

Identification of Herbal Leaf Types Based on Their Image Using First Order Feature Extraction and Multiclass SVM Algorithm

Rohmat Indra Borman
Faculty of Engineering and Computer
Science
Universitas Teknokrat Indonesia
Bandarlampung, Indonesia
rohmat_indra@teknokrat.ac.id

Farli Rossi
Faculty of Engineering and Computer
Science
Universitas Teknokrat Indonesia
Bandarlampung, Indonesia
farli@teknokrat.ac.id

Yessi Jusman
Department of Electrical Engineering,
Faculty of Engineering
Universitas Muhammadiyah
Yogyakarta, Yogyakarta, Indonesia
yjusman@umy.ac.id

Ashrani Aizzuddin Abd. Rahni
Department of Electrical, Electronic &
Systems Engineering
Universiti Kebangsaan Malaysia
Selangor, Malaysia
ashrani@ukm.edu.my

Syahrizal Dwi Putra
Faculty of Computer Science
Universitas Esa Unggul
Jakarta, Indonesia
syahrizal.dwi@esaunggul.ac.id

Arief Herdiansah
Department of Informatics Engineering
Universitas Muhammadiyah
Tangerang, Tangerang, Indonesia
arief_herdiansah@umt.ac.id

Abstract— One way to increase immunity and maintain immunity can be done by consuming herbal plants. This herbal medicine is empirically believed to be useful as a cultural treasure from generation to generation. All parts of the plant can be used as medicine, one of which is the leaves. However, most people do not know the herbal leaves. This herbal leaf can actually be recognized from the characteristics of its shape. This study aims to identify types of herbal leaves using first-order feature extraction and the Multiclass Support Vector Machine (Multiclass SVM) algorithm. First-order feature extraction is able to extract features using the parameters of mean, skewness, variance, kurtosis, and entropy. Meanwhile, Multiclass SVM identifies by obtaining the optimal line in separating the data set of two classes of two-dimensional space points in order to find the maximum hyperplane in separating the data points into classes so that they can be grouped. From the test results, the identification accuracy is an average of 76%. This shows that the algorithm has been able to identify, but needs improvement.

Keywords—image identification, first-order feature extraction, multiclass SVM

I. INTRODUCTION

In the midst of the Covid-19 pandemic that has hit the world, it is important for a person to increase their immunity. One way to increase immunity so that the risk of contracting can be minimized is to consume herbs or herbal plants. Since ancient times, herbs have been used as medicinal plants and are still used for their benefits today. Medicinal plants are all plants that have benefits to support health and maintain health or can treat a disease [1]. There are almost 10% of all vascular plants used as medicinal plants, and there are approximately 350,000 or almost half a million species [2]. These herbal remedies are empirically believed to be efficacy and efficacy and are passed down as cultural treasures from generation to generation through oral tradition [3]. The form of therapy using herbs is a therapy that utilizes plants or plants that are still fresh, or that have been dried [1]. Of a total of about 40 thousand types of traditional medicines that have been known in the world, 30 thousand are believed to be in Indonesia [4]. All parts of the plant can be used as medicine, one of which is the leaf. Leaves are one of the plant organs that grow from twigs, generally green (containing chlorophyll) and mainly function as a catcher of energy from sunlight for

photosynthesis. Based on clinical research, the leaves contain vitamins, minerals and antioxidants that naturally keep the body healthy [5]. In Indonesia, the leaves have been used as traditional medicine. There are several leaves recommended by the Indonesian Ministry of Health based on the Decree of the Minister of Health of the Republic of Indonesia Number HK.01.07/Menkes/187/2017 concerning the formulary of Indonesian traditional medicinal ingredients that can be used as medicine. From some of these leaves there are leaves that can increase immunity, including grass jelly leaves, katuk leaves, cumin leaves, basil leaves and betel leaves. But most people do not know the herbal leaves. This herbal leaf can actually be recognized from its characteristics. So we need a system that can identify and classify leaves based on images so that it is easier to provide information about these plant species. This can be overcome by using digital image processing.

The field of digital image processing studies how an image is formed, managed, and analyzed in order to obtain useful information from the image [6]. In its development, image processing can identify the type of an object by matching several characteristics of a particular image on the image that has been trained. Through the detection of certain objects the image can produce useful information [7]. The image is identified and then classified so that the pixels of an image can be grouped into several classes, so that each class can describe an entity that has recognizable characteristics [8][9]. Before an object can be recognized or distinguished from other objects, feature extraction is carried out first. The extracted features are then used as parameters or input values at the identification or classification stage [10]. Image feature extraction method that can be used is first-order feature extraction. First-order feature extraction performs an extraction based on the histogram characteristics of the image [11]. Image histogram is a diagram that describes the frequency distribution of pixel intensity values in an image. Some of the first order characteristic parameters include mean, skewness, variance, kurtosis, and entropy [12]. These component boundaries can be utilized as contribution to the ID calculation to perceive objects in the picture. Multiclass Support Vector Machine (Multiclass SVM) is one of the strategies utilized in the recognizable proof or order of a picture with an essential idea that utilizes a direct or

hyperplane work that can isolate the preparation information into two classes by amplifying the edge between the two classes [13]. So one might say that the objective in SVM is to attempt to track down the ideal hyperplane or classifier work that can isolate two arrangements of information from two unique classes [14].

Several studies related to feature extraction used first-order feature extraction, including research on the identification of flower species [15]. This paper performs a first-order feature extraction with the parameters of mean, skewness, variance, kurtosis, and entropy. This paper shows that the feature parameters used can be used as input in the identification algorithm to recognize objects in the image. Another study, regarding the identification of fungi with first-order feature extraction [16]. The test results for the mushroom image identification system showed a high level of accuracy of 82% with the first-order statistical feature extraction method. Then research related to the implementation of first-order feature extraction on omega-3 egg images [17]. In this paper, first-order feature extraction with parameter values of mean, standard deviation based on the test obtained a precision level of 87.93%, recall 96.22% and accuracy 85%. First-order feature extraction utilizes histogram equalization operations in extracting the features. Meanwhile, the use of the SVM Multiclass algorithm for identification and classification has shown good results. Such as research on image classification of meat types with the multiclass SVM algorithm [18]. This study resulted in the best recognition rate of 87.5%. SVM has high generalizability without any additional knowledge requirements, even with high dimensions of the input space. The next research is related to the prediction of apple leaf disease by applying Multiclass SVM [17]. In this paper, Multiclass SVM recognizes the type of disease from the original leaf images among 500 images with an accuracy of 96%. Then another study was to identify the quality of tomatoes using multi class SVM [18]. Based on the use of multiclass SVM in classifying healthy or unhealthy conditions in tomatoes, the accuracy rate reaches 99.1%.

In this paper, the identification of herbal leaf species will be carried out using first-order feature extraction using the mean, skewness, variance, kurtosis, and entropy characteristics which are then used as input in the Multiclass SVM identification algorithm. Multiclass SVM is used to map data points to a high dimensional space to obtain a reciprocal linear separation between each of the two classes.

II. METHODOLOGY

This research consists of steps that will be carried out to complete the identification of herbal leaf species based on their image with first-order feature extraction and identification with the Multiclass SVM algorithm. These stages can be seen in Figure 1 below.

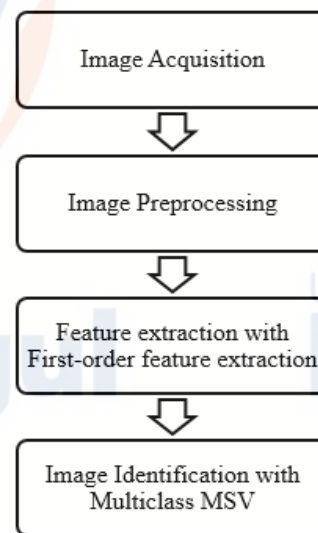


Fig. 1. Research Stages

A. Image Acquisition

The first stage is the image acquisition process of Indonesian herbal leaves which are used for data processing. The herbal leaf data used is the leaf recommended by the Ministry of Health of the Republic of Indonesia based on the Decree of the Minister of Health of the Republic of Indonesia Number HK.01.07/Menkes/187/2017 concerning the Formulary of Indonesian Traditional Medicinal Ingredients that Can be Used as Medicines. From some of these herbal leaves there are leaves that can increase immunity, including grass jelly leaves, katuk leaves, cumin leaves, basil leaves and betel leaves. The process of taking pictures by taking pictures of medicinal plants with the same light level, then conducting data acquisition from the data that has been obtained. At this stage the information is isolated into two sections, specifically:

1. Training image data used to determine the standard pattern of each herbal leaf.
2. Testing image data which will then be used as test data for system work in identifying types of herbal leaves

The training data used are 20 for 6 types of herbs, so the total training images are 120 images. For the test images used on each herbal leaf as many as 15 images, then the total test images are 75 images.

B. Image Preprocessing

The next stage is preprocessing which includes cropping the image and converting the original image to grayscale. Grayscale images are pixel colors that are in the range of black and white gradations. This is done to make it easier to extract features from objects in the image.

C. Feature Extraction with First-order feature extraction

The next stage is feature extraction, where the stage will be carried out to extract features from objects in the image that will be recognized or distinguished from other objects. The extracted features are then used as input parameters at the identification stage. The general characteristics used to identify one or more objects in the image are the size, position or location, and the orientation or angle of inclination of the object. Feature extraction used is first order feature extraction. First-order feature extraction is a feature extraction method based on the histogram characteristics of the image. Some of the first order characteristic parameters include mean,

skewness, variance, kurtosis, and entropy. The mean (μ) feature shows the size of the dispersion of an image. The mean (μ) can be calculated using the following formula:

$$\mu = \sum_{n=0}^N f_n P(f_n) \quad (1)$$

Where, f_n is grayscale level, $P(f_n)$ is histogram value.

The variance feature (σ_2) is the variance in the histogram value. The variance (σ_2) can be calculated by the following formula:

$$\sigma^2 = \sum_{n=0}^N (f_n - \mu)^2 P(f_n) \quad (2)$$

Where, f_n is grayscale level, $P(f_n)$ is histogram value. While is the mean value that has been calculated previously.

Skewness (α_3) Shows the level of relative inclination of the histogram curve of an image. Skewness (α_3) can be calculated by the following formula:

$$\alpha_3 = \frac{1}{\sigma^3} \sum_{n=0}^N (f_n - \mu)^3 P(f_n) \quad (3)$$

Where, f_n is grayscale level, $P(f_n)$ is histogram value, is the mean value that has been calculated previously. While α_3 is the standard deviation of the gray intensity value.

The Kurtosis feature (α_4) indicates the relative sharpness of the histogram curve of an image. Kurtosis (α_4) can be calculated by the following formula:

$$\alpha_4 = \frac{1}{\sigma^4} \sum_{n=0}^N (f_n - \mu)^4 P(f_n) - 3 \quad (4)$$

Where, f_n is grayscale level, $P(f_n)$ is histogram value, is the mean value that has been calculated previously. While α_4 is the standard deviation of the gray intensity value.

Entropy feature (H) Indicates a measure of the irregularity of the shape of an image. Entropy (H) can be calculated by the following formula:

$$H = - \sum_{n=0}^N P(f_n)^2 \log p(f_n) \quad (5)$$

D. Image Identification with Multiclass SVM

From the component extraction that has been done, these element boundaries can be utilized as contribution to the distinguishing proof calculation to perceive objects in the picture. The distinguishing proof calculation utilized is Multiclass Support vector machine. This calculation is the strategy utilized in the distinguishing proof and characterization of a picture with a fundamental idea that utilizes a straight or hyperplane work that can isolate the preparation information into two classes by boosting the edge between the two classes [13]. So one might say that the objective in SVM is to attempt to track down the ideal hyperplane or classifier work that can isolate two arrangements of information from two unique classes [14]. The best isolating hyperplane between the two classes can be found by estimating the edge of the hyperplane and tracking

down its greatest point. Edge is the distance between the hyperplane and the nearest design from each class. This nearest design is known as a help vector. The line between the examples shows the best hyperplane, which is found solidly in the centred of the two classes [18]. The basic form of SVM separation seeks to obtain an optimal line in separating between data sets of two classes of two-dimensional space points. The generalization carried out aims to find the maximum hyperplane in separating data points to potential classes into n-dimensional space. Generalization is defined as the ability of a method to classify a pattern, which does not include the data used in the learning phase of this method. The closest distances to the hyperplane are called Support Vectors.

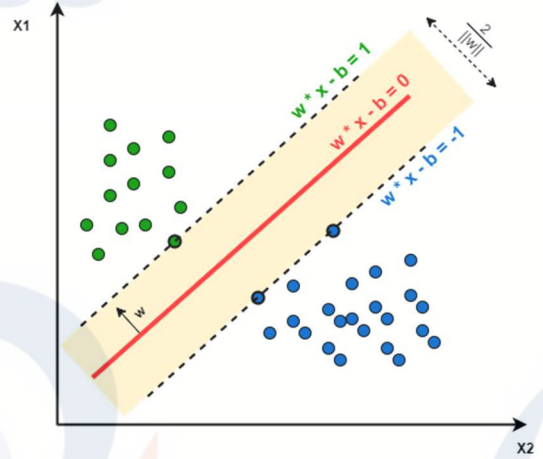


Fig. 2. Overview of Support Vector

In the image above, Support Vectors are 3 dots, namely 2 blue and 1 green. These points are on a line that is scattered and separated by a hyperplane. There are two main approaches to Multiclass SVM. First, find and combine several functions that separate the two class classification problems to solve the multi-class classification problem. The second, directly uses all data from all classes in one optimization problem formulation [4]. One against all (OAA) and one against one, OAO [12]. With this technique it is important to discover the worth of the separator work where each capacity is prepared with information from the two classes. To prepare information from the i -th and j -th classes, it is done by solving the following 2-class classification problems:






$$\min_{w^{ij}, b^{ij}, \xi^{ij}} \frac{1}{2} (w^{ij})^T w^{ij} + C \sum_t \xi_t^{ij} \quad (6)$$

Where w^{ij} is the margin between 2 data groups i and j , b^{ij} parameter i and j to find the value, ξ^{ij} slack variables i and j and C are parameter values.

III. RESULTS AND DISCUSSION

The herbal leaves used in the identification were grass jelly leaves, katuk leaves, cumin leaves, basil leaves and betel leaves. The process of image acquisition by taking pictures by photographing medicinal plants with the same light level, then performs data acquisition from the data that has been obtained. The training data used were 20 for 6 types of herbs, so the total training images were 120 images. For the test images used on each herbal leaf as many as 10 images, then the total test images are 60 images. The following is a sample image of herbal leaves used for training data.

TABLE I. HERBAL LEAF TYPE IMAGE

Photo	Herbal Leaf Type
	Grass Jelly Leaves
	Katuk Leaves
	Cumin Leaves
	Basil Leaves
	Betel Leaves



(a) RGB Image (b) Grayscale Image

Fig. 3. Convert RGB image to grayscale image

After the RGB picture is changed over into a grayscale picture, then, at that point, highlight extraction is done, where these means will be completed to separate elements from objects in the picture that will be perceived or recognized from different articles. Element extraction utilized is first request highlight extraction. Several first-order characteristic parameters include mean, skewness, variance, kurtosis, and entropy. The following is the implementation in Matlab for feature extraction source code using first-order feature extraction:

```
H = imhist(Img)';
H = H/sum(H);
I = [0:255];
MEAN = I*H';
ENTROPY = -H*log2(H+eps)';
VARIANCE = (I-FeatureMEAN).^2*H';
SKEWNESS = (I-MEAN).^3*H'/VARIANCE^1.5;
KURTOSIS = (I-MEAN).^4*H'/VARIANCE^2-3;
```

The output of the calculation of the mean, skewness, variance, kurtosis, and entropy will be displayed in the table properties of the Matlab GUI. Figure 4 below is a sample of the results of feature extraction using first-order feature extraction on the Matlab GUI.

	Feature	Value
1	Mean	107.1308
2	Entropy	7.4557
3	Variance	2127.1496
4	Skewness	0.58262
5	Kurtosis	0.093105

(a) Grayscale Image (b) First Order Feature Extraction Results

Fig. 4. Feature extraction process from grayscale image

Identification of herbal leaf species based on image using first-order feature extraction and Multiclass SVM algorithm implemented using Matlab application. To identify the image, the first step is to convert the RGB image into a grayscale image. This is done to make it easier to extract features from objects in the image. The conversion to grayscale helps in the extraction process, because the information in this image is not more complex than the color image. Below is the source code used to convert an RGB image into a grayscale image:

```
Img = handles.Img;
Img = rgb2gray(Img);
axes(handles.axes2)
```

The grayscale output will be displayed in the axis properties. Figure 3 below is the result of an RGB image that has been converted to grayscale.

After the feature extraction process, the next step is image identification. The Multiclass SVM algorithm uses a linear or hyperplane function that can separate the training data into two classes by maximizing the margin between the two classes. The best dividing hyperplane between the two classes can be found by measuring the hyperplane margin and finding its maximum point. The basic form of SVM separation seeks to obtain an optimal line in separating data sets of two classes of two-dimensional space points. The generalization carried out aims to find the maximum hyperplane in separating data points into potential classes into n-dimensional space. Generalization is defined as the ability of a method to classify a pattern, which does not include the data used in the learning phase of this method. The following is the source code for implementing the Multiclass SVM algorithm in Matlab.


```

while ( (classes~=1) && (itr<=length(u)) &&
size(C,2)>1 && cond>0)
    c1=(C==u(itr));
    newClass=c1;
    svmStruct = svmtrain
    (T,newClass,'
    kernel_function','rbf');
    classes =
svmclassify(svmStruct,tst);

```

From the source above, the while loop is used to generalize to multiclass SVM. Furthermore, so that the dataset can be optimally reduced, the dataset used is reduced. The following is the source code used to reduce the dataset.

```

for i=1:size(newClass,2)
    if newClass(1,i)==0;
        c3(k,:)=T(i,:);
        k=k+1;

```

Figure 5 below is the result of implementing the Multiclass SVM algorithm in the Matlab GUI application.

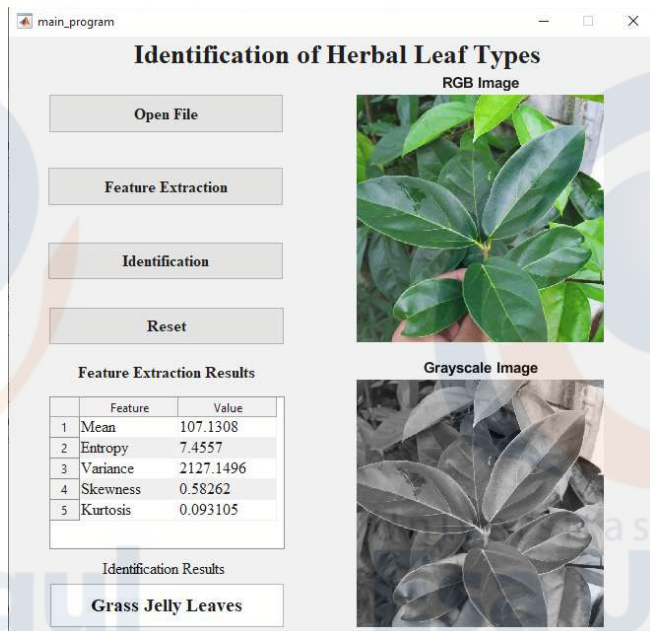


Fig. 5. GUI with Matlab on first-order feature extraction implementation and Multiclass SVM algorithm

To see the effectiveness of the algorithm used, the identification results will be tested for accuracy. Accuracy is used to determine the closeness of a test result or the average test result to the actual value. To test the level of accuracy is done by calculating with the following equation:

$$Accuracy = \frac{CP}{TP} * 100\% \quad (7)$$

CP is the number of correct predictions, namely the number of test data records predicted by the class using the identification method and the results are the same as the actual class. While TP is the total number of predictions, where the total number predicted by the class. The test data used on each herbal leaf are 15 images, so the total test images are 75 images. Each herbal leaf was tested and its accuracy calculated. The results of the percentage of accuracy can be seen in the graph of Figure 6 below.

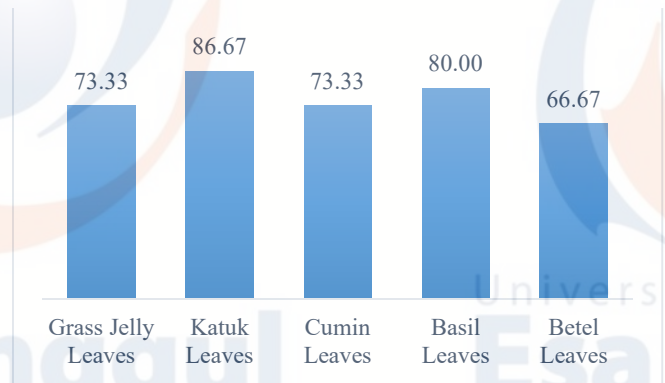


Fig. 6. Identification accuracy percentage chart

The graph above shows that the average accuracy of the accuracy test results obtained from the identification accuracy percentage is 76%. Meanwhile, the average identification error is 14%. First-order feature extraction performs feature extraction based on the histogram characteristics of the image. The histogram of the image pays attention to the frequency distribution of the pixel intensity values in an image. Then the accuracy will be high if the extracted image displays one image with a clear image. If the extracted leaf image is clustered and unclear, it will be difficult to perform feature extraction. The test data used for each type of herbal leaf were 15 images. The percentage classification accuracy criteria are classified into: Good, with a value of 76%-100%; Enough, with a value of 56%-75%; Less Good, with a value of 40%-55%, while Not Good, has a value of less than 40% [19]. If it is seen that the average accuracy obtained from the percentage of identification accuracy is 76%, it shows that the accuracy is quite good, but does not yet have a high value. This is caused by several factors, among others: (1) the amount of training data and test data is still very small; (2) the shape of the picture of the herbal leaf that looks almost the same; (3) for identification with clustered leaf images, the algorithm has difficulties. For further research, improvements are needed, including: (1) increasing the amount of training data and test data; (2) using deep learning algorithms to get better feature extraction and identification.

IV. CONCLUSION

This paper tries to identify types of herbal leaves based on images using first-order feature extraction and the Multiclass SVM algorithm with an average identification accuracy of 76%. First-order feature extraction is able to extract features using the parameters mean, skewness, variance, kurtosis, and entropy. These feature parameters are used as input in the Multiclass SVM algorithm to recognize objects in the image. SVM identifies by obtaining the optimal line in separating the data set of two classes of two-dimensional space points in order to be able to find the maximum hyperplane in separating the data points into classes that can be separated by their characteristics.

ACKNOWLEDGMENT

Special thanks are addressed to the Universitas Teknokrat Indonesia, Universitas Muhammadiyah Yogyakarta, Universiti Kebangsaan Malaysia, Universitas Esa Unggul and Universitas Muhammadiyah Tangerang as well as those who have assisted in the implementation of the service.

REFERENCES

- [1] H. Ivanz *et al.*, "Recommended Medicinal Plants as Source of Natural Products: A Review," *Digit. Chinese Med.*, vol. 1, no. 2, pp. 131–142, 2018.
- [2] E. Salmerón-Manzano, J. A. Garrido-Cardenas, and F. Manzano-Agugliaro, "Worldwide Research Trends on Medicinal Plants," *Int. J. Environmental Res. Public Heal.*, vol. 17, no. 3376, pp. 1–20, 2020.
- [3] A. N. Shikov, A. N. Tsitsilin, O. N. Pozharitskaya, V. G. Makarov, and M. Heinrich, "Traditional and Current Food Use of Wild Plants Listed in the Russian Pharmacopoeia," *Forntiers Pharmacol.*, vol. 8, no. 841, pp. 1–15, 2017.
- [4] Z. Salim and E. Munadi, *Info Komoditi Tanaman Obat*. Jakarta: Badan Pengkajian dan Pengembangan Perdagangan Kementerian Perdagangan Republik Indonesia, 2017.
- [5] M. Silalahi, "Diversity of medicinal plants in homegardens in Tanjung Julu village, North Sumatra, Indonesia," *Int. J. Biol. Res.*, vol. 4, no. 1, pp. 78–82, 2016.
- [6] S. Dix and P. Müller, "Digital Image Processing Methods for The Evaluation of Optical Anisotropy Effects in Tempered Architectural Glass Using Photoelastic Measurements," *Glas. Struct. Eng.*, vol. 6, no. 1, pp. 3–19, 2021.
- [7] A. Mulyanto, R. I. Borman, P. Prasetyawan, W. Jatmiko, and P. Mursanto, "Real-Time Human Detection and Tracking Using Two Sequential Frames for Advanced Driver Assistance System," in *ICICOS 2019 - 3rd International Conference on Informatics and Computational Sciences: Accelerating Informatics and Computational Research for Smarter Society in The Era of Industry 4.0, Proceedings*, 2019.
- [8] R. Ponnusamy, S. Sathyamoorthy, and K. Manikandan, "A Review of Image Classification Approaches and Techniques," *Int. J. Recent Trends Eng. Res.*, vol. 3, no. 3, pp. 1–5, 2017.
- [9] P. Prasetyawan, I. Ahmad, R. I. Borman, A. Ardiansyah, Y. A. Pahlevi, and D. E. Kurniawan, "Classification of the Period Undergraduate Study Using Back-propagation Neural Network," in *Proceedings of the 2018 International Conference on Applied Engineering, ICAE 2018*, 2018.
- [10] S. Ahmed Medjahed, "A Comparative Study of Feature Extraction Methods in Images Classification," *Int. J. Image. Graph. Signal Process.*, vol. 7, no. 3, pp. 16–23, 2015.
- [11] M. Wati, Haviluddin, N. Puspitasari, E. Budiman, and R. Rahim, "First-order Feature Extraction Methods for Image Texture and Melanoma Skin Cancer Detection," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019.
- [12] M. A. Wani, H. F. Bhat, and T. R. Jan, "Position Specific Scoring Matrix and Synergistic Multiclass SVM for Identification of Genes," in *17th IEEE International Conference on Machine Learning and Applications Position*, 2018, pp. 1192–1196.
- [13] A. N. I. Masazhar and M. M. Kamal, "Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM Classifier," in *International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, 2017.
- [14] S. Zgarni and A. Braham, "Classification of Bearing Fault Detection Using Multiclass SVM : A comparative study," in *15th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2018, pp. 888–892.
- [15] M. S. Kara, F. A. Wara, and M. F. Rumba, "Identifikasi Jenis Bunga Menggunakan Ekstraksi Ciri Orde Satu," *J. Creat. (Inovasi dan Kreasi dalam Teknol. Informasi)*, vol. 6, pp. 22–28, 2019.
- [16] S. Gustina, A. Fadlil, and R. Umar, "Sistem Identifikasi Jamur Menggunakan Metode Ekstraksi Ciri Statistik Orde 1 dan Klasifikasi Jarak," *Techno.Com*, vol. 16, no. 4, pp. 378–386, 2017.
- [17] F. Liantoni and A. A. Santoso, "Penerapan Ekstraksi Ciri Statistik Orde Pertama Dengan Ekualisasi Histogram Pada Klasifikasi Telur Omega-3," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 953–958, 2018.
- [18] N. Neneng, K. Adi, and R. R. Isnanto, "Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM)," *J. Sist. Inf. Bisnis*, vol. 1, pp. 1–10, 2016.
- [19] R. I. Borman and Y. Purwanto, "Impelementasi Multimedia Development Life Cycle pada Pengembangan Game Edukasi Pengenalan Bahaya Sampah pada Anak," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 119, 2019.