DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

COMPARISON OF ADVANCED CLASSIFICATION TECHNIQUES IN REMOTE SENSING

by
Hüseyin YAŞAR

November, 2016
İZMİR

COMPARISON OF ADVANCED CLASSIFICATION TECHNIQUES IN REMOTE SENSING

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University In Partial Fulfillment of the Requirements for the Degree of Master of Science in Geographical Information Systems

by
Hüseyin YAŞAR

November, 2016
İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"COMPARISON OF ADVANCED CLASSIFICATION TECHNIQUES IN REMOTE SENSING"** completed by **HÜSEYİN YAŞAR** under supervision of **ASST. PROF. DR A. HÜSNÜ ERONAT** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

AssistantProfessor Dr. A. Hüsnü ERONAT

Supervisor

Doç.Dr.M.Kürşat Özgen

(JuryMember)

Doç.Dr. Özgün Yılmazoğlu

(JuryMember)

Prof.Dr. Emine İlknur CÖCEN
Director
Graduate School of Natural andAppliedSciences

ACKNOWLEDGEMENTS

First, I want to thank my advisor Assistant Professor Dr. A. Hüsnü ERONAT, for his valuable suggestions, encouragement, patience and support during this study.

I would also like to thank Bekir GÜL of General Command of Mapping for his contributions to this study and sharing his ideas during the development of this thesis.

I would also like to thank ASELMAK Company for help with in-situ work outs during the development of this thesis.

Finally, I want to give special thanks to my parents Hasan YAŞAR and Fatma YAŞAR for their endless support, patience and encouragement during my graduate student life.

Hüseyin YAŞAR

COMPARISON OF ADVANCED CLASSIFICATION TECHNIQUES IN REMOTE SENSING

ABSTRACT

Studies on remote sensing have become inescapable for almost all disciplines along with the advances in aviation and space technologies. Such as defence, city planning, forest services, naval surveys, mining etc many disciplines use remote sensing technique.

One of the parameters that makes this discipline important is classification studies. This parameter is divided into two: supervised and unsupervised. Multi and hyper spectral images are processed mathematically in this parameter and they help us become informed regarding land cover.

Basic definitions used in remote sensing are discussed in the first section of this study presented. The mathematical analysis of the support vector machines, artificial neural networks and decision trees, which are also known as further classification algorithms, are included in the second section. Urla semi-island, Izmir, Turkey has been classified by using LANDSAT 8 images by the help of these algorithms in the third and the final section of this study. Classification findings have been taken into verification testing by means of confusion matrix which is formed under favor of the spectral reflection values of the training classes. Besides, which algorithm has an impact on what kind of classes has been determined by the help of confusion matrix.

Keywords: Remote sensing, artificial neural networks, decision tree, support vector machines, confusion matrix.

UZAKTAN ALGILAMADAKİ İLERİ SINIFLANDIRMA TEKNİKLERİNİN KIYASLANMASI

ÖZ

Havacılık ve uzay teknolojilerinin gelişmesiyle birlikte uzaktan algılama çalışmaları hemen hemen birçok disiplin için vazgeçilmez bir unsur haline gelmiştir. Savunma, şehircilik, orman hizmetleri, deniz araştırmaları, madencilik, arazi kullanımı vb birçok alana hizmet etmektedir.

Bu disiplini önemli kılan parametrelerden bir tanesi de sınıflandırma çalışmalarıdır. Eğitimli (supervised) ve eğitimsiz (unsupervised) sınıflandırma olarak ikiye ayrılan bu parametrede çok bantlı ve hiper-spektral görüntüler matematiksel olarak işlenerek arazi örtüsü hakkında bilgi edinmeye yardımcı olur.

Sunulan bu çalışmanın birinci bölümünde, uzaktan algılamada kullanılan temel tanımlar hatırlatılmıştır. İkinci bolümde ileri sınıflandırma algoritmaları olarak adlandırılan destek vektör makineleri, yapay sınıf ağları ve karar ağaçlarının matematiksel analizine yer verilmiştir. Üçüncü ve son bolümde ise bu algoritmalar yardımıyla LANDSAT 8 uydu görüntüleri kullanılarak Urla yarımadası sınıflandırılmıştır. Sınıflandırma bulguları, egıtım sınıflarının spektral yansıma değerleri yardımıyla oluşturulmuş ve hata matrisi yardımıyla doğruluk testine tabi tutulmuştur. Bunun yanı sıra hata matrisi yardımıyla hangi algoritmanın ne tür sınıflar üzerinde etkili olduğu saptanmıştır.

Anahtar Kelimeler: Uzaktan algılama, yapay sinir ağları, karar ağaçları, destek vektör makineleri, hata matrisi.

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Remote sensing is the technique of obtaining information about an object through electromagnetic radiation without physical interaction with the object. It is used in many disciplines such as defense, mining and geographical information systems today. Remote sensing can be used by means of mainly satellites and air platforms in today's life.

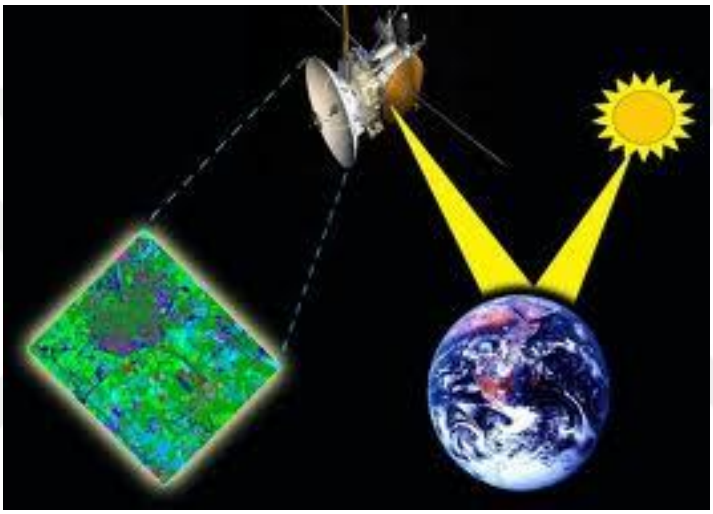

Figure 1.1 As seen above, the sun rays reflected from the earth are sent back to the earth for analysis after have been perceived by the travelling orbital satellites. Data is formed by processing images by means of computers and algorithms.

The pictures that have been achieved through remote sensing include a lot of information about the earth. The whole information is collected through the recording of the electromagnetic energy which is reflected from earth of surface on different tapes after have been perceived by the receptors of the satellites. Each tape has reflection values concerning the features that the relevant tape shows sensitivity

1.1 Images Processing

The data in satellite imaging are unrefined. Various mathematical analysis and statistical interpretation techniques are needed in order to convert these complex appearing data to information. The most common method of converting data to information is image classification. Image classification is the process of producing meaningful numerical subject maps from an image data set. The image which is gained as a result of classification is called as thematic map. There are two methods known as supervised and unsupervised classification that are widely used during classification process.

This section forms the base of remote sensing. The following operations are needed to analyze the data taken from sensitive satellites.

1.2 Preprocessing

Preprocessing is used to correct errors. Here, there are two various error correction types one of which is radiometric correction method. This method is used to overcome errors such as missing line, bit error and destripping which all arise from sensor and errors such as deformation of sun rays arrival angle, atmospheric effects which generate from environmental impacts.

1.3 Numerical Image Processing and Classification

1.3.1 Unsupervised Classification

This classification is combined together with spectral tape grouping algorithms which are obtained from algorithm image data. Grouping algorithms is the approach which is based on data similarities.

Each color of the vision processed through this algorithm corresponds to a class. The algorithm which is commonly used in unsupervised classification is K-

means algorithm. Another one is Iso-Data algorithm which will later be taken into consideration with supervised classification.

Unsupervised classification may be beneficial before the supervised classification of a multi-band image. The density of the classes, which form the land cover is mediated this way. The resolution of a multi-band image may be insufficient during the formation of a training set. Therefore, it may cause a problem in class selection. In-situ type of study is quite beneficial in overcoming the problem in such cases.

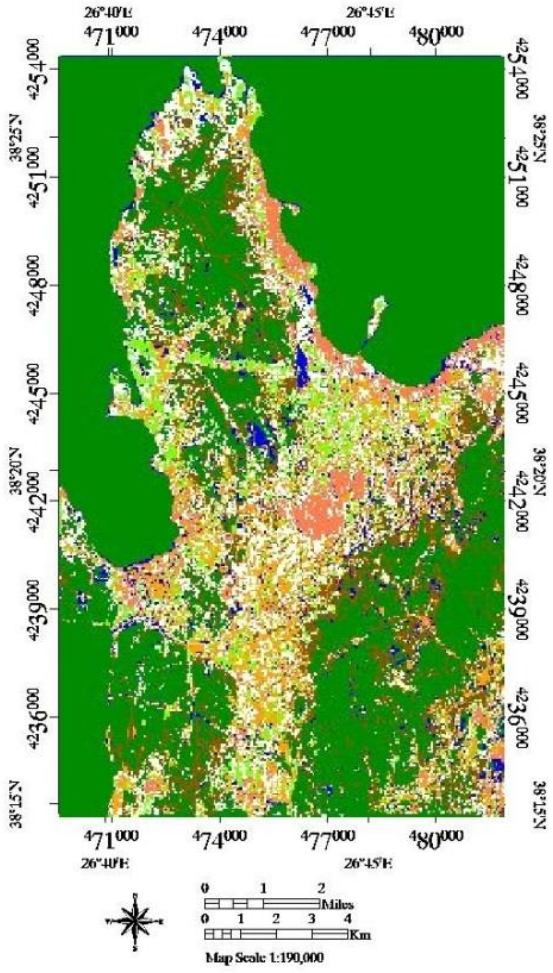

Figure 1.2 The view, classified as unsupervised through K-means algorithm, of Urla county of İzmir that was obtained from LANDSAT 8 satellite. There are a total of 7 classes for each color that represent a class in this image which is composed of 7 different colors. The labels of classes are uncertain as the classification is unsupervised.

1.3.2 Supervised Classification

This is the most effective classification type which is used in data analysis today. The user labels each class before the classification. Then either multi-spectral image is classified by means or classifying algorithm.

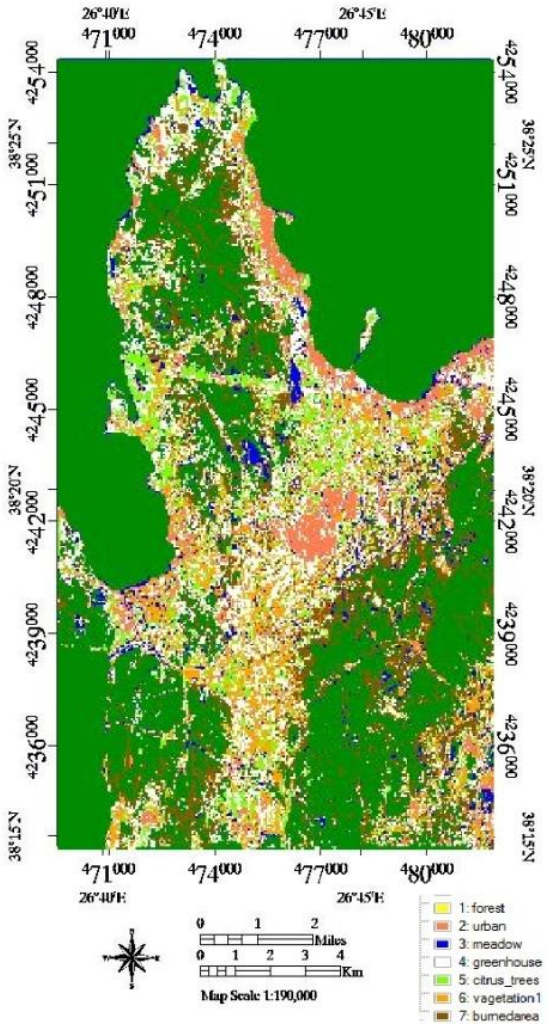

Figure 1.3 Taken by LANDSAT 8 was processed by the help of support vector machine algorithm. Seven classes, which are seen on the left, were determined by the user and mathematical processing period was initiated afterwards.

CHAPTER TWO
THE ANALYSIS OF THE ADVANCED CLASSIFICATION ALGORITHMS

Advanced classification methods, which are found in supervised classification group, were used as a base in this study. They are, in turn, supporting vector machines, artificial neural networks and decision trees.

Classification is a method by which labels or class identifiers are attached to the pixels making up a remotely sensed image on the basis of their characteristics. These characteristics are generally measurements of their spectral response in different wavebands. They may also include other attributes (e.g., texture) or temporal signatures. This labeling process is implemented through pattern recognition procedures, the patterns being vectors of pixel characteristics. The most commonly used classification methodologies used in remote sensing are unsupervised procedures such as ISODATA and supervised methods, the most popular of which is the maximum likelihood (ML) algorithm. The ML classifier is based on a probabilistic classification procedure. Which assumes that each spectral class can be adequately described or modeled by a multivariate normal probability distribution in feature space. The performance of this type of classifier thus depends on how well the data match the pre-defined model. If the data are complex in structure then the model the data in an appropriate way can become a real problem. In order to overcome this problem, which is inherent in statistical approaches, non-parametric classification techniques such as artificial neural networks (ANN) and rule-based classifiers are increasingly being used. Decision tree classifiers have, however, have not been used as widely by the remote sensing community for land use classification despite their non-parametric nature and their attractive properties of simplicity, flexibility, and computational efficiency (Friedl, 1997). The non-parametric property means that non-normal, non-homogeneous and noisy data sets can be handled, as well as, non-linear relations between features and classes, missing values, and both numeric and categorical inputs (Quinlan, 1993).

2.1 Support Vector Machines (SVM)

Support Vector Machines (SVM) classification method is based on statistical learning theory as a supervised classification technique. Its principles were developed by Vapnik & Friends in 1995. Although the mathematical algorithms which are based on SVM have been designed to classify linear data with two classes at first, later they are used in the analysis and classification of non-linear data and these with multi classes by improving the algorithms in use. The anticipation of the most convenient decision function which distinguishes one class from the other lies on the base of SVM. In other words, it is based on the identification of a hyper plane that perfectly separates two classes from each other. SVM are used in various fields such as the identification of optical characters, hand writings, faces as well as fingerprint analysis and text classifications. The most important characteristics of Support Vector Machines, which distinguishes them from other learning algorithms, is that its mathematical algorithm is based on the principle of Structural Risk Reduction (SRR) which is defined as the minimization of the expected risk upper limit at the end of classification process (Kechman, 2001).

Data occur in two different structures that are named as linearly separable and non-linearly separable in classification processes committed with SVM. Classification through linearly separable algorithm is quite easy but is possible to have many non-linearly separable problems in either everyday use or in classifications in remote sensing.

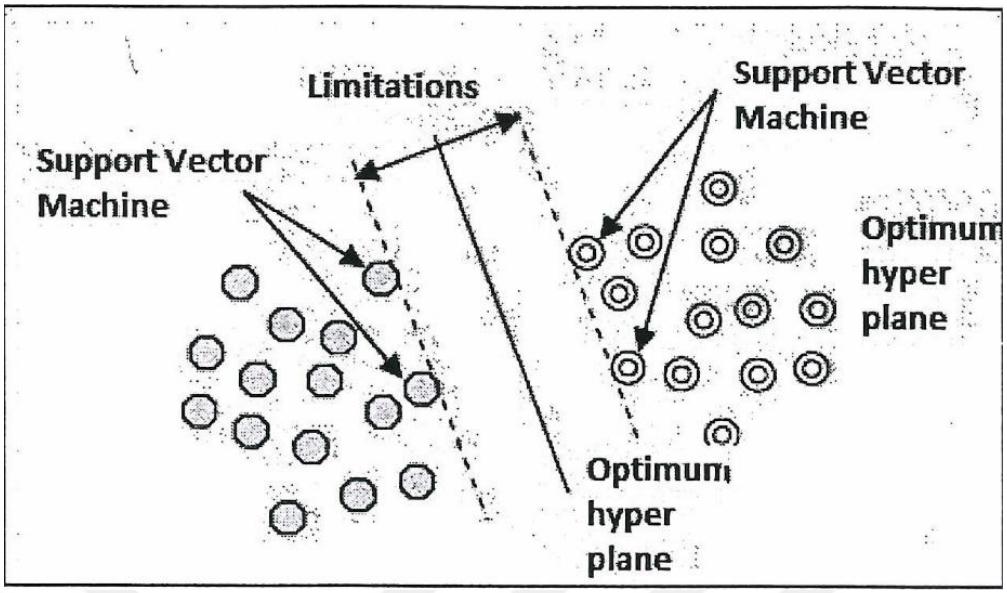

Figure 2.1 Represent SVM, which is one of the supervised classifications. The center in the middle represents the clearest border that separates axis classes from each other. The side axes represent risk axes. The further the inputs are from risk axes, the more reliable results are obtained. Besides, a hyperbolic plane the amount of which is (n-1) is needed so as to classify in different input.

2.1.1 Linear Discriminative Data

Linear discriminatory data are easy and practical to classify as stated in above section. Let us consider a training data set $\{x_i , y_i\}$with a linear structure which is composed of samples with the number of π and two classes and let this set be in series which is equal to J= L,…,K and x ∈ R^ndefined. A space with N dimensions shows class labels in y ∈'{-1,1}. We can state that hyper plane which separates two classes from each other can be determined in case of calculating the numerical size b which represents the diskance of the hyper plane from the origin and weight vector w which belongs to hyper plane as follows :

For $$w.x_1 + b \geq 1,\ y = 1 \quad (2.1)$$

For $$w.x_1 + b \leq 1,\ y = -1 \quad (2.2)$$

If we unite these two equalities in one singlet:

$$y_1(w.x_1 + b) \geq 0 \quad (2.3)$$

In this case, hypothesis set function gains the form of equation in 2.4

$$f_{w.b} = sign(w.x + b) \quad (2.4)$$

We can accept the w and b parameters in the equation 2.4 the same for the decision surface will not change. We can make use of the equation below to simplify the equation:

$$min_{i=l,\ldots,k}|w.x_1 + b| = 1 \quad (2.5)$$

This expression shows the elements of the data set $x_i,\ldots,x_k$. The hyper plane set, which provides equality, is defined as standard hyper plane. In 1995 Vapnik, stated that the Vapnik-Chervonenkis (VC) dimension of the standard hyper plane will be (N+L) which shows the total number of all independent parameters in the case that there will be no limitations except those used for (w,b) pairs. The capacity Vapnik, expressed has the meaning of measuring a complex situation. When a data set, which includes in point is considered the points which form these datasets can be labeled as either positive or negative. On the other hand, a problem with a difference of 2^N can be defined from data sets with an amount of N. When either of these problems is considered, we can say that hypothesis can be broken to pieces by the data set if the hypothesis which separates positive samples from the negative ones can be defined. This case explains that any learning problem which can be defined with N sample is detected accurately with a hypothesis drawn by H. The maximum number of points which can be broken apart by H is known as Vapnik-Chervonekis (VC) dimension. It is essential to form hyper planes set with variable VC dimension in order to make risk reduction principle work. It is also necessary that the experimental risk which expresses training classification error and the VC dimension are supposed to be minimum spontaneously. The distance between an X point and a hyper plane which is formed depending upon (w,b) pair can be expressed with the relation below:

$$d(x, w, b) = \frac{|x.w+b|}{||w||} \quad (2.6)$$

The distance between the standard hyper plane defined with (w, b) in equation 2.6 and the nearest data point to this plane is simply$1/||w||$ (Foody and Mathur, 2004). If sample groups are separated linearly, the best, among standard hyper planes which classifies the data best is the one which has the minimum norm or equally minimum$||w^2||$. This norm and VC dimension will be. When the norm is minimum in the event that linear separation can be done, it is equal to the separator hyper plane that contains the distance calculated among the lines which are perpendicular to the hyper plane among training data classes.

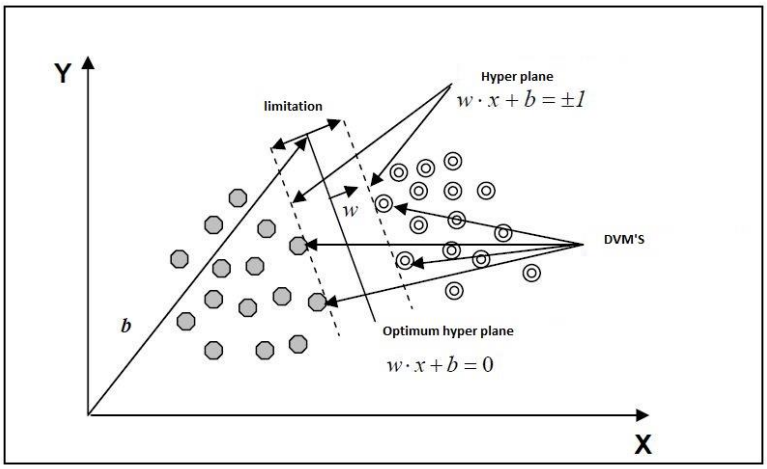

Figure 2.2 Represents the determination of the hyper plane in the classification of the data sets which can be separated linearly. Here, twin perpendicular lines form Support Vector Machines.

In $(x_1, y_1), \ldots., (x_k, y_k), x_i \in R^N, y_i \in \{-1,1\}$ we need to determine the hyper plane which is the most convenient discriminative in this classification. It will be possible by minimizing the expression [[w]] to maximize the border among the hyper planes on which SVM are located. In this case, the determination of the hyper plane with the maximum border is possible on the condition that,

$$min_{w,b} \frac{1}{2} [\![w]\!]^2 \quad (2.7)$$

And with the solution of the equation below,

$$y_i(w.x_i + b) - 1 \geq 0 \; i = 1, \ldots., k \tag{2.8}$$

Lagrange multipliers (λ_ji= 1,… ,k) are used to obtain the solution of the non-linear approaches stated in the equation above. The reason why we use language functions is that we can easily calculate the most convenient hyper plane parameters through Lagrange multiplications in such non-linear problems. Starting from this point, we can define the optimization problem as follows:

$$L(w, b, \lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{k} \lambda_i y_i (w.x_i + b) + \sum_{i=1}^{k} \lambda_i \tag{2.9}$$

When the equation above is extinguished by taking its partial derivatives, the expressions below are obtained:

$$\frac{\partial(w,b,\lambda)}{\partial w} = w^T - \sum_{i=1}^{k} y_i \, \lambda_i x_i = 0$$

$$\frac{\partial(w,b,\lambda)}{\partial b} = \sum_{i=1}^{k} y_i \, \lambda_i = 0 \tag{2.10}$$

Which is one of the equations are as follows:

$$w^T = \sum_{i=1}^{k} y_i \, \lambda_i x_i$$

$$\sum_{i=1}^{k} y_i \, \lambda_i = 0 \tag{2.11}$$

The above expressions are achieved finally. According to this theory, the points which provide the equations 2.1 and 2.2 might either have λ_i^q coefficients which are different than zero or these points form two hyper planes named as support vector. The points $\lambda_i^q \geq 0$ are at same time the ones which provide equations 2.1 and 2.2.

When the equation above is extinguished by taking its partial derivate, the expressions below are obtained:

$$L(w, b, \lambda) = \sum_{i=1}^{k} \lambda_i - \frac{1}{2} \sum_{j=1}^{k} \lambda_i \lambda_j y_i y_j (x_i . x_j) \quad (2.12)$$

Consequently, we can get the separating function for two classes as follows:

$$f(x) = sign(\sum_{i=1}^{k} \lambda_i y_i (x_i . x) + b) \quad (2.13)$$

2.1.2 Data Inseparable Linearly

It is possible to meet with data that are inseparable linearly in many problems in everyday life. The curves with which the most suitable separation is possible are needed to be fit into in cases when such a linear separation cannot be done.

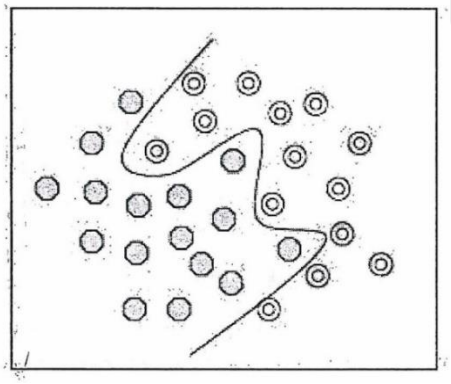

Figure 2.3The separation of a data set inseparable linearly with an ideal curve.

Cortez and Vapnik (1995) added $\xi_i$parameter to optimization model in classifications for the solution of the problem of in such data sets. This parameter is for the correction of the errors in classification. That is to say, if a member (s) of a class is in a different class, the error is corrected by the help of the ξ_i artificial parameter. Figure 2.4 represents the solution of the data sets which are inseparable linearly through Cortez and Vapnik approach.

Figure 2.4 the hyper plane which represents the solution (with ξ_i artificial parameter) of the data sets which are inseperable linearly. Here, the artificial variable takes the values $\xi_i \geq 0$. In the case that the artificial variable is added to equation 2.3 for the non-linear data, the following expression is achieved:

$$y_1(w.x_1 + b) - 1 + \xi_i \geq 0 \quad i = 1, \dots, k \tag{2.14}$$

In this case, while SVM algorithm is scanning the hyper plane which maximizes the border, it at the same aims to minimize the quantity about the amount of wrong classification errors. There is a proportion between maximizing the border and minimizing wrong classification errors. This balance is possible with the definition of a regulation parameter which is represented with C and takes positive values. The power to regulate C parameter is (0, ∞) this parameter expresses the maximum value that Lagrange multipliers can take. This way Lagrange multipliers and multiplications are provided to remain within the definition range of, $(0 \leq \lambda_i \leq C)$. Here, C is a parameter that is mediated by the user and it must be higher than zero. When C parameter is smaller than zero, this causes many data not to be classified. Contrary to this, when it is bigger than zero this provides less errors to occur. If C=∞, all data can be separated linearly. In other words, this permanent term regulates the coefficient of the minimum condition of the solutions which occur as a result when$\xi_i$variable takes big values. Parallel to this, the optimization problem for non-linear data turns to the equations below;

$$\min_{w,b,\xi_1 \dots, \xi_{ik}} \left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k} \xi_i\right] \tag{2.15}$$

$$y_1(w.x_1 + b) - 1 + \xi_i \geq 0 \tag{2.16}$$

$$\xi_i \geq 0, \quad i = 1, \dots, k \tag{2.17}$$

During the solution of this optimization problem when it is rearranged by using Lagrange function and Lagrange multipliers, it can be written as follows:

$$L(w, b, \lambda, \xi, \mu) = \frac{1}{2}\|w\|^2 - \sum_i \xi_i - \sum_i \lambda_i\{y_{i1}(w.x_i + b) - 1 + \xi_i\} - \sum_i \mu_i \xi_i \tag{2.18}$$

The μ and λ parameters in this expression are Lagrande multipliers and the solution of the equation is defined with Lagrange nodal points. While these crucial points are minimizing w, ξ and b parameters, they at the same time maximize the values related to the expressions $\lambda_i \geq 0$ and $\mu_i \geq 0$ when the partial derivatives of the equation above are taken and extinguished.

$$\frac{\partial L}{\partial w} = 0, w = \sum_{i=1}^{k} \lambda_i y_i x_i$$
$$\frac{\partial L}{\partial b} = 0, \sum_{i=1}^{k} \lambda_i y_i = 0 \tag{2.19}$$

$$\frac{\partial L}{\partial \xi}, \lambda_i + \mu_i = 0$$

Starting from this point, the Lagrange function becomes;

$$(\lambda) = \sum_{i=1}^{k} \lambda_{i-} \frac{1}{2} \sum_{i,j=1}^{k} \lambda_i \lambda_j y_i y_j \, (x_i x_j) \tag{2.20}$$

Here, the value range is;

$$\sum_{i=1}^{k} \lambda_i y_i = 0, i = 1, \ldots, k, \tag{2.21}$$

$$\boldsymbol{c} \geq \lambda_i \geq 0, i = 1, \ldots, k$$

2.1.3 Non-Linear Support Vector Machines

A non-linear classifier is needed when training data cannot be separated linearly. We can say that a vector, which can be defined in $x \in R^N$ can be matched in a space with higher dimensions (feature space) through a non-linear vector function in the form of $\emptyset: R^N \rightarrow R^F$

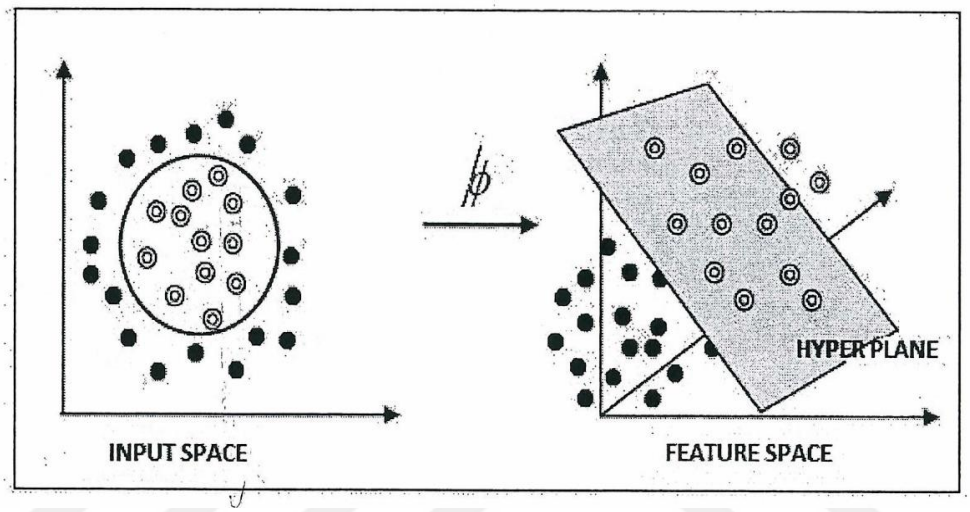

Figure 2.4 Illustrates that data, which are inseparable linearly, can separate linearly by their display in space with higher dimensions**.**

We can write $\emptyset(x_i)$ and $\emptyset(x_j)$ instead of $x_i x_j$ multiplier in equation 2.20 to determine the most suitable hyper plane or the solution for the most convenient border problem in the feature space. In this case, the solution of the optimization problem λ_i is $\lambda_i > 0$ in the feature space where transformation has taken place. The decision function for the solution of SVM will be in the form below by using this matching:

$$f(x) = sign(\sum_i \lambda_i y_i \emptyset(x_i)\emptyset(x) + b) \qquad (2.22)$$

As understood from equation 2.22, function is used for the solution of tha data which are inseperable linearly. This function provides a solution atmosphere by carrying data to a higher space. In this expression, $\emptyset(x_i)\emptyset(x)$ skaler multiplication is needed in kernel functions due to the difficulty of the operation.

$$K(x_i x_j) = \emptyset(x)\emptyset(x_i) \quad (2.23)$$

Is the way for us to define through Kernel function. Mercer theory is applied in the solution of these functions. Kernel function has to meet the following requirements in the solution of Mercer theory.

- There must be a continuous function.
- Symmetrical condition must be provided: $K(x_i x_j) = K(x_j x_i)$
- It must be of positive definition for any x value.

2.1.4 Data Processing in Classification with Support Vector Machine

Training and testing data, which include samples belonging to the data to be classified, are used for the solution of the problem in controlled classifications. Each sample in the training set represents the class in the classification. The purpose of SVM is to function as a decision mechanism which anticipates the labels of the data in the test set the feature values of which are certain.

Two parameters $C, \gamma$and $C, d$are in question in the use of polynom. Which parameter pair gives the best result for any problem is not known. Therefore, the most appropriate parameter must be mediated. A few methods developed with this purpose. The aim in these methods is to mediate the parameter pair which will give the best classification result so the classifier will be able to estimate the unknown class labels concerning the data in a more precise way. One of the most common methods in use is to divide the training data into two sections. The first section is used as training data during the education of the classifier areas the second section is put in process as unknown test data. By the help of this, it will be possible to express the accuracy of estimation for this data set or the classification performance of the unknown data in a more sensitive way.

This process in known as cross validation in literature. The training set is divided into sub-sets in equal dimensions and with a number of v first in v-multi

cross validation. Then another subset is tested by the classifier in training and the remaining v-1 subset. Every sample in all training sets is estimated that way and the percentage of the data classified correctly shows the accuracy of the cross validation. Grid scanning is one of the methods in use in the mediation of C, y and C,d parameters by using cross validation. Parameter pairs are tested by means of cross validation. Parameter pairs are tested by means of cross validation and the pair which has the best cross validation accuracy is mediated in this method. Experimenting the arrays which grow incrementally for the parameter pairs to be used in grid scanning method is the most practical way (Hsu et al., 2008).

Grid scanning in coarse range first will save time and sensitive grid scanning will take time. The pairs which provide the best cross validation accuracy are mediated by means of coarse grid scanning. Then the parameters which give the highest classification accuracy for the data set in this problem and the parameter pairs in lateral clearances are mediated by means of cross validation method. The training of the to the best classification accuracy this way. And then class labels are estimated for the classifier trained and for the whole data set (Çölkesen, 2009).

2.2 Decision Trees

Decision trees are one of the classification and pattern identification algorithms which is widely used. What lies on the base of its widespread use is that the principles used in the formation of tree structures are simple and comprehensible. If the tree structure that is formed according to the use is not very large, the comprehension of the emphasis used in carrying out the classification process and the formation of the tree will be quite simple.

The structure of a tree is composed of three basic parts which are knot, branch and leaf. Every feature in a tree structure is represented by a knot. Branches and leaves are the other elements of the tree structure.

The last part in tree is named as leaf whereas the upmost part is called as root. The parts lie between the root and the leaves are defined as branch (Quinlan, 1993).

There are two approaches in decision trees. They are Manuel Planning Method and Intuitive Assignment Method. Manuel Methods use statistics such as covariance matrix and average vector which are calculated for all classes. We will use covariance statistics of all classes in this thesis. A graphic which is obtained from the averages and variances of all classes is formed for each feature. This graphic is named as coincidental spectral drawing. It is possible to appoint the decision class at a point where all classes are separated in this graphics. The branches of the decision tree are mediated by naming every class one by one. The algorithm which will process the vision is made ready to mediate all classes on the graphics one by one.

It is considered that feature vectors and their class labels are within the training data in order to form a decision tree by using the intuitive assignment method. Then the decision tree turns the training data into a simpler form by separating it continuously. At the same time, test data with one or more features is implemented on every branch or knot point of the tree. This process is carried out in three steps which are separating knots, mediating the end knots and appointing class labels for end knot points to appoint class labels according to the votes of the majority or nominal votes when it is considered that some certain classes have higher possibilities than others.

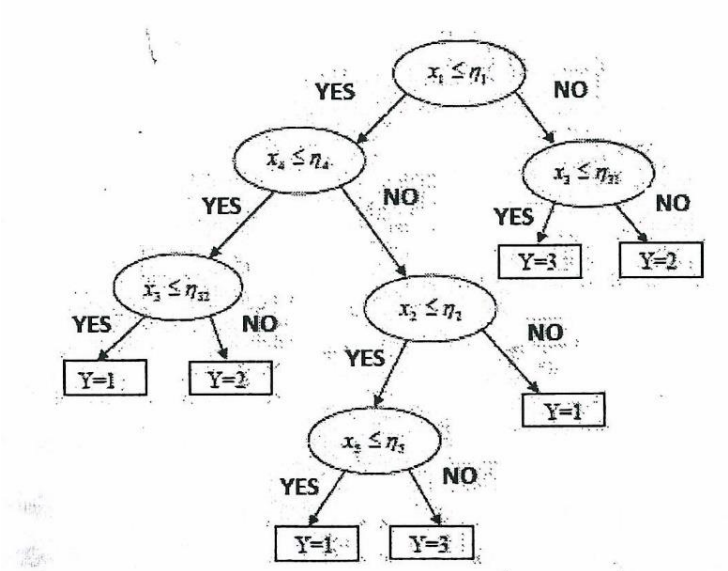

Figure 2.5 Stands for a typical decision tree. Ellipse shaped knots are the decision mechanisms of the tree. As comprehended from the gaps in the knots, x_1 values are determined by $\boldsymbol{\eta_1}$ threshold values. Rectangular knots are terminal knots and hold class labels.

A decision tree is a type of multi-stage classifier that can be applied to a single image or a stack of images. It is made up of a series of binary decisions that are used to determine the correct category for each pixel. The decisions can be based on any available characteristic of the dataset. For example, you may have an elevation image and two different multispectral images collected at different times, and any of those images can contribute to decisions within the same tree. No single decision in the tree performs the complete segmentation of the image into classes. Instead, each decision divides the data into one of two possible classes or groups of classes**.**

2.3 Artificial Neural Networks (ANN)

ANN algorithm is a system which is composed of digital neural which identify human brain uniquely and reflect its decision making ability to computer environment. This system, which is based on the biological theory of the human brain, can play an active role in many fields if defined correctly in accordance with its purpose. We have started to use this system in many fields such as face

detection systems, financial control systems, voice analysis, remote sensing, processing satellite imaging and its effective results have been observed. Important advantages of artificial neural networks are found below in articled form (Kavzoglu, 2001):

- They have nonparametric structure,
- They own arbitrary decision making ability,
- It is easy for them to unite different data types and input structures,
- They are able to generalize better,
- They have tolerance for noise.

The most important advantage is that ANN do not need any pre-information regarding the distribution of the data which makes it nonparametric. ANN can be defined as techniques free from data since they learn the characteristic features of the training in an iterative way. ANN may give better results by using a small number of training sets when compared with traditional statistical classifiers (Blamire, 1994; Kavzoglu, 2009).

Artificial neural networks (ANN) have some negativities since they are a more powerful classification method than traditional statistical classifiers. These negatives can be defined as requiring more training time, they are sensitive to the determination of the most effective network structure and parameter changes of the classification accuracy (Kavzoglu, 2001). Among them the net structure affects the training duration and classification accuracy directly (Dixon and Candade, 2008).

2.3.1 The Classification of Artificial Neural Networks

The array of neurons and the identification of the weights belonging to joints among neurons mediate the characteristics of the network in the formation of artificial neural network structure. When this situation is taken into account.

Artificial neural networks can divide into two groups according to their structure and learning algorithms.

2.3.1.1 Artificial Neural Networks According to Their Structure

Artificial neural networks are divided into two groups according to the integration types of artificial neurons with feed forward and ANN with feedback.

2.3.1.1.1 Artificial Neural Networks with Feedback. The output of a neuron is connected either to the layer before itself or to the neuron which is found in its own layer in artificial neural networks with feedback, which is contrary to those with feedforward. The inputs of a layer or all layers including the output layer in artificial neural networks with feedback are conveyed to the previous layer backward. Thus, inputs are conveyed to the previous layer backward. This, inputs are conveyed in either astern or forward direction. The out at any time in such neural networks is not only a function of the inputs at that moment but it also reflects the previous input and output values. Artificial neural networks with feedback, show non-linear structural feature through its such structure. Weights form a dynamic memory structure in such network types because the dataflow is two-way.

2.3.1.1.2 Artificial Neural Networks with Feedforward. Artificial neurons are generally composed of layers in a network with feedforward. Data are carried from the input layer towards the output layer by one-way joints.

Input data, which are identified according to parameters, are conveyed, distributed and classified towards output layer by one-way joints. There is no connection among artificial neurons in the same layer whereas all artificial neurons of a layer are associated with artificial neurons of the higher layer. The output element at any time is defined as a function of input vector which takes place simultaneously in such networks. Perceptron and (LVQ) Learning Vector Quantization networks are the ideal samples for such ANN mechanisms.

2.3.1.2 Artificial Neural Networks According to Their Learning Algorithms

To obtain output according to input data in artificial neural networks is to learn the data. ANN are divided into three classes according to their learning algorithms. These classes are controlled, uncontrolled and reinforced learning.

2.3.1.2.1 Controlled Learning. Attribute information is appointed to the networks in controlled learning. Then training data, which include class labels belonging to this attribute information are appointed to the networks as input data. The networks standardize its own weights in order to form desired outputs according to inputs given. The error between the outputs of the networks and expected outputs is calculated and the new weights of the networks are arranged according to this error margin. The difference between all outputs of the networks and expected outputs is mediated while calculating the error margin and then the error margin for each neuron is calculated according to this determination. Every neuron updates its own weights according to the calculated errors. Delta rule and LVQ algorithm are the ideal examples for such type of networks.

2.3.1.2.2 Uncontrolled Learning. Attribute data values the class labels of which are uncertain are used as input data during the training of the network in uncontrolled learning algorithm.

The networks mediate its own rules so as to classify every single sample among each other according to beginning data defined in the entrance layer of the network. That network completes its learning process by classifying joint weights and tissues with the same feature. ART and SOM are ideal sample algorithms for uncontrolled classification.

2.3.1.2.3 Reinforced Learning. Class labels belonging to input data in the entrance section are not used in this networks structure as in uncontrolled learning networks. A warning is necessary whether the result that the networks obtained at the end of its each iteration is either good or bad in reinforced learning algorithm. The networks rearranges itself according to this information given. Therefore, the networks continues the operation with any input array while learning and drawing a conclusion. Genetic algorithm can be given as a example for reinforced learning algorithms.

2.3.1.2.4 Multi-Layer Perceptron. It is the model that is widely used today. Multi-storey perceptron models are composed of neurons which are united in the form of flexible and multi-layers. They have a non-linear structure. The reason for this non-linear structure is it has superior neural networks. The algorithm of the networks with feedforward, which multi-layer perceptron networks structure is based on is composed of three different layers. As seen in Figure 2.7 the general structure of multi-layer perceptron are input layer, one or more hidden layer and output layer. The input layer, which represents input data, defines the distribution of the values that belong to each knot are obtained from the sum of the multiplication of the weights which belong to joints coming to this knot and input knot. Output layer is the final process layer that has a group of codes to Show the classes to be defined. All intermediate knot couplings are connected with weights. While passing through intermediate couplings, a value is multiplied by the weights which are connected to intermediate knot couplings (Kavzoglu and Reis, 2008).

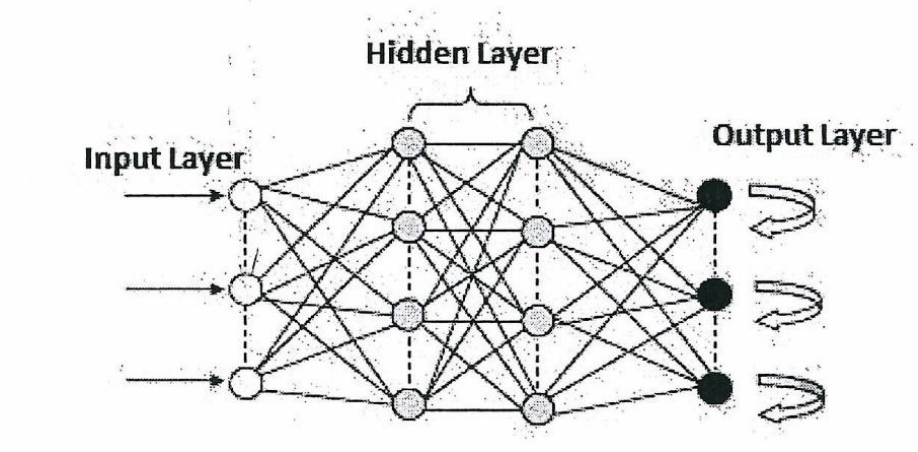

Figure 2.6 A simple multi storey artificial neural network with feedforward

2.3.1.2.5 Learning Algorithms. Learning algorithm forms a base for artificial neural network implementations. Learning algorithms are needed to form neuron networks and to determine the weights. Many learning strategies have been developed for different artificial neural network models (Kavzoglu, 2001).

Back prop learning algorithm in the training of the artificial neural networks with feedforward is one of the most popular of these models (Werbos, 1995).

Back prop learning algorithm, which is also known as generalized delta rule, is the method to learn iterative gradient reduction. Back prop learning algorithm aims to reduce the errors from output to input which is a backward direction. This operation is carried out in two steps. In the first step, the beginning weights, which belong to the whole networks, are mediated randomly, the input data is presented to the networks and they spread forward to mediate the output values. In the second step, error, which is the difference between the output values known and estimated output values is fed backward from the networks and weights are changed to reduce this difference (error). All these operations are repeated with recalculated weights in every iteration until the error is minimized or the pre-mediated threshold value is reached.

Operation knot or artificial neuron collects the input values that have been multiplied by the weights belonging to connections and estimates the output knot

by using the activation function. The equalities regarding the sum of input data and the output knot are as follows:

$$net_{Pj} = \sum_i w_{ji} i_{pi}$$

$$o_{pj} = f(net_{Pj}) \qquad (2.24)$$

In these equations, net_{pj} stands for the sum of input values, w_{ji} represents weight vector, j_{pi} symbolizes the first element value among the input values, o_{pj} shows the output knot for its j sample and f(.) displays non-linear activation function. The one that is used the most frequently among the non-linear activation functions is sigmoid function (Kavzoglu, 2001).

Back prop algorithm aims to minimize the error which is defined as the sum of the difference between the actual output values and calculated ones. This error is calculated for P sample by means of the following equality:

$$E_p = \frac{1}{2}\sum_j (tp_j - op_j)^2 \qquad (2.25)$$

In this equation, tp_j shows the calculated output value belonging to j. Component of the p sample whereas op_j displays the actual output value belonging to j. Component of the p input value. In this case, the total error of the net is calculated by using the following equality:

$$E = \sum E_P \qquad (2.26)$$

New weights are calculated in the following form by updating the previous weights in the form of ΔW_{ji} :

$$w_{ji}^{1} = w_{ji} + \Delta w_{ji}$$

$$\Delta w_{ji} = -n \frac{aE}{aw} \qquad (2.26)$$

N is defined as the rate of learning and it is a parameter mediated by the user. This parameter is used to control the degree of the change in weights concerning the errors calculated in each iteration (Kavzoglu, 2001).

Learning algorithms are used to mediate the characteristics of a data set. These learning algorithms mediate how to adjust the weights among training cycles or stages. The design of the network structure and the mediation of the parameters, which belong to the chosen algorithm, by the user form the most important two steps in ANN practices. The dimension of the hidden layer and the features concerning the number of knots are important from the view of the networks capacity in learning the features of the training data set and in the identification of the net. The number of the knots in the hidden layer directly designates the power and the complexity of the networks (Kavzoglu and Mather, 2003). The adjustment of the training parameters of the networks such as learning rate, momentum, starting values for weights, etc. Is the main factor that affects the working of the learning algorithm ant the performance of the trained network. These parameters are external and internal parameters. External parameters include the characteristics of the input value and work scale, whereas, internal parameters cover limitations such as the selection of the convenient network structure, the designation of the initial values of the network, parameters, iteration number, activation function type ant the mediation of the learning rate (Özkan, 2001). The trading of an artificial networks with feedforward through back prop algorithm requires initially a few parameters such as networks structure, rate of learning, momentum coefficient and activation function. The rate of learning has an important effect from the view of the success of the artificial nerve cells. This parameter states the amount of the change in the values of the weights in the connections. If the rate of learning is chosen very high, global minimum is point reached and a rise in the error is observed. If the rate of learning is chosen small, the training period of the network arises for the period necessary for searching the minimum error will increase. Momentum coefficient is added to the recently formed weights as the rate of weight changes which are calculated in the previous iteration

CHAPTER THREE
IMPLEMENTION

The multi- spectral images taken during the LANDSAT 8+ETM scanning of Urla semi-island of the city of Izmir, Turkey date of 30 April 2013 have been used in the implementation section. Calibration has been provided by doing radiometric and geometric corrections in order to raise the sensitivity before processing the images. These images were processed by the further classification algorithms, which are mentioned widely in Section 2, under the same conditions. The obtained results have been compared with training fields with the help of confusion matrix and the sensitivity of these algorithms has been tested.

3.1 Implementation Region

Urla (Greek: Boupya, Vourla) is a town and the center of a district of the same name in İzmir Province, Turkey. The district center is located in the middle of the isthmus of a small peninsula that protrudes northwards into the Bay of Izmir and which carries the same name as the town (Urla peninsula), but its urban tissue is comparatively loose and extends eastwards to touch the coast and to cover a wide area which also includes a large portion of the peninsula. Sizable parts in the municipal area, owned by absentee landlords, remain uninhabited or are very rural in aspect. The peninsular coastline present a number of compounds constituted by seasonal residences along the coves that are administratively divided between Urla center's municipal area or its dependent villages.

Figure 3.1 Is the view in true color of Urla county which was taken by LANSAT8 satellite in 2013

The eastern neighbor of Urla semi-island is Güzelbahçe, which is denser across this area, contributing to the whole district's average urbanization rate of 75%. With İzmir center (Konak) at a distance of only 35 km (22 mi), an important part of Urla's population is composed of residents, often wealthy, who commute to the city every day, access to and from İzmir and Çeşme, an international center of tourism at distance of 45 km (28 mi) from Urla, having been greatly facilitated by the construction of a six-lane highway. Urla district nevertheless manages to preserve an overall outlook of a pleasant suburb and resort, and as it extends to the West along Karaburun Peninsula, where it borders on the districts of Çeşme and Karaburun. Secondary residences built along the coast or large farms of the interior, as well as native villages, all bearing typical Aegean characteristics, increase in number. To the south, Urla semi-island neighbors that of Seferihisar and the settlement pattern is thinner in that section, with even some empty land, although housing projects targeting. İzmir's commuters start to show a rising interest for that section as well. In economic terms, agricultural products, and especially the fresh produce for the vast nearby market of İzmir, occupy a prominent place in Urla's economy, with fish, poultry and flowers standing out.

3.2 Materials used

3.2.1 LANDSAT 8 Images

LANDSAT 8 is an American Earth observation satellite launched on February 11, 2013. It is the eighth satellite in the LANDSAT program; the seventh to reach orbit successfully. Originally called the LANDSAT Data Continuity Mission (LDCM), it is a collaboration between NASA and the United States Geological Survey (USGS). NASA Goddard Space Flight Center provided development, mission systems engineering, and acquisition of the launch vehicle while the USGS provided for development of the ground systems and will conduct on-going mission operations.

The satellite was built by Orbital Sciences Corporation, who served as prime contractor for the mission. The spacecraft's instruments were constructed by Ball Aerospace and NASA's Goddard Space Flight Center, and its launch was contracted to United Launch Alliance. During the first 108 days in orbit, LDCM Underwent checkout and verification by NASA and on 30 May 2013 operations were transferred from NASA to the USGB when LDCM was officially renamed to LANDSAT 8.

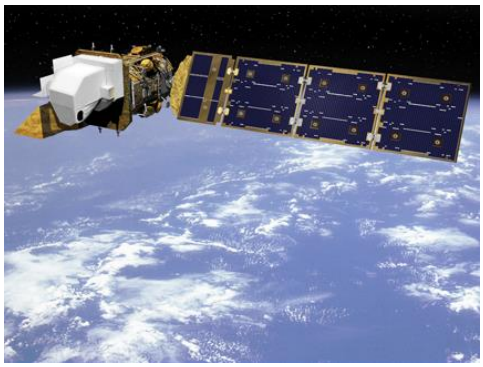

Figure 3.2 LANDSAT 8 satellite

Proving moderate-resolution imagery, from 15 meters to 100 meters, of earth land surface and polar regions, LANDSAT 8 operates in the visible, near

infrared, short wave infrared, and thermal infrared spectrums. LANDSAT 8 captures approximately 400 scenes a day, an increase from the 250 scenes a day on LANDSAT 7. The operational land imager (OLI) and thermal infrared sensor (TIRS) see improved signal to noise (SNR) radiometric performance, enabling 12-bit quantization of data allowing for more bits for better land-cover characterization.

Table 3.1 Technical Specifications of LANDSAT 8 Satellite

Spectral Band	Wavelength	Resolution
Band1- Coastal / Aerosol	0.433 - 0.453 µm	30 m
Band 2 - Blue	0.450 - 0.515 µm	30 m
Band 3 - Green	0.525 - 0.600 µm	30 m
Band 4 - Red	0.630 - 0.680 µm	30 m
Band 5 -Near Infrared	0.845 - 0.885 µm	30 m
Band6-Short Wavelength Infrared	1.560 - 1.660 µm	30 m
Band7-Short Wavelength Infrared	2.100 - 2.300 µm	30 m
Band 8 - Panchromatic	0.500 - 0.680 µm	15 m
Band 9 - Cirrus	1.360 - 1.390 µm	30 m

3.3 Processing Images

Processing images is assembled in four steps. These are respectively, pre-processing, designation of training areas, calculating complex matrix and processing the images via algorithms.

3.3.1 Pre-Processing

The selection of band is quite important in the classification of images. The reason is bands, on which reflection values of classes in process are clearer, are vital in the process of choosing and combining them. This is why the images taken from band 1, band 2, band 3, band 4, band 5, band 6, band 7 channels of LANDSAT 8 have been used in this study. Radiometric and geometric corrections have been done to increase the sensitivity.

Besides, the water section has been masked not only to reduce the time for the operation but also to increase the sensitivity.

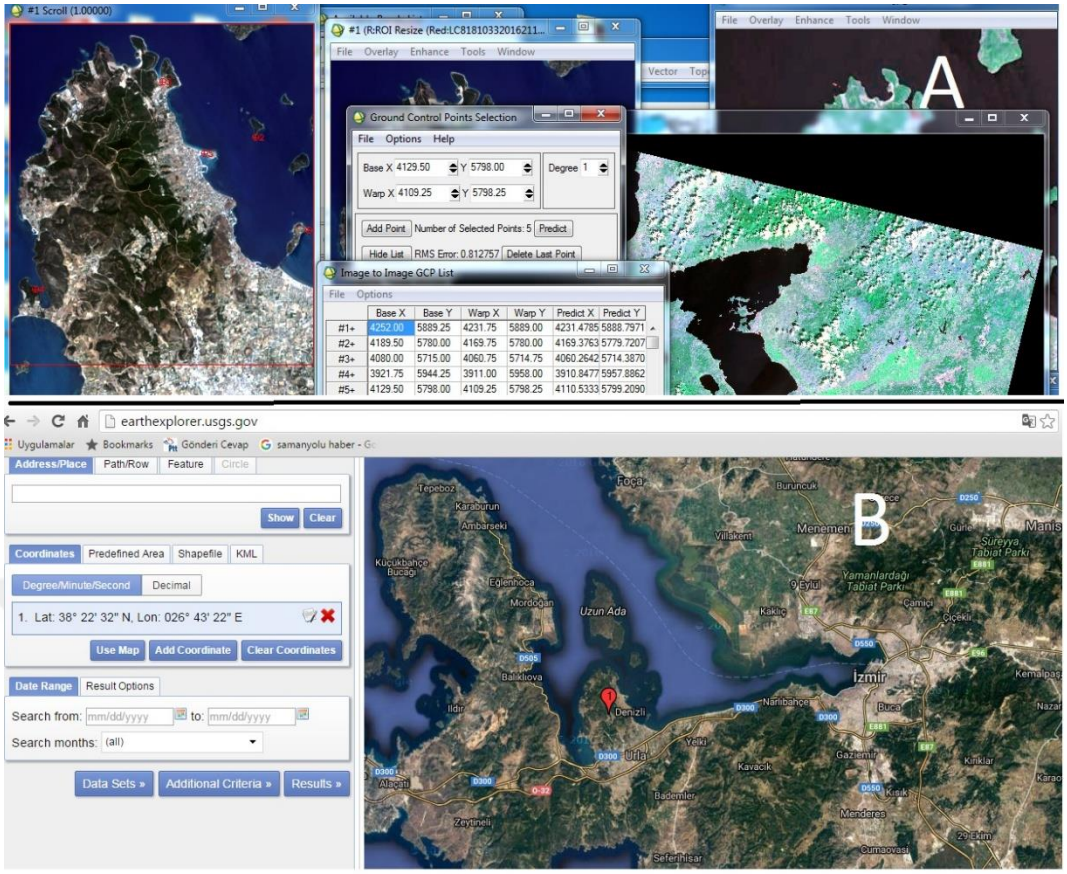

Figure 3.3 A is represent control point that gained from geo referenced map. B is represent of data acquired from USGS.

3.3.2 Designation of Training Areas

Unsupervised classification has been used while choosing training areas. Seven classes, which encompass the land cover the most, have been chosen in classification.

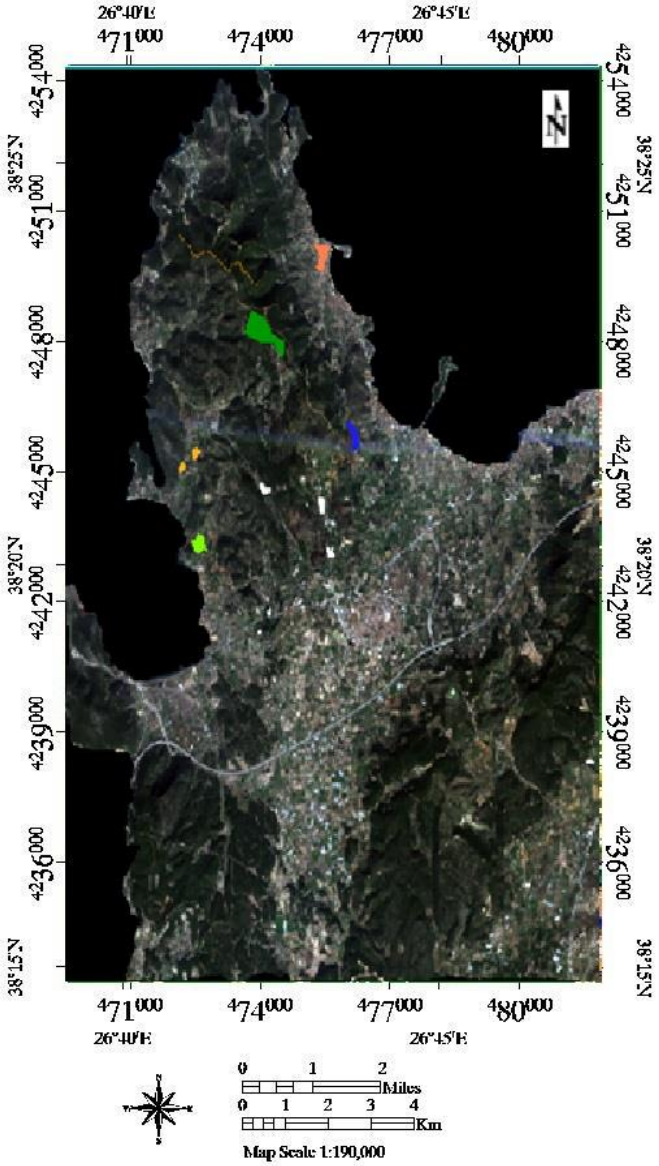

Figure 3.4 The processing of the multi-spectral image aims to determine training fields. Each different color represents one single area.

Moderate resolution images of the satellite of in-situ studies have been used to be sure that what was appointed in the choice of classes were all correct. Therefore, the certainty of the classes appointed has been determined. A GPS device with 2m sensitivity was used to help location justification during in-situ studies. We can see the in-situ and location studies in Figure 3.4.

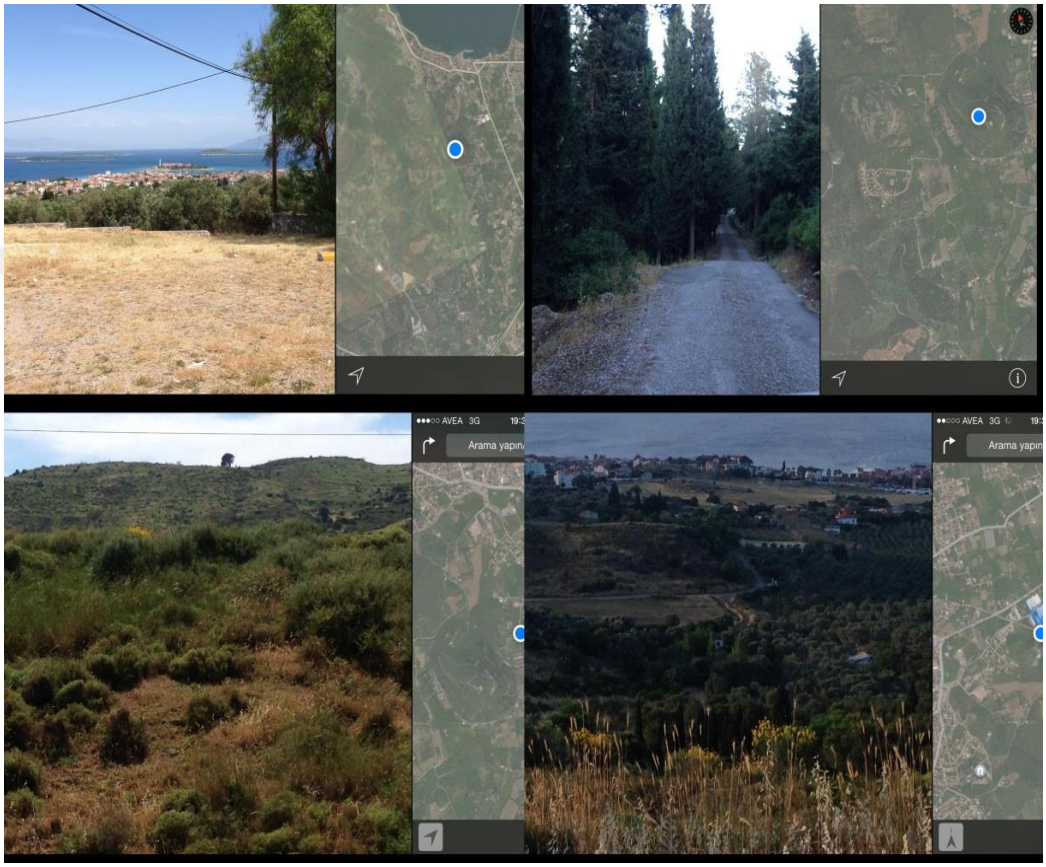

Figure 3.5 The areas where in-situ studies have been made by designating their locations with the help of GPS for the land covers which cannot be detected and for the true colors in satellite imaging.

3.3.3 Calculating Complex Matrix

This operation aims to reveal how accurate the results obtained in a classification study are by testing them with real data. The contribution of this operation to the study will be to test the accuracy of the image processing algorithms.

The overall accuracy is calculated by summing the number of pixels classified correctly and dividing by the total number of pixels. The ground truth image or ground truth ROIs define the true class of the pixels. The pixels classified correctly are found along the diagonal of the confusion matrix table, which lists the number of pixels that were classified into the correct ground truth class.

Kappa Coefficient, the kappa (κ): Coefficient measures the agreement between classification and ground truth pixels. A kappa value of 1 represents perfect agreement while a value of 0 represents no agreement (Jensen, 1986).

$$\kappa = \frac{N \sum_{i=1}^{n} m_{i,i} - \sum_{i=1}^{n} (G_i C_i)}{N^2 - \sum_{i=1}^{n} (G_i C_i)} \tag{3.1}$$

Where:

- i is the class number,

-N is the total number of classified pixels that are being compared to ground truth,

- $m_{i,i}$ is the number of pixels belonging to the ground truth class i, that have also been classified with a class i (i.e., values found along the diagonal of the confusion matrix)

- $C_i$is the total number of classified pixels belonging to class i

-$G_i$is the total number of ground truth pixels belonging to class i

Confusion Matrix (Percent): The ground truth (Percent) table shows the class distribution in percent for each ground truth class. The values are calculated by dividing the pixel counts in each ground truth column by the total number of pixels in a given ground truth class.

Commission: Errors of commission represent pixels that belong to another class that are labeled as belonging to the class of interest. The errors of commission are shown in the rows of the confusion matrix.

Commission: Errors of commission represent pixels that belong to the ground truth class but the classification technique has failed to classify them into her proper class. The errors of omission are shown in the columns of the confusion matrix.

Producer Accuracy: The producer accuracy is a measure indicating the probity that the classifier has labeled an image pixel into Class A given that the ground truth is Class A.

User Accuracy: User accuracy is a measure indicating the probity that a pixel is Class A given that the classifier has labeled the pixel into Class A.

3.3.4 Processing the Images by the help of Support Vector Machines

After training areas were determined, their images were processed first with the help of SVM algorithm. Functions such as linear kernel, polynomial kernel, radial floored kernel and sigmoid kernel were used made use of while the images are processed in this algorithm.

3.3.4.1 The Implementation of Linear Kernel Function

Linear kernel function is part of SVM. Linear kernel function can mathematically be defined as follows:

$$K(X_i, X_j) = X_i^T x_j \quad (3.2)$$

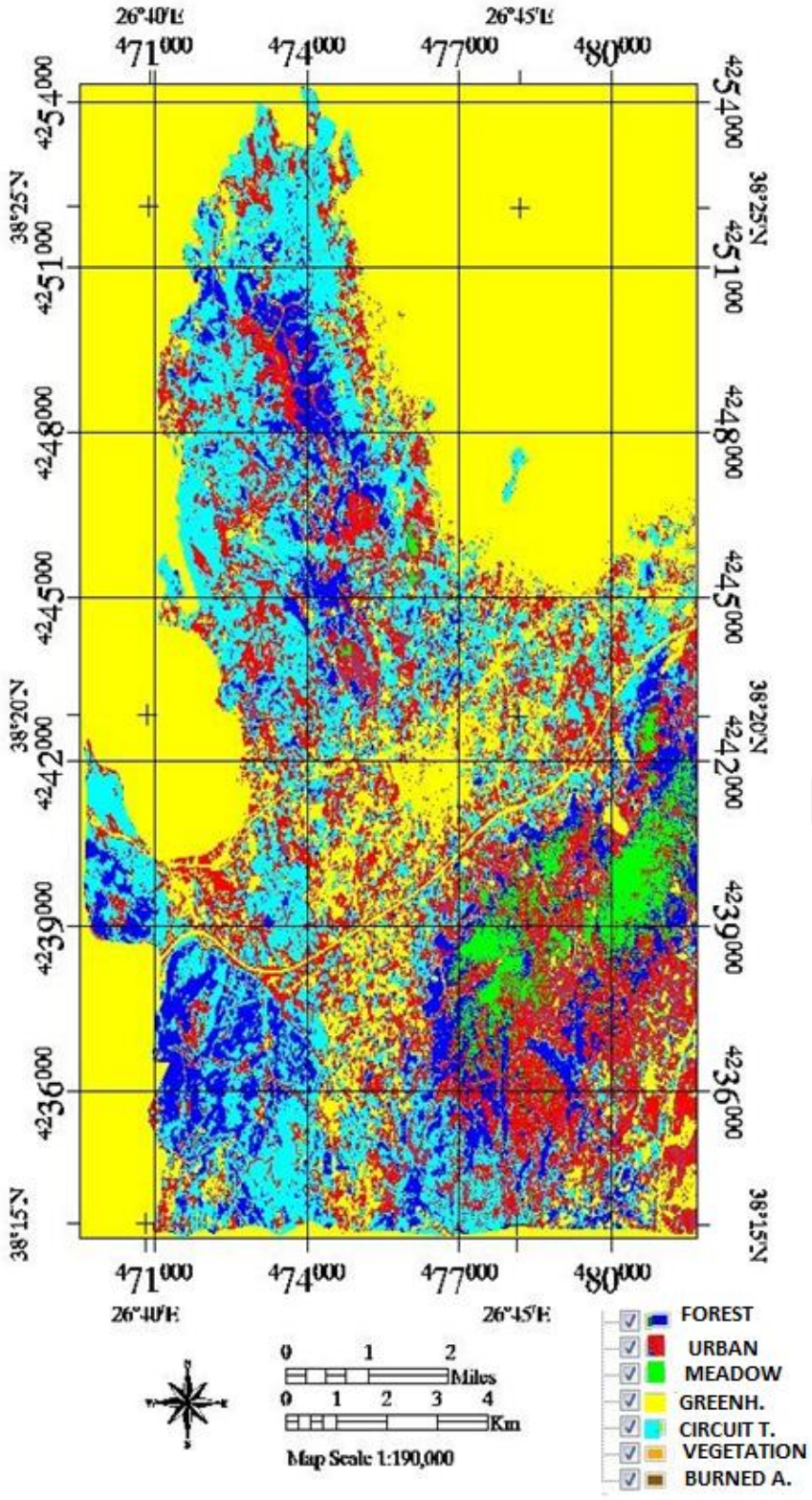

Figure 3.6 The image classified by the help of linear kernel function.

The analysis of the processed image through linear kernel functions is shown given in Table 3.3. The sensitivity of the algorithm used has been tested by the help of the data on Table 3.3 which were obtained by help of complex matrix.

Table 3.3 Data obtained with the help of complex matrix for linear kernel function.

Class	**Prod. Acc.(percent)**	**User Acc.(percent)**	**Prod. Acc.(pixel)**	**User Acc.(pixel)**
Forest	99.30	95.09	426/429	426/448
Urban	91.84	99.26	135/147	135/136
Meadow	87.50	98.82	84/96	84/85
Greenhouse	69.17	71.21	92/133	92/129
Citrus trees	80.73	80.00	88/109	88/110
Vegetation	72.46	72.46	50/69	50/69
Burned area	81.58	75.61	62/76	62/82

Table 3.3 general accuracy of 87.4797% when complex matrix data have been examined. When examined in a more detailed way, the accuracy in the classification of forest areas is high. Controversially, this has no effective role in the exploration of greenhouse areas.

3.3.4.2 The Algorithm of Radial Kernel Function

Radial kernel function part of SVM. Radial kernel function can mathematically be defined as follows:

$$K(X_{\mathrm{i}}, X_j) = \exp\left(-g\left\|x_i - x_j\right\|^2\right), g > 0 \qquad (3.3)$$

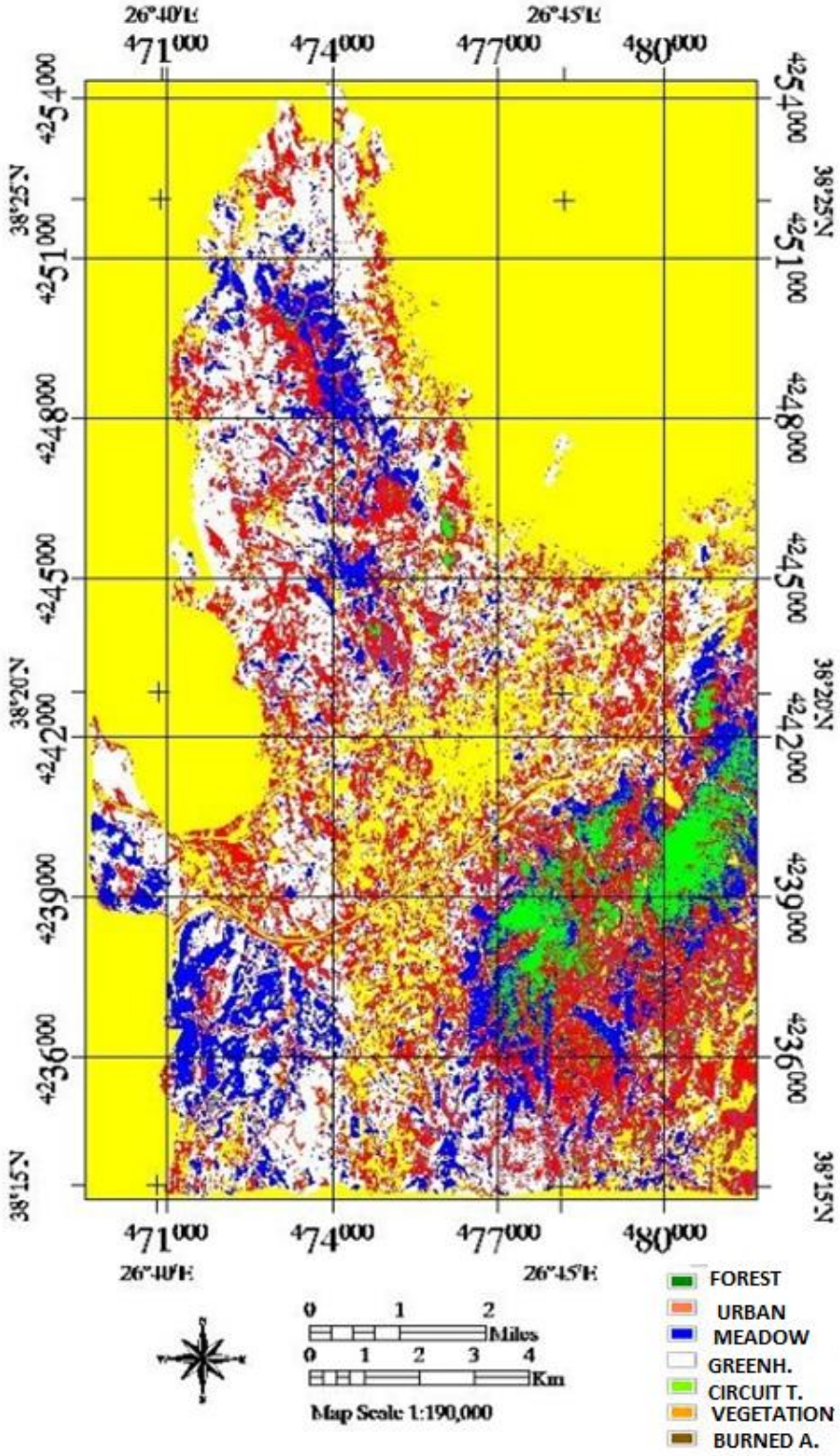

Figure 3.7 The view classified with the help of radial kernel functions

Table 3.4 Data obtained with the help of complex matrix for radial kernel function.

Class	**Prod. Acc.(percent)**	**User Acc.(percent)**	**Prod. Acc.(pixel)**	**User Acc.(pixel)**
Forest	99.07	95.08	425/429	425/447
Urban	91.80	92.96	132/147	132/142
Meadow	87.50	97.67	84/96	84/86
Greenhouse	63.91	73.28	85/133	85/116
Citrus trees	79.82	79.09	87/109	87/110
Vegetation	69.57	68.57	48/69	48/70
Burned area	77.63	67.05	59/76	59/88

After examining the data in Table 3.4, it is seen that the support vector machines which have been formed with radial kernel functions are effective in the classification of forest areas. The value of gamma term in this algorithm has been taken as 0.143. It can be calibrated through gamma term to increase the sensivity of the results obtained when needed. On the other hand, it can be said that this algorithm has not given results with sufficient sensitivity in the designation of greenhouse areas.

3.3.4.3 Polynomial Kernel Function

Polynomial kernel function is part of SVM, polynomal kernel function can mathematically be defined as follows function 3.3:

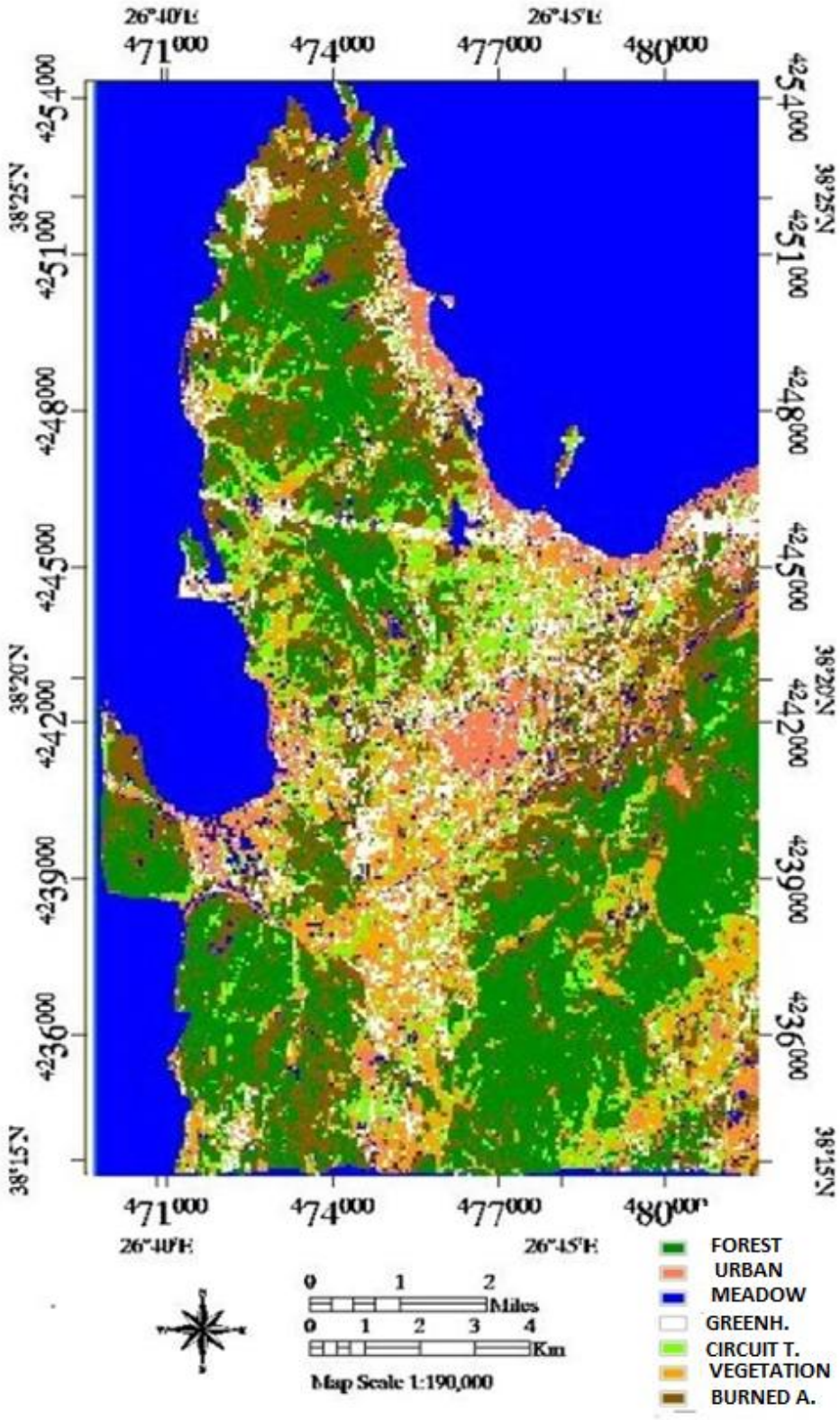

Figure 3.8 The view classified with polynomial kernel functions

$$K(X_i, X_J) = gX_i^T x_j^{\ d}, g > 0 \quad (3.3)$$

Where,

g is the gamma term in the kernel function for all kernel types except linear.

d is the polynomial degree term in the kernel function for the polynomial kernel.

r is the bias term in the kernel function for the polynomial and sigmoid kernels.

The value of gamma term has been taken as 0.143 and bias term as 1 in this classification. It can be seen that the average accuracy is 86.87% according to the results obtained with the help of complex matrix.

Table 3.5 Data obtained through complex matrix for polynomial kernel function

Class	**Prod. Acc.(percent)**	**User Acc.(percent)**	**Prod. Acc.(pixel)**	**User Acc.(pixel)**
Forest	99.07	95.08	425/429	425/447
Urban	89.80	92.96	132/147	132/142
Meadow	87.50	97.67	84/96	84/86
Greenhouse	63.91	73.28	85/133	85/116
Citrus trees	79.82	79.09	87/109	87/110
Vegetation	69.57	68.57	48/69	48/70
Burned area	77.63	67.05	59/76	59/88

Eventhough though it cannot be said that polynomial kernel functions give different results than those of radial kernel functionsare quite sensitive in the classification of urban areas.

3.3.4.4 Image Processing Through Sigmoid Kernel Functions

The values taken for bias is 1 and for gamma is 0.143 during sigmoid classification. General sensitivity has been determined as 82.34% when confusion matrix values are examined. In addition to this, the time spent for computer studies is much more than other algorithms.

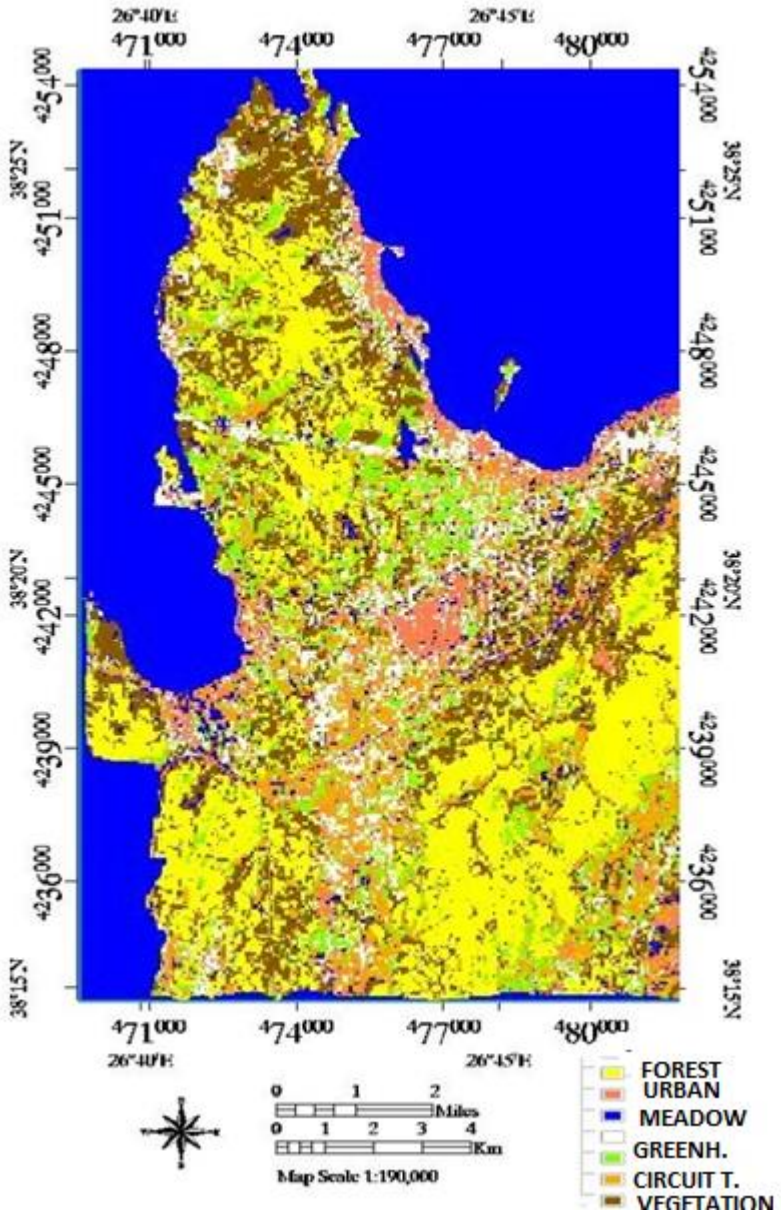

Figure 3.9 The view classified through sigmoid function

$$K(X_i, X_j) = \tanh(g x_i^T x_j + r) \quad (3.4)$$

-g is the gamma term in the kernel function for all kernel types except linear.

-r is the bias term in the kernel function for the polynomial and sigmoid kernels.

Table 3.6 The accuracy analysis of the classification obtained by the help of sigmoid kernel function

Class	**Prod. Acc.(percent)**	**User Acc.(percent)**	**Prod. Acc.(pixel)**	**User Acc.(pixel)**
Forest	98.37	93.78	422/429	422/450
Urban	74.15	84.50	109/147	109/129
Meadow	84.38	94.19	81/96	81/86
Greenhouse	60.90	66.94	81/133	81/121
Citrus trees	78.90	75.44	86/109	86/114
Vegetation	63.77	61.97	44/69	44/71
Burned area	64.47	55.68	49/76	49/88

The SVM, which has been formed with sigmoid function, has given results with a lower accuracy compared with other algorithms. In addition to this the time spent for computer studies is much more than others.

3.3.4.5 Results Obtained by Support Vector Machines

All functions of SVM have shown a great success in processing coniferous areas.

Linear kernel functions have shown high level sensitivity in the determination of burned areas.

Greenhouse areas reflect most of the coming light because they have high reflection parameter. Eventhough it is easy to identify these areas for this reason,

it is hard to distinguish them from white objects such as clouds, a house with a white roof, etc. so thermal support is needed.

Linear kernel functions give sensitive results in the determination of citrus trees.

All SVM have given results with ideal sensitivity in the determination of meadow areas.

Even though linear kernel functions show 72.46% sensitivity in the analysis of vegetation areas, SVM are not an ideal algorithm in the identification of such areas.

3.3.5 Artificial Neural Networks

Another algorithm, which is used in image processing is artificial neural networks. The parameter values which are used in processing images are as follows:

Training Threshold Contribution: 0.9, Training Rate: 0.2, Training Momentum: 0.9, Training RMS:0.1

Training Threshold Contribution: Determines the size of the contribution of the internal weight with respect to the activation level of the node. It is used to adjust the changes to a node's internal weight. The training algorithm interactively adjusts the weights between nodes and optionally the node thresholds to minimize the error between the output layer and the desired response. setting the training threshold contribution to zero does not adjust the node's internal weights. Adjustments of the nodes internal weights could lead to better classifications but too many weights could also lead to poor generalizations.

Training Rate: The training rate determines the magnitude of the adjustment of the weights. A higher rate will speed up the training, but will also increase the risk of oscillations or non-convergence of the training result.

Training Momentum: Entering a momentum rate greater than zero allows you to set a higher training rate without oscillations. A higher momentum rate trains with larger steps than a lower momentum rate. Its effect is to encourage weight changes along the current direction.

Training root meen square (RMS): In the training RMS exit criteria field, enter the RMS error value at which the training should stop. If the RMS error, which is shown in the plot during training, falls below the entered value, the training will stop, even if the number of iterations has not been met. The classification will stop, even if the number of iterations has not been met. The classification will then be executed.

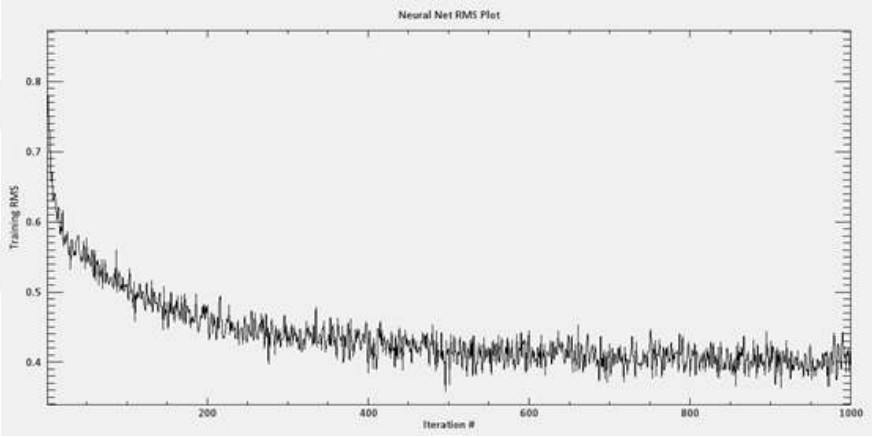

Figure 3.10 RMS curve obtained with 1000 iteration

Table 3.7 Confusion matrix data obtained through artificial neural network

Class	**Prod. Acc.(percent)**	**User Acc.(percent)**	**Prod. Acc.(pixel)**	**User Acc.(pixel)**
Forest	98.83	97.25	424/429	424/450
Urban	94.56	94.50	139/147	139/129
Meadow	91.67	98.88	88/96	88/86
Greenhouse	69.92	73.81	93/133	93/126
Citrus trees	81.65	85.44	89/109	89/104
Vegetation	78.57	61.97	48/69	48/58
Burned area	90.79	55.68	69/76	69/99

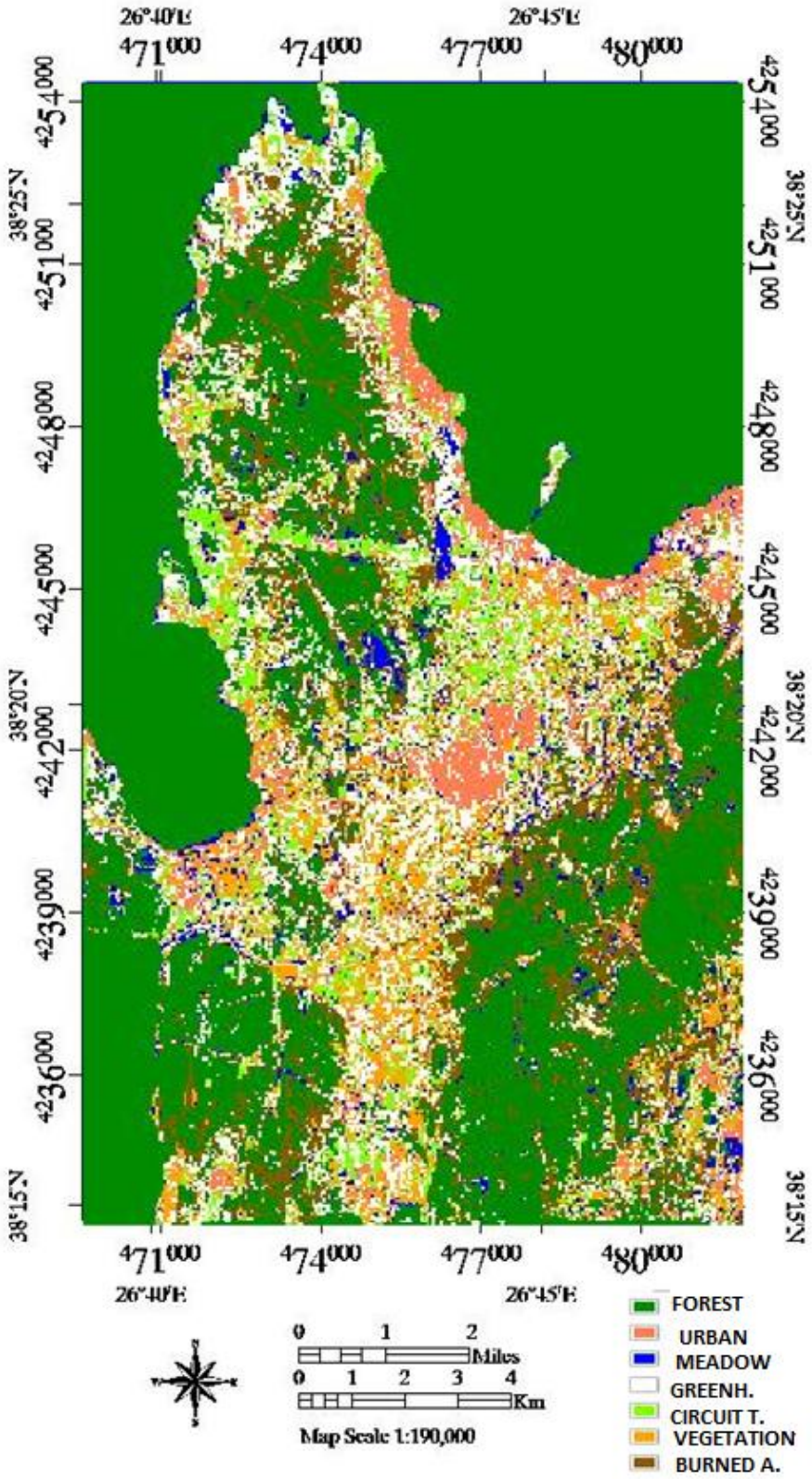

Figure 3.11 View classified through artificial neural network

Sea is excepted from the classification. One of the artificial colors in the classification is automatically appointed to the masked area, which is separate, because of masking. The reason for masking the sea section is to prevent spectral perturbation. Therefore, the sensitivity of the classification has been increased.

It can be seen that the general accuracy is 89.90% when Table 3.5 is examined. This value is quite operative for the classification accuracy. It gives results with high accuracy in the detection of meadow areas, urban regions, coniferous areas and burned regions.

3.3.6 Decision Trees

The basic parameter is the parameter of reflection in the formation of decision trees as in all classification studies. Statistical data base is obtained for each band of multi-spectral image according to this parameter. Proportioning is done according to this data. Critical points for classification are mediated by taking these proportions as a base. Classes are appointed according to these critical points.

The spectral values of the multispectral image are used in Figures 3.12, 3.13, and 3.14

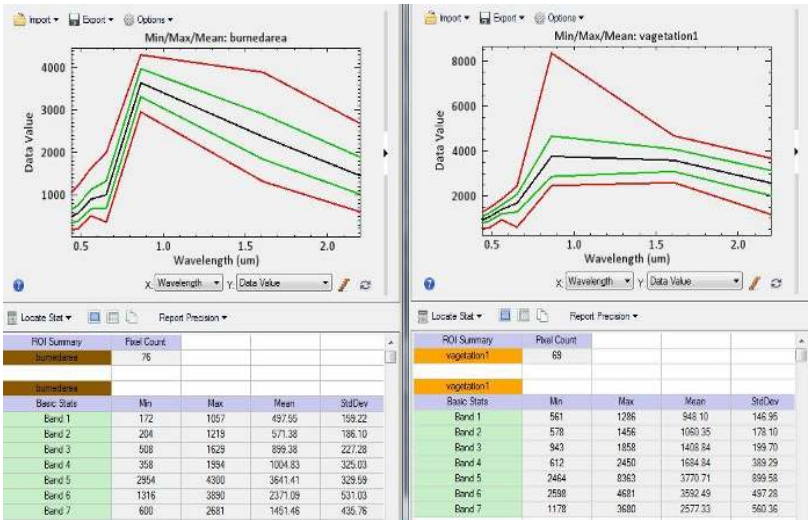

Figure 3.12 The spectral values of burned areas and vegetation regions

The ratio of band 5 and band 4 is taken into consideration to explore the burned areas when statistical averages are taken in Figure 3.12. It is emphasized that ratio must be higher than 3. Besides, it is emphasized that this ratio is lower than 3 after taking the ratios of band 6 and band 4.

The knot, which represents burned areas, has been determined by taking this emphasis into consideration.

In Figure 3.12, it has been emphasized that the ratio of band 7 to band 1 is more than 2.5 as a value when statistical averages are taken to explore the operative areas. It has also been emphasized that the ratio of band 6 to band 4 is less than 2.2. Therefore, a mediating knot has been formed to explore the effective areas when all these stresses are combined.

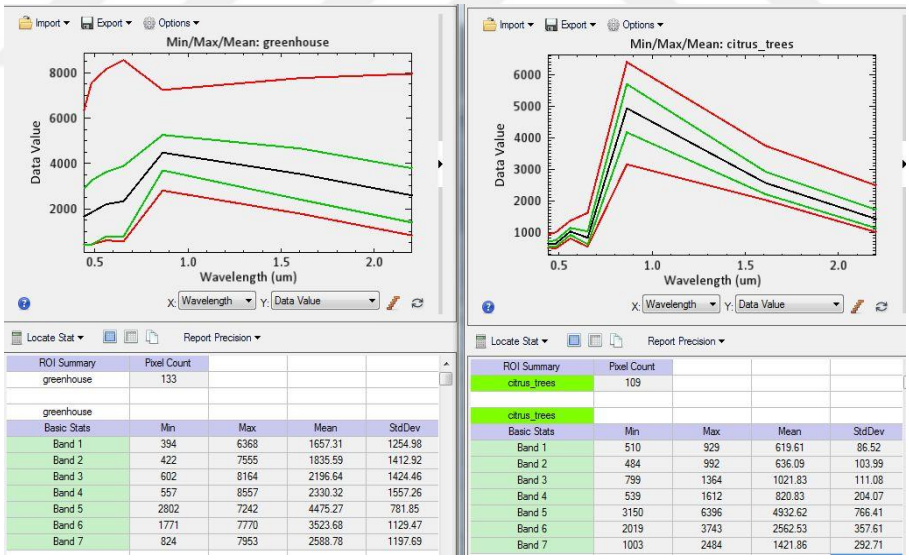

Figure 3.13 The spectral values of greenhouse and citrus trees.

It has been mediated that the ideal band ratio to mediate greenhouse areas is band 6 and band 4 when spectral values in Figure 3.13 are investigated. It has been stressed that this ratio is less than 1.6 the ratio of band 7 to band 1 which is less than 2.5 and the ratio of band 4 to band 3 which is smaller than 0.9 have all

been determined and it has been emphasized that these values are ideal in the designation of citrus trees when the spectral values in Figure 3.12 are taken into consideration.

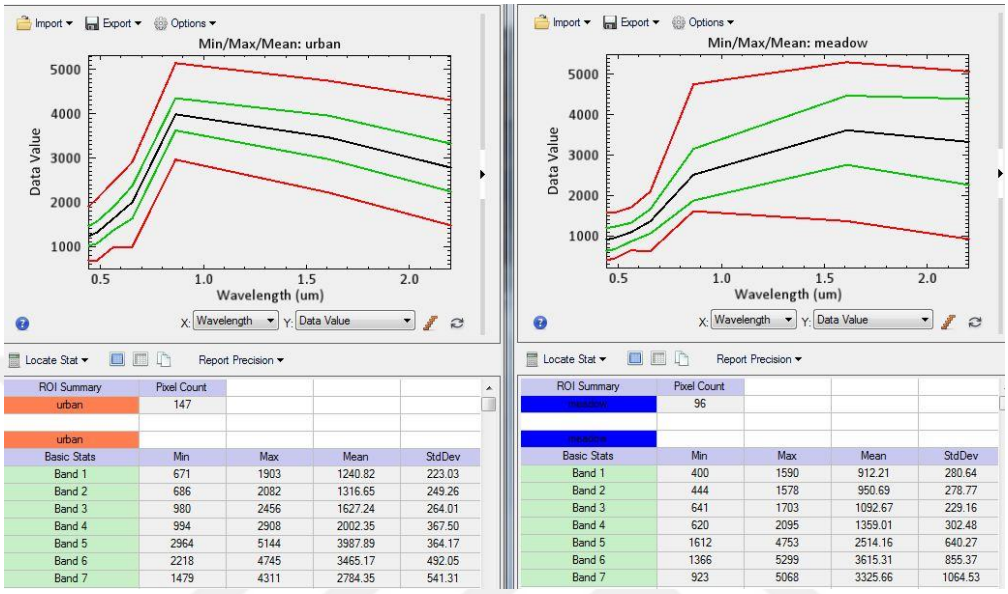

Figure 3.14 The spectral values of urban and meadow area

It has been determined that the ideal band ratio for the meadow areas is that of band 7 to band 1 when Figure 3.14 is examined.

Table 3.8 The accuracy analysis of the classification obtained with the help of decision trees function

Class	**Prod. Acc.(percent)**	**User Acc.(percent)**	**Prod. Acc.(pixel)**	**User Acc.(pixel)**
Forest	98.20	93.78	422/429	422/450
Urban	73.15	84.50	109/147	109/129
Meadow	82.38	94.19	81/96	81/86
Greenhouse	61.90	66.94	81/133	81/121
Citrus trees	82.90	75.44	86/109	86/114
Vegetation	64.77	61.97	44/69	44/71
Burned area	61.47	55.68	49/76	49/88

As seen in Table 3.8, general accuracy is 82%. Although decision tree algorithm is has got high accuracy for forest and meadow areas but not sufficient enough for other classes. As seen in Table 3.8, general accuracy is 82%. Although decision tree algorithm is has got high accuracy for forest and meadow areas but not sufficient enough for other classes.

CHAPTER FOUR
CONCLUSION

Data of classified images of Urla semi-island via algorithms were analyzed. With regard to this analysis, algorithms show diversity over of classes. Partially these algorithms may be effective on determination of subclasses but we couldn't say that for an algorithm suitable for all classes. For example while SVM is more accurate for urban, ANN is more accurate for urban. Results of test show that ANN have the highest success in the analysis or general accuracy (89.90%). Also this algorithm is the most ideal way to discover meadow and urban areas. But considerably compulsive for computer time than other algorithms Linear kernel function test have show that 99.30% for forest, which is a function of support vector machines, is the ideal algorithm in the analysis of forest areas. Decision trees are quite effective and practice in object based studies since they have calibration feature. Also this algorithm is the most effective method in the classification of the vegetation areas. The use of thermal camera images will give more reliable results during classification of Urla semi-island studies in greenhouse areas because the sensitivity in the classification of such areas is quite low.

The thesis inadequate for some stuations. For example algorithms show difference with regard to classes additionally they show difference with regard to tesserial. On the other hands some algorithms show that it can be different accurate for different satellite images. But the thesis may useful target determination or classification of some special areas same Urla semi-island.

REFERENCES

Altun, H. and Curtis, K.M., (1998). Exploiting the statistical characteristic of the speech signals for an improved neural learning in a MLP neural network. *Neural Networks for Signal Processing*, 598, 547-556.

Atkinson, P. M. and Tatnall, A. R. L., (1997). *Neural networks in remote sensing. International Journal of Remote Sensing*, 18, 699–709.

Bauer, E. and Kohavi, R.,(1999). An empirical comparison of voting classification algorithms: bagging, boosting and variants. *Machine Learning*, 36, 105–139.

Blamire, P.A., (1994). *An investigation into the identification and classification of urban areas from remotely sensed satellite data using neural Networks*. M.Sc Thesis, Edinburgh University, UK.

Cortes, C. and Vapnik, V., (1995). Support-vector network. *Machine Learning*, 20, 273–297.

Dixon, B. and Candade, N., (2008). Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? *International Journal of Remote Sensing*, 29,1185-1206.

Foody, G.M., (1995). Using prior knowledge in artificial neural-network classification with a minimal training set. *International Journal of Remote Sensing*, 16, 301-312.

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

Friedl, M.A. and Brodley, C.E., (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61, 399–409.

Hansen, L. and Salamon, P., (1990). Neural network ensembles. *IEEE. Transactions on Pattern Analysis and Machine Intelligence*, 12, 993–1001

Hansen, M., Dubayah, R. and DeFries, R., (1996). Classification trees: an. alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17, 1075–1081

Huang, C., Davis, L. S. and Townshed, J.R.G., (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23, 725–749.

Kavzoglu, T. and Mather, P.M., (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24, 4907-4938.

Kavzoglu, T. and Reis, S., (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. *GIScience & Remote Sensing*, 45, 330-342.

Kavzoğlu, T. and Çetin, M., (2005). Gebze bölgesindeki sanayileşmenin. zamansal gelişiminin ve çevresel etkilerinin uydu görüntüleri ile incelenmesi. *TMMOB Harita ve Kadastro Mühendisleri Odası, 10. Türkiye Harita Bilimsel ve Teknik Kurultayı, Ankara*

Kavzoglu, T., (2001). *An investigation of the design and use of feed-forwad. artificial neural Networks in the classification of remotely sensed images.* P.hD. Thesis, Nottingham University, UK

Kechman, V., (2006). Support vector machines for pattern classification, *S. Abe SIAM Review*, 48, 418-421

Lippman, R. P., (1987.)*An introduction to computing with neural nets*. IEEE ASSP Magazine, 4, 2–22.

Lu, D. and Weng, Q., (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28, 823–870. .

Mathur, A. and Foody, G.M., (2008). Multiclass and binary S.V.M. classification: Implications for training and classification users. *IEEE Geoscience and Remote Sensing Letters 5*, 241-245.

Mathur A. and Foody, G.M., (2008). Crop classification by support vector machines with intelligently selected training data for an operational application. *International Journal of Remote Sensing,* 29, 2227–2240.

Özkan, C., (2001). *Uydu görüntü verisinin yapay sinir ağları ile sınıflandırılması*, P.hD. Thesis, İstanbul Technicial University Fen Bilimleri Enstitüsü, İstanbul.

Pal, M., (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217-222.

Pal, M. and Mather, P.M., (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment,* 86, 554-565

Pal, M. and Mather, P.M., (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26, 1007–1011.

Pal, M., (2002). *Factors influencing the accuracy of remote sensing* classifications: a comparative study. PhD. Thesis, University of Nottingham.