Computers and Electronics in Agriculture 157 (2019) 270–280

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Automated analysis of visual leaf shape features for plant classification

G. Saleem[a], M. Akhtar[b], N. Ahmed[c], W.S. Qureshi[d,*]

[a] Department of Computer and Software Engineering, National University of Sciences & Technology H-12, Islamabad, Pakistan
[b] School of Civil and Environmental Engineering, Research Centre for Integrated Transport Innovation, University of New South Wales, Sydney, NSW 2052, Australia
[c] Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, Pakistan
[d] Department of Mechatronics Engineering, National University of Sciences & Technology H-12, Islamabad, Pakistan

ARTICLE INFO

Keywords:
Segmentation
Feature extraction
Dimensionality reduction
K-nearest neighbours
Naive Bayes

ABSTRACT

A large number of studies have been performed during the past few years to automatically identify the plant type in a given image. Besides common object recognition difficulties arising mainly due to light, pose and orientation variations, the plant type identification problem is further complicated by the differences in leaf shape overage and changing leaf color under different weather conditions. The limited accuracy of existing approaches can be improved using an appropriate selection of representative leaf based features. This study evaluates different handcrafted visual leaf features, their extraction techniques, and classification methods. Towards this end, a new five-step algorithm is presented (comprising image pre-processing, segmentation, feature extraction, dimensionality reduction, and classification steps) for recognition of plant type through leaf images. The proposed algorithm is evaluated on a publicly available standard dataset 'Flavia' of 1600 leaf images and on a self-collected dataset of 625 leaf images. With the proposed algorithm, different classifiers such as k-nearest neighbor (KNN), decision tree, naïve Bayes, and multi-support vector machines (SVM) are tested. The best performing KNN, claimed for the final results, reveals that the proposed algorithm gives precision and recall values of 97.6% and 98.8% respectively when tested on 'Flavia' dataset. The proposed technique is also tested on our self-collected dataset, giving respectively 96.1% and 97.3% precision and recall measure results. Results confirm that our approach, when augmented with efficient segmentation techniques on raw leaf images, can be a significantly accurate plant type recognition method in practical situations. AlexNet, a Convolutional Neural Network (CNN) based approach is also compared for classification on the datasets as oppose to handcrafted feature-based approach and it is found that the later outperforms the former in robustness when the training dataset is small.

1. Introduction

The classification of plants into appropriate taxonomies is of great interest for professionals from different fields such as agronomists, environmental protectors, foresters, land managers and amateur gardeners. Typically, plant identification done by botanists is based on visual inspection of plant fruits, floral parts, and leaves . Since fruits and flowers appear on plants for a small period of time, they cannot be used alone as a suitable feature to identify plant types. The plant leaves based identification is more reliable for a visual cue based plant classification and identification problem. However, the existence of a large number of plant species makes it a quite challenging task to visually identify and recognize all plant species through manual inspection. The identification of the different types of plants can be achieved by their unique leaf texture, shape, venation architecture, and color. The color of the leaf can change in some scenarios with a change in weather but their basic shape, texture, and venation architecture remains almost the same. These unique leaf expressions have created interest for researchers to develop machine based automated plant species identification system. The existing literature describes various types of features which can potentially be used for plant identification. These features include circularity, rectangularity, eccentricity, roundness, and an aspect ratio based features extracted from the leaf. Other features based on wavelets, Fourier descriptors, texture, color, centroid contour distance, venation structure, and sawtooth patterns have also been used (Wu et al., 2007b). Wu et al. (2007b) used twelve features extracted from Flavia dataset to classify the plant types. There is a great variation in the literature on how to extract leaf based features to form the final set that can then potentially be used for further classification between different plant types. Another study (Aakif and Khan, 2015) used Fourier descriptor, morphological features, and shape defining features collectively for plant recognition. The geometric morphological

* Corresponding author.
E-mail addresses: gulshan.saleem14@ce.ceme.edu.pk (G. Saleem), m.akhtar@unsw.edu.au (M. Akhtar), waqar.shahid@alumni.ait.asia (W.S. Qureshi).

https://doi.org/10.1016/j.compag.2018.12.038
Received 23 February 2018; Received in revised form 2 November 2018; Accepted 16 December 2018

features along with tooth features were also used in plant identification systems to identify plant species with significant performance accuracy (Satti et al., 2013).

The selection of appropriate leaf features is considered as the most important aspects of an automated plant identification system. Numbers of techniques are present for the leaf identification including convolution neural network based method although the main essence of this work is to explore the leaf shape features and their importance in identification process. The convolutional neural networks (CNN) require a large amount of manually annotated data for training, which constrains its usage in applications where dataset is limited.

This work studies leaf visual cues and various extraction methods to obtain a highly optimal set of morphological, geometric and texture-based features. The selected features are first normalized and then transformed to principal components to reduce redundant dimensions. The features are then used to classify leaf images to different plant species. Different machine learning classifiers are evaluated including k-nearest neighbor, decision tree, naïve Bayes, and multi-SVM. A new five-step algorithm is presented for recognition of plant type through leaf images. The algorithm is evaluated using the publicly available standard dataset 'Flavia' of 1600 leaf images (Wu et al., 2007a) and a self-collected dataset of 625 images to check the generalization of proposed approach.

The remainder of this paper is organized as follows. The 'Materials and Methods' section reviews existing plant identification work carried out by the research community and gives details of available plant datasets. It also presents our proposed algorithm along with implementation steps (e.g., image resizing, segmentation, feature selection, extraction and normalization, dimensionality reduction and classification). The Results and Discussion section shows experimental results of proposed approach on both datasets and performance differences of used classifiers with explanation followed by the Conclusions section.

2. Materials and methods

2.1. Related work

The visual recognition of plant types may be easy for a botanist but for a machine, it is a complex and computationally expensive task. Wu et al. (2007b) used image processing methods to formulate an automated system for plant recognition through leaves. A dataset known as 'Flavia' was made available publicly for comparison along with their proposed method that used the probabilistic neural network (PNN) as a classifier for classification purposes. Initially, 12 leaf features were extracted and then the dimensions were reduced to only 5 main variables to form an input feature vector for the training and testing of the classifier. The technique has resulted in a high accuracy of 90.3% attained by training PNN with 1800 samples and to classify 32 distinct type of plant leaves. Krishna et al. (2010) used shape features to analyze the performance of three different classifiers. The study is based on using PNN classifier on five geometric features, and ten morphological features along with leaf vein features. For the feature space reduction, the principal component analysis (PCA) was used. The method resulted in an accuracy of 91%. The identification was done using Fourier moments and Support vector machine-BDT giving an accuracy of 71% and 96% respectively. In another study (Tyystjärvi et al., 2011) experiments were performed for maize and barley identification between six weed species. They used chlorophyll fluorescence fingerprinting for identification of both maize and barley. The research is based on the feature set of 17 features which were extracted through fluorescence induction curve and neural network classifier is used to classify results. The study claimed that the proposed method 86.7–96.1% correctly classify as crop (maize or barley) or weed and for single species provide accuracy of 50.2–80.8%. Arun Priya et al. (2012) proposed an algorithm to perform efficient leaf recognition using support vector machine. The algorithm comprised of preprocessing, which includes a gray level based boundary enhancement, feature extraction, and classification. The extracted features comprise twelve digital morphological features (DMFS). Their proposed approach has been tested on the Flavia and local datasets. Their classification takes minimum execution time as compared to KNN and SVM classification methods when tested on the Flavia dataset. According to Kadir et al. (2013) not only the shape of leaf but its color and texture are also important elements. The modified system showed an improved accuracy of 93.75% as compared to 90.31% as obtained by the original work. Their modified method used Fourier descriptors, slimness ratio, roundness ratio, and dispersion for shape. Color based moments such as mean, standard deviation, and skewness were used and the system also used lacunarity for texture, and after completing feature extraction process, a probabilistic neural network (PNN) classifier was applied. The algorithm was tested on the Flavia dataset. Plant tip and the base are important areas for feature extraction and their accurate determination is difficult but they provide significant support to the plant recognition. Puja et al. (2013) designed a leaf recognition system which is pre-step for plant disease identification. The authors used a two-step algorithm, involving segmentation and classification steps. The system utilizes PCA for dimensionality reduction and SVM as a classifier. Scaling is used to handle the varying size of plant images. The proposed method results in an accuracy of 77.96% which can be improved by incorporating more leaf features. Satti et al. (2013) developed a novel approach for plant identification, based on three stages, namely preprocessing (to remove noise and correct damaged data), feature extraction (color and shape based features used), and classification using ANN and Euclidean KNN classifiers. The algorithm was tested on Flavia dataset, the system has shown accuracies of 93.3% and 85.9% respectively under ANN and KNN classifiers. Gwo and Wei (2013) proposed a plant identification algorithm that uses the leaf features calculated from key points on leaf contours. In their feature extraction process, a centroid is found and distance of each contour point from that centroid is computed. These distances are gathered to form a length histogram, which is then normalized. The fuzzy score algorithm is used on that normalized histogram to compensate for the differences in length histograms of same species. The fuzzy score matrix is then fed to a Bayes classifier for identification of plant species. It is claimed that the methodology is robust to rotation and scale changes of the leaf image and is shown to outperform some of the existing methodologies. Their proposed methods show the top accuracy of 92.7%. Lavania and Matey (2014) designed a scale-invariant feature transform based leaf recognition method which aims to avoid human involvement in the process of feature extraction. The technique uses key descriptors of scale invariant feature transform (SIFT) for edge detection and classification. The Mean projection algorithm is used as a second method for contour based edge detection. They have used leaf images as input and carried out feature extraction to extract image features for computation of key points. The computed key points are stored and then matched with leaf images from the dataset for plant classification. The authors have drawn a comparison of their proposed method with SIFT and contour based edge detection method and proposed method giving 87.5% accuracy with the Flavia dataset (Tsolakidis et al., 2014) conducted a study based on Zernike moments and histogram of oriented gradients for plant identification. For shape feature extraction they used Zernike moments, while for texture-based feature extraction they used a histogram of oriented gradients. The SVM-based classification of the system was tested on Flavia and Swedish leaves and the authors claim that their techniques perform better than other existing methods.

Kadir (2014) proposed an approach exploiting gray-level co-occurrence matrix (GLCM), lacunarity, and Shen features. The authors used color, shape, texture, and vein feature for plant identification. With the help of a Bayesian classifier, the proposed approach (containing preprocessing, segmentation of leaf, feature extraction, and classification steps) has been shown to give plant identification accuracies of 97.19%

and 95% respectively on Flavia and Foliage datasets.

Pallavi and Devi (2014) worked on plant identification using feature extraction and Zernike moments. Their framework uses shape, vein, color and texture features to classify plants and a neural network classifier is used for testing and training of dataset. The authors claim that the system works faster as compared to other existing methods used for identification of plants. The proposed method can be modified using features from color moments to classify leaves for disease identification in plants.

Joly et al. (2014) contributed towards identification and classification of plants. The researchers designed web-based and mobile applications of their proposed system making it attractive and friendly for potential users. The system has been tested on above half of the plant species found in France (approximately 2200 species). It has been shown offering an average accuracy of 85% at low brute force despite classification process being run on many different categories at the same time. The authors claim that proposed system works on five different organs of plants not only on leaf features which make their technique more effective and it is also capable of handling user's queries.

Hsiao et al. (2014) reported the implementation of sparse model based dictionary learning approach for plant identification. The authors have claimed to propose a first sparse based leaf recognition system which is also capable of adding new species without going through retraining of the system which has made it a more flexible method as compared to previous approaches. The proposed method also inherits the property of sparse representation (due to sparse coding) which provides dense and comprehensive representation for leaf image and robust false leaf feature extraction. The authors' proposed model has been shown to give plant identification results with a high accuracy of 95.47% and also have drawn a fair comparison of their technique with a bag-of-words approach which has an accuracy of 94.38%. The proposed method is more efficient than previous methods when tested on the Flavia dataset and can be modified by incorporating more leaf features in it.

Munisami et al. (2015) designed an automated system for plant recognition with a mobile application allowing to take and upload a picture of a plant leaf. The system extracted leaf features (such as length, area, perimeter, hull area, hull perimeter, distance map along the vertical and horizontal axes, a color histogram, and a centroid-based radial distance map), and classified with the help of KNN classifier (and refined results using the color histogram) has been shown to give 87.3% accuracy when tested on Folio Dataset of 640 images from 32 different species.

Aakif and Khan (2015) proposed a plants identification algorithm using three steps, namely pre-processing of the dataset, leaf feature extraction, and classification. Their method utilized morphological features, Fourier descriptors, and shape-defining feature. The artificial neural network based (ANN) classification when applied on 817 leaf images from 14 different fruit plants resulted in a high accuracy of 96%. Moreover, authors also tested their proposed methods on two more datasets which are Flavia and ICL datasets.

Salve et al. (2016) used Zernike moments and histogram of the oriented gradient (HOG) as shape descriptor features and tested their method on a dataset of 50 different plants. First, they have done preprocessing on the images and then applied Zernike moments and histogram of oriented gradients for feature extraction. Due to robustness and feature persistence, HOG has shown better performance as compared to Zernike moments. They reported an accuracy of 84.66%. Their results can be further improved by combining above mentioned features with some leaf based features.

The classification of leaf mainly requires a proper description of shape, texture and the color information and the first important task is to represent captured object into mathematical form for theses mentioned parameters. Prasad et al. (2017) have claimed to propose a novel method with low computation which is efficient, accurate rotation-scale-translation invariant shape profile transforms called Angle View Projection (AVP). The method transforms captured images to AVP shape profile curve and compresses using discrete cosine transform for the overall improvement of the system. For evaluation of the proposed method, the algorithm is tested on five different datasets and results have proved that system outperforms existing method and also fast in response. The authors also claim that system work on incomplete leaf images which often happens due to physiological or pathological activity (Prasad et al., 2017).

Another study is based on a novel shape descriptor, they used periodic wavelet descriptor (PWD) to create a database containing features of different plant species. The backpropagation neural network is firstly trained on PWD features and then leaf identification is done which provides 90% accuracy (Zeng et al., 2017).

The single state extraction and simple structure of classifier used in the leaf identification process have made the identification rate low. Sound modification in these two can make identification process efficient and accurate. Liu and Kan (2016) proposed a new method of identification that uses a combination of two different features, which are based on shape and texture cues. Moreover, authors have used deep belief neural networks (DBNNs) as a classifier. The local binary pattern, Gabor filter, and gray level co-occurrence matrix are used to derive texture features and Hu moment invariants and Fourier descriptors are used for shape features. The proposed algorithm has used "dropout" method for training of DBNNs which helps to overcome the overfitting problem. The results showed that the algorithm is robust and have higher recognition rate as compared to the existing approaches of leaf identification (Liu and Kan, 2016).

Table 1 summarizes details such as method adopted by different researchers, features used, pros, and cons of selected existing approaches for plant identification. To be consistant with our work only methods using Flavia dataset are used.

2.2. Dataset

Plant species classification based on leaf data is an important area of research. Several datasets of plant leaves are available. Each dataset has different features, recording conditions, the number of classes, and samples in each class so that the performance of an algorithm tested on one dataset cannot be directly compared with its performance when tested on another dataset. That is why we have evaluated our technique both on a publically available dataset i.e. the Flavia dataset (Wu et al., 2007b) and also on a self-collected dataset to show the real-time performance of the method. Flavia is a widely used dataset which requires minimum preprocessing. It contains controlled images with truncated stem and scanned with a white background. The dataset contains leaf images of 32 species with more than 50 images in each class. The self-collected dataset shown in Fig. 1 contains 625 images of 25 different species with 25 images in each class and also having the plane background. The self-collected dataset is prepared by taking images of leaf with white paper under leaf to make image background clear. Other notable datasets are UCI Plant Leaf Dataset (Silva et al., 2013), UCI One Hundred Plant Species Dataset (Mallah et al., 2013), Herbarium Dataset (Ragupathy, 2016), and LeafSnap Dataset (Kumar et al., 2012). In our evaluation of plant identification approaches, we have used Flavia datasets due to its attractive average sample size and the images are available with white background.

2.3. Proposed method

The proposed algorithm consists of five distinct stages of processing towards the plant species identification (Fig. 2). The detailed operation of each step is explained in the following sections.

Table 1
Summary of existing methods of plant identification using Flavia dataset.

Reference	Technique	Features	Pros and Cons
Wu et al. (2007b)	It is a pioneer research reporting the design of Flavia dataset and classifying proposed work with the help of Probabilistic neural network (PNN) classifier (PCA used for preprocessing)	7 Shape Features: smooth factor, aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of diameter, perimeter ratio of physiological length and width 5 Vein features	**Pros:** High accuracy, simple and easy to apply, general purpose **Cons:** Improvement is required in the selection of feature set
Krishna et al. (2010)	It reports the use of three different classification techniques. It has been shown that the SVM classifier with binary decision tree gives the highest accuracy	Shape Features: Diameter, physiological length, physiological width, perimeter feature, and vein features	**Pros:** High accuracy **Cons:** Lacks novelty
Arun Priya et al. (2012)	This work is based on the use of kernelized support vector machine for classification to identify plants. The system has been evaluated on two datasets of 10 different plant species	Basic geometric, and digital morphological features	**Pros:** Good computational results, high performance, adaptable to new plant species, and minimum execution time **Cons:** A limited number of features, and less feature diversity
Gwo and Wei (2013)	Their system calculates leaf contour and then after selecting key points, it makes use of fuzzy score algorithm to manage the differences in length histograms of same species. Bayes classifier has been used for the classification purposes	Leaf contour normalized length histogram	**Pros:** Great accuracy, outperforms Zernike moments and curvature scale space approaches. **Cons:** Limited leaf features
Satti et al. (2013)	This research is based on color and shape based features to classify leaf types using aritifical neural network (ANN) and KNN classifiers. The system has been tested for 32 different leaf species	**Color:** Average means. **Shape:** Geometric morphological tooth	**Pros:** Good computational results, high performance, adaptable to new species of plants **Cons:** A limited number of features, lack of feature diversity
Kadir et al. (2013)	The research uses color and texture of leaf with its shape as features for identification of plants. After feature extraction, a PNN classifier has been applied	Fourier descriptors, slimness ratio, roundness ratio **Shape:** Dispersion **Color moments:** Mean, standard deviation, and skewness	**Pros:** High accuracy **Cons:** Many features used but little improvement. Not tested with other better classification methods
Lavania and Matey (2014)	The research uses scale invariant features transform (SIFT) for leaf identification. The proposed work is also compared with two algorithms making use of curvature scale space (CSS)	SIFT and CSS	**Pros:** High precision **Cons:** Less accuracy than some existing methods
Tsolakidis et al. (2014)	For shape feature extraction, the approaches employ Zernike moments and histogram of oriented gradients. The plant classification results are shown using the SVM classifier	Zernike moments, histogram of oriented gradients	**Pros:** High accuracy **Cons:** Lacks a number of features, slow execution time mainly due to SVM classifier
Kadir (2014)	The study is based on plant identification using gray-level co-occurrence matrix, lacunarity, and shape based features. The developed system has been tested on two datasets, giving better accuracy on Flavia as compared to Foliage dataset	**Shape:** Polar fourier transform, the central moment, and convex hull. **Color:** Mean (μ), standard deviation (σ), skewness (θ) and kurtosis (δ) **Texture:** Gray-level co-occurrence matrix, and lacunarity **Vein:** Morphology operations	**Pros:** High accuracy, demonstrated better performance of the newly developed system when compared with different existing approaches. **Cons:** A limited number of features, and less feature diversity
Hsiao et al. (2014)	The study is based on comparative analysis of two existing techniques, propose, and implement a sparse model for plant identification	SIFT	**Pros:** Better accuracy , no training required **Cons:** Implementation cost is too high, based on a single feature
Aakif and Khan (2015)	It involves the use of Fourier descriptors, morphological and shapes defining a feature for identification of plant using the leaf. The ANN has been used as a classifier	Fourier descriptors, morphological and shape defining feature	**Pros:** High accuracy. **Cons:** Testing methodology and claimed results are unclear

2.4. Preprocessing and segmentation

Pre-processing is an important step for an image recognition system. The input images may have been captured with different devices in varying ambient conditions, resulting in varying contrast and noise sufficient to cause inconsistent feature extraction. Moreover, the colored images are usually stored in RGB color profile which is a device dependent color profile. Since we are using Flavia dataset for evaluation of our proposed methodology, it contains highly cleaned and restricted images with fixed white background and well-defined color profile. There is no need to apply noise reduction or contrast improvement algorithms as the small effect of noise was reduced in smoothing after the segmentation. The Flavia images come in different sizes but having a fixed aspect ratio. To make all images of equal size (i.e., 800 × 600 pixels), they are transformed. The segmentation process involves the separation of leaf from its background. The extraction of a region of interest (ROI) may involve complexities if the background contains indistinguishable content. However, sophisticated

Fig. 1. Self-collected dataset leaf images.

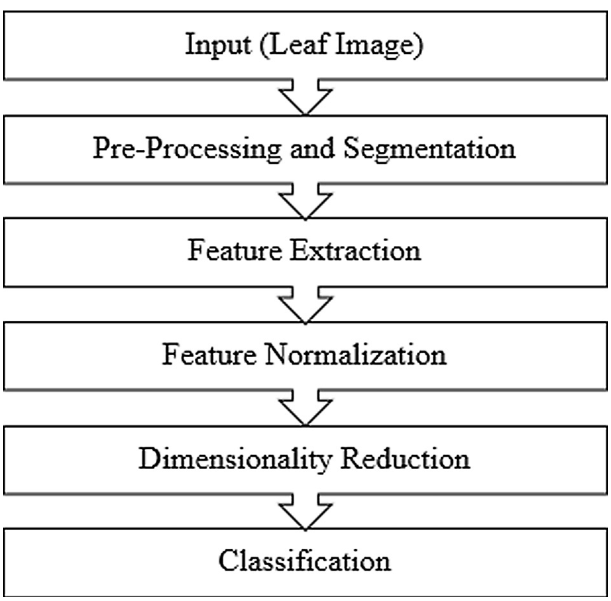

Fig. 2. Flow diagram of proposed plant recognition technique based on five major steps for identification of plant (Preprocessing and Segmentation, Feature Extraction, Feature Normalization, Dimensionality Reduction, and Classification).

$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$

Fig. 3. Laplacian operator to be used for smoothing of the binarized intensity image.

segmentation algorithms can be employed to precisely extract the ROI for further extraction of desired features. In the Flavia dataset, the leaf background is a clear white color making it easy to segment the ROI. The RGB color image of a leaf is converted to L × a × b color space, where 'a' and 'b' are color dependent values. The luminosity 'L' layer is, however, used for further processing of initially processed images. This intensity image is binarized by calculating the threshold using Otsu's algorithm (Otsu, 1975). We only need the leaf boundary in order to extract shape features. The binarized leaf image is processed with a 3 × 3 Laplacian operator (Fig. 3) to smooth leaf edges and regulate small glitches.

The application of Laplacian operator results into an image with only 1-pixel wide leaf edges. The same pre-processing technique is used for the self-collected dataset. The process of pre-processing and segmentation is depicted in Fig. 4.

2.5. Feature extraction

Distinct features of a leaf image can be based on its shape, texture, and color (Wu et al., 2007b). In this work, we have extracted shape features to differentiate different leaves from each other based on their geometry. The texture features are also incorporated to differentiate between the leaves of indistinguishable shapes. After performing an extensive literature review, we have used a careful blend (manual selection) of different features to form the basis of a reasonably accurate plant identification system and then PCA is used to remove redundancy which occurs due to interrelated information of different features i.e. leaf area is dependent on both length and width of leaf. The overall feature set consists of 11 shape features, 7 statistical features, and 5 venation features (and Fourier descriptors). The shape features are related to the leaf geometry and are calculated using morphological operations on binary leaf image. The statistical features are a compact representation of intensity image (L). The venation features are also extracted from the intensity image (L) as morphological features. The Fourier descriptors are shaped approximation of the leaf comprising a feature vector. The parameters related to the calculation of different features are discussed as under.

(i) Leaf Length: It is the longest distance between any two points on the extracted leaf boundary and is denoted by D.
(ii) Leaf Width: It is the longest line perpendicular to the length line. The orthogonality of the line will also be assumed even when the

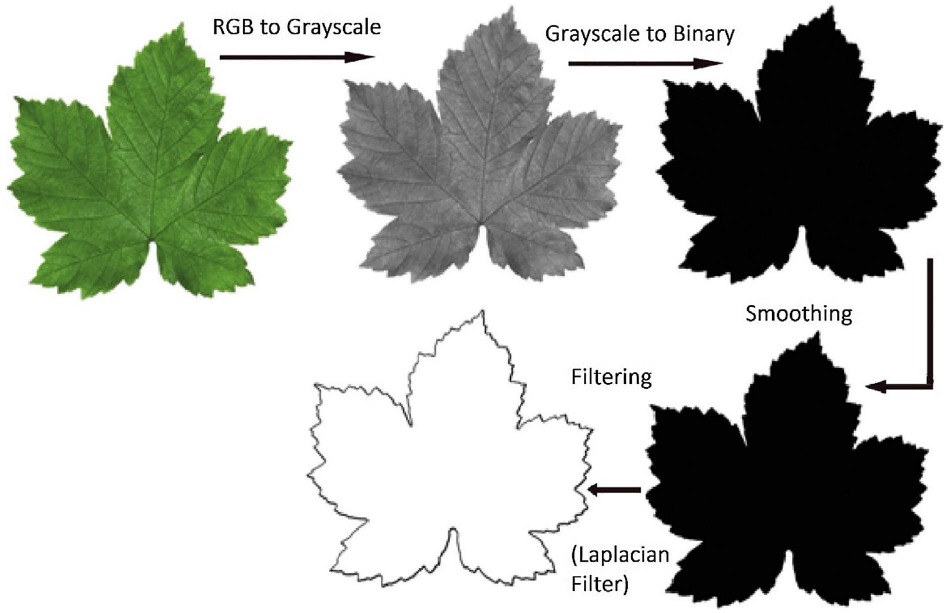

Fig. 4. Results of pre-processing and segmentation processes of our proposed algorithm.

perpendicular is at 90° ± 0.5° due to pixel distribution of leaf and is denoted by W.

(iii) Leaf Area: It is the number of pixels on the smoothed leaf. Every leaf has its own area characteristics as the leaf is completely occupying the 800 × 600 pixels space.

(iv) Leaf Perimeter: It is calculated by counting the number of pixels which lies on the boundary of processed leaf.

(v) Smooth Factor: The binary leaf image is processed to calculate the smooth factor which measures the difference of areas of two leaves processed with a smoothing filter of 5 × 5 and 3 × 3 kernels.

(vi) Aspect Ratio: Aspect ratio is an important criterion which measures the ratio of leaf length and leaf width $AR = \frac{D}{W}$, whereas D, W, and A are respectively the length, width and area of leaf.

(vii) Rectangularity: It is the measure of the similarity between a rectangle and the leaf shape or simply how much leaf shape matches with a rectangle. It is calculated by $\frac{D \times W}{A}$.

(viii) Form Factor: It is a measure of the difference between the circle and leaf shape. It is calculated by $\frac{4\pi A^2}{P^2}$, where A and P are the area and perimeter of the leaf, respectively.

(ix) Elongation: The elongation of the leaf is calculated as: $\frac{D}{W}$, whereas D and W are length and width of leaf.

(x) Longitudinal Spreading: It measures the longitudinal spreading of the leaf as $\frac{P}{D}$, whereas P and D are perimeter and length of leaf.

(xi) Cross-Sectional Spreading: The cross-sectional spreading of the leaf is measured by both length and width of the leaf. It is calculated as $\frac{P}{D \times W}$, whereas P is the perimeter of the leaf and D and W are the length and width of the leaf respectively.

(xii) Average Intensity: It is mean (μ) of intensities in intensity image *(L)*. It can be calculated by averaging all the pixel values of intensity image:

$$\mu = \frac{\sum_{k=1}^{N} L_k}{N} \tag{1}$$

where L is the intensity and N are the numbers of pixel values.

(xiii) Average Contrast: It is the measure of the second central moment of image intensities in intensity image *(L)*. It is defined as variance and can be calculated as:

$$\sigma^2 = \frac{\sum_{k=1}^{N} (L_k - \mu)^2}{N} \tag{2}$$

where L is the image intensities, N is the number of pixel values and μ is the mean of image intensities.

(xiv) Skewness: It is the measure of the third moment of pixel values in intensity image *(L)*. It provides the skewness or asymmetry of the pixel values distribution. Skewness can be positive or negative indicating that histogram is skewed to the right side or the left side respectively. It is:

$$\gamma = E\left(\frac{L}{\sigma}\right)^3 \tag{3}$$

where L is the image intensities, E is the expectation operator, μ is the mean of image intensities and σ is the standard deviation.

(xv) Kurtosis: It is the measure of the fourth moment of pixel values in intensity image *(L)*. If the histogram of an image is normally distributed its kurtosis will be 0 and if it is uniformly distributed its kurtosis would be negative. The positive value of kurtosis is obtained when it is peaked more than the normally distributed histogram. Kurtosis is calculated as:

$$K = n\frac{\sum_{k=1}^{N} (L_k - \mu)^4}{\left(\sum_{k=1}^{N} (L_k - \mu)^2\right)^2} \tag{4}$$

where L is the image intensities, N is the number of pixel values and μ is the mean of image intensities.

(xvi) Smoothness: Smoothness is the measure of relative intensities in the segmented region of the image. It can be calculated by the following equation:

$$R = 1 - \frac{1}{(1 + \sigma^2)} \tag{5}$$

where σ^2 is the variance.

The measure of smoothness will return values of 0 for a region of constant intensities, and 1 when the region exhibit maximum disproportion in intensities.

(xvii) Uniformity: It measures the equality of the grayscale values of the image, as follows:

$$U = \sum_{k=0}^{L-1} p^2(z_i) \tag{6}$$

where Z is a random variable for image intensity and p is its histogram and L is the number of intensity levels and at equal levels of intensity, uniformity attains its maximum value.

(xviii) Entropy: It is a measure of randomness in the values of image pixels and defined as:

$$\sum (p \times \log_2(p)) \tag{7}$$

where p holds the histogram counts.

(xix) Venation Measurement: The arrangement of the vein structure in the leaf is an important piece of information to distinguish between the leaves having closely matching shapes. Venations are calculated from the intensity image by applying the morphological operations. Four disk-shaped morphological operators of radius 1–4 mm are applied to the leaf and the obtained value is subtracted from the leaf area to acquire the vein structure. The areas of the vein structures, denoted by v_1, v_2, v_3 and v_4, are calculated using total number of pixels. Consequentially, the five features are calculated as $\frac{v_1}{A}$, $\frac{v_2}{A}$, $\frac{v_3}{A}$, $\frac{v_4}{A}$ *and* $\frac{v_4}{v_1}$.

(xx) Fourier Descriptors: Fourier descriptors provide a way to encode an image boundary by mapping every pixel position (x, y) into a complex number $(x + \imath y)$. The Fourier transform can be used to recover the original image with fewer terms resulting in a simplified and smooth image. Fourier descriptors inherit some properties of Fourier transform such as translation invariance, scaling, and rotation. In automatic plant identification, these descriptors can effectively be used to encode leaf shape. The Fourier descriptor of a leaf boundary can be calculated as:
 (i) Record the coordinate values of each pixel sequentially (moving clockwise along the shape)
 (ii) Construct a complex-valued vector using coordinate values recorded in step (i) i.e., $(x, y) \to (x + \imath y)$.
 (iii) Take DFT of the complex-valued vector

The inverse Fourier transform would be used to obtain the original shape of the leaf.

Fig. 5 provides correlation analysis of features in the form of linear correlation coefficient. Horizontal and vertical axes contain the features and their correlation with each other is calculated and plotted. It is visible that morphological features have high correlation with each other and dimensionality reduction is a necessary step to reduce the dependencies of features on each other's. The Boxplot of feature set is provided in Fig. 6 and it indicates that values for first 18 features have smaller range and less variability. The feature 19 to 24 have wider range and the values fall much farther than the standard deviation. Fig. 2 indicates that scaling of features will be beneficial and will be helpful in distance based classifiers as the range of values are not consistent among different features.

It refers to the scaling of feature values in a range from 0 to 1, which is usually done before the classification process. It is an important step to avoid biased performance of the classifier as the values of features are calculated by employing different approaches and the range of their values may differ significantly. In this study, we calculated the maximum (max) and minimum (min) value of each feature and normalized using the following equation:

$$y = mx + c \tag{8}$$

where m = 1/(max − min), with × and y being input and output values of features and c is the y-intercept.

3. Dimensionality reduction

There are two different approaches used to reduce the dimensions of the feature vector. The first one is feature subset selection which selects the most optimal features to be used as a final feature set, whereas the second approach applies the dimensionality reduction using orthogonalization. In this work, we have used second approach only to eliminate the redundancy issue in feature set dimension. This employs principal component analysis (PCA) to reduce the dimensionality (Shlens, 2014). The PCA projects data from its existing coordinate system to a new coordinate system. Only the first ten principal components (f: $R^{24} \to R^{10}$, where R is the feature vector, superscript number represent the dimension of vector) are selected to form the feature space which corresponds to nearly 95% of existing feature space. So by applying dimensionality reduction, the feature set is reduced to ten new features which are a combination of all twenty-four features done through PCA. The new data would have reduced dimensionality with slightly reduced variance. The main objective of dimensionality reduction is to improve performance of the proposed algorithm as it uses only high contributing principal components. In the results section, you would note Tables 2 and 3 providing an outlook of results with and without dimensionality reduction.

4. Classification

The final stage is a classification of resultant feature vectors according to species to which they belong. We describe here a number of classification methods that we have used in our work to classify the extracted features into different plant species. In our case, we have employed KNN, decision tree, naïve Bayesian, and multi-support vector machine (SVM) classifiers, as discussed below.

(i) K-Nearest Neighbor

The k nearest neighbor (KNN) is a versatile classifier with a wide range of applications from vision to proteins and computation to graphs. It does not work on assumption but uses only real data which makes it more useful for practical applications. Another advantage of using KNN is that it uses instantaneous training which means as soon as new sample data is written on database in short, it provides fast training. It works on the minimum distance of query instances from training points which further gives k- nearest neighbors. The KNN classifies data based on the majority of its neighbors having most common attributes and also with minimum distance from query object. The neighbors of an object are selected based on distance such that the neighbor on and below k-distance are selected (Weinberger and Saul, 2009). For instance, we have used 3 as K after testing few other values such as 5 and 7 but the results are better with 3.

(ii) Decision Tree

It generates a tree structure regression model. It works through breaking down datasets into a smaller subset which cannot be further sub-divided, which leads to the formation of decision tree incrementally. The resultant tree contains decision as well as leaf nodes. Usually, the decision nodes are further divided into two or more branches and leaf node states the final classification result. The decision tree also has a root node which is the topmost decision defining node related to best predicator. This method works for both numerical and categorical types of datasets. It mostly works on two measures which are entropy and information gain to build a final decision tree (Fayyad and Irani, 1992).

(iii) Naïve Bayes Classifier

The naïve Bayes is a statistical classifier which is based on Bayes'

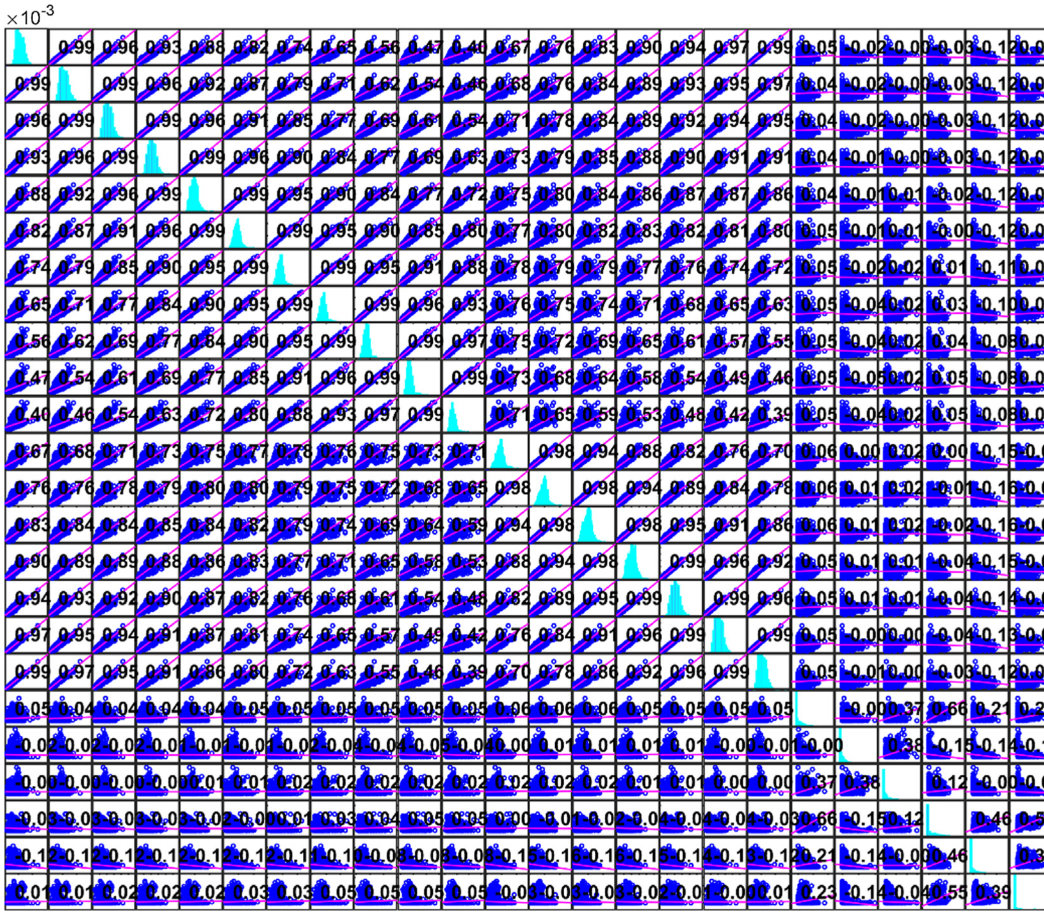

Fig. 5. Linear correlation between features.

theorem. It is used to predict probabilities of a member belonging to a specific class, which means that it gives the probability of occurrence of a given data points within a particular class. The naïve Bayes works by assuming that attribute value of a given class do not depend on values of other attributes which is known as class conditional independence. This makes the computation of algorithm simple and in this way considered as naïve (Lewis, 1998). The following equation is used to explain the principle of Bayes' theorem:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{9}$$

And P(H|X) is the posterior and P(H) is the prior probability of class (target) whereas P(X|H) and P(X) are the likelihood and prior probabilities of predictor respectively.

(iv) Multi Support Vector Machine

The support vector machine (SVM) is a supervised learning model that can be used for data classification and regression analysis (Vapnik, 2013). For non-linear transformation, the optimal hyperplane is created by separating positive and negative classes with maximum margins. The SVM uses a kernel function for data transformation into higher dimension space. It works on support vector points which lie nearest to separating hyperplane points (Weston and Watkins, 1999). A nonlinear support classifier can be created by replacing the product (x, y) with $K(x, y)$. The following equation determines the membership of x.

$$f(x) = sign\left(\sum_{i=1}^{j} a_i y_i K(x_i x) + b\right) \tag{10}$$

The SVM is inherently a binary classifier which classifies data into two different classes. For the leaf based plant identification task, the problem of classification involves more than two classes. There are several methods to deal with the problem of the multiclass classification problem. The approach used in this study is one-vs-all and chooses the class which classifies data with maximum margin. In this approach, a single classifier is trained for every class with test cases of that class as positive and all other as negative (Weston and Watkins, 1999).

5. Evaluation measure

The accuracy of the proposed plant recognition system has been computed using the following expression which uses numerical details of correctly classified plants from total sample space of leaf images in the dataset.

$$\%Accuracy = \left(\frac{no.\ \ of\ \ correctly\ \ recognized\ \ samples}{total\ \ no.\ \ of\ \ samples}\right) \times 100 \tag{11}$$

The precision and recall are also the important measure to consider for system evaluation and are calculated as in (12) and (13) respectively:

$$Precision = \frac{\sum True\ \ Positive}{\sum Predicted\ \ Condition\ \ Positive} \times 100 \tag{12}$$

$$Recall = \frac{\sum True\ \ Positive}{\sum Condition\ \ Positive} \times 100 \tag{13}$$

6. Results and discussion

The proposed plant recognition system was tested on the Flavia

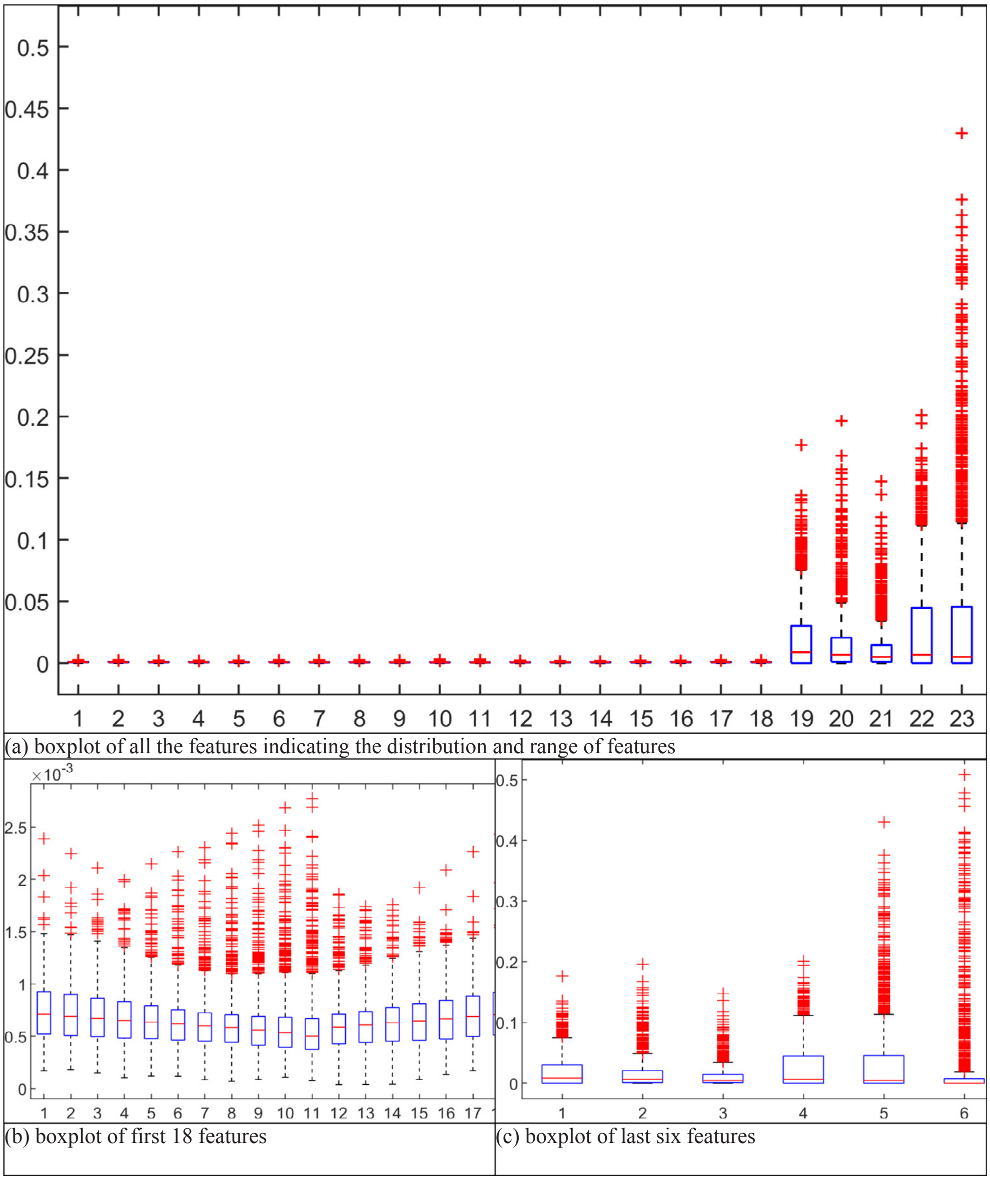

Fig. 6. Boxplot of 24 Feature Normalization.

dataset containing leaf images of 32 plant species with 1600 samples and self-collected dataset of 25 species with 625 samples. We have applied 5-fold cross-validation without holdout and from each class 80% of all used as training and 20% as test samples and gives 96.8% accuracy. Our feature set was comprised of eleven shape features, seven statistical features, and five venation features and Fourier descriptors. We have trained and tested four different classifiers (KNN, decision tree, naïve Bayesian, and multi-SVM). Firstly, we generated results without applying any dimensionality reduction technique and then we applied PCA to reduce dimensions of our feature vector (f^{24} - f^{10}). Table 2 shows recognition results for different species within Flavia dataset (both with and without dimensionality reduction). The Table 2 includes results attained from all 32 species of Flavia dataset and also includes Precision and recall results for all species only with a k-nearest neighbor classifier (KNN). The Output column in Table 2 contains the number of incorrectly classified leaves against every species and then their cumulative score is used to calculate Precision and Recall values for analysis of the proposed method. Here, incorrect recognition contains both false positive and true negative output samples and correct recognition includes all true positive and false negative values after application of proposed method.

Table 3 gives a summary of classification results and comparison of proposed plant recognition system with existing approaches which have used Flavia dataset for evaluation of their technique. Note that there is a slight decrease in accuracy of the proposed algorithm with dimensionality reduction due to reduced feature set but make it computationally better. Our proposed system with KNN classifier gives an accuracy of 98.75% with Flavia dataset and 97.25% with the self-collected dataset. The slight decrease in accuracy of the self-collected dataset is due to the highly refined quality of Flavia dataset. Moreover, since this data set was collected by placing a white background sheet when taking leave images, in practical situations the data can first be pre-processed through segmentation methods to obtain occlusion-free leave before application of our method. The precision and recall on the Flavia dataset are 97.6187% and 98.75% respectively.

The probable reason of good performance of the proposed approach is due to the fact that our system has employed the most comprehensive set of leaf parameters to extract best discriminatory features. It can be seen that the proposed technique with naïve Bayes has shown the second best accuracy of 93.95%, decision tree provides 90.56% and SVM provides 88.93% accuracy. The differences occur in terms of their processing speed as well such as decision tree is the fast one and the

Table 2
KNN classification results for each species of Flavia dataset.

Details of species	Input	Output			
Scientific Name (Each Species has its local name as well)	Testing Sample	Number of Incorrect Recognition (false Positive + True Negative)		Precision %	Recall %
		With Dimensionality reduction	Without Dimensionality reduction		
Phyllostachys edulis (Carr.) Houz.	50	0	0	100	94
Aesculus chinensis	50	0	0	100	100
Berberis anhweiensis	50	0	0	100	94
Cercis c0hinensis	50	1	0	96.0784	96
Indigofera tinctoria L.	50	0	0	100	100
Acer Dalmatum	50	0	0	100	100
Phoebe zhennan S.	50	0	0	100	100
Kalopanax septemlobus (Thunb. ex A.Murr.) Koidz	50	0	0	100	100
Cinnamomum japonicum Siebold ex Nakai	50	2	2	92.3077	100
Koelreuteria paniculata Laxm.	50	0	0	100	100
Ilex macrocarpa	50	0	0	100	100
Pittosporum tobira (Thunb.) Ait. f.	50	0	0	100	100
Chimonanthus praecox L.	50	2	2	92.3077	100
Cinnamomum camphora (L.) J. Presl	50	3	2	88.6792	100
Viburnum awabuki	50	0	0	100	100
Osmanthus fragrans Lour.	50	3	3	88.6792	100
Cedrus deodara (Roxb.) G. Don	50	2	2	92.3077	100
Ginkgo biloba L.	50	0	0	100	100
Lagerstroemia indica (L.) Pers.	50	0	0	100	100
Nerium oleander L.	50	0	0	100	100
Podocarpus macrophyllus (Thunb.) Sweet	50	0	0	100	100
Prunus × yedoensis Matsumura	50	0	0	100	100
Ligustrum lucidum Ait. f.	50	0	0	100	100
Tonna sinensis M. Roem.	50	0	0	100	100
Prunus persica (L.) Batsch	50	1	2	96.0784	98.
Manglietia fordiana Oliv.	50	3	3	88.6792	94.0
Acer buergerianum Miq.	50	0	0	100	100
Mahonia bealei (Fortune) Carr.	50	0	0	100	100
Magnolia grandiflora L.	50	0	0	100	100
Populus × canadensis Moench	50	3	1	0.886792	94.0
Liriodendron chinense (Hemsl.) Sarg.	50	0	0	100	100
Citrus reticulata	50	0	0	100	100
Total Samples	**1600**	**20**	**17**	**97.6187**	
Accuracy	**98.75%**		**98.93%**		**98.75**

Table 3
Summary of classification results and comparison of the proposed approach with different classifiers using Flavia and self-collected dataset.

Technique	Flavia Dataset		Self-Collected Dataset	
	Classifier	Accuracy	Classifier	Accuracy
Proposed Technique (with Dimensionality Reduction)	KNN	98.75%	KNN	97.25%
	Decision Tree	90.5%	Decision Tree	88.9%
	Naïve Bayes	93.95%	Naïve Bayes	92.1%
	Multi SVM	88.7%	Multi SVM	87.45%
Proposed Technique (without Dimensionality Reduction)	KNN	98.93%	KNN	97.75%
	Decision Tree	90.65%	Decision Tree	89.115%
	Naïve Bayes	94.5%	Naïve Bayes	93.33%
	Multi SVM	88.91%	Multi SVM	87.15%

SVM is slowest but here processing speed of different classifier is not the part of study.

The difference between performances of each classifier is due to their few limitations such as SVM provides low accuracy due to data biasing and naïve Bayes performs good but slightly decreased accuracy is due to lack of all posterior probabilities.

The above-mentioned classifiers show different accuracies in comparison to each other and KNN outperforms all others for leaf identification. After experimenting proposed approach, we conclude that KNN performs better than other classifiers for identification of leaf. Table 4 provides classification accuracy of famous existing methods and of the proposed method to give clear picture of how our method outperforms other major leaf identification techniques.

Table 4
Comparison of proposed technique with existing approaches using Flavia dataset.

Technique	Classifier	Recognition Accuracy
Wu et al. (2007b)	PNN	90.3%
Krishna et al. (2010)	PNN	91%
Arun Priya et al. (2012)	KNN, SVM	94.5%
Gwo and Wei (2013)	Bayesian	92.7%
Satti et al. (2013)	ANN	93.3%
Kadir et al. (2013)	PNN	93.75%
Lavania and Matey (2014)	SIFT	87.5%
Tsolakidis et al. (2014)	SVM	97.18%
Kadir (2014)	naïve Bayesi	97.17%
Hsiao et al. (2014)	Sparse based model	95.47%
Aakif and Khan (2015)	ANN	96%
Alex et al. (2012)	AlexNet CNN	**99.48%**
Proposed Technique	**KNN**	98.75%

7. Convolutional neural networks for plant identification

Convolutional Neural Networks has demonstrated state-of-the-art performance in many computer vision tasks. The CNN doesn't require hand-crafted features and they automatically learn discriminative feature representation from images and eliminate the need of hand-crafted features. The limitation with CNN is the need of large image set in order to learn the feature representation and classification. The task under

Table 5
CNN results for plant identification.

Dataset	Accuracy
Flavia	99.48%
Self-Collected	71.12%

consideration have 1600 images and 32 classes which are used for training of final model and it is cross-validated on self-collected image set to check the generalization. There are on average 50 images per class which are not sufficient to fully learn the features. AlexNet is a pre-trained CNN for the task of image classification with demonstrated performance. Transfer learning is used to train AlexNet on Flavia dataset with 80% images used as training data and 20% as test data. For cross-validation on self-collected dataset, complete Flavia dataset with 1600 images is used for training the AlexNet through transfer learning and 625 self-collected images are used for testing. The results of the experiment are provided in the Table 5.

It is noted that cross-validation accuracy on self-collected dataset is higher in case of hand-crafted features as it works by segmenting the leaf image and calculating the shape based features. The segmentation depends on the background texture and occlusion and in our self-collected dataset, images are taken without occlusion and on a plain ground or white paper. On the other hand, the CNN trained on Flavia dataset have learned the particular nature of images and performed good but during cross-validation on self-collected dataset it didn't performed good possibly because images are not captured in perfectly lit condition with plain white background.

8. Conclusion

In this study, a thorough review of different plant identification techniques was presented by outlining their pros and cons. A novel plant identification scheme was proposed, which exploits the use of optimized features extracted from the leaf images. The proposed method employing feature normalization, dimensionality reduction and KNN as a classifier was shown to give promising plant recognition results when compared with other existing approaches. The precision and recall measure values on public available 'Flavia' dataset and one of our self-collected image datasets remained above 97% and 96% respectively. The reported results are quite promising, which suggests that our proposed method can offer a highly accurate practical solution to the plant type identification problem from raw leaves when augmented with sophisticated image segmentation techniques. The current data set was collected by placing a white background sheet when taking leave images, in some practical situations, where the leaf image is taken in a cluttered background, the image can first be pre-processed through state-of-the-art segmentation methods to obtain occlusion-free leave before application of our method. We tested CNN for classification to compare hand crafted features vs CNN based features and found that hand crafted features are more robust when a small dataset is available for training. Our approach can be used for mobile application, where user can take image of a leaf to find the class of plant that the leaf belong. The user however need to take precaution that the leaf image is captured without any occlusion and with the non-cluttered background. Our proposed method can be augmented with efficient segmentation techniques to be applied on leaves images with a cluttered background. Our future work is more focused towards enrichment of our self-collected leaves dataset to include more diverse leaves features to address automatic plant disease identification problem.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2018.12.038.

References

Aakif, A., Khan, M.F., 2015. Automatic classification of plants based on their leaves. Biosyst. Eng. 139, 66–75.

Alex, Krizhevsky, Sutskever, Ilya, Hinton, Geoffrey E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates Inc, pp. 1097–1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Arun Priya, C., Balasaravanan, T., Thanamani, A.S., 2012. An efficient leaf recognition algorithm for plant classification using support vector machine. In: Paper presented at the Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on.

Fayyad, U.M., Irani, K.B., 1992. On the handling of continuous-valued attributes in decision tree generation. Mach. Learn. 8 (1), 87–102.

Gwo, C.-Y., Wei, C.-H., 2013. Plant identification through images: Using feature extraction of key points on leaf contours1. Appl. Plant Sci. 1 (11).

Hsiao, J.-K., Kang, L.-W., Chang, C.-L., Lin, C.-Y., 2014. Comparative study of leaf image recognition with a novel learning-based approach. In: Paper presented at the Science and Information Conference (SAI), 2014.

Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Molino, J.-F., 2014. Interactive plant identification based on social image data. Ecol. Inform. 23, 22–34.

Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I., 2013. Leaf classification using shape, color, and texture features. arXiv preprint arXiv:1401.4447.

Kadir, A., 2014. A model of plant identification system using GLCM, lacunarity and shen features. arXiv preprint arXiv:1410.0969.

Krishna, S., Indra, G., Sangeeta, G., 2010. SVM-BDT PNN and Fourier moment technique for classification of leaf shape. Int. J. Signal Process., Image Process. Pattern Recognition 3 (4).

Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J. V., 2012. Leafsnap: A Computer Vision System for Automatic Plant Species Identification. Retrieved from: http://leafsnap.com/dataset/.

Lavania, S., Matey, P.S., 2014. Leaf recognition using contour based edge detection and SIFT algorithm. In: Paper presented at the Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on.

Lewis, D.D., 1998. Naive (Bayes) at forty: The independence assumption in information retrieval. In: Machine Learning: ECML-98. Springer, pp. 4–15.

Liu, N., Kan, J., 2016. Plant leaf identification based on the multi-feature fusion and deep belief networks method. J. Beijing Forestry Univ. 38 (3), 110–119.

Mallah, C., Cope, J., Orwell, J. (2013). Plant leaf classification using probabilistic integration of shape, texture and margin features.

Munisami, T., Ramsurn, M., Kishnah, S., Pudaruth, S., 2015. Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers. Procedia Comput. Sci. 58, 740–747.

Otsu, N., 1975. A threshold selection method from gray-level histograms. Automatica 11 (285–296), 23–27.

Pallavi, P., Devi, V.S.V., 2014. Leaf recognition based on feature extraction and Zernike moments. Int. J. Innovative Res. Comput. Commun. Eng. 2 (2), 67–73.

Prasad, S., Kumar, P.S., Ghosh, D., 2017. An efficient low vision plant leaf shape identification system for smart phones. Multimedia Tools Appl. 76 (5), 6915–6939.

Puja, D., Saraswat, M., Arya, K., 2013. Automatic Agricultural Leaves Recognition System. In: Paper presented at the Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012).

Ragupathy, S., 2016. OAC-Herbarium, Biodiversity Institute of Ontario from University of Guelph. Retrieved from https://doi.org/10.5886/66f3rsta.

Salve, P., Sardesai, M., Manza, R., Yannawar, P., 2016. Identification of the plants based on leaf shape descriptors. In: Paper presented at the Proceedings of the Second International Conference on Computer and Communication Technologies.

Satti, V., Satya, A., Sharma, S., 2013. An automatic leaf recognition system for plant identification using machine vision technology. Int. J. Eng. Sci. Technol. 5 (4), 874.

Shlens, J., 2014. A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

Silva, Pedro F.B., Marcal, A. R. S., da Silva, Rubim M. Almeida, 2013. Evaluation of Features for Leaf Discrimination.

Tsolakidis, D.G., Kosmopoulos, D.I., Papadourakis, G., 2014. Plant leaf recognition using Zernike moments and histogram of oriented gradients. In: Artificial Intelligence: Methods and Applications. Springer, pp. 406–417.

Tyystjärvi, E., Nørremark, M., Mattila, H., Keränen, M., Hakala-Yatkin, M., Ottosen, C.-O., Rosenqvist, E., 2011. Automatic identification of crop and weed species with chlorophyll fluorescence induction curves. Precision Agric. 12 (4), 546–563.

Vapnik, V., 2013. The Nature of Statistical Learning Theory. Springer Science & Business Media.

Weinberger, K.Q., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. 10, 207–244.

Weston, J., Watkins, C., 1999. Support vector machines for multi-class pattern recognition. In: Paper presented at the ESANN.

Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., Xiang, Q.L., 2007. A leaf recognition algorithm for plant classification using PNN. Retrieved from http://flavia.sourceforge.net/.

Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.-X., Chang, Y.-F., Xiang, Q.-L., 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. In: Paper presented at the Signal Processing and Information Technology, 2007 IEEE International Symposium on.

Zeng, Q., Zhu, T., Zhuang, X., Zheng, M., Guo, Y., 2017. Using the periodic wavelet descriptor of plant leaf to identify plant species. Multimedia Tools Appl. 76 (17), 17873–17890.