Asian Journal of Information Technology 16 (11-12): 810-815, 2017
ISSN: 1682-3915

Feature Selection using Group Search Optimizer for Plant Leaf Classification

A.M. Ravishankkar and P. Amudhavalli
Department of CSE, Karpagam Academy of Higher Education, Karpagam University,
Coimbatore, Tamil Nadu, India

Abstract: Ecosystem provides food, medicine, oxygen, fuel, etc., for plants to play a vibrant role while plant identification has an essential role to play in numerous fields such as agriculture, medicine, culinary art, etc. Due to non-availability of fruits and flowers throughout the year making the classification difficult besides time consuming, plants classification can be done with the help of leaves and flowers. Many researchers have proposed various methods to detect plants besides using various detecting systems to classify neural system. Researchers were attracted towards neural systems to recognise area pattern due to its strength to learn from training dataset. In addition, to recognize pattern, selecting attributes is a vital task to enable fortitude of most pertinent features. Due to its rapid acceleration in coming closer, Group Search Optimizer (GSO), a swarm-based effective algorithm has been successfully put in place to various problems.

Key words: Plants, plant leaf identification, neural networks, feature selection, Group Search Optimizer (GSO), detect plants

INTRODUCTION

Since, ages, human beings have been able to enjoy the fruits of nature, its flora and fauna. However, due to continuous technical progress besides the need for superior roads, bridges, shelters, trees and vegetation have been subjected to thoughtless felling to pave way for such infrastructure. On one hand while such development work has led to vanishing of flora and fauna, on the other hand, it has become essential to maintain the environmental equilibrium. However, at the same time, the human pursuit for detecting and scientific classification of plants and sub-species together with creating procedures to conserve them for future prior to plant species getting extinct have been in vogue for decades (Kumar *et al*., 2016).

Studies on plants have been made for flowers, leaves, seeds besides fruits. Though, there exist millions of various kinds of plant, many sub-types are still unidentified and they will invariably die and become non-existent before they become known (Sumathi and Kumar, 2013). Thus, there is a need for detecting and classifying the plant automatically for speeding up the action of learning the individual plant types. To suggest innovative methods to identify plant types, biologists and computer experts have been constantly playing their roles. Computerized vision methodology has transformed the task of classifying the plant automatically to find ideal distinctive features from the digital images besides ideally classifying them into various types. There requires finding the subset of the data to do the task like the whole data does, since, the data obtained from the digital images is quite gigantic.

Plants classification is fundamentally in accordance with the shapes of the leaves and flowers besides the theory of plant shape classification. The shape of the leaves is invariably around two-dimensional. Neural systems are built with manifold stratums of neurons or computational units with the entire neurons unified together. The input fed on its strata transmits through the system in forward direction through the concealed stratums to produce a yield. The yield signal is derived using weights, bias besides initiation task. The neural systems get trained with the application of reverse transmission rule by reversing the inaccuracies and node's altering weights. The inaccuracy is the variation between the yield has been. The shape being a vital feature in an image, edge identifying methodology is applied for taking out the shape feature.

An important task of classifying leaves system is to take out the common features among the images which belong to the same species of data set for ultimate indexing. This methodology is used to seize image's pictorial content for indexing besides retrieval. The identifying task becomes difficult due to wide unevenness of shapes and dimensions of leaves for

Corresponding Author: A.M. Ravishankkar, Department of CSE, Karpagam Academy of Higher Education, Karpagam University, Coimbatore, Tamil Nadu, India

plants and trees. The methodologies to take out the feature analyse leave's images to obtain the superior noticeable features representing the different species of objects can be identified.

The aim of selecting the attribute is to extract a feature space having low measurable extent, retaining adequate data, enhancing separation in feature space besides ability to compare features among specimens in the same category. The feature selection's goal is to diminish the measureable extent of vectors connected to patterns by choosing a subset of characteristics to a smaller size than the original. Disregarding redundant features, the classifying agent's action is invariably improved.

For reducing the large dataset to a lesser subset, the algorithms of feature selection become essential with daily gradual development. Applying the attributes choosing procedure, there occurs a strong enhancement in the normal foretelling arrangement precision action outcomes, simultaneously requiring time to calculate the foretelling precision from the reduced total dataset due to the existence of smaller number of handling inconsistencies remaining in the dataset has given by Kadir (2015).

Literature review: Divya *et al.* (2015) classification of plant leaves was done using various morphological studies. There are various types of methods to classify which are tree induction, k-nearest neighbour classifier, unnatural neural system, fuzzy logic probabilistic neural system etc. A general appraisal of various classifying besides feature mining procedures in image handling was put in place.

A researcher incorporated contour and strain besides colour and quality attributes for classifying a leaf and the classifying agent used Probabilistic Neural Network (PNN) (Kadir *et al*., 2013). The outcome of the investigation revealed that the procedure for classifying yielded normal precision of 93.75% on testing with Flavia dataset consisting of 32 types of plant leaves, concluding that the procedure used yielded superior performance in comparison to the original task.

While classifying plants, the shape of the leaves plays a prominent role. The greatest prominent part vital for deciding besides processing of data is the recognition of shape. A researcher used a feed-forward neural systems to mechanise for recognising leave for classifying the plants (Sumathi and Kumar, 2013). The precision of classification of intended procedure-Normalized Cubic Spline Feed Forward Neural Network (NCS-FFNN) was equated with RBF, CART and MLP.

Chaki and Parekh (2011) yet, another researcher suggested a mechanised system to identify plant species relied on leaf pictures. The images of plant leaves conforming to three types of plants were analysed applying two different procedures for shape modelling, the first one on Moments-Invariant (M-I) Model and the next one on the Centroid-Radii (C-R) Model. The first four normalized central moments for the M-I Model were studied in multiple groupings separately in joint 2D and 3D feature spaces to yield optimal outputs. An edge detector is used for C-R Model to recognise the periphery of the leaf shape with by using 36 radii at 10° angular parting to build the feature vector. For improving the precision further, a crossbreed set of attributes involving the M-I together with the C-R Models was created and examined to determine if the grouping attribute vector can proceed to improve action. For discriminating classifiers, neural systems were used. The data set comprised of 180 pictures divided is into 3 classes of 60 pictures each when precision ranging from 90-100% were obtained by Tzionas *et al.* (2005) a clearly defined geometrical besides morphological attributes from plant leaves is presented in the proposal and execution of an unnatural visualization structure. A subset of significant picture attributes were recognised with the help of a novel approach of choosing the attributes which diminishes the measurability of the feature space taking forward to a comprehensible classification arrangement suitable for real time classification submissions.

Sumathi and Kumar (2013) suggested a procedure to classify pictures of leaves by taking the advantage of the concept of gaining data and discovered the effectiveness of education algorithms of MLP to classify plant leaf. The outcome will be enhancement in data gain procedure for MLP with batch reverse transmission algorithm-based knowledge acquiring computing efficiency by enhancing the precision of classification. The suggested algorithms outperformed MLP with additional orientation besides getting educated at Levenberg-Marquardt level to classify plant leaf on testing with 9 species. The assessment demonstrated that gaining data helps in handpicked attributes resulting in specific MLP enhancements with batch reverse transmission algorithm classifying action having 94.81% accuracy.

The leaf image surface attributes have been removed through Gabor-based procedures and exposed them to PSO-CFS-based quest procedure to identify the best set of attributes from the whole set of features and classify them applying four classification algorithms KNN, J48, CART and RF (Kumar *et al*., 2016; Mehtre *et al*., 1997). The other aim of this task is to exploit the two faces in place on the plant leaves-dorsal and ventral. The dorsal is used to classify plants on the lines of digital leaf pictures. Both dorsal and ventral phase to analyse the effects on classify precision values for leaf images.

An optimum methodology to extract feature and choosing leave's classification based on Genetic Algorithm (GA) was suggested wherein selecting optimum attribute's subset besides the classification became a significant procedure in classifying leaves (Narayan and Subbarayan, 2014). The causative attribute arrangement is taken off the leaf pictures applying GA procedure and these extracted attributes were put in place for training the Support Vector Machine (SVM). To optimize the features of colour and boundary sequences, GA was used to enhance the total overview action based on the matching precision. SVM was applied to obtain the false positive and negative attributes the outcome of which was applying GA for choosing attribute subset using SVM as classifying agent-established computing efficiency besides improving the precision in comparison to KNN for leaf pattern classifications.

Valliammal and Geethalakshmi (2012), to classify the leaves based on GA and Kernel Based Principle Component Analysis (KPCA), two researchers explained an optimum attitude to select subset attributes. In view of high intricacy in selecting optimum attributes, the classification turned out to be a crucial research to analyse the leaf picture data. Originally, the shape, surface and colour attributes were removed from the leaf pictures and subjected to optimization through the isolated working of GA and KPCA. This approach functions as a crossing process over the subsets extracted from the optimization method. In the end, the most collective corresponding subset is taken forward for SVM training. The experiment's results successfully established that the GA and KPCA applications to select attribute subset using SVM as classifying agent to possess computing efficiency besides enhancing the precision in classifying.

MATERIALS AND METHODS

Dataset: The 9 species with 20 samples each of 197 leaves of similar structures viz., mango, sappota, guava, neem and cotton are used.

Feature extraction: When the initial object is submitted to a specific set of affine formal alterations or an indiscriminate amalgamation of the same, the shape or appearance of the object in a feature should remain unaltered. A 2D shape describing device should not be sensitive to translation, scale changes (uniform in both the X and Y-coordinates) besides rotations. This principle for describing shape denotes that the describing device is capable of carrying out regularization for different looks the object may occur in. The real representation of the object must be able to map a similar model of the pattern.

Chain coded string: It is also called Freeman code, the chain code can be put in place to represent the periphery of any shape. The periphery can be outlined either in clockwise or opposite means with 8 codes for each pixel allocated in accordance with the direction of the next pixel in respect to the current one. A picture is an object having periphery and its periphery is exemplified by chain codes which are strings to describe the shapes. To match a pair of picture peripheries, string illustrations are matched by using string remoteness processes.

Group Search Optimizer (GSO): A Producer-Scrounger (PS) Model is employed by the GSO algorithm as a context. Two foraging approaches within the groups exist-producing, look out for food, freeloading with others uncovering the bonding means. GSO's objective was to control the uninterrupted optimization constraints with inspirations from this model. Population is a group with each individual as a member in GSO. In a group, there are three kinds of members-the best location unearthed by group is referred to as producer with certain individuals being referred to as scroungers going to producer with the aim to locate a better area while others are called as rangers walking randomly for exploring a new search space to escape from local minima set ups.

General animal searching behaviour together with a standard social hunting model such as the PS Model guides the Group Search Optimizer (GSO) with the set up mainly pursuing the Producer-Scrounger (PS) Model which describes the group members either go in search of producer or opportunities to join scroungers. Subject to this model, animal scanning technique concepts are made use of metaphorically to put in place an optimal searching mechanism to solve optimization constraints in incessant dominion (He *et al*., 2009) Saunders.

The population of the GSO algorithm is called a group with each individual therein is referred to as a member. In an n-dimensional search space, the ith member at the kth searching spell (iteration) has a current position $X_i^k \in \mathbb{R}^n$ a head angle $\varphi_i^k = (\varphi_i^k, ..., \varphi_{i(n-1)}^k) \in \mathbb{R}^{n-1}$. The direction of search of the ith member which is a unit vector $D_i^k(\varphi_{i,1}^k, ..., d_{i_n}^k) \in \mathbb{R}^n$ can be calculated from ϕ_i^k via. a polar to Cartesian synchronize conversion as in Eq. 1:

$$
\begin{aligned}
d_{i_1}^k &= \prod_{q=1}^{n-1} \cos(\varphi_{i_q}^k) \\
d_{i_j}^k &= \sin(\varphi_{i_{(j-1)}}^k) \cdot \prod_{q=1}^{n-1} \cos(\varphi_{i_q}^k) (j = 2, ..., n\text{-}1) \\
d_{i_n}^k &= \sin(\varphi_{i_{(n-1)}}^k)
\end{aligned}
\tag{1}
$$

In GSO, a group consists of 3 types of members-producers, scroungers and dispersed members going in random motions whose conducts pursue the PS model. For computing convenience, we suppose a shortened PS Model that there prevails only one producer at each reiteration with the rest of the members being

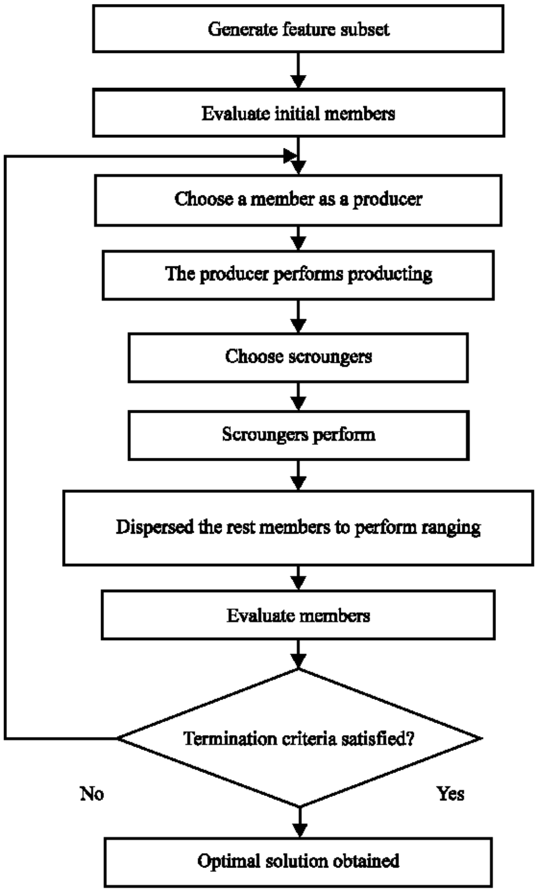

Fig. 1: Flowchart for proposed methodology feature subset

scroungers and dispersed members as mentioned by Roy *et al.* (2013). The naivest joining policy is applied which suppose all scroungers will join the producer-found resource. The flowchart for proposed methodology feature subset is shown in Fig. 1.

In GSO based feature selection, the members represent the candidate solutions in search optimizer. The member is generated by distributing 0 and 1 sec randomly. For every member if rest members is 1 then it is selected and if 0 then it is ignored.

Classifier

Fuzzy classifier: The if-then rules possessed Fussy Classifiers (FCs) that are related to the thinking of human beings which is their main plus points over neural systems. Identifying FCs calls for finding the sufficient configuration and considerations. The configuration discerning comprises various tasks like choosing sufficient inconsistencies, allocating the sufficient quantum of fuzzy sets to each adjustable and describing the number of fuzzy rules applied besides specifying the constraints of fuzzy sets.

Fuzzy classification rules comprise fuzzy sets in the ancestral besides a class label in the sequence. Denoting the data set with D data points and n variables as Z = [Xy], the input matrix X and output vector y are given as in Eq. 2:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots x_{1,n} \\ \dots & \dots & \dots \\ x_{D,1} & x_{D,1} & \dots x_{D,n} \end{bmatrix}, \ y = \begin{bmatrix} y_1 \\ \dots \\ y_D \end{bmatrix} \quad (2)$$

Fuzzy classifuication can be performed as follows:

R_i: If x_1 is $A_{i,1}$, ... and x_n is $A_{i,n}$ then g_i, i = 1,...,R

Where:
R = The number of Rules ($A_{i,j}$)
j = 1, ..., n = A membership function
$g_i \epsilon \{1, ..., C\}$ = The rule consequent and C is the number of different classes in data set

For each data point x_k the degree of fullfilment of a rule is computed as in Eq. 3:

$$\beta_i(x_k) = \prod_{j=1}^{n} A_{i,j}(x_{k,j}) \quad (3)\backslash$$

The rule possessing the highest magnitude of fulfillment is declared as the winner rule, i.e., the winner takes the entire strategy. The classiffer's output is the rule consequently associated with that rule. Besides, there are other types of fuzzy rules and t-norms which are used for thinking and the properties of fuzzy classiffers are discussed in detail given by Pulkkinen and Koivisto (2008).

Multi-Layer Perceptron Neural Network (MLPNN) with back propagation training: MLP have its place on model of feed-forward artificial neural system that draws a set of contributed data onto a set of its suitable yields. Multiple strata of nodes in a fixed graph with each stratum completely linked to the next one are included in MLP. Barring the contributed nodes, each node is reflected as a handling component or a neuron with a straight line initiation task. Generally, MLP avails reverse transmission for training the system. Education takes place in the perceptron by altering the connection weights once every piece of data gets processed depending on the quantum error taking place in the yield while comparing with the anticipated outcome. This specimen of monitored education can be performed through the reverse transmission in other words, generalising the smallest average squares algorithm in the direct perceptron.

A data handling method based on the method of biological nervous system such as the brain is called as Neural Networks (NN), the most common of which is the Multilayer Perceptron (MLP). Such a neural system is called monitored system, since, it requires a yield to get educated. The aim of this system is to create a model to produce the required unfamiliar yield has given by Burger *et al*. (2012). The MLP and many other NNs get educated through an algorithm known as reverse transmission which helps the contributed data that is frequently conveyed to NN. With every submission, the yield of NN is compared to the required yield with an error getting computed. This is then returned to NN with weights getting adjusted to enable diminish the error with every reiteration besides neural model's access getting closer for a desired yield and it is as called training.

Multi-Layer Perceptron Neural Network (MLP NN) with Levenberg-Marquardt (LM) training: Levenberg-Marquardt (LM) algorithm is widely held as an effective algorithm to enable a numerical solution to research and minimise issues. It chiefly presents in finding solution to an Eq. 4:

$$(JtJ + \lambda I)\delta = JtE \tag{4}$$

Where:

J = Jacobian medium for system
λ = Levenberg's checking aspect
δ = Mass fill-in vector that is to be determined with
E = Error vector and yield errors for every contributed vector used on training network

The δ tell us the magnitude of change the network weights should be subjected for achieving a probably superior answer. JtJ matrix is alternatively called the approximated Hessian. The λ checking aspect is attuned at every reiteration to guide the effectiveness. When E lessening is faster, a smaller value is used to get the algorithm closer to Gauss-Newton algorithm. If reiteration gives inadequate lessening in residual, λ is increased to bring it closer to incline lineage course. Fluctuations in algorithm may comprise contradictory values for v, one for decreasing λ and the other for enhancing it has given by Lourakis (2005). Certain merits of LMA includes the capacity to get educated with the LMA and is qualified to be superior, LMA has speedy conjunction merits with the LMA suiting medium-sized datasets besides being the fastest among orientation algorithms.

RESULTS AND DISCUSSION

The 9 spices with each 20 samples are considered for experiments. The features extracted are used to train the classification algorithms MATLAB is used. The features are classified using fuzzy classifier, MLPNN-BP and MLPNN-LM. Figure 2-5 show the classification accuracy, precision, recall and F-measure, respectively.

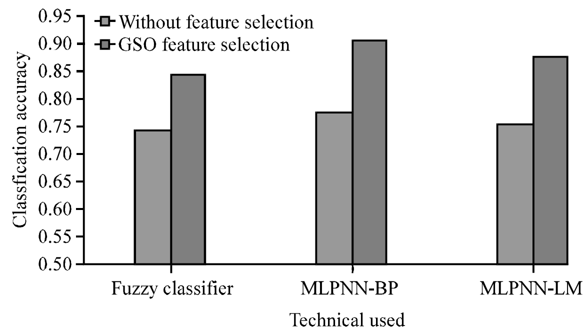

Fig. 2: Classification accuracy for GSO feature selection

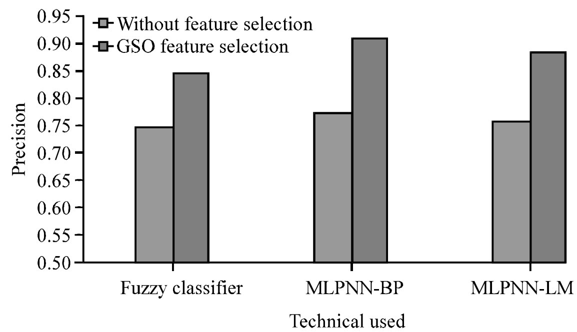

Fig. 3: Precision for GSO feature selection

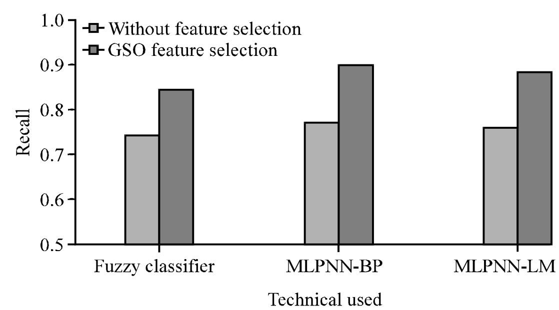

Fig. 4: Recall for GSO feature selection

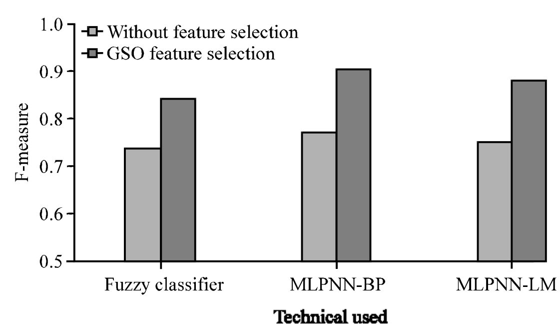

Fig. 5: F-measure for GSO feature selection

From Fig. 2, it is observed that the classification accuracy of proposed GSO feature selection performs better by 12.6, 15.9 and 14.9% than without feature selection for fuzzy classifier, MLPNN-BP and MLPNN-LM, respectively.

From Fig. 3, it is observed that the precision of proposed GSO feature selection performs better by 12.67, 16.2 and 15.4% than without feature selection for fuzzy classifier, MLPNN-BP and MLPNN-LM, respectively.

From Fig. 4, it is observed that the recall of proposed GSO feature selection performs better by 12.6, 15.3 and 14.96% than without feature selection for fuzzy classifier, MLPNN-BP and MLPNN-LM, respectively.

From Fig. 5, it is observed that the F-measure of proposed GSO feature selection performs better by 12.9, 15.9 and 15.4 than without feature selection for fuzzy classifier, MLPNN-BP and MLPNN-LM, respectively.

CONCLUSION

Plant identification systems have been performed by several researchers. A neural network was designed for the classiffcation of the available samples, taking as inputs the features selected by the fuzzy surface model. Neural networks are well-known for their generalization capabilities in classiffcation problems. The proposed feature selection approach results in simpler, faster and easier to train neural network architectures when compared to neural networks used to measure the contribution of individual input features to the output of the neural network. Results show that the classification accuracy of proposed GSO feature selection performs better by 12.6, 15.9 and 14.9% than without feature selection for fuzzy classifier, MLPNN-BP and MLPNN-LM, respectively.

REFERENCES

Burger, H.C., C.J. Schuler and S. Harmeling, 2012. Image denoising with multi-layer perceptrons, part 1: Comparison with existing algorithms and with bounds. J. Mach. Learn. Res., 10: 1-38.

Chaki, J. and R. Parekh, 2011. Plant leaf recognition using shape based features and neural network classifiers. Int. J. Adv. Comput. Sci. Appl., 2: 41-47.

Divya, T., A.C. Korwar and Virupakshappa, 2015. A survey on recognition of plant leaves using Vario us classifiers and various morphological features. Intl. J. Adv. Found. Res. Comput., 2: 89-95.

He, S., Q.H. Wu and J.R. Saunders, 2009. Group search optimizer: An optimization algorithm inspired by animal searching behavior. IEEE Trans. Evolut. Comput., 13: 973-990.

Kadir, A., 2015. Leaf identification using Fourier descriptors and other shape features. Gate Comput. Vision Pattern Recognition, 1: 3-7.

Kadir, A., L.E. Nugroho, A. Susanto and P.I. Santosa, 2013. Leaf classification using shape, color and texture features. Intl. J. Comput. Trends Technol., 1: 225-230.

Kumar, A., V. Patidar, D. Khazanchi and P. Saini, 2016. Optimizing feature selection using particle swarm optimization and utilizing ventral sides of leaves for plant leaf classification. Procedia Comput. Sci., 89: 324-332.

Lourakis, M.I., 2005. A brief description of the Levenberg-Marquardt algorithm implemented by levmar. Found. Res. Technol., 4: 1-6.

Mehtre, B.M., M.S. Kankanhalli and W.F. Lee, 1997. Shape measures for content based image retrieval: A comparison. Inf. Process. Manage., 33: 319-337.

Narayan, V. and G. Subbarayan, 2014. An optimal feature subset selection using GA for leaf classification. Intl. Arab J. Inf. Technol., 11: 447-451.

Pulkkinen, P. and H. Koivisto, 2008. Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. Int. J. Approximate Reasoning, 48: 526-543.

Roy, S., S.M. Islam, S. Das and S. Ghosh, 2013. Multimodal optimization by artificial weed colonies enhanced with localized group search optimizers. Appl. Soft Comput., 13: 27-46.

Sumathi, C.S. and A.S. Kumar, 2013. Plant leaf classification using soft computing techniques. Intl. J. Future Comput. Commun., 2: 196-199.

Tzionas, P., S.E. Papadakis and D. Manolakis, 2005. Plant leaves classification based on morphological features and a fuzzy surface selection technique. Proceedings of the 5th International Conference on Technology and Automation (ICTA'05), October 15-16, 2005, Alexander Technological Educational Institute of Thessaloniki, Thessaloniki, Greece, pp: 365-370.

Valliammal, N. and S.N. Geethalakshmi, 2012. An optimal feature subset selection for leaf analysis. Intl. J. Comput. Commun. Eng., 1: 1-6.