ELSEVIER

Computers and Electronics in Agriculture
24 (1999) 131–151

Computers and electronics in agriculture

www.elsevier.com/locate/compag

Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables

Jock A. Blackard [1], Denis J. Dean*

Remote Sensing and GIS Program, Department of Forest Sciences, 113 Forestry Building, Colorado State University, Fort Collins, CO 80523, USA

Received 29 October 1998; received in revised form 12 February 1999; accepted 3 August 1999

Abstract

This study compared two alternative techniques for predicting forest cover types from cartographic variables. The study evaluated four wilderness areas in the Roosevelt National Forest, located in the Front Range of northern Colorado. Cover type data came from US Forest Service inventory information, while the cartographic variables used to predict cover type consisted of elevation, aspect, and other information derived from standard digital spatial data processed in a geographic information system (GIS). The results of the comparison indicated that a feedforward artificial neural network model more accurately predicted forest cover type than did a traditional statistical model based on Gaussian discriminant analysis.

Keywords: Artificial intelligence; Discriminant analysis; Forest cover types; Geographic information systems (GIS); Neural networks; Spatial modeling

* Corresponding author. Tel.: +1-970-4912378; fax: +1-970-4916754.
E-mail address: denis@cnr.colostate.edu (D.J. Dean)
[1] Present address: US Forest Service, Ecosystem Management Analysis Center, 3825 East Mulberry Street, Fort Collins, CO 80524, USA.

0168-1699/99/$ - see front matter
PII: S0168-1699(99)00046-0

1. Introduction

Accurate natural resource inventory information is vital to any private, state, or federal land management agency. Forest cover type is one of the most basic characteristics recorded in such inventories. Generally, cover type data is either directly recorded by field personnel or estimated from remotely sensed data. Both of these techniques may be prohibitively time consuming and/or costly in some situations. Furthermore, an agency may find it useful to have inventory information for adjoining lands that are not directly under its control, where it is often economically or legally impossible to collect inventory data. Predictive models provide an alternative method for obtaining such data.

Compared to statistical models, artificial neural networks (ANNs) represent a relatively new approach to developing predictive models. Artificial neural networks are "computing devices that use design principles similar to the design of the information-processing system of the human brain" (Bharath and Drosen, 1993, p. xvii). Several recent textbooks describe the mechanics of ANNs (Hertz et al., 1991; Haykin, 1994; Masters, 1994; Bishop, 1995; Ripley, 1996). Recent publications involving artificial neural networks being applied to natural resources topics include: modeling complex biophysical interactions for resource planning applications (Gimblett and Ball, 1995); generating terrain textures from a digital elevation model and remotely sensed data (Alvarez, 1995); modeling individual tree survival probabilities (Guan and Gertner, 1995); and Harvey and Dean (1996), who used geographic information systems (GIS) in developing computer-aided visualization of proposed road networks. Recent comparisons in which ANNs performed favorably against conventional statistical approaches include Reibnegger et al. (1991), Patuwo et al. (1993), Yoon et al. (1993), Marzban and Stumpf (1996), Paruelo and Tomasel (1996), Pattie and Haas (1996), and Marzban et al. (1997).

However, artificial neural networks do not always outperform traditional predictive models. For example, Jan (1997) found a traditional maximum-likelihood classifier outperformed artificial neural network models when classifying remotely sensed crop data. Also, using their best artificial neural network model, Vega-Garcia et al. (1996) obtained only a slight improvement in predicting human-caused wildfire occurrences as compared to their best logit model.

This study examined the ability of an ANN model to predict forest cover type classes in forested areas that have experienced relatively little direct human management activities in the recent past. The predictions produced by the ANN model were evaluated based on how well they corresponded with observed cover types (absolute accuracy), and on their relative accuracy compared to predictions made by a more conventional model based on discriminant analysis (DA).

2. Data description

The study area for this project consisted of the Rawah (29 628 hectares or 73 213 acres), Comanche Peak (27 389 hectares or 67 680 acres), Neota (3904 hectares or

9647 acres), and Cache la Poudre (3817 hectares or 9433 acres) wilderness areas of the Roosevelt National Forest in northern Colorado. As shown in Fig. 1, these areas are located ~70 miles northwest of Denver, Colorado. These wilderness areas were selected because they contained forested lands that have experienced relatively little direct human management disturbances. As a consequence, the current composition of forest cover types within these areas are primarily a result of natural ecological processes rather than the product of active forest management.

In this study, the ANN and DA models utilized a supervised classification procedure to classify each observation into one of seven mutually exclusive forest cover type classes. The seven forest cover type classes used in this study were lodgepole pine (*Pinus contorta*), spruce/fir (*Picea engelmannii* and *Abies lasiocarpa*), ponderosa pine (*Pinus ponderosa*), Douglas-fir (*Pseudotsuga menziesii*), aspen (*Populus tremuloides*), cottonwood/willow (*Populus angustifolia*, *Populus deltoides*, *Salix bebbiana*, *Salix amygdaloides*), and krummholz. The krummholz forest cover type class is composed primarily of Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), and Rocky Mountain bristlecone pine (*Pinus aristata*) in these wilderness areas. These seven cover type classes were chosen for this research since they represent the primary dominant tree species currently found in the four wilderness areas. A few other forest cover types exist in small patches within the study area, however, these relatively minor cover types have been ignored in this analysis. Cover type maps for these areas were created by the US Forest Service, and are based on homogeneous stands varying in size from 2 to 80 hectares (from 5 to 200 acres) that were derived from large-scale aerial photography.

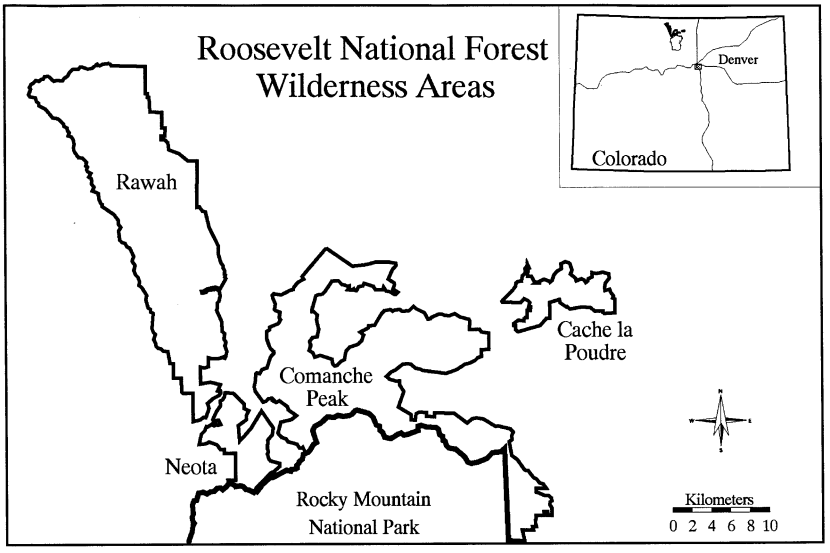

Fig. 1. Study area location map.

For this study, digital spatial data obtained from the US Geological Survey (USGS) and the US Forest Service (USFS) were used to derive independent variables for the predictive models. The following 12 variables (with their units of measure) were used:

1. Elevation (m),
2. Aspect (azimuth from true north),
3. Slope (°),
4. Horizontal distance to nearest surface water feature (m),
5. Vertical distance to nearest surface water feature (m),
6. Horizontal distance to nearest roadway (m),
7. A relative measure of incident sunlight at 09:00 h on the summer solstice (index),
8. A relative measure of incident sunlight at noon on the summer solstice (index),
9. A relative measure of incident sunlight at 15:00 h on the summer solstice (index),
10. Horizontal distance to nearest historic wildfire ignition point (m),
11. Wilderness area designation (four binary values, one for each wilderness area), and
12. Soil type designation (40 binary values, one for each soil type).

Elevation was obtained directly from USGS digital elevation model (DEM) data, based on 30×30-m raster cells (1:24 000 scale). This DEM data was used to standardize all remaining data, so each observation in this study represents a unique 30×30-m raster cell that corresponds to USGS DEM data. Aspect, slope, and the three relative measures of incident sunlight were developed from this DEM using standard GIS-based surface analysis and hillshading procedures (Environmental Systems Research Institute, 1991). Horizontal distance to the nearest surface water feature and horizontal distance to the nearest roadway were obtained by applying Euclidean distance analyses to USGS hydrologic and transportation data. Horizontal distance to the nearest historic wildfire ignition point was determined by using Euclidean distance analyses with a USFS wildfire ignition point coverage, which identified ignition points of forest wildfires occurring over the past 20 years. Vertical distance above or below the nearest surface water feature were calculated using a combination of the DEM, hydrologic data, and a simple custom-built spatial analysis program.

Both soil type information and wilderness area designation were obtained from the USFS. These qualitative variables were treated as multiple binary values. This resulted in a series of variables for each raster cell, where a value of '0' would represent an 'absence' and a value of '1' would represent a 'presence' of a specific wilderness area or soil type (Huberty, 1994, p. 151; Bishop, 1995, p. 300). A total of four wilderness areas and 40 soil type classes were used in this study, producing four wilderness area designator variables, forty soil type designator variables, and ten continuous variables for a total of 54 possible independent variables available for each model.

The data used in this study were produced primarily by using standard GIS procedures with ARC/INFO software, version 7.2.1, and the GRID module

(Environmental Systems Research Institute, 1991). Several software packages were used within a UNIX Sun Sparc workstation environment for data analysis and model development. The feedforward artificial neural network models were developed using PROPAGATOR software, version 1.0 (ARD Corporation, 1993). In addition, the discriminant analysis models were created using the statistical procedures available from SAS software, version 6.11 (SAS Institute, Inc., 1989). These predictive models were then implemented using custom-built GIS macros and embedded 'C' programs.

3. Data set selection

For this study, three mutually exclusive and distinct data sets were created to train, validate, and test the predictive models. A training data set was used to develop classifiers for both the artificial neural network and the discriminant analysis predictive models. The validation data set was used in the development of the feedforward artificial neural network models, to identify the point in the training process where the network begins to 'overfit' (memorize) the training data set, and consequently loses its ability to generalize against unforeseen data inputs. The validation data set, however, is not required in the development of discriminant analysis models, and consequently was not utilized for that purpose in this study. Finally, for both the artificial neural network and the discriminant analysis models, the test data set was used to determine how well each classifier would perform against a data set that was not used in the creation of the predictive model.

These three data sets were selected from a total of 581 012 observations (30×30-m raster cells) encompassing $\sim 52\,291$ hectares (129 213 acres) of the study area. Each observation contained information for every independent and response variable used in the study (e.g. no observations contain 'missing data').

All observations were initially pooled into one very large data set for the whole study area. Then two subsets were extracted from the full data set. The first extracted set contained 1620 randomly selected observations for each of the seven cover types (11 340 total observations) and became the training data set. This number (1620 observations per cover type) was chosen because it represented $\sim 60\%$ of the number of observations available in the least numerous cover type (cottonwood/willow), and utilizing $\sim 60\%$ of the available data for training is an effective use of data (Anderson, Department of Computer Science, Colorado State University, 1996, personal communication).

The second data set extracted from the remaining data contained 540 randomly selected observations for each cover type (3780 total observations) and became the validation data set. This number (540) of observations represents $\sim 20\%$ of the observations available for the cottonwood/willow cover type. Finally, all remaining observations (565 892) were placed in the testing data set. Table 1 lists the number of observations within each cover type for the three data sets. As shown in Table 1, the training and validation data sets are represented by an equal number of randomly selected observations per cover type, while the test data set reflects the more realistic proportions of cover types found in the study area.

Table 1
Number of observations within each forest cover type class for each data set

Forest cover type class	Training data set observations	Validation data set observations	Test data set observations	Total observations per cover type
Spruce/fir	1620	540	209 680	211 840
Lodgepole pine	1620	540	281 141	283 301
Ponderosa pine	1620	540	33 594	35 754
Cottonwood/ willow	1620	540	587	2747
Aspen	1620	540	7333	9493
Douglas-fir	1620	540	15 207	17 367
Krummholz	1620	540	18 350	20 510
Total observations per data set	11 340	3780	565 892	581 012

All variables in the three data sets used by the artificial neural network model were linearly scaled to lie in the range between zero and one. This scaling took place across all three data sets combined, not individually within each data set. Scaling is highly recommended in the development of artificial neural networks where the ranges of values among independent variables are not similar (Marzban and Stumpf, 1996). In addition, the response (dependent) variables in the three artificial neural network data sets were coded as a series of seven binary variables, much like the soil type and wilderness area designator variables described previously. This was done so that the response variables would conform to the architecture of the artificial neural network (seven output nodes). In contrast, the response variable in the discriminant analysis data set was coded as a single variable that assumed integer values from one to seven. Other than this, the data sets used for discriminant and artificial neural network analyses were identical.

4. Artificial neural network specifications

Artificial neural network models require several architectural and training parameters to be selected prior to analysis. The optimal number of hidden layers and the number of nodes per hidden layer are generally not known a priori for a specific data set, and must be empirically determined through an examination of different parameter settings (Haykin, 1994; Marzban and Stumpf, 1996). In this study one hidden layer was used, which past studies have found to be sufficient in most situations (Wong et al., 1995; Fowler and Clarke, 1996).

Determining the optimal number of nodes to place in this single hidden layer is a difficult process. Although several 'rules of thumb' exist concerning the optimal number of hidden nodes for a network, no method is universally appropriate since the number of nodes depends on the complexity of the problem to be solved (Fowler and Clarke, 1996). By systematically experimenting with the number of

hidden nodes in a network, the best fit may be found without making any a priori assumptions (Marzban and Stumpf, 1996).

In addition to parameters defining the network's architecture, training parameters are also required to initialize the learning algorithm used by the network. Backpropagation was the learning algorithm chosen for this study, for no other reason than it is the best known and most common algorithm in use today (Hertz et al., 1991; Gimblett and Ball, 1995; Guan and Gertner, 1995; Markham and Ragsdale, 1995; Wong et al., 1995). Backpropagation requires two initialization parameters, termed the learning rate and the momentum rate. Once again, it is not possible to know a priori optimal values for these parameters for a specific data set, so another trial-and-error process is needed to determine acceptable values.

Once the network architecture and training parameters are selected, an artificial neural network is trained iteratively. Each iteration represents one complete pass through a training data set (an epoch). At the conclusion of each iteration, a measure of discrepancy between observed and predicted values of the dependent variable is calculated. This discrepancy is often expressed as a mean square error (MSE), which for this study was the error function:

$$E(w) = \frac{1}{2N} \sum_{n=1}^{N} \sum_{i=1}^{k} (d_i(n) - y_i(n))^2 \quad (1)$$

where $E(w)$ is the mean square error term, w are the synaptic weights to be estimated, N is the number of observation (input) vectors presented to the network, n is a single observation vector, k is the number of output nodes, i is a single output node, $d_i(n)$ is the observed response and $y_i(n)$ is the predicted response for observation n and output node i (Rumelhart et al., 1986; Hertz et al., 1991; Jan, 1997; Marzban et al., 1997). The N observation vectors constitute a training data set, which is used specifically to 'teach' the network to recognize the relationships between the independent and dependent variables (e.g. to develop a classifier). This classifier will consequently be used to predict class membership for other vectors of input variables not included in the training data set. Theoretically, the backpropagation algorithm ultimately finds a set of weights w that minimizes $E(w)$.

All artificial neural network models in this study had fully connected input, hidden, and output layers (i.e. each node in layer m was connected to all nodes in layer $m+1$). The generalized delta rule with gradient descent (commonly used with the backpropagation learning algorithm) was utilized in each network's learning process. The activation function for each network's input layer was linear [$f(x) = x$], while hidden and output layers utilized logistic activation functions [$f(x) = 1/(1 + \exp(-x))$]. Initial synaptic weights were randomly selected between negative one and positive one, based on a random seed and no input noise. All input variables were linearly scaled to lie in the range between zero and one.

Training patterns were presented to the network in a random order, with an update of the validation data set MSE at an interval of every ten epochs through the training data set. Training was halted after either (1) a minimum of 1000 training epochs had been completed, (2) a validation MSE of 0.05 was reached, or (3) it was subjectively determined that the validation MSE would not significantly

decrease with further training epochs. The stopping criteria listed above are based on previous experience in the development of artificial neural network models with a similar data set (Blackard and Dean, 1996), training results obtained from several preliminary networks utilizing the same data set employed in this current study, and available computer resources.

Finally, after the network parameters were chosen via experimentation, those parameters that produced the artificial neural network judged to be 'best' were used to produce an additional 30 networks, each with the same parameters but with a different set of initial random synaptic weights. These additional networks were developed since artificial neural network methods are, to a degree, based on random processes (e.g. each ANN model is developed based in part on a set of randomly chosen initial synaptic weights). In addition, the method of gradient descent commonly used in artificial neural network strategy does not always provide for the global minimum in error space. Thus, an indication of the nature of the response surface may be obtained from analyzing the results of these 30 networks.

5. Discriminant analysis considerations

Extensive discussions of discriminant analysis may be found in Johnson and Wichern (1992), McLachlan (1992)and Huberty (1994). Discriminant analysis is based upon two main assumptions. The first is that the distributions of all independent variables are normal (Gaussian), which encourages the use of continuous rather than discrete data in the predictive model. The second assumption applies only for linear discriminant analysis, in which the covariance matrices for the different groups of observations are assumed to be equal (homoscedasticity) (Marzban et al., 1997). The second assumption is very restrictive and in practice rarely applies in full (Curram and Mingers, 1994). As McLachlan (1992) points out on p. 132, "…in practice, it is perhaps unlikely that homoscedasticity will hold exactly".

In practice, some amount of violation of these assumptions is common and appears to have minimal impact on results. McLachlan (1992), on p. 132, supports the use of the linear discriminant analysis model in situations where its assumptions are violated: "…its good performance for discrete or mixed data in many situations explains the versatility and consequent popularity of linear discriminant analysis…". McLachlan (1992), on p. 16, also states: "In practical situations, some variables in the feature vector X may be discrete. Often treating the discrete variables, in particular binary variables, as if they were normal in the formulation of the discriminant rule is satisfactory".

For linear discriminant analysis problems having both equal costs and equal prior probabilities between groups (or classes), each observation or unit will be classified into that group which it is nearest in terms of some distance measure applied within the space defined by the independent variables (Johnson and Wichern, 1992, p. 535). Huberty (1994), on p. 55, describes a frequently used

distance measure termed the Mahalanobis index, which determines the squared distance between an observation vector unit u and the centroid for group g as:

$$D_{ug}^2 = (X_u - \overline{X_g})' S_g^{-1} (X_u - \overline{X_g}) \quad (2)$$

where X_u is the $p \times 1$ vector for observation u, $\overline{X_g}$ is the $p \times 1$ vector of means for group g, and S_g is the $p \times p$ covariance matrix for group g. This approach was adapted in this study.

Both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) were investigated in this study. However, the QDA model became unstable when qualitative or discrete variables were considered. The LDA model did not exhibit this type of behavior, and therefore was used more extensively than the QDA model.

6. Experimental design

One hidden layer was used in all of the artificial neural networks developed in this study. The number of nodes in this single hidden layer was systematically changed across 14 possible values while holding constant the learning rate (LR) and momentum rate (MR) training parameters. This procedure identified the number of hidden nodes which produced the minimum error (MSE) of the validation data set under the original LR and MR values. The number of hidden nodes was then held constant and the LR and MR parameters were sequentially changed to find the best combination of network architecture and training parameter values for this data set. While a parallel approach would have been preferable to this sequential approach for determining ANN parameter values (to compensate for interdependent effects), a parallel approach was not practical due to the length of time required for each computer run and the number of runs involved (one run is necessary for each unique combination of network parameter values). Therefore, a sequential approach was selected as a reasonable compromise between completeness and practicality. Table 2 lists the various architectural and training parameter values investigated in this study.

Once the network parameters for the predictive model containing all 54 independent variables were selected, a number of other models with fewer input variables were investigated. The same set of reduced models were investigated via discriminant analysis. This was done to determine if fewer input variables would produce models with similar predictive abilities, thereby identifying variables that did not contribute to the overall predictive capability of the system.

One variable examined in this reduction process was the 'horizontal distance to the nearest wildfire ignition point' measure, since it only provided an ignition point and not a delineation of overall wildfire boundary nor an indication of fire intensity. Another variable investigated was the 'wilderness area designation' indicator, since it was a qualitative measure which may not reflect any true differences between the wilderness areas. Also, the 'soil type designation' variables were either generalized to produce 11 soil types rather than the original 40 types, or excluded

Table 2
Artificial neural network architectural and training parameter values

ANN model with 54 variables														
Step 1: Select the optimal number of hidden nodes parameter from 14 possible values while holding constant LR = 0.05 and MR = 0.9														
Number of hidden nodes	6	12	18	24	30	60	90	120	150	180	210	240	270	300
Step 2: Hold the optimal number of hidden nodes parameter value (selected from step 1) constant, and determine the optimal learning rate (LR) and momentum rate (MR) parameter values by systematically altering between their 42 possible combined values														
Learning rates	0.05	0.10	0.15	0.20	0.25	0.30	–							
Momentum rates	0.3	0.4	0.5	0.6	0.7	0.8	0.9							
ANN model with 53 or fewer variables														
Step 1: Select the optimal number of hidden nodes parameter from 14 possible values while holding constant LR = 0.05 and MR = 0.5														
Number of hidden nodes	6	12	18	24	30	60	90	120	150	180	210	240	270	300
Step 2: Hold the optimal number of hidden nodes parameter value (selected from step 1) constant, and determine the optimal learning rate (LR) and momentum rate (MR) parameter values by systematically altering between their 30 possible combined values														
Learning rates	0.05	0.10	0.15	0.20	0.25	0.30	–							
Momentum rates	0.3	0.4	0.5	0.6	0.7	–	–							

Table 3
Number of input variable subsets examined in this study

Number of independent variables	Description of variables
54	Ten quantitative variables + four wilderness areas + 40 soil types
53	Same as '54' but excluding distance-to-wildfire-ignition-points
21	Ten quantitative variables + 11 generalized soil types
20	Same as '21' but excluding distance-to-wildfire-ignition-points
10	Ten quantitative variables only
9	Same as '10' but excluding distance-to-wildfire-ignition-points

from the model altogether. This was accomplished by objectively grouping those 40 soil types into more generalized categories based solely on their climatic and geologic associations. Table 3 lists the six sets of independent variables examined in this study.

7. Results

The MSE values across all 14 different numbers of hidden nodes from the 54 variable ANN models are shown in Fig. 2. Each of these networks held the LR and

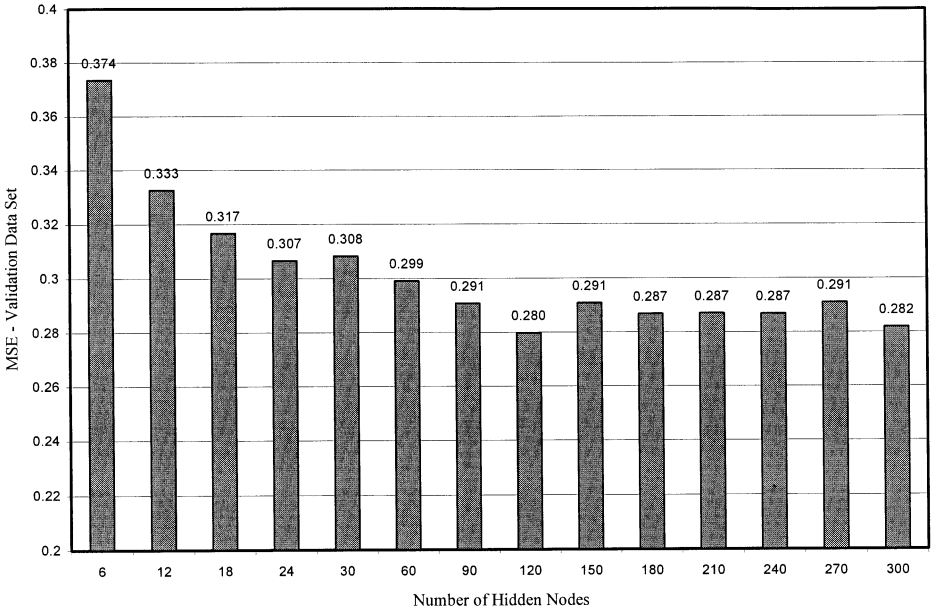

Fig. 2. MSE for the different numbers of hidden nodes for the 54 variable ANN models (optimal value for this model is 120 hidden nodes).

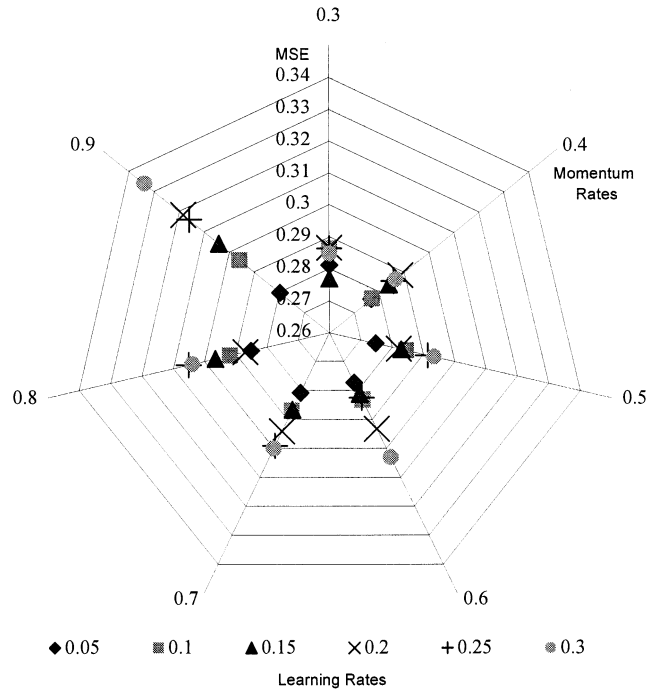

Fig. 3. MSE of the validation data set for the 54 variable ANN model with 120 hidden nodes (optimal values for this model are a learning rate of 0.05 and a momentum rate of 0.5). MSE values are shown by momentum rates (spokes) and learning rates (symbols), and are equal along concentric lines of the plot.

MR values constant at 0.05 and 0.9, respectively. This figure shows that roughly 120 hidden nodes were necessary to minimize the MSE in these networks.

Once the 'best' number of hidden nodes was identified, the learning rates and momentum rates were sequentially changed to determine optimal values for these parameters. Fig. 3 displays a radar plot of the resulting MSE values for 42 different candidate networks, all with 120 hidden nodes. As shown, a learning rate of 0.05 generally produced the lowest MSE for each possible momentum rate. In addition, momentum rate values of 0.3, 0.4, and 0.5 appeared to have the lowest average MSE among the various learning rates. The network model with the lowest MSE was produced with an LR of 0.05 and an MR of 0.5.

From these results, a ANN design of 54 input nodes, 120 hidden nodes, and seven output nodes (symbolized as 54-120-7) with an LR = 0.05 and MR = 0.5 was chosen as 'optimal' for this data set. Optimal parameters for each of the reduced artificial neural network models were found in a similar fashion, and are shown in Table 4.

Classification accuracies produced by each model as calculated from the test data set were also examined. As shown in Table 5, the ANN predictions of forest cover type produced an overall classification accuracy of 70.58%.

Table 4
Artificial neural network parameter values in this study[a]

Number of independent variables	Network architecture	Learning rate	Momentum rate	Validation data set MSE
54	54-120-7	0.05	0.5	0.2747
53	53-120-7	0.05	0.5	0.2908
21	21-60-7	0.05	0.5	0.3061
20	20-60-7	0.05	0.5	0.3363
10	10-90-7	0.10	0.5	0.3312
9	9-60-7	0.05	0.6	0.3699

[a] Network architecture values represent the number of input nodes, number of hidden nodes, and the number of output nodes (respectively) present in the network.

In comparison, Table 6 presents the LDA results obtained from the test data set. The overall classification accuracy for the DA model was 58.38%.

The ANN model was recreated an additional 30 times with randomly selected initial weights to evaluate the nature of the ANN's response surface. The resulting mean classification accuracy was 70.52%, with a 95% confidence interval of 70.26–70.80% and a standard deviation of 0.7293. This narrow confidence interval indicates that the response surface is fairly smooth in the region surrounding the solution found by the ANN.

The classification accuracies of each reduced model for both the ANN and the DA models were also determined. Fig. 4 compares these classification results for the

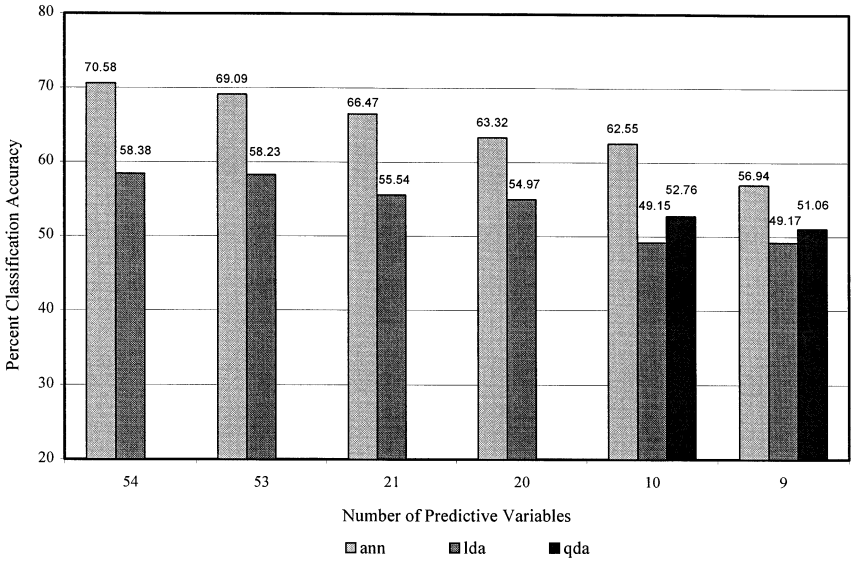

Fig. 4. Comparison of artificial neural network and discriminant analysis classification results (ann = artificial neural network; lda = linear discriminant analysis; qda = quadratic discriminant analysis).

Table 5
Artificial neural network classification matrix for the test data set (54 variables)[a]

Observed	Predicted							Observed totals	Observed percent
	SF	LP	PP	CW	AS	DF	KR		
Spruce/fir (SF)	150 397	39 435	440	0	6683	481	12 244	209 680	37.05
Lodgepole pine (LP)	50 871	186 050	5697	73	27 391	8796	2263	281 141	49.68
Ponderosa pine (PP)	2	258	25 295	2012	625	5402	0	33 594	5.94
Cottonwood/ willow (CW)	0	0	15	563	0	9	0	587	0.10
Aspen (AS)	20	381	138	0	6737	56	1	7333	1.30
Douglas-fir (DF)	0	126	1672	446	161	12 802	0	15 207	2.69
Krummholz (KR)	638	128	1	0	6	0	17 577	18 350	3.24
Predicted totals	201 928	226 378	33 258	3094	41 603	27 546	32 085	565 892	
Predicted percent	35.68	40.00	5.88	0.55	7.35	4.87	5.67		100.0

[a] Network architecture: 54-120-7 (LR = 0.05, MR = 0.5). Overall classification accuracy: 70.58%.

Table 6
Discriminant analysis classification matrix for the test data set (54 variables)[a]

Observed	Predicted							Observed totals	Observed percent
	SF	LP	PP	CW	AS	DF	KR		
Spruce/fir (SF)	136 200	40 835	177	0	9178	1119	22 171	209 680	37.05
Lodgepole pine (LP)	61 466	146 780	5778	121	53 262	12 239	1495	281 141	49.68
Ponderosa pine (PP)	0	101	18 384	3303	1090	10 716	0	33 594	5.94
Cottonwood/ willow (CW)	0	0	93	429	0	65	0	587	0.10
Aspen (AS)	249	1488	581	0	4648	367	0	7333	1.30
Douglas-fir (DF)	0	449	3702	628	1301	9127	0	15 207	2.69
Krummholz (KR)	3440	5	71	0	26	0	14 808	18 350	3.24
Predicted totals	201 355	189 658	28 786	4481	69 505	33 633	38 474	565 892	
Predicted percent	35.58	33.51	5.09	0.79	12.28	5.94	6.80		100.0

[a] Overall classification accuracy: 58.38%.

ANN, LDA, and QDA methods. The QDA accuracies are only provided for the predictive variables subsets which produced valid results (i.e. set including only quantitative variables). As this figure shows, the ANN models outperform the corresponding DA methods across all predictive variable subsets for this data set.

8. Discussion

The ANN models consistently outperformed the DA models in the prediction of forest cover types. One reason for this result may be the assumptions associated with most statistical analysis techniques, including DA. In discriminant analysis, the distributions of both the dependent and independent variables are assumed to be normal (Gaussian), and for LDA the covariance matrices for each group are assumed to be equal (homoscedasticity).

The data set investigated in this study contained both qualitative and quantitative variables, some of which were clearly not normally distributed. This factor could have effected the LDA results. However, previous studies have found that LDA is very robust to mixed data types. As Yoon et al. (1993) states:

> In applied research, data are seldom compatible with the underlying assumptions needed to perform statistical inferences. In many fields, like social and behavioral sciences, business, and biomedical sciences, measurement of hard data is still a problem. Measurements are, at best, of nominal or ordinal nature. Usually, research workers ignore the discrete nature of the data and proceed with classification using Fisher's LDF. In such situations they get a useful, but not an optimal rule of classification.

The non-optimality of the LDA results in mixed data type situations was obvious; in this study, the non-optimality lowered the accuracy of the LDA results to a level lower than that produced by the artificial neural networks.

In addition to parametric discriminant analysis, a single form of non-parametric discriminant analysis was also evaluated in this study. This non-parametric version did not perform noticeably differently from the parametric DA versions. However, there are many forms of non-parametric DA available, each with their own unique strengths and weaknesses. Thus, the single non-parametric DA conducted in this study cannot be considered to be representative of all possible non-parametric DAs.

The ANN model produced higher classification accuracies than the DA approach, both in cases where only quantitative independent variables were used and where both qualitative and quantitative independent variables were employed. This may be due to the fact that the artificial neural network approach makes no explicit assumptions regarding the underlying distributions of the variables involved (Marzban et al., 1997). Another benefit is the ability of the artificial neural network structure to partitioning the input space into regions that allow for classification of linearly inseparable data (Yoon et al., 1993). Consequently, one would expect an

ANN model to outperform a LDA model for most classification tasks involving non-linear data. Non-linearity was certainly present in the data used in this study.

When comparing the individual classification matrices of the best ANN model against that of the best DA model (the two 54 independent variables models), as shown in Tables 5 and 6, many insights may be drawn concerning the respective classification strategies of the two models. In general, both the ANN and DA models seemed to misclassify ponderosa pine, Douglas-fir, and cottonwood/willow cover types primarily with each other. These misclassifications may be due to the actual geographic proximity of the cover types, since they are all principally found in the Cache la Poudre wilderness area. Both predictive models also seemed to misrepresent krummholz as spruce/fir and to a lesser extent as lodgepole pine. All three of these forest cover types are typically considered high elevational species, which might have caused the classification confusion in some instances. The aspen cover type was generally misclassified by both the ANN and DA models as lodgepole pine, but these misclassifications were much more frequent in the DA model (381 observations misclassified for the ANN model and 1488 observations for the DA model). In addition, the lodgepole pine cover type was primarily misrepresented by both the ANN and DA models as either spruce/fir or aspen. However, the DA model misclassified a much larger number of actual lodgepole pine observations as aspen (53 262 observations misclassified for the DA model and 27 391 observations for the ANN model). This classification confusion for the lodgepole pine and aspen cover types and disparity of numbers of misclassifications by the DA model may be due in part to its inability to take advantage of the presumably non-linear influence of the horizontal distance to the nearest wildfire ignition point variable. Another likely factor would be that lodgepole pine and aspen cover types are generally located within the same altitudinal zone (Whitney, 1992) and are frequently found as neighboring forest stands within the study area.

In addition to overall classification accuracies, Tables 5 and 6 also show prediction accuracies for each forest cover type. When comparing correct predictions for each cover type (the major diagonal through each classification matrix) between the ANN model and the corresponding DA model, the ANN model produces more accurate predictions for every cover type.

Furthermore, when considering percent of predicted forest cover types (column totals) against percent of observed forest cover types (row totals), the ANN model was superior to the DA model. Although this measure does include omission and commission errors, the ANN model was closer to the observed percentage of each forest cover type than the corresponding DA model. Overall, these detailed measures of classification accuracy indicate that the ANN model outperformed the DA model for overall classification accuracy, individual forest cover type accuracy, and the percent of predicted forest cover type totals.

An obvious burden of the artificial neural network approach is the need to divide the data into three separate sets for analysis rather than the traditional two sets used in most conventional statistical methods. Fortunately, this was not a problem in this study; a very large number observations were available for analysis and hence all three data sets contained adequate numbers of observations. However, this fact may pose a problem for other applications where data is not so plentiful.

Another consideration when comparing the two classification techniques involves the amount of computational time required to develop each model. The ANN model outperformed the DA model in classification accuracy, but also demanded a greater amount of time to generate a set of network parameters. For example, the ANN model with 54 independent variables required 56 computer runs (i.e. different networks) to determine the 'best' set of network parameter values, with each run taking roughly 45 h to complete (using a UNIX Sun Sparc workstation). In contrast, the DA model with 54 independent variables (using the same workstation) required only a single computer run that lasted only ~ 5 min. This major time difference should be considered by those analysts comparing these two techniques.

In a previous paper (Blackard and Dean, 1996), we described an earlier attempt to predict forest cover types from cartographic variables. In contrast to the findings reported here, this earlier attempt concluded that both the ANN and DA methods performed rather poorly. We feel that these earlier findings reflected the sensitivity of the DA approach to the manner in which the available data is divided into training, validation and testing sets, and to some of the characteristics of the data sets themselves. In the earlier study, the training and validation data sets were created by including every raster cell located within one or more subjectively chosen geographically contiguous areas. Attempts were made to ensure that these contiguous regions represented the full range of variability throughout the study area, but such full representation was not completely possible. Some amount of analyst bias was undoubtedly present by using this type of procedure to develop training and validation data sets. In contrast, this current study randomly assigned raster cells to the various data sets, thereby ensuring a truly random (and presumably representative) sample within each set.

In addition, the sampling scheme used in the current study ensured that within both the training and validation data sets, each forest cover type was represented by an equal number of observations. No effort was made to equalize cover type representations in the data sets used in the previous study. By incorporating a random sampling system along with minor refinement of the predictive variables, this study produced ANN models with greater predictive accuracy than comparable DA models.

9. Conclusions

This study evaluated the ability of an artificial neural network (ANN) model to predict forest cover type classes in forested areas that have experienced relatively little direct human management activities in the recent past. The ANN model was evaluated based both on its absolute accuracy and on its ability relative to a model based on discriminant analysis (DA). In general, the ANN model produced good classification results, with greater accuracies than those produced by the DA model.

For comparison purposes, a study involving land cover classification from remotely sensed data was recently conducted for the Colorado State Forest, which

lies along the western edge of the study area used in this research project. The Colorado State Forest study (Croteau, 1994) used Landsat Thematic Mapper data to classify 30×30-m observations into one of 11 relatively broad land cover classes (water, conifer, aspen, willow/wet meadow, mountain meadow, alpine vegetation, clearcut, range, non-vegetation, sand, and snow). By using remotely sensed data describing actual landforms and vegetation, rather than the cartographic data used in this study that describes only site characteristics, predictive accuracies of 71.1% were achieved (Croteau, 1994). The classification accuracy achieved with the artificial neural network model in this study compares very favorably with this remote sensing accuracy. Along with those results obtained from previous ANN studies, these findings suggest that while the ANN approach does have its own drawbacks, it can still be a viable alternative to traditional approaches to developing predictive models.

References

Alvarez, S., 1995. Generation of terrain textures using neural networks. Unpublished masters thesis. Department of Computer Science, Colorado State University, 44 pp.

ARD Corporation, 1993. Propagator Neural Network Development Software — A User's Manual for Sun. Version 1. ARD Corporation, Columbia, MD.

Bharath, R., Drosen, J., 1993. Neural Network Computing. Windcrest/McGraw-Hill, New York 188 pp.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon, Oxford 482 pp.

Blackard, J.A., Dean, D.J., 1996. Evaluating integrated artificial neural network and GIS techniques for predicting likely vegetative cover types. In: First Southern Forestry GIS Conference. G.J. Arthaud, W.C. Hubbard (Eds.). University of Georgia, Athens, GA, 416 pp.

Croteau, K., 1994. Integrating Landsat Thematic Mapper data and ancillary spatial data for mapping land cover in the Colorado State Forest. Unpublished masters thesis. Department of Forest Sciences, Colorado State University, 137 pp.

Curram, S.P., Mingers, J., 1994. Neural networks, decision tree induction and discriminant analysis: an empirical comparison. J. Oper. Res. Soc. 45 (4), 440–450.

Environmental Systems Research Institute, 1991. Cell-Based Modeling with GRID. Environmental Systems Research Institute, Redlands, CA.

Fowler, C.J., Clarke, B.J., 1996. Corporate distress prediction: a comparison of the classification power of a neural network and a multiple discriminant analysis model. Accounting Forum 20 (3–4), 251–269.

Gimblett, R.H., Ball, G.L., 1995. Neural network architectures for monitoring and simulating changes in forest resource management. AI Appl. 9 (2), 103–123.

Guan, B.T., Gertner, G.Z., 1995. Modeling individual tree survival probability with a random optimization procedure: an artificial neural network approach. AI Appl. 9 (2), 39–52.

Harvey, W., Dean, D.J., 1996. Computer-aided visualization of proposed road networks using GIS and neural networks. In: First Southern Forestry GIS Conference. G.J. Arthaud, W.C. Hubbard (Eds.). University of Georgia, Athens, GA, 416 pp.

Haykin, S., 1994. Neural Networks — A Comprehensive Foundation. MacMillan College, New York 696 pp.

Hertz, J., Krogh, A., Palmer, R.G., 1991. Introduction to the Theory of Neural Computation. Addison-Wesley, Reading, MA 429 pp.

Huberty, C.J., 1994. Applied Discriminant Analysis. Wiley, New York 466 pp.

Jan, J.-F., 1997. Classification of remote sensing data using adaptive machine learning techniques. Unpublished dissertation paper. Department of Forest Sciences, Colorado State University, 140 pp.

Johnson, R.A., Wichern, D.W., 1992. Applied Multivariate Statistical Analysis, 3rd ed. Prentice-Hall, Englewood Cliffs, NJ 642 pp.

Markham, I.S., Ragsdale, C.T., 1995. Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis. Decis. Sci. 26 (2), 229–242.

Marzban, C., Stumpf, G., 1996. A neural network for tornado prediction based on Doppler radar-derived attributes. J. Appl. Meteorol. 35, 617–626.

Marzban, C., Paik, H., Stumpf, G., 1997. Neural networks versus Gaussian discriminant analysis. AI Appl. 11 (1), 49–58.

Masters, T., 1994. Signal and Image Processing with Neural Networks: A C++ Sourcebook. Wiley, New York 417 pp.

McLachlan, G.J., 1992. Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York 526 pp.

Paruelo, J.M., Tomasel, F., 1996. Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. Unpublished research paper. Department of Rangeland Ecosystem Science, Colorado State University, 33 pp.

Pattie, D.C., Haas, G., 1996. Forecasting wilderness recreation use: neural network versus regression. AI Appl. 10 (1), 67–74.

Patuwo, E., Hu, M.Y., Hung, M.S., 1993. Two-group classification using neural networks. Decis. Sci. 24 (4), 825–845.

Reibnegger, G., Weiss, G., Werner-Felmayer, G., Judmaier, G., Wachter, H., 1991. Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees. Proc. Natl. Acad. Sci. 88, 11426–11430.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, New York 403 pp.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. MIT Press, Cambridge, MA Chapter 8.

SAS Institute, Inc., 1989. SAS/STAT User's Guide, Version 6, 4th ed. SAS Institute, Cary, NC Volume 1.

Vega-Garcia, C., Lee, B.S., Woodard, P.M., Titus, S.J., 1996. Applying neural network technology to human-caused wildfire occurrence prediction. AI Appl. 10 (3), 9–18.

Whitney, S., 1992. Western Forests. Knopf, New York 672 pp.

Wong, P.M., Taggart, I.J., Gedeon, T.D., 1995. Use of neural network methods to predict porosity and permeability of a petroleum reservoir. AI Appl. 9 (2), 27–37.

Yoon, Y., Swales, G., Margavio, T.M., 1993. A comparison of discriminant analysis versus artificial neural networks. J. Oper. Res. Soc. 44 (1), 51–60.

Biographies

Jock A. Blackard is a GIS Analyst with the Ecosystem Management Analysis Center for the USDA Forest Service at Fort Collins, CO. His work places emphasis on the role of GIS in the development of forest planning models, with research interests in the application of GIS to natural resource planning issues and the extension of geographic analysis capabilities. He received a Ph.D. degree from the Remote Sensing and GIS Program, Department of Forest Sciences at Colorado State University in 1998.

Denis J. Dean is an Associate Professor of Natural Resource Information Systems in the Remote Sensing and GIS Program, Department of Forest Sciences, Colorado State University. He teaches introductory to advanced courses in the area of spatial data capture, analysis and visualization. His research interests are focused on the development of new methods and techniques of utilizing spatial data in natural resource management.