

Lithological classification of Neoproterozoic rocks, Gabal El-Sabbagh area, South East Sinai Egypt, Using Support Vector Machine (SVM) technique.

Ibrahim Abu El-Leil, Nehal M. A. Soliman**, Mahmoud H. Elyaseer**

** Faculty of Sciences Al-Azhar University.*

*** National Authority for Remote Sensing and Space Sciences.*

Abstract.

Support vector machine (SVM) algorithm is applied in the study area to make an automated lithological classification by using Landsat-8 imagery. A series of SVMs was tested using various combinations of input datasets selected including the original 11 bands and datasets extracted from the original bands in order to determine the optimal inputs that provide the highest classification accuracy. Independent components is the highest overall classification accuracy of 84.05 percent for lithological classes on independent validation samples.

Key Words:

Classification, Gabal El-Sabbagh, Landsat-8, Sinai, SVM.

1. Introduction.

Spectral remote sensing data from space-and air-borne sensors have been widely applied to geological mapping, especially in areas of high out crop density in arid regions. There is a significant body of remote sensing literature that describes various data sets and image processing techniques that are used in geologic remote sensing, in particular for lithological (e.g., Rowan and Mars, 2003; Chen et al.; 2007Massironi et al., 2008), mineral (e.g., Sabins, 1999; Hewson et al., 2005; Hubbard and Crowley, 2005; Hewsonetal., 2006), hydrothermal-alteration (e.g., Crosta et al., 1998;) and structural mapping (e.g., Massironietal., 2008). This approach is now a widely accepted alternative to traditional field work as a cost-effective means of generating

small-scale geological maps especially in inaccessible areas (e.g. Adams et al., 1986; Boardman and Kruse, 1994; Rowan and Mars, 2003; Chen et al., 2007).

In principle, every rock type has a spectral response that can be identified and mapped from satellite images, which is the basis of spectra-based approaches to automated lithological mapping or classification. In practice, however, these spectral responses are conditioned by several factors including atmospheric effects, the spectral and spatial resolution of the image, sub-pixel level heterogeneity in chemical and mineralogical composition of the rock, presence of soil, vegetation cover and weathering. Thus, even for the same rock type, the actual spectra captured from remote sensing measurements can be significantly different from the laboratory spectra of the pure member, even after rigorous pre-processing (i.e. calibration, radiometric correction, geometric correction). This problem is commonly addressed by spectra from field spectrometric observations of homogenous areas for each rock type, which are then used for spectral analysis and matching for lithological interpretations. However, given the coarse spatial resolution of multispectral images, it is often difficult to locate, for every rock type, a sufficient number of pure pixels on the ground that are completely homogenous.

The studied area represents a part of Arabo-Nubian Shield located in Gabal Sabbagh area, South Sinai, bounded by latitude $28^{\circ}10' - 28^{\circ}10'N$ and longitude $34^{\circ}00' - 34^{\circ}15'$ (Fig.1). It is covered by migmatites followed by the metasediments of the Kid group. The Kid group is followed by metagabbros, Older Granitoids, Younger Granitoids, dykes, Alkali Granites and basaltic dykes.

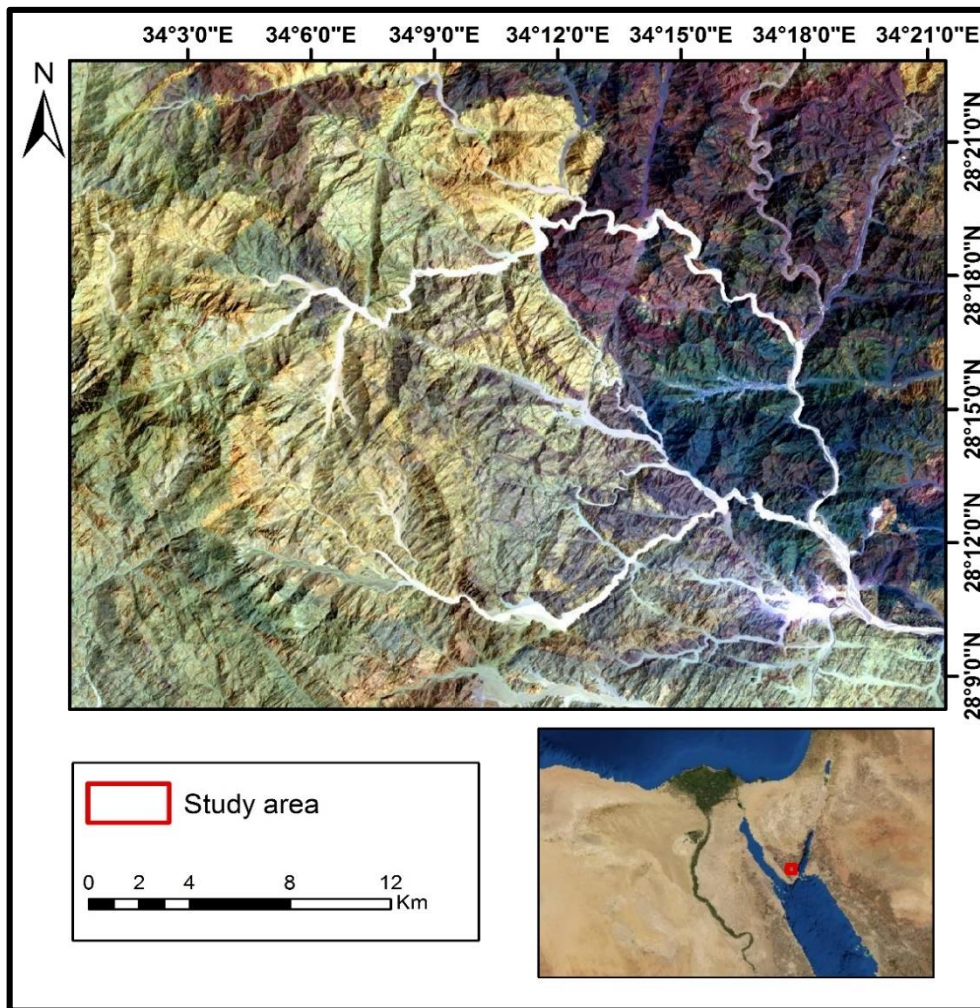


Fig.1. Landsat-8 image (7, 4, 2) in RGB showing the location of the study area.

In this paper, the SVM algorithm was applied to automated lithological classification (for generating a lithological map) of a study area in the south Sinai, Egypt using remote sensing data namely Landsat-8 data, there principal component analysis (PCA), inverse principal component analysis (IPCA), independent component analysis (ICA), inverse independent component analysis (IICA) and some combination between them in the study area using (ROIs).

2. Geologic setting.

The investigated Gabal El-Sabbagh area "Fig.2" is covered mainly by Neoproterozoic rocks that can be arranged from older to younger as the following sequence:

Basalt dikes	Younger
Alkali granite	
Acidic to intermediate dikes	
Younger granites	
Older granites	
Metagabbros	
Metasediments	
Migmatites	Older

The migmatites may relate to the large Feiran- Solaf migmatites belt (Akaad et al, 1967, El Gaby and Ahmed 1980, Abu El-Leil et al, 1995). They are represented by small mass, formed of successive layers of melanosomes and leucosomes. The metasediments cover a considerable area of broadly folded supra structure succession of Kid Group (Shimron, 1980) overlain the migmatites. The metagabbros are represented by relatively small masses, directly intrude the migmatites and the metasediments, whereas the metagabbros are intruded by the older granites and the younger granites (Abu El-Leil et al. 1995). The older granites are represented by highly fractured and foliated rocks directly intruded by the younger granites and by number of dikes particularly along the contact with the younger granites. The younger granites cover a huge area of course grained and medium grained varieties, which gradually grade into each other emphasizing the same age, usually they enclose some clots of mafic aggregation, in addition to some xenoliths, that increase toward the contact with the older granites and the metasediments of the Kid Group,

3. Data used and image processing.

3.1 Landsat-8 data and their image enhancement processing.

The Landsat 8 satellite images the entire Earth every 16 days in an 8 day offset from Landsat 7. Data collected by the instruments onboard the satellites are available to download at no charge from GloVis, Earth Explorer, or via the Landsat Look Viewer within 24 hours of reception. Landsat- 8 carries two instruments:

The Operational Land Imager (OLI) sensor includes refined heritage bands, along with three new bands: a deep blue band for coastal/aerosol studies, a shortwave infrared band for cirrus detection, and a Quality Assessment band.

The Thermal Infrared Sensor (TIRS) sensor provides two thermal bands. These sensors both provide improved signal-to-noise (SNR) radiometric performance quantized over a 12-bit dynamic range. (This translates into 4096 potential grey levels in an image compared with only 256 grey levels in previous 8-bit instruments.) Improved signal to noise performance enable better characterization of land cover state and condition. Products are delivered as 16-bit images (scaled to 55,000 grey levels).

For comparing lithological information discrimination performance, two image enhancement processing algorithms principal component analysis (PCA) and independent component analysis (ICA) were applied to the original 7 Landsat-8 bands. In this paper, PCA and ICA are used as image enhancement techniques, to improve lithology discrimination using Landsat-8 data.

3.2. Support vector machine (SVM).

SVM is a supervised machine learning algorithm that draws on statistical learning theory developed by Vapnik (1995). It functions by separating classes in a dataset with an optimal hyperplane that maximizes the margin between the classes. SVM is a classification system derived from statistical learning theory. The surface is often called the optimal hyperplane, and the data points closest to the hyperplane are called support

vectors. The support vectors are the critical elements of the training set. Consider the simple case of a training dataset that contains only two classes "Fig. 3". An SVM algorithm finds an optimal separating hyperplane by locating two classes bounding hyperplanes (or supporting hyperplanes) and seeking to maximize the margin between them "Fig. 3". The class bounding hyperplanes are parallel to the optimal separating hyperplane and the data points that are located on these supporting hyperplanes are called the support vectors. The algorithm uses a quadratic programming optimization that is constrained by the two supporting hyperplanes to maximize their margin. Note that a so-called hard-margin is applicable when the training dataset is clearly separable in the feature space. However, in real applications, data is often noisy and the classes may not be clearly separable, which can lead to an over the fitting problem. A soft-margin technique, which allows the greater margin may improve generalization abilities but increases the number of outliers (noisy data) that may fall on the wrong side of the margin. To accommodate these outliers, the SVM uses slack variables "Fig. 3" that indicate how far the outlying sample is on the wrong side of the hyperplane that contains the support vectors of its class. Slack variables therefore set the limit to the violation of the constraint imposed by the two supporting hyperplanes. Outlying samples are accommodated by setting slack variables sufficiently large and, at the same time, introducing a penalty term that controls the magnitude of the penalty associated with outlier samples. It is important to tune the value of penalty carefully; a very low penalty value might lead to an inappropriately large number of support vectors, whereas a very high value could lead to over fitting. However, as pointed out by (Belousov et al. 2002), the SVM is generally robust with respect to a relatively wide range of penalty values. Remote sensing data often contain classes that are not linearly separable. In such cases, a non-linear kernel function is used to project the data from input feature space to a higher dimensional feature space and, thus, spread the data points in such a way that they become Separable by a linear hyperplane "Fig. 4". Nonlinear kernels that are commonly used in SVM include a polynomial (homogeneous or heterogenous) function, Gaussian radial basis function, and sigmoid (hyperbolic tangent) function, etc. The radial basis kernel is widely used in remote sensing applications (Melgani and Bruzzone, 2004; Foody and Mathur, 2004; Pal and Mather, 2005; Oommen et al., 2008).

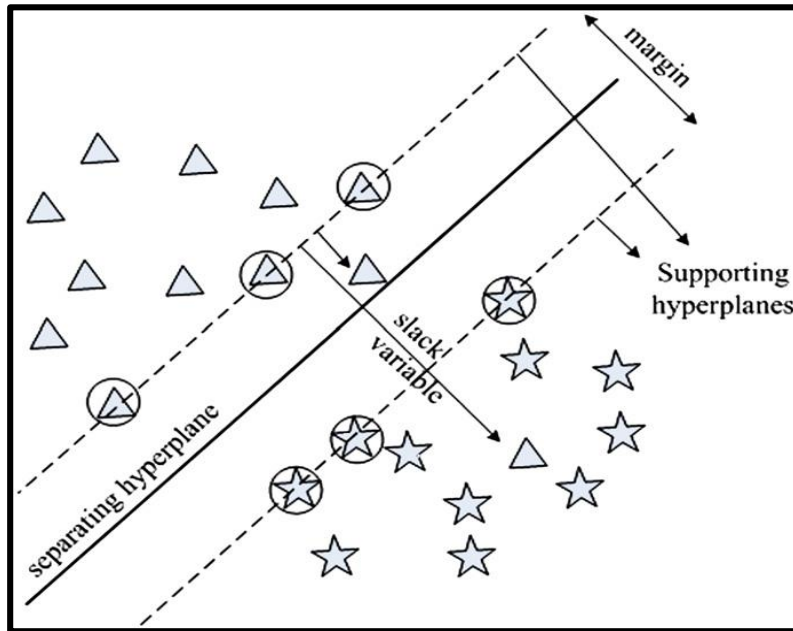


Fig.3. The concept of SVM. In this example, two classes (triangles and stars) need to be separated with a linear hyperplane. The support vectors are bounded with a circle and lie on two class boundary hyperplanes (dashed lines). The optimal separating hyperplane (a solid line) lies between and parallel to the class boundary hyperplanes. The distance between the class boundary hyperplanes is called the “margin” and the distances between a class boundary and outliers of its class indicates “slack variables”.

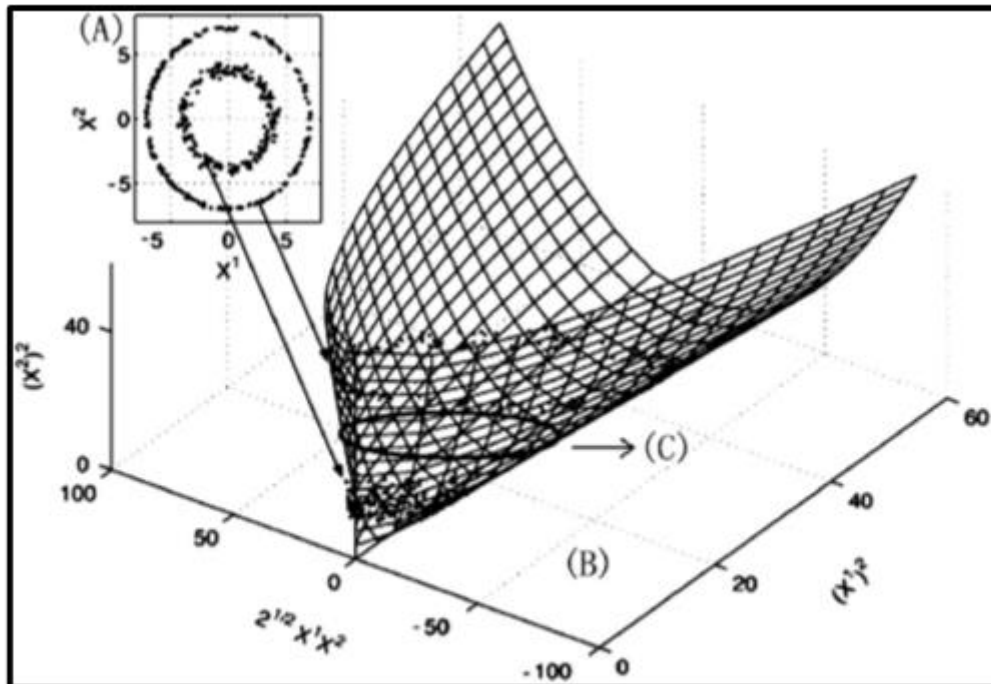


Fig. 4. Projection of data from 2-D to 3-D by means of a polynomial kernel to enable separation of classes using a linear hyperplane; (A) original data pattern, (B) surface to which the data is confined after the mapping and (C) intersection of this surface with the separating hyperplane (From Belousov et al., 2002).

4. Results.

In the present study, SVM was applied to the Landsat8 bands, there principal component analysis (PCA), inverse principal component analysis (IPCA), and independent component analysis (ICA), inverse independent component analysis (IICA) and some combination between them in the study area using (ROIs) "Fig .5".The overall Classification accuracy Table 1 and the results of rock unites in the study area showing in Table 2.

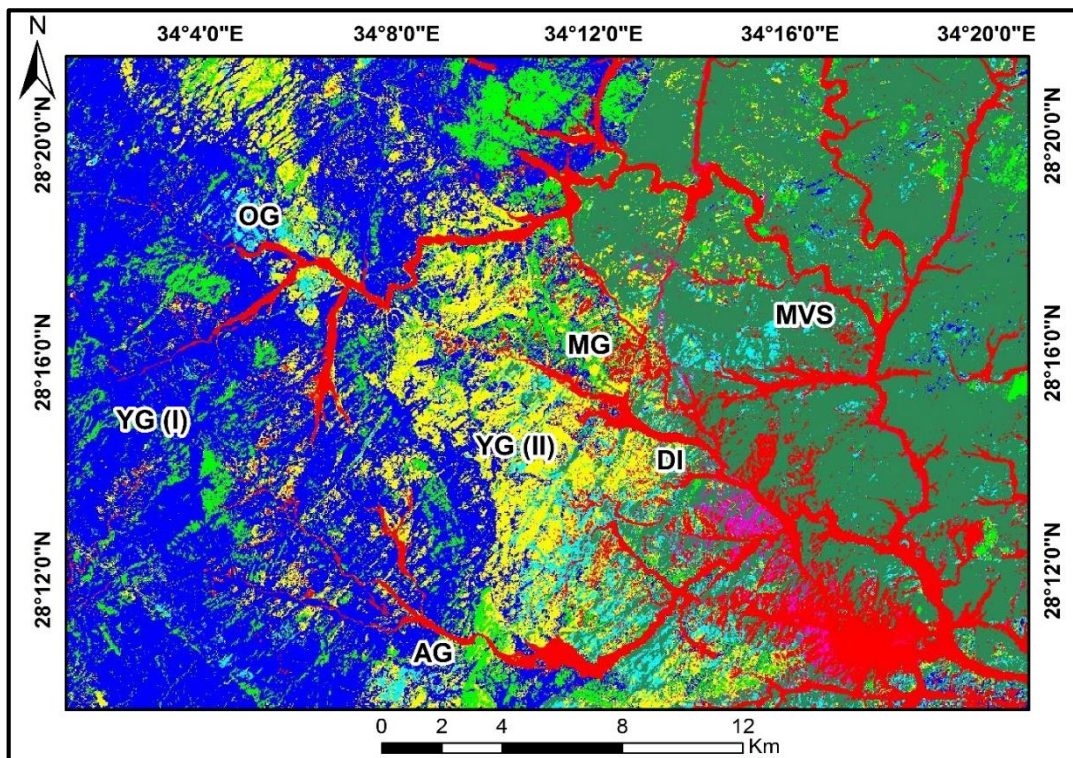


Fig.5: Support vector machine classification for Landsat bands of the study area using (ROIs).

Table.1. Classification accuracies.

Input data	Classification accuracy (%)
ICA	84.05
PC+IC+Landsat8	83.76
PC+ICA	83.73
PC	83.36
IPC+IICA	81.74
IICA	81.61
IPC	81.54
Landsat8	81.52

Table.2. Comparison of SVM classification accuracies (average of producer and user accuracies) for different target classes.

<i>Classes</i>	Landsat-8 Image	PC	IPC	ICA	IICA	PC+ICA	IPC+IICA	PC+IC+Landsat8
Wadi	97.01	97.74	96.95	97.76	97.07	97.77	97.06	97.71
Alkali Granit	71.01	76.09	71.01	76.81	71.01	76.57	70.77	76.09
Younger granitoides II	88.36	88.02	88.52	87.62	88.71	87.65	88.82	87.44
Younger granitoides I	68.38	71.25	67.84	72.38	67.25	71.5	67.81	72.38
Older Granite	44.1	45.63	43.51	46.96	43.98	46.72	43.66	46.92
Meta Gabbros	26.45	33.49	31.59	41.01	31.97	37.96	33.02	49.95
Migmatites	41.23	52.42	43.25	53.86	44.33	53.07	43.75	45.63
Metasediments	86.26	88.59	85.22	89.89	84.85	89.3	85.44	88.66

References.

- [1] Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Remote Sensing of Environment* 84 (3), 350–366.
- [2] Chen, X., Warner, T.A., Campagna, D.J., 2007. Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite, Nevada. *Remote Sensing of Environment* 110, 244–256.
- [3] Massironi, M., Bertoldi, L., Calafa, P., Visona, D., Bistacchi, A., Giardino, C., 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.
- [4] Sabins, F.F., 1999. Remote sensing for mineral exploration. *Ore Geology Reviews* 14, 157–183.
- [5] Hubbard, B.E., Crowley, J.K., 2005. Mineral mapping on the Chilean-Bolivian Altiplano using orbital ALI, ASTER, and Hyperion imagery: data dimensionality issues and solutions. *Remote Sensing of Environment* 99, 173–186.
- [6] Hewson, R.D., Cudahy, T.J., Drake-Brockman, J., Meyers, J., Hashemi, A., 2006. Mapping geology associated with manganese mineralization using spectral sensing techniques at Woodie Woodie, East Pilbara. *Exploration Geophysics* 37, 389–400.
- [7] Crosta, A.P., Sabine, C., Taranik, J.V., 1998. Hydrothermal alteration mapping at Bodie, California, using AVIRIS hyperspectral data. *Remote Sensing of Environment* 65, 309–319.
- [8] Adams, J.B., Smith, M.O., Johnson, P.E., 1986. Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site. *Journal of Geophysical Research* 91, 8089–8112.

- [9] Boardman, J.W., Kruse, F.A., 1994. Automated spectral analysis: a geologic example using AVIRIS data, north Grapevine Mountains, Nevada. Proceedings of the Tenth Thematic Conference on Geologica Remote Sensing, Ann Arbor, MI: ERIM, vol. 1, (pp. 407–418), Ann Arbor, MI: ERIM.
- [10] Belousov, A.I., Verzakov, S.A., von Frese, J., 2002. A flexible classification approach with optimal generalisation performance: support vector machines. *Chermo- metrics and Intelligent Laboratory Systems* 64, 15–25.
- [11] Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing* 42 (8), 1778–1790.
- [12] Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42 (6), 1335–1343.
- [13] Bruzzone, L., Chi, M., Marconcini, M., 2006. A novel transductive SVM for the semisupervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 44, 3363–3373.
- [14] Pal, M., Mather, P.M., 2005. Support vector classification in remote sensing. *International Journal of Remote Sensing* 26, 1007–1011.
- [15] Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., Bandopadhyay, S., 2008. An objective analysis of support vector machine based classification for remote sensing. *Mathematical Geosciences* 40, 409–422.