Natural Resources Research, Vol. 12, No. 2, June 2003 (© 2003)

Use of Noise to Augment Training Data: A Neural Network Method of Mineral–Potential Mapping in Regions of Limited Known Deposit Examples

Warick M. Brown[1,2], Tamás D. Gedeon[1], and David I. Groves[2]

Received 2 December 2002; accepted 14 February 2003

One of the main factors that affects the performance of MLP neural networks trained using the backpropagation algorithm in mineral-potential mapping is the paucity of deposit relative to barren training patterns. To overcome this problem, random noise is added to the original training patterns in order to create additional synthetic deposit training data. Experiments on the effect of the number of deposits available for training in the Kalgoorlie Terrane orogenic gold province show that both the classification performance of a trained network and the quality of the resultant prospectivity map increase significantly with increased numbers of deposit patterns. Experiments are conducted to determine the optimum amount of noise using both uniform and normally distributed random noise. Through the addition of noise to the original deposit training data, the number of deposit training patterns is increased from approximately 50 to 1000. The percentage of correct classifications significantly improves for the independent test set as well as for deposit patterns in the test set. For example, using ±40% uniform random noise, the test-set classification performance increases from 67.9% and 68.0% to 72.8% and 77.1% (for test-set overall and test-set deposit patterns, respectively). Indices for the quality of the resultant prospectivity map, (i.e. D/A, $D \times (D/A)$, where D is the percentage of deposits and A is the percentage of the total area for the highest prospectivity map-class, and area under an ROC curve) also increase from 8.2, 105, 0.79 to 17.9, 226, 0.87, respectively. Increasing the size of the training-stop data set results in a further increase in classification performance to 73.5%, 77.4%, 14.7, 296, 0.87 for test-set overall and test-set deposit patterns, D/A, $D \times (D/A)$, and area under the ROC curve, respectively.

KEY WORDS: Neural networks; multilayer perceptrons (MLP); random noise; mineral prospectivity maps; geographic information systems (GIS); Archean orogenic gold deposits.

INTRODUCTION

The work described here is part of a study on the use of neural networks to predict the potential for orogenic lode-gold deposits in an area near Kalgoorlie, Western Australia (Brown and others, 1997, 2000).

An important problem in applying neural networks to predict the potential for mineral deposits from exploration data is the rarity of deposit training patterns compared to the large number of patterns that correspond to barren cells. In addition, small and poorly documented mineral occurrences typically make up a large proportion of the deposit data. In order to ensure that the network correctly learns the patterns in the data, the training data should sample the entire range of variation in the data population and adequately represent feature vectors that are located close to decision boundaries in multidimensional feature space. Both these requirements are difficult to

[1] School of Information Technology, Murdoch University, South Street, Murdoch 6150, Western Australia; e-mail: wbrown@geol.uwa.edu.au.

[2] Centre for Global Metallogeny, School of Earth and Geographical Sciences, University of Western Australia, Perth, Western Australia.

1520-7439/03/0600-0141/1 © 2003 International Association for Mathematical Geology

fulfil because there are so few deposit examples relative to barren patterns. An additional problem arising from the gradient-descent backpropagation algorithm (Rumelhart, Hinton, and Williams, 1986) used to train multilayer perceptron (MLP) networks is that minimization of the network error function requires that there are approximately equal numbers of training examples from each class (that is deposits and barren cells) irrespective of the actual *a priori* probability of occurrence of the classes (Parikh, Pont, and Jones 1999; Zaknich, 2003). In most instances the solution to this problem is to collect more data. However, this is not feasible for ore deposits and, even in areas well endowed with mineral resources, map grid cells containing deposits are rare compared to barren cells.

PREVIOUS WORK

A number of studies show that adding noise to the input data, outputs, weight connections, and weight changes can improve the generalization ability of trained networks (Sietsma and Dow, 1991; Holmstrom and Koistinen, 1992; Matsouka, 1992; Clay and Sequin, 1992; Krogh and Hertz, 1992). An (1996) demonstrated that input noise is effective in improving the generalization performance for both regression (function approximation) and classification problems. Bishop (1993) showed that the addition of noise is equivalent to error regularization, and Reed, Marks, and Oh (1992, 1995) compared the effects of adding noise on the generalization performance to error regularization, sigmoid gain scaling (that is, reducing the slope of the sigmoid function), and target smoothing. More recently, Wang and Principe (1999) investigated the effect of injecting noise into the target output on learning speed and the ability to avoid local minima. Most papers concentrate on the effect of noise on the generalization ability.

Comparatively little work has been done on the composition of the training sets and how to deal with problems where the number of training examples representing different classes is significantly different. Parikh, Pont, and Jones (1999) studied the performance of a MLP classifier in a condition monitoring and fault diagnosis application, where data representing normal operation are readily available compared to fault data, which may be difficult or expensive to obtain. They showed that the highest overall network performance is obtained where roughly equal numbers of samples from the normal and fault classes are used for training.

GIS DATABASE

The work described here is based on a regional GIS data set, which is being used to examine the prospectivity for orogenic lode-gold deposits in an approximately 100×100 km area of the Archean Yilgarn Block, near Kalgoorlie in Western Australia. Sets of 10 and 17 GIS layers in raster format were used to create feature vectors for each 100 m cell on the map grid. The map area represents an 1140×1100 array consisting of 1,254,000 cells. Because GIS data are captured at a regional scale, only the 120 deposits with a total contained gold resource ≥ 1000 kg are used to train and test the neural networks.

Archean orogenic gold deposits form selectively where fluids have been focussed into dilatant zones or zones of high structural permeability, which, in turn, are produced by heterogeneous stress distributions in areas with complex geometries and strong contrasts in rock strength (Groves and others, 2000). A mineral systems approach, similar to that described by Wyborn, Gallagher, and Raymond (1995), is used to translate these essential ingredients of the mineralizing systems responsible for the formation of orogenic lode-gold deposits into mappable criteria that are likely to have been incorporated into a regional-scale GIS database. Many of the exploration criteria employed in this study were identified by Groves, Ojala and Holyland (1997) and Groves and others (2000), who have emphasized the importance of the late-kinematic timing of orogenic gold deposits to computer-aided exploration techniques. An important consequence of this late timing is that the present regional-scale structural geometries of the deposits and enclosing terranes, as depicted in solid geology maps, are essentially similar to those during gold mineralization and can be used to identify exploration criteria based on repetitive and predictable geometries.

GIS layers used as inputs to neural networks are selected from approximately 60 candidate layers according to the strength of the association between known orogenic gold deposits and the variable forming the thematic layer. Spatial associations are determined using plots of cumulative and interval-based calculations of a bivariate J-function, $(O - E)/E$ (where O is the observed and E is expected number of deposits in a buffer), cumulative contrast of weights, χ^2, and the Kolmogorov-Smirnov cumulative distribution function (see Brown 2002; Brown and others, 2002 for details). The following ten GIS thematic layers are used as inputs for neural networks:

(1) favorable geology (as a fuzzy membership layer), (2) distance to porphyritic felsic intrusions, (3) distance to the nearest crustal-scale fault or shear zone (defined as those exceeding 100 km in length), (4) strike of the nearest NNE-trending regional-scale shear zone, (5) rheological contrast between rock units at the nearest lithological boundary (fuzzy-membership), (6) chemical contrast between rock units at the nearest lithological boundary (sulfidation index based on the ratio, Fe_{total} wt % $\times$ Fe/(Fe + Mg + Ca), which determines if iron sulfides will form and gold complexes will be destabilized), (7) density of crustal and regional-scale faults and shear zones (calculated using a 5 $\times$ 5 km moving window), (8) density of lithological contacts (used as a measure of lithological heterogeneity), (9) favorability of rock-types at the nearest lithological boundary (fuzzy-membership), and (10) the distance to the nearest ground trace of a regional-scale anticlinal axis. A binary deposit layer depicting orogenic lode-gold deposits with a resource of $\geq$1000 kg total contained metal is used to provide the target or desired output values required to train the networks.

With the exception of the rheological-contrast layer, fuzzy-membership layers are based on the statistical association between predictive variables and known deposits. Although rheological contrast at rock boundaries also is modeled using Young's Modulus, uniaxial compressive strength, uniaxial tensile strength and fracture toughness, a subjective estimate of rock competency in the form of fuzzy membership values based on the field observations by one of the authors (Groves) shows the strongest positive association with known deposits (see Brown, Groves, and Gedeon, 2002) and therefore is used as a neural network input.

METHOD

Multilayer Perceptron (MLP) neural networks are trained to perform a pattern recognition task using training data consisting of examples of deposit and barren cell patterns (Brown and others, 1997, 2000). The input patterns consist of the values obtained from the co-located grid cells from each of the input GIS layers. During training, the network is presented with each input pattern in the training data set, together with the corresponding binary value from the deposit layer (1 = deposit present, 0 = deposit absent). Dayhoff (1990) and Masters (1993) provide useful introductions to MLP networks, data preprocessing, and training. Provided the network has correctly learnt the underlying patterns in the training data set, the trained network then can generalize to predict the correct output for completely new patterns. The data processing required to combine a set of GIS layers into a prospectivity map consists of the following six steps:

1. create a feature vector for each grid cell location,
2. create training data sets,
3. train a set of networks,
4. process feature vectors for entire map using best trained network,
5. produce prospectivity maps, and
6. analyse map quality.

GIS layers are prepared using the ArcView GIS package. All other data processing is performed using built-in functions and additional routines written in a C-like scripting language in Matlab.

Training Data Sets

All input values are scaled to the range [0, 1]. As noted, the requirement to have approximately equal numbers of training patterns from each of the deposit and barren classes means that training set size is limited by the number of known deposits. Because the GIS data are captured at a regional scale, only the 120 deposits with a total contained gold resource $\geq$1000 kg, together with an approximately equal number of barren patterns, are used to train and test the neural networks. The original training data (that is those without added noise) are created by randomly splitting the deposit patterns into three approximately equal groups; i.e., training, training-stop, and test sets. Deposits from each rock-type in the solid geology GIS layer are divided approximately equally between the three training sets. A number of barren patterns approximately equal to the deposit patterns also is selected randomly for each rock type. There are slightly more barren than deposit patterns because some rock types do not contain any known deposits. The composition of the training sets is shown in Table 1.

The training set is used to train the network weights, the training-stop set is used to determine when to stop training in order to avoid over-fitting the data, and an independent test data set is used to check the performance of the trained network.

The role of the training-stop data set is to check the generalization ability of the network during training in order to determine an optimal stopping

Table 1. Number of Patterns in the Training Data Sets for Orogenic Gold Deposits in the Kalgoorlie Terrane

	Training set			Stop set			Test set		
Training set type	Deposit	Barren	Total	Deposit	Barren	Total	Deposit	Barren	Total
Original	46	53	99	39	49	88	35	46	81
Noise; train	1058	1053	2111	39	49	88	35	46	81
Noise; train, stop	1058	1053	2111	897	895	1792	35	46	81

point for training. During batch training, the network weights are adjusted iteratively in response to the network error (i.e., the difference between the actual network output and the target output) after all the patterns in the training data set are processed. Target outputs are presented to the network, together with the input pattern during training. Initially, the network learns the underlying patterns in the data. However, if the number of weights (approximately input variables times hidden neurons) is large in relation to the size of the training set, the network eventually may learn the idiosyncrasies in individual training patterns, thus reducing its ability to generalize when presented with new patterns after training is completed. To avoid overfitting the data, the classification performance of the partially trained network is checked using the training-stop data set. Training is halted when the training-stop data set error begins to increase because the network has begun to memorize the patterns in the training set.

Network Topology and Training

The network topology used for all noise experiments is determined by comparing the performance of networks with different numbers of hidden units. A 10-20-1 topology (that is, 10 inputs, 20 hidden neurons and 1 output) gives the best generalization performance (as measured by percentage of correct classifications for the independent test set). Although standard gradient-descent and gradient-descent training algorithms with adaptive learning rates and momentum give the best generalization performance and values of map quality indices (see explanation next), the Levenberg-Marquardt training algorithm (Hagen and Menhaj, 1994) was selected for the experiments. The advantage of the Levenberg-Marquardt algorithm is that it can converge from 10 to 100 times faster than the adaptive and standard gradient-descent algorithms, respectively.

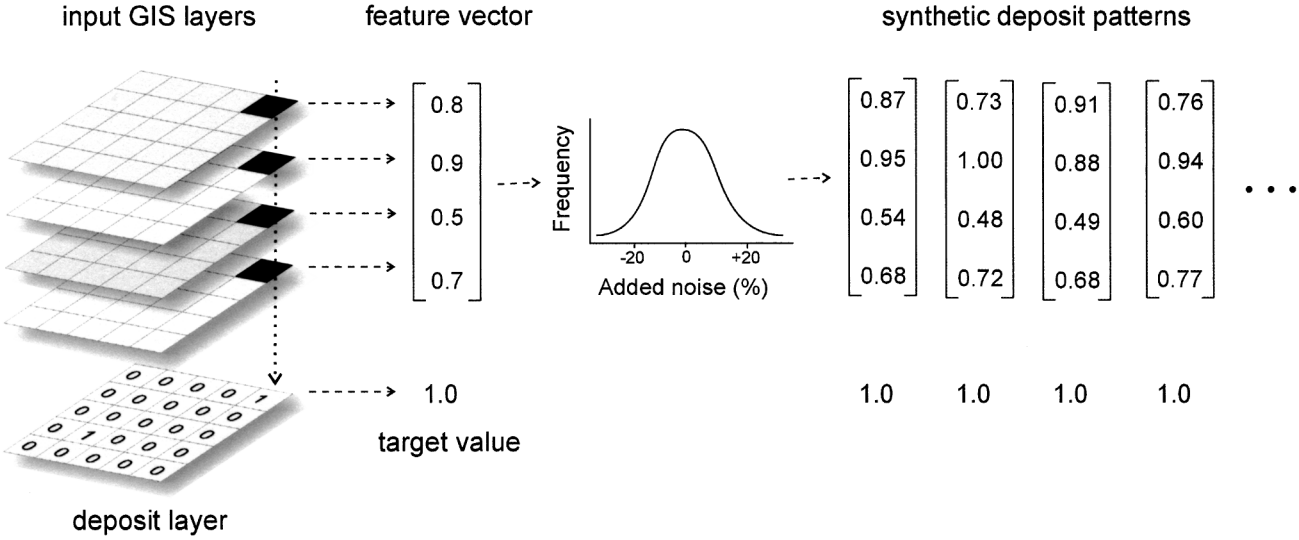

Figure 1. Schematic diagram of data processing steps used to augment the training data set through addition of random noise to deposit patterns. For each original deposit pattern, 22 new deposit patterns were created. This increased the number of deposit patterns in the training set from 46 to 1058.

Ten networks with different initial weights are trained for each experiment and the average and best test-set performances are tabulated. The same ten initial, random-weight settings are used for each set of experiments. Consequently, differences in the results are due to the changed experimental variables (for example, composition of the training set) rather than the starting weights.

Augmentation of Training Data Set With Noise

In order to augment the original training data sets, random noise is added to each component of the original deposit feature vectors in the training data set to create 22 new patterns from each original pattern (Fig. 1). This ratio of new to original patterns results in 1058 deposit patterns compared with 46 in the original training set. New barren patterns are selected randomly from the map grid so that the numbers of deposit and barren patterns are approximately equal and an equal number of barren patterns are selected from each host-rock type (Table 1). For experiments to determine the optimum amount and type of noise, the training-stop and test data sets are not changed. In a final experiment, the training-stop data set is also expanded using additive noise.

Experiments are performed using $\pm 5, \pm 10, \pm 20, \pm 40$, and $\pm 80\%$ noise in an attempt to determine the optimal amount of noise. Two different types of noise; uniform random and normally distributed or Gaussian random noise are tested. In the case of uniform random noise, the percentages represent the maximum amount added or subtracted. For normally distributed noise, the standard deviation is set to a fraction of the value represented by the percentage. The mean is set to the value of the component in the original feature vector.

Measures of Generalization Performance and Prospectivity Map Quality

Results of the experiments were assessed using the classification performance on the test data (independent validation) set as a guide to the generalization ability of the trained network. Activation values produced as the network output are approximately in the range [0.05, 0.95]. Patterns associated with output activations greater or equal to 0.5 were classified as deposits.

Network performance is also assessed on the basis of the quality of the prospectivity maps. Patterns corresponding to the entire map area are processed using the trained network and the network outputs are converted to a nine-class prospectivity map. Each map then is analyzed to obtain a variety of statistical measures of the quality of the prospectivity map. Spearman's correlation coefficient indicates the degree of correlation between increasing prospectivity map class and the probability that the class region contains a known deposit. The capture efficiency ratios, D/A and $D \times (D/A)$, are calculated using the highest prospectivity map class, and measure the degree to which the most prospective map class accounts for the known deposits. D refers to the percentage of known deposits in the area corresponding to the highest class, and A refers to the percentage of the total map area represented by the region. A problem with the first ratio, D/A, is that the value can be high if only a few deposits occur in a small area. In order to account for this, the first capture efficiency ratio is multiplied by the percentage of known deposits that occur in the highest prospectivity class.

The receiver-operating-characteristic (ROC) method has been used to evaluate performance of signal detection and diagnostic systems as well as the performance of neural networks (Uncini and others, 1990; Zaknich, 2003). A ROC curve is generated by plotting the true positive, TP/(TP + FN), and false positive ratios, FP/(FP + TN), together for a range of threshold values (where TP = true positive, TN = true negative, FP = false positive and FN = false negative; see Table 2). In this study, positive and negative classes represent deposits and barren cells, respectively. The ROC curve shows how classification performance varies with the threshold value used to divide feature vectors into the two classes according to the activation at the output unit of the neural network (Figs. 2 and 3). The area under the ROC curve is used as a measure of classification

Table 2. Confusion Matrix Showing the Possible Outcomes for a Classification Decision Involving Two Classes (After Zaknich, 2003)

	True classification	
	Positive	Negative
System		
Positive	True+ve (TP)[a]	False+ve (FP)
Negative	False−ve (FN)	True−ve (TN)

[a] Abbreviations: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). These Variables are Used to Plot the Receiver-Operating Characteristic Curve Shown in Figure 3. The Values are Obtained by Counting the Patterns in Each Category.

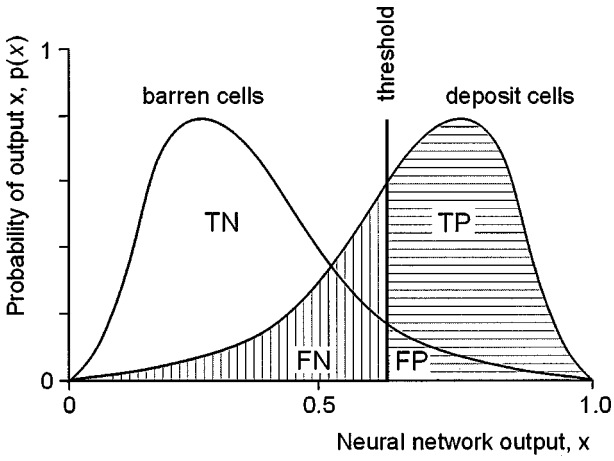

Figure 2. Schematic illustration of probability distribution functions corresponding to barren (left) and deposit classes (right) of input patterns that are classified by neural networks in this study. Range of output values (x-axis) is produced by a neural network in response to input patterns corresponding to both known deposit and barren cells. Probability p(x) that a deposit cell will produce a particular output x is indicated on the y-axis. Abbreviations: TP = true positive, FP = false positive, TN = true negative and FN = false negative. Patterns corresponding to grid cells are assigned to categories by applying threshold to output values. For example, false negative patterns are patterns that are classified as barren but which actually belong to the deposit class. Variables TP, FP, TN, and FN are used to plot receiver-operating-characteristic (ROC) curve shown in Figure 3.

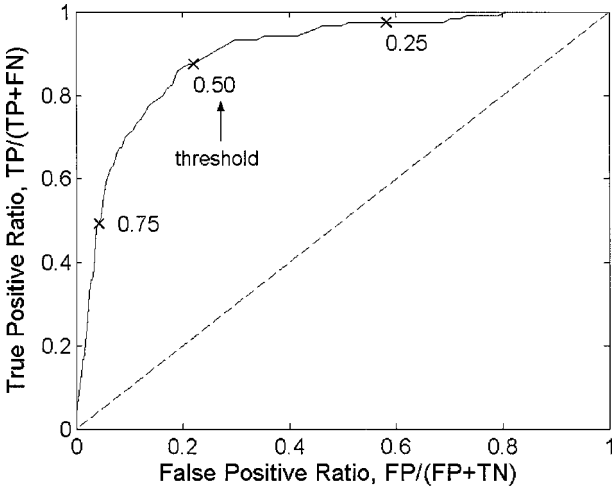

Figure 3. Receiver-operating-characteristic (ROC) curve for a neural network prospectivity map (solid curve). Curve shows how true-positive and false-positive ratios vary with threshold value applied to network output in order to classify cell patterns as either deposits or barren. Area under curve is an indicator of system performance and ranges between 0.5 for a random classifier (dashed line) and 1.0 for perfect discrimination.

performance and ranges from a minimum of 0.5 to a maximum of 1.0 (Fig 3). Harris and Pan (1999) used a similar function in which different threshold values are applied to the outputs from various classifiers to determine which cells to retain for further exploration. They plot the percentage of mineralized cells (TP/(TP + FN)) for which the classifier output is above the threshold (and therefore retained) versus the percentage of total cells that are retained ((TP + FP)/(TP + FP + TN + FN)). Their plot represents a nearly identical function to the one used in this study because of the small proportion of the total number of cells represented by deposits.

RESULTS

Experiment 1. Effect of Numbers of Deposits in Training Set

In order to test the way in which the limited number of deposit patterns available for training affects the performance of the trained network, a series of training data sets are prepared using 5, 10, 20, 40, 80, and 100 deposit patterns. For each of these numbers of deposit patterns, ten training data sets are created by randomly allocating deposits to each of the three training sets. Only the 120 known deposits with a total resource of at least 1000 kg gold are used to train the networks. Where possible, the training and training-stop data are allocated the same number of deposit patterns. If this is not possible, the training-stop and test data sets were allocated equal numbers of deposits. In all cases, the ten data sets contain the same numbers of deposit and barren patterns. The composition of the training sets is shown in Table 3.

The results are presented in Figures 4–7, which show the percentage correct classification for the test set overall, percentage correct classification for deposit patterns in the test set, $D \times (D/A)$ and area under the ROC curve. The first two variables are measures of the generalization ability of the trained networks and the last two variables are measures of the quality of the prospectivity map. Each point on the plots effectively represents the average for 100 networks because ten networks with different random initial weights were each trained using ten different data sets. All of the plots show a trend of significantly increasing generalization ability of the networks and map quality with increasing numbers of deposit patterns and training data set size. The slight drop in

Table 3. Number of Patterns in Training Sets Used to the Test Effect of the Number of Deposit Examples on the Generalization Performance of Trained Networks. Patterns Represent Cells Containing Orogenic Gold Deposits and Barren Cells from Raster GIS Data for the Kalgoorlie Terrane

Number of deposits	Training set			Stop set			Test set		
	Deposit	Barren	Total	Deposit	Barren	Total	Deposit	Barren	Total
5	5	5	10	5	5	10	110	110	220
10	10	10	20	10	10	20	100	100	200
20	20	20	40	20	20	40	80	80	160
40	40	40	80	40	40	80	40	40	80
80	80	80	160	20	20	40	20	20	40
100	100	100	200	10	10	20	10	10	20

performance for 100 training deposits indicated in Figure 4 may be the result that there are only 10 deposits in the training-stop and test data sets. A similar effect appears to occur in Figure 6 for 80 deposit patterns. Because of decreasing size of the test data sets, the reliability of the results decreases with increasing numbers of deposit patterns. Plots of the overall test-set performance, test-set performance for deposit patterns, and the area under the ROC curves suggest that the effect of increasing deposits decreases as deposits are added (Figs. 4, 5 and 7, respectively).

Experiment 2. Amount of Noise

Tables 4 and 5 show the results of experiments using various amounts of uniform random noise. The same network topology (10-20-1) and set of initial weights are used for each of the ten networks trained for each noise level. Each row in Table 4 shows the average results for ten networks, whereas Table 5 shows the network that gives the best overall test-set performance. In Tables 4–7, results for the original training set without added noise are shown in the first row, and the rows in bold show the results obtained by expanding both the training and training-stop set with noise.

The average results (Table 4) clearly show that the increased training-set size, resulting from the addition of noise, results in a significant improvement in the percentage classification of the test set and the four measures of prospectivity map quality (Spearman's correlation coefficient ρ, capture efficiency statistics, D/A, $D \times (D/A)$ and the area under the ROC). In most cases, the values of D/A and $D \times (D/A)$ are approximately doubled. The results for $\pm 40\%$ noise represent a significant improvement in the percentage correct classification for deposit patterns (that is, from 68.0% to 77.1%), although for most other noise levels there is only a slight improvement. The best test-set results (Table 5) differ considerably with different amounts of added noise. Usually the classification performance for barren cells and the $D \times (D/A)$ statistic are poorer than for the network trained without added noise (top row in Table 5). However, the results for ρ, D/A and the area under the ROC are improved significantly and the correct classification of test-set deposit patterns for the $\pm 40\%$ noise level show a similarly large improvement as displayed in the average results (i.e., from 77.1% to 85.7%). An interesting feature of the average results is that there seems to be a weak inverse relationship between correct classification performance for test-set deposit and barren patterns (see, for example, the results for $\pm 20\%$ and $\pm 40\%$ in Table 4).

Examples of mineral prospectivity maps for part of the Kalgoorlie Terrane, Western Australia, showing the potential for orogenic lode-gold deposits are given in Figure 8. The maps were created using a 10-20-1 MLP neural network, which was trained with a data set expanded from 46 to 1058 deposit patterns through the addition of $\pm 40\%$ and $\pm 80\%$ uniform random noise. The effect of training the network with patterns created with $\pm 80\%$ rather than $\pm 40\%$ noise is to reduce the size of areas estimated to have high prospectivity. A similar difference exists between maps produced with 5% and 10% normally distributed noise.

Experiment 3. Type of Noise

The experiments as described are repeated to check the effect of using randomly selected normally distributed rather than uniformly distributed random noise to increase the number of deposit patterns in the training set (Tables 6 and 7). Normally distributed noise does improve the average overall

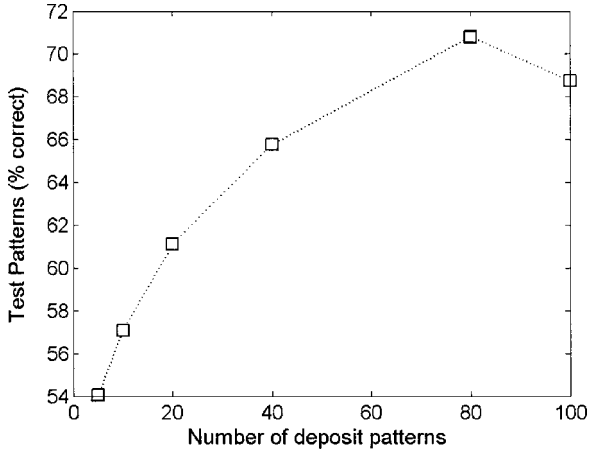

Figure 4. Classification performance for test set versus number of deposit patterns in training set.

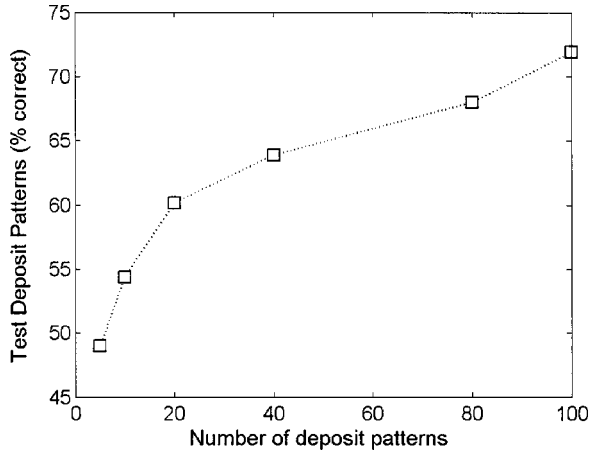

Figure 5. Classification performance for deposit patterns in test set versus number of deposits.

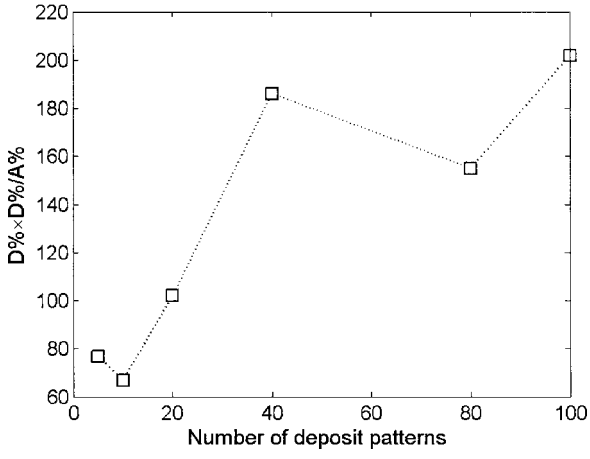

Figure 6. $D \times (D/A)$ for entire map grid versus number of deposit patterns in training set ($D =$ % of total deposits in highest prospectivity map class in 9-class prospectivity map and $A =$ % of total map area corresponding to highest class).

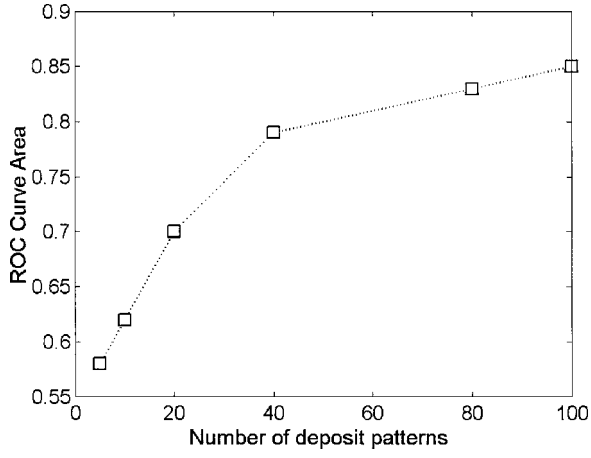

Figure 7. Area under receiver-operating-characteristic (ROC) curve versus number of deposit patterns in training set.

test-set performance as well as measures of map quality, D/A, $D \times (D/A)$, and the area under the ROC, but the percentage correct classification for test-set deposit patterns, and the correlation between probability of a deposit and increasing prospectivity map class (ρ), are generally poorer than the results obtained without using noise (Table 6). An exception is for $\pm 5\%$ noise where the correct classification rate for deposit patterns increases from 68.0% to 71.4%. Despite the fact that the *average* results for normally distributed noise generally are poorer than those for uniformly distributed noise, the results for the networks with the *best* test-set classification performance using $\pm 5\%$, $\pm 10\%$, and $\pm 20\%$ normally distributed noise are slightly better than those for uniform noise. The network with $\pm 10\%$ normally distributed noise represents the best combination of overall test-set and test-set deposit pattern classification.

Experiment 4. Augmentation of Training and Training-Stop Data Set

Although the patterns in the training data set must provide the network with a representative sample of the variation in the data population that the trained network will be required to process, the training-stop data set also determines the

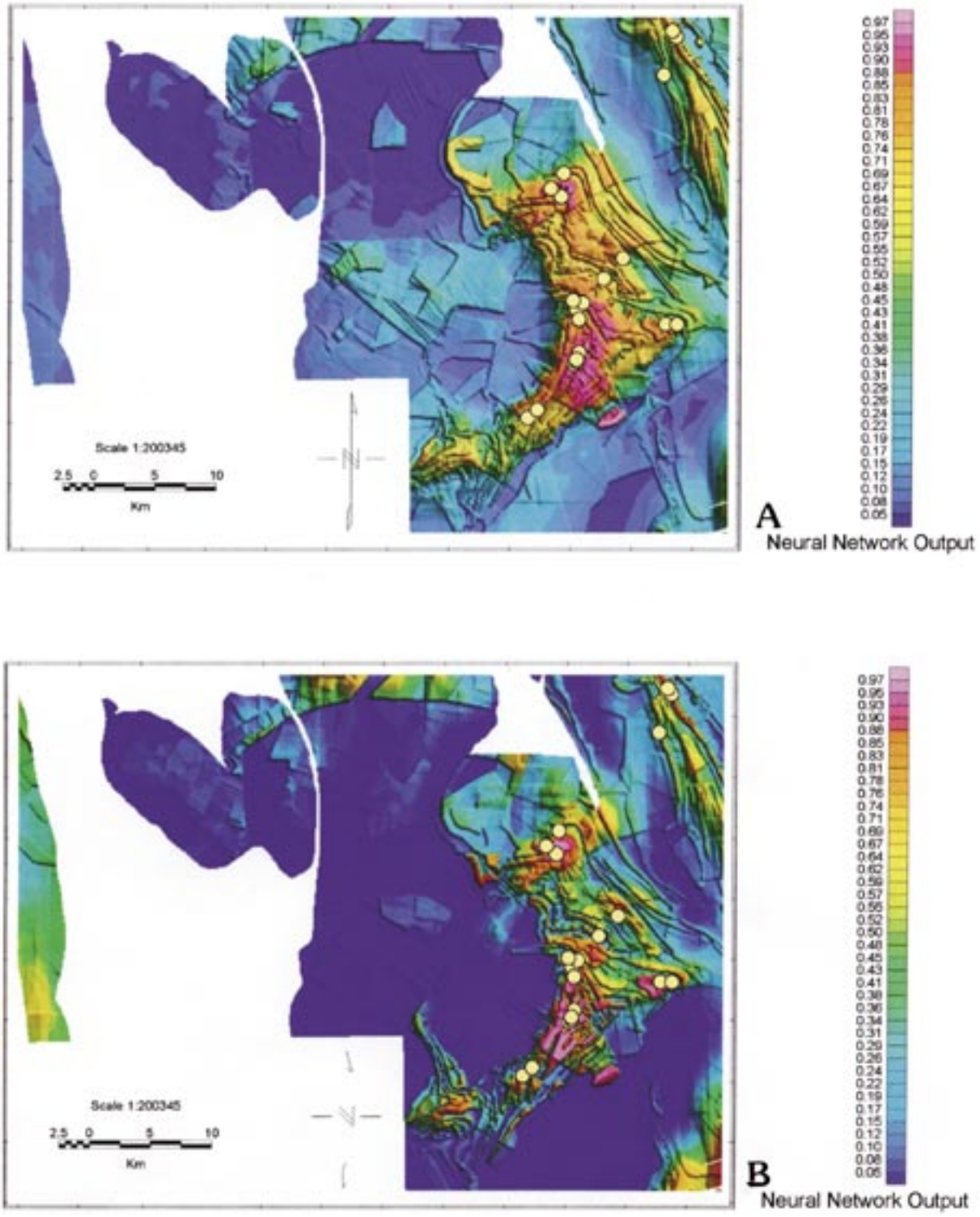

Figure 8. Mineral prospectivity maps of part of Kalgoorlie Terrane, Western Australia, showing potential for orogenic lode-gold deposits. Maps were created using 10-20-1 MLP neural network. Network was trained with training data set which was expanded from 46 to 1058 deposit patterns through addition of A, ±40% and B, ±80% uniform random noise. Effect using ±80% rather than ±40% noise is to reduce size of areas estimated to have high prospectivity. Ten input layers are selected according to evidence-categories criterion. Yellow circles indicate location of deposits containing resource of ≥1000 kg gold.

Table 4. Average Test-Set Performance (Percent Correct): Effect of the Amount of Uniformly Distributed Noise Used to Augment the Training Data Set. Results for Expanded Training and Stop Data Sets are Shown in Bold

	Training sets		Test set			Map quality statistics			
Noise max %	Train	Stop	All	Deps	Barren	ρ	D/A	D × (D/A)	ROC
0	84.4	72.7	67.9	68.0	67.8	0.97	8.2	105.3	0.79
±5	83.3	74.1	74.1	70.3	77.0	0.98	17.3	202.2	0.86
±10	83.4	74.1	72.4	71.4	73.0	0.97	15.3	178.8	0.85
±20	85.2	74.6	71.6	66.6	75.4	0.97	18.9	237.5	0.85
±40	82.5	73.2	72.8	77.1	69.6	0.99	17.9	226.4	0.87
±40	**85.9**	**77.8**	**73.5**	**77.4**	**70.4**	**0.98**	**14.7**	**296.2**	**0.87**
±80	86.5	73.9	71.5	68.9	73.5	0.98	14.6	188.2	0.84

Table 5. Best Test-Set Performance (Percent Correct): Effect of the Amount of Uniformaly Distributed Noise Used to Augment the Training Data Set. Results for Expanded Training and Stop Data Sets are Shown in Bold

	Training sets		Test set			Map quality statistics			
Noise max %	Train	Stop	All	Deps	Barren	ρ	D/A	D × (D/A)	ROC
0	88.9	72.7	76.5	77.1	76.1	0.90	9.6	256.7	0.82
±5	81.3	80.7	79.0	74.3	82.6	0.98	13.2	164.8	0.89
±10	84.0	75.0	76.5	80.0	73.9	0.98	9.7	96.6	0.86
±20	82.7	75.0	76.5	80.0	73.9	0.95	16.7	152.7	0.82
±40	82.9	75.0	76.5	85.7	69.6	1.00	14.4	216.0	0.90
±40	**82.9**	**79.8**	**76.5**	**85.7**	**69.6**	**1.00**	**14.4**	**216.0**	**0.90**
±80	92.9	79.6	80.3	74.3	84.8	0.97	11.0	247.3	0.86

Table 6. Average Test-Set Performance (Percent Correct): Effect of the Amount of Normally Distributed Random Noise Used to Augment the Training Data Set. Results for Expanded Training and Stop Data Sets are Shown in Bold

	Training sets		Test set			Map quality statistics			
Noise max %	Train	Stop	All	Deps	Barren	ρ	D/A	D × (D/A)	ROC
0	84.4	72.7	67.9	68.0	67.8	0.97	8.2	105.3	0.79
±5	83.1	72.8	72.8	71.4	73.9	0.94	12.9	143.0	0.84
±10	83.4	73.6	73.8	66.9	79.1	0.92	13.7	131.2	0.86
±10	**86.2**	**76.4**	**74.1**	**76.6**	**72.2**	**0.98**	**15.0**	**245.6**	**0.87**
±20	84.3	75.0	73.1	69.7	75.7	0.93	17.9	321.5	0.87
±40	84.2	77.1	71.2	68.6	73.3	0.93	13.1	149.8	0.84
±80	84.8	67.4	64.7	47.7	77.6	0.86	19.1	109.9	0.76

Table 7. Best Test-Set Performance (Percent Correct): Effect of the Amount of Normally Distributed Random Noise Used to Augment the Training Data Set. Results for Expanded Training and Stop Data Sets are Shown in Bold

	Training sets		Test set			Map quality statistics			
Noise max %	Train	Stop	All	Deps	Barren	ρ	D/A	D × (D/A)	ROC
0	88.9	72.7	76.5	77.1	76.1	0.90	9.6	256.7	0.82
±5	83.9	77.3	79.0	85.7	73.9	0.97	19.0	269.0	0.87
±10	87.5	71.6	80.3	80.0	80.4	0.98	12.0	130.4	0.85
±10	**87.5**	**76.7**	**80.3**	**80.0**	**80.4**	**0.98**	**12.0**	**130.4**	**0.85**
±20	88.4	78.4	77.8	82.9	73.9	1.00	13.2	439.7	0.90
±40	85.7	78.4	77.8	82.9	73.9	0.98	8.9	110.8	0.87
±80	86.8	68.2	69.1	62.9	73.9	0.97	16.3	81.6	0.76

generalization ability of the trained network. In this last experiment, the same procedure used to augment the number of deposit patterns is applied to the training-stop data set. The composition of the training-data sets is shown in Table 1. The noise levels giving the best test-set performance using uniform and normally distributed random noise are repeated with both noise-augmented training and training-stop data sets. Ten networks are trained using ±40% uniformly distributed random noise and ±10% normally distributed noise. The results are shown in Tables 4–7 in the rows marked in bold. For ±40% uniform noise, there is a small improvement in the percentage of correct classifications for the test-set (overall, deposit and barren patterns) and in the capture-efficiency statistic, $D \times (D/A)$, for the prospectivity map. In the case of the normally distributed noise, there is a small increase in the overall test-set performance and correlation coefficient (ρ), and a large increase in the classification performance for deposit patterns and both capture-efficiency statistics D/A and $D \times (D/A)$. There is no increase in the best test-set results for either uniform or normally distributed noise.

CONCLUSIONS

The use of random noise to create additional synthetic deposit-patterns helps overcome the lack of deposit patterns that limits the training set sizes for neural networks trained using a backpropagation training algorithm. Training with the larger noise-augmented data sets results in significantly increased classification accuracy for the independent test set (overall) and test-set deposit patterns, area under the ROC curve and capture efficiency (D/A and $D \times (D/A)$) statistics. The best overall and deposit-pattern test-set results are obtained with an MLP network with a 10-20-1 topology, which is trained using ±80% and ±40% uniform random noises, respectively. Uniform random noise generally gives better results than normally distributed noise. Increasing the size of the training-stop data set yields a small increase in average overall test-set and test-set deposit pattern classification performance. The method described here could be applied to increase the number of patterns available for training in other applications in which training data are rare, difficult or expensive to obtain.

ACKNOWLEDGMENTS

The principal author would like to thank Scott Halley at Placer Dome (formerly Goldfields Exploration Pty Ltd) for kindly providing the GIS data used in this study. Don Singer is thanked for reviewing the manuscript.

REFERENCES

An, G., 1996, The effects of adding noise during backpropagation training on a generalization performance: Neural Computation, v. 8, no. 3, p. 643–647.

Bishop, C., 1993, Neural networks for pattern recognition: Oxford Univ. Press, Oxford, 482 p.

Brown, W. M., 2002, Artificial neural networks: a new method of mineral prospectivity mapping: unpubl. doctoral dissertation, Univ. Western Australia (Perth), 760 p.

Brown, W. M., Groves, D. I., and Gedeon, T. D., 2002, Use of fuzzy membership input layers to combine subjective geological knowledge and empirical data in a neural network method for mineral potential mapping: Natural Resources Research, v. 12, no. 3, in press.

Brown, W. M., Baddeley, A., Gedeon, T. D., and Groves, D. I., 2002, Bivariate J-function and other graphical statistical methods help select the best predictor variables as inputs for a neural network method of mineral prospectivity mapping, *in* Bayer, U., and Burger, H., and Skala, W., eds., IAMG 2002: 8th Ann. Conf. Intern. Assoc. Mathematical Geology (Berlin), v. 1, p. 263–268.

Brown, W. M., Gedeon, T. D., Groves, D. I., and Barnes, R. G., 2000, Artificial neural networks: a new method for mineral prospectivity mapping: Australian Jour. Earth Sciences, v. 47, no. 4, p. 757–770.

Brown, W. M., Taylor, G. R. T., Jusmady, Groves, D. I., and Knox-Robinson, C. M., 1997, Application of artifical neural networks to prospectivity analysis in a GIS environment: a comparison with statistical and fuzzy logic methods for Au and Sn deposits of the Tenterfield area, NSW (abst.), *in* 14th Australian Geol. Conv. Abstracts, Geol. Soc. Australia, v. 49, p. 57.

Clay, R. D., and Sequin, C. H., 1992, Fault tolerance training improves generalization and robustness, *in* Intern. Joint Conf. Neural Network (IJCNN 1992), IEEE, New York, v. 4, p. 769–774.

Dayhoff, J. E., 1990, Neural network architectures: an introduction: Van Nostrand Reinhold, New York, 259 p.

Groves, D. I., Goldfarb, R. J., Knox-Robinson, C., Ojala, J., Gardoll, S., Yun, G., and Holyland, P., 2000, Late-kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn Block, Western Australia: Ore Geology Reviews, v. 17, no. 1, p. 1–38.

Groves, D. I., Ojala, J., and Holyland, P., 1997, Use of geometric parameters of greenstone belts in conceptual exploration for orogenic lode-gold deposits: AGSO Record 1997/41, p. 103–108.

Hagan, M. T., and Menhaj, M., 1994, Training feedforward networks with the Marquardt algorithm: IEEE Trans. Neural Networks, v. 5, no. 6, p. 989–993.

Harris, D., and Pan, G., 1999, Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis: Natural Resources Research, v. 8, no. 2, p. 93–109.

Holmstrom, L., and Koistinen, P., 1992, Using additive noise in back-propagation training: IEEE Trans. Neural Networks, v. 3, no. 1, p. 24–38.

Krogh, A., and Hertz, J. A., 1992, Generalization in a linear perceptron in the presence of noise: Jour. Physics A (Mathematical & General). v. 25, no. 5, p. 1135–47.

Masters, T., 1993, Practical neural network recipes in C++: Academic Press Inc., San Diego, California, 493 p.

Matsuoka, K., 1992, Noise injection into inputs in back-propagation learning: IEEE Trans. Systems, Man and Cybernetics, v. 22, no. 3, p. 436–440.

Parikh, C. R., Pont, M. J., and Jones, N. B., 1999, Improving the performance of multi-layer perceptrons where limited training data are available for some classes, *in* ICANN 99, Ninth Intern. Conf. Artificial Neural Networks, Conf. Publ. No. 470, v. 1, p. 227–232.

Reed, R., Marks II, R. J., and Oh, S., 1992, An equivalence between sigmoidal gain scaling and training with noisy (jittered) input data, *in* RNNS/IEEE Symp. Neuroinformatics and Neurocomputers: IEEE, New York, v. 1, p. 120–127.

Reed, R., Marks II, R. J., and Oh, S., 1995, Similarities of error regularization, sigmoid gain scaling, target smoothing, and training with jitter: IEEE Trans. Neural Networks, v. 6, no. 3, p. 529–538.

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., eds., Parallel Data Processing, M.I.T. Press, v. 1, Cambridge, Massachusetts, p. 318–362.

Sietsma, J., and Dow, R. J. F., 1991, Creating artificial neural networks that generalize: Neural Networks, v. 4, no. 1, p. 67–79.

Uncini, A., Marchesi, M., Orlandi, G., and Piazza, F., 1990, Improved evoked potential estimation using neural networks: IEEE Conf. Neural Networks (San Diego), p. II, 143–148.

Wang, C., and Principe, J. C., 1999, Training neural networks with additive noise in the desired signal: IEEE Trans. Neural Networks, v. 10, no. 6, p. 1511–1517.

Wyborn, L. A. I., Gallagher, R., and Raymond, O., 1995, Using GIS for mineral potential evaluation in areas with few known mineral occurrences, *in* Second National Forum on GIS in the Geosciences—Forum Proc.: Australian Geol. Survey Organisation Record 1995/46, p. 199–211.

Zaknich, A., 2003, Neural networks for intelligent signal processing: World Scientific, Singapore, 508 p.