

MYLPHerb-1: A Dataset of Malaysian Local Perennial Herbs for the Study of Plant Images Classification under Uncontrolled Environment

Kalananthni Pushpanathan¹, Marsyita Hanafi^{1*}, Syamsiah Masohor¹ and Wan Fazilah Fazlil Ilahi²

¹Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan, Malaysia

²Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan, Malaysia

ABSTRACT

Research in the medicinal plants' recognition field has received great attention due to the need of producing a reliable and accurate system that can recognise medicinal plants under various imaging conditions. Nevertheless, the standard medicinal plant datasets publicly available for research are very limited. This paper proposes a dataset consisting of 34200 images of twelve different high medicinal value local perennial herbs in Malaysia. The images were captured under various imaging conditions, such as different scales, illuminations, and angles. It will enable larger interclass and intraclass variability, creating abundant opportunities for new findings in leaf classification. The complexity of the dataset is investigated through automatic classification using several high-performance deep learning algorithms. The experiment results showed that the dataset creates more opportunities for advanced classification research due to the complexity of the images. The dataset can be accessed through <https://www.mylpherbs.com/>.

Keywords: Deep learning, leaf identification, medicinal plants, perennial herbs, plant dataset

ARTICLE INFO

Article history:

Received: 24 May 2021

Accepted: 17 August 2021

Published: 04 January 2022

DOI: <https://doi.org/10.47836/pjst.30.1.23>

E-mail addresses:

kalananthni19@gmail.com (Kalananthni Pushpanathan)

marsyita@upm.edu.my (Marsyita Hanafi)

syamsiah@upm.edu.my (Syamsiah Mashohor)

wanfazilah@upm.edu.my (Wan Fazilah Fazlil Ilahi)

* Corresponding author

INTRODUCTION

Medicinal plants have been widely used as an alternative to modern medicine due to their lower negative impact on the human body and affordable price (Lulekal et al., 2008). Asia is a region with enormous biodiversity of medicinal plants; hence, expert knowledge is required for plant

identification, which is tedious and time-consuming. Automatic classification is the best solution for the problem mentioned above, and it has been researched extensively. Nevertheless, to achieve high accuracy, identification systems need to be trained with medicinal plant image datasets that reflect real-life scenarios (Sladojevic et al., 2016; Singh & Misra, 2017; Wäldchen et al., 2018). Accessibility of the standard medicinal plant image datasets is limited despite being researched actively. Most available plant datasets consist of images captured under a controlled environment. The datasets that have been widely used are Swedish Leaf Dataset, Flavia Dataset, Leafsnap dataset and ICL dataset. The examples of studies that used these datasets for automatic identification are Arun et al. (2013), Begue et al. (2017), Dahigaonkar and Kalyane (2018), Harsani and Qurania (2016), Janani and Gopal (2013), Pornpanomchai et al. (2011), Vijayashree and Gopal (2017), Wäldchen and Mäder (2018a). The standard datasets for the studies in the uncontrolled environment are very limited, and until now, the only reliable option is the ImageCLEF datasets. Although great variations of leaf images are available in ImageCLEF, most of the images were captured from wild plant species in Western Europe Flora, and North American. The datasets mentioned above are summarised in Table 1.

Table 1
Summary of publicly available datasets

No	Dataset	No. of images	No. of Plant Species	Website
1	Swedish Leaf Dataset	1125	15	https://www.cvl.isy.liu.se/en/research/datasets/swedish-leaf/
2	Flavia Dataset	2621	33	http://flavia.sourceforge.net/
3	Leafsnap dataset	30866	185	http://leafsnap.com/dataset/
4	ImageCLEF	5436	70	https://www.imageclef.org
5	PlantCLEF 2014	60000	500	https://www.imageclef.org/node/179
6	PlantCLEF 2015	113205	1000	https://www.imageclef.org/lifeclef/2015/
7	PlantCLEF 2016	113205	1000	https://www.imageclef.org/node/198
8	PlantCLEF 2017	1100000	10000	https://www.imageclef.org/node/218

METHOD

Plant Organ Selection

Multiple plant characteristics can be used for analysis in classification, namely flowers, stems, leaves, fruits and even the whole plant. Multiple features can be extracted in each characteristic, such as the leaf shape, vein arrangement, texture, colour, and margins from leaves. Leaves are often favoured among researchers for plant identification studies as they do not undergo major changes throughout the year and exist in abundance for each plant (Wäldchen & Mäder, 2018b). Furthermore, there are more than 40 leaf shapes, and at least 12 types of leaf margins and nine types of venation patterns that can be extracted from leaves (OSU, 2021). Among all the features, leaf shape is the most commonly used

as it is considered the most discriminative feature of a plant. Margins are also useful for plant identification as multiple details can be extracted from the margins, namely the tooth spacing, number of teeth per centimetre and shape of the tooth themselves. The leaf venation patterns are useful as the contrast of veins compared to the rest of the leaves makes them reliable as a feature to be considered for automated plant classification. Veins are vital for the survival of a plant as it is embedded with xylem and phloem cells (Sack & Scoffoni, 2013). Factors such as environmental conditions, biomechanical support and even leaf size affect the vein structure, making vein patterns. Meanwhile, the texture of the leaves is the smoothness or roughness on the surface of the leaves and is often analysed in terms of group pixels as it must visually interpret the roughness or smoothness of the leaves (Wäldchen & Mäder, 2018b). The colours of the leaves are also vital features that can be used in plant identification. Different shades of green, edge/margin colours and hues can all be utilised to differentiate the plant species.

Asian Local Perennial Herbs

Asian local perennial herbs grow for more than two years compared to biennial plants and annual plants native to the Asian region. These plants mostly require low maintenance, which makes them very favourable. Some of these plants are often used for cooking or decorative purposes. The selected herbs for this dataset are *Andrographis paniculata* (Green chiretta), *Plectranthus amboinicus* (Mexican mint), *Plectranthus amboinicus* 'Variegatus' (Variegated Mexican mint), *Persicaria minor* (Small water-pepper), *Centella asiatica* (Asiatic pennywort), *Gynura pseudochina* (Chinese gynura), *Gynura procumbens* (Longevity spinach), *Orthosiphon aristatus* (Java tea), *Hydrocotyle javanica* (Java pennywort), *Ocimum tenuiflorum* (Holy basil), *Clinacanthus nutans* (Snake grass) and *Ocimum citriodorum* (Lemon basil) as shown in Figure 1. These herbs need to be brought to the attention of the public for their high medicinal value, which has been described in Table 2.

Data Collection

The introduced MYLPHerbs-1 dataset consists of 34200 images under various imaging conditions. The images were captured at various distances from the camera, illuminations, rotations and scales that reflect real-life scenarios. The acquisition settings for data collection are summarised in Table 3. The settings are based on the situations when capturing the images in real-life scenarios, in which the complexity of the images includes the variation of interclass and intraclass variability that will create more opportunities for developing more advanced machine learning classifiers. The interclass variability is obtained through various plant species, while the intraclass variability is achieved through the various acquisition settings applied to the same plant. Other consideration includes



Figure 1. Plants used for the proposed dataset (a) Green chiretta (b) Mexican mint (c) Variegated Mexican mint (d) Small water-pepper (e) Asiatic pennywort (f) Chinese gynura (g) Longevity spinach (h) Java tea (i) Java pennywort (j) Holy basil (k) Snake grass (l) Lemon basil

Table 2
The medicinal properties of the selected herbs

Plants Species	Medicinal Properties	References
<i>Andrographis paniculata</i> (Green chiretta)	Anti-microbial, anti-inflammatory, antioxidant, anti-diabetic and hepatorenal protective.	Okhwarobo et al. (2014); Kurzawa et al. (2015); Murugan et al. (2020).
<i>Plectranthus amboinicus</i> (Mexican mint)	Antimicrobial, anti-inflammatory, antitumor, antioxidant and analgesic activities.	Bhatt et al. (2013); Swamy et al. (2017); Ashaari et al. (2020)
<i>Plectranthus amboinicus</i> 'Variegatus' (Variegated Mexican mint)	Antimicrobial, anti-inflammatory, antitumor, antioxidant and analgesic activities.	Bhatt et al. (2013); Swamy et al. (2017)
<i>Persicaria minor</i> (Small water-pepper)	Antioxidant, antibacterial, anti-inflammatory, and anticancer.	Vimala et al. (2012); Christopher et al. (2015); Lau et al. (2020)
<i>Centella asiatica</i> (Asiatic pennywort)	Wound healing, effective for leprosy, lupus, eczema psoriasis, relieving anxiety and improving cognition.	Gohil et al. (2010); Deshpande (2017); Giribabu et al. (2020)
<i>Gynura pseudochina</i> (Chinese gynura)	Haemostatic, antipyretic, regulate menses, to treat breast tumours, herpes infections and sore throats.	Siriwatanametanon et al. (2010); Siriwatanametanon & Heinrich (2011); Proklamasiningsih et al. (2020)

Table 2 (continue)

Plants Species	Medicinal Properties	References
<i>Gynura procumbens</i> (Longevity spinach)	Antihypertensive, antihyperglycemic, anticancer, antimicrobial, antioxidant, and anti-inflammatory activity.	Rahman and Asad (2013); Tan et al. (2016); Ashraf et al. (2020)
<i>Orthosiphon aristatus</i> (Java tea)	Antimicrobial, antioxidant, hepatoprotection, cytotoxic, cardioactive, antidiabetic and anti-inflammatory activities	Abdelwahab et al. (2010); Ashraf et al. (2018); Samidurai et al. (2020)
<i>Hydrocotyle javanica</i> (Java pennywort)	Healing common cold, tonsillitis, cephalitis, enteritis, dysentery, zoster, eczema, period pain, hepatitis, and jaundice.	Mandal et al. (2017); Mandal et al. (2016); Karthika (2020).
<i>Ocimum tenuiflorum</i> (Holy basil)	Analgesics, inflammation relievers, anti-asthmatic, hypoglycemic, hepatoprotective, hypotensive and hypolipidemic	Yamani et al. (2016); Singh and Chaudhuri (2018); Sahu et al. (2020)
<i>Clinacanthus nutans</i> (Snake grass)	Relieving skin rashes, scorpion and insect bites, diabetes mellitus, fever, and diuretics	Alam et al. (2016); Khoo et al. (2018); Haida et al. (2020).
<i>Ocimum citriodorum</i> (Lemon basil)	Antimicrobial, antioxidant, hepatoprotection, antigenotoxic, antiplasmodial, cytotoxic, cardioactive, antidiabetic, anti-inflammatory activities	dos Santos et al. (2016); Majdi et al. (2020)

the utilisation of different camera models. The images were captured using three Android phones, namely Samsung A7, Samsung J1 and Samsung J5, with a camera resolution of 24 Megapixels, 9 Megapixels and 5 Megapixels, respectively. The number of images captured by the Samsung J1 phone is 10920 images, which is 910 images for each herb. Nevertheless, a larger number of images were captured by Samsung J5 and Samsung A7, 970 images for each herb, thus providing 11640 images for each phone. The difference is due to the unavailability of the ISO 800 feature in Samsung J1.

Thus, the total number of considered features in the data collection are 91 features for Samsung J1 and 97 features for Samsung J5 and Samsung A7. Each plant has leaves with different details/features that can be learnt. Different species can be differentiated with a combination of features, namely shape, vein, colour, margin, and texture. The visual characteristics of the leaves are explained in Table 4. The characteristics of the leaves in Table 4 are just a simple explanation of what human eyes could decipher from several examples of each plant. With thousands of examples for each plant, the pre-trained deep learning algorithm would gather information from those examples at a high rate with higher sensitivity towards the features and learn to identify those plants during the testing phase. The example of images in the proposed dataset is illustrated in Figure 2. Figure 3 shows the main page of the proposed dataset that can be accessed through <https://www.mylpherbs.com/>.

Table 3
Summary of the image acquisition settings and their description

Settings	Description
Various distances ranging from 5 cm till 30 cm	The distances of the camera facing the leaf are 5 cm, 10 cm, 15 cm, 20 cm, 25 cm and 30 cm.
Distance with a 45° angle at 10 cm, 20 cm and 30 cm	The camera is tilted at 45° relative to the surface of the leaf, where the images were captured at 10 cm, 20 cm and 30 cm, on both rights and left of the leaf.
Optical Zoom: 1×, 1.5×, 2×, 2.5×, 3×, 3.5× and 4×	Besides various optical zooms, the distance between the camera and leaf is 30 cm.
White balance parameters: Sun, Cloudy, Incandescent and Fluorescent:	The process of removing unrealistic colour cast, thereby improving the captured images under a wide range of lighting conditions. Other considerations include the camera facing the surface of the leaf at 10 cm, 20 cm and 30 cm, and the camera tilted at 45° relative to the right-side surface of the leaf.
ISO: 100, 200, 400 and 800	The sensitivity of the camera to light. Other considerations include the camera facing the surface of the leaf at 10 cm, 20 cm and 30 cm and tilted 45° angle relative to the right-side surface of the leaf.
Exposure: -2, -1, 0, +1, +2	The amount of light that reaches the camera sensor. Other considerations include the camera facing the surface of the leaf at 10 cm, 20 cm and 30 cm and tilted at 45° relative to the right-side surface of the leaf.

Table 4
Leaf characteristics of each plant

Plant species	Leaf Characteristics
Green chiretta	Lanceolate-shaped with arcuate veins
Mexican mint	Ovate-shaped and thick with a cross-venulate type of vein and serrated edges.
Variegated Mexican mint	Similar to Mexican mint with a white edge and thinner.
Small water-pepper	Linear-lanceolate shape with pinnate venation.
Asiatic pennywort	Can range from orbicular to reniform shape
Chinese gynura	The colour is purplish with green veins. Pinnate-shaped veins changes into dichotomous shapes as it reaches the edge of the leaf blade.
Longevity spinach	Ovate-shaped and has a smooth green surface. The venation of the leaves is pinnate-shaped, where the secondary veins are paired in the opposite direction extending from the primary midvein.
Java tea	Small and rhomboid-shaped with serrated edges. The veins are pinnate shaped and have a smooth surface.
Java pennywort	Reniform-shaped with rounded margin.
Holy basil	Ovate-shaped at the base and morphs into a rhomboid shape at its tip. The veins are pinnate shaped with serrated edges.
Snake grass	Lanceolate-shaped leaf with a smooth margin. Smooth texture with arcuate veins.
Lemon basil	Smooth and ovate-shaped with pinnate venation.

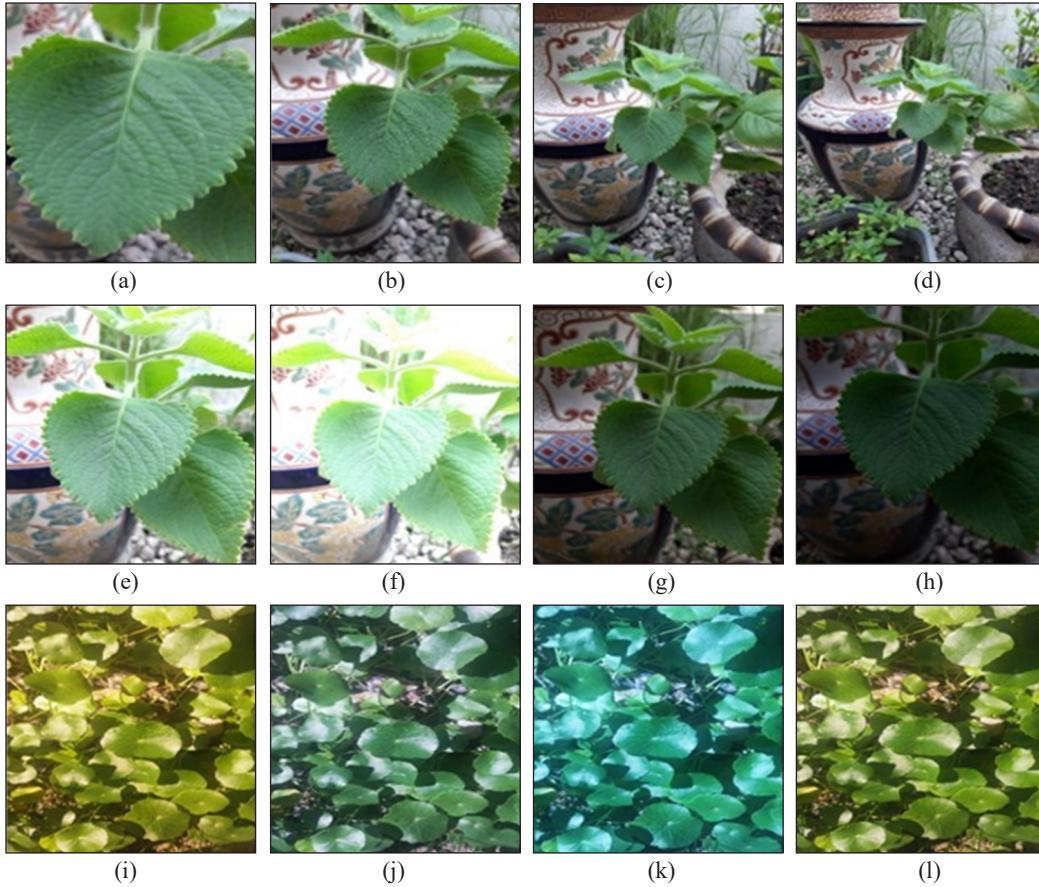


Figure 2. Examples of Mexican mint images that were captured with Samsung A7 at a distance of (a) 5 cm (b) 10 cm (c) 20 cm and (d) 30 cm; Examples of Mexican mints that were captured with Samsung A7 at 10 cm and various exposures namely (e) +1 (f) +2 (g) -1 (h) -2; Examples of Asiatic pennywort captured with Samsung J5 at a distance of 20 cm and various white balance namely (i) cloudy (j) fluorescent (k) incandescent (l) sun

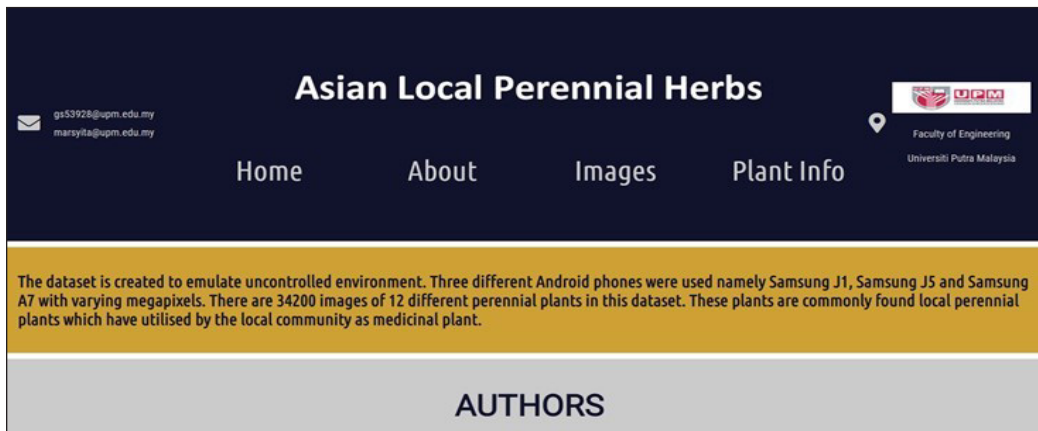


Figure 3. Screenshot of the webpage of the proposed dataset

Experimental Setup

In this paper, the MYLPHerbs-1 dataset is evaluated using five pre-trained deep learning algorithms, namely VGG16, VGG19, ResNet50, EfficientNet B0 and EfficientNet B7. The findings can be used to create the baseline of a more advanced and reliable deep learning architecture. VGG16 and VGG19 algorithms were chosen for their good performance when classifying leaves or plants (Habiba et al., 2019; Rangarajan & Purushothaman, 2020). The VGG algorithms combine convolutional layers, pooling layers between the convolutional and fully connected layers (FC). According to Simonyan and Zisserman (2014), the algorithms produced more than 90% accuracy when tested with ImageNet with 14 million images and more than 1000 classes. ResNet50 also produced an accuracy of more than 90% when tested with the ImageNet dataset (He et al., 2015). This algorithm has a combination of 50 layers deep of convolutional block and identity block, in which each block has three convolutional layers. Meanwhile, EfficientNet utilises Inverted Residual Blocks or MBConv Block (Sandler et al., 2018), in which in the algorithm, a single convolutional filter is used for each input channel to reduce the number of operations and simultaneously increase the efficiency. There are seven versions of EfficientNet, namely EfficientNet B0, EfficientNet B1, EfficientNet B2, EfficientNet B3, EfficientNet B4, EfficientNet B5, EfficientNet B6 and EfficientNet B7. According to (Tan & Le, 2019), EfficientNet B7 outperformed all the EfficientNet variants as it produced the highest accuracy, namely 84.3%, compared to EfficientNet B0 with only 77.1% accuracy when classifying images from ImageNet.

The experiments were implemented using Python in a Jupyter notebook environment, running on Intel®Core™ i5-8250U, where the base frequency is 1.60GHz, and the maximum frequency is 3.70GHz. Transfer learning has been adopted to run the trials. It means that the architecture of the selected deep learning models has similar architecture as mentioned in Simonyan and Zisserman (2014), He et al. (2015), and Tan and Le (2019) except for the final dense layer where it is changed to 8 for the eight plant species. Adam optimiser is an adaptive learning rate technique is utilised to adapt the learning rate for each weight of the neural network. This technique can minimise the loss function. Other than that, the deep learning models were trained using pre-trained weights. The batch size is set to 32, the number of epoch and the steps per epoch are each set to 10.

Nevertheless, the experiment was conducted on eight plants, namely Green chiretta, Mexican mint, Small water-pepper, Asiatic pennywort, Chinese gynura, Longevity spinach, Java tea and Java pennywort. The acquisition process for the remaining four plants was still ongoing during the experiment. The total number of images for these eight plants is 22800, and each plant species has 2850 images. All the input images were resized to 224×224 pixels, and data augmentation was not performed in this experiment. The performance of the deep learning algorithms in classifying a different total number of images is investigated as well. For this purpose, the images were divided into three image classes, namely, full,

half and quarter, in which each class consists of a different number of images, as shown in Table 5. The full class means that there are 22800 images from eight plant species, where there are 2850 images from each plant. As for the half and quarter classes, the total number of images used in the experiment are 1425 images and 713 images, respectively. For these two classes, the total number of images from each plant is 11400 images and 5704 images, respectively. For the purpose of investigating the performance of the deep learning algorithms, the images in each class were split into 70% for training and 30% for testing.

Table 5
Design parameters used in the experiments

Category	Class	Number of images for training	Number of images for testing
1	Full	15960	6840
2	Half	7976	3424
3	Quarter	4000	1704

RESULTS AND DISCUSSION

The performances of the selected deep learning algorithms were evaluated based on the accuracy, sensitivity, specificity, precision, and F1-score, as in Equations 1 to 5. Accuracy is defined as the number of correctly identified samples, while sensitivity is the number of positive samples that are accurately identified. Specificity is the number of negative samples that are correctly identified, and precision is the measurement of accurately identified samples among all the correct samples. As for the F1-score, the value represents a harmonic mean between sensitivity and precision.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

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

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(TP + FN)} \quad (5)$$

where TP, TN, FP and FN are true positive, true negative, false positive and false negative. The validation accuracy produced by each algorithm is illustrated in Figure 4. It is shown that both VGG16 and VGG19 achieved more than 90% of accuracy when tested on

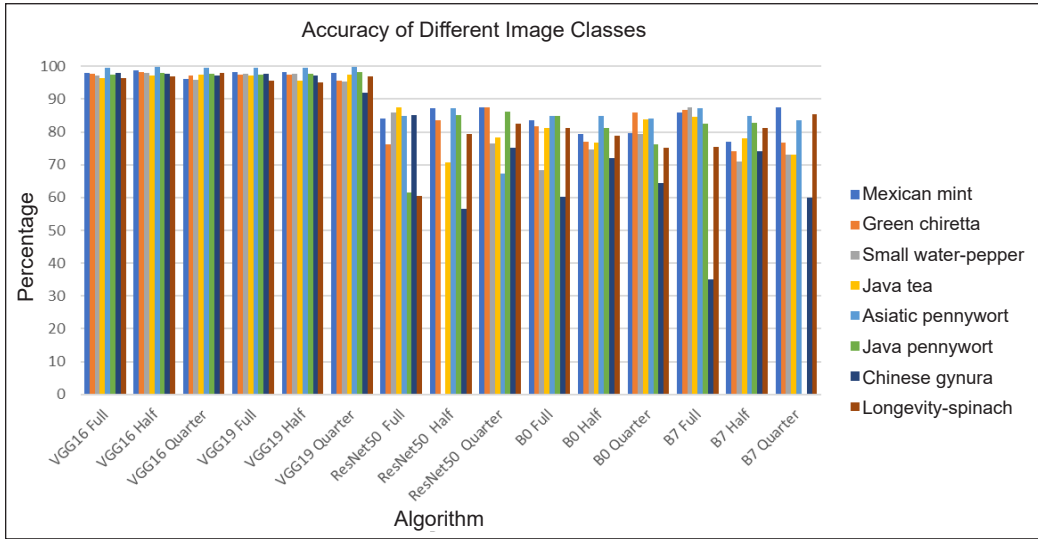


Figure 4. Accuracy performance for different image classes

the full, half and quarter datasets sizes. As for ResNet50, B0 and B7, the accuracy was consistently low. It shows that the dataset's images are very challenging to the three deep learning classifiers, and the classifiers require more images for better performance. It is also shown in He et al. (2016), Simonyan and Zisserman (2014), and Tan and Le (2019), in which the authors showed that ResNet50, EfficientNetB0 and EfficientNetB7 outperformed VGG16 and VGG19 when classifying images from ImageNet dataset that consists of 1000 classes with over 14 million images. Najafabadi et al. (2015) also concluded that some deep learning algorithms are dependent on a high number of data as they need to extract a large number of valuable data points to learn. Their hierarchical architectural nature allows for a better representation of data. Thus, the challenge of classifying images from the MYLPHerbs-1 dataset creates more opportunities for developing advanced classifiers.

The effect of the number of images on the accuracy is further explained through One Way ANOVA Analysis. This test is dependent on the p-value and significance figure. The significance figure set for this experiment is 0.01. The impact of dataset size on accuracy depends on the acceptance or rejection of the null hypothesis. The null hypothesis states that the groups being compared have no significant difference and have the same mean. Therefore, the measured p-value must be lower than the significant figure for the null hypothesis to be rejected and vice versa.

Table 6 shows the p-values for each class. Based on the values, it can be confirmed that the deep learning algorithm affects accuracy. Different deep learning algorithms affect the accuracy differently. As for Table 7, the accuracies achieved by each deep learning algorithm for each class were also compared using ANOVA analysis. The p-values show that the dataset size does not affect the accuracies of each deep learning algorithm.

Table 6
Comparison of the p-value for each image class

p-value for each image class		
Full	Half	Quarter
5.77×10^{-24a}	4.77×10^{-18a}	7.48×10^{-24a}

Note: a denotes values are significantly different

Table 7
Comparison of the p-value for each deep learning algorithm

p-value for each deep learning algorithm				
VGG16	VGG19	ResNet50	B0	B7
0.62 ^b	0.62 ^b	0.96 ^b	0.17 ^b	0.25 ^b

Note: b denotes values are not significantly different

As for sensitivity, specificity, precision and F1-score, the results are shown in Figures 5 to 8. Figures 5 to 8 show that both VGG16 and VGG19 can classify the images quite accurately. The graphs also show that for VGG16 and VGG19, the different total number of images does not affect the ability of the deep learning algorithms to classify the images in the dataset. It shows that both algorithms could classify a small number of images. Nevertheless, a different trend is observed for ResNet50, B0 and B7, in which these algorithms are extremely unstable and have poor classification capabilities for this particular dataset. In Figure 5, it can be seen that ResNet50, B0 and B7 have low sensitivity, and this shows that these three algorithms are unable to correctly identify positive images, in which a low percentage of TP and there is also a possibility of a high percentage of FN. Nevertheless,

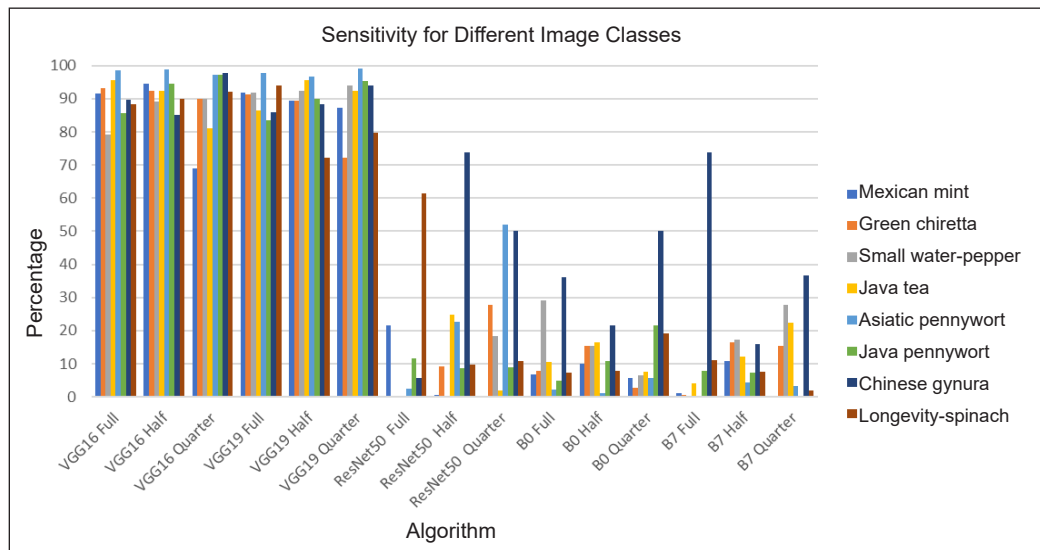


Figure 5. Sensitivity performance for different image classes

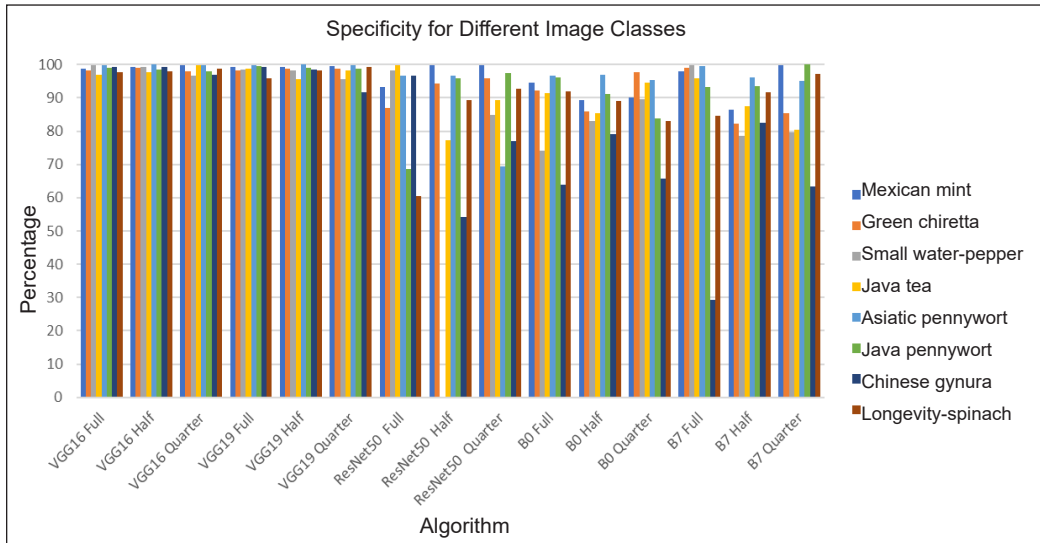


Figure 6. Specificity performance for different image classes

both VGG16 and VGG19 have high accuracy, indicating that these algorithms have high true positives and true negatives. On the other hand, ResNet50, B0 and B7 show a high specificity rate in Figure 6. The specificity rate depends on the value of the true negative.

In Figure 6, it is shown that for most of the herbs, the percentage of specificity produced by all algorithms is much better than the sensitivity rate. It shows that all the algorithms have better performance in identifying negative images. However, for ResNet50, EfficientNet B0 and EfficientNet B7 algorithms, there is a possibility of a high false-positive percentage, especially for Chinese gynura and Asiatic pennywort since the percentage of sensitivity when classifying using ResNet50, EfficientNet B0 and EfficientNet B7 is low. It is also illustrated in Figures 7 and 8, in which the ResNet50, EfficientNet B0 and EfficientNet B7 algorithms produced a low percentage of precision and F1-score. By comparing specificity and precision, it can be seen that the low precision value obtained by ResNet50, EfficientNet B0 and EfficientNet B7 proves that these three algorithms have low true positive and false-positive samples. It shows that the performance of the deep learning models is greatly influenced by the dataset used in the experiment. It must be noted that the challenge faced in this dataset is different from other datasets. Based on the existing works, ResNet50, EfficientNet B0 and EfficientNet B7 models were shown to produce good results using the ImageNet dataset. Hence, the results obtained in this experiment cannot directly reflect the results obtained through the ImageNet challenge as the images used in this dataset are completely different from the images in the ImageNet dataset. Furthermore, the performance of VGG16 and VGG19 models are better than ResNet50, EfficientNet B0 and EfficientNet B7 within the designed parameters in this experiment, where ResNet50 and EfficientNet B0 and B7 may achieve better results if these deep learning models were

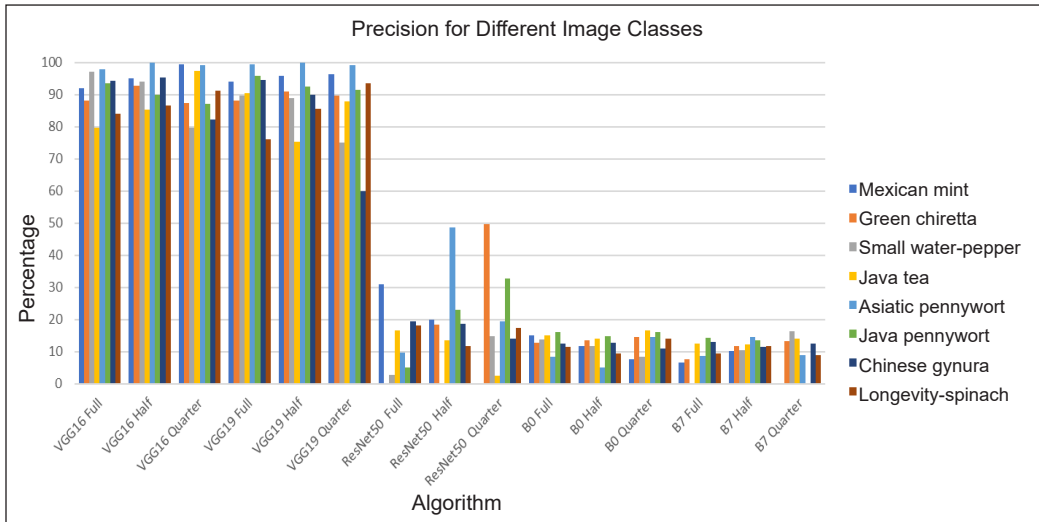


Figure 7. Precision performance for different image classes

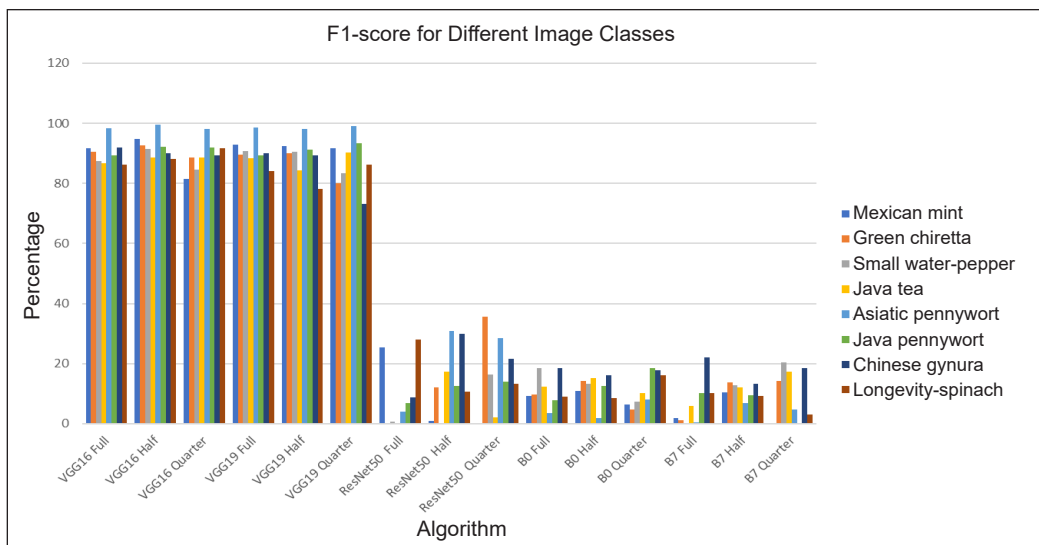


Figure 8. F1-score for different image classes

fine-tuned accordingly. In terms of the architecture of the model, VGG16 and VGG19 models have many weight layers, which could have also contributed to good performance when classifying the images from the proposed dataset.

From the results, it can be concluded that the proposed dataset is challenging. It is due to the complexity of the dataset where some of the plants have approximately similar morphological features, as shown in Figure 9. All four plants have different green tones but have somewhat similar pinnate shapes. The differences among these four plants are the structure of the vein and also the texture. As for Mexican mint, Asiatic pennywort and

Chinese gynura, these three plants have distinctive features, as shown in Figure 10, allowing for deep learning algorithms to learn quickly. Mexican mint is thick and ovate shaped with serrated edges. It also has a deep cross venulate type of vein. As for Asiatic pennywort, it has a very distinctive orbicular shape. Chinese gynura has a purplish shade with light green veins. The vein starting from the midrib are pinnate shaped and subsequently changes to a dichotomous shape as the veins reach the edge of the leaf.

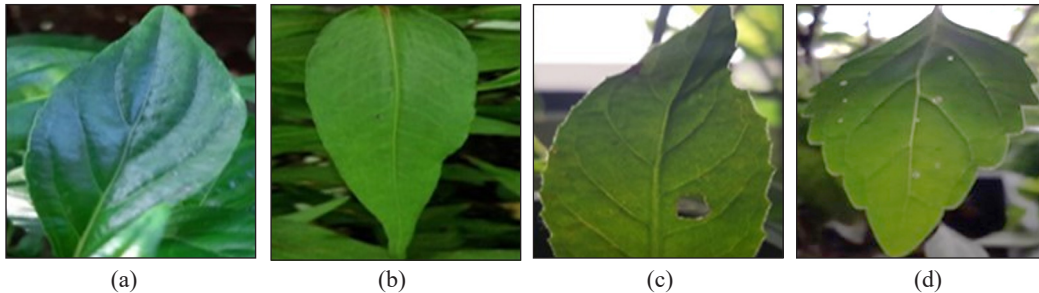


Figure 9. (a) Green chiretta (b) Small water-pepper (c) Longevity spinach (d) Java tea

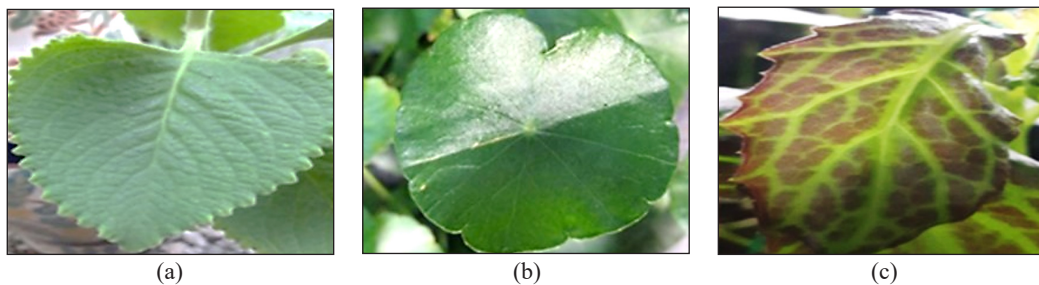


Figure 10. (a) Mexican mint (b) Asiatic pennywort (c) Chinese gynura

CONCLUSIONS

In conclusion, the experiment showed that some of the images in the proposed dataset are very challenging for classification. It is demonstrated by the selected different deep learning algorithms, namely VGG16, VGG19, ResNet50, EfficientNet B0 and EfficientNet B7, where only the VGG16 and VGG19 algorithms can show high performance when classifying a different number of images from eight plant species in the dataset. Overall, the proposed dataset that consists of 34200 images of twelve different Asian local perennial herbs that were captured in an uncontrolled environment under various imaging conditions, such as various distances from the camera, ambient illuminations and poses, may create abundant opportunities for researchers to study further the classifiers that best suited for automatic identification or recognition of medicinal plants in real-time. The contribution to knowledge is not limited to the research community but also the public recognising the Asian local perennial herbs with high medicinal value.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support received from Universiti Putra Malaysia under Putra Graduate Initiative (Vot number:9529000 and 9634300) and colleagues for their insight and expertise.

REFERENCES

- Abdelwahab, S. I., Mohan, S., Elhassan, M. M., Al-Mekhlafi, N., Mariod, A. A., Abdul, A. B., Abdulla, M. A., & Alkharfy, K. M. (2010). Antiapoptotic and antioxidant properties of *Orthosiphon stamineus* benth (Cat's Whiskers): intervention in the Bcl-2-mediated apoptotic pathway. *Evidence-Based Complementary and Alternative Medicine*, 2011, Article 156765. <https://doi.org/10.1155/2011/156765>
- Alam, A., Ferdosh, S., Ghafoor, K., Hakim, A., Juraimi, A. S., Khatib, A., & Sarker, Z. I. (2016). *Clinacanthus nutans*: A review of the medicinal uses, pharmacology and phytochemistry. *Asian Pacific Journal of Tropical Medicine*, 9(4), 402-409. <https://doi.org/10.1016/j.apjtm.2016.03.011>
- Arun, C. H., Emmanuel, W. S., & Durairaj, D. C. (2013). Texture feature extraction for identification of medicinal plants and comparison of different classifiers. *International Journal of Computer Applications*, 62(12), 1-9. <https://doi.org/10.5120/10129-4920>
- Ashaari, N. S., Rahim, M. H. A., Sabri, S., Lai, K. S., Song, A. A. L., Rahim, R. A., Abdullah, W. M. A. N. W., & Abdullah, J. O. (2020). Functional characterization of a new terpene synthase from *Plectranthus amboinicus*. *PloS one*, 15(7), Article e0235416. <https://doi.org/10.1371/journal.pone.0235416>
- Ashraf, K., Halim, H., Lim, S. M., Ramasamy, K., & Sultan, S. (2020). *In vitro* antioxidant, antimicrobial and antiproliferative studies of four different extracts of *Orthosiphon stamineus*, *Gynura procumbens* and *Ficus deltoidea*. *Saudi Journal of Biological Sciences*, 27(1), 417-432. <https://doi.org/10.1016/j.sjbs.2019.11.003>
- Ashraf, K., Sultan, S., & Adam, A. (2018). *Orthosiphon stamineus* Benth. is an outstanding food medicine: Review of phytochemical and pharmacological activities. *Journal of Pharmacy & Bioallied Sciences*, 10(3), 109-118. https://doi.org/10.4103/jpbs.JPBS_253_17
- Begue, A., Kowlessur, V., Mahomoodally, F., Singh, U., & Pudaruth, S. (2017). Automatic recognition of medicinal plants using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 8(4), 166-175. <https://doi.org/10.14569/IJACSA.2017.080424>
- Bhatt, P., Joseph, G., Negi, P., & Varadaraj, M. (2013). Chemical composition and nutraceutical potential of Indian borage (*Plectranthus amboinicus*) stem extract. *Journal of Chemistry*, 2013, 1-7. <https://doi.org/10.1155/2013/320329>
- Christopher, P., Parasuraman, S., Christina, J., Vikneswaran, M., & Asmawi, M. (2015). Review on *Polygonum minus*. Huds, a commonly used food additive in Southeast Asia. *Pharmacognosy Research*, 7(1), 1-6. <https://doi.org/10.4103/0974-8490.147125>
- Dahigaonkar, T. D., & Kalyane, R. (2018). Identification of ayurvedic medicinal plants by image processing of leaf samples. *International Research Journal of Engineering and Technology (IRJET)*, 5, 351-355
- Deshpande, P. (2017). Formulation and evaluation of herbal wound healing formulation of *Centella asiatica*. *World Journal of Pharmaceutical Research*, 1335-1345. <https://doi.org/10.20959/wjpr20176-8658>

- dos Santos, M. S., dos Santos Souza, L. E., Costa, C. A. S., Gomes, F. P., do Bomfim Costa, L. C., de Oliveira, R. A., & da Costa Silva, D. (2016). Effects of water deficit on morpho physiology, productivity and chemical composition of *Ocimum africanum* Lour (Lamiaceae). *African Journal of Agricultural Research*, 11(21), 1924-1934. <https://doi.org/10.5897/AJAR2015.10248>
- Giribabu, N., Karim, K., Kilari, E. K., Nelli, S. R., & Salleh, N. (2020). Oral administration of *Centella asiatica* (L.) urb leave aqueous extract ameliorates cerebral oxidative stress, inflammation, and apoptosis in male rats with type-2 diabetes. *Inflammopharmacology*, 28(6), 1599-1622. <https://doi.org/10.1007/s10787-020-00733-3>
- Gohil, K., Patel, J., & Gajjar, A. (2010). Pharmacological review on *Centella asiatica*: A potential herbal cure-all. *Indian Journal of Pharmaceutical Sciences*, 72(5), 546-556. <https://doi.org/10.4103/0250-474X.78519>
- Habiba, S. U., Islam, M. K., & Ahsan, S. M. M. (2019). Bangladeshi plant recognition using deep learning-based leaf classification. In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)* (pp. 1-4). IEEE Publishing. <https://doi.org/10.1109/IC4ME247184.2019.9036515>
- Haida, Z., Nakasha, J. J., & Hakiman, M. (2020). *In vitro* responses of plant growth factors on growth, yield, phenolics content and antioxidant activities of *Clinacanthus nutans* (Sabah snake grass). *Plants*, 9(8), Article 1030. <https://doi.org/10.3390/plants9081030>
- Harsani, P., & Qurania, A. (2016). Medicinal plant species identification system using texture analysis and median filter. *Jurnal Ilmiah Kursor*, 8(4), 181-188. <https://doi.org/10.28961/kursor.v8i4.112>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep residual learning for image recognition*. arXiv preprint. <https://doi.org/10.1109/CVPR.2016.90>
- Janani, R., & Gopal, A. (2013). Identification of selected medicinal plant leaves using image features and ANN. In *2013 International Conference on Advanced Electronic Systems (ICAES)* (pp. 238-242). IEEE Publishing. <https://doi.org/10.1109/ICAES.2013.6659400>
- Karthika, S. (2020). Investigating apoptotic effects of different extracts of medicinal plants on SH-SY5Y cells. *International Journal of Green Pharmacy (IJGP)*, 14(02), 175-178.
- Khoo, L. W., Kow, S. A., Lee, M. T., Tan, C. P., Shaari, K., Tham, C. L., & Abas, F. (2018). A comprehensive review on phytochemistry and pharmacological activities of *Clinacanthus nutans* (Burm. F.) Lindau. *Evidence-Based Complementary and Alternative Medicine*, 2018, Article 9276260. <https://doi.org/10.1155/2018/9276260>
- Kurzawa, M., Filipiak-Szok, A., Kłodzińska, E., & Szłyk, E. (2015). Determination of phytochemicals, antioxidant activity and total phenolic content in *Andrographis paniculata* using chromatographic methods. *Journal of Chromatography B*, 995, 101-106. <https://doi.org/10.1016/j.jchromb.2015.05.021>
- Lau, H., Shahar, S., Mohamad, M., Rajab, N. F., Yahya, H. M., Din, N. C., & Hamid, H. A. (2020). The effects of six months *Persicaria minor* extract supplement among older adults with mild cognitive impairment: A double-blinded, randomized, and placebo-controlled trial. *BMC Complementary Medicine and Therapies*, 20(1), 1-15. <https://doi.org/10.1186/s12906-020-03092-2>

- Lulekal, E., Kelbessa, E., Bekele, T., & Yineger, H. (2008). An ethnobotanical study of medicinal plants in Mana Angetu District, Southeastern Ethiopia. *Journal of Ethnobiology and Ethnomedicine*, 4(1), 1-10. <https://doi.org/10.1186/1746-4269-4-10>
- Majdi, C., Pereira, C., Dias, M. I., Calhelha, R. C., Alves, M. J., Rhourri-Frih, B., Charrouf, Z., Barros, L., Amaral, J. A., & Ferreira, I. C. (2020). Phytochemical characterization and bioactive properties of cinnamon basil (*Ocimum basilicum* cv. 'Cinnamon') and lemon basil (*Ocimum* × *citriodorum*). *Antioxidants*, 9(5), Article 369. <https://doi.org/10.3390/antiox9050369>
- Mandal, M., Misra, D., Ghosh, N. N., & Mandal, V. (2017). Physicochemical and elemental studies of *Hydrocotyle javanica* Thunb. for standardization as herbal drug. *Asian Pacific Journal of Tropical Biomedicine*, 7(11), 979-986. <https://doi.org/10.1016/j.apjtb.2017.10.001>
- Mandal, M., Paul, S., Uddin, M. R., Mondal, M. A., Mandal, S., & Mandal, V. (2016). *In vitro* antibacterial potential of *Hydrocotyle javanica* Thunb. *Asian Pacific Journal of Tropical Disease*, 6(1), 54-62. [https://doi.org/10.1016/S2222-1808\(15\)60985-9](https://doi.org/10.1016/S2222-1808(15)60985-9)
- Murugan, N. A., Pandian, C. J., & Jeyakanthan, J. (2020). Computational investigation on *Andrographis paniculata* phytochemicals to evaluate their potency against SARS-CoV-2 in comparison to known antiviral compounds in drug trials. *Journal of Biomolecular Structure and Dynamics*, 39(12), 4415-4426. <https://doi.org/10.1080/07391102.2020.1777901>
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1-21. <https://doi.org/10.1186/s40537-014-0007-7>
- Okhwarobo, A., Falodun, J. E., Erharuyi, O., Imieje, V., Falodun, A., & Langer, P. (2014). Harnessing the medicinal properties of *Andrographis paniculata* for diseases and beyond: A review of its phytochemistry and pharmacology. *Asian Pacific Journal of Tropical Disease*, 4(3), 213-222. [https://doi.org/10.1016/S2222-1808\(14\)60509-0](https://doi.org/10.1016/S2222-1808(14)60509-0)
- OSU. (2021). *Plant identification: Examining leaves*. Oregon State University. Retrieved January 2, 2021, from <https://landscapeplants.oregonstate.edu/plant-identification-examining-leaves>
- Pornpanomchai, C., Rimdusit, S., Tanasap, P., & Chaiyod, C. (2011). Thai herb leaf image recognition system (THLIRS). *Agriculture and Natural Resources*, 45(3), 551-562.
- Proklamasiningsih, E., Budisantoso, I., Kamsinah, K., & Widodo, P. (2020). Antioxidant activity and flavonoid contents of daun dewa (*Gynura pseudochina*) in various substrates with humic acid treatment. In *IOP Conference Series: Earth and Environmental Science* (Vol. 593, No. 1, p. 012026). IOP Publishing. <https://doi.org/10.1088/1755-1315/593/1/012026>
- Rahman, A., & Asad, M. (2013). Chemical and biological investigations of the leaves of *Gynura procumbens*. *International Journal of Biosciences* 3(4), 36-43. <https://doi.org/10.12692/ijb/3.4.36-43>
- Rangarajan, A. K., & Purushothaman, R. (2020). Disease classification in eggplant using pre-trained VGG16 and MSVM. *Scientific Reports*, 10(1), 1-11. <https://doi.org/10.1038/s41598-020-59108-x>
- Sack, L., & Scoffoni, C. (2013). Leaf venation: Structure, function, development, evolution, ecology and applications in the past, present and future. *New Phytologist*, 198(4), 983-1000. <https://doi.org/10.1111/nph.12253>

- Sahu, P. K., Singh, S., Gupta, A. R., Gupta, A., Singh, U. B., Manzar, N., Bhowmik, A., Singh, H. V., & Saxena, A. K. (2020). Endophytic bacilli from medicinal-aromatic perennial Holy basil (*Ocimum tenuiflorum* L.) modulate plant growth promotion and induced systemic resistance against *Rhizoctonia solani* in rice (*Oryza sativa* L.). *Biological Control*, 150, Article 104353. <https://doi.org/10.1016/j.biocontrol.2020.104353>
- Samidurai, D., Pandurangan, A. K., Krishnamoorthi, S. K., Perumal, M. K., & Nanjian, R. (2020). Sinensetin isolated from *Orthosiphon aristatus* inhibits cell proliferation and induces apoptosis in hepatocellular carcinoma cells. *Process Biochemistry*, 88, 213-221. <https://doi.org/10.1016/j.procbio.2019.09.031>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520). IEEE Publishing. <https://doi.org/10.1109/CVPR.2018.00474>
- Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint.
- Singh, D., & Chaudhuri, P. K. (2018). A review on phytochemical and pharmacological properties of Holy basil (*Ocimum sanctum* L.). *Industrial Crops and Products*, 118, 367-382. <https://doi.org/10.1016/j.indcrop.2018.03.048>
- Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1), 41-49. <https://doi.org/10.1016/j.inpa.2016.10.005>
- Siriwatanametanon, N., & Heinrich, M. (2011). The Thai medicinal plant *Gynura pseudochina* var. *hispidula*: Chemical composition and *in vitro* NF- κ B inhibitory activity. *Natural Product Communications*, 6(5). <https://doi.org/10.1177/1934578X1100600512>
- Siriwatanametanon, N., Fiebich, B. L., Efferth, T., Prieto, J. M., & Heinrich, M. (2010). Traditionally used Thai medicinal plants: *In vitro* anti-inflammatory, anticancer and antioxidant activities. *Journal of Ethnopharmacology*, 130(2), 196-207. <https://doi.org/10.1016/j.jep.2010.04.036>
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks-based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, 2016, Article 3289801. <https://doi.org/10.1155/2016/3289801>
- Swamy, M., Arumugam, G., Kaur, R., Ghasemzadeh, A., Yusoff, M., & Sinniah, U. (2017). GC-MS based metabolite profiling, antioxidant and antimicrobial properties of different solvent extracts of Malaysian *Plectranthus amboinicus* Leaves. *Evidence-Based Complementary and Alternative Medicine*, 2017, Article 1517683. <https://doi.org/10.1155/2017/1517683>
- Tan, H. L., Chan, K. G., Pusparajah, P., Lee, L. H., & Goh, B. H. (2016). *Gynura procumbens*: An overview of the biological activities. *Frontiers in Pharmacology*, 7, Article 52. <https://doi.org/10.3389/fphar.2016.00052>
- Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105-6114). PMLR.
- Vijayashree, T., & Gopal, A. (2017). Authentication of herbal medicinal leaf image processing using Raspberry Pi processor. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1304-1307). IEEE Publishing. <https://doi.org/10.1109/ICCONS.2017.8250679>

- Vimala, S., Rohana, S., Rashih, A., & Juliza, M. (2012). Antioxidant evaluation in Malaysian medicinal plant: *Persicaria minor* (Huds.) leaf. *Science Journal of Medicine and Clinical Trials*, *1*, 9-16.
- Wäldchen, J., & Mäder, P. (2018a). Machine learning for image-based species identification. *Methods in Ecology and Evolution*, *9*(11), 2216-2225. <https://doi.org/10.1111/2041-210X.13075>
- Wäldchen, J., & Mäder, P. (2018b). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering*, *25*(2), 507-543. <https://doi.org/10.1007/s11831-016-9206-z>
- Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification - Trends and future directions. *PLoS Computational Biology*, *14*(4), Article e1005993. <https://doi.org/10.1371/journal.pcbi.1005993>
- Yamani, H. A., Pang, E. C., Mantri, N., & Deighton, M. A. (2016). Antimicrobial activity of Tulsi (*Ocimum tenuiflorum*) essential oil and their major constituents against three species of bacteria. *Frontiers in Microbiology*, *7*, Article 681. <https://doi.org/10.3389/fmicb.2016.00681>

