View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Queensland University of Technology ePrints Archive

QUT Digital Repository:
http://eprints.qut.edu.au/

Pathak, Virendra and Dikshit, Onkar (2006) *Conjoint analysis for quantification of relative importance of various factors affecting BPANN classification of urban environment.* International Journal of Remote Sensing, 27(21). pp. 4769-4789.

© Copyright 2006 Taylor & Francis

Conjoint Analysis for quantification of relative importance of various factors affecting BPANN classification of urban environment

VIRENDRA PATHAK†, ONKAR DIKSHIT‡

†Institute of Engineering & Technology, Lucknow, U. P., India
‡ Indian Institute of Technology, Kanpur, U. P., India.
*Email: vireniet@hotmail.com

Abstract

This paper is an attempt to suggest an approach for eliminating the lengthy process of selecting various factors while using Backpropagation artificial neural network (BPANN) and to quantify the relative importance of the factors affecting classification results. A novel approach called conjoint analysis has been used here. The paper also presents the classification results of an Indian urban environment using two BPANN approaches and compares them with conventional Gaussian Maximum Likelihood (GML) classification approach. The study showed that conjoint analysis can be successfully used to select various parameters of BPANN prior to carrying out the classifications using any of the BPANN approach. Factors like size of training samples and first hidden layer come out as some of the most important factors while the second hidden layer has the least affect on classification accuracy. Resilient backpropagation method of BPANN is the best and robust method for urban classification. Results also showed that classification obtained using BPANN approach were similar or numerically better than GML classification though the difference was not statistically significantly different.

Keywords: BPANN, GML, RPROP, Conjoint Analysis, Relative Importance

1 Introduction

Land use classes in urban environment are highly deviating from normal distribution and affect results from GML classification. Standard classification techniques like GML usually require assumptions about the underlying statistics of the data, the most common being that the data for each ground class cover is Gaussian distributed (Richards 1994, Schowengerdt 1997). If these assumptions turn out to be correct then the statistical classifier is the optimal choice for the problem otherwise alternative approaches are needed to alleviate problems associated with the aforementioned assumption about frequency distribution of data. In recent years, the artificial neural network has been developed and applied to general pattern recognition problem.

Neural network classifiers are non-parametric and may be more robust when distributions are strongly non-Gaussian (Lippman 1987). The application of neural approaches in remote sensing is advocated mainly due the reasons that they perform more accurately in comparison to the other techniques such as statistical classifiers. They perform more rapidly in comparison to other techniques such as statistical classifiers; they incorporate a priori knowledge and realistic physical constraint into the analysis; it is easy to incorporate different types of data (including those from different sensors) into the analysis, thus facilitating synergistic studies; they can produce considerably better results for small training datasets compared to conventional statistical classifiers (Benediktsson *et al*. 1990, 1993; Paola and Schowengerdt 1994; Cote and Tatnall 1995; Foody 1995 a 1995 b; Atkinson and Tatnall, 1997; Benediktsson 1997).

Although BPANN is being intensively used in research, it has some inherent limitations that may have impact on the performance of the classifier. The use of BPANNs requires some critical decisions on the part of the user, which may influence the accuracy of the resulting classification. Some of these factors affecting results using BPANNs for remotely sensed data include size of the training data, number of hidden layers used in the network and number of neurons in the hidden layers (network architecture), learning rate, momentum factor and number of epochs required for training.

Considering the benefits and various issues involved for selecting various parameters for BPANN classification, this study has investigated some of these issues and has attempted to determine some mechanism for arriving at an optimum combination of various factors for BPANN classification purpose.

2 Objectives

The objectives of the study were:

(i) To suggest an approach for eliminating lengthy process of selecting various factors while using BPANN.

(ii) To determine relative importance of the various factors affecting classification accuracy with BPANN classification approach.

(iii) To determine the aptness of BPANN approach for classification of urban environment.

3 Study Site and Data Resources

The study has been carried on two sites, Kanpur the industrial city and Lucknow the state capital of northern Indian state of Uttar Pradesh. Kanpur is situated on the right side bank of the river Ganges, the geographical extent of this study area lies within North latitudes $26^{\circ}20'$ to $26^{\circ}35'$ and the east longitudes $80^{\circ}10'$ to $80^{\circ}25'$. Lucknow is situated in the upper Gangetic plains of the country, the geographical extent of this study area lies within North latitudes $26^{\circ}45'$ to 27° and the East longitudes $80^{\circ}50'$ to $81^{\circ}5'$. From the study of available maps, field visits and previous knowledge about these study sites, it was observed that for Kanpur study site 15, and for Lucknow 12 classes covered the majority of urban land use features. These classes have been considered for further investigations (Table 1, 2).

The satellite data products used for these cities are images acquired from linear image self scanning (LISS)-III sensors on board IRS-1C satellite through National Remote Sensing Agency (NRSA), Hyderabad, India (Table 3). A central extract of 512 x 512 pixels covering major portion of urban areas from these two cities was extracted from the respective satellite images for the study. Figure 1 and 2 show the false color composite (FCC) of these study areas. In addition to these satellite data products corresponding topographic and land use maps were also used.

4 Literature Review

ANNs have been used in a wide range of scientific disciplines for a variety of applications since the early 1980s. Their application in remote sensing area is relatively new, dated back only from the late 1980s. The first studies established the feasibility of the method (Benediktsson *et al.* 1990, Key *et al.* 1989, Ritter and Hepner 1990). Subsequent studies examined the classifier in more detail and compared it to standard techniques such as maximum likelihood. Some researcher found the statistical classifier to be superior, while a majority found that the network produces similar or superior classifications. Many types of neural network model and learning algorithms have been developed in the recent past.

The most common neural network model is the Multilayer Perceptrons (MLP), which works in a feed forward direction where information moves from an input layer to an output layer in the learning phase. Such network contains an extra layer or layers termed the hidden layer(s). For training a feed forward neural network, the most popular technique is the backpropagation algorithm introduced by Rumelhart *et al*. (1986), so it is also called a backpropagation artificial neural network (BPANN). For multispectral image classification, the researchers have used many variants of artificial neural network but the most widely used method is BPANN.

Some of the factors affecting results using ANNs for remotely sensed data have been discussed in Paola and Schowengerdt (1995), Kanellopoulos and Wilkinson (1997), Foody and Arora (1997). These factors include size of the training data (Benediktsson *et al*. 1990), number of hidden layers used in the network (Benediktsson *et al*. 1990, Civco 1991, Kanellopoulos *et al*. 1992) and number of neurons in the hidden layers, learning rate (Paola and Schowengerdt 1995, Kanellopoulos and Wilkinson 1997), momentum factor (Looney 1997, Riedmiller, and Braun 1993, Zurada 1997) and number of epochs required for training (Gong 1996, Bischof *et al*. 1992, Kanellopoulos *et al*. 1991).

There are few more issues, which have been mentioned in the literature like data encoding, network architecture etc. The various form of data encoding and network structures used in BPANN classification studies have been reported in Paola and Schowengerdt (1995). A few researchers have addressed the use of alternative network architectures (Kanellopoulos *et al*. 1991, Benediktsson *et al*. 1990 Xiao and Liu 1991).

Another issue in the training of a neural network is the initial assignment of random weights. Since this assignment is completely independent of the data, training can be long. Further, repeated training, with the same input data, can have different results (Li and Si, 1992). Thimm and Fiesler (1997) suggested that the best initial weights can be determined by the data set to be used. Network pruning is another issue, which has been reported by some authors (Kavzoglu and Mather 1999).

5 Theoretical Background

The theoretical background of the Artificial Neural Networks is available in all standard texts *i.e.* Rumelhart *et al.* (1986) and Zurada (1997). The following sections present brief theoretical background of the two BPANN approaches used for the study and a novel method called conjoint analysis for assessing relative importance of various factor affecting classification accuracy using BPANN.

5.1 Adaptive Learning Algorithm

In the standard backpropagation technique, weights are adjusted using gradient descent method, which keeps learning rate constant throughout training. Many algorithms have been proposed so far to deal with the problem of appropriate weight-update by doing some sort of parameter adaptation during learning. They can roughly be separated into two categories: global and local strategies. Global adaptation techniques make use of the knowledge of the state of the entire network (*e.g.* the direction of the previous weight-step) to modify global parameters, whereas local strategies use only weight-specific information (*e.g.* the partial derivatives) to adapt weight-specific parameters. Besides the fact, the local adaptation strategies are more closely related to the concept of neural learning and are better suited for parallel implementations. The majority of both global and local adaptive algorithms perform a modification of a (probably weight-specific) learning-rate according to the observed behavior of the error-function. The adaptive learning rate is eventually used to calculate the weight-step (Demuth and Beale 1998). The following sections present adaptive learning algorithms used for the present study.

5.1.1 Backpropagation with Variable Learning Rate (BPVLR). In the standard backpropagation technique, the performance of the algorithm is very sensitive to the proper setting of the learning rate. If the learning rate is set too high, the algorithm may oscillate and become unstable. If the learning rate is too small, the algorithm will take too long to converge. It is not practical to determine the optimal setting for the learning rate before training, and, in fact, the optimal learning rate changes during the training process, as the algorithms moves across the performance of the surface.

The performance of the steepest descent algorithm can be improved if we allow the learning rate to change during the training process. In *BPVLR* method, learning rate is made responsive to the complexity of the local error surface. In this process, the initial network output and error are calculated first. At each epoch new weights are calculated using the current learning rate. New output and errors are subsequently calculated. If the new error exceeds the old error by more than a predefined ratio, the new weights are discarded. In addition the learning rate is decreased. Otherwise the new weights are kept. If the new error is less than the old error, the learning rate is increased. This procedure increases the learning rate, but only to the extent that the network can learn without large error increases. Thus a near optimal learning rate is obtained for the local terrain. The learning rate is increased when a larger learning rate could result in stable learning.. When the learning rate is too high to guarantee a decrease in error, it gets decreased until stable learning resumes (Demuth and Beale, 1998).

5.1.2 Resilient Propagation (RPROP). A great variety of further modifications of the backpropagation procedure have been proposed (*e.g*. the use of modified error functions, or sophisticated weight initialization techniques), which all promise to accelerate the speed of convergence considerably. It has been experienced that some of them worked slightly better on some problems, but for others they did not improve convergence or even gave worse results. What is often disregarded is that the size of the actually taken weight-step Δw_{kj} is not only depended on the (adapted) learning rate, but also on the partial derivative $\partial E / \partial w_{kj}$. So the effect of the carefully adapted learning rate can be drastically disturbed by the unforeseeable behavior of the derivative itself. This was one of the reasons that led to the development of 'resilient propagation' *(RPROP)* to avoid the problem of 'blurred adaptivity'. *RPROP* changes the size of the weight-update Δw_{kj} directly, *i.e*., without considering the size of the partial derivative. So the only modification done was to simply 'clip' the logistic activation function at a value, which could be reasonably distinguished from the asymptotic boundary value. This outcome is an always non-zero derivative, preventing the unit from getting stuck. This technique worked out to be far more stable than adding a small constant value to the derivation of the activation function especially in more difficult problems (Riedmiller and Braun 1993).

The main idea used in *RPROP* algorithm roots in the concept of 'direct adaptation' of the size of the weight-update. In contrast to all other algorithm, only the sign of the partial derivative is used to perform both learning and adaptation. This leads to a transparent and yet powerful adaptation process, that can be straight forward and very efficiently computed with respect to both time and storage computation. Another often discussed aspect of common gradient descent is that the size of the derivative decreases exponentially with the distance between the weight and the output layer, due to the limiting influence of the slope of the sigmoid activation function. Consequently, weights far away from the output-layer are less modified and do learn much slower. Using *RPROP*, the size of the weight-step is only depended on the sequence of signs, not on the magnitude of the derivative. For that reason, learning is spread equally all over the entire network; weights near the input layer have the equal chance to grow and learn weights near the output layer.

5.2 Conjoint Analysis

Conjoint analysis is a family of techniques and methods theoretically based on the models of information integration and functional measurement. The purpose of conjoint analysis is to estimate utility scores, called part-worth of various factors considered for the study. For these factors, utility scores are a measure of how important is each factor. The factors are independent variables. The factor levels are the specific values of these independent variables. Output from conjoint analysis includes importance ratings of the factors, part-worth estimates showing preferences for different alternatives. Reader can refer Green and Wind (1975), Green and Srinivasan (1978) and Oppewal (1995) for further details on this technique. A brief theoretical detail of this technique has been presented in the following paragraphs. Full description of the technique is available in Hair *et al*. (1998).

In conjoint analysis, researcher first constructs a set by combining selected levels of each factor. These combinations are then evaluated. Because the researcher selects different factors and their levels in a specific manner, the influence of each factor and its level on the final result (classification accuracy) can be determined from overall ratings of the particular combination of different factors. Table 4 shows the number of factors and their level considered in the present study. As the study was carried out to see the effects of various factors of BPANN which influence classification accuracy, factors like size of training sample (*SS*), number of neurons in first hidden layer (*HL*1), number of neurons in second hidden layer (*HL*2), initial value of learning rate (*LR*), initial value of momentum factor (MF) and number of epochs (EP) were considered with different levels. Five levels of varying sample size *S*1 to *S*5 were considered for factor *SS*, similarly levels of other factors were considered).

Following paragraphs introduce some terminology associated with the conjoint analysis.

1) Creating stimuli: Once the factors and levels have been selected the researcher turns to the task of creating the combinations of factor levels or stimuli for evaluation. The researcher is always faced with an increasing burden as the number of factors and level increase. The researcher must weigh the benefits of increased task effort versus the additional accuracy gained. The total number of cases needed to represent all possible combinations of factor levels is equal to number of levels of factor 1 times the number of levels of factor 2 times the number of levels of factor *n*. In the present study considering all six factors and their levels simultaneously, total number of stimuli becomes 7500 (5x4x3x5x5x5).

2) The full profile method: Full profile method involves the evaluation of all stimuli at a time. In a simple conjoint analysis, the researcher may evaluate all possible stimuli with a small number of factors and levels. This is known as the factorial design when all combinations are used. But as the number of factors and level increases, the design becomes impractical. If the researcher were interested in assessing the impact of 4 variables with for 4 levels for each variable, 256 stimuli would be created in a full factorial design for the full profile method. What is needed is a method for developing a subset of the total stimuli that can be evaluated and still provides the information needed for making accurate and reliable part-worth estimates.

3) Defining subset of stimuli: A fractional factorial design is the most common method for defining a subset of stimuli for evaluation. The fractional factorial design selects a sample of possible stimuli, with the number of stimuli depending on the type of composition rule assumed to be used. Using the additive model, which assumes only main effects for each factor with no interactions, a study using the full profile method with 4 factors at 4 levels requires only 16 stimuli. The 16 stimuli must be carefully constructed to ensure the correct estimation of the main effects. The designs should be optimal designs, as they should be orthogonal (no correlation among levels across attributes).

The number of factors included in the analysis directly affects the statistical efficiency and reliability of the results. As factors and levels are added, the increased number of parameters to be estimated requires either a large number of stimuli or a reduction in the reliability of parameters. The minimum number of stimuli that must be evaluated if the analysis is performed is equal to: Total number of levels across all factors - Number of factors + 1. For the present study, with 5 factors and 22 (4+3+5+5+5) levels across all factors, the minimum number of stimuli required would be 18 (22-5+1).

Many conjoint studies use only a small subset of all possible combinations, called an orthogonal array. An orthogonal array is a subset of the all-possible combinations that still allows estimation of the part-worth for all main effects. Interactions, where the part-worth for a level of one factor depends on the level of another factor, are assumed to be negligible. In an orthogonal array, each level of one factor occurs with each level of another factor with equal or at least proportional frequencies, assuming independence of the main effects. An orthogonal array represents the most parsimonious way to estimate all main effects. Even though it is true that estimation improves as the number of profile increases, information is not really lost by omitting some combinations. In the present study, only a subset of all possible stimuli (i.e. orthogonal array) was generated using the statistical software SPSS10.1.

4) Part-worth: It is the estimate from conjoint analysis of the overall preference or utility associated with each level of each factor used for the study. For example, in table 6, the first column under the factor sample size (*SS)* shows part-worth of its various levels considered for the study. It shows highest positive value against level 5 which indicate that larger sample size (*S*5) has the highest positive influence on the classification accuracy.

5) Interpreting the result: The most common method of interpretation is an examination of the part-worth estimates for each level of factor used for the study, assessing their magnitude and pattern for both practical relevance as well as correspondence to any theory based relationship among levels. The higher the part-worth or utility score, the more impact it has on overall utility. Part-worth values can be plotted graphically to identify patterns.

6) Assessing the relative importance of factors: In addition to portraying the impact of each level with the part-worth estimates, conjoint analysis can assess the relative importance of each factor. Because part-worth estimates are typically converted to a common scale, the greatest contribution to overall utility-and hence the most important factor is the factor with the greatest range (low to high) of part-worth. The importance values of each factor can be converted to percentage summing to 100 percent by dividing each factor's range by the sum of all range values. For the present study, table 5 and 6 show relative importance of factors considered for the study.

6 Experimental Methodology

Initial experimentations with non-parametric classifier were performed with number of variants of BPANN like standard backpropagation method, conjugate gradient algorithm, scaled conjugate gradient, *BPVLR* and *RPROP*. Out of all these variants, results from *BPVLR* and *RPROP* were found to be satisfactory for urban area classification. Therefore, only these two methods were used for further analysis and their results are presented here.

The entire methodology for experiments was divided into two phases. The first phase was related to finding out relative importance of various factors using conjoint analysis. In the second phase, subsequent experiments were performed with selected factors decided upon after conjoint analysis.

6.1 Relative Importance of Factors

Six factors were considered in this study to assess their relative importance in affecting classification accuracy using *BPVLR* classification. Out of these six, only the first four were considered in *RPROP*. Table 4 presents these factors along with their levels considered for the present study.

1) Sample size (*SS*): To study the effect of variation in training sample size, five training sample set *S1* to *S5* were computed. Sample size for training/ test for all the classes of the study site were calculated using an approach suggested by Congalton and Green (1999).

The training and test pixels for different classes were selected with the help of various maps available for the city, field visits, and by employing the experience of the author about various classes in the city. Two different sets of pixels from every class were selected for training and testing purpose using random sampling approach. These sample sets were also referred as five levels of factor sample set (*SS*). Sample sets for study sites were determined at different reliability with desired precision of ± 5%. Table 5 shows the size of sample sets for both the study areas.

2) Number of neurons in first hidden layer (*HL1*): To study the effect of number of hidden layers and to find out relative importance of each hidden layer, two hidden layers were used in the study. The second factor considered for the study was number of neurons in the first hidden layer (*HL1*). Numbers of neurons in the first hidden layer were kept equal to two, three, four and five times (*i.e* 8, 12, 16 and 20 neurons) the number of input bands (4 number) used for the study. This variation in number of neurons is also referred as four levels of factor *HL1*.

3) Number of neurons in second hidden layer (*HL2*): The third factor considered for the study was number of neurons in the second hidden layer (*HL2*). Numbers of neurons in the second hidden layer were kept in multiple of 1, 2 and 3 times the number of neurons in the first hidden layer (*HL1*). This variation in number of neurons in second hidden layer was also referred as three levels of factor *HL2*.

4) Number of Epochs (*EP*): To study effect of variations in number of epochs on classification accuracy, factor epochs (*EP*) was considered with five levels, varying number of epochs from 1000 to 50000 (*i.e*. 1000, 5000, 10000, 25000 and 50000).

5) Learning rate (*LR*): Five levels of initial learning rates (*LR*) were considered. The values of *LR* considered for the study were 0.1, 0.3, 0.5, 0.7 and 0.9. For conjoint study, these values were referred as five levels of factor *LR*.

6) Momentum factor (*MF*): Five levels of *MF* were considered. The values of *MF* considered for the study were 0.1, 0.3, 0.5, 0.7 and 0.9. For conjoint study, these values were referred as five levels of factor *MF*.

As discussed in the previous section, when a researcher using conjoint analysis selects the factors and the levels according to a specific plan, the combination is known as stimuli. In the present study considering all the factors and their levels simultaneously, total number of stimuli becomes 7500 (5x4x3x5x5x5). A stimuli for this case would be any combination of factor levels taken one each from each factor at a time (*i.e*. sample size S4 with 16 number of neurons in first hidden layer, with 32 (16x2) neurons in the second hidden layer, learning rate 0.3, momentum factor 0.9 and 25000 epochs). As it was not possible and desirable to evaluate all stimuli in a conjoint analysis so a fractional factorial design defining a subset of stimuli was considered. A small subset of all possible combinations, called an orthogonal array was generated. A total of 50 classifications each were evaluated with *BPVLR* and *RPROP* on two study sites.

For all classifications, kappa coefficients (κ) of agreement were derived to measure classification accuracy. The classification accuracy of test samples formed the basis of conjoint analysis to work out relative importance of factors and part-worth of their levels. *Z*-statistic was used to compare the two classifications (If the Z value is greater than 1.96, then two classifications are statistically significantly different from each other with 95% confidence level) (Congalton and Green 1999).

The neural classifications were performed with the help of Matlab Neural network toolbox (Matlab 6.0) and conjoint analysis was carried out using SPSS statistical software (SPSS 10.1).

6.2 *Experiments with Selected Factors*

After finding out relative importance of various factors affecting classification accuracy, experiments were performed with a limited number of factors decided upon at the first stage. These experiments were carried out to validate some findings of conjoint analysis and to find an optimum combination of factors to achieve the best classification accuracy using *BPVLR* and *RPROP*.

7 Results and Analysis

The following section presents results of experiments carried out with *BPVLR* and *RPROP* to find out the relative importance of factors affecting classification accuracy. Results with selected parameter and their comparison with GML classification are presented thereafter.

7.1 Relative Importance of Factors in BPVLR

Results of conjoint analysis for assessing relative importance of factors considered for the study are presented in table 6. This table shows relative importance of various factors affecting classification accuracy along with part-worth of various factor levels. These results are also presented in Figure 3.

Following observations can be made from the results with *BPVLR* classification algorithm.

1) Out of the six factors investigated, sample size (*SS*) has highest relative importance followed by the first hidden layer (*HL1*) and Numbers of epochs (*EP*). Momentum factor (*MF*) and learning rate (*LR*) have lesser relative importance respectively indicating their lesser influence on classification accuracy.

2) Relative importance of second hidden layer (*HL2*) was very low indicating its least effect on classification accuracy.

3) With the increase in size of training samples (*SS*), part-worth score increases, which indicates that higher accuracy is achieved with larger training sets. For Kanpur site, sample size S4 and for Lucknow S5 were having higher values of part-worth indicating higher worth for classification.

4) First hidden layer (*HL1*), with neurons equal to 4 times (for Kanpur) and 5 times (for Lucknow) number of input nodes comes out with higher score of part-worth indicating that number of neurons in the first hidden layer should be 4 to 5 times the number of input bands.

5) Higher initial values of learning rate (*LR*) produced higher part-worth scores. For Kanpur study site, *LR* value of 0.9 and for Lucknow 0.7 shows the higher part-worth.

6) Part-worth scores of momentum factor (*MF*) did not give any specific pattern as for Kanpur a value of 0.1 comes out with higher part-worth and for Lucknow it is 0.9.

7) The highest part-worth score for number of epochs (*EP*) was obtained for 25000 epochs for both the study sites indicating optimum number of epochs with *BPVLR*.

7.2 Relative Importance of Factors in RPROP

Results of conjoint analysis for assessing relative importance of factors considered for the study are presented in table 7. Table shows relative importance of various factors affecting classification accuracy using *RPROP* classifier and part-worth of various factor levels. Figure 3 also shows relative importance of various factors considered for the study.

Following observation can be made from results with *RPROP* classification algorithm.

1) Out of the four factors investigated, sample size (*SS*) has highest relative importance followed by first hidden layer (*HL1*), numbers of epochs (*EP*) and second hidden layer (*HL2*).

2) Relative importance of second hidden layer (*HL2*) was very low indicating its negligible effect.

3) With increase in size of training samples, part-worth score increases, which indicates that higher accuracy is achieved with larger training sets.

4) The first hidden layer (*HL1*), with neurons equal to 2, 3 times number of input nodes comes out with similar values of part-worth, which indicates that the first hidden layer having number of neurons 2 to 3 times that of number of input bands is optimum for classification.

5) For number of epochs (*EP*), part-worth scores were negatively correlated with increase in number of epochs. Highest part-worth was obtained for 1000 epochs for both the study sites indicating that lesser number of epochs is sufficient for classification.

7.3 Experiments with Selected Factors in BPVLR

Based on the findings of the first stage of experiments for determining relative importance of various factors, it was decided to use S5 sample size with a single hidden layer having 12, 16 and 20 neurons for the second stage of experiments. Values of initial learning rate and momentum factor were considered as 0.1 and 0.9. For both the study sites, experiments at this stage were performed with 1000, 10000 and 25000 epochs.

It was observed from the results obtained using selected factors that for both the study sites, optimum number of neurons in hidden layer were found to be four to five times the number of input bands. It was also observed that in case of Kanpur, 20 neurons in hidden layer were found to be optimum, while in case of Lucknow this number was 16. The *BPVLR* algorithm adjusts the learning rate according to error surface encountered during training so the initial choice of learning rate was found to be having insignificant effect on the results at higher number of epochs. Though, higher value of learning rate at lesser number of epochs can have adverse impact on results. For, momentum factor it was observed that change in its value does not have any significant effect on the results. This may also be due to the fact that both of these factors have low relative importance amongst the factors considered for the study. It was observed that the number of epochs certainly has significant effect on results and accuracy with 25000 epochs was found to be significantly higher in comparison to lesser epochs. All of these results were in conformation with the results obtained using conjoint analysis. Table 8 shows results of overall training and test accuracy with different values of learning rate, momentum factor at varying number of epochs for Kanpur and Lucknow study site with 20 and 16 neurons respectively in the hidden layer.

Finally, for the Kanpur study site the following combination of factors was decided as the best combination for *BPVLR* classification. Sample size *S5* with single hidden layer having 20 neurons with learning rate and momentum factor as 0.1 and 0.9 respectively and number of epochs equal to 25000. For Lucknow study site, all factors were kept same except single hidden layer having 16 neurons.

7.4 Experiments with Selected Factors in RPROP

After determining relative importance of various factors in the first stage, further experiments were performed taking lead from the first stage. As the total number of factors considered for *RPROP* were less, so experiments at this stage were performed with all sample sets (*S1* to *S5*) using single hidden layer having two to five times the number of neurons (8, 12, 16 and 20) that of number of input bands. All five values of epochs were considered for this stage of analysis.

It was observed from the results obtained using selected factors that for both the study sites, optimum number of neurons in hidden layer was found to be two to three times (8, 12) that of number of input bands. With the increase in number of epochs, though the training accuracy increases but there is marginal decrease in test accuracy. For Kanpur study site, maximum test accuracy was achieved with 5000 epochs though it was not statistically significantly different from accuracy obtained with 1000 epochs. For Lucknow 1000 epochs were found to be sufficient. Results from the experiments carried out to verify importance of higher sample size show that with increase in sample size, test accuracy increases in a significant manner with the highest value at sample size S5 (Table

9). All of these results were in conformation with the results obtained using conjoint analysis.

Finally, for Kanpur study site the following combination of factors was decided as the best combination for *RPROP* classification. Sample size S5 with single hidden layer having 12 neurons and number of epochs equal to 5000. For Lucknow study site, number of epochs was kept as 1000 and remaining factors were same as for Kanpur.

7.5 Comparison of Classification Accuracy

Figure 4 and 5 show comparison of class wise and overall accuracy obtained using *BPVLR* with GML classification for Kanpur and Lucknow study sites. Results show that except few classes, like barren land, industrial area and medium residential-2 for Kanpur and educational institutes and reserve forest for Lucknow, results of *BPVLR* classifications are similar to or better than GML classification but this change is statistically significantly not different. For test areas, the overall κ-coefficient for Kanpur and Lucknow were 0.88 and 0.76 respectively. These overall results were similar to those obtained using GML classifications.

Figure 6 and 7 show the classification results of Kanpur and Lucknow study sites respectively using *RPROP* classification. Figure 8 and 9 show comparison of class and overall accuracy obtained using *RPROP* with GML classification for Kanpur and Lucknow study sites. Results show that in this case also except for a few classes, like barren land, industrial area and river in Kanpur and commercial, educational institutes and reserve forest in Lucknow, results of *RPROP* classifications are similar or better than GML classification but this difference is again statistically significantly not different. For test areas, the overall κ-coefficient for Kanpur and Lucknow were 0.88 and 0.77

respectively. These overall results were also similar to or marginally better than the results obtained using GML classifications.

8 Conclusions

The following conclusions can be drawn from the various experiments carried out to understand behaviour of BPANN and to study effects of various factors influencing classification accuracy.

1) Conjoint analysis can be successfully used to determine effects of various parameters affecting classification accuracy, as most of the results obtained using this technique were commensurate with the results obtained using detailed experimentations.
2) Sample size has the highest relative importance affecting classification accuracy in BPANN classification. Classification accuracy increases with an increase in the sample size.
3) The second most important factor is the first hidden layer. Relative importance of the second hidden layer is very low indicating that it has the least effect on classification accuracy. This also suggests that only one hidden layer is sufficient for BPANN classification.
4) In case of *BPVLR*, with four input feature the number of neurons in hidden layer should be 4 to 5 times that of number of input features. The effect of variation in learning rate and momentum factor is insignificant at higher number of training epochs.
5) In case of *RPROP*, with four input features the number of neurons in hidden layer could be 2 to 3 times that of number of input features. Further, lesser number of

epochs (1000) is sufficient for training to get good classification results, which is contrary to *BPVLR*.

6) Out of all the variants of BPANN tried in the analysis, *RPROP* comes out as the best and robust method of BPANN classification because it involves less number of factors, lesser number of neurons is required in hidden layers and better results are obtained with lesser number of epochs.
7) Classification results obtained using BPANN classifications are similar or numerically better than GML classification. However, the difference between the results of these two approaches is not statistically significant. Therfore, it can be concluded that for urban areas having mixed spectral classes, the BPANN approach of classification does not add much value to the classification results with medium resolution data, as used in the study.

Reference:

ATKINSON, P. M. and TATNALL, A. R. L., 1997, Neural networks in remote sensing. *International Journal of Remote Sensing,* 18(4), 699-709.

BENEDIKTSSON, J. A., SWAIN, P. H. and ERSOY, O. K., 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Trans. on Geoscience and Remote Sensing*, vol. 28, pp. 540-551.

BENEDIKTSSON, J. A. , SWAIN, P. H. and ERSOY, O. K. ,1993, Conjugate gradient neural networks in classification of multisource and very-high dimensional remote sensing data, *International Journal of Remote Sensing*, vol. 14, pp. 2883-2903.

BENEDIKTSSON, J. A. and SVEINSSON, J. R.,1997, Multisource data classification and feature extraction with neural networks," *International Journal of Remote Sensing*, vol. 18, pp. 727-740.

BISCHOF, H., SCHNEIDER, W. and PINZ, A. J., 1992, Multispectral classification of Landsat images using neural network, *IEEE transactions on Image Processing,* vol. 30, pp. 482-490.

CIVCO, D. L. 1991, Landsat TM image classification with an artificial neural network, *Proceedings, ASPRS-ACSM Annual meeting, Baltimore, MD*, vol.3 (Falls Church, VA: American Society of Photogrammetry and Remote Sensing, pp. 67-77.

CONGALTON, R. G. and GREEN, K., 1999, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press, Inc., USA.

COTE, S. and TATNALL, A. R. L., 1995, A neural network based method for tracking features from satellite sensor images, *International journal of Remote Sensing*, vol. 16, pp. 3695-3701.

DEMUTH, H. and BEALE, M., 1998, *Neural Network Toolbox User's Guide* (*Version 3*), The Math Works Inc.

FOODY, G. M., 1995 a, Using prior knowledge in artificial neural network classification with a minimal training set, *International Journal of Remote Sensing,* vol. 16, pp. 301-312.

FOODY, G. M., 1995 b, Land cover classification using an artificial neural network with ancillary information, *International Journal of Geographical Information Systems*, vol. 9, pp. 527-542.

FOODY, G. M. and ARORA, M. K., 1997, An evaluation of some factors affecting the accuracy of classification by an artificial neural network, *International Journal of Remote Sensing,* vol. 18(4), pp. 799-810.

GONG, P., 1996, Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping, *Photogrammetric Engineering and Remote Sensing,* vol. 62, pp. 513-523.

GREEN, P. E. and WIND, Y., 1975, New Way to Measure Consumers' Judgments, *Harvard Business Review* 53 (July-August): 107-117.

GREEN, P. E. and SRINIVASAN, V., 1978, Conjoint Analysis in Consumer research: Issues and outlook. *Journal of Consumer Research* 5 (September), pp.103-123.

HAIR, J.F., ANDERSON, R.E., TATHAM, R.L. and BLACK, W.C., 1998, *Multivariate Data Analysis*, fifth edition, Prentice-Hall International Inc., NJ, USA.

KANELLOPOULOS, I., VARFIS, A., WILKINSON, G. G. and MEGIER, J., 1991, Neural network classification of a multi date satellite imagery, *Proceedings of the International Geoscience and Remote Sensing Symposium* (*IGRASS' 91*), Espoo, Finland, June, pp. 2215-2218.

KANELLOPOULOS, I., VARFIS, A., WILKINSON, G. G. and MEGIER, J., 1992, Land-cover discrimination in SPOT HRV imagery using an artificial neural network-a 20 class experiment, *International Journal of Remote Sensing,* vol. 13 (5), pp. 917-924.

KANELLOPOULOS, I., WILKINSON, G. G., 1997, Strategies and best practice for neural network image classification, *International Journal of Remote Sensing,* vol. 18 (4), pp. 711-725.

KAVZOGLU, T. and MATHER, P. M., 1999, Pruning artificial neural networks: an example using land cover classification of multi-sensor images, *International Journal of Remote Sensing,* vol. 20, pp. 2787-2803.

KEY, J, MASLANIK, J. A. and SCHWEIGER, A. J., 1989, Classification of merged AVHHR and SMMR arctic data with neural networks," *Photogrammetric Engineering and Remote Sensing,* vol. 55, pp. 1331-1338.

LI, R. and SI, H., 1992, Multispectral image classification using improved backpropagation neural networks, *Proceedings of the International Geoscience and Remote Sensing Symposium* (*IGRASS'92*), Houston, TX, May 1992, pp. 1078-1080.

LIPPMANN, R. P., 1987, An introduction to computing with neural networks, *I.E.E.E. ASSP Magzine*, pp. 4-22.

LOONEY, C.G., 1997, *Pattern recognition using neural networks*, Oxford university press, Oxford, New York.

OPPEWAL, H., 1995, A Review of Conjoint Software” *Journal of Retailing and Consumer Services* 2(1), pp. 55-61.

PAOLA, J. D. and SCHOWENGERDT, R. A., 1994, Comparison of neural networks to standard techniques for image classification and correlation, *Proceeding of the International Geoscience and Remote Sensing Symposium (IGRASS’94),* Pasadena, USA, pp. 1404-1406.

PAOLA, J. D. and SCHOWENGERDT, R. A., 1995, A review and analysis of back-propagation neural networks for classification of remotely-sensed multispectral imagery, *International Journal of Remote Sensing,* vol. 16(16), pp. 3033-3058.

RICHARDS, J. A., 1994, *Remote sensing digital image analysis: an introduction*. II edn, Springer-Verlag, New York.

RIEDMILLER, M. and BRAUN, H., 1993, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, *Proceeding of the IEEE International conference on Neural Networks*.

RITTER, N. D. and HEPNER, G. F., 1990, Application of an artificial neural network to land cover classification of Thematic Mapper imagery, *Computers & Geosciences*, vol. 16, pp. 873-880.

RUMELHART, D. E., HINTON, G. F. and WILLIAMS, J., 1986, Learning internal representations by error propagation, In *Parallel Distributed Processing*: *Exploration in the Microstructure of Cognition, Volume 1: Foundations*, edited by D. E. Rumelhart and J. L. McClelland, MIT Press, Cambridge, pp. 318-362.

SCHOWENGERDT, R. A., 1997, *Remote Sensing: Models and Methods for Image Processing* (Second Edition). Academic Press, MA, USA.

THIMM, G. and FIESLER, E., 1997, High-order and multiplayer perceptron initialization," *IEEE Trans. on Neural Networks*, vol. 8, pp. 349-359.

XIAO, J. Y. and LIU, Z. K., 1991, Classification of remotely–sensed image data using artificial neural networks, *International Journal of Remote Sensing,* vol. 12(11), pp. 2433-2438.

ZURADA, J. M., 1997, *Introduction to artificial neural systems*. Jaico Publishing, Mumbai.

Table 1. Classes in Kanpur study area and their brief description

S. no.	*Name*	*Description*
1	Agriculture-1	Agriculture area having crops at middle stage of growth
2	Agriculture-2	Agriculture area having crops at early stage of growth
3	Airport campus	Civilian aerodrome and its campus
4	Barren land	Few patches of land which are barren
5	Central ordinance depot	Typical shoe shape feature containing houses and depots
6	High residential	Residential areas with more than 600 persons/hectare
7	Industrial area	Area reserved for industrial activity, factories
8	Medium residential-1	Residential areas with 400 to 600 persons/hectare
9	Medium residential-2	Residential areas located at urban-rural fringe
10	Medium residential-3	Residential areas belonging to defense establishment
11	River	Ganges river flowing in the top right corner
12	River sand-1	Dry sand in flood plane area of river Ganges
13	River sand-2	Wet sand in flood plane area of river Ganges
14	Water body	Various small water bodies in the study area
15	Zoo	Zoo located in the top left part of study area

Table 2. Classes in Lucknow study area and their brief description

S. no.	*Name*	*Description*
1	Agriculture-1	Agriculture area having crops at middle stage of growth
2	Agriculture-2	Agriculture area having crops at early stage of growth
3	Commercial	Central business area of the city
4	Educational institutes	Various educational Institutions
5	Government establishment	Different Government establishments
6	Grassy land	Big patches of lands having grass only
7	High residential	Residential areas with more than 600 persons/hectare
8	Medium residential	Residential areas with 400 to 600 persons/hectare
9	Park	Parks for recreational activities
10	Reserve forest	A big portion of land reserved for forest
11	River	River Gomati flowing from left to right
12	Water body	Various small water bodies in the study area

Table 3. Satellite data characteristics for study area

Sensor	*Bands*	*Resolution (m)*	*Size (pixels)*	*Wavelength (μ m)*	*Spectral Region*
Kanpur					
LISS III	B2	23.5	512 x 512	0.52-0.59	Green
	B3	23.5	512 x 512	0.62-0.68	Red
	B4	23.5	512 x 512	0.77-0.86	NIR
	B5	70.5*	512 x 512	1.55-1.70	SWIR
Lucknow					
LISS III	B2	23.5	512 x 512	0.52-0.59	Green
	B3	23.5	512 x 512	0.62-0.68	Red
	B4	23.5	512 x 512	0.77-0.86	NIR
	B5	70.5*	512 x 512	1.55-1.70	SWIR

*resampled to 23.5 m

Table 4. Factors and their levels considered for Conjoint analysis

Level	*Factors*					
	Common for BPVLR and RPROP				*For BPVLR only*	
	SS	*HL1*	*HL2**	*EP*	*LR*	*MF*
1	*S1*	8	1	1000	0.1	0.1
2	*S2*	12	2	5000	0.3	0.3
3	*S3*	16	3	10000	0.5	0.5
4	*S4*	20	------	25000	0.7	0.7
5	*S5*	------	------	50000	0.9	0.9

* Multiple of neurons in first hidden layer

Table 5. Sample size for study area

Sample set (*SS*)	*Reliability*	*Samples/class*	
		A	*B*
S1	50	31	35
S2	75	40	46
S3	85	45	53
S4	95	61	71
S5	99	72	90

A Kanpur study site
B Lucknow study site

Table 6. Relative importance of various factors and part-worth of their levels (*BPVLR*)

Level	*Factors*											
	SS		*HL1*		*HL2*		*LR*		*MF*		*EP*	
	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
	23.17	***39.28***	***22.22***	***17.57***	***7.33***	***5.94***	***12.29***	***9.82***	***13.71***	***13.95***	***21.28***	***13.44***
1	-6.80	-9.40	-3.50	-4.00	-1.63	-1.23	-2.00	-2.40	2.00	-0.80	-5.60	0.00
2	1.00	2.40	-3.70	1.40	1.46	1.06	1.20	0.60	1.60	0.80	1.00	0.20
3	1.00	-2.80	5.70	-0.20	0.16	0.16	0.40	0.40	-0.60	-0.60	-1.20	-0.20
4	3.00	4.00	1.50	2.80	-	-	-2.40	1.40	-3.80	-2.40	3.40	2.60
5	1.80	5.80	-	-	-	-	2.80	0.00	0.80	3.00	2.40	-2.60

* Values shown in bold italics indicate relative importance in percentage
A Kanpur study site
B Lucknow study site

Table 7. Relative importance of various factors and part-worth of their levels (*RPROP*)

Level	*Factors*							
	SS		*HL1*		*HL2*		*EP*	
	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*
	55.89	***45.54***	***17.85***	***18.46***	***3.77***	***8.31***	***22.90***	***27.69***
1	-9.00	-7.40	1.37	2.25	-0.50	1.53	2.80	5.40
2	-4.00	-1.00	1.27	1.85	0.00	-0.36	2.00	-1.60
3	-1.00	-1.60	1.27	-3.75	0.50	-1.16	0.20	0.00
4	6.40	2.60	-3.92	-0.35	-	-	-1.00	-0.20
5	7.60	7.40	-	-	-	-	-4.00	-3.60

* Values shown in bold italics indicates relative importance in percentage
A Kanpur study site
B Lucknow study site

Table 8. Results using *BPVLR* with varying combination of factors

S. no.	*LR*	*MF*	*EP*											
			1000				*10000*				*25000*			
			Training		*Test*		*Training*		*Test*		*Training*		*Test*	
			A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
1	0.1	0.1	0.80	0.65	0.78	0.65	0.87	0.76	0.86	0.73	0.89	0.80	0.86	0.76
2	0.1	0.9	0.84	0.68	0.84	0.67	0.89	0.79	0.86	0.75	0.90	0.80	0.88	0.76
3	0.9	0.1	0.79	0.65	0.77	0.65	0.87	0.79	0.85	0.75	0.90	0.80	0.87	0.76
4	0.9	0.9	0.84	0.68	0.84	0.67	0.89	0.78	0.87	0.75	0.90	0.81	0.88	0.75

A Kanpur study site
B Lucknow study site

Table 9. Overall test accuracy using *RPROP*

(a) With varying neurons and sample sets (*SS*) for 1000 epochs

S. no.	*Neurons*	*S1*		*S2*		*S3*		*S4*		*S5*	
		A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
1	8	0.79	0.72	0.83	0.73	0.85	0.72	0.87	0.73	0.86	0.74
2	12	0.80	0.71	0.86	0.74	0.84	0.72	0.86	0.74	0.86	0.75
3	16	0.82	0.72	0.85	0.75	0.84	0.72	0.86	0.73	0.86	0.76
4	20	0.81	0.72	0.85	0.73	0.85	0.73	0.86	0.74	0.88	0.75

(b) With varying neurons and epochs for sample size *S5*

S. no.	*Neurons*	*1000*		*5000*		*10000*		*25000*		*50000*	
		A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
1	8	0.86	0.74	0.88	0.74	0.88	0.74	0.88	0.73	0.88	0.73
2	12	0.86	0.75	0.88	0.75	0.88	0.74	0.88	0.75	0.88	0.74
3	16	0.86	0.76	0.87	0.75	0.87	0.74	0.87	0.74	0.85	0.73
4	20	0.88	0.75	0.88	0.75	0.88	0.74	0.88	0.75	0.87	0.73

A Kanpur study site
B Lucknow study site

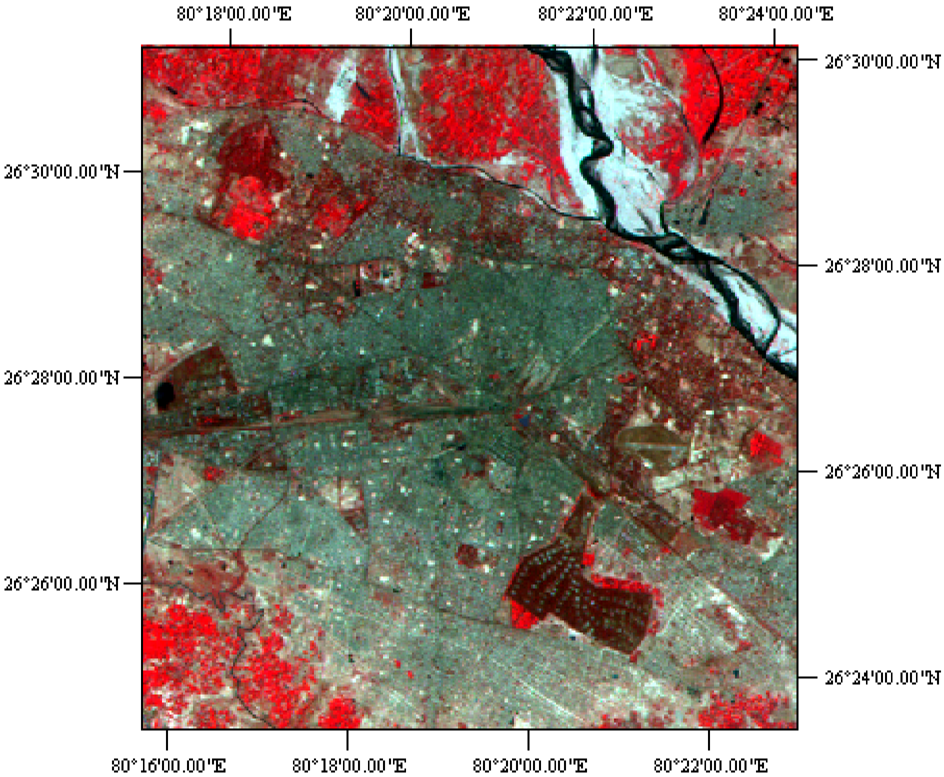

Figure 1 FCC of Kanpur study area

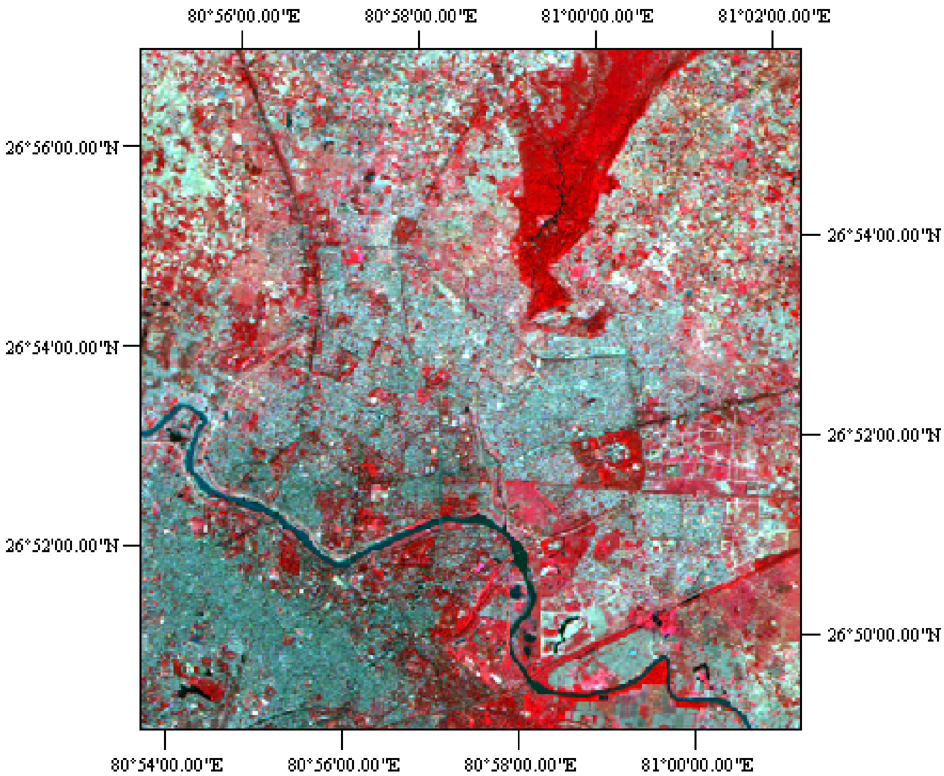

Figure 2 FCC of Lucknow study area

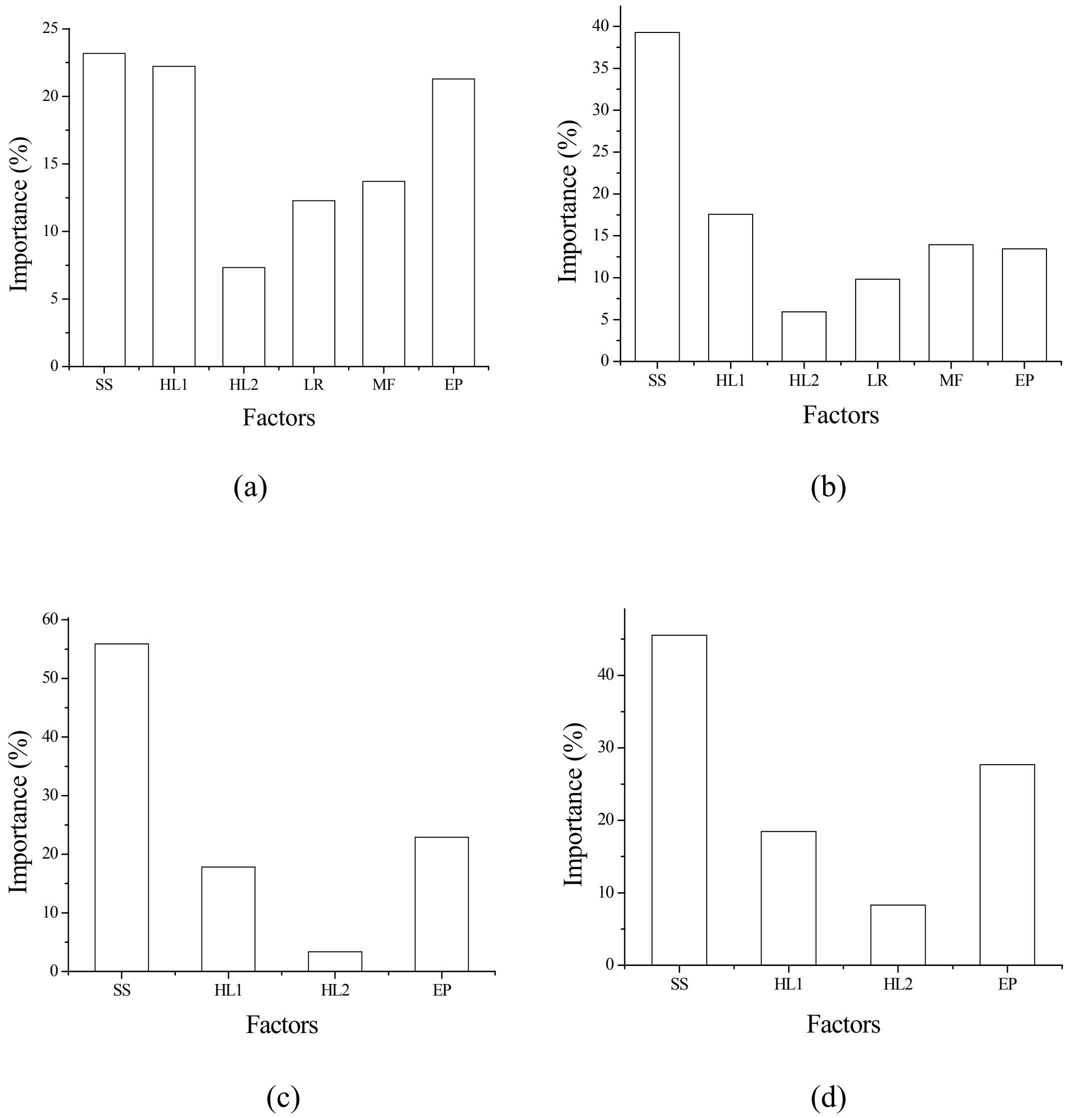

Figure 3. Relative importance of factors with (a) *BPVLR* for Kanpur (b) *BPVLR* for Lucknow (c) *RPROP* for Kanpur (d) *RPROP* for Lucknow

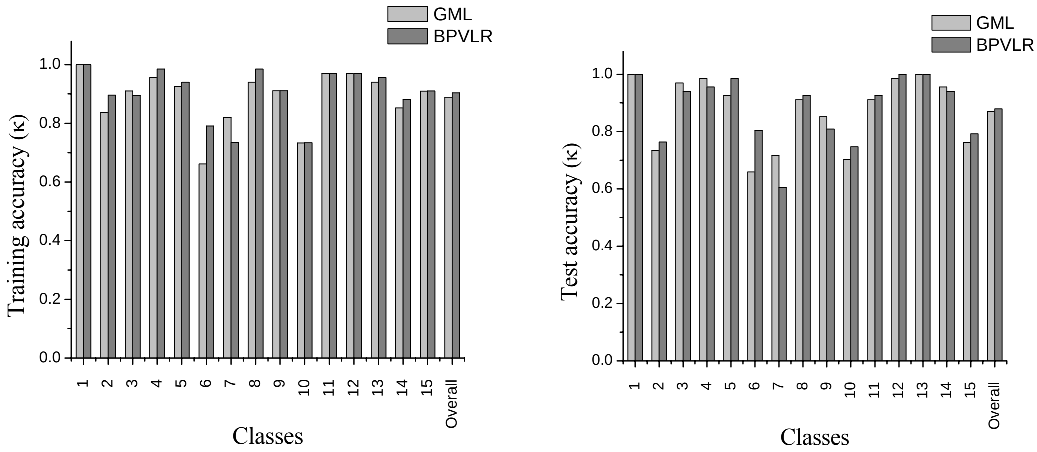

Figure 4. Comparison of *BPVLR* and GML classification accuracy for Kanpur

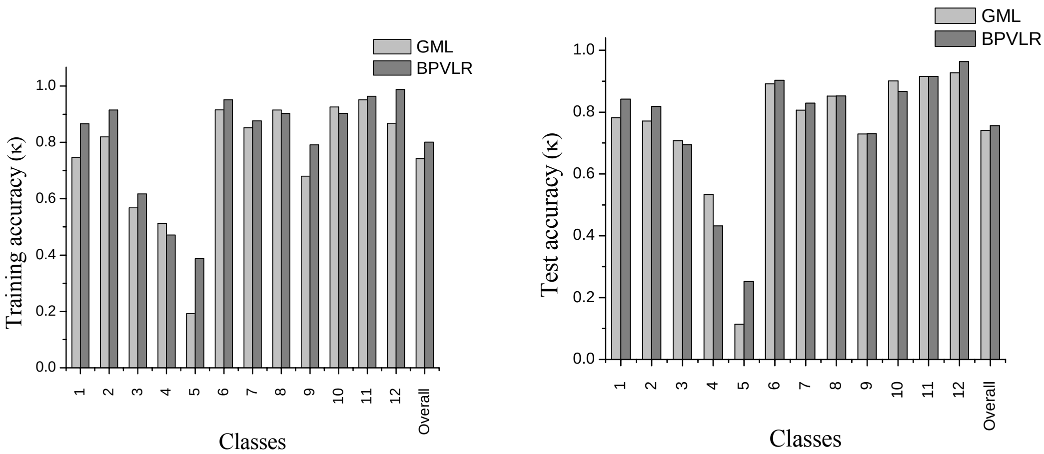

Figure 5. Comparison of *BPVLR* and GML classification accuracy for Lucknow

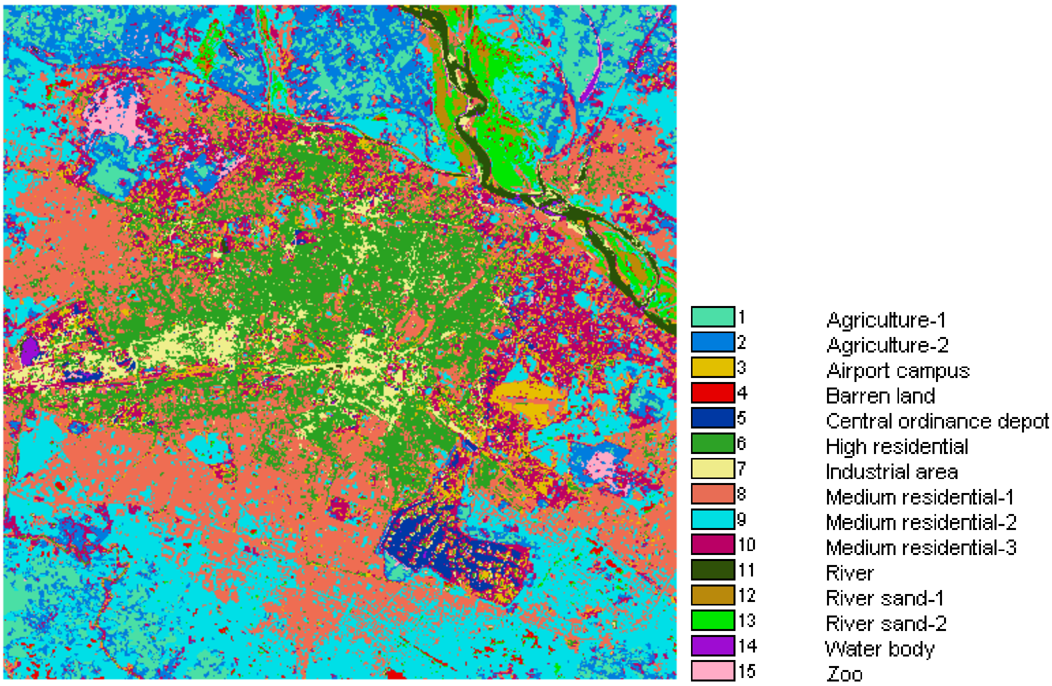

Figure 6 Classified image of Kanpur using *RPROP*

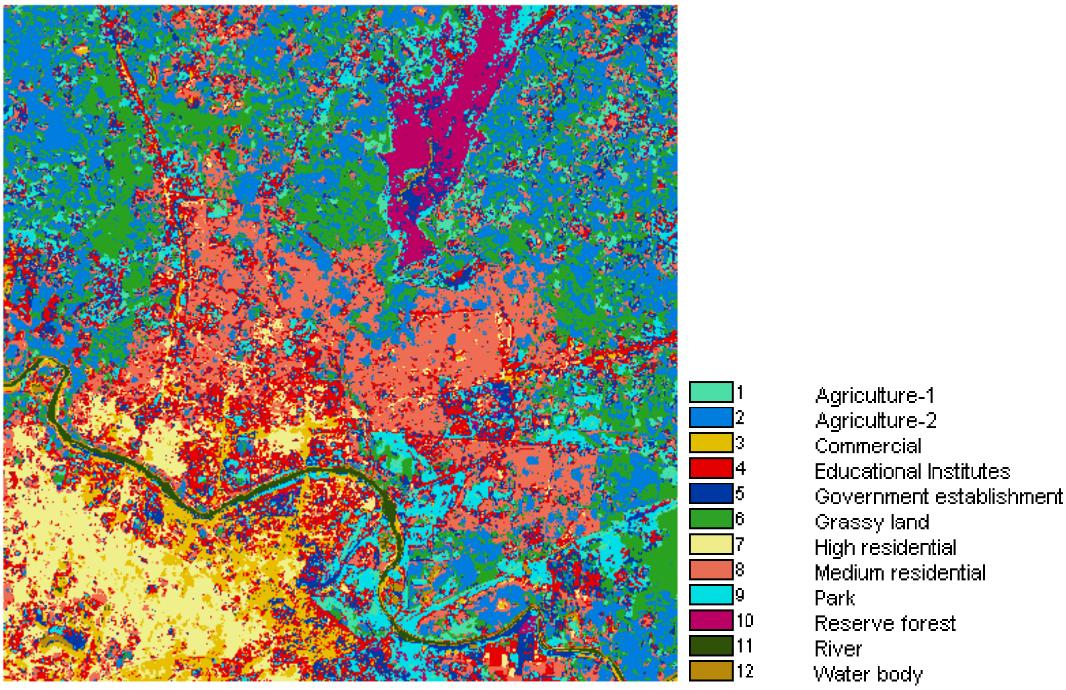

Figure 7 Classified image of Lucknow using *RPROP*

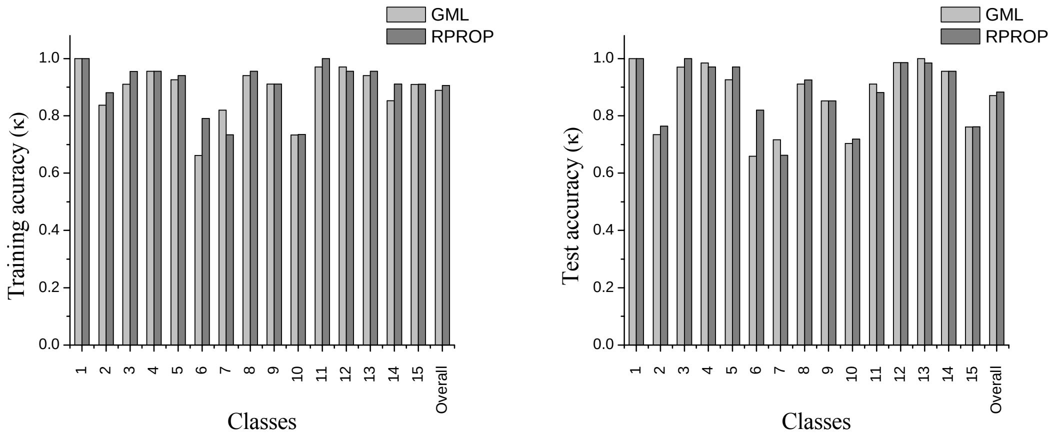

Figure 8. Comparison of *RPROP* and GML classification accuracy for Kanpur

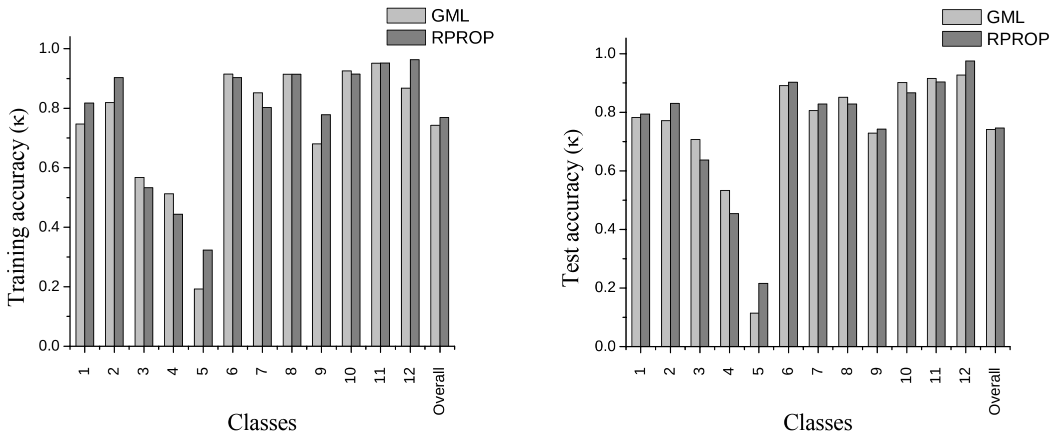

Figure 9. Comparison of *RPROP* and GML classification accuracy for Lucknow