Karbala International Journal of Modern Science

Manuscript 3349

Extraction of Morphometric Features the shape of mangrove leaves based on digital images and classification using the Support Vector Machine

Ishak Ariawan

Della Ayu Lestari

Luthfi Anzani

Tri Yanti

Cakra Rahardjo

See next page for additional authors

Extraction of Morphometric Features the shape of mangrove leaves based on digital images and classification using the Support Vector Machine

Abstract

At present, several botanists still rely on the use of manual estimating methods to assess the carbon content in mangrove. However, these methods have been reported to be extremely time-consuming, showing the need to develop a system for prediction. An effective solution lies in the creation of an artificial intelligence application, which can provide rapid and cost-effective results. In constructing this application, careful consideration must be given to the selection of parameters or attributes. Species is an essential parameter for the assessment of carbon content, but its determination has proven to be challenging due to the similarities of mangrove. The occurrence of errors in identifying species can lead to inaccurate prediction in a given tree. To address this challenge, the identification process can be greatly improved by leveraging plant morphology, particularly leaf. Previous reports have shown that leaf exhibits distinctive morphological features, and the application of geometric mathematics proved instrumental in extracting these characteristics. Therefore, this study aimed to extract the shape of mangrove leaf images using morphometric features. Based on the features obtained, a classification was performed to identify mangrove species using a machine learning algorithm, Support Vector Machine (SVM). The results showed that the geometric method was effective in extracting values for roundness, solidity, eccentricity, convexity, compactness, elongation, rectangularity, and aspect ratio. The analysis of each feature showed that the roundness feature could be used to effectively distinguish the 4 mangrove species. Furthermore, the classification results using SVM obtained the highest average accuracy of 91.26%.

Keywords

Feature extraction; Mangrove; Morphometric; Identification; leaf shape; Support Vector Machine

Authors

Ishak Ariawan, Della Ayu Lestari, Luthfi Anzani, Tri Yanti, Cakra Rahardjo, M. Saleh, Sahril Angga Permana, and Dea Aisyah Rusmawati

RESEARCH PAPER

Extraction of Morphometric Features the Shape of Mangrove Leaves Based on Digital Images and Classification Using the Support Vector Machine

Ishak Ariawan [a,b,*], Della A. Lestari [a], Luthfi Anzani [a], Tri Yanti [a], Cakra Rahardjo [a], M.Saleh [a], Sahril A. Permana [a], Dea A. Rusmawati [c]

[a] Marine Information System, Universitas Pendidikan Indonesia, West Java, 40154, Indonesia
[b] Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia
[c] Department of Food Technology, University of Sultan Ageng Tirtayasa, Serang, Banten, Indonesia

Abstract

At present, several botanists still rely on manual estimating methods to assess the carbon content in mangroves. However, these methods have been reported to be extremely time-consuming, showing the need to develop a system for prediction. An effective solution lies in the creation of an artificial intelligence application that can provide rapid and cost-effective results. In constructing this application, careful consideration must be given to the selection of parameters or attributes. Species are essential parameters for the assessment of carbon content, but their determination has proven to be challenging due to the similarities between mangroves. The occurrence of errors in identifying species can lead to inaccurate prediction in a given tree. To address this challenge, the identification process can be greatly improved by leveraging plant morphology, particularly leaf. Previous reports have shown that leaf exhibits distinctive morphological features, and the application of geometric mathematics proved instrumental in extracting these characteristics. Therefore, this study aimed to extract the shape of mangrove leaf images using morphometric features. Based on the features obtained, a classification was performed to identify mangrove species using a machine learning algorithm, Support Vector Machine (SVM). SVM is able to solve high-dimensional problems, apply Structural Risk Minimization (SRM) strategies, and has a theoretical basis that can be analyzed clearly, so it is possible to use it as an innovative approach to identify mangrove species. The results showed that the geometric method was effective in extracting values for roundness, solidity, eccentricity, convexity, compactness, elongation, rectangularity, and aspect ratio. The analysis of each feature showed that the roundness feature could be used to distinguish the 4 mangrove species effectively. Furthermore, the classification results using SVM obtained the highest average accuracy of 91.26%.

Keywords: Feature extraction, Mangrove, Morphometric, Identification, Leaf shape, Support vector machine

1. Introduction

The sea and coastal areas are integrated ecosystems, which have been reported to have reciprocal correlations. Furthermore, several studies showed that Indonesian forests played an essential role in the sustainability of the world's environmental ecosystems [1]. Indonesian forest area comprises several types of forest, including mangroves, which have a significant role in absorbing and storing carbon [2].

The amount of carbon sequestration stored in mangroves must be calculated to address global climate concerns and increase the function of this forest type [3]. Several botanists still manually estimate the amount of carbon, leading to the consumption of an extended period. To overcome this challenge, there is a need to build a system in the

Received 27 December 2023; revised 22 March 2024; accepted 27 March 2024.
Available online 16 May 2024

* Corresponding author at: Marine Information System, Universitas Pendidikan Indonesia, West Java 40154, Indonesia.
E-mail address: ishak_ariawan@upi.edu (I. Ariawan).
Peer review under responsibility of University of Kerbala.

https://doi.org/10.33640/2405-609X.3349

form of an artificial intelligence application installed on botanists' smartphones. The development of this application comprises several stages, including collecting data on mangrove species, building the best model, and creating the model into an application.

The parameters or attributes for constructing the optimal model must be carefully selected, with species being an essential factor [4]. Species are the main factor due to their role as the basis for choosing the appropriate allometric equation for predicting carbon. However, botanists still experience difficulties in identifying mangrove species accurately because of similarities. Errors during the identification process can lead to inaccurate estimation of the carbon content in a tree. Due to the importance of species identification in calculating the amount of sequestration, an accurate and effective system is needed.

The identification process can be facilitated through the use of plant morphology, which serves to describe the shape or form of plants. Leaves, in particular, emerge as critical components for this process, possessing distinctive characteristics that vary from one plant to another [5]. Furthermore, the main distinguishing characteristics of each plant type, e.g., shape and venation, are readily available in leaves. Extracting the morphometric features of leaf shape and venation comprise the application of geometrical mathematics. The process entails determining the venation point and leaf shape position using the geometry coordinates. Several studies showed that the coordinate values could be used to calculate angles, distances, venation, and leaf shape lengths.

Several findings have been conducted regarding the extraction of leaf-shaped morphometric features. For example [6], performed morphometric feature extraction of the Flavia leaf shape data set. The features extracted from the venation structure include width, length, area, perimeter, aspect ratio, rectangularity, and elongation [7]. Extracted leaf shape morphometric features by calculating Diameter, Convex hull, Solidity, and Roundness values. Furthermore [8], performed extraction using leaf shape morphometric features by calculating the value of position, distance, slope, curvature, and changes in landmark angles. The values calculation is used as a leaf shape feature.

Image classification is a method used to classify information obtained from the observed image, while Support Vector Machine (SVM) is one of the classification methods for grouping images [9]. Previously performed plant identification on *Shorea* species using the SVM algorithm with 84.46% accuracy. Furthermore [10], also conducted classification on Sugar beet, Pigweed, Lamb's quarters, Hare's-ear mustard, and Turnip weed species using the SVM algorithm, with 82.42% accuracy [11]. Studied grape varieties using the same method and obtained an accuracy of 92.94%. SVM has also been used for social media user personality prediction, where good results were obtained with 95.06% accuracy [12]. [13] also explored the classification of diseases in banana plants using the SVM algorithm and obtained 90.9% accuracy.

Therefore, this study aims to extract the shape of a Mangrove leaf image using morphometric features. Leaf shape feature extraction results were expected to produce several characteristics, including leaf area, leaf perimeter, leaf length, leaf width, convex hull area, convex hull perimeter, convexity, aspect ratio, roundness, solidity, rectangularity, elongation, compactness, and eccentricity. From the characteristics obtained, a classification was performed to identify mangrove species using a machine learning algorithm, SVM. SVM is able to solve high-dimensional problems, apply Structural Risk Minimization (SRM) strategies, and has a theoretical basis that can be analyzed clearly, so it is possible to be used as an innovative approach to identify mangrove species [14]. The main contributions of this research can be summarized as follows:

- This research is the first research that uses a specific data set of mangrove species by extracting morphometric features of leaf shape (leaf area, leaf perimeter, leaf length, leaf width, convex hull area, convex hull perimeter, convexity, aspect ratio, roundness, solidity, rectangularity, elongation, compactness, and eccentricity) as parameters and targets for SVM classification.
- This research shows a close to state-of-the-art performance using the SVM algorithm and 14 feature extraction parameters and 4 targets (*Rhizophora stylosa, Rhizophora apiculata, Bruguiera cylindrica,* and *Avicennia marina*).
- The efficient and low-resource nature of machine learning algorithmic models makes them easy to adapt and implement on mobile devices or embedded.

2. Research method

The stages of the study process can be seen in Fig. 1, including data collection, feature extraction, data sharing, parameter optimization, classification model creation, and model evaluation. Each of the stages was explained in detail in the next subchapter.

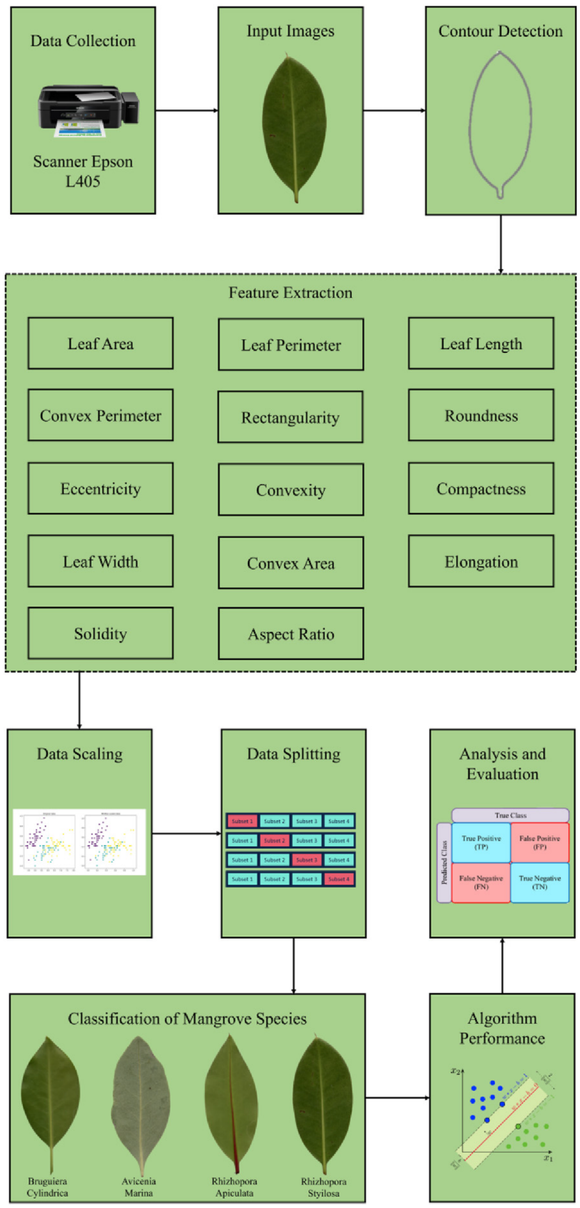

Fig. 1. Research stages.

2.1. Data collection

Data collection was carried out on several mangrove forests in Lontar Village and Burung Island, Serang City, Banten Province, by taking 30–50 leaves per tree, followed by an image process. The image retrieval process was performed using a scanner to obtain a total of 2231 samples, including 528 *R. stylosa* leaves, 573 *R. apiculata*, 544 *B. cylindrica*, and 586 *A. marina* images.

2.2. Feature extraction

The method used for morphometric feature extraction is geometric mathematics. The geometric mathematics method aims to analyze and extract the shape or structure characteristics of objects, such as the shape of mangrove leaves [15]. Some main aspects of geometric mathematics relevant to leaf morphometric extraction involve coordinates, size, comparison, and contour analysis. Initially, geometry deals with coordinates and positioning to represent objects within an area using coordinates as a foundation. The relative positions of points or features on mangrove leaves can be. They are quantified through geometric coordinates [16]. Additionally, geometry introduces concepts like line length distances between points and other geometric measurements to assess the dimensions of mangrove leaves and their characteristics [17]. Such data is commonly utilized in morphometry to enhance our understanding of leaf shapes. Moreover, geometry facilitates comparisons between objects' lengths, widths, and heights [18]. It also aids in examining object contours, measuring curvature, or identifying features within shapes.

Feature extraction was utilized to capture the attributes in each mangrove leaf's morphology. The morphological features of these leaves were used in this study, which consisted of the main features, such as Leaf Area (LA), Leaf Perimeter (LP), Leaf Length (LL), Leaf Width (LW), Convex Area (CA), and Convex Perimeter (CP). Furthermore, the features were assessed using ImageJ, an image analysis software, a free application that performs Java-based digital processing and derived features. The process was carried out through the calculation of the main featured combinations, including Convexity (C), Aspect Ratio (AR), Roundness (R), Solidity (S), Rectangularity (RC), Elongation (E), Compactness (CN), and Eccentricity (EC). The equation of the derived feature formula is presented in Table 1 [19].

2.3. Data scaling

Data Scaling or normalization was done by converting numeric values in a data set to a standard scale without distorting differences in the range of values. Furthermore, data scaling could help assist the learning process in machine learning [20]. Due to the data extraction features of mangrove leaf shapes, the data were scaled using the Standardization (Zero-Mean) method, and it was based on the mean and standard deviation. Standardization of a data set comprised changing the scale of the distribution of values to ensure that the observed mean value was 0 and the standard deviation was 1 [20]. The standard deviation was calculated using Equation (1).

Table 1. Equations of derived feature formula.

Morphological Features	Formula
Roundness	$\frac{4 \times \pi \times Leaf\ Area}{Convex\ Perimeter^2}$
Solidity	$\frac{Leaf\ Area}{Convex\ Area}$
Eccentricity	$\frac{\sqrt{\text{Leaf Length}^2 - \text{Leaf Width}^2}}{\text{Leaf Length}}$
Convexity	$\frac{Convex\ Perimeter}{Perimeter}$
Compactness	$\frac{4 \times \pi \times Leaf\ Area}{Perimeter^2}$
Elongation	$1 - \frac{Leaf\ Width}{Leaf\ Length}$
Rectangularity	$\frac{Leaf\ Width \times Leaf\ Length}{Leaf\ Area}$
Aspect ratio	$\frac{Leaf\ Length}{Leaf\ Area}$

$$x_{std} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - x_{mean})^2} \quad (1)$$

x_{mean} is the average of the data. Data scaling can be calculated with Equation (2).

$$x_i' = \frac{x_i - x_{mean}}{x_{std}} \quad (2)$$

x_i' is the value after the standardization, x_i The initial value of the variables and x_{std} is the standard deviation of the variable x_i.

2.4. Data splitting

Data distribution was performed before classification, and the data were divided into 2 categories: training and testing. The distribution used k-fold cross-validation to ensure the classification results represented the data. First, the data were divided into k parts or folds, and each received an equal allocation. One-fold served as testing data, while the other was training data. The process of distributing testing and training data was iterated k times to derive the average accuracy for each model [21]. An illustration of data sharing is shown in Fig. 2.

2.5. Classification

The model development used SVM, which comprised 3 processes, including optimization of SVM parameters, model selection, and model evaluation.

2.5.1. Optimization of SVM parameters

The first stage in model development was parameter optimization. This stage aimed to determine the best C and gamma (γ) kernel radial basis function (RBF) parameters for data training. C served as the parameter for outlier tolerance. It determined the decision boundary parameter, and γ controlled how much impact one data point has on the formation of the margin in the RBF kernel [22].

2.5.2. Model selection

At this stage, the model selection was performed using the K-fold cross-validation method, which reduced the bias in constructing the SVM model. Employing the K-fold cross-validation method also prevented models from becoming overly fitted to the data. The results showed that the K value given to the K-fold cross-validation was 5. After obtaining 5 models from SVM model development, the selection was based on the accuracy in effectively representing the data.

2.5.3. Model evaluation

The confusion matrix presented information on the class and the classifier's predicted results. Furthermore, it was a valuable tool for assessing the model performance in classifying data [23]. The matrix comprised 4 key terms [24]:

- True Positive (TP): correctly classified positive data.

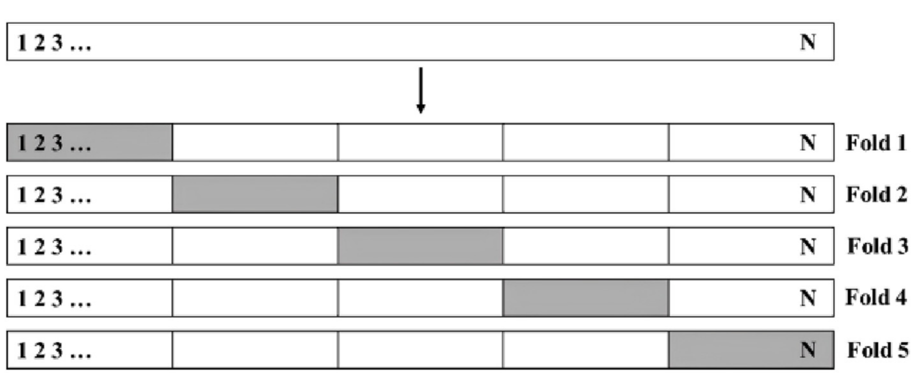

Fig. 2. Illustration of 5-fold cross-validation, gray box as testing data.

- True Negative (TN): the number of correctly classified negative data.
- False Positive (FP): the number of negative data classified as positive.
- False Negative (FN): the number of positive data classified as negative.

TP and TN showed that the classifier identified the data correctly, while FP and FN signified that the classifier identified the data incorrectly. Table 2 presents the confusion matrix based on the procedures proposed by Ref. [24].

Based on the confusion matrix, the results could be evaluated by calculating the accuracy, precision, and sensitivity values using equations (3)–(5), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

3. Results

3.1. Feature extraction

Some features/traits must be extracted first to recognize objects in an image. Furthermore, the features extracted from mangrove leaves included leaf area (LA), leaf perimeter (LP), leaf length (LL), leaf width (LW), convex area (CA), and convex perimeter (CP) (Fig. 3), and the results were presented in Table 3.

The other features used were derived morphological features, which were obtained from the proportion of the main feature, such as roundness (RN), solidity (SL), eccentricity (EC), convexity (CX), compactness (CN), elongation (EG), rectangularity (RC), and aspect ratio (AR). Table 4 presents several examples of extracting derived features from the 4 mangrove species.

The derived characteristics used included RN, SL, EC, CX, CN, EG, RC, and AR. Based on these derived characteristics, the level of roundness of the object was determined from the roundness value, the density of the object shape was assessed using the solidity value, and the ratio of the distance between the foci of the ellipse to the length of the central axis (major axis) of an object was determined with the eccentricity value. Furthermore, the convexity of the object shape was evaluated with convexity value, the convex nature was assessed using the compactness value, and the slenderness of an object was measured based on the elongation value. Describing the similarity of the object shape to a box shape was

Table 2. *Confusion matrix.*

		Prediction Class	
		Positive	*Negative*
Actual class	*Positive*	TP	FP
	Negative	FN	TN

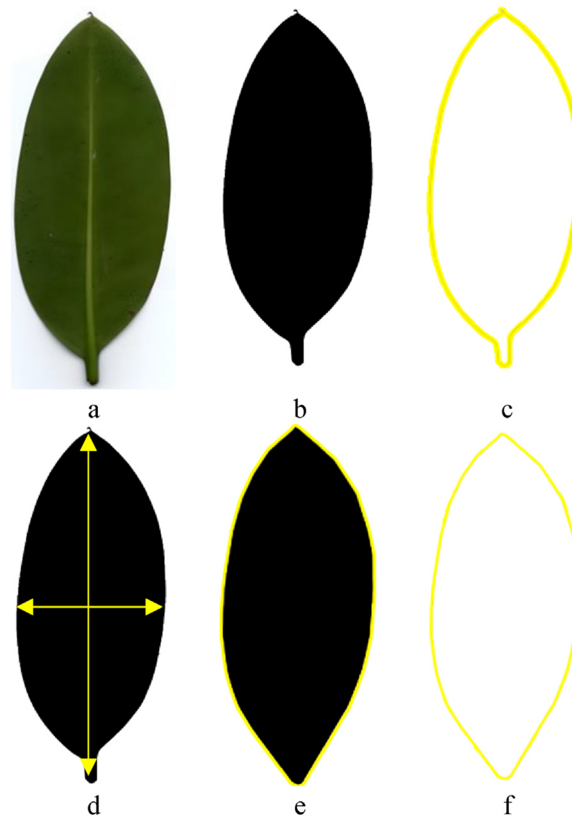

Fig. 3. Main features of mangrove leaf morphology: (a) Original Image, (b) Leaf area, (c) Leaf perimeter, (d) Leaf length and width, (e) Convex area, and (f) Convex perimeter.

Table 3. Feature extraction results for leaf area, leaf perimeter, leaf length, leaf width, convex area, convex perimeter.

LA	LP	LL	LW	CA	CP
8.18	3.79	19.59	20.2	2174.5	30.89
7.92	3.45	16.78	19.07	2083.5	27.21
8.65	3.76	21.33	20.8	2274.8	32.41
7.96	3.06	14.97	18.82	2057.5	24.29
6.76	3.14	13.76	16.48	1797.7	21.16
7.19	3.02	13.19	17.1	1881.5	21.67
7.92	3.4	16.16	19.03	2078.2	26.84
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Table 4. Results of derived feature extraction roundness, solidity, eccentricity, convexity, compactness, elongation, rectangularity, and aspect ratio.

RN	SL	EC	CX	CN	E	RC	AR
0.66	0.015	0.90	2.42	20.2	0.56	0.66	2.28
0.64	0.013	0.91	2.24	18.0	0.59	0.64	2.42
0.55	0.009	0.93	1.81	12.6	0.64	0.55	2.82
0.61	0.009	0.91	1.52	12.6	0.58	0.61	2.36
0.58	0.012	0.89	2.27	16.6	0.55	0.58	2.22
0.61	0.012	0.88	2.04	15.8	0.53	0.61	2.12
0.64	0.006	0.92	1.08	8.80	0.59	0.64	2.49
0.66	0.008	0.81	1.36	11.3	0.56	0.66	2.29
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

based on the rectangularity value, while comparing the length and width of the object was performed using the aspect ratio value. In Fig. 4, a boxplot of the roundness characteristics of the 4 mangrove species was presented. All mangrove species had outliers for the roundness characteristic, while *R. apiculata* species had extremities (546th data). The highest data diversity was shown by *R. stylosa*, and the lowest was *B. cylindrica*. Among the 4 mangrove species, *R. apiculata, B. cylindrica*, and *R. stylosa* had a distribution pattern close to symmetrical, showing that the average data was the same as the median. For *A. marina*, the slope of the data distribution pattern was negative, showing that the average value of *A. marina* was below the median (skewness). The data distribution from all mangrove species showed that the roundness feature could only distinguish *A. marina* and *B. cylindrica* effectively.

In Fig. 5, a boxplot depicted the solidity characteristics of the four mangrove species. Compared to the roundness, all species also exhibited outliers in solidity, with *Rhizophora apiculate* showing extremes at 546th data. The highest data diversity was shown by *R. apiculate*, while the lowest was *A. marina.* Among the 4 species, *R. stylosa*, *apiculate*, and *A. marina* showed a distribution pattern close to symmetry, which indicated concordance between average data and the median. *B. cylindrica* showed a negative slope, implying that the average value was above the median (skewness). Furthermore, the solidity feature effectively distinguished *R. stylosa, B. cylindrica*, and *A. marina* among the 4 mangrove species.

In the analysis of eccentricity characteristics presented for mangrove species (Fig. 6), outliers and extremities were observed across all species, specifically at the 349th, 719th, 944th, 1492nd, 1578th, 1812nd, and 2145th data points. The highest data diversity was shown by *A. marina*, while *R. apiculate* had the lowest. Furthermore, all 4 mangrove species showed a distribution pattern close to symmetrical, implying that the average data were the same as the median. The data distribution from all mangrove species showed that the eccentricity feature could distinguish Rhizophora apiculate from other species.

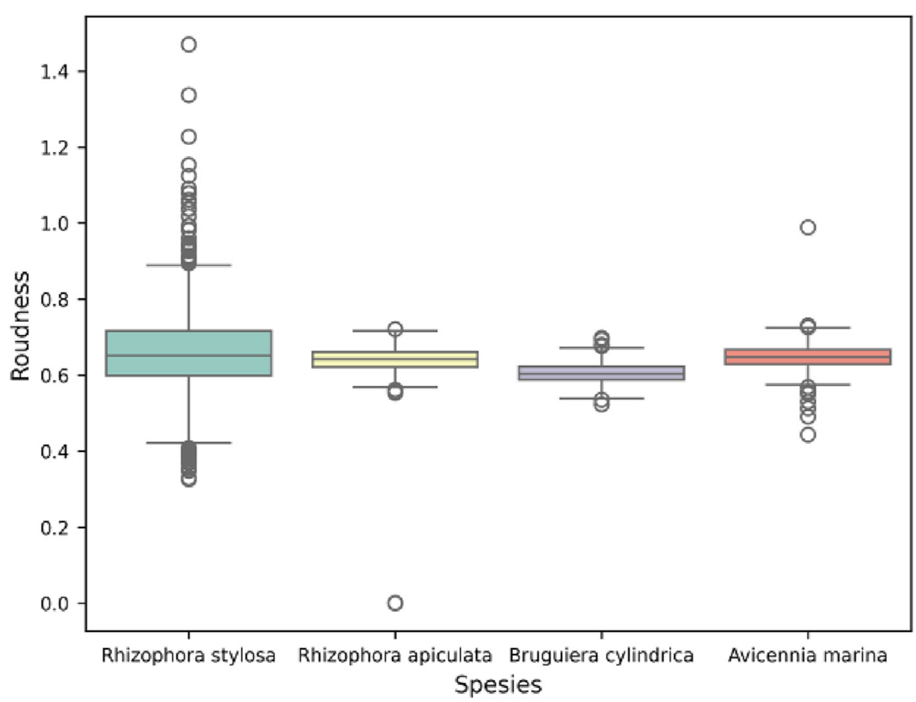

Fig. 4. Boxplot feature roundness.

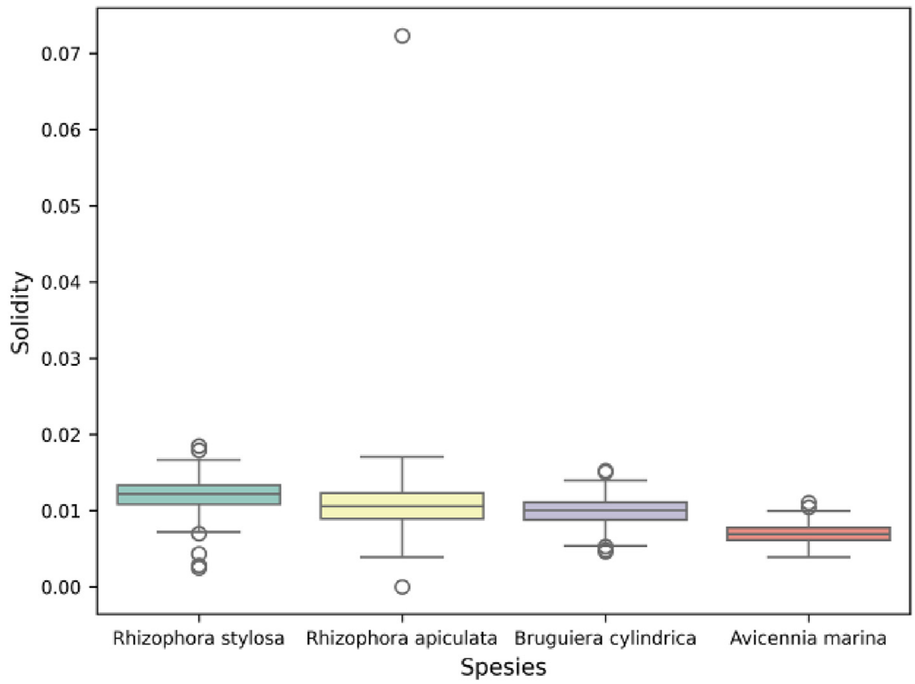

Fig. 5. Boxplot of solidity features.

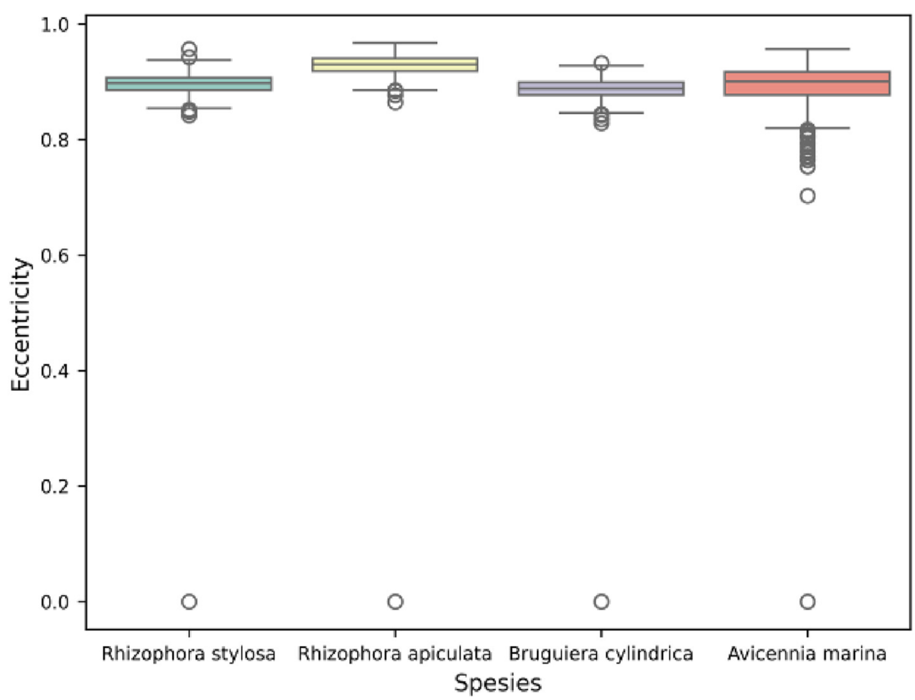

Fig. 6. Boxplot of eccentricity features.

In Fig. 7, a boxplot of the convexity characteristics of the 4 mangrove species was presented. All the species had outliers in convexity characteristics, while *R. apiculate* species had extremities at 546th data. The highest data diversity was shown by *R. apiculate*, while the lowest was observed in *A. marina*. Among the samples, *R. stylosa, apiculate, B. cylindrica*, and *A. marina* had a distribution pattern close to symmetrical, showing that the average data was the same as the median. The distribution

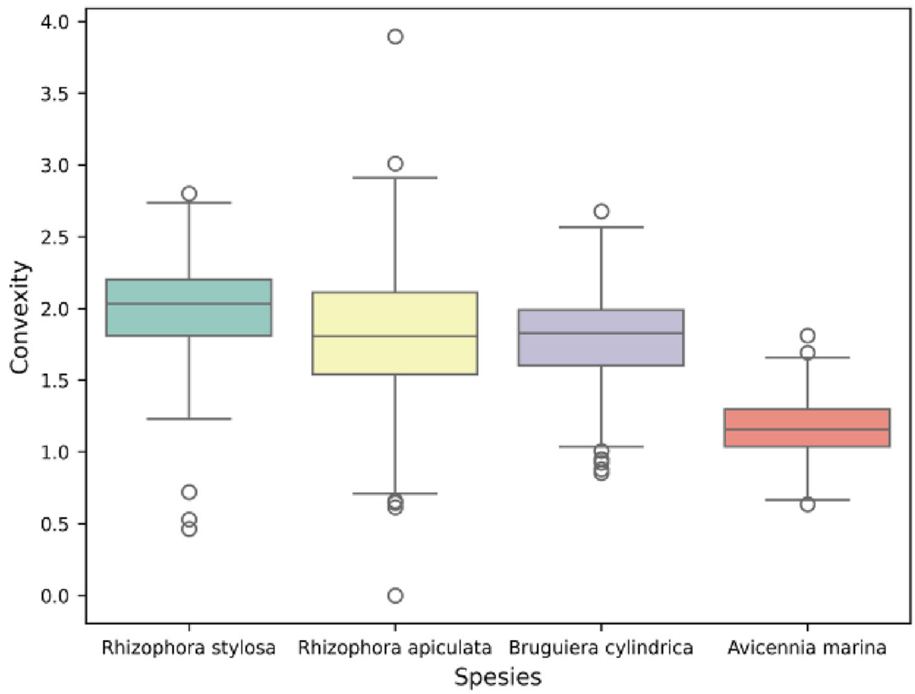

Fig. 7. Boxplot of convexity features.

pattern of the convexity feature in all samples showed effective differentiation between *A. marina* and other species.

As depicted in Fig. 8, a boxplot of the compactness characteristics of the 4 samples was displayed. All species had outliers in compactness, while *R. apiculate* species had extremities (546th data). The highest data diversity was shown by *R. apiculate*, while the lowest was found in *A. marina*. Among the 4 mangrove species, *A. marina* and *apiculate* had a

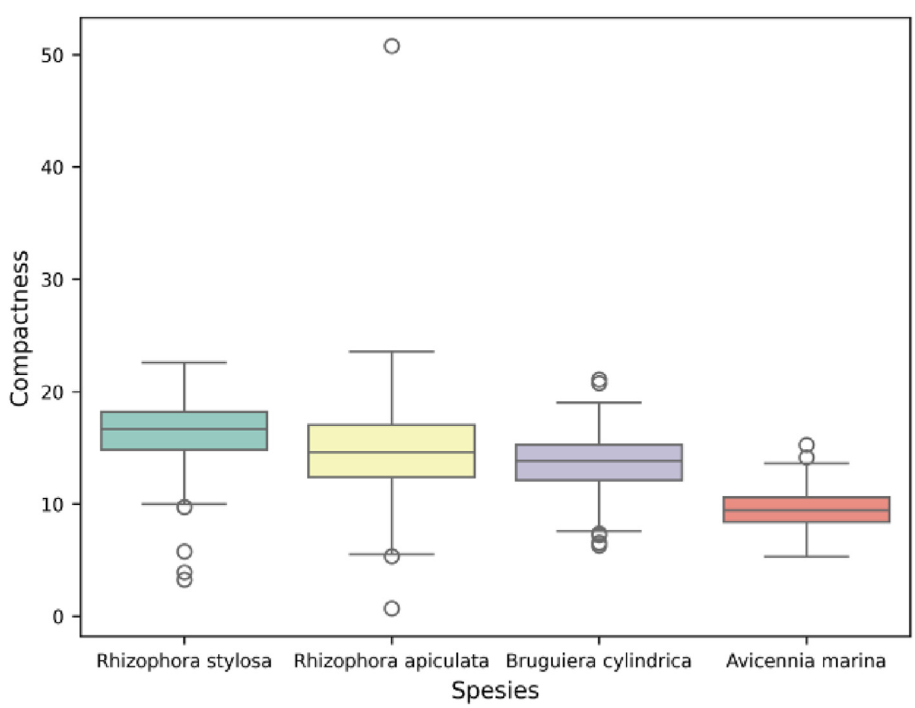

Fig. 8. Boxplot feature of compactness.

distribution pattern close to symmetrical, showing that the average data were the same as the median. For *B. cylindrica* and *R. stylosa*, the slope of the data distribution pattern was negative. This showed that the average value of *B. cylindrica* and *R. stylosa* was above the median (skewness). Furthermore, the data distribution from all mangrove species showed that the compactness feature could distinguish *A. marina* and other species effectively.

Based on Fig. 9, a boxplot of the elongation characteristics presented 4 mangrove species. All the samples had outliers in elongation characteristics, while *A. marina* species had extremities (2136th data). *A. marina* showed the highest data diversity, while the lowest was observed in *B. cylindrica*. Among the 4 mangrove species, *R. stylosa, apiculate*, and *B. cylindrica* had a distribution pattern close to symmetrical, showing that the average data were the same as the median. For *A. marina*, the slope of the data distribution pattern was negative, showing that the average value of *A. marina* was above the median (skewness). The data distribution among the samples showed that the elongation feature effectively distinguished *R. apiculate* and other species.

In Fig. 10, a boxplot showed the rectangularity of the four mangrove species, and all species exhibited outliers in rectangularity, while *R. apiculate* species displayed extremes (546th data). *R. stylosa* and the lowest showed the highest data diversity found in *B. cylindrica*. Among the samples, *A. marina* showed a distribution pattern close to symmetrical, which indicated equality between the average data and the median. For *R. stylosa*, *R. apiculate,* and *B. cylindrica*, the slope of the data distribution pattern was negative, and this showed that the average value of *R. stylosa*, *R. apiculate,* and *B. cylindrica* was skewness. The data distribution from all species showed that the rectangularity feature could distinguish *A. marina* and *B. cylindrica*.

Fig. 11 presents a boxplot that depicted the aspect ratio features of the 4 mangrove species, and all samples exhibited outliers in all aspect ratios, with *R. apiculate, B. cylindrica,* and *A. marina* displaying extremes (719th data, 944th data, 1492nd data, 1578th data, 7th data 1812, and 2145 data). The highest data diversity was shown by *R. apiculate*, while the lowest was observed in *B. cylindrica*. Among the 4 mangrove species, *R. stylosa, B. cylindrica,* and *A. marina* had a distribution pattern that was close to symmetrical, showing that the average data were the same as the median. For *R. apiculate*, the slope of the data distribution pattern was positive. This showed that the average value was above the median, and the distribution of data from all mangrove species indicated the ability of the aspect ratio feature to distinguish *R. apiculate* and other species effectively.

Summarizing all figures (Figs. 4−11), showing the species that can be distinguished based on graph parameters (Table 5).

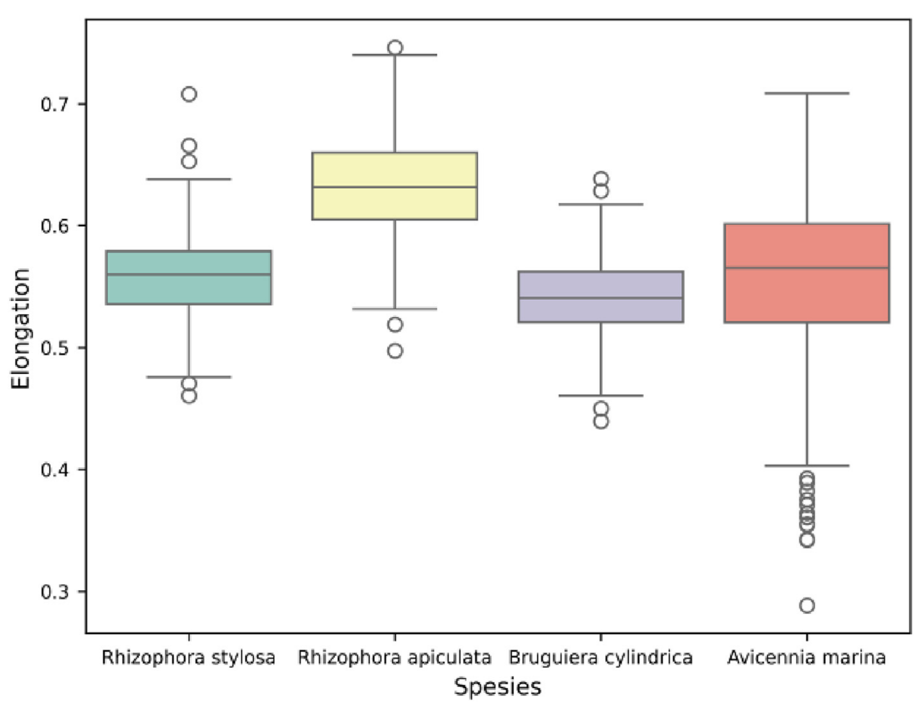

Fig. 9. Boxplot of elongation features.

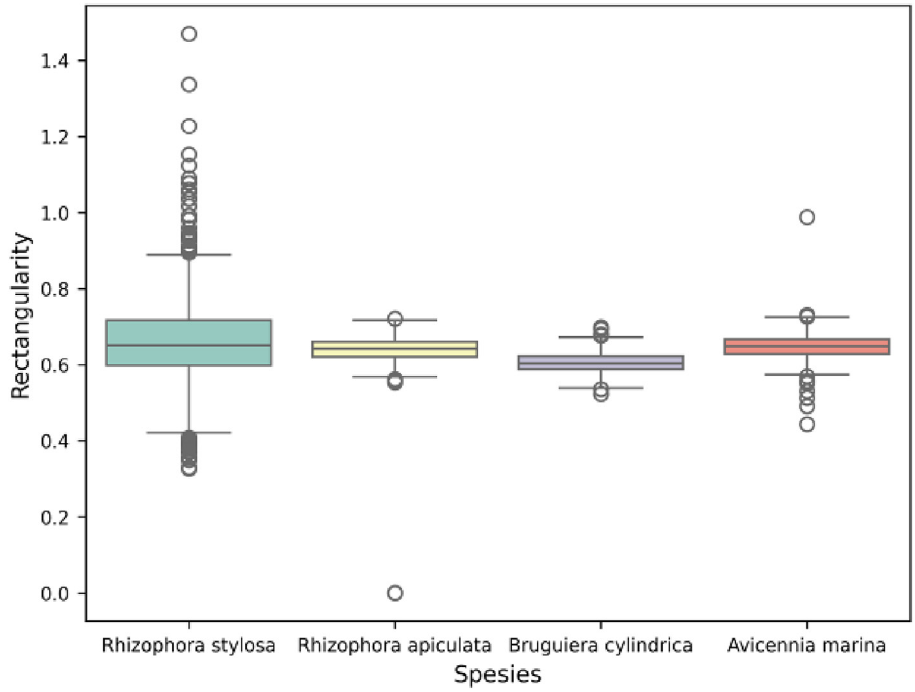

Fig. 10. Rectangularity feature boxplot.

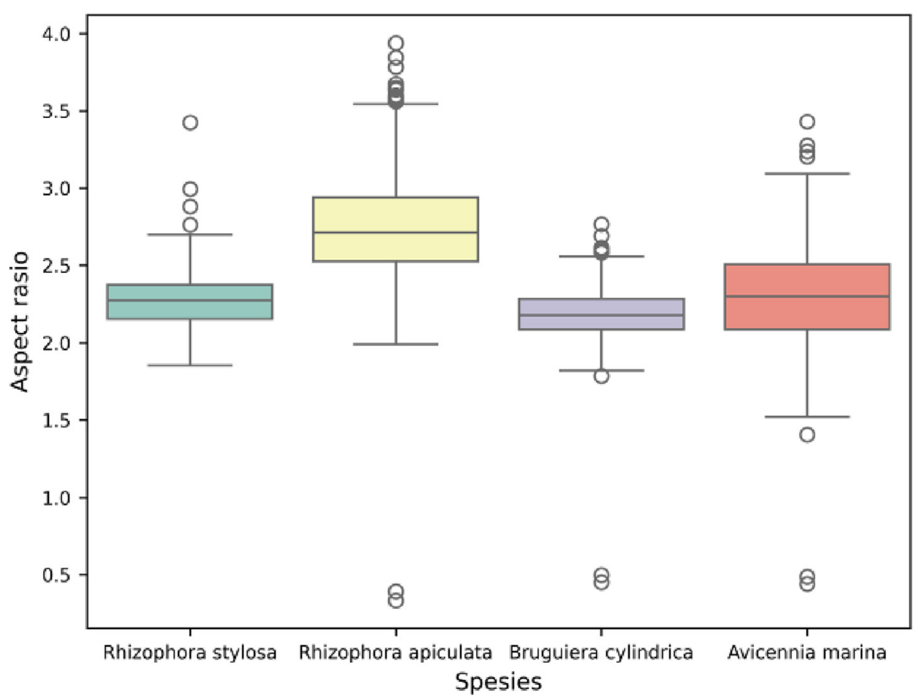

Fig. 11. Boxplot of aspect ratio features.

3.2. Data scaling

In machine learning, only some datasets required normalization, but it became essential when features exhibited varying ranges, as seen in the leaf venation feature data set. Each leaf venation feature had a distinct scale, which ranged from 0 to thousands. Furthermore, data from 14 features was presented in Tables 3 and 4 before normalization, which underscored significant scale differences,

Table 5. Summarizing all figures (Figs. 4–11).

Fig	Feature	Difference
4	Roundness	*Avicennia marina* and *Bruguiera cylindrica*
5	Solidity	*Rhizophora apiculate* and other species
6	Eccentricity	*Rhizophora apiculate* and other species
7	Convexity	*Avicennia marina* and other species
8	Compactness	*Avicennia marina* and other species
9	Elongation	*Rhizophora apiculate* and other species
10	Rectangularity	*Avicennia marina* and *Bruguiera cylindrica*
11	Aspect Ratio	*Rhizophora apiculate* and other species

specifically between convex area and solidity features. These large-scale differences could increase the computations performed by machine learning algorithms. To optimize the performance of the machine learning algorithm, it was necessary to normalize the data when the scale difference in the data was significant. The results of zero-mean normalization on the mangrove leaf shape feature data set were shown in Tables 6 and 7, where it changed the scale of the average value to 0, and the standard deviation became 1. Despite this, each feature maintained a different scale (not in the same interval), and the standardized mangrove leaf shape feature data set was used to test the performance of the SVM algorithm.

Table 6. Results of normalization of leaf area features, leaf perimeter, leaf length, leaf width, convex area, convex perimeter.

LA	LP	LL	LW	CA	CP
−0.53	−0.25	−0.41	−0.43	−0.50	−0.46
−0.63	−0.56	−0.57	−0.60	−0.63	−0.62
−0.36	−0.27	−0.30	−0.34	−0.35	−0.40
−0.62	−0.92	−0.67	−0.64	−0.67	−0.75
−1.07	−0.85	−0.75	−0.99	−1.04	−0.89
−0.91	−0.96	−0.78	−0.90	−0.92	−0.87
−0.63	−0.61	−0.61	−0.61	−0.64	−0.64
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Table 7. Results of normalization of derived features roundness, solidity, eccentricity, convexity, compactness, elongation, rectangularity, and aspect ratio.

RN	SL	EC	CX	CN	E	RC	AR
0.13	1.66	0.03	0.82	1.80	−0.18	0.13	−0.27
−0.02	1.08	0.13	0.61	1.22	0.25	−0.02	0.12
0.08	0.30	0.43	0.13	0.39	1.69	0.08	1.82
0.01	−0.01	0.48	−0.03	−0.72	1.98	0.01	2.27
−0.43	0.70	−0.03	0.65	0.83	−0.42	−0.43	−0.46
−0.19	0.62	−0.12	0.39	0.63	−0.76	−0.19	−0.72
0.02	−1.17	0.18	−0.69	−1.25	0.47	0.02	0.34
0.12	−0.54	0.03	−0.37	−0.57	−0.15	0.12	−0.24
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

3.3. Data splitting

The model development employed 5-fold cross-validation, which divided the data into 80% for training and 20% for testing. The results of each class data distribution are presented in Table 8.

3.4. Classification model selection

The SVM classification model was built using the RBF kernel, with the best C and γ parameter values obtained. The best values of C and γ parameters selected for making the classification model were $C = 10{,}000$ and $\gamma = 0.1$. The value of parameter C determined the significance level of the penalty, and the greater the C value, the greater the penalty given to the classification error. This showed the training process ran tighter and closer to the SVM hard margin (the risk of overfitting increases). When the C value was reduced, the classification error penalty was minor. Consequently, the training process became "tolerant" of data noise, and the risk of underfitting increased. For the parameter γ, the magnitude of γ represented the width of the variance. This meant that the penalty size and the width of the variance used in this study were both small. The detailed accuracy values for each model of the best C and γ pairs are presented in Table 9.

The selection of the classification model from the 5 models built used the best pair of C and γ values, which examined the accuracy of each of them. In this study, the classification model with the best accuracy was selected, which included the fifth fold, with an accuracy of 92.43%. In comparison, the accuracy of the average value generated from the five folds was also relatively good, with an accuracy value of 91.26%.

Table 8. Training data and testing data.

Mangrove Species	Total	Training data	Testing data
Rhizophora stylosa	528	422	106
Rhizophora apiculata	573	458	115
Bruguiera cylindrica	544	435	109
Avicennia marina	586	469	117
Total	2231	1784	447

Table 9. Details of the accuracy values for each model from the best pair C and γ.

Fold	Accuracy
1	91.56%
2	90.76%
3	90.76%
4	90.76%
5	92.43%
Average	91.26%

3.5. Classification model evaluation

At this stage, the classification model was evaluated to test each class's accuracy using 20% of the test data. The amount of test data in each class is presented in Table 8. The average accuracy of the test results for each class differed from the model accuracy in the selected fold. This was because the accuracy of the test results on the fold was based on the classification discriminant function. Meanwhile, the accuracy of testing each class was based on the probability value of the classifier prediction results. A comparison of the accuracy of each class is presented in Fig. 12.

Based on Fig. 12, the average accuracy obtained from the test results for each type of mangrove species was 95.56%. The results showed that *B. cylindrica* was the best class, with an accuracy of 96.01%. Furthermore, the measured performance of a classification model could be performed by calculating the value of precision and sensitivity, and the comparison is shown in Fig. 13.

The average precision value was 92.00%, and the sensitivity was 91.00%, as shown in Fig. 13. The results showed that the difference in value between precision and sensitivity was moderate. This showed that the performance of the classification model built was relatively good. Based on the results, it could be concluded that the geometric features used were quite effective due to the high accuracy.

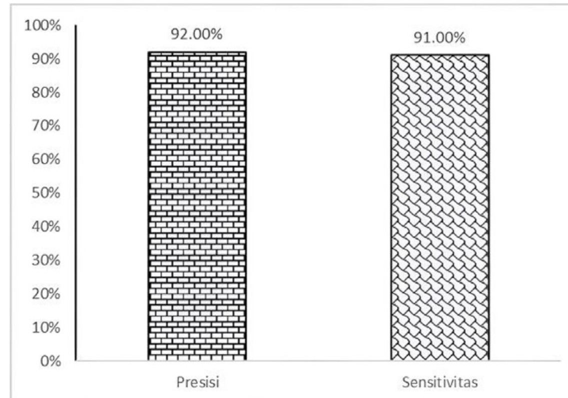

Fig. 13. Comparison graph of mean precision and sensitivity.

4. Discussion

Determining mangrove species has several challenges. Identifying types of mangrove species presents several obstacles. The primary difficulty lies in the variations in characteristics among mangrove species, particularly in leaf shape and size, which can pose a challenge for botanists trying to distinguish between them. This complexity can also lead to issues with classifying the species due to differences among certain types. Moreover, errors in the identification process may stem from factors like lighting conditions, poor image quality, and natural diversity within the mangrove habitat [25]. Uncertainties in botanical knowledge can complicate the identification process, especially when species have similar morphological characteristics [26]. The importance of detailing these challenges lies in developing more effective solutions. The solutions adopted must be able to address these specific challenges, such as by utilizing advanced image processing technologies or artificial intelligence-based approaches that take into account morphological variations and environmental factors [27]. By delving into the difficulties associated with

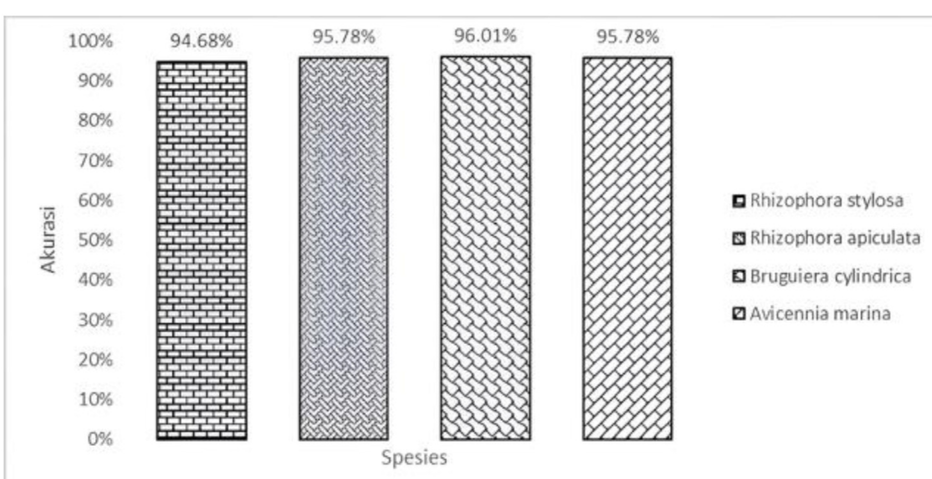

Fig. 12. Accuracy comparison chart for each class.

identifying mangrove species and potential sources of error, this study can substantially enhance the utility and accuracy of identification methods. This effort will increase understanding of the diversity of mangrove ecosystems and open up opportunities for further research and conservation.

By the principles of good classification or model results, the classification or model results achieved should be compared with several others. However, the data set used in this research is a special data set for classifying mangrove species based on leaf images. During this research, this data set has never been used. So, it is impossible to compare the use of the SVM algorithm to classify mangrove species with other studies. Nevertheless, an extensive review shows that SVM has shown excellent results in pattern recognition and image classification. Currently, a commercial solution called Leafsnap [28] uses visual recognition to identify plant species from leaf images, but no mangrove species exist in the Leafsnap data set. Species in the Leafsnap data set include *Acer griseum*, *Acer ginnala, Acer campestre, Abies nordmanniana*, and *Abies concolor*. By comparing our results with other plant species classification methods based on leaf images, it can be said that our model provides better results than research conducted by Refs. [29−31] and approaching the state of the art [15,32,33].

5. Conclusion

In conclusion, featured extraction using a geometric approach successfully extracted roundness, solidity, eccentricity, convexity, compactness, elongation, rectangularity, and aspect ratio values. Furthermore, these features could be used as a marker to identify mangrove species, including *R. stylosa, R. apiculata, B. cylindrica,* and *A. marina*.

Based on the analysis, the solidity feature could effectively distinguish the 4 mangrove species. Furthermore, leaf identification based on mangrove species was carried out using a classification shown, namely SVM. The highest average accuracy obtained from the test results was 91.26% in this study.

Acknowledgements

This study was funded and supported by the Center For Higher Education Funding (BPPT) and Indonesian Endowment Fund for Education (LPDP) under Grant Number 00114/J5.2.3./BPI.06/9/2022.

References

[1] M.L.P. Peñaranda, J.R. Cantera Kintz, E.J. Peña Salamanca, Carbon stocks in mangrove forests of the Colombian Pacific, Estuar. Coast Shelf Sci. 227 (2019) 106299, https://doi.org/10.1016/j.ecss.2019.106299.
[2] I.F. Setiyaningrum, R. Harini, N. Wirasanti, Economic valuation of the mangrove ecosystem in Purwodadi sub-district, Purworejo Regency, central Java, IOP Conf. Ser. Earth Environ. Sci. (2020) 1−8, https://doi.org/10.1088/1755-1315/451/1/012052.
[3] W.L. Farahdita, N. Soenardjo, C.A. Suryono, Drone technology for estimating carbon stocks in mangrove areas of Kemujan Island, Karimunjawa, J. Mar. Res. 10 (2021) 281−290, https://doi.org/10.14710/jmr.v10i2.30466.
[4] G. Angst, K.E. Mueller, D.M. Eissenstat, S. Trumbore, K.H. Freeman, S.E. Hobbie, J. Chorover, J. Oleksyn, P.B. Reich, C.W. Mueller, Soil organic carbon stability in forests: distinct effects of tree species identity and traits, Global Change Biol. 25 (2019) 1529−1546, https://doi.org/10.1111/gcb.14548.
[5] T.-L. Le, D.-T. Tran, V.-N. Hoang, Fully automatic leaf-based plant identification, application for Vietnamese medicinal plant search, in: Proc. Fifth Symp. Inf. Commun. Technol. - SoICT 14, 2014, pp. 146−154, https://doi.org/10.1145/2676585.2676592.
[6] G. Saleem, M. Akhtar, N. Ahmed, W.S. Qureshi, Automated analysis of visual leaf shape features for plant classification, Comput. Electron. Agric. 157 (2019) 270−280, https://doi.org/10.1016/j.compag.2018.12.038.
[7] M. Keivani, J. Mazloum, E. Sedaghatfar, M.B. Tavakoli, Automated analysis of leaf shape, texture, and color features for plant classification, Trait, Signal 37 (2020) 17−28, https://doi.org/10.18280/ts.370103.
[8] Y. Herdiyeni, D.I. Lubis, S. Douady, Leaf shape identification of medicinal leaves using curvilinear shape descriptor, in: Proc. 2015 7th Int. Conf. Soft Comput. Pattern Recognit. SoCPaR 2015, Fukuoka, 2016, pp. 218−223, https://doi.org/10.1109/SOCPAR.2015.7492810.
[9] I. Ariawan, Y. Herdiyeni, I.Z. Siregar, Geometric morphometric analysis of leaf venation in four shorea species for identification using digital image processing, Biodiv. J. Biol. Divers. 21 (2020) 3303−3309, https://doi.org/10.13057/biodiv/d210754.
[10] A. Bakhshipour, A. Jafari, Evaluation of support vector machine and artificial neural networks in weed detection using shape features, Comput. Electron. Agric. 145 (2018) 153−160, https://doi.org/10.1016/j.compag.2017.12.032.
[11] A.M. Fernandes, A.B. Utkin, J. Eiras-Dias, J. Cunha, J. Silvestre, P. Melo-Pinto, Grapevine variety identification using "Big Data" collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks, Comput. Electron. Agric. 163 (2019) 1−9, https://doi.org/10.1016/j.compag.2019.104855.
[12] A.S. Al-Fallooji, A. Al-Azawei, Predicting users' personality on social media: a comparative study of different machine learning techniques, Karbala Int. J. Mod. Sci. 8 (2022) 617−630, https://doi.org/10.33640/2405-609X.3262.
[13] A. Aruraj, A. Alex, M.S.P. Subathra, N.J. Sairamya, S. Thomas George, S.E. Vinodh Ewards, Detection and classification of diseases of banana plant using local binary pattern and support vector machine, in: 2nd Int. Conf. Signal Process. Commun. ICSPC 2019 - Proc., 2019, pp. 231−235, https://doi.org/10.1109/ICSPC46172.2019.8976582.
[14] R.K. Sevakula, N.K. Verma, Support vector machine for large databases as classifier, in: B.K. Panigrahi, S. Das, P.N. Suganthan, P.K. Nanda, eds., Swarm Evol. Memetic Comput., Springer Berlin Heidelberg, Berlin, Heidelberg. 2012, pp. 303−313, https://doi.org/10.1007/978-3-642-35380-2_36.
[15] I. Ariawan, Geometry feature extraction of shorea leaf venation based on digital image and classification using random forest, Int. J. Comput. Digit. Syst. 11 (2022) 141−150, https://doi.org/10.12785/ijcds/110111.
[16] H. Liu, X. Ma, M. Tao, R. Deng, K. Bangura, X. Deng, C. Liu, L. Qi, A plant leaf geometric parameter measurement system based on the android platform, Sensors 19 (2019) 1872, https://doi.org/10.3390/s19081872.

[17] G. Saleem, M. Akhtar, N. Ahmed, W.S. Qureshi, Automated analysis of visual leaf shape features for plant classification, Comput. Electron. Agric. 157 (2019) 270−280, https://doi.org/10.1016/j.compag.2018.12.038.

[18] Y. Liu, Y. Li, J. Song, R. Zhang, Y. Yan, Y. Wang, F.K. Du, Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: quercus dentata Thunberg and Quercus aliena Blume (Fagaceae), Ann. For. Sci. 75 (2018) 90, https://doi.org/10.1007/s13595-018-0770-2.

[19] M.B. Bangun, Y. Herdiyeni, E.N. Herliyana, Morphological feature extraction of Jabon's leaf seedling pathogen using microscopic image, TELKOMNIKA Telecommun. Comput. Electron. Control 14 (2016) 254, https://doi.org/10.12928/telkomnika.v14i1.2486.

[20] W. Li, Z. Liu, A method of SVM with normalization in intrusion detection, Proc. Environ. Sci. 11 (2011) 256−262, https://doi.org/10.1016/j.proenv.2011.12.040.

[21] T.-T. Wong, P.-Y. Yeh, Reliable accuracy estimates from *k* -fold cross validation, IEEE Trans. Knowl. Data Eng. 32 (2020) 1586−1594, https://doi.org/10.1109/TKDE.2019.2912815.

[22] R. Amami, D. Ben Ayed, N. Ellouze, Practical selection of SVM supervised parameters with different feature representations for vowel recognition, Int. J. Digit. Content Technol. Appl. 7 (2013) 418−424, https://doi.org/10.4156/jdcta.vol7.issue9.50.

[23] X. Deng, Q. Liu, Y. Deng, S. Mahadevan, An improved method to construct basic probability assignment based on the confusion matrix for classification problem, Inf. Sci. 340−341 (2016) 250−261, https://doi.org/10.1016/j.ins.2016.01.033.

[24] J. Han, K. Micheline, P. Jian, Data Mining Concepts and Techniques, Morgan Kaufman, New York (US), 2012.

[25] M. Valan, K. Makonyi, A. Maki, D. Vondráček, F. Ronquist, Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks, Syst. Biol. 68 (2019) 876−895, https://doi.org/10.1093/sysbio/syz014.

[26] J.S. Cope, D. Corney, J.Y. Clark, P. Remagnino, P. Wilkin, Plant species identification using digital morphometrics: a review, Expert Syst. Appl. 39 (2012) 7562−7573, https://doi.org/10.1016/j.eswa.2012.01.073.

[27] X. Li, X. Zhao, H. Ma, B. Xie, Image analysis and diagnosis of skin diseases - a review, Curr. Med. Imag. Former. Curr. Med. Imag. Rev. 19 (2023) 199−242, https://doi.org/10.2174/1573405618666220516114605.

[28] P. Barré, B.C. Stöver, K.F. Müller, V. Steinhage, LeafNet: a computer vision system for automatic plant species identification, Ecol. Inform. 40 (2017) 50−56, https://doi.org/10.1016/j.ecoinf.2017.05.005.

[29] A. Ambarwari, Q.J. Adrian, Y. Herdiyeni, I. Hermadi, Plant species identification based on leaf venation features using SVM, TELKOMNIKA Telecommun. Comput. Electron. Control 18 (2020) 726, https://doi.org/10.12928/telkomnika.v18i2.14062.

[30] R.I. Borman, F. Rossi, Y. Jusman, A.A.Abd Rahni, S.D. Putra, A. Herdiansah, Identification of herbal leaf types based on their image using first order feature extraction and multiclass SVM algorithm, in: 2021 1st Int. Conf. Electron. Electr. Eng. Intell. Syst. ICE3IS, IEEE, Yogyakarta, Indonesia. 2021, pp. 12−17, https://doi.org/10.1109/ICE3IS54102.2021.9649677.

[31] B.R. Hussein, O.A. Malik, W.-H. Ong, J.W.F. Slik, Automated classification of tropical plant species data based on machine learning techniques and leaf trait measurements, in: R. Alfred, Y. Lim, H. Haviluddin, C.K. On, eds., Comput. Sci. Technol., Springer Singapore, Singapore. 2020, pp. 85−94, https://doi.org/10.1007/978-981-15-0058-9_9.

[32] E. Yigit, K. Sabanci, A. Toktas, A. Kayabasi, A study on visual features of leaves in plant identification using artificial intelligence techniques, Comput. Electron. Agric. 156 (2019) 369−377, https://doi.org/10.1016/j.compag.2018.11.036.

[33] S. Mahajan, A. Raina, X.-Z. Gao, A. Kant Pandit, Plant recognition using morphological feature extraction and transfer learning over SVM and AdaBoost, Symmetry 13 (2021) 356, https://doi.org/10.3390/sym13020356.