International Journal of Control Theory and Applications

ISSN : 0974-5572

Special Issue, 2016

Enhancing Leaf Classification with Feature Selection

A.M. Ravishankkar[a] and M. Mohanapriya[b]

[a]Research Scholar, Department of CSE,Karpagam University,Coimbatore,Tamilnadu, India

E-mail: ravi662shankkar@gmail.com

[b]Professor, HOD/CSE, Karpagam University, Coimbatore, Tamilnadu, India

Abstract: Feature selection techniques are essential within classification methods for improving accuracy. Leaf images comprising noise because of imaging equipment, operational environments or even positions of images at the time of acquiring images.

Method: In the current study, a technique for classifying leaf images through exploitation of the notion of Minimum Redundancy Maximum Relevance (*m*RMR), chi square as well as looks into the efficiency of learning protocols such as Multi-Layer Perceptron (MLP) for the classification of plant leaves.

Findings: The study reveals that the features selection technique for MLP-NN with back propagation (MLPNN-BP) protocol based leaning improves computational efficacy through enhancement of classification accuracies. It is noted that the suggested method performs better than MLP with incremental training as well as Levenberg Marquardt (LM) based learning for plant leaf classification through evaluation with nine specie.

Results: Experiments prove that features selection with mRMR attains greater classification accuracy for fuzzy classifiers, MLPNN-BP as well as MLPNN-LM when contrasted with features selection utilizing chi square.

Keywords: *Feature Selection, Plant Leaf Classification, Multi-Layer Perceptron (MLP) and Neural Network (NN).*

1. INTRODUCTION

Plants have a critical part to play in the ecology of Earth as they provide sustenance, shelter as well as maintenance of healthy, breathable atmosphere. Constructing a plant database for effective classification as well as recognition is important for the conservation as well as preservation of the plants. This is particularly true as several plant species are almost at the edge of extinction because of constant deforestation for paving the way to urbanization. In recent times, computer vision as well as pattern recognition methods have been used for preparing digital plant catalogue systems for the recognition of plants in effective ways [1].

Various attempts have been carried out for the classification of plants based on flowers, arrangement of leaves, shape, colour and even texture. These studies are necessary for maintaining the ecological balance because few of the plants are on the edge of extinction. For normal humans, characteristic attributes of digital images are essentially textures, shapes or colour. However, for computer systems, there ought to computer recognizable attribute set that can be stored, fine-tuned and analysed for adequate classification [2]. Few typical methods for classifying plant leaves through usage of digital images have their basis in geometrical characteristics, textural as well as shape based attributes and some colour-based attributes.

The attribute is defined as the function of one or more metrics, all of which specify certain quantifiable properties of objects and are calculated so that they quantify certain important properties of the entity. It classifies several attributes currently employed thus [3]

General features: Application independent attributes like colour, texture or shape. As per abstraction levels, they may be further categorized as pixel-level features, local features or global features.

Domain-specific features: Application dependent attributes like human face, fingerprint or conceptual attributes. The attributes are typically a synthesis of low-level attributes for a particular field.

Every feature may also be broadly sorted as either low-level or high-level attributes. The former may be extracted directly from the original image while the latter are to be based on the low-level attributes.

Features extraction refers to the procedure of generation of attributes to be utilized in the selection as well as classification tasks. Features selection has an important role to play in various pattern recognition issues like image classification. Although several attributes may be used for categorizing images, solely certain quantities of them are very effective in classifications. Greater quantity of features does not always results in improved classification performance, and hence, features selection is carried out for selecting compacted as well as relevant features sub set for reducing dimensionality of features space that gradually enhances classification accuracy as well as reduces time consumed [4]

On the basis of various evaluatory criteria, features selection technique may be sorted into 2 groups: filters as well as wrappers. The former typically uses attributes of feature data and are operationally effective. When contrasted with filters, wrappers typically attain greater classification accuracies. The superiority is attained through involvement of classifiers in the selection stage.

The initial step for classification of plant leaves is the acquisition of images. This involves plucking of leaves from plants and then digital colour images of leaves are obtained with digital cameras. After the images are got, certain amount of pre-processing is required. This involves greyscale conversion, image segmentation, binary conversion as well as image smoothing. The goal of image pre-processing is the improvement of image data such that it is capable of suppressing non-desirable distortions and improves image attributes which have relevance for further processing. Colour images of leaves are modified to greyscale images. Several alterations in the atmosphere as well as seasons make the colour attributes very unreliable. Hence, it is better to function with greyscale images. When images are transformed to greyscale, they are segmented from their background and transformed to binary. Utilizing one of the edge detectors, the contour image is identified. Then few morphological attributes are taken from the contour image. The features vector is then provided to the classifiers [5]

Classification issue handles the association of particular input pattern with a distinct class. Patterns are specified by a set of attributes such that it is normal to consider them as d-dimensional vectors, wherein d represents the quantity of various attributes. The abstraction leads to the notion of features space. Patterns are points in the d-dimensional space while classes are subspaces. Classifiers assign a single class to every point in the input space. The issue of classification fundamentally initializes a modification between features as well as classes. Optimum classifiers are those anticipated to yield minimal quantity of misclassifications.

In this paper, the feature selection based plant leaf classifications are proposed. Section 2 details the dataset and the methods used in this investigation. Section 3 presents the results and section 4 concludes the paper.

[6] suggested a novel technique of describing the features of plant leaves on the basis of outline as well as venation fractal dimensions. Experiments proved the efficacy of the latter. [7] extracted distinct attributes from the images of plant leaves and decreased probability of disruption through occlusion, clutters or noises. A new features extraction protocol on the basis of dual-scale decomposition as well as local binary descriptors was suggested. Outcomes of experiments reveal that the suggested method provides improved performance with regard to classification accuracy when contrasted with other techniques.

[8] delineated based an android-based mobile application formulated for the automatic identification of plant species through tree leaf photos. In the application, a single leaf image may be either digital images got from an already present image dataset or photo obtained from cameras. The identification procedure comprises 3 stages: leaf image segmentation, features extraction as well as species identification. The system functions well with excellent identification performance.

[9] suggested a new technique of classifying plants through usage of their leaves. Most plant species possess singular leaves that are distinct from one another in shape or textural attributes. The authors suggested a technique of structural decomposition of edges that extracts structural signatures as well as quantifies the attributes that are not dependent on leaf size or orientation. The protocol has been evaluated on images of forty plant leaves from ten distinct species and it presents accuracy of 67.5%.

[10] suggested an Enhanced Fuzzy Min-Max (EFMM) network for pattern classification. The goal was the overcoming of various restrictions of the original FMM as well as the enhancement of the classification performance. The major contribution was 3 heuristic rules for enhancing the learning protocol of FMM. Efficiency of EFMM was tested through benchmark datasets as well as actual medical diagnosis tasks. The outcomes outperformed those from several FMM-based models, SVM-based, Bayesian-based, decision tree-based, fuzzy-based, as well as neural-based classifiers.

[11] suggested a method for classifying biological images via Rough-Fuzzy Artificial Neural Network (RFANN). The method was utilized for improving learning procedure through Rough Sets Theory (RS) with a focus on features selection, taking into consideration the fact that RS features selection permits the usage of low dimension attributes from the image dataset. For measuring the performance of the suggested RFANN, run times as well as training errors were contrasted with the non-reduced features.

2. METHODOLOGY

In this section, feature extraction using wavelet based texture features, feature selection using MRMR and chi square, and classifiers using fuzzy, MLPNN BP and MLPNN LM are described.

2.1. Wavelet based Texture Features

Wavelets are waveforms of restricted durations that possess average value of 0. Wavelets are neither regular nor symmetric. They have differing frequencies. Wavelet analysis may be employed to 1D data (signals) as well as 2D data (images). The primary reason as well as benefit of employing wavelet transform for detecting edges in images is the potential for selecting the size of image details which will be identified. When processing 2D images, wavelet analyses are carried out distinctly for horizontal as well as vertical directions. Hence, vertical as well as horizontal edges are identified in a separate manner [12].

2D Discrete Wavelet Transform (DWT) splits images into sub images, three details as well as one approximation. The approximation is like the given image but merely one-fourth of the original size. 2D DWT is an expansion of 1D DWT in both horizontal as well as vertical directions. The resultant sub images from an octave are labelled as A, H, V as well as D, as per the filters utilized for generating the sub image. The procedure is iterated through placing first octave's A sub image over one more set of low as well as high pass filters. The iterations build the multi resolution analysis.

Texture is a significant as cue for analysing various kinds of images. The term is utilized for pointing the intrinsic characteristics of surfaces, particularly those which do not possess smoothly varying intensities. It includes intuitive characteristics such as roughness, granulation as well as regularity. Texture may be defined as a set of local neighbourhood characteristics of grey levels of image area. Textural analyses are regarded as a problematic task. The capacity for effective classification as well as segmentation of images on the basis of textural attributes is important in scene analysis, medical image analysis, remote sensing as well as other application domains [13].

A significant issue in wavelet textural analyses is that the quantity of attributes has a tendency to be huge, particularly for wavelet packet decomposition. A huge quantity of features, though they might possess greater amount of information, ensure that classifications as well as segmentations are harder. The phenomenon is famous in pattern recognition as the curse of dimensionality. There exists a typical features reduction technique for handling this. A basic issue is that the pre-dominant scales which possess the most useful data differ from one texture to another. It is beneficial to restrict the quantity of attributes at the level of their generation wherein the nature of the attributes may be taken into consideration.

2.2. Minimum Redundancy Maximum Relevance (MRMR) Feature Selection

MRMR is a filter based feature selection protocol that attempts to choose the most relevant attributes with target class labels as well as decrease redundancies amongst the chosen attributes concurrently, where the protocol utilizes Mutual Information I(X, Y) which assesses the level of similitude between 2 discrete arbitrary parameters X as well as Y [14].

$$\mathrm{I(X, Y)} = \sum_{x \in \mathrm{X}} \sum_{y \in \mathrm{Y}} p(x, y) \log\left(\frac{p(x, y)}{p_1(x) p_2(y)}\right)$$

Wherein $p(x, y)$ refers to the joint probability distribution function of X as well as Y, while $p_1(x)$ as well as $p_2(y)$ refer to the marginal probability distribution functions of X as well as Y correspondingly.

Information theoretic ranking conditions consider nonlinear relations between features as well as targets, but they assess attributes in an independent manner and are not able to handle features redundancies issue. For addressing the problem, for exploring *m*RMR technique that focuses on choosing optimum attributes for classification to give optimal solutions. For features set S with $n0$ attributes $\{xi\}$, $(i = 1 \ldots n0)$. Maximal relevance is to look for attributes so that MI values between individual features as well as targets are to be made maximum. Assume D(S, y) is the mean of MI between individual attributes as well as target y. It is given by [15].

$$\max \mathrm{D}(\mathrm{S}, y) = \frac{1}{|\mathrm{S}|} \sum_{x_i \in \mathrm{S}} \mathrm{I}(x_i, y)$$

Though 2 attributes might possess extreme separate ability on target class, it is not desirable to add them if they are extremely correlated as well. The notion of minimal redundancies is to choose attribute so that they are mutually maximally dissimilar. Assume R(S, y) is the mean of MI between pairs of features in S. It is given by,

$$\min \mathrm{R(S)} = \frac{1}{|\mathrm{S}|^2} \sum_{x_i, x_j \in S} \mathrm{I}(x_i, x_j)$$

The condition fusing the above 2 restrictions is known as MRMR. MRMR features set is acquired through maximization of D(S, y) as well as minimization of R(S) concurrently that needs combination of the 2 metrics into one criterion function.

As features selection is a non-deterministic polynomial time (NP)-hard issue, heuristics are to be utilized for finding adequate as well as suboptimal sets of relevant attributes in higher dimensional data sets. Amongst the heuristics, MRMR features selection method is especially advantageous as it has comparatively lesser computational complexity of the protocol for discovering a set of relevant as well as complementary attributes, from which correct predictive models are formulated. The problem is that MRMR similar to every other features selection protocol in a low sample-to-dimensionality ratio scenario, yields extremely varying outcomes and minor alterations to sample data leads to drastically varying sets of chosen attributes.

2.3. Chi Square (χ^2) Feature Selection

Another common features selection technique is χ^2. In statistics, χ^2 test is employed for testing independence of 2 events, wherein 2 events A as well as B are considered independent P(AB) = P(A)P(B) or P(A | B) = P(A) as well as P(B | A) = P(B). In features selection, the 2 events are occurrence of term as well as occurrence of class. Terms are ranked with regard to the quantity given below:

$$X^2(D, t, c) \;=\; \sum_{e_t \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{(N_{e_t e_c} - E_{e_t e_c})^2}{E_{e_t e_c}}$$

Where *i* e_t as well as e_c are given by above Equation. N represents *observed* frequency in D while E represents *expected* frequency. For instance, E_{11} represents expected frequency of t and occurring together in a document presuming term as well as class which independent with one another are not dependent on one another.

2.4. Fuzzy Classifier

Fuzzy logic denotes an excellent method for taking decisions. Fuzzy logic was originally developed and presented in 1965 by Zadeh. Since then, several studies have been performed for evaluating its employment in several regions of digital image processing like image quality measurement, edges detection, images segmentation and so on. Fuzzy image processing is the set of all methods which comprehend, denote as well as process the images, segments as well as attributes as fuzzy sets. The abstraction as well as processing relies on the chosen fuzzy techniques as well as on the issue to be resolved. Fuzzy image processing has 3 primary phases: 1) Image fuzzification, 2) Alteration of membership values, and, if required, 3) Image defuzzification.

Fuzzification as well as defuzzification stages do not have fuzzy hardware. Hence, coding of image data as well as the decoding of outputs are stages which enable processing of images with fuzzy methods. Fuzzy image processing has several benefits such as: 1) They are powerful methods for representing as well as processing knowledge, 2) They are capable of managing vagueness as well as ambiguity in an efficient manner, 3) Fuzzy logic has tolerance to inaccurate data and 4) Fuzzy logic is simple to comprehend. The mathematical formulations behind fuzzy reasoning are extremely simple. Fuzzy logic is excellent because of its 'natural' character and not its extensive complexity.

Fuzzy classifiers are systems which accept various input features vectors or fuzzy truths that are a part of fuzzy set membership functions. They output code words which provide the class to which features vectors are a part of or else output fuzzy values at k^{th} output node for designating the fuzzy truths of k^{th} class. Typically individual output components are fuzzy truths. The condition is that fuzzy truth is utilized in the procedure of taking decisions [16].

For all pixels in the image, there is an 8D features vector (X1, X2, ..., X8) which comprises grey level differences on their 3x3 neighbourhoods. Fuzzy classifiers operate on the features vectors for determining if the pixel has an edge or not through provision of fuzzy truths for 2 classes of pixels which are edge or non-edge pixels. The two classes are correspondingly mapped to black or white for centre pixel in the new output image. Hence, all image pixels are mapped to black or white in the output image, which is a line drawing image of black lines on white background.

Assume that the pattern classification issue is an nD issue with M classes as well as m specified training patterns $x_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$. With no loss of generality, it is assumed that all attributes of the specified training patterns are normalized into [0, 1]; *i.e.*, the pattern space is an nD unit hypercube $[0, 1]^n$. Fuzzy if-then rules of the kind given below are utilized as the base of the fuzzy rule-based classification systems in the study [17]:

Rule R_j : If x_1 is A_{j1} and ... and x_n is A_{jn}

then Class C_j with CF_j, $\quad j = 1,2, .., N$

Wherein R_j represents label of j^{th} rule, $A_{j1}, \ldots A_{jn}$ represent antecedent fuzzy sets in [0, 1], C_j represents consequent class while CF_j represents grade of certainty of the fuzzy if–then rule R_j.

Fuzzy rule-based classification systems comprise N fuzzy if–then rules. 2 major stages are present in generating fuzzy if-then rules: specification of antecedent part, as well as determination of consequent class C_j as well as grade of certainty CF_j. The antecedent component of the rules is set in a manual fashion. Then consequent part is determined from the specified training patterns. The usage of grade of certainty in fuzzy if–then rules permits the generation of understandable classification systems with excellent classification performance.

2.5. Multi-Layer Perceptron Neural Network with Back Propagation Training (MLPNN-BP)

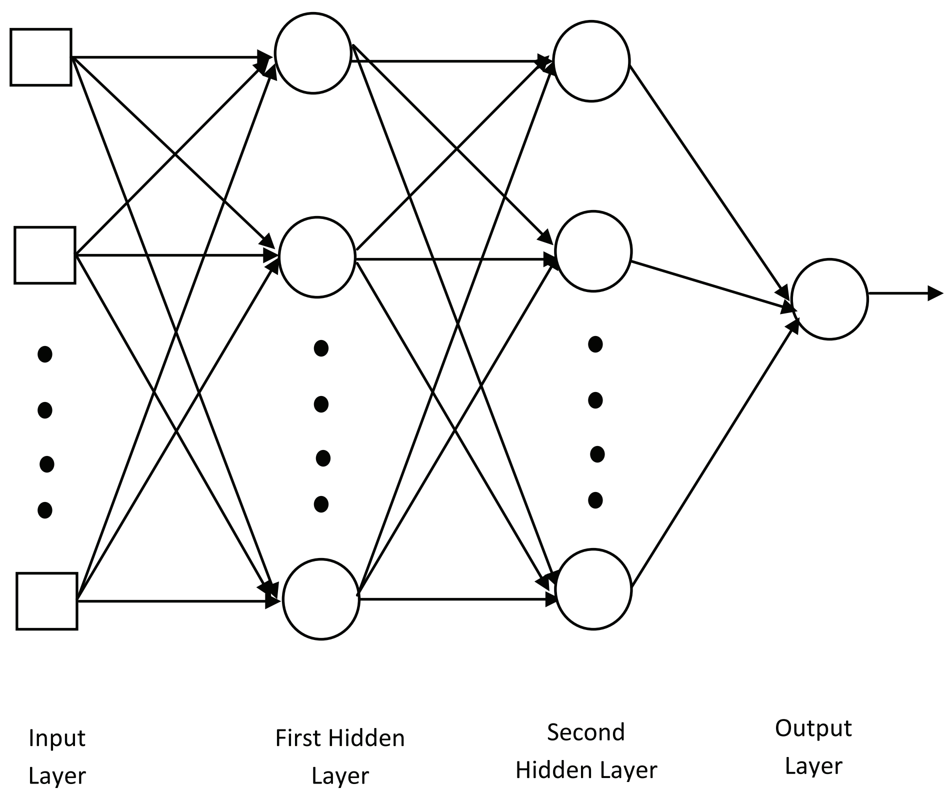

Figure 1: Architecture of an MLP NN

Artificial Neural Networks are information processing paradigms that owe their inspiration to the manner in which biological nervous systems like the brain, process information. The major component of the paradigm is the new structure of the information processing systems. It comprises huge quantities of extremely inter-connected processing elements working together for solving particular issues [18]. The most typical NN model is the MLP. This kind of NN is called supervised network as it needs a favoured output for learning. The aim of this kind of network is the creation of models which may then be utilized for producing outputs when favoured output is not known. Graphical representation of MLP is given in figure 1. The class of networks comprise several layers of computational units, typically inter-connected in a feed-forward manner. All neurons in a single layer have direct connections to the neurons of the next layer. In several applications, the units of the networks employ sigmoid functions as activation functions.

Feed forward BP networks undergo supervised training with finite quantity of pattern pairs comprising input patterns, as well as target patterns. Input patterns are presented at input layers. Neurons there by pass pattern activations to the subsequent layer neurons that are in hidden layers. The outputs of hidden layer neurons are acquired through usage of bias as well as threshold functions with activation defined by weight as well as input utilizing optional bias as well as threshold functions. The final outputs of networks are defined by activation from output layers.

Multilayer networks utilize several learning methods, the most common one being BP. Here, output values are contrasted with accurate answers for computing values of certain pre-determined error functions. Through several methods, errors are then fed back through the networks. Utilizing this information, the protocol alters weights of all connections for reducing values of error functions by a certain amount. After repetition of the procedure for a considerably huge quantity of training cycles, the network typically converges to a certain state wherein errors of computations are small. Then, it can be said that the network has learned a particular target function. For adjusting weights accurately, generic technique is employed for non-linear optimization known as gradient descent. For this, a derivative of error functions with regard to network weights is computed and weights are altered so that errors decrease. For this reason, BP is solely employed on networks with differentiable activation functions [19].

Features vector x is input at input layer with output denoting a discriminator between its class as well as other classes. In training, training samples are fed and predicted outputs are calculated. The output is contrasted with target output and error assessed is reverted through the network and weights modified accordingly.

The training set of size m is denoted as

$$\mathrm{TM} = \{(x_1, y_1), \ldots, (x_m, y_m)\}$$

Wherein $x_i \in \mathrm{R}^a$ represent input vectors of dimension a while $y_i \in \mathrm{R}^b$ represents output vectors of dimension b while R denotes a set of real numbers. Assume fw represents function with weight w for neural network. Supervised learning modifies weight such that:

$$f_w(x_i) = y \; ; \forall \, (x_i, y_i) \in \mathrm{T_M}$$

After NN is trained, as well as evaluated on new instances, the output is correct to a certain extent. A typical training protocol is Error Back Propagation protocol (EBP).

2.6. Multi-Layer Perceptron Neural Network with Levenberg Marquardt Training (MLPNN-LM)

EBP poor convergence rate in NN is a huge problem and great amount of effort is being expended for speeding up the protocol. Although several methods are being attempted, very small enhancements have been seen. 2nd order methods such as Newton's method, conjugate gradients or LM optimization methods attained excellent enhancement of realization performance. LM is extremely effective in realization accuracy attainment. It merges Newton protocol's speed as well as stability of steepest descent technique. LM's primary shortcomings include memory requisite for operating huge Jacobians as well as requirement for inverting huge matrices [20].

For LM, performance index to be optimized is given by

$$\mathrm{F}(w) = \sum_{p=1}^{P} \left[\sum_{k=1}^{\mathrm{K}} (d_{kp} - o_{kp})^2 \right]$$

Wherein $w = [w_1\ w_2\ ...\ w_N]^T$ comprises all network weights, d_{kp} represents favoured value of k^{th} output as well as p^{th} pattern, o_{kp} represents actual value of k^{th} output as well as p^{th} pattern, N represents quantity of weights, P represents quantity of patterns, while K represents quantity of network outputs. The equation may be reformulated thus

$$F(w) = E^T E$$

Wherein

$$E = [e_{11}\ ...\ e_{K1}\ e_{12}\ ...\ e_{K2}\ e_{1p}$$

$$e_{kp} = d_{kp} - o_{kp},$$

$$k = 1, ...\ K\ p = 1, ...\ p$$

Where in E represents cumulative error vector for patterns.

The Jacobian matrix is given by

$$J = \begin{bmatrix} \frac{\partial e_{11}}{\partial w_1} & \frac{\partial e_{11}}{\partial w_2} & \cdots & \frac{\partial e_{11}}{\partial w_N} \\ \frac{\partial e_{12}}{\partial w_1} & \frac{\partial e_{12}}{\partial w_2} & \cdots & \frac{\partial e_{12}}{\partial w_N} \\ . & . & . & . \\ . & . & . & . \\ \frac{\partial e_{K1}}{\partial w_1} & \frac{\partial e_{K1}}{\partial w_2} & \cdots & \frac{\partial e_{K1}}{\partial w_N} \\ . & . & . & . \\ . & . & . & . \\ \frac{\partial e_{1P}}{\partial w_1} & \frac{\partial e_{1P}}{\partial w_2} & \cdots & \frac{\partial e_{1P}}{\partial w_N} \\ . & . & . & . \\ . & . & . & . \\ \frac{\partial e_{KP}}{\partial w_1} & \frac{\partial e_{KP}}{\partial w_2} & \cdots & \frac{\partial e_{KP}}{\partial w_N} \end{bmatrix}$$

And weights are computed through

$$w_{t+1} = w_t - (J^T_i\ J_t + \mu_t I)^{-1}\ J^T_t\ E_t$$

Wherein I represents identity unit matrix, μ represents learning parameter, while J represents Jacobian of m output errors with regard to n weights of NN. For $\mu = 0$ it becomes Gauss-Newton technique. For huge NN, LM protocol becomes steepest decent or EBP protocol. The μ variable is mechanically modified at every cycle for securing convergence. LM protocol requires calculation of Jacobian and Jmatrix at each cycle as well as inversion $J^T J$ of square matrix, dimensions of which are N XN. This is why for huge NN LM is not practicable.

Performance index $F(w)$ to be made minimum in EBP protocol is expressed as sum of squared errors between target output as well as network's simulated output, thus:

$$F(w) = E^T E$$

Wherein $w = [w_1\ w_2\ ...\ w_N]^T$ comprises all network weights, e represents error vector containing error for every training example.

When training with LM technique, increment of weights Δw is acquired thus:

$$\Delta w = [J^T J + \mu I]^{-1}\ J^T e$$

Wherein J represents Jacobian matrix, μ represents learning rate to be updated utilizing β based on outcome. Particularly, μ is multiplied by decay rate $\beta(0 < \beta < 1)$ when $F(w)$ reduces, while μ is split by β when $F(w)$ rises in a novel step.

Although LM is regarded as effective, calculating huge Jacobians requires huge memory. Huge matrixes require inversion for calculation, leading to greater computational time. Therefore, for reducing computational costs, the alterations below are suggested in LM:

Performance index to be optimized in LM is given by

$$F(w) = \sum_{c=1}^{C}\left[\sum_{i=1}^{I}\left(d_{ic} - o_{ic}\right)\right]$$

Where in w represents network weight for every network. d_{ic} represents favoured value of i^{th} output as well as c^{th} pattern. o_{ic} represents actual value.

The performance index below is presented in LM

$$F(w) = \sum_{c=1}^{C}\left[\sum_{i=1}^{I}\left(d_{ic} - o_{ic}\right)\right]^2$$

This results in huge decrease in matrix size, thus decreasing computational costs.

3. RESULTS

Nine species of plant leaves were selected with 20 samples of gained species. Sample image of the plant leaves used is shown in figure 2. The information gain based features were extracted using Matlab and classified using fuzzy classifier, MLP NN with various learning method viz., back propagation and LM.

Figure 2: Sample image of plant leaves

The feature selection using chi square and MRMR are evaluated. Table and Figure shows classification accuracy, average precision, average recall and average f measure.

Table 1
Classification Accuracy

Classifiers used	*Feature selection using Chi Square*	*Feature selection using MRMR*
Fuzzy classifier	0.7889	0.8167
MLPNN-BP	0.8278	0.8556
MLPNN-LM	0.8056	0.8389

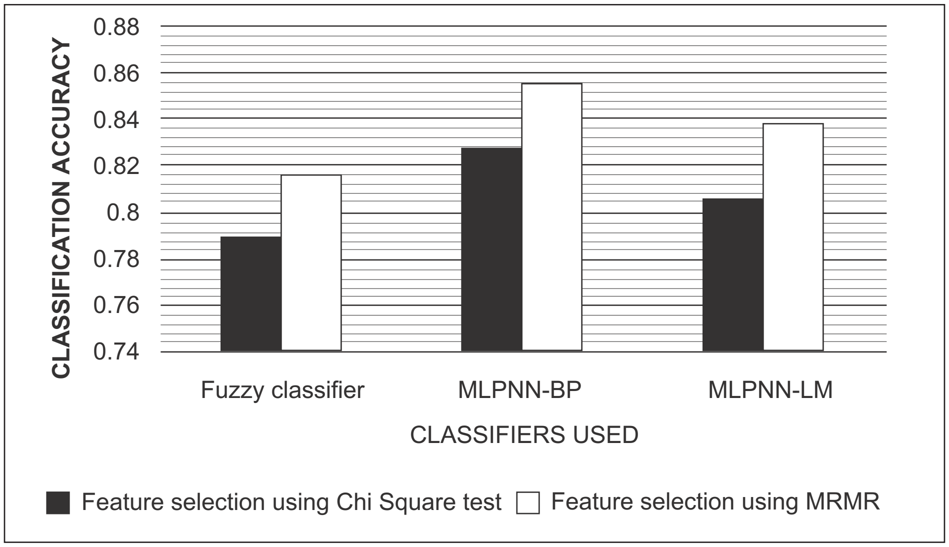

Figure 3: Classification Accuracy

From the table 1 and figure 3, it can be observed that the feature selection using MRMR has higher classification accuracy by 3.46% for fuzzy classifier, 3.3% for MLPNN-BP and by 4.04% for MLPNN-LM when compared with feature selection using chi square.

Table 2
Average Precision

Classifiers used	*Feature selection using Chi Square*	*Feature selection using MRMR*
Fuzzy classifier	0.794344	0.819567
MLPNN-BP	0.830511	0.856833
MLPNN-LM	0.8073	0.839856

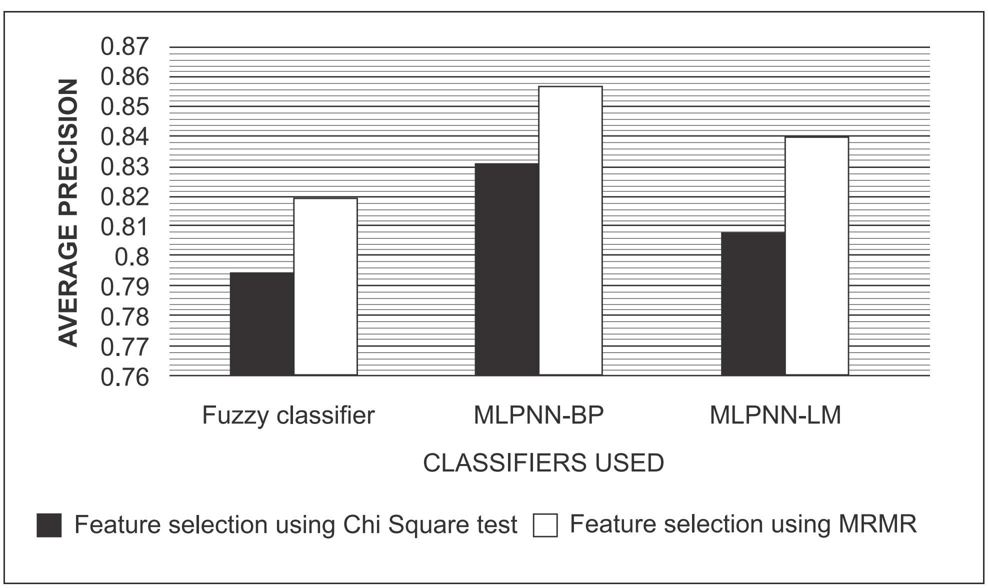

Figure 4: Average Precision

From the table 2 and figure 4, it can be observed that the feature selection using MRMR has higher average precision by 3.12% for fuzzy classifier, 3.12% for MLPNN-BP and by 3.95% for MLPNN-LM when compared with feature selection using chi square.

Table 3
Average Recall

Classifiers used	*Feature selection using Chi Square*	*Feature selection using MRMR*
Fuzzy classifier	0.788889	0.816667
MLPNN-BP	0.827778	0.855556
MLPNN-LM	0.805556	0.838889

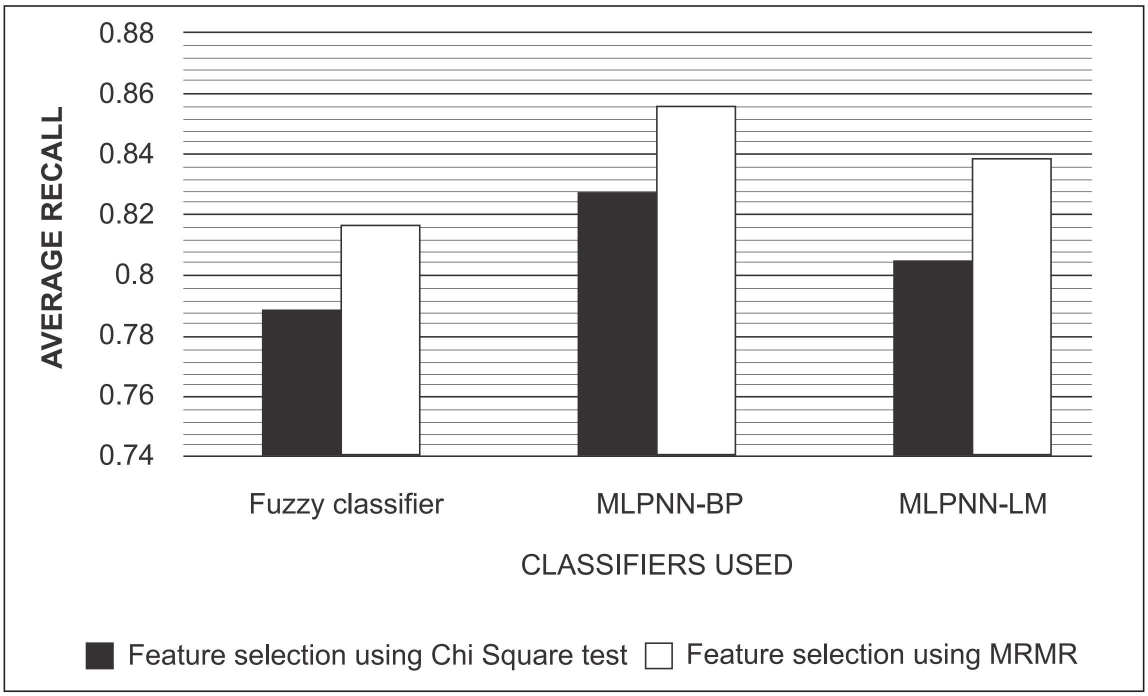

Figure 5: Average Recall

From the table 3 and figure 5, it can be observed that the feature selection using MRMR has higher average recall by 3.46% for fuzzy classifier, 3.3% for MLPNN-BP and by 4.05% for MLPNN-LM when compared with feature selection using chi square.

Table 4
Average F Measure

Classifiers used	*Feature selection using Chi Square*	*Feature selection using MRMR*
Fuzzy classifier	0.787322	0.816789
MLPNN-BP	0.827622	0.855589
MLPNN-LM	0.804711	0.838778

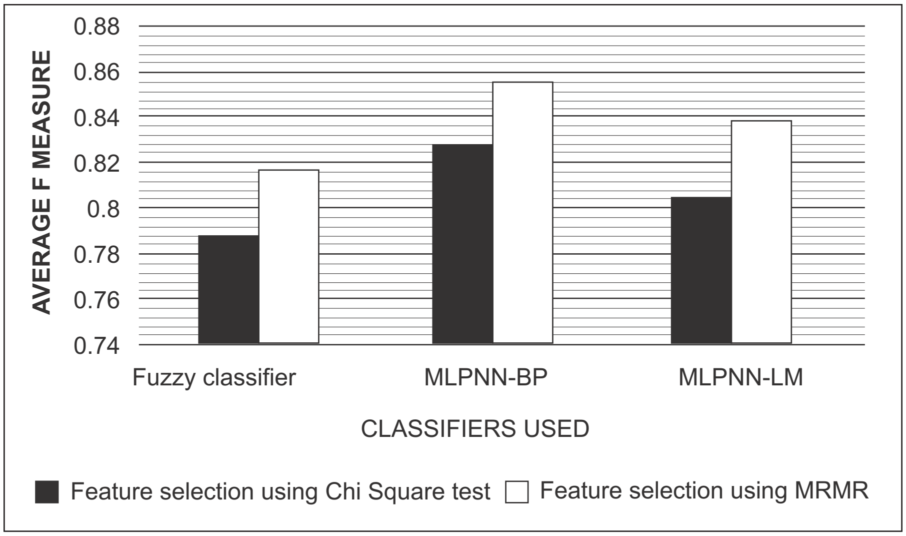

Figure 6: Average F Measure

From the table 4 and figure 6, it can be observed that the feature selection using MRMR has higher average f measure by 3.67% for fuzzy classifier, 3.32% for MLPNN-BP and by 4.14% for MLPNN-LM when compared with feature selection using chi square.

4. CONCLUSION

In this study, for classifying the leaves the wavelet based texture features were extracted and feature selection techniques are used for selecting features. The features are classified using MLP with various learning method such as LM and BP. Nine species of plant leaves were selected with 25 samples of each species. Experimental results show that the feature selection using MRMR has achieved a better performance of classification accuracy, average precision, average recall and average f measures when compared with feature selection using chi square. The feature selection using MRMR has higher classification accuracy by 3.46% for fuzzy classifier, 3.3% for MLPNN-BP and by 4.04% for MLPNN-LM when compared with feature selection using chi square.

REFERENCES & GUIDE BOOKS

[1] Chaki J, Parekh R, Bhattacharya S. Plant leaf recognition using texture and shape features with neural classifiers. Pattern Recognition Letters. 2015;58, pp. 61-8.

[2] Kumar A, Patidar V, Khazanachi D, Saini P. An Approach to Improve Classification Accuracy of Leaf Images using Dorsal and Ventral Features. International Journal of Advanced Computer Science & Applications.;1(6):130-7.

[3] Choras RS. Image feature extraction techniques and their applications for CBIR and biometrics systems. International journal of biology and biomedical engineering. 2007 Sep;1(1):6-16.

[4] Zhou X, Wang J. Feature selection for image classification based on a new ranking criterion. Journal of Computer and Communications. 2015 Mar 17;3(03):74.

[5] Kumar M, Kamble M, Pawar S, Patil P, Bonde N. Survey on techniques for plant leaf classification. International Journal of Modern Engineering Research. 2011;1(2):538-44.

[6] Du JX, Zhai CM, Wang QP. Recognition of plant leaf image based on fractal dimension features. Neurocomputing. 2013 Sep 20;116:150-6.

[7] Wang X, Liang J, Guo F. Feature extraction algorithm based on dual-scale decomposition and local binary descriptors for plant leaf recognition. Digital Signal Processing. 2014 Nov 30;34:101-7.

[8] Zhao ZQ, Ma LH, Cheung YM, Wu X, Tang Y, Chen CL. ApLeaf: An efficient android-based plant leaf identification system. Neurocomputing. 2015 Mar 3;151:1112-9.

[9] Bhandarkar P, Doshi H, Kadakia M, Ahmed R. Leaf identification using morphology and structural decomposition. In Signal Processing and Integrated Networks (SPIN), 2014 International Conference on 2014 Feb 20 (pp. 125-129). IEEE.

[10] Mohammed MF, Lim CP. An enhanced fuzzy min–max neural network for pattern classification. IEEE transactions on neural networks and learning systems. 2015 Mar;26(3):417-29.

[11] Affonso C, Sassi RJ, Barreiros RM. Biological image classification using rough-fuzzy artificial neural network. Expert Systems with Applications. 2015 Dec 30;42(24):9482-8.

[12] Desai P, Pujari J, Goudar RH. Image Retrieval using wavelet based shape features. Journal of Information Systems and Communication. 2012 Jan 1;3(1):162.

[13] Livens S, Scheunders P, Van de Wouwer G, Van Dyck D. Wavelets for texture analysis, an overview.

[14] Gulgezen G, Cataltepe Z, Yu L. Stable and accurate feature selection. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2009 Sep 7 (pp. 455-468). Springer Berlin Heidelberg.

[15] Zhang Z, Khow CK, Liu J, Cheung YL, Aung T. Automatic glaucoma diagnosis with mRMR-based feature selection. J Biomet Biostat S. 2012;7:2.

[16] Liang LR, Basallo EG, Looney CG. Image Edge Detection with Fuzzy Classifier. In CAINE 2001 (pp. 279-282).

[17] Nakashima T, Schaefer G, Yokota Y, Ishibuchi H. A weighted fuzzy classifier and its application to image processing tasks. Fuzzy sets and systems. 2007 Feb 1;158(3):284-94.

[18] Wankhede SB. Analytical Study of Neural Network Techniques: SOM, MLP and Classifier-A Survey. IOSR Journals (IOSR Journal of Computer Engineering). 2014;1(16):86-92.

[19] Saduf MA. Comparative study of back propagation learning algorithms for neural networks. International Journal of Advanced Research in Computer Science and Software Engineering. 2013 Dec;3(12).

[20] Palanivelu L, Vijayakumar P. Improved Neural Network Training Algorithm For Classification Of Compressed And Uncompressed Images. Journal of Theoretical & Applied Information Technology. 2014 Apr 10;62(1).