

## **COMPARATIVE STUDY BETWEEN SUPPORT VECTOR MACHINES AND NEURAL NETWORKS FOR LITHOLOGICAL DISCRIMINATION USING HYPERSPECTRAL DATA**

**A. M. Naguib<sup>1</sup>; M. A. Farag<sup>2</sup>; M. A. Yahia<sup>2</sup>; H.H. Ramadan<sup>3</sup> and M. S. Abd Elwahab<sup>1</sup>**

1. Scientific computing department, Faculty of computer and information sciences, Ain Shams University
2. Geology Department , Faculty of Science, Ain Shams University
3. Basic Sciences department, Faculty of computer and information sciences, Ain Shams University

### **ABSTRACT**

Remote sensing hyperspectral data has many applications especially in the field of earth science. Utilization of this technology has shown a rapid increase in many areas of economic and scientific significance. Hyperspectral sensors capture the detailed spectral signatures that uniquely characterize a great number of diverse surface materials. Classification, clustering, and visualization of these very high-dimensional signatures need untraditional methods. Different approaches for spectral image interpretation have been studied using Artificial Neural Networks (ANNs) and Support Vector Machines (SVM) to meet the challenge of high-dimensionality.

The study used SVMs for geological mapping of hyperspectral imagery at Abu Zenima area, western Sinai, Egypt, the hyperspectral data has been captured in 2003 by Hyperion instrument on the United States Geological survey (USGS) Earth Observing1(EO-1) satellite. Precisely the study compares between the use of SVMs and a neural network built on the concept of SVMs, this network uses the Kernel-Adatron algorithm with the Gaussian kernel for the process of training. The SVMs also uses the Gaussian kernel with different bandwidths to enhance the performance of the interpretation process; the results are compared in details.

The Neural Network was trained with four data sets , the first consists of 11310 samples, gives recognition rate of 84%, the second has 22620 samples, recognition rate was 91.5%; the third has 33930 samples, recognition rate was 94.6%; finally the fourth has 45240 samples, recognition rate of 99.2%. The previous results fall in comparison with the results of SVMs which use two algorithms for training the first is the one against one algorithm which gave a recognition rate of 84% for the first data set, a recognition rate of 76.9% for the second data set, a recognition rate of 95.2% for the third one and 98.5% for the fourth one. and the other is one against many algorithms which gave a recognition rate of 84% for the first data set, a recognition rate of 72.3% for the first data set, recognition rate of 94.6% for the second one and 98.5% for the third one.

**Keywords:** hyperspectral data; artificial neural networks (ANNs); supervised classification; support vector machines (SVMs), kernels, lithological units, Abu Zenima-Egypt.

## **INTRODUCTION**

Hyperspectral systems are known for having dozens to hundreds of narrow contiguous bands. Most are able to collect images starting at about 400 nm which is the edge of the blue visible part of the spectrum (Lillisand, 1999). Typically these systems can measure energy to 1100 or even 2500 nm. Various statistical and artificial intelligence methods have been used for analysis of remote sensing data in the field of geology. Among these methods, artificial neural networks (ANNs) (Merényi, 2002) Self-Organizing Neural Net (SOM), (Rossi, 2005, Karimi, 2005) Support Vector Machine (SVM).

Hyperion data is used in this study, which provides continuous spectral coverage over 220 bands with a 10-nm (average) sampling interval over the contiguous reflected spectrum from 356 to 2577 nm, with a ground sample distance (GSD) of 30 meters for all bands. The Hyperion scene is collected as a narrow strip, covering a ground area approximately 7.7 km in the across-track direction, and 42 km or 185 km in the along-track direction. The image file is provided in scaled at-sensor radiance values, with the data stored as 16-bit signed integers. The radiance units are in  $W / (m^2 * Sr * \mu m)$ .

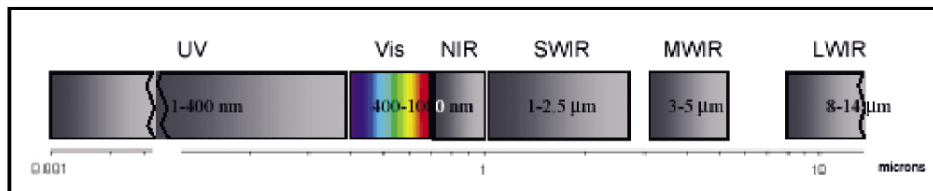
The support vector machine (SVM) algorithm, a supervised machine learning method based on the statistical learning theory (Vapnik, 1995), is a new method in artificial intelligence. SVM is basically a binary classifier, which finds the maximal margin (hyperplane) between two classes. SVM can classify non-linearly separable data sets by plotting the data into a high-dimensional feature space using kernels. In comparison to other data mining techniques such as ANNs, it is easier to use, and only a few parameters need to be adjusted by the users. SVM has been widely applied to classification problems such as weed and nitrogen stress detection in corn (Karimi, 2005) and functional data classification (Rossi, 2005).

The overall objective of this study was to examine the capability of the SVM method in analyzing hyperspectral data for the first time in Egypt, the study area is Abu Zenima area, western Sinai, Egypt. The analysis includes classification and recognition of the different lithological units which are found in the imagery and were denoted by (Moustafa, 2004). The classification results were compared with the results obtained using an ANN method, used by (Merényi, 2002) to classify geological units.

## THE IMAGING SPECTROMETER

Hyperspectral images are produced by instruments called imaging spectrometers. The development of these complex sensors has involved the convergence of two related but distinct technologies: spectroscopy and the remote imaging of earth and planetary surfaces. Spectroscopy is the study of light that is emitted by or reflected from materials and its variation in energy with wavelength (Lillisand, 1999).

As applied to the field of optical remote sensing, spectroscopy deals with the spectrum of sunlight that is diffusely reflected (scattered) by materials at the earth's surface. Instruments called spectrometers (or spectroradiometers) are used to make ground-based or laboratory measurements of the light reflected from a test material. An optical dispersing element such as a grating or prism in the spectrometer splits this light into many narrow, adjacent wavelength bands and the energy in each band is measured by a separate detector (Lillisand, 1999).

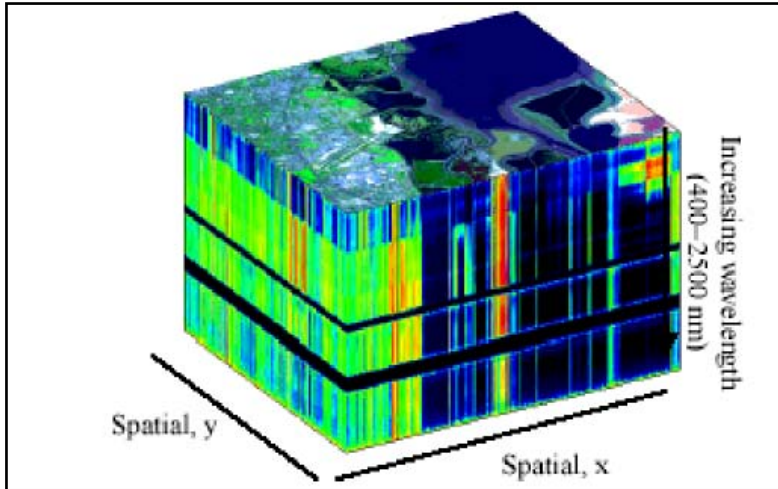


**Figure 1:** Electromagnetic Spectrum.

By using hundreds or even thousands of detectors, spectrometers can make spectral measurements of bands as narrow as 0.01 micrometers over a wide wavelength range Figure (1), typically at least 0.4 to 2.4 micrometers (visible through middle infrared wavelength ranges) Table (1). Remote imagers are designed to focus and measure the light reflected from many adjacent areas on the earth surface. In many digital imagers, sequential measurements of small areas are made in a consistent geometric pattern as the sensor platform moves and subsequent processing is required to assemble them into an image.

### Electromagnetic Spectral Resolution

Spectral resolution is important for detecting fine spectral features that can identify specific materials. Hyperspectral data is difficult to visualize all at once. Because each ground scene can be made up of hundreds of images (bands), one way of understanding the patterns in the data is to create an image cube. The x and y axes are the spatial dimensions showing the ground surface. The z axis is made up of all the other bands as if they were stacked like a ream of paper and placed on its side Figure (2).



**Figure 2:** Hyperspectral data cube.

**Table 1:** The divisions of the electromagnetic spectrum.

Wavelength ( $\mu\text{m}$ )	Frequency Name
0.4-0.68	Visible (Blue, Green, and Red)
0.68-1	NIR (Near Infrared)
1-2.5	SWIR(Short Wavelength Infrared)
3-5	MWIR(Medium Wavelength Infrared )
8-14	LWIR(Long Wavelength Infrared)

Table (2) shows several hyperspectral instruments that are commercially available. All are flown on aircraft. In December of 2000 the first commercial hyperspectral imager (Hyperion) was placed on a satellite that successfully made it into orbit. These systems have varying spectral ranges, bands widths, and spatial (pixel) resolutions.

**Table 2:** Some hyperspectral systems.

Sensor	Wavelength range ( $\mu\text{m}$ )	Band width ( $\mu\text{m}$ )	Number of bands
HYPERION	0.4-2.5	0.01	224
AVIRIS	0.4-2.5	0.01	224
TRWIS III	0.367-2.328	0.059	335
HYDICE	0.4-2.4	0.01	210
CASI	0.4-0.9	0.018	288
OKSI AVS	0.4-1	0.01	61

## ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial neural networks (ANN) have been used in data reduction and data compression for data coding. The usefulness of data compression arises in storage and transmission of data. One can use neural networks to minimize the memory

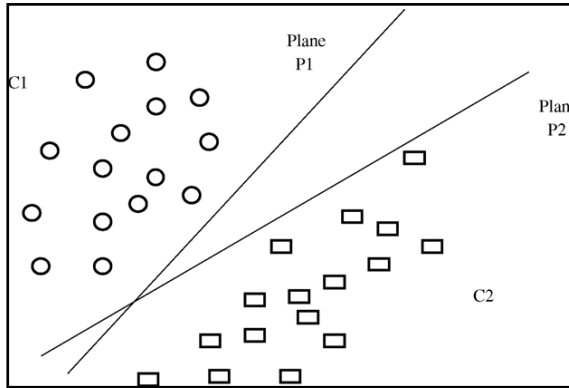
storage of images to recognize the images. Presenting the original (no reduced) images to a neural network is not recommended; this implies long training times. With data reduction, good compact representations of the images can be obtained.

Artificial Neural Networks (ANNs) models have shown great capability in providing redundancy reduction and data compression of the sensory data. Neural networks have an ability to preprocess input patterns to produce simpler patterns with fewer components [8, 5].

### **Support Vector Machines**

Support Vector Machines (SVMs) are pioneered by Vapnik [11]. They were introduced for classification tasks and subsequently extended to regression problems. Their unique features make them particularly suitable for machine discovery and data mining tasks. In particular, the way they present the hypothesis by the most informative patterns (the support vectors) makes them suitable for tasks such as data cleaning, training set reduction, detection of outliers or selective sampling.

Basically, the SVMs are linear machines with some very good properties. To explain how it works, it is perhaps easiest to start with the case of separable patterns that could arise in the context of pattern classification. In this context, the main idea of SVMs is to construct a hyper-plane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized, Figure (3). The machine achieves this desirable property by following a principled approach rooted in statistical learning theory.



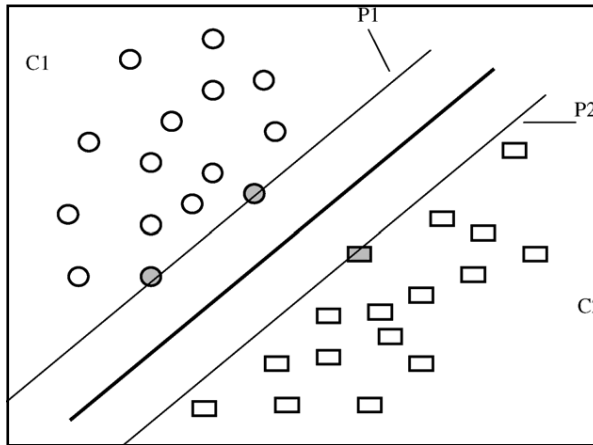
**Figure 3:** Linear separating planes between two classes.

#### ***1. Linearly separable binary classification***

Let us consider a simple binary classification problem with  $M$  training samples that can be represented by a set of pairs  $\{(x_i, y_i), i=1, 2, \dots, M\}$ , where the  $y_i$  are class labels ( $y_i = +1$  (or Class C1) or  $-1$  (or Class C2)) and  $x_i$  are the input vectors. The classification function (i.e., the separating hyperplane in this case, a line) can be written as follows:

$$y_i = \mathbf{w} \cdot \mathbf{x}_i - \mathbf{b} \quad (1)$$

Where  $w$  is a normal vector to the hyperplane,  $|b|/w$  is the perpendicular distance from the hyperplane to the origin, and  $\|w\|$  is the Euclidean norm of  $w$ [1]. It is possible to have an infinite number of planes (or hyperplanes) that can separate the two classes of data; two such planes are shown in Figure (3). Intuitively speaking, one would prefer to select the plane P1 as the classification function (over P2), owing to the fact that a minor change in the data items (of C2) will not introduce any errors in the classification. This is not the case for plane P2. In Figure (4), the supporting planes (marked as P1 and P2) for the two classes and the optimal classifying plane (in bold) are shown. It is clear from the figure that only certain data points (The ones shown in bold called support vectors in SVM) actually influence the equation of the optimal separating plane. Remaining data points are redundant, as far as the formulation/construction of separating plane is concerned. From a geometric perspective, we would select an optimal plane P that is farthest from both the classes. The plane can be determined by computing two parallel supporting planes, one for each of the two classes, and maximizing the distance between them. A supporting plane of a class is defined as any plane such that all points of the class are on one and only one side of that plane. Ideally, for a supporting plane, we would like that  $(w \cdot x_i - b \geq 1)$  for one class ( $y = +1$ ) and  $(w \cdot x_i - b \leq -1)$  for the other ( $y = -1$ ). The points that lie on the hyperplane  $w \cdot x_i - b = 1$  will have a perpendicular distance from the origin of  $|1 - b| / \|w\|$  and those on hyperplane  $w \cdot x_i - b = -1$  would have a corresponding distance of  $|-1 - b| / \|w\|$ . Hence, the distance between the two planes ( $2/\|w\|$ ) can be obtained by subtracting  $|-1 - b| / \|w\|$  from  $|1 - b| / \|w\|$ . For optimal classification, we would like to maximize this distance which is equivalent to following minimization problem[1]:



**Figure 4:** Optimal separating plane between the two classes of data.

$$\begin{aligned}
 &\text{Minimize} && \frac{1}{2} \|w\|^2 \\
 &\text{Subject to:} && w \cdot x_i \geq b + 1 \quad \text{for } x_i \in C1, \text{ and} \\
 &&& w \cdot x_i \leq b - 1 \quad \text{for } x_i \in C2
 \end{aligned} \tag{2}$$

This is a well-known convex quadratic optimization problem and the explicit solution of this equation is rather difficult. Using Lagrangian multipliers, the

solution to this optimization problem is [Burges, C., 1998]:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{W}(\boldsymbol{\alpha}) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{Subject to:} \quad & \sum_{i=1}^M \alpha_i y_i = 0 \quad \mathbf{0} \leq \alpha_i \end{aligned} \quad (3)$$

Where  $\alpha_i$  are the positive Lagrange multipliers. Solution of Eq. (3) provides the equation of the optimal separating plane.

Now, consider a more realistic case when some of the data points belonging to classes C1 and C2 overlap each other. In order to find a robust classification function in such cases, we can relax certain constraints. Ideally, we would have preferred to have not even a single point misclassified. In the presence of outliers, this is not possible; instead, we would prefer to have the “majority” of points correctly classified. In other words, there will be some points in both classes, which will fall on the “wrong” side of the supporting plane. Such points will be treated as outliers or errors.

The optimization problem is then modeled not only to maximize the margin between the supporting planes, but also to minimize the errors of classification. To this end, non-negative “slack variables” (denoted by  $z$ ) are introduced into the optimization problem, which transform the optimization Eq. (2) as follows:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M z_i \\ \text{Subject to:} \quad & \mathbf{W} \cdot \mathbf{x}_i \geq \mathbf{b} + 1 + z_i \quad \text{for } \mathbf{x}_i \in \text{C1, and} \\ & \mathbf{W} \cdot \mathbf{x}_i \leq \mathbf{b} - 1 + z_i \quad \text{for } \mathbf{x}_i \in \text{C2} \end{aligned} \quad (4)$$

Where  $C$  is the tradeoff between the following two factors: the margin between the two supporting planes and the classification error. The equivalent dual problem corresponding to Eq. (3) becomes [Burges, C., 1998]:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{W}(\boldsymbol{\alpha}) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{Subject to:} \quad & \sum_{i=1}^M \alpha_i y_i = 0 \quad \mathbf{0} \leq \alpha_i \leq C \end{aligned} \quad (5)$$

## 2. Non-linearly separable binary classification

For training samples that are not linearly separable, the data need to be transferred onto a space of higher dimensionality (called the “feature space”) so that a reliable linear separation can be computed. Let us denote this mapping onto the feature-space by a function,  $\phi$ , such that  $z = \phi(x)$  is the feature-point corresponding to a data item  $x$ . The equation of the optimal separating hyperplane (OSH) in this feature-space can then be written as [Burges, C., 1998]:

$$\mathbf{F}(\mathbf{x}) = \sum_{i=1}^M \alpha_i y_i \mathbf{x}_i \cdot \boldsymbol{\phi}(\mathbf{x}) - \mathbf{b} \quad (6)$$

Eq. (6) is comparable to Eq. (1) where  $w$  and  $x$  are replaced by  $\sum_{i=1}^M \alpha_i y_i \mathbf{x}_i$  and  $\boldsymbol{\phi}(\mathbf{x})$ , respectively.

Following the procedure explained for the linearly separable case (Eq. (5)), it can be shown that the classification can be performed by solving the following optimization equation [Burgess, C., 1998]:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{W}(\boldsymbol{\alpha}) = \sum_{i=1}^M \alpha_i - 1/2 \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j \boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j) \\ \text{Subject to:} \quad & \sum_{i=1}^M \alpha_i y_i = 0 \quad \quad \quad 0 \leq \alpha_i \leq C \end{aligned} \quad (7)$$

However, direct solution of the above minimization problem is often not feasible. This difficulty is circumvented by employing a “kernel function”,  $K(x, y)$ , which is the dot product of  $\boldsymbol{\phi}(\mathbf{x})$  and  $\boldsymbol{\phi}(\mathbf{y})$ . Replacing the dot product in the above equation by the kernel function, the equation for the OSH becomes:

$$\mathbf{F}(\mathbf{x}) = \sum_{i=1}^M \alpha_i y_i \mathbf{x}_i \cdot \mathbf{k}(\mathbf{x}_i, \mathbf{x}) - \mathbf{b} \quad (8)$$

The commonly used kernels are the radial basis function (RBF) kernels, the sigmoid kernels, and the polynomial kernels. The RBF kernel, most commonly used in SVM classification problems, is given as follows:

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x} - \mathbf{y}\|^2} \quad (9)$$

Where  $\gamma$  is a kernel parameter. The relevant optimization function can then be written as:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^M \alpha_i - 1/2 \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) \\ \text{Subject to:} \quad & \sum_{i=1}^M \alpha_i y_i = 0 \quad \quad \quad 0 \leq \alpha_i \leq C \end{aligned} \quad (10)$$

Furthermore, one also has to choose the parameter  $\gamma$  for the RBF kernel and a parameter  $C$  that determines how severely classification errors are to be penalized. The quality of classification is greatly affected by the parameter  $C$ ; a very large  $C$  value can lead to over fitting of the training data [Cao, L., 2003].

### 3. Data transformation using kernels

The simplest way to divide two groups is with a straight line, flat plane or an  $N$ -dimensional hyperplane. But what if the points are separated by a nonlinear region such as shown below? In this case we need a nonlinear dividing line.

Rather than fitting nonlinear curves to the data, SVM handles this by using a kernel function to map the data into a different space where a hyperplane can be

used to do the separation. The kernel function may transform the data into a higher dimensional space to make it possible to perform the separation.

The concept of a kernel mapping function is very powerful. It allows SVM models to perform separations even with very complex boundaries such as shown below. Many kernel mapping functions can be used – probably an infinite number. But a few kernel functions have been found to work well in for a wide variety of applications. The default and recommended kernel function is the Radial Basis Function (RBF). Some famous kernels such as Linear kernel:  $u \cdot v$  , Polynomial:  $(\gamma u \cdot v + c)^{\text{degree}}$  ,Radial basis function:  $\exp(-\gamma |u-v|^2)$  and Sigmoid (feed-forward neural network):  $\tanh(\gamma u \cdot v + c)$ .

#### **4. Classification with more than two categories**

The idea of using a hyperplane to separate the feature vectors into two groups works well when there are only two target categories, but how does SVM handle the case where the target variable has more than two categories? Several approaches have been suggested, but two are the most popular: (1) “one against many” where each category is split out and all of the other categories are merged; and, (2) “one against one” where  $k(k-1)/2$  models are constructed where  $k$  is the number of categories. The more accurate (but more computationally expensive) technique is “one against one”.

## **THE HYPERION DATA USED FOR TRAINING AND TESTING**

The Hyperion data are provided in standard HDF Version 4.1 (release 5), written as band-interleaved-by-line (BIL) files stored in 16-bit signed integer radiance values.

This study aims to map the different lithological units that are found in the study area Figure (5), those units consist of different sequences of geological units such as sand , shale, limestone, clayely limestone ... (etc); as denoted by (Moustafa, A.R., 2004) .

### **Spatial Characteristics**

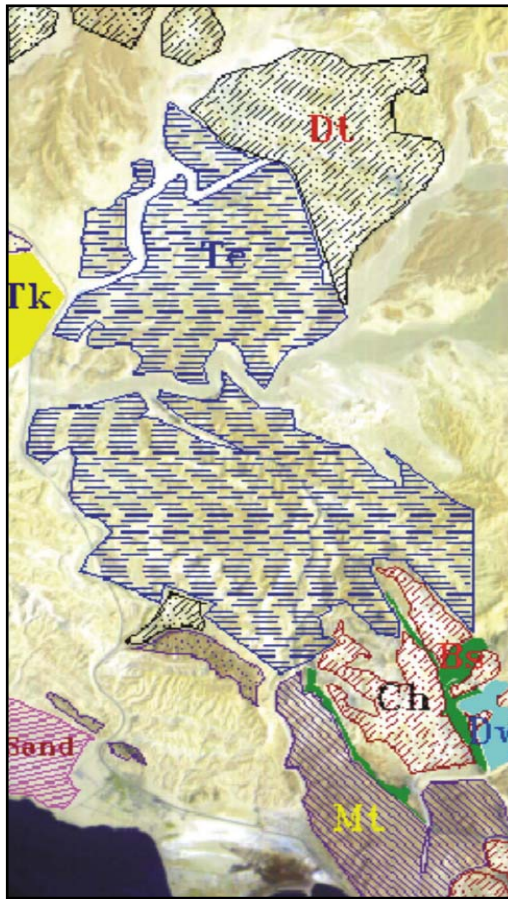
The Hyperion sensor provides continuous spectral coverage over 220 bands, with a ground sample distance (GSD) of 30 meters for all bands. Each Hyperion scene is collected as a narrow strip, covering a ground area approximately 7.7 km in the across-track direction, and 42 km or 185 km in the along-track direction.

### **Spectral Characteristics**

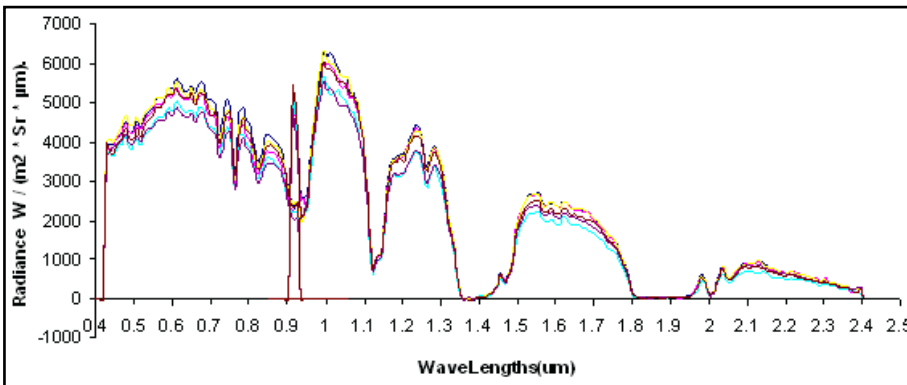
The Hyperion instrument collects a  $0.01 \mu\text{m}$  (average) sampling interval over the contiguous reflected spectrum from 356 to  $2.577 \mu\text{m}$ . The data is acquired from two separate push broom imaging spectrometers: one visible (VNIR) spectrometer and one short-wave infrared (SWIR) spectrometer (sensor description) Figure (6). There is an area of spectral overlap between the two spectrometers from  $0.852$  to  $1.058 \mu\text{m}$ . The Hyperion instrument collects a total of 242 channels. Not all of

the acquired bands are calibrated. This is primarily due to Diminishing detector response at less optimum wavelengths, or an area of overlap between the two spectrometers. Out of the 242 collected channels 1) bands 1-7 (0.356 to 0.417  $\mu\text{m}$ ) and bands 225-242 (2.406 to 2.578  $\mu\text{m}$ ) are not calibrated, Bands 58-70 (collected by the VNIR instrument) and bands 71-76 (collected by the SWIR instrument) are also not calibrated.

The sensor data will provide a total of 198 bands representing continuous spectra from 0.427 to 2.395  $\mu\text{m}$ . The collected and calibrated channels include a small area of spectral overlap between the VNIR and SWIR spectrometers (bands 56-57 and 77-78; from  $\sim 0.912$  to  $\sim 0.923$   $\mu\text{m}$ ). Out of the 198 calibrated bands, there will be 196 unique spectral channels provided. While the average bandwidth for Hyperion is 0.01  $\mu\text{m}$ , the actual bandwidth will vary by band. Hyperion bandwidths are based on the full-width half-max (FWHM) values derived from the instrument spectral response curves.



**Figure 5:** color composite subset image for the used hyperspectral image of Abu Zenima area of study with class labels.



**Figure 6:** Spectral curves extracted from the Hyperspectral image of Abu Zenima area.

## THE PROPOSED METHODS AND THE RESULTS

Learning is performed using three ways (1) the SVM neural network based on kernel Adatron algorithm; (2) one against one algorithm; and (3) one against all algorithm. The neural network is trained using the kernel-Adatron algorithm with the Gaussian kernel  $K(\mathbf{x}_1; \mathbf{x}_2) = \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\|^2 / 2\delta^2)$ . The learning is performed using different number of objects and the testing as well. The network should differentiate between 13 geological units. Four sets of training data have been used, the first consist of 5 examples per each unit, the second has 10 examples, the third has 15 examples, and the fourth has 20 examples. Each example is a sample data vector of 174 elements representing the spectral signature of its own geological unit.

One against one algorithm has been trained using two different data sets the first consist of 15 examples per each class, and the second has 20 examples. The kernel used to transform the data to a separable space was the Gaussian kernel with different values for the Gaussian standard deviation. The one against all algorithm was trained using the same two data sets with the same kernel.

### The Kernel-Adatron (KA) Algorithm

The KA is an efficient algorithm which is forward to implement and train SVM rapidly. The KA is based on the Adatron algorithm [Friess, T.,1998] and adapted by the introduction of kernels to find nonlinear decision boundaries using high-dimensional feature space of SVMs. The original Adatron algorithm comes with theoretical guarantees of convergence to the optimal solution (after a finite number of steps). The rate of convergence is exponentially increases by advancing the number of iterations, provided that a solution exists [Hegazy, D., 2004]. By adapting the Adatron algorithm to include kernels, the algorithm has an attractive fixed point corresponding to the maximal margin hyperplane. For the classification tasks with target label  $y_i = \pm 1$  the algorithm is as follows:

#### The Kernel-Adatron (KA) algorithm

1. Initialize  $\alpha_i = 1$  and  $\theta = 0$ .
2. Starting from pattern  $i = 1$ , for labeled points  $(x_i, y_i)$  calculate:

$$Z_i = \sum_{j=1}^p \alpha_j y_j K(x_i, x_j) + \theta$$

3. For all patterns  $i$ , calculate  $\gamma_i = y_i z_i$  and execute steps 4 to 5 below.
4. Let  $\delta \alpha_i = \eta (1 - \gamma_i)$  be the proposed change to the multiplier  $\alpha_i$ .
- 5.1 .If  $(\alpha_i + \delta \alpha_i) \leq 0$  then  $\alpha_i = 0$ .
- 5.2. If  $(\alpha_i + \delta \alpha_i) > 0$  then  $\alpha_i = \alpha_i + \delta \alpha_i$ .
6. Calculate:

$$\theta = 1/2 (\min (z_i^+) + (\max (z_i^-)))$$

Where  $z_i^+$  are those patterns  $I$  with label  $+1$  and  $z_i^-$  are those patterns  $I$  with label  $-1$ .

7. If a maximum number of presentations of the pattern set has been exceeded or the margin given as

$m = 1/2 (\min (z_i^+) + (\max (z_i^-)))$  has approached 1 then stop, then other case return to step 2.

## The Experiments

We used SVM neural network with the kernel Adatron algorithm with standard deviation for the Gaussian kernel is set to 0.707 and  $p=3$  ( $p$  represents the number of the nearest neighbors processing elements (PEs) in the network that averaged together to compute that variance of the Gaussian transfer function ).

After cross track illumination correction there are 174 representative bands remaining from 224 total bands, excluding the overlapped and not calibrated bands, so the whole remaining bands are used to classify 13 different geological units. Those units are characterized by experts in the study area.

With the four data sets for training the network well trained and tested using 130 different samples, the results are shown in table (4) ;Comparing these results with [Merényi, E., 2002] who used 194 bands AVIRIS data to classify 23 classes with Recognition Rate of 89.8 %, our proposed system is promising if we increase the number of training samples per each class.

**Table 4:** Summary of the proposed methods performance.

Samples per class	Recognition Rate
5	84%
10	91.5%
15	94.6%
20	99.2%

The one against one algorithm was trained using two data sets with the same Gaussian kernel .but to choose the suitable value for the Gaussian standard deviation, a data set (consists of 10 samples per class for training and 5 sample per class for testing) was trained using different values of the standard deviation, the results shows that 2.5 has given the highest recognition rate Table (5). For the proposed system the first data set consists of 15 samples per class for training and 5 samples per class for testing and the second has 20 samples per class for training

and 10 samples per class for testing. For the first the recognition rate was 95.2% and for the second data set the recognition rate was 98.5 %.

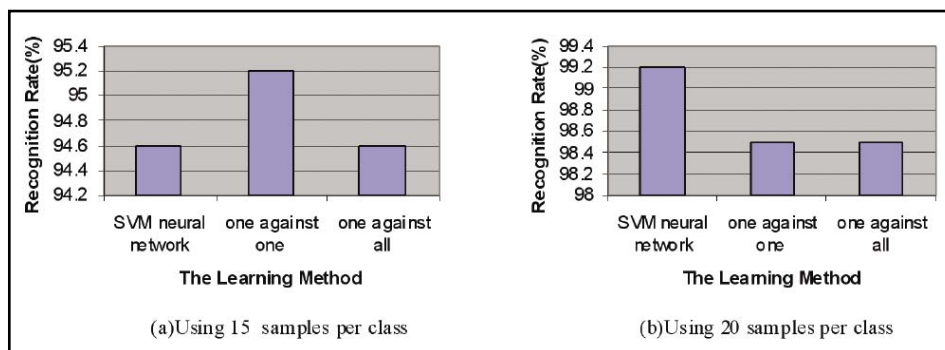
**Table 5:** Choosing the best standard deviation value with the highest recognition rate for one against one algorithm.

Standard deviation	Recognition rate (%)
0.5	46.1
1	66.1
1.5	70.7
2	75.4
2.5	76.9
3	75.4
4	72.3
5	70.7

The same procedure was followed for the one against all algorithms, the best value for the standard deviation was 2 see Table (6). For the proposed system the first data set consists of 15 samples per class for training and 5 samples per class for testing and the second has 20 samples per class for training and 10 samples per class for testing. For the first the recognition rate was 94.6% and for the second data set the recognition rate was 98.5%. As a final comparison between the three systems see Figure(7).

**Table 6:** Choosing the best standard deviation value with the highest recognition rate for one against all algorithm.

Standard deviation	Recognition rate (%)
0.5	50.7
1	66.1
2	72.3
2.5	70.7
3	69.2
4	67.7
5	66.1



**Figure 7:** A comparison between the three algorithms performance.

## CONCLUSIONS

Applications of support vector machines SVM are promising for classification and recognition of very high-dimensional data spaces. Precise and well experimented systems as described in this study allow reliable and robust recognition methodology of high dimensional hyperspectral data (even with using the full spectral resolution). From the previous results the best recognition rate was obtained using SVM neural network with the kernel Adatron algorithm with recognition rate of 99.2%. Current and future work is studying hyperspectral data from functionality point of view. For future work the use of splines to represent the full spectral resolution is suggested, trying to make the training process as easy as possible:

1. One against one algorithm which gave the following results:

- a recognition rate of 84% for the first data set,
- a recognition rate of 76.9% for the second data set,
- a recognition rate of 95.2% for the third one and
- 98.5 % for the fourth one.

2. The one against all algorithm which gave the following results:

- a recognition rate of 84% for the first data set,
- a recognition rate of 72.3% for the second data set,
- recognition rate of 94.6% for the third one and
- 98.5 % for the fourth one.

The ANN model uses:

1. The kernel-Adatron algorithm which gave the following results:

- a recognition rate of 84% for the first data set,
- a recognition rate of 91.5% for the second data set,
- a recognition rate of 94.6% for the third one and
- 99.2 % for the fourth one.

As the results show when the number of samples increases the recognition rates increase for all algorithms, for the data set of 15 samples per class the recognition rates of the three algorithms are nearly equal. From the previous results the best recognition rate was obtained using SVM neural network with the Kernel-Adatron algorithm with recognition rate of 99.2 %.

Although this work presents recognition models that are able to recognize the geological units in the study area, an extensive work is needed to build more wider data base for different geological units exist in Egypt for further classification problems.

Such a data base needs a Hyperspectral sensor to scan the area of interest to collect reflectance of different objects; this provides us with our ground truth in Egypt. The use of other transformation functions is suggested to represent the full spectral resolution of an image. Those functions are such as splines. Also we suggest a model uses the unsupervised learning methodology for classification of Hyperspectral data.

## REFERENCES

- Burges, C., 1998.** A tutorial on support vector machines for pattern recognition. *Data Mining Knowl. Discov.* 2 (2), pp. 121–167.
- Cao, L., and Tay, F., 2003.** Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans. Neural Network.* 14(6), pp. 1506–1518.
- Friess, T.; Cristianni, N. and Campbell, C., 1998.** The Kernel Adatron Algorithm :A fast and simple Learning Procedure for Support Vector Machines, *Machine Learning Proc. Of the fifteenth International Conference*, Morgan Kaufman Publishers, Sanfrancesco, California, pp. 131–177.
- Haykin, S., 1994.** *Neural Networks*, Macllan College Publishing Company.
- Hegazy, D., 2004.** Computerized Recognition of 3-D Objects, Master Thesis, Faculty of Computer & Information Sciences, Ain Shams university, Cairo, Egypt.
- Karimi, Y.; Prasher, S.; Patel, R. and Kimb, S., 2005.** Application of support vector machine technology for weed and nitrogen stress detection in corn, Elsevier.
- Lillisand, T., Ralph W., 1999.** *Remote Sensing and Image Interpretation*, fourth edition, John Wiley & Sons, Inc, New york.
- Merényi, E., and Villmann, T., 2002.** Self-Organizing Neural Net Approaches For Hyperspectral Images, *International Conference on Intelligent Computing and Information Systems*, Cairo, Egypt, June 24 - 26, 2002.
- Moustafa, A.R., 2004.** Geologic maps of the Eastern side of the Suez rift (western Sinai Peninsula), Egypt: AAPG/Data pages, Inc. GIS Series (Geologic maps and cross sections in digital format on CD).
- Rossi, F., and Villa, N., 2005.** Support vector machine for functional data classification, Elsevier.
- Vapnik, V., 1995.** *The Nature of Statistical Learning Theory*, Springer Verlag.

*Received July 12, 2007, Revised July 31, 2007*

## دراسة مقارنة بين اليات تدعيم المتجهات و الشبكات العصبية للتخريط الجيولوجى باستخدام البيانات غزيرة الاطيف

ايات محمد نجيب<sup>1</sup>, محمد عبد الهادى فرج<sup>2</sup>, محمد عادل يحيى<sup>2</sup>, حسن حسن رمضان<sup>3</sup> و محمد سعيد  
عبدالوهاب<sup>1</sup>

1- قسم الحسابات العلمية - كلية الحاسبات والمعلومات - جامعة عين شمس، القاهرة، مصر

2- قسم الجيولوجيا - كلية العلوم - جامعة عين شمس، القاهرة، مصر

3- قسم العلوم الأساسية - كلية الحاسبات والمعلومات-جامعة عين شمس، القاهرة، مصر

تطورت استخدامات بيانات أجهزة الاستشعار عن بعد من المستوى المرئى (3 أطيف) إلى عديدة  
الأطيف (عشرات الأطيف) حتى أصبحت غزيرة الأطيف (مئات الأطيف). وهناك تطبيقات عديدة  
للصور غزيرة الأطيف المستشعرة عن بعد، بشكل عام تطبيقات فى الحياة اليومية (زراعية، طبية... الخ)  
و بشكل خاص فى مجال علوم الأرض. و قد أظهر إستخدام هذه التقنية تقدما ملحوظا على المستويين العلمى  
و الأقتصادى.

ان للمجسات غزيرة الأطيف القدرة على إتقاط بصمات ضوئية مفصلة نستطيع من خلالها التفرقة  
بين المواد المختلفة الموجودة على سطح الأرض. معظم هذه المجسات لها القدرة على إتقاط الصور ابتداء  
من 0.4 ميكرون بداية الجزء الأزرق المرئى من المجال الطيفى، حيث تستطع هذه الأنظمة قياس الأطيف  
حتى 1.1 أو حتى 2.5 ميكرون. ونظرا لغزارة البيانات فإنها تحتاج إلى طرق غير تقليدية لإجراء عمليات  
التصنيف، التقسيم و التعرف على الأجسام.

فى هذا البحث تم دراسة طرق مختلفة لتحليل الصور غزيرة الأطيف باستخدام اليات تدعيم المتجهات  
(SVM) و الشبكات العصبية (ANN) للتغلب على مشكلة كثرة الأبعاد.

و كتطبيق فى هذا البحث قمنا بإستخدام SVM فى التخريط الجيولوجى للصور غزيرة الأطيف  
لمنطقة أبو زنيمة فى سيناء الغربية، مصر. و التى قد تم التقاطها فى 2003 بواسطة مجس مايبيرون على  
القمر الصناعى لمساحة الجيولوجية لولايات المتحدة (USGS) الملاحظة الارضية 1 (EO-1).

فى هذا البحث تم إقتراح نموذجين للتقسيم، الأول يستخدم SVM و الثانى يستخدم الشبكات  
العصبية (ANN) القائمة على SVM. بالنسبة للنموذج الأول قد تم استخدام خوارزمين لعملية التقسيم  
، الخوارزم الأول هو "واحد مقابل واحد" و الخوارزم الثانى هو "واحد مقابل الكل". أما بالنسبة للنموذج  
الثانى (ANN) فقد تم إستخدام خوارزم كيرنل- اداترون. و لعملية تحويل البيانات تم إستخدام جاوسيان  
كيرنل بقيمة مناسبة للانحراف المعيارى لجاوسيان.

قد تم تدريب نموذج الشبكات العصبية باستخدام اربع مجموعات الأولى تتكون من 11310 نموذج  
بمقدار تعرف 84%، الثانية تتكون من 22620 نموذج بمقدار تعرف بنسبة 91.5%، الثالثة تتكون من 33930  
نموذج بمقدار التعرف بنسبة 94.6%، الرابعة وتتكون من 45240 نموذج بمقدار التعرف بنسبة 99.2%.  
وكمقارنة بين مقدار التعرف بالنسبة للنموذجين إتضح أن نموذج SVM وقد أعطى الخوارزم "واحد مقابل  
واحد" مقدار التعرف بنسبة 84% للمجموعة الاولى؛ مقدار التعرف بنسبة 76.9% للمجموعة الثانية؛ مقدار  
التعرف بنسبة 95.2% للمجموعة الثالثة؛ مقدار التعرف بنسبة 98.5% للمجموعة الرابعة. أما الخوارزم  
"واحد مقابل الكل" فقد أعطى مقدار التعرف بنسبة 84% للمجموعة الاولى؛ مقدار التعرف بنسبة 72.3%  
للمجموعة الثانية؛ مقدار التعرف بنسبة 94.6% للمجموعة الثالثة؛ مقدار التعرف بنسبة 98.5% للمجموعة  
الرابعة.