Mathematical Sciences and Informatics Journal
Vol. 2, No. 2, Nov. 2021, pp. 35-43
http://www.mijuitmjournal.com

New Species Orchid Recognition System Using Convolutional Neural Network

Annisa Atikah binti Mohd Fadzil
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Perak, Kampus Tapah, Perak, Malaysia
2017412136@uitm.edu.my

Itaza Afiani binti Mohtar
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Perak, Kampus Tapah, Perak, Malaysia
itaza328@uitm.edu.my

Article Info

Article history:

Received May 15, 2021
Revised Aug. 20, 2021
Accepted Sept 18, 2021

Keywords:

Orchid Recognition
Artificial Neural Network
Convolutional Neural Network
Transfer Learning

ABSTRACT

Orchid is famous for its variety and its beauty. Every year, about a hundred of new species names are published. The study is to determine whether the orchid species can be recognized using a convolutional neural network algorithm and to test the accuracy of the classification model. Transfer learning was also implemented in this project to skip the feature extraction phase that requires many computational resources in the CNN algorithm. The model of transfer learning that is used is the Inception V3 model. This project is to prove that the concept of new orchid species recognition can be done. The web application that created using HTML and Flask was able to recognize new species based on existing species. In this project, 10 existing species with 100 images each was selected in training, validating, and testing phase. The training accuracy reached 97% and the functional testing of orchid recognition results shows 83% accuracy with 1000 datasets. In conclusion, the use of a web system as a prototype tool for the recognition of new orchid species is helpful for the unlicensed persons/organization.

Corresponding Author:
Itaza Afiani Mohtar
Faculty of Computer and Mathematical Sciences,Universiti Teknologi MARA, Cawangan Perak, Kampus Tapah, Perak, Malaysia.
email: itaza328@uitm.edu.my

1. Introduction

Orchids are usually found in the forest by an orchid specialist collector. Normally, the collector identifies the new orchid by taking the specimen to the lab to be further examined by the botanist [1]. In reality, an orchid species cannot simply be collected because it is protected under Biodiversity Conservation Act 2016. Orchid is categorized as protected plant species that require a license for it to be picked or grown for sale as listed in Part 2 of Schedule 6 of the BC Act [2]. Previously, the process of recognizing orchid specimens was done by searching for their species manually in the book. As a result, it will take some time since there are thousands of species orchids ever recorded. If the species is not found through a thorough search, then it proves that the specimen is a new species. This is such a waste of time to go through page by page and manually match the observed species with the one recorded in the book, and surely, this was not an easy task. Therefore, this project aims to allow unlicensed organizations to recognize the species without moving the specimen and using the system to replace manual recognition. The system will be working simply by inserting the image of the specimen as an input to the system and recognition will take part and the name of the species will be displayed as the output. In addition, the existence of the system can help to minimize the time of recognition of the new species in the blink of an eye.

2. Literature Review

Orchids become one of the biggest families in angiosperm besides Asteraceae. The most recent number of orchid species accepted until 1 January 2017 is 28,284 orchid species with a hundred new species names published each year [3]. There will be about 31,000 species of orchids [4]. It has recognized up to 1000 genera and described new orchid genera at a rate increase of about 13 species per year [5]. The structure of the orchid is divided into stem and roots, leaves, flowers, and pollination. Orchidaceae are famous for their numerous structural floral variations. Several orchids have single flowers, but most have a racemose inflorescence, with large numbers of flowers sometimes. Botanical nomenclature is a system for formally classifying and naming organisms according to their genus and species. The starting point of the botanical nomenclature was when Species Plantarum (known as Species of Plant) published in 1753 [6]. If botanical nomenclature is used for naming a plant, taxonomy is more to grouping and classifying the plant. The emerging world of the internet nowadays affects the naming of the wild plant as numerous changes were made at a conference held in Melbourne, Australia in 2011. The naming of wild plants is governed by the International Botanical Congress. Royal Horticultural Society (RHS) in the United Kingdom jobs is to keep track of over 150,000 orchid hybrid names that exist [7].

Image classification is defined as a process in computer vision that classifies the image according to its visual input. The process of image classification is to classify the image based on the group/type provided. The classification of an image is determined by labeling the pixels based on their grey value and the use of multiple features for a set of pixels. Computer classification of remotely sensed images involves the process of the computer program 'learning' the relationship between the data and the information classes.

Supervised learning is defined as a learning process designed to map one variable (data) to another (information class) series of variables. Supervised learning is for training with well-labeled data. Among supervised classification are Decision Tree, Artificial Neural Network, and Convolutional Neural Network. Convolutional Neural Network or known as CNN is part of the deep learning model. It is one of the famous approaches to image classification. A CNN employs a mathematical operation called convolution. Convolution is a linear operation of a specific form. Convolutional networks are neural networks that use convolution in at least one of their layers instead of general matrix multiplication [8].

3. Methodology

The process to achieve a positive outcome of this project requires a specific planning phase. Figure 1 presents a research model for this project which applies the Waterfall model. The model includes five processes: analysis, design, implementation, and testing. The phases selected for this project are only from the Analysis to the Testing phase based on the waterfall model. The Maintenance phase is not chosen because the focus of this project is on development. Figure 2 illustrates the system architecture of the project. In this project, the image of orchid is uploaded into the model and processed by the system classifier. The system classifier then will identify the image and if the system classifier recognized an unfamiliar image, it will declare a new species. Figure 3 shows how to build the system classifier.

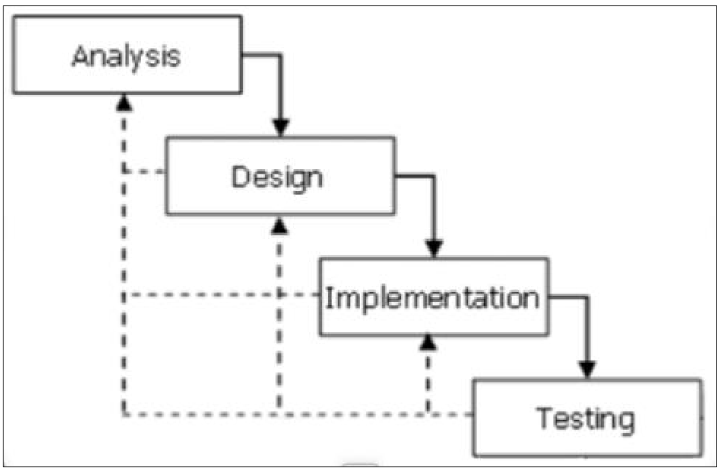

Figure 1. Research model of the project

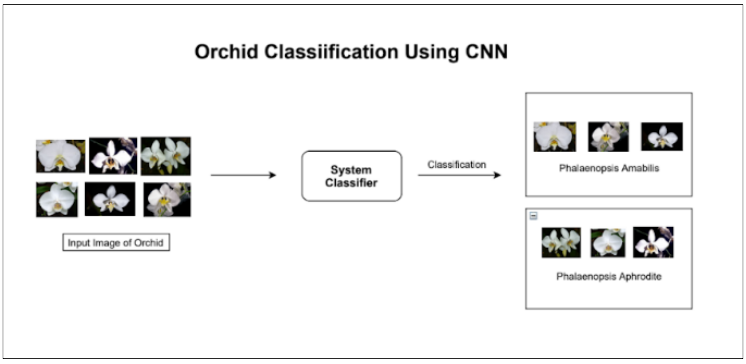

Figure 2. System architecture of the project

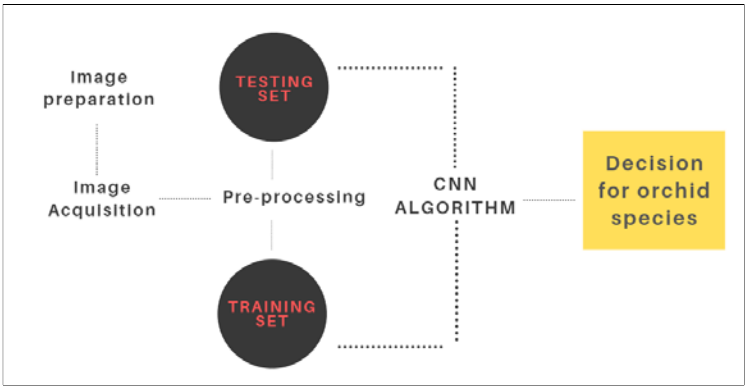

Figure 3. The classifier model

3.1. System Development

3.1.1 Classifier

The base neural network model that will be used in this model is Inception-V3. This project has a smaller dataset to be undergo the Inception-v3 model thus, a transfer learning method that utilized pre-trained Inception-v3 model will be used [9]. Figure 4 shows the Inception-V3 Architecture model. By using transfer learning, the feature extraction part will be using the pre-trained Inception-v3 and the process of training dataset will start at the final layer of Inception-v3 which is the classification layer, the computational resources and training time for small dataset. Transfer learning able to retain the information that the model had gained during its initial training and apply it to the smaller dataset, resulting in highly accurate classifications without the need for intensive training and computing resources [10]. The implementation of transfer learning will be discussed in the Implementation section (pseudocode). The next section will discuss pre-processing stages.

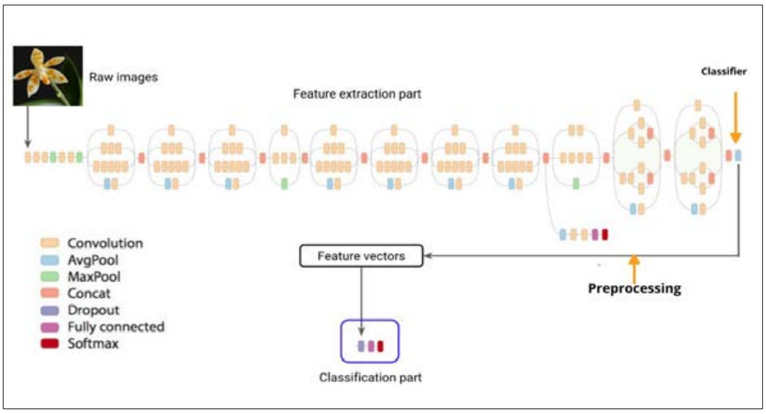

Figure 4. The Inception V3 model

3.2 Pre-processed Image

The aim of pre-processing stage is to produce variety of images that will help to improve the results of the training stage. All the images will be rotated, cropped, flipped to the left and right.

3.2.1 Implementation

The figure 5 shows lines of code implementation of Inception V3 layers and compiling the model.

Figure 5. Code segments to implement Inception V3

Figure 6 shows the lines of code for training model and Figure 7 shows the running process of training model.

Figure 6. Code segment for the training model

```
Epoch 9/30
23/23 [==============================] - 807s 35s/step - loss: 0.5160 - accuracy: 0.8546 - val_loss: 1.0102 - val_accuracy: 0.6
641
Epoch 10/30
23/23 [==============================] - 804s 35s/step - loss: 0.4488 - accuracy: 0.8767 - val_loss: 0.8893 - val_accuracy: 0.7
500
Epoch 11/30
23/23 [==============================] - 813s 35s/step - loss: 0.4239 - accuracy: 0.8764 - val_loss: 0.8993 - val_accuracy: 0.6
719
Epoch 12/30
23/23 [==============================] - 807s 35s/step - loss: 0.3256 - accuracy: 0.9205 - val_loss: 0.6657 - val_accuracy: 0.7
891
Epoch 13/30
23/23 [==============================] - 890s 39s/step - loss: 0.3494 - accuracy: 0.9072 - val_loss: 0.6564 - val_accuracy: 0.8
281
Epoch 14/30
23/23 [==============================] - 953s 41s/step - loss: 0.3336 - accuracy: 0.9018 - val_loss: 0.6972 - val_accuracy: 0.8
047
Epoch 15/30
23/23 [==============================] - 862s 37s/step - loss: 0.2659 - accuracy: 0.9372 - val_loss: 0.5998 - val_accuracy: 0.7
969
Epoch 16/30
23/23 [==============================] - 890s 39s/step - loss: 0.3100 - accuracy: 0.9103 - val_loss: 0.7851 - val_accuracy: 0.8
047
```

Figure 7. The training process

Figure 8 shows the lines of codes for test the functionality of the model which is species prediction. The output is to display the name of species and the confident percentage of the species. The 'conf' is a variable for confident percentage. If the confident of the predicted species is more than 50%, the species is a known orchid species, other than that, the species is categorized as a new orchid species.

Figure 8. Code segments for testing the model

3.3 Interface

The website powered by HTML and Flask is used to implement the model build. As shown in Figure 9, by clicking "Choose file," the user must upload the picture. The image will appear at the view picture box after that. After clicking on the "submit" button, the image will be sent and listed into the classifier. The output will be displayed according to the name of the species with a confident value of how accurate the species is along with origin and the further information of the orchid.

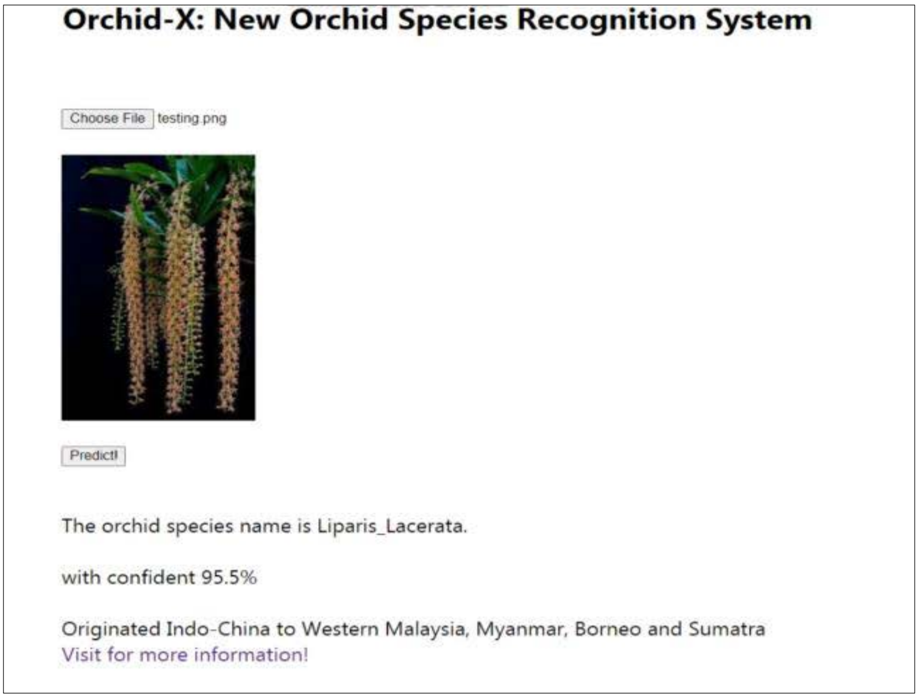

Figure 9. The system interface when a species is identified

Figure 10 shows the user interface if the user upload unknown species to the system. The output message “This is a new species!” will be displayed.

Figure 10. The system interface when a new species is identified

4. Testing and Result Analysis

There are 10 classes used in implementing classification of orchid. The model was tested with 100 images for each class that will sum up to 1000 images in the dataset. The images are randomly selected and not trained.

4.1 Result

Figure 11 shows that the training and validation accuracy after training all the dataset with 30 epochs. The vertical axis represents the accuracy value and the horizontal axis represents the number of epochs. The blue line represents the training accuracy while the orange line represents the validation accuracy. The final accuracy for training set is 0.9768 equivalent to 97 percent while validation set that predicts unknown images (validation) is 0.8438 equivalent to 84 percent.

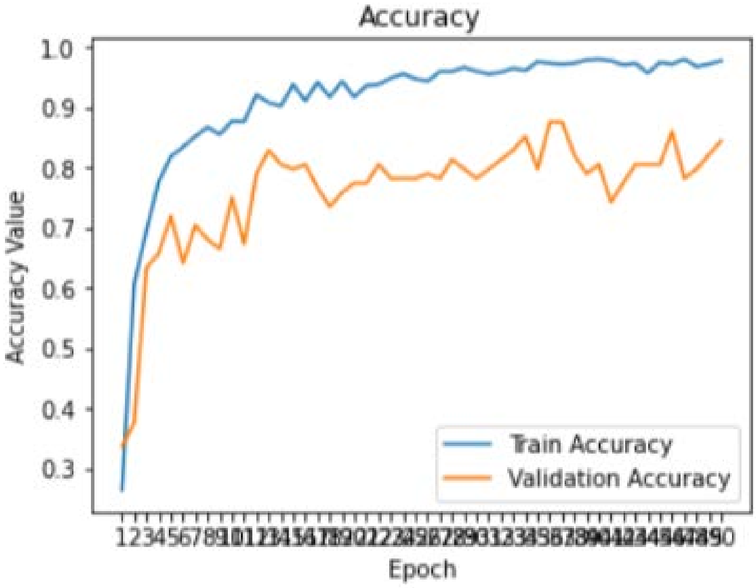

Figure 11: Training and Validation Accuracy

Figure 12 shows the performance of loss value for training and validation. The vertical axis represents the loss value, and the horizontal axis represents the number of epochs. The blue line represents the train loss while the orange line represents the validation loss. The final loss for training set is 0.0717 equivalent to 0.07 percent while validation set that predicts unknown images (validation loss) is 0.6129 equivalent to 61 percent.

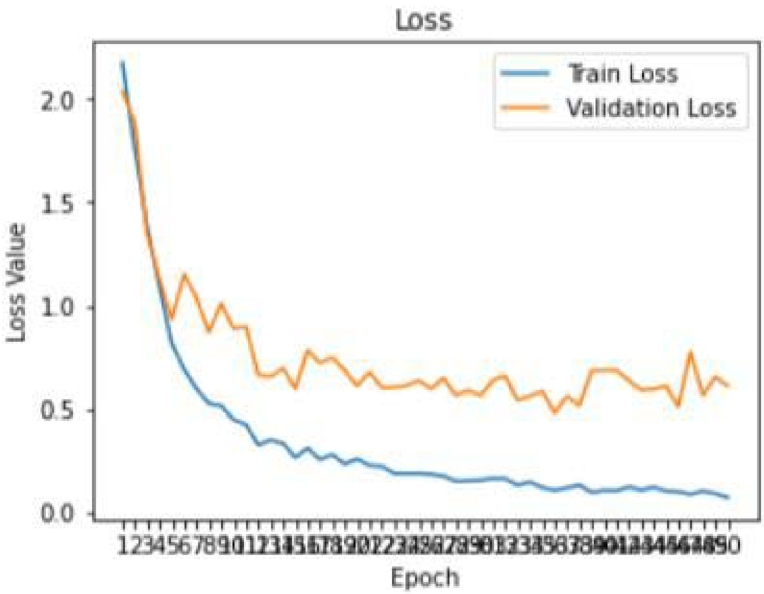

Figure 12. Training and Validation Loss

4.2 Analysis of Functional Testing Performance

The performance is calculated based on accuracy, precision, recall and F1 score metrics. This can be achieved by calculating over True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), which can be defined as:

- (TN) True Negative: The actual value was False, and the model predicted False.
- (FP) False Positive: The actual value was False, and the model predicted True.
- (FN) False Negative: The actual value was True, and the model predicted False.
- (TP) True Positive: The actual value was True, and the model predicted True.

Following are the formula to calculate the performance metrics.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

$$\text{F1 Score} = \left(\frac{2}{2\text{recall}-1+\text{ precision}-1}\right)=2\cdot\frac{\text{precision}\cdot\text{recalll}}{\text{precision+recal}} \quad (3)$$

$$\text{Accuracy} = TP+TN\ TP+TN+FP+FN \quad (4)$$

Table 1 shows the result of the experimental test for accuracy, precision, recall and F1 score from 100 testing images.

Table 1. Experimental Results

Performance Metrics	**Value**
Accuracy	0.83
Precision	0.8481
Recall	0.83
F1 Score	0.8251

Table 2 shows the classification report for the testing of 100 images from 10 class. The class name is the names of orchid species that were trained in the model. The total number of images is number of random images for testing the model. The precision is the calculation of how many correctness for each class.

Table 2. Classification report for orchid species tested.

Class Name	**Number of Images**	**Precision**
Bulbophyllum Purpurascens	10	75%
Liparis Gibbosa	10	100%
Liparis Lacerata	10	83%
Neuwiedia Veratrifolia	10	91%
Paphiopedilum Bullenianum	10	82%
Paphiopedilum Callosum	10	58%
Phalaenopsis Amabilis	10	82%
Phalaenopsis Maculata	10	77%
Renanthera Matutina	10	100%
Taeniophyllum pusillum	10	100%

The results in Table 2 indicate that CNN algorithm that used in the web system is effective to recognize new species of orchids with precision above 70% for all orchid classes.

5. Conclusion

Based on the findings of the study, the use of web system as tool for recognition new orchid species is helpful for the unlicensed organization. In this work, classification of orchid based on its images by using Convolutional Neural Network has accomplished. In this project, the CNN algorithm has been applied to recognize the orchid species. The functional testing show that a satisfactory result has been obtained. Furthermore, the objectives of this project have been achieved, which is to develop a concept proof to recognize the new orchid species using CNN. Overall, this model is effective to recognize the new orchid species.

Acknowledgements

The authors gratefully acknowledge the Universiti Teknologi MARA (UiTM), Perak branch for giving the authors the opportunity, support, and facilities to carry out this project.

References

[1] A. Segovia-Rivas, J. A. Meave, E. J. González, and E. A. Pérez-Garc'ia, "Experimental reintroduction and host preference of the microendemic and endangered orchid Barkeria whartoniana in a Mexican Tropical Dry Forest," *J. Nat. Conserv.*, vol. 43, pp. 156–164, 2018.

[2] C.L Brown, "Can legislation deliver conservation? An assessment of the Threatened Species Conservation Act 1995 (NSW) using two threatened plant species as case studies." PhD Thesis, Wollongong University, 2002.

[3] M. M. Hossain , R. Kant , P. T. Van , B. Winarto , S. Zeng , J. A. Teixeira da Silva "The application of biotechnology to orchids" *Critical Reviews in Plant Sciences*, *32*(2), pp 69-139. 2013.

[4] Hinsley A, de Boer HJ, Fay MF, Gale SW, Gardiner LM, Gunasekara RS, Kumar P, Masters S, Metusala D, Roberts DL, Veldman S, Wong S, Phelps J. "A review of the trade in orchids, and its implications for conservation." Bot J Linn Soc 186, pp 435–455. 2018 .

[5] Mark W. Chase, Kenneth M. Cameron, John V. Freudenstein, Alec M. Pridgeon, Gerardo Salazar, Cássio van den Berg, André Schuiteman, An updated . classification of Orchidaceae, Botanical Journal of the Linnean Society, vol 177, Issue 2, pp 151–174, 2015 https://doi.org/10.1111/boj.12234.

[6] Britannica, T. Editors of Encyclopaedia, "Species Plantarum. Work by Linnaues", *Encyclopedia Britannica*, 8^{th} ed, Oxford UP, 2009.

[7] D. Inayah, T. Artika and E. Arisoesilangsih. "Mapping diverseterestrial orchidhs grown in the orchids garden of the ranu darungan resort, the bromo tengger semeru national park" IOP Conference Series: Earth and Environmental Science, vol 743, The 11^{th} International Conference on Global Resource Conservation 28-29 July 2020, East Java, Indonesia. 2020

[8] I. Goodfellow, Y. Bengio and A. Courville. "Deep Learning." MIT Press, 2016.

[9] R., Poojary, and A. Pai "Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models." In *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)* pp. 1-4. 2019.

[10] N., Mpofu and M., Sears, "Binary View Classification of Echocardiograms of the Heart Using Transfer Learning." In *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)* pp. 46-52, 2020.