Identification of Soursop Leaves Image Based On RGB Color Features Extraction and Gabor Filter Using Backpropagation Artificial Neural Networks

Laily Muntasiroh[1,*] Hendriansyah[2] Fitriyani[3]
[1,2,3] Universitas Muhammadiyah Semarang, Semarang, Central Java 50273, Indonesia
lailymuntasiroh@unimus.ac.id

Abstract. Soursop leaves have a myriad of uses, including boosting the immune system, helping fight cancer cells. The leaf, which has the scientific name Annona Muricata, has become the target of residents who choose alternative paths in natural medicine. Soursop leaves grow abundantly across Indonesia, from lowlands up to 1,000 meters above sea level. Generally, identifying the type of leaf is done visually. Moreover, the shape and colour are very similar to avocado leaves. Therefore, a study was conducted to identify soursop leaves based on RGB color feature extraction. The method chosen in this study was backpropagation neural network as one of the best algorithms that perform well in image identification. In this study, two backpropagation neural network models were utilized to compare the performance between the training and testing systems. The results obtained the best combination of parameter settings which were 6 for the input layers and 5 for the output layer. And the best combination for the Gabor filter was 45º for orientation angle and 8 for wavelength. Thus, this study was able to achieve an average accuracy of 90%. For future works, we recommend using other feature extraction techniques, such as geometry, KK-Nearest Neighbour, histogram, and PCA to then be compared with Gabor filter extraction.

Keywords: Backpropagation, Neural Network, Gabor Filter, RGB Filter.

1. Introduction

Soursop fruit or (Annona Muricata Linn). This fruit includes fruit that is easy to grow in various places. The name soursop itself comes from the Dutch language, namely Zuurzak, which means sour bag . Soursop is an annual plant, which means it can grow all year round, provided that the water in the soil is sufficient for the growth process. Soursop fruit plants are spread in Indonesia, ranging from lowland areas with a fairly hot climate to highlands with a fairly cold climate with an altitude of 1000 meters above sea level. Soursop is classified into[1]:

Kingdom : Plantae

I. Yustar Afif and R. Nindyo Sumarno (eds.), *Proceedings of the 2nd Lawang Sewu International Symposium on Engineering and Applied Sciences (LEWIS-EAS 2023)*, Advances in Engineering Research 234,
https://doi.org/10.2991/978-94-6463-480-8_17

Division : Spermatrophyta

Subdivision : Angiosperms

Class : Dicotyledonea

So far, soursop fruit plants are cultivated for their fruit. Soursop has very high vitamin C tha's way it's widely consumed by many people. In addition, soursop fruit also contains carbohydrates and addition, the content of soursop fruit which is rich in vitamin C can increase stamina and facilitate breast milk.

It is not just the fruit that has benefits, soursop leaf flowers also have efficacy for treating bronchitis and coughs. Additionally, soursop fruit seeds are effective for preventing and treating constipation, flatulence, dizziness, nausea, vomiting, as well as getting rid of head lice, skin parasites, and worms. The bark is utilized for treating asthma, cough, hypertension, and seizures. Equally important are the leaves - soursop leaves are efficacious for treating heart disease, diabetes, and cancer as they contain antioxidant compounds. Soursop leaves have abundant antioxidants consisting of polyphenol compounds, saponins and bioflavonoids. The advantage of these antioxidants is protecting the body from the threats of free radicals and diseases that constantly lurk in the body [2].

The development of computing science is currently experiencing very rapid progress. One of its applications is an automatic object recognition method using a computer, by taking and processing information based on image data (4 from leaf disease journals). Image processing is expected to be an option in the introduction or identification of soursop leaves. One method of image recognition is the RGB color feature extraction method. The leaf image is extracted to get the RGB value with the value taken is the average value of all pixels[3]–[11] .

Various studies were carried out using image processing techniques combined with artificial neural networks. Gray Level Co-Occurrence Matrix (GLCM) feature extraction on surface images of tropical fruit leaves is used as input for training artificial neural networks for the identification process (identification of tropical fruit). Extraction of RGB color features is used as input for artificial neural networks to identify types of medicinal plants for hypertension. (Identification of medicinal plants for hypertension). By using different objects and different methods, this research will carry out the process of identifying soursop leaves using RGB color feature extraction. This research use backpropagation neural network to identification process.

The research method utilized is a backpropagation neural network. Backpropagation is one of the training techniques in artificial neural network algorithms. Aside from backpropagation, there is also forward propagation. Both types of training techniques, backpropagation and forward propagation, iterate to determine the weight and bias values to reduce errors. Artificial neural network algorithms are very adept at handling complex patterns. In research on plant species identification using ANN, a success rate of 93.6% was obtained. Research on tea leaf image identification yielded a 95.45% success rate. Iris identification attained a 100% success rate. Identification of medicinal plant types can reach a rate of 72%. And in other studies, the percentage of success was also obtained with an average above 90% [3]–[7], [9], [12]–[19].

2. Methods

The identification method used is a neural network backpropagation. The backpropagation algorithm is one of the ANN techniques which has 2 (two) phases, namely forward and backward. These two phases are carried out iteratively to change the bias and weight values to reduce errors. Artificial neural networks have 2 (two) stages, namely training and testing. In outline the research stages of this system can be shown in the figure. The required data is 90 (ninety) images of soursop leaves and avocado leaves consisting of 70 images for data training and 20 images for data testing.

2.1. Input Data Image

The data used in this research is divided into 2, namely training data and testing data. The training data uses 50 images of soursop leaves and 50 images of avocado leaves. Soursop leaves were obtained from gardens in Semarang, Indonesia. Meanwhile, avocado leaves were obtained from Simo, Central Java. Images of soursop leaves and avocados were taken using the Samsung M12 camera device with camera specifications, 2 Mega Pixel macro camera, 5 Mega Pixel ultra wide camera, 48 Mega Pixel main camera, and 2 Mega Pixel depth camera. In the process of taking pictures using only sunlight and using a light yellow paper background.

2.2. Resize Images

The first process to be performed is image resizing which is an operation to change a pixel size in a digital image. In this process, the images of soursop leaves and avocado leaves from the camera device are converted to a size of 256 x 256 pixels to minimize computing time so that the computer can work fast.

2.3. Extraction of RGB Color Features

The next step is to extract the Red Green Blue color characteristics to take the value of each Red Green Blue component. This research pays attention to color content as an input value. However, in general color feature extraction does not pay attention to the size or location of the image but only pays attention to the pixels in the image.

2.4. Gabor Filter Feature Extraction

The computation of the six parameters in the Gabor Filter involves two calculation orders. In this study, a first-order Gabor filter is employed, initiating the calculation of the entropy formula value for an irregular shape by utilizing the image histogram value.

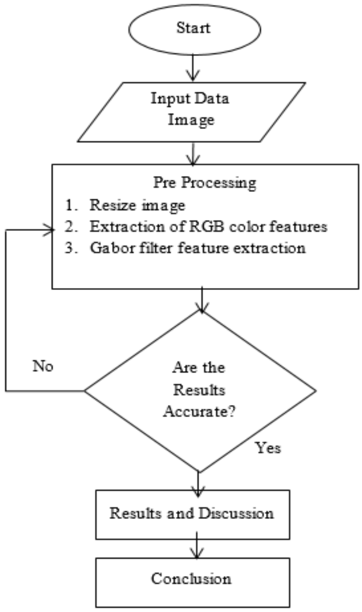

Fig 1. Flow Chart

2.5. Training

The training procedure involves utilizing a constructed network architecture designed to acquire the most effective weights. Once these optimal weights are achieved, they are stored for subsequent testing, specifically for the identification of soursop leaf images. During the testing phase, users input test data to conduct evaluations, yielding results in the form of extracted RGB color features. The outcomes of these tests are then compared to predefined targets to assess the system's proficiency in comprehending the training data it has been exposed to.

Table 1. Data Image

No	Data Image	Name
1		Avocado Leaves
2		Avocado Leaves
3		Soursop leaves

No	Data Image	Name
4		Soursop leaves
5		Soursop leaves

2.6. Testing

Following the completion of the training phase, evaluations are conducted on the trained network. The implementation of neural network algorithms involves the execution of forward propagation, succeeded by backward propagation, for all training patterns. This step signifies the process of training a neural network, achieved by adjusting the weights. Upon the conclusion of the training process, a subsequent stage involves addressing new problems, known as the mapping or testing phase. In this phase, the network's efficacy in recognizing novel input patterns is assessed through a testing process that involves comparing the output values with the desired targets to determine the success rate.

3. Result and Discussion

In this research, the Lavenberg Marquardt network's training function, known as train-lm, is employed. Trainlm is not only the default for the forward propagation training function but is specifically utilized for problems involving approximation functions where the network possesses fewer than one hundred weights and precise approximation is essential. However, it should be noted that trainlm demands substantial memory resources during the training process. Table 1 outlines the training parameters for the artificial neural network using the Lavenberg-Marquardt (trainlm) network training function. These parameters require adjustment prior to training to achieve optimal training outcomes.

The artificial neural network's architecture comprises 6 inputs, including R, G, B values, entropy, mean, and variance, with 2 hidden layers and 2 outputs representing images of soursop leaves and avocado leaves. The first hidden layer incorporates 30 neurons utilizing the tantig activation function, while the second hidden layer comprises 15 neurons employing the logsig activation function. The output layer utilizes the purelin activation function with 2 neurons. Additionally, neurons in the hidden layers use the logsig activation function, and the output layer deploys the purelin activation function with 2 neurons.

Table 2. Artificial neural network architectural design

Characteristic	Specification
Neuron Input Layer	6
Neuron Hidden Layer 1	20
Neuron Hidden Layer 2	15
Neuron Output Layer	1
Activation function	Tansig-logsig-purelin
Algorithm Training	Traingdx
Error Maximum	10^{-5}
Set Epoch Maximum	1000

Table 3. Testing Scenario

No.	Gabor Filter		Accuracy	
	Wavelength	Θ	Training	Testing
1	2	45°	100%	0%
2	3	45°	100%	0%
3	4	45°	100%	0%
4	5	45°	100%	35%

No.	Gabor Filter		Accuracy	
	Wavelength	ϴ	Training	Testing
5	6	45°	100%	60%
6	7	45°	100%	70%
7	8	45°	100%	90%
8	2	90°	100%	0%
9	3	90°	100%	0%
10	4	90°	100%	5%
11	5	90°	100%	5%
12	6	90°	100%	60%
13	7	90°	100%	55%
14	8	90°	100%	65%
15	2	135°	100%	0%
16	3	135°	100%	0%
17	4	135°	100%	0%
18	5	135°	100%	5%
19	6	135°	100%	45%
20	7	135°	100%	55%
21	8	135°	100%	60%

The examination was conducted utilizing a network architecture derived from the optimal training outcomes identified in Table 2. A set of 20 data points was employed for the test, comprising 10 test images of soursop leaves and 10 test images of avocado leaves. Notably, this dataset had not been utilized in any prior stages of network training. The primary aim of the testing phase is to assess the network's ability to generate outputs in alignment with the specified targets, particularly when presented with different input data-data that has not been introduced to the network during the training process.

The tests encompassed a systematic exploration of Gabor filter parameters, involving alterations in wavelength and angle orientation. This iterative process aimed to identify the most effective combination, as illustrated in Table 2 showcasing the various test scenarios. The objective was to evaluate the network's proficiency in recognizing diverse input data patterns.

Table 4. Results

Image	**Results**
Image 1-1	Recognized
Image 1-2	Recognized

Image	Results
Image 1-3	Recognized
Image 1-4	Recognized
Image 1-5	Recognized
Image 1-6	Recognized
Image 1-7	Recognized
Image 1-8	Recognized
Image 1-9	Recognized
Image 1-10	Not Recognized
Image 2-1	Recognized
Image 2-2	Recognized
Image 2-3	Recognized
Image 2-4	Recognized
Image 2-5	Recognized
Image 2-6	Recognized
Image 2-7	Recognized
Image 2-8	Not Recognized
Image 2-9	Recognized
Image 2-10	Recognized

The validation scores for the ninth test are as follows:

$$\text{Validation} = \frac{18}{20} \text{ x } 100\ \% = 90\% \tag{1}$$

The validation metric is calculated by dividing the number of accurately classified images by the total number of test images and then multiplying the result by 100%. This value is derived from 18 instances of testing, each conducted under different conditions. The average time taken for these tests is 4.1 seconds, utilizing the training set. The training and testing processes were repeated 18 times, as indicated in Table 2.

The testing phase involves the evaluation of 20 images, excluding those present in the training dataset. The outcomes, based on varying test parameters, are detailed in Table 4.

4. Conclusion

The conclusion that can be drawn based on the research results is that the image processing method was successfully carried out through the RGB color feature extraction process. The use of feature extraction using a Gabor filter of order 1 is carried out through 6 (six) parameters, namely the values of R, G, B, mean, variance and entropy. The performance of the Gabor filter and the artificial neural network are able to work well as the training and testing process is relatively fast, which is around 4 seconds. So that the obtained accuracy value is 90%. Future research is expected to be able to use other feature extraction, such as geometry, KK-Nearest, histogram, and PCA to then be compared with Gabor filter extraction.

References

1. N. Kurniasih, M. Kusmiyati, Nurhasanah, R. Puspita Sari, and R. Wafdan, "Potensi Daun Sirsak (*Annona Muricata Linn*), Daun Binahong (*Anredera Cordifolia (Ten) Steenis*), Dan Daun Benalu Mangga (<i>Dendrophthoe Pentandra) Sebagai Antioksidan Pencegah Kanker," *J. Istek*, vol. 9, no. 1, pp. 162–184, 2015, [Online]. Available: https://journal.uinsgd.ac.id/index.php/istek/article/view/182
2. N. Kurniasih, M. Kusmiyati, Nurhasnah, R. Puspita Sari, and R. Wafdan, "Potensi daun sirsak, daun binahong, dan daun benalu sebagai antioksidan pencegah kanker," *J. Istek*, vol. 9, no. 1, pp. 162–184, 2015.
3. R. I. SAA Bowo, A Hidayatno, "Analisis deteksi tepi untuk mengidentifikasi pola daun," *Undergrad. thesis, Diponegoro Univ.*, pp. 1–7, 2011.
4. R. Rahmadewi, E. Purwanti, and V. Efelina, "Identifikasi Jenis Tumbuhan Menggunakan Citra Daun Berbasis Jaringan Saraf Tiruan (Artificial Neural Networks)," *J. Media Elektro*, vol. VII, no. 2, pp. 38–43, 2018, doi: 10.35508/jme.v0i0.427.
5. Y. Sari, M. Alkaff, and M. Arif Rahman, "Identifikasi Penyakit Tanaman Ubi Kayu Berdasarkan Citra Daun Menggunakan Metode Probabilistic Neural Network (PNN)," *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 1–9, 2021, doi: 10.31603/komtika.v5i1.4605.
6. N. Sivi Anisa and T. Herdian Andika, "Sistem Identifikasi Citra Daun Berbasis Segmentasi Dengan Menggunakan Metode K-Means Clustering," *Aisyah J. Informatics Electr. Eng.*, vol. 2, no. 1, pp. 9–17, 2020, doi: 10.30604/jti.v2i1.22.
7. C. C. Kusumadewa and S. Supatman, "Identifikasi Citra Daun Teh Menggunakan Metode Histogram untuk Deteksi Dini Serangan Awal Hama Empoasca," *JMAI (Jurnal Multimed. Artif. Intell.*, vol. 2, no. 1, pp. 27–36, 2018, doi: 10.26486/jmai.v2i1.71.
8. A. Novitasari, E. P. Purwandari, and F. F. Coastera, "Identifikasi Citra Daun Tanaman Jeruk Dengan Local Binary Pattern Dan Moment Invariant," *J. Inform. dan Komput.*, vol. 3, no. 2, pp. 76–83, 2018.
9. I. Jamaliah, "Identifikasi Jenis Daun Tanaman Obat Hipertensi Berdasarkan Citra Rgb Menggunakan Jaringan Syaraf Tiruan," *Penelit. Ilmu Komput. Sist. Embed. dan Log.*, vol. 5, no. 1, pp. 1–11, 2017.
10. B. Sitohang and A. Sindar, "Analisis Dan Perbandingan Metode Sobel Edge Detection Dan Prewit Pada Deteksi Tepi Citra Daun Srilangka," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 314–322, 2020, doi: 10.32672/jnkti.v3i3.2511.

11. M. A. Hasan, Y. Riyanto, and D. Riana, "Grape leaf image disease classification using CNN-VGG16 model," *J. Teknol. dan Sist. Komput.*, vol. 9, no. 4, pp. 218–223, 2021, doi: 10.14710/jtsiskom.2021.14013.
12. R. R. Isnanto, "Identifikasi Iris Mata Menggunakan Tapis Gabor Wavelet Dan Jaringan Syaraf Tiruan Learning Vector Quantization (LVQ) oleh : Program Studi Sistem Komputer Fakultas Teknik Universitas Diponegoro (Eye Iris Identification Using Tapis Gabor Wavelet And Neura," *Univ. Diponegoro, Semarang*, pp. 1–12, 2009.
13. admi syarif, A. R. TANJUNG, R. ANDRIAN, and F. R. LUMBANRAJA, "Implementasi Metode Ekstraksi Fitur Gabor Filter dan Probablity Neural Network (PNN) untuk Identifikasi Kain Tapis Lampung," *J. Komputasi*, vol. 8, no. 2, pp. 1–9, 2020, doi: 10.23960/komputasi.v8i2.2641.
14. R. A. Setiawan, R. R. Isnanto, and A. Hidayatno, "Menggunakan Metode Tapis Gabor 2-D Dan Jaringan Syaraf Tiruan Learning Vector Quantization (Lvq)".
15. M. A. Agmalaro, A. Kustiyo, and A. R. Akbar, "Identifikasi Tanaman Buah Tropika Berdasarkan Tekstur Permukaan Daun Menggunakan Jaringan Syaraf Tiruan," *J. Ilmu Komput. dan Agri-Informatika*, vol. 2, no. 2, p. 73, 2013, doi: 10.29244/jika.2.2.73-82.
16. R. R. Isnanto *et al.*, "Teknik Pencocokan Template Tapis Gabor," *Telekomnika*, vol. 5, no. 1, pp. 1–8, 2007.
17. M. A. Wiratama and W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 1, p. 1, 2022, doi: 10.23887/janapati.v11i1.45282.
18. A. Mustakim, I. Santoso, and A. A. Zahra, "Pengenalan Ekspresi Wajah Manusia Menggunakan Tapis Gabor 2-D Dan Support Vector Machine (Svm)," *Transient*, vol. 6, no. 3, p. 232, 2017, doi: 10.14710/transient.6.3.232-238.
19. S. Sudianto, A. D. Sripamuji, I. Ramadhanti, R. R. Amalia, J. Saputra, and B. Prihatnowo, "Penerapan Algoritma Support Vector Machine dan Multi-Layer Perceptron pada Klasisifikasi Topik Berita," *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 11, no. 2, pp. 84–91, 2022, [Online]. Available: https://ejournal.undiksha.ac.id/index.php/janapati/article/view/44151