Automatic Recognition of Medicinal Plants using Machine Learning Techniques

Adams Begue, Venitha Kowlessur
Department of Computer Science and Engineering,
Faculty of Engineering,
University of Mauritius

Fawzi Mahomoodally
Department of Health Sciences,
Faculty of Science,
University of Mauritius

Upasana Singh
School of Management, Information Technology and Governance,
University of KwaZulu-Natal (UKZN),
Durban, South Africa

Sameerchand Pudaruth*
Department of Ocean Engineering & ICT,
Faculty of Ocean Studies,
University of Mauritius

***Abstract*—The proper identification of plant species has major benefits for a wide range of stakeholders ranging from forestry services, botanists, taxonomists, physicians, pharmaceutical laboratories, organisations fighting for endangered species, government and the public at large. Consequently, this has fueled an interest in developing automated systems for the recognition of different plant species. A fully automated method for the recognition of medicinal plants using computer vision and machine learning techniques has been presented. Leaves from 24 different medicinal plant species were collected and photographed using a smartphone in a laboratory setting. A large number of features were extracted from each leaf such as its length, width, perimeter, area, number of vertices, colour, perimeter and area of hull. Several derived features were then computed from these attributes. The best results were obtained from a random forest classifier using a 10-fold cross-validation technique. With an accuracy of 90.1%, the random forest classifier performed better than other machine learning approaches such as the k-nearest neighbour, naïve Bayes, support vector machines and neural networks. These results are very encouraging and future work will be geared towards using a larger dataset and high-performance computing facilities to investigate the performance of deep learning neural networks to identify medicinal plants used in primary health care. To the best of our knowledge, this work is the first of its kind to have created a unique image dataset for medicinal plants that are available on the island of Mauritius. It is anticipated that a web-based or mobile computer system for the automatic recognition of medicinal plants will help the local population to improve their knowledge on medicinal plants, help taxonomists to develop more efficient species identification techniques and will also contribute significantly in the protection of endangered species.**

Keywords*—*leaf recognition; medicinal plants; random forest; Mauritius

I. Introduction

The world bears thousands of plant species, many of which have medicinal values, others are close to extinction, and still others that are harmful to man. Not only are plants an essential resource for human beings, but they form the base of all food chains. To use and protect plant species, it is crucial to study and classify plants correctly. Identifying unknown plants relies much on the inherent knowledge of an expert botanist. The most successful method to identify plants correctly and easily is a manual-based method based on morphological characteristics. Thus many of the processes involved in classifying these plant species is 'dependent on knowledge accumulation and skills of human beings' [1]. However, this process of manual recognition is often laborious and time-consuming. Hence many researchers have conducted studies to support the automatic classification of plants based on their physical characteristics [2][3]. Systems developed so far use varying number of steps to automate the process of automatic classification, though the processes are quite similar. Essentially, these steps involve preparing the leaves collected, undertaking some pre-processing to identify their specific attributes, classification of the leaves, populating the database, training for recognition and finally evaluating the results. Although, leaves are most commonly used for plant identification, the stem, flowers, petals, seeds and even the whole plant can be used in an automated process. An automated plant identification system can be used by non-botanical experts to quickly identify plant species quite effortlessly.

II. Related Works

Several studies have been conducted in order to develop tools for the identification of plants during the last 10 years. One of the most authoritative works in the field of plant classification has been done by Wu *et al*. [2]. From five basic geometric features, twelve morphological features are derived and then Principle Component Analysis (PCA)) is used for dimension reduction so that fewer inputs could be sent to a probabilistic neural network (PNN). They achieved an average accuracy of 90.3% with the Flavia dataset, which is their own creation. Using a different dataset but the same classifier, Hossain and Amin (2010) achieved a similar level of accuracy with similar features [4]. Using similar features but a different dataset with only 20 species, Du *et al*. (2007) attained 93% with the k-nearest neighbour classifier [5]. Using a new distance measure called '*isomap*', Du *et al*. (2009) reached an accuracy of 92.3% on a dataset of 2000 images containing 20 different types of leaves [6].

Herdiyeni and Wahyuni (2012) used a fusion of fuzzy local binary pattern and fuzzy colour histogram and a probabilistic neural network (PNN) classifier on a dataset of 2448 leaf images (270 *240 pixels) obtained from medicinal plants from the Indonesian forests to achieve a classification accuracy of 74.5% [7]. Prasvita and Herdiyeni (2013) developed a corresponding mobile application based on the previous research [8]. Using the kernel descriptor (KDES) as a new feature extraction technique, Le *et al*. (2014) developed a fully automated plant identification system [9]. The proposed technique was tested on a dataset of 55 medicinal plants from Vietnam and a very high accuracy of 98.3% was obtained with a support vector machines (SVM) classifier. Furthermore, their algorithm achieved an accuracy of 98.5% on the Flavia dataset, which is the best result published so far on this dataset [9].

Using the discrete wavelet transform to extract translation invariant features from a collection of 8 different ornamental plants in Indonesia, Arai *et al*. (2013) achieved an accuracy of 95.8% using a support vector machines (SVM) classifier [10]. The size of each image was 256*256 pixels. Du *et al*. (2013) proposed an approach based on fractal dimension features based on leaf shape and vein patterns for the recognition and classification plant leaves [11]. Using a k-nearest neighbour classifier with 20 features, they were able to achieve a high recognition rate of 87.1%. Using a volumetric fractal dimension approach to generate a texture signature for a leaf and the Linear Discriminant Analysis (LDA) algorithm, Backes *et al*. (2009) was able to beat traditional approaches which were based on Gabor filters and Fourier analysis [12].

TABLE. I. SUMMARY OF RELATED WORKS

Reference	**Features**	**Classifier**	**Accuracy (%)**	**Dataset**	**Training**	**Testing**	**Species**
Wu *et al*. (2007)	Shape, Veins	PNN	90.3	1800	1480	320	32
Du *et al*. (2007)	Shape	kNN, MMCH	93, 90	400	200	200	20
Du *et al*. (2009)	Shape	kNN	92.3	2000+	1000	1000+	20
Backes *et al*. (2009)	Texture	LDA	89.6	2000	1200	800	10
Hossain and Amin (2010)	Shape	PNN	91.4	1200	Ten-fold cross-validation		30
Du *et al*. (2013)	Curvature, Veins	kNN	87.1	2422	1695	727	30
Amin and Khan (2013)	Curvature	kNN	71.5	1600	1120	480	100
Herdiyeni and Wahyuni (2012)	Texture, Colour	PNN	74.5	2448	1938	510	51
Arai *et al*. (2013)	Wavelets	SVM	95.8	120	96	24	8
Hernandez-Serna and Jimenez-Segura (2014)	Shape, Texture	ANN	92.9	1800	1620	180	32
Le *et al*. (2014)	Kernel Descriptor	SVM	98.5	1905	1585	320	32
			98.3	1312	649	663	55
Munisami *et al*. (2015)	Shape, Colour	kNN	87.3	640	Leave-one-out cross-validation		32
Chaki *et al*. (2015)	Shape, Texture	NFC	97.6	930	310	620	31
Siravenha and Carvalho (2015)	Shape	ANN	97.5	1865	Ten-fold cross-validation		32
Carranza-Rojas and Mata-Montero (2016)	Curvature, Texture	kNN	87.2	2345	Leave-one-out cross-validation		66

Using a k-nearest neighbour (kNN) classifier, Munisami *et al*. (2015) achieved an accuracy of 87.3% on a dataset of 640 leaves taken from 32 different plant species [13]. They used shape and colour information only. The images were acquired using a smartphone camera with a resolution of 1980*1024.

Hernandez-Serna and Jimenez-Segura (2014) reached an accuracy level of 92.9% using the Flavia dataset [14]. Sixteen inputs (6 geometrical, 8 texture and 2 morphological features) were fed to an artificial neural network (ANN) with 60 nodes in the hidden layer and a learning rate of 0.1 over 50000 generations. Using the same dataset, Chaki *et al*. (2015) achieved an overall accuracy of 97.6% using a Neuro-Fuzzy classifier (NFC) with a 44-element texture vector and a 3-element shape vector [15]. Using shape features only on the Flavia dataset and Pattern Net (a flavour of neural network), Siravenha and Carvalho [16] reached a similar accuracy as Chaki *et al*. [15]. Their feed-forward neural network had two hidden layers with 26 neurons in each and it was trained over 100 epochs.

An interesting work was done by Carranza-Rojas and Mata-Montero (2016) in which they created two datasets: a clean one and a noisy one [17]. They implemented the Histogram of Curvature over Scale (HCoS) algorithm to extract contour information and the local binary pattern variance (LBPV) to extract texture information. In the best case, the clean dataset outperformed the noisy dataset by only 7.3%. This suggest that images taken directly using a smartphone can produce satisfactory levels of accuracy compared with images which are manually processed in a lab and then classified. Earlier, Amin and Khan (2013) have used a distributed hierarchical graph neuron (DHGN) to capture curvature information using 64 feature vectors and the k-nearest neighbour classifier with Canberra distance to obtain an accuracy of 71.5% [18].

Table I summarises some of the works that have been done in the automated recognition of medicinal and non-medicinal plant species during the last decade. Babatunde *et al*. (2015) has done a good survey on different computer vision techniques and machine learning classifiers that have been used in this field during the last ten years [19]. Furthermore, Mata-Montero and Carranza-Rojas (2016) has provided a good introduction to the field and also discussed the challenges and opportunities in this domain [20].

III. METHODOLOGY

A database of medicinal plants which are available on the tropical island of Mauritius was created. Using a Samsung Galaxy J1 Mini smartphone, thirty (30) images of different leaves were taken for twenty-four (24) different plant species. The petiole of each leaf was removed and then placed one by one on a sheet of white paper before being photographed. The size of each image was 1024x600 pixels. The images are stored in the jpeg format.

No manual pre-processing was done on the images in order to enhance them. However, a number of post-processing operations were performed automatically on each image. From the basic attributes (width, length, area, perimeter, area of white space, area of bounding box, area of hull, perimeter of hull), 40 different attributes were derived for each leaf. These values are stored in a csv file. A Java programming environment with the open source Weka machine learning workbench was then used to assess the performance of the system [21].

A. Automatic pre-processing steps

One drawback of taking pictures using a camera, instead of using a scanner, is the presence of shadows on the image. If the shadow is not removed, this will affect all measurement. Thus, to remove the shadow, the image must first be converted to the HSV format and then split into its different colour channels. Only the second channel (saturation) is kept. This has the effect of removing the shadow from the image. To reduce noise in the image, a median blur filter with a window size of 25 is applied to the resulting image.

The next step is to perform a thresholding operation which will convert the image into a binary image with only two values: black and white pixels. This is achieved using the Otsu thresholding method. An opening operation is then performed on the images. This is an erosion operation followed by a dilation. Erosion has the effect of reducing the size of foreground (white) pixels while dilation enlarges them. This operation is important in order to clear the image from many small noisy pixels, which are the artefacts of the thresholding operation.

Fig. 1. Original Giant Bramble Leaf

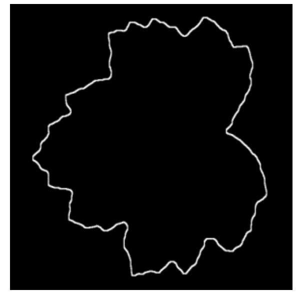

Fig. 2. Binary Image of Leaf

B. Feature extraction

A number of base features were extracted from the images in Figures 1 and 2. These are: length, width, area of the bounding box, area of leaf, perimeter of leaf, hull area, hull perimeter, number of vertices, horizontal & vertical distance maps, 45^{o} radial map and the original RGB values of each pixel.

Fig. 3. Bounding Box & Contour

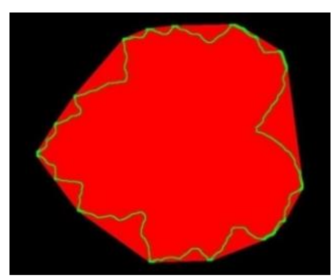

Fig. 4. Area & Perimeter of Hull

Figure 3 shows the bounding box (in red) and the contour line (in green) around a giant bramble leaf. Using the bounding box, the length and the width of the leaf can easily be computed. The perimeter of the leaf is obtained from the contour line. The area of the leaf corresponds to the white space inside the green contour line. Figure 4 shows the convex hull which can be used to compute the hull perimeter and the hull area. The hull is the smallest polygon that can contain the leaf. The convex hull is also used to calculate the number of vertices in the leaf. Although the algorithm which is used to calculate the number of vertices is not very accurate, it was still a good differentiator. This is mainly because it is a raw attribute which is independent of the size of the leaf.

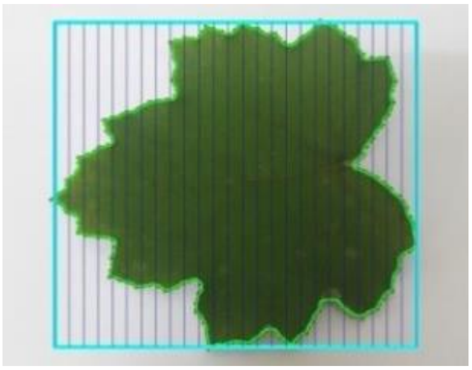

Fig. 5. Vertical Distance Map

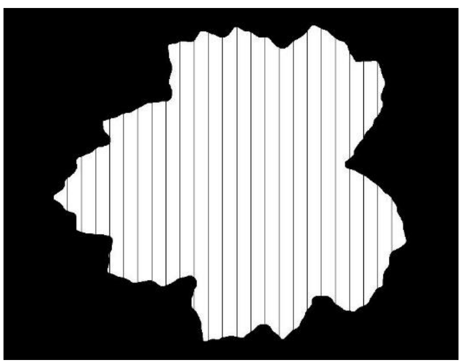

Fig. 6. Vertical Distance Map

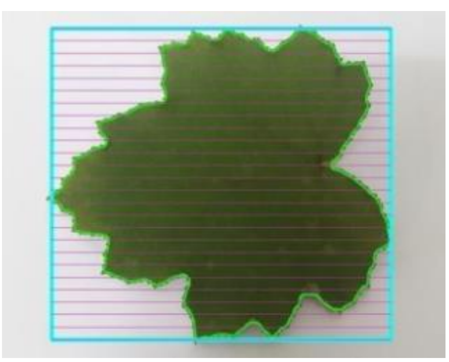

Fig. 7. Horizontal Distance Map

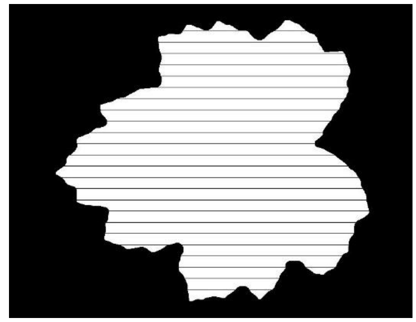

Fig. 8. Horizontal Distance Map

Figure 5 shows the vertical distance maps in which the image is divided into 24 equal strips. The aim is to find where each vertical line intercepts the contour line of the leaf. The distances between the intercepts are then computed. This is shown in Figure 6. Figure 7 shows the horizontal distance maps in which the image is divided into 24 equal strips. Again, the objective of this procedure is to locate where each horizontal line touches the boundary of the leaf, as displayed in Figure 8. To avoid overfitting, only 12 alternate values are used for both directions. Similarly, the radial distances are computed as shown below in Figure 9 and Figure 10.

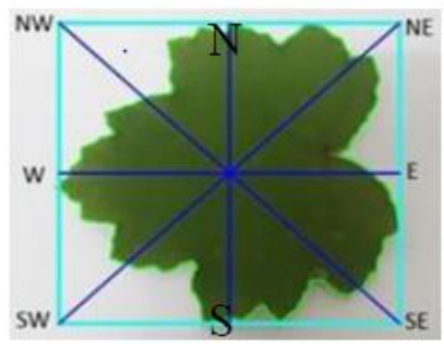

Fig. 9. Radial Map

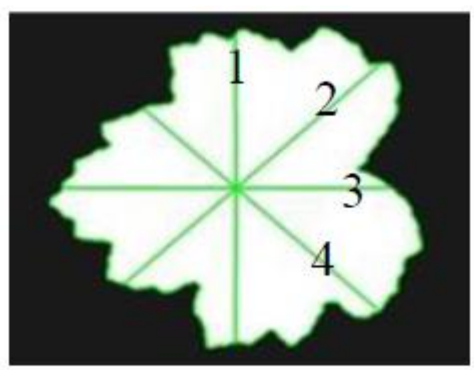

Fig. 10. Radial Map

C. Derived features

Using the base features which are extracted directly from the image, a number of derived features are calculated [22]. Ratios are more suitable for comparison as they are independent of the actual size of the image in pixels. The ratios are shown below in Table II. The different plant species studied are summarized in Table V.

TABLE. II. DERIVED FEATURES

#	Derived Feature	Method
1	aspect ratio	width/length
2	circularity/roundness	area/(perimeter*perimeter)
3	solidity	area/hull area
4	convexity	hull perimeter/perimeter
5	rectangularity	width*length/area
6	hydraulic radius	area/perimeter
7	lobidity	perimeter/(width+length)
8	white area ratio	(area of bounding box–area)/area of bounding box
9	hull ratio	hull area/hull perimeter
10	red to green ratio	red/green
11	red to blue ratio	red/blue
12	blue to green ratio	blue/green
13	N_S ratio	1/N_S (refer to Figure 9 & 10)
14	NE_SW ratio	2/NE_SW (refer to Figure 9 & 10)
15	E_W ratio	3/E_W (refer to Figure 9 & 10)
16	SE_NW ratio	4/SE_NW (refer to Figure 9 & 10)
17	vertical distance ratios (12)	length of line in leaf/width
18	horizontal distance ratios (12)	length of line in leaf/length

IV. RESULTS AND DISCUSSION

Medicinal plants have received much attention since they are generally perceived as safe and accessible for human utilization. However, the proper identification of plant species has major benefits for a wide range of stakeholders ranging from consumers, forestry services, botanists, taxonomists, physicians, pharmaceutical laboratories, organization fighting for endangered species, government, and the public at large.

Five different machine learning classifiers were used to assess the recognition rate. The results are shown in Table III. The Random Forest classifier achieves the best performance with an accuracy of 90.1%, i.e., out of 720 leaves, 649 leaves were classified correctly while 71 were not.

The Multilayer Perceptron produced the second best accuracy at 88.2%. However, due to resource constraints, the potential of neural networks has not been fully exploited and it is still possible to achieve even higher accuracy with this classifier. The k-Nearest Neighbour (kNN) classifier had the lowest accuracy. A 10-fold cross-validation technique was used in all the experiments. The main parameters of each classifier was varied to find the ones producing the highest accuracy.

TABLE. III. PERFORMANCE OF MACHINE LEARNING CLASSIFIERS

#	Classifier	Accuracy (%)
1	Random Forest (numTrees=100)	90.1
2	Multilayer Perceptron Neural Network (Epochs=500)	88.2
3	Support Vector Machine (PolyKernel and c=4.0)	87.4
4	Naïve Bayes	84.3
5	k-Nearest Neighbour (k=1)	82.5

Figure 11 shows the confusion matrix obtained when using the Random Forest classifier with 100 trees and 6 attributes in each iteration. The value 25 (first number in the second row) in the matrix indicates that 25 Antidesma leaves were correctly classified. Out of the five remaining leaves, one was incorrectly classified as an Avocado leaf, one as a Fandamane leaf, another as a Jackfruit, and last two as Guava. From the first column, it can be seen that two Bigaignon Rouge and one Pomegranate leaves were misclassified as Antidesma. The high values in the diagonal line indicate that the recognition was very successful. Another important observation is that six Coriander leaves were misclassified as Bitter Gourd and the only two Bitter Gourd leaves that were not correctly carried were predicted as Coriander leaves. These observations could be explained by the fact that both these plants have highly lobed leaves.

```
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q  r  s  t  u  v  w  x   <-- classified as
 25  1  0  0  0  0  0  0  0  0  0  0  0  1  1  0  0  0  0  0  0  0  2  0 |  a = Antidesma
  0 28  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  1  0  0  0 |  b = Avocado
  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |  c = Ayapana
  0  0  0 29  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0 |  d = Balloon Plant
  2  0  0  0 24  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  3  0  0  0 |  e = Bigaignon Rouge
  0  0  0  0  0 28  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0 |  f = Bitter Gourd
  0  0  1  0  0  0 25  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  3  0 |  g = Bois Carrote
  0  0  0  0  0  0  0 27  0  0  0  0  0  0  0  0  0  0  3  0  0  0  0  0 |  h = Bois Cerf
  0  0  0  0  0  0  0  0 26  0  0  0  0  3  0  1  0  0  0  0  0  0  0  0 |  i = Bois de rat
  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |  j = Bramble
  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0 |  k = Chinese Okra
  0  0  0  0  0  6  0  0  0  1  0 21  2  0  0  0  0  0  0  0  0  0  0  0 |  l = Coriander
  0  0  0  0  0  0  0  0  0  0  0  0 26  1  0  0  1  0  0  0  0  0  1  1 |  m = Curry Leaf
  0  0  0  0  0  0  0  0  0  0  0  0  0 27  0  0  1  0  0  0  0  2  0  0 |  n = Fandamane
  0  0  0  0  0  0  2  0  0  0  0  0  0  0 23  0  0  0  0  0  0  2  3  0 |  o = Jackfruit
  0  0  0  0  0  0  1  1  3  0  0  0  0  0  0 24  0  1  0  0  0  0  0  0 |  p = Mango
  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0 28  0  0  0  1  0  0  0 |  q = Moringa
  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  1  0 27  0  0  0  0  0  0 |  r = Neem
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0 |  s = Orange Climber
  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 29  0  0  0  0 |  t = Parsley
  0  0  0  0  1  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0 28  0  0  0 |  u = Peppermint
  1  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0 27  0  0 |  v = Pomegranate
  0  0  0  0  0  0  1  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0 28  0 |  w = Strawberry Guava
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0 29 |  x = Tulsi
```

Fig. 11. Confusion Matrix for Random Forest with 100 Trees

Besides the overall accuracy, the performance of the automated system was also assessed on a class-wise basis. Recall is the proportion of leaves, for each class, that was correctly picked out from the entire set. Precision is the proportion of correctly identified leaves out of all the leaves that are predicted to be of a specific plant while F-measure can simply be considered as the average of these two values. Table IV shows that Ayapana, Bramble, Chinese Okra and Orange Climber has a perfect recall of 100% while Coriander has the lowest recall of 70%. Only the Chinese Okra and Parsley has a precision value of 100% while the Strawberry Guava has the lowest precision at 76%. Thus, Table IV provides us with much useful information which can be used to both gauge the strengths of the system and address its weaknesses as well. Plants which have low recall and low precision must be relooked into. For example, new features must be designed and extracted that bring out the uniqueness in such leaves and are determinative of their species.

TABLE. IV. PERFORMANCE ASSESSMENT BY SPECIES USING A RANDOM FOREST WITH 100 TREES

#	Species	Precision	Recall	F-Measure
1	Antiderma	0.89	0.83	0.86
2	Avocado	0.97	0.93	0.95
3	Ayapana	0.97	1.00	0.98
4	Balloon Plant	0.97	0.97	0.97
5	Bigaignon Rouge	0.96	0.80	0.87
6	Bitter Gourd	0.82	0.93	0.87
7	Bois Carotte	0.81	0.83	0.82
8	Bois Cerf	0.90	0.90	0.90
9	Bois de rat	0.90	0.87	0.88
10	Bramble	0.97	1.00	0.98
11	Chinese Okra	1.00	1.00	1.00
12	Coriander	0.91	0.70	0.79
13	Curry Leaf	0.87	0.87	0.87
14	Fandamane	0.84	0.90	0.87
15	Jackfruit	0.89	0.77	0.82
16	Mango	0.89	0.80	0.84
17	Moringa	0.93	0.93	0.93
18	Neem	0.93	0.90	0.92
19	Orange Climber	0.88	1.00	0.94
20	Parsley	1.00	0.97	0.98
21	Peppermint	0.85	0.93	0.89
22	Pomegranate	0.84	0.90	0.87
23	Strawberry Guava	0.76	0.93	0.84
24	Tulsi	0.97	0.97	0.97
Average		0.91	0.90	0.90

Besides investigating the effect of different classifiers, the effect of the number of plant species in the dataset on the overall accuracy of the system was also studied. As expected, with only eight features, a very high accuracy of 97.9% is obtained, as shown in Figure 12. Next, the number of plant species is doubled to 16 but the accuracy decreases by only 3.1%. After an additional set of eight new types of leaves are added, the accuracy drops by 4.7%. Munisami *et al*. (2015) reported very similar results but with a different dataset containing 32 plants [13].

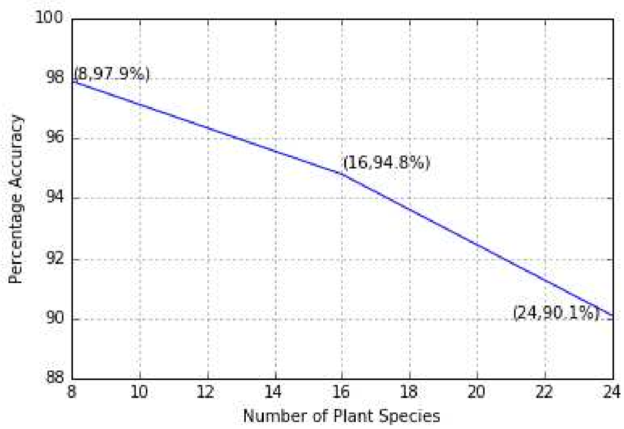

Fig. 12. Number of Plant Species v/s Percentage Accuracy

As shown in Figure 13, increasing the number of leaves per plant species has a positive impact on the classification accuracy. The peak performance is achieved when using 25 leaves per plant. There is no improvement in accuracy beyond this threshold. This is an important result which can be used by researchers and scientists to decide on the number of samples that they must collect in their studies. Munisami *et al*. (2015) performed a similar assessment [13]. However, they collected only 20 samples per species and therefore they could not arrive at this threshold value [13].

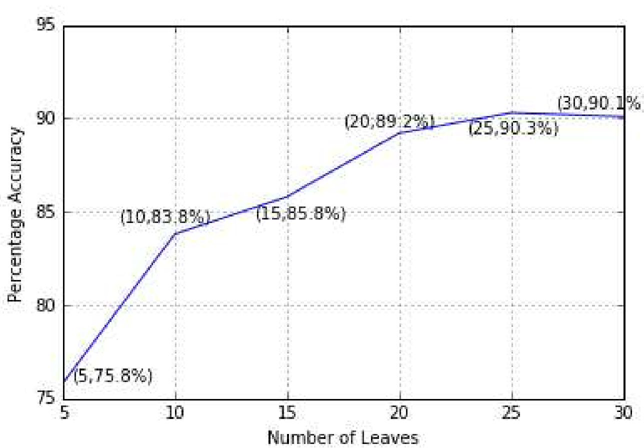

Fig. 13. No. of Leaves per Plant v/s Percentage Accuracy

Using the chi-squared ($\chi 2$) statistical test in Weka, the k best features from the dataset were selected. The results are shown in Figure 14. The eight (k=8) best features were: white area ratio (first position), rectangularity, number of dents, hull ratio, hydraulic radius, aspect ratio, lobidity, and solidity at the eight position. An accuracy of 77.1% was obtained using these 8 features. The next best features were: convexity (ninth position), NE_SW ratio, SE_NW ratio, N_S ratio, E_W ratio, circularity, blue to green ratio and the red to green ratio (sixteenth position). Using the 16 best features lead to a significant boost in accuracy by 12.1%. The accuracy was only minimally better with the 24 best features and using more than 32 features did not bring any improvement in the accuracy.

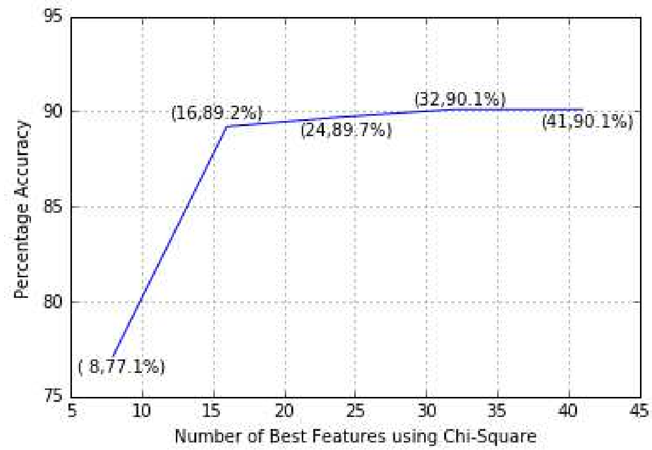

Fig. 14. Number of Best Features v/s Percentage Accuracy

Although literature on leaf-based recognition of plant species using image processing and data mining techniques are abundant, only a handful of researchers have applied these techniques on medicinal plants. Countries like China, India, Indonesia, Malaysia, and Vietnam have vast repositories of medicinal plants and therefore it is no wonder that many of the research works on medicinal plants come from these countries [23]. The novelty of this work resides in the creation of a unique image dataset for medicinal plants that are available on the island of Mauritius.

A new measure called lobidity has also been proposed. This is the seventh best differentiating feature according to the chi-squared test. Although the majority of researchers have been able to achieve accuracies above 90%, improvements are still possible. The work of Mata-Montero and Carranza-Rojas (2016) outlines the different approaches that have been used for plant classification, including morphometrics (curvature, texture and venation), DNA barcoding, and crowd sourcing [20]. More importantly, they outlined the challenges of collecting, classifying and sharing huge datasets and discuss the opportunities that crowd sourcing and deep learning offer to this community of researchers.

V. Conclusion

A new dataset on medicinal plants of Mauritius has been made publicly available on the machine learning repository portal. In this paper, computer vision techniques have been used to extract several shape-based features from the leaves of medicinal plants. Machine learning algorithms were then used to classify the leaves from 24 different plant species into their appropriate categories. The highest accuracy of 90.1% was obtained from the random forest classifier. This excellent performance indicates the viability of such computer-aided approaches in the classification of biological specimens and its potential applicability in combatting the '*taxonomic crisis*'. A web-based or mobile computer system for the automatic recognition of medicinal plants will help the local population to improve their knowledge on medicinal plants, help taxonomists to develop more efficient species identification techniques and will also contribute significantly in the protection of endangered species. For future research, in an attempt to achieve even higher accuracies, probabilistic neural networks and deep learning neural networks would be investigated.

REFERENCES

[1] Gao, W. and Lin, W., 2012. Frontal Parietal Control Network Regulates the Anti-Correlated Default and Dorsal Attention Networks. Human Brain Mapping, 33(1), 192–202.

[2] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2007. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11-16.

[3] Zhang X., Liu Y., Lin H., Liu Y. (2016) Research on SVM Plant Leaf Identification Method Based on CSA. In: Che W. et al. (eds) Social Computing. ICYCSEE 2016. Communications in Computer and Information Science, Vol 624, Springer, Singapore.

[4] Hossain, J. and Amin, M.A., 2010. Leaf Shape Identification Based Plant Biometrics. 13th International Conference on Computer and Information Technology, Dhaka, Bangladesh, 458-463.

[5] Du, J.X., Wang, X.F. and Zhang, G.J., 2007. Leaf shape based plant species recognition. Applied Mathematics and Computation, 185, 883-893.

[6] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. 5th International Conference on Emerging Intelligent Computing Technology and Applications, Ulsan, South Korea, 627-634.

[7] Herdiyeni, Y. and Wahyuni, N.K.S., 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. International Conference on Advanced Computer Science and Information Systems (ICACSIS), West Java, Indonesia, 301-306.

[8] Prasvita, D.S. and Herdiyeni, Y., 2013. MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. International Journal of Advanced Science, Engineering and Information Technology, 3, 5–8.

[9] Le, T.L., Tran, D.T. and Hoang, V.N., 2014. Fully Automatic leaf-based plant identification, application for Vietnamese medicinal plant search. Fifth Symposium on Information and Communication Technology, Hanoi, Vietnam, 146-154.

[10] Arai, K., Abdullah, I.N. and Okumura, H., 2013. Identification of Ornamental Plant Functioned as Medicinal Plant Based on Redundant Discrete Wavelet Transformation. International Journalof Advanced Research in Artificial Intelligence, 2(3), 60-64.

[11] Du, J.X., Zhai, C.M. and Wang, Q.P., 2013. Recognition of plant leaf image based on fractal dimension features. Neurocomputing, 116, 150-156.

[12] Backes, A.R., Casanova, D. and Bruno, O.M., 2009. Plant Leaf Identification based on Volumetric Fractal Dimension. International Journal of Pattern Recognition and Artificial Intelligence, 23(6), 145-1160.

[13] Munisami, T., Ramsurn, M., Kishnah, S. and Pudaruth, S., 2015. Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers. Procedia Computer Science, 58, 740-747.

[14] Hernandez-Serna, A. and Jimenez-Segura, L.F., 2014. Automatic Identification of species with neural networks. PeerJ 2:e563; doi:10.7717/peerj.563.

[15] Chaki, J., Parekh, R. and Bhattacharya, S, 2015. Plant leaf recognition using texture and shape features with neural classifiers. Pattern Recognition Letters, 58, 61-68.

[16] Siravenha, A.C.Q. and Carvalho, S.R., 2015. Exploring the use of Leaf Shape Frequencies for Plant Classification. 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, Brazil, 297-304.

[17] Carranza-Rojas, J. and Mata-Montero, E., 2016. Combining Leaf Shape and Texture for Costa Rican Plant Species Identification. CLEI Electronic Journal, 19(1), Paper 7.

[18] Amin, A.H.M. and Khan, A.I., 2013. One-Shot Classification of 2-D Leaf Shapes using Distributed Hierarchical Graph Neuron (DHGN) Scheme with k-NN Classifier. Procedia Computer Science, 24, 84-96.

[19] Babatunde, A., Armstrong, L., Diepeveen, D. and Leng, J., 2015. A survey of computer-based vision systems for automatic identification of plant species. Journal of Agricultural Informatics, 6(1), 61-71.

[20] Mata-Montero, E. and Carranza-Rojas, J., 2016. Automated Plant Species Identification: Challenges and Opportunities. IFIP Advances in Information and Communication Technology, 481, 26-36.

[21] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H., 2009. The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11(1).

[22] Russ, J.C., 2011. The Image Processing Handbook. CRCPress, USA.

[23] FAO, 2016. Food and Agriculture Organisation of the United Nations (online). Available from: http://www.fao.org/docrep/005/AA010E/AA010e02.htm [Accessed 30th December 2016].

TABLE. V. PLANT SPECIES

Image	
Scientific name	*Antidesma madagascariense*
English name	Antidesma
Mauritian name	Bigaignon batard
Medicinal uses	Effective against dysentery and skin infections
Image	
Scientific name	*Persea americana*
English name	Avocado
Mauritian name	Avocat
Medicinal uses	Reduce the risk of eye cataract; Control cholesterol level
Image	
Scientific name	*Ayapana triplinervis*
English name	Ayapana
Mauritian name	Ayapana
Medicinal uses	Stomachic; anti-ulcerous; Fights cough
Image	
Scientific name	*Cardiospermum halicacabum*
English name	Balloon plant

Mauritian name	Pocpoc
Medicinal uses	Management of Type II diabetes mellitus; Healing skin infections
Image	
Scientific name	*Psiloxylon mauritianum*
English name	Bigaignon wood/Red bigaignon
Mauritian name	Bois bigaignon/Bigaignon rouge
Medicinal uses	Treatment and management of amenorrhea, dysentery and Type II diabetes mellitus
Image	
Scientific name	*Momordica charantia*
English name	Bitter gourd
Mauritian name	Margose
Medicinal uses	Management of Type II diabetes mellitus
Image	
Scientific name	*Pittosporum senacia*
English name	Not available
Mauritian name	Bois carotte
Medicinal use	Treatment of stomach disorders
Image	
Scientific name	*Olea lancea*
English name	Not available
Mauritian name	Bois cerf
Medicinal uses	Effective against cough and hypertension
Image	
Scientific name	*Coptosperma borbonicum*
English name	Not available
Mauritian name	Bois de rat
Medicinal uses	Treatment of fever
Image	
Scientific name	*Rubus alceifolius*
English name	Giant bramble/Wild raspberry
Mauritian name	Framboisier/Piquant loulou

Medicinal uses	Management of Type II diabetes mellitus
Image	
Scientific name	*Luffa acutangula*
English name	Chinese okra/Silky gourd
Mauritian name	Pipengaille
Medicinal uses	Prevention of cardiovascular disease; Effective against hypertension
Image	
Scientific name	*Coriandrum sativum*
English name	Coriander
Mauritian name	Cotomili/Coriandre
Medicinal uses	Management of Type II diabetes mellitus
Image	
Scientific name	*Murraya koenigii*
English name	Curry tree
Mauritian name	Cari poulet
Medicinal uses	Management of hypertension
Image	
Scientific name	*Aphloia theiformis*
English name	Fandamane
Mauritian name	Fandamane
Medicinal uses	Management of Type II diabetes mellitus; Prevention of eye cataract
Image	
Scientific name	*Artocarpus heteophyllus*
English name	Jackfruit
Mauritian name	Jack
Medicinal uses	Management of Type II diabetes mellitus
Image	
Scientific name	*Magnifera indica*
English name	Mango
Mauritian name	Mangue
Medicinal uses	Management of Type II diabetes mellitus

Image	
Scientific name	*Moringa oleifera*
English name	Moringa
Mauritian name	Brede mouroum
Medicinal uses	Management of Type II diabetes mellitus; Control cholesterol level
Image	
Scientific name	*Azadirachta indica*
English name	Neem
Mauritian name	Neem
Medicinal uses	Management of Type II diabetes mellitus
Image	
Scientific name	*Toddalia asiatica*
English name	Orange climber
Mauritian name	Patte poule
Medicinal uses	Treatment of asthma
Image	
Scientific name	*Petroselinum crispum*
English name	Parsley
Mauritian name	Persil
Medicinal uses	Management of Type II diabetes mellitus; Control cholesterol level; Treatment of kidney diseases

Image	
Scientific name	*Mentha x piperita*
English name	Peppermint
Mauritian name	Menthe
Medicinal uses	Effective against indigestion
Image	
Scientific name	*Punica granatum*
English name	Pomegranate
Mauritian name	Grenade
Medicinal uses	Prevention of cardiovascular disease; Control cholesterol level; Effective against diarrhea
Image	
Scientific name	*Psidium cattleyanum*
English name	Guava
Mauritian name	Goyave de chine
Medicinal uses	Management of Type II diabetes mellitus; Control cholesterol level
Image	
Scientific name	*Ocimum tenuiflorum*
English name	Holy basil/Tulsi
Mauritian name	Tulsi
Medicinal uses	Management of Type II diabetes mellitus; Control cholesterol level