

Quantitative evaluation of mining geo-environmental quality in Northeast China: comprehensive index method and support vector machine models

Xue Jiang · Wen-xi Lu · Hai-qing Zhao ·
Qing-chun Yang · Mo Chen

Received: 25 January 2014 / Accepted: 9 December 2014 / Published online: 18 December 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract The long-term mining of mineral resources has contributed to damaging the geo-environment, drawing significant attention to the evaluation of geo-environmental quality. This paper presents an original system for evaluating mining geo-environmental quality in the Changjitu economic zone, Northeast China. The proposed evaluation framework considers five criteria, namely basic mining information, the geo-environmental background of the mining area, mining-related geological problems, the importance of the evaluation area, and the difficulty of geo-environmental recovery. Objective weighting methods, such as the variation coefficient, entropy, and Kantiray weighting methods, and subjective weighting methods such as the analytic hierarchy process are developed to determine the comprehensive weights of the elements and indicators. A common comprehensive index method and a new support vector machine (SVM) model are then proposed and compared to evaluate mining geo-environmental quality. The findings show that the accuracy of the linear SVM model is 93.10 %, demonstrating that the SVM is appropriate for the evaluation of mining geo-environmental quality. Compared with existing common methods, the SVM model, which classifies mining geo-environmental

quality into multiple groups, adopts the structural risk minimization principle. The evaluation results also show that mining geo-environmental quality tends to rank as level II in the study area, accounting for 75.86 % of the total eligible mines compared with 2.59 and 21.55 % for levels I and III, indicating that most mining geo-environments are moderately affected by mining activities.

Keywords Comprehensive index method · SVM · Mining geo-environment · Quality evaluation · Changjitu economic zone

Introduction

The rapid growth in economic benefits along with the increased capacity to extract mineral resource around the globe has contributed to damaging the geo-environment (Monjezi et al. 2009; Mayes et al. 2009; Jordan and Project 2009; Schellenbach and Krekeler 2012; Huang et al. 2012; Howladar 2013). For example, mining and related industries have long operated in the Changjitu economic zone (CEZ) of Jilin Province, Northeast China. However, while the long-term exploitation of mineral resource has created enormous economic benefits, it has also caused serious environmental problems such as geological disasters, resource damage, and environmental pollution that cannot be ignored (Das 1999; Jiang et al. 2014). Geological disasters caused by underground mining include ground subsidence, ground fissures, and the inrush of mine water, while open pit mining has led to collapses, landslides, and debris flows (Xu et al. 2006; Liu et al. 2008). Further, mining activities have damaged land/water resources and the local landscape (Brown 2005), while the widespread discharging of mining waste

X. Jiang · W. Lu (✉) · Q. Yang · M. Chen
Key Laboratory of Groundwater Resources and Environment,
Ministry of Education, Jilin University, Changchun,
Jilin 130021, People's Republic of China
e-mail: luwx999@163.com

X. Jiang · W. Lu · Q. Yang · M. Chen
College of Environment and Resources, Jilin University,
Changchun, Jilin 130021, People's Republic of China

H. Zhao
Shenyang Institute of Geology and Mineral Resources,
Shenyang, Liaoning 110032, People's Republic of China

residues, wastewater, and waste gas represents a major source of pollution that threatens the health of residents (Fessler et al. 2003; Ghose 2003; Mars and Crowley 2003; Pierce et al. 2004; Ryser et al. 2005; Mendez et al. 2006; Krekeler et al. 2008, 2010; Geise et al. 2011; Schellenbach and Krekeler 2012).

The evaluation of mining geo-environmental quality (MGEQ) has important theoretical and practical value based on the comparative analysis of the major geological problems caused by different types of mineral development, different regions, and different mining approaches in the CEZ. It can provide not only a scientific basis for restoring the geological environment and rationally exploiting mineral resources, but also geo-environmental conditions for the suitable distribution of economic and social benefits in the CEZ, which in turn promotes the coordinated development of the economic, social, and geological environments (Li 2013).

In the past few years, numerous authors have researched the selection of evaluation factors, establishment of evaluation indicators, and choice of evaluation methods (Mejía-Navarro et al. 1994; Sarkar et al. 2007; Turer et al. 2008; Huang et al. 2012). For example, in China, Cai et al. (1998) used the analytic hierarchy process (AHP) to determine the weights of evaluation factors (Cai et al. 1998), while Liao and Wu (2004) applied a fuzzy mathematics method to evaluate the mining environment (Liao and Wu 2004). Xu et al. (2003, 2006) classified mining geo-environmental problems in Northwest China by applying the multi-level fuzzy synthetic evaluation, gray situation assessment, and comprehensive evaluation methods to assess MGEQ (Xu et al. 2003, 2006). Further, Ma (2013) determined an MGEQ evaluation system based on five aspects, namely natural geography, basic geology, mining development area, mining-related geological disasters, and mining environment, and selected AHP combined with a fuzzy mathematics method to evaluate MGEQ in two mining areas (Ma 2013).

In the UK, Legg (1990) used remote sensing technology to qualitatively evaluate the environmental and land reclamation problems caused by surface mining (Legg 1990). In the mid-1970s, the famous operational research expert Saaty (1980) proposed the AHP (Saaty 1980), which combines quantitative and qualitative analysis, to calculate the weights of the elements in a pairwise manner and ultimately evaluate MGEQ. Similarly, Zadeh (1965) created a fuzzy evaluation method based on the maximum membership and fuzzy transform principles to evaluate MGEQ (Zadeh 1965). In addition, various researchers have examined a number of other methods for evaluating MGEQ such as graphics overlay, scoring superposition, comprehensive index method (CIM), and Delphi methods (Ma 2013).

While significant achievements and progress in mining geo-environmental assessment have been made by previous studies, some questions remain to be resolved. First, the evaluation results presented thus far are not objective enough, because the selection of the evaluation factors has not been sufficiently comprehensive (Liang 2012). Second, previous works have focused on the qualitative research of mining geo-environments and there has been a lack of quantitative studies (Wang and Chen 2011). Third, unreasonable comprehensive evaluation methods with subjective factors and complex weight calculations have led to the evaluation results lacking comparability. Finally, the complex mapping between the evaluation factors and MGEQ has not been processed well. Therefore, it is of utmost importance to formulate a scientific, comprehensive, reliable, and practical quantitative evaluation system as well as a multi-index, multi-level evaluation model that can handle the complex mapping between the evaluation factors and MGEQ.

The present study establishes such a quantitative evaluation system for the CEZ based on the approaches of previous researchers and the characteristics of the mining geo-environment. Based on the CIM, which was developed to evaluate MGEQ, we propose a new model termed the support vector machine (SVM) model. Objective weighting methods (e.g., the variation coefficient, entropy, and Kantiray assignment methods) and subjective weighting methods (e.g., the AHP) are combined to determine the integrated weights of the evaluation indicators. The proposed SVM model is capable of mapping out the relationship between input and output variables by using a kernel function in a high-dimensional space (Liang 2009) as well as treating pattern recognition. SVM has been widely used in the fields of human face recognition, signal processing, the prediction of rainfall runoff (Dibike et al. 2001), water quality assessment, text classification, remote sensing image analysis (Vapnik 1995; Yoon et al. 2011), stream flow or stage (Liong and Sivapragasam 2002; Asefa et al. 2006; Yu et al. 2006), and lake water level assessment (Khalil et al. 2006; Khan and Coulibaly 2006). However, it has not thus far been applied in the field of MGEQ evaluation. Consequently, the present study applies the SVM model to assess MGEQ and explores the modeling process to evaluate its effects.

Specifically, the aims of the present study are fourfold: (1) to rationally construct a more holistic and scientific evaluation framework; (2) to compare the advantages and disadvantages of the CIM and SVM models for the MGEQ evaluation in the study area, where the geo-environment has been damaged over the past decade because of mining activities; (3) to provide some guidance and be of practical importance for the popularization and application of SVM technology in the field of MGEQ assessment; and (4) to

provide reliable data and a basis for the reasonable exploitation and sustainable utilization of mineral resource in the CEZ.

This research is novel in the following three ways. First, the evaluation framework constructed in this study is original. Second, we apply the SVM model to assess MGEQ and obtain good evaluation results. Third, the variation coefficient, entropy, and Kantiray weighting methods as well as the AHP are combined to determine the optimal weights, which not only reflect expert knowledge and opinions, but also use objective information.

Study area

The study area is located in the east of Jilin Province, Northeast China, which includes the cities of Changchun, Dehui, Jiutai, Jilin, and Jiaohe, the counties of Nongan and Yongji, and the Yanbian Korean Autonomous Prefecture. It lies between the geographic coordinates of 42°00'20"N and

44°56'24"N latitude, and 124°32'24"E and 131°18'36"E longitude (Fig. 1), with an area of 73,000 km². The study area is characterized by a temperate continental monsoon climate. Average annual temperature is 2–6 °C and average annual rainfall is 600–1,400 mm, with 80 % or more of the rainfall occurring between June and September. Mining has become a pillar industry of economic development in the Changjitu region, but it also causes complex and serious mining-related geological problems, which can be divided into three categories: mining-related geological disasters, resource damage, and environmental pollution.

Materials and methods

MGEQ evaluation system

The MGEQ evaluation system is a complex system involving many factors. Mining, dressing, smelting, and other types of engineering activities can induce and

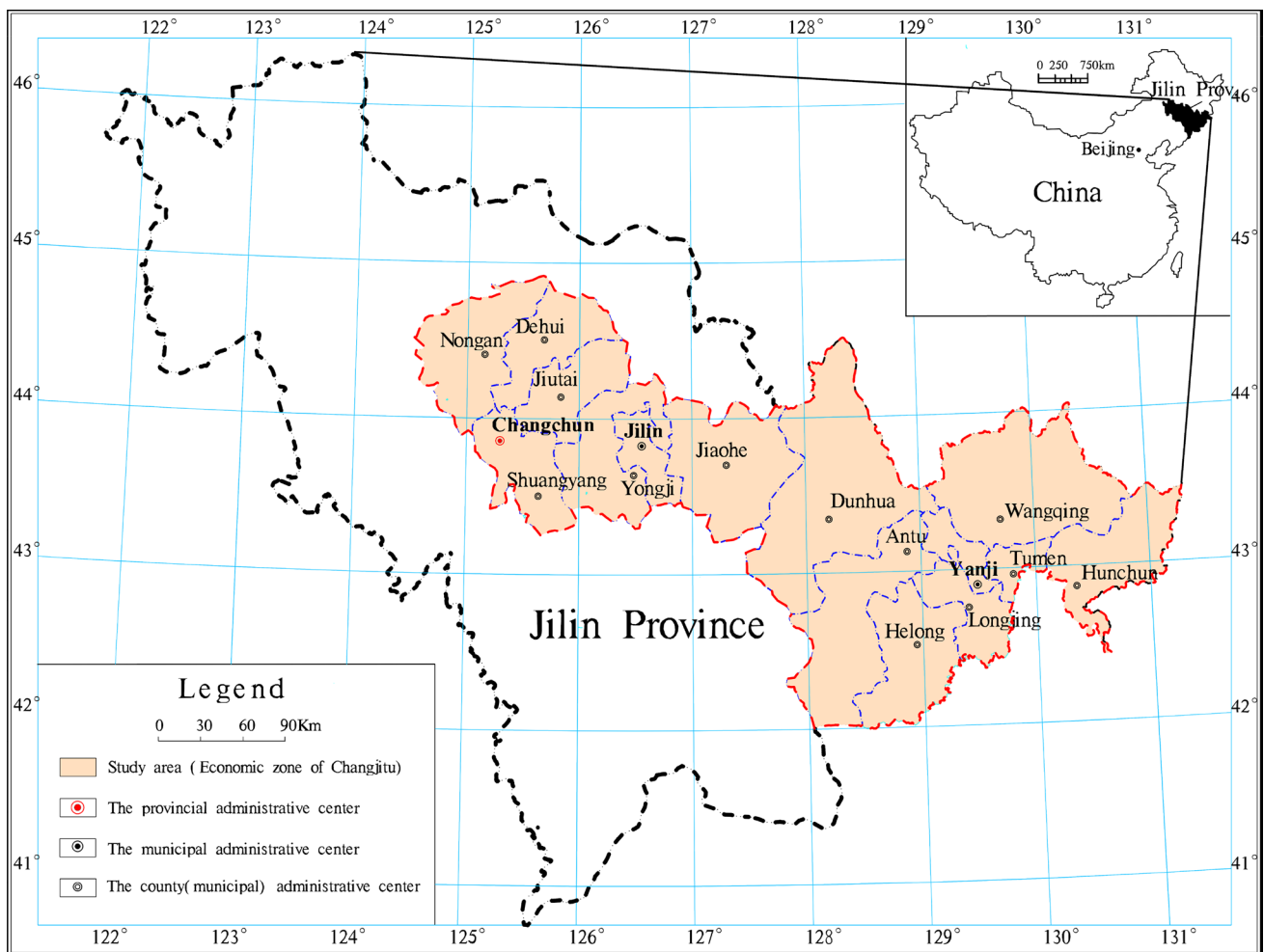


Fig. 1 Location of the study area

aggravate various types of geological problems at different levels, while geological background and natural conditions are additional factors that may lead to mining-related geological problems. Simultaneously, the difficulty of mining geo-environmental recovery and management is a key factor for MGEQ evaluation (Liu et al. 2008).

The MGEQ evaluation system proposed herein was established based on the comparative analysis of the major mining-related geological problems in the CEZ. The evaluation framework considers five criteria, namely basic mining information, the geo-environmental background conditions of the mining area, mining-related geological problems, the importance of the evaluation area, and the difficulty of mining geo-environmental recovery, along with 15 indicators such as production scale, mining methods, topography, geological conditions, geological disasters, resource damage, environmental pollution, average rainfall, land use, population density, important buildings, nature reserve, water resources, types of destroyed land, security deposit, and mining-related geological disasters (Fig. 2).

The determination of geological environment elements and indicator weights

Objective weighting methods, such as the variation coefficient, entropy, and Kantiray weighting methods, and subjective weighting methods such as the AHP were used to determine the comprehensive weights of the elements and indicators. The comprehensive weights not only include expert knowledge and opinions, but also use objective information. Therefore, this method reflects the diversity of the practical data, overcoming the shortcomings of using either objective or subjective weights, making subjective and objective weights consistent, and leading to evaluation results that are sound and scientific (Zhu 2012).

Variation coefficient method

The variation coefficient evaluation method can be used when each factor is relatively independent. Suppose there are n participating samples, with each sample described by p indicators (X_1, X_2, \dots, X_p). The mean (\bar{X}_i) and variance (S_i^2) are calculated by using Eqs. (1) and (2):

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^p X_{ji} \quad (1)$$

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^p (X_{ji} - \bar{X}_i)^2. \quad (2)$$

The variation coefficient of each indicator is

$$V_i = S_i / \bar{X}_i \quad i = 1, 2, \dots, p. \quad (3)$$

The weight of each indicator W_i is obtained after normalizing V_i :

$$W_i = V_i / \sum_{j=1}^p V_j \quad j = 1, 2, \dots, p. \quad (4)$$

Entropy method

The entropy method determines its weights based on the amount of information that the various indicators transmit to policymakers. The greater the difference in the evaluation indicators, the smaller is the entropy, while the weights scale is proportional to the information content of the indicators. The detailed steps are as follows:

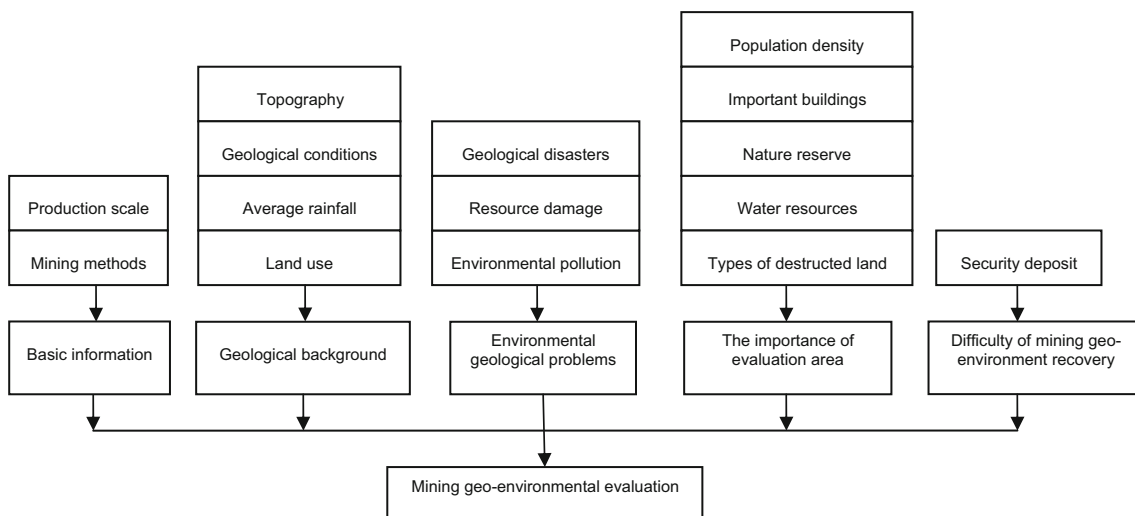


Fig. 2 The MGEQ evaluation system

Step 1. Normalize the value of the indicators:

$$a_{ij} = X_{ij} / \sum_{i=1}^n X_{ij} \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, m. \quad (5)$$

Step 2. Calculate the entropy of the evaluation indicators:

$$H = -K \sum_{i=1}^n a_{ij} \ln a_{ij} \quad (K = 1/\ln n). \quad (6)$$

Step 3. Transfer the entropy into the weight:

$$W_i = (1 - H_i) / (m - \sum_{j=1}^m H_j). \quad (7)$$

Although the entropy method requires a certain number of sample units, it has a strong mathematical theory and provides a close relationship between the entropy and indicator value itself (Ma 2009).

Kantiray weighting method

This approach was proposed by Kantiray (1989). The weight of each indicator can be derived using the following equation:

$$(RS - \lambda I)W = 0, \quad (8)$$

where R is the correlation matrix of the original variables, S is the diagonal matrix of the standard deviation, and λ is the largest eigenvalue. W is the eigenvector that corresponds to the maximum eigenvalue, which is affected by the standard deviation and correlation coefficient. Therefore, the matrix RS contains both the interaction information between the indicators in the original data and the degree of indicator variation (Kantiray 1989; Ma 2009).

AHP

The AHP has been proven to be an effective decision analysis method in multiple-criteria assessment. The present study uses the AHP to determine the subjective weights of the elements and indicators. First, a judgment matrix is used and then the weights of the indicators are obtained after calculating the eigenvalue and eigenvectors. Next, the normalized eigenvector that corresponds to the maximum eigenvalue-passed consistency check is taken as the weight of each index factor (Hao 2011). The equations describing the above-mentioned procedures are as follows:

$$AW = \lambda_{\max} W, \quad (9)$$

$$CI = \frac{\lambda_{\max} - N}{N - 1}, \quad (10)$$

$$CR = \frac{CI}{RI}, \quad (11)$$

where W is the eigenvector as well as the weight vector, λ_{\max} is the maximum eigenvalue of matrix A , CI is the consistency index, N is the order of the judgment matrix, RI is the average random index of the order of the matrix, and CR is the consistency ratio. CR should be lower than 0.1 to consider the eigenvector W to be an acceptable weight. Otherwise, the comparison and calculation should be redone.

Comprehensive weight method

First, an appropriate objective and subjective weighting method is chosen based on the samples and data and then comprehensive weights are calculated according to a certain ratio of the column (Wang 2013). The equation is written as follows:

$$w^* = \alpha w + (1 - \alpha)\phi, \quad (12)$$

where w^* is the comprehensive weight, w is the subjective weight, ϕ is the objective weight, and α is the subjective preference factor. In this study, the value of α is set as an empirical value of 0.5, while the comprehensive weight is the average of the subjective and objective weights.

Proposed evaluation model

The developmental characteristics, distribution, and extent of damage of mining-related geological problems are analyzed comprehensively based on the investigation of typical mine fields and data collection in the CEZ. Specifically, the CIM and SVM are used to evaluate MGEQ.

CIM

According to this method, we first determine the level of the mining geo-environment by using a weighted sum. This sum is calculated based on the status of the geological environment after assigning basic mining information, the geo-environmental background of the mining area, mining-related geological problems, the importance of the evaluation area, and the difficulty of mining geo-environmental recovery as well as the other 15 indicators. The grading standards for the MGEQ evaluation criteria are given in Table 1. The mathematical formulation is as follows (Liu et al. 2008):

$$F = \sum_{i=1}^n F_i W_i, \quad (13)$$

where F_i is the individual score for each indicator and W_i is the comprehensive weight of each indicator.

Mines that have a comprehensive quality index greater than or equal to 6.0 are classified into level III, between

Table 1 Indicator scores and ratings of the MGEQ evaluation criteria

Elements	Indicators assignment	Scores			
		2	6	10	
Basic information	Production scale	Small scale	Medium scale	Large scale	
	Mining methods	No mining	Open pit	Underground mining	
Geological background	Topography	Simple	Moderate	Complex	
	Geological conditions	Simple	Moderate	Complex	
	Average rainfall	Less than 400 mm	400–800 mm	Greater than 800 mm	
	Land use	Wasteland, barren land	Grassland, woodland, and construction sites	Arable land	
Environmental geological problems	Geological disasters	Lighter	Moderate	Serious	
	Resource damage	Destruction of land resource	Lighter	Moderate	Serious
		Destruction of water resource	Lighter	Moderate	Serious
		Destruction of landform landscape	Lighter	Moderate	Serious
	Environmental pollution	Lighter	Moderate	Serious	
The importance of evaluation area	Population density, important buildings, nature reserve, water source, and the types of destruction land	General	Moderate	Important	
Difficulty of mining geo-environment recovery	Security deposit	Less than 1 million	1–5million	More than 5 million	

Table 2 The influence degree of each class

Level	I	II	III
Influence degree	Lighter	Moderate	Serious

3.32 and 6.0 level II, and less than 3.32 level I. The influence degree corresponding to each class is shown in Table 2.

SVM model

(1) Theory of SVM

SVM (Vapnik 1995) is an emerging machine learning technology that has been extensively used as a classification tool. The theory is based on the structural risk minimization principle (Liang et al. 2011; Yoon et al. 2011). SVM attempts to find the optimal separating hyperplane between classes by maximizing the class margin (Harris 2013). SVM can be divided into linear and nonlinear models, with the former a special case of the latter. Therefore, this article only describes the nonlinear SVM.

Nonlinear problems can be transformed into linear problems by using a kernel function and finding the optimal separating hyperplane in the transform space. Because

the kernel function that corresponds to an inner product function is $K(x_i, x_j) = \psi(x_i) \times \psi(x_j)$, $K(x_i, x_j)$ can avoid this so-called “dimension disaster,” which primarily relates to the large number of possible nonlinear mappings and to the computational complexity associated with any high-dimensional space (Baly and Hajj 2012).

The method can be described as follows. Input vector x is first mapped to a high-dimensional feature space by using the nonlinear mapping pre-selected ϕ and then finding the optimal separating hyperplane in the high-dimensional space (Keerthi et al. 2000; Cristianini and Shawe-Taylor 2000; Sun et al. 2009; Wu and Wang 2009). The sample set is (x_i, y_i) , $i = 1, 2, \dots, n$; $y = \{1, -1\}$ are class labels. The hyperplane can thus be expressed by the following equation:

$$\omega \times x_i + b = 0. \quad (14)$$

Next, we add a slack variable to the constraints $\xi_i \geq 0$, so that the maximum interval hyperplane is now called the generalized optimal separating hyperplane (Leng et al. 2007; Baly and Hajj 2012; Harris 2013). The constraint becomes

$$\text{s.t } y_i[(\omega \times x_i) + b] \geq 1 - \xi_i. \quad (15)$$

The optimization problem is written as

$$\min \frac{1}{2}(\omega \times \omega) + C \sum_{i=1}^n \xi_i \tag{16}$$

$$\text{s.t } y_i[(\omega \times x_i) + b] \geq 1 - \xi_i \quad i = 1, 2, \dots, n,$$

where ω is the weight vector, b is the bias, and ξ_i is the slack variable. In addition, $C > 0$ is the penalty factor, where the larger the C , the greater is the penalty. To develop the dual form of the problem, we introduce the Lagrange multipliers α, β

$$L(\omega, \xi, b, \alpha, \beta) = \frac{1}{2}(\omega \times \omega) + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\omega \times x_i + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i. \tag{17}$$

Equation (17) is then minimized with respect to ω, ξ , and b by taking the partial derivatives with respect to ω, ξ , and b and setting them to zero. Hence, the dual problem is as follows:

$$\begin{aligned} \max_{\alpha, \beta} Q(\alpha) &= L(\omega, \xi, b, \alpha, \beta) \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\psi(x_i) \times \psi(x_j)) \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \end{aligned} \tag{18}$$

$$\text{s.t } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0.$$

The optimal judgment function is

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b). \tag{19}$$

Many types of kernels have been used to establish the SVM model, such as the linear, polynomial, Gaussian radial basis function (RBF), and sigmoid kernels (Chen et al. 2013). The respective equations are listed as follows:

Linear kernel,

$$K(x, x_i) = (x \times x_i). \tag{20}$$

Polynomial kernel,

$$K(x, x_i) = (g(x \times x_i) + r)^d. \tag{21}$$

RBF kernel,

$$K(x, x_i) = \exp(-g\|x - x_i\|^2). \tag{22}$$

Sigmoid kernel,

$$K(x, x_i) = \tanh(g(x \times x_i) + r). \tag{23}$$

(2) Construction, application, and validation of the SVM model

According to the classification standard of MGEQ in this study, random training samples were generated by using the rand function in MATLAB. There are three levels of MGEQ, and each of the elements can be evaluated by the following standard levels: less than or equal to 2.0 is level I, between 2.0 and 6.0 is level II, and between 6.0 and 10.0 is level III. We randomly generated 200 pairs of training samples under these three levels (i.e., 600 pairs of training samples), while the test samples were 116 groups of data on the important mines in the CEZ.

The SVM model was built as follows:

Step 1. Normalize the data (Eq. 24)

$$\tilde{X}_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}. \tag{24}$$

Step 2. Determine the structure of the SVM model. In this study, basic information (X_1), geological background (X_2), mining-related geological problems (X_3), the importance of the evaluation area (X_4), and the difficulty of mining geo-environmental recovery (X_5) are the input variables, with the level of MGEQ as the output variable.

Step 3. Determine the parameters. The present study uses the C-support vector classification model, while the penalty parameter C and kernel parameter g need to be set. To optimize the two SVM parameters (C and g) simultaneously, a cross-validation parameter optimization program was performed. First, training data were separated into several folds. Sequentially, a fold is considered to be the validation set and the rest are used for training. The average accuracy for predicting the validation sets represents the cross-validation accuracy. We provided a possible interval of $\log_2 C$ (or $\log_2 g$) with the grid space $(-10, 10)$. Then, all the grid points of $(C; g)$ were examined to assess which one provided the highest cross-validation accuracy. After training, the optimal penalty parameter C and optimal kernel parameter g were obtained with values of 1,024, and 64, respectively. Then, we used the best parameter to train the whole training set and generated the final model.

Step 4. Determine the kernel function and establish the evaluation model. By using the MATLAB program to train the random learning samples, the optimal penalty and optimal kernel parameters as well as the linear, polynomial, Gaussian RBF, and sigmoid kernels were developed to establish the SVM model to evaluate MGEQ. The 116 groups of test samples were then placed into the evaluation model to obtain the MGEQ evaluation results, and the evaluation results were compared by using these different kernel functions.

Results and discussions

Results of the MGEQ evaluation based on the CIM

Because each evaluation indicator is relatively independent, the Kantiray weighting method is unsuitable in this case; therefore, the variation coefficient and entropy methods were applied as the objective weighting methods and the AHP as the subjective weighting method. The five evaluation elements of the MGEQ were correlated in some areas, but also showed certain differences. The Kantiray weighting method was more appropriate than the variation coefficient method. Therefore, we selected the Kantiray and entropy weighting methods as the objective weighting methods and the AHP as the subjective weighting method to determine the comprehensive weights of the five elements. The weights of the elements and indicators are listed in Table 3. We finally obtained the evaluation results of CIM using Eq. 13.

The vector represented the relative importance of the assessment criteria for the geological environment: $A3 > A5 > A1 > A4 > A2$. The assessment for the indicators under level A1 (i.e., basic information) was $B2 > B1$, that for the indices under level A2 (i.e., geological background) was of the order of $B3 > B4 > B5 > B6$, and that under level A3 (i.e., mining-related geological problems) was of the order of $B7 > B9 > B8$. In the present study, we thus determined the optimal weights that not only reflect expert knowledge and opinions, but also use objective information.

Table 3 The comprehensive weights of the elements and indicators

Evaluation elements	Weights	Indicators	Weights
Basic information [A1]	0.1500	Production scale [B1]	0.4507
		Mining methods [B2]	0.5493
Geological background [A2]	0.0942	Topography [B3]	0.4602
		Geological conditions [B4]	0.2467
		Average rainfall [B5]	0.1608
		Land use [B6]	0.1323
Environmental geological problems [A3]	0.4263	Geological disasters [B7]	0.5091
		Resource damage [B8]	0.2244
		Environmental pollution [B9]	0.2665
The importance of evaluation area [A4]	0.0966	Population density, important buildings, nature reserve, water source, and types of destruction land [B10]	
Ease of mining geo-environment recovery [A5]	0.2329	Security deposit [B11]	

Results of the MGEQ evaluation based on the SVM model

The accuracy of the four SVM models with their different kernel functions is illustrated in Fig. 3. The selection of kernel function was significant for pattern recognition. Figure 3 also shows that the accuracy of the linear, polynomial, RBF, and sigmoid SVM models was 93.10, 92.24, 52.59, and 0 %, respectively, illustrating that the linear SVM model is slightly more accurate than the polynomial SVM model, with both well ahead of the RBF and sigmoid SVM models. Therefore, this finding proves that the most appropriate method for evaluating the mining geo-environment in the study area is the linear SVM.

Discussion

Conventional evaluation methods such as the CIM, fuzzy mathematics method, and gray clustering method cannot address the complex relationships between the evaluation elements and MGEQ, and the evaluation results are greatly affected by subjective factors (Liao and Wu 2004; Huang et al. 2012). Compared with conventional evaluation methods, SVM is a small-sample machine learning method based on statistical learning theory that uses the structural risk minimization principle. This approach allows generalization and can map the evaluation elements and MGEQ using a kernel function in a high-dimensional space. Hence, it can overcome the shortcoming of conventional methods when evaluating MGEQ as well as the defects of slow training speed, poor network generalization, and low learning accuracy in artificial neural networks (Liao et al. 2012). In addition, the SVM model calculates weights automatically and the evaluation results are comparable. Therefore, as an important pattern recognition method, SVM is appropriate for MGEQ evaluation, which is a typical pattern recognition issue.

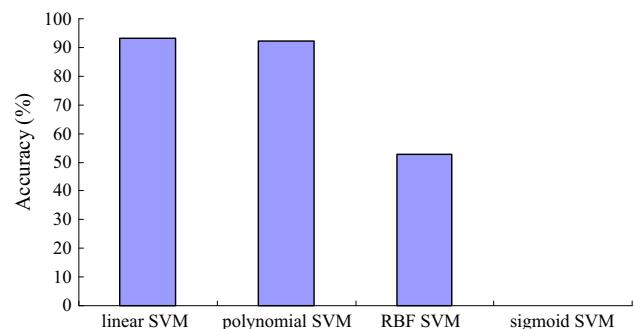


Fig. 3 The accuracy of linear SVM, polynomial SVM, RBF SVM, and sigmoid SVM

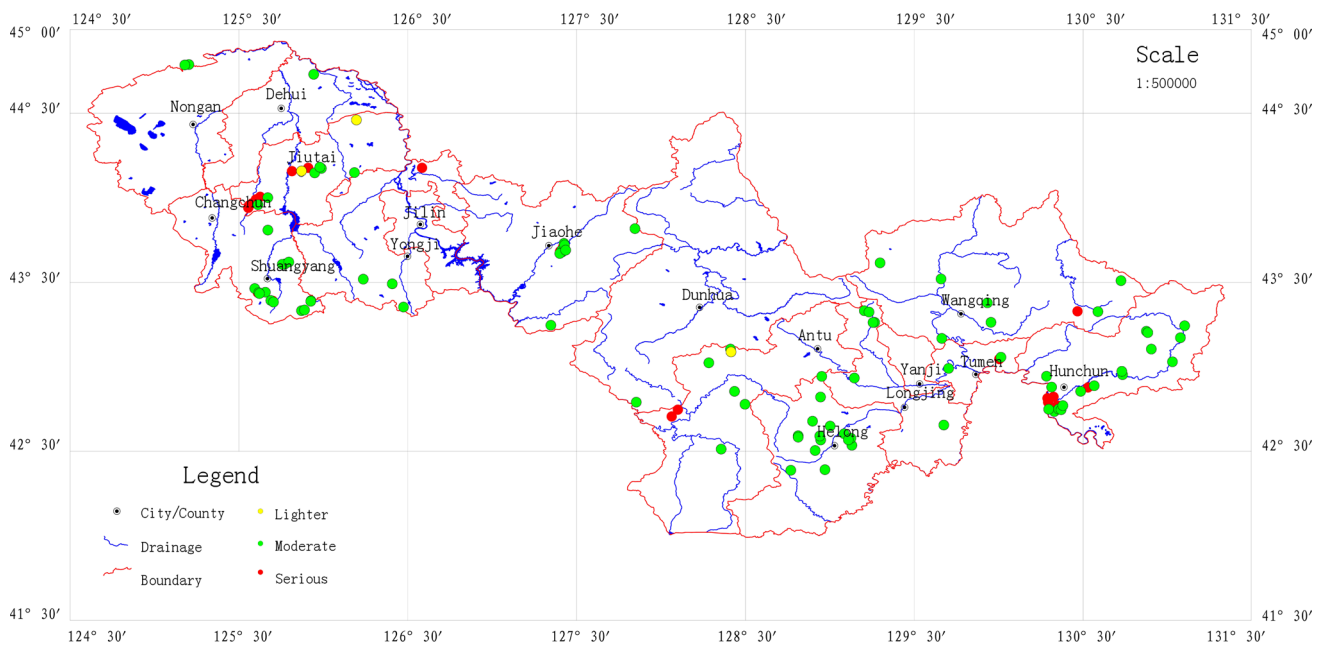


Fig. 4 The levels of the mining geo-environment in the CEZ

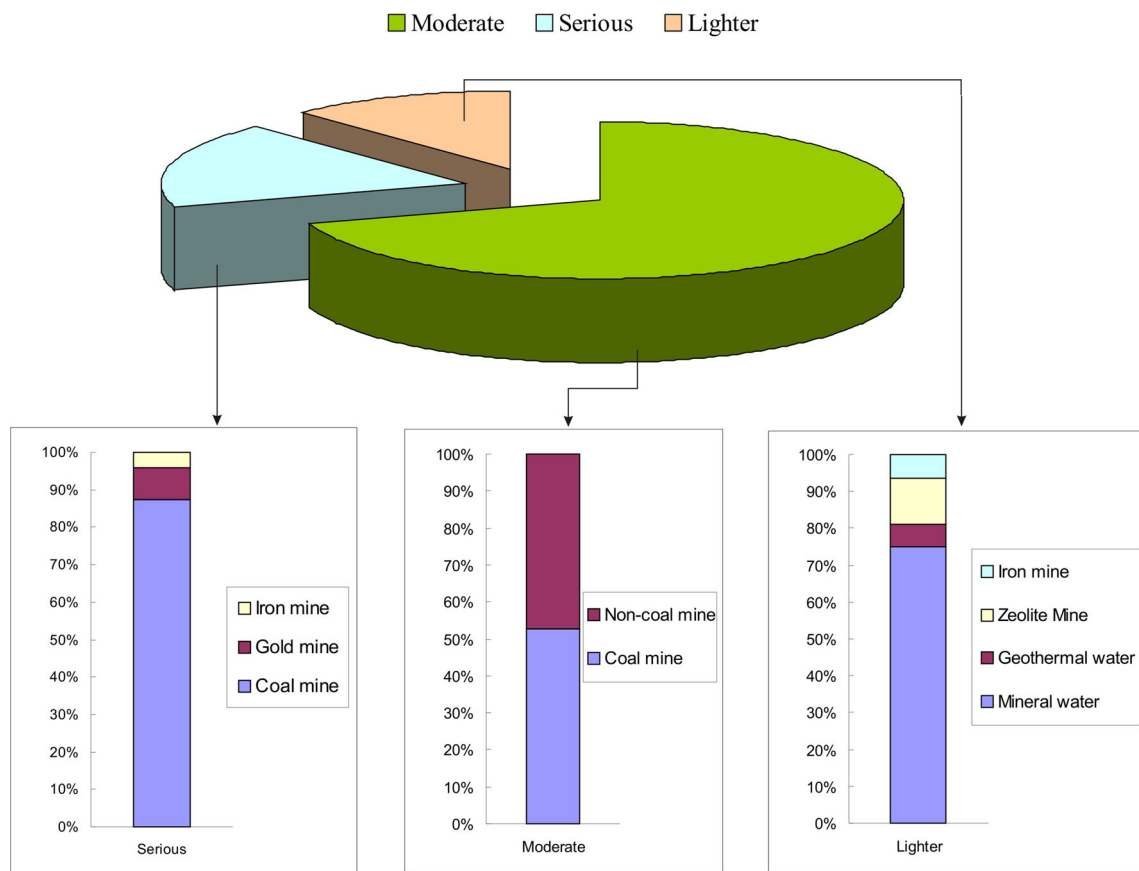


Fig. 5 The statistical results of the MGEQ evaluation

According to the CIM and SVM evaluation results presented herein, Figs. 4 and 5 show that the majority of the MGEQ in the study area ranks as level II, which

corresponds to moderately affected, accounting for 75.86 % of total eligible mines and 52.6 % of those located in coal mining areas. By contrast, level III (seriously

affected) accounts for 21.55 % (87.50 % in coal mining areas) and level I (lightly affected) 2.59 %, located in mineral water, zeolite, iron, and geothermal mines.

In summary, these results indicate that most mining geo-environments in the region are moderately or seriously affected by mining activities, especially coal mining areas (Fig. 5). The long-term exploitation of coal in the study area has created enormous mined-out areas, causing surface collapses and ground fissures. Indeed, there are more than 35 collapsed areas of different sizes, the largest up to 2.346 km². These surface collapses and ground fissures have damaged the mining geo-environment, threatening the safety of residents and destroying the local landscape. Therefore, related recovery and management measures should be implemented.

Conclusions

Based on the findings of domestic and foreign research and the characteristics of the mining geo-environment of the CEZ, a new MGEQ evaluation system was proposed in this study. The evaluation framework considered five criteria, namely basic mining information, the geo-environmental background conditions of the mining area, mining-related geological problems, the importance of the evaluation area, and the difficulty of mine geo-environmental recovery. The evaluation framework established in this study is thus a relatively comprehensive system for evaluating MGEQ in the CEZ.

CIM and SVM models based on different kernel functions were developed to evaluate MGEQ. The results showed that the accuracy of the linear, polynomial, RBF, and sigmoid SVM models was 93.10, 92.24, 52.59, and 0 %, respectively, confirming that the linear SVM is the most appropriate method for evaluating the mining geo-environment in the CEZ.

The study also demonstrated that the SVM model was appropriate for MGEQ assessment. Compared with existing MGEQ assessment methods, SVM adopts the structural risk minimization principle, which classifies quality into multiple groups. This research thus provides some guidance and is of practical importance for the popularization and application of SVM technology in the field of MGEQ assessment.

The evaluation results also showed that most of the MGEQ is classified into level II in the study area (75.86 % of total eligible mines), with levels I and III accounting for 2.59 and 21.55 %. This finding suggests that the mining geo-environment in the region is moderately affected by mining activities. The results also provide reliable data and a basis for the reasonable

exploitation and sustainable utilization of mineral resource in the CEZ.

Acknowledgments This research was supported by China Geological Survey Project (1212011140027, 12120114027401) and the National Natural Science Foundation of China (41372237). Scientific Research Foundation for Returned Scholars of Jilin University (419080500024) is also greatly appreciated.

References

- Asefa T, Kemblowski M, McKee M, Khalil A (2006) Multi-time scale stream flow predictions: the support vector machines approach. *J Hydrol* 318:7–16. doi:10.1016/j.jhydrol.2005.06.001
- Baly R, Hajj H (2012) Wafer classification using support vector machines. *IEEE Trans Semicond Manuf* 25:373–383
- Brown MT (2005) Landscape restoration following phosphate mining: 30 years of co-evolution of science industry and regulation. *Eco Eng* 24:309–329
- Cai HS, Zhou AG, Tang ZH (1998) Expert-analytic hierarchy weighting process in geological environmental quality assessment (in Chinese). *Earth Sci (J Chin Univ Geosci)* 23:299–302
- Chen M, Lu WX, Hou ZY, Huang H, Li P (2013) The assessment of groundwater quality based on support vector machine in western Jilin (in Chinese). *Water Saving Irrig*:29–33
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, New York
- Das BK (1999) Environmental pollution of Udaisager lake and impact of phosphate mine, Udaipur, Rajasthan, India. *Environ Geol* 38:244–248
- Dibike YB, Velickov S, Solomatine D, Abbott MB (2001) Model induction with support vector machines: introduction and applications. *J Comput Civ Eng* 15:208–216. doi:10.1061/(ASCE)0887-3801(2001)15:3(208)
- Fessler AJ, Moller G, Talcott PA, Exon JH (2003) Selenium toxicity in sheep grazing reclaimed phosphate mining sites. *Vet Human Tox* 45:294–298
- Geise G, LeGalley E, Krekeler MPS (2011) Mineralogical and geochemical investigations of silicate-rich mine waste from a kyanite mine in central Virginia: implications for mine waste recycling. *Environ Earth Sci* 62:185–196
- Ghose MK (2003) Indian small-scale mining with special emphasis on environmental management. *J Cleaner Prod* 11:159–165
- Hao HB (2011) Research of mine geologic environment evaluation and prevention and curing countermeasure of typical mining area in Panzhuhua of Sichuan province (in Chinese). Dissertation, Chengdu University of Technology
- Harris T (2013) Quantitative credit risk assessment using support vector machines: broad versus narrow default definitions. *Expert Syst Appl* 40:4404–4413. doi:10.1016/j.eswa.2013.01.044
- Howladar MF (2013) Coal mining impacts on water environs around the Barapukuria coal mining area, Dinajpur, Bangladesh. *Environ Earth Sci* 70:215–226. doi:10.1007/s12665-012-2117-x
- Huang SB, Li X, Wang YX (2012) A new model of geo-environmental impact assessment of mining: a multiple-criteria assessment method integrating Fuzzy-AHP with fuzzy synthetic ranking. *Environ Earth Sci* 66:275–284. doi:10.1007/s12665-011-1237-z
- Jiang X, Lu WX, Zhao HQ, Yang QC, Yang ZP (2014) Potential ecological risk assessment and prediction of soil heavy-metal pollution around coal gangue dump. *Nat Hazards Earth Syst Sci* 14:1611–1624. doi:10.5194/nhess-14-1599-2014

- Jordan G, Project JP (2009) Sustainable mineral resources management: from regional mineral resources exploration to spatial contamination risk assessment of mining. *Environ Geol* 58:153–169. doi:[10.1007/s00254-008-1502-y](https://doi.org/10.1007/s00254-008-1502-y)
- Kantiray A (1989) On the measurement of certain aspects of social development. *Soc Indic Res* 21:35–92
- Keerthi SS, Shevade SK, Bhatacharyya C, Murthy KR (2000) A fast iterative nearest point algorithm for support vector machine classifier design. *IEEE Trans Neural Netw* 11:124–136
- Khalil AF, McKee M, Kembrowski M, Asefa T, Bastidas L (2006) Multiobjective analysis of chaotic dynamic systems with sparse learning machines. *Adv Water Resour* 29:72–88. doi:[10.1016/j.advwatres.2005.05.011](https://doi.org/10.1016/j.advwatres.2005.05.011)
- Khan MS, Coulibaly P (2006) Application of support vector machine in lake water level prediction. *J Hydrol Eng* 11:199–205. doi:[10.1061/\(ASCE\)1084-0699\(2006\)11:3\(199\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:3(199))
- Krekeler MPS, Morton J, Lepp J, Tselepis CM, Samsonov M, Kearns LE (2008) Mineralogical and geochemical investigation of clay-rich mine tailings from a closed phosphate mine, Bartow Florida, USA. *Environ Geol* 55:123–147
- Krekeler MPS, Allen CS, Kearns LE, Maynard JB (2010) An investigation of aspects of mine waste from a kyanite mine, Central Virginia, USA. *Environ Earth Sci* 61:93–106
- Legg CA, Institution of Mining and Metallurgy (1990) Application of remote sensing to environmental aspects of surface mining operations in the UK. Remote sensing: an operational technology for the mining and petroleum industries. Kluwer Academic Publishers, London, pp 159–164
- Leng B, Qin Z, Li L (2007) Support vector machine active learning for 3D model retrieval. *J Zhejiang Univ Sci A* 8:1953–1961
- Li MJ (2013) Evaluation of the mine geological environment in Yangjiazhangzi District of Huludao (in Chinese). Dissertation, Jilin University
- Liang HX (2009) Support vector machine model research and application (in Chinese). Dissertation, Liaoning Normal University
- Liang W (2012) The research of mine geological environment assessment of Hegang based on RS and GIS (in Chinese). Dissertation, Jilin University
- Liang XC, Gong YB, Xiao D (2011) Novel method for water quality prediction based on multi-kernel weighted support vector machine (in Chinese). *J S Univ (Nat Sci Edit)* 41:14–17
- Liao GL, Wu C (2004) Application on integrated mine environmental assessment by fuzzy mathematics approach (in Chinese). *Environ Sci Trends* 29:15–17
- Liao Y, Xu JY, Wang ZW (2012) Application of biomonitoring and support vector machine in water quality assessment. *J Zhejiang Univ Sci B* 13:327–334. doi:[10.1631/jzus.B1100031](https://doi.org/10.1631/jzus.B1100031)
- Liong SY, Sivapragasam C (2002) Flood stage forecasting with support vector machines. *J Am Water Resour Assoc* 38:173–186. doi:[10.1111/j.1752-1688.2002.tb01544.x](https://doi.org/10.1111/j.1752-1688.2002.tb01544.x)
- Liu YC, Shen B, Chen JB, Chen BY (2008) Mineral development and environmental geology in southwest China (in Chinese). Geological Publishing House, Beijing
- Ma H (2009) Objective weighting method of comprehensive evaluation system (in Chinese). *Cooper Techn Sci*:50–51
- Ma LL (2013) The method of mine geological environment assessment based on GIS and RS in the different coal mine (in Chinese). Dissertation, China University of Geosciences
- Mars JC, Crowley JK (2003) Mapping mine wastes and analyzing areas affected by selenium-rich water runoff in southeast Idaho using AVIRIS imagery and digital elevation data. *Rem Sens Env* 84:422–436
- Mayes WM, Johnston D, Potter HAB, Jarvis AP (2009) A national strategy for identification, prioritisation and management of pollution from abandoned non-coal mine sites in England and Wales. I.: methodology development and initial results. *Sci Total Environ* 407:5435–5447. doi:[10.1016/j.scitotenv.2009.06.019](https://doi.org/10.1016/j.scitotenv.2009.06.019)
- Mejía-Navarro M, Wohl EE, Oaks SD (1994) Geological hazards, vulnerability, and risk assessment using GIS: model for Glenwood Springs, Colorado. *Geomorphology* 10:331–354
- Mendez L, Palacios E, Acosta B, Monsalvo-Spencer P, Alvarez-Castaneda T (2006) Heavy metals in the clam *Megapitaria squalida* collected from wild and phosphorite mine-impacted sites in Baja California, Mexico—considerations for human health effects. *Biol Trace Elem Res* 110:275–287
- Monjezi M, Shahriar K, Dehghani H, Namin FS (2009) Environmental impact assessment of open pit mining in Iran. *Environ Geol* 58:205–216. doi:[10.1007/s00254-008-1509-4](https://doi.org/10.1007/s00254-008-1509-4)
- Pierce RH, Wetzel DL, Esteves ED (2004) Charlotte harbor initiative: assessing the ecological health of Southwest Florida's Charlotte Harbor Estuary. *Ecotoxic* 13:275–284
- Ryser AL, Strawn DG, Marcus MA, Johnson-Maynard JL, Gunter ME, Moller G (2005) Micro-spectroscopic investigation of selenium-bearing minerals from the Western US phosphate resource area. *Geochem Trans* 6:1–11
- Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill International Book Company, New York
- Sarkar BC, Mahanta BN, Saikia K, Paul PR, Singh G (2007) Geo-environmental quality assessment in Jharia Coalfield, India, using multivariate statistics and geographic information system. *Environ Geol* 51:1177–1196. doi:[10.1007/s00254-006-0409-8](https://doi.org/10.1007/s00254-006-0409-8)
- Schellenbach WL, Krekeler MPS (2012) Mineralogical and geochemical investigations of pyrite-rich mine waste from a kyanite mine in central Virginia with comments on recycling. *Environ Earth Sci* 66:1295–1307
- Sun A, Lim EP, Liu Y (2009) On strategies for imbalanced text classification using SVM: a comparative study. *Decis Support Syst* 48:191–201. doi:[10.1016/j.dss.2009.07.011](https://doi.org/10.1016/j.dss.2009.07.011)
- Turer D, Nefeslioglu HA, Zorlu K, Gokceoglu C (2008) Assessment of geo-environmental problems of the Zonguldak province (NW Turkey). *Environ Geol* 55:1001–1014. doi:[10.1007/s00254-007-1049-3](https://doi.org/10.1007/s00254-007-1049-3)
- Vapnik VN (1995) The nature of statistical learning theory. Springer, Berlin
- Wang HW (2013) Reliability comprehensive evaluation research of Italy Rambaudi five-axis machining center (in Chinese). Dissertation, Jilin University
- Wang HQ, Chen L (2011) Evaluation of mine geological environment in Linglong-Laixi area (in Chinese). *Resour Ind* 13:72–76
- Wu KP, Wang SD (2009) Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. *Pattern Recogn* 42:710–717. doi:[10.1016/j.patcog.2008.08.030](https://doi.org/10.1016/j.patcog.2008.08.030)
- Xu YN, Yuan HC, He F, Chen SB, Zhang JH (2003) Comprehensive evaluation index system of the environmental geological problems of mines (in Chinese). *Geol Bull Chin* 22:829–832
- Xu YN, He F, Yuan HC, Zhang JH, Chen SB (2006) Survey and assessment on environmental geology problems of mine in northwest China (in Chinese). Geological Publishing House, Beijing
- Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396:128–138. doi:[10.1016/j.jhydrol.2010.11.002](https://doi.org/10.1016/j.jhydrol.2010.11.002)
- Yu PS, Chen ST, Chang IF (2006) Support vector regression for real-time flood stage forecasting. *J Hydrol* 328:704–716. doi:[10.1016/j.jhydrol.2006.01.021](https://doi.org/10.1016/j.jhydrol.2006.01.021)
- Zadeh LA (1965) Fuzzy sets. *Inf control* 8:338–353
- Zhu Y (2012) Research on optimization of combination weighting based on item discrimination (in Chinese). Dissertation, Nanjing University of Science and Technology