Carpathian Journal of Earth and Environmental Sciences, February 2015, Vol. 10, No. 1, p. 15 - 26

OPTIMIZED DATA INPUT FOR THE SUPPORT VECTOR MACHINE CLASSIFIER USING ASTER DATA. CASE STUDY: WADI ATALLA AREA, EASTERN DESERT, EGYPT

Safaa M. HASSAN[1], Omar S. SOLIMAN[2], & Amira S. MAHMOUD[1]

[1]National Authority for Remote Sensing and Space Sciences. 23 Joseph Tito Street, El-Nozha El-Gedida (P.O. Box : 1564 Alf Maskan), Cairo, Egypt. E-mail: Safaamh2002@yahoo.com

[2]Cairo University, Faculty of Computers and Information, Cairo, Egypt. Email:Dr.omar.soliman@gmail.com

Abstract: Satellite imagery provides an advanced and cost-effective way to emphasize the lithological information of any poorly mapped area. In this study, a lithological identification method, based on the support vector machine (SVM) classifying algorithm, is proposed to discriminate the widely exposed lithological features around Wadi Atalla area, Central Eastern Desert of Egypt. The SVM classifier has been applied to a series of datasets derived from the Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) imagery with an ASTER-derived Digital Elevation Model (DEM) in order to find the best set of data input for the optimal classification results. Combinations of various input datasets including; the Visible Near Infrared (VNIR) and Shortwave Infrared (SWIR) of the ASTER bands as well as some of its derivatives e.g. Principal Component Analysis (PCA), Independent Component Analysis (ICA), the stacking of both PCA and ICA data (PC/IC-Stack) as well as the ASTER generated DEM are tested for best classification accuracy. A combination of the ASTER-(PCA/ICA and DEM stack) data input provided the highest overall classification accuracy of 95% for the independently validated samples of the lithological classes using the SVM classifier. Results indicate that this particular dataset input can help producing a good lithological distribution map for any remote area that have some background information about its lithology. This new proposed method successfully differentiated between ophiolitic assemblage, highly deformed rocks of Meatiq group, intrusive rocks and Hammamat molasse sediments in the study area.

Keywords: Lithological mapping, ASTER, Classification, Support Vector Machine, Wadi Atalla, Eastern Desert, Egypt.

1. INTRODUCTION

Development in Satellite image processing techniques has reached advanced levels in lithological mapping as well as identifying mineral deposits (Abrams et al., 1983; Abrams & Hook, 1995). The Advanced Space borne Thermal Emission and Reflection radiometer (ASTER) is one of the most widely used sensors in that purpose. It has three separate instrument subsystems (Table 1) that makes the ASTER data of high spatial, spectral and radiometric resolutions (Yamaguchi et al., 1998).

One of the main principal applications of remote sensing data is to create maps of identifiable ground features through assigning image pixels to distinguishable real-world classes using an automated process, the process that is called “image classification”. Using a suitable classification technique is significant in determining the quality of the classification results. Geological studies using multi-spectral remote sensing imagery is usually based on the differences in both physical and chemical properties of each rock type (Yu, et al., 2012). The need of efficient image classifier to classify remote sensing images with high accuracy is essential (Perumal & Bhaskaran, 2011). This is why many extensive works had been done relevantly to this issue (Ninomiya & Fu, 2010; Li, et al., 2011; Salati, et al., 2011; James & Daniel, 2002; Lu, & Weng, 2007). Boser et al., (1992) was the first to introduce the support vector machine (SVM)

classifier. Comparing the SVM with the mostly and widely used, maximum likelihood classifier (MLC), showed that the SVM did provide higher accuracy in terms of the independently validated samples (Yu, et al., 2012; Mondal et al., 2012). Another important factor that is been widely used in the lithological discrimination processes is the elevation data.

Topographic informations have been widely used for lithological mapping, either by themselves (e.g. Barnett, 2004) or in combination with spectral data (Yu, et al., 2012). Different rock units have differently topographical expressions, which depends on their relative susceptibility to weathering. Yu, et al., (2012) implemented a spatial image processing mining for lithological classification using SVM. In their procedure, the SVM algorithm was applied to an automated lithological classification using ASTER imagery, ASTER-derived digital elevation model (DEM) in order to determine the optimal inputs that provide the highest classification accuracy.

Table 1.Characteristics of ASTER Satellite data

Subsys tems	Bands	Spectral range(μm)	Spatial Resolution (m)
VNIR	Band1 Band2 Band3NBand3 B	0.52-0.60 0.63-0.69 0.78-0.86 1.60-1.70	15
SWIR	Band5 Band6 Band7 Band8 Band9	2.145-2.185 2.185-2.225 2.235-2.285 2.295-2.365 2.360-2.430	30
TIR	Band10 Band11 Band12 Band13 Band14	8.125-8.475 8.475-8.825 8.925-9.275 10.25-10.95 10.95-11.65	90

The aim of this study is to develop a classification method for remote sensing ASTER satellite imagery using SVM with Radial base function (RBF) and evaluate the accuracy of classification using various combinations of parameters of RBF in order to get the best lithological identification at Wadi Atalla area. The suitability of combining ASTER derivatives for lithological classification using SVM is evaluated. In addition, a full comparison of the different data inputs to identify the most discriminatory data layers for lithological classification at Wadi Atalla area is applied. Varieties of image processing techniques are applied to different datasets to generate the most enhanced lithological information of the study area. The performance of different SVMs (with different layers) is evaluated by classification accuracy on independent validation samples as well as the similarity with the field work derived, similarity or non- similarity with the published lithological map.

2. GEOLOGIC SETTING OF THE STUDY AREA

The study area as shown in (Fig. 1a) is located along the fold and thrust belt of the Nubian Shield as part of the Neoproterozoic basement rocks of the Central Eastern Desert (CED) of Egypt forming very rugged terrains.

The ophiolitic sequence of the area is composed mainly of ultramafic rocks (serpentines), mafic plutonic rocks (meta gabbro) and meta volcanic (met basalts) (El-Gaby et al., 1984; Habib et al., 1985; Hassan & Hashad, 1990; El-Sayed et al., 1999; Abd El-Rahman, et al., 2009). The best exposures of these ophiolitic units occur in the Fawakhir area, which is located between the Meatiq gneissic dome to the east and the Hamammat sedimentary rocks to the west (Nasseef et al., 1980) along the Qift-Qusier road. Highly deformed meta-sedimentary sequences including metapelites and metasammites and subordinate amphibolites, which suffered amphibolite-facies metamorphic conditions (Neumayr et al., 1995), are also recorded with amphibolite being dispersed in the sheared sedimentary matrix (Ries et al., 1983; El-Gaby et al., 1984; Fowler & Osman, 2001).

The Ophiolite lies between the Hammamat conglomerates to the west and the Meatiq Dome to the east. The western contact between the ophiolite suite and the Hammamat conglomerates is an easterly dipping thrust zone, while its eastern contact with the Meatiq rocks is defined by a zone of tectonic melange, mylonitised rocks and highly deformed gabbro. The contacts between the ultramafics and adjacent rocks are sharp and distinct as shown in (Fig. 2).

The basal contact between the serpentinite and the underlying mélange rocks is sharp and marked by a deep thrust fault trending north-northwest-south-southeast (El-Sayed et al., 1999) and a relatively narrow band of dark green schistose amphibolite is located between the ultramafic and the mélange zone (Hassanen, 1985). Metagabbros (located at the mouth of Wadi Atalla) constitute a major part of the ophiolite suite and have suffered regional metamorphism up to green schist-facies.

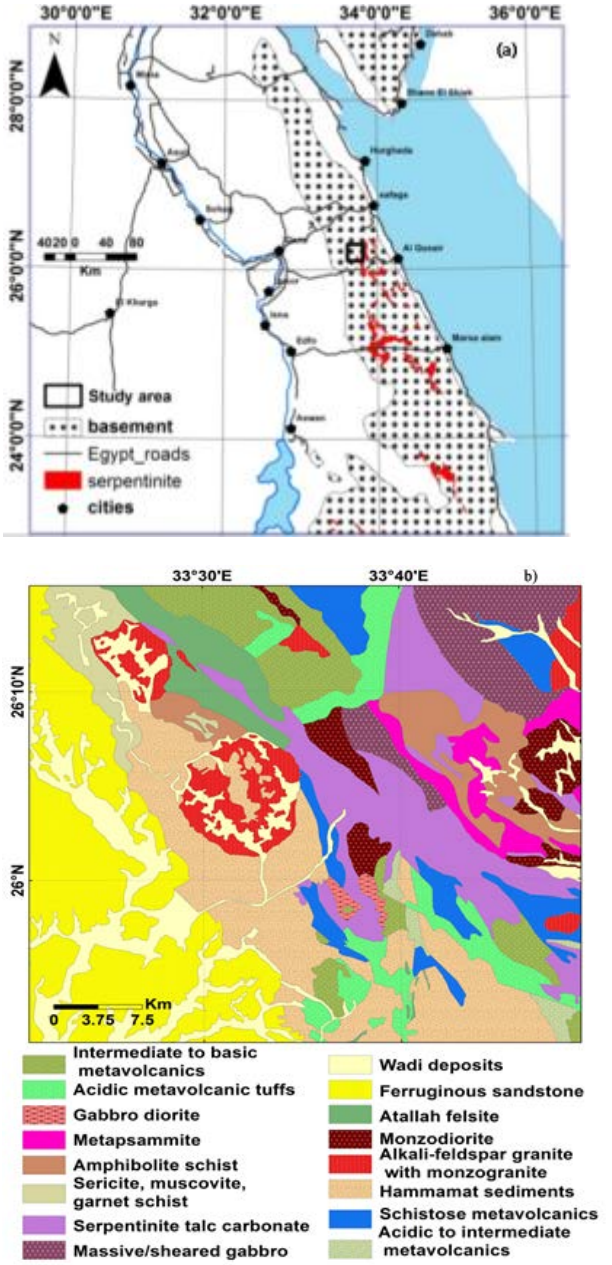

Figure 1. (a) Location map of the study area. (b)Lithological map of the study area (modified after EGSMA, 1992).

Two ophiolitic metagabbro bodies display layering on a regional scale. The contact with the serpentinites is highly sheared and show development of schistose amphibolite rocks. Small irregular bodies of basic metavolcanics are overly the ultramafic/mafic rocks (El-Sayed et al., 1999). The contact between the metavolcanic and ultramafic rocks is fairly sharp, although small fragments of the basic metavolcanics occur in the ultramafic serpentinite. Low-grade regional metamorphism, up to greenschist-facies, has affected the volcanic rocks. The post-ophiolite granitic pluton occupies the western central part of the mapped area, and intrudes both the ultramafic and metagabbro rocks (Harraz & Ashmawy, 1994; El-Sayed et al., 1999; Fowler, 2001). The granitic pluton is characterised by the presence of numerous rounded to elliptical xenoliths with different sizes, which have sharp contacts with the enveloped granitic rocks.

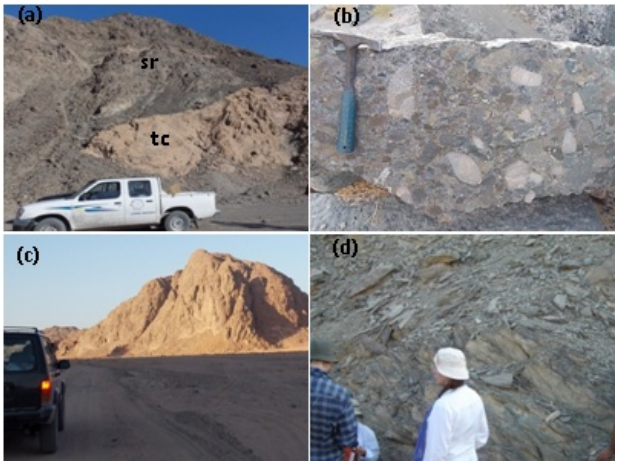

Figure 2. (a) Talc carbonate rocks (tc) emplaced within serpentinites (sr). (b) Green conglomerate in Hammamat sediments. (c) General view of the southern side of Um Had granitic pluton. (d) Sheared Meta Volcanic, Wadi Atalla.

3. MATERIAL AND METHODS

Eight steps of a methodological process have been applied as followings; 1) Identify the study area to determine area of interest (AOI) from remote sensing satellite imagery and gathering the required SWIR and VNIR ASTER bands for it, 2) Generate ASTER DEM of the AOI of study area, 3) Associate the DEM layer with the nine ASTER layers, 4) image segmentation for partitioning a digital image into multiple segments (sets of pixels) or setting region of interest (ROI), 5) Features extraction to transform the input image into a set of features using PCA and ICA algorithms, 6) Image classification for sorting pixels into a finite number of individual classes, or categories of data, based on their data file values. If a pixel satisfies a certain set of criteria, the pixel is assigned to the class that corresponds to those criteria using SVM with RBF functions, 7) Apply post classification using siev, clumb and majority windows techniques, 8) Accuracy assessments calculation of various combination of the data sets, as shown in (Fig. 3).

3.1. ASTER Data

In the present study, the ASTER Level 1B image data acquired on June, 2003 is used. The ASTER data is in terms of radiance at sensor data with radiometric and geometric corrections applied. A spectral stacking of the VNIR and SWIR

(resampled to 15 x15 m per pixel) absorption bands, covering the study area, are processed and analyzed using ERDAS Imagine 2013 and Arc GIS 10.1 softwares. Fieldwork is carried out to mainly check the occurrence, spatial distribution of the exposed rock types where samples are collected for further analysis.

3.2. Digital Elevation Model (DEM)

Topographic attributes have been widely used for lithological mapping, either by themselves (Barnett, 2004) or in combination with spectral data. A Digital Elevation Model (DEM) is a regularly spaced raster grid of elevation values of a terrain.

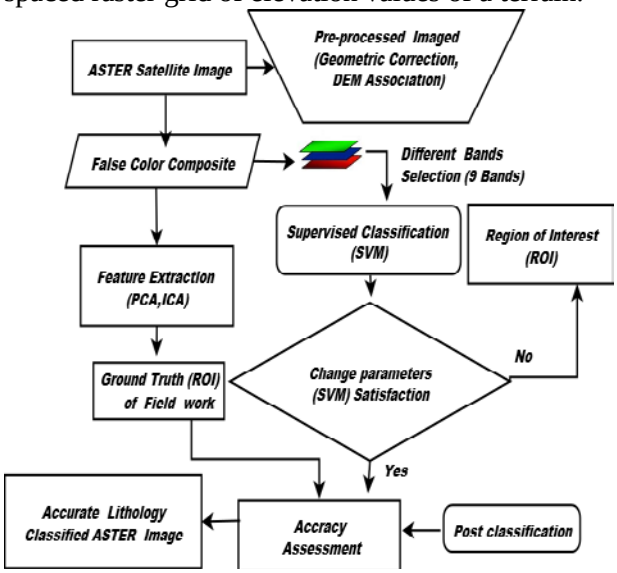

Figure 3. Flowchart shows the applied methodological processes

In remote sensing, DEMs are used in mapping, ortho-rectification, and land classification. Using ASTER band 3N (Nadir view) and 3B (Backward view) as stereo pair for DEM generation. Digital elevation model are the most suitable tool for geological structure and rock units discrimination. Combination between ASTER satellite images and DEM show relationship between topography and geology (Cries et al., 1995). Rock types have topographical terms that depend on their relative susceptibility to weathering. Slope defined as the rate of change of elevation each grid cell in the DEM while the curvature describes how much a surface is curved at a particular point in the landscape. Tightly folded terrain has large curvature values, while flat terrain has zero curvature (Roberts, 2001). In this paper, topographical slope and curvature are calculated from the ASTER DEM (resampled to 15x15 m per pixel). ENVI Topographic Modeling function was used to produce slope and curvature in order to improve the geological information and increase overall accuracy of the classification results.

3.3. Principal and independent Component Analysis (PCA and ICA)

Principal Component Analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. PCA-based spectral enhancements have been successfully used for geological applications and used to enhance the classification results While, the Independent Components (IC) analysis transforms an input dataset into a new dataset containing new bands comprised of a linear combination of the input bands. IC transforms a set of mixed, random signals into components that are mutually independent. The benefit is that IC can distinguish features of interest even when they occupy only a small portion of the pixels in the image. ICA is a type of spectral unmixing method that does not require a priori knowledge of targeted surface materials, e.g. rocks (Gómez et al., 2007). For comparing lithological information discrimination performance the two image enhancement processing algorithms (PCA) and (ICA) were applied to 9 ASTER bands SWIR and VNIR. The combination between PCA and ICA features (18 new layers) were applied and combined with 9 ASTER bands (VNIR and SWIR) as well as the DEM (slope and curvature) in order to improve the performance of classification results, for all the used numbers of components instead of using PCA or ICA separately (Mercier & Lennon, 2003).

3.4. Support vector machine (SVM)

Kernel-based techniques (i.e. SVM, Bayes point machine, kernel principal component analysis, and Gaussian processes) represent a major development in Machine Learning (ML) algorithms. SVMs were first suggested for supervised classification and have recently become an area of intense research owing to developments in the techniques and theory coupled with extensions to regression and density estimation (Burges, 1998). SVM are a group of supervised learning methods that can be applied to classification or regression. SVM is considered as one of the standard tools for ML and data mining. It is applied on many real applications as text categorization, hand-written character recognition, image classification, bio-sequences analysis, etc. It employs kernel to map the

input data into some much higher dimensional feature space implicitly in which data becomes linear separable. The SVM approach seeks to find the optimal separating hyper plane between classes by focusing on the training cases that are placed at the edge of the class descriptors. These training cases are called support vectors. Training cases other than support vectors are discarded. This way, not only is an optimal hyper plane fitted, but also less training samples are effectively used; thus high classification accuracy is achieved with small training sets (Perumal & Bhaskaran, 2011).

Given a training dataset with n samples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where x_i is feature vector in a v-dimensional feature space and with labels $y_i \in -1, 1$ belonging to either of two linearly separable classes C1 and C2. Geometrically, the SVM modeling algorithm finds an optimal hyperplane with the maximal margin to separate two classes, which requires solving the optimization problem;

$$\text{Maximize} \sum_{i=1}^{n} \propto_i - \frac{1}{2}\sum_{j=1}^{n} \propto_i \propto_j y_i y_j . K(x_i, x_j) \quad (1)$$

$$\text{Subject} - \text{to}: \sum_{i=1}^{n} \propto_i y_i, \qquad 0 \leq \propto_i \leq C \quad (2)$$

Where $\propto_i$ is the weight assigned to training sample x_i.If > 0, x_i is called support Vector C is a regulation parameter used to trade-off the training accuracy and the model complexity K is a kernel function, which is used to measure the similarity between two samples. The SVM algorithm provides a choice of four kernel types: Linear, Polynomial, Radial Basis Function, and Sigmoid. These kernel functions are defined as:

Linear Kernel function

$$K(x_i, x_j) = x_i^T x_j \quad (3)$$

Polynomial kernel function:

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \qquad \gamma > 0. \quad (4)$$

Radial Basis kernel function (RBF):

$$K(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}, \qquad \gamma > 0. \quad (5)$$

Sigmoid kernel function:

$$K(x_i, x_j) = \tan(\gamma x_i^T x_j + r) \quad (6)$$

Several kernel functions help the SVM obtain the optimal solution. The most frequently used such kernel functions are the polynomial, sigmoid and radial basis kernel function (RBF). The RBF is generally applied most frequently, because it can classify high-dimensional data, unlike a linear kernel function. Additionally, the RBF has fewer parameters to set than a polynomial kernel.

Parameter C represents the cost of the penalty. The choice of value for C influences on the classification outcome. If C is too large, then the classification accuracy rate is very high in the training phase, but very low in the testing phase. Parameter γ has a much greater influence on classification outcomes than C, because its value affects the partitioning outcome in the feature space (Ding, 2011). Large value for parameter γ results in over-fitting, while a small value leads to under-fitting. Both parameters, the γ and C, depend on the data range and distribution and they differ from one classification problem to another. A common strategy to search for adequate values for g and C is a two-dimensional grid search with internal validation. This strategy is implemented in image SVM in EnMap-Box tool (Linden et al., 2010). Grid search is the most common method to determine appropriate values for C and γ. Values for parameters C and γ that lead to the highest classification accuracy rate in this interval can be found by setting appropriate values for the upper and lower bounds. The optimal value of C and γ in this study, as (100, 100) in the same order for ASTER-DEM (slope, curvature)-PCA-ICA data layer.

4. APPLICATION OF SVM TO THE STUDY AREA

4.1. Input data

For optimized lithological classification of Wadi Atalla area from remote sensing data, we experimented by training a series of SVMs using various combinations of input data selected from among 20 data layers including the original 9 ASTER bands and 11 derivative data layers extracted from the ASTER and DEM data. The derivative data sets contain enhanced spectral and topographic responses of various lithological units to provide better classification accuracy than the original data sets. The target classes for the SVM classification included the sixteen lithological units listed in (Table 2).

4.2. Training procedure and classification data sets

The geological map figure 1a is a helpful tool to select the locations of training sets for different lithological classes.

A suitable number of training pixels were selected for each class, depending on its area extent.

A total of 19470 training samples of the exposed lithological units (e.g. serpentine-talc carbonate, metagabbro, metavolcanic and metapasmmite) have been detected. Training involved determining the optimal values of the SVM parameters, namely, the penalty term "c" and the basis width of the kernel "gamma", which yield the best classification results for the independent validation samples. After identifying the training sets, the SVMs have been applied to classify features extracted from all of the pixel locations within different input data set. A total of 19470 training samples were selected and randomly divided into two parts; first part (75% of the data) was used for training the SVMs and the second part (25%) was used as an independent validation dataset also called test data set for evaluating the classification performance of the trained SVMs that shown in (Table 2)

Table 2. Training and Independent validation data for SVM

Class Name (lithological unit)	No. of samples	
	Training	Independent validation (Test)
Wadi Deposits Qp	1547	516
Monzogranite /alkai feldspar granite Gk	1292	431
Amphibolite schist Ma	1648	947
Hammamat sediments (greywackes & conglomerates), Ha	2921	549
Sericite muscovite garnet schist Msm	152	51
Acidic to intermediate metavolcanics Vi-b	675	225
Atallah felsite Af	841	280
Ferruginous sandstone KuNT	710	237
Monzodiorite Gd	545	182
Gabbro-diorite Mgb	407	136
Intermediate to basic metavolcanics Vb	844	281
Schistose metavolcanics Mvtac	323	108
Serpentinite talc - carbonate rocks Os	853	284
Acidic metavolcanic tuffs Vp-c	377	221
Ophiolitic metagabbro Ohgb	809	126
Metapsammite Mqf	663	270

4.3. Accuracy Assessment

The overall accuracy is calculated as the total number of correctly classified pixels divided by the total number of test pixels. The pixels classified correctly are found along the diagonal of the confusion matrix (Table 3) which lists the number of pixels that were classified into the correct ground truth class. The kappa coefficient (K) is another measure of the accuracy of the classification that takes into account the off-diagonal elements as well. It ranges from –1 (maximum disagreement) to 1 (best agreement), while 0 means that the agreement between reference and classification data is the same that can be obtained by chance. The mathematical equation for calculating kappa coefficient is given by K= (N*A-B) / N2-B

Where N is the total number of pixels, A is the number of correctly classified pixels (Sum of diagonal elements in the confusion matrix), B is the sum of product of row and column total in confusion matrix.

5. RESULTS AND DISCUSSIONS

Lithological classification of remote sensing data requires large number of datasets where each data set contains some relevant information on different rock types at a pixel-level. We evaluate the accuracy of SVM through changing parameters gamma and penalty of Radial Base Function (RBF) for all sixteen lithological classes which represent the widely exposed rock units at Wadi Atalla area. The obtained overall accuracy of all lithological units using ENVI software, confusion matrix and Kappa Coefficient of SVM is calculated. The combination of the nine ASTER bands (SWIR+VNIR), DEM (slope and curvature), Independent Component (IC) and Principal Component (PC) layers give the highest classification accuracy about 95%. Using quantitative analysis, the accuracy assessment focused on the comparison of classifications made with and without the DEM Layer. An error matrix (confusion matrix) as shown in (Table 3) was generated for each dataset and used to derive the overall accuracy and an estimate of Kappa coefficient referred to KHAT (Getman, et al., 2008). Overall accuracy is the most commonly used estimate of accuracy in satellite image classification. Calculations of the producer's and user's accuracy were included to distinguish the contributions for individual lithology classes to the overall accuracy of the classification results. Average accuracy (producer's and user's accuracy) is calculated from different data set layers and listed in (Table 4) The result shows that the use of all derivatives layers performed better, which are useful for lithological discrimination.

The average accuracy comparisons of each rock units in different SVMs are listed in (Table 4) and (Fig. 4). A comparison of the performance of

various SVMs for different classes indicates that the ASTER and ASTER-derived data sets are most useful in classification of the most exposed rock units, In particular, acidic metavolcanic tuffs, basic to intermediate metavolcanics, ferruginous sandstone, Hammamat sediments, amphibolite schist, serpentinite, talc-carbonate, ophiolitic metagabbro and metapasmmite. ASTER (VNIR-SWIR)-DEM-PC-IC SVM has better overall accuracy than ASTER-DEM (Table 5). ASTER (VNIR-SWIR)-DEM-PC-IC SVM has better overall accuracy than ASTER-DEM (Table 5). The ASTER DEM ICA has the best overall accuracy for Sericite muscovite garnet schist, monzodiorite and metapasmmite. ASTER-DEM SVM and ASTER (VNIR-SWIR)-DEM-PC-IC SVM also have equal accuracy for some classes (e.g. amphipolite schist, serpentinite talc-carbonate rocks and ophiolitic metagabbro). Also, sericite muscovite garnet schist has the worst overall accuracy in all the grouped layers ranging from 63 % to 75 % (Table 4).

5.1. Overall classification results

The overall classification accuracies of different data sets for all sixteen classes are listed in (Table 5). All three ASTER-derived products (PC, IC and combination of PC and IC) are compared. The use of the ASTER-derived products, PCA individually or combined with DEM provide higher classification accuracies 93.1% and 95% respectively than directly using the original ASTER data (Table 5).

Table 3. Confusion matrix calculated for the exposed lithological units in the study area using ASTER DEM –PC-IC-SVM, (see table 2 for lithological abbreviations).

	Reference Data																Total
Class	QP	Gk	Ha	Ma	Msm	Vi-b	Af	KuNT	Gd	Mgb	Vb	Mvtac	Os	Vp-c	Ohgb	Mqf	
QP	521	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	521
Gk	0	391	12	1	1	0	5	0	22	7	0	0	0	2	0	0	441
Ha	0	5	903	3	20	0	30	2	0	1	0	0	0	6	0	0	970
Ma	0	0	0	558	0	4	0	0	0	0	0	3	0	0	0	0	565
Msm	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	23
Vi-b	0	0	0	4	0	201	0	0	0	0	5	0	0	0	2	0	212
Af	0	3	18	0	0	0	242	0	1	3	0	0	0	1	0	0	268
KuNT	0	0	1	0	6	0	0	256	1	0	0	0	0	0	0	0	264
Gd	0	17	2	0	0	0	1	1	179	0	0	0	0	0	0	0	200
Mgb	0	7	0	0	0	0	3	0	0	109	0	0	0	0	0	0	119
Vb	0	0	0	0	0	7	0	0	0	0	266	3	0	0	4	0	280
Mvtac	0	0	0	0	0	0	0	0	0	0	6	93	1	0	0	0	100
Os	0	0	0	0	0	0	0	0	0	0	0	0	281	0	0	0	281
Vp-c	0	0	4	0	0	0	2	0	0	0	0	0	3	213	0	0	222
Ohgb	0	0	0	0	0	1	0	0	0	0	4	0	0	0	122	0	127
Mqf	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	271	275
	521	423	940	566	50	213	283	259	203	120	281	102	286	222	128	271	4868

Table 4. Average (producer's and user's accuracy) calculated for the exposed lithological units in the study area using different datasets.

Class Name / lithological unit	ASTER (SWIR-VNIR)	ASTER -DEM	DEM-PCA	ASTER DEM-ICA	ASTER DEM-PCA-ICA
Qp	99	99.5	99	99	99.9
Gk	88	87.8	89	90	90.5
Ma	98.5	98.6	98.6	97.5	98.6
Ha	93.3	93.4	93.4	94	94.5
Msm	64.2	66.7	69.7	75	63
VI-b	93.5	93.5	95	93.7	94.5
Af	86	85.9	86	87.3	87.8
KuNT	96.4	97	97	97.4	97.8
Gd	87	86.7	87.7	90.5	88.8
Mgb	86	86	85.2	85.7	91.2
Vb	93	93.2	93.3	93.2	94.8
Mvtac	90	92	89	89.4	90
Os	99.4	99.4	99.4	99.4	99.4
Vp-c	93.7	94.8	94.2	94.9	96
Ohgb	98.3	99	99	98.3	99
Mqf	95.4	92.6	92.3	96.2	95.2

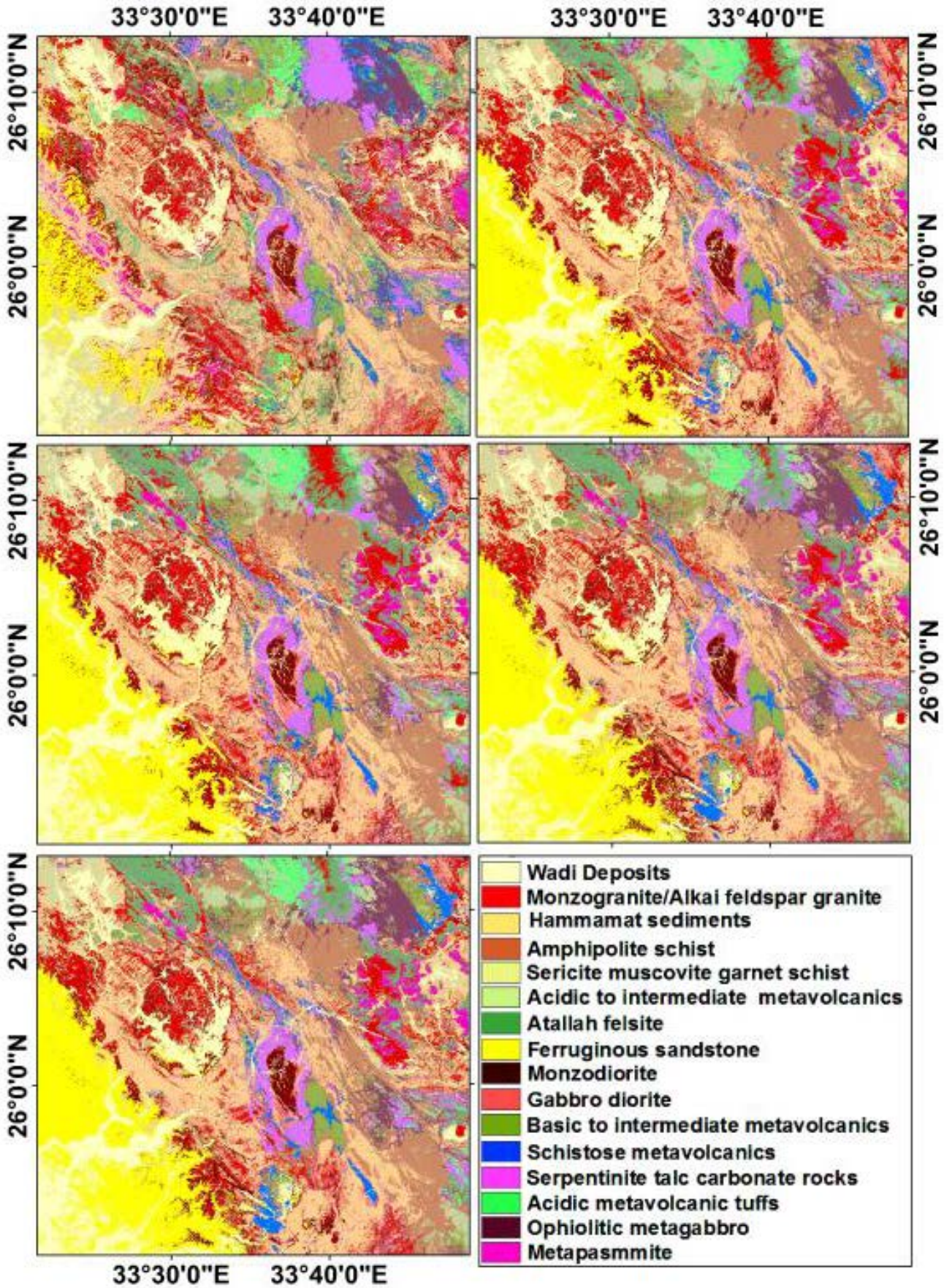

Figure 4. Lithological classification result of the exposed rock units using SVM Classifier (a) ASTER (SWIR-VNIR)-SVM, (b) ASTER -DEM-SVM , (c) ASTER- DEM -PC-SVM, (d) ASTER-DEM-IC-SVM and (e) ASTER-DEM-PC-IC-SVM, (f) Colored scale of the extracted classified images including figs (a-e).

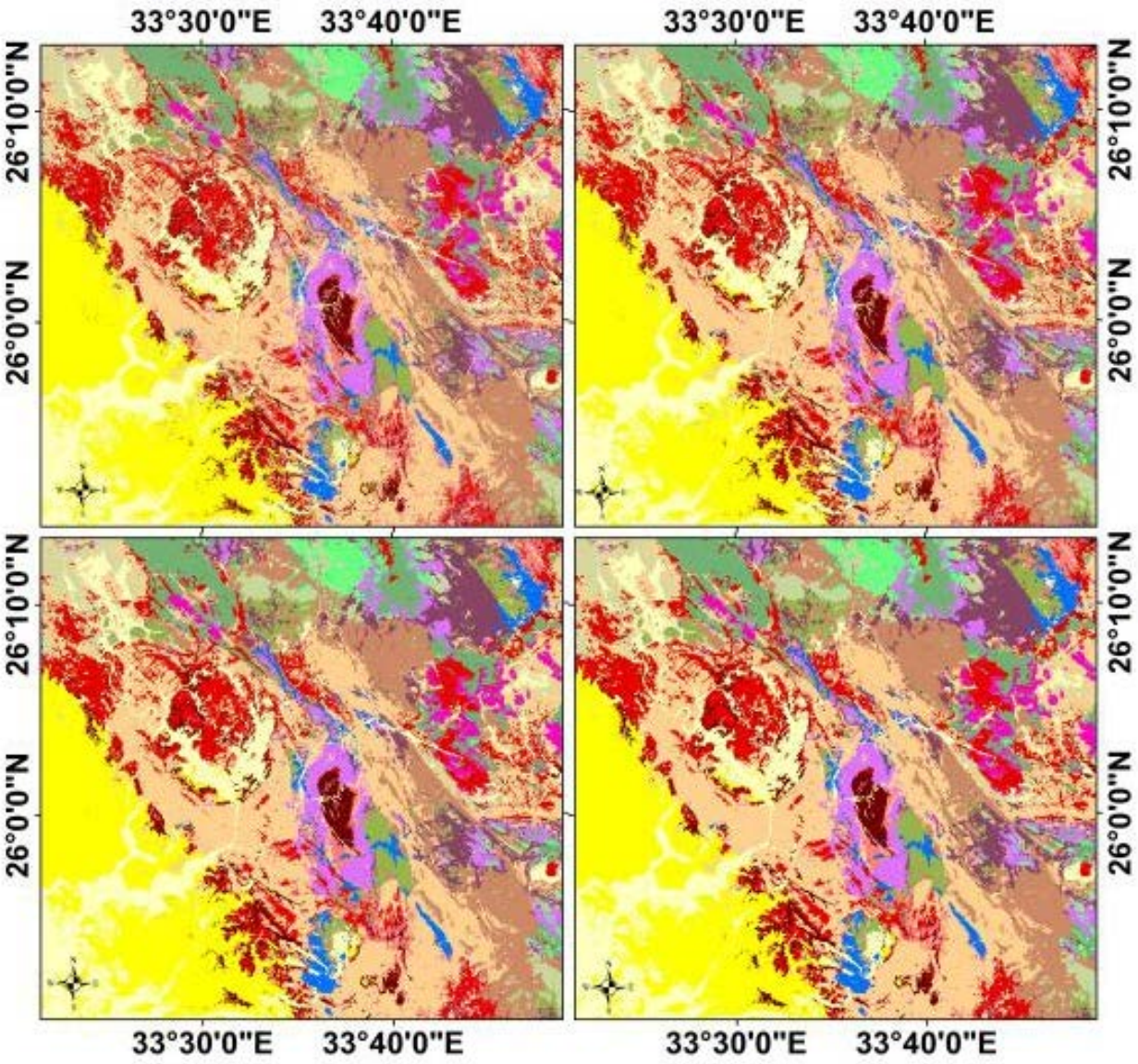

Figure 5. The post-processing result of the ASTER-DEM-PC-IC-SVM, (a) 3×3 pixel majority filter. (b) 5×5 pixel majority filter. (c) 7×7 pixel majority filter, (d) 9×9 pixel majority filter.

Table 5. Overall classification accuracy calculated for various combinations of data sets using SVMs and penalty parameter, gamma for each data set.

Data Set Name for SVM classifier	No. of Data Layers	Penalty parameter (c)	Gamma (g)	Overall Accuracy	Kappa Coefficient
ASTER (VNIR+SWIR)	9	100	100	93.1%	0.93
ASTER (VNIR+SWIR)+DEM (slope and curvature)	11	100	100	94%	0.93
ASTERDEM (slope and curvature) -PCA	11	100	1000	94.1%	0.93
ASTER DEM (slope and curvature)-ICA-	11	10	100	94.8%	0.94
ASTER DEM (slope and curvature)-PCA-ICA-	20	100	100	95%	0.95

The use of combined datasets produced better results than the use of a single dataset and the highest classification accuracy for the independent validation dataset is achieved when using the combination of ASTER DEM+PCA+ICA SVM. The integrative SVMs (ASTER+DEM SVM and PCA, ICA+DEM) perform slightly better than all other SVMs for all classes. The SVM proposed method

produce a good distribution of the lithological units for the widely exposed rock units and the resulted classified image (Fig. 4) successfully emphasizes the spatial distribution of the lithological units around Wadi Attalah area. The produced image show differences in the distribution of some rock units and contacts between them compared to the published geological map of EGSMA (1992) (Fig. 1b) which has been verified by field work. The Image processing of the remote sensing data as proposed in the present study proves its high capability in detecting and emphasizing the spatial distribution of the lithological units in other geologically similar arid regions in the Arabian-Nubian Shield (ANS) and worldwide.

5.2. Post-processing procedures

Some processing techniques have been used for post classification as (siev, clumb and Majority Analysis) to improve overall accuracy and remove noisy pixels as well as to enhance the visual interpretation of the classified image. Sieving classes removes isolated classified pixels using blob grouping. Clump classes to clumps adjacent similar classified areas together using morphological operators. Use of majority/minority analysis applies to a classification image. Use majority analysis to change spurious pixels within a large single class to that class. The kernel size is specified and the center pixel in the kernel will be replaced with the class value that represent majority of the pixels in the kernel.

It has been found that the best window size is 9 by 9 pixels that filter out spurious pixels. From the post-classified image, we can see that the visual appearance is improved and the distribution of the lithological units is enhanced compared with the published geological map.

6. CONCLUSIONS

In this study, we used high accurate remote sensing data (ASTER spectral bands, ASTER DEM) for Wadi Atalla area to emphasize and identify the lithological distribution detected of this part of the Arabian-Nubian Shield.

The SVM algorithm was applied to automated lithological classification of Wadi Attalah area in order to emphasize and discriminate the widely exposed rock units (i.e. ophiolitic group, Meatiq group, intrusive rocks and Hammamat molasse sediments) in Eastern Desert of Egypt for the first time using ASTER multispectral data. Several digital image processing techniques were used to produce ASTER derivative data sets (i.e. DEM, PCA and ICA) that contained enhanced information related to lithological discrimination.

Series of SVMs were tested using various combinations of input datasets to select the optimal inputs that provide the highest classification accuracy. We conclude that combination of 9 ASTER VNIR and SWIR as well as ASTER derived PC+IC+DEM provide the highest overall classification accuracy of 95% for the widely exposed rock units (Table 5). The average accuracy calculated for each lithological units of the study area using 9 ASTER bands+PC+IC+ DEM datasets provide a perfect accuracy for seven lithological units (e.g. Serpentinite talc-carbonate rocks, acidic metavolcanic tuffs, ophiolitic metagabbro, metapasmmite) (Table 4).

The Image processing of the remote sensing data as proposed in the present study can therefore be used to produce more accurate distribution for the exposed lithological unit than the previous published geological map. These results have been verified by field work. The classification accuracy for each lithological unit demonstrates that the technique can be used by exploration companies in such areas for a lithological mapping that can be further distinguished using field investigation.

The newly developed methods are used to distinguish between ophiolitic rocks which include (serpentinite and metagabbro), granitic rocks monzogranite/alkali feldspar granite and monzodiorite) of the Fawakhir area and Meatiq group (amphibolite schist, sericite muscovite garnet schist and metapasmmite) as well as Hammamat sediments (Greywackes and conglomerate). It is recommended using ASTER-PC+IC+DEM SVM for mapping serpentinites and to discriminate between the different types of granitic rocks, ophiolitic rocks, Meatiq group, Attalah felsite and Hammamat molasse sediments. Comparison between the results derived from the proposed new methods and field work demonstrate that the new methods were successful in emphasizing the lithological information at Wadi Attalah area. Therefore, we suggest that these techniques may be used as time - and cost-effective approach for lithological feature identification in other geologically similar arid regions in the Arabian-Nubian Shield (ANS) and worldwide.

REFERENCES

Abd El-Rahman, Y., Polata, A., Dilekb, Y., Fryera, B. J., El-Sharkawy M., & Sakran, S., 2009. *Geochemistry and tectonic evolution of the Neoproterozoic incipient arc–fore arc crust in the Fawakhir area, Central Eastern Desert of Egypt*". Precambrian Research, 175, 116-134.

Abrams , M. J. & Hook , S. J., 1995. *Simulated ASTER data for geologic studies*. IEEE Transactions on Geoscience and Remote Sensing, 33, 692-699.

Abrams, M., Brown, D., Lepley & Sadowski, L., R., 1983. *Remote sensing of porphyry copper deposits in Southern Arizona*. Economic Geology, 78, 591-604.

Barnett, W. S., 2004. *Does Head Start have lasting cognitive effects?: The myth of fade-out. In E. Zigler&S.Styfco (Eds.*). The Head Start debates Baltimore: Paul H. Brookes Publishing Co., 221-249.

Boser, E.B., Guyon, M.I & Vapnik, N.V., 1992. *A Training Algorithm for Optimal Margin Classifiers*. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, 144-152.

Burges, C. J., 1998. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery. 2, 121-167.

Cries, j. Marturia, A. De Paz, J. Casnovas & Lieopart, A., 1995**.** *Digital Elevation Model, A useful tool for Geological Mapping. Some Examples from Catalonia*. 3-9.

El-Gaby, S., El-Nady, O. & Khudeir, A., 1984. *Tectonic evolution of the basement complex in the Central Eastern Desert of Egypt*. GeologischeRundschau, 73, 1019-1036.

El-Sayed, M.M., Furnes, H. & Mohamed, F.H., 1999. *Geochemical constraints on the tectono magmatic evolution of the late Precambrian Fawakhir ophiolite, Central Eastern Desert, Egypt.*. Journal of African Earth Science, 29, 515-533.

Fowler, T.J. & Osman, A.F., 2001. *Gneiss-cored interference dome associated with two phases of late Pan-African thrusting in the Central Eastern Desert, Egypt.* Precambrian Research, 108, 17-43.

Getman, D. J., Harbor, J. M., Johannsen, C. J., Engel, B. A. & Shao, G. (n.d.)., 2003. *Improving the Accuracy of Historic Satellite Image Classification by Combining Low-Resolution Multispectral Data with High-Resolution Panchromatic Data*, 1, 70-87.

Gómez, A. J., Montero-Pau, D. H., Lunt, M. Serra, & Campillo, S., 2007**.** *Persistent genetic signatures of colonization in Brachionusmanjavacas rotifers in the Iberian Peninsula*, Mol. Ecol. 16, 3228-3240.

Habib, M. E., Ahmed, A. A. & El Nady, O. M. 1985. Two orogenies in the Meatiq Area of the Central Eastern Desert, Egypt. Precambrian Research 30, 83-111.

Hassan, M. A. & Hashad, A. H., 1990. *The anorogenic alkalic rocks, South Eastern Desert, Egypt*". Annal Geological Survey of Egypt, 9, 81-101.

Hassanen, M. A, 1985. *Petrology and geochemistry of ultramafic rocks in the Eastern Desert, Egypt, with special reference to Fawakhirarea*. Ph.D.dissertation, Alexandria University, Egypt, 348p.

Harraz, H.Z. & Ashmawy, M.H., 1994. *Structural and lithogeochemica constraints on the localization of gold deposits at the El Sid-Fawakhir Gold Mine area, Eastern Desert, Egyp*t. Egyptian Journal Geology 38, 629-648.

James, A., Shine & Daniel, B. Carr, 2002. *A Comparison of Classification Methods for Large Imagery Data Sets*, JSM Statistics in an ERA of Technological Change-Statistical computing section, 3205-3207.

Li, N., Frei, M. & Altermann, W., 2011. *Textural and knowledge-based lithological classification of remote sensing data in Southwestern Prieska sub-basin*, Transvaal Supergroup, South Africa. Journal of African Earth Sciences, 60(4), *237–246.*

Lu, D. & Weng, Q., 2007**.** *A survey of image classification methodsand techniques for improving classification performance*, International Journal of Remote Sensing, 28 (5), 823-870.

Mercier, G. & Lennon, M., 2003**.** *Support vector machines for hyperspectral image classification with spectral-based kernels*, Proceedings of Geoscience and Remote Sensing Symposium, 1, 288-290.

Mondal, A., Kindu, S., Chandniha, K.S., Shukla, R. & Mishra P.K, 2012.*Comparison of Support Vector Machine an d Maximum Likelihood Classification Technique using Satellite Imagery*, International Journal of Remote Sensing, 1 (2), 116-123.

Nasseef, M. O., Bakor, A. R. & Hashad, A. M., 1980. *Petrography of possible ophiolitic rocks along the Qift-Quseir road Eastern Desert Egypt, evolution and mineralization of the Arabian Shield*. Inst. Appl. Geol.Jeddah Bull, 3, 157-168.

Ninomiya, Y., & Fu, B., 2010. *Regional Scale Lithologic Mapping In Western Tibet Using Aster Thermal Infrared Multispectral Data*, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, 8, 454–458.

Neumayr, P., Hoinkes, G. & Puhl, J., 1995. *Constraints on the P–T–t evolution of a poly metamorphic Panafrican basement dome in the Central Eastern Desert (Egypt),* Terra Abstracts 7, 316.

Perumal, K. & Bhaskaran, R., 2011. *Supervised Classification Performance Of Multispectral Images*, journal of computing, 2, 2151-9617.

Ding, S., 2011. *Spectral and Wavelet-based Feature Selection with Particle Swarm Optimization for Hyperspectral Classification*, Journal of Software, 6(7), 1248-1256.

Linden, S. Van Der, Rabe, A., Wirth, F., Suess, S. & Hostert, P., 2010**.** *Image SVM Classification Manual for Application*, 1-26.

Ries, A.C., Shackleton, R.M., Graham, R.H. & Fitches, W.R., 1983. *Pan-African structures, ophiolites and mélanges in the Eastern Desert of Egypt*, a traverse at 26 N. Journal of Geological Society, 140, 75–95.

Roberts, A., 2001. *Curvature attributes and their application to 3D interpreted horizons*, First Break, 19, 2, p. 85-100.

Salati, S., Van Ruitenbeek, F. J. a., Van der Meer, F. D., Tangestani, M. H., & Van der Werff, H, 2011. *Lithological mapping and fuzzy set theory:"Automated extraction of lithological boundary from ASTER imagery by template matching and spatial accuracy assessment*. International Journal of Applied Earth Observation and Geoinformation, 13(5), 753-765.

Yamaguchi, T., Takamura, Matoba H., Terao T. & HPLC, J., 1998. *method for evaluation of the free radical-scavenging activity of foods by using 1, 1-diphenyl-2-picrylhydrazyl*, Biosci. Biotechnol. Biochem, 62, 1201-1204.

Yu, L., Porwal, A., Holden, E.-J., & Dentith, M. C., 2012**.** *Towards automatic lithological classification from remote sensing data using support vector machines*|, Computers and Geosciences, 45 , 229-239.

Received at: 29. 04. 2014
Revised at: 21. 10. 2014
Accepted for publication at: 27. 10. 2014
Published online at: 13. 11. 2014