



## Research article

# Recognizing the Types of Beans Using Artificial Intelligence

Nur Nafiyah<sup>a,\*</sup>, Endang Setyati<sup>b</sup>, Yosi Kristian<sup>c</sup>, Retno Wardhani<sup>d</sup>

<sup>a,d</sup> Department of Informatics, Universitas Islam Lamongan, Jl. Veteran 53A Lamongan, East Java, 62211, Indonesia

<sup>b,c</sup> Department of Informatics, Institute of Sciences and Technology Integrated Surabaya, Jl. Ngagel Jaya Tengah No.73-77 Gubeng, Surabaya, East Java, 60284, Indonesia

email: <sup>a,\*</sup> [mynaiff26@gmail.com](mailto:mynaiff26@gmail.com), <sup>b</sup> [esetyati@gmail.com](mailto:esetyati@gmail.com), <sup>c</sup> [yosi@stts.edu](mailto:yosi@stts.edu), <sup>d</sup> [retzno@yahoo.com](mailto:retzno@yahoo.com),

\* Correspondence

## ARTICLE INFO

### Article history:

Received 28 January 2023

Revised 6 May 2023

Accepted 31 October 2023

Available online 21 November 2023

### Keywords:

Optimal Feature;

Selection Feature;

Correlation;

Bean Leaves;

Backpropagation;

Please cite this article in IEEE style as:

N. Nafiyah, E. Setyati, Y. Kristian, and R. Wardhani, "Recognizing the Types of Beans Using Artificial Intelligence," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 9, no. 2, pp. 134-143, 2023.

## ABSTRACT

Many studies have previously addressed the recognition of plant leaf types. The process of identifying these leaf types involves a crucial feature extraction stage. Image feature extraction is pivotal for distinguishing the types of objects, thus demanding optimal feature analysis for accurate leaf type determination. Prior research, which employed the CNN method, faced challenges in effectively distinguishing between long bean and green bean leaves when identifying bean leaves. Therefore, there is a need to conduct optimal feature analysis to correctly classify bean leaves. In our research, we analyzed 69 features and explored their correlations within various image types, including RGB, L\*a\*b, HSV, grayscale, and binary images. The primary objective of this study is to pinpoint the features most strongly correlated with the recognition of bean leaf types, specifically green bean, soybeans, long beans, and peanuts. Our dataset, sourced from farmers' fields and verified by experienced senior farmers, consists of 456 images. The most highly correlated feature within the bean leaf image category is STD b in the L\*a\*b image. Furthermore, the most effective method for leaf type recognition is Neural Network Backpropagation, achieving an accuracy rate of 82.28% when applied to HSV images.

Register with CC BY NC SA license. Copyright © 2023, the author(s)

## 1. Introduction

Utilizing image processing and artificial intelligence can effectively classify leaf types. Numerous studies have focused on identifying leaf species and leaf diseases. In this study, we aim to evaluate the most informative features within leaf images for recognizing leaf species and analyzing effective image types. In previous research, the analysis of optimal features for distinguishing between various leaf types has been emphasized, as it facilitates the automation of leaf type recognition [1]. The referenced research [1] specifically seeks to determine the best feature correlations among 22 image features to identify leaf types.

Identifying plant species based on their leaves is a crucial task for botany experts, but it can be challenging for the general public to distinguish plant types solely from their leaves. Therefore, there is a need for a fast and precise system to classify plant types based on leaf characteristics [2], [3], [4], [5], [6]. In a study conducted by [2], the utilization of shape features derived from image morphology and Support Vector Machines (SVM) to differentiate between leaf types yielded an impressive accuracy of 94.5%. A study conducted by [3] carried out analysis of various features, including shape, Fourier, and morphology, to assess the accuracy of the Artificial Neural Network (ANN) method in classifying leaf types.

Several CNN methods, such as in [7], [8], [9] and Machine Learning approaches such as ANN, SVM, and *k*-NN [10], [11] have been employed to evaluate the identification of plant leaf species and the most suitable classification of leaf diseases. These methods make use of diverse features, including color,

texture, and shape features [12], [13], [14]. In some studies, researchers have further advanced leaf type recognition by proposing innovative features such as LBP-HF (Local Binary Patterns Histogram Fourier) [15]. This feature extraction technique derives information from the Fourier histogram values, resulting in an impressive accuracy of 95%. In another study [16], the combination of LBP (Local Binary Patterns) and ELM (Extreme Learning Machine) features has demonstrated the ability to achieve a remarkable accuracy rate of 98% in identifying leaf types.

Several studies focused on the identification of leaf disease types or the characterization of different leaf types have sought to achieve optimal feature extraction. This includes the proposal of leaf geometry features [17] and the introduction of features derived from wavelet and leaf edges [18]. In addition to conducting extensive feature analysis, some studies aim to identify the types of images that are most effective in recognizing different leaf types [19]. It is worth noting that different types of images can yield varying discriminant information [20]. For example, several RGB, L\*a\*b, and HSV images that have undergone Gaussian or Wavelet filters are analyzed to determine which image type is optimal for classification.

Many studies have explored feature analysis and feature selection, recognizing the pivotal role of feature extraction in identifying image objects. In previous research, leaf types were identified using the CNN method, but there was room for improvement in distinguishing between long and mung bean leaves [21]. Therefore, we propose to conduct an in-depth analysis and selection of 69 features and seek correlations for each feature across various image types [20], including L\*a\*b, HSV, grayscale, and binary, with the goal of optimizing the recognition of bean leaf types. Unlike previous studies, which primarily focused on feature analysis using a single image type [1], our approach utilizes several image types, namely RGB, L\*a\*b, HSV, grayscale, and binary [20]. The objective of this study is to analyze and select 69 image features within the categories of RGB, L\*a\*b, HSV, grayscale, and binary for the purpose of identifying different bean leaf types.

## 2. Materials and Methods

The data for this research was collected directly from farmers' fields and subsequently verified by senior farmers. The data comprises images of bean leaves, as illustrated in Figure 1. There are a total of 456 data, divided into 228 for training and 228 for testing, as outlined in Table 1. Specifically, Figure 1(a) depicts various types of green bean leaves, Figure 1(b) shows soybean leaves, Figure 1(c) features long bean leaves, and Figure 1(d) displays bean leaves.

Table 1. Dataset

No	Leaf type	Training	Testing	Total
1	Green beans	51	51	102
2	Soybeans	66	66	132
3	Long beans	59	59	118
4	Peanuts	52	52	104
Total		228	228	456

This study aims to identify the features that yield the highest accuracy in distinguishing different types of bean leaves. Various types of color images, including RGB, L\*a\*b, HSV, grayscale, and binary, were employed to retrieve features. These features encompass a range of statistical attributes, such as mean, standard deviation, variance, skewness, kurtosis, entropy, contrast, energy, correlation, homogeneity, area, perimeter, metric, major axis, minor axis, eccentricity, and circularity. The complete list of features can be found in Table 2. Figure 2 provides an illustration of the leaf image utilized for feature retrieval. The original RGB image was transformed into L\*a\*b, HSV, grayscale, and binary formats, with the aim of extracting the features as listed in Table 2. The research process flow is visualized in Figure 3. It commences with the conversion of the initial RGB image into L\*a\*b, HSV, grayscale, and binary formats, followed by the extraction of features. Subsequently, the features from RGB, L\*a\*b, HSV, grayscale, and binary images were analyzed by calculating their correlations with the various types of bean leaves. Furthermore, training and classification trials were carried out using the Backpropagation and SVM methods, with a thorough evaluation at the final stage (as depicted in Figure 4). Equations 1 and 2 are the formulas for calculating the correlation between each feature or attribute and the type of bean leaf, as well as the accuracy formula.

$$r_{x,y} = \frac{(n\sum x_i y_i) - (\sum x_i \sum y_i)}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}} \tag{1}$$

$$accuracy = \frac{\sum y_i = \bar{y}_i}{N} \tag{2}$$

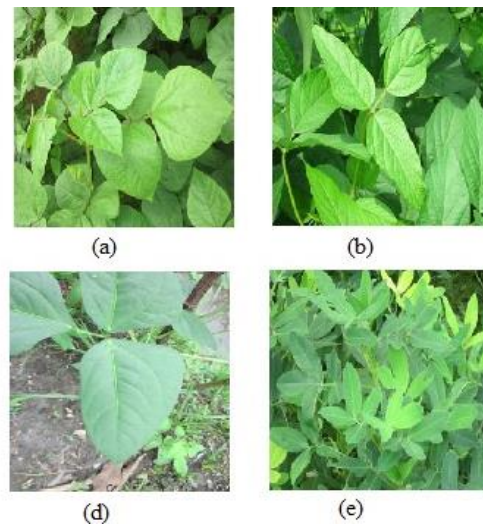


Fig. 1. Bean leaf types

Eq.(1)  $r(x, y)$  describes is the correlation of each image feature ( $x$ ) with the bean leaf type class ( $y$ ). Eq.(2)  $y$  indicates the actual leaf species class, and  $\bar{y}$  represents the prediction. We considered a total of 17 RGB image features, encompassing parameters such as average R, average G, average B, standard deviation R, standard deviation G, standard deviation B, variance R, variance G, variance B, skewness R, skewness B, kurtosis R, kurtosis G, kurtosis B, entropy R, and entropy B. Additionally, we included 17  $L^*a^*b$  image features, consisting values such as average L, average a, average b, standard deviation L, standard deviation a, standard deviation b, variance L, variance a, variance b, skewness L, skewness a, skewness b, kurtosis L, kurtosis a, kurtosis b, entropy a, and entropy b. We included 18 HSV image features, featuring H mean, S mean, V mean, standard deviation H, standard deviation S, standard deviation V, variance H, variance S, variance V, skewness H, skewness V, kurtosis H, kurtosis S, kurtosis V, entropy H, entropy S, and entropy V. Furthermore, we utilized ten grayscale image features: contrast, energy, correlation, homogeneity, entropy, mean, standard deviation, variance, skewness, and kurtosis. In addition, we considered seven binary image features: area, perimeter, metric, major axis, minor axis, eccentricity, and circularity.

Table 2. Feature extraction

Type Image Feature	RGB	$L^*a^*b$	HSV	Grayscale	Biner
1	Mean R	Mean L	Mean H	Contrast	Area
2	Mean G	Mean a	Mean S	Energy	Perimeter
3	Mean B	Mean b	Mean V	Correlation	Metric
4	STD R	STD L	STD H	Homogeneity	Major Axis
5	STD G	STD a	STD S	Entropy	Minor Axis
6	STD B	STD b	STD V	Mean	Eccentricity
7	Var R	Var L	Var H	STD	Circularity
8	Var G	Var a	Var S	Var	
9	Var B	Var b	Var V	Skew	
10	Skew R	Skew L	Skew H	Kurtosis	
11	Skew G	Skew a	Skew S		
12	Skew B	Skew b	Skew V		
13	Kurtosis R	Kurtosis L	Kurtosis H		
14	Kurtosis G	Kurtosis a	Kurtosis S		
15	Kurtosis B	Kurtosis b	Kurtosis V		
16	Entropy R	Entropy a	Entropy H		
17	Entropy B	Entropy b	Entropy S		
18	-	-	Entorpy V		

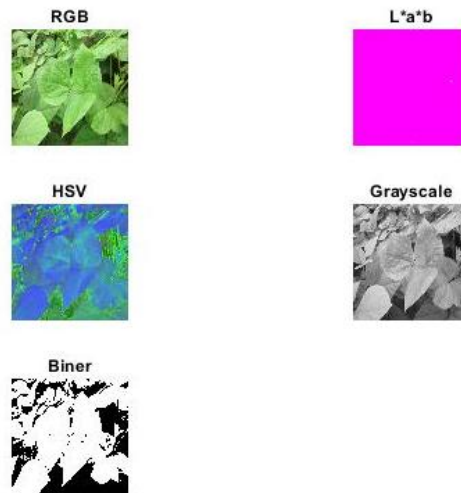


Fig. 2. Image of leaf types

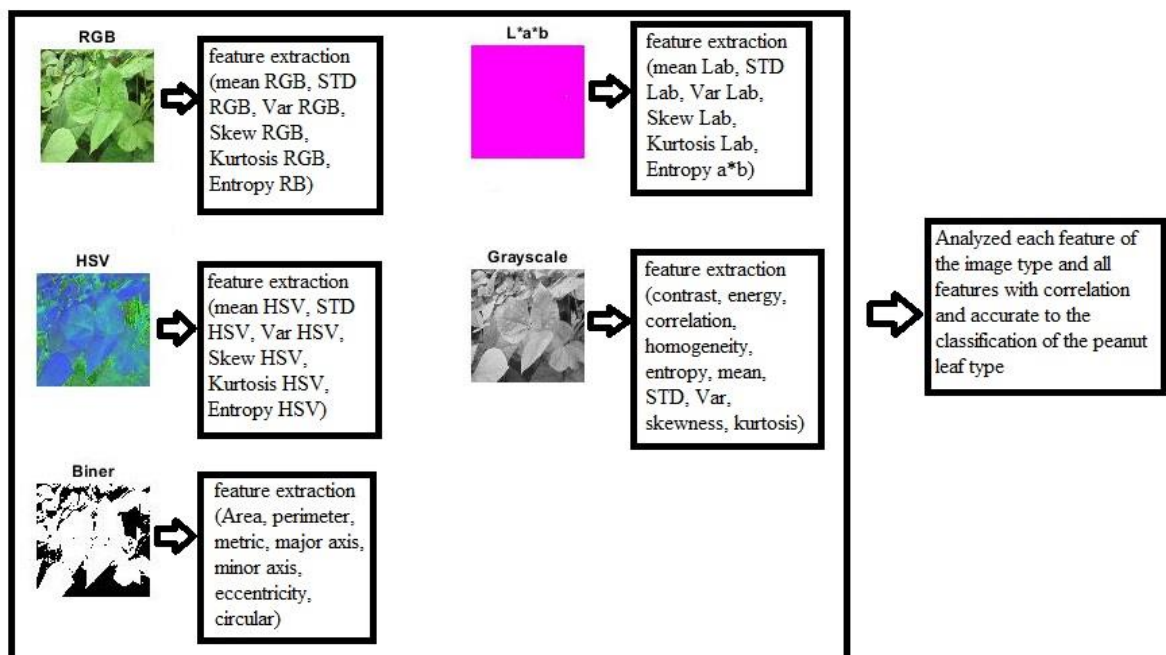


Fig. 3. Research proposal

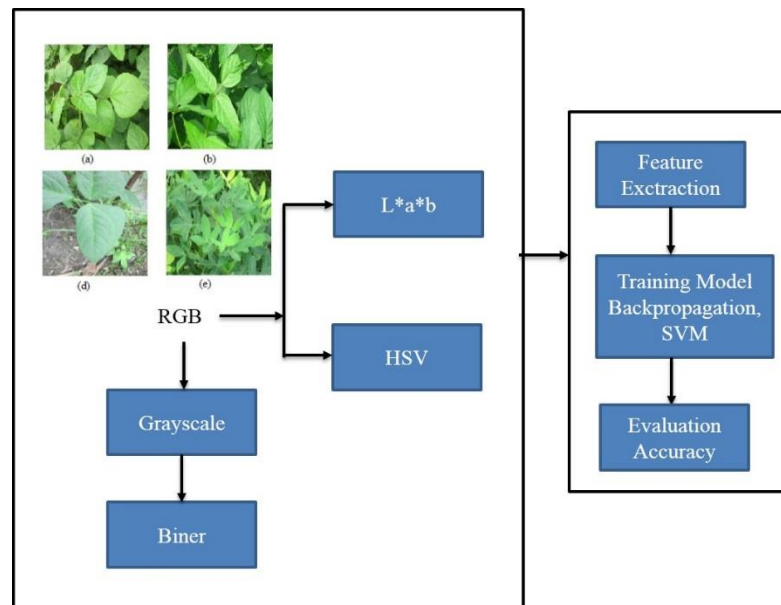


Fig. 4. Research diagram

### 3. Results and Discussion

The initial step in our process involves feature extraction from RGB, L\*a\*b, HSV, grayscale, and binary images, as outlined in Table 2. Subsequently, each image feature is evaluated for its correlation with the type of bean leaf, and the results are presented in Table 3. In this table, some features are abbreviated for clarity, such as Var (Variance), STD (Standard Deviation), and Skew (Skewness). Table 3 reveals the features from the RGB images that exhibit the highest positive correlations, including standard deviation red (0.1321), standard deviation L\*a\*b feature b (0.3522), standard deviation HSV feature H (0.324), and in grayscale, the energy feature (0.1803). Meanwhile, in binary images, the feature with the highest positive correlation is circularity (0.0693). Overall, when considering all the features, it is worth noting that the standard deviation b in the L\*a\*b image demonstrates the most substantial and positive correlation with the bean leaf type.

Table 3. Correlation of image features

No	RGB	Lab	HSV	Grayscale	Biner
1	Mean R 0.0441	Mean L 0.0448	Mean H -0.0017	Contrast -0.1986	Area -0.0560
2	Mean G 0.0377	Mean a 0.0009	Mean S -0.0734	<b>Energy 0.1803</b>	Perimeter -0.1200
3	Mean B 0.0143	Mean b 0.0129	Mean V 0.0431	Correlation -0.0006	Metric 0.0693
4	<b>STD R 0.1321</b>	STD L 0.0826	<b>STD H 0.3240</b>	Homogeneity 0.1059	Major axis -0.1164
5	STD G 0.1114	STD a 0.3281	STD S 0.1390	Entropy -0.2005	Minor axis -0.1052
6	STD B -0.006	<b>STD b 0.3522</b>	STD V 0.1125	Mean 0.0402	Eccentricity 0.0047
7	Variance R 0.0598	Variance L -0.0101	Variance H 0.1660	STD 0.0836	<b>Circularity 0.0693</b>
8	Variance G 0.0287	Variance a 0.2716	Variance S 0.0612	Variance -0.0204	
9	Variance B -0.0303	Variance b 0.2635	Variance V 0.0286	Skewness 0.0146	
10	Skewness R 0.0487	Skewness L 0.0441	Skew H -0.0321	Kurtosis 0.0687	
11	Skewness G 0.0943	Skewness a 0.0344	Skew S -0.0903		
12	Skewness B 0.0156	Skewness b 0.0730	Skew V 0.0895		
13	Kurtosis R -0.0247	Kurtosis L 0.1363	Kurtosis H -0.0226		
14	Kurtosis G 0.0759	Kurtosis a -0.0141	Kurtosis S -0.0022		
15	Kurtosis B 0.1154	Kurtosis b 0.0349	Kurtosis V 0.0791		
16	Entropy R -0.1067	Entropy a 0.2934	Entropy H 0.2519		
17	Entropy B -0.1471	Entropy b 0.2444	Entropy S -0.1712		
18			Entropy V -0.0527		

During the experiment, we constructed two Backpropagation network architectures with the goal of achieving high accuracy in the identification of bean leaf types. Each Backpropagation network architecture employs various input features for RGB images (17 features), L\*a\*b (17 features), HSV (18 features), grayscale (10 features), and binary (7 features). The hidden layer is designed with numerous neurons to optimize accuracy, while the output layer consists of two neurons. The feature extraction results from RGB, L\*a\*b, HSV, grayscale, and binary images were trained using the Neural Network Backpropagation method. In the first experiment, we utilized RGB features with architecture [17, 34, 2] and employed both the 'trainscg' and 'trainlm' functions, with their corresponding accuracy results presented in Tables 4 and 5. Based on the results in Tables 4 and 5, the 'trainlm' function yielded a higher average accuracy value. Consequently, the 'trainlm' function was adopted for all subsequent experiments involving L\*a\*b, HSV, grayscale, and binary images. Table 6 presents the results of the L\*a\*b feature experiment with architecture [17, 34, 2]. Similarly, Table 7 shows the outcomes of the HSV feature experiment in two scenarios: architecture [18, 34, 2] and architecture [18, 36, 2], as shown in Table



8. Furthermore, Table 9 and Table 10 display the results of the grayscale feature experiment in two scenarios: architecture [10, 34, 2] and architecture [10, 20, 2], respectively.

Table 4. RGB feature accuracy results (trainscg)

Accuracy	
Mean	58.38
STD	3.97
Max	64.04
Min	52.19

Table 5. RGB feature accuracy results (trainlm)

Accuracy	
Mean	72.68
STD	5.43
Max	82.02
Min	61.4

Table 6. L\*a\*b feature accuracy results (trainlm)

Accuracy	
Mean	73.07
STD	2.50
Max	77.19
Min	69.74

Table 7. First scenario HSV feature accuracy results

Accuracy	
Mean	82.28
STD	6.82
Max	90.35
Min	67.11

Table 8. Second scenario HSV feature accuracy results

Accuracy	
Mean	81.62
STD	6.57
Max	88.60
Min	71.05

Table 9. First scenario grayscale feature accuracy results

Accuracy	
Mean	50.57
STD	3.42
Max	55.70
Min	46.49

Table 10. Second scenario grayscale feature accuracy results

Accuracy	
Mean	49.65
STD	4.38
Max	56.14
Min	45.61

Table 11 presents the results of binary feature experiments under two different scenarios: architecture [7, 34, 2] and architecture [7, 14, 2], as shown in Table 12. We combined all the features to be tested in classifying the types of bean leaves. The outcomes of these comprehensive experiments are detailed in Table 13, with results from the first scenario (architecture [69, 138, 2]) and the second scenario (architecture [69, 69, 2]) provided in Table 14.

Table 11. First scenario binary feature accuracy results

Accuracy	
Mean	38.73
STD	5.24
Max	45.61
Min	28.51

Table 12. Second scenario binary feature accuracy results

Accuracy	
Mean	33.95
STD	1.66
Max	35.53
Min	32.02

Table 13. First scenario all feature accuracy results

Accuracy	
Mean	87.37
STD	1.94
Max	89.04
Min	82.46

Table 14. Second scenario all feature accuracy results

Accuracy	
Mean	82.46
STD	8.98
Max	91.23
Min	64.04

Table 15. Accuracy evaluation

	RGB	L*a*b	HSV	Grayscale	Biner	All
Backpropagation	72.675	<b>81.623</b>	<b>82.28</b>	50.571	38.728	<b>87.37</b>
SVM	28.51	40.35	<b>61.4</b>	21.05	32.46	27.63

The subsequent process involves evaluating the accuracy of the Backpropagation and SVM methods based on the image features, and the results are detailed in Table 15. Notably, in Table 15, the RGB features that demonstrate the highest accuracy values are the HSV features, whether employing the Backpropagation or SVM methods. The accuracy evaluation of bean leaf species classification is depicted in Figure 5. It is evident from the results that the Backpropagation method attains the highest accuracy when utilizing all 69 features. However, it is important to note that the SVM method, when applied to all 69 features, yields very low accuracy.

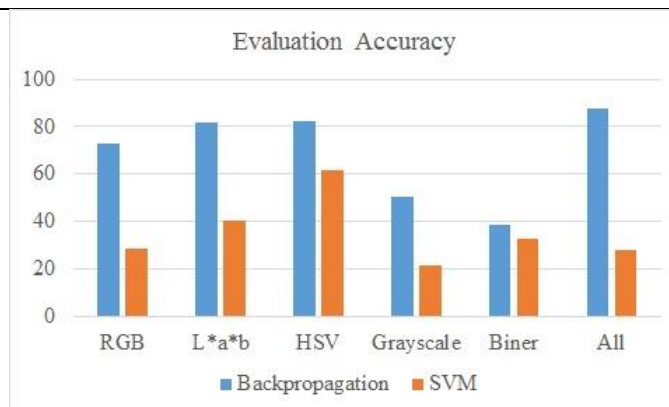


Fig. 5. Evaluation

In previous research [1], leaf recognition involved the utilization of 22 features, and it was determined that the most optimal feature was the edge Fourier transform feature. This recognition method employed SVM, achieving an accuracy rate of 92.53% [1]. In another study [2], feature extraction was carried out using 16 features, and a subset of the best five features was identified for leaf recognition, resulting in an impressive accuracy of 96.8% with the SVM method. Research [3] adopted morphological features and the Fourier transform method for leaf recognition, employing the Neural Network, and reached a 96% accuracy rate. Meanwhile, in a separate investigation [16], the LBP feature was used for leaf recognition in conjunction with the Extreme Learning Machine (ELM) method, achieving an accuracy of 92.92%. Research [15] involved feature extraction using LBP-HF and employed a recognition method based on calculating similarity distances.

#### 4. Conclusion

This study proposes the optimization of feature selection for bean leaf images. The feature selection process involves evaluating features based on their correlation with the type of bean leaf image. The study utilizes images in different color spaces, including RGB, L\*a\*b, HSV, Grayscale, and Binary, and extracts a total of 69 features from these images. Among these features, the most strongly correlated one in bean leaf images is STD b in the L\*a\*b image. The classification methods used for leaf image features are Backpropagation and SVM. When evaluating the classification methods, the study found that the two highest-performing methods were applied to HSV images. Backpropagation achieved an accuracy of 82.28, while SVM achieved an accuracy of 61.4. For future research, it is suggested to explore additional features and their potential impact on improving the classification of bean leaf images.

#### Author Contributions

N. Nafiiyah: Conceptualization, methodology, software, writing – original draft, and writing - review & editing. E. Setyati: Writing – original draft and writing - review & editing. Y. Kristian: Writing - review & editing. R. Wardhani: Funding acquisition, resources, and supervision.

#### References

- [1] E. Yigit, K. Sabanci, A. Toktas, and A. Kayabasi, "A study on visual features of leaves in plant identification using artificial intelligence techniques," *Comput. Electron. Agric.*, vol. 156, 2019, doi: 10.1016/j.compag.2018.11.036.
- [2] C. A. Priya, T. Balasaravanan, and A. S. Thanamani, "An efficient leaf recognition algorithm for plant classification using support vector machine," 2012, doi: 10.1109/ICPRIME.2012.6208384.
- [3] A. Aakif and M. F. Khan, "Automatic classification of plants based on their leaves," *Biosyst. Eng.*, vol. 139, 2015, doi: 10.1016/j.biosystemseng.2015.08.003.
- [4] Ji X. du, de S. Huang, X. F. Wang, and X. gu, "Computer-Aided Plant Species Identification (CAPSI) Based on Leaf Shape Matching Technique," *Trans. Inst. Meas. Control*, vol. 28, no. 3, 2006, doi: 10.1191/0142331206tim176oa.
- [5] S. W. Zhang, Y. J. Shang, and L. Wang, "Plant disease recognition based on plant leaf image," *J. Anim. Plant Sci.*, vol. 25, no. 3, 2015.
- [6] U. Padma, S. Jagadish, and M. K. Singh, "Recognition of plant's leaf infection by image processing approach," *Mater. Today Proc.*, 2021, doi: 10.1016/j.matpr.2021.06.297.
- [7] L. Nanni, M. Paci, G. Maguolo, and S. Ghidoni, "Deep learning for actinic keratosis classification,"

- AIMS Electron. Electr. Eng.*, vol. 4, no. 1, pp. 47–56, 2019, doi: 10.3934/ElectrEng.2020.1.47.
- [8] S. Anubha Pearline and V. Sathiesh Kumar, "Performance analysis of real-time plant species recognition using bilateral network combined with machine learning classifier," *Ecol. Inform.*, vol. 67, no. August 2021, p. 101492, 2022, doi: 10.1016/j.ecoinf.2021.101492.
- [9] F. Jiang, Y. Lu, Y. Chen, D. Cai, and G. Li, "Image recognition of four rice leaf diseases based on deep learning and support vector machine," *Comput. Electron. Agric.*, vol. 179, 2020, doi: 10.1016/j.compag.2020.105824.
- [10] J. F. I. Nturambirwe and U. L. Opara, "Machine learning applications to non-destructive defect detection in horticultural products," *Biosystems Engineering*, vol. 189. 2020, doi: 10.1016/j.biosystemseng.2019.11.011.
- [11] J. Parraga-Alava, K. Cusme, A. Loor, and E. Santander, "RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition," *Data Br.*, vol. 25, 2019, doi: 10.1016/j.dib.2019.104414.
- [12] C. H. Arun and D. Christopher Durairaj, "Identifying Medicinal Plant Leaves using Textures and Optimal Colour Spaces Channel," *J. Ilmu Komput. dan Inf.*, vol. 10, no. 1, 2017, doi: 10.21609/jiki.v10i1.405.
- [13] T. Beghin, J. S. Cope, P. Remagnino, and S. Barman, "Shape and texture based plant leaf classification," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, vol. 6475 LNCS, no. PART 2, doi: 10.1007/978-3-642-17691-3\_32.
- [14] J. Huixian, "The Analysis of Plants Image Recognition Based on Deep Learning and Artificial Neural Network," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2986946.
- [15] C. Yang, "Plant leaf recognition by integrating shape and texture features," *Pattern Recognit.*, vol. 112, 2021, doi: 10.1016/j.patcog.2020.107809.
- [16] M. Turkoglu and D. Hanbay, "Leaf-based plant species recognition based on improved local binary pattern and extreme learning machine," *Phys. A Stat. Mech. its Appl.*, vol. 527, 2019, doi: 10.1016/j.physa.2019.121297.
- [17] S. Zhang, W. Huang, and Z. Wang, "Plant species identification based on modified local discriminant projection," *Neural Comput. Appl.*, vol. 32, no. 21, 2020, doi: 10.1007/s00521-018-3746-0.
- [18] J. xiang Du, C. M. Zhai, and Q. P. Wang, "Recognition of plant leaf image based on fractal dimension features," *Neurocomputing*, vol. 116, 2013, doi: 10.1016/j.neucom.2012.03.028.
- [19] Y. A. Nanekaran, D. Zhang, J. Chen, Y. Tian, and N. Al-Nabhan, "Recognition of plant leaf diseases based on computer vision," *J. Ambient Intell. Humaniz. Comput.*, 2020, doi: 10.1007/s12652-020-02505-x.
- [20] A. Morinaga, K. Hara, K. Inoue, and K. Urahama, "Classification between natural and graphics images based on generalized Gaussian distributions," *Inf. Process. Lett.*, vol. 138, pp. 31–34, 2018, doi: 10.1016/j.ipl.2018.05.010.
- [21] N. Nafiiyah, N. S. Fatonah, R. Wardhani, B. Jokonowo, and T. A. Y. Siswa, "Majority voting transfer learning CNN for peanut leaf types identification," in *ICOn EEI*, 2022, pp. 16–20.