Identification of Hydrothermal Alteration Minerals for Exploring Gold Deposits Based on SVM and PCA Using ASTER Data: A Case Study of Gulong

Kai Xu 1,2,3, Xiaofeng Wang 1, Chunfang Kong 1,2,4, Ruyi Feng 1, Gang Liu 1 and Chonglong Wu 1,2

Abstract: Dayaoshan, as an important metal ore-producing area in China, is faced with the dilemma of resource depletion due to long-term exploitation. In this paper, remote sensing methods are used to circle the favorable metallogenic areas and find new ore points for Gulong. Firstly, vegetation interference was removed by using mixed pixel decomposition method with hyperplane and genetic algorithm (GA) optimization; then, altered mineral distribution information was extracted based on principal component analysis (PCA) and support vector machine (SVM) methods; thirdly, the favorable areas of gold mining in Gulong was delineated by using the ant colony algorithm (ACA) optimization SVM model to remove false altered minerals; and lastly, field surveys verified that the extracted alteration mineralization information is correct and effective. The results show that the mineral alteration extraction method proposed in this paper has certain guiding significance for metallogenic prediction by remote sensing.

Keywords: gold deposit; alteration information; ASTER image; support vector machine (SVM); principal component analysis (PCA)

1. Introduction

Surrounding rock alteration is one of the important interpretation markers of mineral exploration using remote sensing technology [1,2]. The metasomatism of hydrothermal mineralization can cause the minerals to alter and generate groups or ions such as Fe3+, magnesium hydroxyl group and aluminum hydroxyl group, which show different hue and spectral characteristics compared to the non-altered rock in remote sensing images [3–5]. In the process of mineral exploration, according to the different hue and reflectance spectra of altered minerals, the composition and spatial distribution of altered minerals can be analyzed using remote sensing technology, and the favorable metallogenic areas can be found out.

With the development of remote sensing technology, most scholars have focused on the mechanism of the mineral spectral [6–9]. Others have studied extraction method of altered mineral [10–12]. Still, some have researched the identification of hydrothermal alteration minerals using multi- and hyper-spectral images [13–16]. In recent years, altered mineral identification has been investigated by
using different methods in different locations for mineral exploration, including alteration mineral mapping in the Northwestern Junggar basin using Landsat thematic mapper (TM) data and principal component analysis (PCA) [17], predictive mineral prospectivity modeling for Cu deposits in Varzaghan district, NW Iran based on the support vector machine (SVM) [18], as well as predictive models for Rodalquilar mining district mineral prospectivity in the southeast of Spain with machine learning [19]. All of these studies contribute to the development of metallogenic prognosis in arid and semi-arid areas. However, research on altered mineral extraction in high vegetation-covered areas is still limited, particularly in a specific region as an important metal mineral area [20,21], where many old mines are facing resource depletion after long-term exploitation, such as Gulong, China.

In response to the above problems, the objectives of this paper are to: (1) select the image identification marks of altered minerals according to the distribution of gold deposits and the ore-bearing strata in Dayaoshan; (2) remove vegetation interference by using mixed pixel decomposition method with hyperplane and genetic algorithm (GA) optimization; (3) extract altered mineral information based on PCA and SVM methods by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data; and (4) verify the altered mineral extraction result using field survey method. Lastly, the research results not only delineate the favorable metallogenic area of gold deposits, but also provide reasonable suggestions for mineral exploration with remote sensing data.

2. Study Area and Materials

The study area, located at the junction of the Yangtze paleo-plate and Cathaysia paleo-plate, is part of the Guangxi Dayaoshan Au ore belt, dotted with gold deposits [22–25] (Figure 1). It belongs to the subtropical monsoon climate region, where the annual average temperature is 21°C, the frost-free days are about 320 days, and the average annual rainfall is 1600 mm. There are a lot of pristine forests and all kinds of herbaceous plants, and forest cover can reach more than 90%.

![Figure 1. Cont.](image-url)
Figure 1. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) image of the study area (a), regional geological sketch map (b) (from the Remote Sensing Center of Guangxi), located in Guangxi (c) and in China (d).

The study area is strongly influenced by multistage structural evolution, such as sedimentation, uplift, and folds and faults in different directions in a long process of geological evolution, which makes
the structures of the research area very complex. It is located in the anticline axis of the Gulong-Ligang Indochina period. The north-south anticline strikes and superimposes on the Caledonian near-east-west fold, which has a length of 28 km and width of about 2–5 km. There are two circular structures in the study area, which are distributed in the Dacun and Gulong. The single ring is approximately circular and about 2-5km in diameter. There are fault structures in the study area, which are near the south-north direction of the Limu-Majiang fault, with a total length of more than 200 km. The southern section of the Limu-Majiang fault plays an important role in controlling gold deposits. The northeast direction of Pingxiang-Dali fault, which passes through study area, plays an important role in controlling altered rock type and volcanic type gold deposit. The strata exposed are mainly Cambrian, Cretaceous, Devonian and a spot of Quaternary. The main intrusive rocks are mainly Caledonian granitic porphyry veins and granodiorite veins; Yanshanian granite-porphyry, granodiorite, quartz diorite, and monzonitic granite; and Indosinian gabbro diorite, which plays a vital role in formation of porphyry deposits of the whole study area (Figure 1b).

At present, the endophytic metal deposits found in the study area can be divided into porphyry deposits, broken zone altered rock deposits, and quartz vein deposits. Those three kinds of deposits are often associated with each other, that is to say, there are altered rock deposits and quartz vein deposits of different scales in the outer part of the porphyry deposits (Figure 1b).

The ASTER images adopted in this paper are from the Remote Sensing Center of Guangxi, acquired in September 2001, and the RGB band combination of ASTER image is as follows: R is B2, G is B1, and B is (3*b1 + b3)/4 (Figure 1a). Before identifying the altered mineral, the images were pretreated, such as atmospheric correction by the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [26,27], crosstalk correction [28], water disturbance removed by using a water index. The flow chart of ASTER image preprocessing is shown in Figure 2.

3. Methods

The extraction of altered mineral information is usually based on the spectral characteristics of minerals [9,29–31]. Although minerals have their own spectral characteristics, rocks are usually formed by various mineral assemblages, so the spectral information of rocks on remote sensing images is very complex. In addition, due to the influence of other surface features, such as soil and vegetation, the altered mineral information is very weak in the remote sensing images [32–34]. Therefore, this paper firstly focuses on eliminating vegetation interference, extracting mineralized information samples with remote sensing images, and then establishing PCA and SVM methods to improve the accuracy of altered mineral information extraction.

3.1. Mixed Pixel Decomposition with Hyperplane Optimized by GA

In areas with high vegetation coverage, the spectral characteristics of pixels in remote sensing images are not a single ground object, but a mixed reflection of the several objects. The altered mineral information is seriously hided by high vegetation coverage, so it is very difficult to extract the information which is needed. For mixed pixels, an unmixed pixel decomposition method with hyperplane and an optimized genetic algorithm [35–39], which is established in this paper, will eliminate the influence of vegetation, and obtain altered mineral information from the sub-pixel level.

In the feature space of a remote sensing image, the same kind of mineral is assembled in one region because of its similar spectral characteristics, while other similar minerals are gathered in another
Mixed pixel decomposition with genetic optimization for hyperplane classification model

\[ f(x) = M \cdot X > 0, \quad f(x) = M \cdot X \leq 0 \] (1)

where \( M = m_1, m_2, ..., m_n \), \( X = (x_1, x_2, ..., x_n) \), \( n \) is the dimension of feature space.

The position of hyperplane in multidimensional remote sensing data space can be well represented by iterative recursive mode, and the equation is expressed as Equation (2):

\[
d = \alpha_N \cos \alpha_{N-1} + \beta_{N-1} \sin \alpha_{N-1}
\]
\[
\beta_{N-1} = \alpha_{N-1} \cos \alpha_{N-2} + \beta_{N-2} \sin \alpha_{N-2}
\]
\[
\beta_{N-2} = \alpha_{N-2} \cos \alpha_{N-3} + \beta_{N-3} \sin \alpha_{N-3}
\]
\[
\vdots
\]
\[
\beta_1 = \alpha_1 \cos \alpha_0 + \beta_0 \sin \alpha_0
\] (2)

where \( \alpha_{N-1} \) is the angle between the hyperplane unit normal and the \( X_N \), \( \alpha_{N-2} \) is the angle between the normal projection in the \( X_1, X_2, ..., X_{N-1} \) space and the \( X_{N-1} \) axis, \( \alpha_{N-3} \) is the angle between the normal projection in the \( X_1, X_2, ..., X_{N-1} \) space and the \( X_{N-2} \) axis, and \( \alpha_1 \) and \( \alpha_0 \) is the angle between the normal projection of 2D space and the 2nd feature axis, and 1D space and the 1st feature axis, respectively. If \( \alpha_0 = 0 \), \( d \) is the vertical distance between the hyperplane and the origin, in this way, a hyperplane of N-dimensional space can be determined by \( N - 1 \) angle \( \alpha_1, \alpha_2, ..., \alpha_{N-1} \) and a vertical distance \( d \) [40].

It can be seen in Equation (2) that if hyperplane is used to classify the feature space composed of multi-band remote sensing images, the first problem that needs to be solved is how to determine the \( N \) parameters of the hyperplane equation, which can best achieve the classification effect. In this paper, to establish a hyperplane classification model optimized by the genetic algorithm for classification, firstly, the pattern description and pattern matching of the samples in the training sample set are carried out; secondly, comparison and selection of different pattern schemes with evolution are performed, and the best pattern classification scheme is chosen; and finally, this is extended to the whole image to achieve the purpose of pattern classification. This process is shown in Figure 3.

Figure 3. Mixed pixel decomposition with genetic-hyperplane flow chart.
3.2. Mineral Altered Samples Collection with Ratio Analysis and PCA

By using the ratio of reflection and absorption band, the spectral difference among various geological information can be enhanced, and the influence of topography can be reduced [41–44]. Therefore, according to the characteristics of the spectrum, the selection of appropriate ASTER band for ratio operation can enhance the weak altered mineral.

Because of high correlation among the bands of images, some of the data are redundant and repetitive [13, 45–47]. In this paper, PCA of images is described, and the independent quantities obtained in order to reduce the correlation influence in the process of alteration information extraction are recorded.

PCA combines the original n features linearly and establishes m principal components, which are not correlated with each other, and the variance decreases gradually. The expression of the transformation is as follows (Equation (3)):

\[ Y = AX = \sum_{i=1}^{n} A_i X_i \]  

Transformation is to reduce the dimension of the original data, using m-dimensional vector can cover the original n-dimensional information, and make its standard deviation within the allowable range, then there is Equation (4):

\[ Y = AX = \sum_{i=1}^{m} A_i X_i + \sum_{i=m+1}^{n} A_i X_i \]  

where \( X_i \) is denoted as \( b_i \), then the Equation (4) becomes Equation (5):

\[ Y(m) = \sum_{i=1}^{m} A_i X_i + \sum_{i=m+1}^{n} A_i b_i \]  

Let the error is \( \varepsilon \) between \( Y \) and \( Y(m) \),

\[ \varepsilon = \sum_{i=m+1}^{n} A(X_i - b_i) \]  

3.3. Altered Mineral Extraction Using SVM and ACA

To improve the extraction accuracy of mineral alteration information, the SVM model, an excellent remote sensing image classification technology [48–52], was trained using alteration sample data by PCA. SVM firstly maps samples to high-dimensional spaces by nonlinear transformation; and then finds out the optimal classification hyperplane in high-dimensional space; and finally, classifies the sample data.

The samples are set \((X_i, Y_i)\), \(i = 1, 2, 3, \ldots, n\), \(Y_i \in \{-1, 1\}\), \(\omega X + b = 0\) is the classified hyperplane of the sample, so construct discriminant function (7):

\[ g(x) = \omega X + b \]  

where \( \omega \) represents the normal vector, \( b \) represents the displacement, which determines the distance between the hyperplane and the origin. For the optimal classification hyperplane, the SVM is:

\[ \min_{\omega, b} \frac{1}{2} ||\omega||^2 \]  

s.t. \( y_i (\omega^T X_i + b) \geq 1, i = 1, 2, \ldots, l \)
For nonlinear data samples, relaxation variables ($\xi_i$) and penalty factors (C) are introduced to deal with the error problem, Equation (8) becomes Equation (9).

$$\min_{\omega,k} \frac{1}{2}||\omega||^2 + C \sum_{i=1}^{l} \xi_i$$

s.t. $y_i(\omega^T\cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, l$  

In order to map the sample to the higher dimensional feature space, and to avoid the difficulty of calculation, the Gaussian kernel function is introduced:

$$k(x_i, x_j) = \exp\left(-\gamma||x_i, x_j||^2\right)$$  

(10)

where $\gamma = \frac{1}{2\sigma^2}$

Classification decision function of optimal classification hyperplane is Equation (11).

$$h(x) = \text{sgn}\left(\sum_{i=1}^{l} a_i y_i k(x_i, x) + b\right)$$

(11)

In summary, determining the parameters C and $\gamma$ of SVM is the key to improve the classification accuracy. So the ACA is introduced to solve the problem of SVM parameter optimization [53–57]. ACA, as a heuristic algorithm based on global optimization, can solve the parameter optimization problem very well. The optimization operation steps are as follows: (1) parameter search interval, $C \in [0, 10], \quad \gamma \in [0, 1]$ is determined; (2) the process is initialized, each ant carries on the random allocation operation, named $[C, \gamma]$, the grid point on each space corresponds to one state, and each state corresponds to a solution on the space, the movement trajectory of ants in the grid is recorded, and the pheromone is arranged according to the objective function value between each grid point, so that the ant can search according to the size of the pheromone on the grid point; and (3) the root mean square error is set as the objective function, and then the best combination is selected among the many parameter combinations. The optimization algorithm flow of ACA is shown in Figure 4.

Figure 4. Flow chart of the support vector machine (SVM) parameters with the ant colony algorithm (ACA).

To sum up, the steps of remote sensing altered mineral information extraction with PCA and SVM methods are as follows:

(1) The remote sensing images in the study area are preprocessed, including radiometric calibration, atmospheric correction and descrambling, and eliminate the influence of vegetation with mixed pixel decomposition method.
(2) The threshold density segmentation of pyrite, sericite and chlorite principal components is carried out with ratio analysis and PCA, and the training samples are collected according to the known deposits.

(3) The kernel parameters and penalty factors of SVM with ACA are optimized.

(4) The SVM classifier trained by training samples is used to process the remote sensing image to obtain the abnormal information of altered minerals.

4. Results and Discussion

4.1. Vegetation Disturbance Elimination

In the study area, there are three main land cover types, which are vegetation, rock and water. The selection of training samples was based on the field survey, remote sensing image, and topographic map. According to the results of the field survey, 52 field rock sampling regions in the image were used as sampling region of rock, and 50–200 pixels were selected as rock samples in each region, which have a total of 6320 rock samples. Then in the vegetation coverage region surrounding the 52 rock sampling regions, and 50–500 pixels were selected as vegetation samples in each region, which have a total of 8460 vegetation samples. Finally, 50–100 pixels of water samples were selected for each of the 50 water regions on the image, which have a total of 4210 water samples. The training samples and reference samples were carried out in ENVI4.8, with 80% of them used as training samples and the remaining 20% as test samples. The details are shown in Table 1.

<table>
<thead>
<tr>
<th>Band</th>
<th>Type</th>
<th>Number of Samples</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>vegetation</td>
<td>8460</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>rock</td>
<td>6320</td>
<td>18,990</td>
</tr>
<tr>
<td>C3</td>
<td>water</td>
<td>4210</td>
<td></td>
</tr>
</tbody>
</table>

Three bands of remote sensing image were selected, and three kinds of training samples were classified by using two-dimensional feature space. According to Equation (2), the hyperplane equation can be simplified to:

\[
d = X_2 \times \cos \alpha + X_1 \times \sin \alpha
\]  

(12)

Because a chromosome represents a set of hyperplane sets, and there are three kinds of objects to be classified, a group of hyperplanes has three chromosomes. A hyperplane consists of an angle \(\alpha\) and a distance \(d\), and there are 30 chromosomes in the genetic initialization population in this paper.

The parameter setting of GA is very important to the result accuracy and time efficiency of hybrid pixel decomposition. In this study, appropriate parameters were determined through the following three tables (Tables 2–4).

<table>
<thead>
<tr>
<th>Variation Rate</th>
<th>0.01</th>
<th>0.02</th>
<th>0.05</th>
<th>0.07</th>
<th>0.09</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Number of Classifications</td>
<td>432</td>
<td>456</td>
<td>391</td>
<td>416</td>
<td>381</td>
<td>372</td>
<td>365</td>
<td>371</td>
</tr>
<tr>
<td>Training Accuracy</td>
<td>0.864</td>
<td>0.912</td>
<td>0.782</td>
<td>0.832</td>
<td>0.762</td>
<td>0.744</td>
<td>0.73</td>
<td>0.742</td>
</tr>
<tr>
<td>Training Time (s)</td>
<td>261</td>
<td>259</td>
<td>302</td>
<td>431</td>
<td>462</td>
<td>294</td>
<td>309</td>
<td>192</td>
</tr>
</tbody>
</table>

(Number of samples: 500; Type: 3; Crossing rate: 0.99).
When the variation probability is less than 0.1, the training accuracy is high. When the variation probability is set to 0.02, the relationship between the cross rate and the training accuracy and training time is shown in the following table when the other parameters remain unchanged (Table 3).

**Table 3.** The relationship between Crossing rate and training accuracy & time.

<table>
<thead>
<tr>
<th>Crossing Rate</th>
<th>Generations</th>
<th>Correct Number of Classifications</th>
<th>Training Accuracy</th>
<th>Training Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>711</td>
<td>457</td>
<td>0.914</td>
<td>149</td>
</tr>
<tr>
<td>0.05</td>
<td>623</td>
<td>451</td>
<td>0.902</td>
<td>138</td>
</tr>
<tr>
<td>0.1</td>
<td>658</td>
<td>451</td>
<td>0.902</td>
<td>141</td>
</tr>
<tr>
<td>0.3</td>
<td>110</td>
<td>445</td>
<td>0.89</td>
<td>92</td>
</tr>
<tr>
<td>0.4</td>
<td>1120</td>
<td>472</td>
<td>0.906</td>
<td>245</td>
</tr>
<tr>
<td>0.6</td>
<td>812</td>
<td>472</td>
<td>0.944</td>
<td>193</td>
</tr>
<tr>
<td>0.8</td>
<td>421</td>
<td>465</td>
<td>0.94</td>
<td>121</td>
</tr>
<tr>
<td>0.99</td>
<td>1128</td>
<td>471</td>
<td>0.942</td>
<td>207</td>
</tr>
</tbody>
</table>

(Number of samples: 500; Type: 3; Variation rate: 0.02).

When the cross rate is 0.3, the training time is the least and the training accuracy is high. When the cross rate is equal to 0.3 and the variation rate is equal to 0.02, the relationship between the number of training points and the training accuracy & time is shown in the following Table 4.

**Table 4.** The relationship between training points and training accuracy & time.

<table>
<thead>
<tr>
<th>Number of Samples</th>
<th>Generations</th>
<th>Correct Number of Classifications</th>
<th>Training Accuracy</th>
<th>Training Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>631</td>
<td>94</td>
<td>0.94</td>
<td>0.0548</td>
</tr>
<tr>
<td>200</td>
<td>292</td>
<td>192</td>
<td>0.96</td>
<td>0.089</td>
</tr>
<tr>
<td>300</td>
<td>273</td>
<td>273</td>
<td>0.91</td>
<td>0.132</td>
</tr>
<tr>
<td>500</td>
<td>446</td>
<td>359</td>
<td>0.718</td>
<td>0.184</td>
</tr>
<tr>
<td>900</td>
<td>68</td>
<td>465</td>
<td>0.517</td>
<td>0.231</td>
</tr>
<tr>
<td>1200</td>
<td>2000</td>
<td>864</td>
<td>0.72</td>
<td>0.734</td>
</tr>
<tr>
<td>2000</td>
<td>2000</td>
<td>1187</td>
<td>0.594</td>
<td>0.891</td>
</tr>
<tr>
<td>4000</td>
<td>2000</td>
<td>2739</td>
<td>0.685</td>
<td>1.632</td>
</tr>
</tbody>
</table>

(Type: 3; Variation rate: 0.02; Crossing rate: 0.3).

When the cross rate is equal to 0.3 and the variation rate is equal to 0.02, the increase of training points will not lead to the increase of training time in a certain range.

In this paper, the genetic algorithm uses binary coding. The initial population is randomly selected, and then according to the above research, cross rate is equal to 0.3, and variation rate is equal to 0.02. After 212 generations of genetic evolution, 18,250 training samples were successfully classified, and the training success rate was 96.1%. The three hyperplane equations and angles acquired by genetic training were as in Table 5:

**Table 5.** Hyperplane set.

<table>
<thead>
<tr>
<th>Number</th>
<th>Distance</th>
<th>Angle</th>
<th>Hyperplane Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>1.546207</td>
<td>44.136729</td>
<td>$1.546207 = x_2 \times \cos 44.136729 + x_1 \times \sin 44.136729$</td>
</tr>
<tr>
<td>H2</td>
<td>0.760832</td>
<td>161.892638</td>
<td>$0.760832 = x_2 \times \cos 161.892638 + x_1 \times \sin 161.892638$</td>
</tr>
<tr>
<td>H3</td>
<td>1.546207</td>
<td>14.834792</td>
<td>$1.546207 = x_2 \times \cos 14.834792 + x_1 \times \sin 44.14.834792$</td>
</tr>
</tbody>
</table>

The mixed pixel decomposition with hyperplane and GA was used to process the remote sensing images, the categories were determined according to the final assignment, the vegetation interference were removed, and results are as in Figure 5.
Because vegetation is the main disturbance factor in extracting mineral alteration information in a high vegetation area, removing or weakening vegetation will greatly improve the accuracy of extracting alteration information. The spectral profiles of the study image before and after vegetation removal were as in Figure 6.

Figure 6 indicates the image has basically no vegetation spectral information after vegetation removal treatment.

Figures 5 and 6 certify the image of the rock is highlighted in place of the original green part of the vegetation cover area after vegetation removal by unmixing with hyperplane optimized by GA.

### 4.2. Collection of Altered Mineral Samples

Through the analysis of the distribution of gold deposits and the ore-bearing strata of the study area, the pyrite, sericite and chlorite are the most closely related to altered mineral of surrounding rock, which is a good prospecting indicator of gold ore [58–60].

The main basis for the selection of bands was based on the characteristic spectrum of minerals to select the two bands with the largest reflectivity difference in ASTER image for ratio operation and to generate the ratio image, which can enhance the spectral difference among minerals and eliminate the interference of topography and shadow.

In the light of the minerals reflectance spectra of pyrite, B1, B2, B5/B3, and B4 of ASTER image were selected as input bands for PCA. The eigenvector of principal component was as shown in Table 6.
Table 6. Pyrite eigenvector matrix.

<table>
<thead>
<tr>
<th>Eigenvector</th>
<th>B1</th>
<th>B2</th>
<th>B5/B3</th>
<th>B4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.166731</td>
<td>0.120621</td>
<td>0.822539</td>
<td>0.530171</td>
</tr>
<tr>
<td>2</td>
<td>0.106507</td>
<td>0.233433</td>
<td>−0.562449</td>
<td>0.786013</td>
</tr>
<tr>
<td>3</td>
<td>0.599849</td>
<td>0.734171</td>
<td>0.024860</td>
<td>−0.317107</td>
</tr>
<tr>
<td>4</td>
<td>−0.775267</td>
<td>0.626062</td>
<td>0.080392</td>
<td>0.023353</td>
</tr>
</tbody>
</table>

According to the spectral curve characteristics of pyrite, the spectral curve at B1 of ASTER image is absorption valley, and the coefficient symbol should be negative in the eigenvector matrix. The spectral curve at B2 and B4 of ASTER image are reflection peak, and the coefficient symbol should be positive in the eigenvector matrix. The PC4 in Table 6 satisfies the above conditions, so the PC4 reflects the alteration information of iron mineralization and was selected as the extraction layer of pyrite mineralization.

In order to get pyrite mineralization, PC4 needs to be further processed. Firstly, the PC4 was filtered by using $3 \times 3$ median filter to eliminate the noise. Then, the result of the filter was stretched linearly by using 0–255 grey value. Finally, the mineralization information was segmented by using the average value plus standard deviation of 3.0 times as the threshold, and the segmentation result was as shown in Figure 7:

Figure 7 shows that the altered zone of pyrite has obvious ribbon and planar distribution characteristics, which are mainly distributed in the western of the study area, and a small amount scattered in the northeast of the study area, forming a fourth-step ladder spatial distribution pattern along the northwest-southeast direction. Along the northwest-southeast direction are the alteration belts formed in the northeast, southeast and southwest of Dacun rock mass; in the east and south of Gulong rock mass; in the surrounding of Pingtoubei, Sitai, Shedong, Dapo rock mass; as well as in the surrounding of Wujie, Ludong, Fenghuang, Fuqing, and Siwei rock mass (Figure 7).

In the same way, in the light of the minerals reflectance spectra of sericite, B1, B4/B6, B7, and B9 of the ASTER image were selected as input bands for PCA. The eigenvector of principal component was as shown in Table 7.
Table 7. Sericite eigenvector matrix.

<table>
<thead>
<tr>
<th>Eigenvector</th>
<th>B1</th>
<th>B4/B6</th>
<th>B7</th>
<th>B9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16443</td>
<td>0.811045</td>
<td>0.523520</td>
<td>-0.202710</td>
</tr>
<tr>
<td>2</td>
<td>0.070343</td>
<td>-0.569964</td>
<td>0.690800</td>
<td>0.439304</td>
</tr>
<tr>
<td>3</td>
<td>0.859394</td>
<td>0.038299</td>
<td>-0.353052</td>
<td>0.367872</td>
</tr>
<tr>
<td>4</td>
<td>-0.479012</td>
<td>0.126017</td>
<td>0.352244</td>
<td>-0.794908</td>
</tr>
</tbody>
</table>

According to the spectral curve characteristics of sericite, the spectral curve at B1 and B9 of ASTER image are an absorption valley, and the coefficient symbol should be negative in the eigenvector matrix. The spectral curve at B7 of the ASTER image is a reflection peak, and the coefficient symbol should be positive in the eigenvector matrix. The PC4 in Table 7 satisfied the above conditions, so the PC4 was selected as the extraction layer of sericite mineralization.

In order to get sericite mineralization, the PC4 needed to be further processed in the same way as pyrite, and the segmentation result was as shown in Figure 8:

Figure 8. Alteration distribution of sericite mineralization.

Figure 8 indicates the altered zone of sericite has obvious ribbon and planar distribution characteristics, which are mainly distributed in the western of the study area, forming a fourth-step ladder spatial distribution pattern along the northwest-southeast direction. Along the northwest-southeast direction are the alteration belts formed in the northeast of Dacun rock mass; in the interior of Gulong rock mass; in the surrounding of Pingtoubei, Sitai, Shedong, Dapo rock mass; as well as in the surrounding of Wujie, Ludong, Fenghuang, Fuqing, and Siwei rock mass (Figure 8).

Similarly, in the light of the minerals reflectance spectra of chlorite, B1, B2, B5/B8 and B9 of ASTER image were selected as input bands for PCA. The eigenvector of the principal component was as shown in Table 8.

Table 8. Chlorite eigenvector matrix.

<table>
<thead>
<tr>
<th>Eigenvector</th>
<th>B1</th>
<th>B2</th>
<th>B5/B8</th>
<th>B9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.162074</td>
<td>0.799186</td>
<td>0.516097</td>
<td>0.262064</td>
</tr>
<tr>
<td>2</td>
<td>0.063939</td>
<td>-0.592936</td>
<td>0.688779</td>
<td>0.412216</td>
</tr>
<tr>
<td>3</td>
<td>0.909014</td>
<td>0.073300</td>
<td>-0.309002</td>
<td>-0.269885</td>
</tr>
<tr>
<td>4</td>
<td>0.378598</td>
<td>-0.065995</td>
<td>0.404654</td>
<td>0.829797</td>
</tr>
</tbody>
</table>
According to the spectral curve characteristics of chlorite, the spectral curve at B2 of the ASTER image is an absorption valley, and the coefficient symbol should be negative in the eigenvector matrix. The spectral curve at B1 and B9 of the ASTER image are reflection peaks, and the coefficient symbol should be positive in the eigenvector matrix. The PC4 in Table 8 satisfies the above conditions, so the PC4 is selected as the extraction layer of chlorite mineralization.

In order to get chlorite mineralization, the PC4 needs to be further processed in the same way as pyrite and sericite, and the segmentation result are as shown in Figure 9:

![Figure 9](image-url)  
**Figure 9.** Alteration distribution of chlorite mineralization.

Figure 9 illustrates that the altered zone of chlorite has obvious ribbon distribution characteristics in Figure 8, which are mainly distributed in the western of the study area, and a small amount scattered in the northeast. Specifically they are distributed in the Dacun rock mass; in the interior of Gulong rock mass; in the surrounding of Pingtoubei, Sitai, Shedong, Dapo rock mass; as well as in Ludong, Fenghuang, Fuqing, and Siwei rock mass (Figure 9).

In general, the altered mineral formation of pyrite, sericite and chlorite are closely related to the hydrothermal alteration, so the altered zone has the characteristics surrounding rock mass distribution [60,61]. In addition, the alteration area is distributed both sides of the Limu-Majiang fault near the north-south direction, and in the south of the Pingxiang-Dali fault near the northeast direction, so the spatial distribution of the altered zone is closely related to the intrusive rocks and faults.

4.3. Altered Mineral Information Extraction

In order to obtain a mineralization alteration sample, the mineralization alteration information of pyrite, sericite, and chlorite were loaded into the previous survey results map (Figure 10).

The selection of training samples was based on the field survey, remote sensing image and topographic map. According to the results of the field survey and 14 gold deposits in Figure 10, 300 pixels each altered minerals were selected from the image where 14 gold mining areas overlap with altered minerals. At the same time, 900 pixels were selected in the non-mining area and the non-alteration mineral area, with a total of 1800 pixels as the training and test samples to train the SVM model. Among them, 80% of the samples (half of them were mineralization alteration, 1/3 of per mineralization alteration) were randomly selected as training samples, and the remaining 20% were used as the test samples.
Figure 10. Distribution map of gold deposits in Dayaoshan area (Remote Sensing Center of Guangxi).

On the basis of selecting SVM kernel function, the SVM was trained by the above sample data, and at the same time, the SVM was optimized by the ACA, in which the maximum number of iterations K = 10, ant colony size N = 20, pheromone evaporation coefficient Rho = 0.8, pheromone increasing intensity Q = 0.9, ant crawling speed Lambda = 0.3.

In order to verify the correctness of the SVM model, the area under curve (AUC) of the receiver operating characteristic (ROC) curve was introduced in this research (Figure 11). Figure 11 indicates that the AUC was 0.889 and the forecast accuracy was 88.9%. This result indicates that the SVM has higher prediction accuracy. Therefore, the model built in this paper can be used to extract the altered mineral.

Figure 11. Receiver operating characteristic (ROC) curve of the SVM model.

The altered mineral obtained using the PCA method in the study area were input into the trained SVM, the false information was removed, and the mineralization alteration results were obtained, as shown in Figure 12.
Figure 12. Extraction results of altered mineral information.

Figure 12 indicates that the mineral alteration zone has obvious ribbon and planar distribution characteristics, and forming three regions in space: located in the area of Dacun and Gulong rock mass; located surrounding Pingtoubei, Sitai, Shedong, and Dapo rock mass, as well as located surrounding Fenghuang, Fuqing, and Siwei rock mass. Compared with Figures 7–9, it can be seen that the range of the alteration area is smaller, which is due to the removal of some false anomalies. Figure 12 also indicates that mineral alteration does not exist alone, but is associated with each other, which is a good prospecting indicator for gold deposits.

In summary, there are obvious reflection characteristics near 0.67 μm and an absorption valley near 0.81 μm for pyrite, and absorption characteristics at 2.17 μm for sericite, as well as absorption characteristics at 2.35 μm for chlorite. Those characteristics in combination can be used to identify the alteration zone and support mineral prediction. At present, many scholars have used multi-spectral data to identify the alteration zone based on above combination of characteristics \([60,62,63]\).

4.4. Verification of Altered Mineral Information Extraction Results

In order to verify the accuracy of the mineral alteration information extracted, a special field investigation was carried out. First of all, the working scope of field verification was determined according to the spatial distribution of remote sensing alteration information. Then, the field survey route was planned according to the regional geological characteristics. Finally, the rock specimens with alteration anomalies were collected in different locations. After 15 days of field sampling and verification in Gulong of Dayao Mountain, 116 rock samples were collected at 52 sampling points (Figure 13), and the rock sample verification analysis was as follows (Figures 14–16).
are quartz and feldspar, and sometimes biotite and hornblende. Quartz, which is consistent with the results of mineralization and...

Figure 13. Spatial distribution of 52 field sampling points.

Samples at Sanxianding (110.945° E, 23.641° N; 7-2-2015) using ASTER images (Fig 14. (a)) View of the surrounding rock for pyritization, (b) The rock sample of the pyritization.

Figure 14. Samples at Sanxianding (110.945° E, 23.641° N; 7-2-2015) (a) View of the surrounding rock for pyritization, (b) The rock sample of the pyritization.

Samples at Sunjia (111.027° E, 23.540° N; 8-2-2015) View of the granitic porphyry, (b) The rock sample of the granitic porphyry.

Figure 15. Samples at Hecun (111.027° E, 23.540° N; 8-2-2015) (a) View of the granitic porphyry, (b) The rock sample of the granitic porphyry.
Figure 15. Samples at Hecun (111.027° E, 23.540° N; 8-2-2015) (a) View of the granitic porphyry, (b) The rock sample of the granitic porphyry.

(1) The exposed rocks in Sanxianding have obvious pyritization with a light brass color and bright metallic luster, which shows that the altered minerals contained in the outcrop rocks in the field are consistent with the results extracted using ASTER images (Figure 14). In hydrothermal deposits, pyrite is symbiotic with other sulfides, oxides, quartz and so on. This phenomenon indicates that such rock composition often contains gold, silver and other elements.

(2) The rocks in Hecun sampling area are mainly granite porphyry with 15–20% phenocryst, the main rock components are quartz and feldspar, and sometimes biotite and hornblende. Quartz phenocrysts usually to be hexagonal bipyramid, and biotite and hornblende sometimes were existed darkening edges (Figure 12). This phenomenon indicates that there are intrusive rocks in the sampling area, which directly proves the existence of mineral alteration in surrounding rock (Figure 15). Metallic minerals related to granitic porphyry are gold, silver, copper, and so on.

(3) The sampling point is mainly meso-basic igneous rock and metasomatic rock, which is the surrounding rock related to chloritization (Figure 16). Chloritization rarely occurs alone, often accompanied by pyritized, sericitization, propylitization, and carbonate and so on.

The results of field verification certified that at 40 of the 52 sampling sites rocks containing pyrite, sericite and chlorite can be collected, which is consistent with the results of mineralization and alteration extraction using remote sensing image.

In order to further verify the reliability of the proposed method, the spectral curves of rock samples collected in the field were measured and analyzed in laboratory, and the results were as shown in Figure 17.

By comparing spectral curves between Figure 17 and Spectroscopy Lab of the United States Geological Survey (USGS) [64], the results show that the rock samples collected in the field have altered mineral characteristic of pyrite, sericite, and chlorite, respectively, which are the main signs for gold prospecting.

The results indicate that the extraction results of alteration information in this paper have very high accuracy and could play a vital role in the exploration of metal minerals in the study area. According to the extracted mineralization alteration and the results of field verification, the favorable areas of gold deposit can be delineated (Figure 18).
The results indicate that the extraction results of alteration information in this paper have very high accuracy and could play a vital role in the exploration of metal deposits. The methodological reference for areas with high vegetation coverage and strong disturbance information can be mapped well with remote sensing ground object spectral data, laboratory mineral spectrum data, rock mineral characteristic of pyrite, sericite, and chlorite, respectively, which are the main signs for gold prospecting. Simultaneously, the alteration mineral is the weak information in the image, which is very vulnerable to interference from high vegetation coverage. The genetic algorithm is used to remove the influence of vegetation interference, and geological big data metallogenic prediction, especially in areas of high vegetation cover.

Figure 17. The spectral curves of samples collected in the field (a-Feldspar quartz sandstone from Sanxianding), (b-Quartz porphyry from Sunjialing), (c-Granite-porphyry from Hecun).

Figure 18. The favorable areas of gold deposit.
5. Conclusions

This study describes a methodology for gold ore deposit identification based on ASTER data using SVM and PCA in areas of high vegetation. Gulong in the Dayaoshan metallogenic belt of Guangxi was taken as the study area for field verification. The results show that the mineral alteration regions of the Gulong can be mapped well with field investigations. At the same time, the results also show that the mineral alteration extraction method proposed in this paper has certain guiding significance for metallogenic prediction by remote sensing, especially in areas with high vegetation cover. This study results in the following three main conclusions:

(1) Altered mineral is the weak information in the image, which is very vulnerable to inundation in the background. At the same time, the study area is located in the gold ore-forming zone of the southern margin of the Dayaoshan-Daguishan, where exposed intrusive rocks and surrounding rock alteration can be found on the surface by field exploration. That altered mineral information can be disturbed by high coverage of vegetation. The mixed pixel decomposition method with hyperplane optimized using the genetic algorithm is used to remove the influence of vegetation interference information and highlight altered mineral [65] (Figure 5). Thus, it provides a methodological reference for areas with high vegetation coverage and strong disturbance information.

(2) The extraction model of altered mineral information was constructed based on PCA and SVM optimized by ACA [66], and successfully extracted the altered mineral in the study area, by comprehensively processing the multi-source heterogeneous data, such as remote sensing ground object spectral data, laboratory mineral spectrum data, rock mineral distribution data, and so on. On this basis, the favorable metallogenic areas were mapped in the study area (Figure 18).

(3) By comparing spectral curves between field rock samples and Spectroscopy Lab of USGS [64], the results show that the rock samples collected in the field have altered mineral characteristic of pyrite, sericite, and chlorite, respectively, which are the main signs for gold prospecting. Simultaneously, field investigations verified that the distribution of extracting alteration areas from ASTER display a good correlation with the alteration regions, which further demonstrates that the method of extracting alteration mineralization information adopted in the study is feasible.

In general, the results can provide technical support for multi-source heterogeneous data fusion and geological big data metallogenic prediction, especially in areas of high vegetation cover.

Author Contributions: All the authors made significant contributions to the work. K.X., X.W., R.F., and C.K. designed the research, analyzed the results, and accomplished the validation work. K.X., and X.W. completed field investigation, sampling, analysis, and validation; G.L., and C.W. provided advice for the revision of the paper.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the Remote Sensing Center of Guangxi for providing the various data sets used in this paper. This work has been supported by the National Natural Science Foundation of China (No: 41201193); the Research on evaluation method of Oil and Gas Resources investigation (prediction) in “Belt and Road Initiative” Spatial Information Corridor; Guizhou Science and Technology Planning Project: Research and development application of big data Management and Intelligent processing system for Mine Exploration ([2017]2951); the Open research project of key laboratory of Tectonics and Petroleum Resources (China University of Geosciences), Ministry of Education (No: TPR-2019-11); and the Open fund project of National-Local Joint Engineering Laboratory on Digital Preservation and Innovative Technologies for the Culture of Traditional Villages and Towns (CTCZ19K01). The authors would like to thank the anonymous reviewers for providing valuable comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References
Remote Sens. 2019, 11, 3003


16. Alimohammadi, M.; Alirezaei, S.; Kontak, D.J. Application of ASTER data for exploration of porphyry copper deposits: A case study of Daraloo-Sarmeshk area, southern part of the Kerman copper belt, Iran. *Or Geol. Rev.* 2015, 70, 290–304. [CrossRef]


42. Aydal, D.; Ardal, E.; Dumanlilar, O. Application of the Crosta technique for alteration mapping of granitoidic rocks using ETM+ data: Case study from eastern Tauride belt (SE Turkey). Int. J. Remote Sens. 2007, 28, 3895–3913. [CrossRef]


58. Liu, G.H. Extraction and Analysis of Information of Alteration with Multi-Source Remote Sensing Data in High Vegetation Coverage Area; Chengdu University of Technology: Chengdu, China, 2017.


63. Mars, J.C.; Rowan, L.C. Regional mapping of phyllic- and argillic-altered rocks in the Zagros magmatic arc, Iran, using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data and logical operator algorithms. Geosphere 2006, 2, 161–186. [CrossRef]


© 2019 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 (http://creativecommons.org/licenses/by/4.0/).