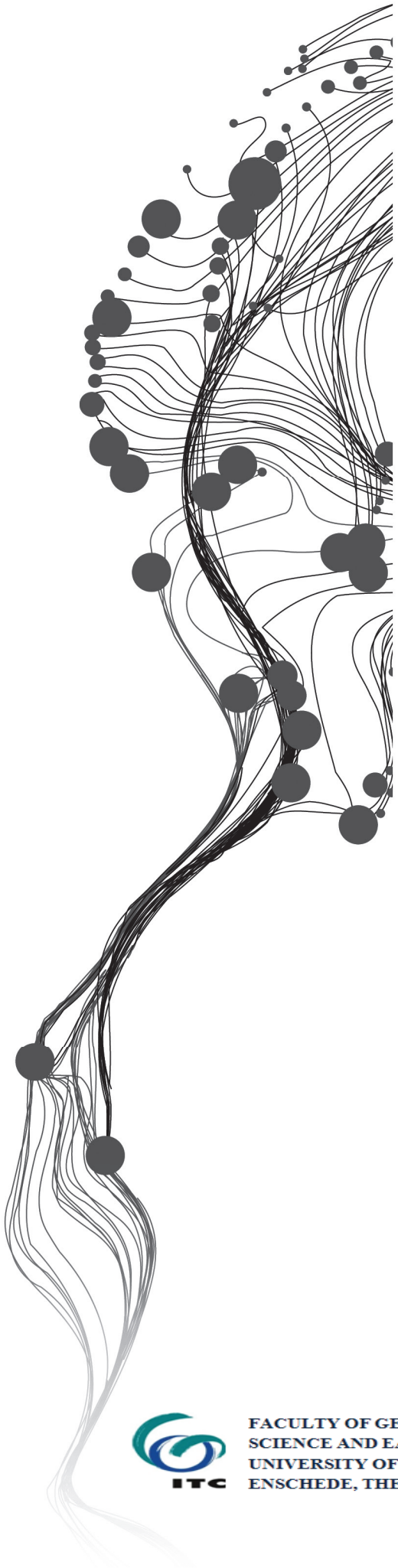
# SVM Based Dimensionality Reduction and Classification of Hyperspectral Data

PAVAN KUMAR KOLLURU
April, 2013

**IIRS SUPERVISORS**
Mr. Kamal Pandey
Dr. Hitendra Padalia

**ITC SUPERVISOR**
Prof. Dr. Ir. Alfred Stein



# SVM based Dimensionality Reduction and Classification of Hyperspectral Data

PAVAN KUMAR KOLLURU
Enschede, The Netherlands [April, 2013]

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

**THESIS ASSESSMENT BOARD:**

| | |
|---|---|
| Chairperson | : Dr. Raul Zurita-Milla |
| External Examiner | : Dr. Debashish Chaudhuri (DEAL) |
| IIRS Supervisor | : Mr. Kamal Pandey |
| IIRS Supervisor | : Dr. Hitendra Padalia |
| ITC Supervisor | : Prof. Dr. Ir. Alfred Stein |

**OBSERVERS:**

| | |
|---|---|
| IIRS Observer | : Dr. S. K. Srivastav |
| ITC Observer | : Dr. Nicholas Hamm |

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION, UNIVERSITY OF TWENTE, ENSCHEDE, THE NETHERLANDS

INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Government of India

## DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, The Netherlands. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Dedicated to "**Eshwara**" who helped and was constantly with me in the form of my Parents, Supervisors, Friends and well wishers

# ABSTRACT

The processing of hyperspectral remote sensing data, for a variety of natural resource applications, is challenging due to its higher dimensionality and non-linear characteristics. Classification techniques based on machine learning algorithms such as Support Vector Machine is preferably applied for performing classification of high dimensionality data. However, in practice, prior to applying SVM classifier, performing dimensionality reduction on hyperspectral data is a conventional step. There is a requirement of a single-step unified framework, which can decide the intrinsic dimensionality of data and achieve higher classification accuracy through SVM.

This research work contemplates on developing a unified framework for dimensionality reduction and classification of hyperspectral remote sensing image using Support Vector Machine (SVMDRC). The study also evaluates the influence of dimensionality reduction on the feature separability. A comparative analysis of the classification accuracies using the two methods viz., SVMDRC and SVM completes the scope of the study.

There are four classes in the study area namely alunite, kaolinite, illite and limestone mineral mines which were to be classified. Separability analysis was applied by using Jeffries-Matusita (JM) distance method, where, it was shown that dimensionality reduction does not influence the feature separability. Intrinsic dimensionality is calculated using modified broken stick method. The accuracy of the hyperspectral image classified by the framework has shown better results than the image classified using SVM alone. The accuracy of classification of SVM classified image was 64.70% (k=0.4361) whereas, the accuracy of the SVMDRC classified image was 82.35% (k=0.7197). The results thus indicate that SVM takes care of dimensionality to a limited degree. The complete framework is a single-step process written in an open-source language R.

KEYWORDS: Unified Framework (SVMDRC), Dimensionality Reduction, SVM classification, separability analysis, intrinsic dimensionality.

# ACKNOWLEDGEMENTS

First and foremost I would thank Mr. Kamal Pandey, my IIRS supervisor for his support and enthusiasm towards my research work. His expertise has guided me in writing the code and evaluating the algorithm written in this research work. His fervor towards my research work from the stage of the proposal was very valuable. His comments, recommendations were really helpful in completing the research work and the thesis.

I extend my gratitude towards Dr. Hitendra Padalia, my co-supervisor from IIRS for his encouragement, support, and invaluable guidance. As an application and environmental scientist, he was very helpful in making me understand the science aspects of hyperspectral remote sensing and tune my mind constantly towards science part of my research work. I would like to appreciate his patience for making me understand the concepts.

I would like to thank and express my sincere gratitude to my ITC supervisor Prof. Dr. Ir. Alfred Stein, for his invaluable guidance and encouragement throughout the research work. His critical comments, suggestions and recommendations were very useful for shaping up this thesis. Thank you very much Prof. Stein for your encouraging support and having belief in me.

I would like to thank specially my course coordinator Dr. S.K. Srivastav, for his support. His comments and suggestions on my research work form the stage of proposal to the thesis are treasured. His encouragement and belief in me is unforgettable. I am also thankful to Shri P.L.N. Raju for providing all necessary support during the research work.

I would give my special thanks to Dr. Nicholas Hamm, for his constant support in both the fore and background, throughout my research work. He also made our ITC trip the most memorable by taking care of each and every aspect from the day we landed and took off from Enschede. I thank Mr. Wim H Bakker for his critical comments and support in the initial stages of formulation of research work. I thank Dr. V.A. Tolpekin, for helping me to solve problems regarding R. I thank Dr. C.A. Hecker, and Dr. F.J.A Van Ruitenbeek, for providing me with the datasets and essential field data and make me understand them.

I thank Dr. Y.V.N. Krishna Murthy, Director IIRS and extend my sincere gratitude towards him for his support and encouragement throughout my research period at IIRS. I am also grateful to Dr. P. S. Roy, Former Director IIRS for giving me an opportunity to undergo M.Sc. course at IIRS.

I would also like to thank all my dearest friends and M.Sc classmates, Ankur Singh, Jayson Jariwala, Deepak Choudary, Hemanth, Sai Bharadwaj, Shankaracharya, Pavan Vijappu, Anukesh, Mrinal, Ravi, Chetan, Bhavya, Sharath, Dipima Sarma, all my PGD friends especially Sangeeta Sarma, Priyanka Arya, Sindhuja Chaudary, all my M.Tech. friends specially Anudeep, Suman, Rajtantra for their encouragement and giving me a wonderful time at IIRS. I specially thank all my M.Sc. batchmates who gave me strong and constant support when I was in very critical and crucial period of my MSc and also giving me a memorable time in IIRS and at ITC. I would never forget their help, if so, achieving my master's degree would have no ethical value.

I want to thank my lovely parents, my brother Praveen Kumar Kolluru, my best friends Sai Krishna GV and Abhinaya Katta for their unconditional love and support during my studies at IIRS and ITC. Without them I would not have gone this far.

-Pavan Kumar Kolluru

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# 1. INTRODUCTION

## 1.1. Hyperspectral Remote Sensing

Hyperspectral remote sensing is a fast growing technology in the field of remote sensing. In the past few years, many advances in hyperspectral technology have taken place. Hyperspectral remote sensing increases the perception and knowledge of the earth's surface (Muhammad et al., 2012). It is a passive type of remote sensing technology. Hyperspectral remote sensing, also known as imaging spectroscopy, is a study and measure of spectra obtained by reflection of the electromagnetic radiation (light) from a target. Hyperspectral remote sensing combines imaging and spectroscopy into a single system, which results in large data sets. Hyperspectral imagery is typically as a data cube with spatial information collected in the X-Y plane, and spectral information in the Z-direction (Fig 1.1). Each pixel in the image represents the spectral signature of the material imaged (Burgers et al., 2009).

A hyperspectral image is a set of contiguous co-registered spectral bands. In this sense, it is different from multispectral images that have discrete broad spectral bands. The bands in the hyperspectral images are very narrow, mostly in the range of 5μm - 20μm, depending upon the imaging sensor. They range from ultraviolet to thermal infrared regions (Muhammad et al., 2012). Because of these narrow bands, hyperspectral images have a much higher spectral resolutions as compared to multispectral images. With these high spectral resolutions the chances of uncovering subtle objects by the hyperspectral sensors are superior to multispectral remote sensing, thus leading to better discrimination and identification of the target. This in turn provides a higher potential for deriving information from the area imaged than the conventional multispectral imaging systems. In many fields of study hyperspectral images are applied. They capture the spatial and spectral information of the target. Usage of hyperspectral images are mostly in the fields of geologic purposes, atmospheric analysis, land cover analysis, forestry analysis, agricultural mapping applications and surveillance applications like military and land mine mapping.

For classification of hyperspectral image, large number of bands could aid the classification of features, as the information available in them is large compared to other type of remote sensing data sets. However, at some point adding large number of bands to the classifier will deteriorate accuracy of classification unless addition of more number of training samples takes place. This is "Hughes phenomenon". With insufficient training sets, the estimation of statistical parameters decreases hence, large amount of training samples are required. This requirement of increased number of training sets, as the dimensions increases is referred as "curse of dimensionality".

Processing of hyperspectral data cube leads to lofty computational costs as the number of dimensions in the image acquired is high (150 – 300 bands). Data redundancy is a challenge in classification of a hyperspectral image (Burgers et al., 2009). Hyperspectral images can mapped therefore into less number of dimensions because a small portion of data can explain most of the variance of the hyperspectral image, while the original features of the data are preserved. Such a process is known as dimensionality reduction (Burgers et al., 2009).

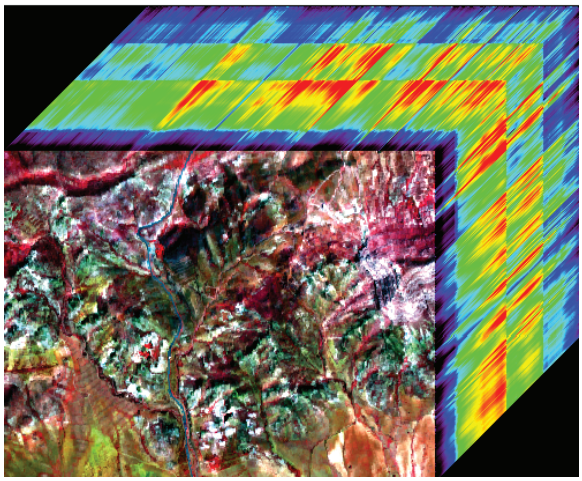


Figure 1:1. Hyperspectral data cube of HyMap with 126 bands

Analysing hyperspectral data is not an easy task. Some of the most important factors that make it complex are atmospheric distortions, land cover class spectral signature variability and curse of dimensionality. It has several challenges like:

1. Data storage is a big challenge as the volume and the size of the images are huge compared to other remote sensing datasets.
2. Due to such immense data volumes, the processing of the data is a firm task.
3. Hughes phenomenon and curse of dimensionality.

There are studies describing dimensionality reduction as an important pre-processing step for high dimensional data (Duda et al., 2009; Hastie et al., 2009). Dimensionality reduction is to be carried out in order to handle the so-called curse of dimensionality (Bellman, 2001). Due to high dimensionality, data becomes extremely sparse. Hence, reduction of dimensions can be an effective by removing the irrelevant, redundant and noisy features. Dimensionality reduction can be separated into supervised and unsupervised approaches ( Fukunga, 1990; Fisher, 2009). Generally, a supervised approach is superior to unsupervised one.

There are two types of feature reduction methods for remote sensing data viz., feature extraction and feature selection methods. In feature extraction method, original dataset is transferred into a smaller dataset by transforming the image into a new space, also called as 'dimensionality reduction'. Feature selection methods identify a subset that maintains the information which is useful to separate the classes with highly correlated and redundant features of the original image which were excluded during classification analysis (Pal and Foody, 2010)

Processing of hyperspectral data involves the following steps:

- Conversion of digital number to radiance.
- Atmospheric correction.
- Dimensionality reduction.
- Pure end-members selection though pixel purity index.
- Classifier training using the selected end-members for performing classification.

The commonly used classification algorithms for hyperspectral data classification are spectral angle mapper (SAM) and the spectral feature fitting (SFF). In these methods, classification is a non-iterative process. Therefore, optimization of classification accuracy from misclassified pixels is not

taken care of (Soman, 2009). To overcome the drawbacks of SAM and SFF, an iterative process based classification algorithm, support vector machine (SVM) is used.

## 1.2. Research Identification

Previous studies have shown that dimensionality reduction increases the classification accuracy (Burgers et al., 2009; Fong, 2007). So far, there are no clear guidelines for selecting dimensionality reduction procedures for using along with SVM classifier. In real-time most of the high dimensional datasets, do not follow normal distribution. Hence, there is a need to treat such high dimensional data with a nonlinear SVM based dimensionality reduction procedures for improving feature extraction and classification accuracy.

**Motivation and Problem Statement**

In recent studies, SVM as a dimensionality reducer and classifier has been used for non-spatial dataset (Yang, 2009) that have lesser dimensions compared to the hyperspectral images. Hyperspectral images are nonlinear and are of high dimension. Motivated by this an algorithm is introduced that classifies an image along with dimensionality reduction, using SVM. This algorithm implements dimensionality reduction and classification in a unified framework. Training samples from selected features are included into the algorithm so that it iteratively performs the assigned task.

## 1.3. Research Objectives

The main goal of the research is to develop a modified SVM based dimensionality reduction and classification algorithm in a unified framework for hyperspectral datasets and to evaluate its performance.

**Sub-objectives**

- To develop a modified SVM based algorithm for dimensionality reduction and classification in hyperspectral datasets by introducing a nonlinear function.
- To find the influence of the dimensionality reduction on the feature extraction.
- To compare the classification accuracies derived from proposed approach vis-à-vis conventional approach of support vector machine (SVM) classification.

**Research Questions**

- How effective is the application of nonlinear function for dimensionality reduction?
- Does a dimensionality reduction technique show any impact on extraction of different features types from hyperspectral data?
- How does SVM based dimensionality reduction and classification algorithm perform as compared to SVM classification on the hyperspectral data?

## 1.4. Study Area and Dataset

The study area of this research work is the Los Tollos area that is a part of the Rodalquilar district in the Sierra del Cabode Gata, in south-eastern Spain. The area has volcanic rocks of different compositions form pyroxene-bearing andesites to rhyolites. The intense alteration of rocks is due to two reasons viz., volcanic geothermal activity known as hypogene alteration and chemical weathering also known as supergene alteration. Because of volcanic activity and alterations, there are deposits of different minerals. The most interesting mineral mines in this area are Gold deposits. This area is the first documented example of caldera-related gold deposit mineralization in Europe (Arribas et al., 1995). There are five distinguished hydrothermal alteration zones in this area classified by Arribas et al., (1995) viz., silicic, advanced argillic, intermediate argillic, sericitic, and propylitic zones (table 1.1). In addition to hypogene advanced argillic alteration, supergene advance argillic alteration, also known as stage 2 alunite (Arribas et al., 1995), is present in the area, which is the interest of this research work. Large-scale mining of alunite has taken place in the area. A generalized geologic map is shown in figure 1.2.

The Los Tollos, Rodalquilar, Spain was selected as study area of the research thesis work because of availability of ground data and area remaining relatively undisturbed by previous mining activities. Gold (Au) mining has been abandoned and restricted in this area as so far 10 tonnes of gold has been mined (Bedini, 2005). The sparse vegetation cover in the area will allow better surface reflectance from the mineral deposits, however non-photosynthetic vegetation exists in few areas, which have low to moderate effect on the spectra of the surface reflection. Mineral mine area with similar climatic conditions (semi-arid) like that of Los Tollos Rodalquilar, also occur in Rajasthan and Selam (Tamil Nadu) regions of India. The problem for not selecting these areas is non-availability of airborne hyperspectral data. The availability of hyperspectral data for Indian region is only through space-borne Hyperion sensor, which is selectively available and the spatial resolution is coarse (30m). Atmospheric correction of Hyperion data also has challenges. The methodology developed in the Rodalquilar area will be useful in future studies in India once the better resolution, atmospherically corrected hyperspectral datasets become available.

Table 1-1: Summary of alteration zones and dominant minerals in the Rodalquilar area (Arribas et al, 1995)

| Alteration Zone | Alteration Minerals |
|---|---|
| Silicic | Quartz; Chalcedony; Opal |
| Advanced argillic | Quartz; Alunite; Kaolinite; Pyrophyllite; Illite; Illite - Smectite |
| Intermediate argillic | Quartz; Kaolinite; Illite; Illite - Smectite |
| Sericitic | Quartz; Illite |
| Propylitic | Quartz; Illite; Montmorillonite |
| Stage 2 Alunite | Alunite; Kaolinite; Jarosite |

A HyMap hyperspectral image shown in figure 1.3, is used in this research work. The image obtained from the HyMap sensor, has 126 contiguous spectral bands, covering 0.45 – 2.5μm of electromagnetic spectrum at spectral resolution between 15 – 20nm. Spectral coverage is nearly continuous in the SWIR and VNIR regions with small gaps in the middle at atmospheric water absorption bands (1.4 and 1.9μm) (table 1.2). The HyMap image of the area is a sub-scene of

285X375 pixels, covering the Los Tollos area. This subset is considered because the area is mostly covered with the four minerals alunite, illite, kaolinite and lime stone cover.

HyMap is an airborne hyperspectral imaging system operated by HyVista Corporation and owned by Integrated Spectronics, Sydney, Australia. It is flown at an altitude of 2.5km on a fixed wing aircraft. The study area was imaged on 11.07.2003 in 126 narrow bands, from 0.45 to 2.48μm with a pixel size of 5m. The subset image shown in the figure 1.3 is used in this thesis. One problem with the data set is that the SWIR-1 data is not available as there were technical complications at the time of imaging the area, hence data form that particular part of spectrum is missing.

Table 1-2. HyMap Instrument details (Cocks et al, 1998)

| Spectrum | Wavelength Range(μm) | Bandwidth(nm) | Spectral Sampling(nm) |
|---|---|---|---|
| VIS | 0.45-0.89 | 15-16 | 15 |
| NIR | 0.89-1.35 | 15-16 | 15 |
| SWIR1 | 1.40-1.80 | 15-16 | 13 |
| SWIR2 | 1.95-2.48 | 18-20 | 17 |
| IFOV | 2.5m along track<br>2.0m across track | | |
| FOV | 60°(512 pixels) | | |
| Swath | 2.3km at 5m IFOV<br>4.6km at 10m IFOV | | |

The HyMap scene was atmospherically corrected by using parametric geocoding procedure (PRAGE), Airborne Atmospheric, and Topographic Correction Model (ATCOR4) software by German Aerospace Centre. Where the scanning geometry of the image has been reconstructed by using PRAGE with the aid of the pixel positions, altitude and terrain elevation data (Schlapfer and Ritcher, 2002).

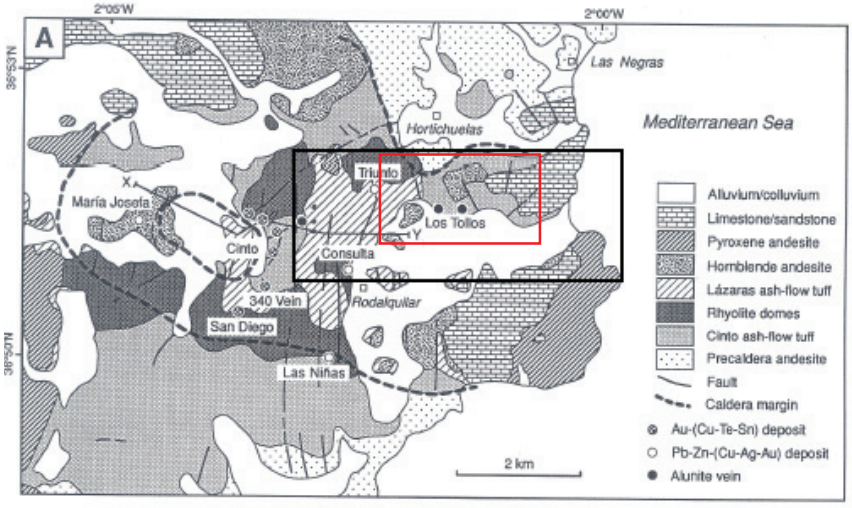



Figure 1:2. Generalized geologic map of Rodalquilar and outline of the HyMap image (after Arribas et al, 1995)and the subset image(Los Tollos in red box) . Image Courtesy (Bedini , 2005)

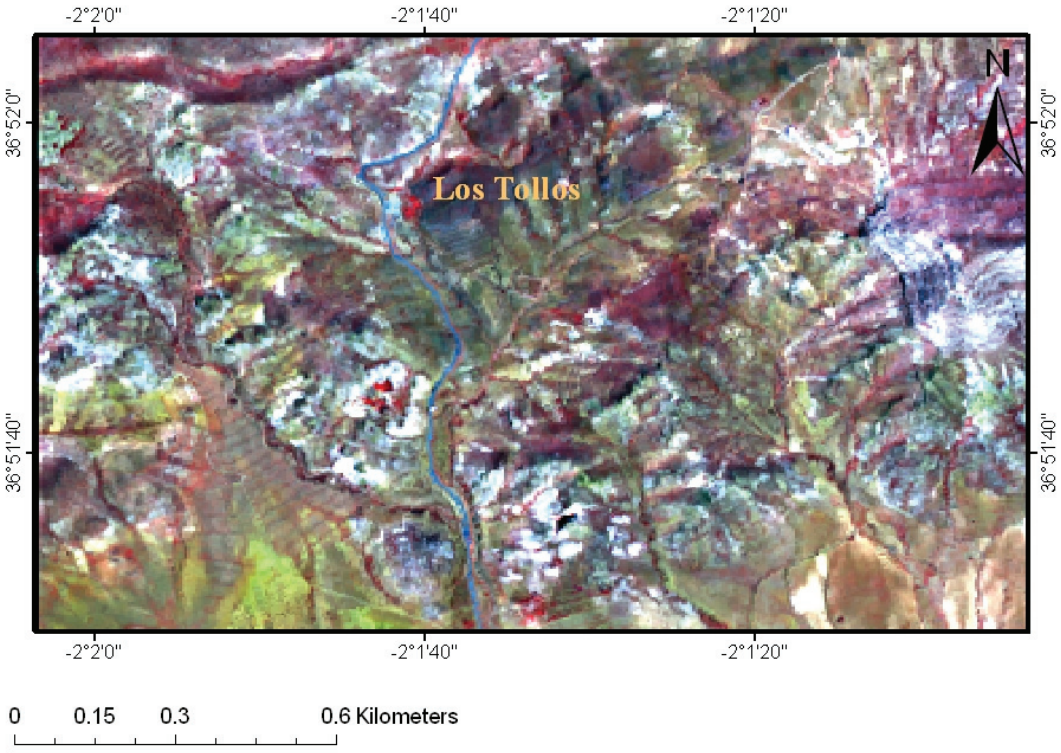



Figure 1:3. HyMap image of Study area Los Tollos, Rodalquilar in FCC (R:22,G:17,B:4)

**Validation Data**

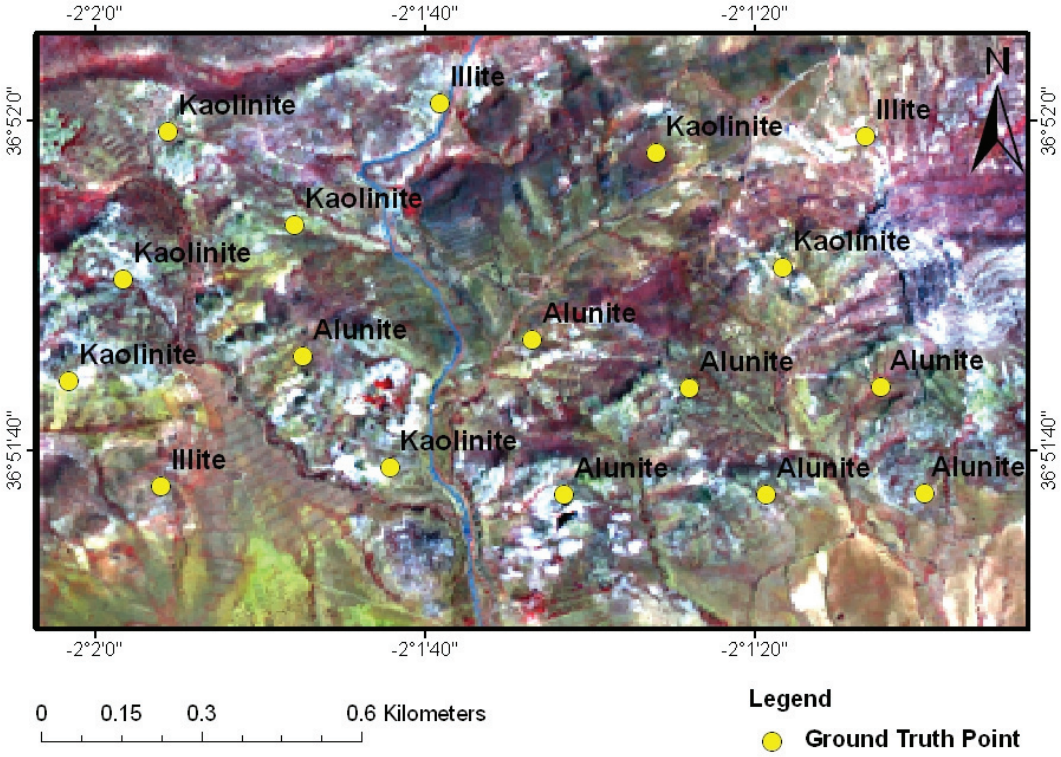



Figure 1:4. HyMap Image of the study area showing the positions of validation points

Collection of field spectra from some parts of the study area (shown in figure 1.4) was performed during the over-flight using the Analytical Spectral Device (ASD) fieldspec-pro spectrometer. This spectrometer covers the 0.35–2.50µm wavelength range with a spectral resolution

of 3nm at 0.7μm and 10nm at 1.4μm and 2.1μm. The spectral sampling interval is 1.4nm in the 0.35–1.05μm wavelength range and 2nm in the 1.0–2.5μm wavelength range.

As shown in table 1.3, 17 validation points were available in the study area. Of the available 17 validating points, 7 points are of alunite, 7 for kaolinite and 3 points for illite mineral.
There were no points available for limestone.

Table 1-3. Reference data from the ground

| Station | X | Y | Determinant | Secondary |
|---|---|---|---|---|
| LT04-25 | -2.019393 | 36.860440 | Alunite | |
| LT04-15 | -2.022093 | 36.860410 | Alunite | |
| LT04-11 | -2.025467 | 36.860381 | Alunite | |
| LT04-12 | -2.028400 | 36.860880 | Kaolinite | |
| LT04-04 | -2.032273 | 36.860528 | - | Illite |
| LT04-10 | -2.020156 | 36.862229 | Alunite | |
| LT04-20 | -2.023354 | 36.862200 | Alunite | |
| LT04-6 | -2.026024 | 36.862992 | Alunite | |
| LT04-3 | -2.029867 | 36.862699 | Alunite | |
| LT04-14 | -2.021799 | 36.864195 | Kaolinite | |
| LT04-17 | -2.030014 | 36.864958 | Kaolinite | |
| LT04-23 | -2.032860 | 36.864019 | Kaolinite | |
| LT04-7 | -2.020391 | 36.866425 | Illite | |
| LT04-1 | -2.023941 | 36.866161 | Kaolinite | |
| LT04-24 | -2.027579 | 36.866982 | Illite | Kaolinite |
| LT04-9 | -2.032126 | 36.866483 | Kaolinite | |

# 2. LITERATURE REVIEW

## 2.1. Feature Extraction and Dimensionality Reduction

Feature selection for classification of hyperspectral data by using Support vector machine has been performed by Pal and Foody (2010). The study had principally focussed on feature selection method by using SVM on the hyperspectral datasets. An attempt was made to addresses the key aspect of uncertainty over the sensitivity of the SVM and accuracy of classification of dataset to the dimensionality of the dataset. Four main feature selection algorithms have been used for analysis viz., Recursive Feature Elimination (SVM-RFE), Correlation based feature selection (CFS), Minimum redundancy – Maximum Relevance (mRMR*)* and Random forest (RF). It was noticed that the accuracy of the SVM classification varied as a function of the number of features used and the size of the training set used. As the number of features were increased the accuracy of the SVM classification also increased. When a fixed size of training set were used the accuracy had initially rose when features to the peak were added but thereafter decreased with the addition of more features. However, the decrease in the accuracy was significant statistically. When small training sets were used, the curse of dimensionality reduction and the Hughes effect were observed with SVM classification. Finally, a conclusion was made that when larger training sets are used mostly the effect of the Hughes phenomenon could be reverted. Also as the features increases accuracy of the classification will be reduced. These points are useful and help in the research for selecting the training sets.

Burgers et al., (2009) have performed a comparative analysis of dimensionality reduction techniques aiming to evaluate the performance of the dimensionality reduction algorithms. Eight different algorithms viz., Principal Component Analysis, Kernel Principal Component Analysis, Isomap, Diffusion maps, Laplacian Eigen maps, Independent Component Analysis, LMVU and LTSA have been evaluated for their performance on the dimensionality reduction and determination of the intrinsic dimensionality of the hyperspectral images. Nonlinear methods had given comparably better results but had a major setback of taking very long runtimes. Thus increasing the cost of the processes run. When the high dimensional data sets were used, their runtimes was very high compared to linear methods resulting in increase of the computational cost. Different hyperspectral data sets were used in the experiment and the performance evaluation has been done both on the classification accuracy and on the runtime of the algorithm. In this process, PCA was observed to be fastest in running and gave the most accurate results. But the dimensionality reduction algorithm performance depends on the image. After investigation has been performed, of all the odds PCA had outperformed and has been proved as the best dimensionality reduction technique giving the best results when performed. KPCA works best with the images, which have multiple edges, but PCA and ICA had performed comparably on the images without many edges. Target detection was comfortably performed by PCA, KPCA and ICA. PCA had the least error rates in the processes and outperformed in all the tasks compared to other methods.

A similar kind of comparative work has also been performed by Fong (2007), where, different dimensionality reduction techniques like Principal Component Analysis, Fast ICA (Independent component Analysis), Laplacian Eigenmaps, Local Linear Embedding (LLE), Local Tangent Space Analysis (LTSA), Linear Local Tangent Space Analysis (LLTSA) and diffusion maps are compared for their performances. According to Fong, Laplacian Eigenmaps LLE and LTSA are

local nonlinear techniques. They preserve the properties of small neighbourhoods around the data. LLTSA uses a linear technique to minimize the cost function of LTSA. The major disadvantage of these methods is that they are incapable to handle the images larger than 70X70 pixels. Kernel PCA (KPCA) is a nonlinear version of PCA having a disadvantage of drastically increasing the computation time and process as the size and dimensions the image increases. It gives a poor performance on the hyperspectral images as the size of the images and the dimensions are large.

Hyperspectral data dimensionality reduction and end member extraction has been performed by Muhammad et al., (2012). To present an algorithm for overcoming the computational complexities of hyperspectral data to detect the multiple targets and end members effectively with less computational time was the main aim. Standard deviation and chi square distance metrics methods are considered. The end member estimation was done by unbiased iterative correlation method.

Dimensionality reduction using sparse Support machines was performed on the hyperspectral datasets by Bi et al., (2003). A method for performing variable ranking and selection using support vector machines (SVM), by constructing a series of sparse linear SVM's to generate linear models that could be generalized was described. A subset of nonzero weighted variables was used, found by the linear models to find a final non-linear model. In addition, it claims the work exploits a fact that a linear SVM with $l_1$–norm regularization inherently performs variable selection as a side effect of SVM model capacity minimization. The method is known as variable selection via sparse support vector machine (VS-SSVM). It consists of two parts: variable selection and non-linear induction where first part serves as a pre-processing step for the final SVR kernel induction. The VS-SSVM has five components they are a linear model with sparse **w**, an efficient search using "pattern search" for optimal hyper parameter C and v in linear SVM, usage of bagging for reduce the variability of the variable selection, a method for discarding the least significant variables by comparing t with the random variables and a nonlinear model obtained by training the LP's with RBF kernel on the final subsets. It concluded that VS-SSVM is effective on the specific problem and the number of the variables was reduced while maintaining the generalization ability. It is not a general method suitable for all types of problems. Demonstration for its effectiveness on very high dimensionality problems with little data was performed; it was proved that where the linear models cannot capture the relationships the method would fail.

## 2.2. Classification

Classification of hyperspectral images using SVM has been performed by Melgani and Bruzzone, (2004). A brief discussion was made on SVM and its application to hyperspectral Images. SVM is a binary classification method, classifying only two classes. For multi class classification, this can be overcome by using certain strategies viz., parallel architectural approach and hierarchical tree bases architectural approach. Hierarchical approach is further divided into two type's viz., Balanced Branches strategy and one against all strategy. Two experiments were performed such as classification in the original hyper dimensional feature space and Feature reduction and classification. The two major aims were one with assessment of SVM hyper dimensional space properties and the second is assessment of the effectiveness of strategies based on ensembles of binary SVM's used to solve multiclass problems in hyperspectral data. The conclusions of the work were

- SVM is best classifier compared to the other nonparametric classifiers in terms of the classification accuracy and computational cost
- SVM is more effective than the traditional pattern recognition approaches
- SVM exhibit low sensitivity to the Hughes phenomenon.

Four different multiclass strategies were considered from which each other differ from the manner in which the complexity of the multiclass classification is distributed over the single members (SVM's) of the architecture. The parallel architectures showed better results than the hierarchical architecture. This is because the hierarchical approaches propagate the error to the next levels, because the final result is the combination of several hierarchal approaches and the error gets accumulated at the last level which gives the results. In terms of computational time hierarchical approaches were faster compared with the parallel approaches. So, depending on the application the multiclass strategies must be selected keeping the trade off in mind. Finally it was showed that the multiclass problem does not significantly affect the analysis of the hyperspectral data. All the SVM approaches showed the better results than the non-parametric classification approaches.

A unified framework for generalized linear discriminant analysis was developed by Ji and Ye, (2008). The work proposes a unified framework for generalized LDA through a transfer function. Linear discriminate analysis is a classical statistical approach for dimensionality reduction. It computes a projection by minimizing the in class distance and maximizing the between class distance simultaneously, thus achieving the maximum class discrimination. However, LDA has a major drawback of having the total scatter matrix used in the discrimination to be a non-singular matrix. But generally the matrix is a singular matrix for high dimensional data. This is known as singularity problem. However, a systematic study has not been implemented to know the common features in the algorithms and their intrinsic relationships. The proposed framework is basically a four step algorithm which computes a series of Eigen values and Eigen vectors and achieves an orthogonalization. This framework elucidates the properties and functionalities of different algorithms.

Hyperspectral image classification by performing dimensionality reduction was performed by Harsanyi and Chang (1994). A method, which performs dimensionality reduction and detecting signatures of interest from the hyperspectral images, was development and demonstrated. It is a combination of two linear operators, optimal rejection interference process and optimal detector in the maximum SNR sense, into a single classification operator. This approach could be applied to the images with both mixed pixels and spectrally pure pixels. Representative signatures of interest could be detected by this method, which could be as low as few percent of the SNR having a spectral resolution of less than 10nm. The performance could be varied with the varying datasets but this could be used for analyzing the sensor capabilities for solving a classification and detection problem. This method produces component images, which represent the class maps of various materials within the scene, which were almost comparable to the geological maps.

Reduction of the dimensionality of hyperspectral data for the classification of agricultural scenes was performed Silva et al., (2008). Usage of genetic algorithms (tournament and elitism) for yielding better classification accuracies and its feasible for hyperspectral images was established. A comparative study of the sequential and genetic algorithms with the same datasets having different bands and groups have been performed. The results showed that by performing the genetic

algorithms the accuracy and the kappa indices have been increased drastically. Genetic algorithms have outperformed in this case. The best results have been given by the elitism genetic algorithms, which have given the kappa values of around 0.9218 which is a very good value.

A review of support vector machines (SVM) in remote sensing was given by Mountrakis et al., (2011). A discussion was made on SVM and its importance in remote sensing. SVM has an ability generalize the data with very less training samples and give better accuracies compared to other training methods. As SVM is non-parametric, for classification it does not assume a statistical distribution. It needs and always sticks to global minima as it deals with quadratic problems. As the remote sensing data have unknown distributions this property of SVM is very much useful allowing to outperform than the other type for classification techniques. The main limitation of SVM is that the selection of SVM key parameters and the kernel function to be used. An optimal value must be chosen so that over fitting and over smoothing, might be avoided, which is usually made manually. This drawback holds good for all the methods involving kernels and hence holds good even for SVM. It was claimed that the one-against all type of strategy for multiclass classification is a problematic issue and needs some serious attention, leading to unclassified instances of data. In SVM kernel mapping is more vulnerable to dimensionality reduction, as the dimensions are high for the hyperspectral data. Mostly SVM's are not made to deal with the noise component hence leading to outliers in the data. As, the training and validation sets used in SVM are smaller compared to those used for other machine learning algorithms the quality of them must be maintained. The performance of SVM could be abridged if the data have any mislabels. The work conclude by saying that SVM's self-adaptability, swift learning pace and limited training size has become a reliable intelligent data processing technique in the field of remote sensing.

Hyperion hyperspectral image analysis combined with machine learning classifiers have been performed by Petropoulos et al., (2012). A comparison between SVM and artificial neural network (ANN) classification was performed. The results obtained by this work are as follows. Both the methods have produced comparable results in terms of spatial distribution and cover density of each land cover category. The work has also highlighted the important point of SVM that it has been designed to identify the optimal hyperplane for class separation with the least error among all the separating hyperplanes, which the other classifiers cannot. This produces the accurate classes at the end of classification addressing all ill-posted problems providing high classification accuracies even when the small training data sets were used. A similar pattern has been shown by both the methods as per the single class accuracies are concerned. However, as a whole, SVM has outperformed than ANN in their method as per the accuracy assessment reports.

Comparison of methods for multi class SVM has been performed by Hsu and Lin, (2002). One of the authors is involved in the development of lib-SVM, which is most used algorithm for SVM applications. SVM is developed for binary classification and can be extended to multi class classification. There are approaches like one-many type of classification approach where the user can use SVM for multiclass classification. This field is still under development and new techniques evolve as the time lapses. However, as of now this topic is not yet stabilized. The main aim of considering this paper in my literature is that even the research involved in this thesis is also involved with multi classes and uses SVM for it.

Multiclass approach for SVM classification was performed by Pal, (2008) describing different types of multi class techniques and comparison of the results obtained by them. Six multi class approaches viz., one vs. one, one vs. rest, Directed Acyclic Graph (DAG) and Error Corrected

Output Coding (ECOC) based multiclass approaches were compared. All the approaches were created from the binary classifier. Classification of the image was done by using all the above techniques and the kappa values were calculated. All the methods have given considerable results except ECOC (dense coding approach). It gave the least accuracy compared to all the other types of approaches. The highest accuracy was for ECOC (exhaustive approach). One against rest approach has a problem, that its produces unclassified data which leads to lower accuracies. One against one approach is the best approach for multiclass classification using SVM. A similar kind of approach was used in this research work.

## 2.3. Other Related Work

A Recursive support vector machine (RSVM) for dimensionality reduction was discussed by Tao et al., (2008). A multidimensional maximum margin feature extraction approach was discussed extensively which is used for constructing an orthogonal based dimensionality reduction. The analysis shows that as the number of recursive components increases the objective function of the SVM decreases. RSVM shows better accuracy than the regular SVM and linear discriminant analysis (LDA) and have no singularity problems. The analysis was carried on standard benchmark non-spatial data sets. The main idea of considering this literature is to test the same on the spatial high dimensional hyperspectral dataset.

Sukens, (2001) have focused on SVM for classification and nonlinear function estimation and on least square SVM, which involves in the solution of the linear systems and nonlinear function estimation problems. Standard SVM's are used for classification, regression etc that are standard static problems but the LS-SVM are developed for even more recurrent and optimal control problems. They also have good computational advantages. The disadvantages of using them are the cost function involved has lack of sparseness in the solution vector and Gaussian assumptions. Infinite number of weights can be possessed by LS-SVM systems as they are characterized by KKT systems in a primal weight space. By these views of the author on SVM, involvement of SVM is done in this research work to check the potential usage of it for hyperspectral imagery.

SVM has been used as a tool for mapping mineral prospectively by Zuo and Carranza, (2011). The work proved that SVM is the best geo-computational tool for spatial analysis. SVM was subjected to multiple variables for mineral prospective mapping. SVM algorithm with different kernel functions was tested with the mineral area. The results obtained were satisfactory and indicated that it is a useful tool for integrating multiple evidence layers in mineral prosperity mapping. These results encouraged the usage of SVM in this research work as the study area is occupied with different mineral mines.

The above literature review is extractions of the essence of individual works done by different authors. However, the most common points in them are discussed. SVM is a binary type of classifier and most powerful to the present. It is a machine learning technique, which makes the model trained with minimum number of training samples rather than large number required by the other type of classification methods. Apart from many advantages, there are also disadvantages that it is a kernel-based type of classifier hence the parameter selection is an issue and usually is done by trial and error method, the quality of the training samples must be the finest, and that cannot be possible in all the conditions and for all type of datasets. It gives the classification results as per the dataset,

which means the same method applied for different datasets will give different results. Apart from these drawbacks SVM has been chosen in this research work as it is gives best results by outperforming over the other classifiers and is also simple to execute.

# 3. METHODOLOGY

The methodology followed in this research is shown in the flowchart (Figure 3.1). The research method is divided into two parts, SVMDRC and SVMC. SVMDRC is the work carried in this research, which includes the developing of the framework and applying it on the hyperspectral image, and SVMC part is performing SVM classification on the hyperspectral image using the same training samples used by the left part of the work. Here by the words SVMDRC and SVMC will be used for the two processes explained for ease of understanding and readability. The difference between the SVMDRC and is that, in the SVMC part no dimensionality reduction is performed and in the SVMDRC part dimensionality reduction is performed along with classification. Training samples are taken from the image, which are the endmembers of the class, which has to be classified. With the training samples provided, the SVM classifier decides the hyperplane and the support vectors are generated to separate the classes of the image.

The methodology in brief is as follows. An air borne hyperspectral image is taken which has to be classified. The image is preprocessed by atmospherically and geometrically correcting it. The pre processed image id then subjected to two types of classifications one using the unified framework and the second by conventional SVM. The classification is trained by giving the training samples. The classified images are then validated. The accuracies obtained from the validation report are compared. The feature separability is evaluated by using JM distance method.

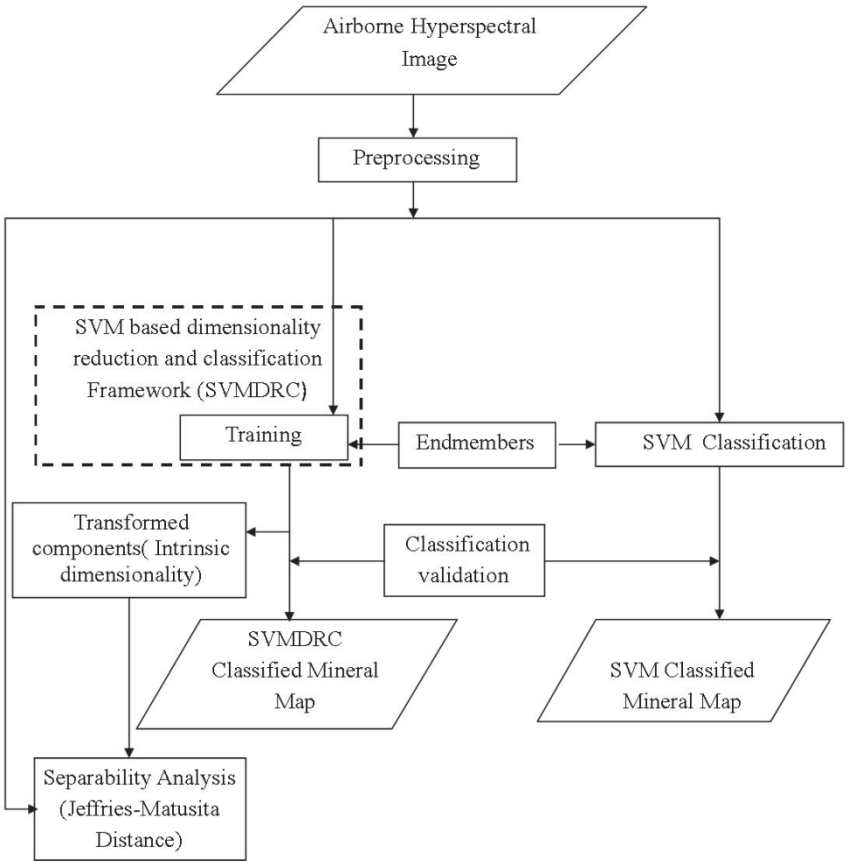



Figure 3.1. Methodology flowchart

## 3.1. Data Pre-processing and Training Dataset

The level 1 HyMap data was atmospherically corrected but not georeferenced. Atmospheric correction was performed using the Atmospheric and Topographic Correction (ATCOR 4) model at the time of receiving the data. The image is georeferenced and has been converted into geotiff format, for further processing. The image is converted to geotiff format because it is easy for performing computation.

The hyperspectral image has four classes, alunite, kaolinite, limestone and illite. For classification, the training samples are extracted from the image and with these training samples, the classification model is run to perform classification of the image. Endmembers of these four classes are collected and the training set is made. Region of interest (ROI) containing the pure endmembers are identified in the image and extracted. These roi's are further converted into tiff format for further processing.

In this research work a unified framework is developed which performs dimensionality reduction and classification in a single process. If the framework is provided with the hyperspectral image and the corresponding region of interest (roi's) of the classes to be classified, it performs dimensionality reduction and classification using support vector machine (SVM).

## 3.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) concept was introduced by Cortes and Vapnik, (1995) to solve the regression and classification problems. SVM is based on statistical learning theory and structural risk minimization. It finds an optimal hyperplane that maximizes the margin between the classes by using a small number of training samples known as support vectors (Cortes and Vapnik, 1995). As supposed by Silva et al., (2008) it has become a very popular method for image classification. SVM has a property of simultaneously minimizing the empirical classification error and maximizing the geometric margin (Yang, 2009).

Support vector machine uses kernel method to perform regression and classification by transforming the data to the higher dimensional space by nonlinear transformation techniques. It separates the two classes by finding a linear spacing between them. This linear spacing is achieved, as the data is transformed into the higher dimensions it tends to spread the data out which makes a way to find the linear spacing between the classes to get separated (Gualtieri and Cromp, 1998). Thus, the hyperplane is the greatest margin between the two classes. Figure 3.2 shows the concept of a hyperplane. Bold line shows the acceptable hyperplane which separates the data.

SVM is a supervised machine learning algorithm, where it is given a set of inputs with the corresponding labels. The inputs are in the form of attribute vectors. SVM constructs a hyperplane that separates two classes to achieve maximum separation between the classes. By separating the classes with a large margin, generalization error is minimized. The objective of achieving the minimum generalization error is to predict the correct class of the data without any error or minimal error, when it arrives for classification (Soman, 2009).

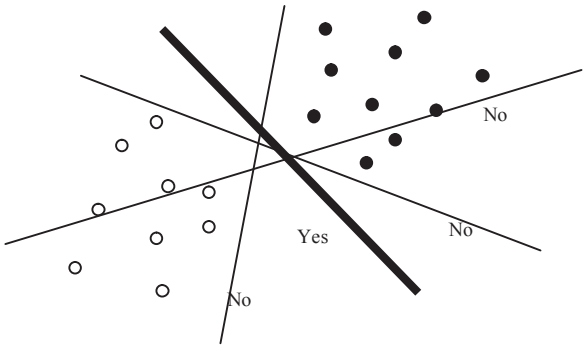



Figure 3:2. SVM Hyperplanes between two classes

The two planes parallel to the classifier and which passes through one or more points in the data are called '*bounding planes*'. The distance between these bounding planes is called '*margin*'. By the process of learning hyperplane, which maximizes this margin, is evaluated. The points of the corresponding class, which falls on the bounding planes, are called '*support vectors*'. These points are crucial in forming a hyperplane hence the name support vector machine (Soman, 2009). Figure 3.3 shows the concept of support vectors, bounding planes and maximum margin.

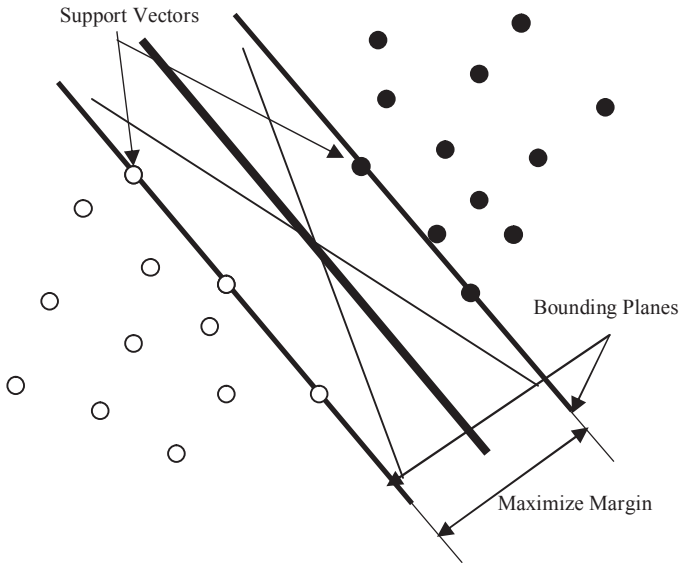



Figure 3:3. Bounding planes, Support vectors and Maximum Margin in SVM

If the optimal hyperplane separates the training vectors without any errors, the ratio of the expectation of the support vectors to the number of training vectors limits the expected error rate. A good generalization is guaranteed if a small set of support vectors is found because this ratio is independent of the dimension of the problem (Cortes and Vapnik, 1995).

In spite of taking all the required measures for classification, there are chances likely for misclassifications. SVM takes care of them, by allowing misclassifications of pixels between classes. Figure 3.4 shows two classes for classification, class A with white dots and class B with black dots. The hyperplane gives its maximum efforts in all the possible ways to classify the image with very less

misclassification. The hyperplane is a straight line in this classification. It would be much more interesting if the hyperplane is a twisted line such that it surpasses the pixels of other class and classify the distinct classes without misclassification errors. Such type of classification with twisted separating boundary is known as nonlinear SVM classification.

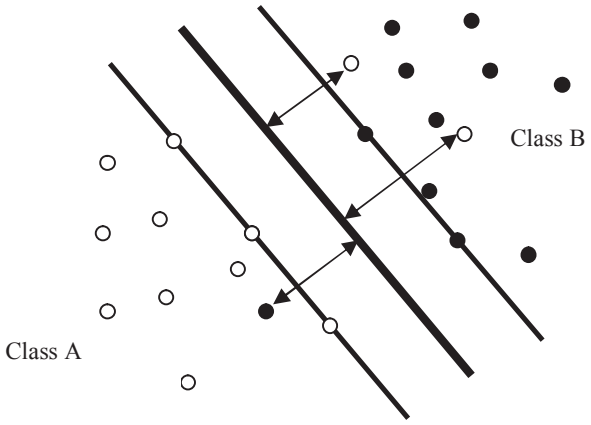



Figure 3:4. Misclassifications in SVM

Broadly, SVM is of two types. They are linear and nonlinear type of SVM. If the hyperplane in the SVM classification is linear in nature, it is known as linear SVM and if the hyperplane in the SVM is a nonlinear equation, it is known as nonlinear SVM. A non-linear SVM is achieved by using a kernel trick.

### 3.2.1. Linear SVM

If the hyperplane of the SVM is linear then such an SVM is known as linear SVM. Linear SVM is applicable for two types of data. They are separable and non-separable type of data as discussed below (Gualtieri and Cromp, 1998).

**a) For Separable Data**

Consider $l$ training pairs $(y_i, \boldsymbol{x}_i)$ where $i = 1,2, \ldots l$ having class labels $y_i \in \{1, -1\}$ and $x_i \in \boldsymbol{R}^2$.
Figure 3.3 shows two classes A and B. In this equations class A is represented as +1 and class B as -1. Main aim of the SVM classifier is to introduce a hyperplane, which separates all the points belonging to -1 on one side and +1 on the other side as shown.
The hyperplane is defined as a plane separating the classes such that the closest vector in the two classes are farthest from the plane separating them as shown in figure 3.3. It is denoted by the equation 3.1

$$\mathbf{w}.\boldsymbol{x} + b = 0 \tag{3.1}$$

Where, $\boldsymbol{x}$ is a point on the hyperplane and $b$ is the distance of the closest point on the hyperplane to origin and $\mathbf{w}$ is a two dimensional vector pointing perpendicular to the hyperplane.
The classifier for the data is represented by a function $y = f(\mathbf{x}; \boldsymbol{\alpha})$, where, $\boldsymbol{\alpha}$ is the parameter of the classifier. Hence the classifier for the hyperplane in Equation (3.1) is

$$f(\mathbf{x}; \mathbf{w}, b) = sgn\,(\mathbf{w}.\boldsymbol{x} + b) \tag{3.2}$$

Let $d_i$ be the perpendicular distance of vector $\boldsymbol{x_i}$ from any point $\boldsymbol{x}$ on the hyperplane. It is given by the equation (3.3)

$$d_i = y_i \frac{w}{|w|}.(\boldsymbol{x_i} - \boldsymbol{x}) \tag{3.3}$$

This is further simplified by placing the hyperplane equation and the distance $d_i$ as in equation (3.4)

$$d_i = y_i \frac{\mathbf{w}.\boldsymbol{x_i}+b}{|w|} \tag{3.4}$$

The distance of the hyperplane from all the vectors must be minimum and the distances over the entire hyperplane placement must be maximum. Hence, the classifier becomes

$$\max_{\mathbf{w},b} \min_{i=1,\dots l} \left[ y_i \frac{\mathbf{w}.\boldsymbol{x_i}+b}{|w|} \right] \tag{3.5}$$

If $i$ is a support vector which is nearest to the hyperplane, then

$$y_i(\mathbf{w}.\boldsymbol{x_i} + b) - 1 \ = 0$$

If $i$ it is not a support vector the value is >0.

$$y_i(\mathbf{w}.\boldsymbol{x_i} + b) - 1 \geq 0 \text{ where } i = 1,2, \dots l \tag{3.6}$$

Equation 3.5 is further simplified and the optimal hyperplane for separable data is given by equation 3.7

$$\min_{\mathbf{w},b} \frac{1}{2}|\mathbf{w}|^2 \tag{3.7}$$
$$y_i(\mathbf{w}.\boldsymbol{x_i} + b) - 1 \geq 0 \text{ where } i = 1,2, \dots l$$

To solve the hyperplane the optimization problem is solved by Lagrangian variables, where

$$\lambda_i \geq 0 \ \ and \ i = 1,2, \dots l \tag{3.8}$$

are Lagrangian multipliers.

$$\mathcal{L}(\mathbf{w}, b, \lambda_1, \dots, \lambda_l) = \frac{|\mathbf{w}|^2}{2} - \sum_{i=1}^{l} \lambda_l [y_i(\mathbf{w}.\boldsymbol{x_i} + b) - 1] \tag{3.9}$$

and the problem becomes

$$\left.\begin{array}{c} \max_{\lambda_1,\dots,\lambda_l} \min_{\mathrm{w},b} \mathcal{L}(\mathbf{w}, b, \lambda_1, \dots, \lambda_l) \\ \lambda_i \geq 0 \ \ and \ i = 1,2, \dots l \\ y_i(\mathbf{w}.\boldsymbol{x_i} + b) - 1 \geq 0 \text{ where } i = 1,2, \dots l \end{array}\right\} \tag{3.10}$$

From equation (3.10) and (3.12) the $\lambda_i$ value must be maximized because by putting the Lagrangian multipliers minimum, the Lagrangian undetermined constraints reach the equality. Thus

$$\lambda_i[(\mathbf{w}.\boldsymbol{x_i} + b) - 1] = 0 \ where \ i = 1,2, \dots l \tag{3.11}$$

Equation (3.13) is known as complimentary condition. This can be further minimized by differentiating the function with $\mathbf{w}$ and $b$ and we have the conditions as

$$\frac{d\mathcal{L}}{d\boldsymbol{w}} \tag{3.12}$$

$$\frac{d\mathcal{L}}{db} \tag{3.13}$$

The conditions in equations (3.6), (3.7), (3.11), (3.12), (3.13) are known as Karsh – Kuhn – Tucker (KKT) optimality conditions.
By solving the Lagrangian dual with the above conditions the dual problem optimization eliminating **w** and $b$ is given as equation (3.16)

$$\left.\begin{array}{c} \max_{\lambda_1,\ldots,\lambda_l}\left[-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i y_i(\boldsymbol{x_i}.\boldsymbol{x_j})y_j\lambda_j+\sum_{i=1}^{l}\lambda_i\right] \\ \lambda_i \geq 0 \ \ and\ i=1,2,\ldots l \\ \sum_{i=1}^{l}\lambda_i y_i = 0 \end{array}\right\} \quad (3.14)$$

**b) For Non Separable Data**

SVM classifier for non separable data is a relaxed version of the separable data known as soft margin classifier (Gualtieri and Cromp, 1998). A new variable known as slack variable ($\xi_i$) is introduced which allows a certain amount of misclassification. Where $\xi_i \geq 0$ and $i = 1,2,\ldots l$.
Hence the equation of the hyperplane with the slack variable can be written from equation (3.6)

$$y_i(\mathbf{w}.\boldsymbol{x_i}+b)-1+\xi_i \geq 0 \text{ where } i=1,2,\ldots l \quad (3.15)$$

The equation of the optimal hyperplane is derived by solving the equation

$$\min_{\mathbf{w},b,\xi_1\ldots\xi_l}\left[\frac{1}{2}|\mathbf{w}|^2+\ C\sum_{i=1}^{l}\xi_i\right] \quad (3.16)$$

where, $C$ is a constant which minimizes the solution for which $\xi_i$ get larger. Hence, C is an important parameter, which decides the appropriate hyperplane of the classifier.
The dual optimization for non-separable data is given by the equation (3.17)

$$\left.\begin{array}{c} \max_{\lambda_1,\ldots,\lambda_l}\left[-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i y_i(\boldsymbol{x_i}.\boldsymbol{x_j})y_j\lambda_j+\sum_{i=1}^{l}\lambda_i\right] \\ C \geq \lambda_i \geq 0 \quad \text{and } i=1,2,\ldots l \\ \sum_{i=1}^{l}\lambda_i y_i = 0 \text{ and } i=1,2,\ldots l \end{array}\right\} \quad (3.17)$$

The only difference between the separable and non-separable dual problems is that in non-separable data the Lagrangian dual variables are bounded by the constant C making an impression that the non-separable data is valid only when $\xi_i = 0$. The classifier becomes soft when $\xi_i > 0$.
The points, which are non-separable, are calculated and segregated by applying the condition $\lambda_i = C$.

### 3.2.2. Non-Linear SVM

Not always non-separable data have a solution using liner SVM classifier. There exist certain cases where linear classifier fails to find an optimal solution of classification. In such situations, nonlinear type of classification is used. The decision surface is nonlinear in this type of classification unlike linear in the linear type of classification. A nonlinear hyperplane is achieved by introducing a

nonlinear kernel function into the SVM dual problem, which is responsible for deriving the hyperplane (Gualtieri and Cromp, 1998).

From equations (3.14) and (3.17) the training data is entering the optimization as a dot product. That means it is linear in nature. If a modification is done at that stage we can achieve a non-linear function. That is performed by introducing a nonlinear function $\Phi: \boldsymbol{R^2} \rightarrow \boldsymbol{T}$ where T is a Euclidian space where the feature vectors are mapped. Hence, the optimal problem can be replaced by $\boldsymbol{\Phi(x_i).\Phi(x_j)}$. Then the equation is used to solve the optimization solution, and the classifier function is derived is given in the equation (3.18)

$$f(\mathbf{x}, \lambda_1, \ldots, \lambda_l) = sgn(\textstyle\sum_{i=1}^{l} \lambda_i y_i \, \Phi(\boldsymbol{x}_i).\Phi(\boldsymbol{x}_j) + b) \tag{3.18}$$

The function $\boldsymbol{\Phi(x_i).\Phi(x_j)}$ is denoted by $\boldsymbol{K(x_i, x_j)}$ known as a kernel. This kernel function is known as a nonlinear kernel. Depending upon the nonlinear function used the hyperplane shape changes and generalization of the SVM could be achieved. The kernel matrix **K** for a non-linear mapping is:

$$\boldsymbol{K} = \begin{bmatrix} \phi(\boldsymbol{x}_1)^T \phi(\boldsymbol{x}_1) & \phi(\boldsymbol{x}_1)^T \phi(\boldsymbol{x}_2) & . & \phi(\boldsymbol{x}_1)^T \phi(\boldsymbol{x}_m) \\ . & . & \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j) & . \\ . & . & . & . \\ \phi(\boldsymbol{x}_m)^T \phi(\boldsymbol{x}_1) & \phi(\boldsymbol{x}_m)^T \phi(\boldsymbol{x}_2) & . & \phi(\boldsymbol{x}_m)^T \phi(\boldsymbol{x}_m) \end{bmatrix}$$

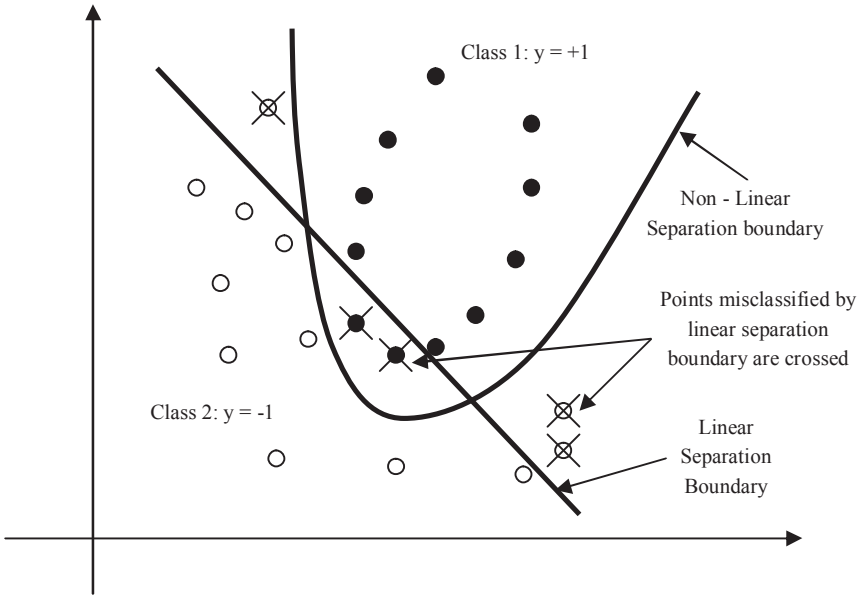



Figure 3:5. A nonlinear SVM classifier and linear SVM classifier

The above figure 3.5 shows the difference between a linear and a nonlinear hyperplane of a SVM classifier. A nonlinear classifier solves a non-separable case of a linear classifier. The function $\boldsymbol{\Phi}$ in a nonlinear classifier is not explicitly computed. This leads to very high computational cost. Instead the kernel $\boldsymbol{K(x_i, x_j)}$ is computed directly. This is known as "kernel trick", which SVM utilizes to reduce the computational cost making it to be the fast and efficient in performing classification. A nonlinear mapping of feature space is shown in figure 3.6.

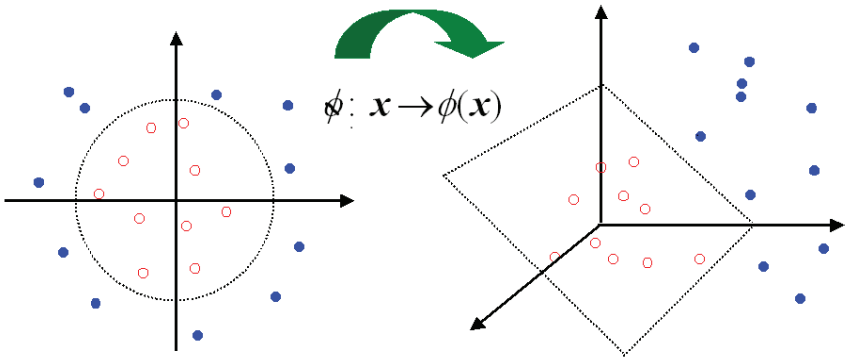



Figure 3:6. Non-linear mapping into feature space

The above description is for two class SVM classification where only two values of classes are considered +1 and -1 as basically SVM is a binary classifier. But the binary classification can also be extended for multi class classification. This can be following way.

In multi class classification, consider there are **K** classes, which are to be classified. Perform $\binom{K}{2} = \frac{K(K-1)}{2}$ binary classifications on all the pairs of the training data and apply to each vector of the test data.

## 3.3. Types of Non-Linear Kernels

**Polynomial kernels**

Let $x \in \Re^2$ i.e., $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and if we choose $\Phi(x) = \begin{bmatrix} {x_1}^2 \\ \sqrt{2}x_1x_2 \\ {x_2}^2 \end{bmatrix}$ (i.e., there is an $\Re^2 \to \Re^3$ mapping, kernel function is $K(x_i, x_j) = \left[x_{i1}^2 x_{j1}^2 + \quad 2x_{i1}x_{i2}x_{j1}x_{j2} + \quad x_{i2}^2 x_{j2}^2\right] = (x_i^T x_j)^2$

The kernel function is a polynomial function. To calculate the scalar product in a feature space $\phi(x_i)^T \phi(x_j)$, we need not perform the mapping $\Phi(x) = \begin{bmatrix} {x_1}^2 \\ \sqrt{2}x_1x_2 \\ {x_2}^2 \end{bmatrix}$, the function is directly calculated by computing $(x_i^T x_j)^2$

**Radial Basis Function Kernel**

The radial basis function is given by $k(x, y) = \exp(-\sigma \|x - y\|^2)$ where $\sigma$ is a positive parameter controlling the radius. Expanding $\exp(-\sigma \|x - y\|^2)$ we obtain

$$\exp(-\sigma \|x - y\|^2) = \exp(-\sigma \|x\|^2) \exp(-\sigma \|y\|^2) \exp(2\sigma x^T y).$$

Since $\exp(2\sigma x^T y) = 1 + 2\sigma x^T y + \frac{(2\sigma)^2}{2!}(x^T y)^2 + \frac{(2\sigma)^3}{3!}(x^T y)^3 + \ldots..,$

$\exp(2\sigma x^T y)$ is an infinite summation of polynomials . Thus it is a kernel whose mapping function maps a point to an infinite dimensional space. Also since $\exp(-\sigma \|x\|^2)$ and $\exp(-\sigma \|y\|^2)$ are

proved to be kernels and the product of two kernels is also a kernel, $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\sigma \|\boldsymbol{x} - \boldsymbol{y}\|^2)$ is a kernel (Soman, 2009). The reason for using RBF kernel is it performs well and gives considerably better results when data has large number of features. All the above discussed methods are done for two dimensions $n$=2 however, this also satisfies for even larger dimensions.

## 3.4. Dimensionality Reduction and Classification

**Process of Dimensionality Reduction**

Dimensionality reduction is performed by Eigen decomposition of the covariance matrix. In this process covariance matrix of the original image is calculated and the Eigen values and the Eigen vectors are calculated for each band. The scores for each band are calculated and the corresponding components are obtained. These components have the data in the order of decreasing variability. As per the above statements, it means that the first component obtained will have the data, which have maximum variability, and so the variability decreases. In this way, the initial components will have the maximum amount of data. However, selecting the optimum number of transformed bands for obtaining the dimensionally reduced image is real task, as there are no standards to calculate them.

**Modified Broken Stick Rule for Intrinsic Dimensionality**

After the reduced bands of the complete hyperspectral image are obtained and their variability are calculated here comes the important part where the number of bands are to be selected for the further processing and analysis of the image. This process of selecting optimum bands where most of the data is contained is known as the intrinsic dimensionality of the image. The traditional statically methods used are to find the number of bands by calculating the number of bands falling into the threshold set on the percentage variability of the image. This is generally set between 98 – 99% of variability. By these methods, we achieve the intrinsic dimensionality mostly between 2 to 5. This would give better results if the number of bands were low. However, if a dataset like hyperspectral image are used which have very high dimensions the number of dimensions achieved by that above method is not satisfactory. As the dimensions are very high, such a low intrinsic dimensionality will not be promising and the complete feature detection chances would be less. Therefore, by setting a bigger threshold could increase the chance of apt feature selection and detection form the hyperspectral image. However, which threshold must be set. It cannot be an arbitrary value. To find out that broken stick rule has been used to calculate the intrinsic dimensionality of the hyperspectral image (Bajorski, 2009).

Virtual Dimensionality, because of some undesirable properties, produces unreasonable results in case of hyperspectral images. Second moment linear dimensionality technique avoids the pitfalls of virtual dimensionality and is successful in identifying a certain number of components. It locates exceptionally large gaps in Eigen values and gives a unique solution if the recommended level is used (Jackson, 2003). The results will depend upon the user-defined threshold, which in all cases may not be optimum, but Modified Broken-Stick Rule (MBSR) avoids it. In MSBR method, k is the number of principal components out of total dimension p and 'λ' are Eigen values of various dimensions.

The value of k is defined as

$$\frac{\lambda_j}{\sum_{i=j}^{p}\lambda_i} > b_j \text{ for } j = 1,2,\dots,k \text{ and } \lambda_{k+1} \leq b_{k+1}$$

Where, $b_j = \left(\frac{1}{p} - j + 1\right)\sum_{i=1}^{p-j+1}\left(\frac{1}{i}\right)$ is a fair share of total variability represented by $\lambda_j$ within $\lambda_j, \dots, \lambda_p$

SVM is a binary classifier. It classifies only two classes. Now, how to classify multi classes? For that, we have techniques like one – many and ad-hoc classification methods. Although the application of SVMs to multiclass classification problems remains an open issue, in practice the one-versus-the-rest approach is the most widely used in spite of its ad-hoc formulation and its practical limitations (Bajorski, 2011).

SVM is the classifier, which is used in this research work. Detailed description of SVM has been done in the previous chapters of this thesis. SVM used in the framework is a nonlinear SVM. Radial basis function (RBF) kernel is used in SVM. For dimensionality reduction the complete data is transformed into the lower dimensions using the Eigen decomposition method.

## 3.5. Proposed SVMDRC Framework

The SVM based dimensionality framework developed in this study is an integration of dimensionality reduction procedure based on Eigen vector analysis, automatic selection of optimal transformed components and simultaneously classification of hyperspectral image using non-linear SVM. The proposed framework has been mathematically briefed as follows:

Let $\mathrm{X}(\mathrm{x_i}, \mathrm{y_i})\ \mathrm{where\ i} = 1, \dots, \mathrm{n} \in \mathrm{R}^2$ is the input data and $\mathrm{X^T}(\mathrm{y_i}, \mathrm{x_i})\ \mathrm{where\ i} = 1, \dots, \mathrm{n} \in \mathrm{R}^2$ be the transposed matrix of the input data.

Calculate the covariance of the data denoted by K

$$\mathrm{K} = \mathrm{cov}(\mathrm{X}, \mathrm{X^T}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(\mathrm{x_i} - \overline{\mathrm{X}})(\mathrm{y_i} - \overline{\mathrm{Y}})}{(\mathrm{n} - 1)}$$

Apply Eigen decomposition to K and obtain the Eigen vector and Eigen values

$$\mathrm{V^{-1}KV} = \mathrm{D}$$

where V is the Eigen values matrix and D is the Eigen vector Matrix

Sort the values of V in the decreasing order

$\mathrm{K}(\mathrm{p}, \mathrm{q}) = \mathrm{V}(\mathrm{p}, \mathrm{q})$ where $\mathrm{p} = 1, \dots, \mathrm{m}$ and $\mathrm{q} = 1, \dots, \mathrm{l}$ and $1 \leq \mathrm{l} \leq \mathrm{m}$

$$\frac{\lambda_\mathrm{j}}{\sum_{\mathrm{i=j}}^{\mathrm{p}}\lambda_\mathrm{i}} > \mathrm{b_j} \text{ for } \mathrm{j} = 1,2,\dots,\mathrm{k} \text{ and } \lambda_{\mathrm{k+1}} \leq \mathrm{b_{k+1}}$$

Where, $\mathrm{b_j} = \left(\frac{1}{\mathrm{p}} - \mathrm{j} + 1\right)\sum_{\mathrm{i}=1}^{\mathrm{p-j+1}}\left(\frac{1}{\mathrm{i}}\right)$ is a fair share of total variability represented by $\lambda_\mathrm{j}$ within $\lambda_\mathrm{j}, \dots, \lambda_\mathrm{p}$. J gives the intrinsic dimensionality.

$\mathrm{C} = \frac{(\mathrm{x_i} - \overline{\mathrm{X}})}{\mathrm{s.h}}$ where c is the normalized value.

The dimensionally reduced and projected data is

$$Y = W^T.C$$

Applying SVM for the projected data

$$y_i(w.x_i + b) - 1 + \xi_i \geq 0 \text{ where } i = 1,2,\dots l$$

$$f(x, \lambda_1, \dots, \lambda_l) = \text{sgn}(\sum_{i=1}^{l} \lambda_i y_i \, \Phi(x_i).\Phi(x_j) + b)$$

where $\Phi()$ is the Radial Basis Function.

## 3.6. Separability Analysis and The Jeffries-Matusita (JM) Distance

JM distance measure is used to define the distances (class separability) between two distributions. Consider two distributions $p(x|\omega_i)$ and $p(x|\omega_j)$ which are polynomial populations each having N classes. The sum of the values is equal to 1 as per laws of probability.

The JM distance is defined as in equation (3.19)

$$J_{ij} = \int_x \{\sqrt{p(x|\omega_i)} - \sqrt{p(x|\omega_j)}\}^2 \, dx \quad (3.19)$$

It is the measure of average distances between two class probability density functions.
If the classes are normally distributed then the distance is given by the equation (3.20)

$$J_{ij} = 2(1 - e^{-B}) \quad (3.20)$$

Where, B is Bhattacharyya distance. The importance of the exponential factor is that it gives the decreasing weight for the increasing separability between the spectral classes. The values of the distance are scaled between 0 and 2.0. The distance with a value 2.0 indicates that the classes are 100% separable and a value 0 that the classes are not separable (Richards and Jia, 2006).

## 3.7. Validation

Validation is an essential step for any classification, which assesses its accuracy. By performing validation one can evaluate the accuracy of the classification performed, whether all the classes in the image are classified correctly or not. For performing validation, we need to have the ground truth data. By using this ground truth data, pixel wise check is performed on the classified image and the accuracy is calculated.

Validation samples are collected and a list of the pixels is made. The pixels in the classified image for which the correct ground data available are taken and then the validation is performed. The result of the validation test is obtained which has the class wise accuracy of classification, over all accuracy of classification and the kappa '*k*' value of the classification. The kappa statistic is the difference measured between the change of agreement between the reference data and a random classifier to that of actual agreement between the reference data and an automated classifier. This statistic is an indicator for degree to which the error matrix percentage values are true to that of true agreement versus chance agreement. The value of *k* lies between 0 and 1. The kappa value '*k*' is mathematically defined as:(Lillesand et al., 1994)

$$k = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})}$$

where,

$r$ = number of rows in the error matrix

$x_{ii}$= the number of observations in row i and column i

$x_{i+}$= total observations in row i

$x_{+i}$= total observations in column i

$N$ = total number of observations in the matrix

Validation for the both the classified images, obtained from the two methods viz., SVMDRC and SVM, is performed.

## 3.8. Software

The framework for the research work is completely written in R, a statistical programming language (R core team, 2008). As open source software, it is distributed under GNU Affero General Public License. Just like any other programming language, R also requires to import the required packages. R has many packages of which user have a liberty to choose the package of choice, which serves his need in accomplishing the tasks. In the same way certain packages has been imported for performing the task of developing the framework. Five packages have been used in the development of the framework. They are as follows

- kernlab
- rgdal
- sp
- raster
- rioja

The above used packages have their own roles in the development of the framework. The "kernlab" (Karatzoglou, 2012) package is used to import the function ksvm which is the main function used for classifying the image. The "rgdal" (Timothy, 2013) is used to real the lat/long information present in the input image. The "sp" (Pebesma, 2005) is used to read the classes and methods of the spatial data. It is the main function which identifies the frames and grids in the images and creates the output frame for the final classified image. The "raster" (Robert, 2012) package is used to read/ write raster files of the image. These raster files created are further processed in Erdas IMAGINE for validation and calculation of the accuracy of the classification. The "rioja" (Juggins, 2012) is a package which is used to perform advanced statistical analysis.

# 4. RESULTS AND DISCUSSIONS

By using the framework designed for dimensionality reduction and classification, the following results were obtained. All the results are discussed below:

## 4.1. Pre-processing

As described in the methodology left part of it is the framework and is completely written in R. The code for the frame work is placed in appendix A. The code is divided into several parts for the ease of understanding, but the code runs in a single process. It is a combination of dimensionality reduction and classification using SVM. The first part of the code is the data reading part. Where the hyperspectral image is read by the code, form the folder containing the input data. Since the image is a spatial data which is georeferenced, 'rgdal' package is used to read the spatial content of the image. Then the image read by the code is displayed. The displayed image is shown in the figure 4.1.

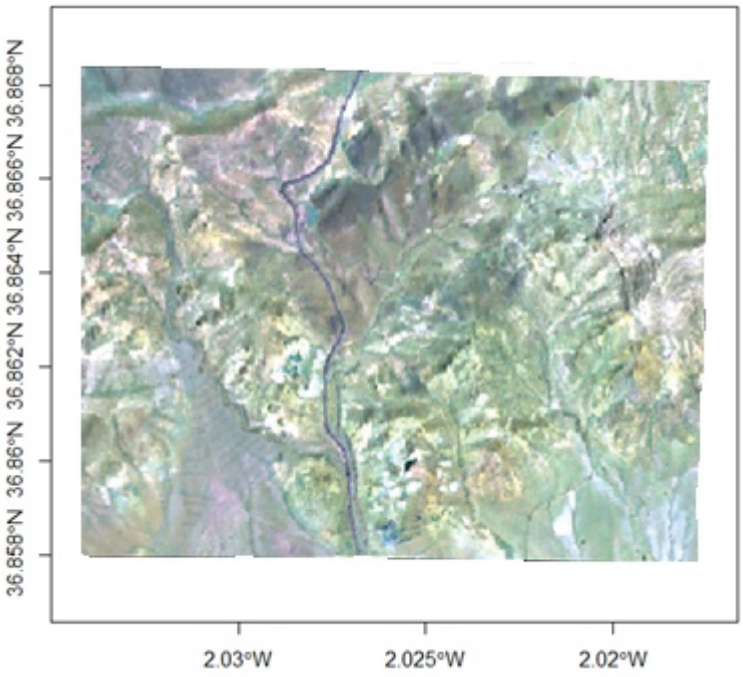



Figure 4:1. Input HyMap image displayed by the framework in RGB (bands 90, 60, 20)

The inputs required for it are the georeferenced hyperspectral image and the corresponding training sets of the classes. The input image is converted to tiff format because computations performed on the tiff image will be faster and near to accurate results are obtained. The framework automatically reads the images once the path of the folder containing the input image and the roi's are specified.

Once the input image is given and the training data sets are given to the framework it reads the inputs and the training sets for further processing. Figure 4.1 shows the image displayed in bands 90, 60 and 20 of the input dataset read by the framework. The selection of the bands to be displayed is user defined and is not fixed to certain values. Then the image is forwarded to the second part where the dimensionality reduction and classification of the data is performed using SVM. The

Training data of the four classes to be classified is collected in the form of roi's and are converted into geotiff image format. This step is performed to read the training sample by far for further processes. The framework for training the classifier reads the training samples.

## 4.2. Dimensionality Reduction and Classification

The framework next performs dimensionality reduction and classification. The covariance matrix of the image is calculated and Eigen decomposition is performed. By this procedure, Eigen values and Eigen vectors for the image are obtained. They are named as scores and loadings. This is a standard followed by the developers and is incorporated in this research work. The Eigen values for each band are calculated and a list of them for all the bands is generated. Table 4.1, depicts the Eigen values of the first ten bands of the image. The remaining values are very small hence, they are not displayed. The first five bands contain 98.8% of data in it. So, in the traditional way the intrinsic dimensionality of the reduced transformed image is considered to be 5. Nevertheless, this is not true in the case of hyperspectral imagery. Due to high correlation between the bands and the data shown in the first 5 bands as per the traditional way is limited. To overcome this, modified broken stick rule is implemented to calculate the intrinsic dimensionality of the reduced image.

Table 4-1. Eigen values of the first 10 transformed bands of the image

| Transformed Bands | Eigen Value |
|---|---|
| 1 | 114.49 |
| 2 | 6.16 |
| 3 | 2.40 |
| 4 | 1.53 |
| 5 | 0.41 |
| 6 | 0.33 |
| 7 | 0.19 |
| 8 | 0.16 |
| 9 | 0.07 |
| 10 | 0.04 |

### 4.2.1. Modified Broken Stick Rule

By the MSBR method, the intrinsic dimensionality is the number of dimensions out of total dimension of the image, which is calculated and found out to be 27 for the dataset used in this research. This is how intrinsic dimensionality is calculated and the reduced data with 27 bands is classified. It is to be noted that the complete process of dimensionality reduction and classification is a single stepped processes, but the intrinsic dimensionality can be retrieved just for knowledge by passing the arguments in the code. In the figure 4.2, the transformed components selected by the broken stick rule along with remaining noisy transformed bands are displayed. After the 27$^{th}$ transformed band the noise content in the image is increased drastically, even though the transformed band have information the noise content is dominating it, which could be visually interpreted. Name written in yellow colour is transformed band considered in intrinsic dimensionality and red one is the band with noise.

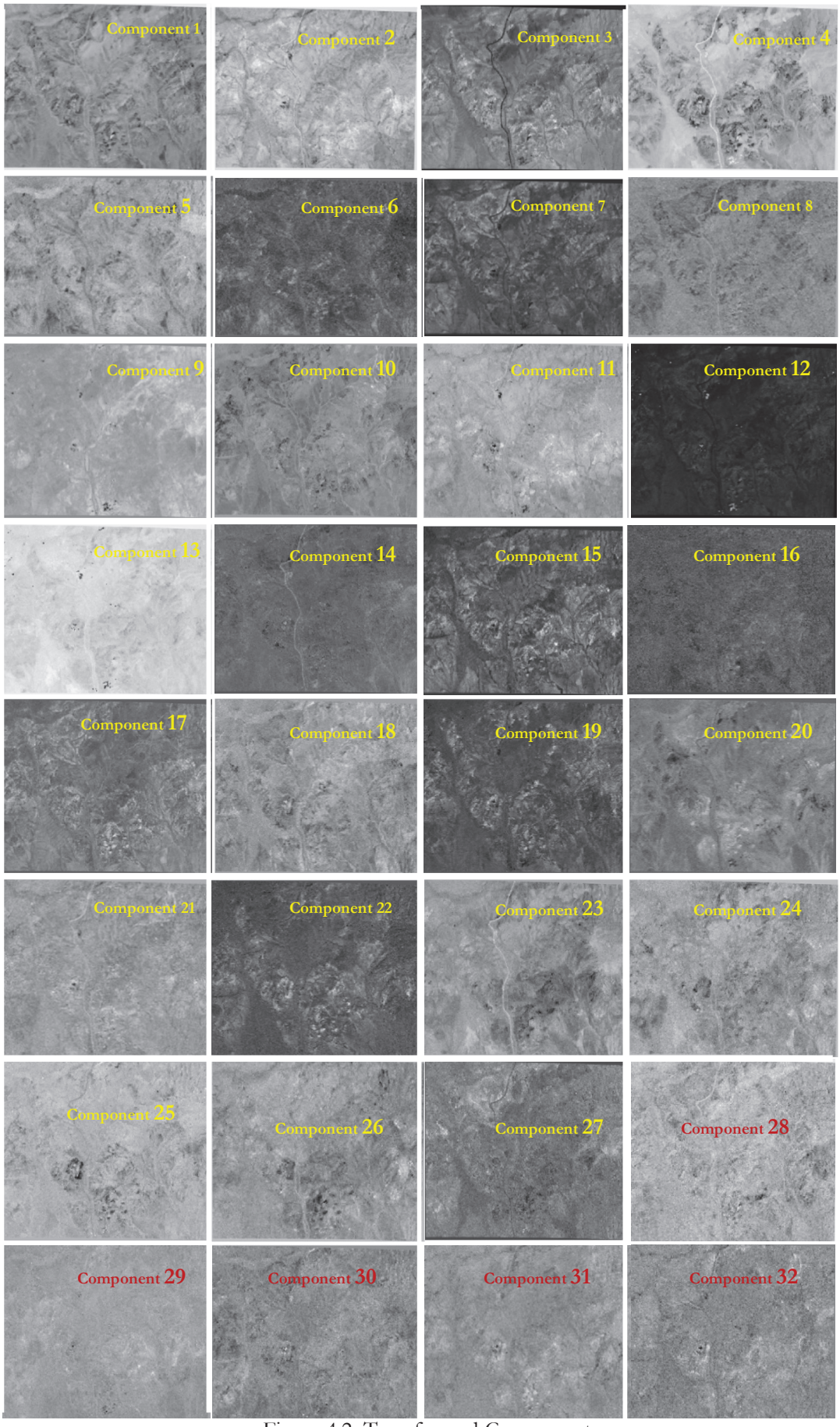



Figure 4:2. Transformed Components

## 4.3. Training Set

Endmembers of all the four classes, which are to be classified, are collected from the image in the form of roi's. These roi's are used as training set for classifying the minerals alunite, illite, kaolinite and limestone. The training samples are renamed as C1, C2, C3, and C4 and converted to tiff format for the ease of programming. Figure 4.7, shows the HyMap image of study area displaying the roi's used and table 4.2 gives their description.

Table 4-2. Details of training data roi's

| Region | Color | Number of Pixels |
|---|---|---|
| Alunite (Region 1) | Red | 63 |
| Illite (Region 2) | Green | 99 |
| Kaolinite(Region 3) | Blue | 115 |
| Limestone (Region 4) | Yellow | 161 |

Figures 4.3, 4.4, 4.5, 4.6 shows the spectra of the pure endmember spectra extracted which are processed as the training samples for further work.

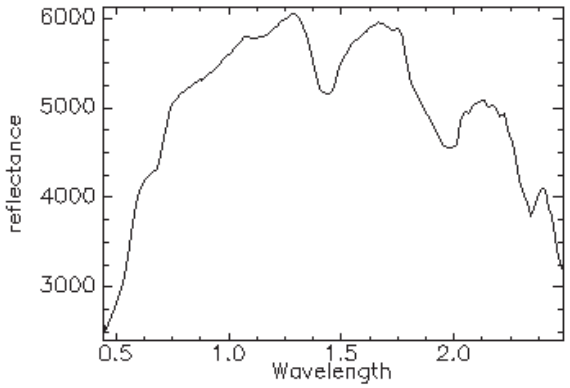



Figure 4:3. Spectral profile of limestone

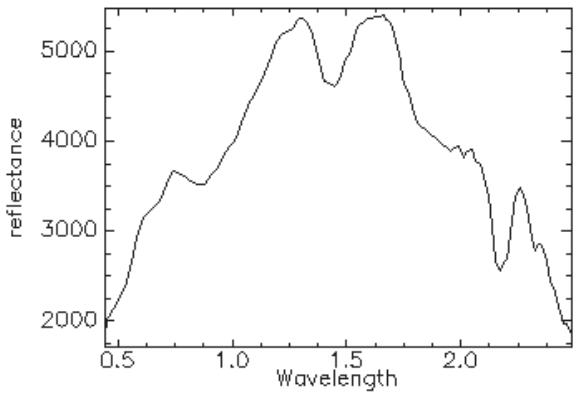



Figure 4:4. Spectral profile of alunite

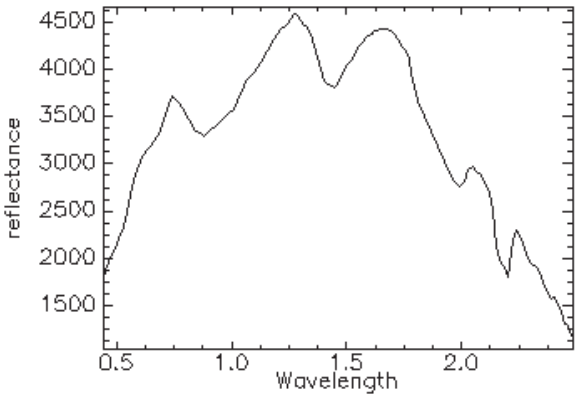



Figure 4:5. Spectral profile of kaolinite

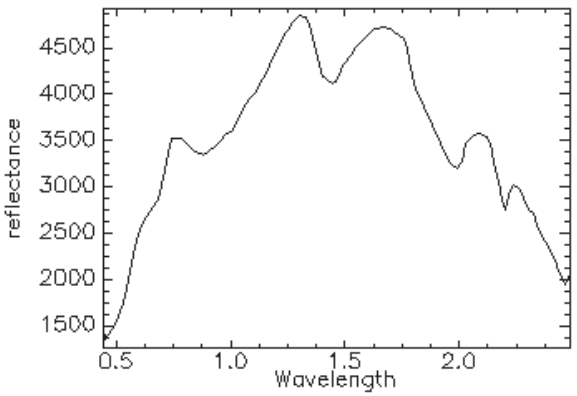



Figure 4:6. Spectral profile of illite

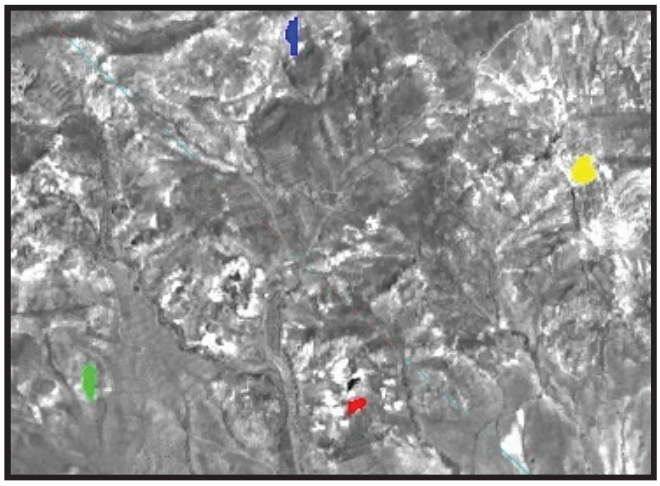


Figure 4:7. Band 35 of the dataset displaying roi's used for training

## 4.4. Dimensionally Reduced and Classified Image

The framework produces a raster image, which is dimensionally reduced and classified. There are two parts in the research work, the SVMDRC and SVMC part. Both of them give a classified image as output. However, the differences between them are one is dimensionally reduced and other is not dimensionally reduced and the other difference is the accuracy of the classification. The accuracy is checked by performing accuracy assessment.

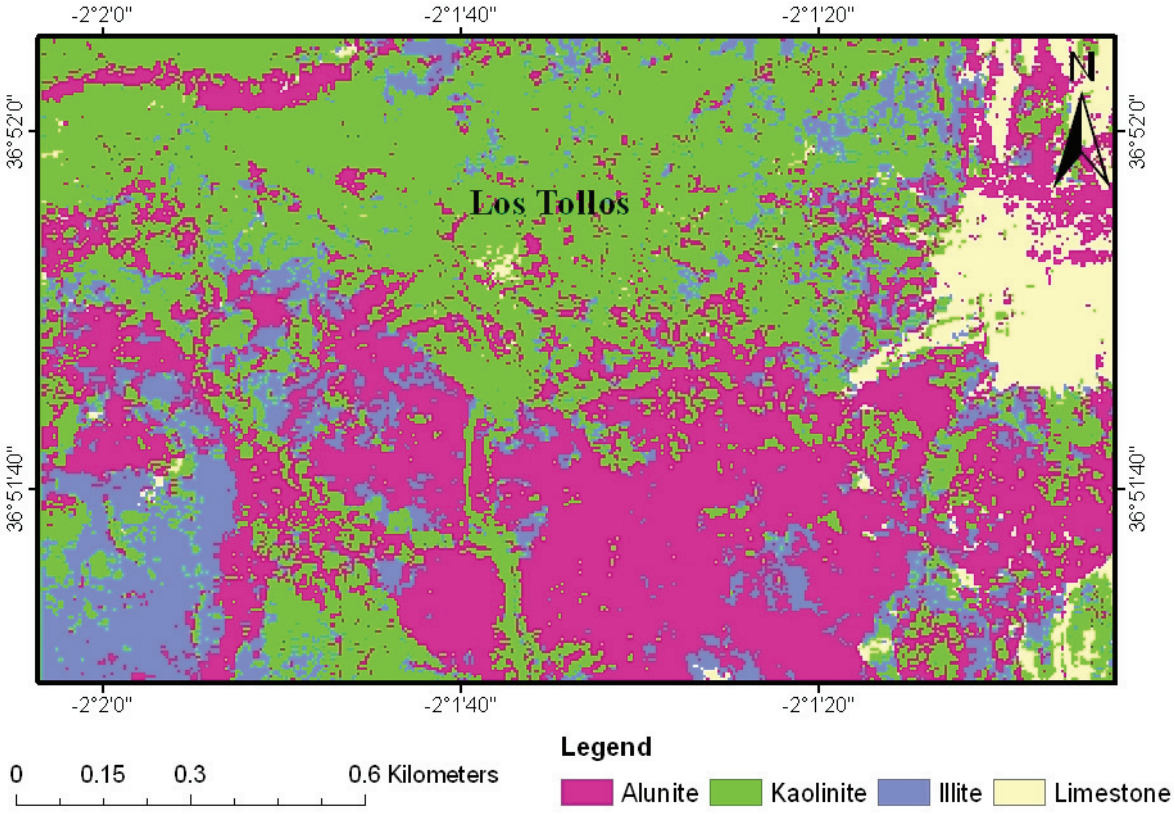



Figure 4:8. SVMDRC classified Image

The figures (4.8 and 4.9) shows the outputs obtained by two different processes of classification using SVM one with dimensionality reduction and other without dimensionality

reduction. The framework performs dimensionality reduction. A part of the framework is written for performing the dimensionality reduction. This part focuses on the reducing the dimensions of the image by projecting the original input image into a higher dimensional space. In dimensionality reduction, the intrinsic dimensionality of the image is estimated and calculated by modified broken sick rule method and the value is observed to be 27. The first 27 reduced bands are used for classification of the image. The dimensionally reduced and classified image is shown in the figure 4.8.

## 4.5. Classified Image without Dimensionality Reduction

Figure 4.9 shows the image obtained by performing SVM classification without dimensionality reduction of the hyperspectral image. The training of the SVM classifier is done with the same training sets used for SVM dimensionality reduction and classification process, to make sure that same type of classification occurs.

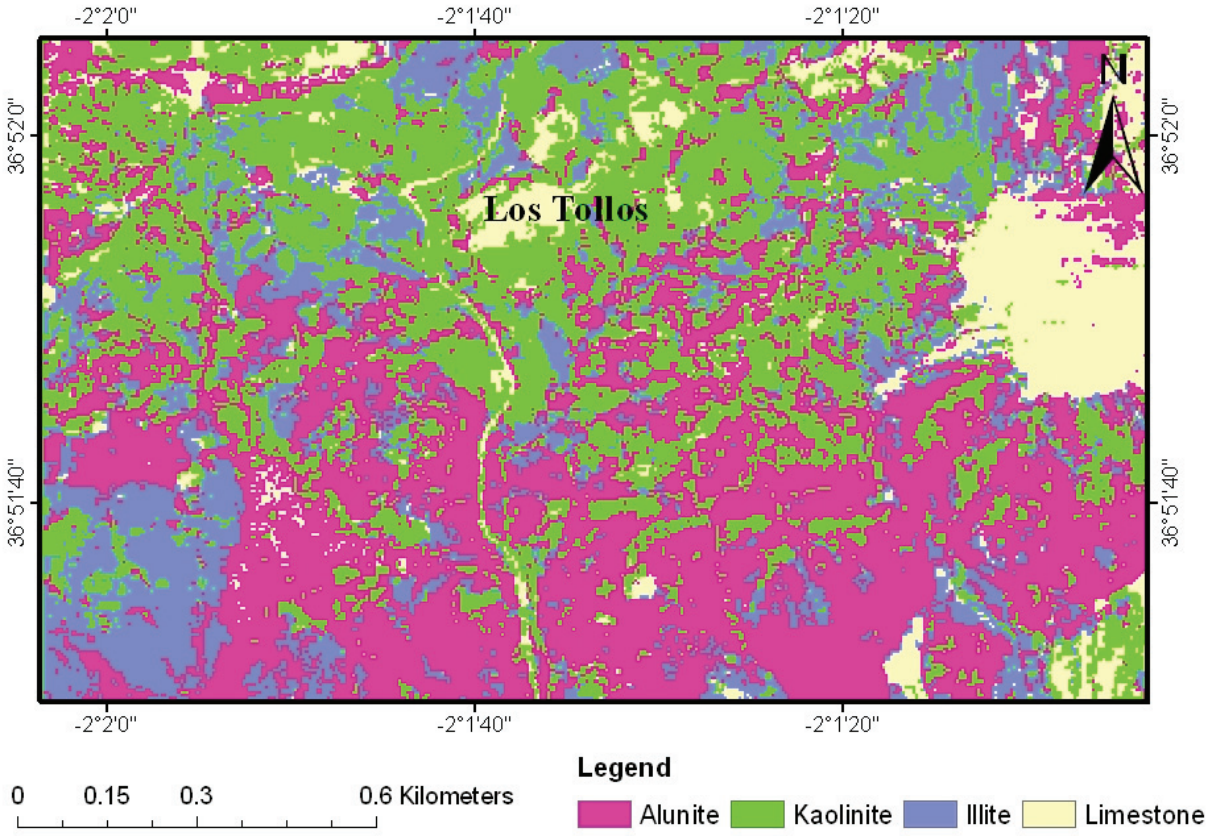



Figure 4:9. Conventional SVM classified image without dimensionality reduction (SVMC)

## 4.6. Validation

Validation is an important step, which gives the true assessment of the results obtained, i.e., it gives the accuracy of the classified image. Validation data has been collected from different parts of the study area. By this validation data also known as ground truth data, validation of the classified image is performed.

**Accuracy Assessment Result**

Accuracy assessment is performed by calculating the confusion matrix between the ground truth data and the classified image. This process is performed on both the classified images. The results of the SVM classified image and dimensionally reduced and classified image and are given in the table 4.3 and 4.4 respectively.

Table 4-3. Validation result obtained from SVM classified image

| **Accuracy report (Confusion Matrix)** | | | | | | |
|---|---|---|---|---|---|---|
| | | Ground Truth | | | | |
| | Class Name | Alunite | Kaolinite | Illite | Total | Users Accuracy |
| Classified Image | Alunite | **6** | 1 | 0 | 7 | 85.71% |
| | Kaolinite | 1 | **3** | 1 | 5 | 60.00% |
| | Illite | 0 | 3 | **2** | 5 | 40.00% |
| | Total | 7 | 7 | 3 | **17** | |
| | Producer Accuracy | 85.71% | 42.85% | 66.66% | | |

Overall Accuracy of classification is = $\frac{6+3+2}{17} = \frac{11}{17} = 0.6470 \text{ X } 100 = 64.70\%$

The kappa value $k$ is = $\frac{(11X17)-[(7X7)+(5X7)+(5X3)]}{17^2-[(7X7)+(5X7)+(5X3)]} = \frac{88}{190} = 0.4631$

$k = 0.4361$

Table 4-4. Validation result obtained from dimensionally reduced and SVM classified image

| **Accuracy report (Confusion Matrix)** | | | | | | |
|---|---|---|---|---|---|---|
| | | Ground Truth | | | | |
| | Class Name | Alunite | Kaolinite | Illite | Total | Users Accuracy |
| Classified Image | Alunite | **6** | 0 | 0 | 6 | 100% |
| | Kaolinite | 1 | **6** | 1 | 8 | 75.00% |
| | Illite | 0 | 1 | **2** | 3 | 66.66% |
| | Total | 7 | 7 | 3 | **17** | |
| | Producer Accuracy | 85.71% | 85.71% | 66.66% | | |

Overall Accuracy of classification is = $\frac{6+6+2}{17} = \frac{14}{17} = 0.8235 \text{ X}100 = 82.35\%$

The kappa value $k$ is = $\frac{(14X17)-[(6X7)+(8X7)+(3X3)]}{17^2-[(6X7)+(8X7)+(3X3)]} = \frac{131}{182} = 0.7197$

$$k = 0.7197$$

The above two tables show the accuracy assessment results of the two classified images. It is clearly shown that the accuracy of the SVM dimensionally reduced and classified image is higher that SVM classified image. There are four classes in the image, which have been classified. The accuracy of the individual classes has increased when dimensionality reduction is performed. The overall accuracy of the SVM classified image is 64.70% and the accuracy has increased to 82.35% when dimensionality reduction is performed on the image. This proves that even though SVM takes care of dimensionality of the image, performing dimensionality reduction will increase the classification accuracy. The individual accuracies of the classes have also been increased. The major change was with the class kaolinite. The accuracy of kaolinite was 42.85% before dimensionality reduction was performed and after it became 85.71%. Before dimensionality reduction, kaolinite has been classified into illite class. This is because of much similarity between both the classes. The spectra of kaolinite and illite classes showed in the figure 4.5 and 4.6 gives an estimate of spectral similarity between them. There is only a small change in the spectral values between the classes. This has lead to misclassification of pixels. By performing dimensionality reduction the separation between the two classes have increased leading to better classification of the pixels. This is explained in the next section.

The classification accuracy of the kaolinite mineral has led to the drastic change in the levels of accuracy and the framework has successfully separated them and gave a better level of classification accuracy.

## 4.7. Influence of Dimensionality Reduction on Feature Extraction

Jeffries-Matusita (JM) distance Method

In this analysis, it is observed that before dimensionality reduction, there was a moderate separability between the classes. The value was 1.34 and the separability values of remaining class pairs were. After dimensionality reduction, the same tests were performed on the image and the vales of all the class separability between the class pair kaolinite-illite was 1.38 and the remaining classes were between 1.3 and 1.41, which is a measure of good separability. Hence, it is shown that performing dimensionality reduction shows a positive index on the feature extraction

Table 4-5. JM distances between the class pairs before dimensionality reduction

| Signature Name | Alunite | Illite | Kaolinite | Limestone |
|---|---|---|---|---|
| Alunite | 0.0 | 1.39 | 1.41 | 1.41 |
| Illite | 1.39 | 0.0 | 1.34 | 1.41 |
| Kaolinite | 1.41 | 1.34 | 0.0 | 1.38 |
| Limestone | 1.41 | 1.41 | 1.38 | 0.0 |

Table 4-6. JM distances between the class pairs after dimensionality reduction

| Signature Name | Alunite | Illite | Kaolinite | Limestone |
|---|---|---|---|---|
| Alunite | 0.0 | 1.41 | 1.41 | 1.41 |
| Illite | 1.41 | 0.0 | 1.38 | 1.41 |
| Kaolinite | 1.41 | 1.38 | 0.0 | 1.38 |
| Limestone | 1.41 | 1.41 | 1.38 | 0.0 |

## 4.8. Discussions and Limitations

The processing of hyperspectral remote sensing data, for variety a of natural resource applications, is challenging due to its higher dimensionality and non-linear characteristics. Considering intrinsic dimensionality by taking the first few bands having maximum variability in the data produces unreasonable results in case of hyperspectral images. Second moment linear dimensionality technique avoids the pitfalls of virtual dimensionality and is successful in identifying a certain number of intrinsic dimensions. It locates exceptionally large gaps in Eigen values and gives a unique solution if the recommended level is used (Bajorski, 2011). The results of the intrinsic dimensionality usually depend upon the user-defined threshold, which in all cases may not be optimum, but Modified Broken-Stick Rule (MBSR) avoids it.

Separability analysis by JM distance method gave an interesting result. Before dimensionality reduction of the image, the separability analysis showed that two pairs of classes were having very less separability between them. This was the reason for the intermixing of the class pixels. There was poor separability between kaolinite and illite. Once dimensionality reduction is performed on the image the separability has increased between the class pair kaolinite and illite. The confusion matrix shows that the misclassification rate has decreased and hence the classification accuracy increases. This result gives an impression that if dimensionality reduction is performed the chances of having misclassified pixels will low.

After validation is performed on the classified images by using, the ground truth data the results obtained were rather interesting. The overall accuracy of the classified image was 64.70% without dimensionality reduction is performed and after dimensionality reduction the overall accuracy was 82.35%. This could be treated as positive response of the dimensionality reduction. This step depicts an important point that when dimensionality reduction is performed on the image the classification accuracy increases and hence this research work suggests that this step is necessary.

The key feature of this research work is the unified framework of dimensionality reduction and classification. There are certain advantages of using a unified framework. First is that the process of working will become shorter. Instead of a two-step process, the work is performed in a single step, this makes the work process faster and more efficient, the results of this research work have proved this point. The second advantage is the usage of SVM in the framework. SVM is a very effective powerful tool for classification. If SVM is applied on ordinary images, having lower dimensions than hyperspectral images there is no need of performing dimensionality reduction. Our results show that use of dimensionality reduction techniques in combination with SVM improves classification accuracy. The framework is completely written in open-source software hence it is accessible to the scientific community for the usage and further improvements.

The major limitations of the research work are unavailability of sufficient ground truth data. The study area selected was in Spain and the feasibility to perform ground truth was very less and practically not possible as the research work is conducted in India. With the limited available ground truth data the accuracy assessment was performed. Out of the four classes classified only three classes had ground truth data to perform validation. With the very less points, the results obtained were satisfactory. If good amount of ground truth data were available validation process might have been conducted with stringent measures. In spite of very sparse ground truth, data the results obtained in this research work are satisfactory compared to the classified outputs of the same area performed in other related works. This shows the strength of framework developed in this research work. By this situation it can be said that the framework can work and produce good results with less training and validation data nevertheless it must be tested.

Another limitation of this research work is regarding the classes classified in the hyperspectral image. Out of four classes classified in the image, three classes have high similarity between them. As the classes classified are minerals, there exists a high similarity between the spectra of the minerals. Kaolinite and illite are having very high similarity between them and another pair kaolinite and alunite pair has moderate similarity between them. Despite of these limitations the framework has successfully classified the classes with reasonably good classification accuracy.

Similar kind of climate (semi-arid) and mineral mining areas like that of Los Tollos in Rodalquilar district of Spain are also there in India. But, availability of the hyperspectral datasets is an issue at this current scenario. There are no hyperspectral datasets for those areas except Hyperion. Once the datasets gets available, the framework could be tested on those datasets.

# 5. CONCLUSION AND RECOMMENDATIONS

As intended, the framework developed in this research work proved to be fruitful. The accuracy of the hyperspectral image classified by the framework has shown better results than the classified using SVM alone. This has also proved that SVM take care of dimensionality to a limited degree. Hence, dimensionality reduction is a compulsory step rather than taking it as a complementary step. Nonlinear type SVM is best suited for hyperspectral data. But the results for non-linear type dimensionality reduction were not conclusive. The framework developed was aimed for working with all types of hyperspectral datasets, subjected to a condition that atmospheric correction is done at prior. The framework is developed using open-source language R which can be widely assessable to the scientific community.

This research work mostly aims on dimensionality reduction and classification using support vector machine (SVM) and next is to prove whether the framework is comparable to that of the conventional method. This chapter summarizes the results obtained by following the methodology adopted for this research work.

## 5.1. How effective is the application of nonlinear function for dimensionality reduction?

Application of nonlinear function for dimensionality reduction on the hyperspectral data was not conclusive. The computational cost tremendously increases if a nonlinear function is used for dimensionality reduction, because hyperspectral images have extremely huge data for which nonlinear functions make the process much more complex and increasing the computational cost. A small 50X50 pixel image takes more than 20min to perform dimensionality reduction where the same image get reduced within 50sec while a linear function is used. Hence, it is suggested to perform dimensionality reduction by using a linear function.

## 5.2. Does a dimensionality reduction technique show any impact on extraction of different features types from hyperspectral data?

Yes, dimensionality reduction shows a positive impact on feature extraction. Separability between the classes of the hyperspectral image before and after dimensionality reduction is calculated. JM distances method is tested between the classes. The same training samples, which were used for classification, are used for this analysis. .

## 5.3. How does SVM based dimensionality reduction and classification algorithm perform as compared to SVM classification on the hyperspectral data?

The framework in this research work have given better accuracy results compared to performing SVM classification directly on the hyperspectral image. The framework shows a good accuracy of classification. Performing SVM on the hyperspectral image have given an accuracy of 64.70% whereas accuracy after the framework is applied was 82.35%. By this, it proves that comparatively the framework gives better results compared to the traditional method followed.

## 5.4. Recommendations

- Despite the positive results of this study, it is still recommended that the performance of proposed framework should be tested with sufficient ground truth data at similar study site in order to adjudge its operationality.
- Spectral similarities within mineral types poses difficulty in their discrimination, hence, narrow band width (< 5 nm) data should be tested to improve their discrimination by capturing subtle spectral differences.

# REFERENCES


Arribas, Jr., A., Cunningham, C. G., Rytuba, J. J., Rye, R. O., Kelley, W. C., Podwysocki, M. H., McKee, E. H. and Tosdal, R. M., (1995). Geology, geochronology, fluid inclusions and isotope geochemistry of the Rodalquilar gold alunite deposit, Spain. *Economic Geology*, 90, 795–822.

Alexandros, K., Smola, A., Hornik, K., Zeileis, A., (2004). The ' kernlab' - An S4.Package for Kernel methods in R. *Journal of Statistical Software* ,11(9), 1-20.

Bajorski, P., (2009). Does Dimensionality work in hyperspectral images? In Proc. *SPIE, Algorithms and technologies for Multispectral, Hyperspectral and Ultrasepectral Imagery XV*.

Bajorski, P., (2011). Second moment linear dimensionality as an alternative to virtual dimensionality. *IEEE Transactions on Geosciences and Remote Sensing*, 49(2), 672-678.

Bedini,E., (2005). Mapping Mineralogy in the Rodalquilar Caldera, Spain with Airborne Imaging Spectrometer Data using Spectral Information Divergence and Multiple Endmember Spectral Mixture Models*, Msc Theis*, ITC, The Netherlands.

Bellman, R. E., (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

Bi, J., Bennett, K., Embrechts, M., Breneman, C., and Song, M., (2003). Dimensionality Reductio Reduction via Sparse Support Vector Machines. *The Journal of Machine LearningResearch,*3, 1229–1243.

Burgers, K., Fessehatsion, Y., Rahmani, S., Seo, J., and Wittman, T., (2009). A Comparative Analysis of Dimension Reduction Algorithms on Hyperspectral Data. *LAMDA Research Group*.

Cocks, T., Jenssen, R., Stewart, W.I., and Sheilds, T.,(1998). The HyMap airborne hyperspectral sensor: The system, calibration and performance. *First EARSEL workshop on imaging spectroscopy*. Zurich, October, 1998

Cortes, C., and Vapnik, V., (1995). Support-vector Networks. *Machine Learning,* 20 (3), 273–97.

Cunningham, C. G., Arribas, Jr., A., Rytuba, J. J., and Arribas, A., (1990). Mineralized and unmineralized calderas in Spain; Part I, evolution of the Los Frailes Caldera. *Mineralium Deposita*, 25, 521–528.

Debba,P., (2005). Sampling scheme optimization from hyperspectral data, *Ph.D. Thesis*, ITC, The Netherlands.

Duda, R.O., Hart, P.E., and Stork, D. G., (2001). *Pattern Classification*, 2nd ed. New York: Wiley.

Fisher, R., (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179 - 188.

Fong, M., (2007). Dimension Reduction on Hyperspectral Images. *Univ. California, Los Angeles, CA*.

Fukunaga, K., (1990). *Introduction to statistical pattern recognition.* 2nd edition Academic Press Prefessional, Inc.

Gualiteri, A.J. and Cromp R.F. (1998). Support Vector Machine for hypersperspectral remotesensing classification. *Advances in computer Assisted Recognition*, Vol 3584.

Harsanyi, J. C., and Chang C. I., (1994). Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach. *IEEE Transactions on Geoscience and Remote Sensing,* 32 (4), 779–785.

Hastie, T., Tibishirani, R.,and Friedman, J. H., (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. NewYork: Springer-Verlag,

Hsu,C.W., and Lin,C.J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Tranactions on Neural Networks*, 13(2) : 415–425.

Hijmans.R.J., and Etten, J.V., (2012). raster: Geographic data analysis and modeling. R package version 2.0-41.

Jackson, J.E., (2003). *A User's Guide to Principal Components*. Wiley.

Ji, S., and Ye. J., (2008). A Unified Framework for Generalized Linear Discriminant Analysis. In *IEEE Conference On Computer Vision and Pattern Recognition, CVPR 2008.* 1–7.

Juggins, S., (2012). rioja: Analysis of Quaternary Science Data, R package version 0.7-3.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (1994). *Remote Sensing and Image Interpretation*. New York, John Wiley & Sons.

Melgani, F., and Bruzzone. L,. (2004). Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing,* 42 (8), 1778–1790.

Mountrakis, G., J., and Ogole C., (2011). Support Vector Machines in Remote Sensing: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66 (3), 247–259.

Muhammad.A., Lee.S., Haq,E., Mushtaq.Q., (2012). Hyperspectral Remote Sensing: Dimensional Reduction and End member Extraction. *International Journal of Soft Computing and Engineering*, 2(2), 170 - 175.

Pal.M., and Foody,G.M., (2010). Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Transactions on Geoscience and Remote Sensing,* 48 (5), 2297–2307.

Pal.M., (2008). Multiclass Approaches for Support Vector Machine Based Land Cover Classification, *arXiv preprint arXiv:0802.2411.*

Petropoulos, G. P., Arvanitis, K., and Sigrimis, N., (2012). Hyperion hyperspectral imagery analysis combined with machine learning classifier for land use/cover mapping, Expert systems and applications, 39, 3800-3809.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richards.J.A, Jia. X., (2006). *Remote Sensing Digital Image Analysis An Introduction*, 4th Edition, Springer.

Schlapfer D. and Ritcher R. (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: parametric orthorectification. *International journal of Remote Sensing*, Vol 23 (14), 2609 - 2630

Soman, K.P., Ajav. V, Loganathan, R., (2009). *Machine Learning with SVM and other Kernel Methods*, Prentice-Hall India.

Silva, C. R., Centeno, J. A., and Ribeiro, S. R. (2008). Reduction of the dimensionality of Hyperspectral data for the classification of agricultural scenes. P*resented at the LNEC*, Lisbon.

Suykens, J. A .K. (2001). Nonlinear modelling and support vector machines. *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, IMTC 2001*. 1, 287-294.

Tao.Q., Chu.D., and Wang.J., (2008). Recursive Support Vector Machines for Dimensionality Reduction. *IEEE Transactions on Neural Networks,* 19 (1), 189–193.

Keith.T.H., Bevan.R., Pebesma.E., and Rowlingson.B., (2013). rgdal: Bindings for the Geospatial Data Abstraction Library, R package version 0.8-4.

Pebesma, E.J., Bevan, R.S., 2005. Classes and methods for spatial data in R. R News 5 (2)

Yang, B., (2009). SVM-Induced Dimensionality Reduction and Classification. In *proceedings of Second conference on intelligent computation technology and Automation*, 275–278.

Renzuang. Z., Carranza .M., (2011). Support vector machine: A tool for mapping mineral prosperity, *Computers and geosciences* 37, 1967-1975.

# APPENDIX

**Appendix A**

Code for dimensionality reduction and classification framework written in R

```
rm(list=ls(all=TRUE))
require(kernlab)
require(rgdal)
require(sp)
require(raster)
require(rioja)
require(Matrix)
Root <- 'C:\\Users\\Pavan Kolluru\\Desktop\\trainingclasses'
Path_in <- 'C:\\Users\\Pavan
Kolluru\\Desktop\\trainingclasses\\input'
Path_start <- getwd()
setwd(Path_in)

histstretch<-function(data)
{
  cur.lim<-quantile(data,c(0.025,0.975),na.rm=TRUE)
  data<-pmax(cur.lim[1],pmin(cur.lim[2],data))
  data<-floor(255*(data-cur.lim[1])/(cur.lim[2]-cur.lim[1]))
  data[is.na(data)]<-0
  data
}

image.fn <- "projj.TIF"
A <- readGDAL(image.fn)

# Image dimensions
Nb <- dim(A)[2]
Npix <- nrow(A)
d <- A@grid@cells.dim
M <- d[1]
N <- d[2]
A.image <- A

#  RGB composition for viewing the input image data
A.image@data$red   <-histstretch(A.image@data$band90)
A.image@data$green <-histstretch(A.image@data$band60)
A.image@data$blue  <-histstretch(A.image@data$band20)
D <- array(0, c(Npix,Nb))

for(k in 1:Nb) D[,k]<- as.matrix(A@data[,k])
xyr <- coordinates(A)

Colour_palette <- c("yellow","green","magenta","red")
Class_names <- c("Class1","Class2","Class3","Class4")
Ncl <- length(Class_names)

windows()
image(A.image,red="red",green="green",blue="blue",axes=TRUE)
title("RGB=90:60:20")

Dtr <- data.frame(array(0,c(0,Nb+1)))
```

```
for(k in 1:Ncl)
{
  image.fn <- paste(Path_in,"/c",k,".TIF",sep="")
  temp <- readGDAL(image.fn)
  P <- temp@data
  p <- temp@grid@cells.dim
  Dtr <- rbind(Dtr,cbind(temp@data,rep(k,nrow(temp))))
}

names(Dtr)[1:Nb]<- names(A)
names(Dtr)[Nb+1]="class"


C_SVM <- 100
sigma_SVM <- .008

# SVM model nonlinear ==>RBF
svm_model <- ksvm(class~.,data=Dtr,type="C-svc",cache
=2000,kernel="rbfdot",kpar=list(sigma=sigma_SVM),C=C_SVM,prob.mod
el=TRUE)

# SVM
SVM <- A
temp <- predict(svm_model, D)
SVM@data <- data.frame(class=temp)
PP = SVM@data

## displaying the SVM classified which is dimensionally not
reduced.
windows()
image(SVM, col=Colour_palette,axes=TRUE)
title("SVM classification")
SVMM = raster(SVM)
## storing the classified image in raster Format
rf <- writeRaster(SVMM, filename="SVM.tif",
format="GTiff",overwrite=TRUE)


covmat <- cov.wt(D)
n.obs <- covmat$n.obs
cv <- covmat$cov *(1 - 1/n.obs)
cen <- covmat$center
sds <- sqrt(diag(cv))
cv <- cv/(sds %o% sds)
edc <- eigen(cv, symmetric =TRUE)
ev <- edc$values
cn <- paste("eigVal", 1L:ncol(cv), sep = "")
names(ev)<- cn
sdev <- sqrt(ev)
sc <- rep(1, ncol(cv))
names(sc)<- colnames(cv)

scr <-    scale(D, center = cen, scale = sc) %*% edc$vectors
edc <- list(sdev = sdev, loadings = structure(edc$vectors,class =
"loadings"), center = cen, scale = sc, n.obs = n.obs,
```

```
              scores = scr)
class(edc)<- "princomp"

covr = cov(D)
l = eigen(covr)
lam = as.matrix(l$values)

cork <- cor(D)
ro = eigen(cork)
rho = as.matrix(ro$values)

KKK = bstick(cv)

rhoo = round(rho, digits =3)
lamm = round(lam, digits =3)

zz <- intersect(rhoo,lamm)
which(zz=TRUE)
ZZ


Nc <- 40
redu <- edc$scores[,1:Nc]
redu <- data.frame(redu)
names(redu)<-
c("rb1","rb2","rb3","rb4","rb5","rb6","rb7","rb8","rb9","rb10","r
b11","rb12","rb13","rb14","rb15","rb16","rb17","rb18","rb19","rb2
0","rb21","rb22","rb23","rb24","rb25","rb26","rb27","rb28","rb29"
,"rb30","rb31","rb32","rb33","rb34","rb35","rb36","rb37","rb38","
rb39","rb40")

B <- A

B@data <- B@data[,-(1:Nb)]


B.image <- B
B.image@data$red   <-histstretch(B.image@data$band1)
B.image@data$green <-histstretch(B.image@data$band2)
B.image@data$blue  <-histstretch(B.image@data$band3)

windows()
image(B.image,red="red",green="green",blue="blue",axes=TRUE)
title("Dim reduced image")


for(i in 1:Nc)
{
  windows()
  image(B.image,attr=i,col=gray((0:255)/255),axes=TRUE)
  title(paste("rb ",i,sep=""))
}


Dtr_rb <- predict(edc,Dtr[,1:Nb])
```

```
Dtr_rb<- Dtr_rb[,c(1:Nc)]
Dtr_rb <- cbind((Dtr_rb),Dtr$class)
Dtr_rb <- data.frame(Dtr_rb)
names(Dtr_rb)<-
c("rb1","rb2","rb3","rb4","rb5","rb6","rb7","rb8","rb9","rb10","r
b11","rb12","rb13","rb14","rb15","rb16","rb17","rb18","rb19","rb2
0","rb21","rb22","rb23","rb24","rb25","rb26","rb27","class")



C_SVM <- 100
sigma_SVM <- .008

svm_model <- ksvm(class~.,data=Dtr_rb,type="C-svc",cache =
2000,kernel="rbfdot",kpar=list(sigma=sigma_SVM),C=C_SVM,prob.mode
l=TRUE)
svm_model

SVM <- B
temp <- predict(svm_model, redu)
SVM@data <- data.frame(class=temp)


windows()
image(SVM, col=Colour_palette,axes=TRUE)
title("SVMred")
SVc = raster(SVM)
## storing the classified image in raster Format
rf <- writeRaster(SVc, filename="SVMred.tif",
format="GTiff",overwrite=TRUE)
```