Proceedings of the
1st International Conference on Environmental Health, Socioeconomic and Technology 2022
Lamongan, Indonesia, August 10-11 2022
Vol. 1, No. 1 (https://conference.unisla.ac.id/index.php/icehst/)

CONTOUR BASED AND REGION BASED FEATURE EXTRACTION FOR CLASSIFICATION OF TUBERS BASED ON LEAF IMAGE

Miftahus Sholihin[1], M. Rosidi Zamroni[1], Erry Anggraini[1], Mohd Farhan MD Fudzee[2]
[1]Universitas Islam Lamongan, Jl. Veteran No 53 A, Lamongan, Indonesia
[2]Universitas Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia

miftahus.sholihin@unisla.ac.id (+62 838-3234-7831); rosidizamroni@unisla.ac.id (+62 857-0718-0857); eryanggraini17@gmail.com; farhan@uthm.edu.my

ABSTRACT

This study aims to classify the types of tubers. The amount of data used in this research is 411, which is divided into training data and testing data. The number of training data is 317, which is divided into 3 classes, namely cassava class 116, taro class 86, and uwi class 115. Total testing data used is 148, which is divided into 3 classes, namely cassava class 40, taro class 21, and uwi class 33. This research begins with segmenting the image using Otsu thresholding, then features extraction using contour-based and region-based methods. The last step is classification using K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) methods. This study provides the highest accuracy of 93% with the KNN algorithm and a neighboring value of 1 with contour-based feature extraction.

KEYWORDS: *segmentation, features extraction, classification, accuracy*

1 INTRODUCTION

Plants have a very important role in human life and the surrounding environment. Plants have great benefits for human life. This is because plants are able to produce oxygen that humans and animals need to breathe. In addition, plants can also be used as a source of food for humans. One type of plant that can be used as food is tubers. In Indonesia itself, tubers are used as the main food substitute for rice.

Plants can be identified by looking at their fruit. However, this method is considered less efficient. This is because the fruit is not always there throughout the season. Another way that is considered more efficient in recognizing plant species is to look at the leaves. This is because the leaves are available throughout the season and are easy to obtain.

In recent years, there have been many studies on the classification of plant species using leaves as objects (Mittal et al., 2018; Thanikkal et al., 2018; Wang et al., 2018). (Amlekar et al., 2015) uses shape features to classify plant species. Leaf shape extraction for classification using Sobel, Prefit, Robert, and Canny operators. Research conducted by (Murat et al., 2017) uses four shape features, namely Morphological Shape Descriptors (MSD), Histogram of Oriented Gradients (HOG), Hu invariant moments, and Zernike moments. Other research uses Centroid Contour Distances as Shape Features for the Classification of Mango Species. An SVM was used for classification with an accuracy of 67.3% (Prasetyo et al., 2018). (Donesh & Piumi Ishanka, 2020) conducted a comparison of Contour-based and Region-Based Feature Extraction for the classification of planting types with SVM used for classification. This study provides the highest accuracy of 72.25%. Other research uses shape features (rectangularity, roundness, aspect ratio, deviation degree, sawtooth degree) and texture features (contrast, energy, homogeneity) (Yang, 2021). Other research uses (Anitha et al., 2021) classified plants based on their leaves. This study uses mean filtering, statistical feature extraction, SVM, and Binary Decision Tree (BDT) methods for classification.

This study aims to classify the types of tubers based on leaf images. This research begins by segmenting leaf images using Otsu thresholding (Otsu, 1979). The next process is feature extraction with contour-based and region-based methods. The last process is classification using K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) methods. With this research, it is hoped that the community will be able to recognize the types of tubers based on the images of the leaves.

2 MATERIAL AND METHODS

2.1 Data Collection

The data used in this study is image data of tubers taken by farmers. The amount of data used in this study is 411, which is divided into training data and testing data. The number of training data is 317, which is divided into 3 classes, namely Singkong class 116, Talas class 86, and Uwi class 115. The amount of testing data used is 148, which is divided into 3 classes, namely Singkong class 40, Talas class 21, and Uwi class 33. Figure 1 is an example of the data used in the study. Figure 1 shows the data used in this study.

a) singkong b) talas c) uwi

Figure 1: Image data

2.2 Desain System

Figure 2 is the system design that was built in this research. Based on Figure 2, there are several processes as follows:

a) Segmentation

Image segmentation aims to separate an area in the image from other areas. Segmentation refers to the operation of separating an image into parts or dividing an image into desired parts. Image segmentation is generally based on one of two intensity values: discontinuity or similarity. In the first category, the approach is to break up or sort images based on rough changes in intensity, such as edges in an image. The main approach of the second category is based on splitting the image into equal regions according to a number of defined criteria, for example, thresholding, region expansion, region splitting, and merging.

In this research, segmentation uses the thresholding method. Thresholding is basically a process used to produce a binary image, which is an image that only has 0 and 1 or black and white. The process changes the image to black or white, depending on the threshold value (T). If a pixel has a value greater than the T value, it will be changed to white, while a pixel with a value less than the T value will be changed to black. The thresholding method used in this research is Otsu.

b) Feature Extraction

Feature extraction aims to extract features from an image. Feature extraction is carried out in this study using contour-based and region-based methods. Feature data will be used for the classification process.

c) Classification

Classification aims to group objects into certain classes based on the attribute values associated with the observed object. Each particular object has certain characteristics, so that classification can distinguish an object from other objects. In this study, the classification methods used were KNN and SVM.

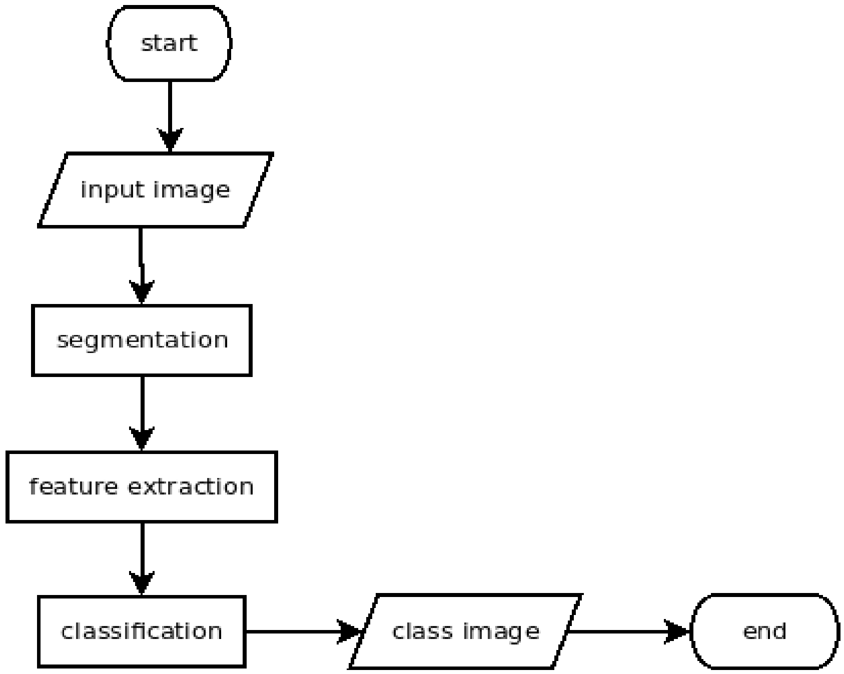


Figure 2: Desain System

2.3 Evaluation Metrics

To evaluate all methods, we use commonly classification metrics, which are sensitivity, specificity, precision factor, and accuracy determined using the following equations:

Sensitivity= TP/(TP+FP) (1)

Specificity= TP/(TP+FN) (2)

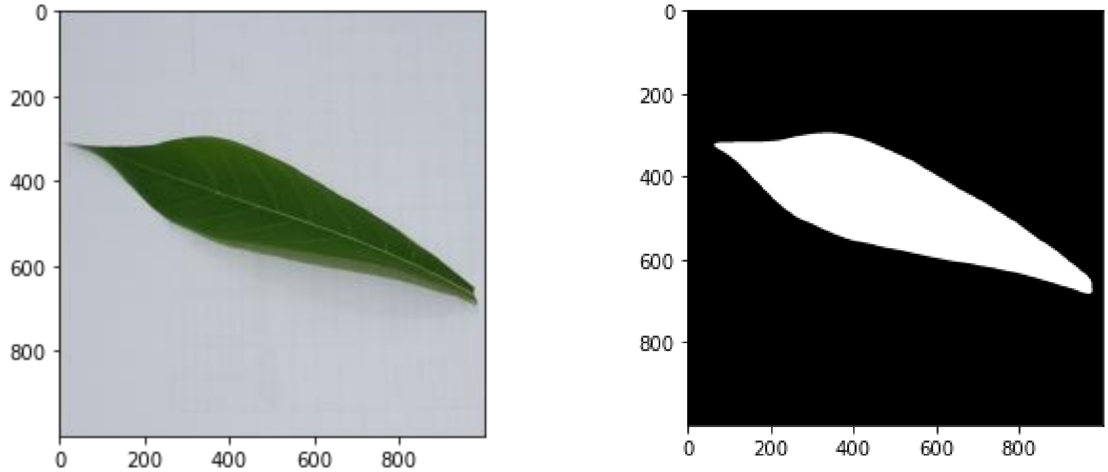
Precisoin factor = TP/(TP+FP) (3)

Accuracy= (TP+TN)/(TP+FNTP+FN+FP+TN) (4)

Where TP, FP, FN, and TN correspond to True Positives, False Positives, False Negatives and True Negatives. Sensitivity measures the proportion of correctly classified images to all classified images. Specificity measures how well the algorithm predicts other classes. While accuracy is to measure the total level of prediction the system makes.

3 RESULT

In this study, the trial aims to determine the class of the image entered by the user. This process begins by providing input in the form of an image of tubers into the system. The next process carried out is segmentation of the input image, as shown in Figure 3. After the segmentation process is carried out, the next process is to perform feature extraction. This feature extraction process is carried out to train the system. After feature extraction, the last process is classification, where this process aims to determine the class of the image entered by the user.



a) original image

b) image segmentation

Figure 3: Image segmentation

The results of the experiments carried out are shown in the form of a confusion matrix. Figure 4 shows the result of the confusion matrix using contour-based feature extraction and the SVM algorithm for classification. Based on Figure 4, it can be seen that the system being built has not been fully able to recognize the image of tubers. From 40 images, Singkong class can recognize as many as 39 images correctly, while 1 image is recognized in another class, namely Uwi class. The Talas class of 21 data was correctly recognized as many as 17 images, while 4 images were recognized as Uwi class. The Uwi class itself, from 33 image data, can be correctly recognized as many as 26 images, while 2 images are recognized as the Taro class and 5 are recognized as the Singkong class.

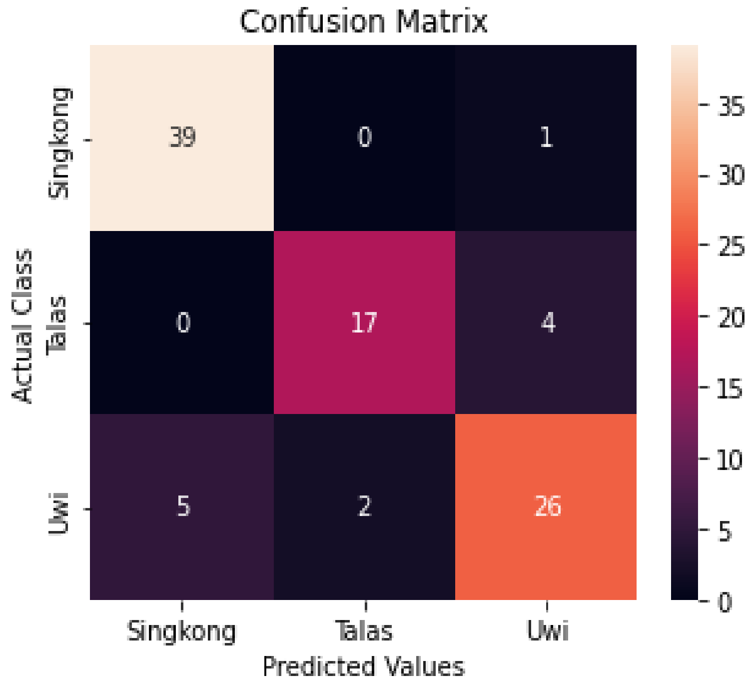


Figure 4: Contour based extraction feature with SVM classification

Figure 5 shows the results of the confusion matrix with contour-based features and the KNN algorithm for classification. Based on Figure 5, it can be seen that the system being built has not been fully able to recognize the image of tubers. From 40 data points, the Singkong class can recognize as many as 38 images correctly, while only 2 images are recognized in another class, namely the Uwi class. The Talas class of 21 data can be identified completely correctly. From 33 image data, the Uwi class can correctly recognize 28 images, and 5 are recognized as Singkong class. KNN gives maximum results when the neighbor value is 1.

Figure 6 shows the results of the confusion matrix with region-based features and the SVM algorithm for classification. Based on Figure 6, it can be seen that the system being built has not been fully able to recognize the image of tubers. From 40 data points, the Singkong class can recognize as many as 27 images correctly, and 13 images are recognized in another class, namely the Uwi class. From 21 image data , the Talas class 20 can be recognized correctly, and there is 1 data point that is recognized as the uwi class. The uwi class itself, from 33 image data, can be correctly identified as many as 21 images; 8 are recognized as the Talas class, and 4 are recognized as the Singkong class.

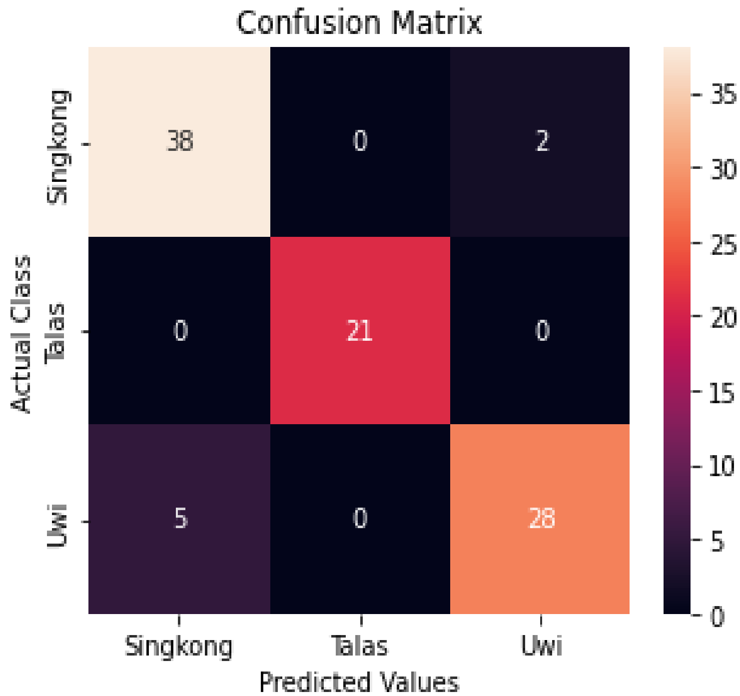


Figure 5: Contour based extraction featur with KNN classification

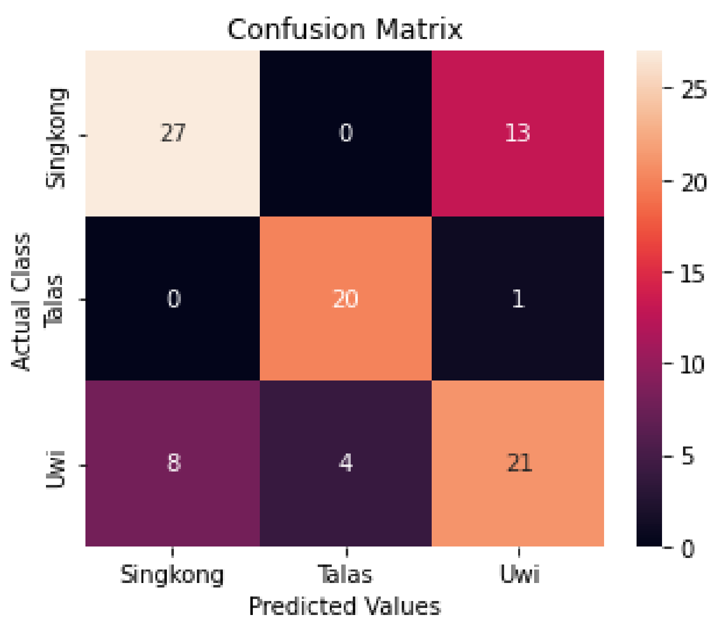


Figure 6: Region based extraction featur with SVM classification

Figure 7 displays the classification outcomes of the confusion matrix using the KNN method and region-based features. Figure 7 shows that the method under development has not been entirely successful in identifying the images of tubers. From 40 images, the Singkong class can be accurately identified in 32 images, the Talas class in 1, and the Uwi class in 7. The Talas class of 21 data can be precisely recognized. Up to 28 images out of 33 image data may be accurately identified as belonging to the Uwi class; one image is identified as belonging to the Talas class and four as belonging to the Singkong class.

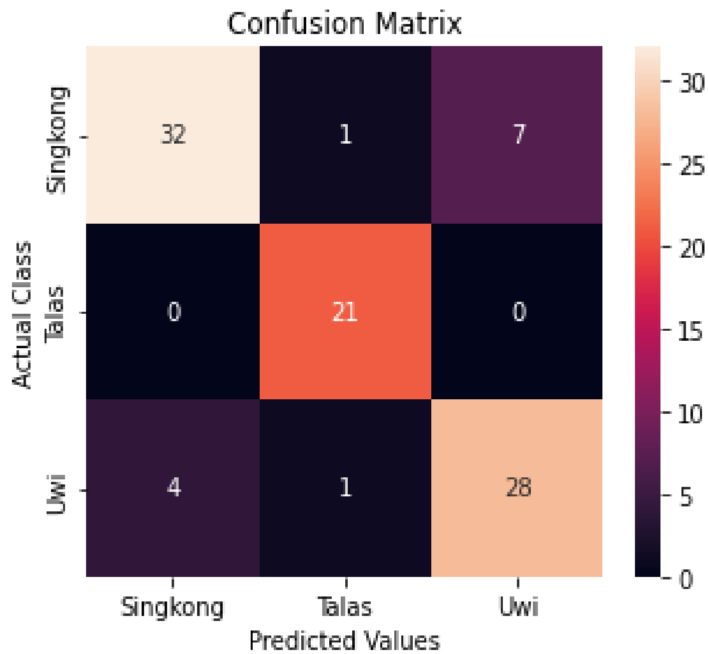


Figure 7: Region based extraction featur with KNN classification

The classification results of the confusion matrix using region-and contour-based features and the SVM technique are shown in Figure 8. Figure 8 demonstrates that the method under development has not been entirely successful in identifying the images of tubers. 40 data from the Singkong class can accurately identify up to 39 images, and 1 image is identified as being from the Uwi class. The 21 data points in the Talas class were correctly identified as 16. The Uwi class was identified by 4 data points, and the Singkong class by 1 data point. From 33 picture data, the Uwi class can accurately identify up to 26 images; three images are recognized as belonging to the Talas class, and four images are Singkong class.

The classification results of the confusion matrix using region- and contour-based features and the KNN algorithm are shown in Figure 9. Figure 9 illustrates how the system under development hasn't been able to adequately distinguish the image of tubers. Uwi class can accurately identify two images, while Singkong class using 40 data can correctly identify up to 38 images. It is possible to accurately identify the Talas class of 21 data. 27 images out of 33 image data may be correctly classified as belonging to the Uwi class; one image is identified as belonging to the Talas class and five as belonging to the Singkong class.

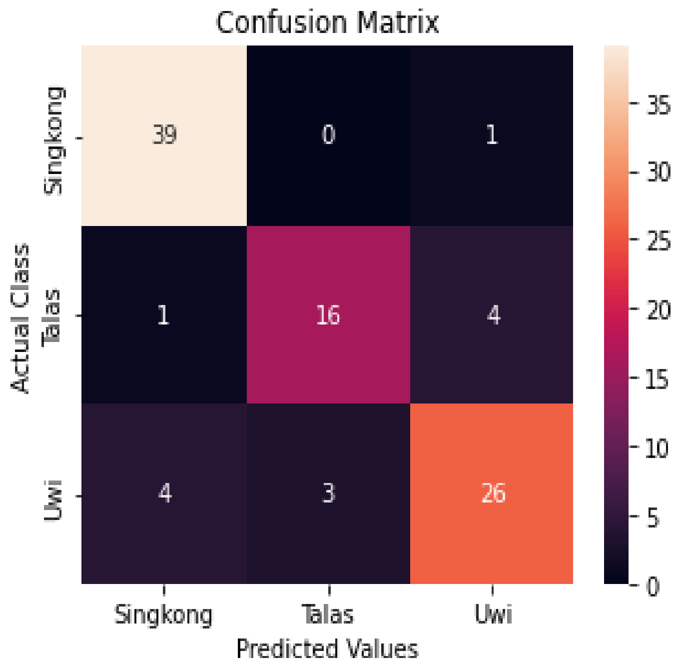


Figure 8: Contour based and region based extraction featur with SVM classification

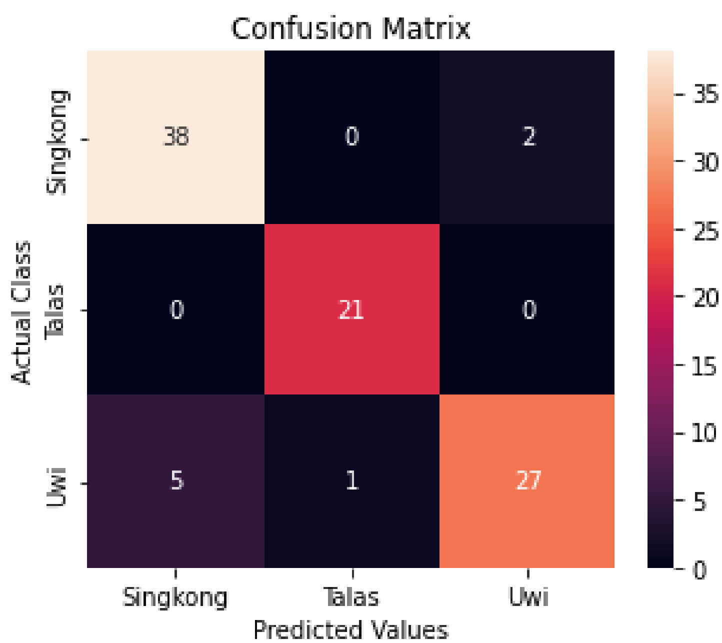


Figure 9: Contour based and region based extraction featur with KNN classification

Based on the trials that have been done, The system built provides the highest accuracy of 93% with contour based feature extraction and the KNN algorithm with a neighbor value of 1. Based on the results obtained, the KNN algorithm gives better results when compared to the SVM algorithm, and contour-based feature extraction gives better results when compared to region-based feature extraction. The use of contour features extraction also produces better results when compared to the combination of contour-based and region-based features extraction. For identification of leaves, KNN performs better than other classifiers such as Decision Tree, Nave Bayes Classifier, and Multi Support Vector Machine (Saleem et al., 2019).

4 CONCLUSION

Based on trials that have been carried out, the system provides the highest accuracy of 93% using contour-based feature extraction. The use of the KNN algorithm gives better results when compared to using the SVM algorithm.

5 ACKNOWLEDGEMENTS

A big thank you to the Islamic University of Lamongan, which has provided funding for this research.

REFERENCES

Amlekar, M. M., Gaikwad, A. T., Manza, R. R., & Yannawar, P. L. (2015). Leaf shape extraction for plant classification. *2015 International Conference on Pervasive Computing (ICPC)*, 1–4. https://doi.org/10.1109/PERVASIVE.2015.7087088

Anitha, G., Rao, E. G., & Kumar, V. P. (2021). Automatic Classification of Plant Leaves Along With Mean Filtering and Support Vector Machine. *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, 207–211. https://doi.org/10.1109/ICOEI51242.2021.9452995

Donesh, S., & Piumi Ishanka, UA. (2020). Plant Leaf Recognition: Comparing Contour-Based and Region-Based Feature Extraction. *2020 2nd International Conference on Advancements in Computing (ICAC)*, *1*, 369–373. https://doi.org/10.1109/ICAC51239.2020.9357152

Mittal, P., Kansal, M., & kaur Jhajj, H. (2018). Combined Classifier for Plant Classification and Identification from Leaf Image Based on Visual Attributes. *2018 International Conference on Intelligent Circuits and Systems (ICICS)*, 184–187. https://doi.org/10.1109/ICICS.2018.00046

Murat, M., Chang, S.-W., Abu, A., Yap, H. J., & Yong, K.-T. (2017). Automated classification of tropical shrub species: A hybrid of leaf shape and machine learning approach. *PeerJ*, *5*, e3792. https://doi.org/10.7717/peerj.3792

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Prasetyo, E., Adityo, R. D., Suciati, N., & Fatichah, C. (2018). Mango Leaf Classification with Boundary Moments of Centroid Contour Distances as Shape Features. *2018 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 317–320. https://doi.org/10.1109/ISITIA.2018.8711115

Saleem, G., Akhtar, M., Ahmed, N., & Qureshi, W. S. (2019). Automated analysis of visual leaf shape features for plant classification. *Computers and Electronics in Agriculture*, *157*, 270–280. https://doi.org/10.1016/j.compag.2018.12.038

Thanikkal, J. G., Dubey, A. K., & Thomas, M. T. (2018). Advanced Plant Leaf Classification Through Image Enhancement and Canny Edge Detection. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 1–5. https://doi.org/10.1109/ICRITO.2018.8748587

Wang, Z., Chen, L., Li, H., Li, J., Zhu, Y., & Wang, J. (2018). Comparative Study on Different PCNN Models in Plant Leaf Classification. *2018 IEEE International Conference on Information and Automation (ICIA)*, 560–564. https://doi.org/10.1109/ICInfA.2018.8812573

Yang, C. (2021). Plant leaf recognition by integrating shape and texture features. *Pattern Recognition*, *112*, 107809. https://doi.org/10.1016/j.patcog.2020.107809