Indian Journal of Traditional Knowledge
Vol 19(4), October 2020, pp 813-820

Enrich Ayurveda knowledge using machine learning techniques

S Roopashree[a,b,*,+] & J Anitha[c,$]

[a]Dayananda Sagar University , Bengaluru, India
[b]Visvesvaraya Technological University, Belagavi, Karnataka, India
[c]RV Institute of Technology and Management, Bengaluru, India
E-mail: [+]roopashaily@gmail.com, [$]anitha.jayapalan@gmail.com

Received 29 July 2019; revised 17 October 2019

In India, every region, urban or rural the whole population is dependent on plants for life sustenance in the form of food, shelter, clothes and medicines. Due to inflation, synthetic medicines have become less affordable and their side effect has led in seeking alternative medication system. Indian medicinal herbs and its uses are good alternates for curing many common ailments and diseases. Using computer vision and machine learning techniques, the Indian medicinal herbs can be classified based on their leaves and thus promote the Indian traditional system – Ayurveda to a great extent. In this paper, a systematic approach consisting of Scale Invariant Feature Transform (SIFT) which is uniform in nature to scale, illumination and rotation is combined with different classifiers. Different models are built using SIFT as the common feature extractor in combination with Support Vector Machine (SVM), K-Nearest Neighbor (kNN) and Naive Bayes Classifier. Finally, the proposed method consists of SIFT features with dimension reduction using Bag of Visual Words and classified by SVM. The work is carried over in comparison with newly built herb dataset and Flavia dataset. The model shows an accuracy of 94% with newly built dataset which consists of six Indian medicinal herbs.

Keywords: BoVW, Indian medicinal herbs, Machine learning, SIFT, SVM, Traditional medicine

IPC Code: Int. Cl.[20]: A61K 36/9066, G06N 20/20, A61K 36/00

Ayurveda, is an Indian traditional medicine from ancient times. Presently, huge side-effects and inflation of English medicine for common ailments and life-threatening illness has given wider scope for showing interest in Ayurveda studies. Ayurveda aims in promotion of good health and enhancing the quality of life by treating diseases with therapeutic approaches. Ayurvedic herbs have progressively gained public attention as they are the main source for traditional medicine. But, today a lot of these herbs are endangered and neglected as people are unaware of its existence and usage. Even today, the traditional herbs for Ayurvedic medicine are identified manually, where the process is time consuming and requires years of rigorous training for accurate identification. Automatic classification of Indian traditional herbs using the techniques of computer vision and machine learning can offer complementary assistance to accuracy of manual herb identification and can also reduce time and effort.

Currently, there are several websites, books and CD's which provide the details of Indian traditional herbs. But many fail to provide the complete information on numerous species. They offer diverse details with similar information. A great number of researches on identification and classification on general images under computer vision techniques have established outstanding results. In that perspective, the system of robust recognition and classification of Indian medicinal herbs that is put forward consists of using feature extraction of the input leaf image to classify to its respective family and provide the essential herb information useful to researchers, botanists, drug designers and many more.

Plants are recognized using its characteristics such as flowers, stem, fruits and leaves. Identifying plants through its leaves is more dominant. The local features of leaf such as margin, shape, texture, pattern, vein and many more are used for manual classification and can be used for automatic classification process. Automatic identification consists of few main features and their descriptors as shown in the Table 1.

Shape is one of the main features in plant classification. Scale Invariant Feature Transform (SIFT) descriptors has stability with following

*Corresponding author

Table 1 — Categorization and summary of important dominant feature descriptors in Plant Classification

Features	General	Shape	Contour	1. Simple & morphological (e.g. perimeter, compactness) 2. Shape signature 3. Shape context 4. Curvature scale space 5. Fourier 6. Fractal Dimension
			Region	1. Simple & morphological (e.g. convex hull, area etc.) 2. Image moments 3. Local features (SIFT, HOG, DSIFT, LBP, SURF)
		Color		1. Color moments 2. Color histogram 3. CSIFT
		Texture	Spatial	1. Co-occurance matrix 2. Fractal dimension 3. Local binary pattern 4. SURF
			Spectral	1. Gabor filter 2. Fourier 3. Wavelet 4. Curvelet
	Organ Specific	Leaf	Vein	1. Morphological 2. Graph representation 3. Fractal dimension 4. Shape context 5. SIFT
			Margin	1. Morphological 2. Curvature & scale space 3. Shape context

characteristics such as invariant to geometric transformation, rotation, scaling and translation mainly preferred for accuracy[1]. It is shown in Table 1 that SIFT descriptors can be used to extract both shape and vein features of leaf. It is shown that vein architecture of leaf consists of extracting vein median - volume, orientation, width and number of areole features in the work implemented on dataset of two species with an accuracy of 95%[2]. To accomplish better accuracy on larger dataset such as Flavia dataset using only Vein feature extraction involves segmentation, is expensive in respect to time, includes repeated threshold operations and is prone to errors[3] and for detecting the edges involves exploiting numerous threshold parameters[4].

Literature Review

In contrast to botanist categorization, computer vision techniques for image classification uses texture, shape, color etc. as features to identify the plants. Gray Level Co-occurrence Matrix (GLCM) method to pull-out the texture features of 9 Indian herb varieties namely Neem (*Azadirachta indica*), Thulasi (*Ocimum tenuiflorum*), Hibiscus (*Hibiscus rosa-sinensis*), Omavalli (*Plectranthus amboinicus*), Henna (*Lawsonia inermis*), Thudhuvalai *(Solanum trilobatum*), Vana-thulasi (*Ocimum gratissimum*), Curry leaves (*Murraya koenigii*) and Nochi (*Vitex negundo*) with an accuracy of 94%[5].

A qualified study on automatic identification of leaves on Flavia dataset using shape, color, texture and vein features on SVM classifier for recognition is shown in the Table 2. This proves that the classification accuracy is satisfactory by extracting the vein and other features on Flavia dataset and can be improved.

Herbal leaf analysis by extracting shape, color and venation features on five different leaf samples of Neem alone. The classifiers such as Probability

Table 2 — Comparison of Classification Accuracy on Flavia Dataset[6-8]

Leaf features extracted	Classifier	Accuracy	Reference
Shape, color, texture and vein	SVM	87%	Ghasab *et al*. 2015
Shape and vein	SVM	94%	Priya *et al*. 2012
Vein		95%	Chakkaravarthy *et al*. 2016

Neural Network (PNN), SVM and Principal Component Analysis (PCA) used for categorizing the input[9]. Color and shape features such as aspect ratio, compactness, dispersion, centroid and eccentricity on ten different medicinal herbs and calculated least dissimilarity technique for classification with an accuracy of 92%[10]. Canny edge detector, morphological and neural network algorithm on four Indian medicinal herbs with classification accuracy of Hibiscus leaf - 70.87%, Castor - 64.78%, Betel - 65.5% and Manathakali - 68.57%[11]. SIFT feature extraction and SVM classifier on five different Indian herbs such as Circinatum, Garryana, Glabrum, Kelloggii and Macrophyllum with an accuracy of 92%[12].

A comparative study on Malaysian herbal plants by extracting the features such as Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG) and Speeded-up robust features (SURF) separately and distinguished using SVM (multi-class) with an accuracy of 99% for HOG and LBP where as 74% accuracy for SURF features on the newly constructed dataset of 50 samples from each of ten different Malaysian herbs[13]. Contribution on Indian medicinal herbs such as peepal, betel, castor oil leaf, bilva and hibiscus has been made by implementing Artificial Neural Network (ANN) on the herbs[14]. Ayurvedic leaf classification has been done by extracting the morphological leaf features and using leaf factor for classification as well as showcased the medicinal properties of the identified leaves[15].

Classification of herbs based on leaf consists of various challenges in choosing novel combination of feature extraction and classification algorithms. Classification accuracy of 90% on general digital online images by using the SIFT feature extraction technique, BoVW and SVM[16].

In this paper, four models are proposed which include data acquisition of Indian herbs by using a digital camera and leaves from Flavia Dataset followed by SIFT feature extraction of the input leaf image. The models comprise of classifying the extracted SIFT features by using different classifiers such as BoVW and SVM, kNN, SVM (Linear Kernel) and Naive Bayes.

The manuscript is divided as follows: In the next section (Methodology) we discuss on building the newly-built dataset, the feature extraction technique and classification process of four different models. In results and discussion, we present the comparison study of all the models and herb classification discussing the significances so obtained. Conclusion section offers the efficacy of one of the proposed models and the future enhancements to improvise the methodology further.

Methodology

Construction of new herb dataset

Non-availability of online digital images of Indian medicinal herbs motivated for a new dataset to be built with few locally available medicinal herbs to check on the novel machine learning model consisting of extracting SIFT feature with different classification techniques (kNN, SVM and Naive Bayes). The dataset is built as per the steps below:

1. Eight Indian medicinal herbs namely Malabar Spinach (*Basella alba*), Amarnath (*Amaranthus*), Mint (Mentha), Betel (Piper betle), Neem (*Azardirachta indica*), Curry (*Murraya koenigii*), Tulsi (*Ocimum tenuiflorum*) and Hibiscus (*Hibiscus rosa - sinensis*) are considered.
2. The petiole of the leaf is removed before capturing on a bright sunny day through DSLR camera over a white background.
3. Thirty images per species are captured for training the classifiers. Altogether 240 images were used for training each classifier.
4. The images are preprocessed / cleaned for any noise.

The background of every image is removed and placed on 1600x1200 white canvas to maintain all images of same size.

Feature extraction technique

Image processing is a part of signal processing which works on domain of images. In Image processing, numerous operations are performed on the images to acquire enhanced images. From these enhanced images, valuable and non-redundant information also known as features can be extracted. Features are "interesting points" on an image used for image analysis in machine learning and computer vision techniques.

The matching of images is a prime task in the domain of computer vision. To extract features for image matching with different scales and rotation, Scale Invariant Feature Transform (SIFT) is very useful. SIFT is invariant to scale, rotation and illumination of images[17]. Generally, SIFT feature algorithm consists of two important steps: The detection of key points[18] and extraction of a descriptor associated to each key point.

To summarize, the SIFT is outlined as:

1. Construction of scale space: The scale space is created by blurring the original images. Size is reduced for further blur. Ideal is to create four octaves with five blur levels. Gaussian blur is applied to each pixel as shown in (1).

$$L(p, \mathrm{q}, \sigma) = G(p, \mathrm{q}, \sigma) * I(p, \mathrm{q}) \quad \dots (1)$$

Where,

$$\mathrm{G}(\mathrm{p}, \mathrm{q}, \sigma) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-(\mathrm{p}+\mathrm{q}^2)/2\sigma^2} \quad \dots (2)$$

Here in (1) and (2),

- L is a blurred image
- G is the Gaussian Blur operator
- I is an image
- p, q are the location coordinates
- σ is the "scale" variable indicates the amount of blur. Greater the value, greater the blur

2. Laplacian of Gaussian (LoG) approximation: The blurred images from the previous stage are used to generate another set of images called the Difference of Gaussians (DoG) for finding the key points. Here, the difference between the two consecutive scale spaces (Difference of Gaussians) is calculated as shown in the Figure 1. These invariants are scale invariants.

   The convolution operation on x and y is applied by the operator ‘*’. The gaussian blur G is applied onto image I.

3. Key points Detection:It is a two-step process 1. Coarsely locating minima and maxima from previous DoG images and 2. Finding subpixel minima and maxima. Locating the minima and maxima consists of iterations through every pixel and checking on all its neighbors in the current image and also the image above and below it. Subpixel maxima and minima is calculated around the approximated key points by Taylor expansion of the image. Increasing the chances of matching and algorithm stability is given by these calculated subpixel key point values.

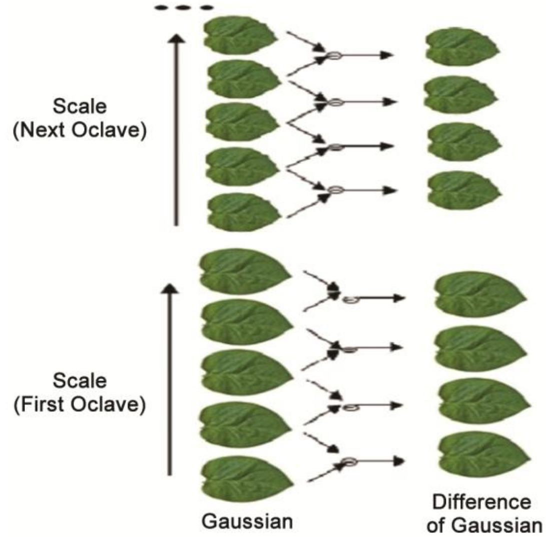

Fig. 1 — Difference of Gaussian Images

4. Removal of unwanted key points: Discarding those key points which are lying at the edges and are of low contrast. As these points are of no interest in matching the images.

5. Orientation assignment to key points: At this step we have the valid key points of the image. Now, to achieve rotation invariance, orientation to be applied to each key point derived from the previous step. Orientation parameter can be achieved by calculating the magnitude and gradient directions for all the pixels around each key point using (3) and (4). Then, histogram is created.

$$\mathrm{m}(\mathrm{p}, \mathrm{q}) = \sqrt{(\mathrm{L}(\mathrm{p}+1, \mathrm{q}) - \mathrm{L}(\mathrm{p}-1, \mathrm{q}))^2 + (\mathrm{L}(\mathrm{p}, \mathrm{q}+1) - \mathrm{L}(\mathrm{p}, \mathrm{q}-1))^2} \quad \dots (3)$$

$$\theta(p, q) = \tan^{-1} \frac{(L(p, q+1) - L(p, q-1))^2}{(L(p+1, q) - L(p-1, q))^2} \quad \dots (4)$$

6. Generating SIFT features: For this step, a 16 × 16 window is split into sixteen 4 × 4 windows to which the gradient magnitude and orientations will be calculated. The calculated orientations are put into a histogram of 8 bins. The above step is repeated for all 16, 4 × 4 pixels. Finally, the feature vector is determined. Before finalizing the feature vector, few hitches like rotational independence and illumination independence should be achieved.

Bag of Visual Words (BoVW)

BoVW is an extension of Natural Language Processing (NLP) and for image classification it is the Bag of Words algorithm[19,20], a good supervised learning model. It defines the histogram of visual words of an image. Outline of BoVW consists of firstly, sample selection of the extracted features that is key points and its descriptors from the above SIFT algorithm for the images in the dataset as in Figure 2.

Next, the visual calculated words are clustered around centroids from the extracted descriptors using K-Means algorithm technique[21]. This forms the visual vocabulary. In K-Means clustering, X objects are split into K clusters where the input is a set of features X= $\{x_1, x_2, x_3, \ldots x_n\}$. Minimizing the distance between each feature and assigning the centroids is the optimal goal as shown in (5).

$$\arg\min S \sum_{i=1}^{k} \sum_{p \in s_i} ||p - \mu_i||^2 \qquad \ldots (5)$$

Where, μ denotes mean of points for every cluster S_i and S the good set of points segregated into clusters of $\{S_1, S_2, \ldots, S_i\}$. Initially the cluster centroids are placed randomly within the bounds of points. Later K-Means iterates over the input features for further deciding on the closest centroid. The process is repeated until no further movement of cluster centroid is possible. SIFT features detected and computed for every image produces m × 128-dimensional array, where m indicates the number of extracted features. We then group similar features with vocabulary.

Lastly, visual words histogram is calculated based on the number of descriptors assigned to each visual word. These histograms are the Bag of Visual Words of same length. The histogram size will be number-of-images × number-of-clusters.

Classification Techniques

Support Vector Machine (SVM)

A well-known supervised algorithm for classification is SVM which is defined by a separating hyperplane and kernels[22]. The algorithm outputs an optimal hyperplane which results in categorization of the data. The objective is to find a plane to obtain maximum distance between the classes.

To classify using SVM, model uses the histogram array of size - no. of images × no. of features which are the samples to be trained. A trained multiclass SVM will classify the images into different species. SVM consists of many kernels such as Radial basis function (RBF), linear, nonlinear, polynomial, Gaussian kernel, sigmoid etc. The proposed system uses one of the kernels i.e RBF for training. RBF kernel is used classify non-linear points into right classification. The SVM classifier is tuned for classifying into different species of Indian herbs. The Figure 3. Shows the test image matching keypoints.

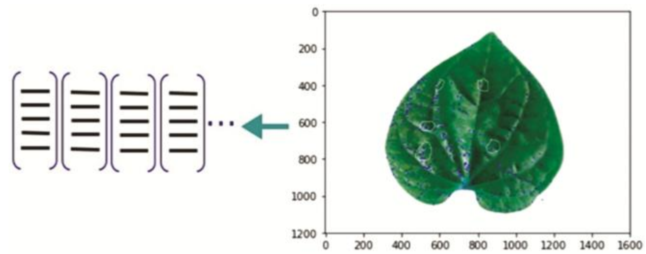

Fig. 2 — Detecting features and extracting descriptors

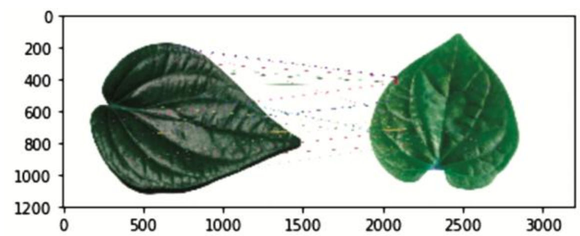

Fig. 3 — Matching keypoints

K-Nearest Neighbor (kNN)

It is used for both predicting and classifying problems[23]. The factor of 'K' is what drives the algorithm as it distinguishes between the training error rate and validation error rates. The 'K' value determines the nearest neighbors using the distance technique like Euclidean or cosine between every training and test points. The top 'K' matches are picked for classification.

Naive Bayes

This classifier algorithm is based on Bayes theorem for calculating probabilities including the conditional probability[24]. It yields good results on vast data sets. In brief, the Bayes theorem calculates the posterior probability P(c|y) from P(c), P(y) and P(y|c) as shown in (6).

$$P(c|y) = \frac{P(y|c)P(c)}{P(y)} \qquad \ldots (6)$$

Where, P(c|y) = Posterior Probability, P(y|c) = Likelihood, P(c) = Class Prior Probability, P(y) = Predictor Prior Probability.

Initially, the dataset is converted into frequency tables to which the Naive Bayes equation is applied for every class. The class with maximum probability is the outcome. It shows good performance over categorical data than numerical data. Limitation includes the assumptions on independent predictors. It is well suited for real time prediction because of fast performance.

Some of the potential Ayurveda herbs[25-27] used for medicinal purpose which needs concentration using ML techniques.

Materials and Methods

Proposed Models

In our work, different models are built to showcase the comparative study between different supervised learning classifiers and SIFT feature functioned on both newly-constructed and Flavia dataset as described below.

- *Model-01*: The model comprises of extracting SIFT features and then clustering using K-Means which results in visual dictionary. Finally, implemented classification using SVM.
- *Model-02*: It incorporates techniques such as SIFT feature extraction and classification using SVM.
- *Model-03*: This model extracts SIFT features with k-NN classification.
- *Model-04*: The model is built by extracting SIFT features and classified by Naive Bayes Classification technique.

Details of the phases included in all the above models are listed below:

- Phase-1: Dividing the newly-built database and Flavia database into 80:20 for training and testing the images.
- Phase-2: Image Acquisition – the query image is read through the camera on a white background.
- Phase-3: Feature Extraction – SIFT features are extracted from the images in the training folder
- Phase-4: BoVW Classification – K-Means clustering algorithm will group the descriptors and build the visual dictionary.
- Phase-5: Classification – Training the classifier using the histogram from previous phase and classify the leaf images in train folder.
- Phase-6: Predicting – Extracting SIFT features of the query images and predict their respective species.
- Phase-7: Display Results – Finally, displaying the recognized Indian medicinal herb.

The phases 1,2,3,6 and 7 remains same for all the models. The classification step varies for different classifiers.

The Figure 4 shows the conceptual diagram of Model-01 in detail. As in the next section, we prove that the efficient model for real-time classification of Indian medicinal herbs is Model-01. The accuracy obtained from the aforementioned model is highest compared to all other models with respect to both new dataset and the Flavia dataset.

Results and Discussion

The paper provides a good contrast between the four models. From the results obtained Model-01 shows better results in terms of accuracy and classification time. The models used the leaves from newly-constructed dataset and Flavia dataset. The four models are worked on 240 leaf images of eight species. The Indian herbs collected to build the dataset are Malabar Spinach, Betel, Neem, Curry, Amaranth, Mint, Tulsi and Hibiscus. Each species consists of 30 different leaves for training. The train set and test set are divided in 80:20 ratio. The newly-built dataset using DSLR camera is used to analyze the robustness of the four models as shown in Figure 5.

In all the models, the SIFT algorithm generates a huge number of features. But the classifiers could not handle such huge feature vector. Hence, in order to reduce the feature vector some features are dropped in the second, third and fourth model whereas in the first model the features are not dropped but the vector dimension is reduced. The features of size number-of-keypoints × 128 are reduced to row vector 1 × number-of-features in the dictionary. The BoVW uses the K-Means clustering where the features are clustered into k clusters. The cluster mean defines the

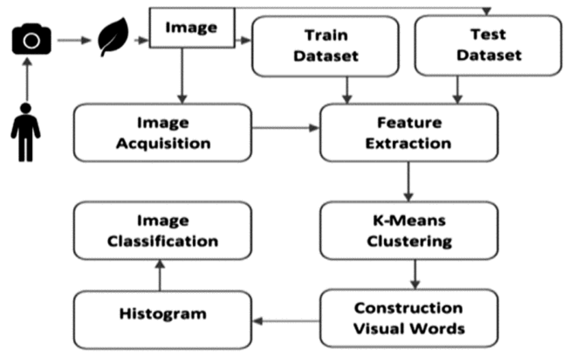

Fig. 4 — Conceptual diagram of proposed Model-01

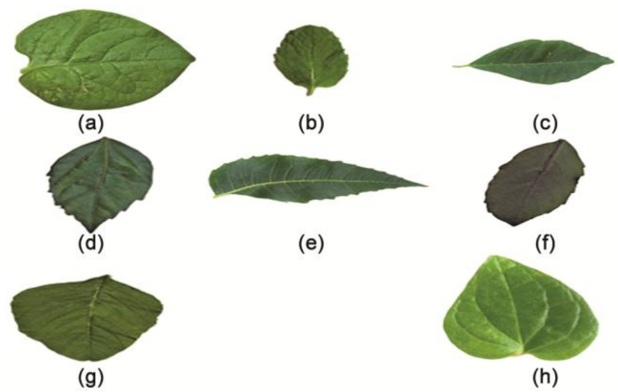

Fig. 5 — Samples of newly-built Dataset (left to right) (a) Malabar Spinach (b) Mint (c) Curry Leaf (d) Hibiscus (e) Neem (f) Tulsi (g) Amaranth (h) Betel

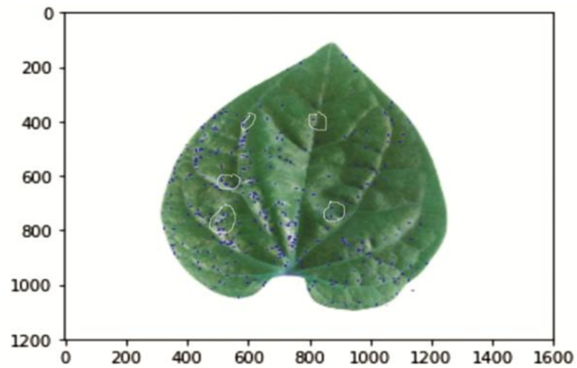

Fig. 6 — SIFT keypoints shown in blue color on Betel Leaf

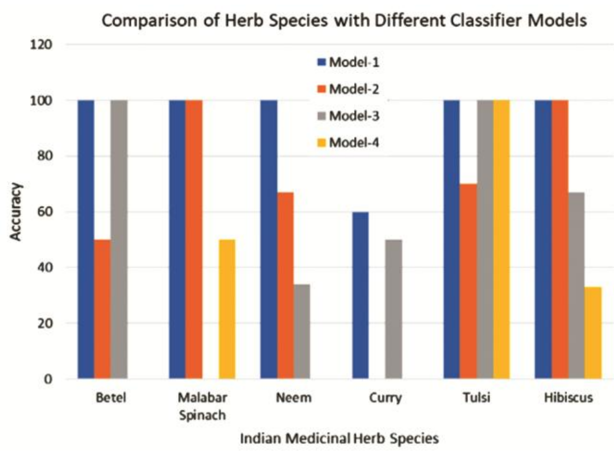

Fig. 7 — Accuracy of the four models for six different Indian Herbs

centroid. Thus, cluster centroids define the visual dictionary. Figure 6 shows the SIFT feature key points extracted from the query image.

Accuracy of both models are calculated as

$$Accuracy = \frac{x}{y} * 100 \qquad \dots (7)$$

Where,

x – denotes the no. of correctly classified category

y – denotes the total no. of images in the test folder of that category.

The experiment purely focuses on showing the performance of the all the models in terms of accuracy on new dataset and Flavia dataset. The Figure 7 shows the performance predicted for the six herb types in comparison with other four models with respect to new dataset and reveals that Model-01 is outperforming. The Figure 8 shows classification performance of the models on Flavia dataset and discloses that Model-01accuracy is 100% which is slightly better than on real dataset as the number of images in Flavia is 60 images per species. If the Real-Time dataset increases the totalnumber of images per

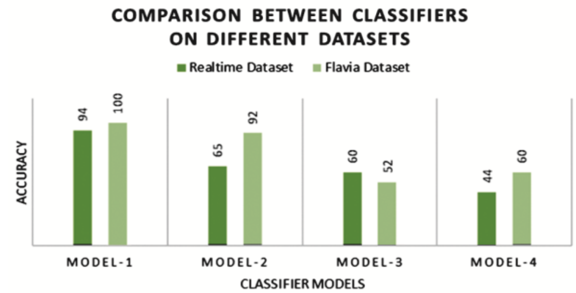

Fig. 8 — Comparison between different classifiers on newly-constructed (Realtime) and Flavia dataset

species from 30 to 60 or 100 then the accuracy can be expected to improve further.

Comparison chart of all models in Figure 7 shows that Model-01 with BoVW and SVM outperforms with 94% accuracy on newly-constructed dataset when compared to other models.

Conclusion

In this paper, a new Machine learning technique of using the visual dictionary to reduce the dimension of the SIFT feature vector without eliminating any features has been proposed on the Indian medicinal herb dataset. As there is no online dataset available for the proposed method for robust recognition of Indian Herbs, a self-built database is used. We have obtained satisfactory results and showcased that our first model – Model-01: SIFT-BoVW-SVM outperforms the other models with an accuracy of 94%. The other models such as, Model-02: SIFT and SVM shows an accuracy of 65%, Model:03-SIFT and kNN confirms accuracy of 60% with k parameter set to 5 and the last Model-04 with 44%. The same models outperformed well when fed with Flavia dataset. This is because Flavia dataset consists of 60 images for each species.

Future work consists of extending the construction of the dataset for the Indian herbs concentrating more on endangered species, increasing each species images in training to 60 or 100 in number, incorporating upcoming technologies of machine learning and deep learning technique on cloud and Internet of Things for Indian Herb dataset. Construction of digital herb dataset should look into parameters, such that digital leaf images are captured without plucking them, as many herbs are not available in bulk and to avoid herb extinction. Also, ensuring to collect and display the relevant details of the recognized herb to form a common platform to experts of various domains.

Conflict of Interest

Authors declare there is no conflict of interest

Author Contributions

Roopashree S Conceived and designed the analysis, Collected the data and wrote paper.

Anitha J conceived and investigated the analysis. Both authors discussed the results and commented on the manuscript.

References

1 El-Gayar M M, & Soliman H, A comparative study of image low level feature extraction algorithms, *Egyptian Inform J*, 14 (2) (2013) 175-181.

2 Larese M G, Craviotto R M, Arango M R, *et al*., Legume identification by leaf vein images classification. In Iberoamerican Congress on Pattern Recognition, (Springer, Berlin, Heidelberg),2012, p. 447-454).

3 Di Ruberto C & Putzu, L, A fast leaf recognition algorithm based on svm classifier and high dimensional feature vector, In 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Vol. 1, 2014, p. 601-609). IEEE.

4 Du J X, Huang D S, Wang X F, *et al*., Computer-aided plant species identification (CAPSI) based on leaf shape matching technique, *Transac Inst Measure Control*, 28 (3) (2006) 275-285.

5 Sathwik T, Yasaswini R, Roshini V, *et al*., "Classification of Selected Medicinal Plant Leaves Using Texture

6 Analysis", 4th ICCCNT – 2013.

7 Ghasab M A J & Khamis S, Mohammad F, *et al*., Feature decision-making ant colony optimization system for an automated recognition of plant species. Expert Systems with Applications, 42 (5) (2015) 2361-2370.

8 Priya C A, Balasaravanan T & Thanamani A S, An efficient leaf recognition algorithm for plant classification using support vector machine, International conference on pattern recognition, informatics and medical engineering (PRIME-2012), *IEEE, 2012*, p. 428-432.

9 Chakkaravarthy S S, Sajeevan G, Kamalanaban E, *et al*., Automatic leaf vein feature extraction for first degree veins. In: Advances in Signal Processing and Intelligent Recognition Systems, (Springer, Cham.), 2016, p. 581-592).

10 Venkataraman D & Mangayarkarasi N, Computer vision based feature extraction of leaves for identification of medicinal values of plants, IEEE International Conference on Computational Intelligence and Computing Research, 2016.

11 Gopal A, Prudhveeswar Reddy S & Gayatri V, Classification of selected medicinal plants leaf using image processing, International Conference on Machine Vision and Image Processing (MVIP), IEEE, 2012.

12 Ananthi C, Azha Periasamy & Muruganand S, Pattern recognition of medicinal leaves using image processing techniques, *J Nanosci Nanotechnol*, 2 (2) (2014) 214-218.

13 Maheswari A, Bharathi N, Neelamegam P, *et al*., Classification and recognition of herbal leaf using SVM algorithm, *Res J Pharmac, Biol Chem Sci*, 5 (5) (2014) 515-523.

14 Ibrahim Z, *et al*., Leaf recognition using texture features for herbal plant identification, *Indo J Elect Eng Comp Sci*, 9 (1) (2018) 152-156.

15 K Nithiyanandhan & Bhaskara Reddy P, Analysis of the medicinal leaves by using image processing techniques and ANN, *Int J Adv Res Comp Sci*, 8 (5) (2017) 1801-1805.

16 Pushpa B R, Anand C & Mithun Nambiar P, Ayurvedic plant species recognition using statistical parameters on leaf images, *Int J Appl Eng Res*, 11 (7) (2016) 5142-5147.

17 Li Qilong & Xiaohong Wang, Image classification based on SIFT and SVM, *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 2018, p. 762-765.

18 Lowe D, Object recognition from local scale-invariant features, *Proc IEEE Conf Computer Vision*, vol. 2, 1999, p. 1150-1157.

19 Lowe D, Distinctive image features from scale-invariant key points, *Int'l J Computer Vision*, 60 (2) (2004) 91-110.
Yin Z, Rong J & Zhi-Hua Z, Understanding bag-of-words model: A statistical framework, *Int J Mach Learn Cybern*, 1 (1) (2010) 43-52.

20 Hiroharu K & Tatsuya H, Image reconstruction from Bag-of-Visual-Words, *arXiv:1505.05190v1*, 2015.

21 Hartigan J A & Wong M A, A K-Means clustering algorithm, *Appl Stat*, 28 (1) (1979) 100-108.

22 Goh K S, Chang E Y & Li B T, Using one-class and two-class svms for multiclass image annotation, *IEEE Trans Knowl Data Eng*, 17 (10) (2005) 1333-1346.

23 Giuseppe A & Fabrizio F, KNN based image classification relying on local feature similarity, Third International Workshop on Similarity Search and Applications, SISAP 2010, 18-19 September 2010, Istanbul, Turkey.

24 Shih-Chung H, Chieh C & Chung-Lin H, Image classification using naive Bayes classifier with pairwise local observations, *J Inf Sci Eng*, 2017.

25 Santhosh Kumar S, Suclarshan M & Chakraborty A, The energy dispersive X-ray fluorescence spectroscopic study of *Solanum rubrun* Mill: A potential ayurvedic traditional medicinal plant, *Indian J Biochem Biophys*, 55 (2018) 280-285.

26 Boddupally S, Kameswari S V, Shaik M N, *et al*., Application of machine learning tools for evaluating the impact of premenopausal hysterectomy on serum anti-mullerian hormone levels, *Indian J Biochem Biophys*, 57, (2020) 245-251.

27 Das U, Saha T, Ghosh R, *et al*., *Trianthema portulacastrum* L.: Traditional medicine in healthcare and biology, *Indian J Biochem Biophys*, 57, (2020) 127-145.