Lettuce classification using convolutional neural network

[1*]Hassim, S.A. and [2]Chuah, J.H.

[1]*Engineering Research Center, MARDI Head Quarters, 43400 Serdang, Selangor, Malaysia.*
[2]*Department of Electrical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

Abstract

Determining the varieties of lettuce through image processing and pattern recognition is a part of precision farming. Automatic classification is becoming vital for precision farming practice as it is rapidly sprouting field with the emergence of many applications in agriculture. It is a hassling process to differentiate and identify the lettuce varieties through human capabilities as it is time-consuming and also prone to errors in the identification process. Hence, there is a need to perform this task assisted by a machine capability which makes it faster with even greater accuracy. The objective of this research work is to design lettuce varieties recognition using Convolutional Neural Network (CNN) in MATLAB with an accuracy of at least 90%. CNN was employed to classify seven types of most commonly found lettuce. The CNN model was trained with 7000 leaves and tested with 1800 leaves for the classification of 7 varieties of lettuce. The overall classification accuracy is 97.8%; meanwhile, individual classification accuracies for the selected lettuce varieties, i.e. Butterhead, Celtuce Love, Italian, Red Coral, Lactuca Sativa Lettuce, Red Oakleaf and Salad Grand Rapid are 97%, 99.3%, 98.7%, 96%, 100%, 99.3%, and 94%, respectively. The results from this study have proven the high effectiveness of using a machine learning technique, i.e. CNN, to identify a particular variety of lettuce.

1. Introduction

Lettuce is one of the healthier vegetables that has market value where some of the varieties only can be obtained from hypermarkets. Preservation of the product's qualities, proper plant monitoring is crucial during the growth to get a high yield. However, before developing an application for plant monitoring or get other information related to the lettuce varieties, the recognition process with the correct name is the essential part of the application. Designing the system that able to recognise lettuce varieties is necessary to facilitate fast classifying lettuce and allowing researchers or farmers to have a proper method to manage them.

Artificial Intelligence (A.I.) is part of a universal field which is relevant to any intellectual task, ranging from the general right to the specific application. Pannu (2015) claimed that sectors or areas that adapting Artificial Intelligence resulted in an incremental in quality and efficiency. The adoption of A.I. gives an impact on various fields as expert system widely use it to solve complex problems such as in science, engineering, business, medicine, weather forecasting.

Some of the research has done to classify the plant species based on leaf and use a different method to recognise them. Wu *et al*. (2007) used a Probabilistic Neural Network (PNN) approach for automated leaf recognition for plant classification. The writer used feature extraction to allow the computer to obtain feature values automatically. The feature extraction involved five basic geometry features that can be defined as digital morphological features for leaf recognition. The Principal Component Analysis (PCA) was used to represent the information of original data as a linear combination of certain linear irrelevant variables. The author also mentioned that PNN used in this research because of its simple structure and the training part was easy and instantaneous. PNN derived from Radial Basis Function (RBF) which scales the variable non-linearly. One thousand eight hundred (1800) leaves were trained and has an accuracy percentage greater than 90% to classified 32 types of plants.

Lee *et al*. (2015) studied CNN for 44 different plant

species to learn unsupervised feature. The author compared the performance of M.K. leaf dataset with different classifier. The result obtained one of the essential elements to identify plant species is the venation structure that has an accuracy of 99.5%. The author also justified that for a better representation image for leaf, it is better for learning features through CNN compared to hand-crafted features. Priya *et al.* (2012) applied the Support Vector Machine (SVM) for plant classification. Three critical phases involved in this approach which is preprocessing, feature extraction and classification. Twelve (12) features obtained are extracted and processed by PCA to form the input vector of SVM. From the result received, the author stated that the proposed algorithm produces better accuracy and required less time for execution compared to the k-NN method.

Chaki and Parekh (2011) use Neural Network classifiers and shape-based features for plant leaf recognition. Three different plant types are analysed using the Moments-Invariant (M-I) model and CentroidRadii (C-R) model. Between both models, C-R method gets better accuracy compared to M-I model where C-R performed 100% accuracy. Du *et al.* (2005) proposed a Radial Basis Probabilistic Neural Network (RBPNN) recognise the shape. Orthogonal least square algorithm (OLSA) is to train RBPNN, and recursive OLSA is used to optimise the structure of RBPNN. The author also compared the RBPNN classifier with multi-Layer perceptron network (MLPN). The 20 species from different plants are used as leaf image dataset where 40 leaves images for each species. From the result obtained, the percentage of recognition rate for both methods are 96.2 and 94.4 for RBPNN and MLPN, respectively. However, the training time for RBPNN is less than MLPN, which only took 48 seconds for RBPNN and 272 seconds for MLPN.

Pan and He (2008) proposed the recognition of plants using leaves digital image and neural network. The data divided into two parts, one is for training, and the other is for validation. The author took shots of soybean, goosegrass and alligator alternanthera at the fields. Two types of detection were applied, which are border segmentation and area segmentation. The author chose a Radial Basis Network (RBN) as it has a strong classification power. The layer consist of a hidden radial basis layer and linear output layer. The dataset used for this experiment is about 145 blocks which 100 blocks were use as the training dataset, and 45 blocks were used to check the validation of the model. The result for this model, correctly achieved classification by more than 80%.

Satti *et al.* (2013) compared the classification of leaf recognition for plant identification using Artificial Neural Network (ANN) and Euclidean (KNN) classifier. The proposed approach consists of preprocessing, feature extraction and classification. The extraction phase features based on the colour and shape of leaf images. The accuracy for both classifier ANN and KNN are 93.3% and 85.9% respectively.

Another research in the agricultural sector is the classification type of rice. Silva and Sonnadara (2013) using MLP for classification of rice grains. The model was developed for feature set individually and combined. Combined feature model gets an overall accuracy of 92% while individual feature gets the overall accuracy 51%, 63 % and 34% for the morphological model, texture model and colour model respectively. Patela and Joshib (2017) also research rice type classification by using CNN with transfer learning. The 2-class model trained 1600 images in order to classify a broken and regular rice, whereas the 5-class model trained 4000 images in order to classify the rice types. The accuracy achieved with model 5-class is 86.8% with transfer learning of classification and accuracy of 94.32% achieved without transfer learning of classification.

Another research using CNN is to detect plant disease identification. Ferentinos (2018) identify plant diseases using healthy leaves images and plant diseases. The author compared the result with the different CNN model architecture such as AlexNet, AlexNetOWTBn, GoogLeNet. Overfeet and VGG. All of the architecture achieved the success rate of more than 97%, and VGG obtained the highest accuracy of 99.53%in classification. The author stated that the model could be used as an early warning notification or as a support to an integrated plant disease identification system.

Maa *et al.* (2018) recognise cucumber diseases based on Deep Convolutional Neural Network (DCNN) using leaf symptom images under field condition. Data augmentation methods are utilised in order to decrease of overfitting. Author conduct experimentation using DCNN, AlexNet, Random Forest and Support Vector Machines. From the result obtained, the accuracy for DCNN and AlexNet achieved 93.4 and 94 percentage whereas R.F. and SVM achieved 81.9 and 84.8 percentage respectively. There are some researches done for image classification at the field. Dyrmann (2017) doing an automatic detection and classification of weed seedlings under natural light conditions. The experiment can handle weed detection and classification in natural environments. Thus, the methods applied able to help and reduce the operational cost involved which in turn lead to the higher potential rate of adoption compared

with existing precision techniques of weed control that lead to a potential saving in regards to consumption of herbicide. Based on the results of accuracy percentage, VGG 19 architecture achieved the highest accuracy of 87.3% with a total of 2967 plants spread over 17 weed species.

CNN is a type of Deep Neural Networks (DNN) that consists of many layers such as the Convolution layers, Pooling layer, and Fully-connected layer. It is mainly used for image classification purposes. CNN architecture successively applying layers of convolutional onto input which follows the same design principles, the spatial dimensions in downsampling while the number of feature map is increasing.

From the past review papers, researchers choose A.I. in order to achieve high accuracy in classification. There are many uses of A.I. in agricultural research such as leaf classification, plant recognition, leaf disease detection, plant disease detection and many more. These researches can be applied and provide impact to the agriculture sector in terms of product quality, proper plant arrangement and reduction in labour cost and many more. This research focuses on design leaf lettuce recognition with an accuracy of more than 90% using CNN in MATLAB.

2. Materials and methods

The implementation method to recognise the lettuce varieties was use MATLAB (2018) and ran on Intel Core i5 with the clock of 2.5GHz and the RAM of 4G under Microsoft Windows 10 environment software. Figure 1 below shows the architecture of the CNN of this paper. Each image's size is 32 x 32 pixel and depth is three (3) referring to the Red, Green, Blue (RGB) channel. The output of neurons that connects to local regions in the input will be computed by the Convolutional Layer with each computing a dot product between its weights and small region connected in the input volume. For this paper, the Convolutional Layer creates 32 filters of size [5 5]. Then, each of the Convolutional Layer will proceed with the Pooling and ReLU Layer.
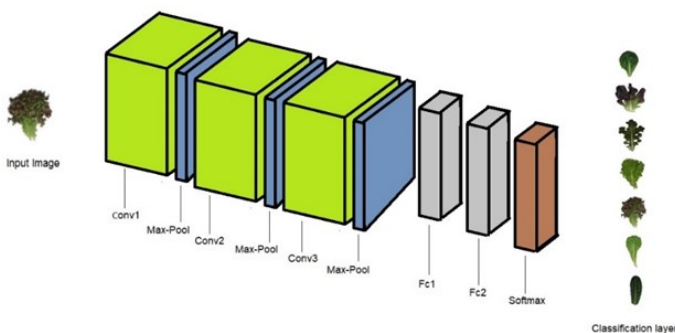

Figure 1. The architecture of the CNN

The network then proceeds with the Fully-Connected layer that will compute the class scores, resulting in a volume of size [1x1x7], where seven correspond to the categories of the leaf lettuce recognition. Optimisers used is Stochastic Gradient Descent with Momentum (SGDM). Qian (1998) stated that SGD with momentum is an approach to assist acceleration SGD in the relevant direction and dampens oscillations. This optimiser was selected because it is able to increase the speed of learning and make updates from the stored velocity of all parameters.

2.1 Lettuce samples

The total images involved in lettuce recognition is 7000, which is 70% and 30% from the total image used for training and testing dataset respectively. The number of images for each class is as listed in Table 1. The cultivation of seven varieties lettuce was planted inside of a plant factory and greenhouse that is available at Malaysia Agricultural Research and Development Institute (MARDI). The verities considered in this research work are; Butterhead, Celtucelove, Italian, Red Coral, Red Lettuce, Red Oakleaf and Salad Grand Rapid (SGR). Figure 2 shows the images of lettuce varieties used for this study.

Table 1. Distribution of images for leaf lettuce dataset for training and test.

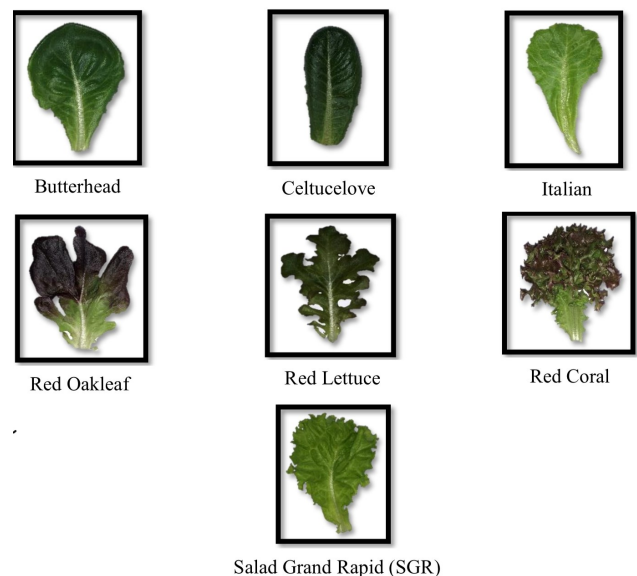| No	Names	Training	Testing	Total
1	Butterhead	700	300	1000
2	Celtucelove	700	300	1000
3	Italian	700	300	1000
4	Red Coral	700	300	1000
5	Red Lettuce	700	300	1000
6	Red Oakleaf	700	300	1000
7	Salad Grand Rapid (SGR)	700	300	1000

Figure 2. Lettuce varieties

2.2 Image acquisition

The images of lettuce were captured by using a smartphone dual-camera and a resolution of 12 megapixels and optical image stabilisation. The image taken is in RGB (Red, Green, blue) image. Wang *et al.* (2008) stated that the background image needs to be clean either in white colour or any colour that contrast with the sample. The background chosen for the image sample is white since it has good colour contrast with the sample. The camera's height set at 1.5 feet for clear image visibility of the leaf lettuce.

2.3 Image pre-processing

Some of the image datasets have different sizes due to camera orientation selected during the photoshoot. The original image size is 3024 by 3024 pixel and 3024 by 4024 pixel. Hence, the images needed to resize in order for all images has the same size. The image size was changed to 32 by 32 pixel because to reduce the time taken for training whereas training time will be prolonged if maintained the original size of the images.

The size of the image was set to 32 by 32 as training input. Number of training sample to work through before internal parameter was updated was control by the batch size, which is a hyperparameter. The leaf lettuce recognition tested from batch size 80 until 32. The epoch number is the number of complete passes through the training dataset.

3. Result and discussion

Dataset was tested with several parameters, such as the number of epochs and mini-batch size, to get high accuracy. From Figure 3 it shows that the parameter setting was the best since the accuracy percentage achieved for each variety of lettuce was above 90%. The number of epochs set to 10 and the mini-batch size is 32.

Figure 4 is the confusion matrix of the model, and it shows that the validation reduced to 2.2%. The total number of correct and wrong prediction is represented by green and red boxes, respectively. The total number of correct prediction was improved with five varieties of
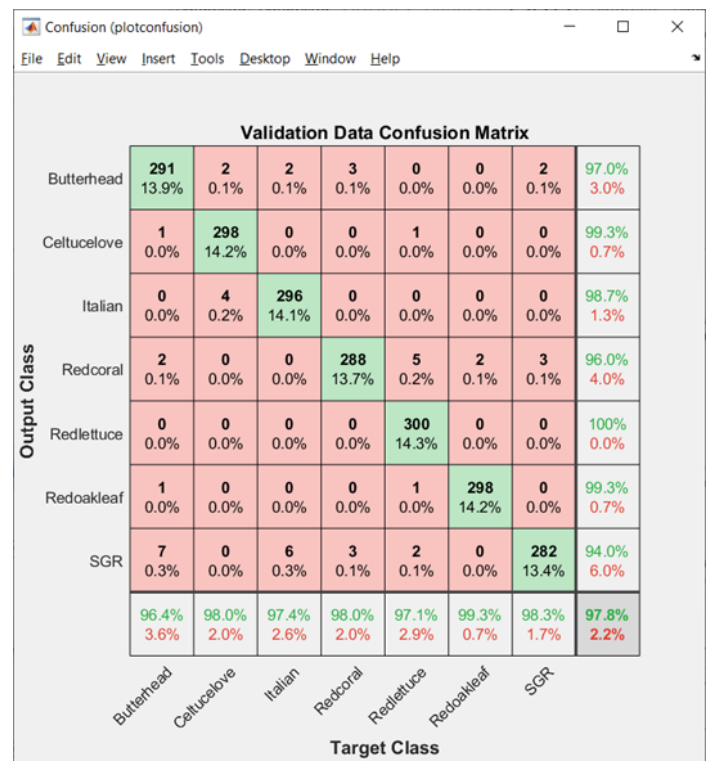

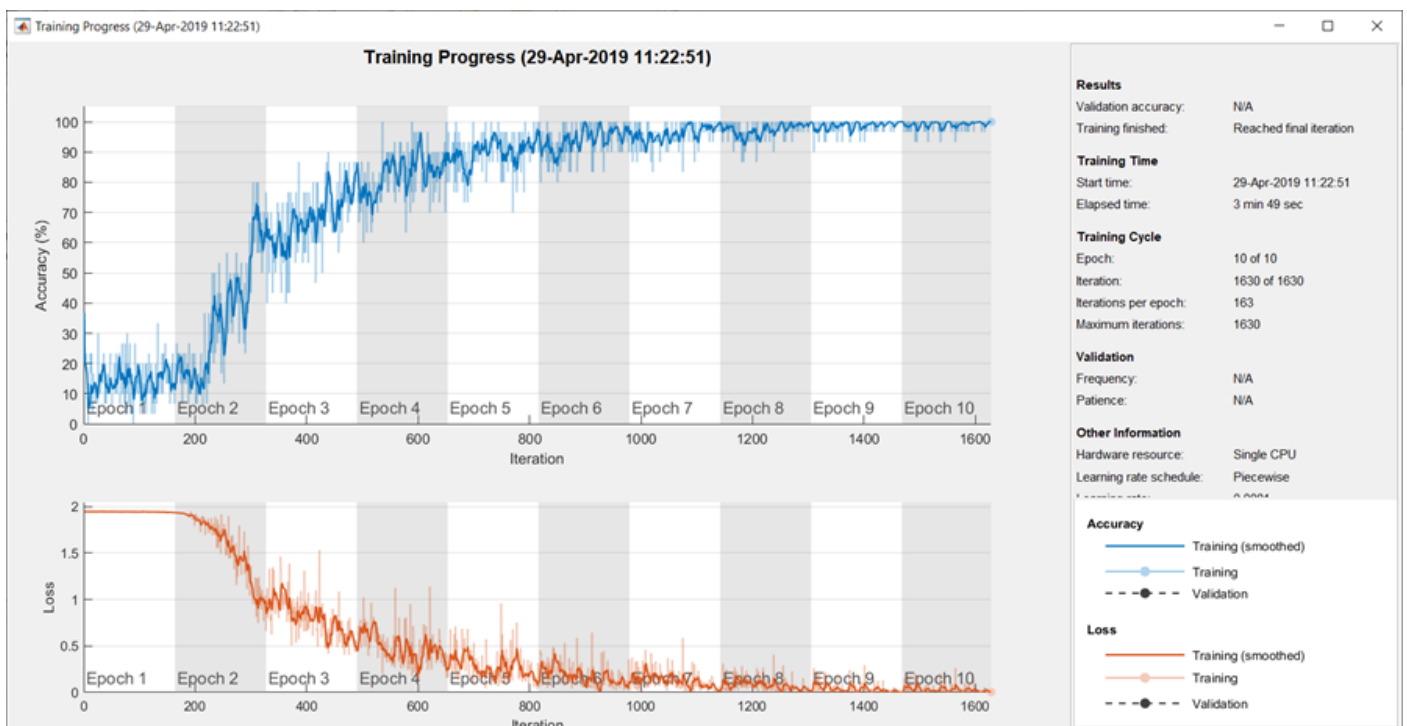Figure 4. Confusion matrix for mini-batch size 32 and number of epoch 10

Figure 3. Training progress with mini-batch size 32 and number of epoch 10

Table 2. Summary of all parameter setting for the model

No of epochs	Mini batch size	Average accuracy (%)	Validation error (%)	Training time
10	80	79.8	20.2	2 mins 44 s
15	80	83.7	16.3	6 mins 20 s
20	80	91.7	8.3	6 mins 9 s
10	60	92.8	7.2	1 min 30 s
20	60	93	7	5 mins 30 s
10	40	96.5	3.5	2 mins 39 s
15	40	96.8	3.2	4 mins 02 s
10	32	97.8	2.2	3 mins 49 s

lettuce have the correct prediction more than 290 out of 300 leaf sheets and the others two types are correctly predicted above 282 out of 300 leaf sheets of the testing image after the mini-batch size reduced to 32. From the confusion matrix, each variety gets the accuracy of more than 95% with the average accuracy for the model is 97.8%.

Based on Table 2, the number of epochs of 10 and mini-batch size of 32 get the highest accuracy compared to the other setting of the parameter.

Figure 5 shows the result of the average accuracy and validation error percentage. It shows that the accuracy increased when the size of the mini-batch reduced while the validation error can be reduced until 2.2%. Therefore, the parameter setting is suitable to the model since the average accuracy and average of each variety of lettuce is more than 90%.
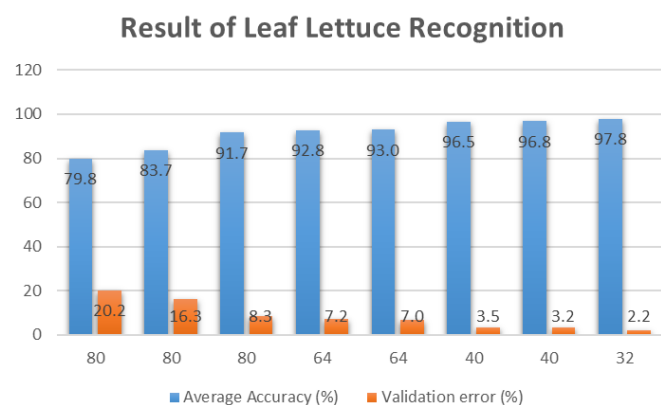

Figure 5. Results of leaf lettuce recognition based on different mini-batch sizes.

4. Conclusion

From the results obtained, it is justified that CNN was able to determine and obtained high accuracy percentage prediction of leaf lettuce images in the MATLAB simulation window. Based on leaf images, the model was able to differentiate seven different varieties of lettuce with a high accuracy percentage. For future improvement, more images are needed for testing and training the model to verified and certified that the model was able to predict varieties of lettuce correctly. The preliminary work presented in this paper could further be improved with images taken at the field without neglecting other obstacles for the image's background. From the model, the application to recognize varieties of lettuce can be developed further, and other information related to lettuce such as significant name, family name, the benefit of lettuce or market price can be incorporated. Besides that, other features such as differentiation size of lettuce based on growth day. Farmers can observe the growth of lettuce real-time or notified the best time to harvest.

Acknowledgement

References

Chaki, J. and Parekh, R. (2011). Plant leaf recognition using shape-based features and neural network classifiers. *International Journal of Advanced Computer Science and Applications,* 2(10), 41-47. https://doi.org/10.14569/IJACSA.2011.021007

Du, J., Huang, D., Wang, X. and Gu, X. (2005). Shape recognition based on radial basis probabilistic neural network and application to plant species identification. In Wang, J., Liao, X.F. and Yi, Z. (Eds.). Advances in Neural Networks - ISNN 2005. Lecture Notes in Computer Science, Vol. 3497, p. 281-285. Berlin, Germany: Springer. https://doi.org/10.1007/11427445_45

Dyrmann, M. (2017). Automatic detection and classification of weed seedlings under natural light conditions. Denmark: University of Southern Denmark, PhD Dissertation.

Ferentinos, K.P. (2018). Deep learning models for plant

disease detection and diagnosis. *Computers and Electronics in Agriculture,* 145, 311-318. https://doi.org/10.1016/j.compag.2018.01.009

Lee, S.H., Chan, C.S., Wilkin, P. and Remagnino, P. (2015). Deep-plant: Plant identification with convolutional neural networks, presented at IEEE International Conference on Image Processing (ICIP), 2015. Canada: IEEE. https://doi.org/10.1109/ICIP.2015.7350839

Maa, J., Dua, K., Zhenga, F., Zhangb, L., Gongc, Z. and Suna, Z. (2018). A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Computers and Electronics in Agriculture,* 154, 18-24. https://doi.org/10.1016/j.compag.2018.08.048

Pan, J. and He, Y. (2008). Recognition of plants by leaves digital image and neural network, presented at International Conference on Computer Science and Software Engineering, 2009. Hubei, China: IEEE. https://doi.org/10.1109/CSSE.2008.918

Pannu, A. (2015). Artificial intelligence and its application in different areas. *International Journal of Engineering and Innovative Technology (IJEIT)*, 4(10), 79-84.

Patel, V.A. and Joshi, M.V. (2017). Convolutional neural network with transfer learning for rice type classification, presented at Tenth International Conference on Machine Vision, 2017. Vienna, Austria: SPIE. https://doi.org/10.1117/12.2309482

Priya, C.A., Balasaravanan, T. and Thanamani, A.S. (2012). An efficient leaf recognition algorithm for plant classification using support vector machine, presented at Conference on Pattern Recognition, Informatics and Medical Engineering, 2012. India: SPIE. https://doi.org/10.1109/ICPRIME.2012.6208384

Qian, N. (1999). The momentum term in gradient descent learning algorithms. *Neural Networks,* 12(1), 145–151. https://doi.org/10.1016/S0893-6080(98)00116-6

Satti, V., Satya, A. and Sharma, S. (2013). An automatic leaf recognition system for plant identification using machine vision technology. *International Journal of Engineering Science and Technology*, 5(4), 874-879.

Silva, C.S. and Sonnadara, U. (2013). Classification of rice grains using neural networks, presented at Proceedings of Technical Sessions, 2013. Sri Lanka.

Wang, X.F., Huang, D.S., Du, J.X., Xu, H. and Heutte, L. (2008). Classification of plant leaf images with complicated background. *Applied mathematics and computation*, 205(2), 916-926. https://doi.org/10.1016/j.amc.2008.05.108

Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network, presented at IEEE international symposium on signal processing and information technology, Giza, Egypt: IEEE. https://doi.org/10.1109/ISSPIT.2007.4458016