Development of GUI for Malaysian herbs plant image identification

Suhaila Abd Halim¹ and Syafiqah Md Lazim²

¹²Centre of Mathematics Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

¹E-mail: suhaila@tm.ukm.edu.my

Abstract. In the previous era, sickness and diseases were treated using traditional herbs since medical technology had yet to be developed. Due to massive development in medicine, more herbs species are facing the risk of extinction due to not being well conserved. This paper describes the development of a graphical user interface (GUI) for identification of Malaysian Herbs Plant from image data using MATLAB. Besides, the GUI development will contribute to the database development of herbs species that are much needed to systematically stores relevant information. It can provide valuable knowledge for society. It is quite challenging for people including the herbs’ expert, to differentiate herbs’ plants since there are too many of them and limited knowledge and experiences. Fortunately, this can be overcome by using image processing approaches specifically image segmentation and classification. A set of image data consists of five types of Malaysian herbs plants is used in this study. The types are Rerama, Sirih, Mexican Mint, Belalai Gajah dan Senduduk. Sobel edge detection operator is a segmentation approach that works by computing the gradient of image intensity at each point and measuring the pixel value changes in horizontal and vertical directions by using two convolution masks. The shape and texture features are used in multiclass Support Vector Machine (SVM) model to classify the species of the herbs. Results show that performance of model are 95.56% for accuracy and sensitivity, 98.89% for specificity, 95.78% for precision, and 94.53% for Matthews correlation. The demonstration of image data shows the effectiveness of the developed GUI in classifying the five types of herbs.

1. Introduction
Traditional herbs have once become the main source for medicine in ancient days. In fact, some people are still practicing this way to treat some diseases or to maintain their health. However, over the years, more herbs species are being forgotten and in fact, some of them are in jeopardy of being extinct as the result from not being well conserved [1]. This indicates the needs of development of databases and preservation places for herbs species to document and gather the information regarding herbs plants systematically to provide valuable knowledge for the society.

Taking this as a part of the motivation for the study, image processing specifically image segmentation and classification are used to further support the development of databases and preservation places for Malaysian medicinal herbs plant, specifically. Image segmentation was carried by using edge detection method under gradient based method that detect the edges by finding the maximum and minimum in the first derivative of the image. Sobel operator is one of the gradient based edge operators that works by computing the gradient of image intensity at each point, giving the
direction of the largest possible increase from light to dark and rate of change in that direction. It also emphasizes in measuring the changes of pixel value in two directions which are horizontal and vertical directions, and it uses 2 convolution masks to represents both directions [2].

Edge detection method or specifically gradient based method has been used widely in image segmentation. Canny operator is undeniably having the most superior performance among all edge detection method [3]. However, if being analyzed closely, Canny has a very complex computation which makes it time consuming whereas on the other hand, Sobel has the advantage of being simple and at the same can produce fairly good results especially when being compared to other gradient based method [4]. Sobel also has been used as a base for a new set of filters that aimed to lessen the presence of false edges [5].

There is wide selection of features that can be extracted from leaves which can be comprised into two broad categories consists of geometric and visual features [6]. However, the selection of feature extraction method needs to be done carefully as it can affect the outcome of the extraction as well as the classification process [7]. Gray Level Co-occurrence Matrix (GLCM) is one of the texture features that extract the statistical measures. The GLCM is calculated from grayscale image with direction analysis of horizontal, vertical, and diagonal.

Classification process needs to be assisted by extracting features information from the leaves that can help in identifying the species of the respective herbs. Each leaf has its own significant features that differentiate them from one species to another that can be divided into geometric and visual features. In this study, image classification will be focused on support vector machine (SVM) classifier that was first developed by Vapnik and Cortes in the late of 1990s [8]. SVM classifies images by finding an optimal hyperplane which splits the instances spaces [7].

The classification by using support vector machine (SVM) has been receiving positive remarks over the years as it can produce high accuracy of classification even when to be compared to other classifiers [8]. On the contrast, not all type of dataset will yield the same quality of classification as the optimality of each classifier is differ from one another [9].

The selection of SVM in plant recognition application has been done numerous times with some of them emphasize the combination of feature extraction method such as the scale-invariant feature transform (SIFT) with SVM which produces remarkable classification results [10]. Apart from the differentiation of plant classes, the recognition of plant diseases by using SVM is one of the plant-related recognition problems that can be done with the assistance of several method such as principal component analysis (PCA) to lessen the number of significant data and yield well classification result [11].

The objectives of this study are to apply Sobel edge detector to extract object boundary and apply SVM based on selected shape and texture features. In addition, the model developed based on SVM is deployed in a MATLAB graphical user interface (GUI). The GUI composed all the processes in detecting types of herbs based on leaf image.

2. Support Vector Machine

The SVM is one of the supervised machine learning models. The SVM find the optimal separating hyperplane between classes. The output of a linear SVM is in equation (1).

\[ u = w \cdot x - b \]  

(1)

where \( w \) is the normal vector to the hyperplane and \( x \) is the input vector. The separating hyperplane is the plane \( u=0 \). The nearest points lie on the planes \( u = \pm 1 \). The margin \( m \) is as in equation (2).

\[ m = \frac{1}{\|w\|} \]  

(2)

The kernel nonlinearly maps samples into a higher dimensional space. There are 4 types of kernels for training and classification as in equation (3), (4), (5) and (6) respectively.

**Linear:** \( K(x,x_j) = x^T x_j \)  

(3)
Polynomial: \( K(x_i, x_j) = (\gamma \cdot x_i^T x_j + r)^d, \quad \gamma > 0 \) \tag{4}

Radial Basis Function (RBF): \( K(x_i, x_j) = \exp(-\gamma \cdot \|x_i - x_j\|^2), \quad \gamma > 0 \) \tag{5}

Sigmoid: \( K(x_i, x_j) = \tanh(\gamma \cdot x_i^T x_j + r), \quad \gamma > 0 \) \tag{6}

where \( \gamma, r \) and \( d \) are the kernel parameters. The \( \gamma \) is the parameter controlling the performance of the kernel which affects the behavior of SVM.

3. Methodology

There are seven main processes of methodology that are carried out in this study as shown in Figure 1. The first process is data collection. A set of herbs images are collected for this study purposes. Then proceed with preprocessing to enhance the quality of image data. Then, edge detection is completed by segmentation process using Sobel edge operator. Next, compilation of various features by using feature extraction is carried out. The overall data extracted from feature extraction process is used in the next process of image classification using multiclass SVM model. Then, the performance of the model is evaluated in performance evaluation. Lastly, the developed model is deployed in MATLAB GUI.

![Figure 1. Processes in methodology.](image_url)
The detail explanation is given in the next section and the implementation of the processes in methodology is carried out using MATLAB R2018a.

3.1. Data Collection
In this process, a dataset comprises of 30 random selected leaves for each of 5 species of herbs leaves which are Sirih, Mexican Mint, Rerama, Belalai Gajah and Senduduk are collected and used in this study. The leaf images are captured via Sony Alpha a6000 mirrorless camera. These herbs leaves are chosen due to their abilities and benefit to cure health issues as shown in Table 1 [12]-[15]. The original size of the image is 1616 x 1080 pixel.

| Class | Leaf Image | Name          | Scientific Name  | Benefit                                      |
|-------|------------|---------------|------------------|----------------------------------------------|
| 1     |            | Sirih         | Piper Betle      | Treating diabetes and digestion problem      |
| 2     |            | Mexican Mint  | Plectranthus Amboinicus | Stimulate urination as well as keeping kidney healthy |
| 3     |            | Rerama        | Christia vespertilionis | Anticancerous characteristics that can help to treat cancer |
| 4     |            | Belalai Gajah | Clinacanthus nutans | Treat cancer                                  |
| 5     |            | Senduduk      | Melastoma malabathricum | Ability to help in diarrhoea problem         |

3.2. Preprocessing
In this process, the original size of the image which is 1616 x 1080 pixel then reduced to 768 x 513 pixel that purposely aim to reduce the processing time. The original RGB color image is also then converted to grayscale. Thresholding process is applied to convert the image to binary image. A threshold value is required in this process. The value is varying depends on the input image.

3.3. Edge Detection
The image edge can be detected by taking the derivative of the intensity value across the image and the points where the derivative is maximum. In this study, the object edge is detection using Sobel
edge detection. The process of Sobel operator starts by forming $\Delta x$ and $\Delta y$ to form two convolution masks as in equation (7) and (8) followed by calculating the magnitude and direction of gradient:

$$\frac{\partial f(x,y)}{dx} = \Delta x = \frac{f(x+dx,y) - f(x,y)}{dx}$$

(7)

$$\frac{\partial f(x,y)}{dy} = \Delta y = \frac{f(x,y+dy) - f(x,y)}{dy}$$

(8)

In which $\Delta x$ highlights the edges in the horizontal direction and $\Delta y$ highlights the edges in vertical direction.

In discrete image, $dx$ and $dy$ is considered in terms of numbers of pixel between two points which is 1 pixel spacing at which pixel coordinates $(i,j)$. In which equation (7) and (8) become equation (9) and (10).

$$\Delta x = f(i+1,j) - f(i,j)$$

(9)

$$\Delta y = f(i,j+1) - f(i,j)$$

(10)

The magnitude measures the changes in the gradient at $(i,j)$ as in equation (11).

$$M = \sqrt{\Delta x^2 + \Delta y^2}$$

(11)

The gradient direction is given by equation (12).

$$\theta = \tan^{-1} \frac{\Delta x}{\Delta y}$$

(12)

3.4. Feature Extraction

There are 14 shape and texture features used in this study. Several texture features are chosen to be extracted by using an image analysis technique called as GLCM. The contents of GLCM are used to calculate texture feature by giving a measure of the variation in intensity at the pixel of interest. In detail, all of 7 shape features are of area, perimeter, major axis length, minor axis length, roundness, slimness, and smoothness and the other 7 texture features are contrast, correlation, energy, entropy, homogeneity, kurtosis and skewness. Table 2 illustrates the shape features and their definition used in this study.

| Feature          | Definitions                        |
|------------------|------------------------------------|
| Area             | The number of pixels of value 1 inside the object |
| Perimeter        | The distance of the line forming boundary of the object |
| Major Axis Length| The length of the major axis of the ellipse |
| Minor Axis Length| The length of the minor axis of the ellipse |
| Roundness        | The ratio of length width of leaf |
| Smoothness       | The flatness of leaf                |

The area and perimeter are extracted directly from the leaf whereas major and minor axis length are extracted by using major and minor ellipses around the leaf. The other three features are extracted using formula in equation (13), (14) and (15).

$$\text{Roundness} = \frac{4\pi A}{p^2}$$

(13)

$$\text{Slimness} = \frac{l_1}{l_2}$$

(14)
Smoothness = \frac{1}{1 + a} \quad (15)

where

- $A$ = area of the leaf.
- $P$ = perimeter of the leaf.
- $l_1$ = The major axis length of the leaf.
- $l_2$ = The minor axis length of the leaf.
- $a$ = total value of pixels on the leaf boundary.

The flatness of image leaf is extracted by using image analysis technique called as GLCM. It is used to tabulate the regularity of distinct combination of grayscale to co-occur whether in an image section or in the whole image. Table 3 shows the texture features with their definition.

### Table 3. Explanation and definition of texture features.

| Features    | Definition                                                                 |
|-------------|----------------------------------------------------------------------------|
| Contrast    | The intensity of a pixel and its neighbor over the image                   |
| Correlation | The linear dependency of grey levels of neighboring pixels                |
| Energy      | The uniformity of the image                                               |
| Entropy     | The measure of randomness that is used to characterize the texture of the input image |
| Homogeneity | The closeness of the distribution of elements in the GLCM to the GLCM diagonal |
| Kurtosis    | The measure of whether the data are heavy-tailed or light-tailed relative to the normal distribution |
| Skewness    | The lack of symmetry of leaf                                               |

Five of the texture features that consist of contrast, correlation, energy, entropy and homogeneity depend solely to the computation of GLCM. For both of kurtosis and skewness, the calculations are done in separate manner. The formulas for the measures are listed as in equation (16)-(22).

\[
\text{Contrast} = \sum_{i,j=0}^{N-1} P_{ij} (i - j)^2 \quad (16)
\]

\[
\text{Correlation} = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i - \mu)(j - \mu)}{\sigma^2} \quad (17)
\]

\[
\text{Energy} = \sum_{i,j=0}^{N-1} (P_{ij})^2 \quad (18)
\]

\[
\text{Entropy} = \sum_{i,j=0}^{N-1} -\ln(P_{ij}) P_{ij} \quad (19)
\]

\[
\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i - j)^2} \quad (20)
\]

\[
\text{Kurtosis} = \frac{E(x - \mu)^4}{\sigma^4} \quad (21)
\]

\[
\text{Skewness} = \frac{E(x - \mu)^3}{\sigma^3} \quad (22)
\]

where

\[
\mu = \sum_{i,j=0}^{N-1} iP_{ij}
\]
\[ \sigma^2 = \sum_{i,j=0}^{N-1} P_{ij} (i-\mu)^2 \]

\( P_{ij} \) = element \( i, j \) of the normalized symmetrical GLCM.

\( N \) = number of grey levels in the image.

\( \mu \) = the GLCM mean.

\( \sigma \) = the intensities variance of all references pixels in the relationship that contributed to GLCM.

3.5. Image Classification

In image classification, the supervised learning model of multiclass SVM using RBF kernel is applied. There are two phases involved that are training phase and testing phase. The calculated features from previous process are collected and divided the data to training and testing dataset. The training dataset consists of training patterns from features values and their corresponding class labels. In testing phase, the classifier decides the class from the sample based on features value. The dataset is divided into 70\% of samples for training and 30\% for testing. All the

3.6. Performance Evaluation

The output result of segmented image that has been going through classification process are analyzed by evaluating the result of classification rate. The classification results are first to be illustrated in a confusion matrix that consists of the value of actual class and predict class. Next, it is converted into multi confusion matrix that shows the value for true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Lastly, by using all the evaluation, 8 quality measures for each approach that consists of Accuracy, Sensitivity, Specificity, Precision, and Mathews Correlation Coefficient are explained in Table 4.

| Measure                        | Definition                                           |
|-------------------------------|------------------------------------------------------|
| Accuracy                      | The closeness of the measurement to a specific value. |
| Sensitivity                   | The correct predicted values that the value of actual class is true and the predicted class is also true. |
| Specificity                   | The correctly predicted negative values that the value of actual class is false and the predicted class is also false. |
| Accuracy                      | A ratio of correctly predicted positive observations to the total predicted positive observation. |
| Matthews Correlation Coefficient | The balanced measure that considers TP, TN, FP, and FN. |

The formula related to the measurements are as in equation (23) – (27).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (23)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (24)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (25)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (26)
\]

\[
\text{Mathews Correlation Coefficient} = \frac{TP(TN) - FP(FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (27)
\]
where

\[ TP = \frac{The \ correctly \ predicted \ positive \ values \ that \ the \ value \ of \ actual \ class \ is \ true \ and \ the \ value \ of \ predicted \ class \ is \ also \ true}{}
\]

\[ TN = \frac{The \ correctly \ predicted \ negative \ values \ that \ the \ value \ of \ actual \ class \ is \ false \ and \ value \ of \ predicted \ class \ is \ also \ false}{}
\]

\[ FP = \frac{When \ actual \ class \ is \ false \ and \ predicted \ class \ is \ true}{}
\]

\[ FN = \frac{When \ actual \ class \ is \ true \ but \ predicted \ class \ is \ false}{}
\]

3.7. Model Deployment in MATLAB GUI

Figure 2 shows the layout design of MATLAB GUI for classification model deployment that has been developed using MATLAB R2018a. MATLAB is used as it is the commonly used program for scientific computing [16]. Input image of herbs are used as input and displayed in GUI. The preprocessing starts with the conversion process of the RGB color image to a grayscale image. This is to ensure smaller data will be processed instead of colors and hence make processing much simpler. Next is to determine the threshold value of the herbs image. The values might be varying depends on the image. Then Sobel edge detection is applied to detect the leaf contour or edge. Features of object are extracted using shape and texture features. The features act as input for classification process. The multiclass SVM model is used to classify the herbs leaf as there are 5 types of classes. Lastly, the GUI displays the herbs image with it label.

Figure 2. Layout design of GUI for Malaysian herb leaf classification.

There are two axes in the GUI design which is to presents the input image and the other is to present the detected edge image. LOAD IMAGE button is used to load the input data of herbs image. Then, PROCESS IMAGE button is used to complete the preprocessing, edge detection processes. For IDENTIFICATION button, classification using SVM is applied by using training and testing dataset. Training dataset is used to train the image data. Once it is completed, testing dataset is loaded to test the data. The RESET button is to clear all the image data and the calculated features lastly the CLOSE button is to close the GUI.
4. Results and Discussion
As discussed earlier, there are 5 types of herbs leaves that are used in this study. Few examples of Sobel segmentation results for each class of herbs species are illustrated in Table 5.

| Class No | Leaf image | Binary Image | Detected Image Edge |
|----------|------------|--------------|---------------------|
| 1        | ![Image](image1) | ![Image](image2) | ![Image](image3) |
| 2        | ![Image](image4) | ![Image](image5) | ![Image](image6) |
| 3        | ![Image](image7) | ![Image](image8) | ![Image](image9) |
| 4        | ![Image](image10) | ![Image](image11) | ![Image](image12) |
| 5        | ![Image](image13) | ![Image](image14) | ![Image](image15) |

Table 5. Sample results.

Based on Table 5, these are few of examples of successfully segmented leaf images by using Sobel edge detection operator. The first column of the figure shows the original leaf image before undergoing segmentation process. The second column illustrates the results of leaf segmentation by using Sobel edge operator and the third column shows the images that are processed through thresholding. The figure is arranged in the manner of five examples for each class started with Class no. 1 (Sirih), Class no. 2 (Mexican Mint), Class no. 3 (Rerama), Class no. 4 (Belalai Gajah) and Class no. 5 (Senduduk).

Table 6 illustrates the performance measurement in percentage. It can be observed that the measurements are above 85% for all values. It is considered good measurement.
Table 6. Results of performance measurement.

| Measurement               | Performance value (%) |
|---------------------------|------------------------|
| Accuracy                  | 95.56                  |
| Sensitivity               | 95.56                  |
| Specificity               | 98.89                  |
| Precision                 | 95.78                  |
| Matthews Correlation      | 94.53                  |

Figure 3 shows a sample result of GUI for input image of Sirih. It shows the GUI able to detect and classify the leaf and label it correctly. In addition, the benefit of leaf is also highlighted as additional information to user.

![GUI for input image of Sirih](image)

**Figure 3.** Sample result of Sirih.

This GUI can lead to a possible collaboration with any agricultural-based organisations such as MARDI and MoA in which it is focusing in the recognition process of numerous plant species in order to gather the data in a systematic way. In line with this, in order to gain some profit, the development of a mobile application that is linked to these organisations also can be considered to be commercialised.

5. Conclusion

All of the herbs images have been successfully segmented by using Sobel first order derivative edge operator and classified by using multiclass SVM model where almost all of the images have been successfully classified according to the respective classes despite some cases of misclassification. The results show that the model has achieved fairly good accuracy rate of 95.56%. As for future work, different machine learning algorithms can be applied and hence provide other alternative algorithm as comparison with the existing method.
Acknowledgement
The authors would like to express our gratitude to Persatuan Sains Matematik Malaysia (PERSAMA) for the financial support in funding the conference fee and Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Shah Alam for the technical support in completing this study.

References
[1] Chaki J 2017 Template based Shape Feature Selection for Plant Leaf Classification International Journal of Engineering and Technology (IJET) Vol 9 No 3
[2] Shrivakshnan G and Chandrasekar C 2012 A Comparison of various Edge Detection Techniques used in Image Processing IJCSI International Journal of Computer Science Issues, 9(5), No 1, ISSN (Online): 1694-0814
[3] Muthukrishnan R and M Radha 2011 Edge Detection Techniques for Image Segmentation International Journal of Computer Science & Information Technology (IJCST) Vol 3, No 6
[4] Maini R and Aggarwal H 2009 Study and Comparison of Various Image Edge Detection Techniques International Journal of Image Processing (IJIP), (3), (1). Retrieved from 27 November, 2017
[5] Chaudhary A and Gulati T 2013 Segmenting digital images using edge detection International Journal of Application of Innovation in Engineering & Management, 2(5)
[6] Faizal A J, Suhardi H, Salehuddin S and Illiasak A 2013 Leaf Features Extraction and Recognition Approaches to Classify Plant Journal of Computer Science 9(10): 1295-1304
[7] Medjahed S A 2015 A Comparative Study of Feature Extraction. IJ. Image, Graphics and Signal Processing, 2015, 3, 16-23
[8] Kadhm M S and Alia K A 2015 Handwriting Word Recognition Based on SVM Classifier International Journal of Advanced Computer Science and Applications, Vol. 6, No. 11, 2015
[9] Joshi A, Bapna S and Chunduri S 2012 Comparison Study of Different Pattern Classifiers. Retrieved 18 August, 2017
[10] Poudel P, Kumar S, Philip V S, Kishore P and Roopashree S 2016 Robust Recognition and Classification of Herbal Leaves International Journal of Research in Engineering and Technology (5) (4).
[11] Sudha V P 2017 Feature Selection Techniques for the Classification of Leaf Diseases in Turmeric International Journal of Computer Trends and Technology 43(3)
[12] Handoo S 2021 What Are the Benefits and Side Effects of Betel Leaf?. Retrieved 13 February, 2021 from http://www.stylecraze.com/articles/medicinal-uses-of-betel-leaf/#gref
[13] 8 Health Benefits Of Indian Borage (Mexican Mint). (2020, October). In Organic Facts. Retrieved 13 February, 2021 from https://www.organicfacts.net/health-benefits/herbs-and-spices/indian-borage-mexican-mint.html
[14] Azman Z (2017, March). UPM Runs Stage Two of Anti-Cancerous Red Butterfly Wing Research. Retrieved 13 February, 2021 from http://www.upm.edu.my/news/upm_runs_stage_two_ofAnti_cancerous_red_butterfly_wing_research-25072
[15] Straits Rhododendron (Melastoma malabathricum). (2012, September). In Green Tropicals. Retrieved 13 February, 2021 from https://greentropicals.wordpress.com/2012/09/11/senduduk-putih-melastoma-malabathricum/
[16] Lu Y and Lu R 2020 Siritool: A Matlab Graphical User Interface for Image Analysis in Structuredillumination Reflectance Imaging for Fruit Defect Detection, 2020 American Society of Agricultural and Biological Engineers, Vol. 63(4): 1037-1047