


Article

Mapping Mineral Prospectivity Using a Hybrid Genetic Algorithm–Support Vector Machine (GA–SVM) Model

Xishihui Du ¹ , Kefa Zhou ^{2,3,*}, Yao Cui ⁴, Jinlin Wang ^{2,3} and Shuguang Zhou ^{2,3}

¹ School of Transportation Engineering, Shenyang Jianzhu University, Shenyang 110168, China; daisy_duxi@126.com

² Xinjiang Research Center for Mineral Resources, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China; wangjinlin@ms.xjb.ac.cn (J.W.); zhoushuguang@ms.xjb.ac.cn (S.Z.)

³ Xinjiang Key Laboratory of Mineral Resources and Digital Geology, Urumqi 830011, China

⁴ British Columbia Geological Survey, Victoria, BC V8W 9N3, Canada; yao.cui@gov.bc.ca

* Correspondence: zhoukf@ms.xjb.ac.cn

Abstract: Machine learning (ML) as a powerful data-driven method is widely used for mineral prospectivity mapping. This study employs a hybrid of the genetic algorithm (GA) and support vector machine (SVM) model to map prospective areas for Au deposits in Karamay, northwest China. In the proposed method, GA is used as an adaptive optimization search method to optimize the SVM parameters that result in the best fitness. After obtaining evidence layers from geological and geochemical data, GA–SVM models trained using different training datasets were applied to discriminate between prospective and non-prospective areas for Au deposits, and to produce prospectivity maps for mineral exploration. The F1 score and spatial efficiency of classification were calculated to objectively evaluate the performance of each prospectivity model. The best model predicted 95.83% of the known Au deposits within prospective areas, occupying 35.68% of the study area. The results demonstrate the effectiveness of the GA–SVM model as a tool for mapping mineral prospectivity.

Keywords: support vector machine; genetic algorithm; mineral prospectivity mapping; Au deposits



Citation: Du, X.; Zhou, K.; Cui, Y.; Wang, J.; Zhou, S. Mapping Mineral Prospectivity Using a Hybrid Genetic Algorithm–Support Vector Machine (GA–SVM) Model. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 766. <https://doi.org/10.3390/ijgi10110766>

Academic Editor: Peter Blišťan

Received: 31 August 2021

Accepted: 6 November 2021

Published: 12 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mineral prospectivity mapping (MPM) is a critical step in mineral exploration and exploitation, as it reduces uncertainty and risk by narrowing the target area [1–3]. In MPM, multiple datasets (e.g., geological, geophysical, geochemical, and remote sensing data) are collected, analyzed, and integrated to delineate target areas most likely to contain mineral deposits of interest. To achieve this goal, a variety of MPM approaches have been proposed, which can be categorized into knowledge-driven and data-driven methods [4,5]. (1) Knowledge-driven MPM methods use expert knowledge to qualitatively assess the importance of each evidence layer for known deposits of the type sought. Index overlay [6,7], fuzzy logic [8–10], and multiple-criteria decision-making methods [11–13] are examples of knowledge-driven MPM methods, which are used in frontier or less-explored areas (so-called “greenfields”) with no or very few known mineral deposits of the desired type. (2) Data-driven MPM methods analyze and quantify spatial associations between each evidence layer and the locations of known deposits that share a common genesis, and include weights of evidence [14,15], evidence belief functions [16,17], and logistic regression [18,19]. These methods are commonly applied in well-explored areas with sufficient known mineral deposits of the desired type.

Over the last decade, some machine learning methods as data-driven methods have been developed for MPM. These include support vector machine (SVM) [20–22], a margin-based classifier based on small sample learning that has good generalization capabilities [23] and is an effective tool to model the complex nonlinear relationships between

evidence layers and mineral occurrences. However, in standard SVM, classification performance is heavily dependent on parameter selection (hyper-parameters and kernel parameters) in cases with no criteria or principles to follow when setting SVM parameters. Genetic algorithm (GA) [24] is a well-known and widely used method for variable selection [25,26]. GA provides a search technique that solves optimization problems by employing simulated evolution via “survival of the fittest” using various genetic functions. Therefore, an SVM classifier incorporating GA for parameter optimization has great potential in MPM, making full use of the unique merits of these two data-mining approaches.

With respect to training data, SVM, as a supervised algorithm, is different from the traditional data-driven methods used in MPM (e.g., weights of evidence), which usually require both mineralized and non-mineralized training datasets. Because an optimal separating hyperplane between the mineralized and non-mineralized locations is affected by both the mineralized and non-mineralized training datasets, learning bias can be caused by imbalanced training datasets, increasing misclassification [27]. Consequently, balancing the number of mineralized and non-mineralized training datasets is an efficient way to obtain more reliable classifications [28]. However, the selection of non-mineralized samples is challenging, as it is not possible to identify whether all non-mineralized samples are truly non-mineralized, because of the complexity of geological conditions [29]. In this context, Carranza et al. [30] summarized four criteria for the selection of non-mineralized samples: (1) non-mineralized samples must be randomly distributed in the study area; (2) non-mineralized samples should be distal to any known mineralized samples; (3) non-mineralized samples must have values for all the univariate geoscience spatial data; and (4) the number of non-mineralized samples must be equal to the number of mineralized samples. In addition, point pattern analysis was applied to evaluate the spatial pattern of non-mineralized samples and determine the optimal distance between the mineralized and non-mineralized samples. Nykänen et al. [31] proposed that other types of known deposits can be used as non-mineralized samples in well-explored areas, whereas random locations that are geologically constrained could represent non-mineralized samples in greenfields. In recent years, various sampling techniques, such as undersampling and oversampling, have been used to select non-mineralized samples [1,32,33]. Prado et al. [33] used the synthetic minority over-sampling technique and random under-sampling to create 400 training datasets with proportions of mineralized-to-non-mineralized samples ranging from 600:30 to 30:600.

In this study, SVM and GA were combined to optimize parameter design and develop a predictive model for mapping Au prospectivity zones in Karamay, NW China. For this purpose, after constructing five evidence layers from geological and geochemical data using spatial data processing methods and a prediction-area (P-A) plot [34,35], point pattern analysis was employed to estimate and randomly select non-mineralized samples based on the selection criteria. Subsequently, GA–SVM models trained using different training datasets were employed to delineate target areas and generate binary prospectivity maps. Ultimately, the F1 score and spatial efficiency were compared between different prospectivity models to evaluate their performance.

2. Methodology

2.1. Support Vector Machine

Support vector machine (SVM), introduced by Vapnik [23] and proposed for classification and regression tasks, is a novel type of machine learning method. SVM is constructed on the Vapnik–Chervonenkis dimension theory and the structural risk minimization principle. In essence, it employs a nonlinear transformation of the inner product function definition to transform the input space into a high-dimensional space, where it finds the optimal linear separating hyperplane. A detailed description of SVM can be found in Cristianini and Shawe-Taylor [36]. Here, a brief summary of SVM is provided.

Given a training set of instance–label pairs (x_i, y_i) , $i = 1, 2, \dots, n$, where $x_i \in R^d$ and $y_i \in \{+1, -1\}$:

In linearly separable cases, the following optimization problems need to be solved to find an optimal separating hyperplane:

$$\text{Minimize : } \frac{1}{2} \|\omega\|^2 \quad \text{Subject to : } y_i(\omega \cdot x_i + b) \geq 1, i = 1, 2, \dots, n \quad (1)$$

where ω is a vector normal to the hyperplane, and b is a scalar quantity.

The Lagrangian multipliers method was introduced to solve the aforementioned problem and obtain a classifying determination function:

$$f(x) = \text{sign}[(\omega \cdot x_i) + b] \quad (2)$$

In linear non-separable cases, a non-negative slack variable $\xi_i \geq 0, i = 1, 2, \dots, n$, was introduced, and the equation to be solved became:

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \\ \text{Subject to : } & y_i(\omega \cdot x_i + b) + \xi_i \geq 1, i = 1, 2, \dots, n \end{aligned} \quad (3)$$

where C is the penalty parameter, which has an important effect on the accuracy of the SVM classifier. This should be predetermined by the user. Similar to the linearly separable cases, this optimization model can be solved using the Lagrangian multipliers method.

In nonlinear separable cases, the input features are mapped into a new high-dimensional feature space using a kernel function $K(x_i, x_j)$, transforming it into a linearly separable case. Several kernel functions, including the polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel, are popular. This study used the RBF kernel function (Equation (4)), which is an effective kernel function with fewer parameters and provides excellent overall classification performance:

$$K(x_i, x_j) = \exp\left\{-\|x_i - x_j\|^2 / 2\sigma^2\right\} \quad (4)$$

where σ is the kernel parameter, which is always greater than zero.

2.2. GA-SVM Model

Genetic algorithm (GA) was first introduced by Holland [24] as an adaptive optimization technique based on the Darwinian evolutionary hypothesis of natural selection. The aim of GA is to find optimum solutions within the potential areas by defining a fitness function and applying the biological processes of natural selection, crossover, and mutation to individuals in the population. Compared to traditional algorithms, GA can handle large search spaces efficiently and is less prone to converging on a locally optimal solution. Recently, GA has been progressively developed in conjunction with other techniques and has been applied to many optimization problems [37–39].

Therefore, GA is used to optimize the SVM parameters σ and C based on the process of natural selection, in which *Accuracy* is adopted as the fitness function to evaluate the quality of the solutions. For a two-class mapping of the mineral prospectivity problem, the classified results can be represented as a confusion matrix (Table 1), which was defined using Equation (5). To determine the optimal parameters for SVM model, k-fold cross-validation [40,41] was used to construct a series of independent test datasets for the GA-SVM model, which was trained with the remaining $k - 1$ subsets of the training dataset, to search for the best fitness; $k = 5$ could achieve an adequate balance between the reliability of the calculation time and parameter estimation [42]. After repeating the cross-validation process, the fitness was obtained by calculating the accuracy of each test dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Here, true positives (TP) and true negatives (TN) are the numbers of known mineralized samples and known non-mineralized samples, respectively; FP (false positives) and FN (false negatives) are the numbers of predicted mineralized samples and predicted non-mineralized samples, respectively.

Table 1. Confusion matrix.

Prediction Known	‘Mineralized’	‘Non-Mineralized’
‘Mineralized’	TP	FN
‘Non-mineralized’	FP	TN
Total	$TP+FP$	$FN+TN$

The proposed GA–SVM model was employed to extract the optimal combined parameters of SVM to distinguish between prospective and non-prospective areas. The procedure involved in the GA–SVM model for MPM is divided into three parts (Figure 1), as follows:

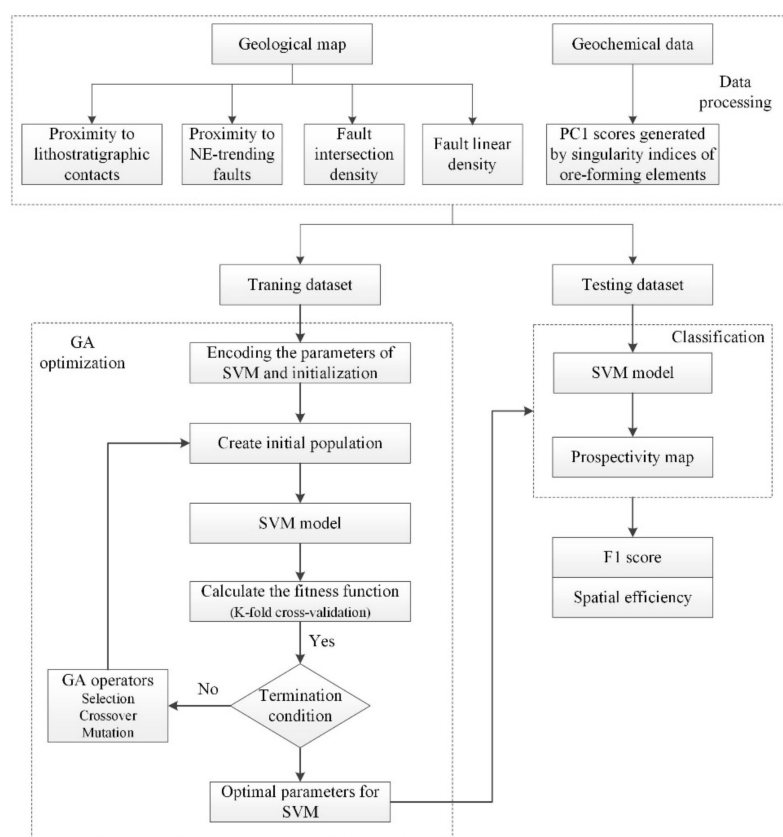


Figure 1. The procedure involved in the GA–SVM model for MPM.

- **Data processing.** After constructing a geospatial database, geological maps and geochemical data were analyzed to map five evidence layers and generate training and testing datasets.
- **GA optimization.** After setting the initial parameters for GA and SVM, the training dataset was used to train an SVM model, while the fitness was calculated by k-fold cross-validation classification accuracy. If the termination conditions were satisfied, the optimal parameters of SVM were determined. Otherwise, the selection, crossover, and mutation operations were performed to create a new population, and the GA optimization process was repeated.

- **Classification.** An SVM model was trained with the optimal parameters, and a prospectivity map was produced. Ultimately, the F1 score and spatial efficiency were combined to evaluate the classification ability of the GA-SVM model.

2.3. Performance Evaluation

In MPM, although evaluating the prospectivity model's ability to identify mineralized and non-mineralized locations is equally important, evaluating mineralized locations often has greater significance for the following reasons: (1) MPM aims to identify and distinguish mineralized locations; (2) a mineralized location misclassified as a non-mineralized location can result in the loss of important mineralization and incur high costs [43]; and (3) non-mineralized locations always introduce uncertainty, which is often randomly selected. Therefore, the F1 score [44], which is the harmonic mean of precision and recall, was used to measure the ability of the prospectivity model to identify mineralized locations:

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Here, precision represents the probability of known deposits being correctly classified as deposits, and recall represents the probability of known deposits in the total number of classified deposits, as follows:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{TP}{TP+FN} \end{aligned} \quad (7)$$

The relevant parameters are mentioned in Table 1 and Equation (5).

3. Study Area and Evidence Layers

3.1. Study Area

The study area is located in the eastern part of the Tangbale–Hatu belt (western Junggar region, China), part of the eastern extension of the Balkhash–Junggar metallogenic domain, in which a number of Au deposits were discovered, including Hatu, Qi-II, Qi-III, Qi-IV, and Qi-V. The study area covers approximately 11,784 km². The region is characterized by ophiolitic mélanges, Carboniferous volcanic-sedimentary rocks, and granitoid intrusions (Figure 2). Several ophiolitic mélanges show contact relationships with the Lower Carboniferous volcanic-sedimentary strata via faults [45]. Extensive Carboniferous volcanic-sedimentary rock outcrops, which are mostly distributed on both sides of the Darabut, occur in three successive formations: the Tailegula, Baogutu, and Xibeikulasi formations (ordered from bottom to top) [46]. The regional structure is dominated by a series of NE-trending faults. The larger NE-trending faults, namely, the Darabut, Anqi, and Hatu, constitute the basic framework of the region. Distributed between these large faults are various-sized granite bodies, including the Akebasitao, Hatu, Miaoergou, and Karamay plutons, formed 300–280 Ma [47–49]. These granite plutons provided a favorable tectonic environment for Au mineralization.

Based on the zonal distribution or clustered occurrence of metallogenesis, the study area can be roughly divided into two metallogenic belts: the Hatu metallogenic belt and Baogutu metallogenic belt [50,51]. The Hatu metallogenic belt, located north of the Darabut fault, is an extremely important Au-producing base. The Au occurrences in this belt are primarily controlled by the fault basin formed by the north-east-oriented Anqi fault and Hatu fault. Currently, Au deposits with large proven reserves are all located in this metallogenic belt and the Hatu Au deposit is a representative large-scale deposit. The Baogutu metallogenic belt is located in the south of the Baogutu fault. The Kuogashaye Au deposit is representative of its medium-sized Au deposits. The veins in this belt are primarily distributed around small intermediate-acid intrusive and intermediate dike rocks and essentially follow the direction of the intermediate dike rock group's distribution [50]. Thus, based on the synthetic studies of the typical Hatu and Kuogashaye Au deposits,

a conceptual model for Au mineralization in the study area was proposed, as shown in Table 2.

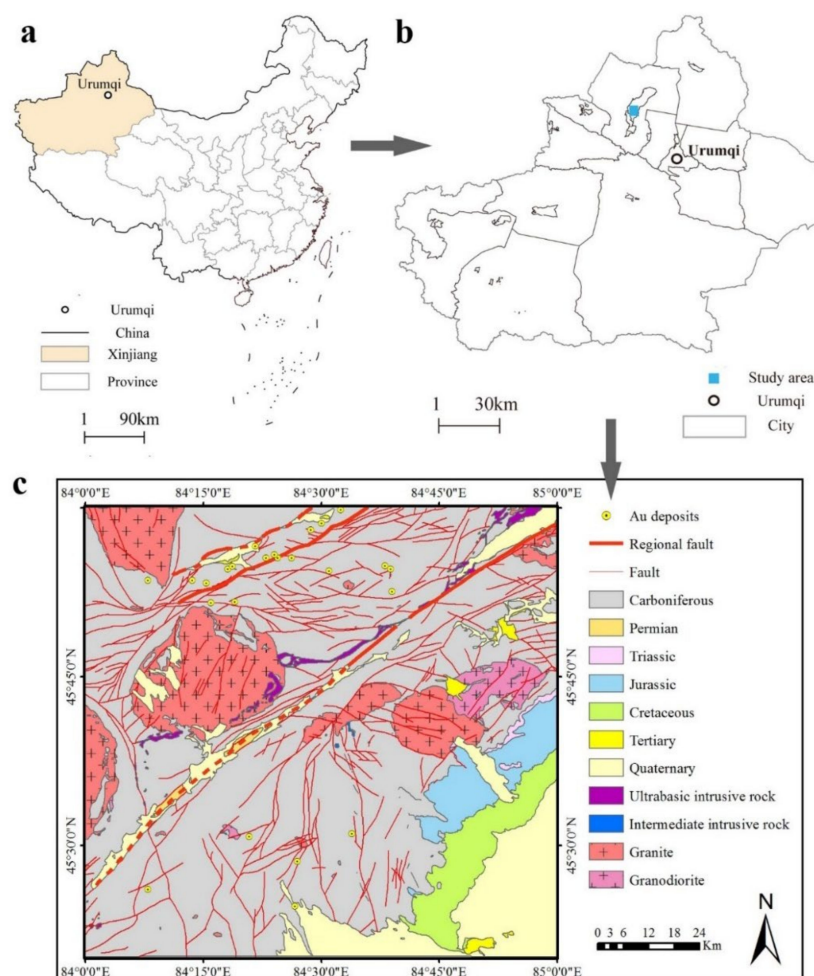


Figure 2. Spatial context of the study area: (a) location of Xinjiang in China; (b) location of the study area in Xinjiang; (c) schematic geological map of the study area (modified after 1:200,000 geological map).

Table 2. Conceptual Model of Au Deposits in the Study Area.

Metallogenic Factor		Description
Regional geological background	Tectonic environment	The north-east fault is the main tectonic line in the region. The crustal uplift and depression transitional zone to the north of the Darabut fault shows evidence of intense magmatic and volcanic activities and is the main ore-forming material source of Au deposits.
	Intrusive rocks	Intermediate-acid intrusive rocks are closely spatially related to mineral deposits.
	Ore-bearing strata	The vast majority of Au deposits are located in the Tailegula and Baogutu Formations of the upper Carboniferous in the Paleozoic.
	Ore-forming epoch	Middle and late Variscan age
	Wallrock alteration	Common forms of wallrock alteration include silicification, pyritization, arsenopyritization, and sericitization.
Regional geochemical field		The geochemistry of this region is dominated by Au anomalies. High concentrations of Au exist distributed between Toli and Karamay, with clear concentration centers and zoning.

3.2. Mapping Evidence

3.2.1. Data

In this study, a spatial dataset was derived from established multi-source geological spatial databases containing geological and geochemical data. Geological maps at a scale of 1:200,000 were collected from the Bureau of Geology and Mineral Resources of Xinjiang. Stream sediment geochemical data at a scale of 1:200,000 were obtained from the National Geochemical Mapping Project of China, with a sampling density of 1 per 4 km² [52].

3.2.2. Evidence Layers

The selection of evidence layers requires consideration of the characteristics of Au deposits, favorable conditions for Au mineralization, and the available data in the study area. Five evidence layers were used to produce a potential map, as shown in Table 3.

Table 3. Summary of Evidence Layers Used in this Study.

Criteria	Evidence Layer	Relevance
Geology	Proximity to lithostratigraphic contacts	The ore-forming elements migrate to the lithostratigraphic contacts and accumulate, resulting in precipitation, enrichment, and mineralization.
	Proximity to NE-trending faults	The region's main tectonic line runs NE and provides the driving force, the migration channel, and the depositing space for the mineral flow.
	Fault intersection density	Fault density reflects the location of frequent magma and hydrothermal activity, and the frequent superimposition of ore-forming elements.
	Fault linear density	
Geochemistry	PC1 scores generated by singularity indices of ore-forming elements	Ag, As, Au, and Sb are present in high concentrations above ore bodies. These elements can be used to differentiate provenance characteristics, understand the migration and evolution patterns of elements, and distinguish geochemical anomalies.

In this area, Au deposits are mainly hosted in Carboniferous formations or in contact with granitoid intrusions. The contacts of the granitoid intrusions and the stratigraphic units between the Tailegula, Xibeikulasi, and Baogutu Formations were extracted from a 1:200,000 geological map of the study area, and a map of proximity to lithostratigraphic contacts was produced using Euclidean distance in the ArcGIS environment (Figure 3a). Similarly, a map of proximity to NE-trending faults was generated (Figure 3b), because NE-trending faults play a dominant ore-controlling role during Au mineralization. Fault density, consisting of fault intersection density and fault linear density, reveals the spatial relationship between the development and accumulation levels of faults and Au deposits. Here, the fault intersection density and fault linear density were analyzed, and the corresponding maps were generated using point density and line density in the ArcGIS environment, respectively (Figure 3c,d).

The element contents of Ag, As, Au, and Sb obtained from 1:200,000 stream sediment geochemical data were analyzed using the singularity mapping technique [53] and principal component analysis (PCA) [54,55]. The singularity mapping technique was implemented to identify local anomalies of Ag, As, Au, and Sb from geochemical background fields based on sliding windows in MATLAB. To highlight the inherent relevance of multiple elements and reduce the uncertainty of each single element, PCA was used to integrate multi-element singularity indices, based on their correlations to delineate comprehensive anomalous areas. As shown in Figure 4a, the first principal component (PC1) accounted for 40.7% of the overall variance, whereas PC2, PC3, and PC4 modeled an additional 23.1%, 21.0%, and 15.2% of the total variance, respectively, indicating the importance of each component. In addition, the resulting PC1 shows positive loadings on the singular association of Ag, As, Au, and Sb, which is consistent with the characteristics of the geochemical anomalies (Figure 4b). Accordingly, PC1 can be used as an evidence layer

for Au deposits. As shown in Figure 4c, low PC1 scores have a strong spatial association with most of the known Au deposits. More details on the analytical methods and data processing can be found in Zhou et al. [56].

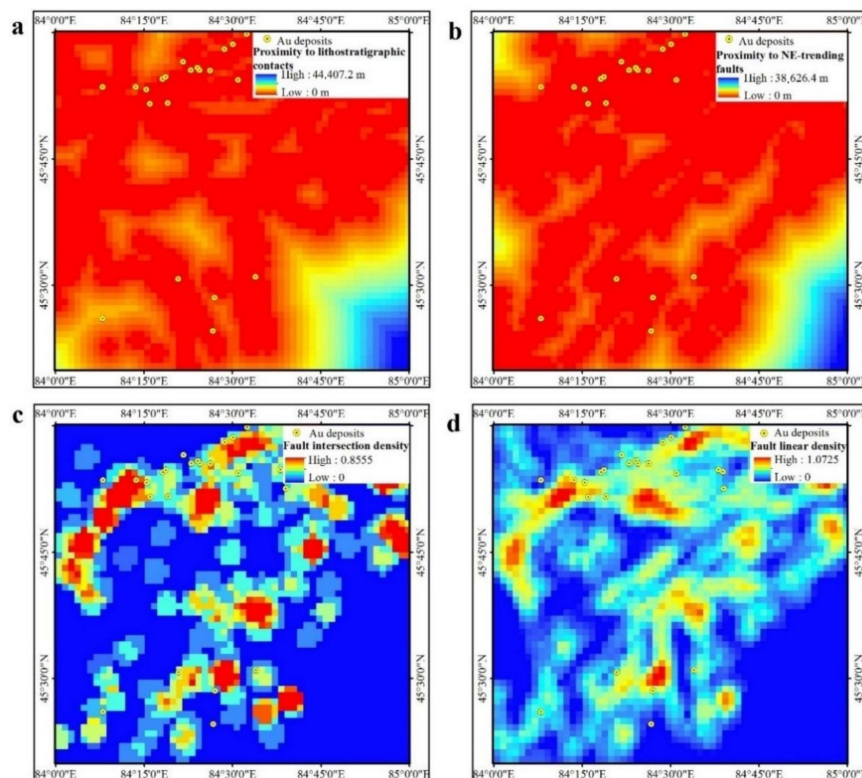


Figure 3. Derived geo-evidential layers: (a) proximity to lithostratigraphic contacts; (b) proximity to NE-trending faults; (c) fault intersection density; (d) fault linear density.

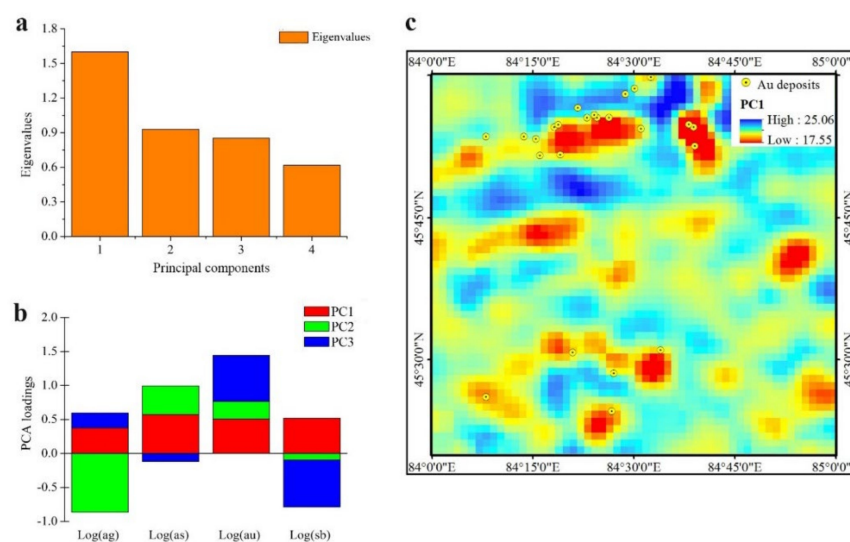


Figure 4. PCA results for the geochemical dataset: (a) Scree plot of the principal component (PC1–PC4) eigenvalues for singularity indices of ore-forming elements; (b) loadings on principal components (PC1–PC3); (c) PC1 scores generated by singularity indices of ore-forming elements.

4. Application of GA-SVM Model

4.1. Target Variable and Feature Vectors

The application of the GA-SVM model for MPM requires a training dataset with geological feature vectors of five evidence layers and a target variable to represent mineral prospectivity. The target variable expresses mineralized locations or non-mineralized locations with scores of 1 and 0, respectively. For mineralized locations, we used 24 known Au deposits to ensure classification accuracy. For non-mineralized locations, we used point pattern analysis [57,58] to analyze the nearest-neighbor distances between every pair of deposits within the study area to determine the optimal distance from known deposits at which the probability of finding a deposit decreased. In this study, most of the nearest-neighbor distances were less than 14.5 km, and there was only one outlier. Thus, 14.5 km is regarded as the optimal distance for the selection of non-mineralized samples. In addition, to obtain a balanced dataset, the number of non-mineralized samples should be the same as the number of known Au deposits. To this end, we created four training datasets, each consisting of 24 randomly selected non-mineralized samples, according to the selection criteria used by Carranza et al. [30]. Figure 5 shows the spatial distribution of the four training datasets.

The feature vector is a multidimensional numeric vector representing a combination of the attributes of evidence layers in a specific location. In this study, the attributes of the five evidence layers were encoded as either 1 or 0, where 1 and 0 indicate favorable and unfavorable conditions for Au mineralization, respectively. Consequently, to obtain binary patterns for the evidence layers used in the GA-SVM model, it was necessary to define the optimum threshold for classifying the maps. The P-A plot, which is a simple prediction rate-occupied area plot [34,35], was employed to determine the optimum thresholds with respect to the evidence layers. When the intersection point of two curves is high in a P-A plot, it portrays a small area containing a large number of mineral deposits. In this study, the P-A plot consisted of the curve of the percentage of known mineral occurrences corresponding to the classes of the evidence layer and the curve of the percentage of occupied areas corresponding to the classes of the evidence layer. Therefore, the location of the intersection point in the P-A plot could guide us in finding the binary pattern of evidence layers for Au mineralization. Figure 6 shows that (1) the optimum distances between the location of Au deposits and lithostratigraphic contacts and NE-trending faults are 1028.63 and 1108.10 m, respectively; (2) the optimum densities between the location of Au deposits and fault intersection density and fault linear density are 0.10 and 0.35, respectively; and (3) the optimum cutoff value between the location of Au deposits and PC1 scores was 22.1. According to the above optimal values and spatial associations between each evidence layer and the known Au deposits, all the evidence layers were encoded and combined to generate 2968 feature vectors.

4.2. Mineral Prospectivity Mapping

The GA-SVM model implemented in this study was programmed using the LIBSVM package [59] as a supplementary tool in MATLAB. The GA-SVM model with the parameters shown in Table 4 was used to search for the kernel parameter σ of RBF and the penalty parameter C of SVM (Figure 7), and the best fitness values corresponding to the optimal parameters of SVM were obtained (Table 5). Prospectivity models were established using these optimal parameters to determine the spatial associations between evidence layers and mineralized locations and produce prospective maps for mineral exploration (Figure 8).

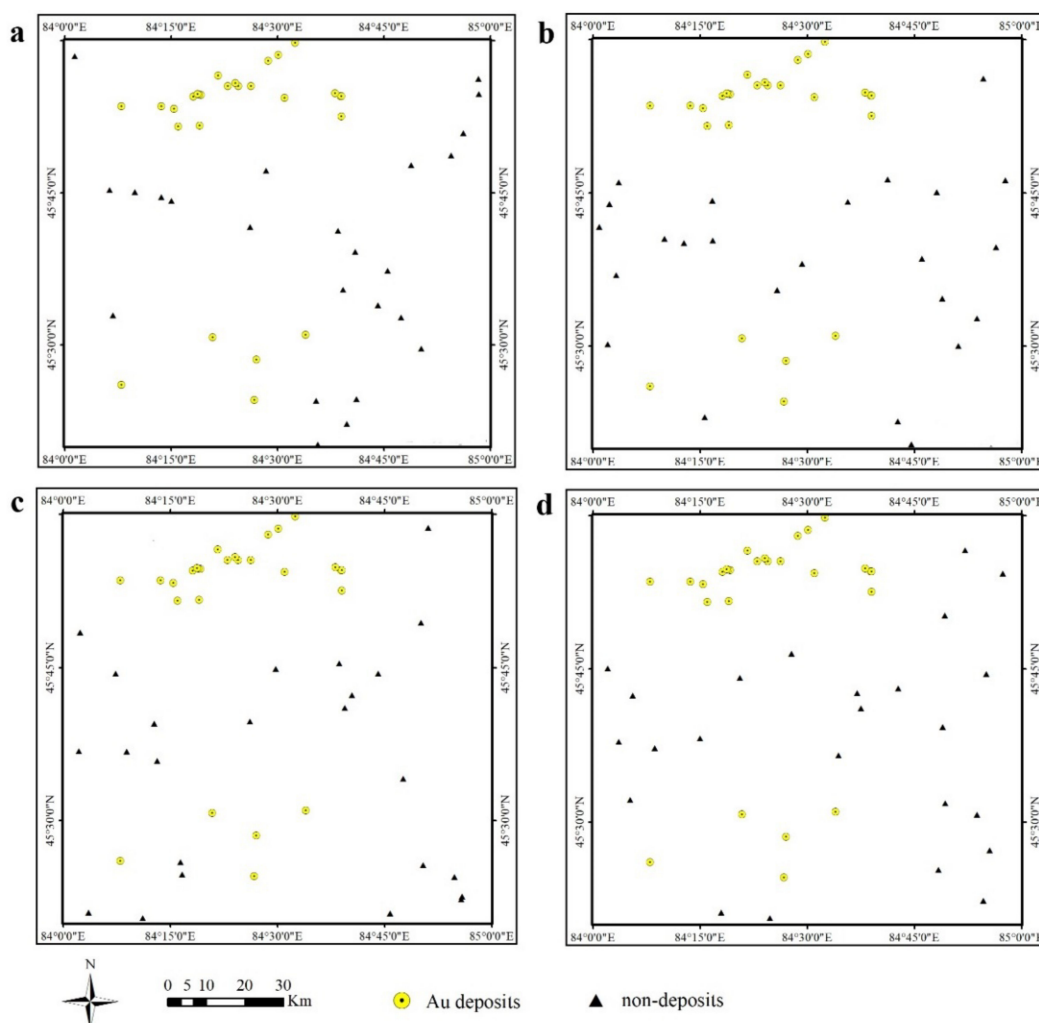


Figure 5. Maps show the locations of four training datasets: (a) training dataset 1; (b) training dataset 2; (c) training dataset 3; (d) training dataset 4.

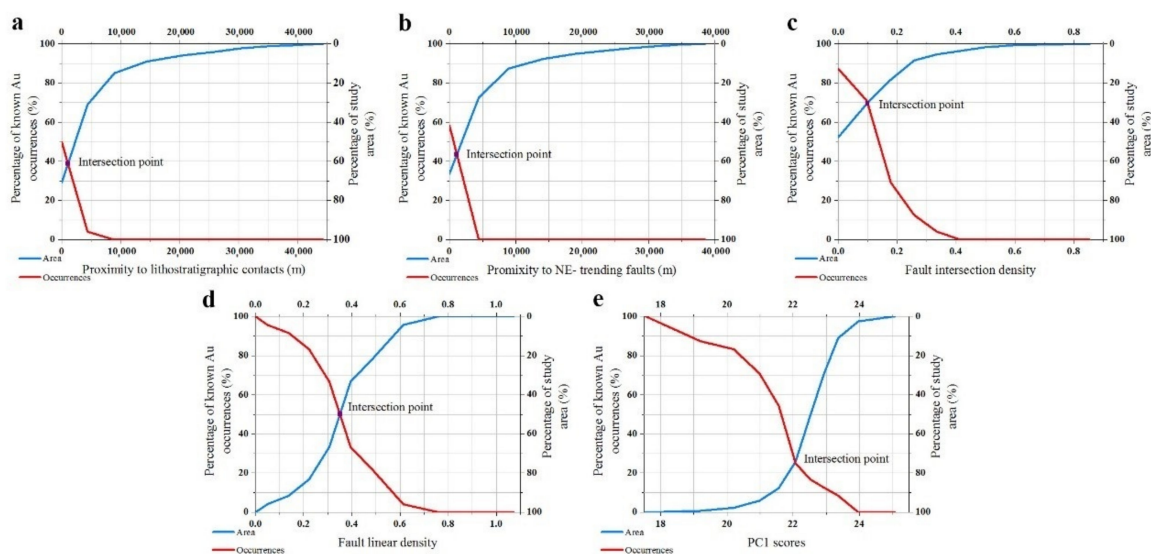


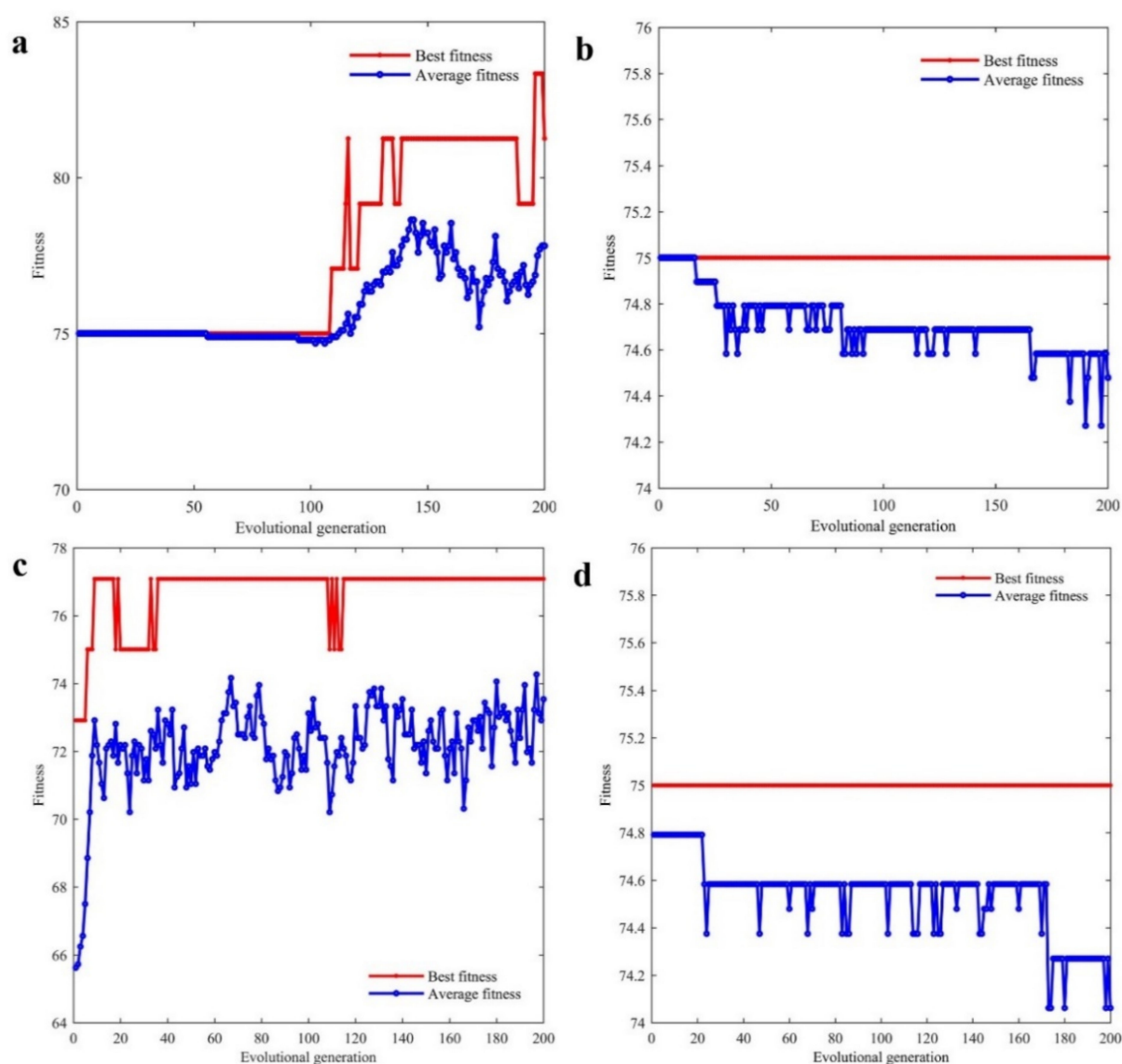
Figure 6. Prediction-area (P-A) plots for evidence layers: (a) proximity to lithostratigraphic contacts; (b) proximity to NE-trending faults; (c) fault intersection density; (d) fault linear density; (e) PC1 scores generated by Ag, As, Au, and Sb.

Table 4. Parameters used for the GA–SVM models.

Parameter	Description	Value
maxpop	Maximum number of population	50
maxgen	Maximum number of Generation	200
C	The penalty parameter of SVM	0–100
σ	The kernel parameter of RBF for SVM	0–100
k	k-fold cross-validation	6

Table 5. The optimal parameters of the SVM model, based on GA.

Parameters	Training Dataset 1	Training Dataset 2	Training Dataset 3	Training Dataset 4
Best	0.0947	93.0015	0.1940	50.5152
Best C	2.3241	2.1234	0.0547	1.2698
Accuracy	83.33%	75%	77.08%	75%

**Figure 7.** The parameters searching process of GA: (a) training dataset 1; (b) training dataset 2; (c) training dataset 3; (d) training dataset 4.

The confusion matrices and F1 scores for the individual models are presented in Table 6 to demonstrate the performance evaluation results of the GA-SVM model presented in this study. In terms of the confusion matrices, most of the known deposits were classified accurately, and the highest precision was 0.96. In addition, the results show that the F1 score based on training dataset 2 was the highest, indicating that it had the greater ability to distinguish mineralized locations than the other models.

Although the F1 score provides a proxy for measuring the classification ability of the GA-SVM models, it cannot assess the spatial efficiency of the prospectivity model classifications [33]. Therefore, the number of known deposits in the prospectivity area and the percentage of occupied areas corresponding to the prospectivity area for each prospectivity model were calculated to measure the relative spatial efficiency. From the statistical comparison between the GA-SVM models with the four training datasets (Table 7), it is obvious that the GA-SVM model that used training dataset 2, which occupied the largest area of the study area, had a larger number of known Au deposits, accounting for the poor spatial efficiency of the prospectivity model, although the F1 score was the highest. This may have resulted from overfitting. In contrast, although the lowest F1 score was obtained for the GA-SVM model trained using training dataset 3, it was more efficient in its classification than the other models. This illustrates that the prospectivity model is sensitive to randomly selected non-mineralized samples.

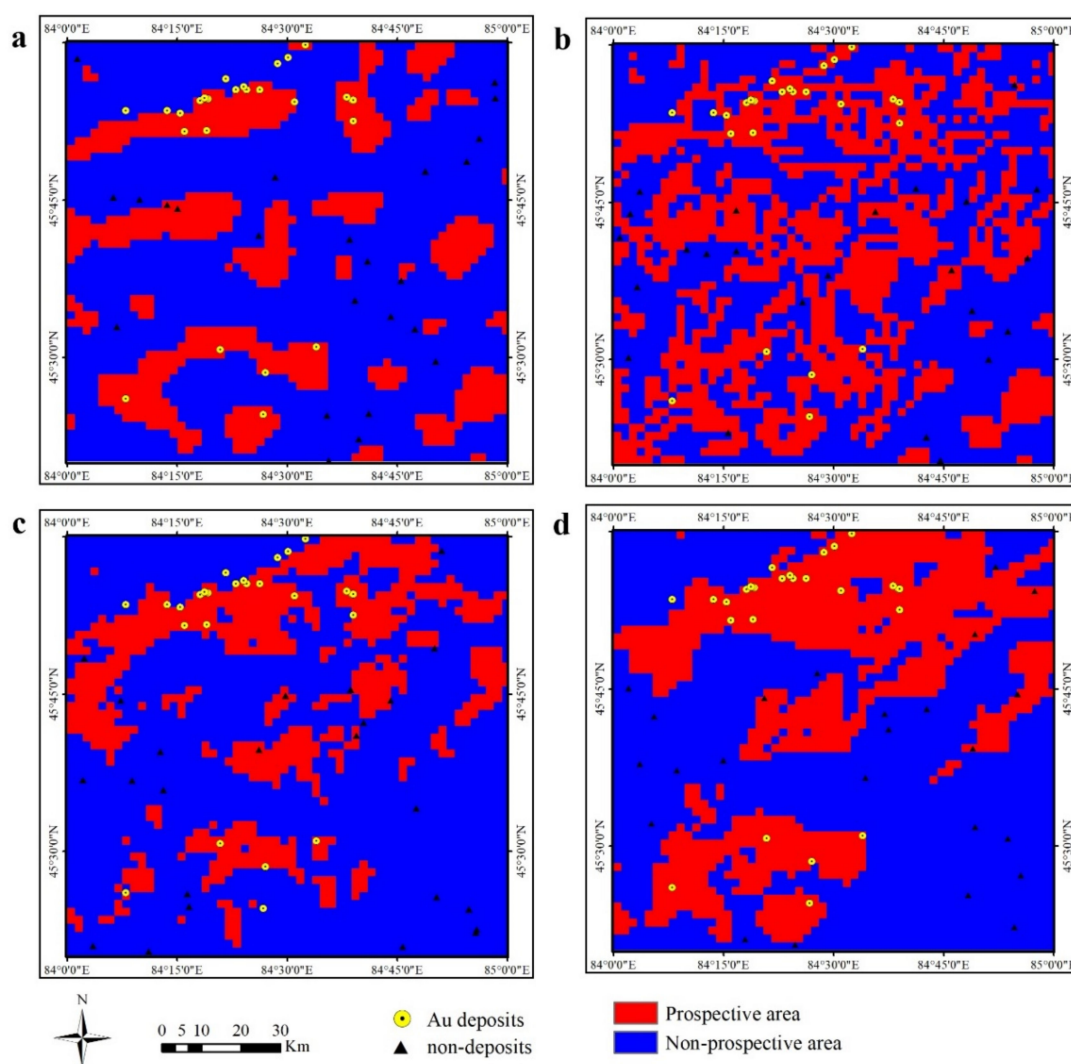


Figure 8. Prospectivity maps generated using GA-SVM model: (a) training dataset 1; (b) training dataset 2; (c) training dataset 3; (d) training dataset 4.

Table 6. Performance metrics for the GA–SVM models.

Known	Training Dataset 1		Training Dataset 2		Training Dataset 3		Training Dataset 4	
Prediction	A	B	A	B	A	B	A	B
A	18	2	22	2	18	4	23	5
B	6	22	2	22	6	20	1	19
Precision	0.75		0.92		0.75		0.96	
Recall	0.9		0.92		0.82		0.82	
F1	0.82		0.92		0.78		0.88	

A: Mineralized samples; B: Non-mineralized samples.

Table 7. Statistical results for the GA–SVM models.

	Training Dataset 1	Training Dataset 2	Training Dataset 3	Training Dataset 4
Number of known deposits	18	22	18	23
Prospective area (%)	25.24%	43.56%	25.13%	35.68%

The prospectivity model with training dataset 4 was the best in terms of both the F1 score and the spatial efficiency of the classification, as it reduced the target area of the study area while predicting the same number of known deposits and exhibited good performance in identifying mineralized locations. The prospective areas in Figure 8d occupied 35.68% of the study area and contain 95.83% of the known Au deposits. From the perspective of the spatial domain, the spatial distribution of the best prospectivity map (Figure 8d) showed a spatial correlation with proximity to NE-trending faults, which is consistent with the model of Au mineralization.

5. Conclusions

This study employed a hybrid support vector machine (SVM) model with genetic algorithm (GA) to discriminate between prospective and non-prospective areas for Au deposits in Karamay, northwest China. The findings support the following conclusions:

- Since SVM generalization performance is heavily dependent on parameters σ and C , it is necessary to adopt GA as an objective function to select better combinations of the two parameters for SVM.
- Owing to the characteristic of P-A plot, it can be used for classifying evidence layers into binary patterns. It is important to note that the knowledge of the metallogenic model should be applied to differentiate favorable and unfavorable areas in the binary maps.
- A key procedure in implementing the GA–SVM model was the selection of the training dataset, especially the ‘non-mineralized’ locations. In complex geological environments, it is impossible to identify non-mineralized locations; thus, point pattern analysis is a useful measure for determining the optimal distance at which non-mineralized locations can be randomly selected based on the selection criteria.
- The performance of the GA–SVM model for distinguishing prospective areas in the study area was evaluated using both the F1 score and spatial efficiency. The best prospectivity model predicted 95.83% of the known Au deposits within prospective areas, occupying 35.68% of the study area.
- The best prospectivity map, as classified by the GA–SVM model, displayed a strong spatial correlation between prospective areas and proximity to NE-trending faults. This conforms to the characterization of spatial associations between geological features and Au deposits, indicating that the results emphasize the strong control of Au mineralization by NE-trending faults within the study area.

Author Contributions: Conceptualization, Xishihui Du and Kefa Zhou; methodology, Xishihui Du and Shuguang Zhou; software, Jinlin Wang and Shuguang Zhou; resources, Kefa Zhou and Jinlin Wang; writing—original draft preparation, Xishihui Du; writing—review and editing, Kefa Zhou and Yao Cui; visualization, Xishihui Du; supervision, Kefa Zhou and Yao Cui. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Foundation of the Educational Department of Liaoning Province (lnqn202018) and West Light Foundation of the Chinese Academy of Sciences (2017-XBQNXZ-B-019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank Li Dong and the contribution of all the anonymous reviewers that improved the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Li, T.; Xia, Q.; Zhao, M.; Gui, Z.; Leng, S. Prospectivity mapping for Tungsten Polymetallic mineral resources, Nanling Metallogenic Belt, South China: Use of random forest algorithm from a perspective of data imbalance. *Nat. Resour. Res.* **2020**, *29*, 203–227. [\[CrossRef\]](#)
- Parsa, M.; Maghsoudi, A. Assessing the effects of mineral systems-derived exploration targeting criteria for random Forests-based predictive mapping of mineral prospectivity in Ahar-Arasbaran area, Iran. *Ore Geol. Rev.* **2021**, *138*, 104399. [\[CrossRef\]](#)
- Porwal, A.; Carranza, E.J.M. Introduction to the Special Issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. *Ore Geol. Rev.* **2015**, *71*, 477–483. [\[CrossRef\]](#)
- Bonham-Carter, G.F. *Geographic Information Systems for Geoscientists Modelling with GIS*; Pergamon: Oxford, UK, 1994.
- Carranza, E.J.M. Geochemical anomaly and mineral prospectivity mapping in GIS. In *Handbook of Exploration and Environmental Geochemistry*; Elsevier: Amsterdam, The Netherlands, 2008.
- Carranza, E.J.M.; Mangaoang, J.C.; Hale, M. Application of mineral exploration models and GIS to generate mineral potential maps as input for optimum land-use planning in the Philippines. *Nat. Resour. Res.* **1999**, *8*, 165–173. [\[CrossRef\]](#)
- Sadeghi, B.; Khalajmasoumi, M. A futuristic review for evaluation of geothermal potentials using fuzzy logic and binary index overlay in GIS environment. *Renew. Sustain. Energy Rev.* **2015**, *43*, 818–831. [\[CrossRef\]](#)
- An, P.; Moon, W.M.; Rencz, A. Application of fuzzy set theory for integration of geological, geophysical and remote sensing data. *Can. J. Explor. Geophys.* **1999**, *27*, 1–11.
- Knox-Robinson, C.M. Vectorial fuzzy logic: A novel technique for enhanced mineral prospectivity mapping, with reference to the orogenic gold mineralisation potential of the Kalgoorlie Terrane, Western Australia. *Aust. J. Earth Sci.* **2000**, *47*, 929–941. [\[CrossRef\]](#)
- Mutele, L.; Billay, A.; Hunt, J.P. Knowledge-driven prospectivity mapping for Granite-Related Polymetallic Sn–F–(REE) mineralization, Bushveld Igneous Complex, South Africa. *Nat. Resour. Res.* **2017**, *26*, 535–552. [\[CrossRef\]](#)
- Abedi, M.; Mohammadi, R.; Norouzi, G.H.; Mohammadi, M.S.M. A comprehensive VIKOR method for integration of various exploratory data in mineral potential mapping. *Arab. J. Geosci.* **2016**, *9*, 482. [\[CrossRef\]](#)
- Abedi, M.; Ali Torabi, S.; Norouzi, G.H.; Hamzeh, M.; Elyasi, G.R. PROMETHEE II: A knowledge-driven method for copper exploration. *Comput. Geosci.* **2012**, *46*, 255–263. [\[CrossRef\]](#)
- Pazand, K.; Hezarkhani, A.; Ataei, M. Using TOPSIS approaches for predictive porphyry Cu potential mapping: A case study in Ahar-Arasbaran area (NW, Iran). *Comput. Geosci.* **2012**, *49*, 62–71. [\[CrossRef\]](#)
- Liu, Y.; Cheng, Q.; Xia, Q.; Wang, X. Mineral potential mapping for tungsten polymetallic deposits in the Nanling metallogenic belt, South China. *J. Earth Sci.* **2014**, *25*, 689–700. [\[CrossRef\]](#)
- Zuo, R. Regional exploration targeting model for Gangdese porphyry copper deposits. *Resour. Geol.* **2011**, *61*, 296–303. [\[CrossRef\]](#)
- Carranza, E.J.M. Data-driven evidential belief modeling of mineral potential using few prospects and evidence with missing values. *Nat. Resour. Res.* **2015**, *24*, 291–304. [\[CrossRef\]](#)
- Carranza, E.J.M.; Hale, M. Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. *Ore Geol. Rev.* **2003**, *22*, 117–132. [\[CrossRef\]](#)
- Zhang, D.; Ren, N.; Hou, X. An improved logistic regression model based on a spatially weighted technique (ILRBSWT v1.0) and its application to mineral prospectivity mapping. *Geosci. Model Dev.* **2018**, *11*, 2525–2539. [\[CrossRef\]](#)
- Xiong, Y.; Zuo, R. GIS-based rare events logistic regression for mineral prospectivity mapping. *Comput. Geosci.* **2018**, *111*, 18–25. [\[CrossRef\]](#)

20. Abedi, M.; Norouzi, G.H.; Bahroudi, A. Support vector machine for multi-classification of mineral prospectivity areas. *Comput. Geosci.* **2012**, *46*, 272–283. [\[CrossRef\]](#)
21. Shabankareh, M.; Hezarkhani, A. Application of support vector machines for copper potential mapping in Kerman region, Iran. *J. Afr. Earth Sci.* **2017**, *128*, 116–126. [\[CrossRef\]](#)
22. Zuo, R.; Carranza, E.J.M. Support vector machine: A tool for mapping mineral prospectivity. *Comput. Geosci.* **2011**, *37*, 1967–1975. [\[CrossRef\]](#)
23. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2000.
24. Holland, J.H. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975.
25. Ding, S.; Su, C.; Yu, J. An optimizing BP neural network algorithm based on genetic algorithm. *Artif. Intell. Rev.* **2011**, *36*, 153–162. [\[CrossRef\]](#)
26. Huang, C.; Wang, C.J. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* **2006**, *31*, 231–240. [\[CrossRef\]](#)
27. Sun, Y.; Wong, A.K.; Kandrewamel, M. Classification of imbalanced data: A review. *Int. J. Pattern Recognit. Artif. Intell.* **2009**, *23*, 687–719. [\[CrossRef\]](#)
28. Rahimi, H.; Abedi, M.; Yousefi, M.; Bahroudi, A.; Elyasi, G.R. Supervised mineral exploration targeting and the challenges with the selection of deposit and non-deposit sites thereof. *Appl. Geochem.* **2021**, *128*, 104940. [\[CrossRef\]](#)
29. Zuo, R.; Wang, Z. Effects of random negative training samples on mineral prospectivity mapping. *Nat. Resour. Res.* **2020**, *29*, 3443–3455. [\[CrossRef\]](#)
30. Carranza, E.J.M.; Hale, M.; Faassen, C. Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping. *Ore Geol. Rev.* **2008**, *33*, 536–558. [\[CrossRef\]](#)
31. Nykänen, V.; Lahti, I.; Niiranen, T.; Korhonen, K. Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geol. Rev.* **2015**, *71*, 853–860. [\[CrossRef\]](#)
32. Hariharan, S.; Tirodkar, S.; Porwal, A.; Bhattacharya, A.; Joly, A. Random forest-based prospectivity modelling of Greenfield Terrains using sparse deposit data: An example from the Tanami Region, Western Australia. *Nat. Resour. Res.* **2017**, *26*, 489–507. [\[CrossRef\]](#)
33. Prado, E.M.G.; de Souza Filho, C.R.; Carranza, E.J.M.; Motta, J.G. Modeling of Cu–Au prospectivity in the Carajás mineral province (Brazil) through machine learning: Dealing with imbalanced training data. *Ore Geol. Rev.* **2020**, *124*, 103611. [\[CrossRef\]](#)
34. Yousefi, M.; Carranza, E.J.M. Data-driven index overlay and boolean logic mineral prospectivity modeling in Greenfields exploration. *Nat. Resour. Res.* **2016**, *25*, 3–18. [\[CrossRef\]](#)
35. Yousefi, M.; Carranza, E.J.M. Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Comput. Geosci.* **2015**, *79*, 69–81. [\[CrossRef\]](#)
36. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines*; Cambridge University Press: Cambridge, UK, 2000.
37. Cao, Y.; Yin, K.; Zhou, C.; Ahmed, B. Establishment of landslide groundwater level prediction model based on GA-SVM and influencing factor analysis. *Sensors* **2020**, *20*, 845. [\[CrossRef\]](#)
38. Huang, S.; Zheng, X.; Ma, L.; Wang, H.; Huang, Q.; Leng, G.; Meng, E.; Guo, Y. Quantitative contribution of climate change and human activities to vegetation cover variations based on GA-SVM model. *J. Hydrol.* **2020**, *584*, 124687. [\[CrossRef\]](#)
39. Min, S.H.; Lee, J.; Han, I. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Syst. Appl.* **2006**, *31*, 652–660. [\[CrossRef\]](#)
40. Anguita, D.; Ghelardoni, L.; Ghio, A.; Oneto, L.; Ridella, S. The “K” in k-fold cross validation. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium, 25–27 April 2012.
41. Bengio, Y.; Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.* **2004**, *5*, 1089–1105.
42. Chen, K.Y.; Wang, C.H. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Syst. Appl.* **2007**, *32*, 254–264. [\[CrossRef\]](#)
43. Xiong, Y.; Zuo, R. Effects of misclassification costs on mapping mineral prospectivity. *Ore Geol. Rev.* **2017**, *82*, 1–9. [\[CrossRef\]](#)
44. Rijsbergen, C. *Information Retrieval*, 2nd ed.; Butterworth-Heinemann: London, UK, 1979.
45. An, F.; Zhu, Y. Native antimony in the Baogutu gold deposit (west Junggar, NW China): Its occurrence and origin. *Ore Geol. Rev.* **2010**, *37*, 214–223. [\[CrossRef\]](#)
46. Shen, Y.; Jin, C. *Magmatism and Gold Mineralization in Western Junggar*; Science Press: Beijing, China, 1993. (In Chinese)
47. Geng, H.; Sun, M.; Yuan, C.; Xiao, W.; Xian, W.; Zhao, G.; Zhang, L.; Wong, K.; Wu, F. Geochemical, Sr–Nd and zircon U–Pb–Hf isotopic studies of Late Carboniferous magmatism in the West Junggar, Xinjiang: Implications for ridge subduction? *Chem. Geol.* **2009**, *266*, 364–389. [\[CrossRef\]](#)
48. Han, B.; Ji, J.; Song, B.; Chen, L.; Zhang, L. Late Paleozoic vertical growth of continental crust around the Junggar Basin, Xinjiang, China (Part I): Timing of post-collisional plutonism. *Acta Petrol. Sin.* **2006**, *22*, 1077–1086. (In Chinese)
49. Su, Y.; Tang, H.; Hou, G.; Liu, C. Geochemistry of aluminous A-type granites along Darabut tectonic belt in West Junggar, Xinjiang. *Geochimica* **2006**, *35*, 55–67. (In Chinese)
50. Dong, C. Characteristics of Forming Fluids and Metallogenic Prediction in the Depth of Hatu Gold Deposit. Master’s Thesis, Xinjiang University, Urumqi, China, 2012. (In Chinese)

-
51. Liu, S.; Tao, M.; Zheng, G.; Zeng, K. Experiment study on mineral processing of some complex gold ore. *Met. Mine* **2009**, *39*, 98–103. (In Chinese)
 52. Xie, X.; Mu, X.; Ren, T. Geochemical mapping in China. *J. Geochem. Explor.* **1997**, *60*, 99–113.
 53. Cheng, Q. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geol. Rev.* **2007**, *32*, 314–324. [[CrossRef](#)]
 54. Cheng, Q. Non-Linear theory and Power-Law models for information integration and mineral resources quantitative assessments. *Math. Geosci.* **2008**, *40*, 503–532. [[CrossRef](#)]
 55. Xiao, F.; Chen, J.; Zhang, Z.; Wang, C.; Wu, G.; Agterberg, F. Singularity mapping and spatially weighted principal component analysis to identify geochemical anomalies associated with Ag and Pb-Zn polymetallic mineralization in Northwest Zhejiang, China. *J. Geochem. Explor.* **2012**, *122*, 90–100. [[CrossRef](#)]
 56. Zhou, S.; Zhou, K.; Cui, Y.; Wang, J.; Ding, J. Exploratory data analysis and singularity mapping in geochemical anomaly identification in Karamay, Xinjiang, China. *J. Geochem. Explor.* **2015**, *154*, 171–179.
 57. Diggle, P.J. A kernel method for smoothing point process data. *Appl. Stat.* **1985**, *34*, 138–147. [[CrossRef](#)]
 58. Diggle, P.J. *Statistical Analysis of Spatial Point Pattern*; Academic Press: Cambridge, MA, USA, 1983.
 59. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]