

Alternate Neural Network Models as Supervised Classifiers for Satellite Data

P. S. Sangle^{1*} and S. M. George²

¹Information Technology, National Institute of Industrial Engineering, Vihar Lake, Powai, Mumbai 400087, India

²Care Sustainability, A-29, AWHO, Dara Enclave, Sector 9, Nerul, New Mumbai 400 706, India

ABSTRACT. Investigations on the effectivity of different neural network architectures, viz. number of hidden neurons, constrained neuronal connections (hierarchical network), and fuzzy aggregation based synaptic neuronal functions (fuzzy neural network) for satellite data classification are presented. Performance of networks trained with varied number of training sizes for classification in large spatial extensions are used as illustration through two case studies, viz. land use/land cover classification of Delhi Ridge and species classification of floral resources in Shimla and Chopal regions in India. The results have been compared with statistical methods.

Keywords: Floral species classification, fuzzy neural network, hierarchical network, land use classification, Neural Network, satellite imagery

1. Introduction

Use of neural networks to remote sensing applications continues to receive attention in research, despite apprehensions regarding its effectivity in all situations. The effectivity of artificial neural networks (ANN) in satellite image classification is due to its intrinsic ability to generalize the non-dependence on statistical distribution of data, and capability to form highly nonlinear decision boundaries in the feature space (Lippmann, 1987; Hush et al., 1993).

Also, ANN models do not require any a priori knowledge of the class distribution in a remotely sensed image unlike the statistical classifiers wherein distributions may change with data class as well as with data structure. The ANN models are distribution free (Benediktsson et al., 1990). In multisource remote sensing data classification, ANN does not require any a priori specifications of the weight of each data source models. The procedures involved with neural network spectral classification are more closely associated with "supervised" classification as opposed to "unsupervised" classification techniques. With supervised techniques, the analyst initially defines several training areas in the imagery, which are representative of the desired classes.

Lippmann (1987) concluded that neural network classifiers are non-parametric and are more robust when distributions are strongly non-Gaussian. Neural networks have been used in a number of studies to classify remotely sensed satellite data (Benediktsson et al., 1990; Tzeng et al., 1997). Hepner and others (Hepner et al., 1990) examined the application of neural networks for land cover classification and demonstrated that the ANN classifies imagery better than conventional

supervised statistical classification procedures. The accuracy of classification using ANN models with single training site-per-class was found to be comparable to the accuracy obtained using statistical methods with four training-sites-per-classes and the statistical supervised classification using the single minimal training set was very inferior to the ANN classification. These results suggest that the neural network based techniques have a potential advantage of minimizing training sets and hence ground truth data requirement. However, the investigations thus far are not comprehensive and do not cover complex cases of large up-scaling ratios, viz. ratio between numbers of classification pixels to training pixels. Pao (1989) hypothesized that additional non-linear processing of the data before presenting it to the network, can result in linearly separable classes that dramatically decrease effectivity of training.

Dreyer (1993) calculated a number of textural features based on gray-level statistics and found that the use of these features increased the accuracy of a 'field' class, has without any effect for 'urban' and 'water' classes while decreasing the accuracy in case of a 'forest'. Key et al. (1990) used texture calculations such as second momentum and entropy to produce a single texture measure for each pixel in the classification of land cover and cloud types, with classification results superior to those using spectral pixel values alone. However, Geaga et al. (1992) have reported that the use of seven principal component images instead of the original LANDSAT spectral values does not result in faster training or higher accuracy. Yoshida et al. (1994) proposed a neural network classification method for remotely sensed data analysis in order to improve neighborhood relation between pixels and to decrease error probability for pattern classification. Civco (1993) investigated the effectiveness of training with a single mean vector for each class in land use classification and con-

* Corresponding author: purnima@nitie.edu

cluded that non-inclusion of variability in the data leads to spurious convergence of the network. Paola et al. (1995) have demonstrated the advantages of the neural network method over traditional classifiers for classification of remotely sensed multi spectral imagery.

Despite the distinct advantages offered by neural network methods, their success is contingent upon further research to achieve improvements in convergence speed, sensitivity to architecture and fewer requirements of initializations. Several investigations to remedy these drawbacks (Bernard et al., 1997; Paek et al., 1997) are reported in the literature. Significant reduction in training time was achieved by Tzeng and others using a dynamic learning neural network and Kalman filtering technique (Tzeng et al., 1997).

Chen et al. (1997) adopted a neural network classification approach using non-backpropagation learning leading to a fast convergence and built-in optimization function. Such a result demonstrates that fractal information significantly improves the discrimination capability of heterogeneous area such as an urban region, while it degrades accuracy for homogeneous areas such as open water.

Foody et al. (1997) investigated the effects of the dimensionality of the remotely sensed data, the neural network architecture, and the characteristics of the training and testing set on the accuracy of classification using neural networks. The salient conclusions were:

- Variation in the dimensionality of the data set, as well as the training and testing set had a significant effect on classification accuracy;
- The network architecture, specifically the number of hidden neurons and layers, did not have a significant effect on classification accuracy.

Paola et al. (1997) found that there was a range of optimal hidden layer sizes below which the accuracy is decreased and above which the training time is increased. However, they have also reported that, for a fairly wide range, the hidden layer size made little difference to the final classification accuracy.

The suitability of Back Propagation Neural Network (BPNN) for classification of remote sensing images is proposed by Aria et al. (2003). Their approach consists of three steps to classify IRS-1D images is proposed. In the first step, features are extracted from the first order histogram measures. The next step is feature classification based on BPNN, and in the finally step the results are compared with the maximum likelihood classification (MLC) method.

In summary, the review of the literature indicates that despite the potential benefits of approaches based on neural networks, the entire potential of ANN is yet to be harnessed in its applications to remote sensing.

Following conclusions are made from the literature review on neural network approaches for classification of satellite imageries:

- training algorithms for classification are slow and thus, noneffective in practical implementation though accu-

rate in many cases;

- wide variation exists in effectivity of classification with training algorithms and architectures;
- incorporation of the expert knowledge and the ancillary information is possible;
- dependence of training size on accuracy of classification is not conclusive, and is based on small spatial extensions, and lack validation using ground truths;
- absence of demonstration of effectivity of neural networks for finer classification such as forest classification based on dominant species, and built-up land use of different densities.

This study investigates the performances of neural network trained with varied number of training sizes, and its usage in classification for different spatial extensions and fine classification categories. The research will demonstrate the application of neural networks trained with considerably small training size in the analyses of large spatial areas and fine classification with two case studies, viz. land use/Land cover classification of Delhi Ridge in India, and species classification of floral resources in Shimla and Chopal Regions in India. The results will be compared with the statistical methods, viz. the maximum likelihood classifier, the Mahalonobis distance classifier, and the minimum distance classifier.

The paper also reports the investigations carried out with different neural network architectures viz. different number of hidden neurons, constrained neuronal connections (hierachical networks) and fuzzy aggregation based synaptic neuronal functions (fuzzy neural network).

2. Field Survey and Ground Truth Data Collection

2.1. The Ridge Forest Area in the National Capital Territory of Delhi

The Delhi Ridge is the northern extension of Aravalli hill range, the oldest mountain chain in the country, entering the National Capital Territory of Delhi in the Southeast at the Tughlakabad – Bhatti mines – Dera Mandi axis and terminating in the north near Wazirabad on the right bank of river Yamuna.

The Delhi Ridge is characterized by tropical, thorny, secondary forest. The bulk of the vegetation consists of co-dominant, spinous scrubs and trees, and is capable of drought resistance. The ground flora is seasonal, and the thorny shrubs occur in widely spaced clumps supporting a number of twines and climbers. The species composition has undergone changes over the decades, and the climax forest has been replaced by thorny secondary forest with *Prosopis juliflora* as the dominant tree species.

The topography of the ridge is undulating with uniformly distributed drainage network. The ridge has 16 perennial tanks and 56 pits, which are good groundwater recharge points. The drainage is hindered by about 30 hectares of built up area mainly on the central ridge. Delhi Ridge is divided into four administrative regions viz. Northern, Central, Southern Cen-

tral and Southern ridges.

The Northern Ridge covers a total area of 87.0 hectares out of which 42% of landmass is covered by built up area. The scrub, fallow and degraded forests together cover an area of 22 hectares and dense forest covers an area of about 20 hectares. A major portion of the forest is managed and the eastern flank having extensive cover of *Prosopis juliflora* with *Acacia nilotica*. The Central Ridge envelops an area of 864 hectares. The built up area and cropland constitute 28 and 3% of the landmass respectively. The dense forests cover 116 hectares. There are 13 natural perennial tanks and 34 water holding pits in this part of the ridge. The South Central part of the ridge constitutes an area of 626 hectares out of which 15, 7, 20, and 20% of landmass is under land use categories of built up, cropland, scrub and fallow respectively. The vegetation in this central part is very poor and is limited to Sanjay Van that covers only 37% of the landmass. The undergrowth in Sanjay Van is very poor.

The Southern Ridge envelops an area of 6200 hectares out of which built up, plantation; crop, scrub and fallow land use categories constitute 5, 10, 7, 48, and 28% of landmass respectively. The sandy area extends over 84 hectares; and there is no forest vegetation in this part of the ridge. A mosaic of *Prosopis juliflora* with occasional bushes of *Carissa spinarum*, and *Zizyphus mauritiana* characterizes the vegetation. The vegetation in the Asola Wild Life Sanctuary located in the Southern Ridge is mainly scrub interspersed with plantations.

2.2. Shimla and Chopal Regions

The Shimla and Chopal regions are located in three administrative districts in the state of Himachal Pradesh, namely, Solan, Shimla, and Shimla. This region is mountainous and constitutes the headwaters of the Yamuna river tributaries. The river system in this region is constituted by Pabbar and Giri Rivers. Alpine grasslands to grasslands of semi arid tropics constitute the biological features. The region and its contiguous areas are characterized by wide climatological and topographical variations. The forest types and its composition vary widely from the altitudes 34 meters to 250 meters across the region.

The Shimla region has four forest types viz. upland hardwoods mixed with coniferous, mixed coniferous, blue pine, fir/spruce besides exposed rocks, barren lands; and agricultural land and habitation. The details of four forest types that were based on the dominant floral species classification are delineated in Table 1. The detail of codes viz. A, B, C, D, E, and F used for representing the types of land cover forming the classification categories was presented in Table 1.

Field surveys of both Delhi Ridge and Shimla & Chopal Regions were carried out within the experimental area with the grid sizes of 50 m × 50 m and 250 m × 250 m respectively. The purpose of field survey was to establish the main characteristics and variability of the land cover and dominant species based forest classification. Field survey also provided good ground truth data for training and testing the classifiers.

Table 1. Dominant Floral Species in Shimla and Chopal Regions

Class Code	Class Name
A	Upland hardwoods mixed with coniferous
B	Exposed rocks, barren lands
C	Mixed coniferous
D	Agricultural land and habitation
E	Blue pine (Kail)
F	Fir, spruce

The sizes of the grids were chosen to ensure that each grid was large enough to be easily identifiable on the satellite imageries. The land cover and dominant species in each grid were recorded in detail, which included the main land cover class and dominant species, the forest types and crown density, the topographical and geological conditions and man made features including roads.

The boundaries of each grid were also recorded precisely on topographic maps. The survey was designed to capture the details of as many different kinds of land cover and vegetation details as possible. All of the field data after requisite digitization were managed in a Geographic Information System (ARC/INFO Version 6.1) as geocoded polygon coverage with appropriate polygon attribute information. The association of ground truths obtained in grids with those of remote sensing information in pixels has been established using union operation in Geographical Information System (ARC/INFO 6.1).

Ground truths in the regions - Ridge Forests in National Capital Territory of Delhi, and Shimla and Chopal forests in India – have been grouped as training and testing sets for ascertaining the effectivity of neural network models for land use classification and species classification of crown cover respectively.

3. Satellite Data

To carry out the land cover classification in *Delhi Ridge*, data from the Indian Remote Sensing Satellite (IRS IA), LISS II (Linear Imaging Self-Scanner) with the ground resolution of 36.5 sqm. is used. The LISS II digital image is of August 1995 which has seasonal correspondence with the times of field surveys. The image was obtained in raw digital format, and was analyzed and geocoded to the field data in Polyconic projection system and further processed using Digital Image Processing software *ERDAS Imagine 8.2*. Likewise, the dominant species based classification of floral resources in *Shimla* and *Chopal* regions, was carried out with LISS I data from Indian Remote Sensing Satellite (IRS IB) having ground resolution of 72.5 sqm. The LISS I digital image was also obtained in raw format (not geo-coded). Geometrically Registration using the image using ground control points was conducted out for both the images to obtain pixel registration accuracy (RMS error 0.5). The data from Indian Remote Sensing Satellites (IRS IA & IB with LISSII and LISS I) sen-

sors cover the 0.45 - 0.52, 0.52 - 0.59, 0.62 - 0.68, and 0.77 - 0.86 μm spectral ranges. The spectral (band) information form inputs for the neural network. Training and testing set constitutes sets of data of different pixels corresponding to different ground locations.

The data set was split, by random selection, into independent training and testing sets comprising different pixels of each class. The distribution characteristics of spectral responses are chosen to be different for different spectral bands and classes. The distribution characteristics are also chosen to be different for training and test data sets. The choice of training and testing data sets is chosen to test the following hypotheses:

- Neural networks based classification methods are not dependent on statistical distributions of spectral response of different classes in the test and training data sets;
- Neural networks based classification methods are not dependent upon choice of training data set sizes and spatial characteristics, and upscaling ratio;
- Neural Networks based classification methods are accurate with small sizes of training data set and large upscaling ratio.

4. Neural Network Models

4.1. Neural Network Models

In the present study, the following neural network models for classification of satellite imageries of Delhi Ridge and Shimla and Chopal regions in India have been investigated:

- Crisp Neural networks
- Neural network with fuzzy synaptic operation
- Neural networks with constrained connections based on spectral relevance (hierarchical neural networks)

The classification results obtained using the models are compared to those obtained using parametric decision rule, viz. Maximum likelihood classifier, Mahalonobis distance classifier and Minimum distance classifier.

The issue of choice of different representations including data transformation is not addressed in this research due to, conclusive negative evidence reported in the literature. However, the investigations on effectiveness of different models of neural networks, dependence on architecture (number of hidden neurons), size of training set, and ratio of the size of testing and training data set (scale-up-factor) have been the focus of this investigation. Scale-up-factor provides a measure of the geographical extent over which a neural network trained with a training set shall be usable for classification.

4.1.1. Crisp Neural Networks

The crisp neural network model was characterized by three-layered structure with simple backpropagation learning algorithm without momentum term (Werbos, 1974; Rumelhart

et al., 1986) and sigmoidal synaptic function. The architecture of the neural network model is varied to identify the model with higher accuracy in learning and smoother convergence. The learning rate is similarly varied to achieve improvements in convergence.

In the case of crisp neural network, the spectral responses of the four channels (bands) constitute the input to the neurons in the input layer and each output neuron represents a class. The input and output neurons are allowed to take continuous values and the value of output neuron that is between 0 and 1 is taken to represent confidence level of the class assignment. In the case of crisp neural networks, the number of neurons in the single hidden layer is varied from 1 to 8. Convergence characteristics have been measured for all different training and prediction sets. The convergence of the neural networks with six hidden neurons in the case of land use classification in Delhi Ridge and with eight hidden neurons in the case of Shimla and Chopal regions are found to be the best. Within 10,000 iterations, and convergence has been faster and smoother for small sizes of training sets.

4.1.2. Fuzzy Neural Networks

Neural information processing comprises of three basic steps: synaptic processing, aggregation of the synaptic information by the neuron; and encoding of the aggregated information into the neural activation. Fuzzy logic fused with neural networks promise a system to deal with the cognitive uncertainties in a manner more like humans.

The following are salient alternative approaches to fuzzy neural networks:

- To incorporate fuzzy aggregation of information arriving at each node (Bezdek, 1991)
- To introduce fuzziness into the input data, weights and target vector (Yamakawa, 1989)

The input-output architecture of fuzzy neural networks has been chosen with four input neurons and one output neuron. Different ranges of values at the output neuron representing classes are delineated in Table 2. The fuzzy neural neural networks, in this study, use additive-operator based aggregation operator and back propagation learning algorithm. In the case of fuzzy neural networks, the parameter governing the relative importance of and or operations for neuronal aggregation has been varied, in addition to the number of hidden neurons to study the relative performances.

4.1.3. Hierarchical Neural Networks

In a decision making situation, where inputs to the neural network are of different hierarchical levels and can be consistently related to one another, the massively connected multi-layered networks, crisp or fuzzy, may be redundant. In such cases, the hierarchical network is considered as suitable choice (Krishnapuram and Lee, 1992). The fuzzy hierarchical aggregation networks are fuzzy extensions of hierarchical networks where neuronal aggregation is performed using fuzzy connectives.

The hierarchical aggregation is frequently encountered in decision-making processes that involve aggregation of criteria at several levels.

Table 2. The Representation of Output Class in Fuzzy Neural Networks

Class Name	Representation of Output
Dense forest	0.8
Open forest	0.7
Degraded forest	0.6
Built-up Area	0.5
Forest blank	0.4

In hierarchical networks, in order to aggregate and propagate degrees of satisfaction of criteria, proper type of connective at each node is to be chosen. The type of aggregation chosen at each node depends on the type of behavior the network is expected to exhibit. Krishnapuram and Lee studied the utility of various fuzzy connectives in fuzzy aggregation networks. The iterative algorithm introduced by them determines the aggregation function and its parameter at each node of the network. The behavior of the network can also be modified depending on the decision making situation and convergence of the algorithm in few sample cases is also illustrated (Krishnapuram and Lee, 1992). The hierarchical aggregation networks require:

- Hierarchical system structure
- Known neuronal aggregation function

In the present study, in the case of hierarchical neural networks, the spectral responses of the four channels (bands) constitute the input to the neurons in the input layer; and each output neuron represents a class. The input and output neurons are allowed to take continuous values; and the value of output neuron between 0 and 1 is taken to represent confidence level of the class assignment.

The configuration of the hierarchical neural network incorporates prior knowledge regarding spectral relevance. The three neurons in the hidden layer represent vegetation, mixed vegetation, and non-vegetation classes of land cover. The models for crisp neural networks for land cover classification, neural network with fuzzy synaptic operation, and hierarchical networks are presented in Figures 1 to 3 respectively.

4.2. Parametric and Non Parametric Decision Rules

A parametric decision rule is trained by the parametric signatures. These signatures are defined by the mean vector and covariance matrix for the data file values of the pixels in the signatures. When a parametric decision rule is used, every pixel is assigned to a class since the parametric decision space is continuous (Kloer, 1994).

A nonparametric decision rule is not based on statistics; therefore, it is independent of the properties of the data. If a pixel is located within the boundary of a nonparametric signa-

ture, then this decision rule assigns the pixel to the signature's class. Basically, a nonparametric decision rule determines whether or not the pixel is located inside of nonparametric signature boundary.

In the present study, the classification results obtained using the models are compared to those obtained using parametric decision rules, viz. the Maximum likelihood classifier, the Mahalanobis distance classifier and the Minimum distance classifier.

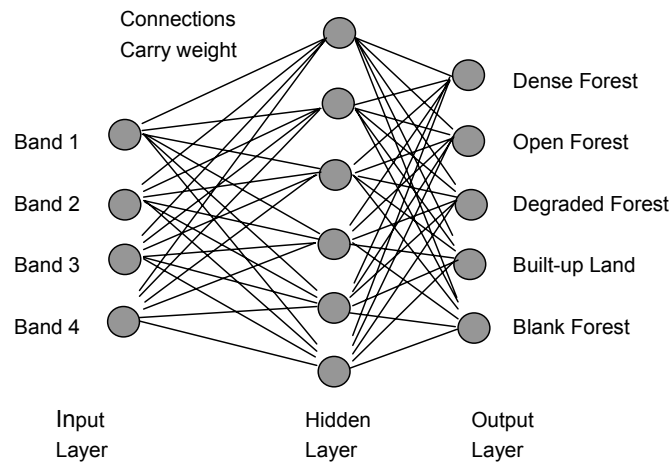
5. Evaluation of Classification Accuracy

Many methods for assessing classification accuracy have been proposed (Congalton, 1991; Foody, 1995). Lewis and others (Lewis et al., 2000) proposed a generalized confusion matrix for assessing area estimates from remote sensing imagery. Ideally, classification accuracy should be expressed in the form of a single index which is readily interpretable and which allows the relative performance of different classifications to be evaluated. The most widely used measures are derived from classification confusion or error matrix. This matrix gives the predicted and actual classes of membership for a set of pixels sampled from the classified image or the test data. In confusion matrix, main diagonal shows those pixels, which have been allocated correctly whilst the off-diagonal elements represent incorrect allocations. If desired, this could be calculated for individual classes from the producer's and user's perspectives (Story et al., 1986). In this investigation, classification accuracy computed using the confusion matrices for each class, is used for assessing the effectivity of classification methods. The confusion matrix lists the correct classification against the predicted classification for each class. The number of correct predictions for each class falls along the diagonal of the matrix. All other numbers are the number of errors for a particular type of misclassification error. For example, in Table 4 Dense forest class is correctly classified as 402 times, but is erroneously classified as Open forest class 17 time(s), Degraded forest class 35 times, Builtup area class as 4 time and Forest Blank class as 25 times and unclassified pixels as 33 times. The accuracy of dense forest is 77%. Similarly, the Open forest class is correctly classified as 562 times, but is erroneously classified as Dense forest class 7 time(s), Degraded forest class 1 time, Built-up area class as 6 times and Unclassified as 19 times. The accuracy of Open forest is 94%. In the present study, based on the confusion matrix the individual accuracy of each class has been measured and also the average accuracy has been measured.

6. Results and Discussion

The current research was restricted to the following objectives:

- Comprehensive study of effectiveness of supervised neural models for the classification of satellite imageries, and comparison of classification accuracies with those



Note: Network is fully-connected; for clarity not all connections are shown here

Figure 1. Architecture of crisp neural network classifier for 5-class forest and land cover classification.

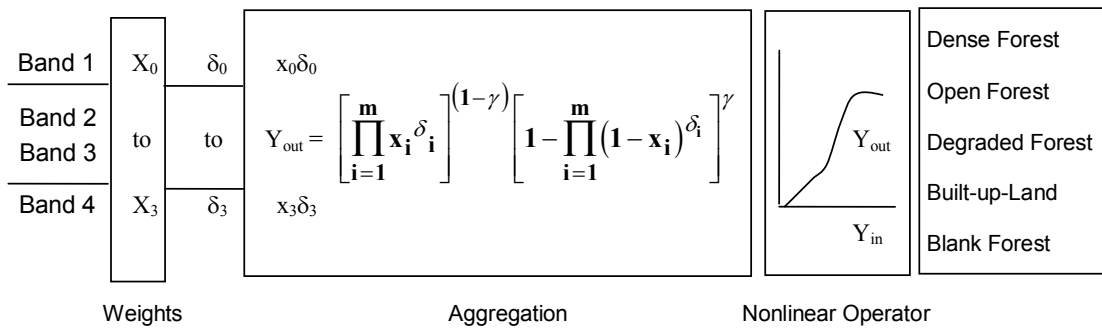


Figure 2. Schematic diagram of fuzzy neural networks based classifier delineating the configuration of input and output layers.

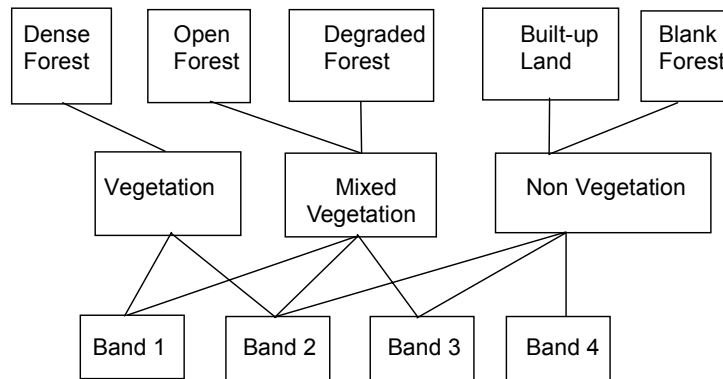


Figure 3. Architecture of hierarchical network classifier for 5-class forest and land cover classification.

obtained with statistical classifiers;

- Investigations on different sizes of training sets, scale-up factor ratios, and neural network architectures in terms of classification accuracy and convergence time.

The investigations have been conducted for two cases of analysis: (i) Land use/Land Cover classification in Delhi Ridge, and (ii) Dominant species based forest classification in Shimla and Chopal Regions. In both the cases, the effectivity of neural network models to perform a finely tuned classification has been investigated.

In order to arrive at a preferred neural network for classification of satellite imageries, following approaches have been investigated.

6.1. Crisp Neural Networks

In the case of land use classification in Delhi Ridge, the neural network with number of hidden neurons varied from eight to one has been used. Performances of classification using crisp neural network for Delhi Ridge have been evaluated for the cases wherein:

- Locations of predictions/verification sets have been var-

ied;

- Size of training sets has been varied from 252 to 50;
- Ratio of prediction set size to training set size, scale-up factor has been varied from 2 to 60 in contrast to statistical classifiers wherein the reliable maximum ratio recommended is 10.

Representative confusion matrices for these cases are shown in Tables 3, 4, and 5. Table 3 is for crisp neural networks with one constant size for training data, but with varied sizes, viz., 3000, 2000, 1500, 1000 and 500 for prediction data. Both training data and prediction data correspond to geographical locations distributed across the study area. Table 4 is with the same training set but for different verification sets, whose geographical locations are different and selected from larger geographical spread. The average accuracy for Table 3 is above 86.8% while for Table 4 the average accuracy is above 77.6%. The verification pixels corresponding to Table 3 are more clustered, and have less geographical spread compared to the verification sets corresponding to Table 4.

Table 5 is for neural networks with a minimum size of 50 for training, with prediction sets same as of Table 3. The aver

Table 3. Confusion Matrix of Neural Network Classifier for 6 Hidden Neurons and 252 Training Pixels in the Delhi Ridge

Class Name	Unclassified	Dense forest	Open forest	Degraded forest	Built-up Area	Forest Blank	Accuracy (%)	Average Accuracy (%)
<i>Verification Set 3000</i>								
Dense forest	33	402	17	35	4	25	77	87.4
Open forest	19	7	562	1	6	0	94	
Degraded forest	39	19	13	560	24	9	84	
Built-up Area	27	2	0	1	944	1	96	
Forest Blank	19	8	0	0	6	217	86	
<i>Verification Set 2000</i>								
Dense forest	22	244	10	29	4	17	74	87.2
Open forest	14	6	336	1	6	0	92	
Degraded forest	28	15	5	438	13	6	86	
Built-up Area	17	1	0	0	617	1	97	
Forest Blank	13	3	0	0	5	149	87	
<i>Verification Set 1500</i>								
Dense forest	17	193	6	24	2	10	76	87.8
Open forest	11	4	279	1	5	0	93	
Degraded forest	21	13	4	296	9	5	85	
Built-up Area	10	1	0	0	461	0	97	
Forest Blank	10	3	0	0	2	113	88	
<i>Verification Set 1000</i>								
Dense forest	12	106	1	20	2	7	71	86.8
Open forest	8	2	151	0	4	0	91	
Degraded forest	20	8	2	234	6	4	85	
Built-up Area	5	1	0	0	322	0	98	
Forest Blank	6	2	0	0	1	76	89	
<i>Verification Set 500</i>								
Dense forest	6	67	0	8	1	1	80	86.8
Open forest	5	2	95	0	3	0	90	
Degraded forest	11	6	1	92	3	1	80	
Built-up Area	2	1	0	0	157	0	98	
Forest Blank	3	2	0	0	0	33	86	

Table 4. Confusion Matrix of Neural Network Classifier for 6 Hidden Neurons and 252 Training Pixels

Class Name	Unclassified	Dense forest	Open forest	Degraded forest	Built-up Area	Forest Blank	Accuracy (%)	Average Accuracy (%)
<i>Verification Set 2000</i>								
Dense forest	39	147	12	56	1	22	53	77.8
Open forest	5	0	314	0	0	0	98	
Degraded forest	56	37	83	354	24	15	62	
Built-up Area	7	0	6	3	557	32	92	
Forest Blank	20	5	1	2	8	194	84	
<i>Verification Set 1500</i>								
Dense forest	32	121	6	45	1	12	55	77.8
Open forest	4	0	199	0	0	0	98	
Degraded forest	45	34	61	289	19	3	64	
Built-up Area	5	0	6	3	480	20	93	
Forest Blank	12	5	1	1	5	91	79	
<i>Verification Set 1000</i>								
Dense forest	22	76	6	35	0	9	51	77.6
Open forest	2	0	134	0	0	0	98	
Degraded forest	31	20	43	198	13	2	64	
Built-up Area	3	0	4	2	312	16	92	
Forest Blank	7	3	0	0	2	60	83	
<i>Verification Set 500</i>								
Dense forest	10	31	5	15	0	4	47	79.4
Open forest	0	0	69	0	0	0	100	
Degraded forest	12	5	24	110	7	1	69	
Built-up Area	1	0	3	0	160	8	93	
Forest Blank	3	1	0	0	0	31	88	

Table 5. Confusion Matrix of Neural Network Classifier for 8 Hidden Neurons and 50 Training Pixels

Class Name	Unclassified	Dense forest	Open forest	Degraded forest	Built-up Area	Forest Blank	Accuracy (%)	Average Accuracy (%)
<i>Verification Set 3000</i>								
Dense forest	4	289	25	64	3	2	74	75.0
Open forest	4	12	548	56	9	0	87	
Degraded forest	73	64	34	418	24	6	67	
Built-up Area	68	1	0	5	903	13	91	
Forest Blank	60	101	0	20	4	211	56	
<i>Verification Set 2000</i>								
Dense forest	6	176	16	45	2	1	71	76.0
Open forest	4	9	364	6	8	0	94	
Degraded forest	57	40	21	330	13	5	70	
Built-up Area	51	1	0	5	584	9	89	
Forest Blank	40	50	0	13	4	140	56	
<i>Verification Set 1500</i>								
Dense forest	6	176	16	45	2	1	71	76.2
Open forest	4	9	364	6	8	0	94	
Degraded forest	57	40	21	330	13	5	70	
Built-up Area	51	1	0	5	584	9	89	
Forest Blank	40	50	0	13	4	140	56	
<i>Verification Set 1000</i>								
Dense forest	4	94	9	11	2	1	77	76.8
Open forest	1	6	139	25	4	0	79	
Degraded forest	36	18	13	182	5	2	71	
Built-up Area	26	1	0	3	303	5	89	
Forest Blank	10	18	0	7	0	75	68	
<i>Verification Set 500</i>								
Dense forest	1	52	4	9	2	0	76	75.6
Open forest	0	5	89	18	2	0	78	
Degraded forest	7	12	8	71	3	1	69	
Built-up Area	11	1	0	2	148	3	89.00	
Forest Blank	10	3	0	4	0	34	66.00	

age accuracy for all cases with 50 training pixels is above 70%.

The minimum scale-up factor and the corresponding confusion matrices are given in Table 3 (Verification Set 500) and Table 4 (Verification Set 500) wherein the accuracies are 86.8 and 78.4% respectively. The maximum scale-up factor ratio is 60 and the corresponding confusion matrix is given in Table 5 (Verification Set 3000) wherein the accuracy is 75%. Applicability of neural network with smaller size of training set for classification is thus evident.

Similarly, in the case of Shimla region, the performance of classification has been studied for cases:

- The training set size has been varied from 408 pixels to 60 pixels and the corresponding confusion matrices are given in Table 6;
- The location of training/prediction set has also been changed.

6.2. Fuzzy Neural Networks

The performance of the fuzzy neural network as an alternative neural network model has been evaluated for the classification and for the selection of best network architecture:

- the representations of land use at output layer have been varied;
- the number of hidden neurons has been varied;
- the value of gamma, viz. the relative weight of union and intersection operations in fuzzy neuronal synaptic function, has been varied.

The classification accuracies obtained using fuzzy neural networks are not good. In all cases of different verification sets with a constant 252 training set size, the accuracies have not been above 19.0%.

6.3. Constrained Connectivity or Hierarchical Neural Networks

The performance of Constrained Connectivity or Hierarchical Neural Networks has been calculated. The classification accuracies of crisp neural network model have been compared to those of the best results obtained using different supervised and unsupervised parametric and non-parametric statistical classifiers viz. Maximum likelihood, Mahalanobis distance, Minimum distance and Parallelepiped Classifier. Comparison of performance of crisp networks for different architectures and different training and verification sets with the statistical classifiers has been given in Tables 7 and 8 for Delhi Ridge. Similarly, in the case of Shimla and Chopal regions have been given in Tables 9 and 10. The corresponding results for fuzzy neural networks and hierarchical networks are presented in Tables 11 and 12.

The classification accuracy is sensitive to the number of hidden neurons and it varies from 87.4 to 20.0% with the variation of hidden neurons from 8-1 in the case of land use classification in Delhi Ridge. This observation is independent of number of training pixels, sizes of verification sets and

geographical locations of training and verification sets, and has been observed in both the cases of Delhi Ridge and Shimla & Chopal Regions. This is a significant result and is in variance to the earlier observations in the literature (Foody et al., 1995; Paola et al., 1995). The classification accuracy decreases by 6 to 15% as the size of the training set is reduced by five times in the case of land use classification in Delhi Ridge depending on the location of the choice of training/testing sites. The decrease in classification accuracy is 10 to 12% for the reduction in training size by 6.8 times in the case of dominant species in Shimla and Chopal regions. The loss in classification accuracy with the reduction in training set size is independent of neural network architecture but dependent on application and choice of location of training set.

The scale-up factor, the ratio of the size of verification/prediction set to the size of training set has no discernible effect on classification accuracy. However, it decreases by a maximum of 10% for certain neural network configurations.

The variations of classification accuracy with number of hidden neurons, training set sizes and up-scaling ratios for the two cases, viz. land cover and floral species have been presented in this study. The corresponding results for statistical classification are also presented. It is interesting to note that the mean squared error of converged and trained neural network beyond a certain limit does not improve the classification accuracy. The classification accuracy for different classes varies from 71 to 98% in the case of crisp neural network with 6 hidden neurons. Such wide variations in the classification accuracies for different classes are features in the case of statistical classifiers too. The built-up areas and open forests are classified best and confusions amongst forest classes do not exhibit a pattern.

The classification accuracy using statistical classifiers varies from 64 to 74% and 69 to 77% in the cases of Delhi Ridge and Shimla region respectively (Tables 19 and 22). These statistical classification accuracies are for training set sizes of 252 and 408 pixels respectively. The corresponding classification accuracies obtained using the best neural network architectures range from 87.8 to 77.6% and 80.1 to 75.6% for Delhi Ridge and Shimla region with 252 and 408 training set sizes respectively. The corresponding classification accuracies with five times reduced training set size of 50 and 60 for Delhi Ridge and Shimla Region, range from 76.8 to 66.0% and 73.6 to 63.8% respectively. The accuracies obtained with the reduced training sizes are better than the accuracies obtained using statistical classifiers with four times enlarged training sizes.

The classification accuracy obtained using fuzzy neural network model is very poor. As indicated in the earlier published work, the utility of fuzzy neural networks is more in identifying systems endowed with complex logics and small sizes. The classification accuracy using constrained connectivity network is comparable to that of statistical classifiers. The classification accuracy varies from 64.1 to 67.2% with a specific advantage of faster convergence due to reduced number of network connections.

Table 6. Confusion Matrix of Neural Network Classifier for 8 Hidden Neurons and 408 Training Pixels in the Shimla and Chopal Regions

Class Name	Unclassified	A	B	C	D	E	F	Accuracy (%)	Average Accuracy (%)
<i>Verification Set 2000</i>									
A	19	281	4	20	28	14	8	75	80.16
B	14	1	318	1	12	4	0	90	
C	40	3	0	190	2	17	4	74	
D	21	6	3	4	413	13	2	89	
E	48	1	3	4	8	162	0	71	
F	5	7	3	19	3	21	274	82	
<i>Verification Set 1500</i>									
A	12	245	4	15	26	11	7	76	79.16
B	9	1	243	1	10	2	0	91	
C	27	2	0	105	0	12	4	70	
D	17	6	3	3	338	10	1	89	
E	31	1	3	3	8	91	0	66	
F	2	4	3	13	3	17	207	83	
<i>Verification Set 1000</i>									
A	6	168	1	8	17	8	5	78	79.16
B	7	1	154	1	10	1	0	88	
C	17	1	0	77	0	10	1	72	
D	13	5	2	2	215	7	1	87	
E	22	1	0	3	5	64	0	67	
F	2	2	1	9	3	11	139	83	
<i>Verification Set 500</i>									
A	2	86	1	4	5	5	3	81	78.66
B	5	0	77	0	4	1	0	88	
C	8	1	0	29	0	4	1	67	
D	8	3	2	2	107	36	0	85	
E	10	0	0	2	3	4	0	70	
F	0	1	1	7	3		72	81	

Table 7. Performance of the Neural Network Classifier for Different Neural Network Architectures, Training Sizes, and Geographical Land Locations (Clustered, Distributed, and Varied Sizes) in the Delhi Ridge

Hidden Neurons	Number of Pixels	Verification Accuracy (percentage of agreement with the ground truth)				
		Set I (3000 pixels)	Set II (2000 pixels)	Set III (1500 pixels)	Set IV (1000 pixels)	Set V (500 pixels)
8	252	74.0	74.0	74.2	72.2	74.0
6 (best)	252	87.4	87.2	87.8	87.8	86.8
3	252	67.0	68.0	68.0	66.4	68.0
1	252	21.2	20.2	21.0	20.0	20.4
8	95	76.1	75.0	75.0	74.3	72.0
6 (best)	95	78.0	77.2	77.0	76.0	74.0
3	95	71.0	71.0	71.3	71.6	68.0
1	95	21.4	20.0	21.0	20.0	20.3
8 (best)	50	75.0	76.0	76.2	76.8	75.6
6	50	69.0	68.0	68.0	68.2	66.2
3	50	70.0	70.0	70.0	69.0	64.4
1	50	22.6	21.6	22.4	20.6	21.6

Table 8. Performance of Different Statistical Classifiers, for Different Geographical Land Locations (Clustered, Distributed, and Varied Sizes) in the Delhi Ridge

Name of Classifier	Number of Pixels	Verification Accuracy (percentage of agreement with the ground truth)			
		Set I (2000 pixels)	Set II (1500 pixels)	Set III (1000 pixels)	Set IV (500 pixels)
Maximum Likelihood	252	71.0	71.0	70.0	72.0
Mahalanobis Distance	252	74.0	70.0	69.0	67.0
Minimum Distance	252	65.0	64.0	64.0	64.0

Table 9. Performance of the Neural Network Classifier for Different Neural Network Architectures, Training Sizes and Geographical Land Locations (Clustered, Distributed, and Varied Sizes) in the Shimla and Chopal Reigions

Hidden Neurons	Number of Pixels	Verification Accuracy (percentage of agreement with the ground truth)			
		Set II (2000 pixels)	Set III (1500 pixels)	Set IV (1000 pixels)	Set V (500 pixels)
8	408	80.1	79.1	79.1	78.6
6 (best)	408	77.0	76.3	76.0	75.1
3	408	70.8	69.5	69.8	70.3
1	408	17.0	17.0	17.0	16.0
8	210	70.2	70.1	70.0	71.4
6 (best)	210	70.0	70.0	69.0	71.0
3	210	67.1	67.6	65.0	65.0
1	210	17.0	16.8	17.0	16.0
8 (best)	60	69.1	70.5	70.6	73.6
6	60	66.6	64.6	64.6	62.8
3	60	60.7	60.2	58.0	59.0
1	60	17.0	17.0	16.8	17.0

Table 10. Performance of Different Statistical Classifiers for Different Geographical Land Locations (Clustered, Distributed, and Varied Sizes) in the Shimla and Chopal Regions

Name of Classifier	Number of Pixels	Verification Accuracy (percentage of agreement with the ground truth)			
		Set I (2000 pixels)	Set II (1500 pixels)	Set III (1000 pixels)	Set IV (500 pixels)
Maximum Likelihood	408	77.0	77.0	77.0	76.0
Mahalanobis Distance	408	71.0	69.0	70.0	72.0
Minimum Distance	408	74.0	74.0	74.0	75.0

Table 11. Performance of the Fuzzy Neural Network Classifier for Different Training Sizes, γ -Values and Geographical Land Locations (Clustered, Distributed, and Varied Sizes) in the Delhi Ridge

Number of Pixels	γ - Value	Verification Accuracy (percentage of agreement with the ground truth)				
		Set I (3000 pixels)	Set II (2000 pixels)	Set III (1500 pixels)	Set IV (1000 pixels)	Set V (500 pixels)
252	0.3	18.4	17.0	18.2	18.4	19.0
252	0.6	17.2	16.8	18.0	17.9	18.1
252	0.8	18.1	17.0	17.9	17.6	17.8
95	0.3	16.2	16.3	17.2	16.2	15.2
95	0.6	17.0	15.2	16.3	16.7	16.2
95	0.8	17.3	16.2	16.0	17.0	17.1
50	0.3	15.2	15.1	15.6	17.2	15.7
50	0.6	15.3	14.2	13.0	15.1	14.2
50	0.8	15.1	14.4	14.2	17.3	13.3

Table 12. Performance of the Hierarchical Neural Network Classifier for Different Training sizes and Geographical Land Locations (Clustered, Distributed, and Varied Sizes) in the Delhi Ridge

Number of Pixels	Verification Accuracy (percentage of agreement with the ground truth)				
	Set I (3000 pixels)	Set II (2000 pixels)	Set III (1500 pixels)	Set IV (1000 pixels)	Set V (500 pixels)
252	67.2	66.6	67.2	65.6	66.6
252	66.3	66.0	66.2	64.2	66.7
252	65.0	64.2	64.3	64.1	66.3
95	67.1	66.2	67.5	64.2	65.2
95	67.0	66.3	66.8	65.2	66.3
95	66.0	65.0	64.2	65.1	66.4
50	65.1	66.3	64.0	65.0	66.0
50	64.2	66.7	64.7	62.8	61.0
50	62.3	61.5	61.3	62.7	61.4

7. Conclusions

The study conclusively establishes the effectivity of crisp neural network based classifiers for satellite imagery analysis. The specific advantage with this set of classifiers is higher accuracy and reduced size of ground truth data in comparison to statistical classifiers. The above conclusion is not dependent on application, geographical locations of training and verification data sets, and scale-up factor. The reduction in the requirement of ground truth data is a significant advantage in practical applications. In addition, the crisp neural network is effective in fine tuning the classification as demonstrated by the forest classification based on dominant species in Shimla and Chopal regions.

The hierarchical neural networks, in despite of their inability in providing higher accuracies, are effective with reduced sizes of training set. As their convergence is faster, hierarchical neural network could be used as a placing them as an alternative to neural network classifier in practical and routine applications not requiring very high accuracies.

The fuzzy neural network classifier has no specific advantage. It has poor accuracy of prediction and is sensitive to parameters governing fuzzy neuronal aggregation and exhibits slow convergence.

In summary, the effectivity of crisp neural network in the analysis of remotely sensed data is demonstrated in this study with the help of the two distinct cases viz. Delhi ridge and Shimla and Chopal Regions in India.

References

- Aria, E.H., Amini, J. and Saradjian, M.R. (2003). Back propagation neural network for classification of IRS-1D satellite images, *Joint Workshop of ISPRS WG I/2, I/5 and IC WG II/IV and EARSeL Special Interest Group 3D Remote Sensing on High Resolution Mapping from Space 2003*.
- 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 Tran. Geosci. Remote Sens.*, 28, 540-551.
- Bernard, A.C., Wilkinson, G.G. and Kanellopoulos, I. (1997). Training strategies for neural network soft classification of remotely sensed imagery. *Int. J. Remote Sens.*, 18(8), 1851-1856.
- Bezdek J. C. (1991) Pattern Recognition with Fuzzy Objective Function Algorithms, New York Plenum Press.
- Chen, M.W., Zalza, A.M.S.S. and Sharkey, N.E. (1997). Towards a comparative study of neural networks in inverse model learning and compensation applied to dynamic robot control, *IEE Conference Publication No. 440, IEE Stevenage, Engl.*, pp. 146-151.
- Civco, D.L. (1993). Artificial neural networks for land-cover classification and mapping. *Int. J. Geogr. Inf. Syst.*, 7, 173-186.
- Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.*, 37, 35-46.
- Dreyer, P. (1993). Classification of land cover using optimized neural nets on SPOT data. *Photogramm. Eng. Remote Sens.*, 59, 617-621.
- Foody, G.M. (1995). Land cover classification by an artificial neural network with ancillary information. *Int. J. Geogr. Inf. Syst.*, 9(5), 527-542.
- Foody, G.M. and Arora, M.K. (1997). Evaluation of some factors affecting the accuracy of classification by an artificial neural network. *Int. J. Remote Sens.*, 18(4), 799-810.
- Geaga, J.V., Le, T.C. and Sailer, C.T. (1992). Neural net spectral pattern recognition. *SPIE, Neural Stochastic methods Image Signal Proc.*, 1766, 609-618.
- Heermann, P.D. and Khazenie, N. (1992). Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Tran. Geosci. Remote Sens.*, 30, 81-88.
- Hepner, G.F., Logan, T., Ritter, N. and Bryant, N. (1990). Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogramm. Eng. Remote Sens.*, 56, 469-473.
- Howald, K.J. (1989). Neural network image classification, in *Proc. American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping Fall Convention*, American Society for Photogrammetry and Remote Sensing, Cleveland, OH, pp. 207-215.
- Hush, D.R. and Horne, B.G. (1993). Progress in supervised neural networks. *IEEE Signal Proc. Mag.*, 10, 8-39.
- Key, J., Maslanik, J.A. and Schweiger, A.J. (1990). Neural Network Vs. Maximum likelihood classifications of spectral and textural

- features in visible, thermal, and passive microwave data, in *Proc. of the International Geoscience and Remote Sensing Symposium, IGRASS'90*, College Park, MD, USA, pp. 1277-1280.
- Lewis, H.G. and Brown, M. (2000). A generalized confusion matrix for assessing area estimates from remote sensing imagery. *Int. J. Remote Sens.*, 13, 917-924.
- Lippmann, R.P. (1987). An introduction to computing with neural nets. *IEEE Acoust. Speech Signal Proc. Mag.*, 4, 4-22.
- Murai, H. and Omatu, S. (1997). Remote sensing image analysis using a neural network and knowledge based processing. *Int. J. Remote Sens.*, 18(4), 811-828.
- Kloer B. R. (1994). Hybrid Parametric / Non-Parametric Image Classification, Proceedings of *ASPRS/ACSM*, pp. 307-316.
- Krishnapuram R. and Lee J. (1992) Fuzzy connective based Hierarchical Aggregation Networks for Decision Making, *Fuzzy Sets and Systems*, Vol. 46, pp. 1-17.
- Paek, K.N., Young, S., Hyo, S. and Kim, K.E. (1997). Study on the characteristics of the supervised classification of remotely sensed data using artificial neural networks, in *Proc. of the International Geoscience and Remote Sensing Symposium, IGARSS'97*, part 2 (of 4), Piscataway, NJ, USA, pp. 528-530.
- Pao, Y.H. (1989). *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, MA, USA.
- Paola, J.D. and Schowengerdt, R.A. (1995). A review and analysis of backpropagation neural networks for classification of remotely sensed multi spectral imagery. *Int. J. Remote Sens.*, 16(16), 3033-3058.
- Paola, J.D. and Schowengerdt, R.A. (1997). The effect of neural network structure on a multispectral land-use/land-cover classification. *Photogramm. Eng. Remote Sens.*, 63(5), 535-544.
- Rumelhart, D.E., Hinton, G. E., and Williams, R.J. (1986). Learning internal representations by error propagation. In Rumelhart, D.E. and McClelland, J. L., editors, *Parallel Distributed Processing*, MIT Press, volume 1, pp. 318-362.
- Story, M.M. and Congalton, R.G. (1986). Accuracy assessment: A user's perspective. *Photogramm. Eng. Remote Sens.*, 52, 397-399.
- Tzeng, Y.C. and Chen, K.S. (1997). Wide reconstruction from ERS-1 scatterometer data using neural network, in *Proc. of the International Geoscience and Remote Sensing Symposium IGARSS'97*, part 3 (of 4), Piscataway, NJ, USA, pp. 1208-1210.
- Werbos, P.J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D thesis, Harvard University.
- Yamakawa, T. (1989). A Fuzzy Neuron and its Application to pattern Recognition. In Proceedings of Third IFSA Congress, Seattle, WA, pp. 30-38.
- Yoshida, T. and Omatu, S. (1994). Neural Network approach to land cover mapping. *IEEE Tran. Geosci. Remote Sens.*, 32(5), 1103-1109.