Novel Outline Tracing Techniques for Leaf Species Identification from Shrouded Leaves

Dipak Pralhad MAHURKAR, Hemant PATIDAR

Electronics and Communication Engineering, Oriental University Indore, India

dipsmahu11@gmail.com, hemantpatidar08@gmail.com

Abstract: Life on Earth depends heavily on plants. Various methods have been developed to extract a plant leaf's attributes. In order to extract the characteristics, several classification techniques are used. Because there are so many different kinds of plants in the world, it may be difficult to remember their names; a method for identifying plant types was developed. The conventional way to identify a plant is by its leaves. To extract the leaf-based characteristics from the leaf picture, image processing techniques are applied. Eventually, leaf identification was accomplished using machine learning techniques. This article contrasts the methods used by other authors that have worked on various occluded leaf species identification methodologies with this recommended method for identifying leaf species.The unique method for expressing forms described in this article is termed "Tracing Leaf Outline" (TLO). The TLO approach has a low processing complexity and is also reasonably easy to apply. It is demonstrated that this technique can identify the leaf species although there is a sizable proportion of leaf occlusion (let's say, about 50% occlusion). Standard datasets for Flavia leaf species have been used to test this method. The efficiency of the proposed technique is evaluated using accuracy, precision, recall, and F1 score. These measures have produced a result of 99.6%, which is greater than all prior findings, and have significantly exceeded the traditional methodologies for identifying plant leaf species.

Keywords: Shrouded leaf, Feature extraction, Leaf Classification, Plant Species Identification, Tracing Leaf Outline (TLO)

1. Introduction

The importance of plants to human existence on Earth cannot be overstated. Vegetables, fruits, flowers, and other types of plants are among the enormous diversity of plants. The leaves are considered while identifying a plant. The same may be captured with a camera, a cell phone, or scanners since leaves are just accumulated everywhere. Through computer-based plant identification, the species of leaves are identified. For safeguarding biodiversity, species knowledge is crucial. For non-experts, plant identification using traditional keys is difficult, expensive in time, and frustrating owing to the usage of certain agricultural words. For those who are just getting started learning about species, this presents a challenging hurdle. The idea of automating the procedure for identifying species is becoming more and more popular today. Thanks to the accessibility and pervasiveness of technology trends like mobile phones, digital cameras, remote database access, and updated Algorithms for image processing and pattern recognition, the concept of automated species identification is already a reality. The problem addressed in the paper is the identification of plant species from images of leaves that may be partially shrouded or occluded.

The problem addressed in the paper is the identification of plant species from images of leaves that may be partially shrouded or occluded. The paper proposes a computational approach that combines image processing techniques and machine learning algorithms to overcome the challenges posed by shrouded leaf images. The authors propose an innovative solution that leverages their novel Tracing Leaf Outline (TLO) techniques. These innovative techniques enable the extraction of partial contours from full leaf contours that correspond to shrouded regions. By effectively identifying and matching these partial contours, the proposed method significantly enhances the ability to recognize plant species from shrouded leaf images. Through extensive experimentation and meticulous comparisons with existing methods, the researchers demonstrate the good performance of their approach in identifying plant species accurately. The method surpasses the utilization of the TLO technique on renowned public leaf image datasets such as Flavia, which serve as benchmarks in the field. The author’s method not only outperforms existing techniques but also showcases superior results across various evaluation metrics. By employing the

https://doi.org/10.33436/v33i2y202306

TLO technique, the proposed approach excels in capturing the unique characteristics of shrouded leaf images, enabling more precise and reliable plant species identification

The subsequent sections of the paper are outlined as follows. The related work is discussed in Part 2, the proposed methodology is included in Part 3, the experimental results are included in Part 4, and the conclusion and suggestions for further developing the suggested techniques are included in Part 5.

2. Related work

In the publication (Truong et al., 2020), In the publication (Truong et al., 2020), two methods for classifying the picture leaves are proposed by combining a shallow and deep architecture for image processing. For shallow architecture identification, For shallow architecture identification, there are two possibilities: either manually identify the parameters or use feature extraction that was manually produced. By including the effects of horizontal reflection and rotation into datasets, deep architecture significantly improves recognition results. How accurate shallow architecture depends on both the feature vector length and the image input size. The deep architecture features a shorter feature vector that is found by the model itself for more accuracy with the same input picture stimulation. The process of creating the model and selecting the right parameters only requirements for the recognition procedure. In terms of the effectiveness of the categorization process, identification and rearrangement of picture features a labour and time-intensive procedure no longer strongly affect the discovery. Vein Features, Colour Descriptors, Fourier Descriptors, and the GLCM method for feature extraction are all used in the paper's (Turkoglu & Hanbay, 2019) suggested system in addition to the ELM proposed technique as a classifier. In this study, numerous viewpoints were produced in order to characterise the plant. With the use of feature extraction techniques, division procedures have been applied to leaf image datasets to identify properties. Each piece is divided into 2 and 4 sections, from which the leaf characteristics are taken independently. The feature vectors that make up the whole leaf picture are made using the parameters acquired from each component together. Additionally, certain leaf image properties might be extracted without segmentation by the use of vein characteristics. The ELM approach was later employed to carry out the categorising procedure. While colour features and the GLCM approaches had the best precision of the four-section approach, the Fourier descriptor (FD) approach had the highest accuracy in the case of the bisection technique. Additionally, during the classification step, the characteristics derived from the methods of feature extraction were integrated. These findings demonstrated that non-splitting obtained 97.68% identification accuracy whereas bisection achieved 99.10%. A powerful method for recognising plant leaves based on morphological traits and adaptive boosting approaches has been developed (Kumar et al., 2019). The experimental findings are produced using three distinct categorization methods: decision trees, multilayer perception, and k-NN. The Ada Boost method is believed to improve the recommended system's accuracy. The best accuracy rate that was able to be attained was 95.42% for 32 different varieties of plant leaves.

The greatest outcome in this investigation, according to the authors (Tan et al., 2020), was a testing accuracy of 94.88% when employing classifier ANN and D-Leaf for feature extraction. Furthermore, the D-Leaf's performance approach is verified using CV techniques, and fresh datasets from Flavia, Swedish, and Malaya. The validation performance of the D-Leaf approach (>93%) showed that it is appropriate for automated plant species categorization. It concluded that the DLNN method beat the SVM algorithm after the authors (Muneer & Fati, 2020) studied two algorithms, DLNN and SVM, using the same dataset. As a result, it can be concluded that, except the oral backdrop, changing the background colour has no impact on the categorization system. The accuracy reached in various backgrounds testing is also fairly similar to one another. The suggested approach can reach an accuracy of 52.50%. The suggested method also showed 99% and 97% accuracy on FLAVIA and Malaysian dataset tests, respectively. The suggested approach was the most accurate since it took into account the many texture and form features included in the FLAVIA dataset. The authors (Hu et al., 2018) investigate plant leaf identification and propose the Multi-Scale Fusion Convolutional Neural Network (MSFCNN). The proposed MSF-CNN divides an input image into a sequence of low-resolution pictures, which are then successively fed into the MSF-CNN architecture to train discriminative features at different levels. A concatenation process combines feature maps learned on pictures of multiple scales from a channel view to accomplish feature fusion between two scales. MSF-final The CNN layer gathers all discriminative information to produce the final feature, which is then used to determine the kind of plants in the input picture. The image-to-image translation model of the Generative Adversarial Networks (GAN) was investigated by the authors (Kanda, Xia & Sanusi, 2021to improve the performance of an automated plant classification system using photographs of plant leaves. The precision with which the model was able to produce artificial pictures, as well as the conditional Generative Adversarial Network's accuracy in creating augmented data. During the first 300 epochs of training, the network struggled to produce pictures with the proper shape that were as similar to the source images as was practical, but it eventually made progress in terms of leaf shape. Integral Contour Angles (ICA), a brand-new approach to form description, was put out by the authors (Feng & Wang, 2018). The suggested ICA descriptions all have the following desired qualities: (1) Its intrinsic invariance to group transformations, such as translation, rotation, and uniform scaling, has been shown in theory. (2) Because the ICA descriptors are produced using integral operation, they are noise-resistant and unaffected by undesirable geometric perturbations (3). Because it is multi-scale, it naturally extracts shape information from coarse to fine scales, making it a powerful discriminator for leaf identification. Additionally, because they only depend on the area around the contour point, the recommended ICA descriptors can handle the recognition of occluded forms because they define the good local geometric quality of shape.

The authors (Jeon & Rhee, 2017) utilized Google Net to generate 2 models by varying the depth of the network and developed a unique method for categorizing leaves using the CNN model. Based on the damage to the leaves, the performance of all models was compared. The identification rate was better than 94% even with 30% of the leaf damaged. The authors (Lukic, Tuba & Tuba, 2017) described a plant categorization method that uses leaf recognition An SVM that was fine-tuned using hierarchical grid search as a classifier was employed. In other cases, invariant Hu moments are insufficient as a means of categorization since various species' leaves have extremely similar shapes. Additionally, metrics from the histogram of a uniform local binary pattern, such as entropy, mean, energy and standard deviation were employed, to resolve these instances. The suggested approach was evaluated using the industry-accepted benchmark, the accuracy of the Flavia data set was 94.13%, which is superior to other methods described in the literature. The writers (Li et al., 2019) by fusing 3D filtering with facet region growth, provide a technique for non-overlapping segmentation of individual plant leaf point clouds. The method is divided into three stages. An innovative Operator for 3D joint filtering is utilised to eliminate occluded and overlapping areas from the point cloud's leaves before pre-segmenting the remaining leaf centres into fundamental leaf sets. Second, a facet over-segmentation technique is used for the previously filtered areas of the 3D joint filtering to obtain a set of facets that organise the filtered points into clusters. The previously pre-segmented leaf centre is then given the over-segmented facets back. Finally, starting at the centre of the labelled leaf and scanning for unmarked neighbouring facets, labels from the inside out are grown. Once a label has been given to each facet, segmentation is complete. The investigations make use of four distinct plant point clouds generated by three different types of 3D imaging hardware. According to the findings, the proposed approach for extracting individual leaves from dense plant point clouds is both feasible and effective. Also, techniques for determining the breadth, length, and area of leaves are provided.

The authors (Wang, Pan & Pan, 2019), to increase the accuracy of a KNN classifier, a classification ability ranking (CAR)-based TDC technique was suggested. The proposed classification ability function scores training samples based on their contribution to correctly classifying other training samples as a KNN using the leave-one-out (LV1) approach. As a result, the training sample is removed from the original training data set due to its inferior classification capacity. This training sample most likely incorrectly identifies the other samples when using the LV1 technique with KNN classifications. Extensive testing on 10 real-world data sets demonstrates that the proposed CAR-based TDC technique may significantly cut the classification error rates of KNN-based classifiers while reducing computational complexity to a smaller training data set. The authors (Yang & Wei, 2019) presented TDR as a fresh method for describing shapes. Two matrices are used to express the proposed TDR: A sign matrix with a triangular centre distance matrix. The sign matrix describes the concave and convex properties of a shape, whereas the triangle centre distance matrix provides the bending degree and spatial distribution information of a shape. The TDR descriptor is a portable multi-scale representation approach that can describe both a shape's local and global properties and is resistant to similarity modifications. Only data on plant leaf shape was utilised. The authors (Zeng et al., 2019) explain: a) To boost the description capacity of local descriptors, a multi-scale curvature integral descriptor is developed. b) Encoding of the curvature descriptor, resulting in a semantic description feature that is intermediate in nature. This gets past the limitation that the shape-matching sample points must be in the same relationship. b) Pooling equal-curvature integral ranks to improve the effectiveness of middle-level descriptors and feature identification. The authors (Choudhary & Barron, 2020) demonstrate a method for categorising plant species from an overlapping leaf picture by using a database of known species with discrete types of leaves. The writers (Twum et al., 2022) utilised using the two techniques, textural characteristics for categorization were retrieved. The Gabor and Log Gabor filters were extracted using two separate filter banks. Multi-feature fusion was recommended by the authors (Wäldchen & Mäder, 2018) in order to increase classification accuracy. The authors (Zhang et al., 2018) proposed SVD and SR as a method for automatically identifying plant species. The proposed technique calculates the estimated SR of a given input leaf picture by first creating an over-complete lexicon from plant leaf photos without any training. The method may meet the requirements for a true system of plant categorization. The results of the experiments on two datasets of plant leaf photos show how useful and practical the recommended technique is. The writers (Mahurkar & Patidar, 2022) examining every categorization method, it was discovered that the closest neighbour strategy was the most straightforward (Flavia, 2019). Suggests that you may identify a plant by merely entering its leaf image, which was taken with a digital camera or scanner. 32 plants can be classified by this approach. The average accuracy for all of them is 93%.

3. Proposed methodology

The suggested machine learning-based technique for recognising plant leaf species is depicted in Figure 1. The input dataset often undergoes preprocessing procedures including contrast enhancement, normalisation, data augmentation, etc. The finished dataset is then split into training and testing halves, which generally includes thousands of photographs. Using training data, which generally comprises about 80% of the whole dataset, thousands of parameters are learned. The trained model is then tested using test data that was omitted (usually 20% of the test data), allowing the performance of the learnt model to be approximated. Using these performance indicators, an assessment of how well the model fits the supplied data before generating a result.

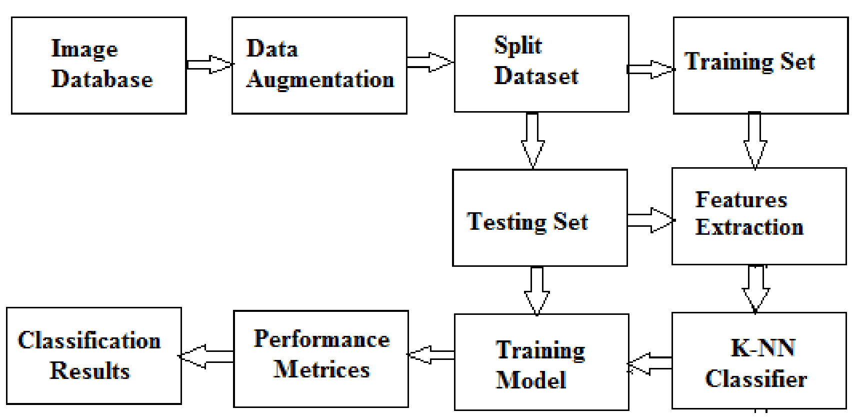


Figure 1. Block diagram of a Machine Learning based Plant leaf species identification model

3.1. Applied Classification

Bypassing the issue of probability densities entirely, the single closest neighbour strategy simply assigns an unknown sample point in the same class as the closest comparable or "nearest" sample point in the training set of data, sometimes referred to as a reference set. The term "nearest" can be interpreted to indicate the lowest Euclidean distance, which is the typical an unknown sample point in the same class as the closest comparable or "nearest" sample point in the training set of data space a= $a_1, a_2\_\_\_\_\_a_n$ and b= $b_1, b_2\_\_\_\_\_b_n$ is defined by

$$d_e(a,b) = \sqrt{\sum_{i=1}^{i=n}(b_i - a_i)^2} \tag{1}$$

n refers to the number of features.

Despite being the distance function or measure of dissimilarity between feature vectors, euclidean distance is not always the best metric that is most frequently utilised. The characteristics for which the dissimilarity is considerable are heavily stressed due to the fact that the distances in each dimension are squared prior to summing. It could be more reasonable to utilise the total of the absolute differences between each attribute as the overall indicator of dissimilarity rather than the squares of those differences. Additionally, it would speed up computation. Then, this distance metric would

$$d_{cb}(a,b) = \sum_{i=1}^{n} |\, bi - ai| \tag{2}$$

The number of features is n. The city block distance is the total of the absolute distances in each dimension. By using a nonlinear function, such as the square root of the absolute values of the individual feature differences before summing, a metric that would downplay single large feature differences and be more impacted by several minor ones might be developed. The maximum distance measure is an extreme metric that only takes into account the most different pair of attributes.

$$d_m(a,b) = \max_{i=1}^{n} |\, bi - ai| \tag{3}$$

The Minkowski distance, which is defined by (1), (2), and (3), is a generalization of the three distances.

$$d_{cb}(a,b) = \left[\sum_{i=1}^{n} |\, bi - ai|^r\right]^{\frac{1}{r}} \tag{4}$$

Where *r* is an adjustable parameter.

The K-NN employed in picture 2 may be explained using the following algorithm (KNN: Algorithm):

Step 1: Decide on the neighbor's K-numbers.
Step 2: Calculate the Euclidean distance between K neighbors' in step two.
Step 3: Based on the determined Euclidean distance, select the K closest neighbors.
Step 4: Count the number of data points in each category among these k neighbors.
Step 5: Assign the fresh data points to the category where the neighbor count is highest.
Step 6: Our model is complete.

Figure 2. K-NN Algorithm

Take into account the situation when a fresh data point must be classified to use it, as indicated in Figure 3 below.

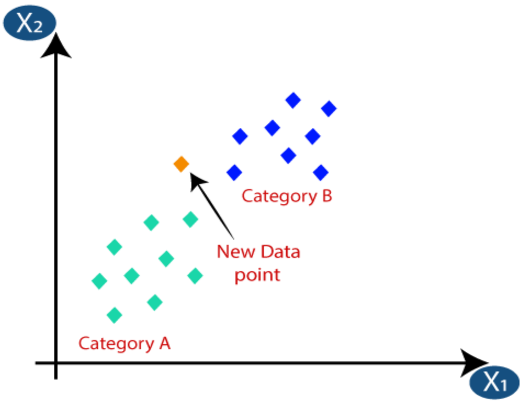


Figure 3. When a new data point is added

- K=5 to represent the number of neighbours.
  Then, the Euclidean separation between the data points will be determined and it is mentioned in the equation number 5. The distance between two points is known as the Euclidean distance, which is studied in geometry.

Euclidean distance between A1 and B1 = $\sqrt{(X2 - X1)^2 + (Y2 - Y1)^2}$ (5)

- The Euclidean distance was computed to determine "the nearest neighbours." Figure 4 shows that Category B had two nearest neighbours while Category A had three.

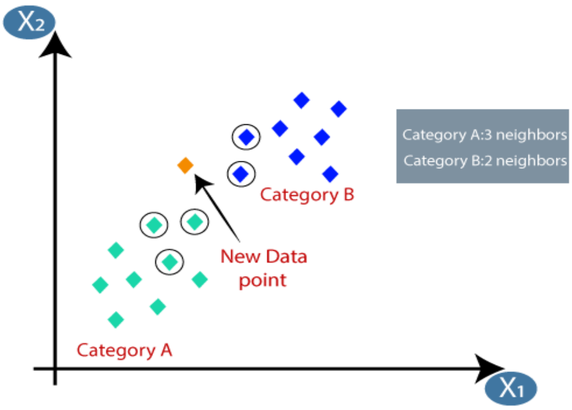
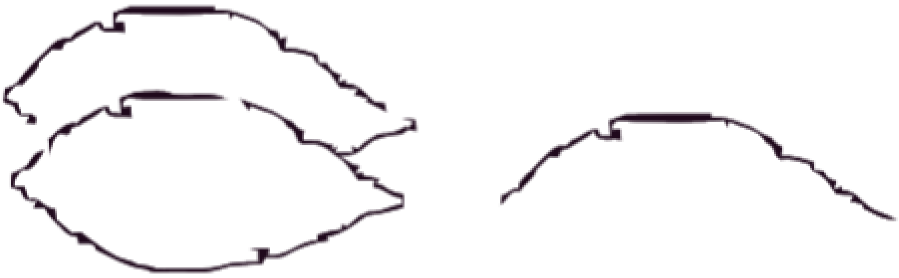


Figure 4. Data point added in category A

- Considering that the three closest neighbours are in category A, this new data point must be in the same category (KNN: Algorithm KNN Algorithm was chosen due to various following reasons:
  a) KNN provides interpretable results by directly associating the prediction with the most similar instances in the training dataset.
  b) KNN allows the use of various distance metrics, such as Euclidean distance, Manhattan distance, or Minkowski distance, to measure the similarity between data points. This flexibility allows the algorithm to handle different types of data and capture relevant patterns effectively.
  c) KNN is relatively robust to outliers since it considers the neighbours'votes when making predictions. Outliers have a minimal impact on the decision boundary or the regression line as long as the number of neighbours (k) is appropriately chosen.

3.2. Proposed technique

Tracing Leaf Outline (TLO) shape description technique is a newly suggested approach. For this procedure, a digital pen pad is used to sketch the shape of the leaf. The first step is to obtain the standard leaf database from the website The Flavia leaves database was chosen, which is publicly available on the Internet (Flavia, 2019). The Flavia leaf database was chosen based on the highest number of species as well as the overall quantity of leaves. When compared to other databases, such as Swedish, the Flavia database has more species, having a total of 1907 leaves. The concept of finding reliable results requires a big number of leaf species, along with a huge number of total leaves, which is accessible in the Flavia leaf database. The next step is to find species kinds with comparable orientation patterns, which implies the leaf should be parallel to the x-axis. Ten different sorts of leaf species have been identified from Flavia's leaf database. Then, in the graphic editor paint, open one by one leaf picture. Then connect the digital pen pad to the computer and draw an outline around 50% of the leaf portion, i.e. 0 to 180. Because 50% of the leaf is obscured, as shown in Figure 5. Next, draw a point on the contour of a leaf that will be useful for determining the gradient of a line, as illustrated in Figure 6. Then, for each line, calculate the gradients. This gradient is nothing but a set of characteristics for the algorithm, which is saved as a database. The same procedure is followed for all of the chosen species' leaves. Using a data augmentation strategy, enlarge the database to 1000 total leaves with 10 different species by varying the gradient values from +/- 1% to +/- 3%. Now, save this database on Google Drive. After execution of the machine learning program output is generated and this is just the identification of leaf species.



(a) (b)

Figure 5. Leaf No. 2187 of (a) 3D view of occluded 50% part of the leaf (b) Leaf outline trace by the digital pen

Figure 6. Leaf No. 2187 - Point mark on the outline of the leaf

4. Experimentation findings

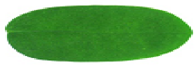
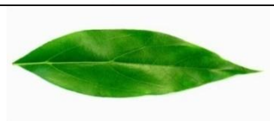
The suggested approach is implemented and tested on a Windows 8.1 64-bit platform equipped with an i3-3120m CPU running at 2.50 GHz and 8GB of RAM The experiment was based on the leaf database, which is available on the internet. Table 1 displays information on the available leaf databases.

Table 1. Available Leaf Dataset

Sr. No	Name of the Dataset	No. of Species	Total no. of Leaves
1	Flavia	32	1907
2	Swedish	15	1125

Flavia dataset was picked from the database since it has a greater number of species. Then, from the Flavia database, choose the leaf numbers 1131, 1211, 1463, 1498, 2023, 2089, 2123, 2187, 2518, 3447, as shown in Table 2.

Table 2. Selected Leaves from the Flavia dataset

Sr. No	Leaf Species Diagram	Leaf No in Database
1		1131
2		1211
3		1463
4		1498
5		2023
6		2089
7		2123
8		2187
9		2518
10		3447

In Table 3 calculated selective leaf species slops are mentioned. Use this data for data augmentation method then create a leaves database.

Table 3. Leaf slops details

TYPE	Species No.	AB	BC	BD	DE	EF	FG	GH	HI	IJ	JK
1	1131	-2.43	6	2.5	0.78	0.35	0	-0.58	-1.5	-1	-0.55
2	1211	0.8	0.2	0	-0.33	-0.69	--	--	--	--	--
3	1463	0.13	0.47	0.2	-0.14	-0.5	--	--	--	--	--
4	1498	1.75	0.92	0.45	-0.04	-0.44	-0.83	-1	-0.5	--	--
5	2023	0.43	0.44	0.05	-0.33	-0.8	-1	--	--	--	--
6	2089	3	1.33	0.67	0.08	-0.08	-0.75	-1	-3.4	--	--
7	2123	8	0.67	0.13	-0.04	-0.3	-1.71	--	--	--	--
8	2187	1.5	0.73	0.33	0	-0.25	-0.81	-0.27	-1.5	--	--
9	2518	1.05	0.6	0.11	-0.12	-0.62	-0.44	-6	--	--	--
10	3447	-6.25	1.28	0.28	-0.09	-0.37	-0.73	-1	-0.73	--	--

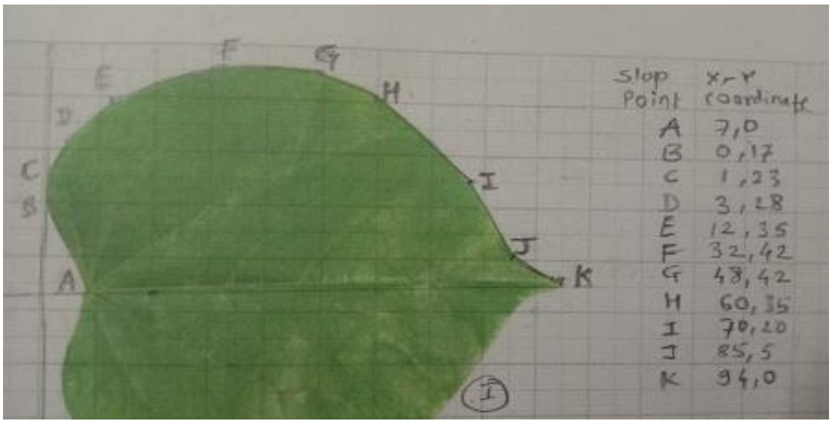


Figure 7. Marked slop point

The first derivative (FD) and second derivative (SD) of the selected first leaf no1131 slopes are presented in Figure 8 and 9, respectively. In Table 4 slope details of sample leaf no 1131 are mentioned and Figure 7 shows after the marked slope,the point on the contour of the leaf.

Table 4. Leaf no. 1131 slopes details

Slope	AB	BC	BD	DE	EF	FG	GH	HI	IJ	JK
FD	-2.43	6	2.5	0.78	0.35	0	-0.58	-1.5	-1	-0.55
SD		-8.43	3.5	1.72	0.43	0.35	0.58	0.92	-0.5	-0.45

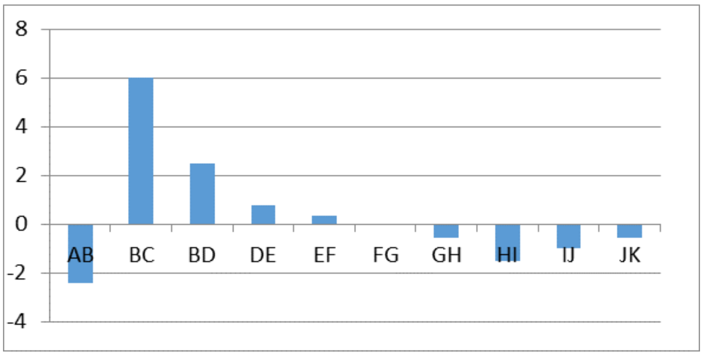


Figure 8. First Derivative of Leaf no. 1131

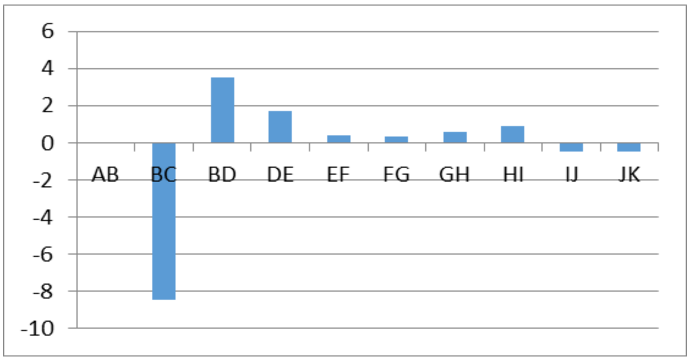


Figure 9. Second Derivative of Leaf no. 1131

The impulse function is the derivative of a discontinuous function, and its magnitude is equal to the difference between the two amplitudes at a particular location of the discontinuity. In the graph, a magnitude greater than zero indicates a positive slope, and the leaf contour is growing, while a value less than zero shows a negative slope, and the leaf contour is dropping. In Fig.10a, 11b, and 11c, all leaf species characteristics are listed. The slops point's markings were made on the obscured area of the leaf with a digital pen pad. Then compute the slopes of each leaf species, then determine the second derivative specified in the figure below, and then plot the graphs to conclude that when the graph magnitude is positive, the contour is growing, and vice versa for negative.

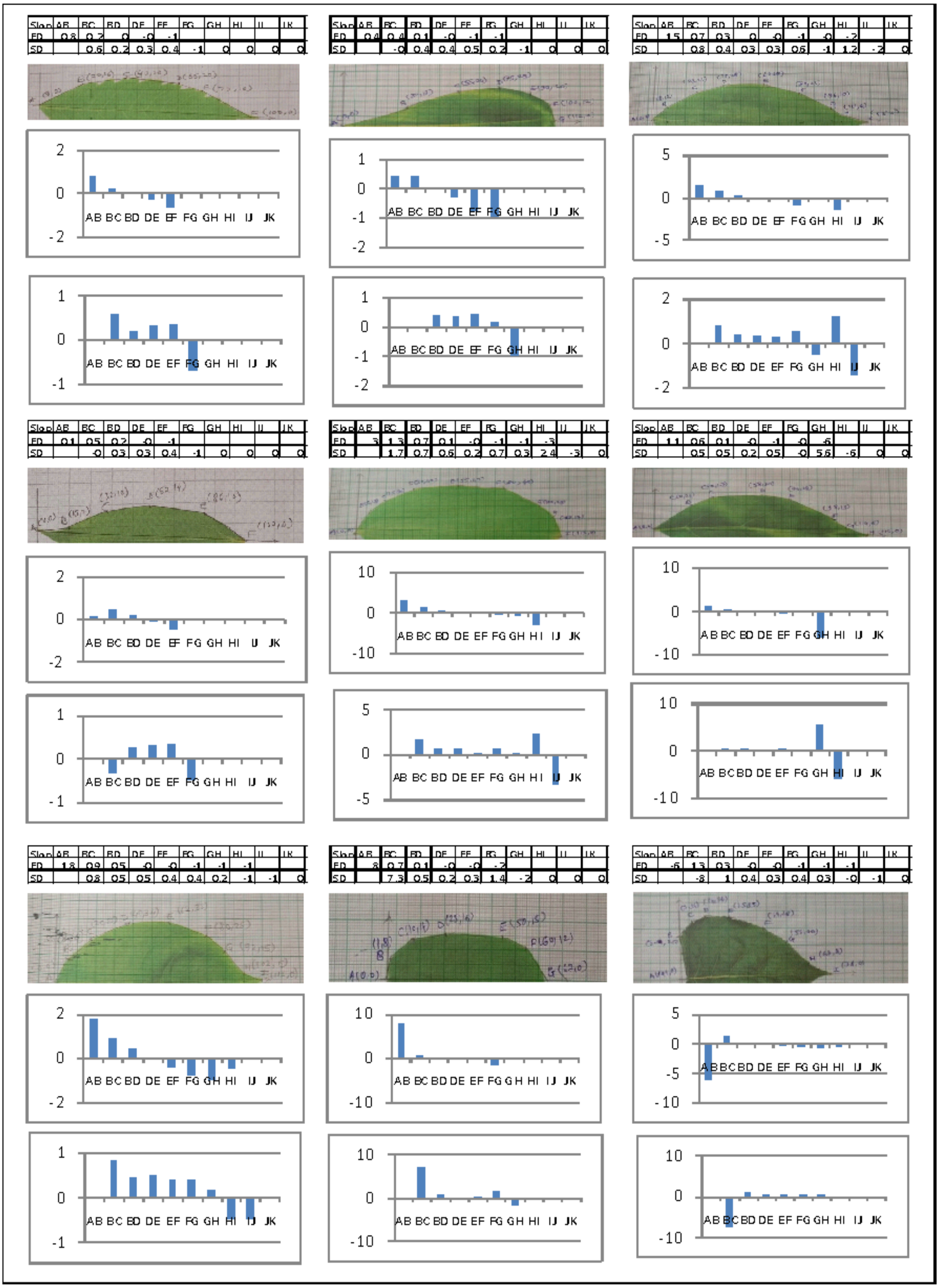


Figure 10. a) Slopes of Leaf numbers 1463, 1498, 2023 with FD and SD graph;
b) Slopes of Leaf numbers 2089, 2518, 1211 with FD and SD graph;
c) Slopes of Leaf numbers 2123, 2187, and 3447 with FD and SD graph

As the final step of the outcome, data augmentation was employed and produced a 1000-leaf database, which was split into an 80:20 ratio for training and testing. Table 5 displays the results after the applied K-NN classifier.

Table 5. Accuracy Rate comparison of several classifiers applied on the Flavia leaf dataset

Method	Feature	Classifier	Training data	Testing data	Accuracy
(Tan et al., 2020)	Leaf Vein Morphometric	CNN	80%	20%	94.63%
(Muneer & Fati, 2020)	Shape features	DLNN	60%	40%	99%
(Turkoglu & Hanbay, 2019)	Fourier Descriptors, Vein features, and Color features, with Gray-Level Co-occurrence Matrix	ELM	90%	10%	99.10%
(Kanda, Xia & Sanusi, 2021)	Discriminative features	DL+LR	75%	25%	99.58%
(Twum et al., 2022)	Shape Features	SVM	Not Mentioned	Not Mentioned	97%
Proposed Method	Geometrical and shape	K-NN	80%	20%	99.6 %

From Table 5, the author (Tan et al., 2020) exhibited 94.63% accuracy using the CNN classifier, (Muneer & Fati, 2020) demonstrated 99% accuracy using the DLNN classifier, (Turkoglu & Hanbay, 2019) displayed 99.10% accuracy using the ELM classifier, and **(**Kanda, Xia & Sanusi, 2021) demonstrated 99.58% accuracy using the DL+LR classifier. According to (Twum et al., 2022), the SVM classifier is 97% accurate.

Based on the tracing contour of the leaf, geometrical and form characteristics are derived. Following that, K-NN was chosen from the available classifiers since it is a straightforward method to understand, effective for regression and classification, and has high accuracy. Another consideration is data separation for training and testing. After experimenting with several split ratios, the 80:20 split ratio had the better output. After running the programme, the accuracy of the proposed approach utilising K-NN Classifier was 99.6%. Figure 12 depicts the outcome in comparison to the prior results. Figure 11 shows a rate comparison for the Flavia leaves dataset utilising the recommended approach with different classifiers such as CNN, DLNN, ELM, DL+LR, ANN, and SVM.

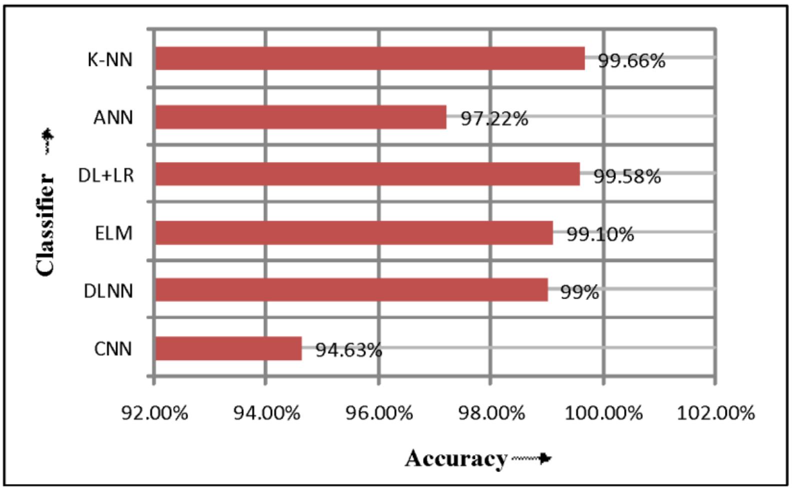


Figure 11. A rate comparison for the Flavia leaves dataset is shown

5. Conclusion and future scope

This study established a system for automated plant species identification. The existing authors' papers acknowledge the limitation of a limited dataset for training and evaluation To address the constraints of a limited dataset, data augmentation techniques are proposed. Data augmentation involves generating new training samples by applying varying the gradient values from +/- 1% to +/- 3%. For recognizing distinct plant species, a form description approach called Tracing Leaf Outline (TLO) is suggested. Tracing the shape of a leaf on a computer screen using a digital pen pad. TLO performs similarly to a swype keyboard on mobile phones. The TLO technique is an efficient method for accurately identifying leaf species, even in situations where the leaves are partially shrouded. The study's phases comprised picture image segmentation, feature extraction, and finally picture classification during pre-processing. Here classification was carried out using a K-NN classifier. The outcome of the machine learning programme running is then generated. The Flavia dataset was utilised for testing, and the system attained an average accuracy of 99.6%. The experiment produces a result of 99.6%, which is greater than all prior findings. This approach is likewise simple to develop and has a minimal amount of processing complexity. The applications of outline tracing techniques extend beyond leaf species identification and can be applied in various fields, supporting research, conservation, agriculture, and ecological studies.

REFERENCES

Choudhary, A. & Barron, J. L. (2020) Plant Species Identification from Occluded Leaf Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 17(3), 1042-1055. doi: 10.1109/tcbb.2018.2873611.

Feng, N. & Wang, B. (2018) Integral contour angle: An invariant shape descriptor for classification and retrieval of leaf images. In *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP).* Athens, Greece, 7–10 October 2018, pp. 1223–1227. doi: 10.1109/ICIP.2018.8451605.

Flavia user manual, version alpha, (2019), http://flavia.sf.net.

Hu, J. Chen, Z. Yang, M. Zhang, R. & Cui, Y. A (2018) Multiscale fusion convolutional neural network for plant leaf recognition. *IEEE Signal Process. Lett.* 25(6). 853-857. doi: 10.1109/lsp.2018.2809688.

Jeon, W.S. & Rhee, S.Y. (2017) Plant Leaf Recognition Using a Convolution Neural Network. *International Journal of Fuzzy Logic and Intelligent Systems (IJFIS).* 17(1), 26–34. doi: 10.5391/IJFIS.2017.17.1.1.

Kanda, P.S. Xia, K. & Sanusi, O.H. (2021). Deep Learning-Based Recognition Technique for Plant Leaf Classification. *IEEE Access.* 9, 162590 – 162613. doi: 10.1109/access.2021.3131726.

KNN: Algorithm, https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

Kumar, M. Gupta, S. Gao, X.-Z. & Singh, A. (2019) Plant Species Recognition Using Morphological Features and Adaptive Boosting Methodology. *IEEE Access*. 7, 163912-163918, doi: 10.1109/access.2019.2952176.

Li, D. Cao, Y. Shi, G. Cai, X. Chen, Y. Wang, S. & Yan, S. (2019) An Overlapping-Free Leaf Segmentation Method for Plant Point Clouds. *IEEE Access,* 9, 129054 – 129070. doi: 10.1109/access.2019.2940385.

Lukic, M. Tuba, E. & Tuba M. (2017) Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns. In *Proceedings of the 2017 IEEE 15th International*

Symposium on Applied Machine Intelligence and Informatics, Herl0any, Slovakia, 2017, pp. 000485–000490. doi: 10.1109/sami.2017.7880358.

Mahurkar, D. & Patidar Hemant. (2022) A Study of Image Characteristics and Classifiers Utilized for Identify Leaves. *Intelligent Sustainable Systems Proceedings of ICISS 2022,* 458, 2022, pp.559-568. https://doi.org/10.1007/978-981-19-2894-9_42.

Muneer, M. & Fati, S. M. (2020) Efficient and Automated Herbs Classification Approach Based on Shape and Texture Features using Deep Learning. *IEEE Access,* 8, 196747-196764. doi: 10.1109/access.2020.3034033.

Tan, J. W. Chang, S. W. Sameem, A.-K. Yap, H. J. & Yong, K. T. (2020) Deep Learning for Plant Species Classification using Leaf Vein Morphometric. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 17(1). 82-90. doi: 10.1109/tcbb.2018.2848653.

Truong Q. B., Nguyen T. T. K., Truong Q. D. & Huynh X. H. (2020) Plant species identification from leaf patterns using histogram of oriented gradients feature space and convolution neural networks. *Journal of Information and Telecommunication*. 4(2), 140–150. doi: 10.1080/24751839.2019.1666625.

Turkoglu, M. & Hanbay, D. (2019) Recognition of plant leaves: An approach with hybrid features produced by dividing leaf images into two and four parts. *Applied Mathematics and Computation.* 352(3), 1-14. doi: 10.1016/j.amc.2019.01.054.

Twum, F. Missah, Y. M. Oppong, S. O. & Ussiph. N. (2022) Textural Analysis for Medicinal Plants Identification Using Log Gabor Filters. *IEEE Access*. 10, 83204-83220. doi: 10.1109/access.2022.3196788.

Wäldchen, J. & Mäder, P. (2018) Plant species identi_cation using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering.* 25(2), 507–543. doi: 10.1007/s11831-016-9206-z.

Wang, Y. Pan, Z. & Pan, Y. (2019) A Training Data Set Cleaning Method by Classification Ability Ranking for the k-Nearest Neighbor Classifier. *IEEE Transactions on Neural Networks and Learning Systems.* 31(5), 1544 – 1556. doi: 10.1109/tnnls.2019.2920864.

Yang, C. & Wei, H. (2019) Plant Species Recognition Using TDR. *IEEE Access*. 7, 178108-178120. doi: 10.1109/access.2019.2958416.

Zeng, J., Liu, M. Fu, X. Gu, R. & Leng. L. (2019) Curvature Bag of Words Model for Shape Recognition. *IEEE Access*. 7, 57163 – 57171. doi: 10.1109/access.2019.2913688.

Zhang, S. Zhang, C. Wang, Z. & Kong, W. (2018) Combining sparse representation and singular value decomposition for plant recognition. *Applied Soft Computing.* 67, 164-171. doi: 10.1016/j.asoc.2018.02.052.

Dipak Pralhad MAHURKAR earned a Bachelor's degree in Electronics and Telecommunication Engineering from North Maharashtra University in Jalgaon and a master's degree in Electronics Engineering from Dr. Babasaheb Ambedkar Marathwada University in Aurangabad. He is presently enrolled as a research scholar at Oriental University in Indore, India, in the Electronics and Communication Department. He is currently an Assistant Professor at Sanjivani College of Engineering in Kopargaon (Shirdi), Maharashtra, India, in the Department of Electronics and Computer Engineering. His areas of expertise include image processing, digital electronics, and communication systems.

Hemant PATIDAR was born in Ujjain, India. He completed a Bachelor of Engineering degree in Electronics & Communication Engineering from Swami Vivekananda College of Engineering, Indore, India in 2012. He completed his Ph.D. from Electronics & Communication Department, NIT Durgapur, India in 2018. Presently, he is working as an Assistant Professor, at the Department of Electronics & Communication Engineering, Oriental University, Indore, India. He has published more than 30 research papers in International Journals and Conferences. Also he is the reviewer of many reputed international journals and conferences. His research interest includes array antenna synthesis, evolutionary algorithms, electromagnetics, soft computing, etc.