

COMPARATIVE ANALYSIS USING VARIOUS IMAGE PROCESSING TECHNIQUE FOR THE MEDICAL QUALITIES IN NEEM LEAF

**Submitted in partial fulfilment of the requirements for the award of Bachelor
of Engineering Degree in Electronics and Communication Engineering**

By

GAURAV RATNU (37130140)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade “A” by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI – 600**

119 MARCH 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide
Work of

GAURAV RATNU(37130140) who carried out the project entitled "**COMPARATIVE ANALYSIS USING VARIOUS IMAGE PROCESSING TECHNIQUE FOR THE MEDICAL QUALITIES IN NEEM LEAF**" under our supervision from September 2020 to April 2021.

Internal Guide

Ms. Vijayashree T

Head of the Department

Dr. T. RAVI, M.E., Ph.D.

Submitted for Viva voce examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **Gaurav Ratnu** hereby declare that the Project Report entitled “Comparative analysis using various image processing technique for the medical qualities in neem leaf” done by me under the guidance of **Ms.Vijayashree** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

1)

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of SATHYABAMA for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. N. M. NANDHITHA, M.E., Ph.D.** Dean, School of Electrical and Electronics Engineering and **Dr. T. RAVI, M.E., Ph.D.** Head of the Department, Department of Electronics and Communication Engineering for providing us necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to our Project Guide Ms. Vijayashree T, Department of Electronics and Communication Engineering for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of the project work.

I wish to express my thanks to all teaching and Non-teaching staff members of the Department of Electronics and Communication Engineering who were helpful in many ways for the completion of the project. I express our gratitude to my parents for their constant encouragement and support for the completion of the project.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO.
1.	Introduction.	10
2.	Pre-extraction of plant samples.	14
3.	Existing system.	22
4.	Image processing in leaf pattern recognitions	26

Abstract

Plant systematics can be classified and recognized based on their reproductive system (flowers) and leaf morphology. Neural networks are one of the most popular machine learning algorithms for plant leaf classification.

The commonly used neural networks are artificial neural network (ANN), probabilistic neural network (PNN), convolutional neural network (CNN), k-nearest neighbour (KNN) and support vector machine (SVM), even some studies used combined techniques for accuracy improvement.

The utilization of several varying reprocessing techniques, and characteristic parameters in feature extraction appeared to improve the performance of plant leaf classification.

The findings of previous studies are critically compared in terms of their accuracy based on the applied neural network techniques. This paper aims to review and analyse the implementation and performance of various methodologies on plant classification.

Each technique has its advantages and limitations in leaf pattern recognition. The quality of leaf images plays an important role, and therefore, a reliable source of leaf database must be used to establish the machine learning algorithm prior to leaf recognition and validation.

CHAPTER -1

Introduction

Plants are essential for mankind. In particular, herbs have been used as folk medicines by indigenous people since ancient times.

Herbs are usually identified by practitioners based on years of experiences through personal sensory or olfactory sense. Recent advances in analytical technology have significantly assisted in herbal recognition based on scientific data. This eases many people, especially those that are lacking experience in herbal recognition.

Laboratory-based testing requires skills in sample treatment and data interpretation, in addition to time consuming procedures. Therefore, a simple and reliable technique for herbal recognition is needed.

Computation combined with statistical analysis is likely to be a powerful tool for herbal recognition. This non-destructive technique shall be the method of choice to rapidly identify herbs, particularly for those who cannot apply expensive analytical instrumentation.

Medicinal plants are gaining much interest recently because their use in ethno medicine treating common disease such as cold, fever and other medicinal claims are now supported with sound scientific evidences.

The study on medicinal plants started with extraction procedures that play a critical role to the extraction outcomes (e.g. yield and phytochemicals content) and also to the

consequent assays performed. A wide range of technologies with different methods of extraction is available nowadays. Hence, this review aim to describe and compare the most commonly used methods based on their principle, strength and limitation to help evaluating the suitability and economic feasibility of the methods.

Literature Survey

Although a significant amount of research has been done studying various aspects of leaf identification in inventory systems, most of it deals with semi-automated systems. A state-of-the-art system which is fully automated and requires least human interaction is yet to be developed.

Pavan et al. [3] proposed an algorithm for identification using multiclass classification based on color, shape volume and cell feature. They performed three stage comparisons: first stage compares redness, greenness, blueness index feature, second stage compare shape feature and the last stage compares cell feature and volume fraction feature. Experiment is performed on a sample of diverse collection of 1000 leaf and flower images.

Limitations of this approach is that it semi-automatic approach and its recognition rate is up to 85% percent on an average.

Arun Priya [4] the proposed approach consists of three phases such as preprocessing: transforming to gray scale and boundary enhancement, feature extraction: derives the common DMF from five fundamental features and classification: Support Vector Machine (SVM) classification for efficient leaf recognition. 12 leaf features which are extracted and orthogonalized into 5 principal variables are given as input vector to the SVM.

Valliamal et al. [5] A probabilistic curve evolution method with particle filters is used to measure the similarity between shapes during matching process. The experimental results prove that the preferential image segmentation can be successfully applied in leaf recognition and segmentation from a plant image.

Dr. H.B.Kekre et al. [6] the method of CBIR is discussed in this paper to filter images based on their content. In this paper feature vector is generated using color averaging technique, similarity measures and performance evaluation. Precision –Recall cross over plot is used as the performance evaluation measure to check the algorithm.

The effect due to the size of database and number of different classes is seen on the number of relevancy of the retrievals.

Javed et al. [7] used PNN to classify the plants with broad flat leaves. In this algorithm there were few select point where the user needs to specify the leaf blades and a base point according to which the image is then aligned and compared with other images on the basis of some features like area, eccentricity, etc. They used 1200 sample leaves belonging to 30 different plants to train their system. This system is also semi automatic and 91.41 percent accurate.

In literature survey it has been observed that most of the systems emphasized basically on morphological features only, some have used tooth features also. So, in this paper the new algorithm has been designed by extracting 5 geometric feature, 12 morphological features, tooth features and color features also to increase the efficiency of proposed leaf recognition system. The algorithm is explained in the next section.

3. Proposed Methodology

A typical image based plant identification system is shown in fig. 1 and the major steps are explained in consecutive sub-sections.

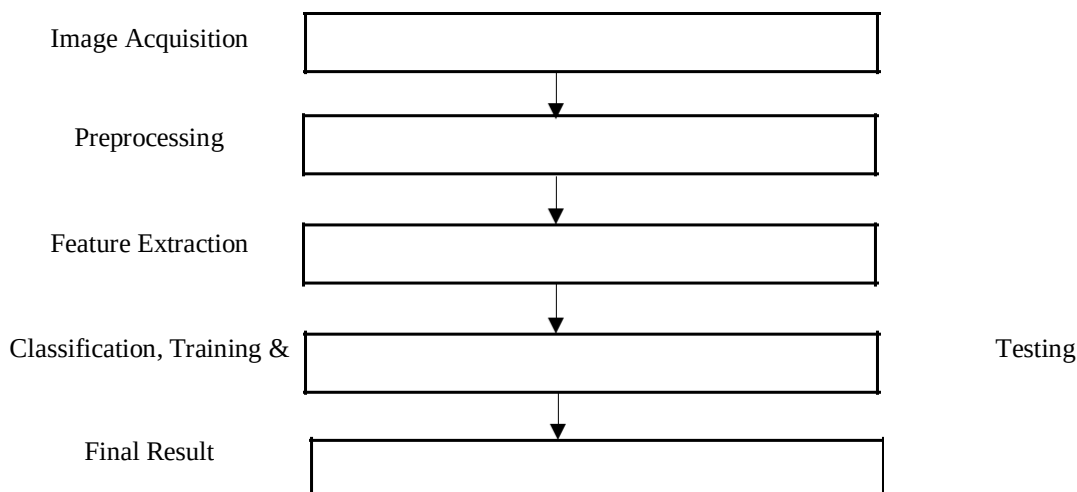


Fig. 1. Flow diagram of proposed scheme.

CHAPTER -2

Pre-extraction preparation of plant samples

The initial stage in studying medicinal plants is the preparation of plant samples to preserve the biomolecules in the plants prior to extraction.

Plants samples such as leaves, barks, roots, fruits and flowers can be extracted from fresh or dried plants material. Other pre-preparation of plant materials such as grinding and drying also influences the preservation of phytochemicals in the final extracts.

Fresh vs. dried samples: Both fresh and dried sample is used in medicinal plants studies. In most cases, dried sample is preferred considering the time needed for experimental design.

Sulaiman et al limit the interval between harvest and experimental work at the maximum period of 3 hours to maintain freshness of samples, as fresh samples are fragile and tend to deteriorate faster than dried samples.

Comparison between fresh and dried *Moringa oliefera* leaves showed no significant effect in total phenolics but with higher flavonoids content in dried sample [5].

Grinded vs. powdered samples: Lowering particle size increases surface contact between samples and extraction solvents.

Grinding resulted in coarse smaller samples; meanwhile powdered samples have a more homogenized and smaller particle, leading to better surface contact with extraction solvents. This particular pre-preparation is important, as for efficient extraction to occur, the solvent must make contact with the target analyses and particle size smaller than 0.5 mm is ideal for efficient extraction [6].

This particular size of particle was mentioned in Suleiman et al, preparing vegetable samples that was ground to 400 μm (0.4 mm) in size.

Conventional mortar and pestle or electric blenders and mills are commonly used to reduce particle size of sample. Investigation of nanoparticles powder of Centella

Feature extraction

Our method takes into account the color and shape features of the leaf. Leaves of different plants are invariably similar in color and shape therefore a single feature alone may not produce expected results.

3.3.1. Color features

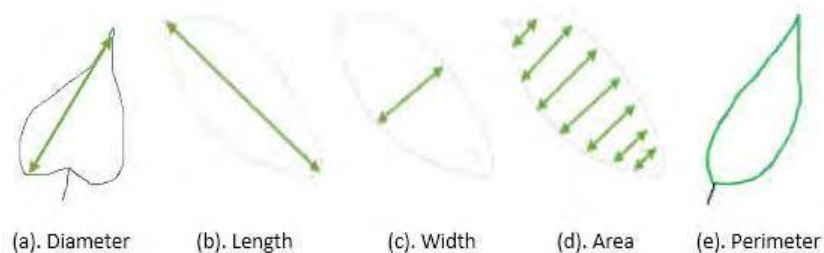
The method of image searching and retrieval proposed by Dr. H.B. Kekre et al. [7] mainly focuses on the generation of the color feature vector by calculating the average means. In the proposed algorithm, first the three color planes namely Red, Green and Blue are separated. Then for each plane row mean and column mean of colors are calculated. The average of all row means and all columns means is calculated for each plane. The features of all 3 planes are combined to form a feature vector. Once the feature vectors are generated for an image, they are stored in a feature database.

3.3.2. Shape features

We defined shape features on the basis of morphological features and tooth features:

- ***Geometric features***

We used the similar (as described in [8]) commonly used 5 geometric features (DMFs), illustrated in fig. 3, derived from following 5 basic features:



basic morphological features.

Fig. 3. The five

- (1). *Diameter*: The diameter of the leaf is the longest distance between any two points on the closed contour of the leaf.
- (2). *Physiological Length*: It is the length of the line connecting the two terminal points of the main vein in the leaf.
- (3). *Physiological Width*: It refers to the distance between the two endpoints of the longest line segment perpendicular to the physiological length.
- (4). *Leaf Area*: It is the number of pixels of binary value 1 on smoothed leaf image.
- (5). *Leaf Perimeter*: It is the number of pixels along the closed contour of the leaf.

- *Morphological Features*

Based on above 5 basic geometric features, we can define following 12 digital morphological features:

- (1). *Smooth Factor* : This is defined as the ratio between area of leaf image smoothed by 5x5 rectangular averaging filter and the one smoothed by 2x2 rectangular averaging filter.
- (2). *Aspect Ratio*: This is defined as the ratio of physiological length to physiological width, i.e., L/W .
- (3). *Form Factor*: It is defined as the difference between a leaf and a circle and is calculated by the formula $4\pi A/P^2$.
- (4). *Rectangularity*: It describes how similar a leaf is to a rectangle and is computed as $L.W/A$
- (5). *Narrow Factor*: It defines the narrowness of the leaf and is calculated as D/L .
- (6). *Perimeter Ratio of Diameter*: It is defined as the ratio of the perimeter of the leaf to the diameter of the leaf, i.e., P/D .
- (7). *Perimeter Ratio of Physiological Length and Physiological Width*: It is defined as the ratio of the perimeter of the leaf to the sum of its physiological length and physiological width, i.e., $P/(L+W)$.
- (8). *5 Vein Features*: Leaf vein forms the basis of leaf characterization and classification as they define the skeletal structure of the leaf. Different species have different leaf vein patterns which can be used in distinguishing the leaves that have similar shape. The standard procedure for computing the vein features is to perform a morphological opening operation on the grayscale image. A flat, disk shaped structuring element of radius 1,2,3,4 is used and the resultant image is then subtracted from the contour of the leaf. The output resembles to the vein structure of the leaf on the basis of which following 5 vein features are calculated: A_1/A , A_2/A , A_3/A , A_4/A , A_4/A_1 where A_r is the remaining leaf obtained using a structuring element of radius r and A is the area of the leaf.

- *Tooth features*

A tooth [1] in a leaf is a pixel that is serrated and toothed around the margins of the leaf. Fig. 4. (a) represents a tooth point of a leaf. To determine whether a point P_i on the margin of the leaf is a tooth point or not, we examine the angle θ subtended at P_i by its neighbors P_{i-k} and P_{i+k} (where k is the threshold). If the angle θ is within a particular range, then P_i is a tooth; otherwise, it is not. Fig. 4. (b) represents a standard toothed leaf. In proposed algorithm total number of tooth points are calculated in each image and stored against feature vector.

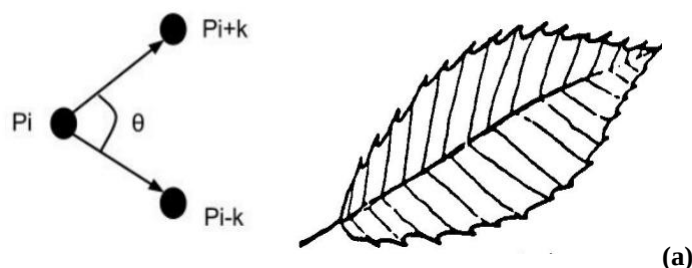


Fig. 4. (a). A tooth point (b). Toothed leaf

3.4. *Classification, training & testing*

General statistical classification is the process of identifying a set of categories, or classes, to which a new observation belongs, on the basis of prior knowledge such as a training dataset. More specifically, classification in this work will be the process used to assign a certain plant species to an image, based on its feature set. It is also a subset of the more general classification problem in statistics and machine learning, namely supervised learning. We formalize the classification elements as follows:

Where C represents the class set, f a feature vector in the corresponding m -dimensional feature space F , a_i is a feature attribute and T the training set of feature vectors and their respective class labels. Hence, we are looking for a function $Class(f) : F \rightarrow C$ that assigns a class label from C to a given feature vector f based on the data from T . Throughout this work we will be looking for a classification method not only capable of attributing a class label to a feature vector f_s but also a confidence vector describing the probability that a given sample belongs to a certain class:

3.4.1. Classifier selection

We followed two approaches to classify our dataset, i.e., Neural Networks and Euclidean Distance Method.

A neural network, illustrated in fig. 5, consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand. This is the learning phase.

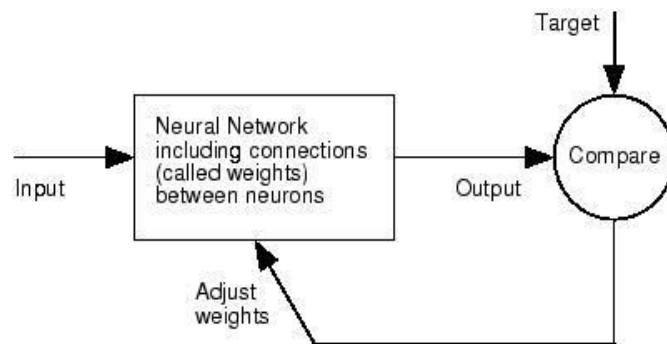


Fig. 5. A typical Neural Network

The Euclidean or KNN classifier based on the distance is direct and simple. Special interest was given to KNN due to its Simplicity and Efficiency. It is one of the simplest classifiers with characteristics fitting our requirements. It requires no training computations and is easily handled by weak processors. Its testing time, however, grows linearly with the size of the training set, limiting the scalability of the classifier. It is based on distance measures in feature space but instead of comparing f_s to a class representative value, it compares it to all samples of the training set f_i selecting the first k closest ones. We call the subset of k -closest training samples K .

3.4.2. Training & Testing

Flavia dataset contains a total of 1907 images of 33 different plant species. These images were used to train the classifier. For each type of plant in flavia dataset, we selected 5 species of leaves from testing sets which are then used to test the efficiency of the proposed algorithm in terms of accuracy and execution time.

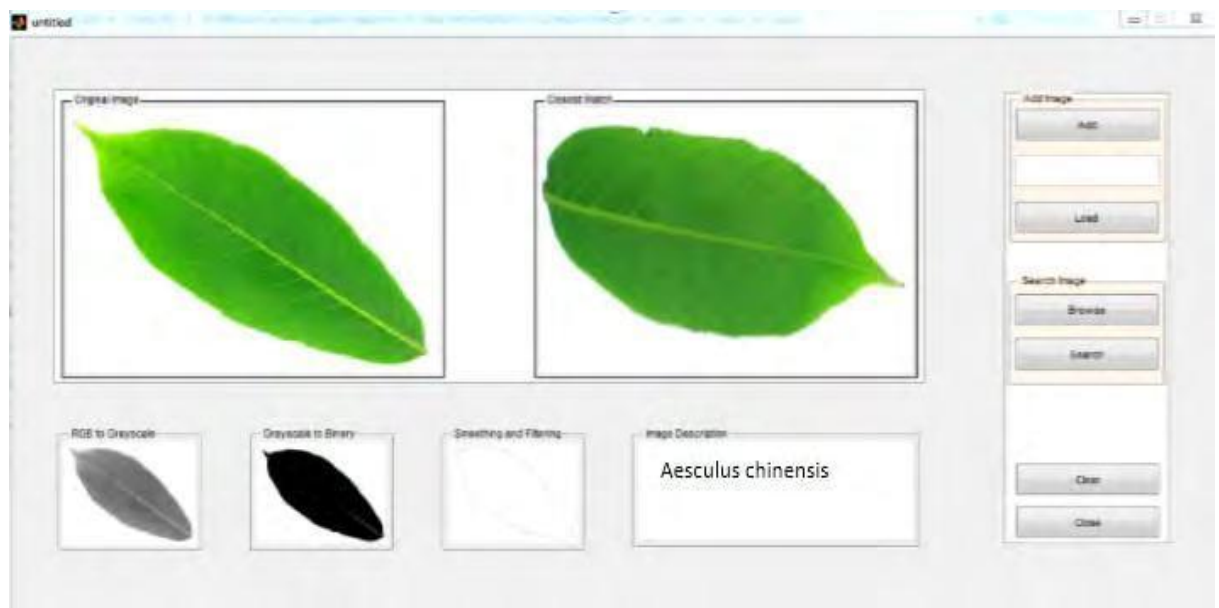
4. Experimental results

The classification was performed using Neural Networks and Euclidean classifier. The results obtained with these schemes were used to compare the classification technique and to conclude this study. Table I shows the accuracy of the system for KNN & ANN classifier.

Table I. Accuracy for the KNN & ANN classifier.

Classifier	Accuracy (%)
KNN	85.9
ANN	93.3

The time taken and the accuracy attained with the classification scheme are the important viewpoints of any user. It is found that ANN classification scheme is better than KNN for a dataset with large number of images whereas, KNN outperforms the ANN approach for a smaller dataset. The efficiency is calculated not only in terms of accuracy but also in terms of time taken by the classifier used. KNN classifier is faster for smaller dataset whereas, ANN is a good choice for a scaled dataset. Fig. 6. Shows the graphical user interface (GUI) of our plant identification system.



ig. 6. Snapshot of a sample run of program.

OBJECTIVE

An automatic leaf recognition system for plant identification using machine vision technology

CHAPTER -3

Existing System

Air-drying, microwave-drying, oven-drying and freeze-drying (lyophilisation) of plants samples: Air-drying usually takes from 3-7 days to months and up to a year depending on the types of samples dried (eg. leaves or seed). Plant samples, usually plants leaves with stem were tied together and hang to expose the plant to air at ambient temperature. This drying method does not force dried plant materials using high temperature; hence, heat-labile compounds is preserved. However, air-drying take longer time in comparison to microwave-drying and freeze drying and may be subjected to contamination at unstable temperature condition.

Microwave-drying uses electromagnetic radiation that possesses both electric and magnetic fields. The electric field causes simultaneous heating through dipolar rotation; alignment on the electric field of the molecules possessing a permanent or induced dipole moment (e.g. solvents or samples), and ionic induction, that produce oscillation of the molecules [8]. Oscillation causes collisions between molecules and resulted in fast heating of the samples simultaneously. This method can shorten drying time but sometimes causes degradation of phytochemicals.

Oven-drying is another pre-extraction method that uses thermal energy to remove moisture from the samples. This sample preparation is considered as one of the easiest and rapid thermal processing that can preserved phytochemicals. Oven-drying at 44.5°C for 4 hours using 80% methanol resulted in highest antioxidants activities in *Cosmos caudatus* extracts and similar result were observed in optimized 80% methanol extracts at 44.12°C for 4.05 hours [9]. Shorter period of extraction time was obtained using this method. However, effect of drying on *Orthosiphon stamineus* showed no significant effect on the antioxidant activity but the bioactive phytochemicals; such as sinensetin and rosmarinic acid content were affected by the oven-and sunlight-drying, suggesting the sensitivity of the compounds to temperature [10].

Freeze-drying is a method base on the principle of sublimation. Sublimation is a process when a solid is changed into gas phase without entering the liquid phase. Sample is frozen at -80°C to -20°C prior to lyophilisation to solidify any liquid (eg. solvent, moisture) in the samples. After an overnight (12 h) freezing, sample is immediately lyophilized to avoid the frozen liquid in the sample from melting. Mouth of the test tube or any container holding the sample is wrapped with needle-poked-parafilm to avoid loss of sample during the process. Most of the time, sample was lost by splattering out into the freeze-flask (Figure 1a and 1b). Freeze-drying yielded to higher level of phenolic contents compared to air-drying as most of the phytochemicals are preserved using this method. However, freeze-drying is a complex and expensive methods of drying

compared to regular air drying and microwave-drying. Thus, the use is restricted to delicate, heat-sensitive materials of high value.

Maceration, infusion, percolation and decoction: Maceration is a technique use in wine making and has been adopted and widely used in medicinal plants research. Maceration involved soaking plant materials (coarse or powdered) in a stoppered container with a solvent and allowed to stand at room temperature for a period of minimum 3 days with frequent agitation [11]. The processed intended to soften and break the plant's cell wall to release the soluble phytochemicals. After 3 days, the mixture is pressed or strained by filtration. In this conventional method, heat is transferred through convection and conduction and the choice of solvents will determine the type of compound extracted from the samples. Infusion and decoction uses the same principle as maceration; both are soaked in cold or boiled water. However, the maceration period for infusion is shorter and the sample is boiled in specified volume of water (eg. 1:4 or 1:16) for a defined time for decoction [11]. Decoction is only suitable for extracting heat-stable compounds, hard plants materials (e.g. roots and barks) and usually resulted in more oil-soluble compounds compared to maceration and infusion. Unique equipment called percolator (Figure 1c and 1d) is used in percolation, another method that shares similar fundamental principle. Dried powdered samples are packed in the percolator, added with boiling water and macerated for 2 hours. The percolation process is usually done at moderate rate (e.g. 6 drops /min) until the extraction is completed before evaporation to get a concentrated extracts [12].

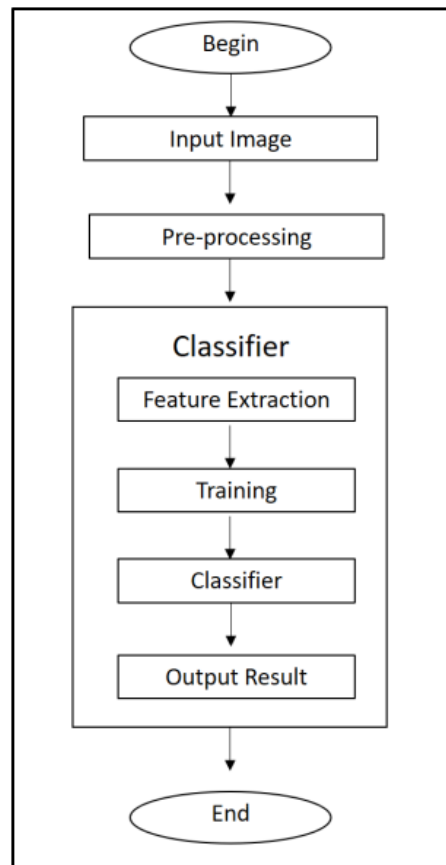
Strength and limitation: This technique is the easiest and simple method. However, organic waste come into an issue as large volume of solvents is used and proper management of the waste is needed. Alteration in temperature and choice of solvents enhance the extraction process, reduce the volume needed for extraction and can be introduced in the maceration technique, when such alteration is not objectionable. Boiling *Centella asiatica* at 90°C showed to increase phenolics content and antioxidant activities, but jeopardized the pH of the extracts with increase extraction time [13]. In this method, solvents used in the soaking process play a critical role.

Studies: Extraction of *Psidium guajava* L. leaves using ethanolic and hydroalcohol extracts (4:1 v/v) resulted in highest extraction yield with maximum presence of phytoconstituents (alkaloids, saponins, carbohydrates, tannins and flavonoids) compared to the other solvents such as petroleum ether, chloroform and water [14]. Non-polar solvents such as petroleum ether and chloroform showed no active compounds preserved and very little tannins presence respectively in the extracts. Water showed to have similar efficiency as ethanol except no trace of alkaloids was presence in water extracts [14]. Polar solvents are more effective in the extraction of bioactive molecules from *Psidium guajava*. Methanol extracts of *Garcinia atriviridis* (1:10 w/v) exhibited higher antioxidant activities compared to the aqueous (1:10 w/v) extracts, but the aqueous showed higher anti-hyperlipidemic activity [15]. Effect of different solvents using maceration at 1:10 w/v sample to solvent ratio for 1 hour showed 70% acetone as efficient solvent for *Portulaca oleracea* based on total phenolics, and 70% methanol as efficient solvent for flavonoids in *Cosmos caudatus* [16]. In the case of *Moringa oliefera*, maceration with 70% ethanol powdered dried samples at 1:40 w/v exhibited highest phenolics and flavonoids content compared to Soxhlet extraction and percolation using similar solvent [5].

CHAPTER -4

Image Processing in Leaf Pattern Recognition

Leaf pattern recognition usually follows the steps as shown in Figure 1. The most challenging part of this study is to extract distinctive features of leaves for plant species recognition. In this case, different classifiers using high performance statistical approaches have been used to perform leaf features extraction and classification. The advancement in computer vision and artificial intelligence features extraction and classification. The advancement in computer vision and artificial intelligence.



In Kurtz algorithm, a hierarchical approach was used to extract segments of interest from the lowest to the highest resolution data, which was the first cluster image material sharing the common colour properties into a group of coarse image patches [20]. The individual patches are depicted as hierarchical construct with a binary partition tree (BPT). This cluster of images can be assumed to be like a forest of BPTs. When the construction of this forest is completed, each tree will be gradually segmented. Eventually, the progressive segmentation will produce a global segmentation of the image. This method could produce precision up to 85.1%. The

data of Kurtz algorithm only utilized colour distance map without input stroke as opposed to the two previously reported methods; SLIC and GAC [21].

A comparative result had been established using power watershed for preprocessing of the image. Power watershed used the concept of magnitude reduction to get more accurate data from the adjacent region, which is identical to the Graphcut method, hence improving segmentation of the image. Power watershed technique included multi labelling, contrast, and ratio invariant stage [22]. The Power watershed method compared to the Graphcut technique yielded 63.5% precision.

Leaf Feature Extraction

Leaf features such as shape, size, and colour are important characteristics in a computational recognition system. The feature extraction can be performed by contour-based or region-based extraction. The contour-based extraction is expressed in length, width, aspect ratio, and leaf diameter as descriptors. The length descriptor is measured based on the main vein of leaf, as it stretches from the main vein to the end tip. The width descriptor is the span of leaf viewed from one side to the other, from the leftmost point to the rightmost point of leaf. Aspect ratio can be determined by dividing the length of leaf to its respective width. The longest distance between two points inside the covered area of leaf is the leaf diameter, while the maximum distance of the area covered by leaf is the perimeter. Alternatively, leaf perimeter can also be obtained by computing the number of pixels that contains leaf margin.

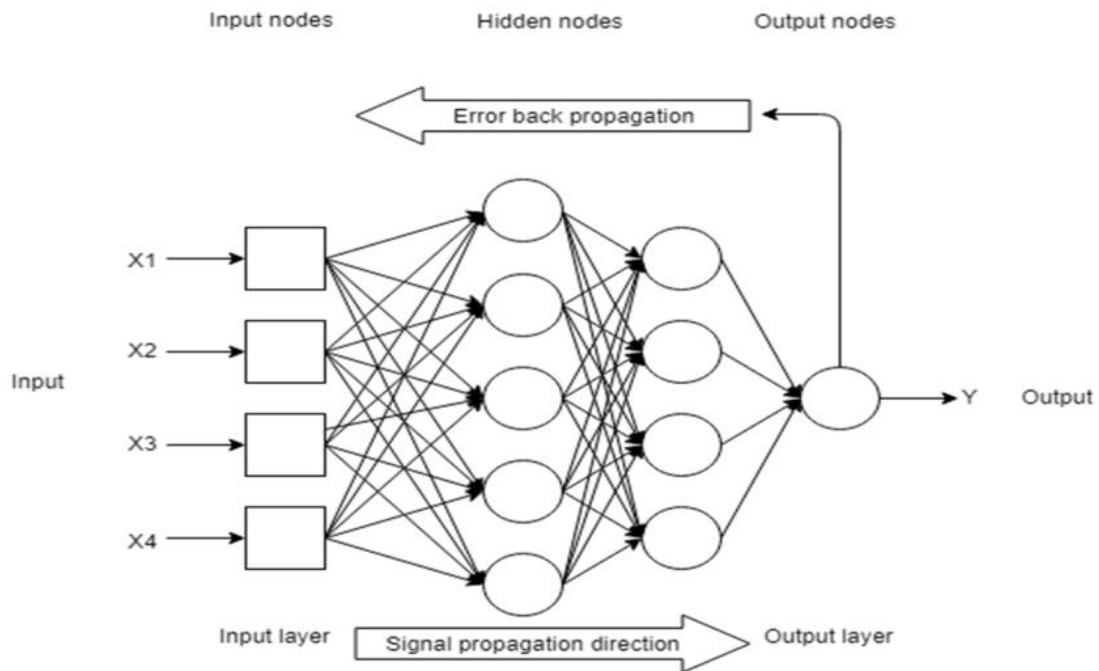
On the other hand, leaf features extracted from the region-based technique uses the descriptors such as shape, rectangularity, compactness, area, and eccentricity. Shape (Convex Hull) is the coordinates of points where the whole area of leaf can be conveniently determined using the convex hull algorithm. Rectangularity (R) is a rectangle which just fits the image drafted outside the image in Equation (1).

$$R = LpWp/A$$

where Lp , Wp , and A represent the length, width, and area of the leaf, respectively.

Mathematical Classifiers

Artificial Neural Network (ANN)

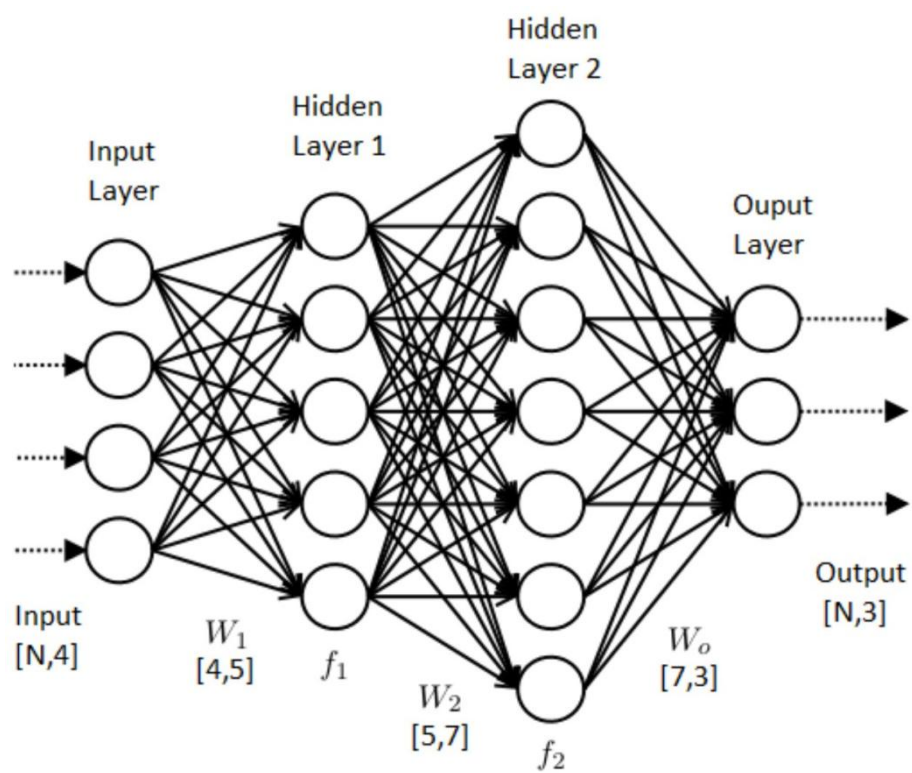


Artificial neural networks are a main component of machine learning and they are designed to spot patterns in data. This makes ANNs an optimal solution for classifying (sorting data into predetermined categories), clustering (finding like characteristics among data and pulling that data together into categories) and making predictions from data (such as helping determine infection rates for COVID, the next catastrophic weather event or box-office smash).

In everyday life, ANNs are powering the “watch next” feature of YouTube videos, creating realistic CGI faces, helping detect fraud, giving us the ability to chat with chatbots and more. In fact, there are probably not many tasks an artificial neural network can’t do as long as it’s trained to do it.

While ANNs can tackle most tasks if they are allowed to train for it, the biggest obstacle to overcome is the amount of time it takes to train ANNs and the computing power required for a complex task.

In addition, it’s impossible for humans to fully understand what happens in the hidden layers of an artificial neural network. Although researchers are actively working on this, there is still a lot to learn even though we’ve come so far in helping machines think and act like a human.



Support Vector Machine (SVM)

The idea of SVM is to define decision boundaries of feature vectors on decision plane which separates features unanimously.

Since the distinction in feature between the images are evident, the images will be classified into their respective class with little to no complication. New sample is appointed into one of the classes that match their characteristics in this training algorithm.

The samples are then charted onto the same region of class where they shall be. A fine separation between classes is naturally attained by the hyperplane that marks the furthest distance to the nearest data point of each group. This is also due to the fact that the larger margin results in small generalization error of the classifier.

Applying invariant feature to the SVM will vastly increase the accuracy of the classifier. The application of the invariant features goes with the contour-based descriptor.

This classifier is exceptionally accurate because it is capable to recognize images without the problem of image rotation, translation, scale, and inversion or mirroring. The only drawback is that the system is not capable to differentiate leaves that have almost the same shape. In response to this situation, the principal component analysis (PCA) could be applied to enhance the performance of this system.

PCA operates by converting data set into a new set of data and the variables are arranged according to their priorities. This causes the components to be uncorrelated and sorted in the order that the components with the largest variation occurs first, followed by the components with the smallest variation which is then terminated.

Finally, one of the more superior classifiers would be SVM which integrates the Hu moment invariant algorithm to its model

which could achieve up to 94.1% accuracy [45]. Zhang et al. classified plant species using leaf shape and texture. They used SVM classifier for a dataset of 1900 leaves belonging to 32 different species.

A new classification method based on the generation of space features by combining local texture features using wavelet decomposition, co-occurrence matrix statistics, and global shape features.

The method was able to extract features of the plant leaf images and yielded an accuracy of over 93.8%.

A method of extracting 15 features of leaf images via canny edge detector, and SVM as the classifier was reported by Salman et al.

The extracted features were convex area, filled area, perimeter eccentricity (E), solidity, perimeter ratio of diameter, orientation, narrow factor, extent, Euler number, diameter, circularity, rectangularity, perimeter ratio of length and width, complexity, and compactness.

The study used 2220 images from 22 plant species in the Flavia dataset and achieved an accuracy of 85 to 87% only.

Khmag et al. aimed to create a recognition system for leaf images based on leaf contour and centroid. An image processing algorithm was developed by utilizing the variant to scaling shift, spin technique, scaling approach, and filtering processes, as well as SVM as the classifier for the leaf contour.

The proposed method used 70 samples taken from the Flavia dataset and their geometrical and shape features were extracted and yielded an accuracy of 97.7% which is the highest achievement reported in literature.

Different classifiers had been used by Srivastava et al. who did a comparative analysis for leaf classification and recognition.

The study extracted 14 leaf features using shape detector and SVM as the classifiers of the system. Sixteen different plant species from the Flavia database comprising of 480 images were used as the training dataset and 14 features extracted for optimal performance.

Quadratic SVM yielded the highest accuracy of 90.9% as compared to medium Gaussian SVM and cubic SVM which gave 89.4% and 89.8% test accuracy, respectively.

A multiple classifier system was also used by Araujo et al based on texture and shape features of leaf images. SVM and neural network classifiers were used to train four different features, namely local binary pattern (LBP), histogram of gradients (HOG), speed of robust features (SURF), and Zernike moments (ZM).

Then, a static classifier selection method was used to search for the ensembles that maximize the average classification score. The ImageCLEF 2011 and Image CLEF 2012 datasets were used for the experiments.

The results found that the proposed method was able to improve for 11.7% and 4.2% in the scan category and 4.06% and 5.87% for the scan-like category in the average classification score relative to the best results reported in the literature for ImageCLEF 2011 and ImageCLEF 2012 datasets, respectively.

Hence, the multiple classifier system overcame the performance of monolithic methodologies.

A combined classifier consisted of Naïve Bayes and a decision tree along with SVM was discovered to yield maximum accuracy, 93.1% in

leaf image classification. The classification was based on visual attributes such as shape, colour, texture, and vein pattern.

The accuracy was slightly lower when using the individual classifier, namely 91.8% for SVM, 88.3% for Naïve Bayes, and 85.7% for decision tree. The high performance of Naïve Bayes classifier can also be seen from the findings of Eid et al. .

The computational model was proposed using digital plant images based on biometric features such as shape and vein patterns.

The model could achieve an average accuracy, 97% of leaf identification based on 10 different biometric information extracted from 1907 leaf images of 32 plant species taken from the Flavia dataset.

The system used three plant species with 300 leaf images for the experiments with 86.7% accuracy in identification. The images were captured and segmented using a marker-controlled watershed segmentation, which was generated automatically using morphological operations.

The shapes were then converted into features via Hu moments. The accuracy of the system can be further improved by the addition of more features extracted via the system operation, as well as increasing the dataset used for the experiments.

The significant advantage of SVM is its robustness as a feature vector classifier mechanically, which is based on digital morphological paper. It is a potent classifier with its simplicity in terms of feature extraction.

However, this system needs to be trained for long duration because a new decision boundary is drawn each time when new data is tested.

To the best of our knowledge, this classifier could achieve the highest accuracy (85%–98%) in leaf identification. This could be due to the high number of leaf images which usually ranges from hundreds to thousands

used in model development. Probably, this also explains that many studies would like to use the SVM classifier in recent years.

The classifier of SVM also proved to be effective to detect plants in their natural habitat, which was purposely designed in an environment with heavy interferences and overlapping issues .

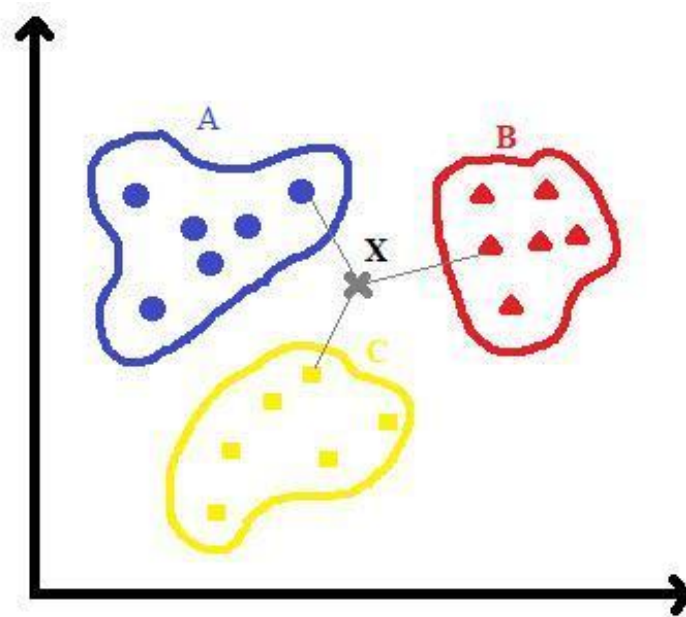


Table 1. Comparison between advantages and disadvantages of classifiers.

Classifiers.	Advantages	Disadvantages
Artificial Neural Network (ANN)	1 . Capable in distinguishing complex nonlinear relationship between independent and dependent variables. 2 Simplistic statistical . training.	1 . Great tendency of data overfitting. 2 . Bigger computational load.
Convolutional Neural Network (CNN)	1 . Multiple features can be extracted simultaneously. 2 . Robust to noise.	1 . High computation level. 2 No capable for . generalization.
Probabilistic Neural Network (PNN)	1 . High distortion resistance. 2. Flexible to changing data. 3. Specimen can be classified into multiple output.	1 . Long training duration. 2 Complicated network . layout. 3 Have tendency for . overfitting with too many traits.
Support Vector Machine (SVM)	1 Great generalization . potential. 2 . Exceptionally robust.	1 Speed and size constraint . for both training and testing. 2 Complex algorithm . structure. 3 . Slow training.
K-Nearest Neighbor (KNN)	1 . No training needed. 2. Robust in term of research space. 3 . Simplest classifier.	1 . Susceptible to noise. 2 . Lazy learning. 3 Costly testing for each . instance.

ANN is a unique classifier to identify the complex nonlinear relationship between independent and dependent variables. Since leaf pattern recognition is considered as a complex nonlinear problem, ANN would be able to interpret the variables effectively. ANN is also considered to be simple in terms of its statistical training. The iteration of the training mechanism is conducted one at a time, and the result is feed-forwarding into the system as a new information. Consequently, this system may cause data overfitting due to the fact that it iterates a large sum of data. This will end with huge computational load to sustain the feed-forward back-propagated data.

CNN may involve multiple features extraction, and at the same time, providing detail and quick detection. Therefore, this mechanism makes CNN be more resilient towards unnecessary noise. This classifier also utilizes a high computation level to extract multiple features. This system is not appropriate in generalizing features. Similar to CNN, PNN has a high noise resistance too. With its capability to foresee potential information, it is flexible to do data changing. Since the weight of the vector is predetermined, the specimen could be classified into multiple outputs depending on the closeness of the sample to the vector.

The drawback for this classifier is the needs for long duration of training when new weight vector is applied. This system has an intricate layout structure. In case excessive traits are involved, it is prone to overfitting also.

SVM is different from other neural networks, it has great generalization potential and is exceptionally robust.

The disadvantages of this system are its speed and size which are restricted for both training and testing.

This classifier has a complicated algorithm structure to determine the feature vector of each sample individually. Since its generalization potential is great, it takes time for the training process before it could classify a particular set of samples.

The research team of Shah et al. utilized a dual-path deep CNN to learn leaf image features and optimize for classification in order to classify the leaf based on marginalized shape context and shape-texture dual-path deep CNN.

The study revealed that the dual-path CNN method outperformed other CNN methods such as uni-patch CNN, texture-patch CNN, marginalized shape context with SVM classifier, multiscale distance matrix with SVM classifier, curvature histogram. The method gave near-perfect-top-1-match result using Flavia dataset and top-three-match results for other datasets.

A more distinct approach brought by KNN classifier is to emphasize the differences between features and compare the close feature vector, and thus make training absolutely unnecessary.

This is also robust in terms of research space.

The characteristics make KNN the simplest classifier. KNN classifier is very susceptible to noise which may alter the results during the classification process. Lack of training makes this classifier likely to be a lazy learning system.

The prediction step in KNN is rather expensive, since every prediction is made, the classifier determines the nearest neighbor by comparing to the entire dataset.

5. Pattern Recognition Method

Various pattern recognition methods have been proposed and new techniques will be discovered continuously to meet the needs of technological advancement. Depending on the situation, different issues require different image classification methods.

.

In the study of Pankaja et al who stated that grey level co-occurrence matrix, discrete wavelet transform, and hierarchical centroid could assist in obtaining special information of leaves accurately and improve the accuracy of plant leaf recognition. Three hundred leaf samples of 30 different species sourced from the Flavia dataset were used for the experiment and the model achieved an accuracy of 96.7%. This result was obtained after pre-processing, feature extraction, and classification of leaf samples based on shape and texture, via grey level co-occurrence matrix and hierarchical centroid based technique.

Chaki et al.purposed a methodological recognition of 32 pre-defined classes of plants using a neuro-fuzzy classifier and compared it to other classifiers such as KNN and neural network classifiers. In the study, 640 leaf sample images with various sizes, shapes, and orientations were tested and found to show improvement compared to k-NNC and NNC with an accuracy of 97.5% and 79.7%, respectively. A limitation of the proposed method is that it works reliably only when deformations do not alter the major and minor axis lengths.

In the work of neural network-based leaf recognition, the implementation of pulse coupled neural network in the feature extraction of leaf images was used to obtain the entropy sequence, aspect ratio, Zernike moments, Hu's invariants, form factor, rectangularity, circularity, and area. The accuracy of the system was better than the other methods, probably because the entropy was taken as the key feature in the leaf classification, yielding an accuracy of 92.0%.

The study of Chaki et al. developed a plant leaf recognition using a layered approach. The layered architecture with each layer handles a specific type of visual characteristic using a set of features to form a customized data model.

The different layers are then fed into an array of custom classifiers for robust recognition.

Colour-based modelling is used for non-green leaves, while shape-based modelling is used for green, simple, and compound leaves, using a layered approach.

The database of 600 leaf images from 30 different classes of leaves were used to test the layered model and yielded an accuracy of over 90% for all the different approaches. The utilization of different visuals, segmentation to reduce computational load, different customization in the classifiers, and addition of different layers make the model more robust than all the other models proposed in previous studies.

The advancement of vision technology paralleled with deep learning has brought upon a remarkable benchmark in artificial intelligence, primarily in the field of visual classification. In spite of all the advantages that visual classification has granted to various fields, improvement can be seen in order to have a robust system.

For example, a lot of classification systems are still having problems with overfitting which greatly affects how the system views the subject. The machine learning system should provide short processing time in this computational age.

Table A1. Previous pattern recognition methods.

No.	Goals	Method	Results	References
Artificial Neural Network (ANN)				
1	Identification Of Selected Medicinal Plant Leaves Using Image Features And ANN	Method utilizes feature extraction based on shape, color, and texture of leaf images and training via ANN classifier for leaf classification of the system. The image processing and neural network toolboxes were obtained from MATLAB.	Implementation of 8 input features into the system were found to be optimal in terms of the complexity as it demands minimum input as well as less computational time required. The accuracy of the system was 94.4% against 63 leaf images in the dataset.	[30]
2	Neural Network Based Leaf Recognition	Implementation of Pulse Coupled Neural Network into the feature extraction of leaf images to obtain the entropy sequence, aspect ratio, Zernike moments, Hu's invariants, form factor, rectangularity, circularity, and area. The classifier used is Artificial Neural Network.	The accuracy of the system proposed was better compared to the other methods in the study, where the entropy was taken as the key feature in the leaf classification, yielding an accuracy of 91.98%.	[63]
3	Combined Thresholding and Neural Network Approach for Vein Pattern Extraction From Leaf Images	A proposed model that extracts leaf vein patterns via combining thresholding methods and an ANN classifier.	The experimental results of the method proposed have shown to possess the capability of extracting a more accurate venation pattern of the leaf samples for pattern recognition with 97.3% accuracy while reducing computing time as well compared to direct neural network approach which achieved an accuracy of 84.4%.	[31]
	Classification of Plant Leaves	using Morphological Features and Zernike Moments	The proposed model utilizes leaf morphological features and Zernike moments which are independent of leaf growth and image translation,	

rotation and scaling and then classified with several different classifiers to achieve optimal result.

The dataset tested were Flavia dataset of 32 species

(1907 images, 50-60 leaves each species) and a medicinal plants dataset of 6 species (30 images each). SVM and

[65]

PNN have comparably higher accuracy than k-NN and naïve Bayes classifiers due to the latter being lazy learners algorithms.

Convolutional Neural Network (CNN)

Leaf Classification based on Marginalized Shape Context and Shape-Texture Dual-Path Deep Convolutional Neural Network

Utilizing a dual-path deep CNN to learn leaf image features and optimize them for classification compared to vanilla convolutional network systems.

Dual-path CNN method outperforms various other CNN methods, being: uni-patch CNN, texture-patch CNN, marginalized shape context with SVM classifier, Multiscale Distance Matrix with SVM classifier, [60] curvature histogram, giving near-perfect-top-1-match results on Flavia dataset and top-3-match results on all other datasets.

Table A1. Cont.

No.	Goals	Method	Results	References
Probabilistic Neural Network (PNN)				
6	Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram	Utilizing the combined methods of Fuzzy Local Binary Pattern (FLBP) and the Fuzzy Color Histogram (FCH) in order to identify the medicinal plants found in Indonesia. The combination was done via Product Decision Rule (PDR) method, and is capable of extracting leaf image texture and color. The classifier utilized was PNN classifier.	Fusion of methods FLBP and FCH has shown to be capable of increasing the overall accuracy of plant identification methods and yielded an accuracy of 74.51%. The accuracy of the system without the fusion features were 59.61% for FLBP features and 50.78% for FCH feature respectively.	[42]
7	Plant Leaf Classification Using Centroid Distance and Axis of Least Inertia Method	The utilization of centroid distance and axis of least inertia for plant classification. Canary operator was implemented for the binary images converted from RGB (Red, Green, Blue) to detect and thin out the edges of the leaf images before the shape is traced. The centroid distances of these points as well as the distance of sampling points from the axis of least inertial lines were computed. The classifier used was the	Combination of both these methods, which are invariant to translation, rotation and scale into the extraction of feature vectors for leaf image classification has been proven to be able to improve the accuracy. Experimentation results have shown that the accuracy of the system from the Flavia dataset was 82.05% and Swedish leaf dataset was 80.10%, an improvement from previous methods proposed	[44]

PNN
classifier for two public leaf databases,
Flavia
and the Swedish leaf dataset.

Support Vector Machine (SVM)		
8	Combined Classifier for Plant Classification and Identification from Leaf Image based on Visual Attribute	The identification of the plant images based on varying features such as shape, color, texture, and vein pattern with combined classifier using majority voting technique has been proposed for recognizing leaf's category along with SVM, Naïve Bay's and Decision tree as the combined classifiers Utilization of the combined classifiers with several features extraction of the leaf images have been shown to yield maximum accuracy output as compared to them individually and all existing methodologies with accuracy of 93.11% while with separately, SVM is 91.8576, Naïve Bayes is 88.2736%, and Decision tree is 85.6678% respectively. [52]
9	Comparative Analysis of Leaf Classification and Recognition by different SVM Classifiers	16 different plants species from the Flavia database comprising of 480 images were used as the training dataset with 14 features extracted for optimal performance. Quadratic SVM yielded the highest accuracy of 90.9% as compared to Medium Gaussian SVM and Cubic SVM gave 89.4% and 89.8% test accuracy separately. [48]

Table A1. Cont.

No.	Goals	Method	Results	References
10	Multiple Classifier System for Plant Leaf Recognition	The methodology presents a multiple classifier system (MCS) of plant identification based on texture and shape features of the leaf images. SVM and Neural Network classifiers are trained on four different feature sets, namely, Local Binary Pattern (LBP), Histogram of Gradients (HOG), Speed of Robust Features (SURF) and Zernike Moments (ZM). Then, a static classifiers selection method is used to search for the ensembles that maximize the average classification score.	ImageCLEF 2011 and ImageCLEF 2012 datasets were used for the experiments with the proposed MCS methods showing an improvement of 11.76% and 4.23% in the scan category and 4.06% and 5.87% for the scan-like category in the average classification score relative to the best results reported in the literature for ImageCLEF 2011 and ImageCLEF 2012 datasets respectively. MCS approach also overcomes the performances of monolithic methodologies compared in the study.	[49]
11	Identifying leaf in a natural image using morphological characters	Proposed methodology of plant detection in their natural habitat, designed in an environment with heavy interference and overlapping. The images are captured and segmented using marker-controlled watershed segmentation which is generated automatically	3 species with 300 different leaf image samples were used for the experimentation of the system and captured in real time with 86.7% identification accuracy. The accuracy of the system can be further improved by the addition of more features extracted via the system's	[54]

using morphological operations and the shapes obtained are converted into features via Hu moments. The classifier utilized is SVM classifier.

operation as well as increasing the dataset used for the experiment.

k-Nearest Neighbor (KNN)

12	Recognition of Whole and Deformed Plant Leaves using Statistical Shape Features and Neuro-Fuzzy Classifier	A method-logical recognition of 32 pre-defined classes of plants using a Neuro-fuzzy classifier and comparing it to other classifiers such as k-Nearest Neighbor and Neural Network classifiers.	640 leaf sample images of varying sizes, shapes, and orientation were tested in the experiment and have been proven to show improvements compared to k-NNC and NNC with an accuracy of 97.5% and 79.7%. A limitation of the proposed method is that it works reliably only when deformations do not alter the major and minor axis lengths.	[62]
13	Leaf Classification based on Shape and Edge feature with k-NN Classifier	A proposed approach via the utilization of edge based and shape based features extraction for the classification of leaf images.	The method was tested on Flavia dataset of 32 plant species and yielded an average classification accuracy rate of 94.4%, an improvement compared to existing methods, with size and rotation invariant being the novelty of the algorithm.	[59]

Table A1. Cont.

No.	Goals	Method	Results	References
14	Leaf Recognition and Classification using GLCM and Hierarchical Centroid Based Technique	Utilization of various techniques for preprocessing, feature extraction, and classification of leaf based on shape and texture, via Gray Level Co-occurrence Matrix and hierarchical centroid based technique.	No Classifier	[61]
			Gray Level Co-occurrence Matrix, Discrete Wavelet Transform, and hierarchical centroid have been proven to be able to obtain the special information of leaves accurately and improve on the accuracy. 300 leaf samples of 30 different species from Flavia dataset were used for the experiment and the model achieved an accuracy of 96.66%.	
15	Plant Leaf Recognition Using a Layered Approach	A proposed methodology using layered architecture with each layer handling a specific type of visual characteristics using a set of features to form a customized data model. The different layers are then fed into an array of custom classifiers for robust recognition. Color based modeling is used for non-green leaves while shape based modeling is used for green simple and compound leaves, using a layered	The database of 600 leaf images from 30 different classes of leaves were used to test the layered model and yielded an accuracy of over 90% for all the different approaches. The utilization of different visuals, segmentation to reduce computational load, different customization in the classifiers, and addition of different layers make the model more robust than all the other models proposed in previous studies.	[64]

approach.

16	Leaf Plant Identification System based on Hidden naïve Bays Classifier	A proposed computational model in plant identification system of digital images of plants utilizing biometric features such as shape and vein patterns via hidden naïve Bays classifier.	1907 leaf image samples from 32 different species of plants taken from Flavia dataset were used in the study in determining the accuracy of the system and the model has proven to have an average identification accuracy of 97% based on 10 different biometric information extracted.	[53]
The				
17	Leaf species classification based on a botanical shape sub-classifier strategy	implementation of fusion strategy and its corresponding random-forest-based sub-classifiers as part of a leaf recognition system.	The fusion technique provides a significant accuracy enhancement compared to other proposed methods, while providing necessary information for educational purposes.	[43]

. Conclusion

In this paper a new robust & computationally efficient system is presented that takes into consideration the color features and tooth features of the leaf in addition to the shape features. We finally used a combination of color, shape, morphological and tooth features. The system was tested on Flavia dataset by using two classifier and the results were admissible as can be seen in experimental results.

The proposed work can be further extended to identify complex images with petiole and clustered leafs and real time images of leaf.

References

Arora A., Gupta A., Bagmar N., Mishra S., Bhattacharya A.: A Plant Identification System using Shape and Morphological Features

on Segmented Leaflets: Team IITK, CLEF 2012 In: CLEF (Online Working Notes/ Labs/Workshop). (2012)

Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares, "Leafsnap: A Computer Vision System for Automatic Plant Species Identification", Computer Vision – ECCV 2012, 2012, pp 502-516.

Pavan Kumar Mishra, Sanjay Kumar Maurya, Ravindra Kumar Singh, Arun Kumar Misra, "A semi-automatic plant identification based on digital leaf and flower Images", International Conference On Advances In Engineering, Science And Management (ICAESM), 2012, pp. 68-73.

ArunPriya C., Balasaravanan T., Antony Selvadoss Thanamani, "An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 428-432.

N.Valliammal, Dr. S.N.Geethalakshmi, "Automatic Recognition System Using Preferential Image Segmentation for Leaf and Flower Images", Computer Science & Engineering: An International Journal (CSEIJ), 2011, pp. 13-25.

Dr. H.B.Kekre, Mr. Dhirendra Mishra, Ms. Stuti Narula, Ms. Vidhi Shah, "Color Feature Extraction for CBIR", International Journal of Engineering Science and Technology (IJEST), 2011, pp. 8357-8365.

Javed Hossain, M. Ashraful Amin, "Leaf Shape Identification Based Plant Biometrics", Proceedings of 13th International Conference

on Computer and Information Technology (ICCIT), 2010, pp. 458-463.

[8] Jiazhi Pan and Yong He "Recognition of plants by leaves digital image and neural network", International Conference on Computer Science and Software Engineering, 2008, pp 906-910.

Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, Qiao-Liang Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", International Symposium on Signal Processing and Information Technology, IEEE, 2007, pp 11-16.

Ji-Xiang Du, Xiao-Feng Wang and Guo-Jun Zhang, "Leaf shape based plant species recognition", Applied Mathematics and Computation 185 , 2007 pp 883-893.

Qingfeng Wu, Changle Zhou, and Chaonan Wang, "Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial Neural Network", Avances en Ciencias de la Computación, 2006, pp 5-12.