Use of a Neural Network to Integrate Geoscience Information in the Classification of Mineral Deposits and Occurrences

Singer, D.A.[1] and Kouda, R.[2]

1. U.S. Geological Survey, Menlo Park, California, U.S.A.
2. Geological Survey of Japan, Tsukuba, Ibaraki-ken, Japan

ABSTRACT

A test of the ability of a probabilistic neural network to classify deposits into types based on a simple representation of mineralogy and six broad rock types is conducted here. The purpose is to examine whether this kind of system might serve as a basis for integrating geoscience information available in large mineral databases to classify sites by deposit type. Benefits of proper classification of many sites in large regions are identification of terranes permissive for deposit types and recognition that a few specific sites might be worth exploring extensively.

Probabilistic neural networks can provide mathematically sound confidence measures based on Bayes theorem and are relatively insensitive to outliers. Founded on Parzen density estimation, they require no assumptions about distributions of random variables used for classification, even handling multimodal distributions. They train quickly and work as well as, or better than, multiple-layer feedforward networks. Tests were performed with a probabilistic neural network employing a Gaussian kernel and separate sigma weights for each class and each variable.

Ore and alteration mineralogy and six rock types in 28 well-typed deposits were used to train the network. To reduce the number of minerals considered, analyzed data were restricted to minerals present in at least 50% of at least one deposit type. The training set was reduced to the presence or absence of 58 reported minerals and six generalized rock types from a total of 1005 deposits.

Two kinds of independent tests are performed with 2751 deposits and occurrences from Nevada, U.S.A., that were not used in the training set. The first test is a deposit-type by deposit-type comparison of the neural network's classification of 989 deposits with that of experts. Overall, the 53% agreement between the experts and the neural network is quite low compared to the 98% success reported in other studies.

In the other kind of test, deposit types identified by the neural network are grouped and plotted into terranes determined by experts to be permissive for the grouped deposit types. Comparison of the spatial distribution of the neural network's estimated deposit classes and permissive tracts determined by experts show that the probabilistic neural network is able to perform well at generalization. Classifying correctly over 98% of the sites in a large mineral database into the broad pluton-related and epithermal classes suggests that the probabilistic neural network can efficiently identify terranes permissive for grouped deposit classes.

INTRODUCTION

In exploration or mineral–resource assessments of large areas, the fundamental problem is to integrate information about geology, geochemistry, geophysics, and exploration history, as well as the known deposits and occurrences. The purpose of this integration is to determine what kinds of deposits might exist in some parcel of land or to determine the probability that some specific deposit type exists at some specific site. The problem can be reformed into distinguishing whether mineralized rocks of some kind could exist in some parcel of land or determining what type of mineralization does exist at a specific site. Thus in the first situation, the determination is one of whether there is any possibility of any deposit type existing, whereas in the second situation it is to determine the probability of a specific type of mineralization, given that mineralization already is known to exist. There is a parallel distinction of these two situations in that the first typically must deal with making

In "Proceedings of Exploration 97: Fourth Decennial International Conference on Mineral Exploration" *edited by* A.G. Gubins, 1997, p. 127–134

Table 1: ***Minerals used in the training and validation data.***

actinolite	cerargyrite	garnet	marcasite	silver
adularia	cerussite	goethite/limonite	molybdenite	sphalerite/wurtzite
alunite	chalcocite/digenite	gold	montmorillonite/ smectite	stibnite
anhydrite/gypsum	chlorite	graphite/organics	muscovite/sericite	tetrahedrite/fahlore/ tennantite/freibergite
apatite	cinnabar	hematite/specularite	opal	topaz
argentite/acanthite	copper	huebnerite	pyrite	tourmaline
arsenopyrite	covellite	jarosite	pyrrhotite	wolframite
azurite/malachite	electrum	jasper	realgar/orpiment	wollastonite
barite	engarite/luzonite/ famatinite	kaolinite/illite/dickite	rhodochrosite	uranophane/autunite/ pitchblende/coffinite/ gummite/uraninite
biotite	epidote	magnetite	Sb oxides	
bismuthinite	fluorite	manganite/ psilomelane/ pyrolucite/ wad	scheelite	
bornite	galena		siderite	
cassiterite				

estimates for polygons whereas in the second, the estimates are typically for a specific site. Knowing the kinds of mineralization at discovered mineral occurrences provides information about the specific sites considered and about the types of mineralization possible in the broader geologic setting of the occurrences. Knowledge of deposit types would benefit exploration in the identification of specific sites of interest and the possibilities of associated deposit types. In addition, classification of deposit types in a region aids in the identification of permissive terranes for specific and related deposit types.

If it were possible to classify correctly a large proportion of deposits and occurrences into deposit types based on the kinds of information frequently available in the geologic literature, then perhaps a system could be built that would automatically screen large data files. In such a system, the necessary and sufficient information would exist to discriminate among deposit types. Extensions to this kind of system might serve as a basis for integrating geological, geophysical, and geochemical information for estimating and managing risk. The key issue is how should these diverse kinds of information be combined.

Barton (1986) provided estimates of the frequency of mineral occurrence by deposit type. His subjective estimates for over 150 minerals in about 80 deposit types were used by McCammon (1992) in conjunction with subjective estimates of frequencies of rock types, ages, alteration, geophysical and geochemical signatures in an attempt to classify deposits with a system called Prospector II. McCammon's test of this system (1992) resulted in 83% of 124 Alaskan deposits correctly classed.

In expert systems like Prospector II, a human expert's knowledge, in the form of qualitative principles as perceived by the expert, is encoded. Performance of these systems depends on the quality of the expert's knowledge and the care taken in the representation of that knowledge. Such expert systems are desirable where the underlying model relations or information are not known. Expert systems have difficulties where the experts are internally inconsistent or rely on inconsistent information.

Where information is available, inductive learning systems exist that can use pre-classified samples as a training set to learn the appropriate classification rule. These learning systems can be successful at classifying previously unseen samples, that is, at generalization. Examples of inductive learning systems are decision trees (Quinlan, 1986), artificial neural networks (Masters, 1995), and statistical pattern recognition (Fukunaga, 1990). Features of statistical pattern recognition such as probabilistic estimates of class membership and ability to handle contradictory examples are integral to probabilistic neural networks.

The correct classification of 98% of 267 mineral deposits into eight deposit types using mineralogy and two rock types by Singer and Kouda (1997) suggested that a probabilistic neural network might do an excellent job of integrating the diverse geoscience information. To test the ability of a probabilistic neural network to classify deposits into types based on a simple representation of mineralogy and six broad rock types is conducted here. The nature and sources of data are discussed first. Following this, probabilistic neural networks and their implementation in this study are discussed. Classification of deposits into types by the neural network is tested in the next section with data in a large database containing deposits typed by experts and with a comparison of three classes of deposit types to tracts of land delineated by experts as permissive for these types in Nevada. Finally, classification errors are examined and possible improvements identified.

TRAINING DATA

Initially, it might appear that geochemical and geophysical data would be ideal for geoscience information integration projects because they are so numerous. However, the probabilistic neural network needs many examples from each group to be classed in order to reflect the variability within each group. It is not clear that the large effort required to find and record such geochemical or geophysical information, if it exists for more than a few deposit types, would be fruitful. A more tractable problem is to try to integrate mineralogy and a few rock types as reported in large mineral databases; this is the approach taken here.

Information on the mineralogy of mineral deposits varies widely in quantity and quality. Depending on the purpose of a study and its researcher's interest, a report on a mineral deposit might contain a detailed list of alteration minerals and a mention of unnamed sulphide and sulphosalt minerals, a detailed list of ore minerals and mention of alteration in broad terms, a complete list of all minerals, or a sparse list of minerals. In some studies, the author attempted to list the relative or absolute amounts of each mineral. Unfortunately, these attempts were not common and frequently not comparable with many other reports because of different standards. Thus, it was decided to use only the presence or absence of minerals in our study.

Both ore and alteration minerals were recorded for this study. Rock-forming minerals such as varieties of quartz and feldspars (except adularia) were not recorded, even if they locally represent alteration. General statements about mineralogy such as "clays", "carbonates", or "phyllic alteration" present were ignored because multiple minerals were possible. These decisions were made to keep minerals not related to the mineral deposit type out of the analysis, to reduce the number of minerals considered, and to keep the data objective. Even with these restrictions and the exclusion of clearly single-case listings, the presence or absence of 155 minerals was recorded. Closely related minerals such as the tellurides, manganese oxide minerals, anhydrite-gypsum, and enargite-luzonite were combined to further reduce the number of minerals to 119. To further reduce the number of minerals considered, the analyzed data were restricted to minerals present in at least 50% of at least one deposit type used in the study. An advantage of this restriction is that rare occurrences of minerals cannot dominate the results. The data were reduced to the presence or absence of 58 reported minerals (Table 1) in 28 deposit types (Table 2). Deposit types are classed based on the models in U.S. Geological Survey Bulletins 1693 (Cox and Singer, 1986), 2004 (Bliss, 1992), and 1811 (Mosier and Page, 1988).

Also coded into presence or absence were six rock-type categories. The broad types, marine felsic to intermediate volcanic, marine mafic volcanic, granitoid, subareal felsic to intermediate volcanic, subareal mafic volcanic, and carbonate rocks are intended to capture information about many geologic settings. Rocks not represented would be coded as the absence of all six types. These rock types and the 58 minerals were coded for each of the 28 deposit types (Table 2) to train the neural network to recognize the different deposit types. The number of deposits available for training varied by deposit type. Some deposit types, like tungsten veins, only had 12 deposits for training due to the difficulty of obtaining data. For other types, such as porphyry copper, the authors limited the training to 75 deposits to reduce the size of the problem for the neural network. Some types varied so little, such as rhyolite-hosted tin, that only 50 deposits were used for training. A total of 1005 deposits were used in the training of the neural network.

PROBABILISTIC NEURAL NETWORK

The goal here is to be able to make an estimate of the probability that an unknown mineral deposit belongs to a given deposit type. Standard statistical classification methods assume some knowledge of the distribution of the variables used to classify. Typically a multivariate normal distribution is assumed and the training data are used to estimate the means and variances. Large deviations from normality or multimodal distributions cause these methods to fail. Neural networks can typically handle complex distributions. The three-layer feedforward network (Singer and Kouda, 1996) is an excellent classifier (Masters, 1995); however, it trains slowly and does not produce probabilities.

Table 2: ***Deposit types and number of deposits used in training.***

Deposit type	number of deposits
Bedded barite	19
Besshi massive sulphide	12
Cu skarn	16
Cyprus massive sulphide	49
Distal disseminated Ag–Au	10
Epithermal Mn	21
Epithermal quartz alunite Au–Ag	32
Epithermal quartz-adularia Au–Ag	75
Fe skarn	50
Franciscan Mn	50
Hot–spring Au–Ag	12
Hot–spring Hg	39
Kuroko massive sulphide	50
Low–sulphide Au–quartz vein	75
Polymetallic replacement	25
Polymetallic veins	48
Porphyry Cu	75
Porphyry Mo, low–F	25
Replacement Mn	21
Rhyolite–hosted Sn	50
Sediment–hosted Au	27
Sedimentary exhalative Zn–Pb	36
Silica–carbonate Hg	50
Simple Sb	50
Volcanic–hosted magnetite	22
W skarn	28
W veins	12
Zn-Pb skarn	26
TOTAL	1005

Probabilistic neural networks were designed to be classifiers. If we know the true probability density function, $f_k(x)$, for all populations, then there is a Bayes optimal decision rule for classifying unknown sample x into population *i*:

$$p_i \, c_i \, f_i(x) > p_j \, c_j \, f_j(x) \qquad [1]$$

for all populations j not equal to i. Generalizing, p_k is the prior probability of the general class *k*, and c_k is the cost associated with misclassification of population *k*. Under these conditions, a Bayes decision rule will minimize the expected cost of misclassification. The problem is that we do not know the true probability density function, $f_k(x)$. Standard statistical classification methods, such as discriminant analysis, typically assume that the variables follow a multivariate normal distribution

Table 3: ***Confusion matrix showing the number of mineral deposits correctly (in bold) and incorrectly classified from the validation set.***

		Neural net predicted class																						
		Bedded barite	Besshi massive sulfide	Cu skarn	Cyprus massive sulfide	Distal disseminated Ag-Au	Epithermal Mn	Epith. qtz.-adularia Au-Ag	Fe skarn	Franciscan Mn	Hot-spring Hg	Hot-spring Au-Ag	Kuroko massive sulfide	Low-sulfide Au-qtz. vein	Polymetallic replacement	Polymetallic vein	Porphyry Mo, low-F	Porphyry Cu	Epith. quartz-alunite Au-Ag	Replacement Mn	Rhyolite-hosted Sn	Sed. exhalative Zn-Pb	Sediment-hosted Au	Silica-carbonate Hg
	Bedded barite	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Besshi massive sulfide	0	**0**	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Cu skarn	0	1	**13**	0	0	1	1	5	15	0	0	1	4	2	1	0	0	0	3	0	1	1	0
	Cyprus massive sulfide	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	Distal disseminated Ag-Au	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
	Epithermal Mn	1	0	0	0	0	**5**	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Expert class	*Epith. qtz.-adularia Au-Ag*	2	0	0	0	0	18	**153**	0	0	1	6	0	9	0	5	0	0	1	1	0	0	2	0
	Fe skarn	0	0	0	0	0	0	0	**9**	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0
	Franciscan Mn	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Hot-spring Hg	1	0	0	0	0	2	0	0	0	**63**	0	0	0	0	0	0	0	0	0	0	0	0	0
	Hot-spring Au-Ag	0	0	0	0	0	0	4	0	0	1	**2**	0	1	0	0	0	0	0	1	0	0	0	0
	Kuroko massive sulfide	0	0	0	0	0	0	0	1	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0
	Low-sulfide Au-qtz. vein	0	0	0	0	0	0	2	0	0	0	0	0	**2**	0	1	0	0	0	0	0	0	0	0
	Polymetallic replacement	7	0	0	0	1	0	0	1	16	1	0	0	1	**3**	51	0	0	0	7	0	8	2	0
	Polymetallic vein	4	0	0	2	0	0	7	1	6	0	0	0	17	0	**82**	18	0	0	6	0	4	3	1
	Porphyry Mo, low-F	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	**6**	1	0	0	0	0	0	0
	Porphyry Cu	0	0	1	0	0	0	0	0	1	0	0	0	0	2	1	1	**0**	0	0	0	0	0	0
	Epith. quartz-alunite Au-Ag	0	1	0	0	0	0	3	0	0	0	2	2	0	0	0	0	0	**10**	0	0	0	0	0
	Replacement Mn	1	0	0	0	0	5	0	0	2	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0
	Rhyolite-hosted Sn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0
Expert class	*Sediment-hosted Au*	0	0	0	0	0	0	0	0	0	0	0	0	11	0	1	0	0	0	0	0	0	**24**	0
	Simple Sb	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	Volcanic-hosted magnetite	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	W skarn	0	0	0	0	0	0	4	2	11	1	0	0	1	0	21	8	0	0	0	0	0	0	0
	W vein	0	0	0	0	0	0	0	0	2	0	0	0	1	0	2	0	0	0	0	0	0	0	0
	Zn-Pb skarn	0	0	0	0	0	0	0	1	1	0	0	0	0	1	8	1	0	0	1	0	0	0	0
	Total by network	17	2	14	4	1	31	184	20	61	67	10	5	49	8	175	35	1	11	23	4	15	34	1
	Percent agreement	6	0	93	0	0	16	83	45	8	94	20	20	4	38	47	17	0	91	9	100	0	71	0

or that the nearest neighbor is the appropriate class regardless of the density of other samples near the unknown.

The development by Parzen (1962) of a general way to estimate a univariate probability density function from a random sample, even when the parent density function is unknown, provides a necessary tool to free us from these constraints. Parzen's estimator is essentially a sphere-of-influence weighting function, frequently termed a kernel, and the scaling weight, σ, controls the width of the area of influence. The weighting function has its largest values at sample points and decreases toward zero as the distance increases.

For this study separate weights (σ) were used for each class and each variable and a Gaussian kernel was used for the weighting ($\mathbf{W}$) function (Masters, 1995). The choice of the Gaussian function is based on its excellent performance and has nothing to do with assumptions of normal distributions. Specht (1990) constructed a neural network form of Parzen's estimation procedure. In the present study, the algorithms for a probabilistic neural network developed by Masters (1995) were employed. Masters' algorithms find the scale weights, σ, that minimize the error of misclassification of the training data using the standard statistical technique called jackknifing in which every case is sequentially held back from training.

Probabilistic neural networks require no assumptions about distributions of random variables used to classify; they even can handle multimodal distributions. They train quickly and as well as, or better than, multiple-layer feedforward networks. They have the ability to provide mathematically sound confidence levels and are relatively insensitive to outliers. Mathematically sound Bayesian confidence levels require that the classes are mutually exclusive and exhaustive (i.e., no case can possibly fall into more than one population and the training set encompasses all populations fairly). When these conditions exist, Bayes' Theorem can be used to compute the probability that an observation, $\mathbf{x}$, was the member of a population. Each density estimate could be multiplied by prior probabilities and cost constants, if desired. These features are not used in this study, however.

In many practical situations, the mutually exclusive and exhaustive class conditions might not exist. The unknown sample used in testing

Table 3: ***Confusion matrix (continued).***

	Neural net predicted class	Simple Sb	Volcanic-hosted magnetite	W skarn	W vein	Zn-Pb skarn	**Total by experts**	**Percent agreement**
Expert class	*Bedded barite*	0	0	0	0	0	1	100
	Besshi massive sulfide	0	0	0	0	0	2	0
	Cu skarn	0	0	3	0	0	52	25
	Cyprus massive sulfide	0	0	0	0	0	1	0
	Distal disseminated Ag-Au	0	0	0	0	0	3	0
	Epithermal Mn	0	0	0	0	0	10	50
	Epith. qtz.-adularia Au-Ag	7	0	0	0	1	206	74
	Fe skarn	0	0	0	0	0	12	75
	Franciscan Mn	0	0	0	0	0	5	100
	Hot-spring Hg	3	0	0	0	0	69	91
	Hot-spring Au-Ag	0	0	0	0	0	9	22
	Kuroko massive sulfide	0	0	0	0	0	2	50
	Low-sulfide Au-qtz. vein	0	0	0	0	0	5	40
	Polymetallic replacement	1	0	0	0	0	99	3
Expert class	*Polymetallic vein*	44	1	2	4	0	202	41
	Porphyry Mo, low-F	0	0	0	0	0	9	67
	Porphyry Cu	0	0	0	0	1	7	0
	Epith. quartz-alunite Au-Ag	0	0	0	0	0	18	56
	Replacement Mn	0	0	0	0	0	10	20
	Rhyolite-hosted Sn	0	0	0	0	0	4	100
	Sediment-hosted Au	1	0	0	0	0	37	65
	Simple Sb	**54**	0	0	0	0	55	98
	Volcanic-hosted magnetite	0	**1**	0	0	0	11	9
	W skarn	0	0	**66**	8	0	122	54
	W vein	0	0	0	**16**	0	21	76
	Zn-Pb skarn	0	0	0	1	**3**	17	18
	Total by network	110	2	71	29	5	989	53
	Percent agreement	49	50	93	55	60	53	

might be from a population different from any of the training classes. For example, if the mineralogy of a carbonatite deposit were tested in the network developed in this study, Bayesian confidence estimates could not be computed properly. The neural network program will estimate the probabilities that the unknown deposit belongs to the deposit classes it has been taught; thus, careless use of a neural network could lead to mistaken classifications.

TESTING THE NEURAL NETWORK

Because of the ability of neural networks to learn the training data well, validation data, not used in any training, is the proper data set to test the efficiency of classification. A data set large enough to test the ability of a probabilistic neural network to classify many different deposit types is reported in recent papers on Nevada's resources by Sherlock *et al.* (1996) and Cox *et al.* (1996). Approximately 1400 metallic and nonmetallic mineral locations in Nevada were classed into mineral deposit types (Sherlock *et al.,* 1996) following Cox and Singer (1986). These typed locations and many untyped mineral sites are available in the Mineral Resources Data System, which contains over 6100 entries for mineral sites throughout Nevada (Sherlock and Tingley, 1985).

After removal of both nonmetallic and duplicate entries, and entries without at least one rock and one of the 58 reduced set of minerals or two of the 58 minerals, the remaining 2751 sites were prepared for testing in two different kinds of tests. One kind of test is a deposit-type by deposit-type comparison of the neural network's classification of a set of deposits with that of experts. In the other kind of test, deposit types identified by the neural network are grouped and plotted into terranes determined by experts to be permissive for the grouped deposit types and the exceptions are counted.

To make the comparison with an expert classification, the neural network was trained to recognize the 28 deposit types in Table 2 using the minerals in Table 1 and the six rock types discussed earlier. Two of the deposit types used in training, sedimentary exhalative Zn–Pb and silica–carbonate Hg, are not known to exist in Nevada but were added to the training set to make the test more realistic. The test data were the Mineral Resources Data System sites from Nevada that contained at least the minimum number of minerals and that had been classed by Sherlock *et al.* (1996) into one of the deposit types used in training the neural network.

Results of the test on the 989 mineral sites in Nevada that had been classed by the experts and had the required minerals in the data file are reported in Table 3. The table can be used to examine how well the neural network agrees with the experts for each deposit type by the number of deposits in bold and by the percent agreement in the last column. If the neural network and the experts were in complete agreement, all of the deposits would be counted in the bold positions only. For example, in the epithermal quartz–adularia Au–Ag row, 153 of the 206 deposits (or 74%) that the experts classed as epithermal quartz-adularia Au–Ag were also classed as epithermal quartz-adularia Au–Ag by the neural network (Table 3). The table also shows that 18 of the epithermal quartz–adularia Au–Ag deposits according to the experts were classed as epithermal Mn deposits by the network. The largest number of classification errors are polymetallic replacements classed by the neural network as polymetallic veins. Overall, the 53% agreement between the experts and the neural network is quite low compared to the 98% success reported by Singer and Kouda (1997).

There are several possible reasons for the relatively low success rate of correct classification reported in Table 3. Data used in the training are typically reported in scientific studies conducted on large deposits, whereas information in large regional databases are dominated by small unstudied occurrences. The 98% success rate reported in an earlier study (Singer and Kouda, 1997) was for a situation where the training and validation data were predominantly from scientific studies. Few minerals are reported and alteration minerals are rarely reported in regional mineral deposit databases. Unfortunately, this problem cannot be easily fixed because the sparse information on most occurrences also prevents them being used for training because there is not enough information to class them confidently. Another possible reason is that experts use data not in the database. A third possible reason for the low success rate is that experts use more information than the neural network was given. For example, experts frequently plot occurrences on geologic maps which provides information about geologic settings and relationships to other possible nearby mineral deposits. Related to this problem is the different nature of data for some deposit types. For example, the polymetallic replacements training data really represent districts whereas the entries used in testing represent individual sites and occurrences. Thus, the training data can represent minerals from many nearby sites whereas the test data represent only one site. With some care in design, it might be possible to incorporate this kind of information in a neural network. A fourth possible reason for the low success rate is that the experts might have made mistakes and the neural network correctly classed deposits. This reason probably only accounts for some of the neural network's apparent misclassifications.

For comparison of the distribution of estimated deposit classes and permissive tracts, the same training set used in the expert comparison was used. The test data consists of all 2751 entries available, including the sites typed by experts. Deposits classed as the following types are considered epithermal: epithermal Mn, epithermal quartz–alunite Au–Ag, epithermal quartz–adularia Au–Ag, hot-spring Au–Ag, hot-spring Hg, rhyolite-hosted Sn, and silica–carbonate Hg deposits. Deposits classed as the following types are considered pluton-related: Cu skarn, distal disseminated Ag–Au, Fe skarn, polymetallic replacement, polymetallic veins, porphyry Cu, porphyry Mo, low–F, replacement Mn, W skarn, W veins, and Zn–Pb skarn deposits.

Most deposits not within a tract delineated as permissive for pluton-related deposits by Cox *et al.* (1996), such as the deposits in southeast Nevada (Figure 1), are correctly classed as replacement Mn or W vein deposits. However, these were not delineated by Cox *et al.* (1996) because they were either not economically important or were associated with Proterozoic plutons. A few deposits apparently outside delineated tracts, such as the one in northeast Nevada, are within tracts too small to be seen at the scale of the figures. When these apparent errors are properly counted, the probabilistic neural network successfully classified 99% of the 907 deposits and occurrences grouped as pluton-related in Cox *et al.* (1996).

The neural network put 112 deposits and (or) occurrences into the sediment-hosted gold (also known as Carlin-type) class. About 96% of these deposits plot (Figure 2) within the tracts designated permissive for sediment-hosted gold by Cox *et al.* (1996). The majority of deposits classed as sediment-hosted gold in southwestern Nevada have not generally been recognized as Carlin-like. Given that the information available to the neural network for these deposits was typically gold in carbonate rocks, the classification makes sense.

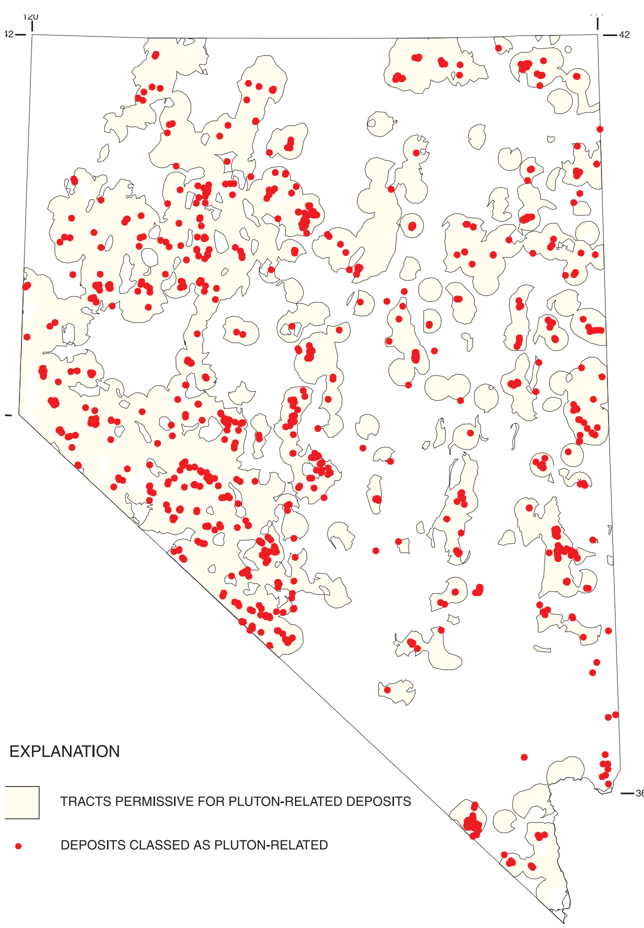

Figure 1: *Tracts permissive for pluton related deposits with deposits classed as pluton-related.*

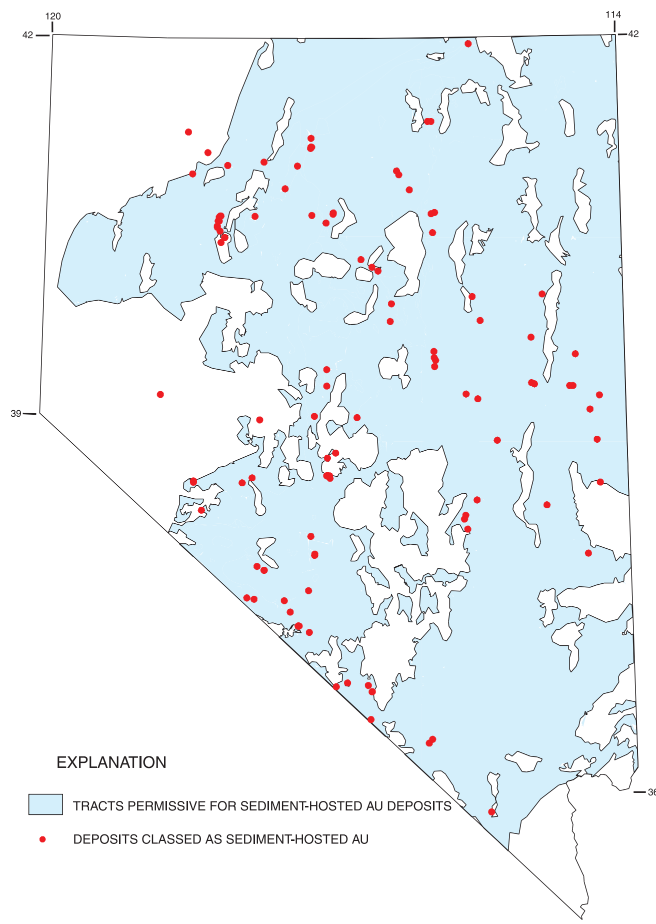

Figure 2: *Tracts permissive for sedmiment-hosted Au deposits with deposits classed as sediment hosted (Carlin type)*

Over 98% of the 825 deposits classed as epithermal by the neural network fall within the epithermal permissive tracts (Figure 3) of Cox *et al.* (1996). At least one of the occurrences plotted outside permissive tracts is clearly epithermal and must either have the wrong location in the database or must be misclassified by the experts.

Based on these tests on both pluton-related deposits and epithermal deposits, the probabilistic neural network is able is perform well at generalization. The neural network's error rates are probably no worse, and may be better, than experts' for classifying such large data sets into the broad classes.

CONCLUSIONS

From previous studies we know that it is possible to classify correctly a large proportion of deposits and occurrences into eight deposit types based on the kinds of information frequently available in the scientific literature. Here we examine whether this kind of system might serve as a basis for integrating geoscience information available in large mineral databases to classify these sites by deposit type. In well–explored regions, a large proportion of such sites are occurrences. The benefits of proper classification of many sites in these regions are the identification of terranes permissive for deposit types, and the recognition that a few specific sites might be worth exploring extensively.

Comparison of the spatial distribution of the neural network's estimated deposit classes and permissive tracts determined by experts shows that the probabilistic neural network is able is perform well at generalization. Classifying correctly over 98% of the sites in a large mineral database into the broad pluton-related and epithermal classes suggests that the probabilistic neural network can efficiently identify terranes permissive for grouped deposit classes.

The 53% agreement between the experts and the neural network on specific deposit types is low compared to the 98% success reported by Singer and Kouda (1997). The primary reason for the different outcomes lies in the differences between economic deposits and mineral occurrences. It is the economic deposits that receive detailed study and, therefore, we feel confident in classifying. Typically, few minerals are reported in and alteration minerals rarely are reported in regional mineral deposit databases that are dominated by non–economic occurrences. Another reason for the low success rate is that experts use more information than the neural network was given, such as geologic settings and relationships to other possible nearby mineral deposits, as well as personal knowledge. Related to this problem is that for some deposit types, training data really represent districts whereas entries used in testing represent individual sites and occurrences. Thus, the training data can represent minerals from many nearby sites and the test data only one site. With some care in design, however, it should be possible to incorporate many of these kinds of information in neural networks.

ACKNOWLEDGEMENTS

Nichelle Waller and Dan Mosier gathered much of the data used to train the neural network to recognize deposit types. Jim Bliss provided the mineral data on polymetallic vein and low–sulphide Au–qtz. vein deposits. Barry Moring plotted the figures.

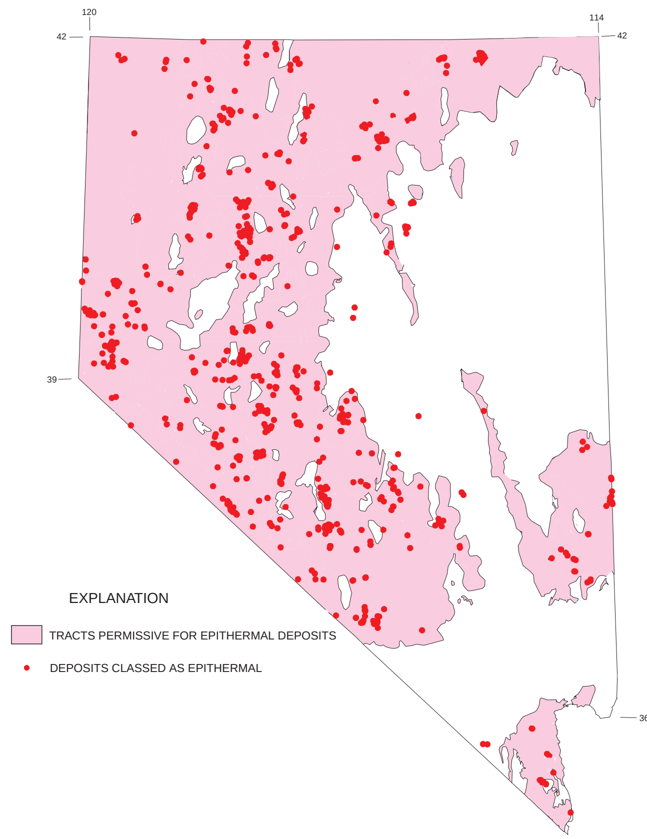

Figure 3: *Tracts permissive for epithermal deposits with deposits classed as epithermal.*

REFERENCES

Barton, P.B., 1986, Mineralogical index, *in* Cox, D.P., and Singer, D.A., eds., Mineral deposit models: U.S. Geological Survey Bulletin 1693, 318–348.

Bliss, J.D., ed., 1992, Developments in mineral deposit modeling: U. S. Geological Survey Bulletin 2004, 168.

Cox, D. P., Berger, B.R., Ludington, S., Moring, B.C., Sherlock, M.G., Singer, D.A., and Tingley, J.V., 1996, Delineation of mineral resource assessment tracts and estimation of number of undiscovered deposits in Nevada, *in* Singer, D.A., ed., An analysis of Nevada's metal–bearing mineral resources: Nevada Bureau of Mines and Geology Open-File Report 96-2, Chapter 12, p. 12.1-12.25, three sheets, scale 1:1 000 000. (Also available at: www.nbmg.unr.edu/ofr962.)

Cox, D.P., and Singer, D.A., eds., 1986, Mineral deposit models: U.S. Geological Survey Bulletin 1693, 379.

Fukunaga, K., 1990, Introduction to statistical pattern recognition, (2nd ed.): Academic Press.

Masters, T., 1995, Advanced algorithms for neural networks: a C++ sourcebook: John Wiley & Sons.

McCammon, R.B., 1992, Numerical mineral deposit models, *in* Bliss, J.D., ed., Developments in deposit modeling: U.S. Geological Survey Bulletin 2004, 6-12.

Mosier, D.L., and Page, N.J, 1988, Descriptive and grade-tonnage models of volcanogenic manganese deposits in oceanic environments—a modification: U. S. Geological Survey Bulletin 1811, 28.

Parzen, E., 1962, On estimation of a probability density function and mode: Annals of Mathematical Statistics, **33**, 1,065–1,076.

Quinlan, J.R., 1986, Induction of decision trees: Machine Learning, **1**, 81–106.

Sherlock, M.G., Cox, D.P., and Huber, D.F., 1996, Known mineral deposits and occurrences in Nevada, *in* Singer, D.A., ed., An analysis of Nevada's metal–bearing mineral resources: Nevada Bureau of Mines and Geology Open–File Report 96–2, Chapter 10, p. 10.1–10.38, 2 sheets, scale 1:1 000 000. (Also available at: www.nbmg.unr.edu/ofr962.)

Sherlock, M.G., and Tingley, J.V., 1985, Nevada mineral-resource data; Information available through the U.S. Geological Survey Mineral Resources Data System: U.S. Geological Survey Circular 966, 35.

Singer, D.A., and Kouda, R., 1996, Application of a feedforward neural network in the search for kuroko deposits in the Hokuroku District, Japan: Mathematical Geology, **28**, 1,017–1,023.

Singer, D.A., and Kouda, R., 1997, Classification of mineral deposits into types using mineralogy with a probabilistic neural network: Nonrenewable Resources, **6**, 27-32.

Specht, D., 1990, Probabilistic neural networks: Neural Networks, **3**, 109–118.