Effect of Fusion of Statistical and Texture Features on HSI based Leaf Images with both Dorsal and Ventral Sides

Poonam Saini[1], Arun Kumar[2]

Sir Padampat Singhania University
Udaipur, India

Abstract—The present work involves statistically analyzing and studying the overall classification accuracy results using Hue channel images of different plant species using their dorsal and ventral sides, and then subjecting them to the process of feature extraction using first order statistical features and texture based features. These extracted features have been subjected to the classification process using KNN and Random Forest algorithms. Further, this work studies the fusion of two different kinds of features extracted for dorsal and ventral plant leaf images and studying the effect of fusion on the overall classification accuracy results. This work also delves into the feature selection task using random forest algorithm and studies the effect of reduced dataset with unique features on the overall classification accuracy results. The most important outcome of this investigation is that the ventral leaf images can be a suitable alternative for plant species classification using digital images and further, the fusion of features does improve the classification accuracy results.

Keywords—Dorsal; ventral; leaf classification; random forest; texture features; statistical features

I. INTRODUCTION

The human beings have been gifted with brain and eyes for proper discrimination of objects found in their daily lives. While discriminating the objects, the images formed in the eye seek important valuable features from the objects of concern and store those cues for future reference. The eyes often seek color, shape and texture features for identification process. The color is a very important discrimination parameter, as one always sees objects with color as almost first criteria for discrimination. As the picture is formed in the eyes, the brain interprets these visual cues, the color is one of the most important factors that is stored as valuable information for discrimination purpose. Secondly, the shape of the object being regular or irregular polygon, forms the other important parameter for discrimination by the human brain. Thirdly, the property of smoothness or roughness of the surface of the object, i.e. the sense of touch has its role in the discrimination of the objects by the human brain.

The plants have been studied for their fruits, seeds, roots, flowers and leaves etc. The scientific methods for the proper taxonomic classification of the plants and their species have been there with the biologists for quite a long time, but the role of image processing professionals by applying the latest state of the art machine learning techniques in discriminating the plants through their digital images is a recent one, and this

has become possible with the improvement of computational efficiency of the computers and also the decrease in the prices of the hardware. There are millions of plants and their subspecies and not all of them have been taxonomically classified and this all accounts for the slow identification tools and techniques of the past and vastness of the flora on this earth. Therefore, in order to preserve the plants for future generation, one needs to understand them and taxonomically classify them for any future reference.

The research work carried out by [1] has prepared a mobile biodiversity informatics tools for identifying and mapping Indonesian medicinal plants. The system - called MedLeaf – has been developed as a prototype data resource for documenting, integrating, disseminating, and identifying of Indonesian medicinal plants. The results indicate that combination of leaves features outperform than by using single features with an accuracy value of 88.5%.

The researchers in [2] have implemented an image feature extraction method using morphological techniques and have worked on leaf diseased dataset containing leaf spot disease and leaf blight. The results have been obtained using morphological features.

The researchers in [3] have carried work on identification of whitefly on the plants using image processing technique. The segmentation results using triangle method has an accuracy of 75.36%.This suggests that triangle method can be used for the whitefly segmentation process in vegetable crops leaf images.

The leaf image texture features have been extracted by [4], through Gabor based techniques and then subjecting them to PSO-CFS based search method for identifying the best set of features from the complete feature set and then classifying them using four classification algorithms like KNN, J48, CART and RF. Another objective of [4] was to utilize the two faces available on the plant leaves (Dorsal and Ventral), instead of one (i.e. Dorsal) for classification of plants on the basis of digital leaf images and to analyze the effects on classification accuracy values for dorsal and ventral sides of leaf images. The accuracy achieved using this work has been 92.09%.

The researchers in [5] have worked on extraction of standard tobacco leaves based on HIS color space, wherein the H, S and I components have been quantified and extracted by

using color histogram, and then the average value corresponding to every color component was calculated and the grading for tobacco leaves was done on this basis.

The researchers in [6], have utilized eight varieties of crop images. In this work, the authors have worked with texture features obtained using GLCM method and color features using HSV have been deployed. The artificial neural network (ANN) has been used for classification of plants in this work. The highest value of classification accuracy achieved by this work is 84.375%.

The GLCM and FFT techniques for feature extraction from the color images have been used by [7] and have achieved an accuracy value of 86% for the GLCM based color feature set.

The researchers in [8] have used the concept of chromaticity moments for the images and has extracted the texture features for the images. The accuracy value achieved is as high as 90% using SVM algorithm for classification of diseased plants.

The work carried out by [9] considers the classification of a household plant used for its medicinal properties by using the shape and color features obtained from the leaves of the plant. This work has utilized the shape descriptors represented by Dyadic Wavelet transform and Zernike complex moments along with the HSV based color features and have obtained a classification accuracy value of 81.77%.

In the present work, the section II explains about the methodology adopted for carrying out the research work. It explains the role of HSI color models for extraction of color features. This section also explains the extraction of first order statistical features and texture features and the preparation of feature database and along with it, it also describes about the computation of variance values between various features using PCA. The section III describes about the feature selection methodology adopted to prepare the feature subsets. The section IV discusses the results obtained for various subsets of data created in the previous sections along with their statistical analysis and comparison of the present results with other works of similar nature which is followed by the conclusion drawn from this work in section V.

II. Materials and Methods

The plant leaves have two faces viz.: dorsal and ventral, therefore there is a need for paying critical attention on both the sides of the leaves, as both have independent and unique set of features. The existing leaf image databases available on the internet contain the leaf images of dorsal sides, but to achieve the objective of this work, there is a need to create an independent leaf image database with dorsal as well as ventral sides of the leaves. Therefore, 25 dorsal side and 25 ventral side leaf images were clicked for each plant species and a database of ten plant species has been created with 250 dorsal and 250 ventral sides of leaf images and 500 images in totality. A sample of such colored leaf images with dorsal and ventral sides is shown in Fig. 1 by [10]. The 500 images captured were subjected to the process of background removal, size reduction to 256 X 256.

Fig. 1. A Sample of Collected Leaf Images with Dorsal and Ventral Sides.

The human eye considers the color as important characteristic feature for the recognition of the objects visible through naked eye in the images as well as for various image processing tasks like segmentation and detection etc. The color features are defined subject to a particular color space or model. There are a number of color spaces that have been used by different researchers, such as RGB, LUV, HIS. By using the different color spaces, the different color features can be extracted from images or the regions of interest in the images.

A. Extraction of Color Features using HSI Color Model

There has been numerous works carried out on the principles of binary and gray image processing techniques. The image processing tasks carried out on the basis of color image processing have high computational dependency and the earlier works reduced the images into two dimensions color maps to reduce the time consumed in computing. Certainly by doing so, there is loss of information upon the conversion of color images to binary or gray levels. The growth of computing power, the storage capacity and the reduced costs of capture systems and printing images have made the image acquisition and further processing in color easier and reasonably computationally cheaper. The RGB color model has three color components (Red, Green or Blue) in varying proportions. To overcome the disadvantages of RGB model in various image processing tasks, HSI model fares better and is a simple substitute of human visual system. The HSI model contains three matrices of *mXn* dimensions containing values with respect to hue, saturation and intensity. Each pixel of the image prepared according to this model has hue and saturation values indicating the color information contained in it and intensity indicates the brightness.

Therefore, the HSI color space model is suitable for detection and analysis of the color characteristic properties. By using (1), (2) and (3), the five hundred RGB images captured for ten different plant species were subjected to the process of RGB to HSI conversion process. The HSI images prepared using above equations were converted into a stack of images and then the stack was split into three independent channels (H, S and I) respectively.

As discussed earlier, the color information is contained in the hue channel, therefore the hue channel images were further subjected to the process of feature extraction using two techniques namely the general first order statistical features and second order texture features using the co-occurrence

matrix. There were ten statistical features: Mean, Standard Deviation(StdDev), Max, XM, YM, Integration Density (InteDen), Median,Skewness (Skew), Kurtosis (Kurt), Area_Percent. that were extracted for this study using Fiji by [11] and the different values were stored in a CSV file.

Texture is another important property of images. It is generally believed that human visual systems use texture for recognition and interpretation. The texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. The texture, on its own does not have the capability of finding similar images, but it can be used to classify textured images from non-textured ones and then be combined with another visual attribute like color to make the retrieval more effective.

The researchers in [8] have proposed 14 texture features, of which 11 have been chosen and they are Contrast(Contr), Homogeneity, Angular Second Moment (ASM), IDM, Energy, Entropy, Variance, correlation, Inertia, Shade and Prominence. The GLCM matrix was prepared for $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$ with an offset of unity in ImageJ using the GLCM plugin for batch processing given by [11]. The GLCM matrix of the dorsal and ventral sides of the leaf image was prepared and combined together. The complete matrix contains 600 rows per degree with unit offset of data for the above fourteen attributes and in totality 2400 rows for all the four degrees and with unit offset values was prepared.

$$I = \frac{1}{3}\left(R + G + B\right) \tag{1}$$

$$S = 1 - \frac{3}{\left(R + G + B\right)}\left[\min\left(R, G, B\right)\right] \tag{2}$$

$$H = \cos^{-1}\left\{ \frac{\frac{1}{2}\left[\left(R - G\right) + \left(R - B\right)\right]}{\left[\left(R - G\right)^2 + \left(R - B\right)\left(G - B\right)\right]^{1/2}} \right\} \tag{3}$$

B. *Preparation of Database with Different Feature Set*

The first order statistical database has been prepared using the 10 first order features mentioned in section (IIA). Two independent features sets have been prepared for dorsal and ventral side images of leaves using hue channel based images. The ten features extracted are Mean(SD1 or SV1), StdDev(SD2 or SV2), Max(SD3 or SV3), XM(SD4 or SV4), YM(SD5 or SV5), IntDen(SD6 or SV6), Median(SD7 or SV7), Skew(SD8 or SV8), Kurt(SD9 or SV9), Area_Percent(SD10 or SV10). Here D in SD_i indicates the dorsal and V in SV_i indicates ventral leaf images.

The two datasets are called HDIM (Hue-Dorsal-Istorder-Measure) and HVIM (Hue-Ventral-Istorder-Measure) as represented through (4) and (5) respectively. Here $SD_1, SD_2, ..., SD_{10}$ indicate all the 10 different values of first order statistical features obtained for dorsal images and the $SV_1, SV_2, ..., SV_{10}$ indicate all the 10 different values of statistical

first order features obtained for the ventral images.

$$HDIM = \left(SD_1, SD_2, SD_3, SD_4, SD_5, SD_6, SD_7, SD_8, SD_9, SD_{10}\right) \tag{4}$$

$$HVIM = \left(SV_1, SV_2, SV_3, SV_4, SV_5, SV_6, SV_7, SV_8, SV_9, SV_{10}\right) \tag{5}$$

By using the gray level co-occurrence feature extraction methodology as mentioned in section (IIA), the texture feature dataset has been prepared for the dorsal as well as the ventral leaf images for which the Hue channel has been extracted.

$$HDGM = \left(TD_1, TD_2, TD_3, TD_4, TD_5, TD_6, TD_7, TD_8, TD_9, TD_{10}, TD_{11}\right) \tag{6}$$

$$HVGM = \left(TV_1, TV_2, TV_3, TV_4, TV_5, TV_6, TV_7, TV_8, TV_9, TV_{10}, TV_{11}\right) \tag{7}$$

The two datasets are called HDGM(Hue-Dorsal-GLCM-Measure) and HVGM(Hue-Ventral-GLCM-Measure) as represented through (6) and (7) respectively. Here $TD_1, TD_2, ..., TD_{11}$ indicate all the 11 different values of texture features obtained for the dorsal images and the $TV_1, TV_2, ..., TV_{11}$ indicates all the 11 different values of texture features obtained for the ventral images.

There are two more datasets prepared by combining all the dorsal and ventral features together and named as HDGIM(Hue-Dorsal-GLCM-Istorder-Measure) and HVGIM(Hue-Ventral-GLCM-Istorder-Measure) and are represented through (8) and (9) respectively, where HDGIM has been obtained by combining the features obtained through (4) and (6), whereas HVGIM has been obtained by combining features of (5) and (7).

$$HDGIM = HDIM + HDGM \tag{8}$$

$$HVGIM = HVIM + HVGM \tag{9}$$

C. *Studying the Variance amongst the Various Features Extracted using PCA*

The feature sets prepared in subsection B have been thoroughly studied for the correlation or variance amongst each other so that only those features could be selected which are least correlated and are unique in nature. By using unique features, the size of the dataset is reduced and the overall computation time is reduced as well for discrimination of the datasets using images.

The PCA component plots have been shown in the Fig. 2, 3, 4 and 5 for different datasets. These plots demonstrate the importance of unique features required for the classification process. The concept of PCA (Principal Component Analysis) algorithm as discussed by [11] and [12], is an unsupervised method, it has been in vogue for the dimensionality reduction process in almost all the literature concerning the classification of data.

This technique is used for extracting important variables (in form of components) from a large set of variables available in a data set. It extracts low dimensional set of features from a high dimensional data set with a sole motive to capture as much information as possible. By using a fewer variables, the process of visualization of data also becomes much more meaningful and the behavior of the data can be studied in a

better way. The concept of PCA is more useful when dealing with three or higher dimensional data variables. It is a technique to combine similar or correlated items or variables. The normalized data is subjected to the process of PCA and only those dimensions are selected which have the highest value of variance and this process churns out the highly correlated items from the variable set. The Fig. 2, 3, 4 and 5 show the plots with two PC's (principal components, PC1 and PC2) obtained for the various variables used. The principal component is a normalized linear combination of the original predictors in a data set.

The first principal component (PC1) is a linear combination of original predictor variables which captures the maximum variance in the data set. It determines the direction of highest variability in the data. The larger the variability captured in first component PC1, the larger is the information captured by that component and no other component can have variability higher than first principal component. The first principal component results in a line which is closest to the data i.e. it minimizes the sum of squared distance between a data point and the line. **The second principal component (PC2),** is also a linear combination of original predictors which captures the remaining variance in the data set and is uncorrelated with PC1. In other words, the correlation between first and second component should is zero. If the two components are uncorrelated, their directions should be orthogonal. The Fig. 2 and 3 show the plot for principal component and their values. The Fig. 2 shows that the first principal component of HDIM dataset has a value of 42.08% and the values for other components are lower than this value. For HVIM dataset the Fig. 3 shows, the highest value of principal component is 52.79%. In Fig. 4 for HDGM dataset, it is 55.51% and that of HVGM, it is 60.16% as shown in Fig. 5.
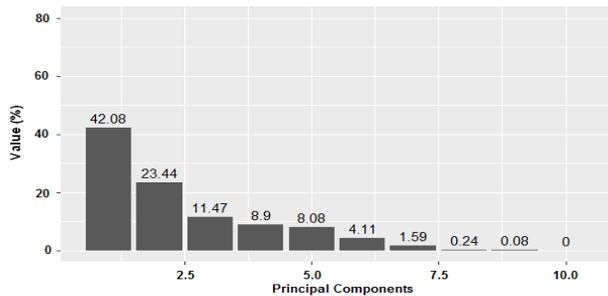
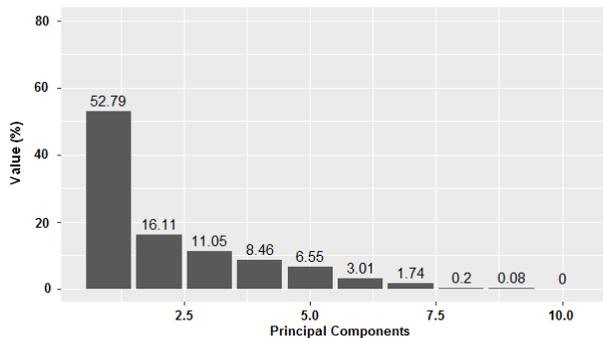Fig. 2.   PCA Component of HDIM.
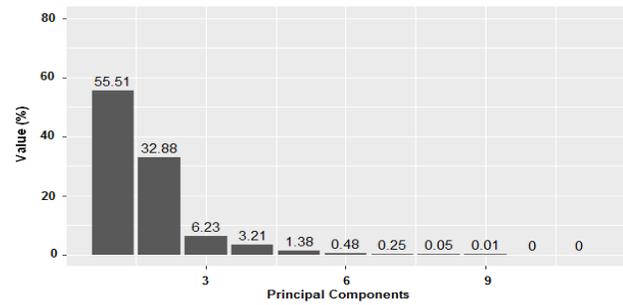
Fig. 3.   PCA Component of HVIM.
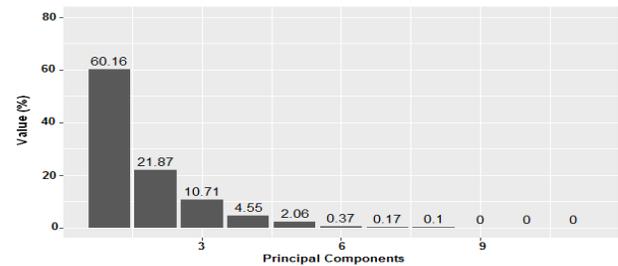
Fig. 4.   PCA Component of HDGM.

Fig. 5.   PCA Component of HVGM.

III.   ADOPTION OF FEATURE SELECTION METHODOLOGY FOR FEATURE SUBSET SELECTION

The concept of feature selection also known as variable selection, feature reduction, attribute selection or variable subset selection, is a widely used dimensionality reduction technique, which has been the focus of much research in machine learning and data mining and has found applications in text classification, web mining, and so on . It allows faster model building by reducing the number of features, and also helps removing irrelevant, redundant and noisy features. This begets simpler and more comprehensible classification models with better classification performance results.

Hence, selecting relevant attributes are a critical issue for competitive classifiers and for data reduction. In this present work, random forest technique has been used for feature subset selection which identifies unique features from large datasets.

A. Use of Random Forest as a Feature Selector for Feature Subset Selection

Random Forest directly performs feature selection while a classification rule is built. The two commonly used variable importance measures in RF are Gini importance index and permutation importance index (PIM) .In this paper, two step approach has been used for feature selection. In first step, permutation importance index are used to rank the features and then in second step, Random Forest is used to select the best subset of features for classification. This reduced feature set is then subjected to the process of plant species classification using images of their leaves.

The high dimensional nature of many tasks in pattern recognition has created an urgent need for feature selection techniques. The goal of feature selection in this field is manifold, where the two most important are: to avoid over fitting and improve model performance, and to gain a deeper

insight into the underlying processes that generated the data [12]. The interpretability of machine learning models is treated as important as the prediction accuracy for most life science problems. Unlike most other classifiers, Random Forest directly performs feature selection while a classification rule is built by [13]. Permutation importance measure (PIM) given by [14] is arguably the most popular variable's importance measure used in RF.

B. Preparation of Different Feature Subsets using Random Forest Feature Selection Technique

By suitably applying the RF, one can examine which variables are working the best or worst in each of the trees. In this study, CARET package developed by [13] has been used for finding the importance of the features which builds up a model. The decision trees help in determining the variable importance.
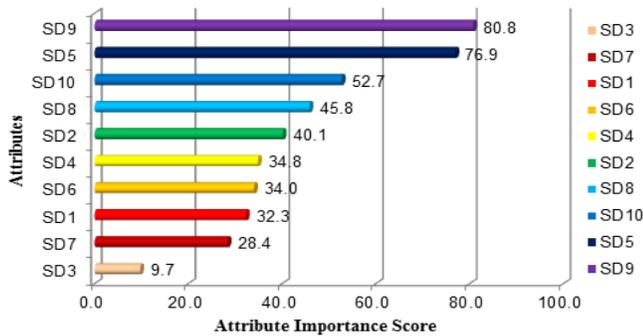

Fig. 6.   Attribute Importance Plot for HDIM.

Fig. 7.   Attribute Importance Plot for HVIM.

Fig. 8.   Attribute Importance Plot for HDGM.

Fig. 9.   Attribute Importance Plot for HVGM.

The cross validation technique has been utilized for identifying the error rate and this helped in identifying the fitness of the individual features. The Fig. 6, 7, 8, 9, 10 and 11 clearly depict the attribute importance value for each variable which helps in choosing the unique variables and forming the feature subsets to be utilized for the classification process and represent HDIM, HVIM, HDGM, HVGM, HDGIM and HVGIM feature datasets respectively and arranged in ascending order of variable importance. In the present work; seven unique features have been selected from each of the feature datasets mentioned in Fig. 6, 7, 8, 9, 10 and 11 for further classification process. These features have been selected on the basis of feature score.
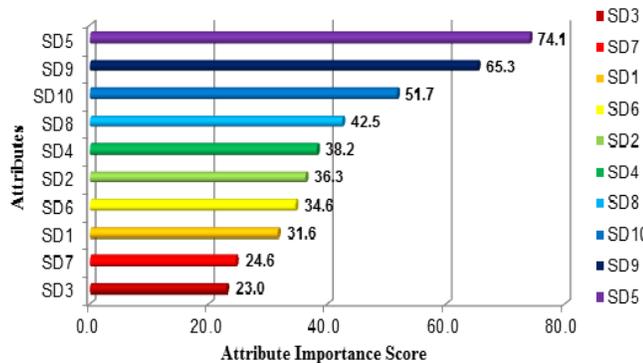

Fig. 10. Attribute Importance Plot for HDGIM.

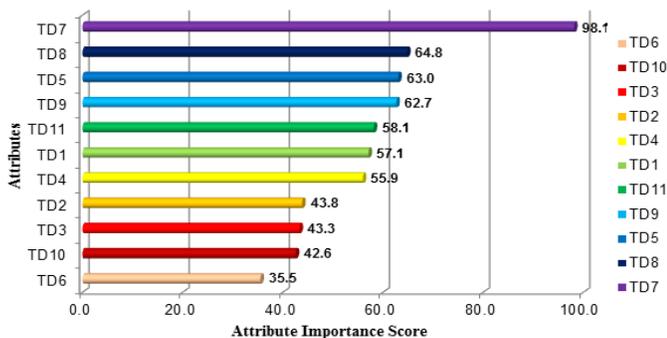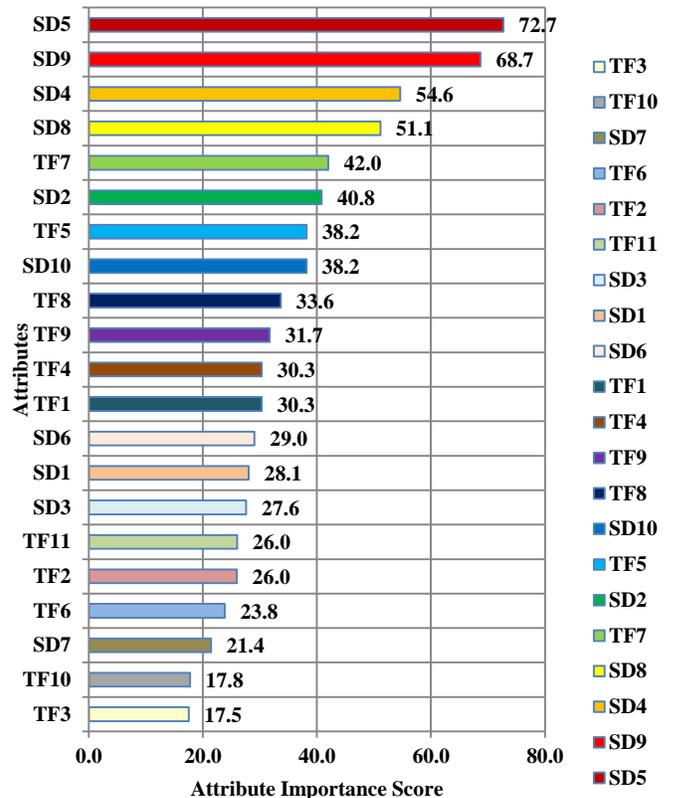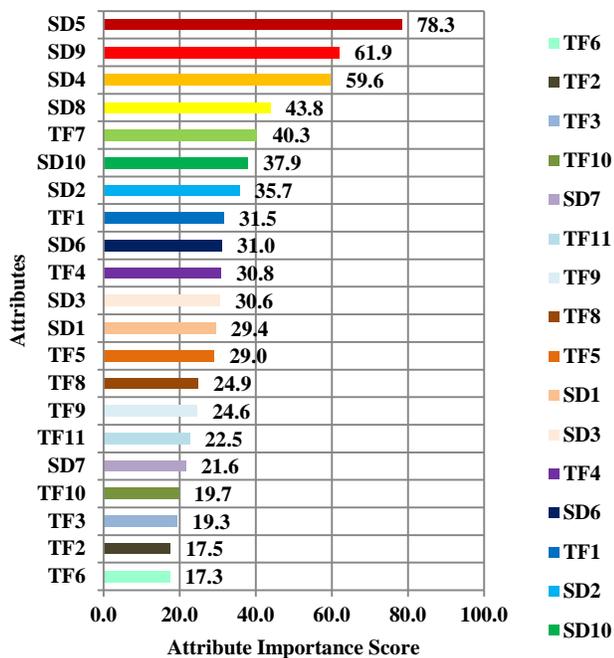Fig. 11. Attribute Importance Plot for HVGIM.

In the present work, the statistics based features, texture based features have been computed for the dorsal and ventral leaf images of the different plant species.

By using the random forest feature selection method, six more datasets have been prepared and mentioned through (10), (11), (12), (13), (14), (15) and each of the subset datasets contain only seven most important features which are unique in nature and thereby reducing the datasets considerably.

$$HDIM\_S = \left(SD_2, SD_4, SD_5, SD_6, SD_8, SD_9, SD_{10} \right) \quad (10)$$

$$HVIM\_S = \left(SV_2, SV_4, SV_5, SV_6, SV_8, SV_9, SV_{10} \right) \quad (11)$$

$$HDGM\_S = \left(TD_1, TD_4, TD_5, TD_7, TD_8, TD_9, TD_{10} \right) \quad (12)$$

$$HVGM\_S = \left(TV_1, TV_4, TV_5, TV_7, TV_8, TV_9, TV_{11} \right) \quad (13)$$

$$HDGIM\_S = HDIM\_S + HDGM\_S \quad (14)$$

$$HVGIM\_S = HVIM\_S + HVGM\_S \quad (15)$$

HDIM_S(Hue-Dorsal-Istorder-Measure_Subset), HVIM_S (Hue-Ventral-Istorder-Measure_subset), HDGM_S(Hue-Dorsal-GLCM-Measure_Subset), HVGM_S(Hue-Ventral-GLCM-Measure_Subset), HDGIM_S(Hue-Dorsal-GLCM-Istorder-Measure_Subset), HVGIM_S(Hue-Ventral-GLCM-Istorder-Measure_Subset) are four feature subsets prepared with unique features which have the highest scores for attribute importance as shown through Fig. 6, 7, 8, 9, 10, 11 and through (10), (11), (12), (13), (14) and (15). HDGIM_S and HVGIM_S are the two datasets prepared by using the combined features.

IV. RESULTS

The classification accuracy results have been computed for all the twelve different kinds of datasets prepared in the present work and has been compared with [4, 15, 16] and have been shown through Fig. 15 and 16.The predictive accuracy values have been calculated using KNN and Random Forest algorithms. The corresponding Kappa accuracy results have been calculated as well and shown through Fig. 12 and 13. The result Fig. 12 shows the Kappa results calculated for the complete dataset for HDGIM and HVGIM which have fared better with values of 98.61% and 99.45% respectively and these results substantiate the point that the fusion of features improves the predictive accuracy results. The Fig. 13 shows the kappa values for the feature subsets prepared and HDGIM_S and HVGIM_S datasets have fared better as compared to other datasets with values 98.91% and 99.51% respectively. Further it can be substantiated that the ventral leaf image dataset has proved to be better performer.

By using seven features in dorsal as well as ventral feature subset of datasets, ten features in HDIM and HVIM, eleven features in case HDGM and HVGM and twenty one features in case of HDGIM and HVGIM; a difference of maximum percentage accuracy results achieved by using Random forest algorithm, through the creation of different datasets and has been shown in Fig. 14. By using subset of features, there is a rise in percentage accuracy values, but exception in the case of HVIM_S and HVGM_S. Therefore, it has been observed that there is a positive effect on predictive accuracy results, by using a small subset of feature datasets as compared to datasets with more number of features.
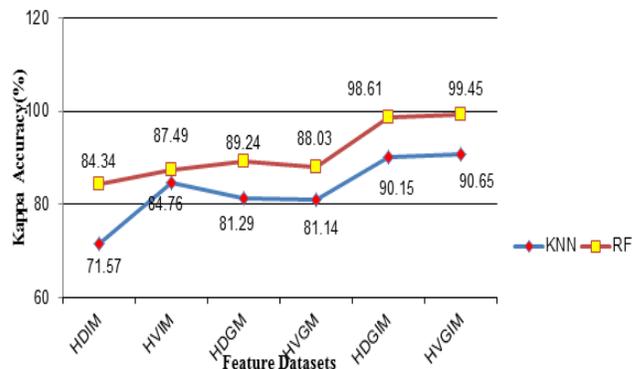

Fig. 12. Predictive Kappa Accuracy Results for Complete Feature Set.
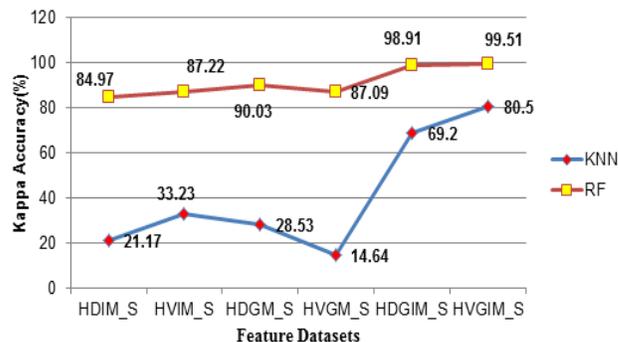

Fig. 13. Predictive Kappa Accuracy Results for Feature Subset.

On observing Fig. 15, on fusing the dorsal datasets (HDIM and HDGM), the rise in percentage predictive accuracy is approximately 10.63% and in the case of fusion of ventral datasets (HVIM and HVGM), the rise in predictive accuracy is approximately 10.52%. Therefore, in the result Fig. 15, HDGIM (98.75%) and HVGIM (99.56%) have shown the maximum value of predictive accuracy results using Random forest based classification algorithm amongst all the complete datasets created. This shows that the fusion of features have comparative effect on the overall classification accuracy results.

On observing Fig. 16, on fusing the subset features obtained for dorsal datasets (HDIM_S and HDGM_S), the rise in percentage predictive accuracy is approximately 10.26% and in the case of fusion of subset features obtained for ventral datasets (HVIM_S and HVGM_S), the rise in predictive accuracy is approximately 11.10%. The result Fig. 16 is for the subsets of features and HDGIM_S (99.02%) and HVGIM_S (99.56%) have shown better results and these results are even better than the combined feature subsets as well.

The Fig. 15 and 16 also depicts the comparison of results with [4, 15, 16]. Complete datasets prepared in the present work have been compared with [4, 15, 16] as shown in Fig. 15 and it has been observed that the HDGIM and HVGIM fare better over all the datasets. On the other hand, the results obtained in the present work with HDGIM_S and HVGIM_S have fared better than mentioned by [4] as shown in Fig. 16.
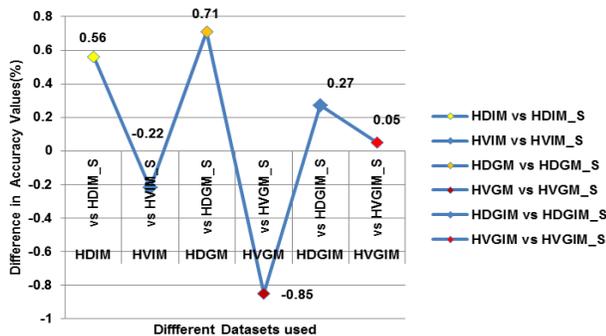

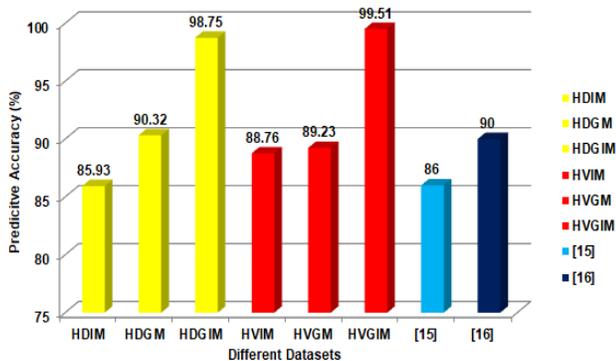Fig. 14. Difference in Maximum Percentage Accuracy Values of Different Datasets.

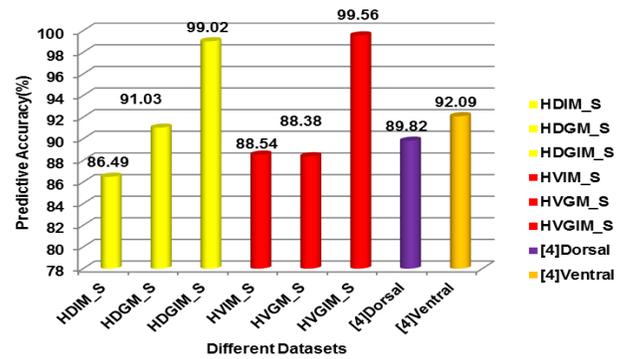Fig. 15. Comparison of the Complete Dataset Results with the Work of [15, 16].

Fig. 16. Comparison of the Feature Subset Results with the Work of [4] having Dorsal and Ventral Sides of Leaf Images.

V. CONCLUSIONS

The present work portrays the efficacy of ventral leaf image dataset based classification of plant species over the dorsal image datasets, which are in vogue. The concept of color as used in the present work, especially the hue channel which carries the color information, is based on human perception of classification of objects into classes based on visual perception of color in human minds and the concept of texture features based on feeling of touch and their subsequent fusion to obtain a combined set of features, has resulted in a new understanding of the fact that the combination of features can be utilized for improved classification accuracy results in the case of plant species classification using digital images. The application of PCA on results in understanding various features thereby helping in overall reduction of the dataset.

The random forest based feature selection technique for minimizing the datasets has provided improved classification accuracy results.

REFERENCES

[1] Herdiyeni, Y., Ginanjar, A. R., Anggoro, M. R. L., Douady, S., & Zuhud, E. A. (2015, November). MedLeaf: Mobile biodiversity informatics tool for mapping and identifying Indonesian medicinal plants. *Proceedings of 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR),* 54-59. IEEE, Fukuoka, Japan.

[2] Manik, F. Y., Herdiyeni, Y., & Herliyana, E. N. (2016). Leaf morphological feature extraction of digital image anthocephalus cadamba. *TELKOMNIKA (Telecommunication Computing Electronics and Control), 14*(2), 630-637.

[3] Hutomo, D. T., Herdiyeni, Y., MKom, M., & Rauf, A. (2013). Whitefly (Bemisia tabaci) Calculation based on image processing using triangle method. *Proceedings of International Conference on Computer, Control, Informatics and its Applications*, 165-168, Indonesia.

[4] Kumar, A., Patidar, V., Khazanchi, D., & Saini, P. (2016). Optimizing feature selection using particle swarm optimization and utilizing ventral sides of leaves for plant leaf classification. *Procedia Computer Science, 89*, 324-332.

[5] Ji, J. T., Deng, M. L., He, Z. T., Jia, S. T., Du, X. W., He, Y. K., & Liu, J. J. (2014). Extraction of tobacco leaves color features based on HSI color space. *Applied Mechanics and Materials,* Trans Tech Publications, 651, 2424-2429.

[6] Blotta, E., Bouchet, A., Ballarin, V., & Pastore, J. (2011). Enhancement of medical images in HSI color space. *Journal of Physics: Conference Series, 332*(1), 012041. IOP Publishing.

[7] Gonzalez and Woods (1992). The HSI color model, derivations of HSI to RGB and RGB to HSI conversion equations. In: Digital Image Processing, (1st ed.). Addision Wesley.

[8] Haralick, R. M., and Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610-621.

[9] Arai, K., Abdullah, I. N. and Okumura H.(2013) Image Identification Based on Shape and Color Descriptors and Its Application to Ornamental Leaf. *Interbational Journal of Image, Graphics and Signal Processing,* 10, 1-8.

[10] Kumar, A., Patidar, V., Khazanchi, D., Purohit, G. & Saini, P. (2016) An approach to improve the classification accuracy of leaf images with dorsal and ventral sides by adding directionality features with statistical feature sets. In R. Choudhary, J. Mandal, N. Auluck, H. Nagarajaram (Eds), *Advanced Computing and Communication Technologies. Advances in Intelligent Systems and Computing* (pp. 89-97), 452. Springer, Singapore

[11] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T. & Tinevez, J. Y. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, *9*(7), 676-682.

[12] Charpentier, A. Variable importance plot and variable selection (2015, *June, 17)*. Retrieved from https://feakonometrics.hypothesis.org/19835

[13] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Retrieved from http://www.R-project.org.

[14] Latte, M. V., Shidnal, S., Anami, B., & Kuligod, V. (2015). A combined color and texture features based methodology for recognition of crop field image. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *8*(2), 287-302.

[15] Ghazali, K. H., Mustafa, M. M., & Hussain, A. (2007). Color image processing of weed classification: a comparison of two feature extraction technique. *Proceedings of the International Conference on Electrical Engineering and Informatics,* 607-610.

[16] Tian, Y., Niu, Y., & Li, T. (2008, October). A study on the method of image processing and feature extraction for cucumber diseased. *Proceedings of International Conference on Computer and Computing Technologies in Agriculture,* Springer, Boston, MA. 967-972.