

Application of Discriminant Analysis and Support Vector Machine in Mapping Gold Potential Areas for Further Drilling in the Sari-Gunay Gold Deposit, NW Iran

Hamid Geranian,¹ Seyed Hassan Tabatabaei,^{1,4} Hooshang H. Asadi,^{1,2}
and Emmanuel John M. Carranza³

Received 11 March 2015; accepted 27 May 2015
Published online: 12 June 2015

In this contribution, we used discriminant analysis (DA) and support vector machine (SVM) to model subsurface gold mineralization by using a combination of the surface soil geochemical anomalies and earlier bore data for further drilling at the Sari-Gunay gold deposit, NW Iran. Seventy percent of the data were used as the training data and the remaining 30 % were used as the testing data. Sum of the block grades, obtained by kriging, above the cutoff grade (0.5 g/t) was multiplied by the thickness of the blocks and used as productivity index (PI). Then, the PI variable was classified into three classes of background, medium, and high by using fractal method. Four classification functions of SVM and DA methods were calculated by the training soil geochemical data. Also, by using all the geochemical data and classification functions, the general extension of the gold mineralized zones was predicted. The mineral prediction models at the Sari-Gunay hill were used to locate high and moderate potential areas for further infill systematic and reconnaissance drilling, respectively. These models at Agh-Dagh hill and the area between Sari-Gunay and Agh-Dagh hills were used to define the moderate and high potential areas for further reconnaissance drilling. The results showed that the nu-SVM method with 73.8 % accuracy and c-SVM with 72.3 % accuracy worked better than DA methods.

KEY WORDS: Classification, Discriminant analysis, Support vector machine, Gold potential mapping, Sari-Gunay deposit.

INTRODUCTION

A key element in an exploration project is evaluation of the results in different exploration phases. This is important to obtain scientific and

economic implications to move to the next steps of exploration. This problem is more obvious when the volumes of exploration operations are increased and the quality levels of the explorations are different. Comparative, integration, and modeling methods have been used to solve this problem (Bonham-Carter et al. 1988; Venkataraman et al. 2000; Moon et al. 2006). Classification methods are important in exploration data modeling methodologies. In exploration data modeling methodologies, deposits are divided into training and testing classes. Using the training class, a model is built on the basis of exploration datasets. Then, based on this model and the exploration data of the testing class, probabilistic

¹Department of Mining Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran.

²Centre for Exploration Targeting, Australian Research Council Centre of Excellence for Core to Crust Fluid Systems (CCFS), School of Earth and Environment, The University of Western Australia, Crawley WA 6009, Australia.

³Department of Earth and Oceans, James Cook University, Townsville QLD, Australia.

⁴To whom correspondence should be addressed; e-mail: tabatabaei@cc.iut.ac.ir

results of the model are estimated. In case of positive results in estimation, extension of exploration is proposed. Classification methods are divided into two categories (Duda et al. 2000; Theodoridis and Koutroumbas 2009): (i) for data with specified probability density function and (ii) for data without specified probability density. In this research, discriminant analysis (DA) representing the first category and support vector machine (SVM) representing the second category are used.

The application of DA in mineral exploration was first introduced by Harris (1965) to evaluate the role of geological variables in determining the probability of mineralization. This method has also been used to identify geochemical anomalies (Rose 1972; Whitehead and Govett 1974), to identify exploration criteria to separate ore-bearing rocks from barren rocks (Divi et al. 1979; Fedikow et al. 1991), to separate rock units and minerals (McKinley et al. 2014; Tibljas et al. 2002; Varadanchari and Mukherjee 2004), and to map hydrothermal alterations (Tahmasebi et al. 2010). Roshani et al. (2013) used the DA method to correlate surface geochemical anomalies with borehole data in the Kuh-Panj porphyry copper deposit, SE Iran. Belkhiri and Mouni (2014) used the DA for geochemical characterization of surface water and groundwater resources. Modeling of mineral potential using DA has been demonstrated by Agterberg (1974), Chung (1977), Prelat (1977), Bonham-Carter and Chung (1983), Harris and Pan (1999), Pan and Harris (2000), and Harris et al. (2003). Carranza (2009) has demonstrated a GIS-based technique for spatial evidence representation in modeling of mineral potential through DA.

The SVM was first proposed for pattern recognition by Vapnik in 1982. Then, it was used by Li (2005) to identify optimum locations for drilling the oil and gas fields in China. This method was also used for classification, regression and data ranking (Yu and Kim 2012), lithologic classification (Al-Anazi and Gates 2010; Yu et al. 2012), geochemical fingerprinting of ores (Savu-Krohn et al. 2011), identification of hydrothermal alteration associated with mineralization (Abbaszadeh et al. 2013), mineral resource estimation (Chatterjee and Bandopadhyay 2011), and modeling of mineral potential (Wu et al. 2010; Zuo and Carranza 2011; Abedi et al. 2012; Rodriguez-Galiano et al. 2014).

With regard to modeling of mineral potential, the performance of DA has been compared and contrasted with classification methods suitable for

data with specified probability density [e.g., logistic regression (LR)] and with classification methods suitable for data without specified probability density [e.g., artificial neural networks (ANN)] (cf. Agterberg 1974; Harris and Pan 1999; Harris et al. 2003). A common conclusion of these comparative studies is that, with regard to modeling of mineral potential, classification methods suitable for data without specified probability density (e.g., ANN) generally outperform classification methods suitable for data with specified probability density (e.g., DA, LR), and that LR generally outperforms DA. In this study, we compare and contrast DA with SVM, which has been around since just a decade or so ago. Both DA and SVM make use of a hyperplane for classification (see below), and this commonality is the basis for comparing their performance in this study.

The study area (20 km²) covers the Sari-Gunay and Agh-Dagh hills at the Sari-Gunay epithermal gold deposit, NW Iran. The exploration, including extensive drilling, has been completed at the Sari-Gunay hill, but the Agh-Dagh hill and its surrounding areas still need more drilling. In this research, DA and SVM are applied to the soil geochemical and borehole data to obtain algorithms for mapping potential gold mineralization. The final potential map can be used for further drilling in the Sari-Gunay exploration area.

METHODS

Classification methods can be divided into two types, one in which the probability distribution of data must be specified and the other without the need to know the probability distribution of data. The aim of the first type is to minimize empirical risk, while the second type is used to minimize structural risk (Cristianini and Taylor 2000). In empirical risk minimization, the output is a model with a minimum error in the training data, whereas in the structural risk minimization, the output is a model with generalization properties that has a minimum error in both the training and the testing data (Theodoridis and Koutroumbas 2009).

The DA and LR methods belong to the first type, and the SVM and ANN methods are attributed to the second type. The DA method has been chosen because, compared to LR, it is less sensitive to the number of samples and the size of training set (Davis 2002). It has been shown that when sample sizes are equal and homogeneity of variance/covariance holds,

DA is more accurate than LR (Hu and Yu 2011). Discriminant analysis has more assumptions and restrictions than LR (Davis 2002), but when the assumptions of DA are met it is more powerful than LR (e.g., Bökeoğlu Çokluk and Büyüköztürk 2008; Cohen et al. 2002; Hu and Yu 2011). The SVM method was also selected because of its four main advantages (Auria and Moro 2008; Byvatov et al. 2003; Shawe-Taylor and Cristianini 2004): (1) it has a regularization parameter, which makes the user think about avoiding over-fitting; (2) it uses the kernel trick, so it can build an expert knowledge about the problem via the kernel; (3) SVM is defined by a convex optimization problem (no local minima), which makes it more efficient than ANN; and (4) it is an approximation bound on the test error rate, and there is a substantial body of theory behind it that suggests to choose SVM over other methods.

Discriminant Analysis

DA obtains a hyperplane that describes the separation between the observed groups. It allows classification of new observations into one of the known and predefined groups. This method is based on the mean and pooled matrix of the data. To perform DA, two conditions are required, namely (Davis 2002; McLachlan 1992): (i) all variables must have a normal distribution; and (ii) the matrix of pooled variance–covariance for all groups should be homogeneous. If these two conditions are satisfied, DA with a linear function can be used, whereas if the first condition is satisfied but the second condition is not, DA with quadratic function can be used instead of a linear function (Croux and Joossens 2005; Friedman 1989). If, in a new observation point, the p variables are measured, then

$$x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})^T \quad i = 1, 2, \dots, n, \quad (1)$$

where x_i is a column array of the data and n is the number of samples. To classify the new observation point between K classes that are separated by a hyperplane, computation of the classification score of that point in each class is required. The value of the score is obtained from the following equation:

$$cf(x_i) = (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) + \ln |\Sigma_k| - 2 \log \pi_k \quad (2)$$

In this equation, μ_k is the group k mean vector ($1 \leq k \leq K$), Σ_k^{-1} is the inverse covariance matrix of

group k , $|\Sigma_k|$ is the covariance matrix determinant, and π_k is the initial probability of group k . Therefore, the new observation point belongs to the group that has the lowest classification score. This equation is called quadratic function discriminant analysis (QDA) (Wu et al. 1996).

If the covariance matrix of the groups is homogeneous in Eq. (2), then the matrix of pooled covariance between groups can be used. This pooled matrix is calculated from the following equation:

$$\Sigma = \frac{1}{N} \sum_{k=1}^K n_k \Sigma_k \quad (3)$$

By placing the matrix in Eq. (2), we will have

$$\begin{aligned} cf(x_i) &= (x_i - \mu_k)^T \Sigma^{-1} (x_i - \mu_k) - 2 \log \pi_k \\ &= 2 \mu_k^T \Sigma^{-1} x_i - \mu_k^T \Sigma^{-1} \mu_k - 2 \log \pi_k \end{aligned} \quad (4)$$

This equation is called linear function discriminant analysis (LDA) (George and Fernandez 2002; Wu et al. 1996).

In DA, three aspects are considered for function validation (Srivastava 2002): (1) significant differences between the classes; (2) proper selection of significant variables in the discriminant function; and (3) estimation of the probability of misclassification.

Support Vector Machine

SVM is a tool based on the statistical learning theory (Vapnik 1999). In this method, the input vectors are mapped by kernel functions into a multi-dimensional space with larger dimension than the original dimension. Then, a hyperplane is created that separates the input vectors at the maximum possible distance. This hyperplane is called the hyperplane with the maximum separating boundary. Two other parallel hyperplanes with no data between them exist on both sides of this hyperplane (Fig. 1). As the distance between two parallel hyperplanes is increased, the classification error is decreased (Ben-Hur and Weston 2010; Burges 1998).

Suppose the training data include N data pairs in Eq. (5); therefore, the independent variable x_i is classified in two classes of $y_i = 1$ and $y_i = -1$ (Kavzoglu and Colkesen 2009).

$$D = \{(x_i, y_i) | x_i \in R^n, y_i \in \{-1, 1\}\}_{i=1}^N \quad (5)$$

Then, the separating hyperplane equation of the two classes is

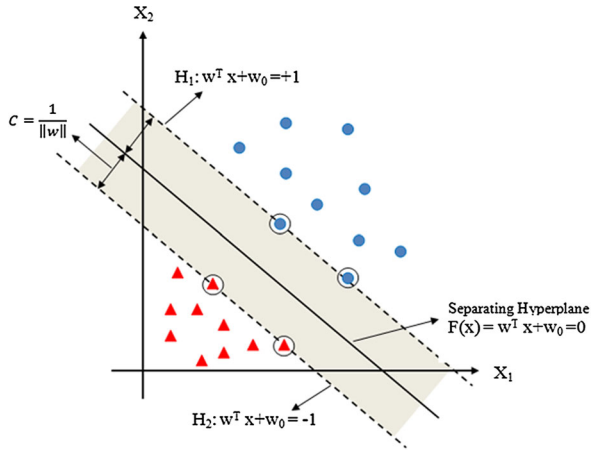


Figure 1. A separating hyperplane with a maximum possible margin to distinguish classes in a dataset in a linear case. The support vectors, the data fall in the two hyperplanes (dash lines) parallel to the separating hyperplane, are encircled.

$$\{x \in R | f(x) = w \cdot x + b = 0\}, \quad w \in R, \quad b \in R, \quad (6)$$

where w is the vector of hyperplane coefficient and b is the vector representing the hyperplane distance away from the origin of coordinates. Selection of unique solutions for these two parameters would lead to a maximum distance between the two parallel hyperplanes and increase in generalization of the separating hyperplane, and therefore, the optimization is

$$\min_{w,b} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^N \xi_i \quad (7)$$

subject to the following constraints:

$$y_i(\varphi(x_i) \cdot w + b) \geq (1 - \xi_i) \forall i, \quad \xi_i \geq 0, \quad (8)$$

where $\varphi(x_i)$ is the kernel function, c is the fixed value of capacity, and ξ_i is the auxiliary parameter to separate overlapping data. Larger values of c represent a further consideration to the points close to the decision boundary, and smaller values that are points away from the decision boundary are optimized in solving the problem. This SVM method is called c-SVM approach. But in nu-SVM approach, the optimization is

$$\min_{w,b} \frac{1}{2} \|w\|^2 - nu \times \rho + \frac{1}{N} \sum_{i=1}^N \xi_i \quad (9)$$

subject to the following constraints:

$$y_i(\varphi(x_i) \cdot w + b) \geq \rho - \xi_i \forall i, \quad \xi_i \geq 0, \quad \rho \geq 0, \quad (10)$$

where ρ is half of the distance between the two parallel hyperplanes and nu is the control parameter

Table 1. Common Kernel Functions in SVM Method

Type of Kernel	Formula
Linear	$x_i \times x_j$
Polynomial	$[\gamma x_i x_j + \text{coefficient}]^d$
RBF	$\exp\ (-\gamma x_i - x_j^2)\ $
Sigmoid	$\tanh(\gamma x_i x_j + \text{coefficient})$

effect of the second part of the optimization function. The value of nu is selected between zero and one. In order to solve Eqs. (7) and (9), the Wolfe dual Lagrange function is used by keeping the Karush–Kuhn–Tucker conditions. Detailed information can be found in Theodoridis and Koutroumbas (2009). By obtaining the w and b vectors by maximizing the Wolfe dual Lagrange function, the decision function to classify a new sample is (Yang et al. 2008):

$$G(x) = \text{sign}[f(x)] = \text{sign}[\varphi(x) \cdot w + b] \quad (11)$$

To obtain appropriate results in classification using SVM, two aspects must be considered. First, to obtain optimal values for the parameters c and nu , respectively, the methods c-SVM and nu-SVM are used by cross-validation method. Second, to obtain the appropriate kernel function, the trial-and-error method is used. In the SVM method, four functions of linear kernel, polynomial kernel, radial basis function, and the sigmoid function are used additionally (Table 1).

GEOLOGY OF THE STUDY AREA

Sari-Gunay deposit is located in 60 km west of Hamadan city in the northwestern part of Iran. The geology of this area in terms of tectonic setting is part of the Tethyan Ocean formed between the Eurasian and Gondwana supercontinents. Evidence for closure of at least two oceans (i.e., Paleotethys in Paleozoic and Neotethys in Cenozoic) has been recorded in this area (Alavi 1994; Bagheri and Stampfli 2008; Stampfli and Borel 2004; Stöcklin 1968). The Sanandaj–Sirjan magmatic–metamorphic zone and Urumieh–Dokhtar magmatic arc (Fig. 2a) are two important geologic subdivisions of the Zagros orogenic belt in this area (Aliyari et al. 2012; Kouhestani et al. 2012). Epithermal gold and porphyry copper mineralizations in these two zones are associated with subduction and obduction processes from Jurassic to Miocene (Dargahi et al. 2010; Mirnejad et al. 2011; Moosavi et al. 2008).

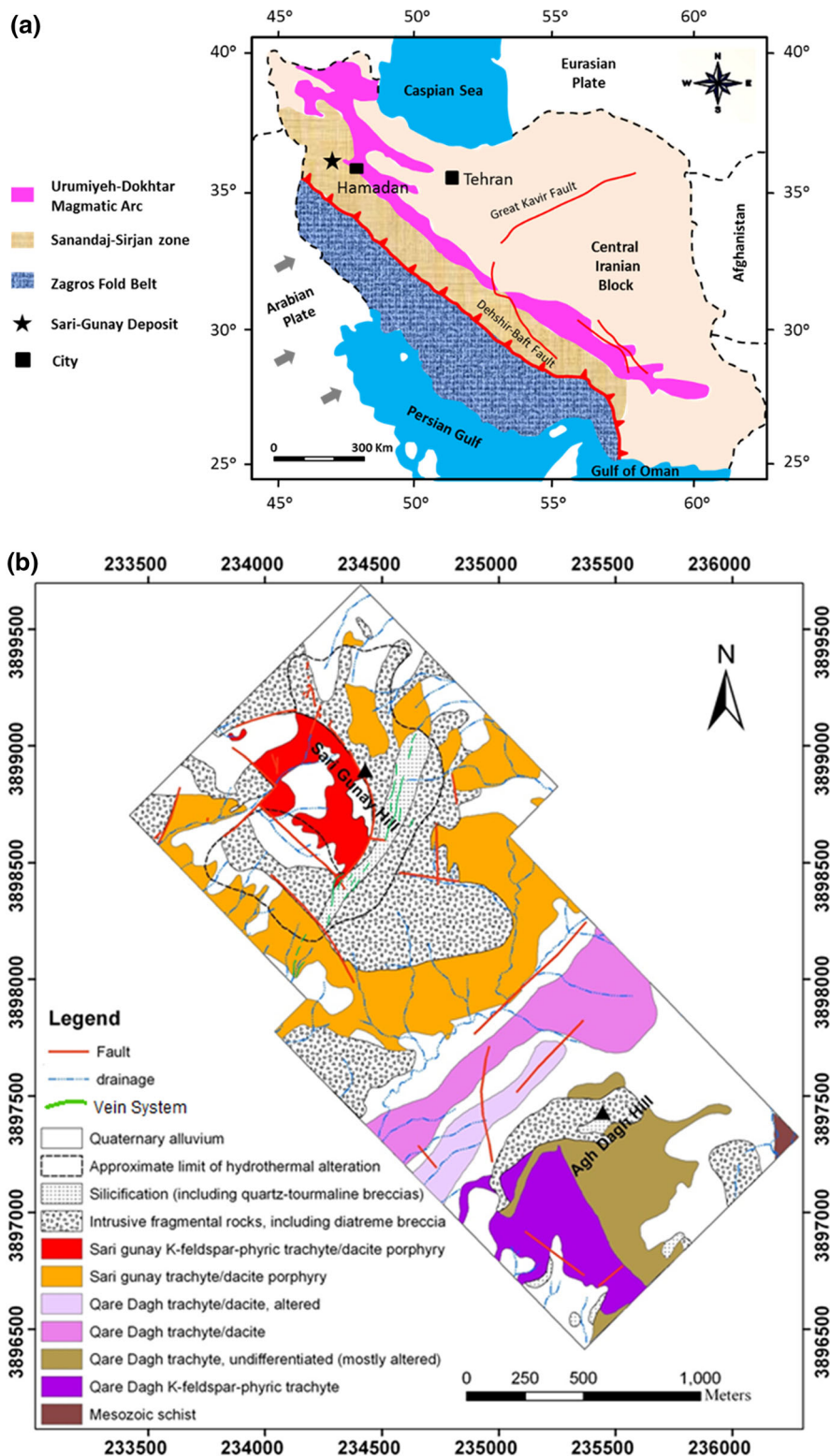


Figure 2. a Sanandaj–Sirjan zone and Urumieh–Dokhtar magmatic arc in Iran. b Geological map of the Sari-Gunay epithermal gold deposit (Wilkinson 2005).

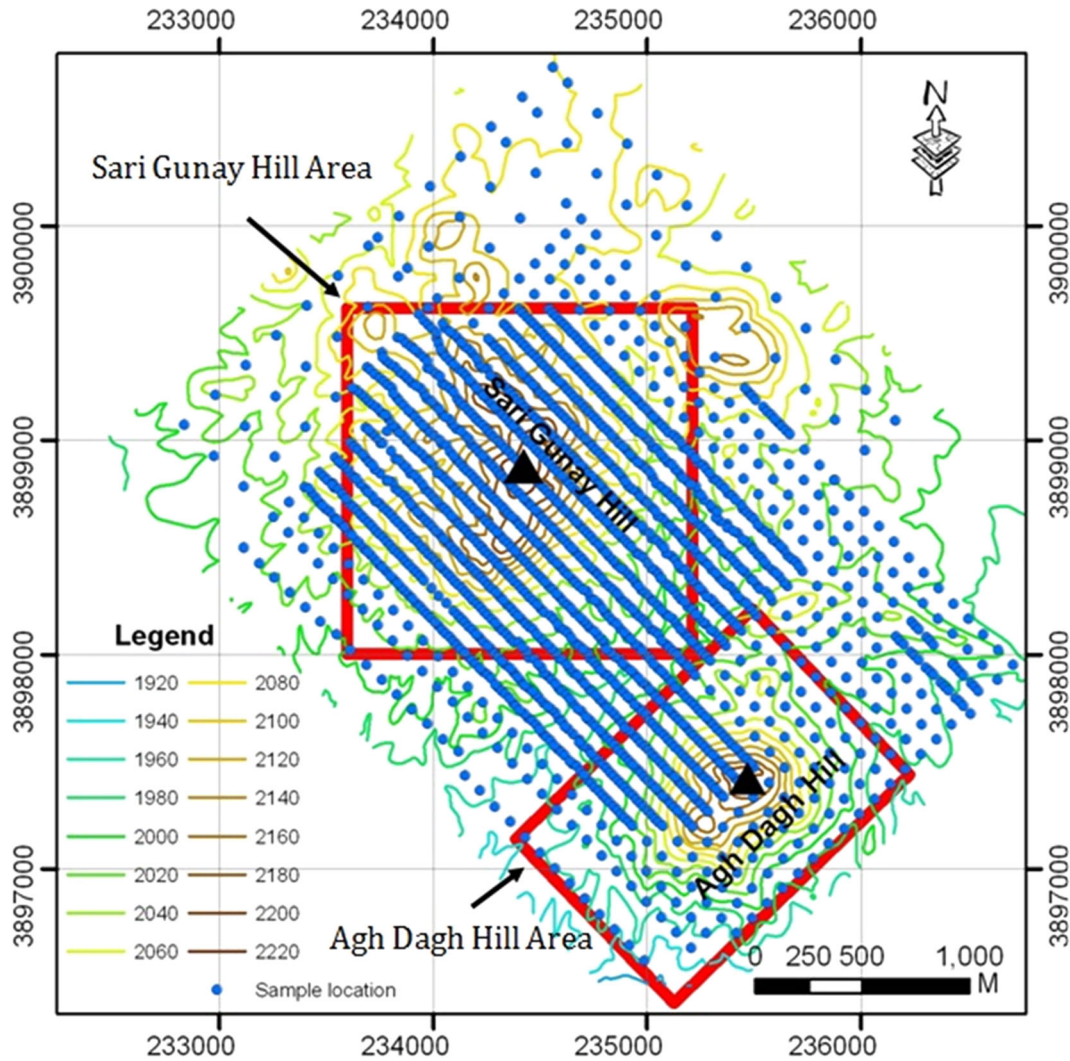


Figure 3. Location map of the soil geochemical samples on the topography map of the study area.

The Sari-Gunay gold deposit is composed of two mineralized zones (Sari-Gunay hill in the northwest and Agh-Dagh hill in the southeast) (Figs. 2b, 3). It lies on the northern western part of Sanandaj–Sirjan zone (Fig. 2a). This deposit is related to sub-volcanic intrusion in a high potassium sub-alkaline volcanic complex of the Middle Miocene (Richards et al. 2006). The bedrock in this area is mostly Jurassic limestone and some clastic sediments as well as metamorphic sedimentary rocks. These rocks are covered by andesitic volcanic and Oligo-Miocene to Pleistocene pyroclastic rocks. The mineralized host rocks are strongly altered diatreme breccia with sub-alkaline dacite porphyry (Fig. 2b). These rock units are intersected by a large number of antimony sulfide and arsenic veins (Asadi et al. 2014).

According to Richards et al. (2006), Sari-Gunay is mostly a large low sulfidation epithermal system showing strong vertical and lateral gold mineralization associated with silicified and brecciated volcanic and sub-volcanic rocks. The epithermal system is characterized by a gold–stibnite–mercury-rich core, surrounded by arsenic halos and further away by lead–silver–zinc–copper halos.

Two distinct mineralization phases of epithermal gold and porphyry Cu–Au have been recorded at Sari-Gunay (Richards et al. 2006; Wilkinson 2005). The first mineralization phase is a porphyry system associated with potassic, sericite, and quartz–tourmaline–feldspar–potassic alteration assemblages. This porphyry Cu–Au mineralization is low grade and deeper than 500 m and is mostly restricted

Table 2. Statistical Parameters of the Dependent Variable and Independent Variables Involved in the Classification Functions

Variable (Unit)	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
Productivity Index (g m/t)	0	16,830	1754.40	2931.90	4.19	2.08
Ag (g/t)	0.50	150.54	2.18	9.15	9.69	114.58
As (g/t)	16.00	9882.00	331.73	570.04	8.42	109.94
Ba (g/t)	57.00	1963.00	427.36	199.17	2.17	9.15
Be (g/t)	0.50	3.00	1.07	0.51	1.09	.65
Cd (g/t)	0.50	60.00	0.95	2.01	17.98	462.51
Co (g/t)	2.00	51.00	16.24	4.33	0.62	4.23
Cu (g/t)	7.00	513.00	36.98	39.80	5.79	45.55
Hg (g/t)	0.50	770.00	3.09	21.40	28.26	970.67
K (%)	0.11	1.18	0.51	0.17	-0.23	0.54
La (g/t)	17.00	152.00	50.17	12.37	0.87	4.92
Na (%)	0.01	0.09	0.02	0.01	2.03	7.78
Ni (g/t)	7.00	109.00	55.23	17.14	0.10	-0.33
P (%)	0.041	0.55	0.09	0.03	5.18	51.38
Pb (g/t)	15.00	11436.00	337.49	660.44	8.00	89.19
S (%)	0.01	3.60	0.08	0.14	12.26	254.15
Sb (g/t)	2.50	11115.00	147.72	510.61	13.28	229.33
Sc (g/t)	2.00	17.00	6.40	1.87	1.23	3.21
Th (g/t)	3.00	133.00	13.55	7.06	5.86	70.02
W (g/t)	2.50	590.00	4.43	16.68	27.40	905.71
Y (g/t)	5.00	29.00	15.00	2.86	0.42	1.04
Zr (g/t)	1.00	17.00	3.61	1.89	2.08	7.77
Au (mg/t)	1.10	7580.00	157.24	486.05	7.26	73.44

to the sub-volcanic dykes showing weak to moderate potassic alteration. The second phase of mineralization is low-temperature epithermal gold system associated with strong silicification and sericite-fengite alteration and formation of gold that is mostly hosted in arsenian pyrite (Wilkinson 2005; Richards et al. 2006).

The porphyry mineralization mostly occurred as quartz-pyrite-chalcopyrite-tennantite-magnetite vein/veinlets (with average 0.25% Cu and ≤ 0.5 g/t Au grade), quartz-tourmaline veins and breccia cements (with minor amounts of sulfide minerals and low gold grade), quartz-pyrite-stibnite-realgar-orpiment veins (with dominantly gold mineralization), and quartz-calcite-pyrite \pm galena \pm sphalerite veins (with low grade of base metals). The dominant vein trend is 20°–30° (NNE-SSW) with dips from vertical to 70° WNW. Sari-Gunay with 120–130 Mt resource at an average grade of 2 g/t gold is so far the largest discovered gold deposit in Iran (Wilkinson 2005; Richards et al. 2006).

Mineralization at Agh-Dagh hill is mainly associated with diatreme breccia on northwest slopes of the hill (Fig. 2b). Arsenic mineralization at Agh-Dagh hill is stronger, and gold and antimony mineralizations are weaker than Sari-Gunay hill (Wilkinson 2005). Mineralization at Agh-Dagh de-

posit is associated with sills that contain the suite of metals Fe–As–Sb–Hg–Au–Ag–Tl with outer zones enriched in Fe–Pb–Ag–Zn elements.

DATA PREPARATION

Within the Sari-Gunay exploration area, 1724 soil geochemical samples were collected and analyzed for 46 elements by ICP-MS method and for Au by fire assay method. The soil sample density varied from 100 m 25 m to 100 m 100 m, depending on the evidences of surface mineralization (Fig. 3). The most effective elements (listed in the Table 2 with their statistical parameters) have been selected from 47 elements by using stepwise and feature selection methods (see Sect. 5.1 and 5.2 for details) and then used for DA and SVM classifications. After removing the outlier and censored geochemical data, the results were normalized by using normal score transformation method (Tabachnick and Fidell 2012). In the normal score transformation, in addition to data normalization, standardization was also considered (Siddiqui and Syed Osman 2013). Therefore, the mean of the total normal score of geochemical samples per cell is considered as the normal score of each element in that cell.

Forty six exploration boreholes data have been used in this study (Fig. 4). Core samples were analyzed like surficial soil geochemical samples and in OMAC laboratory in Ireland with supervised by Rio-Tinto Mining and Exploration Limited. In-house reference materials were used in every batch of 50 samples and sub-sample duplicates for each 25 measurements to check the analytical accuracy. This accuracy has been evaluated by calculating the relative standard deviation (RSD) with Thompson and Howarth (1976) method. Analytical precision (2 RSD) exhibited better than $\pm 10\%$ for all elements.

In this research, ordinary kriging was used to interpolate and calculate the average gold grade of blocks of 10 m 25 m 25 m as the optimal dimensions for reserve estimation (Bartram 2005). Then, for each 25 m 25 m cell at the surface (Fig. 4a), the sum of block grades above cutoff grade (0.5 g/t) multiplied by the thickness of the blocks was calculated and considered

as an index of the cell productivity (Solovov 1985). In Table 2, the statistical parameters of the productivity index (PI) are shown. From 1767 cells, 660 cells were below economic gold mineralization (i.e., the PI is zero), and regarded as the background class.

To identify the number of classes of the remaining cells with economic gold mineralization, the number-size fractal method was used (Mandelbrot 1983; Deng et al. 2009). In the fractal method, the cumulative number of grade has a power-law function with the mineralization variables such as block thickness and grade that can be used to calculate the PI (Sanderson et al. 1994; Roberts et al. 1998; Wang et al. 2010). The number-size plot showed that the PI data were composed of two classes (Fig. 4b). The inflection point (corresponding to a value of 4050 g/t) was considered as the threshold for separating the classes. Thus, the PI variable is classified into three populations, namely class A with PI of zero, class B with intermediate PI, and class C with high PI. Obviously, high PI in each cell represents gold mineralization in that cell. Figure 4a shows the distribution of these cells in the Sari-Gunay hill area.

CLASSIFICATION

From 1767 cells, with their PIs estimated, 387 cells overlap with the soil geochemical samples. This is due to the geochemical sampling with 100 m line spacing. Seventy percent of these soil geochemical samples were selected randomly as the training data to construct the classification function, and the remaining thirty percent were used as the testing data. The populations of these two datasets are shown in Table 3.

The first step in classification is to reduce the number of variables involved in the classification function, which is also called the reduced dimensions. This is due to the reason that some of the independent variables can be a linear or nonlinear combination of other independent variables. The

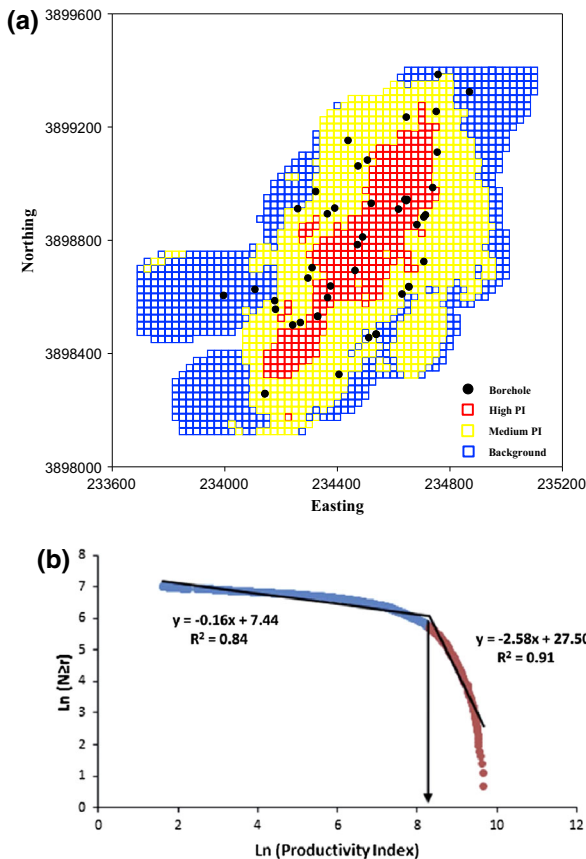


Figure 4. a Locations of the mineralized zones and boreholes in the Sari-Gunay hill area. b Number-size fractal plot of the PI variable.

Table 3. The Number of Cells Overlapped with Geochemical Samples Based on Each Population

Population	Productivity Index	Total N	Training N	Testing N
A	0	135	93	42
B	$5 < PI \leq 4050$	185	135	50
C	$4050 < PI \leq 16,830$	67	48	19
Sum		387	276	111
				z

presence of these variables in the function is useless and they will just increase the computations. Besides, reducing the number of variables would increase generalizability of the function and reduce errors in classification of the test data.

Classification by DA

In order to determine the effective elements in the discriminant function, the stepwise method with Mahalanobis distance was used (Roshani et al. 2013). The results showed that only seven out of the 47 elements (i.e., Au, Na, W, Hg, K, Y, and Co) are the most effective for DA of the present data. The Au as the major element in mineralization and Hg as the minor element in mineralization are appropriately included in the discriminant function. In addition, Co, W, and Y (low mobility elements) and Na and K were considered effective as they are typically associated with low sulphidation epithermal hydrothermal alteration (Pirajno 2009; Robert et al. 2007). Therefore, linear and quadratic discriminant functions by using the training data were computed for the seven significant elements. The leave-one-out cross-validation method (Hastie et al. 2009) was used to determine the accuracy of discriminant functions on the classification of the training data. Then, the obtained discriminant functions were used to classify the testing data, and the results were compared with the earlier classes. The results of DA classification of the training and testing data are shown in Tables 4 and 5.

The classification accuracy of the testing data was almost close to the classification of the training data (Tables 4, 5). This demonstrates an acceptable validity of the discrimination functions. The best accuracy of classification in both linear and nonlinear methods was attributed to the B population, and the lowest accuracy was attributed to the data of the C population. In general, the QDA function with classification accuracy of 68.2% showed a better performance than the LDA function with classification accuracy of 61.5 %.

Classification by SVM

First, the effective elements were selected through the feature selection method, with nearest neighbor criteria (Hastie et al. 2009). Gold, Ag, As, Ba, Be, Cd, Co, Cu, La, Ni, P, Pb, S, Sb, Sc, Th, Y,

and Zr are these elements that involved in the classification functions as the independent variables. Gold, Ag, Cu, and Pb were considered the key elements, and As, Sb, and S are also effective elements as they are typically associated with low sulphidation epithermal gold mineralization (Pirajno 2009; Robert et al. 2007). Classification of the training data was done by each four kernel functions, and the best results were achieved with radial basis function kernel. The 10-fold cross-validation was used to optimize parameters c and ν , as well as parameter γ for the selected radial basis function kernel. The results of data classification by two methods of c -SVM and ν -SVM are listed in Tables 6 and 7.

The results in Tables 6 and 7 indicated that in both methods of SVM, the best classification was attributed to the population A, and the lowest accuracy was attributed to the population C. The considerable difference between the accuracy of the classification of the training data and the testing data in both methods was due to the usage of the classification model obtained by the training data for classification of the given data. However, the accuracy of the obtained classification by the cross-validation method in both SVM methods was close to the classification accuracy of the testing data. In total, classification of the training data and the testing data, both methods with classification accuracy of 88.9% revealed the same behavior.

DISCUSSION

It is very important to define the location of the boreholes accurately in different stages of exploration. In most exploration projects, the borehole locations are identified by integrated analysis of the surface exploration data. In this research, we provide a new method to model the geometry of the mineralization by using a combination of surface geochemical data and subsurface drilling data to suggest further drilling locations. For this purpose, two classification methods were used in Sari-Gunay gold deposit.

By comparing the results of the classification by the DA method with the SVM method on the testing data, the ν -SVM method revealed to have the best ability and the QDA method showed the lowest ability to separate the data of the population A. The LDA method showed the weakest performance and three other methods showed the best performances in classification of the data of the population B. The

Table 4. Results of Classification of the Training and Testing Data by LDA

Put into Group	True Group in Training Data			True Group in Testing Data		
	A	B	C	A	B	C
A	54	28	0	29	13	2
B	39	96	27	13	29	8
C	0	11	21	0	8	9
Total <i>N</i>	93	135	48	42	50	19
Proportion correct	0.58	0.71	0.44	0.69	0.58	0.47
	Total <i>N</i> = 276			Total <i>N</i> = 111		
	Total <i>N</i> correct = 171			Total <i>N</i> correct = 67		
	Proportion correct = 0.62			Proportion correct = 0.60		

Table 5. Results of Classification of the Training and Testing Data by QDA

Put into Group	True Group in Training Data			True Group in Testing Data		
	A	B	C	A	B	C
A	65	27	2	28	8	2
B	27	97	19	14	37	7
C	1	11	27	0	5	10
Total <i>N</i>	93	135	48	42	50	19
Proportion correct	0.70	0.72	0.56	0.67	0.74	0.53
	Total <i>N</i> = 276			Total <i>N</i> = 111		
	Total <i>N</i> correct = 189			Total <i>N</i> correct = 75		
	Proportion correct = 0.69			Proportion correct = 0.68		

Table 6. Results of Classification Using c-SVM with Optimized Parameters $\gamma = 0.01$ and $c = 151$

Put into Group	True Group in the Training Data			True Group in the Testing Data		
	A	B	C	A	B	C
A	88	2	0	34	9	2
B	5	131	4	8	37	7
C	0	2	44	0	4	10
Total <i>N</i>	93	135	48	42	50	19
Proportion correct	0.946	0.970	0.917	0.809	0.740	0.526
	Total <i>N</i> = 276			Total <i>N</i> = 111		
	Total <i>N</i> correct = 263			Total <i>N</i> correct = 81		
	Proportion correct = 0.95			Proportion correct = 0.72		

Table 7. Results of Classification Using nu-SVM with Optimized Parameters $\gamma = 0.015$ and $\nu = 0.320$

Put into Group	True Group in the Training Data			True Group in the Testing Data		
	A	B	C	A	B	C
A	89	2	1	34	6	1
B	4	130	4	8	37	7
C	0	3	43	0	4	11
Total <i>N</i>	93	135	48	42	50	19
Proportion correct	0.957	0.963	0.896	0.810	0.740	0.580
	Total <i>N</i> = 276			Total <i>N</i> = 111		
	Total <i>N</i> correct = 292			Total <i>N</i> correct = 82		
	Proportion correct = 0.95			Proportion correct = 0.74		

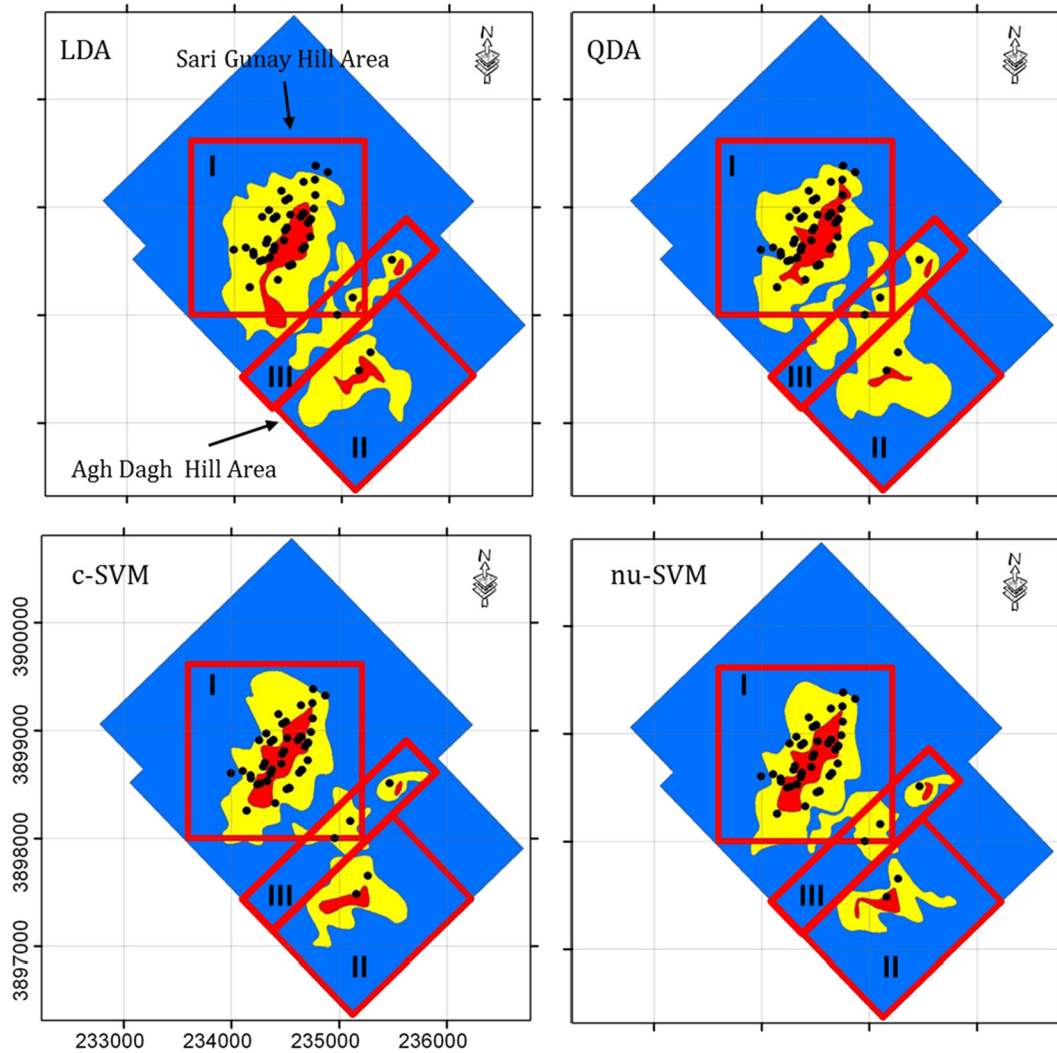


Figure 5. Predicting models of gold mineralization by four classification functions in the exploration area (The background, moderate, and high potential areas are shown as blue, yellow, and red, respectively; black points indicate the location of the drilled exploration boreholes).

best and weakest methods for classification of the data of the population C were attributed the nu-SVM and LDA methods, respectively. Therefore, it can be concluded that the methods of SVM perform a better classification than DA methods. Also, the nu-SVM method was the best method of classification, and after that *c*-SVM, QDA, and LDA methods showed better performances, respectively.

These results can be interpreted with two mathematical and geological aspects. In mathematical view, SVM method can be applied irrespective to data distribution and outliers. This method is also known as the generalization per-

formance of the classifier (Duda et al. 2000; Theodoridis and Koutroumbas 2009). Therefore, it can be operated satisfactorily with training and testing data to gain better results. But, in geological view, Ag, As, Cd, Cu, Pb, S, and Sb independent variables are chalcophile elements that have a significant positive correlation with gold mineralization. Ba, Be, La, and Zr independent variables are also lithophile elements that have a significant negative correlation with gold mineralization in the study area. These two groups of elements have major role in the classification functions of the SVM method.

The above-mentioned four classification functions were applied to the geochemical data to predict the extension of gold mineralization at surface and to determine the optimum locations for further drillings in Sari-Gunay hill. These data were also used to define the appropriate locations for reconnaissance drilling at the Agh-Dagh hill. The results of surface modeling are shown in Figure 5. The classification results indicated three gold potential areas:

- I. The strongest and largest gold potential area is located in the central part of the Sari-Gunay hill. As shown in Figure 5, the four methods of classification separated the areas with high and moderate PI values from the background. However, the comparison of the obtained models with kriging indicates that the shape and size of these two potential areas in the obtained models by SVM method are more consistent with actual results (Fig. 4a). In addition, the mineralization model of QDA method is reasonably consistent with the kriging model.
- II. The second important gold potential area is located in Agh-Dagh hill. All four classification functions revealed two strong and moderate gold mineralized zones (high and moderate PI values) at the Agh-Dagh hill (Fig. 5). In terms of shape and size of these two zones, two methods of LDA and c-SVM were similar. The QDA method showed the largest gold mineralized area and the nu-SVM method showed the smallest gold mineralized area at the Agh-Dagh hill.
- III. The third important gold potential area is located between the Sari-Gunay and the Agh-Dagh hills. This area is mostly covered by soil. There are a few stibnite veins trending in a NE-SW direction in this area. All classification models showed moderate mineralized zones in this area (Fig. 5). The LDA method determined the continuity of the mineralization better than the other methods in this zone. It is concluded that LDA method is more useful in determining small areas of oriented structural control mineralization, while SVM works better to determine large areas of disseminated mineralization in this study area.

Two and three boreholes already successfully drilled at the Agh-Dagh hill and the area between

the Sari-Gunay and Agh-Dagh hill, respectively. Four of these boreholes lie on the moderate and one hole lies in the high potential area. Lithology and chemical analysis of cores, obtained from these boreholes, shown that gold mineralization occurred in lithic tuff breccia and dacite-andesite porphyritic rocks in dispersion vein-like shape with an average grade of 3.1 g/t gold in II area (Fig. 5) and dacite crystalline tuff and dacite-andesite porphyritic rocks in narrow veins with an average grade of 4.1 g/t gold in III area (Fig. 5). Silicification and sericite alterations are the principal hydrothermal alterations in these rocks. Therefore, both moderate and high potential areas (moderate and high PI) are proposed as potential areas for further reconnaissance drilling. The high potential areas (high PI) at Sari-Gunay hill is proposed for further systematic drilling, and moderate potential areas here are proposed for further reconnaissance drilling.

CONCLUSIONS

The main conclusions obtained from the applications of two classification methods, DA and SVM, to the surface soil and drilling data to locate the high gold potential areas for further drilling in the Sari-Gunay gold deposit are as follows:

1. SVM methods classified the PI in all three classes of background, moderate, and high areas for the training and testing data due to the lower structural errors.
2. All four models of gold mineralization prediction in Sari-Gunay hill showed reasonable agreement with the model obtained by the kriging method. Nevertheless, the SVM models showed a better fit with kriging model in terms of shape and extent of the mineralization with different quantities.
3. All four models predicted the mineralization in Agh-Dagh hill, which is very similar to Sari-Gunay hill. However, the size of the strong mineralization zone in the Agh-Dagh hill is much smaller than the Sari-Gunay hill.
4. The obtained gold mineralization prediction model by the LDA method showed good continuity of gold mineralization along the northeast-southwest trending valley between the two hills.
5. Zones of high PI are proposed as the first priority for infill systematic drilling at the

Sari-Gunay hill, and zones with medium PI are proposed for further reconnaissance drilling at Agh-Dagh hill as well as the area between Sari-Gunay and Agh-Dagh hills.

ACKNOWLEDGMENTS

We would like to thank the Zar Kuh mining company, the owner of the Sari-Gunay gold deposit, for providing the geochemical data. We acknowledge Prof. Renguang Zuo, Associate Editor of NRR, for expert handling of our manuscript and we thank two anonymous reviewers for their constructive comments.

REFERENCES

- Abbaszadeh, M., Hezarkhani, A., & Soltani-Mohammad, S. (2013). An SVM-based machine learning method for the separation of alteration zones in Sungun porphyry copper deposit. *Chemie der Erde: Geochemistry*, 73, 545–554.
- Abedi, M., Norouzi, G. H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences*, 46, 272–283.
- Agterberg, F. P. (1974). Automatic contouring of geological maps to detect areas for mineral exploration. *Mathematical Geology*, 6, 373–395.
- Al-Anazi, A., & Gates, I. D. (2010). On the capability of support vector machines to classify lithology from well logs. *Natural Resources Research*, 19, 125–139.
- Alavi, M. (1994). Tectonics of Zagros Orogenic belt of Iran, new data and interpretation. *Tectonophysics*, 229, 211–238.
- Aliyari, F., Rastad, E., & Mohajjel, M. (2012). Gold deposits in the Sanandaj-Sirjan zone: Orogenic gold deposits or intrusion-related gold systems. *Resource Geology*, 62, 296–315.
- Asadi, H. H., Kianpouryan, S., Lu, Y., & McCuaig, T. C. (2014). Exploratory data analysis and C-A fractal model applied in mapping multi-element soil anomalies for drilling: A case study from Sari-Gunay epithermal gold deposit, NW Iran. *Journal of Geochemical Exploration*, 145, 233–245.
- Auria, L., & Moro, R. A. (2008). *Support vector machines (SVM) as a technique for solvency analysis*. Discussion Papers No. 811 of DIW, Berlin, pp. 1–16.
- Bagheri, S., & Stampfli, G. M. (2008). The Anarak, Jandaq and Posht-e-Badam metamorphic complexes in central Iran: New geological data, relationships and tectonic implications. *Tectonophysics*, 451, 123–155.
- Bartram, J. (2005). *Pre-feasibility compilation of reports and memo*. Rio-Tinto Mining and Exploration 421 Ltd Technical Report, p. 106.
- Belkhir, L., & Mouni, L. (2014). Geochemical characterization of surface water and groundwater in Soummam Basin, Algeria. *Natural Resources Research*, 23, 393–407.
- Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines, Chap. 13. In O. Carugo & F. Eisenhaber (Eds.), *Data mining techniques for the life sciences* (pp. 223–239). Totowa, NJ: Humana Press.
- Bökeoğlu Çokluk, Ö., & Büyüköztürk, Ş. (2008). Discriminant function analysis: Concept and application. *Eğitim Araştırmaları Dergisi*, 33, 73–92.
- Bonham-Carter, G. F., Agterberg, F. P., & Wright, D. F. (1988). Integration of geological datasets for gold exploration in Nova Scotia. *Photogrammetric Engineering and Remote Sensing*, 54, 1585–1592.
- Bonham-Carter, G. F., & Chung, C. F. (1983). Integration of mineral resource data for Kasmere Lake area, Northwest Manitoba, with emphasis on uranium. *Mathematical Geology*, 15, 25–45.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 121–167.
- Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of Chemical Information and Computer Sciences*, 43, 1882–1889.
- Carranza, E. J. M. (2009). Geochemical anomaly and mineral prospectivity mapping in GIS. In M. Hale (Ed.), *Handbook of exploration and environmental geochemistry* (Vol. 11). Amsterdam: Elsevier.
- Chatterjee, S., & Bandyopadhyay, S. (2011). Goodness Bay Platinum resource estimation using least squares support vector regression with selection of input space dimension and hyperparameters. *Natural Resources Research*, 20, 117–129.
- Chung, C. F. (1977). An application of discriminant analysis for the evaluation of mineral potential. In: Ramani, R. V. (Ed.), *Application of Computer Methods in the Mineral Industry, Proceedings of the 14th APCOM Symposium*, Society of Mining Engineers of American Institute of Mining, Metallurgical, and Petroleum Engineers, New York, pp. 299–311.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioural sciences* (3rd ed.). New York: Routledge.
- Cristianini, N., & Taylor, J. S. (2000). *An introduction to support vector machines and other Kernel-based learning methods*. Cambridge: Cambridge University Press.
- Croux, C., & Joossens, K. (2005). Influence of observations on the misclassification probability in quadratic discriminant analysis. *Journal of Multivariate Analysis*, 96, 384–403.
- Dargahi, S., Arvin, M., Pan, Y., & Babaei, A. (2010). Petrogenesis of post-collisional A-type granitoids from the Urumieh-Dokhtar magmatic assemblage, Southwestern Kerman, Iran: Constraints on the Arabian-Eurasian continental collision. *Lithos*, 115, 190–204.
- Davis, J. C. (2002). *Statistics and data analysis in geology* (3rd ed.). New York: Wiley.
- Deng, J., Wang, Q., Wan, L., Yang, L., Gong, Q., Zhao, J., & Liu, H. (2009). Self-similar fractal analysis of gold mineralization of Dayingezhuang disseminated-veinlet deposit in Jiaodong gold province, China. *Journal of Geochemical Exploration*, 102, 95–102.
- Divi, S. R., Thorpe, R. I., & Frankli, J. M. (1979). Application of discriminant analysis to evaluate compositional controls of stratiform massive sulfide deposits in Canada. *Mathematical Geology*, 11, 391–406.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). New York: Wiley.
- Fedikow, M. A. F., Parbery, D., & Ferreira, K. J. (1991). Geochemical target selection along the Agassiz Metaltect utilizing stepwise discriminant function analysis. *Economic Geology*, 86, 558–559.
- Friedman, J. (1989). Regularized discriminant analysis. *Journal of America Statistical Association*, 84, 165–175.
- George, C., & Fernandez, J. (2002). Discriminant analysis, a powerful classification technique in data mining. *Statistics and data analysis* pp. 244–247.

- Harris, D. P. (1965). Multivariate statistical analysis—a decision tool for mineral exploration. In J. C. Dotson & W. C. Peters (Eds.), *Symposium on computers and computer applications in mining and exploration* (pp. C1–C35). Ariz: College of Mines, University of Arizona, Tucson.
- Harris, D. P., & Pan, G. (1999). Mineral favorability mapping: A comparison of artificial neural networks, logistic regression, and discriminant analysis. *Natural Resources Research*, 8, 93–109.
- Harris, D. P., Zurcher, L., Stanley, M., Marlow, J., & Pan, G. (2003). A comparative analysis of favorability mappings by weights of evidence, probabilistic neural networks, discriminant analysis, and logistic regression. *Natural Resources Research*, 12, 241–255.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Berlin: Springer.
- Hu, Y., & Yu, D. (2011). The comparison of five discriminant methods. In *International Conference on Management and Service Science (MASS)*, pp. 1–4.
- Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11, 352–359.
- Kouhestani, H., Ghaderi, M., Zaw, K., Meffre, S., & Hashem Emami, M. (2012). Geological setting and timing of the Chah Zard breccia-hosted epithermal gold–silver deposit in the Tethyan belt of Iran. *Mineralium Deposita*, 47, 425–440.
- Li, J. (2005). Multiattributes pattern recognition for reservoir prediction. *CSEG National Convention*.
- Mandelbrot, B. B. (1983). *The fractal geometry of nature*. San Francisco: Freeman.
- McKinley, J. M., Roberson, S., Cooper, M., & Tolosana-Delgado, R. (2014). Discriminant analysis of palaeogene Basalt Lavas, Northern Ireland, Using Soil Geochemistry. In *Mathematics of Planet Earth*. Lecture Notes in Earth System Sciences, p. 103–106.
- McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. New York: Wiley.
- Mirnejad, H., Simonetti, A., & Molasalehi, F. (2011). Pb isotopic compositions of some Zn–Pb deposits and occurrences from Urumieh-Dokhtar and Sanandaj-Sirjan zones in Iran. *Ore Geology Reviews*, 39, 181–187.
- Moon, C. J., Whateley, M. K. G., & Evans, A. M. (2006). *Introduction to mineral exploration*. Oxford: Blackwell Publishing.
- Moosavi, S. A., Heidari, S. M., Rastad, E., Esfahaninejad, M., & Rashidnejad Omran, N. (2008). A brief review of mineral deposit types and geodynamic settings related to Neotethys in Iran. *Geosciences*, 17, 132–142.
- Pan, G. C., & Harris, D. P. (2000). *Information synthesis for mineral exploration*. New York: Oxford University Press Inc.
- Pirajno, F. (2009). *Hydrothermal processes and mineral systems*. Australia: Springer Publication.
- Prelat, A. E. (1977). Discriminant analysis as a method of predicting mineral occurrence potentials in central Norway. *Mathematical Geology*, 9, 343–367.
- Richards, J. P., Wilkinson, D., & Ullrich, T. (2006). Geology of the Sari-Gunay epithermal gold deposit northwest Iran. *Economic Geology*, 101, 1455–1496.
- Robert, F., Brommecker, R., Bourne, B. T., Dobak, P. J., McEwan, C. J., Rowe, R. R., & Zhou, X. (2007). Models and exploration methods for major gold deposit types. In B. Milkereit (Ed.) *Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration*, pp. 691–711.
- Roberts, S., Sanderson, D. J., & Gumiel, P. (1998). Fractal analysis of Sn–W mineralization from central Iberia, insights into the role of fracture connectivity in the formation of an ore deposit. *Economic Geology*, 93, 360–365.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2014). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*. doi:10.1016/j.oregeorev.2015.01.001.
- Rose, A. W. (1972). Favorability for Cornwall-type magnetite deposits in Pennsylvania using geological, geochemical and geophysical data in a discriminant function. *Journal of Geochemical Exploration*, 1, 181–194.
- Roshani, P., Mokhtari, A. R., & Tabatabaei, S. H. (2013). Objective based geochemical anomaly detection: Application of discriminant function analysis in anomaly delineation in the Kuh Panj porphyry Cu mineralization (Iran). *Journal of Geochemical Exploration*, 130, 65–73.
- Sanderson, D. J., Roberts, S., & Gumiel, P. (1994). A fractal relationship between vein thickness and gold grade in drill core from La Codosera, Spain. *Economic Geology*, 89, 168–173.
- Savu-Krohn, C., Rantitsch, G., Auer, P., Melcher, F., & Graupner, T. (2011). Geochemical fingerprinting of Coltan ores by machine learning on uneven datasets. *Natural Resources Research*, 20, 177–191.
- Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge: Cambridge University Press.
- Siddiqui, F. I., & Syed Osman, S. B. A. B. (2013). Simple and multiple regression models for relationship between electrical resistivity and various soil properties for soil characterization. *Environmental Earth Sciences*, 70, 259–267.
- Solovov, A. P. (1985). *Chemical prospectivity for mineral deposits*. Amazon.
- Srivastava, M. S. (2002). *Methods of multivariate statistics*. New York: Wiley.
- Stampfli, G. M., & Borel, G. D. (2004). The TRANSMED transects in space and time: Constraints on the paleotectonic evolution of the Mediterranean domain. In W. Cavazza, F. Roure, W. Spakman, G. M. Stampfli, & P. Ziegler (Eds.), *The TRANSMED atlas: the mediterranean region from crust to mantle* (pp. 53–80). Berlin: Springer Verlag.
- Stöcklin, J. (1968). Structural history and tectonics of Iran: A review. *The American Association of Petroleum Geologists Bulletin*, 52, 1229–1258.
- Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston: Pearson Publisher.
- Tahmasebi, P., Hezarkhani, A., & Mortazavi, M. (2010). Application of discriminant analysis for alteration separation; Sungun Copper Deposit, East Azerbaijan, Iran. *Australian Journal of Basic and Applied Sciences*, 6, 564–576.
- Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition* (4th ed.). Amsterdam: Elsevier.
- Thompson, M., & Howarth, R. J. (1976). Duplicate analysis in geochemical practice. Part 1: Theoretical approach and estimation of analytical reproducibility. *Analyst*, 101, 690–698.
- Tibljias, D., Loparic, V., & Belak, M. (2002). Discriminant function analysis of Miocene volcanoclastic rocks from North-Western based geochemical data. *Geologia Croatica*, 55, 39–44.
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10, 988–999.
- Varadanchari, C., & Mukherjee, G. (2004). Discriminant analysis of clay mineral composition. *Journal of Clay and Clay Minerals*, 52, 311–320.
- Venkataraman, G., Babu Madhavan, B., Ratha, D. S., Antony, J. P., Goyal, R. S., Banglani, S., & Sinha Roy, S. (2000). Spatial modeling for base-metal mineral exploration through integration of geological data sets. *Natural Resources Research*, 9, 27–42.
- Wang, Q., Deng, J., Liu, H., Yang, L., Wan, L., & Zhang, R. (2010). Fractal models for ore reserve estimation. *Ore Geology Reviews*, 37, 2–14.

- Whitehead, R. E. S., & Govett, G. J. S. (1974). Exploration rock geochemistry—detection of trace element halos, Heath Steele Mines (N.B., Canada), by discriminant analysis. *Journal of Geochemical Exploration*, 3, 371–386.
- Wilkinson, L. D. (2005). *Geology and mineralization of the Sari-Gunay gold deposits*. Kordestan province Iran, Rio-Tinto Ltd Technical Report.
- Wu, C., Lv, X., Cao, X., Mo, Y., & Chen, C. (2010). Application of support vector regression to predict metallogenic favorability degree. *International Journal of the Physical Sciences*, 5, 2523–2527.
- Wu, W., Mallet, Y., Walczak'va, B., Penninckx, W., Massarta, D. L., Heuerdingb, S., & Ernib, F. (1996). Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data. *Analytica Chimica Acta*, 329, 257–265.
- Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, 34, 592–602.
- Yu, H., & Kim, S. (2012). *SVM tutorial—classification, regression and ranking*. *Handbook of natural computing* (pp. 479–506). Berlin: Springer.
- Yu, L., Porwal, A., Holden, E. J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, 45, 229–239.
- Zuo, R., & Carranza, E. J. M. (2011). Support vector machine: A tool for mapping mineral prospectivity. *Computers & Geosciences*, 37, 1967–1975.