Defence Science Journal, Vol 51, No 3, July 2001, pp. 263-278

Neural Network Parameters Affecting Image Classification

K.C. Tiwari
Army Headquarters, Kashmir House, New Delhi – 110 011

ABSTRACT

This study is to assess the behaviour and impact of various neural network parameters and their effects on the classification accuracy of remotely sensed images which resulted in successful classification of an IRS-1B LISS II image of Roorkee and its surrounding areas using neural network classification techniques. The method can be applied for various defence applications, such as for the identification of enemy troop concentrations and in logistical planning in deserts by identification of suitable areas for vehicular movement. Five parameters, namely training sample size, number of hidden layers, numbers of hidden nodes, learning rate and momentum factor were selected. In each case, sets of values were decided based on earlier works reported. Neural network-based classifications were carried out for as many as 450 combinations of these parameters. Finally, a graphical analysis of the results obtained was carried out to understand the relationship among these parameters. A table of recommended values for these parameters for achieving 90 per cent and higher classification accuracy was generated and used in classification of an IRS-1B LISS II image. The analysis suggests the existence of an intricate relationship among these parameters and calls for a wider series of classification experiments as also a more intricate analysis of the relationships..

Keywords: Remote sensing, digital image classification, artificial neural network technique, knowledge-based classification techniques, fuzzy techniques, multi-band remote sensing data

1. INTRODUCTION

Our planet, Earth is endowed with rich natural resources that are widespread, but mostly inaccessible. The advancements in the field of remote sensing has made access to these areas somewhat easy. However, the remote sensing data made available through the satellites has to be put through a process called digital image classification before it can be put to any meaningful use.

Conventionally, statistical classification algorithms, such as maximum likelihood classifier, minimum distance-to-means classifier, parallelopiped classifier, etc. have been used for supervised image classification[1]. These classification algorithms are based on certain statistical distribution assumptions, i.e. normal or Poisson distribution of data. The data in practice, however, rarely satisfies these assumptions and thus results in erroneous classifications. Nevertheless, if these assumptions are met, accurate classifications can be obtained. With significant improvements in space and satellite technologies, huge amount of multi-band data is encountered which need to be classified to retrieve useful information. The conventional classification techniques fail miserably both in terms of their capability to handle such a large amount of data as also in producing accurate classifications.

Revised 7 March 2001

The limitations of the conventional classification techniques have heralded an era of newer and better techniques capable of handling complex, multi-band data[1]. Some of these most popular techniques are knowledge-based classification techniques[2,3], fuzzy techniques[4], and artificial neural network techniques[5,6]. Unlike conventional classifiers, these are non-parameteric in nature, as they do not depend on statistical pre-assumptions about the distribution of data. In addition, these classifiers have shown the capability of handling multi-source data and so have found increasing acceptance in classification of multi-band remote sensing data, particularly when data from other ancillary sources are also used to generate qualitative classifications. Various parameters involved in one such new technique currently under active research, namely the artificial neural network technique, has been studied.

2. STUDY AREA & DATA

An IRS-1B LISS II 512 × 512 image of 19 January 1992 in four spectral bands (three in visible and one in near-infrared region) of Roorkee and its surrounding areas has been selected for this study (Fig. 1). The area is largely dominated by vegetation of various types with a river meandering in its vicinity. The river has produced large patches of sandy areas. The centre of the city is largely built-up area and spreads right up to the peripheries of the city.

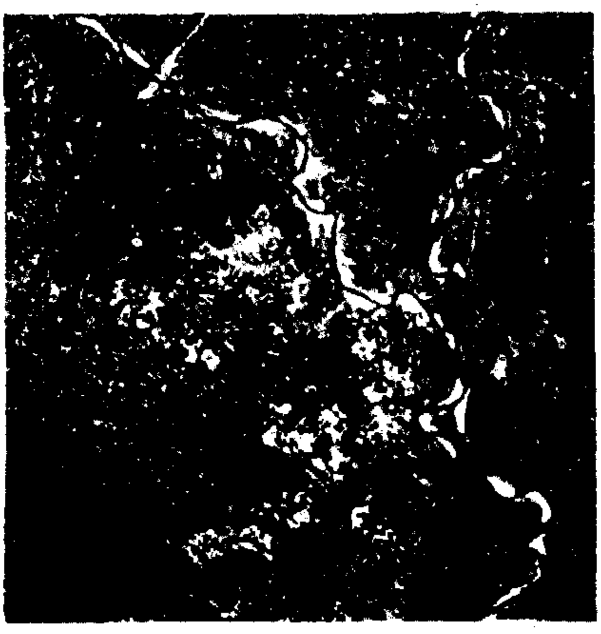

Figure 1. IRS-1B LISS II image of Roorkee and its surrounding area.

Survey of India map of Roorkee area at scale 1:50,000 (map sheet No 53-G/13) and aerial photographs at scale 1:25,000 of the same area were used as reference data for selecting training and testing data. The same was also confirmed through ground visits. Selection of training and testing data was done displaying the image on a computer monitor using intergraph's image processing system. Five different classes dominant in the area, namely built-up, water, sand, vegetation and miscellaneous were extracted from the image using the reference data. The miscellaneous class constituted pixels, which did not fall in any of the classes selected. Extraction of data for various classes was done by generating a number of polygons well spread over the entire image. The identity of a particular class was determined using the reference data. A total number of pixels contained in the polygons for different classes varied from 600 to 1,000.

The data thus generated was segregated randomly into training and testing data. While the testing sample size was fixed for each class, the training sample size was varied since it was one of the various parameters under investigation. A total of 300 pixels were selected in each class as testing samples.

2.1 Artificial Neural Network Parameters

Artificial neural networks (ANNs) have found several applications around the globe and adequate literature is available to prove its efficacy and potential. From remote sensing point of view, ANNs have been found to produce accuracy as high as 99 per cent for a range of classification problems, such as land use, landcover mapping[6], soil mapping[1], cloud classification[1], geological mapping[7], etc. However, the accuracy of a neural network classification depends upon several factors[8] as mentioned above. Improper selection of any of these factors may result in inaccurate classifications. Therefore, there is a need to study the effect of various neural network parameters and their interdependency to make a suitable recommendation for their judicious selection. Although, there are several parameters which affect the classification performance of ANN as classifiers, the effect of only the following parameters have been investigated in the present study:

2.1.1 *Training Sample Size*

Initially 300 pixels per class (viz., built-up, water, sand, vegetation and miscellaneous class) were selected as testing pixels and then from the rest of the data, five different training sample sizes were created varying in size from as low as 10 pixels per class to as high as 200 pixels per class and designated as

Training sample size TSS 10 = 10 pixels per class

Training sample size TSS 50 = 50 pixels per class

Training sample size TSS 100 = 100 pixels per class

Training sample size TSS 150 = 150 pixels per class

Training sample size TSS 200 = 200 pixels per class

The data presented a case of nonlinear classification and the miscellaneous class further introduced sufficient confusion because of the presence of many different classes. This was done so that the selected data could act as a true litmus test for the ANNs capability as a classifier.

2.1.2 *Number of Hidden Layers & Hidden Nodes*

In remote sensing literature, studies using both one and two hidden layers have been reported. This study therefore had a set of experiments conducted for both one and two hidden layers. Although, there is no well-defined rule for the selection of number of hidden nodes, these have been selected using a simple logic in relation to the number of input nodes, four in the present case. The number of hidden nodes selected for this study was in multiple of the number of input nodes. Since for the present study, the input nodes were four, one each corresponding to four input bands, the number of hidden nodes selected were 2, 4, 8, 12 and 16 (i.e. ½, 1,2,3,4th multiple of number of input nodes). Further, a similar strategy was adopted for selecting the number of hidden nodes for the case of two hidden layers. The architecture for two hidden layers was kept as 2/2, 4/4 8/8, 12/12 and 16/16 (number of hidden nodes in first and second hidden layers, respectively).

2.1.3 *Learning Rate & Momentum Factor*

Being faster and superior, adaptive learning rate approach of back propagation learning rule was adopted for training the nodes. This required input of only the initial value of the learning rate. Thereafter, at the end of each epoch, if the learning rate produced error more than a pre-defined ratio, it was discarded and a lower value of learning rate was selected, otherwise learning rate value was increased to speed up learning. Three different values of starting learning rates were tested. The upper and lower range of these learning rate values were determined using the empirical formula given by Heerman and Khazenie in 1992. The formula is $Lr = Co/P.N$, where Lr is learning rate, P is number of samples (or patterns), N is total number of nodes in the network and Co is a factor experimentally determined[9] which is equal to 10. The learning rates, thus determined and selected for the study were 0.04, 0.004 and 0.0004. Three momentum factor values chosen in the range 0 to 1 were 0.15, 0.55 and 0.95.

3. METHODOLOGY

The main objective of this study was to assess the behaviour and impact of various ANN parameters on image classification. Having selected the range of values for different parameters, the implementation was carried out as follows:

3.1 Architecture

The number of input and output nodes is governed by external parameters. Since for the present study, four spectral bands were used as input data and their classification sought into five different classes, the number of input and output nodes were fixed as four and five, respectively. Different ANN architectures were developed, using these as input and output nodes, for both one and two hidden layers with various combinations of hidden units.

3.2 Input Data Encoding

Different data files for various training and testing sample sizes were created. To ensure that the input and output data matched well with each other, input data was normalised in the range 0 to 1. The normalisation of input data was done separately for each training sample size at the time of training by multiplying the input data by an encoding factor. This encoding factor (EF) was determined using the following formula:

$$EF = \frac{DN_i - DN_{min}}{DN_{max} - DN_{min}}$$

where DN_i denotes the individual pixel value and DN_{max} and DN_{min} are the maximum and minimum pixel values, respectively.

3.3 Target Data Encoding

Although in remote sensing literature, various schemes of target data encoding have been reported, however, for the present study, it was decided to constitute the target data set using only two numbers, 1 and 0. The code consisted of five digits (one for each class). The leftmost digit represented first class and the second leftmost digit represented second class and so on. A target value of 1 signified identification with a particular class and a target value of 0 indicated otherwise. Thus, a value of 1 as first leftmost digit meant identification with first class. It was not possible to have value of 1 at any two locations in the five digit code. For example, an output node representing class 3 was encoded as 00100 in a five-class experiment. Data sets were grouped class by class, i.e. all the pixels of class 1 were included first and then rest of the classes were appended to it in their respective class groups for both input as well as output data.

3.4 Initial Weights

A typical ANN classification process starts with a choice of a set of initial weights. Since the aim of the present study was to evaluate the effect of various parameters, initial weights were kept fixed for logical comparison. Fixed weights for different number of hidden node categories under investigation were generated by training the complete test data set for 10,000 epochs. The training was done using adaptive learning rate method with a fixed starting learning rate of 0.000267 and momentum factor of 0.95. Separate training was carried out for different hidden node categories. These weights were then used as fixed starting weights in all subsequent classification experiments.

3.5 Other Parameters

Separate experiments using all the three values of learning rate and momentum factors were carried out. All the training were commenced by randomly pegging the error goal at 0.01. Sum squared error and last iteration error were noted for each classification. In all the experiments, training was carried out for a fixed number of epochs (2500 epochs) and no attempt was made to seek convergence to the error goal fixed.

3.6 Classification

In the remote sensing literature, depending upon the type of input and output data encoding, several classification strategies have been proposed[9].

In the present study, the target outputs for training were constituted using two numbers, i.e. 0 and 1 only. However, the values of classified output data ranged between 0 and 1(in decimals) for each class. Therefore, a pixel was assigned to a class which had the highest output value.

3.7 Accuracy Assessment

The accuracy of classifications were assessed by generating error matrices at the end of each classification and overall accuracy (OA) calculated using the formula:

$$OA = \frac{\text{Sum of pixels correctly classified}}{\text{Total number of testing pixels}}$$

3.8 Computer Program

The neural network toolbox of the MATLAB software was used for all computational tasks. However to carry out classifications as desired, necessary computer programs were developed in MATLAB (MATLAB version 5.2). These programs were implemented on pentium P-I, PC-233/PC-266 MHz with 32 MB RAM. Processing was carried out simultaneously in 2-3 windows.

4. RESULTS & DISCUSSION

The objective of the study was to investigate the effects of various neural network parameters on the accuracy of image classification. Although, there are several parameters which affect the classification accuracy, only five parameters were investigated in this study. These parameters and the range in which these parameters were studied were arrived at after a thorough literature survey of neural network applications in the field of remote sensing. A total of 450 neural network classifications were performed (225 classifications each in one and two hidden layer categories). The overall accuracy was assessed for each classification with a testing sample size of 300 pixels per class. The results were analysed using simple graphical analysis carried out separately for one and two hidden layer categories. When effect of training sample size was studied, experiments for a given number of nodes (in case of two layers, the number of nodes were same in first and second layer) were grouped together. Since there were three values each of learning rate and momentum factor, there were nine different experiments. These were differentiated using three different colours and three different line types (Figs 2 and 3). A similar approach was adopted when effect of number of nodes was studied except that here experiments for a given training sample size were grouped together and plotted (Figs 4 and 5). However, when learning rate and momentum factor were studied, these produced 15 different experiments which were differentiated using five different colours and three different line types (Figs 6 to 9).

4.1 Effect of Training Sample Size

Five categories of training sample sizes were selected and classifications performed. In all, 45 classifications, each in one and two hidden layer categories using various combinations of the rest of the parameters were performed. These classifications were plotted on five different graphs, one each for a given number of hidden node, with training sample sizes on the X- axis and overall accuracy on the Y-axis [Figs 2(a) to 2(e) and 3(a) to 3(e)]. The following observations regarding the behaviour of training sample size were made:

- In general for one hidden layer category, overall accuracy shows improvement with the increase in training sample size. However, this improvement is not gradual and continuous. Between two sample sizes, it decreases temporarily at times but the frequency of decrease reduces at higher number of hidden nodes where the improvement in overall accuracy is continuous throughout with minor fluctuations. Hence, there exists a threshold value for sample size for different hidden node categories below which overall accuracy cannot give good results.

- A similar trend is noticed in case of two hidden layer category also. However, the curves in this case are relatively smoother than in the case of one hidden layer category. In other words, there is a near-continuous improvement in overall classification accuracy with the increase in sample size.

4.2 Effect of Number of Hidden Nodes

The number of hidden nodes in one hidden layer category were selected as multiples of the number of input nodes. Thus, 2, 4, 8, 12 and 16 hidden nodes (1/2, 1, 2, 3, 4^{th} multiples) were selected for one hidden layer category. The number of hidden nodes were kept same in both the first and second hidden layer for the two hidden layer category. Thus, the number of hidden nodes were 2,4,8,12 and 16 in each of the hidden layers.

Here again 45 classifications, each in one and two hidden layer category, were performed. These classifications were then sub-grouped training samplewise, for generating graphs. Hence, five different graphs corresponding to five training sample sizes were generated following a similar approach, as done earlier [Figs. 4(a) to 4(e) and 5(a) to 5(e)]. The following observations were made:

- The behaviour of number of hidden nodes is similar to training sample size. In general, there is an improvement in the overall accuracy with the increase in the number of hidden nodes. Here again, the improvement is not continuous but it fluctuates between two different number of hidden nodes.

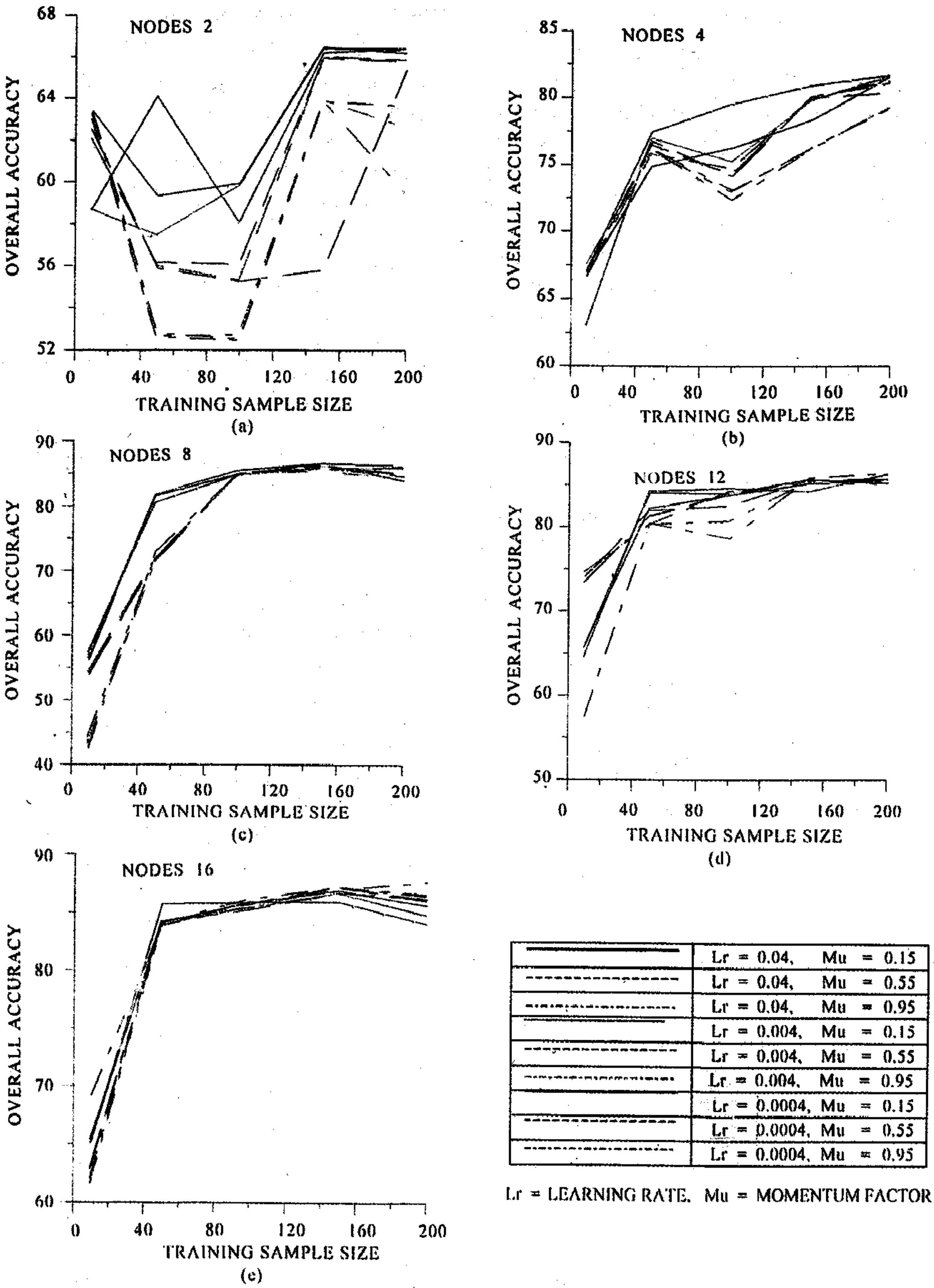

Figure 2. Effect of training sample size in one hidden layer category

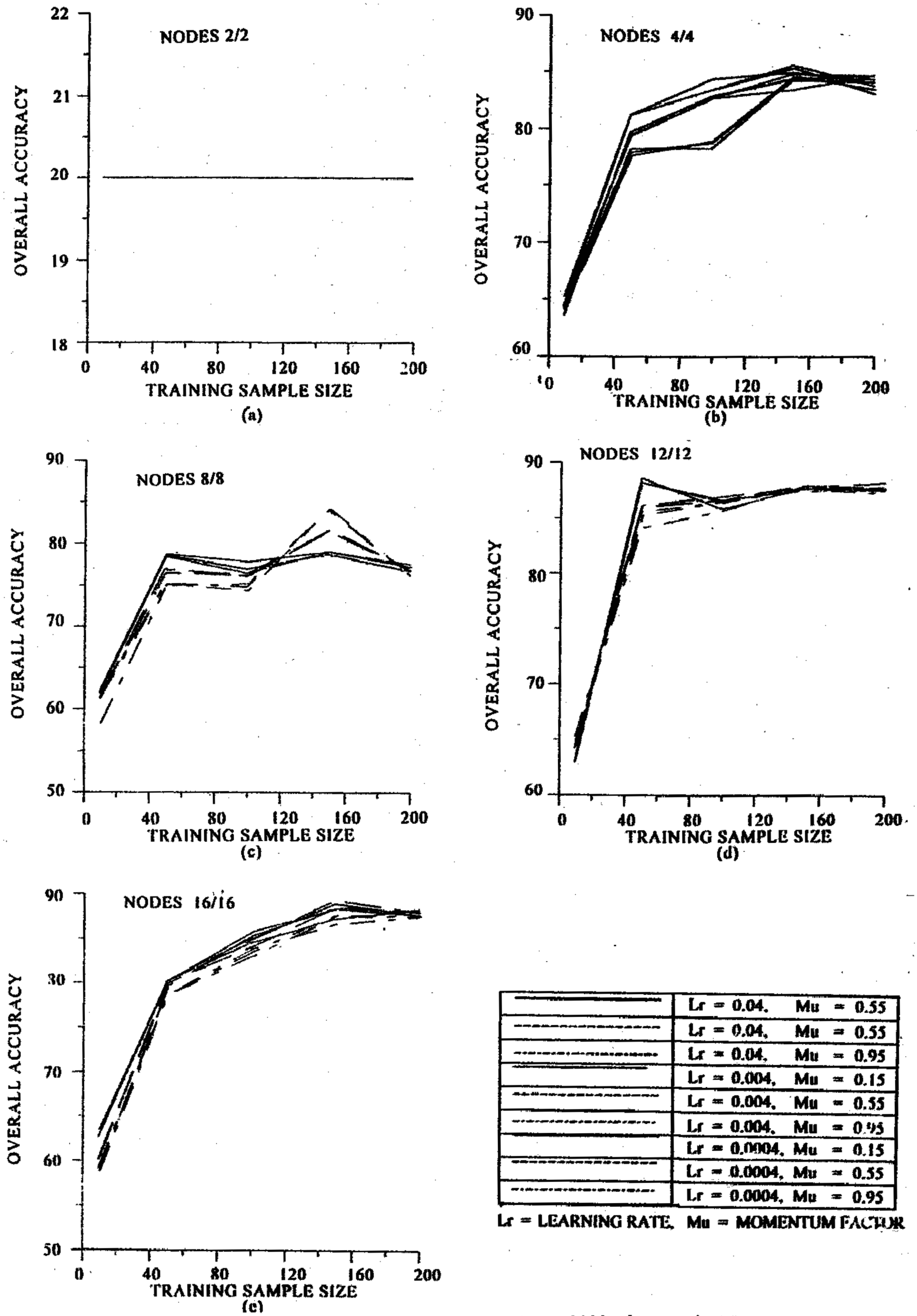

Figure 3. Effect of training sample size in two hidden layers category

This fluctuation is high at lower sample sizes and reduces on increasing the sample size.

- In case of two hidden layers, a similar pattern is observed. Here, performance is better as compared to one hidden layer category and contains smaller and lesser number of fluctuations.

- In case of two hidden layers, 2/2 hidden node category fails to classify at all. In general, it can be concluded that if number of hidden nodes is less than the number of input nodes, it gives poor classifi- cations. It is safe to deduce that a minimum of three times as many hidden nodes as the number of input nodes be selected for obtaining satisfactory classification results.

4.3 Effect of Learning Rate

For the present study, classifications were carried out with three different values of learning rates. The range of values for learning rate was fixed based on an empirical formula as given earlier. A total of 75 classifications, each in one and two hidden layer category, have been conducted. These classifications were then sub-grouped hidden nodewise. Thus, 15 curves in each graph for all the five graphs were plotted. When any one graph is considered, there appears to be very little increase in the overall accuracy with the decrease in learning rates for all sets of classifications (see each curve in any graph). However, when all three curves of any one sample size (shown in one colour) are compared with any other three curves for a different sample size in the same graph, it is noticed that the overall accuracy increases with the increase in sample size. But this is true only for lower values of learning rates. In other words, it appears that for higher overall accuracy, in case training sample size and number of hidden nodes are increased, then learning rate should be decreased by a corresponding value. A similar behaviour is seen in two hidden layer category. The miniscule improvement in performance due to decrease in learning rate for a given hidden node and sample size group may also be attributed to the fact that this range was fixed using an empirical formula which by itself had brought the learning rate closer to the desired value. However, this narrow range of learning rate and its dependency on training sample size and number of hidden nodes highlights the importance of finding an accurate learning rate value for any classification task. The observations are summarised as

- Learning rate appears to be dependent on the number of hidden nodes and also on the training sample size. Further, in order to achieve higher accuracy, as the training sample size and number of hidden nodes are increased, learning rate must be decreased by a corresponding value.

- For a given sample size and number of hidden nodes, there is a very narrow optimal range for learning rate, only within that range can a neural network classify and give satisfactory results.

4.4 Effect of Momentum Factor

Three momentum factors in the range 0 to 1 (0.15, 0.55 and 0.95) were used in these classifications. Here again 45 classifications, each in one and two hidden layer category, were conducted. Five graphs each for one and two hidden layers [Figs 8 (a) to 8(e) and 9(a) to 9(e)] were plotted. The behaviour of momentum factor for one layer with two hidden nodes category [Fig 8(a)] and 2/2 hidden nodes for two hidden layers category [Fig. 9(a)] is erratic and does not match well with the corresponding graphs plotted for other higher hidden nodes categories [Figs 8(b) to 8(e) and Figs 9(b) to 9(e)]. For a given sample size, (consider a small sample size of 10), an improvement in performance with increase in momentum factor can be seen in all the graphs. However in any one graph, it can also be seen that when sample size is increased, the performance decreases with increase in momentum factor. Now consider the graph for one layer category with four hidden nodes [Fig. 8(b)]. Here, it is seen that for small sample sizes, performance increases slightly on increasing momentum factor. But on increasing the number of hidden nodes, for all lower sample sizes, the behaviour of momentum factor remains erratic. However at higher sample sizes, it shows a pattern, e.g. for sample sizes exceeding 50 (say 100), for lower number of hidden nodes, the

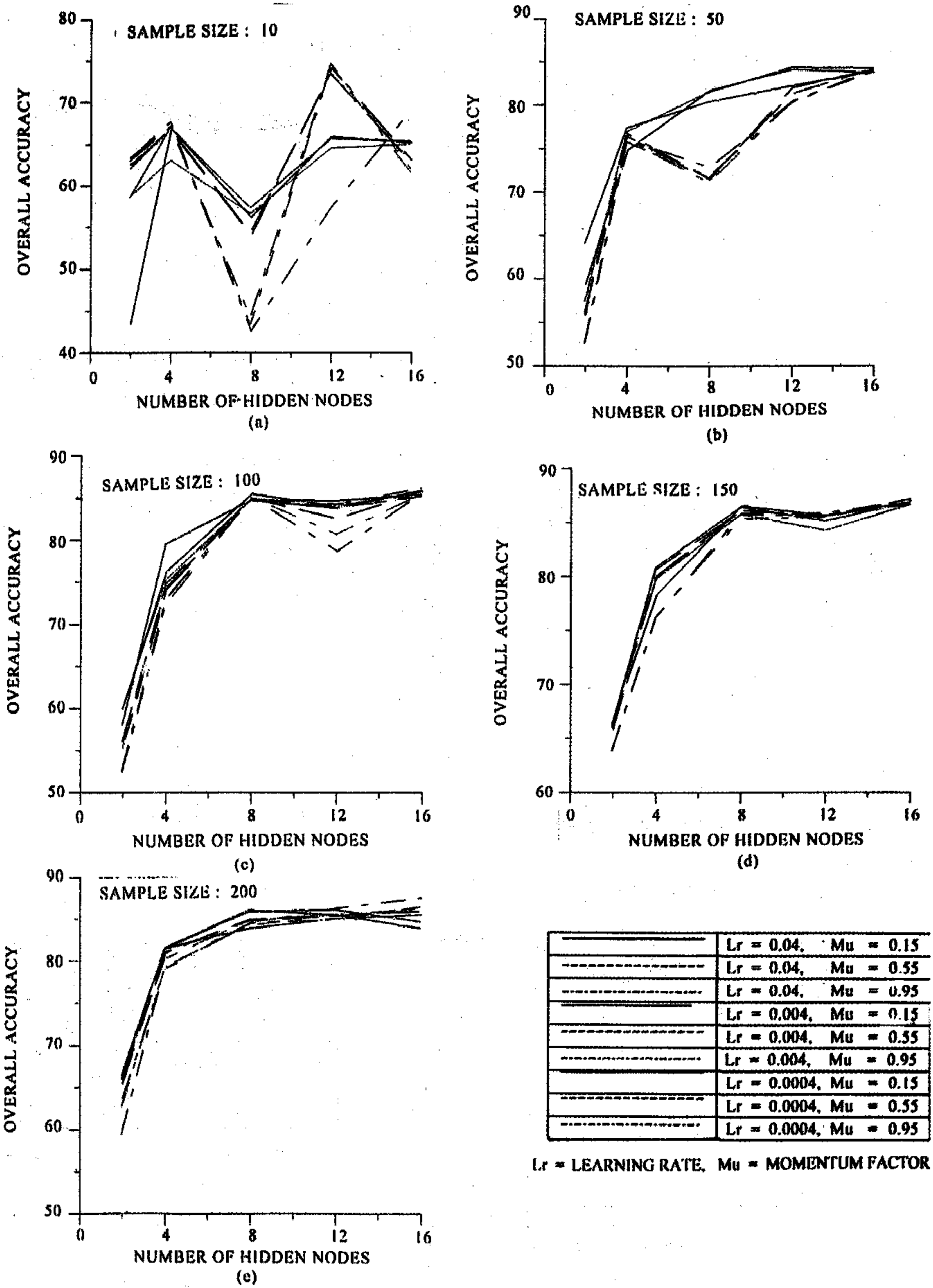

Figure 4. Effect of number of hidden nodes in one hidden layer category

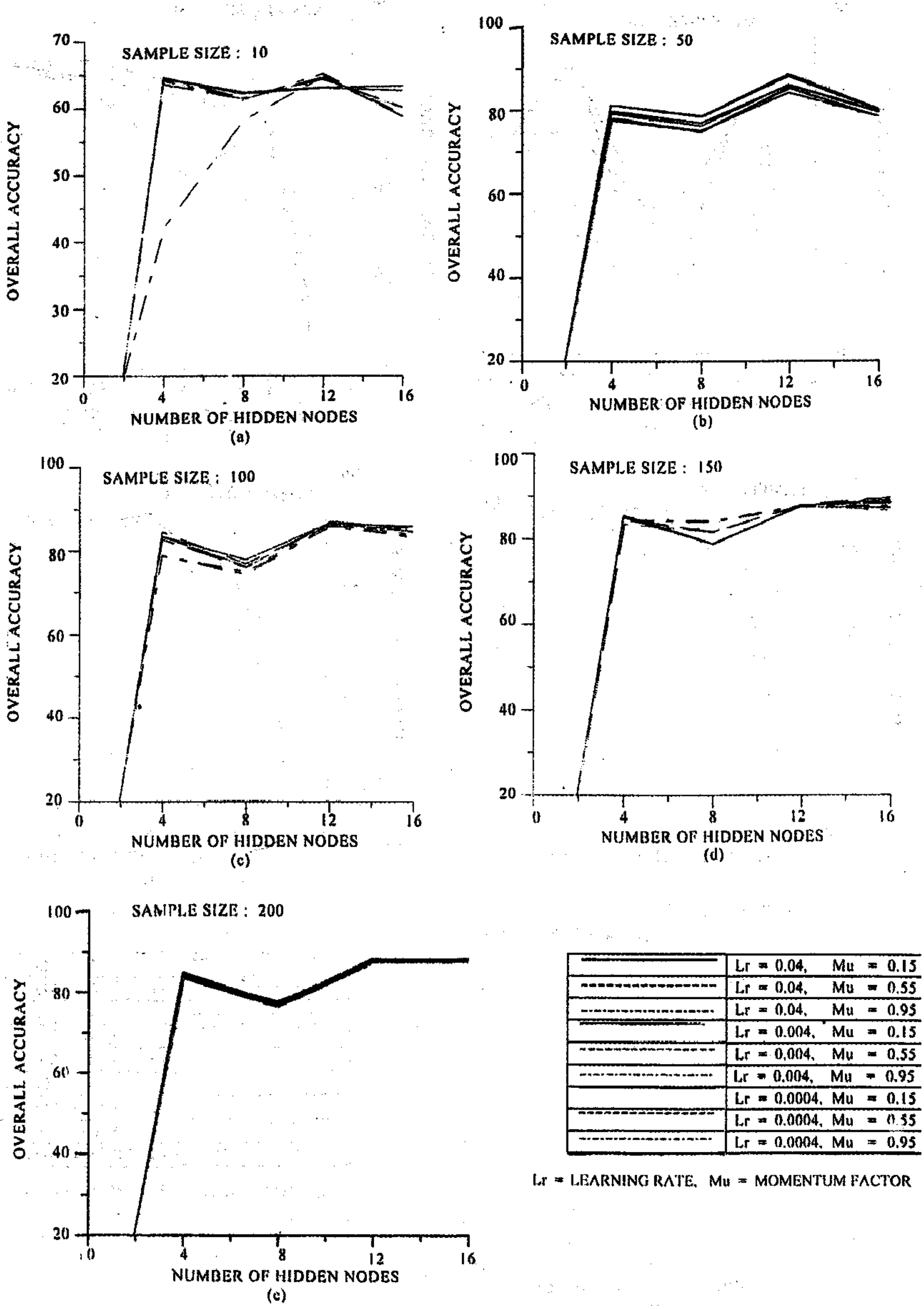

Figure 5. Effect of number of hidden nodes in two hidden layers category

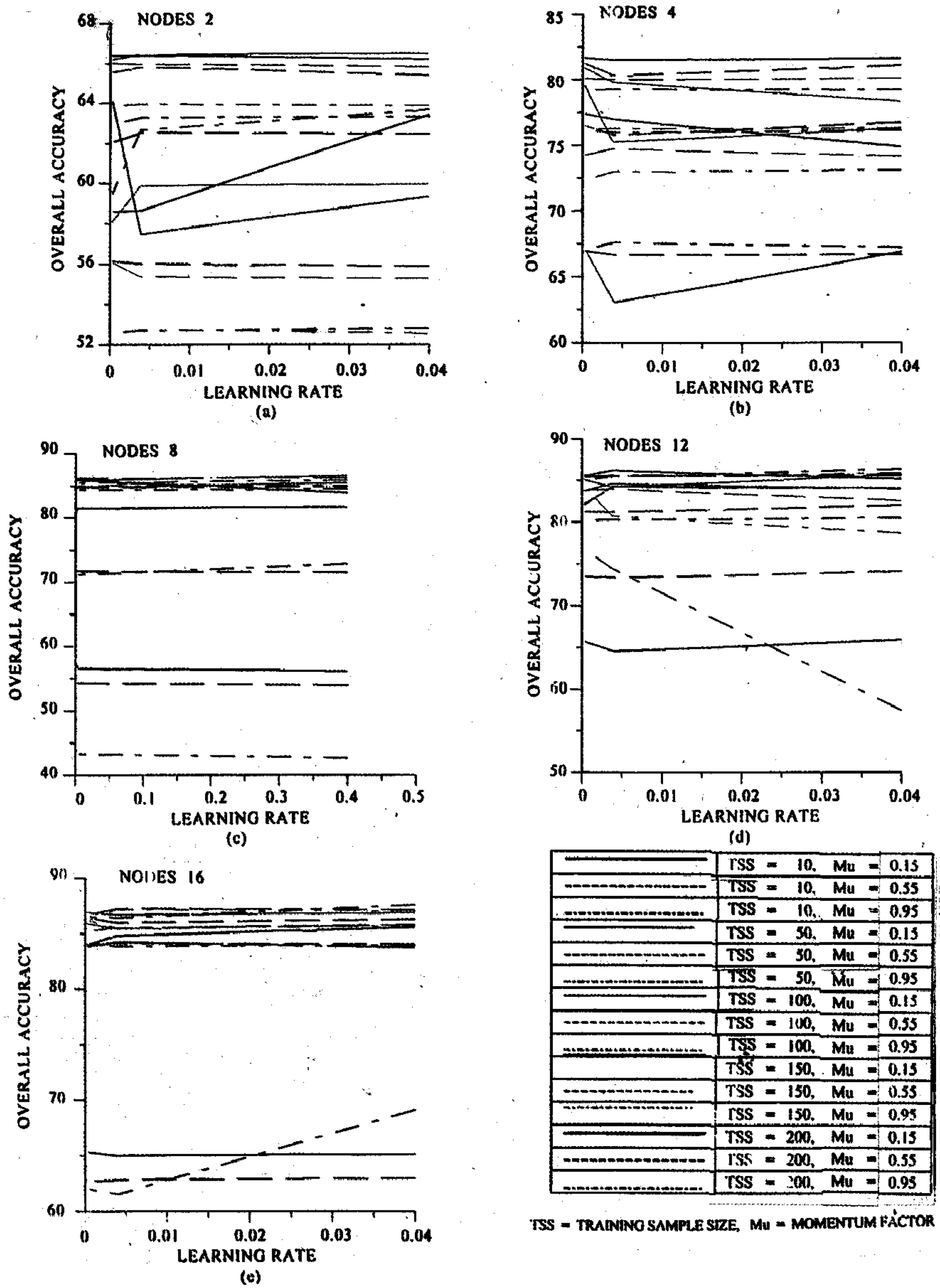

Figure 6. Effect of learning rate in one hidden layer category

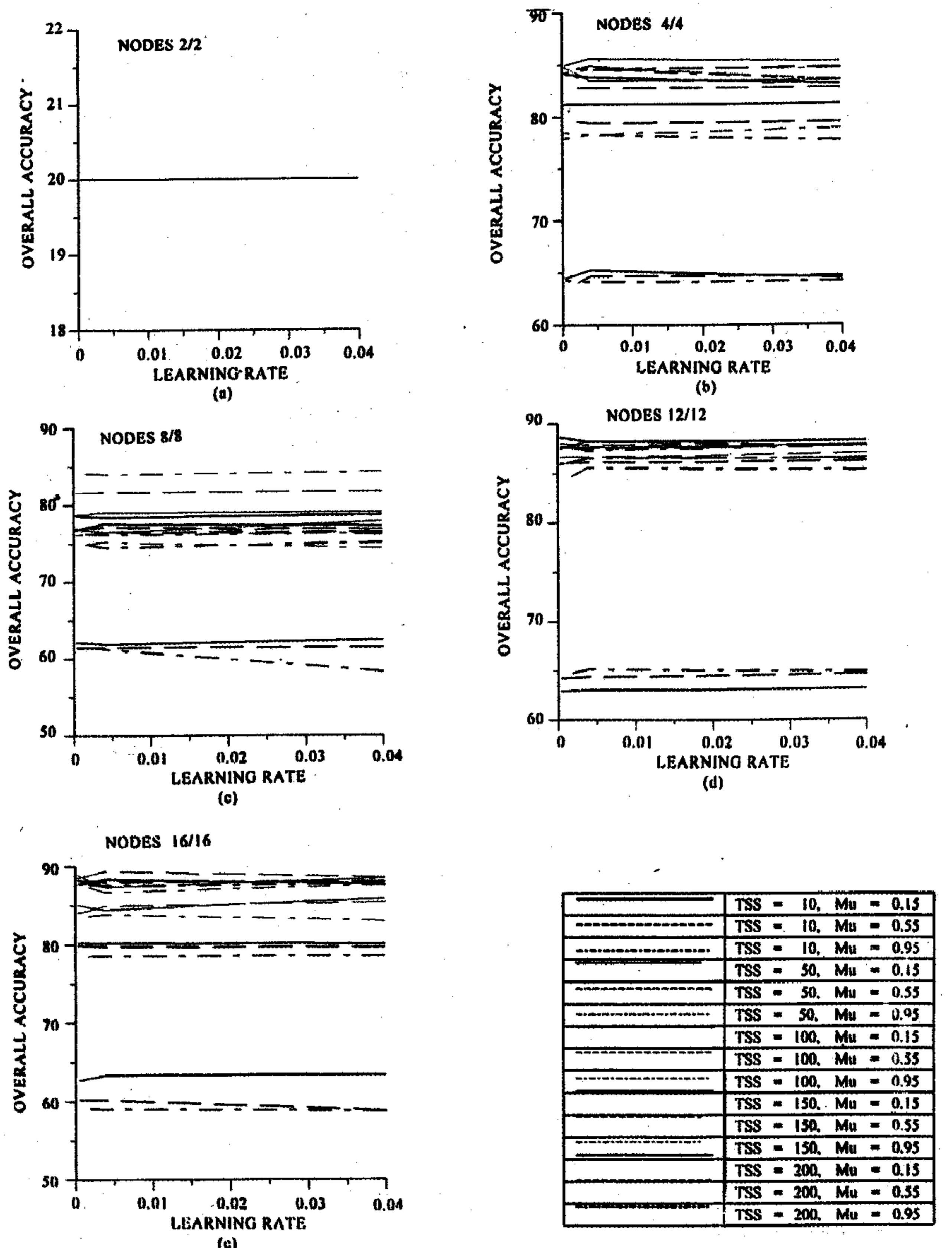

Figure 7. Effect of learning rate in two hidden layers category

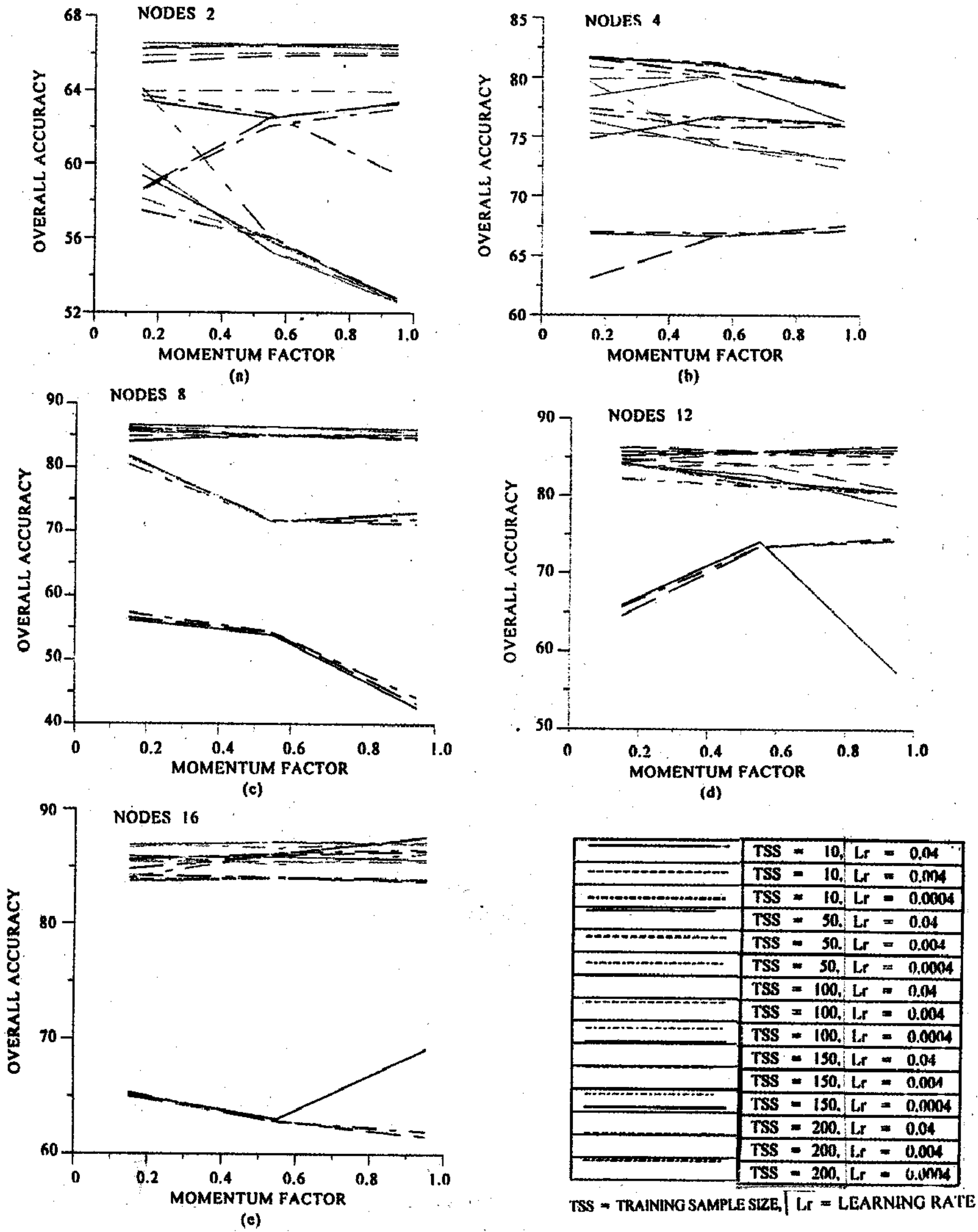

Figure 8. Effect of momentum factor in one hidden layer category

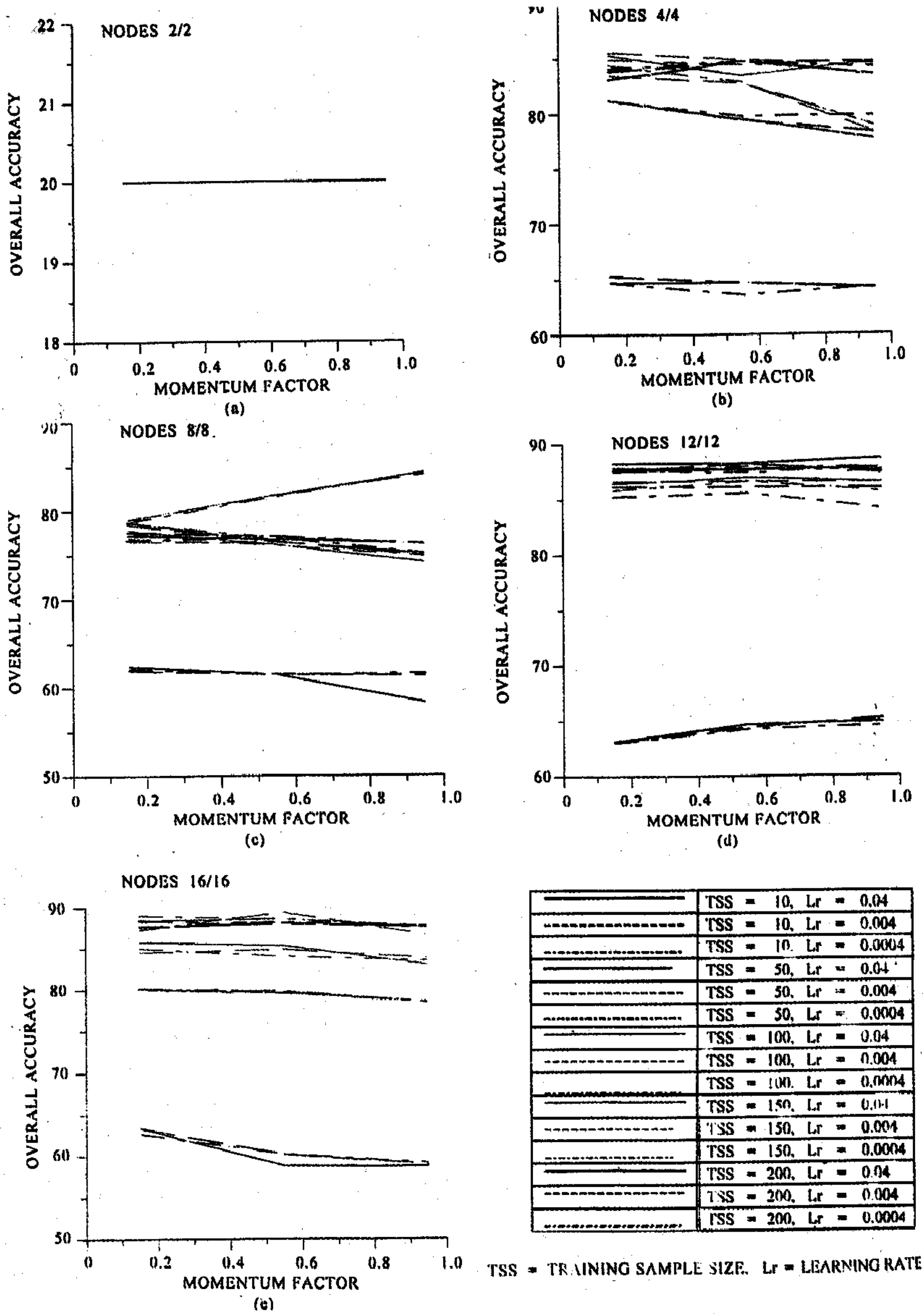

Figure 9. Effect of momentum factor in two hidden layers category

performance dips with increase in momentum factor and as the number of hidden nodes are increased, the performance gradually becomes constant and remains unaltered with increase in momentum factor. Therefore, it appears that beyond a threshold value of sample size, momentum factor is an inverse function of the number of hidden nodes. The behaviour is same for both one and two hidden layer categories. The main observations in this case are:

- The behaviour of momentum factor at all lower sample sizes remain erratic and does not improve even if number of hidden nodes are increased.

- For higher sample sizes exceeding 100 with lower number of hidden nodes (up to 8 or twice the number of input nodes), lower values of momentum factor (< 0.5) give better results while with higher number of hidden nodes (12 or three times the number of input nodes), higher values of momentum factor, (> 0.5) give better results.

Based on the above analysis a set of values for the above parameters as given in Table 1 can be recommended for achieving classifications with reasonably high accuracies.

Table 1. Recommended values for various parameters

No. of hidden layers	No. of hidden nodes	Training sample size	Momentum factor
1	4	160 or higher	0.25
	8	100 or higher	0.40
	12	100 or higher	0.95
	16	50 or higher	0.95
2	4/4	120 or higher	0.25
	8/8	80 or higher	0.40
	12/12	60 or higher	0.95
	16/16	40 or higher	0.95

Learning rate – 0.0004
Overall accuracy expected – 90 per cent or higher
Remarks – Training for 10,000 –15,000 epochs may be required.

4.5 Classification of an IRS-1B LISS II Image

Finally, to validate the results obtained an IRS-1B LISS II image of Roorkee and its surrounding areas was classified using neural network classifier. Classifications were successful with various combinations as listed in Table 1.

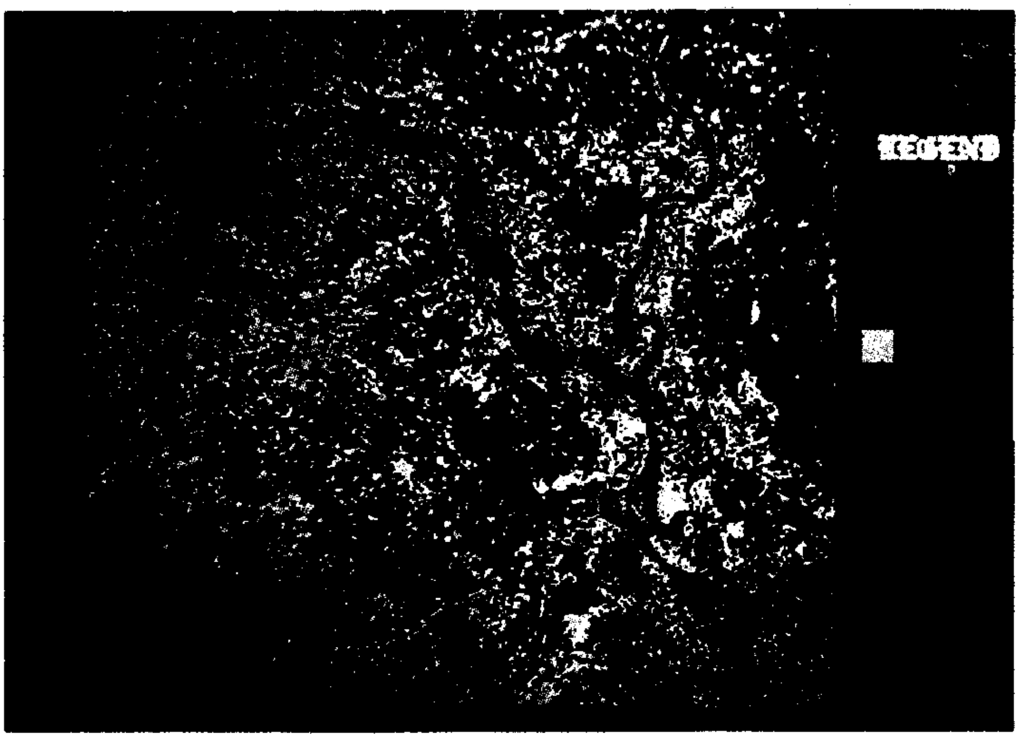

Figure 10. A classified image of an IRS-IB LISS II image of Roorkee and its surrounding area

However, best classification with 92.53 per cent accuracy was obtained for a two-layered neural network with 16/16 nodes in each layer, training sample size of 200 pixels per class, learning rate of 0.000267 and momentum factor 0.95. The classified image is shown in Fig. 10.

5. CONCLUSIONS

Five parameters, namely a number of hidden layers, and hidden nodes, training sample size, learning rate and momentum factor were investigated with the help of input data extracted from an IRS-1B LISS II 512×512 image of Roorkee and its surrounding areas in four spectral bands three in visible and one in near-infrared region). A total of 450 experiments were conducted and the results analysed graphically. On the basis of results obtained, certain conclusions have been arrived at. However, there is a requirement for a wider series of experiments and a more intricate analysis to determine the exact relationship among these parameters.

REFERENCES

1. Richards, J.A. Remote sensing digital image analysis – An Introduction. Australian Defence Force Academy, Campbell, Australia, 1993.
2. Katariana, J. Segment-based land-use classification from SPOT satellite data. *Photogrammetric Eng. Remote Sens.*, 1994, **60**(1), 47-53.
3. Mathieu-Marni, Lemario & Berthod. Removing ambiguities in a multispectral image classification. *Inter. J. Remote Sens.*, 1995, **17**(8), 1493-04.
4. Wang, G. Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Eng. Remote Sens.*, 1990, **56**(7), 1163-69.
5. Foody, G.M. Approaches for the production and evaluation of fuzzy landcover classification from remotely sensed data. *Inter. J. Remote Sens.*, 1996, **17**, 1317-40.
6. Kanellopolllos, T. & Wilkinson, G.G. Strategies and best practice for neural network image classification. *Inter. J. Remote Sens.*, 1997, **18**(4), 711-26.
7. Gong, P. Integrated analysis of spatial data from multiple sources using evidential reasoning and artificial neural network for geological mapping. *Photogrammetric Eng. Remote Sens.*, 1996, **62**(5), 513-23.
8. Foody, G.M. & Arora, M.K. An evaluation of some factors affecting the accuracy of an artificial neural network. *Inter. J. Remote Sens.*,1997, **18**(4), 789-10.
9. Paola, J.D. & Schowengerdt, R.A. A review and analysis of back propagation neural networks for classification of remotely sensed multispectral imagery. *Inter. J. Remote Sens.* ,1995, **16**.