A Comparative Study of MobileNetv2 and VGG-16 Convolutional Neural Network Architectures for Identification of Medicinal Mushrooms

Mark Jayson Y. Sutayco
School of Electrical, Electronics, and Computer Engineering
Mapua University
Manila, Philippines
mjysutayco@mapua.edu.ph

Meo Vincent C. Caya
School of Electrical, Electronics, and Computer Engineering
Mapua University
Manila, Philippines
mvccaya@mapua.edu.ph

***Abstract*— Convolutional Neural Networks (CNNs) have considerably enhanced the field of Artificial Intelligence (AI) during the past ten years through deep learning. The efficacy of CNNs has been established in research, particularly in agricultural automation. The researchers in this study assessed the performance of two CNN architectures - MobileNetv2 and VGG-16 - in predicting six medicinal mushrooms - Lions Mane, Oyster, Reishi, Shiitake, Shimeji, and Volva. The aim was to compare the accuracy levels achieved by each architecture. A dataset comprising of 600 image samples of six medicinal mushrooms was employed to train the two architectures, with both training procedures employing an 80 by 20 ratio. 80 percent of the total data was utilized to train the models, while the remaining 20 percent was reserved for the validation set. The results show that MobileNetv2 yielded a testing accuracy rate of 97.3 percent while the VGG-16 achieved testing accuracy of 72.6 percent. This implies that accuracy in predicting medicinal mushroom using MobileNetv2 is higher than VGG-16 by 24.7 percent. Hence, using MobileNetv2 architecture will provide optimal results over VGG-16 in identifying medicinal mushrooms. For future studies, the researchers aim to train these medicinal mushrooms dataset in a continual learning scenario and evaluate the extent of catastrophic forgetting.**

Keywords*— *Medicinal Mushroom, Convolutional Neural Network, MobileNetv2, VGG-16

I. Introduction

The success of Convolutional Neural Networks (CNN) in relation to the task of image identification, detection, and classification have been proven very well in the past decade. CNNs are based on the concept of learning from data, recognizing patterns, and making decisions with minimal human involvement. This powerful deep learning algorithm has been successfully applied to various practical applications, including agriculture. It has helped reduce human labor costs and increased efficiency in various agricultural tasks.

Mushrooms on the other hand, have shown its importance especially in the cycle of nutrients and the flow of energy in ecosystems [1]. Owing to numerous health benefits of mushrooms, several kinds of mushrooms are also used as medicine. For a hundred years, it has been used to treat infections, used as antioxidants, and even used for skincare. At present, medicinal mushrooms are also beneficial in curing lung diseases and cancer. In China and Japan, medicinal mushrooms have been an addition to cancer treatments by either combining it with radiation or chemotherapy (National Cancer Institute, 2021).

Different medicinal mushrooms are used for specific diseases; hence the correct identification and classification of these mushrooms is important. Image classifications using deep learning algorithms have become very popular in this field wherein these algorithms motivated scientists and researchers to automate the process of identification, detection, and classifications of medicinal mushrooms.

Based on ensemble image processing learning such as VGGNet 16, ResNet18, and GoogleNet architecture was integrated by bagging algorithm, researchers in [2] focused on identifying wild mushrooms. Experiments have shown that the developed model has better performance compared to a single CNN model. The study used 10% holdout validation datasets which acquired an accuracy rate of 93.1%. In [3], the study focused on using deep learning algorithms, namely Inception-V3, VGG-16, and ResNet50 to classify edible, inedible, and poisonous mushrooms. The results show that Resnet50 achieved the highest accuracy among the algorithms used in the study with an accuracy rate of 88.40%. These algorithms were tested using 8190 dataset samples, and the researchers split the data into 80/20 ratio. In a previous study referenced as [4], a comparable methodology was employed, albeit using different algorithms. Classifiers including Support Vector Machine (SVM), K-Nearest Neighbor, Logistic Regression, and The Classification tree were utilized to differentiate edible from poisonous mushrooms based on enzymatic browning behavior. The study was evaluated using 135 samples of mushrooms wherein SVM outperformed the other architectures used with an accuracy rate of 80%. Identification of edible mushrooms were given focused in [5] among Agaricus and Lepiota mushrooms. The result of the study shows that both the decision tree algorithm and SVM have acquired the same accuracy. Moreover, to quantify the quality of enoki mushrooms in [6], LeNet architecture was utilized for quality inspection. In [7], researchers diverted their focus on identifying whether the mushrooms are healthy or have been damaged. The damage considered in this study was either caused by microbial or

979-8-3503-8117-7/23/$31.00 ©2023 IEEE

2023 IEEE 15th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) | 979-8-3503-8117-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/HNICEM60674.2023.10589123

mechanical. In this study, the developed model which achieved 90% accuracy rate using support vector machine algorithm was embedded in the microprocessor to be deployed as a device for identification of healthy and damaged mushrooms. Convolutional Neural Network become popular also for developing models for quality inspections of different fruits. This deep learning algorithm was utilized in [8] for fruit grading based on fruits color, size, texture, and shape. CNN has achieved the highest accuracy over competing algorithms such as K-NN, SVM and ANN. CNN was also embedded in the Raspberry Pi microprocessor in [9-13] for grading abaca fiber, cacao classifications, cloud-based signature validation, grading broccoli, and mangosteen, and for plant identification.

Based on reviewed literature and studies, to the researchers knowledge, none of them have provided a study to compare two convolutional neural network architectures applied on the task of identifying medicinal mushrooms. In deep learning tasks, it is important to find the optimal model which can give optimal accuracy in every set of tasks. Hence, given this new target problem, the primary aim of this study was to compare the performance of two existing convolutional neural network architecture namely, VGG-16 and MobileNetv2 in the task of identifying medicinal mushrooms. Specifically, this study will investigate which architecture will generate higher accuracy. To do this, sub objectives are laid: 1) development of a device for image acquisition, 2) data gathering, 3) training of two convolutional neural network architectures, and 4) evaluation and comparison of the results of the performance of two CNN architectures.

The results of this study will be beneficial, especially in identifying which architecture is suitable for the task of identifying medicinal mushrooms. Optimal architecture can be embedded in a system using microprocessors to be deployed and used in medicinal mushroom farms. In general, the study focused on training two models using MobileNetv2 and VGG-16 convolutional neural network architecture. The datasets utilized in this study were restricted to only six categories, specifically Lion's Mane, Shiitake, Reishi, Shimeji, Oyster, and Volva. The study's restriction only pertains to the assessment of the two architectures' performance.

II. MATERIALS AND METHODS

This part of the study outlined the process done to achieve the objective of this study.

A. Image Acquisition Device Development

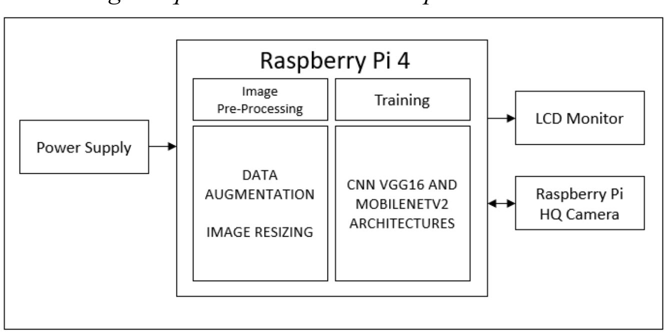

Fig 1. System Block Diagram

Fig. 1 presents the block diagram of the proposed system wherein it also demonstrates how the major component of the device relates to each other.

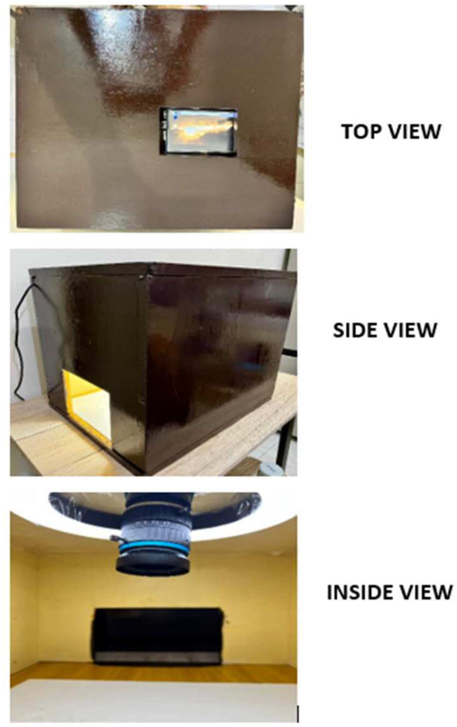

Fig 2. Sample Image of Developed Device

Fig. 2 shows the developed image acquisition device. The actual dimensions of the device in terms of length, width and height are 12x12x12 inches. To easily gather raw data of medicinal mushrooms, an image acquisition device was developed. The newly created device consists of several essential components, such as a Raspberry Pi 4 that acts as the device's microcontroller, an LCD monitor, Raspberry HQ Camera, and a ring light used for illumination.

B. Data Gathering

Fig 3. Sample Images of Gathered Dataset

Fig. 3 shows the sample images of datasets gathered with its labels. To gather data for this study, the researchers manually collected data using the newly developed image acquisition device. The sample medicinal mushrooms used in the study were obtained from farms in Marilao, Bulacan, and were labeled by the farmers themselves. A total of 600 training images were obtained, with each class of medicinal mushroom having 100 training images. These training images were captured using the image acquisition device and had a dimension of 4056 x 3040 pixels.

C. *Software Development*

The software development of this study concentrated on creating a model using two convolutional neural network architectures such as MobileNetv2 and VGG-16.

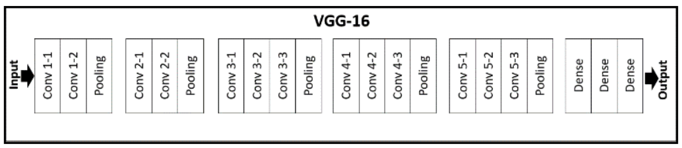

Fig 4. Sample Image of VGG-16 CNN architecture [14]

VGG-16 CNN architecture become popular in 2014 when it bagged the award as best architecture on ImageNet Large Scale Visual Recognition competition. This architecture was able to achieve localization error of 25%. Since then, many studies focused on adapting this architecture in different machine learning projects, including detection, identification, and classifications where it demonstrated state-of-the-art results [14-16].

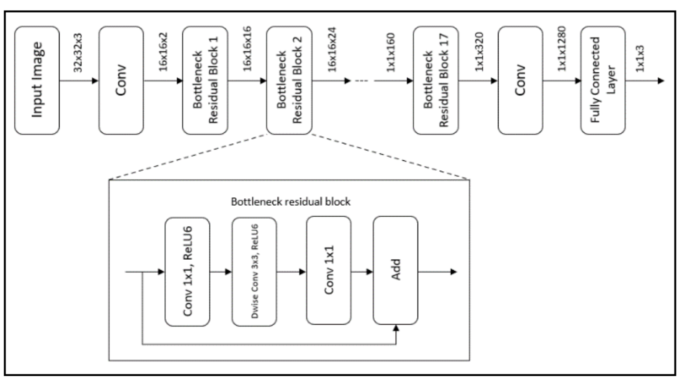

Fig 5. Sample Image of MobileNetv2 CNN architecture [17]

On the other hand, MobileNetv2 is lightweight and can be embedded even on mobile devices. The focus of this architecture is to reduce the required memory, that is why in embedding it does not require high memory requirements. In addition, this architecture is advantageous in terms of feature extraction for tasks such as object detection, segmentation, and classification. Numerous studies have also exhibited the effectiveness of this architecture in classification tasks [18-21].

1. *Model Training*

For the training of the model, both experiments used the pre-trained model of the two architectures. To evaluate the difference in performance of these two architectures, both models were trained with the same epoch number, learning rate, activation functions and optimizers.

In addition, both architecture training adapted the 80/20 training ratio wherein 80% of the training data is used to train the model and 20% is used to validate whether the results of the training are within acceptable values. Since training data are also limited to 600 images only, both training performed data augmentation to create more training data. For MobileNetv2, data is pre-processed using its pre-processing syntax wherein instead of making the RGB pixels into 0-255, the value was scaled to -1.0 to 1.0. The same process was done in pre-processing VGG-16, the data was also pre-processed the way the model expects it to be.

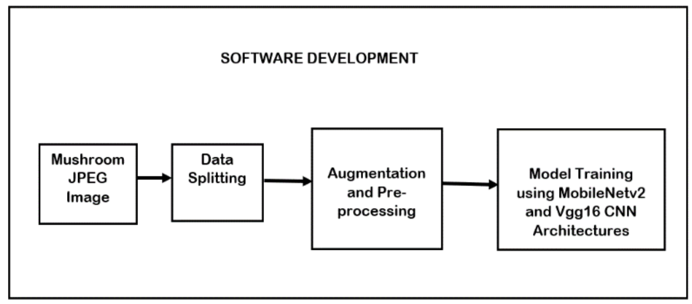

Fig 6. Software Development

Fig. 6 shows the block diagram of software development. In training, both architectures were trained using 5 epochs for base training, and another 5 epochs for final training. The learning rate was set to 0.001, and both training uses adam as the optimizer and softmax as an activation function.

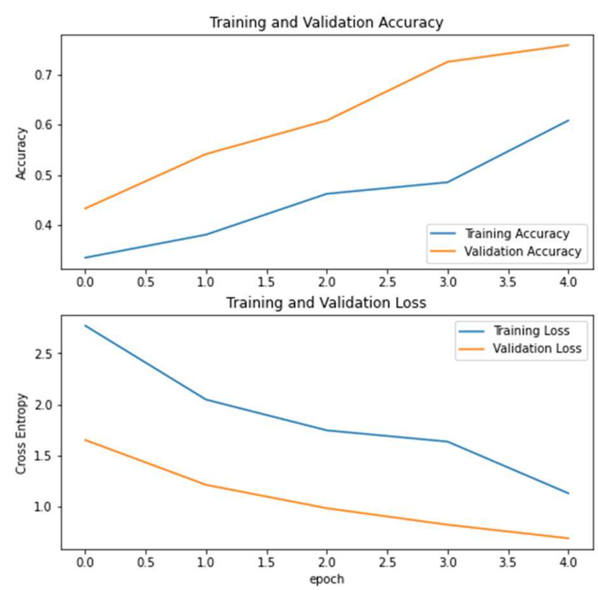

Fig 7. VGG-16 Base Training and Validation Accuracy and Loss

The results of base training the VGG-16 architecture is shown in figure 7. The training and validation accuracy and loss are shown on the plot. The VGG-16 architecture achieved 60% training accuracy, and 75.8% validation accuracy, according to the base training results. After training the base layer, the VGG-

16 network's final layer was fine-tuned using the same learning rate, optimizer, and activation function. An extra five epochs of training were performed on the VGG-16 final layer. The model was able to achieve a training accuracy rate of 66% and a validation accuracy rate of 78%, according to architecture's performance.

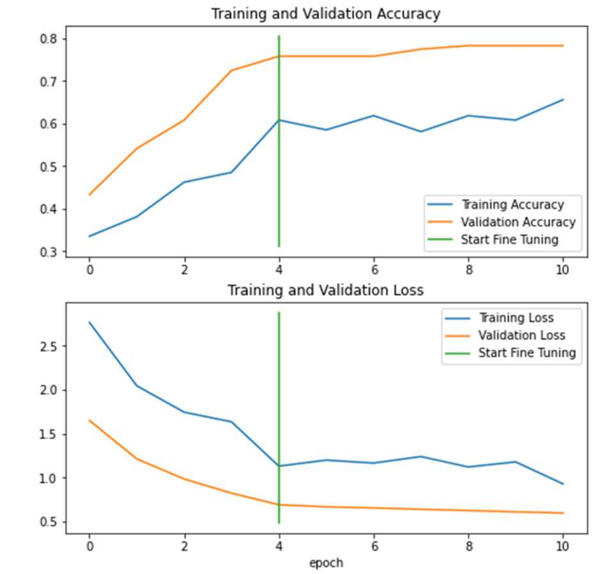

Fig 8. VGG-16 Training and Validation Accuracy and Loss after Fine-Tuning

Although the training used the pre-trained model of VGG-16, the performance of architecture when applied to the task of identifying medicinal mushrooms does not demonstrate good performance.

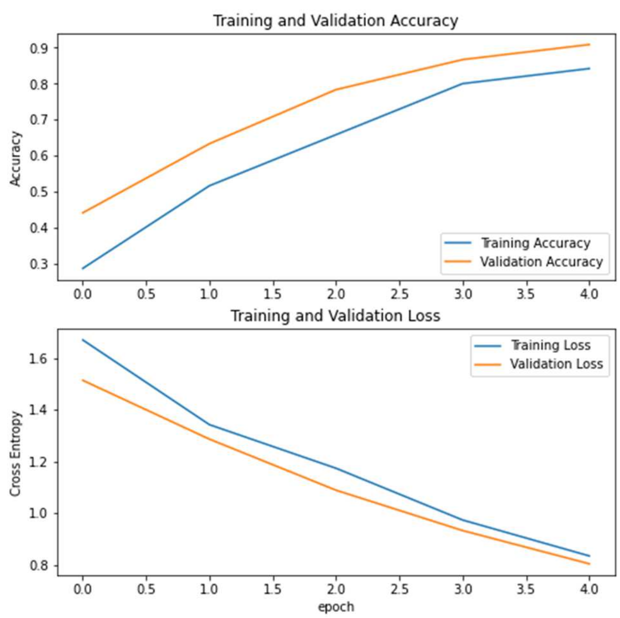

Fig 9. MobileNetv2 Base Training and Validation Accuracy and Loss

When it comes to using MobileNetv2 architecture, the based training demonstrated in fig. 9 shows that the base training achieved training accuracy rate of 84% and validation accuracy of 90%.

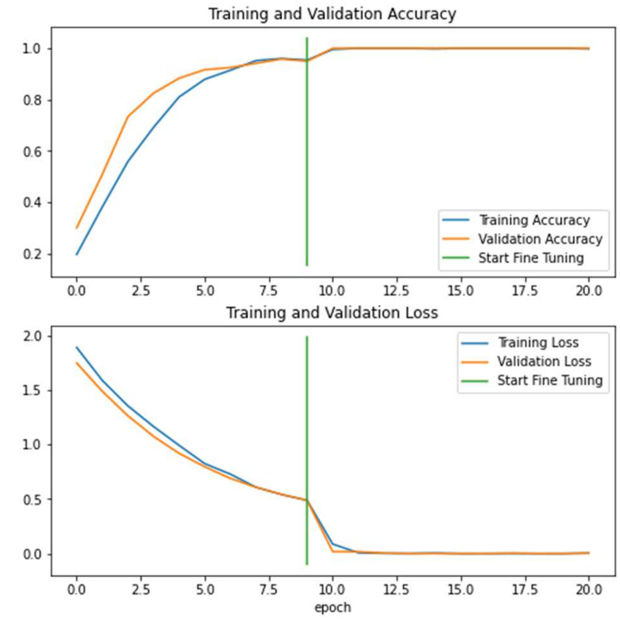

Fig 10. MobileNetv2 Training and Validation Accuracy and Loss after Fine-tuning

The final layer of MobileNetv2 architecture was also trained using fine tuning. After final training in another 5 epochs, both training and validation accuracy achieved a relatively good performance. Although both architectures were trained using the same dataset, MobileNetv2 has overpowered VGG-16 when it comes to identifying medicinal mushrooms.

2. *Model Testing*

To evaluate whether the training results are correct, developed models using VGG-16 and MobileNetv2 were tested using another set of datasets. For testing, the researchers gathered another set of datasets wherein in every class there are 25 testing images for a total of 150 testing samples. Both testing results are plotted in confusion matrices wherein using the data in generated confusion matrices, accuracy and recall of the two developed model was computed using the following formulas.

III. Results and Discussion

TABLE I
CONFUSION MATRIX FOR VGG-16 CNN ARCHITECTURE

		PREDICTED CLASS					
	n=150	***Lion's Mane***	***Oyster***	***Reishi***	***Shiitake***	***Shimeji***	***Volva***
ACTUAL CLASS	***Lion's Mane***	19	6	0	0	0	0
	Oyster	0	16	0	0	9	0
	Reishi	0	0	21	4	0	0
	Shiitake	4	0	0	18	0	3
	Shimeji	0	10	0	0	15	0
	Volva	0	0	0	5	0	20

Table 1 presents the results in testing 150 testing sample using the model obtained for VGG-16 CNN architecture. Upon testing, the results show that 109 testing samples were correctly classified by the developed model.

$$\text{Accuracy} = \frac{\textit{Sum of correct prediction}}{\textit{Total number of samples}} \qquad (1)$$

By utilizing the formula specified in the equation (1), the algorithm's efficacy can be assessed by calculating the proportion of accurate predictions to the total number of samples. This indicates that the model trained with VGG-16 achieved a testing accuracy rate of 72.6%. However, the low performance rate observed could be due to the limited amount of training data used.

This is not the case, on the other hand, with the performance of the model trained using MobileNetv2.

TABLE II
CONFUSION MATRIX FOR MOBILENETV2 CNN ARCHITECTURE

		PREDICTED CLASS					
ACTUAL CLASS	*n=150*	***Lion's Mane***	***Oyster***	***Reishi***	***Shiitake***	***Shimeji***	***Volva***
	Lion's Mane	25	0	0	0	0	0
	Oyster	0	24	0	0	1	0
	Reishi	0	0	24	1	0	0
	Shiitake	0	0	5	23	0	2
	Shimeji	0	3	0	0	25	0
	Volva	0	0	0	0	0	25

Table 2 displays the results obtained in training the developed model using MobileNetv2 architecture. The results show that out of 150 testing samples, 146 of them were correctly classified. This means that using this architecture yielded a testing accuracy rate of 97.3% in identifying and classifying medicinal mushrooms.

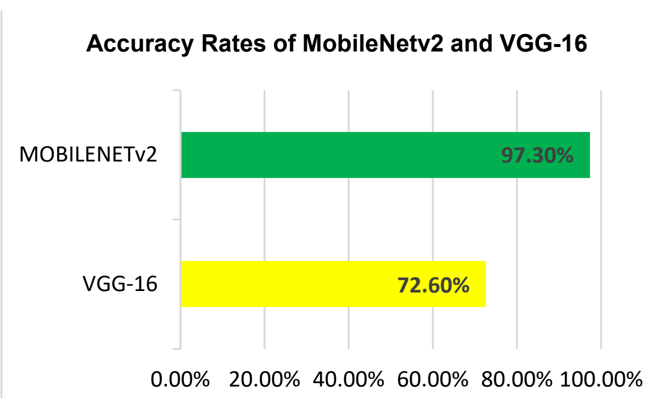

Fig 11. MobileNetv2 and VGG-16 Accuracy Rates

Fig. 11 shows the comparison of accuracy obtained from MobileNetv2 and VGG-16 based on testing. It is evident that MobileNetv2 outperformed VGG-16 in terms of solving the task in identifying six medicinal mushrooms by 24.7%. This difference in performance shows that using MobileNetv2 over VGG-16 will perform better in classifying medicinal mushrooms.

$$\text{Recall} = \frac{\textit{Correct Prediction for each class}}{\textit{Total number of sample per class}} \qquad (2)$$

TABLE III
RECALL FOR EACH CLASS BASED ON TABLE 1 (VGG-16)

Medicinal Mushrooms	Recall
Lions Mane	76%
Oyster	64%
Reishi	84%
Shiitake	72%
Shimeji	60%
Volva	80%

Table 3 presents the recall calculated from the results in table 1. Using the formula in (2), we can compute for the recall by getting the ratio of the accurate prediction for each class and the total quantity of samples per class. Based on the results of calculations, Lions Mane achieved a recall of 76%, 64% for oyster, 84% for Reishi, 72% for shiitake, 60% for Shimeji, and 80% for Volva.

TABLE IV
RECALL FOR EACH CLASS BASED ON TABLE 2 (MOBILENETV2)

Medicinal Mushrooms	Recall
Lions Mane	100%
Oyster	96%
Reishi	96%
Shiitake	92%
Shimeji	100%
Volva	100%

Table 4 presents the calculated recall based on the results shown in Table 2. Based on the results, Lions Mane, Shimeji and Volva yielded a recall of 100% which implies that all the testing samples for this class were correctly classified. Oyster Reishi both achieved 96% of recall and for Shiitake, 92% of recall was obtained.

IV. CONCLUSION AND FUTURE WORKS

Although both were trained using the same size of dataset, VGG-16 failed to reach optimal results while MobileNetv2 architecture masters the task even with a few amounts of data. With about 24.7% difference in testing performance, MobileNetv2 outperformed VGG-16 in solving the task of identifying medicinal mushrooms. Hence, based on these experiments conducted, the researchers believed that using MobileNetv2 over VGG-16 is more appropriate. Furthermore, for future studies, we aim to evaluate the performance of other convolutional neural network architecture when it is applied on the task of identifying medicinal mushroom. Experiments where classes will be added in the developed model is also one of the considerations of the researchers for future study to attain a robust model which can identify other types of mushrooms with continual learning.

ACKNOWLEDGMENT

This study owes its realization to the invaluable support and collaboration of numerous individuals. The researcher wishes to convey sincere gratitude to his technical adviser, Engr. Meo Vincent C. Caya, for unwavering support throughout the entire research journey. Additionally, profound appreciation is extended to Engr. Noel B. Linsangan, Engr. Cyrel O. Manlises, and Dr. Jocelyn F. Villaverde for their insightful contributions, which significantly enriched the refinement of this research.

REFERENCES

[1] Boddy, L., “Fungi, Ecosystem, and Global Change”, Retrieved October 17, 2021, from https://doi.org/10.1016/B978-0-12-382034-1.00011-6., 2021.

[2] Owaid, M. and Muslat M., “First Collection and Identification of Wild Mushrooms in Western Iraq”, University of Anbar, 2014.

[3] Anil et al., “Computer Vision-Based Method for Identification of freshness in Mushrooms”, International conference in Intelligent Computing”, International Conference in Intelligent Computing, 2019.

[4] Zehan et al., “A Deep Learning Approach for Edible, inedible and Poisonous mushroom Classification, 2021.

[5] Wibowo, A. et al., “Classification algorithm for Edible Mushrooms Identification”, 2018 International Conference on Information and Communication Technology (ICOIACT), 2018.

[6] Zhao et al., “Identification of Wild Mushrooms based on Ensemble Learning”, 4th International Conference on Big Data and Artificial Intelligence, 2019.

[7] Dong, J and Zheng, L., “Quality Classification of Enoki Mushroom Caps Based on CNN”, 2019 IEEE 4th International Conference on Image, Vision, and Computing, 2019.

[8] N.B. Linsangan and R.E. Angelia, Scopus: “Fermentation Level Classification of Cross Cut Cacao Beans Using K-NN Algorithm” 5th International Conference on Bioinformatics Research and Applications, 2018.

[9] N.B. Linsangan et al., Scopus: “Plant Identification by image processing of leaf veins”, ACM International Conference, 2017.

[10] D.S Bumacod et al., Scopus: “Image-Processing-Based Digital Goniometer using OpenCV”, 12th IEEE International Conference on Humanoid, 2020.

[11] N.B. Linsangan et al., Scopus: “Valgus and Varus Disease Detection using Image Processing”, 12th IEEE International Conference on Humanoid, 2020.

[12] C. Ishikawa, J. A. U. Marasigan and M. V. C. Caya, "Cloud-based Signature Validation Using CNN Inception-ResNet Architecture," 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Manila, Philippines, 2020.

[13] J. Hong and M. V. C. Caya, "Development of an Abaca Fiber Automated Grading System Using Computer Vision and Deep Convolutional Neural Network," TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON), Auckland, New Zealand, 2021

[14] S. Naik and B. Patel, “Machine Vision based Fruit Classification and Grading – A Review, International Journal of Computer Applications, 2017.

[15] Masoudian, A. and McIsaac, K, “Application of Support Vector Machine to Detect Microbial Spoilage of Mushrooms”, 2013 International Conference on Computer and Robot Vision, 2017.

[16] Brital, A., “Inception-V3 CNN Architecture explained”, Retrieved June 18, 2022 https://medium.com/@AnasBrital98/inception-v3-cnn-architecture-explained-691cfb7bba08, 2019.

[17] Narien, A. (n.d.), “Inception-V3 Model Architecture”, Retrieved June 18, 2022 from https://iq.opengenus.org/inception-v3-model-architecture/

[18] Sgedy, C et al., “Rethinking the Inception Architecture for Computer Vision”, CVPR, 2022.

[19] Frasca, F. et al., “SIGN: Scalable Inception Graph Neural network”, 2020.

[20] Ma. C., et al., “TS-LSTM and Temporal-Inception Exploiting Spatiotemporal Dyamics for Activity Recognition”, 2017.

[21] Deliwala, S. “Streamlit: An in-depth Introduction”, Retrieved June 18, 2022 from https://towardsdatascience.com/streamlit-101-an-in-depth-ad9492f2, 2019.