Shahrood University of Technology
Iranian Society of Mining Engineering (IRSME)
Journal of Mining and Environment (JME)
journal homepage: www.jme.shahroodut.ac.ir
THOMSON REUTERS

Vol. 11, No. 1, 2020, 49-61.
*DOI: 10.22044/jme.2019.7956.1664.*

# A Comparative Study of SVM and RF Methods for Classification of Alteration Zones Using Remotely Sensed Data

N. Mahvash Mohammadi and A. Hezarkhani*

*Department of Mining and Metallurgy Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran*

Received 12 January 2019; received in revised form 9 April 2019; accepted 23 July 2019

**Keywords**
*Classification*
*Machine learning*
*Random forest*
*Support vector machine*
*Advanced space borne thermal emission and reflection radiometer*
*Alteration*
*Porphyry copper*

**Abstract**

Identification and mapping of the significant alterations are the main objectives of the exploration geochemical surveys. The field study is time-consuming and costly in order to produce the classified maps. Therefore, processing of the remotely sensed data, which provides timely and multi-band (multi-layer) data, can be substituted for the field study. In this work, the Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) imagery is used for alteration classification by applying two new methods of machine learning including random forest and support vector machine. The 14 band ASTER and 19 derivative data layers extracted from ASTER including band ratio and PC imagery are used as the training datasets in order to improve the results. Comparison of the analytical results achieved from the two mentioned methods confirm that the SVM model has a sufficient accuracy and a more powerful performance than the RF model for alteration classification in the studied area.

## 1. Introduction

Identification of geochemical alterations is necessary to identify and recognize deposits due to the association of alteration with mineralization [1,2] The porphyry and porphyry copper deposits, which are characterized by low grades and high tonnage, are considered as primary sources of copper in the world [3, 4]. The rock alteration formed during the mineralization and intrusion is a key parameter to explore the porphyry deposits. Hydrothermal alterations originated from the metasomatic processes changing the composition, mineralogy, and texture of rocks [5]. Potassic, phyllic, advanced argillic, intermediate argillic, propylitic, sodic–calcic and sodic, greisen, and skarn have been presented as major hydrothermal alteration types in the porphyry copper deposits [5, 6]. Mapping of the porphyry copper alterations using remotely sensed data offers a variety of benefits such as availability, multi-band, data repetition in short time intervals, and detection ability to depict various units [7]. Additionally, using satellite imagery leads to saving time, money, and human resource. Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) is a Japanese sensor placed on the Terra platform by NASA in 1999; the information has been acquired from the instrument since 2000 [8]. ASTER provides high-resolution images in 14 spectral bands from the visible to the thermal infrared (TIR) regions of the electromagnetic spectrum. Visible Near Infrared (VNIR) is a sensor image in the three bands between 0.52 and 0.86 micrometers, Short Wave Infrared (SWIR) bands including six bands between 1.6 and 2.430 micrometers, and the last five bands are between 8.125 and 11.65 micrometers (TIR), where the spatial resolutions are 15, 30, and 90 m, respectively [9]. Using the

*Corresponding author: ardehez@aut.ac.ir (A. Hezarkhani).*

remotely sensed data, especially the ASTER data, has been applied in geological research works, for instance, geological mapping [10-16] and alteration mapping [17-20]. Nowadays, the classification algorithms, converting data to meaningful information, have been widely used in geological mapping. Various classification techniques, both parametric and non-parametric, have been developed in different contexts—remote sensing inclusive [21]. The Maximum Likelihood Classifier (MLC), as a parametric method, assumes the class signature in normal distribution [22]. The most significant MLC issue is invalid for classes including several sub-classes or classes having different spectral features [23]. The non-parametric classification techniques have been introduced to solve this problem in remotely sensed data such as Neural Network (NN), Support Vector Machine (SVM), and Random Forest (RF) classifier [24-27]. The SVM algorithm is based on the basic statistical learning theory introduced by Vapnik in 1995 [28]. This method is widely used in a variety of pattern recognition applications such as copper potential mapping [29], separation of alteration zones [30], lithological classification [25], and land cover classification [31]. The RF algorithm is one of the advanced machine learning techniques classifying the data via classification and regression tree (CART: [32]). The potential use of the RF approach, as an efficient classifier, has been broadly investigated in the remotely sensed data and geological purposes [24, 33, 34]. Therefore, in this research work, two effective algorithms, SVM and RF, were used for alteration classification of the studied area located in SE Iran. Three different types of remotely sensed data including 14 ASTER band, 4 PCA bands, and 4 band ratios were used as the training data sets.

## 2. Geology of studied area

The subduction of the Arabian Plate beneath central Iran ceases to form the Central Iranian Volcanic belt, which hosts many known porphyry copper deposits such as Meiduk, Sarcheshmeh, and Takht-e-Gonbad [35-38] (Figure 1). The 1:100000 sheet of Chahar-Gonbad was selected as an exploration sheet because of the importance of the exploration of porphyry copper deposits such as the Takht-e-Gonbad deposit. This sheet is situated in the Urumieh-Dokhtar magmatic arc and between the longitudes $56^{\circ}\ 06'\ 35''$ and $56^{\circ}\ 28'\ 10''$ and the latitudes $29^{\circ}\ 27'\ 19''$ and $29^{\circ}\ 42'\ 12''$. The rock units in this district include Eocene volcanic and pyroclastic rocks, where the striking N-S and the Neogene deposits are mainly exposed as carbonate or volcanic rocks in the southwestern, east, and center of the Chahar-Gonbad sheet. There is a colorful melange in the west and southwestern of the studied area, while the Quaternary alluvial deposits are in the north and northeastern. Only a few of the Paleozoic metamorphic rocks are found in the south of the studied area. In this area, igneous rocks are exposed as volcanic units, intrusive igneous rocks, and ophiolite complexes. 70% of the outcrop rocks are formed from volcanic rocks with ages of Eocene and Pliocene [39, 40]. Quartz-sericitic or phyllic alteration is the most common type of alteration in aluminum-rich rocks at moderate depths. Major minerals are sericite (muscovite, illite), kaolinite +/- quartz, tourmaline, and sometimes carbonate minerals, and pyrite often accompanies them. The main area of the sericitic altered area is closely related to the rhyolitic tuff unit, and the second reason is their connection with hydrothermal processes. Argillic or intermediate argillic alteration is the most extensive and most common type of alteration for many hydrothermal and mineralization systems. The alteration of rock plagioclase causes the occurrence of this alteration with the presence of kaolinite, illite, smectite, and montmorillonite minerals. At this point, partial values of alonite were observed. Also Tuff-sand stone and Marn exist in these sections according to the geological map of the studied area. The propylitic alteration is the first and most external alteration around the deposit in the large area. The propylitic alteration is characterized by chlorite, epidote, calcite and lower amounts of smectite, zeolite, and amphibole. Chlorite, epidote, and calcite minerals are abundant in the studied area but many of them are not related to the propylitic alteration [40, 41].

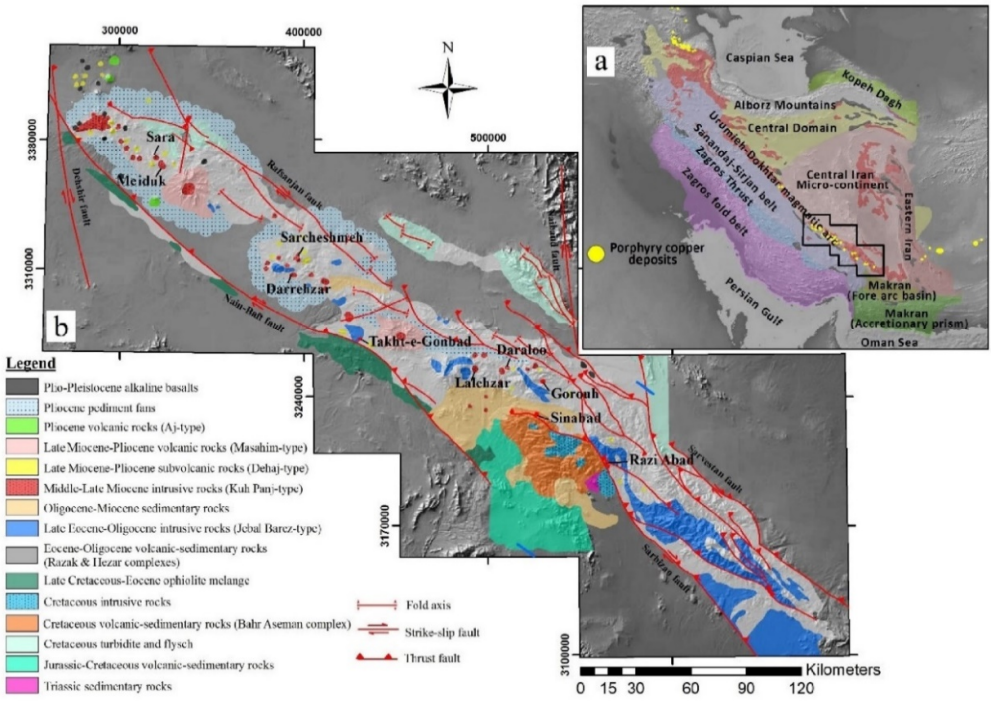


**Figure 1. a) Geographical location of studied area and Urumieh Dokhtar magmatic arc in Iran, b) simplified geological map of Kerman belt and the locations of Takht-e-Gonbad and other main porphyry copper deposits, i.e. Meiduk and Sarcheshmeh (original map from [42] and modified by [40]).**

### 3. Support vector machine (SVM)

Different classes are classified by SVM based on finding an optimal separating hyperplane, where the hyperplane has the maximum distance from the points close to the decision boundary. Support vectors are the points near the optimal separating hyperplane; all the other points are irrelevant for identifying the binary class boundaries [28]. When the data is not linearly separable, the kernel function is applied to solve a non-linear problem. This method can provide a single solution with a strong feature that is more appropriate for classification problems with poor conditions [29]. The SVM algorithm for a two-class problem is discussed. The equation of line separator two-class or two groups can be written by sgn(x) that is a sign function as a decision function:

$$f(x) = \text{sgn}(wx_i + b)$$

Where $w$ is a decision hyperplane normal vector and $x_i$ is the data point $i$ [22]. Assume that all data is at least distance 1 from the hyperplane; then the following two constraints follow for a training set $\{(x_i, y_i)\}$, where $y_i$ is the class of data point *i* (i.e. $y_i \in \{-1, +1\}$).

$$wx_i + b \geq +1 \quad for \ y_i = +1$$

$$wx_i + b \leq -1 \quad for \ y_i = -1$$

which is equivalent to:

$$y_i(wx_i + b) \geq 1 \quad i = 1, 2, ..., n$$

For calculating $f(x)$, the two parameters of decision function $w$ and $b$ should be found by solving the Lagrange quadratic function. There are objective and minimization functions in the quadratic function that according to the above assumptions are:

Minimize $$\tau(w) = \frac{1}{2}\left\|w^2\right\|$$

subject to:

$$y_i((wx_i) + b) \geq 1 \quad i = 1, ..., l$$

The Lagrange function is constructed and minimized as follows:

$$L(w, b, \alpha) = \frac{1}{2}\left\|w^2\right\| - \sum_{i=1}^{l} \alpha_i (y_i((x_i w) + b) - 1$$

Then the derivative Lagrange function is taken with respect to *w*, *b*, and $\alpha$ afterwards it is equated to 0.

$$\frac{\partial}{\partial b}L(w,b,\alpha)=0\ ,\ \frac{\partial}{\partial w}L(w,b,\alpha)=0$$

where $\alpha$ is a Lagrange multiplier that is called dual variables. The objective function of the dual problem is required to maximize. Dual variables are calculated by the following optimization function [43]:

$$Maximize\ \sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{ij=1}^{l}\alpha_i\alpha_j y_i y_j(x_i x_j)$$

subject to:

$$\alpha_i\geq 0,\ \ i=1,...,l,and\ \sum_{i=1}^{l}\alpha_i y_i=0$$

Finally, the new decision function based on the rule of optimal hyperplane is as follows [43]:

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{l}y_i\alpha_i(xx_i)+b)$$

Dual variables are obtained by solving the quadratic and training set; afterwards, SVM is used for classifying the testing data. An important assumption is that the data with linearly separable can be correctly classified by a linear SVM, and if the data is not linearly separable, it could not be correctly classified by the linear decision [30]. Therefore, this problem is solved using various kernel functions, applied for transforming the input features to a higher dimensional space that is called feature space. The original feature space can always be mapped to some higher dimensional feature space, where the training set is separable. The linear classifier relies on an inner product between vectors $K(x_i,x_j)=x_i x_j$. If every data point is mapped into a high-dimensional space via some transformation $\phi: x\rightarrow\phi(x)$, the inner product becomes:

$$k(x_i,x_j)=\varphi(x_i).\varphi(x_j)$$

In this function, $\varphi(x_i).\varphi(x_j)$ are used instead of $x_i,x_j$ Now, kernel function $K$ ( $x_i,x_j$ ) is used such that:

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{l}y_i\alpha_i(\varphi(x).\varphi(x_i))+b)$$

Finally, the following function, as decision function, is presented:

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{l}y_i\alpha_i K(x_i,x)+b)$$

There are some common kernel functions such as 1. Linear, 2. Polynomial, 3. Radial basis function (RBF: [43]). In this work, the following kernel function was considered:

The radial basis function (RBF)

$$k(x_i,x_j)=e^{-\gamma(x_i-x_j)^2}$$

where $\lambda$, as the kernel parameter, is referred to the RBF width [43]. The RBF kernel commonly used in SVM classification due to high-dimensional data can be analyzed, and just the two parameters $C$ (penalty parameter) and $\lambda$ are required. The two stages of design SVM are as follow: 1. Choosing a kernel function and 2. Determining $C$ margin constant parameter [44]. The SVM classifier uses the one-against-one method for multiple-class problems to classify the data by considering the class with the greatest number of votes. In this method, for m+1 classes, m (m+1)/2 of two class classifier is produced to classify the combination of two classes of data [45].

**4. Random forest (RF)**

One of the most developed and non-parametric machine learning methods is the RF technique, which is based on a decision tree. RF is an ensemble learning method that has been used for various applications such as classification and regression [46]. The samples were randomly selected from the observation data by the bootstrap method, and a tree was grown for each sample by the recursive partitioning algorithms. These algorithms were stopped when the tree reached a large size and one node was produced for each observation. This process was continued until an RF model was developed. Variable importance (VI) is an index that is used for ranking the variables; the Gini and Permutation importance indices are known as the VI index. The Gini index was applied to construct the RF trees based on determining the impurity of each node in a decision tree [32, 47]. The RF classifier uses a voting system, whereby the class of forest has been selected by the majority vote of the class. The final classification map is formed based on the majority

vote of each class that has been calculated from all the trees and assigned to each pixel basis [32, 48].

## 5. Datasets

### 5.1. Aster data

The ASTER level L1T scene that covered the 1:100000 sheet of Chahar-Gonbad was recorded in summer on 21 April, 2007 at 07:02:51.793 AM GMT. The cloud cover of the ASTER scene is almost equal to zero, and also $133.66^0$ and $65.98^0$ are the solar azimuth angle and the solar elevation angle, respectively. Data pre-processing including applying the corrections and masking vegetation is an important step in the data mining process. Therefore, the image was mapped to the Universal Transverse Mercator (UTM) zone 40N and was georeferenced to the World Geodetic System 1984 for applying the geometric correction. The radiometric correction is performed for this data based on the Log-Residual and LAR reflectance corrections [49].

### 5.2. Principal component analysis (PCA)

PCA is one of the most powerful multivariate techniques used for data analysis and processing [50]. PCA is broadly used in many fields such as data compression, image processing, visualization, exploratory data analysis, pattern recognition, and time series prediction. The initial data is transformed to a new set of variables (principal components) using PCA, which leads to summarize the features of the data and to reduce the dimensions [51]. In this work, PCA was used on the original 14 ASTER bands including the SWIR, VNIR, and TIR ones.

### 5.3. Band ratio

Band ratio is one of the multi-spectral image processing methods that is calculated by dividing the image pixel values or one spectral band on the corresponding image pixel or the other bands. This method can enhance the spectral reflectance by reducing the effect of topography [52]. The hydrothermal zone is characterized by the mineral assemblages, which indicates the diagnostic spectral absorption features [18]. According to the spectral characteristics of the index minerals (Figure 2), the band ratios were calculated. For example, the chlorite, epidote, and calcite minerals are known as the index minerals of the propylitic alteration. On the other hand, Fe, Mg-OH absorption features and CO3 features determine the propylitic reflectance spectra by molecular vibrations in the index minerals of the propylitic alteration [53]. The spectra for epidote, chlorite, and carbonate display the absorption features in the 2.35 µm region, which coincide with ASTER band 8 [18]. The kaolinite and alunite spectra, which are the index minerals of the argillic alteration, show an absorption in ASTER band 5 situated at the 2.17 µm region. The reflectance spectra of Muscovite and illite, as the index minerals of the phyllic alteration, indicate an intense Al-OH absorption feature at 2.20 µm that coincide with ASTER band 6 [18]. According to these spectral features and the previous studies, band ratios 9/8, (7+9)/8, (4+6)/5, 7/6, and (5+7)/6 have been determined for enhancing the alterations in this area [54, 55].

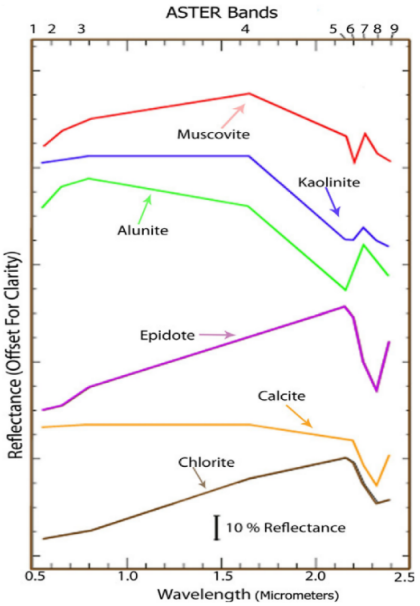


**Figure 2. Reflectance spectra for popular hydrothermal minerals [18, 56].**

## 6. Preparing dataset

The three training datasets include14 band ASTER and 19 derivative data of original ASTER by applying the band ratio and PC techniques (Table 1). These derivative datasets are expected to classify better than the original datasets [25]. Three alteration zones, namely argillic, propylitic, and phyllic are defined as the target classes in the classification.

**Table 1. Input data for SVM- and RF-based alteration classification.**

| Dataset name | Description | Number of data layers |
|---|---|---|
| ASTER | Original aster band | 14 |
| PC | Principal component of ASTER | 14 |
| Band ratio | Band ratio of ASTER | 5 |

The training samples were selected from known locations and ASTER imagery representing the mentioned alterations. The ASTER imagery with RGB (4, 6, 8) that expose the alteration zones are used as the representative training areas with the geological knowledge of the studied area. A sufficient number of training samples were selected for each one of the target classes. The number of training samples is 10*n* pixels for each class was proposed by [57], whereas 10*n* to 30*n* pixels were suggested by [58] and [59], where *n* is the number of bands. Therefore, in this work, the number of pixels for each class is almost 10*n* (Table 2).

**Table 2. Number of pixels for training and validation areas used in the Chahar Gonbad area for classification of ASTER data by SVM and RF.**

| Alteration class | SVM and RF methods | |
|---|---|---|
| | No. of training pixels | No. of validation pixels |
| Propylitic | 2067 | 886 |
| Argillic | 642 | 275 |
| Phyllic | 567 | 243 |
| | Total = 3276 | Total = 1404 |

## 7. Application of SVM to studied area

In this section, the data is classified by the SVM algorithm by the one-against-one strategy. The argillic, propylitic, and phyllic are used as the target classes, and 14 band ASTER, 5 band ratios, and 14 band PC are used as the training datasets (Table 1). The values for all the input data layers were extracted for each training point location and formatted as the feature vectors to create the training datasets [25]. The process of training data is controlled in order to prevent over-learning error in the training process and to validate the performance of the SVM classifier by the validation data [25, 29]. Therefore, 70% of the total samples was randomly selected as the training data and 30% of them as the independent validate data. A non-linear kernel function is often applied since the remotely sensed data is not linearly separable. Also the Radial Basis Function (RBF) kernel is widely used in remote sensing methods [26, 27, 60]. The optimization SVM parameters (γ, C) are obtained by the grid search method in order to achieve the best model. A grid search was performed by increasing the values of the parameters, while cross-validating each set of values by error and truth so that the values 0.5 and 2 were determined as γ and C, respectively. Finally, the alteration map of the total samples of the studied area was produced by the SVM technique with the optimum parameters (Figure 3).

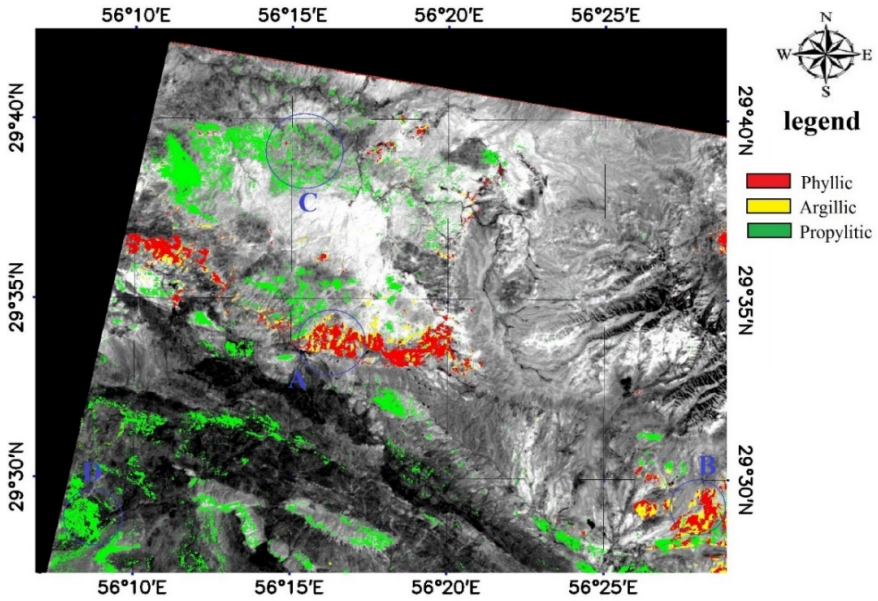


**Figure 3. Alteration map derived from the classification ASTER image using SVM.**

In order to evaluate the performance of the SVM method, the confuse matrix and the overall accuracy were calculated and presented in Table 3. According to the results obtained, the best classification was assigned to the propylitic alteration and the lowest accuracy was assigned to the argillic alteration. As indicated, a large amount of the argillic alteration was detected as the phyllic

alteration; these alterations had the same behavior and occurred very close together. Additionally, the phyllic alteration is often associated with the argillic alteration, which occurs at lower temperatures and dominantly affects plagioclase [61].

**Table 3. Confusion matrix for validation dataset by SVM.**

| | Argillic | Phyllic | Propylitic |
|---|---|---|---|
| Unclassified | 0 | 0 | 47 |
| Argillic | 180 | 67 | 3 |
| Phyllic | 89 | 176 | 23 |
| Propylitic | 6 | 0 | 813 |
| Total | 275 | 243 | 886 |
| Proportion correct | 0.654 | 0.724 | 0.917 |
| Overall accuracy | (1169/1404) = 0.83 | | |
| Kappa coefficient | 0.6994 | | |

## 8. Application of RF to studied area

In this section, the RF method is applied for classification alteration by an IDL-based remote sensing software called EnMAP-Box, which exists at Environmental Mapping and Analysis Program ([62]; EnMAP, http://www.enmap.org). All the three input datasets including ASTER data, PC images, and band ratio were used for RF classification. The number of variables in each node of the tree (*Mtry*) and the number of trees (*Ntree*) are required as the RF parameters to develop the model. The square root of the number of variables is suggested to select for *Mtry* [32, 47, 63]. There were 33 variables (14 ASTER band, 14 PC, 5 band ratios) in this work, and therefore, *Mtry* was set to 6 and 100 was set as the default for *Ntree*. Three input datasets including 14 band ASTER, PC, and band ratio were applied as the training datasets for classification by the RF model; then the accuracy of the model was calculated (Table 4). The confuse matrix and Kappa coefficient were calculated for validating RF classification. According to the results of the accuracy, the highest classification accuracy (0.855) is referred to propylitic and the lowest accuracy (0.567) was referred to argillic, the same as the result for the SVM accuracy. Comparing Tables 4 and 5, it can be concluded that the accuracy of argillic and propylitic in SVM is greater than their equivalent values in RF as well as the overall accuracy, whereas the proportion correct phyllic in RF is almost higher than the equivalent value in SVM.

Finally, the alteration map of the total samples of the studied area was produced using the RF model (Figure 4).

**Table 4. Confusion matrix for validation dataset by RF.**

| | **Argillic** | **Phyllic** | **Propylitic** |
|---|---|---|---|
| Unclassified | 0 | 0 | 95 |
| Argillic | 156 | 46 | 30 |
| Phyllic | 73 | 197 | 3 |
| Propylitic | 46 | 0 | 758 |
| Total | 275 | 243 | 886 |
| Proportion correct | 0.567 | 0.810 | 0.855 |
| Overall accuracy | (1111/1404) = 0.791 | | |
| Kappa coefficient | 0.5971 | | |

## 9. Discussion

The two efficient methods RF and SVM are able to separate the phyllic, argillic, and propylitic alteration zones. These two methods have responded well to the classification and identification of alterations due to the complexity of the trigonometric algorithm. The best classification method was selected based on the amount of misclassification and accuracy. According to Tables 3 and 4, the overall accuracy of the RF method is lower than that for the SVM method (about %4 less), which is not very significant. The results obtained show that the performance of SVM is more powerful than that for RF in the classification of alterations in this research work. In addition, the value of the Kappa coefficient for SVM is approximately 0.1 more than the RF Kappa coefficient. Some important findings are easily extracted from the alteration maps accordingly; three main alterations were determined, which were mainly located in the center, northwest, and southeast part of the studied area; therefore, these regions require further studies. However, the performance of the SVM and RF algorithms were analyzed statistically by individual, overall accuracies, and Kappa coefficient, alteration maps were not reviewed. In order to meet this objective, some places of the alteration maps derived from the RF and SVM algorithms show different behaviors. According to point A in Figures 3 and 4, the RF model has a good ability in allocating the phyllic classes, while some

of the argillic pixels are allocated to the phyllic class mistakenly by the SVM model representing point B in Figures 3 and 4. It can be concluded that the accuracy of SVM is stronger than RF in this case (Tables 3 and 4). Despite the high-intensity propylitic alteration in the southern part of the alteration map produced by SVM, there is a low intensity of them in the northern part of the studied area, although it has conversely occurred in the alteration map produced by RF, which is illustrated in points C and D in Figures 3 and 4. According to the achieved results and comparing the performance of the two methods, the SVM method was selected as a powerful method with the RBF kernel function for the alterations classification in this work.

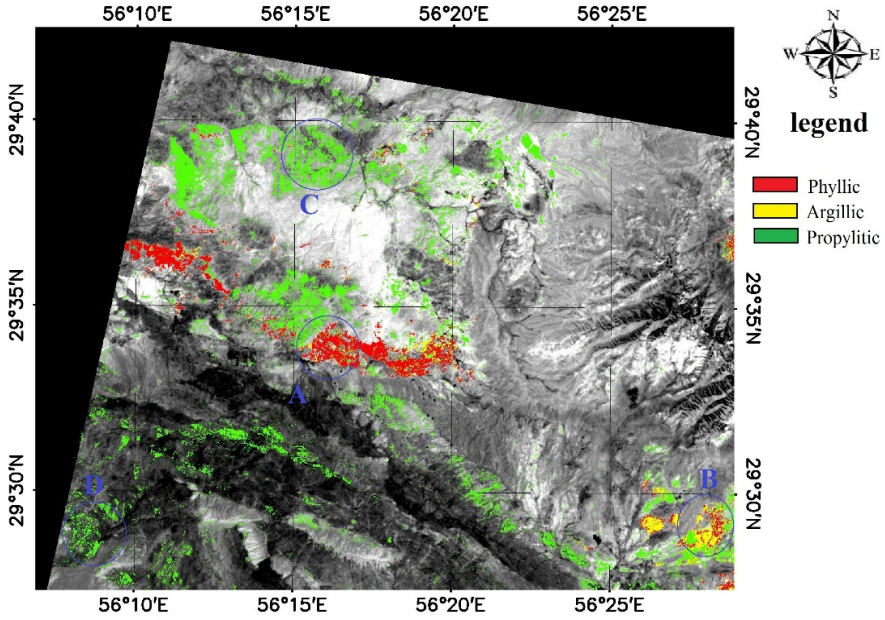


**Figure 4. The alteration map derived from the classification ASTER image using RF.**

## 10. Validation of results

Ground control points, which are very important and sensitive, directly affect the accuracy of the mathematical model. Ground control points can be selected from various sources such as previously corrected images with proportional scale or digital base maps. Some notes must be considered in choosing the points such as their stability and precision on the image and their distribution in the whole image. An adequate number of ground control points (x, y, z) was selected in the UTM coordinate system and the WGS elliptical, 40 N zone on the image and layers, based on the layers extracted from the 1:25000 maps of the area. Field checking was performed using field observation and geological maps. It was concluded that the representative samples of the index alterations were matched with ground control points with their location recorded using a global positioning system (GPS).

**Table 5. Location of ground control points with the observation results.**

| ID | X | Y | Alteration type |
|---|---|---|---|
| G1 | 433033 | 3269764 | Argillic, Phyllic, Propylitic |
| G2 | 430458 | 3269965 | Argillic, Phyllic |
| G3 | 428994 | 3274913 | Phyllic, Propylitic |
| G4 | 432262 | 3272909 | Phyllic, Propylitic |
| G5 | 426203 | 3271522 | Phyllic, Propylitic |
| G6 | 423228 | 3281357 | Propylitic |
| G7 | 419204 | 3275438 | Argillic, Phyllic |
| G8 | 423598 | 3273757 | Argillic, Phyllic |
| G9 | 445721 | 3262087 | Argillic, Phyllic |
| G10 | 426419 | 3280648 | Propylitic |
| G11 | 427159 | 3280818 | Propylitic |

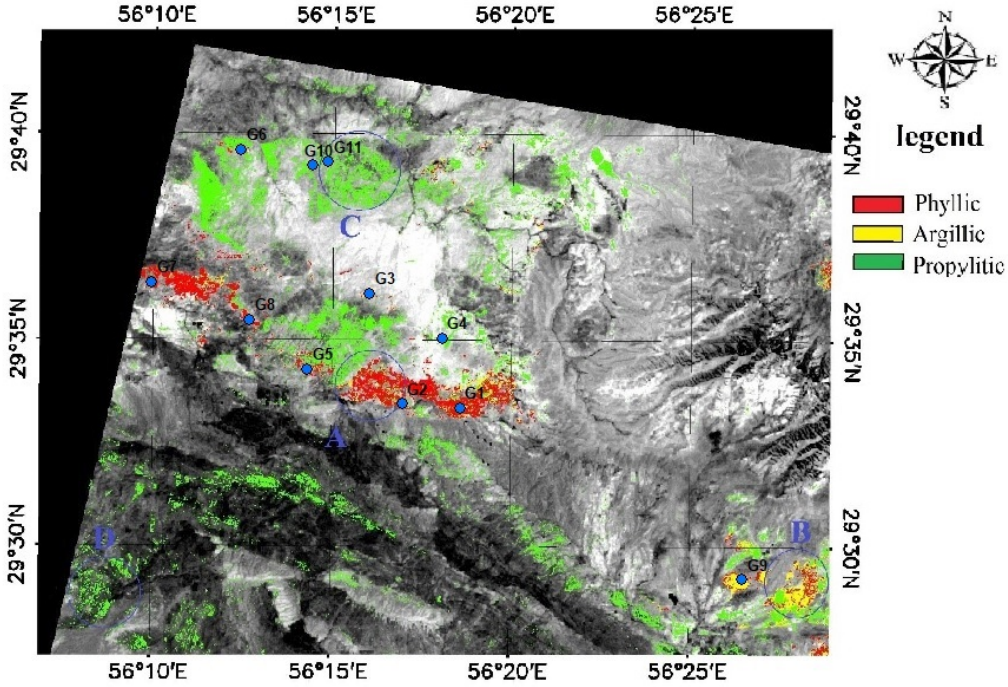


**Figure 5. Location of ground control points on the alteration map achieved by the RF algorithm.**

## 11. Conclusions

This work suggests the machine learning-based methods such as SVM and RF for remote sensing data classification. The classification of remotely sensed data is requested to use a high-resolution data; therefore, these methods were implemented using the ASTER data. After reasonable processing on the data, the 14 bands original ASTER and derived ASTER data such as PC and band ratio as the training data were used to classify. Radial basis function kernel and optimum parameters were selected for the SVM algorithm to classify the alterations of the studied area. The results obtained illustrate that the machine learning methods can achieve acceptable accuracies and help to save time and cost. Also these results indicate that SVM is competitive with the RF algorithm in terms of classification accuracy of remotely sensed data in the studied area. The results illustrate that the SVM method outperform the RF classifier in terms of the overall accuracy by about 4% and Kappa coefficient of about 0.1. Therefore, SVM can be used to increase the classification accuracy.

## References

[1]. Zhu, Y., An, F. and Tan, J. (2011). Geochemistry of hydrothermal gold deposits: a review. Geoscience Frontiers. 2(3): 367-374.

[2]. Mahvash Mohammadi, M.N., Hezarkhani, A. and Saljooghi, B.S. (2016). Separation of a geochemical anomaly from background by fractal and U-statistic methods, a case study: Khooni district, Central Iran. chemie der erde-geochemistry. 76 (4): 491–499.

[3]. Cooke, D., Hollings, P. and Walsh, J. (2005). Giant porphyry deposits: characteristics, distribution, and tectonic controls. Economics geology. 100, 801–818.

[4]. Li, J.W., Zhao, X.F., Zhou, M.F., Vasconcelos, P., Ma, C.Q. and Deng, X.D. (2008). Origin of the Tongshankou porphyry–skarn Cu–Mo deposit, eastern Yangtze craton, Eastern China: geochronological, geochemical, and Sr–Nd–Hf isotopic constraints. Mineralium Deposita. 43 (3): 315–336.

[5]. John, D.A., Ayuso, R.A., Barton, M.D., Blakely, R.J., Bodnar, R.J. and Dilles, J.H. (2010). Porphyry copper deposit model, chap. B of Mineral deposit models for resource assessment: U.S. Geological Survey Scientific Investigations Report. pp: 169.

[6]. Elliott, J.E., Wallace, C.A., Lee, G.K. Antweiler, J.C., Lidke, D.J. and Rowan, L.C. (1992). Maps showing mineral resource assessment for skarn deposits of gold, silver, copper, tungsten, and iron in the Butte 1◦ × 2◦ Quadrangle, Montana: U.S. Geological Survey Miscellaneous Investigations Series. I-2050-E.

[7]. Pohl, C. and Van Genderen, J.L. (1998). Review article multisensor image fusion in remote sensing: concepts, methods and applications. International journal of remote Sensing. 19(5): 823-854.

[8]. Abrams, M. Tsu, H. Hulley, G. Iwao, K., Pieri, D., Cudahy, T.J. and Kargel, J. (2015). The advanced spaceborne thermal emission and reflection radiometer (ASTER) after fifteen years: review of global products.

International Journal of Applied Earth Observation. Geoinf. 38: 292–301.

[9]. Ninomiya, Y. Fu, B. and Cudahy, T.J. (2005). Detecting lithology with Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) multispectral thermal infrared "radiance-at-sensor" data. Remote Sensing and Enviroment. 99:127–139. http://dx.doi.org/10.10.1016/j.rse.2005.06.009.

[10]. Pour, B.A., Hashim, M., Park, Y., and Hong, J.K. (2018). Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. Geocarto International, doi.org/10.1080/10106049.2017.1347207.

[11]. Pour, B.A., Park, T.S., Park, Y., Hong, J.K. Zoheir, B., Pradhan, B., and Ayoobi, I Hashim, M. (2018). Application of Multi-Sensor Satellite Data for Exploration of Zn-Pb Sulfide Mineralization in the Franklinian Basin, North Greenland. Remote Sensing, 10, 1186. Doi: 10.3390/rs10081186.

[12]. Aubakar,A.J., Hasim, M., and Pour, B.A. **(**2018). Identification of hydrothermal alteration minerals associated with geothermal system using ASTER and Hyperion satellite data: a case study from Yankari Park, NE Nigeria. Geocarto International, DOI: 10.1080/10106049.2017.1421716.

[13]. Massironi, M., Bertoldi, L., Calafa, P., Visona, D., Bistacchi, A., Giardino, C., and Schiavo, A. (2008). Interpretation and processing of ASTER data for geological mapping and granitoids detection in the Saghro massif (eastern Anti-Atlas, Morocco). Geosphere. 4 (4), 736–759.

[14]. Madani, A.A. and Emam, A.A. (2011). SWIR ASTER band ratios for lithological mapping and mineral exploration: a case study from El Hudi area, southeastern desert, Egypt, Arab Journal of Geoscience 4:45-52. https://doi.org/10.1007/s12517-009-0059-8

[15]. Fuchs, M., Awan, A.A., and Akhtar, S.S. (2017). Lithological mapping with multispectral data – setup and application of a spectral database for rocks in the Balakot area, Northern Pakistan., J MT SCI. 14: 948–963.

[16]. Yousefi, T., Aliyari, F., Abedini, A. and Calagari, A. A. (2018). Integrating geologic and Landsat-8 and ASTER remote sensing data for gold exploration: a case study from Zarshuran Carlin-type gold deposit, NW Iran. Arabian Journal Geoscience. 11: 482. https://doi.org/10.1007/s12517-018-3822-x

[17]. Ranjbar, H., Honarmand, M., and Moezifar, Z. (2004). Application of the Crosta technique for porphyry copper alteration mapping, using ETM þ data in the southern part of the Iranian volcanic sedimentary belt, 24, 237–243. https://doi.org/10.1016/j.jseaes.2003.11.001

[18]. Mars, J.C., and Rowan, L.C. (2006). 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 2, 161–186.

[19]. Salati, S., Ruitenbeek, F. Van, Meer, F. Van Der and Naimi, B. (2014). Detection of Alteration Induced by Onshore Gas Seeps from ASTER and World View-2 Data. 3188–3209. doi: 10.3390/rs6043188.

[20]. Mahvash Mohammadi, N. M., and Hezarkhani, A. (2019). A comparison of two classification methods, Support vector machine and maximum likelihood, for separation alteration units in the Takht-e-Gonbad district. Iranian Journal of geology.

[21]. Mountrakis, G., Im, J. and Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal Photogrammetry and Remote Sensing. Elsevier B.V. 66(3): 247–259. doi: 10.1016/j.isprsjprs.2010.11.001.

[22]. Huang, C., Davis, L.S., and Townshend, J.R.G. (2002). An assessment of support vector machines for land cover classification. International journal of remote sensing. 23: 725-749.

[23]. Kavzoglu, T., and Reis, S. (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. GIS science and remote sensing. 45: 330–342.

[24]. Masoumi, F., Eslamkish, T., Abkar, A.A., Honarmand, M. and Jeff, R. (2017). Integration of spectral, thermal, and textural features of ASTER data using Random Forests classification for lithological mapping. Journal African Earth science. Elsevier Ltd. doi: 10.1016/j.jafrearsci.2017.01.028.

[25]. Yu, L., Porwal, A., Holden, E. and Dentith, M.C. (2012), Towards automatic lithological classification from remote sensing data using support vector machines, Computer Geoscience. 45: 229–239. doi: 10.1016/j.cageo.2011.11.019.

[26]. Pal, M., and Mather, P.M. (2005). Support vector machines for classification in remote sensing. International journal of remote sensing. 26: 1007–1011.

[27]. Foody, G.M., and Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. Transactions on Geoscience and Remote Sensing. 42: 1335–1343.

[28]. Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York.

[29]. Shabankareh, M. and Hezarkhani, A. (2016). Application of support vector machines for copper potential mapping in Kerman region, Iran. J African Earth Science. Elsevier Ltd. doi: 10.1016/j.jafrearsci. 11.032.

[30]. Abbaszadeh, M., Hezarkhani, A. and Soltani-mohammadi, S. (2013). An SVM-based machine learning method for the separation of alteration zones in

Sungun porphyry copper deposit. chemie der erde-geochemistry. 73(4): 545–554.

[31]. Kavzoglu, T. and Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. International journal Applied Earth Observation. 11: 352–359. doi: 10.1016/j.jag.2009.06.002.

[32]. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[33]. Jhonnerie, R., Siregar, V. P., Nababan, B., and Budi, L. (2015). Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries. Procedia Environment science. 24: 215–221. https://doi.org/10.1016/j.proenv.2015.03.028

[34]. Krishna, G., Sahoo, R.N., and Pradhan, S. (2017), Hyperspectral satellite data analysis for pure pixels' extraction and evaluation of advanced classifier algorithms for LULC classification. Earth science Information. 1-12.

[35]. Hezarkhani, A. (2006). Mineralogy and fluid inclusion investigations in the Reagan Porphyry System, Iran, the path to an uneconomic porphyry copper deposit. Journal Asian Earth Science. 27: 598-612.

[36]. Hezarkhani, A., (2006). Petrology of the intrusive rocks within the Sungun porphyry copper deposit, Azerbaijan, Iran. Journal Asian Earth Science. 27: 326-340.

[37]. Khaleghi, M., and Ranjbar, H. (2011). Alteration Mapping for Exploration of Porphyry Copper Mineralization in the Sarduiyeh Area, Kerman Province, Iran, Using ASTER SWIR Data. Australian Journal Basic and Applied Science. 5(8): 61–69.

[38]. Alizadeh, M., Ngah, I., Hashim, M., Pradhan, B., and Pour, B.A. (2018). A Hybrid Analytic Network Process and Artificial Neural Network (ANP-ANN) model for urban Earthquake vulnerability assessment. Remote Sensing. 10 (6): 975.

[39]. Afzal, P., Tehrani, M.E., Ghaderi, M. and Hosseini, M.R. (2016). Delineation of supergene enrichment, hypogene and oxidation zones utilizing staged factor analysis and fractal modeling in Takht-e-Gonbad porphyry deposit, SE Iran. Journal Geochemical Exploration. 161: 119–127.

[40]. Hosseini, M.R., Ghaderi, M., Alirezaei, S. and Sun, W. (2017). Geological characteristics and geochronology of the Takht-e-Gonbad copper deposit, SE Iran: a variant of porphyry type deposits, Ore Geology Review. Elsevier B.V. doi: 10.1016/j.oregeorev.2017.03.003.

[41]. Mohammadi, N.M. and Hezarkhani, A. (2018). Application of support vector machine for the separation of mineralised zones in the Takht-e-Gonbad porphyry deposit, SE Iran. Journal of African Earth Sciences. 143: pp.301-308.

[42]. Dimitrijevic, M.D. (1973). Geology of Kerman region. Geological Survey of Iran, Report no. 52, pp: 334.

[43]. Zuo, R. and Carranza, E.J.M. (2011). Support vector machine: A tool for mapping mineral prospectivity, Computer Geoscience. Elsevier, 37(12), 1967–1975. doi: 10.1016/j.cageo.2010.09.014.

[44]. Huang, C.L., and Wang, C.J. (2006). A GA-based feature selection and parameters optimization for support vector machines. Expert System Applied. 31: 231–240.

[45]. Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer Science. pp. 738.

[46]. Verikas, A., Gelzinis, A., and Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. Pattern Recognition. 330-349.

[47]. He, J., Harris, J.R., Sawada, M., and Behnia, P. (2015). A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada's Arctic. International Journal Remote Sensing. 36, 2252–2276. http://dx.doi.org/ 10.1080/014311 61. 20 15.1035410.

[48]. Eisavi, V., Homayouni, S., Yazdi, A.M., and Alimohammadi, A. (2015). Land cover mapping based on random forest classification of multitemporal spectral and thermal images. Environment Monitoring Assessment. 187, 1–14. http://dx.doi.org/10.1007/s10661-015-4489-3.

[49]. Green, A.A., and Craig, M.D. (1985). Analysis of aircraft spectrometer data with logarithmic residuals, proceedings of the airborne imaging spectrometer data analysis workshop. 111-119.

[50]. Jolliffe, I.T. (1986). Principal Component Analysis. New York: Springer-Verlag.

[51]. Jolliffe, I. (2002). Principal component analysis. Springer. 2nd edition.

[52] Sabins, F. (1999). Remote sensing for mineral exploration, Ore Geology Review. 14:157-183.

[53]. Spatz, D.M., and Wilson, R.T. (1995). Remote sensing characteristics of porphyry copper systems, western America Cordillera. In: Pierce, F.W., Bolm, J.G. (Eds.), Arizona Geology Society Digest. 20: 94–108.

[54]. Beiranvnd Pour, A. and Hashim, M. (2011). Identification of hydrothermal alteration minerals for exploring of porphyry copper deposit using ASTER data, SE Iran, Journal Asian Earth Science. 42: 1309–1323. doi: 10.1016/j.jseaes.2011.07.017.

[55] Hewson, R.D., Cudahy, T.J., Mizuhiko, S., Ueda, K., and Mauger, A.J. (2005). Seamless geological map generation using ASTER in the Broken Hill-Curnamona Province of Australia. Remote Sensing Environment. 99: 159–172.

[56]. Clark, R.N., Swayze, G.A., Gallagher, A., King, T.V.V., and Calvin, W.M. (1993). The US Geological Survey, Digital Spectral Library. Version 1: 0.2 to 3.0 microns. US Geological Survey Open File Report 93-592. pp: 1340.

[57]. Jensen, J.R. (2005). Introductory digital image processing: a remote sensing perspective. Pearson Prentice Hall. 3rd edition. Upper Saddle River, NJ.

[58] Mather, P.M. (1999). Computer Processing of Remotely Sensed Images. John Wiley and Sons. 2nd Edition. Chichester. P: 352.

[59]. Piper, J. (1987). The effect of zero feature correlation assumption on maximum likelihood based classification of chromosomes. Signal Process. 12: 49–57. http://dx.doi.org/10.1016/0165-1684 (87)90081-8.

[60]. Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., and Bandopadhyay, S. (2008). An objective analysis of support vector machine based classification for remote sensing. Math Geology Science. 40: 409–424.

[61]. Parry, W.T. Jasumback, M. and Wilson, P.N. (2002). Clay Mineralogy of Phyllic and Intermediate Argillic Alteration at Bingham, Utah. Economic Geology. 97: 221–239.

[62]. Van der Linden, S. Rabe, A.; Held, M. Jakimow, B. Leitão, P.J. Okujeni, A. Schwieder, M. Suess, S. and Hostert, P. (2015). The EnMAP-Box—A Toolbox and Application Programming Interface for EnMAP Data Processing Remote Sensing .7: 11249-11266.

[63] Gislason, P.O., Benediktsson, J.A., and Sveinsson, J.R. (2006). Random forests for land cover classification. Pattern Recognition Letter. 27: 294–300. http://dx.doi.org/10.1016/j.patrec.2005.08.011.

# مقایسه روش‌های ماشین بردار پشتیبان و جنگل تصادفی به منظور طبقه‌بندی زون‌های دگرسانی با استفاده از داده‌های سنجش از راه دور

**ندا ماهوش محمدی، اردشیر هزارخانی***

**دانشکده مهندسی معدن و متالورژی، دانشگاه صنعتی امیرکبیر، تهران، ایران**

ارسال 2019/1/12، پذیرش 2019/7/23

* نویسنده مسئول مکاتبات: ardehez@aut.ac.ir

**چکیده:**

یکی از مهمترین مراحل اکتشاف ذخایر، شناسایی و به نقشه درآوردن دگرسانی‌ها منطقه است. مطالعات صحرایی برای تهیه نقشه‌های طبقه‌بندی، وقت گیر و پر هزینه می‌باشد، بنابراین پردازش داده‌های سنجش از راه دور که داده‌های سریع و چند باندی است، می‌تواند جایگزین مطالعات صحرایی شوند. در این تحقیق، از تصاویرASTER برای طبقه‌بندی دگرسانی با استفاده از دو الگوریتم جدید یادگیری ماشین از جمله روش‌های جنگل تصادفی و ماشین بردار پشتیبان استفاده شده است. 14 باند ASTER و 19 لایه استخراج شده از ASTER که شامل تصاویر نسبت باندی وآنالیز مؤلفه‌های اصلی است، به عنوان مجموعه داده‌های آموزشی برای بهبود نتایج استفاده شده است. مقایسه نتایج عددی حاصل از دو روش ذکر شده، تأیید کرد که مدل ماشین بردار پشتیبان دارای دقت و عملکرد بهتری نسبت به مدل جنگل تصادفی برای طبقه‌بندی دگرسانی‌های منطقه مورد مطالعه است.

**کلمات کلیدی:** طبقه‌بندی، یادگیری ماشین، جنگل تصادفی، ماشین بردار پشتیبان، ASTER، دگرسانی، مس پورفیری.