,14]. The essence of the study is to identify the most suitable machine learning model for the Hot pepper disease detection and classification.

2. METHODOLOGY

This study is focused on comparing the effectiveness of four different pre-trained machine learning models in detecting and predicting hot pepper diseases. Specifically, the following four different machine learning models are considered: Residual Neural Network 50, Inception V3, Residual Neural Network 152 V2 and Mobile Net V2.The flow diagram for the hot pepper leave diseases detection and classification machine learning models training and evaluation is shown in Figure 1. The diagram in Figure 1 shows the key steps that were followed while carrying out the study; it highlighted the different stages or phases in the model training and evaluation.

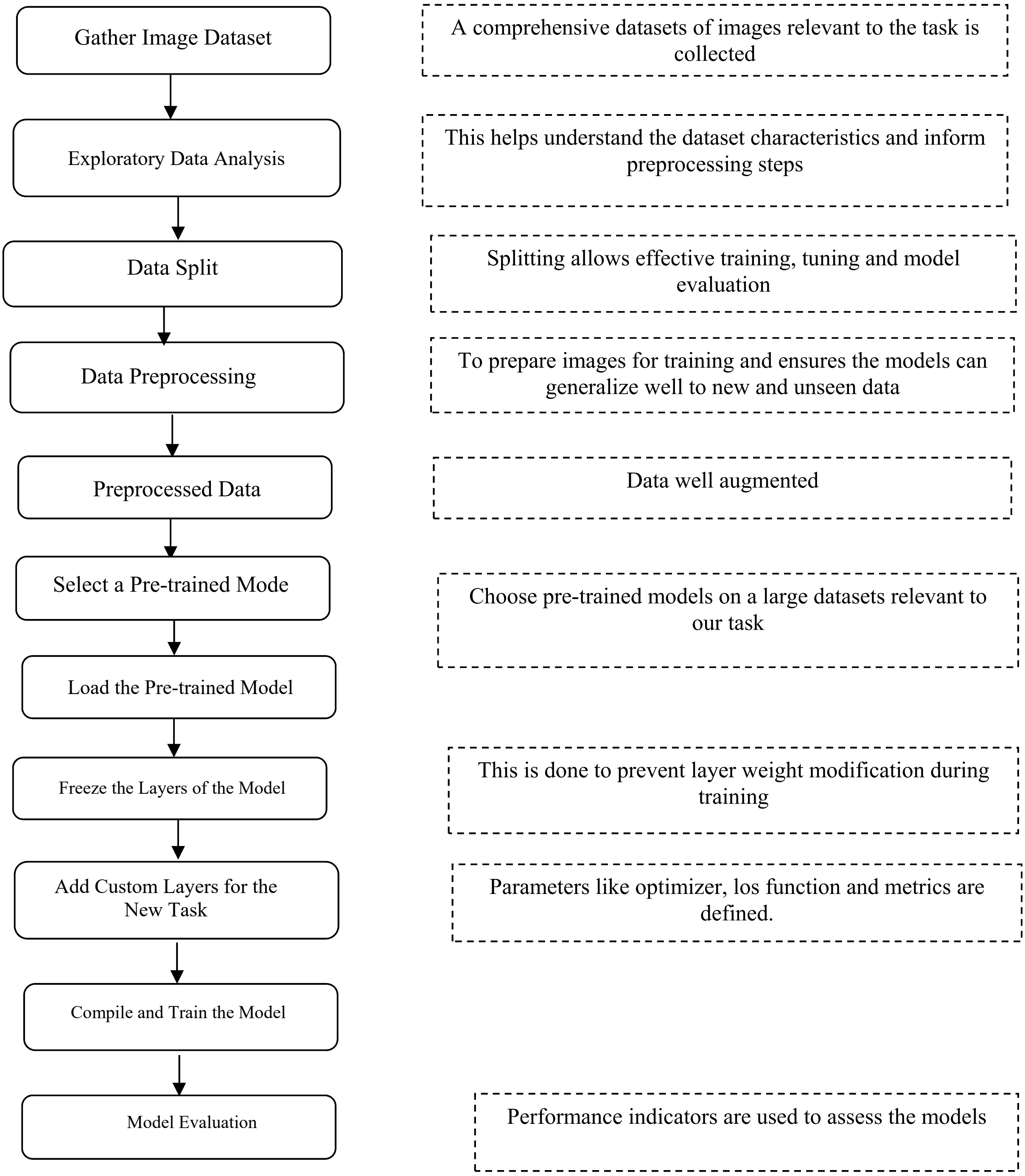


Figure 1 The flow diagram for the machine learning models training and evaluation

2.2 Data collection and preprocessing

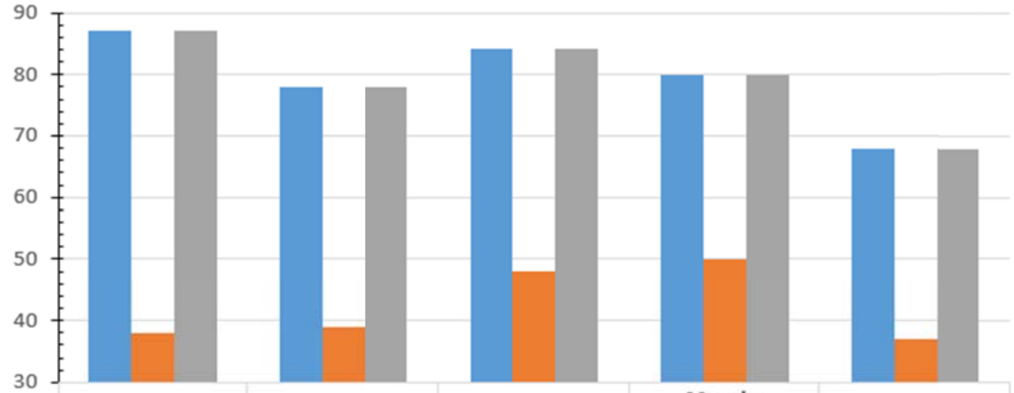
The image dataset used in this work were obtained from the Hot Pepper disease images documented by the Syiah Kuala University, Indonesia. The enhanced bar chart in Figure 2 shows the number of image samples in the dataset which has four different diseases and a healthy pepper dataset. Some image samples of the diseased peeper leaves are displayed in Figure 3**.**

The cropping of images carried out in this work was manually done, auto-orientation of pixel data (with EXIF-orientation stripping) was done to ensure the diseased areas are effectively captured as accurate as possible; each image was cropped to 640 x 640 pixels from each original image.

The augmentation steps applied to the source image were as follow;

i. 50% possibility of vertical flip
ii. 50% possibility of horizontal flip
iii. Random rotation of between -45 and +45 degrees, as shown in Figure 4.

In order to train and validate the models, the 70%:10%:20% data split combination, as shown in Figure 4, was applied. Dataset splitting was done as follows: 70% of training dataset, 10% of validation dataset, and 20% of testing dataset, as shown in Figure 5. The hyperparameter settings for the training of the four models are presented in Table 1. The number of model training parameters in million and model size in MB are shown in Figure 6. According to Figure 6, the Residual Neural Network 152 V2 is the heaviest model with size of 232 MB and 60.4 M training parameters while the Mobile Net V2 is the lightest model with size of 14 MB and 3.5 M training parameters.



	Healthy	Leaf Curl	Leaf Spot	Mouche Blanche,	Yellowish
Original Images	87	78	84	80	68
Augmented Samples	38	39	48	50	37
Cropped Samples	87	78	84	80	68

Figure 2 The number of image samples in the 4 different disease and healthy Hot Pepper images

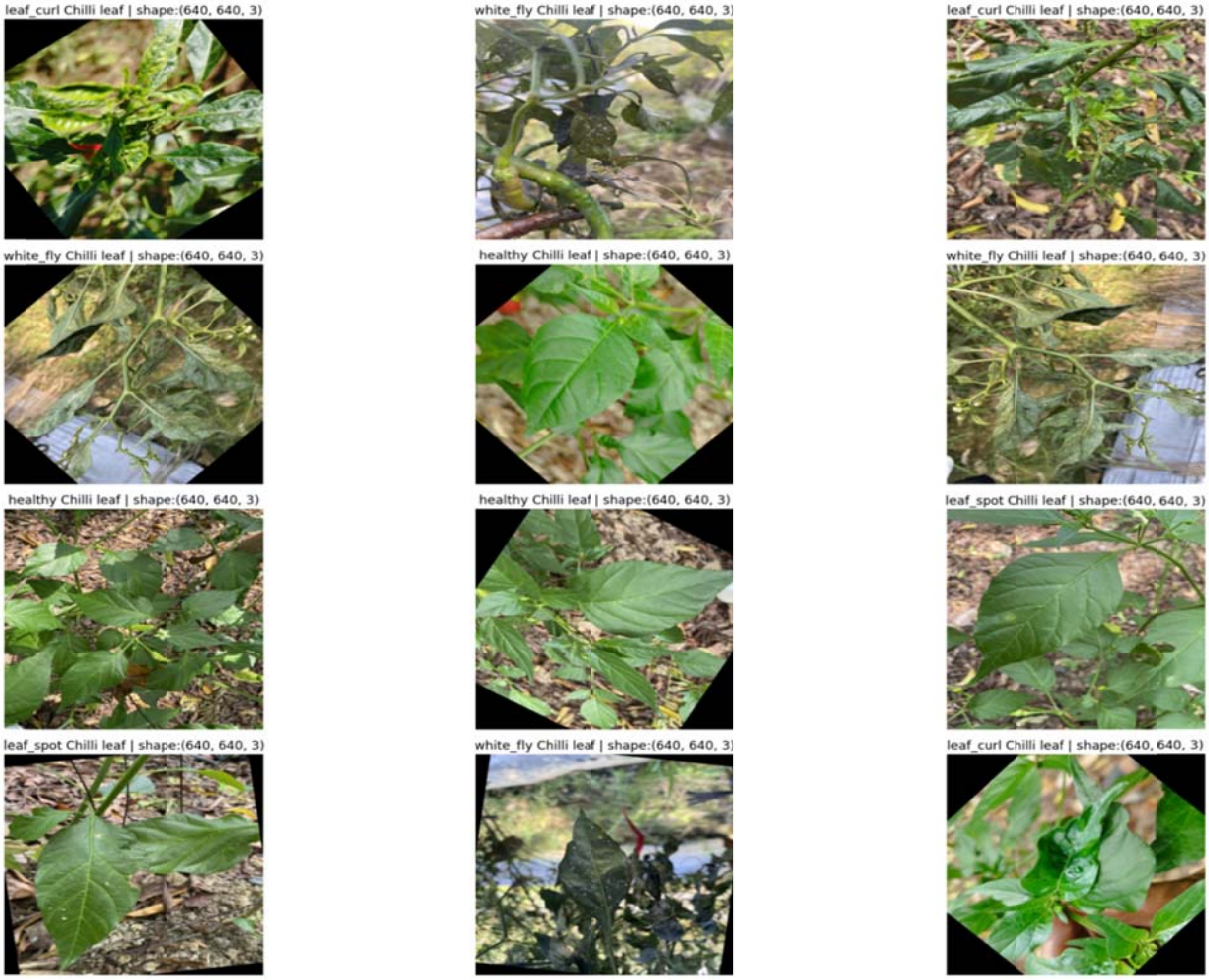


Figure 3 Sample Images and Diseases Source: Syiah Kuala University, Indonesia (2024)

Figure 4 Sample of Mouche Blanche disease image augmentation (Rotation of between -45 and +45 degrees)

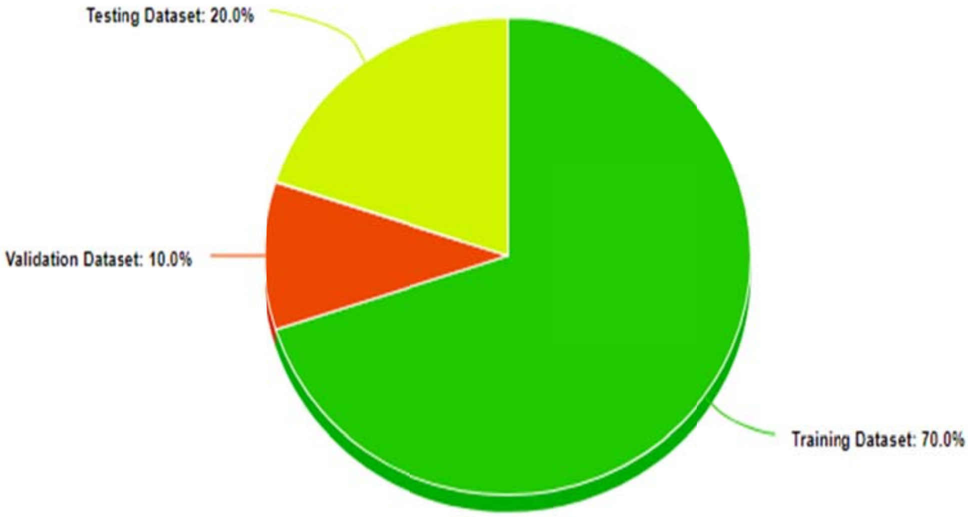


Figure 5 The bar chart showing the data splitting in percentages

Table 1: The hyperparameter settings for the Training of the four models

	ResNet50	InceptionV3	ResNet152V2	MobileNetV2
Parameters	Training Configuration Values	Training Configuration Values	Training Configuration Values	Training Configuration Values
Optimizer	Adam	Adam	Adam	Adam
Loss Function	categorical_crossentropy	categorical_crossentropy	categorical_crossentropy	categorical_crossentropy
Metrics Tracked	Accuracy, Precision and Recall	Accuracy, Precision and Recall	Accuracy, Precision and Recall	Accuracy, Precision and Recall
Batch Size	32	32	32	32
Epochs	20	20	20	20
Training Parameters	25.6M	23.9M	60.4M	3.5M
Model size	98MB	92MB	232MB	14MB

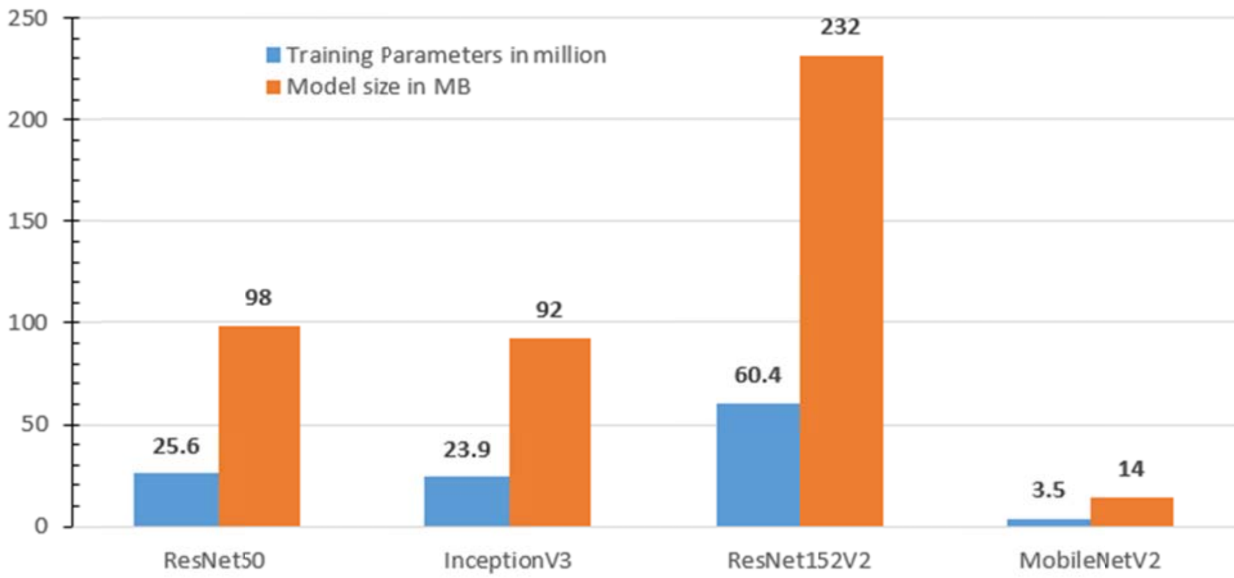


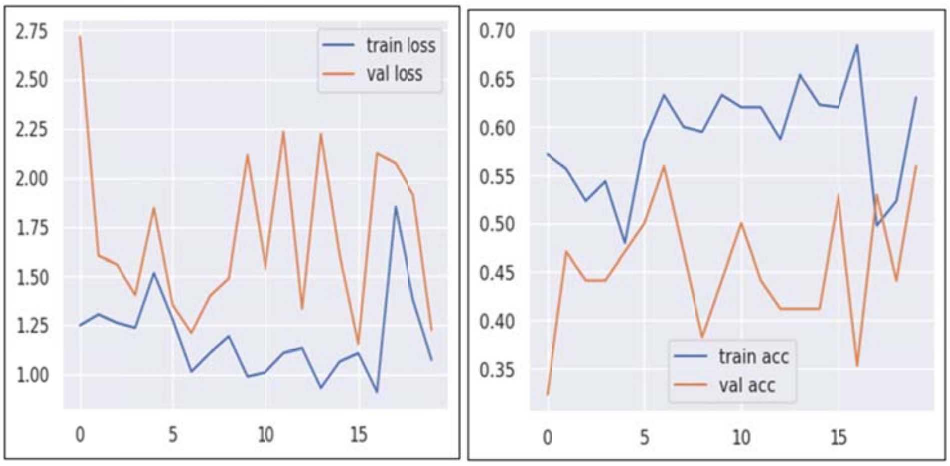
Figure 6 Number of model training parameters in million and model size in MB

3 RESULT AND DISCUSSION

The models were trained and validated over 20 epochs. The training loss and validation loss, as well as the training accuracy and validation accuracy results were captured for each of the four models and the results are as shown in Figure 7, Figure 8, Figure 9 and Figure 10. Also, the overall training and validation accuracy results are captuered for the four models and the comparison of the overall training accuracy and validation accuracy for the four pre-trained models is presented in Figure 11 and it showed that MobileNetV2 has the highest training accuracy of 99.74 % but the validation accuracy of 62 % whereas the InceptionV3 has the training accuracy of 94.39 % and validation accuracy of 78 %. It can be concluded that based on accuracy performance, the InceptionV3 model is the best model. This is because, the sharp drop in the accuracy of the validation test for the MobileNetV2 model can be an indication of model over fitting. In any case, the ResNet50 model has the worst accuracy performance with a value of 63.03 % on the training dataset.

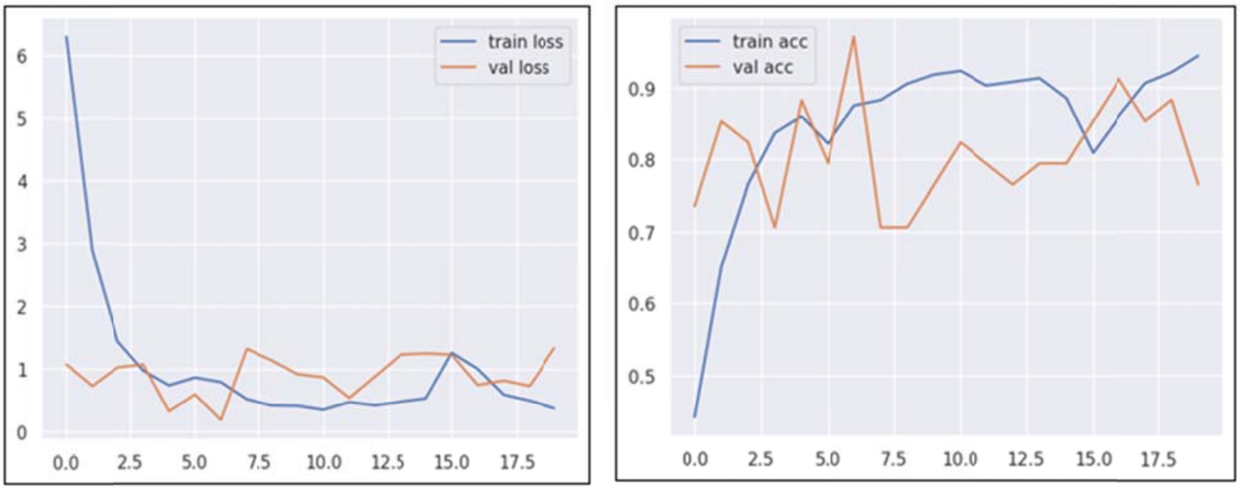
Comparison of the precision and recall for the four pre-trained models is presented in Figure 12 and it showed that the InceptionV3 has the highest precision value of 80.0% and the highest recall value of 79.0%. It can be concluded that based on precision and recall performance, the InceptionV3 model is the best model.

Comparison of the precision and F1 score for the four pre-trained models is presented in Figure 13 and it showed that the InceptionV3 has the highest F1 score value of 78.0%. It can be concluded that based on precision and F1 score l performance, the InceptionV3 model is the best model.



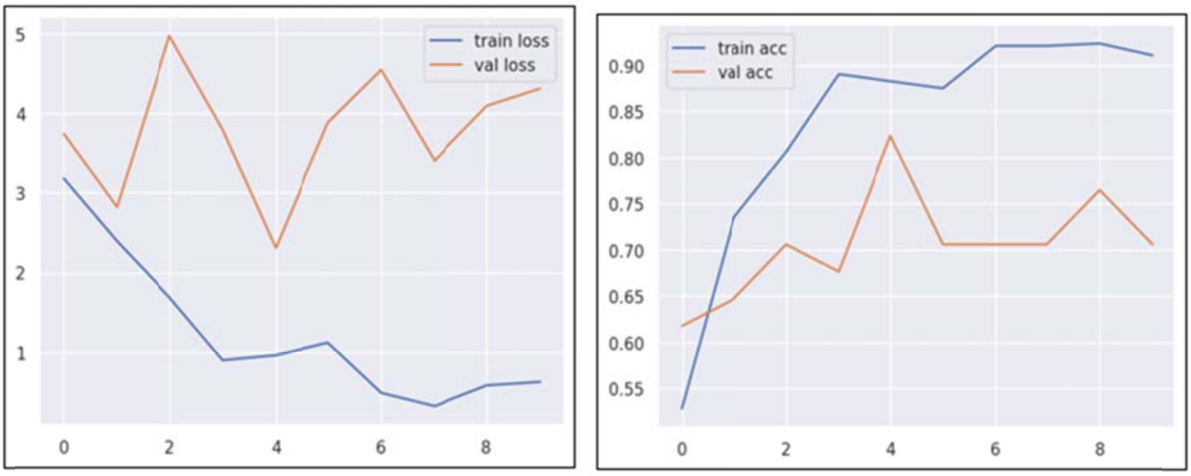
(a) (b)

Figure 7 (a) The training and validation loss for Residual Neural Network 50 model (b) The training and validation accuracy for Residual Neural Network 50 model



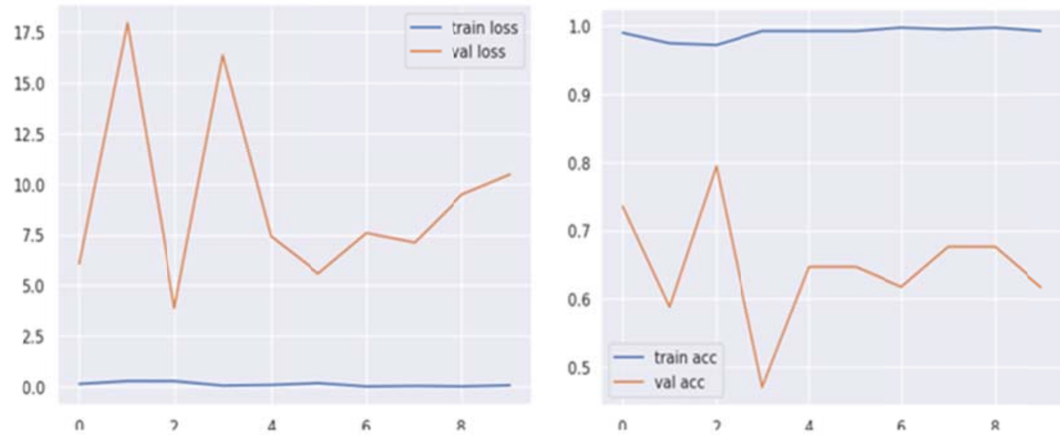
(a) (b)

Figure 8(a) The training and validation loss for Inception V3 model (b) The training and validation accuracy for Inception V3 model



(a) (b)

Figure 9 (a) The training and validation loss for Residual Neural Network 152 V2 model (b) The training and validation accuracy for Residual Neural Network 152 V2 model



(a) (b)

Figure 10 (a) The training and validation loss for Mobile Net V2model (b) The training and validation accuracy for Mobile Net V2model

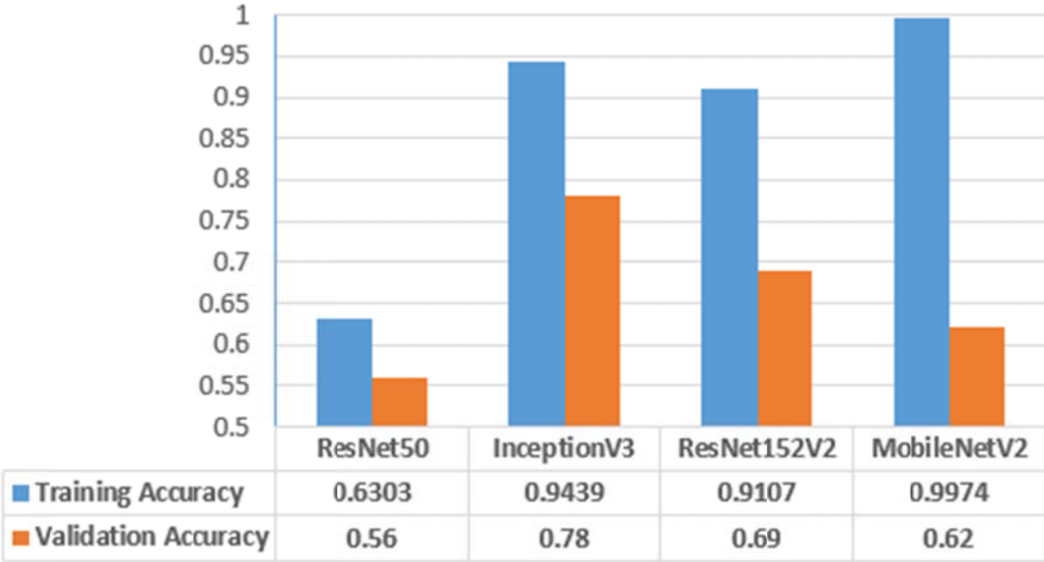


Figure 11 Comparison of the training accuracy and validation accuracy for the four pre-trained models

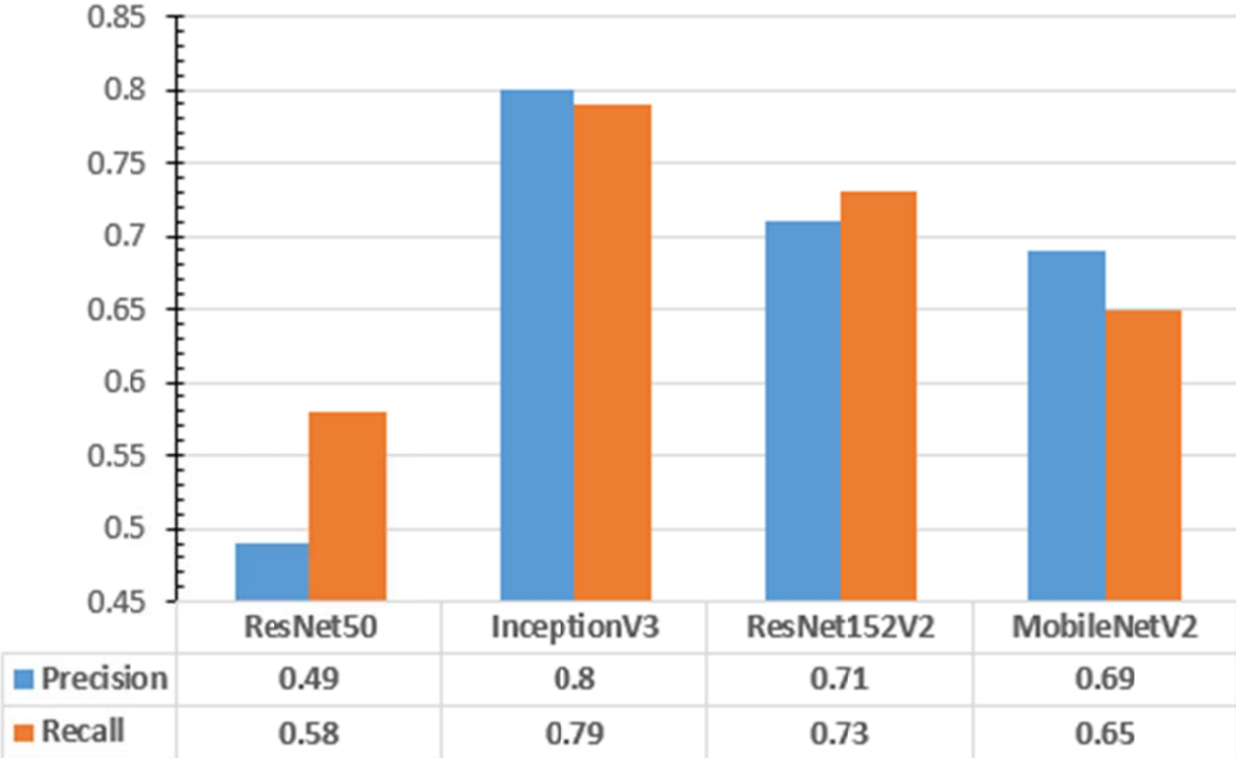


Figure 12 Comparison of the precision and recall for the four pre-trained models

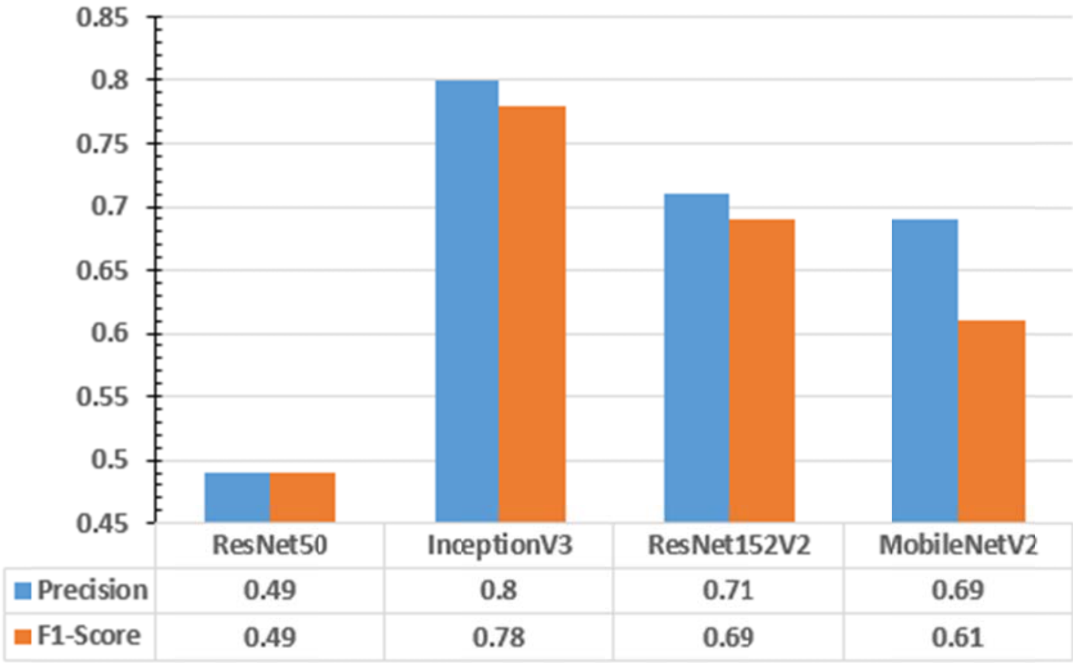


Figure 13 Comparison of the precision and F1 score for the four pre-trained models

4. CONCLUSION

Four different pre-trained models are employed for the detection and classification of Hot Pepper disease. The study used to models in a transferred learning approach which enabled the use of a small case study dataset in fine-tuned pre-trained models and still achieve good prediction performance. This is because the pre-trained models had been trained using large dataset such that the use of small dataset in a transferred learning setting still achieve good results. The models considered in the study are Residual Neural Network 50, Inception V3, Residual Neural Network 152 V2 and Mobile Net V2. The results showed that the Inception V3 model had the best performance.

REFERENCES

1. Giller, K. E., Delaune, T., Silva, J. V., Descheemaeker, K., van de Ven, G., Schut, A. G., ... & van Ittersum, M. K. (2021). The future of farming: Who will produce our food?. *Food Security*, *13*(5), 1073-1099.
2. Barrett, C. B. (2021). Overcoming global food security challenges through science and solidarity. *American Journal of Agricultural Economics*, *103*(2), 422-447.
3. Brenton, P., Chemutai, V., & Pangestu, M. (2022). Trade and food security in a climate change-impacted world. *Agricultural Economics*, *53*(4), 580-591.
4. Husaini, A. M., & Sohail, M. (2023). Robotics-assisted, organic agricultural-biotechnology based environment-friendly healthy food option: Beyond the binary of GM versus Organic crops. *Journal of Biotechnology*, *361*, 41-48.
5. Maraveas, C., Konar, D., Michopoulos, D. K., Arvanitis, K. G., & Peppas, K. P. (2024). Harnessing quantum computing for smart agriculture: Empowering sustainable crop management and yield optimization. *Computers and Electronics in Agriculture*, *218*, 108680.
6. Raj, M., Harshini, N. B., Gupta, S., Atiquzzaman, M., Rawlley, O., & Goel, L. (2024). Leveraging Precision Agriculture Techniques using UAVs and Emerging Disruptive Technologies. *Energy Nexus*, 100300.
7. Jones, R. A. (2021). Global plant virus disease pandemics and epidemics. *Plants*, *10*(2), 233.
8. Gullino, M. L., Albajes, R., Al-Jboory, I., Angelotti, F., Chakraborty, S., Garrett, K. A., ... & Stephenson, T. (2022). Climate change and pathways used by pests as challenges to plant health in agriculture and forestry. *Sustainability*, *14*(19), 12421.
9. Tariq, H., Xiao, C., Wang, L., Ge, H., Wang, G., Shen, D., & Dou, D. (2024). Current Status of Yam Diseases and Advances of Their Control Strategies. *Agronomy*, *14*(7).
10. Terentev, A., Dolzhenko, V., Fedotov, A., & Eremenko, D. (2022). Current state of hyperspectral remote sensing for early plant disease detection: A review. *Sensors*, *22*(3), 757.
11. Liu, X., Min, W., Mei, S., Wang, L., & Jiang, S. (2021). Plant disease recognition: A large-scale benchmark dataset and a visual region and loss reweighting approach. *IEEE Transactions on Image Processing*, *30*, 2003-2015.

12. Sabanci, K., Aslan, M. F., Ropelewska, E., & Unlersen, M. F. (2022). A convolutional neural network-based comparative study for pepper seed classification: Analysis of selected deep features with support vector machine. *Journal of Food Process Engineering*, *45*(6), e13955.
13. Kodukula, S., Vidyullatha, P., Ghantasala, G. P., Vysali, P., Dubba, N., & Nassar, I. (2024). Assessing the Efficacy of IoT-driven Machine Learning Models in Enhancing Chili Crop Growth and Yield Quality.
14. Kaur, A., Kukreja, V., Gopal, L., Verma, G., & Sharma, R. (2024, January). Next-Gen Hybrid Deep Learning for Accurate Pepper Leaf Mosaic Virus Classification. In *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)* (pp. 1-4). IEEE.