Heaven's Light is Our Guide

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

Target Class Oriented Subspace Detection for Effective Hyperspectral Image Classification

Author

Md. Tanvir Ahmed

Roll No. 123086

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology

Supervised by

Dr. Md. Ali Hossain

Assistant Professor

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology

ACKNOWLEDGEMENT

I would like to express my special appreciation and thanks to my supervisor **Dr. Md. Ali Hossain**, Assistant Professor, Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, you have been a tremendous mentor for me. Again I would like to thank you for encouraging this research. Your advice on both research as well as on my career have been priceless.

I would also like to express my sincere appreciation & deepest sense of gratitude to my honorable teacher **Dr. Md. Rabiul Islam**, Head of the Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology for his support regarding this research. Finally, thanks to all of my honorable teachers, friends & well-wishers for their great role to do complete this research.

Date: December 4, 2017 Md. Tanvir Ahmed

RUET, Rajshahi

Heaven's Light is Our Guide

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

CERTIFICATE

With immense pleasure, it is hereby certified that the thesis ***Target Class Oriented Subspace Detection for Effective Hyperspectral Image Classification****, is prepared by* ***Md. Tanvir Ahmed****, Roll. 123086, has been carried out under my supervision. The thesis has been prepared in partial fulfillment of requirements for degree of Bachelor of Science in Computer Science and Engineering.*

Supervisor	**External Examiner**
____________________	____________________
Dr. Md. Ali Hossain Assistant Professor Department of Computer Science & Engineering Rajshahi University of Engineering & Technology Rajshahi-6204	Department of Computer Science & Engineering Rajshahi University of Engineering & Technology Rajshahi-6204

Abstract

Achieving high classification accuracy in hyperspectral image classification is a challenging task. This problem can be addressed by reducing the irrelevant features for the task of classification. Principal Component Analysis (PCA) is a popular feature extraction technique but it depends solely on global variance which makes it limited for some application. To address this, a target class oriented feature reduction method is proposed which incorporates the normalized Mutual Information (NMI) over PCA images to maximize the relevance of the selected subspace. Experimental analysis is performed to assess the effectiveness of the proposed method and the selected subspace is evaluated using kernel Support Vector Machine (KSVM) classifier. The proposed approach can achieve 96.57% classification accuracy on real hyperspectral data which is better than the standard approaches studied.

Contents

Acknowledgement i

Certificate ii

Abstract iii

List of Figures vi

List of Tables viii

1 Introduction 1

1.1 Overview 1

1.2 Remote Sensing 2

1.3 Challenges of Remote Sensing 3

1.4 Motivation 4

1.5 Objectives 5

1.6 Outline 5

2 Hyperspectral Imaging 6

2.1 Introduction 6

2.2 Spectral Image Basics 6

2.3 Hyperspectral Data 9

2.4 Imaging Sensors 10

2.5 Application of Hyperspectral Images 12

2.6 Challenges of Hyperspectral Data Processing 13

3 Literature Review 15

3.1 Introduction 15

3.2 Feature Mining . . . 15
3.3 Related Works . . . 18
3.4 Feature Relevancy and Redundancy . . . 20
3.5 General Approach for Feature Selection . . . 20
3.6 Feature Mining for Hyperspectral Image . . . 22
3.7 Classification Techniques . . . 23
3.7.1 Unsupervised Classification . . . 24
3.7.2 Supervised Classification . . . 25

4 Subspace Detection and Classification 26
4.1 Introduction . . . 26
4.2 Feature Extraction . . . 27
4.3 Feature Selection Based on Normalized Mutual Information . . . 28
4.4 Proposed Subspace Detection Method . . . 29
4.5 Experimental Procedure . . . 29
4.5.1 Feature Extraction . . . 29
4.5.2 NMI Based Feature Selection . . . 30
4.5.3 Proposed TCOSD Method . . . 32
4.6 Experimental Results . . . 34
4.6.1 Indian pines (92AV3C) Dataset . . . 34
4.6.2 Feature Extraction Result . . . 35
4.6.3 NMI Based Feature Selection Result . . . 37
4.6.4 Experimental Results . . . 39
4.7 Classification Results . . . 40

5 Conclusion and Future Works 43
5.1 Conclusion . . . 43
5.2 Limitations . . . 43
5.3 Future Works . . . 43

References 44

List of Figures

1.1 Active remote sensing [1] 2

1.2 Passive remote sensing [1] 2

1.3 3D view of Hyperspectral cube [2] 3

1.4 Technical characteristics of digital image data [3] 4

2.1 The electromagnetic spectrum [4] 7

2.2 Reflectance spectra for several common Earth surface materials [4] 7

2.3 Reflectance spectra for important minerals [4] 8

2.4 The concept of hyperspectral imagery. Image measurements are made at many narrow contiguous wavelength bands, resulting in a complete spectrum for each pixel. [4] 8

2.5 Reflectance spectra of the three materials as they would appear to the hyperspectral AVIRIS sensor. The gaps in the spectra are wavelength ranges at which the atmosphere absorbs so much light that no reliable signal is received from the surface. [4] 9

2.6 Curse of dimensionality [5] 14

3.1 Four key steps for the feature selection process 17

3.2 A view of feature relevance 20

3.3 Filter Method for feature selection 21

3.4 Wrapper Method for Feature selection 21

3.5 Embedded Method for Feature selection 22

3.6 Possible operations for feature mining (only one path is used for any specific approach). [12] 23

3.7 Unsupervised Classification 24

3.8 Supervised Classification 25

4.1 Flow chart of the proposed TCOSD Method 33
4.2 AVIRIS 92AV3C dataset for experiment 34
4.3 Ground truth of AVIRIS 92AV3C dataset 35
4.4 PCA image 1, 2, 7 and 9. 36
4.5 Cumulative variance of bands obtained via PCA 37
4.6 Normalized mutual information between class labels and original bands 38
4.7 Normalized mutual information between class labels and PC1-PC20 39
4.8 Classification accuracy of AVIRIS 92AV3C data 42

List of Tables

2.1 Landsat 1, 2, 3 orbit characteristics 10
2.2 Current and Recent Hyperspectral Sensors and Data Providers 11

4.1 Selected features with Org+NMI 37
4.2 Selected features with PCA+NMI 38
4.3 Selected features for each target class with proposed method (TCOSD) 40
4.4 Details of the train and test samples 41
4.5 Details of parameter for KSVM(RBF kernel) 41
4.6 Classification result for 8 features 42

Chapter 1

Introduction

1.1 Overview

Hyperspectral sensors simultaneously measure hundreds of continuous spectral bands with a fine resolution to form a three dimensional hyperspectral image data cube. For instance, the AVIRIS sensor simultaneously measures 224 bands with a fine resolution of 0.01μm. This high data volume presents many challenges which creates opportunity for research. The data captured are highly correlated and contains a significant amount of redundant data. All the image bands are not equally important for specific application. Also, as the feature space dimension increases, if the size of the training data does not grow correspondingly, a reduction in the classification accuracy of the testing data is observed due to poor parameter estimation of the supervised classifier. This effect is known as the Hughes phenomenon [6]. So it is required to extract only relevant features from the input dataset. Therefore an effective and efficient technique to find this relevant features is a major interest in current literature. Principal component analysis (PCA) [7] is one of the most popular feature extraction technique, though its components is not always suitable for better classification accuracy [8]. It is also not sensitive to input classes and consider only the global variance of the dataset. There are few techniques for finding relevant features in current literature. For example, mutual information based feature selection in combination of principal component analysis [9] and normalized mutual information based feature selection [5]. A target class oriented subspace detection technique (TCOSD) is proposed as an alternative for the effective subspace detection in collaboration with kernel support vector machine (SVM) [10] to achieve better classification accuracy.

1.2 Remote Sensing

Remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object. Remote sensing is used in numerous fields, including geography, land surveying and most Earth Science disciplines (for example, hydrology, ecology, oceanography, glaciology, geology); it also has military, intelligence, commercial, economic, planning, and humanitarian applications.

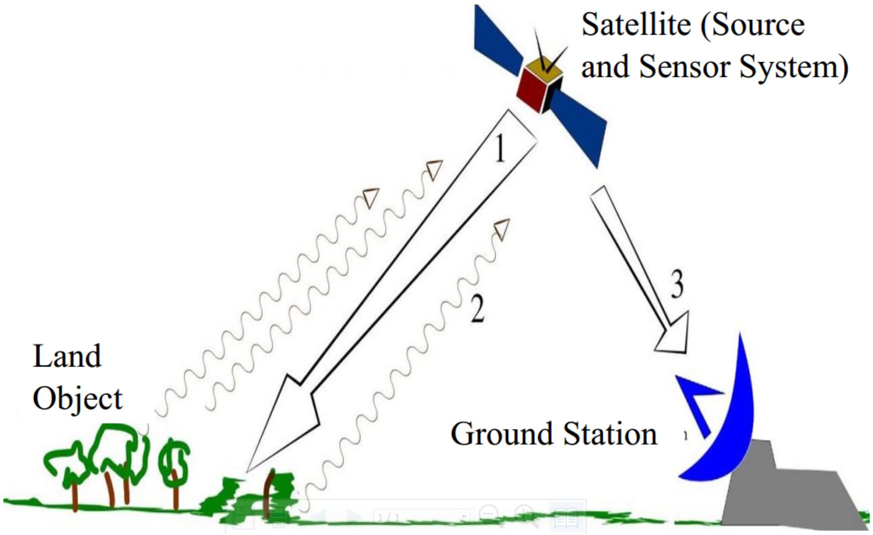

Figure 1.1: Active remote sensing [1]

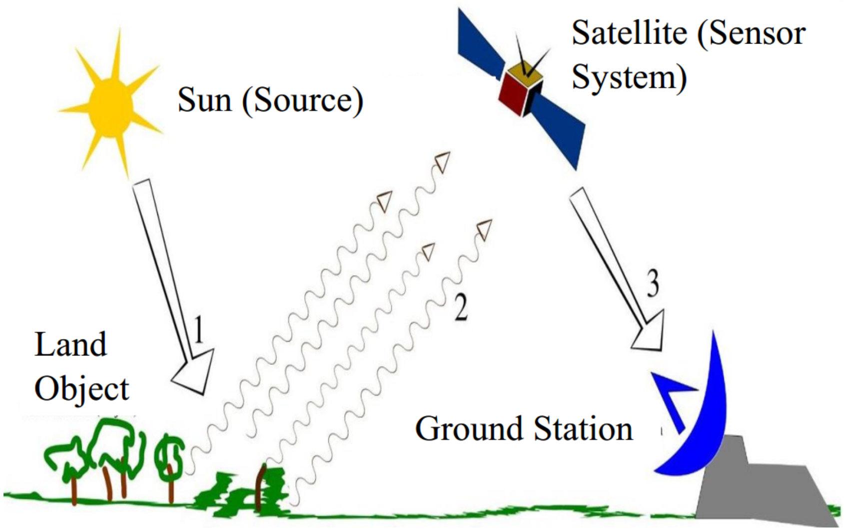

Figure 1.2: Passive remote sensing [1]

In current usage, the term ”remote sensing” generally refers to the use of satellite- or aircraft-based sensor technologies to detect and classify objects on Earth, including on the surface and in the atmosphere and oceans, based on propagated signals (e.g. electromagnetic radiation). It may be split into ”active” remote sensing (i.e., when a signal is emitted by a satellite or aircraft and its reflection by the object is detected by the sensor) and ”passive” remote sensing (i.e., when the reflection of sunlight is detected by the sensor). Fig. 1.1 and Fig. 1.2 shows a scenario with active and passive remote sensing platform. Fig. 1.3 presents a 3D view of data obtained by these sensors.

Figure 1.3: 3D view of Hyperspectral cube [2]

1.3 Challenges of Remote Sensing

As we have entered an era of high resolution earth observation, the RS data are undergoing an explosive growth. The proliferation of data also give rise to the increasing complexity of RS data, like the diversity and higher dimensionality characteristic of the data. RS data are regarded as RS “Big Data”. In ground object detection, remote sensing faces several challenges. A few challenges of remote sensing in ground object detection is given below:

- Field sampling is difficult and expensive
- The geographical orientation is difficult in the images

- The upwelling signal to the sensor is low and disturbed

- The water and its constituents limits the usable spectral region

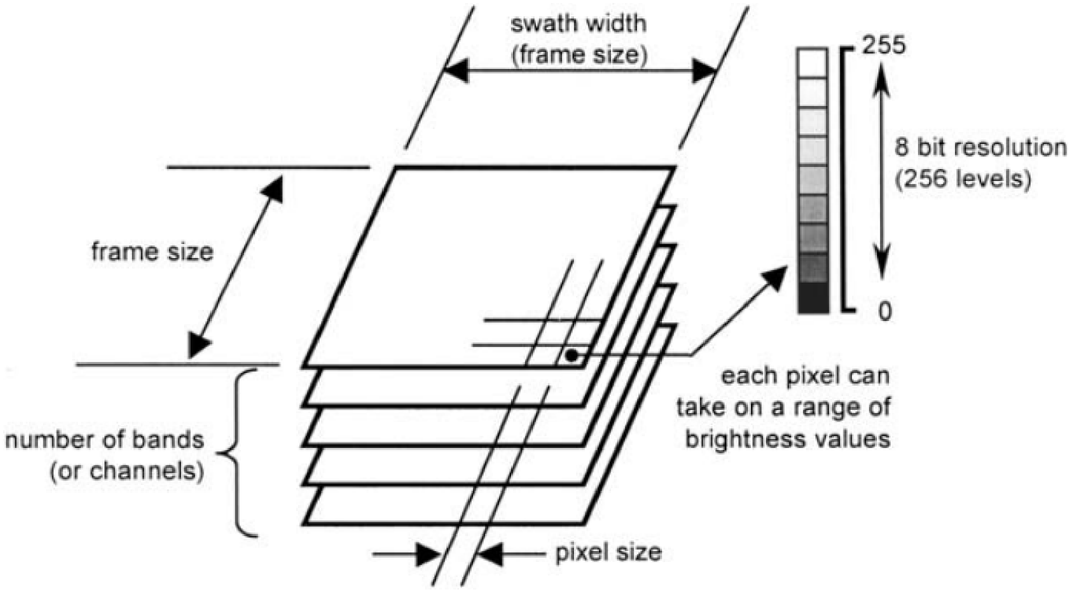

Figure 1.4: Technical characteristics of digital image data [3]

Fig. 1.4 represents the technical characteristics of the digital data form of reflections received by the remote sensing sensors.

1.4 Motivation

Hyperspectral image processing is one of the fast growing research field because of its various applications. Remotely sensed ground areas are classified into different groups of object by classification of the hyperspectral image. For this purpose, feature extraction and selecting only relevant features before classification is an important task to achieve high classification accuracy. So effective subspace detection is very important for classification of hyperspectral images. Few motivation for this research is given below:

- Extract only relevant features and find an effective subspace

- Address curse of dimensionality [6]

- Improve classification accuracy

1.5 Objectives

The main objective of this research work is to illustrate a target class oriented subspace detection method (TCOSD) which is a combination of feature extraction using principal component analysis (PCA) and relevant feature selection by normalized mutual information (NMI) to achieve a better classification accuracy. Objectives of this research will be:

- Create a method of feature extraction by removing correlation among bands.
- Create an approach to select relevant features after feature extraction.
- Create a method with collaboration of feature mining and classification.
- Improving classification accuracy.

1.6 Outline

This report is organized in 5 chapters discussing all related topics that may be helpful in reproducing a subspace method for effective ground object detection and identification.

- **Chapter 1:** A short overview of the whole research work and topics related to this research is discussed in this chapters.
- **Chapter 2:** Basics of hyperspectral remote sensing, sensors, applications and problems of hyperspectral images are discussed in this chapter.
- **Chapter 3:** This chapter is focused on the current literature of this research field. Recent and relevant feature mining techniques are discussed in this chapter.
- **Chapter 4:** Features extraction based on PCA and relevant features selection using NMI, their implementation with experimental results on real hyperspectral image is discussed in this chapter. Details explanations, algorithm, flow chart and experimental results of our proposed target class oriented subspace detection method (TCOSD) is also discussed in this chapter. Chapter 4 is the main part of this thesis report book.
- **Chapter 5:** Conclusion of this research is noted in this chapter. Drawbacks of the proposed TCOSD method and possible ways of improvement is also discussed here.

Chapter 2

Hyperspectral Imaging

2.1 Introduction

The most significant recent breakthrough in remote sensing has been the development of hyperspectral sensors and software to analyze the resulting image data [3]. Fifteen years ago only spectral remote sensing experts had access to hyperspectral images or software tools to take advantage of such images. Over the past decade hyperspectral image analysis has matured into one of the most powerful and fastest growing technologies in the field of remote sensing. The ”hyper” in hyperspectral means ”over” as in ”too many” and refers to the large number of measured wavelength bands. Hyperspectral images are spectrally overdetermined, which means that they provide ample spectral information to identify and distinguish spectrally unique materials. Hyperspectral imagery provides the potential for more accurate and detailed information extraction than possible with any other type of remotely sensed data.

2.2 Spectral Image Basics

To understand the advantages of hyperspectral imagery, it may help to first review some basic spectral remote sensing concepts. We know that each photon of light has a wavelength determined by its energy level. Light and other forms of electromagnetic radiation are commonly described in terms of their wavelengths. For example, visible light has wavelengths between 0.4 and 0.7 microns, while radio waves have wavelengths greater than about 30 cm (Fig. 2.1).

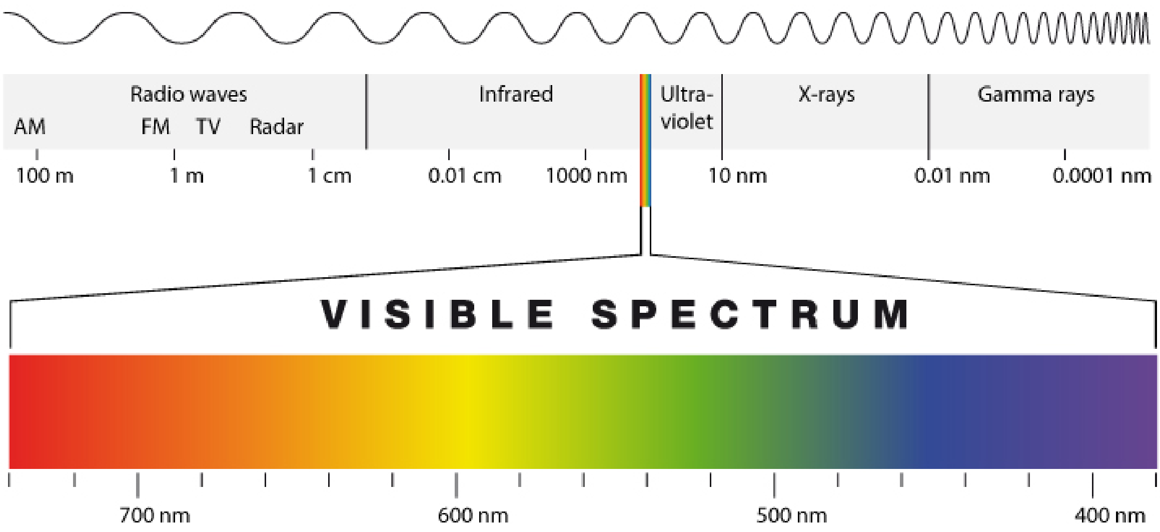

Figure 2.1: The electromagnetic spectrum [4]

Reflectance is the percentage of the light hitting a material that is then reflected by that material (as opposed to being absorbed or transmitted). A reflectance spectrum shows the reflectance of a material measured across a range of wavelengths (Fig. 2.2). Some materials will reflect certain wavelengths of light, while other materials will absorb the same wavelengths. These patterns of reflectance and absorption across wavelengths can uniquely identify certain materials.

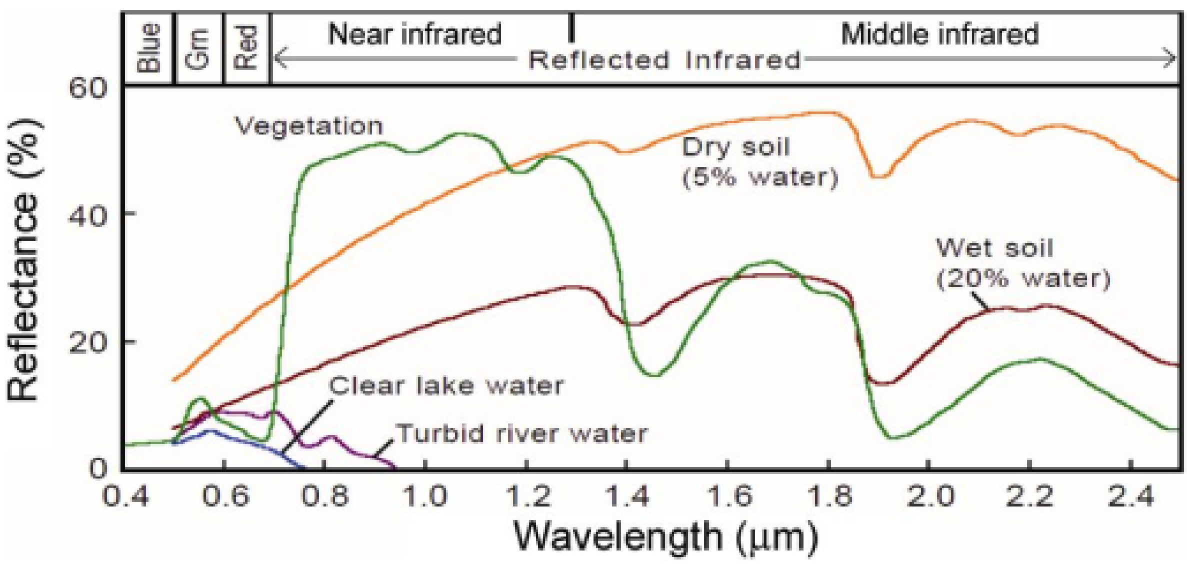

Figure 2.2: Reflectance spectra for several common Earth surface materials [4]

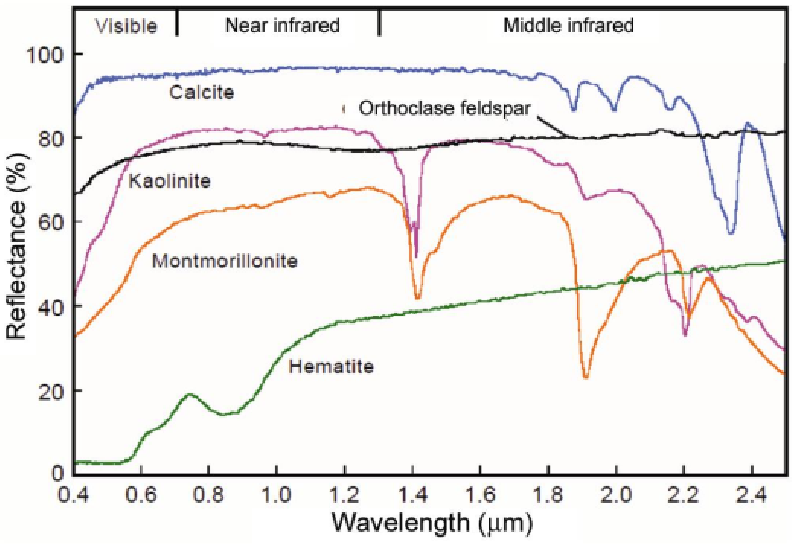

Figure 2.3: Reflectance spectra for important minerals [4]

Fig. 2.3 shows a reflectance spectra for some important minerals. Field and laboratory spectrometers usually measure reflectance at many narrow, closely spaced wavelength bands, so that the resulting spectra appear to be continuous curves (Fig. 2.2). When a spectrometer is used in an imaging sensor, the resulting images record a reflectance spectrum for each pixel in the image (Fig. 2.4).

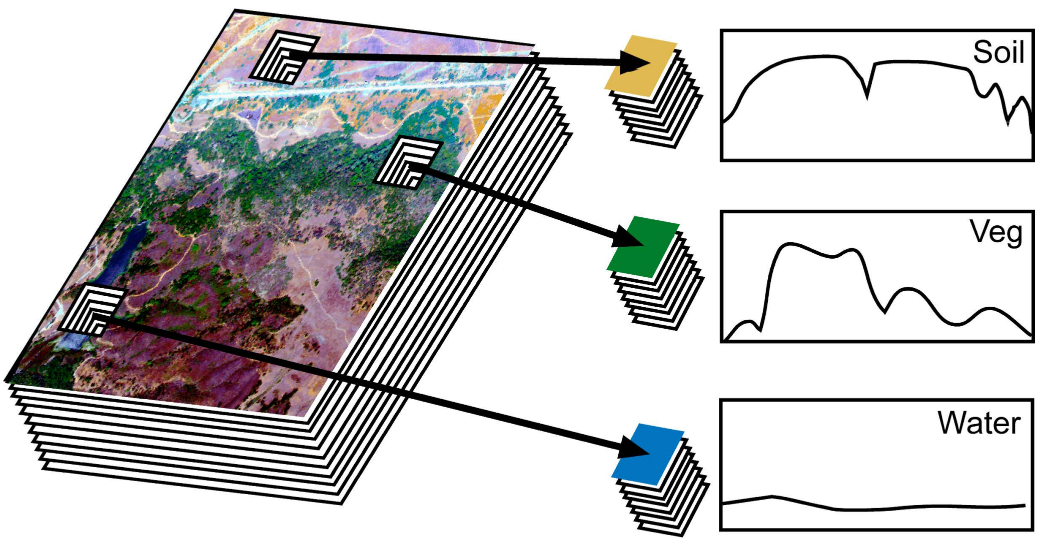

Figure 2.4: The concept of hyperspectral imagery. Image measurements are made at many narrow contiguous wavelength bands, resulting in a complete spectrum for each pixel. [4]

2.3 Hyperspectral Data

Most multispectral imagers (e.g., Landsat, SPOT, AVHRR) measure radiation reflected from a surface at a few wide, separated wavelength bands. Most hyperspectral imagers, on the other hand, measure reflected radiation at a series of narrow and contiguous wavelength bands. When we look at a spectrum for one pixel in a hyperspectral image, it looks very much like a spectrum that would be measured in a spectroscopy laboratory. This type of detailed pixel spectrum can provide much more information about the surface than a multispectral pixel spectrum.

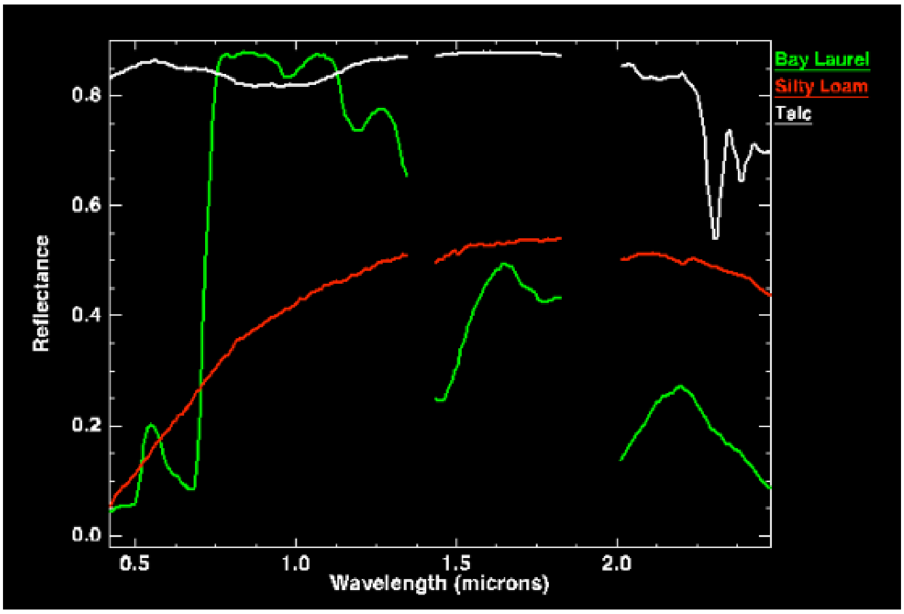

Figure 2.5: Reflectance spectra of the three materials as they would appear to the hyperspectral AVIRIS sensor. The gaps in the spectra are wavelength ranges at which the atmosphere absorbs so much light that no reliable signal is received from the surface. [4]

Although most hyperspectral sensors measure hundreds of wavelengths, it is not the number of measured wavelengths that defines a sensor as hyperspectral. Rather it is the narrowness and contiguous nature of the measurements. For example, a sensor that measured only 20 bands could be considered hyperspectral if those bands were contiguous and, say, 10 nm wide. If a sensor measured 20 wavelength bands that were, say, 100 nm wide, or that were separated by non-measured wavelength ranges, the sensor would no longer be considered hyperspectral. Standard multispectral image classification techniques were generally developed to classify multispectral images into broad categories. Hyperspectral imagery provides an opportunity for more detailed image analysis. For example, using hyperspectral data, spectrally similar materials can be distinguished [11] and sub-pixel scale information can be extracted. To fulfill this potential, new image processing techniques have been developed.

2.4 Imaging Sensors

The Landsat earth resources satellite system was the first designed to provide near global coverage of the earth's surface on a regular and predictable basis. The first three Landsats had identical orbit characteristics, as summarised in Table 2.1. The orbits were near polar and sun synchronous – i.e., the orbital plane precessed about the earth at the same rate that the sun appears to move across the face of the earth. In this manner data was acquired at about the same local time on every pass.

Table 2.1: Landsat 1, 2, 3 orbit characteristics

Orbit:	Sun synchronous,near polar and nominal 9:30 am descending equatorial crossing; inclined at about 99deg to equator
Altitude:	920km (570 mi)
Period:	103 min
Repeat Cycle:	14 orbits per day over 18 days (251 revolutions)

Most past and current hyperspectral sensors have been airborne, with two recent exceptions: NASA's Hyperion sensor on the EO-1 satellite, and the U.S. Air Force Research Lab's FTHSI sensor on the MightySat II satellite. Several new spacebased hyperspectral sensors have been proposed recently. Unlike airborne sensors, space-based sensors are able to provide near global coverage repeated at regular intervals. Therefore, the amount of hyperspectral imagery available should increase significantly in the near future as new satellite-based sensors are successfully launched.

Table 2.2: Current and Recent Hyperspectral Sensors and Data Providers

Sensors	Manufacturer	Number of Bands	Spectral Range
AVIRIS (Airborne Visible Infrared Imaging Spectrometer)	NASA Jet Propulsion Lab makalu.jpl.nasa.gov/	224	0.4 to 2.5 μm
HYDICE (Hyperspectral Digital Imagery Collection Experiment)	Naval Research Lab	210	0.4 to 2.5 μm
FTHSI on MightySat II	Air Force Research Lab www.vs.afrl.af.mil /TechProgs/MightySatII	256	0.35 to 1.05 μm
Hyperion on EO- 1	NASA Goddard Space Flight Center eo1.gsfc.nasa.gov	220	0.4 to 2.5 μm
PROBE- 1 Earth Search Sciences Inc.	www.earthsearch .com	128	0.4 to 2.5 μm
CASI (Compact Airborne Spectrographic Imager)	ITRES Research Limited www.itres.com up to	228	0.4 to 1.0μm

2.5 Application of Hyperspectral Images

Multi spectral images like Landsat thematic Mapper and SPOT XS produce few broad wavelength bands but the hyperspectral image sensor will produce narrower wavelength bands. These bands can range from 100 to 200 or more. These measurement make it possible to derive continuous spectrum for each image cell. After sensor is adjusted, terrain and atmospheric condition are applied, then the results is verified against the collected spectrum value to categorize the type of vegetation or minerals or other features. Hyperspectral images contain ton of information, surface information and its spectrum behavior should be understand deeply and how it related to the hyperspectral images. This type of image are finding their importance in different fields as before it was just used for remote sensing application. Here are few applications of hyperspectral images.

1. **Ground Object Detection:** Hyperspectral images are blessing in the field of ground object detection. Ground object can be detected more accurately with the help of hyperspectral image analysis than all other conventional methods. In this research work, a subspace detection method of hyperspectral data is proposed for ground object identification.

2. **Remote Sensing:** In remote sensing technology it is very important to distinguish earth surface features, each features have different spectrum band. Multi spectral satellite can capture image up few bands for example Landsat 7 have 8 bands. But multi spectral imaging satellite can capture earth surface in more than 200 bands which helps scientist to differentiate objects that were not possible in multi spectral imaging because of spectral resolution.

3. **Seed Viability Study:** By using the hyperspectral image and plotting the reflectance spectrum one can conclude that whether those seed are viable or not viable. Seed might be looking same through naked eyes but its viability will be trace down by the hyperspectral image.

4. **Biotechnology:** Hyperspectral technology has become popular in the biological and medical applications. It is easy and quick to acquire the data that can be used in the laboratory. Mostly they are used in the study of the wound analysis, fluorescence microscopy, and cell biology.

5. **Environmental Monitoring:** Hyperspectral imaging is becoming widely popular for tracking changes in the environment. It is commonly used to understand surface CO2 emissions, map hydrological formations, tracking pollution levels, and more.

6. **Food:** Hyperspectral imaging is widely used in the food sector. It is used in different discipline of food industry, bruise detection in apples, freshness of the fish, citrus fruit inspection, distribution of sugar in melons, and sorting of potatoes. For example apple bruise is not visible on the early stage and it takes few days to show dark color mark. In this type of scenario hyperspectral imaging techniques can be used to track the early stage of bruise for the quality control.

7. **Pharmaceuticals:** Hyperspectral imaging technique is widely used to enhance the quality control. It is used widely to control the counterfeit or illegal drugs, managing the packaging of medicine and mixing of the powder.

8. **Medical Diagnose:** Early disease detection and disease prevention are very important for the healthy body. Hyperspectral imaging technology can be used to detect the early of various types of cancer or retinal disease.

2.6 Challenges of Hyperspectral Data Processing

Optical sensing has come a long way from grayscale to multispectral and now to hyperspectral images. The advances in imaging hardware over recent decades have enabled availability of high spatial, spectral, and temporal resolution imagery for a variety of applications. Hyperspectral imagery, also called imaging spectroscopy, entails acquiring images using a large number (typically a few hundreds) of narrow and often contiguous spectral bands, covering a wide range of the electromagnetic spectrum from the visible to the infrared regions. Compared to conventional color imagery (with 3 spectral bands covering the red, green and blue wavelengths, respectively), or compared to conventional multispectral imagery (typically a few spectral bands), hyperspectral data provide a very fine spectral characterization of the sensed materials, which facilitates their detection and characterization. Advances in hardware to acquire hyperspectral imagery have made such data easily accessible to a wide variety of application domains, but have also created unique challenges for researchers working on algorithms for the representation, exploitation, and analysis of such data. Unfortunately, this is often a double-edged sword. A direct consequence of the dense spectral sampling implies that each measurement corresponds to a vector with several hundreds of values. Consequently, the data are evolving in a vector space with several hundreds of dimensions. Traditional information-processing techniques cannot be used to process such data effectively. While it is a curse from an analytical, theoretical, and statistical point of view, the very high dimensionality of the data is also a blessing. But the main challenge of hyperspectral image is its highly correlated huge data set. On the other hand, this large number of spectral bands has a direct impact on the required computational cost for classification. Also, as the feature space dimension increases, if the size of the training data does not grow correspondingly, a reduction in the classification accuracy of the testing data is observed due to poor parameter estimation of the supervised classifier. This effect is known as the Hughes phenomenon [6]. Challenges of hyperspectral images are given below:

- Curse of dimensionality.
- High correlation.
- Irrelevant information for classification.

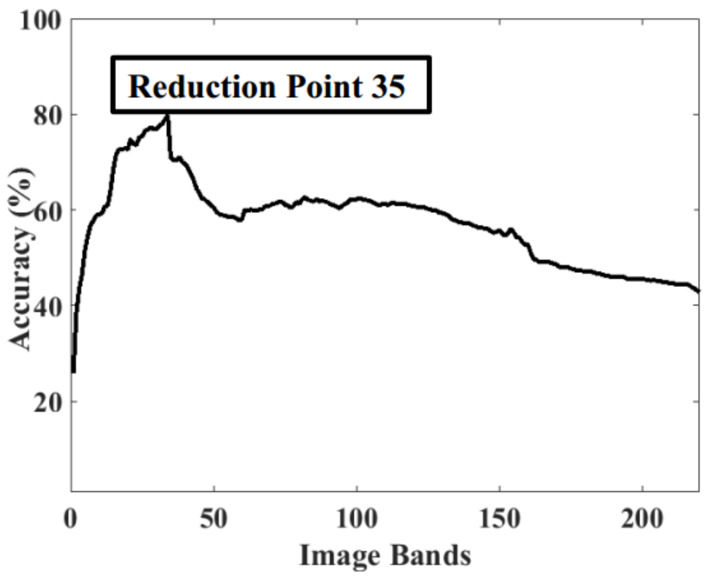

Figure 2.6: Curse of dimensionality [5]

In Fig. 2.6 the problem of hyperspectral data, curse of dimensionality is present in image band number 35. From image band 35, the accuracy of the classification task is gradually decreasing as the train sample of the experiment remains constant.

Chapter 3

Literature Review

3.1 Introduction

Relevant feature identification has become an essential task to apply data mining algorithms effectively in real-world scenarios. Therefore, many feature selection methods have been proposed to obtain the relevant feature or feature subsets in the literature to achieve their objectives of classification and clustering. The amount of high-dimensional data that exists and is publically available on the internet has greatly increased in the past few years. Therefore, machine learning methods have difficulty in dealing with the large number of input features, which is posing an interesting challenge for researchers. In order to use machine learning methods effectively, pre-processing of the data is essential. Feature selection is one of the most frequent and important techniques in data pre-processing [12], and has become an indispensable component of the machine learning process. It is also known as variable selection, attribute selection, or variable subset selection in machine learning and statistics. It is the process of detecting relevant features and removing irrelevant, redundant, or noisy data. This process speeds up data mining algorithms, improves predictive accuracy, and increases comprehensibility. Irrelevant features are those that provide no useful information, and redundant features provide no more information than the currently selected features.

3.2 Feature Mining

Hyperspectral sensors record the reflectance from the Earth's surface over the full range of solar wavelengths with high spectral resolution. The resulting high-dimensional data contain rich information for a wide range of applications. However, for a specific application, not all the measurements are important and useful. The original feature space may not be the most effective space for representing the data. Feature mining [13], which includes feature generation, feature selection (FS), and feature extraction (FE) [14], is a critical task for hyperspectral data classification. Significant research effort has focused on this issue since hyperspectral data became available in the late 1980s. The feature mining techniques which have been developed include supervised and unsupervised, parametric and nonparametric, linear and nonlinear methods, which all seek to identify the informative subspace. In the process of feature mining, irrelevant and redundant features or noise in the data may be hinder in many situations, because they are not relevant and important with respect to the class concept such as microarray data analysis . When the number of samples is much less than the features, then machine learning gets particularly difficult, because the search space will be sparsely populated. Therefore, the model will not able to differentiate accurately between noise and relevant data . There are two major approaches to feature selection. The first is Individual Evaluation, and the second is Subset Evaluation. Ranking of the features is known as Individual Evaluation . In Individual Evaluation, the weight of an individual feature is assigned according to its degree of relevance. In Subset Evaluation, candidate feature subsets are constructed using search strategy. The general procedure for feature selection has four key steps as shown in Figure 3.1.

- Subset Generation
- Evaluation of Subset
- Stopping Criteria
- Result Validation

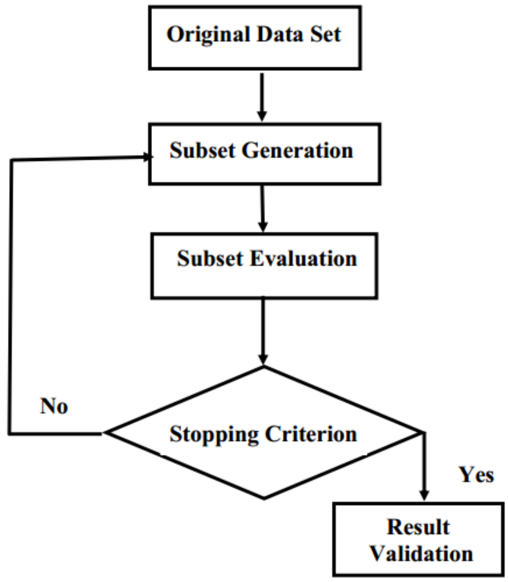

Figure 3.1: Four key steps for the feature selection process

Subset generation is a heuristic search in which each state specifies a candidate subset for evaluation in the search space. Two basic issues determine the nature of the subset generation process. First, successor generation decides the search starting point, which influences the search direction. To decide the search starting points at each state, forward, backward, compound, weighting, and random methods may be considered . Second, search organization is responsible for the feature selection process with a specific strategy, such as sequential search, exponential search or random search . A newly generated subset must be evaluated by a certain evaluation criteria. Therefore, many evaluation criteria have been proposed in the literature to determine the goodness of the candidate subset of the features. Base on their dependency on mining algorithms, evaluation criteria can be categorized into groups: independent and dependent criteria . Independent criteria exploit the essential characteristics of the training data without involving any mining algorithms to evaluate the goodness of a feature set or feature. And dependent criteria involve predetermined mining algorithms for feature selection to select features based on the performance of the mining algorithm applied to the selected subset of features. Finally, to stop the selection process, stop criteria must be determined. Feature selection process stops at validation procedure. It is not the part of feature selection process, but feature selection method must be validate by carrying out different tests and comparisons with previously established results or comparison with the results of competing methods using artificial datasets, real world datasets, or both. The relationship between the inductive learning method and feature selection algorithm infers a model. There are three general approaches for feature selection. First, the Filter Approach exploits the general characteristics of training data with independent of the mining algorithm. Second, the Wrapper Approach explores the relationship between relevance and optimal feature subset selection [15]. It searches for an optimal feature subset adapted to the specific mining algorithm. And third, the Embedded Approach is done with a specific learning algorithm that performs feature selection in the process of training.

3.3 Related Works

Many feature selection methods have been proposed in the literature, and their comparative study is a very difficult task. Without knowing the relevant features in advance of the real data set, it is very difficult to find out the effectiveness of the feature selection methods, because data sets may include many challenges such as the huge number of irrelevant and redundant features, noisy data, and high dimensionality in term of features or samples. Therefore, the performance of the feature selection method relies on the performance of the learning method. There are many performance measures mentioned in the literature such as accuracy, computer resources, ratio of feature selection, etc. Most researchers agree that there is no so-called ”best method”. Therefore, the new feature selection methods are constantly increasing to tackle the specific problem (as mentioned above) with different strategy.

- To ensure a better behavior of feature selection using an ensemble method
- Combining with other techniques such as tree ensemble and feature extraction [16]
- Reinterpreting existing algorithms
- Creating a new method to deal with still-unresolved problems
- To combine several feature selection methods

Dimensionality reduction prior to classification is advantageous in hyperspectral data analysis because the dimensionality of the input space greatly affects the performance of many supervised classification methods [17]. Further, there is a high likelihood of redundancy in the features and it is possible that some features contain less discriminatory information than others. Moreover, the high-dimensionality imposes requirements for storage space and computational load. The analysis in [6] supports this line of reasoning and suggests that feature selection may be a valuable procedure in preprocessing hyperspectral data for classification by the widely used SVM classifier. Two preferred preprocessing technique in current literature are given below:

- Feature Extraction
- Feature Selection

In hyperspectral image analysis, feature selection is preferred over feature extraction for dimensionality reduction [16], [17]. Feature extraction methods involve transforming the data and hence, crucial and critical information may be compromised and distorted. In contrast, feature selection methods strive to discover a subset of features which capture the fundamental characteristics of the data, while possessing sufficient capacity to discriminate between classes. Hence, they have the advantage of preserving the relevant original information of the data. There are various studies which establish the usefulness of feature selection in hyperspectral data classification. There are various feature selection methods for hyperspectral data such as the SVM Recursive Feature Elimination (SVM-RFE), Correlation based Feature Selection(CFS), Minimum Redundancy Maximum Relevance(MRMR) [18] feature selection and Random Forests [19] in current literature. In [14], a band prioritization scheme based on Principal Component Analysis (PCA) and classification criterion is presented. Two main feature extraction technique in this field of hyperspectral image analysis is given below:

- Principal Component Analysis (Unsupervised approach)
- Linear Discriminant Analysis (Supervised approach)

Mutual information is a widely used quantity in various feature selection methods. In [20] and [21] uses mutual information to extract most informative features from the original huge data set. In a general setting, features are ranked based on the mutual information between the spectral bands and the reference map(also known as the ground truth). In [9], mutual information is computed using the estimated reference map obtained by using available a priori knowledge about the spectral signature of frequently-encountered materials. In [5], normalized mutual information is computed between class reference and the bands obtained via PCA to select relevant features for better classification accuracy.

3.4 Feature Relevancy and Redundancy

The optimal feature subset is a subset of all relevant features. Therefore, the relevance of the features must be properly defined according to their relevance. In the literature, features are classified by their relevancy with three qualifiers: irrelevant, weakly relevant, and strongly relevant. A graphical representation is shown in Figure 3.2. Many definitions have been proposed to answer a question "relevant to what?". In [5], a greedy search approach is applied to select the most relevant and less redundant features to the class labels.

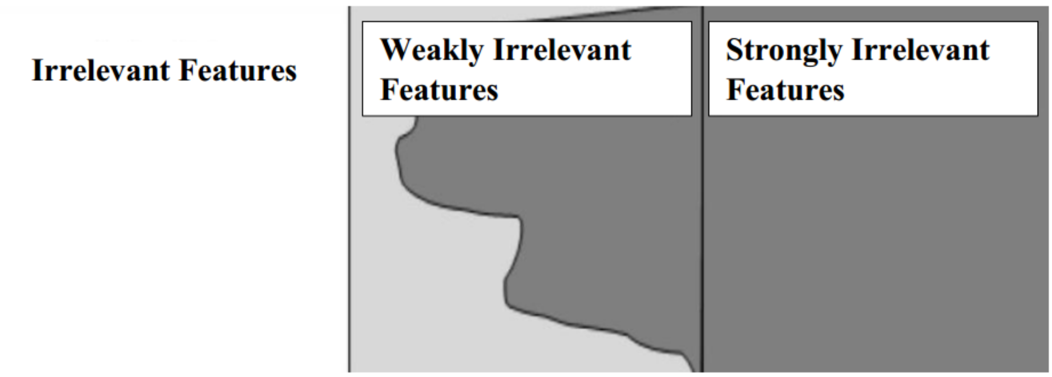

Figure 3.2: A view of feature relevance

3.5 General Approach for Feature Selection

The feature selection methods are typically presented in three classes based on how they combine the selection algorithm and the model building.

1. **Filter method:** Filter type methods select variables regardless of the model. They are based only on general features like the correlation with the variable to predict. Filter methods suppress the least interesting variables. The other variables will be part of a classification or a regression model used to classify or to predict data. These methods are particularly effective in computation time and robust to overfitting.

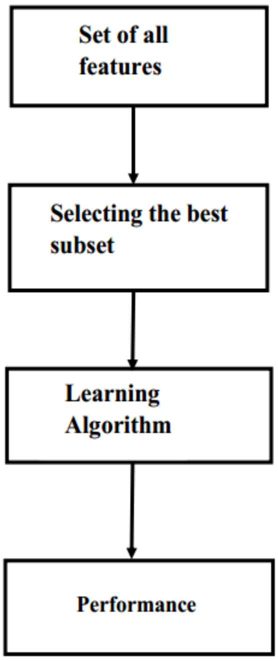

Figure 3.3: Filter Method for feature selection

However, filter methods tend to select redundant variables because they do not consider the relationships between variables. Therefore, they are mainly used as a pre-process method.

2. **Wrapper method:** Wrapper methods evaluate subsets of variables which allows, unlike filter approaches, to detect the possible interactions between variables. The two main disadvantages of these methods are :

   - The increasing overfitting risk when the number of observations is insufficient.
   - The significant computation time when the number of variables is large.

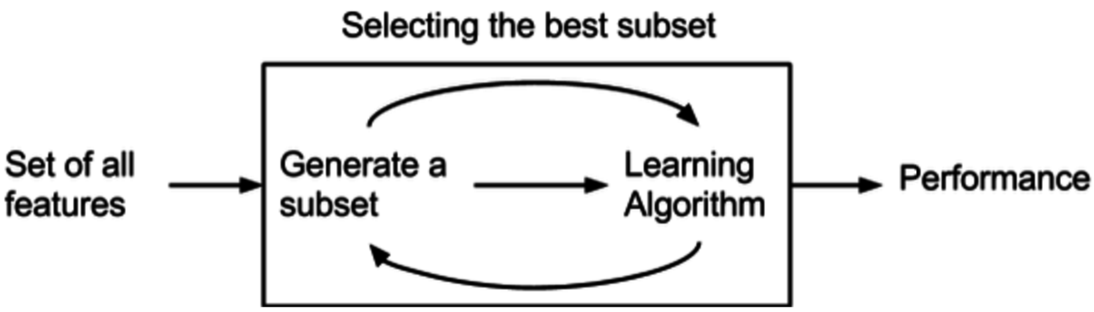

Figure 3.4: Wrapper Method for Feature selection

3. **Embedded method:** Embedded methods have been recently proposed that try to combine the advantages of both previous methods. A learning algorithm takes advantage of its own variable selection process and performs feature selection and classification simultaneously.

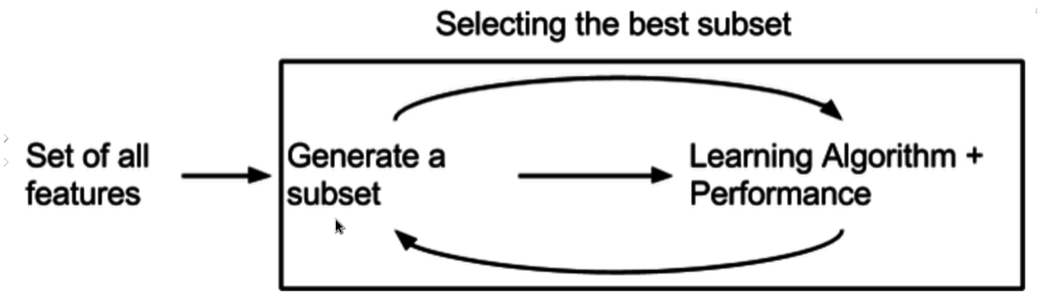

Figure 3.5: Embedded Method for Feature selection

3.6 Feature Mining for Hyperspectral Image

Hyperspectral sensors record the reflectance from the Earth's surface over the full range of solar wavelengths with high spectral resolution. The resulting high dimensional data contain rich information for a wide range of applications. However, for a specific application, not all the measurements are important and useful. The original feature space may not be the most effective space for representing the data. Feature mining, which includes feature generation, feature selection (FS), and feature extraction (FE), is a critical task for hyperspectral data classification.

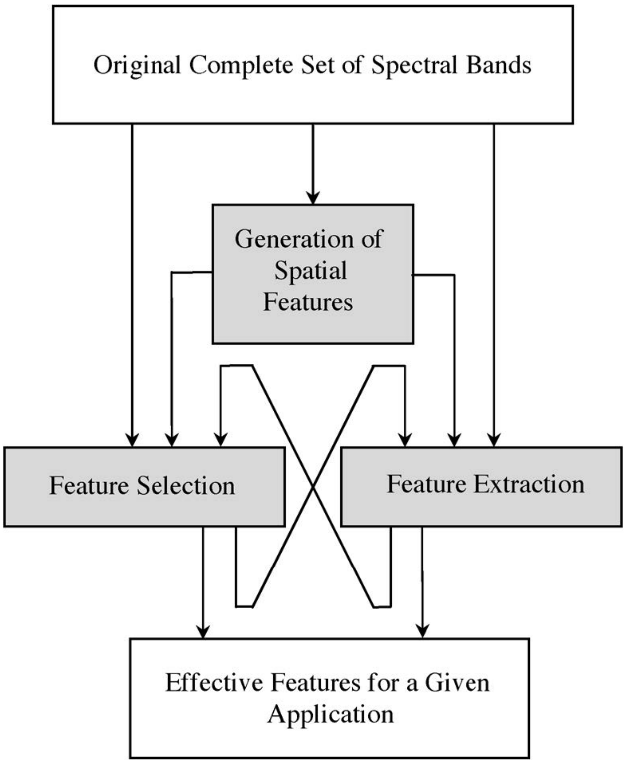

Figure 3.6: Possible operations for feature mining (only one path is used for any specific approach). [12]

Significant research effort has focused on this issue since hyperspectral data became available in the late 1980s. The feature mining techniques which have been developed include supervised and unsupervised, parametric and nonparametric, linear and nonlinear methods, which all seek to identify the informative subspace.

3.7 Classification Techniques

Digital image classification techniques group pixels to represent land cover features. Land cover could be forest, urban, agricultural and other types of features. There are two main image classification techniques.

- Unsupervised Classification
- Supervised Classification

3.7.1 Unsupervised Classification

Pixels are grouped based on the reflectance properties of pixels. These groupings are called ”clusters”. The user identifies the number of clusters to generate and which bands to use. With this information, the image classification software generates clusters. There are different image clustering algorithms such as K-means and ISODATA. Unsupervised classification is a method which examines a large number of unknown pixels and divides into a number of classed based on natural groupings present in the image values. unlike supervised classification, unsupervised classification does not require analyst-specified training data. The basic premise is that values within a given cover type should be close together in the measurement space (i.e. have similar gray levels), whereas data in different classes should be comparatively well separated (i.e. have very different gray levels) (PCI, 1997; Lillesand and Kiefer, 1994; Eastman, 1995) The classes that result from unsupervised classification are spectral classed which based on natural groupings of the image values, the identity of the spectral class will not be initially known, must compare classified data to some from of reference data (such as larger scale imagery, maps, or site visits) to determine the identity and informational values of the spectral classes. Thus, in the supervised approach, to define useful information categories and then examine their spectral separability; in the unsupervised approach the computer determines spectrally separable class, and then define their information value. (PCI, 1997; Lillesand and Kiefer, 1994)

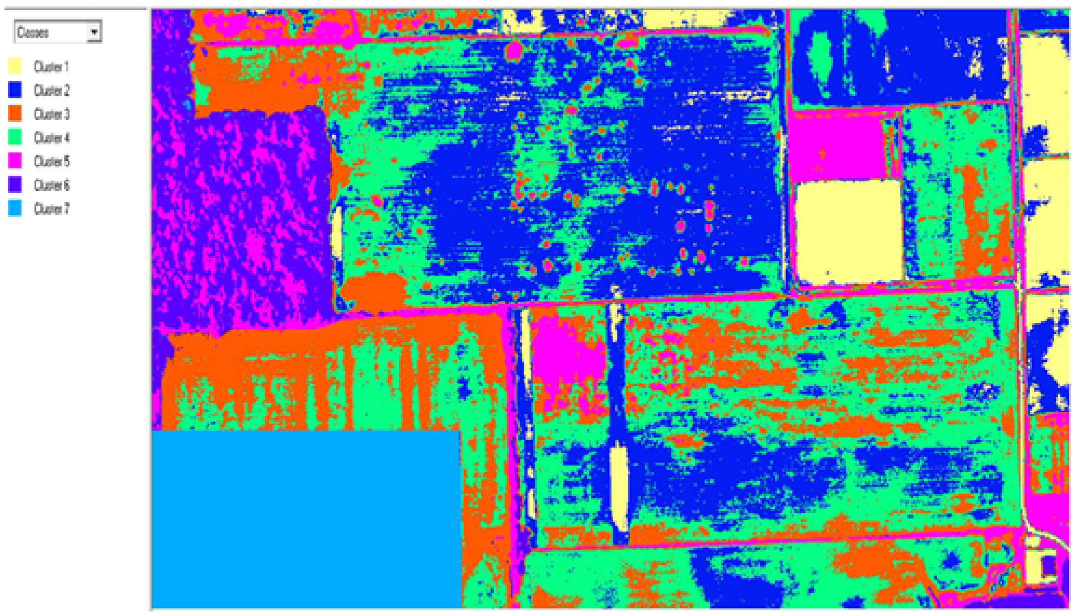

Figure 3.7: Unsupervised Classification

Unsupervised classification is becoming increasingly popular in agencies involved in long term GIS database maintenance. The reason is that there are now systems that use clustering procedures that are extremely fast and require little in the nature of operational parameters. Thus it is becoming possible to train GIS analysis with only a general familiarity with remote sensing to undertake classifications that meet typical map accuracy standards. With suitable ground truth accuracy assessment procedures, this tool can provide a remarkably rapid means of producing quality land cover data on a continuing basis.

3.7.2 Supervised Classification

The user selects representative samples for each land cover class in the digital image. These sample land cover classes are called ”training sites”. The image classification software uses the training sites to identify the land cover classes in the entire image. The classification of land cover is based on the spectral signature defined in the training set. The digital image classification software determines each class on what it resembles most in the training set. The common supervised classification algorithms are maximum likelihood [22] and support vector machine [23].

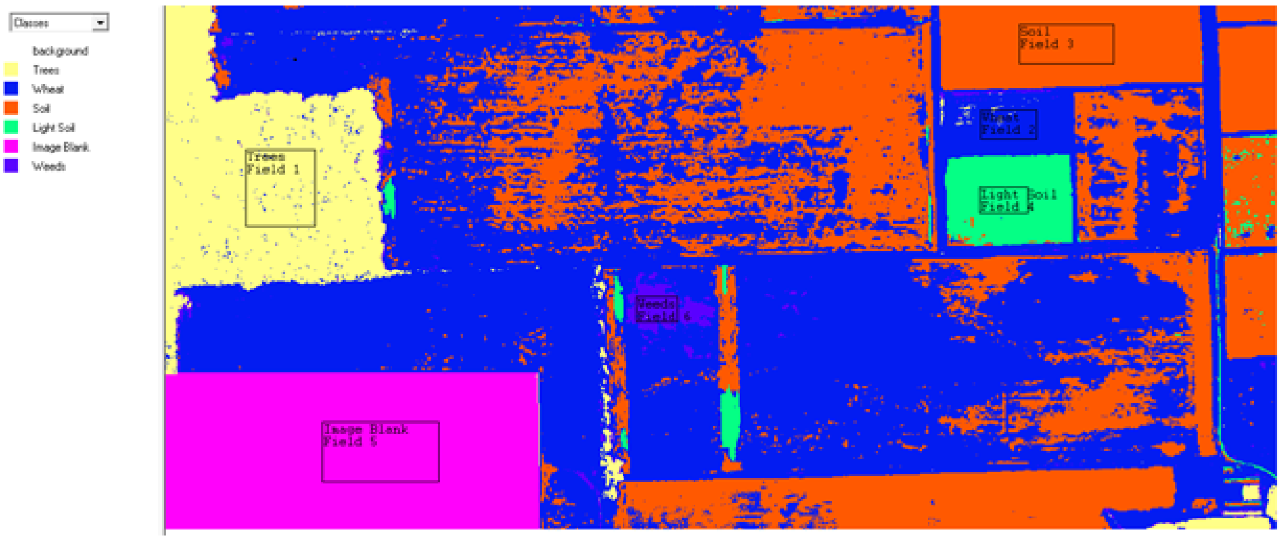

Figure 3.8: Supervised Classification

Chapter 4

Subspace Detection and Classification

4.1 Introduction

Over the past few decade hyperspectral image sensor has become one of the most powerful and fastest growing technologies in the field of remote sensing [3]. Hyperspectral sensors collect land cover reflectance in many narrow contiguous spectral bands and form a large data cube [3]. For instance, the hyperspectral imaging sensors like NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) which simultaneously measures 224 bands with a fine resolution of 0.01 μm in the spectral range of visible to near infrared. This data cube contains rich information for wide range of application including effective land cover detection and classification [24]. This large data also presents some challenges while classifying into constituent objects. A major classification challenge is when the training data set is small in compare to feature space dimensions, the classification starts to decrease. This problem is known as the curse of dimensionality or Hughes phenomenon [16] [25] [26]. On the other hand, some bands are highly correlated and can be removed. In addition not all the bands are important for a specific classification problem. Therefore, an efficient and reliable technique to extract most informative and relevant features for classification is a major interest in current literature. The aforementioned feature reduction problem can be addressed using two basic approaches [12]. One is the extraction of limited number of features after mapping to a new feature space based on some criteria like variance [27, 28] [8] [7]. For instance, PCA reduces the dimensionality of the input hyperspectral image which mitigates the curse of dimensionality. In PCA, an orthogonal transformation is performed to uncorrelate the input data set and the components are selected with higher variance. However, PCA doesn't consider the class structure of input data and depends solely on the global variance. Therefore, some noisy bands with high variance can be selected as a resultant feature. This limits the performance of PCA in classification task. An alternative approach to overcome the limitation of PCA and to find an effective subspace, some sorts of feature selection on PCA can be applied [29]. Normalized Mutual Information (NMI) is one of the most popular feature selection technique [5, 9, 30].

Thus, feature extraction and feature selection gives a significant improvement in the classification accuracy. In some application of hyperspectral image only one class may be needed to identify at a time. For example, in land cover identification, if the objective is to identify forest land, we may not be interested in any other class except forest of the data set. In this research, a target class oriented feature selection method is proposed which uses normalized mutual information measure over PCA data to maximize the relevance of the input selected subset. Thus, feature selection over PCA based on target class samples also overcomes the limitation of the PCA. Finally, the Kernel Support Vector Machine (KSVM) classifier is used to assess the effectiveness of the proposed method.

Feature mining generally refers to transform a high dimensional correlated dataset into a low dimensional uncorrelated data space. It can be done by following two steps:

- Feature extraction
- Feature selection

4.2 Feature Extraction

Principal Component Analysis produces a new set of reduced and uncorrelated features. PCA works well on hyperspectral image as the neighboring bands of hyperspectral images are highly correlated and often convey similar information about the land cover object. PCA performs the eigenvalue decomposition of covariance matrix of the input hyperspectral image. The new transformed uncorrelated images is called principal component (PCs). Let $\mathbf{X}$ be the input data cube and $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be the bands of the image. The covariance matrix is calculated as:

$$cov(\mathbf{X}_2, \mathbf{X}_2) = \frac{\sum_{j=1}^{n}(X_1^j - \mu_1)(X_2^j - \mu_2)}{n} \tag{4.1}$$

From this covariance matrix eigenvalue and eigenvector is determined. The eigenvectors are ordered according to the highest eigenvalue. This first few eigenvectors are selected as principal component as they contain most of the variance of the data set. The new features can be obtained as:

$$\mathbf{Y} = \mathbf{E}^T\mathbf{X} \tag{4.2}$$

Where $\mathbf{Y}$ is the transformed PCA data and $\mathbf{E}^T$ is the matrix of normalized eigenvectors ordered by the higher eigenvalues. The first few components can be used as the extracted features.

4.3 Feature Selection Based on Normalized Mutual Information

Mutual information (MI) measures the amount of information one band of hyperspectral image holds about the other band [30]. Mutual information between two image bands $\mathbf{X}$ and $\mathbf{Y}$ can be formulated as below:

$$I(\mathbf{X},\mathbf{Y}) = \sum_{y=1}^{L_1}\sum_{x=1}^{L_2} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{4.3}$$

where x, y are the pixel values and L1, L2 are the maximum pixel values. P(x), P(y) is the marginal and P(x,y) is the joint probability distribution function of $\mathbf{X}$ and $\mathbf{Y}$ respectively [30]. A method to normalize mutual information [9] is available to give a new value between 0 and 1, where 0 is the value for the independent case and 1 for the identical or one-to-one relationship case [5]. Normalized mutual information of $\mathbf{X}$ and $\mathbf{Y}$ (NMI) is defined as follows:

$$\hat{I}(\mathbf{X},\mathbf{Y}) = \frac{I(\mathbf{X},\mathbf{Y})}{\sqrt{I(\mathbf{X},\mathbf{X})}\sqrt{I(\mathbf{Y},\mathbf{Y})}} \tag{4.4}$$

The aim of feature selection is to find only informative features k < n for a dataset. To find uncorrelated informative features, Normalized Mutual Information is applied over PCA images. If the reference image is $\mathbf{C}$ and PCA image is $\mathbf{Y}$ then the NMI is determined between $\mathbf{Y}$ and $\mathbf{C}$ as follows:

$$\hat{I}(\mathbf{Y},\mathbf{C}) = \frac{I(\mathbf{Y},\mathbf{C})}{\sqrt{I(\mathbf{Y},\mathbf{Y})}\sqrt{I(\mathbf{C},\mathbf{C})}} \tag{4.5}$$

The NMI from equation (4.5) has been used to determine the relevancy of selected subset. The 2nd term of Eq. (4.6) ensures that the selected features do not have redundancy among them.

$$\hat{R}(\mathbf{Y}_i,k) = \hat{I}(\mathbf{Y}_i,\mathbf{C}) - \frac{1}{k}\sum_{\mathbf{Y}=S_k}\hat{I}(\mathbf{Y}_i,\mathbf{Y}), \mathbf{Y}_i \notin S_k \tag{4.6}$$

4.4 Proposed Subspace Detection Method

The proposed method is a target class oriented subspace detection technique which select relevant features for each class of the dataset using principal component analysis (PCA) as feature extraction [7] and normalized mutual information (nMI) as features selection [5]. At first, the unsupervised feature extraction technique PCA is applied to find uncorrelated dataset in smaller data space. Then normalized mutual information (nMI) to maximize general relevance and minimize redundancy on the components obtained via PCA is applied for feature selection [31]. Here we consider one class of the dataset at a time and make all others as the background class. For each class of the dataset feature selection method is applied to obtain a relevant feature space to classify that class. Finally kernel support vector machine (SVM) is applied for classification purpose of the data set [10]. Here first SVM use those selected features of each class to train a model for that specific class and test data to find test result for that specific class. RBF kernel [32] is used in SVM as the relationship between class label C and datase X_i is nonlinear. K-fold cross validation is applied to find correct parameter C and γ of KSVM for each target class of the dataset. Finally average of those accuracy for each class is the final classification accuracy.

4.5 Experimental Procedure

4.5.1 Feature Extraction

- **Step 1: Get input data.**

  For processing hyperspectral image, I have converted 3 dimensional data cube into 2 dimensional spaces which I will use as input data for PCA.

  Input data = [Band1 image,band2 image, ...,Band 220 image]

- **Step 2: Calculating the mean of each image band and then subtracting the mean from the original.**

  Calculating the mean across each dimension and then subtracting along the dimensions. This produce images with mean value of 0. This mean subtracted data is represented by adjusted data.

  Adjusted data = [Band1 image-mean1, band2 image-mean2 , ...,Band220 image-mean220]

- **Step 3: Calculating the covariance of the mean subtracted image.**
  In the 3rd step, I have calculated the covariance of the data got from step 2. From the non-diagonal value of the covariance matrix, I got the information about how the dimensions of data are related to each other.

- **Step 4: Calculating the eigenvector and eigenvalues of the covariance matrix.**
  From the covariance matrix, I have calculated the eigenvectors and eigenvalues matrix. This eigenvectors information is very important as they provide information about the characteristics of input data. Eigenvalues indicate the variance and eigenvectors indicate to which direction. For 220 dimensions I have got 220 eigenvalues and 220 eigenvectors. Among the 220 vectors one provides a best fitting of the input data.

- **Step 5: Forming feature vector by choosing the significant component.**
  The eigenvectors are rearranged according to the higher value of the eigenvalues. The eigenvector with the highest eigenvalue is the principal component of the data set and it has the largest relationship with the input data and dimension. This component contains the maximum feature of the input patterns. A subset of the eigenvectors with significant eigenvalues collected from all the eigenvectors is called feature vector.

- **Step 6: Calculating the reduced data set and getting back the original data.**
  This is the final step of principal component analysis. From the chosen components defined as feature vectors I have got the data set with reduced dimensionality from the following equation.
  Final data (reduced in dimensionality) = Feature vector $\times$ Adjusted dataset
  Where, the adjusted dataset means dataset with mean value of 0. The original data is obtained by the following equation. Data obtained from the following equation will be slightly different from the original input data as some of the noise of the input data is removed by PCA when subset of the eigenvectors are used.
  Original Dataset = ($Featurevector^T$ $\times$ Final data) + mean of each band image.

4.5.2 NMI Based Feature Selection

- **Step 1: Normalizing dataset into a range**
  In my analysis I have normalized the dataset in 1 to 64 range. Then calculated a 64 $\times$

64 probability table, from which I have calculated the marginal probability and the joint probability of variable(s).

- **Step 2: Calculating mutual information between class labels and bands**
  To select or rank best features for classification of the dataset I have calculated mutual information according to equation (4.3) between class labels and each band of the dataset.

- **Step 3: Calculating normalized mutual information**
  To normalize mutual information values between 0 to 1 I have used equation (4.6). Then I have used this normalized mutual information values to select effective features for the classification of the dataset.

Original Dataset Plus NMI Approach

- **Step 1: Get input data**
  In this analysis I have used original hyperspectral data set. I have taken all 220 bands into consideration. Then I generate class labels of the sample. I have taken fourteen classes which label from 1 to 14.

- **Step 2: Normalizing dataset into a range**
  I have normalized the dataset in 1 to 64 range. Then calculated a 64×64 probability table from which I have calculated the marginal probability and the joint probability of variable(s).

- **Step 3: Calculating mutual information between class labels and bands**
  To select or rank best features for classification of the dataset I have calculated mutual information according to equation (4.3) between class labels and each band of the dataset.

- **Step 4: Calculating normalized mutual information**
  To normalize mutual information values between 0 to 1 I have used equation (4.6). Then I have used this normalized mutual information values to select effective features for the classification of the dataset.

PCA Dataset Plus NMI Approach

- **Step 1: Get input data**
  In this analysis I have used dataset obtained via principal component analysis. I have

taken first 20 principal components into consideration. Then I generate class labels of the sample. I have taken fourteen classes which label from 1 to 14.

- **Step 2: Normalizing dataset into a range**
  I have normalized the dataset in 1 to 64 range. Then calculated a 64×64 probability table from which I have calculated the marginal probability and the joint probability of variable(s).

- **Step 3: Calculating mutual information between class labels and bands**
  To select or rank best features for classification of the dataset I have calculated mutual information according to equation (4.3) between class labels and each band of the dataset.

- **Step 4: Calculating normalized mutual information**
  To normalize mutual information values between 0 to 1 I have used equation (4.6). Then I have used this normalized mutual information values to select effective features for the classification of the dataset.

4.5.3 Proposed TCOSD Method

In this research, a target class oriented subspace detection method is proposed by applying NMI over PCA data. At first, PCA is applied on hyperspectral data cube to uncorrelate the input images. Then for each class of the data set NMI is determined to obtain a relevant feature space. During the feature selection, one class is considered as target and the remaining classes are considered as the background/second class. The algorithm of the proposed target class oriented subspace detection method is summarized as follows:

- Step 1: Perform PCA and extract features of the input data set.
- Step 2: Extract the training data and training label from the PCA features.
- Step 3: Consider one class as target at a time, i.e, make a class as target and all other class as background.
- Step 4: Apply NMI between the training labels and the principal components.
- Step 5: Select the best principal component based on the high value of NMI. List the selected subset.

- Step 6: Apply the Eq. (4.6) for selecting multiple features based on NMI for the target class.

- Step 7: Apply the selected features to the KSVM classifier.

- Step 8: Repeat step 3 to 6 by making another class as target and the remaining as background.

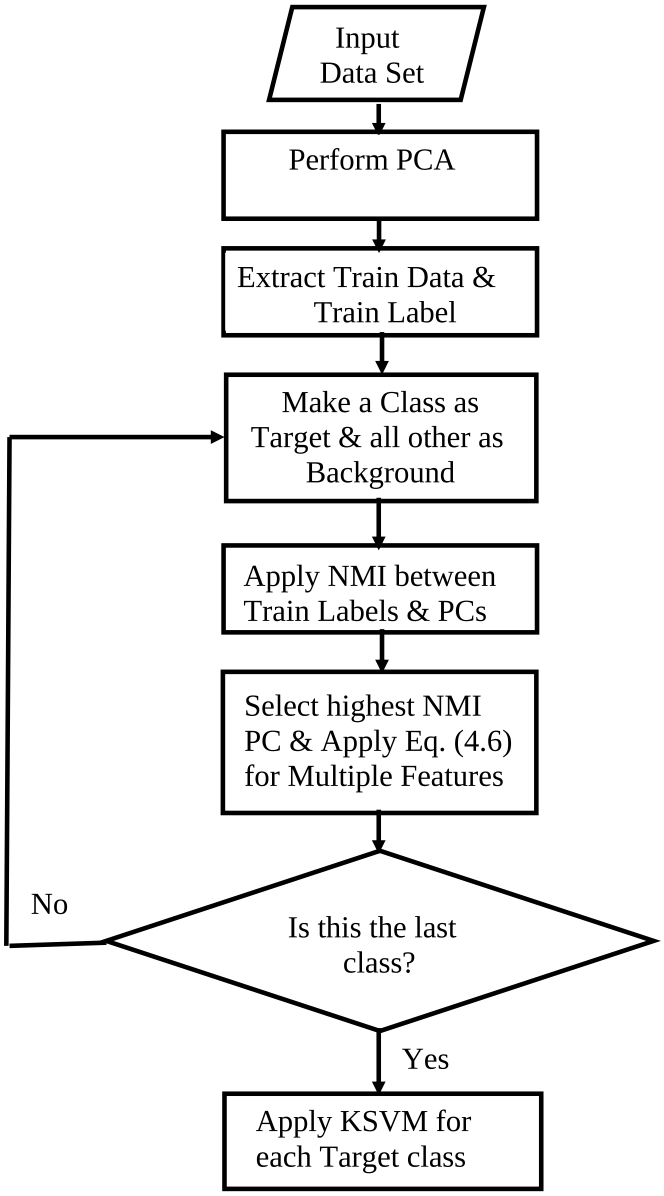

Figure 4.1: Flow chart of the proposed TCOSD Method

All the classes of the data set are not important for some specific application and this is why a target class oriented NMI method applied that select relevant features only for that target class. We call this method as Target Class Oriented Subspace Detection (TCOSD). In this work, Kernel Support Vector Machine (KSVM) is applied to assess the performance as it is the most popular classifier in the current literature [10].

4.6 Experimental Results

4.6.1 Indian pines (92AV3C) Dataset

This scene was acquired by the AVIRIS sensor. Indian Pines is a 145×145 scene containing 220 spectral reflectance bands in the wavelength range 0.4-2.5μm. The Indian Pines scene contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. A random band with training and testing sample pixels which is used in this research along with ground truth is shown in Figure 4.2 and 4.3. The ground truth available is designated into sixteen classes. The corrected Indian Pines data-set contains 200 bands, obtained after removing bands covering the region of water absorption: (104- 108), (150-163), 220.

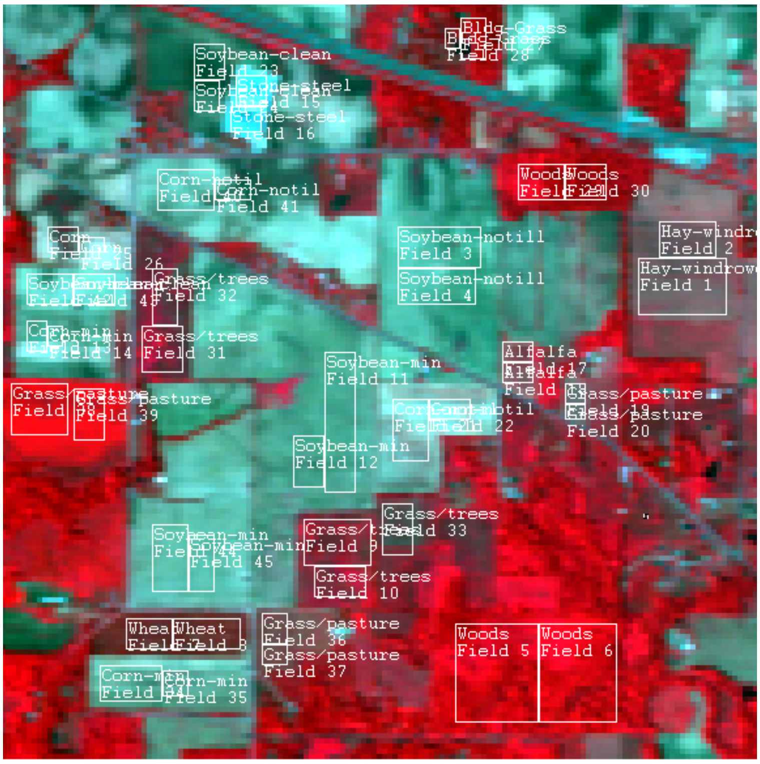

Figure 4.2: AVIRIS 92AV3C dataset for experiment

Figure 4.3: Ground truth of AVIRIS 92AV3C dataset

4.6.2 Feature Extraction Result

The first few PCA images have 99% of the total variance of the input data set. The first component is the brightest and sharpest. Fig. 1 shows few PCA images.It is not always true that higher the variance higher the information. In Fig. 4.4, it is visually shown that the PCA image 7 has high variance but less spatial information than PCA image 9.

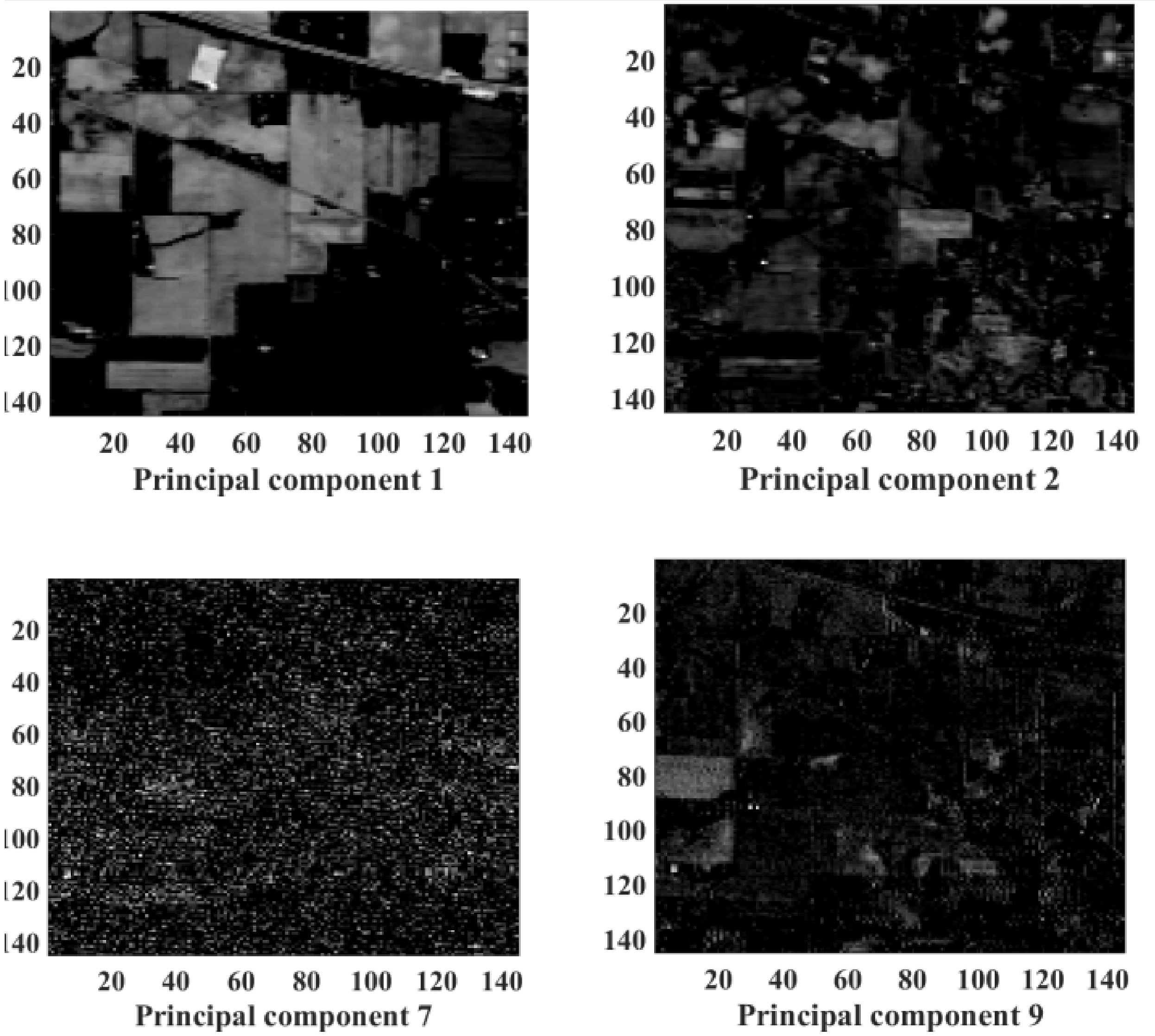

Figure 4.4: PCA image 1, 2, 7 and 9.

Figure 4.5 represents the cumulative relations between variance and each component. Here we can see than first 8 components of PCA has contain 99.69% variations of the dataset.

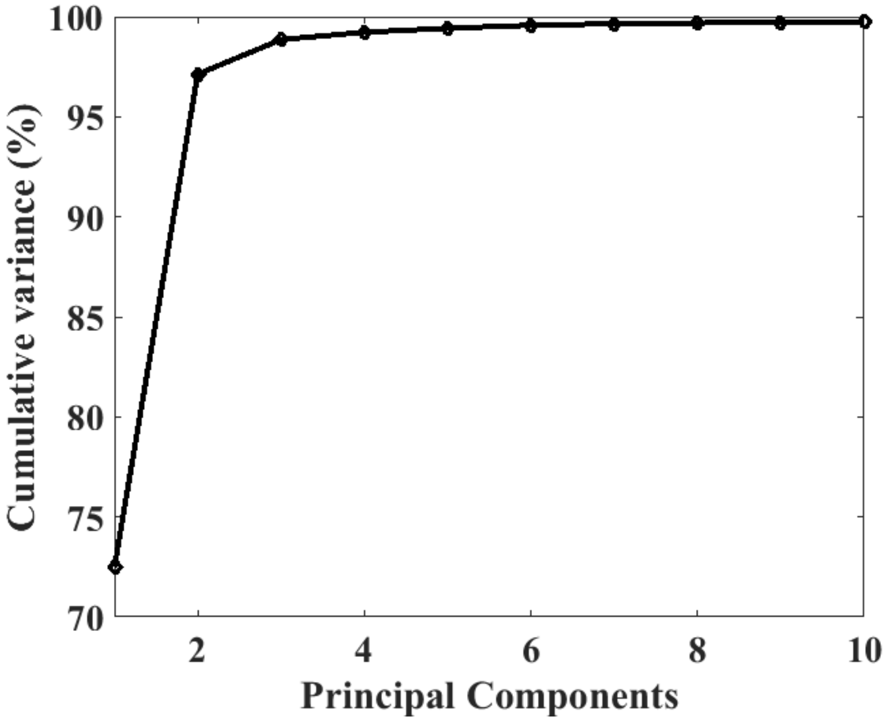

Figure 4.5: Cumulative variance of bands obtained via PCA

4.6.3 NMI Based Feature Selection Result

To address the limitation of PCA, feature selection method is used. Normalized mutual information as selection technique is applied to overcome the problem of PCA. Table-4.1 shows selected features applying NMI over original bands and Table-4.2 shows the selected features applying NMI over PCA images. It is seen that the PC9 is selected but not PC7.

Table 4.1: Selected features with Org+NMI

Method	Order of selected features
Org+NMI	Band: 27,1,167,3,139,14,2,8

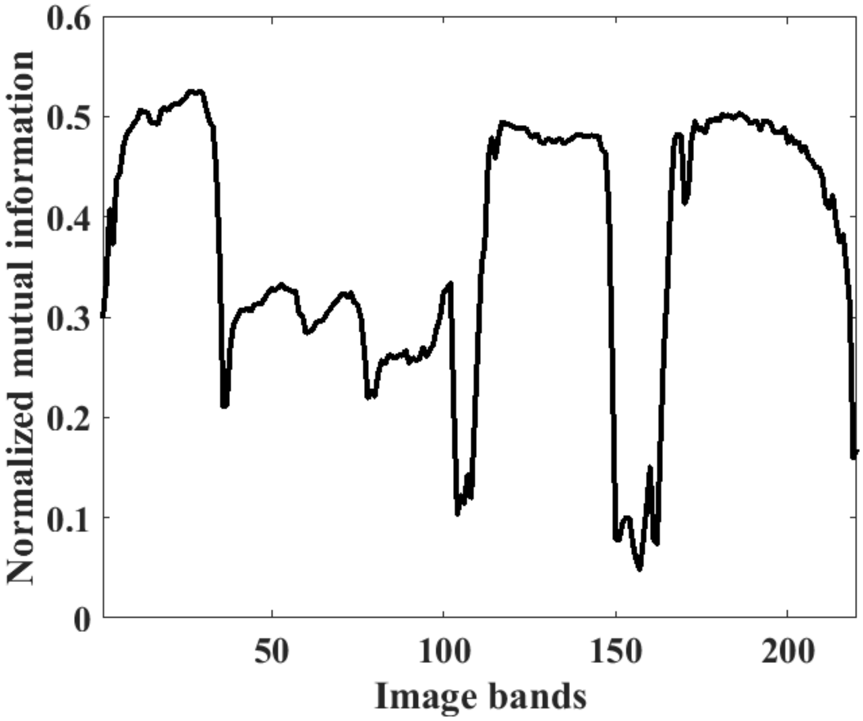

Figure 4.6: Normalized mutual information between class labels and original bands

Fig. 4.6 shows the NMI between class labels and original data bands. The higher the NMI value of a band, the band is more important for the classification. Fig. 4.7 represents NMI between class labels and first 20 principal components. Therefore, in PCA+NMI, principal components are selected based on high value of NMI.

Table 4.2: Selected features with PCA+NMI

Method	Order of selected features
PCA+NMI	PC: 1,4,3,5,9,6,2,15

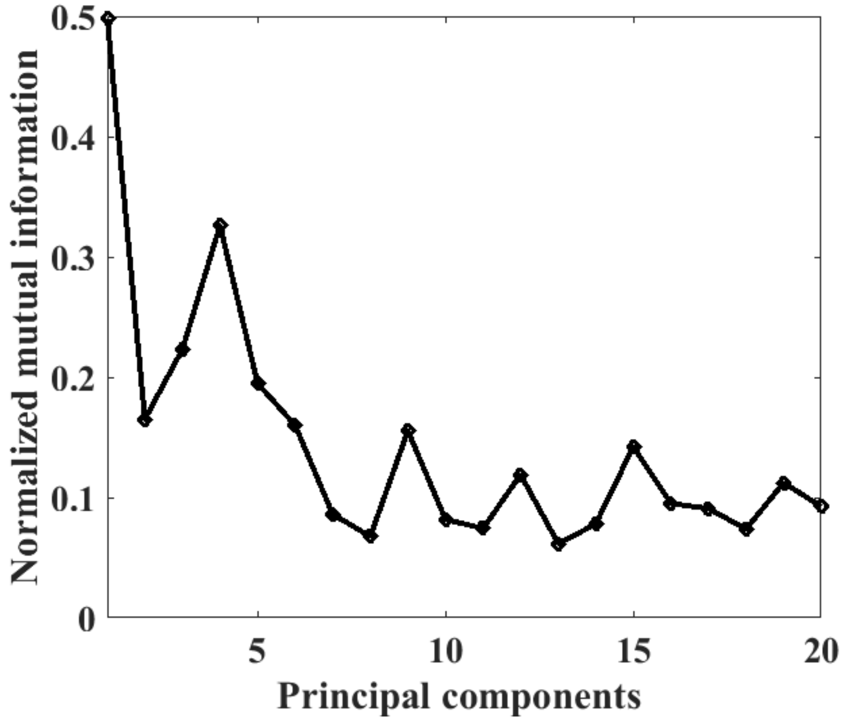

Figure 4.7: Normalized mutual information between class labels and PC1-PC20

4.6.4 Experimental Results

In this experiment, we proposed a target class oriented feature selection method (TCOSD) using NMI measure for maximizing the relevance of the selected subspace.The objective is to show that the target class oriented feature selection is more effective than the traditional feature selection while considering all the classes together. The proposed method manages this situation by considering one class as target at a time and the remaining as the background. Table-4.3 shows the selected features for each target class.

Table 4.3: Selected features for each target class with proposed method (TCOSD)

Target class	Order of selected features
Hay-windrowed	PC: 4,1,17,5,12,3,16,2
Soybean-notil	PC: 1,17,16,11,20,14,13,3
Woods	PC: 1,16,17,3,15,11,5,20
Wheat	PC: 6,19,17,3,12,16,11,5
Grass/trees	PC: 4,9,5,6,16,17,2,1
Soybean-min	PC: 1,17,16,3,5,12,11,9
Corn-min	PC: 1,17,16,5,11,3,20,19
Stone-steel	PC: 3,4,1,17,16,10,5,11
Alfalfa	PC: 4,17,15,20,3,16,6,12
Grass/Pasture	PC: 9,3,17,6,16,11,1,15
Corn-notill	PC: 1,17,2,16,11,18,20,19
Soybean-clean	PC: 1,15,17,3,16,20,12,11
Corn	PC: 1,19,17,16,18,8,11,3
Bldg-Grass	PC: 15,16,17,3,18,11,19,4

4.7 Classification Results

The hyperspectral image named 92AV3C which was captured over the indian pine test site in Northwestern indiana by the NASA AVIRIS sensor [33] is used as the input data set. It has 220 image bands which contains 145×145 pixels. The ground truth available have sixteen classes. In the experiment, 14 classes are used, because the oats and another class are too small. Details training and testing samples of AVIRIS 92AV3C data set is shown in Table 4.4.

Table 4.4: Details of the train and test samples

Class name	Train	Test
Hay-windrowed	187	77
Soybean-notil	128	105
Woods	367	341
Wheat	54	78
Grass/trees	249	115
Soybean-min	253	115
Corn-min	253	115
Stone-steel	36	30
Alfalfa	24	24
Grass/Pasture	156	92
Corn-notill	172	60
Soybean-clean	96	78
Corn	30	15
Bldg-Grass	40	12
Total	1900	1179

The Selected features is applied to a KSVM classifier. A comparison of the classification accuracy is presented in Table-4.6 and it can be seen that the proposed method outperforms the baseline approaches.

Table 4.5: Details of parameter for KSVM(RBF kernel)

Method	C	γ
Org+NMI	10	2.44
PCA	19	2.40
PCA+NMI	16	0.7
TCOSD (Proposed)	5	0.75

The kernel hyperparameter such as the best cost C and kernel width γ are calculated using 10 fold cross-validation [34]. Table-4.5 shows kernel hyperparameter used in this experiment.

Table 4.6: Classification result for 8 features

Method	Classification Result
Org+NMI	72.26%
PCA	89.90%
PCA+NMI	93.12%
TCOSD (Proposed)	96.57%

Fig. 4.8 shows a visualization of comparison among the proposed approach and some standard approach studied when the classification accuracy is determined by adding the features progressively.

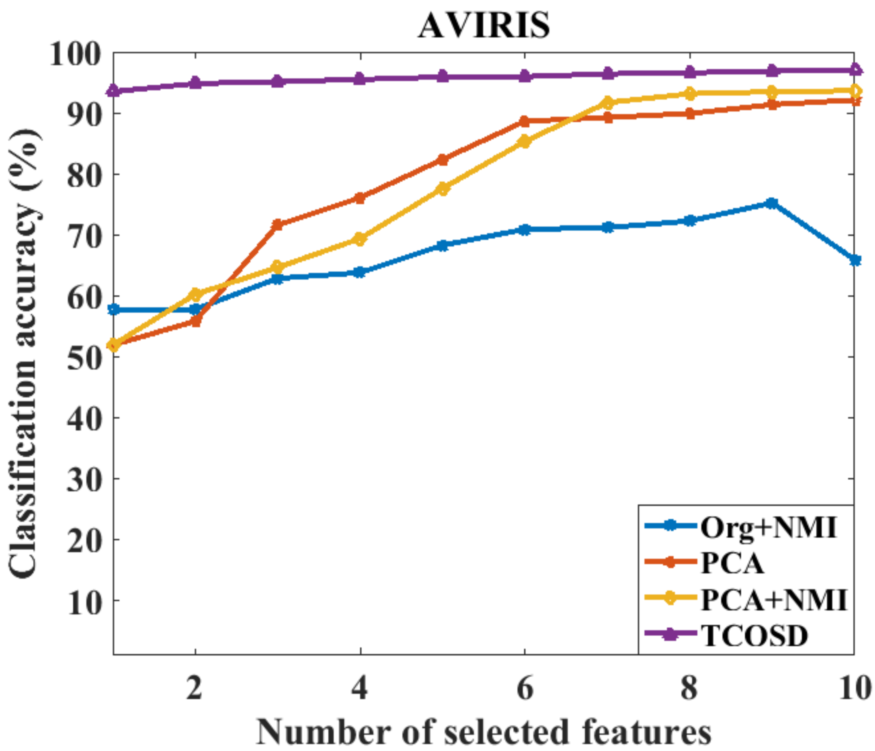

Figure 4.8: Classification accuracy of AVIRIS 92AV3C data

Chapter 5

Conclusion and Future Works

5.1 Conclusion

The proposed target class oriented subspace detection (TCOSD) provide improved subspace for the task of classification. The classification performance of the proposed method is justified using KSVM. It can be seen that the proposed method is able to detect a better subspace which can provide the best classification accuracy among the standard approaches studied. This is because the proposed method selects the most relevant subset of images for required classes. It also consider the spatial information while performing the feature selection which was absent in PCA only method. Similarly, PCA+NMI considers all the classes together and come up with suboptimal solution. Therefore, the proposed method is most suitable for ground object detection and its classification.

5.2 Limitations

The proposed method needs a huge amount of computational time and processing power. This approach may become very costly when the number of training and testing samples and number of classes are very large.

5.3 Future Works

The proposed method (TCOSD) needs some further improvement to handle the complex class relationships where only a few feature may not capable to complete the task.

References

[1] *Available: https://en.wikipedia.org/wiki/Remote_sensing*.

[2] *Available: https://en.wikipedia.org/wiki/Hyperspectral_imaging*.

[3] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed. Germany: Springer-Verlag, 2006.

[4] P. Shippert, "Introduction to hyperspectral image analysis," *Online Journal of Space Communication*, Jan 2003.

[5] M. A. Hossain, X. Jia, and M. Pickering, "Subspace detection using a mutual information measure for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, pp. 424–428, 2014.

[6] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, p. 55–63, Jan. 1968.

[7] L. I. Smith, "A tutorial on principal components analysis," *IEEE Geoscience and Remote Sensing Letters*, pp. 2–8, 2002.

[8] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surveying and Land Information Systems*, vol. 62, no. 2, pp. 115–200, 2002.

[9] M. A. Hossain, X. Jia, and M. Pickering, "Improved feature selection based on a mutual information measure for hyperspectral image classification," pp. 3058–3061, 2012.

[10] C. Chang and C. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.

[11] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, p. 674–677, 2007.

[12] X. Jia, B. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, p. 676–697, Mar. 2013.

[13] M. A. Hossain, Hasin-E-Jannat, B. Ahmed, and M. A. Mamun, "Feature mining for effective subspace detection and classification of hyperspectral images," *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 544–547, 2017.

[14] L. Ying, G. Yanfeng, and Z. Ye, "Hyperspectral feature extraction using selective pca based on genetic algorithm with subgroups," *in Proc. ICICIC*, p. 652–656, 2006.

[15] A. E. Akadia, Amine, and D. A. A. E. Ouardighi, "A two-stage gene slection scheme utilizing mrmr filter and ga wrapper," *Knowledge Information Systems*, vol. 26, no. 3, pp. 487–500, 2011.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* USA: Academic, 1990.

[17] S. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 45, no. 2, February.

[18] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relecvance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, August 2005.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] K. Torkkola, "Feature extraction by non-parametric mutual information maximization," *J. Mach. Learn. Res.*, vol. 3, pp. 1415–1438, Mar. 2003.

[21] M. L. Murillo and A. A. Rodriguez, "Maximization of mutual information for supervised linear feature extraction," *IEEE Trans. Neural Netw*, vol. 14, no. 5, pp. 1433–1441, Sep. 2007.

[22] X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 274–281, Mar. 1994.

[23] F. Melgani and L. Bruzzone, "Support vector machines for classification of hyperspectral remote-sensing images," vol. 1, pp. 506–508 vol.1, 2002.

[24] J. Xia, "Multiple classifier systems for the classification of hyperspectral data," *Université Grenoble Alpes*, 2014.

[25] . AVIRIS, "Airborne visible/infrared imaging spectrometer, *[Online]*, vol. Available: http://aviris.jpl.nasa.gov/.

[26] B. Guo, R. Gunn, R. Damper, and J. Nelson, "https://engineering.purdue.edu/ multispec/hyperspectral.html."

[27] S. Sindhumol and M. Wilscy, "Hyperspectral image analysis–a robust algorithm using support vectors and principal components," *in Computing: Theory and Applications, 2007. ICCTA '07. International Conference on*, pp. 389–395, 2007.

[28] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Germany: Springer, 2002.

[29] M. A. Hossain, B. Ahmed, S. S. Ghosh, and M. N. I. Mondal, "Effective subspace detection based on cross cumulative residual entropy for hyperspectral image classification," *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 548–551, 2017.

[30] M. A. Hossain, M. Pickering, and X. Jia, "Unsupervised feature extraction based on a mutual information measure for hyperspectral image classification," pp. 1720–1723, 2011.

[31] M. A. Hossain, X. Jia, and M. Pickering, "Improved feature selection based on a mutual information measure for hyperspectral image classification," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany*, pp. 3058–3061, Jul. 2012.

[32] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," *IEEE Trans. Signal Process.*, vol. 45, no. 11, p. 2758–2765, Nov. 1997.

[33] D. A. Landgrebe, *[Online]. Available: https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.h tm.*

[34] M. Joo, S. Wu, and J. Lu, "Face recognition with radial basis function (rbf), neural networks," *In: IEEE Transactions on Neural Networks*, no. 13, May 2002.