Dimensionality Reduction of Hyperspectral Images Using Combinations of Extracted and Selected Features

A thesis

submitted in partial fulfillment of the requirement for the Degree of

Master of Computer Science and Engineering

of

Jadavpur University

By

Payel Pramanik

Registration No.: 140746 of 2017 - 18

Examination Roll No.: M4CSE19015

Under the Guidance of

Prof. Susmita Ghosh

Department of Computer Science and Engineering

Jadavpur University, Kolkata-700032

India

2019

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

Certificate of Recommendation

This is to certify that the dissertation entitled "Dimensionality Reduction of Hyperspectral Images Using Combinations of Extracted and Selected Features" has been carried out by Payel Pramanik (University Registration No.:140746 of 2017-18, Examination Roll No.: M4CSE19015) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Computer Science and Engineering. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree in any other University or Institute.

……

Prof. Susmita Ghosh (Thesis Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-32

Countersigned

…….

Prof. Mahantapas Kundu
Head, Department of Computer Science and Engineering
Jadavpur University, Kolkata-32.

…….

Prof. Chiranjib Bhattacharjee
Dean, Faculty of Engineering and Technology
Jadavpur University, Kolkata-32.

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

Certificate of Approval*

This is to certify that the thesis entitled "Dimensionality Reduction of Hyperspectral Images Using Combinations of Extracted and Selected Features" is a bona fide record of work carried out by Payel Pramanik in partial fulfillment of the requirements for the award of the degree of Master of Computer Science and Engineering in the Department of Computer Science, Jadavpur University during the period of June 2018 to May 2019. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

……..

Signature of Examiner 1

Date:

……..

Signature of Examiner 2

Date:

*Only in case the thesis is approved

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis entitled "Dimensionality Reduction of Hyperspectral Images Using Combinations of Extracted and Selected Features" contains literature survey and original research work by the undersigned candidate, as part of her Degree of Master of Engineering in Computer Science and Engineering.

All information have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Payel Pramanik
Registration No: 140746 of 2017 - 18
Exam Roll No.: M4CSE19015
Thesis Title: Dimensionality Reduction of Hyperspectral Images Using Combinations of Extracted and Selected Features

…..……..

Signature with Date

Acknowledgement

I would like to start by thanking the holy trinity for helping me deploy all the right resources and for shaping me into a better human being.

I would like to express my deepest gratitude to my advisor, Prof. Susmita Ghosh, Department of Computer Science and Engineering, Jadavpur University for her admirable guidance, care, patience and for providing me with an excellent atmosphere for doing research. Our numerous scientific discussions and her many constructive comments have greatly improved this work.

I am grateful to Dr. Aloke Datta (Department of Computer Science and Engineering, NIT Meghalaya) for sharing the data used in this study as well as for his guidance regarding the program implementation part.

I am thankful to Mr. Abhishek Dey (Department of Computer Science, Bethune College) for helping me by discussing various methods regarding this work. I am also thankful to my lab partners for always encouraging me to complete this work. Without all of them it would surely be a very lonely lab.

Most importantly none of this would have been possible without the love and support of my family. I extend my thanks and my love to my parents, their forbearance and whole hearted support helped this endeavor succeed.

……..

Payel Pramanik

Registration No.: 140746 of 2017 – 18

Exam Roll No.: M4CSE19015

Department of Computer Science & Engineering

Jadavpur University

Contents

List of Figures

List of Tables

Abstract

In hyperspectral images, each pixel has a very large number of features which is equal to the number of bands present in the image. It needs a very complex procedure to perform any classification task with this high dimension. So reduction of dimension of hyperspectral images is an important task. In this thesis, a combination of extracted and selected features is used for classification of hyperspectral images. In doing so, we have proposed a new approach to combine these two types of features. Three different techniques are used on the set of combined features to reduce the dimension of hyperspectral images. Two different heuristic methods are considered for combining both types of features. In the first method, extracted features are combined with a set of selected features, where the cardinalities of both the feature sets are different. While in other they are same. Thereafter, on this combined set of features, two popular feature selection algorithms are employed to reduce dimensionality. It has been found that, with such reduced combined features, pixel classification accuracy of hyperspectral images is higher in contrast to working only with the set of extracted features both having same cardinality.

Date:

Place:

_____

(PAYEL PRAMANIK)

Chapter 1

Introduction

A number of earth satellites have been launched to advance our understanding regarding the environments of the Earth. The satellite sensors capture data from visible to microwave regions of the electromagnetic spectrum. When spectral measurement is performed using hundreds of narrow contiguous wavelength intervals, the resulting image is called a hyperspectral image which is represented as an image cube. In this cube, the x and y axes specify the size of the images and z axis denotes the number of bands in the hyperspectral data. This image gives the detailed spectral signatures for each pixel of an image to identify and distinguish spectrally unique materials [1]. But it is a difficult task to exploit such a large amount of data and thus classification of the images is required to summarize the information. It is another difficult task to perform classification with this high dimension of hyperspectral images and thus dimensionality reduction is required. A brief description of electromagnetic spectrum, remote sensing, hyperspectral sensors, hyperspectral image acquisition, its characteristics and applications of hyperspectral images and a small introduction on dimensionality reduction are described in this chapter.

1.1 Electromagnetic spectrum

Electromagnetic radiation consists of electromagnetic waves which are synchronized oscillations of electric and magnetic fields at right angles (both the waves are perpendicular to the direction of propagation) that propagate at the speed of light in vacuum [2, 3]. The entire range of frequencies and wavelengths of electromagnetic radiation makes up the electromagnetic spectrum. The common categorization of electromagnetic waves is by their wavelength location within the electromagnetic spectrum. The categorizations in order of increasing frequency or decreasing wavelengths are radio waves, microwaves, infrared radiation (IR), visible light, ultraviolet radiation (UV), X-rays and gamma rays. The diagrammatic representation of this is shown in fig. 1.1. The unit used to measure the wavelengths along the spectrum is micrometer, denoted by μm (i.e., 1×10^{-6} m).

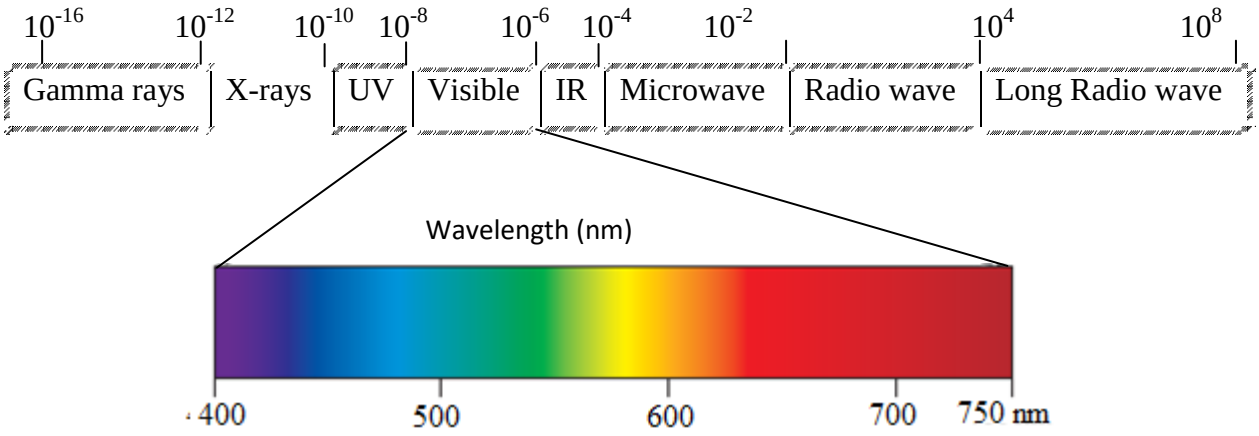


Fig.1.1: Categorization of Electromagnetic Spectrum

The radio wave is used to broadcast radio and television. Microwaves are used in cooking, radar, telephone and other signals. Infrared wave transmits heat from sun, fires, and radiators. Visible wave makes things able to be seen. Ultraviolet rays can be absorbed by the skin and it is used in fluorescent tubes. X-rays used to view inside of bodies and objects, Gamma rays used in medicine for killing cancer cells [3].

According to particle theory in physics, electromagnetic radiation is composed of many discrete units called photons or quanta. The energy of a quantum is inversely proportional to its wavelength. The longer the wavelength, the lower is its energy contents. Radio waves

have the longest wavelength and least energy whereas gamma rays have the shortest wavelength having massive energy. From the remote sensing perspective, it can be found that long wavelength radiation is more difficult to sense than radiation of shorter wavelengths [2]. The sun is the most obvious source of electromagnetic radiation for remote sensing. The reflectance based remote sensing imaging technique uses visible to near infrared (IR) region of electromagnetic wave. The other wavelengths of electromagnetic spectrum are not useful for remote sensing because of the following reasons. The gamma rays, X-rays, UV ray are not used because these ranges of wavelengths are completely absorbed by the atmosphere or there are severe atmospheric scattering. The mid IR and thermal IR are dominated by radiation emitted by the Earth. Microwave wavelength is used for radar images. The energy of the longest wavelength portion i.e., TV and radio wave (>10nm), is lower range for remote sensing. Table 1.1 shows the detail characteristics of visible to near IR range of electromagnetic spectrum (400nm to 2.4 μm) [1, 4].

1.2 Remote Sensing

Remote sensing has been defined as the field of study of recording, measuring, and analyzing information about an object, area, or phenomenon by a device without coming into physical contact with the object, area, or phenomenon [1]. Humans with the aid of their eyes, noses, and ears are constantly seeing, smelling, and hearing things from a distance as they move through an environment. Thus, humans are naturally designed to be remote sensors. In order to study large areas of the Earth's surface geographers use devices known as remote sensors. These sensors are mounted on platforms such as helicopters, planes, and satellites that make it possible for the sensors to observe the Earth from above [6]. So remote sensing is the measurement of object properties on Earth's surface to capture specific information to make decisions. For example, a weather satellite is used to measure the global atmospheric parameters that ultimately help to take decision for weather forecasting [1, 8]. The main three components of remote sensing are: energy source, platforms and sensors. Energy source means the reflectance/ radiated signals from an object or phenomena. It can be a passive system (Sun, irradiance from earth's materials) or an active system (irradiance from artificially generated energy sources such as radar).

Name	Spectral Range	Characteristics
Visual Blue	0.45-0.52 µm	As water increasingly absorbs electromagnetic (EM) radiation at longer wavelengths, band 1 provides the best data for mapping depth-detail of water-covered areas. It is also used for soil-vegetation discrimination, forest mapping and distinguishing cultural features.
Visual Green	0.50-0.60 µm	The blue-green region of the spectrum corresponds to the chlorophyll absorption of healthy vegetation and is useful for mapping detail such as depth or sediment in water bodies. Cultural features such as roads and buildings also show up well in this band.
Visual Red	0.60-0.70 µm	Chlorophyll absorbs these wavelengths in healthy vegetation. Hence this band is useful for distinguishing plant species, as well as soil and geologic boundaries.
Near IR	0.70-0.80 µm	This band corresponds to the region of the EM spectrum, which is sensitive to varying vegetation biomass. It also emphasizes soil-crop and land-water boundaries.
	0.80-1.10 µm	This band is used for vegetation discrimination, penetrating haze and water-land boundaries.
Mid IR	1.55-1.74 µm	This region is sensitive to plant water content, which is a useful measure in studies of vegetation health. It also used for distinguishing clouds, snow and ice.
	2.08-2.35 µm	This region is used for mapping geologic formations and soil boundaries. It is also responsive to plant and soil moisture content.

Table 1.1: Characteristics of Electromagnetic Spectrum

Platforms mean the vehicles to carry the sensor e.g., aircraft, space shuttle, or satellite. The sensors are the devices to detect electromagnetic radiation e.g., camera or scanner. Also another major task is accumulating knowledge or information through analysis and displaying these measurements over a two-dimensional spatial grid means images. Depending on the platform from which the remote sensing sensor captures images, the images have different categorizations. If an image is captured from sensor on the platform of

aircraft then it is called air-borne image. On the other way, if it is taken from satellite then it is called space-borne image [1, 2]. The main objective of remote sensing systems is to provide a repetitive and consistent view of the Earth facilitating the ability to monitor the Earth system and the effects of human activities on Earth. A few application areas of remote sensing images are weather prediction, agricultural forecasting, resource exploration, land cover mapping etc. [3, 5 and 6].

Fig.1.2: Remotely sensed image: A hyperspectral image - University of Pavia acquired by the ROSIS sensor

Remotely sensed image analysis involves the source for radiating energy (sun light), interactions among such energy and the matters present on the Earth's surface, traversal of reflected/radiated energy through atmosphere, capturing the reflected/ radiated energy by the sensor, storage of such energy information as remote sensing images, processing and analysis of these images, and finally, the decision made by the users on these images [8]. Remote sensing images have different types of resolutions: spatial, spectral, radiometric and temporal. Spatial resolution is the distance between two nearest objects that a sensor can record distinctly. It is basically the geometric properties of imaging devices and is measured in units of length (e.g., meter). Spectral resolution is defined by the wavelength of the channel over which radiation is recorded. High Spectral resolution is achieved by narrow bandwidths which provide more accurate spectral signature for discrete objects than broad bandwidths. Radiometric resolution, which is also called quantization, is determined by the least

difference in intensity of the signal that can be distinguished. It is nothing but the sensitivity of the sensor to differences in strength of electromagnetic radiation. The temporal resolution refers to the frequency of observations of a particular scene by remote sensing sensors [5, 6]. Figure 1.2 shows a remotely sensed image (a hyperspectral image) of University of Pavia acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy.

1.3 Hyperspectral Sensors:

Hyperspectral imaging sensors capture the amount of visible and infrared radiance, reflected/emitted from the surface of an object. With respect to the type of energy resources, remote sensing technology is classified into two categories: active and passive remote sensing. Figure 1.2 shows the illustration.

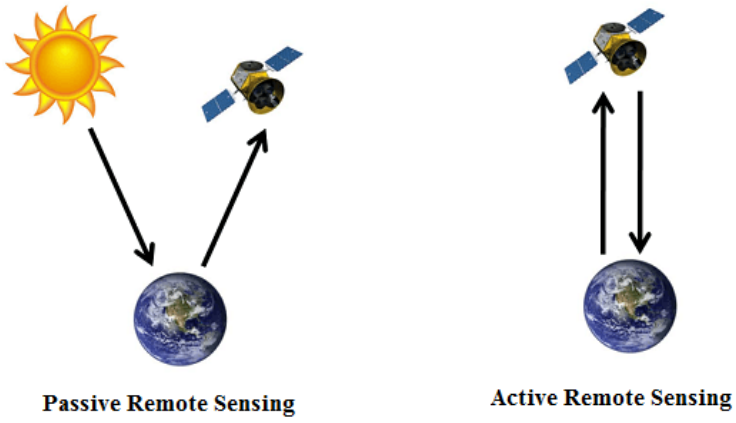


Fig.1.3: Passive and Active Remote Sensing

Active sensors emit energy in order to scan objects and areas whereupon a sensor then detects and measures the radiation that is reflected or backscattered from the target. RADAR (Radio Detection and Ranging) and LiDAR (Light Imaging detection and Ranging) are examples of active remote sensing where the time delay between emission and return, location, speed and direction of an object is measured. Passive sensors gather radiation that is emitted or reflected by the object or surrounding areas. Reflected sunlight is the most common source of radiation measured by passive sensors. Examples of passive remote sensors include film photography, charge-coupled devices and radiometers [10].

However, in remote sensing, sunlight is reflected from the surface of the Earth and this reflected energy is captured by hyperspectral sensors. So this is a kind of passive sensing. Hyperspectral images may not be collected at night, due to the unavailability of reflected

energy. Though there are some exceptions with thermal infrared energy, which is naturally emitted and can be detected both in day or night. But there are very limited applications of thermal energy in hyperspectral images till date. Hyperspectral sensors are equipped with advanced digital cameras which provide both spatial and spectral details of objects. They collect images of hundreds of band and measure the reflected radiation of a particular wavelength in each band image. The primary characteristics of hyperspectral sensors are that they can capture images with narrow spectral and relatively larger spatial resolution. Spectral resolution can be determined by the range of spectral bands which are used to measure radiation [1]. Availability of Hyperspectral images is increasing day by day and analysis of these images is becoming more important due to recent advancements of remote sensing sensors. A brief detail of few of the existing hyperspectral sensors are described below [1].

Hyperion sensor – Satellite EO-1, Number of bands: 220, Spectral range: 0.4 to 2.5 µm.

AVIRIS (Airborne Visible Infrared Imaging Spectrometer) – Airborne, Number of bands: 224, Spectral range: 0.4 to 2.5 µm.

ROSIS (Reflective Optics System Imaging Spectrometer) – Airborne, Number of bands: 115, Spectral range: 0.4 to 0.86 µm.

HYDICE (Hyperspectral Digital Imagery Collection Experiment) – Airborne, Number of bands: 210, Spectral range: 0.4 to 2.5 µm.

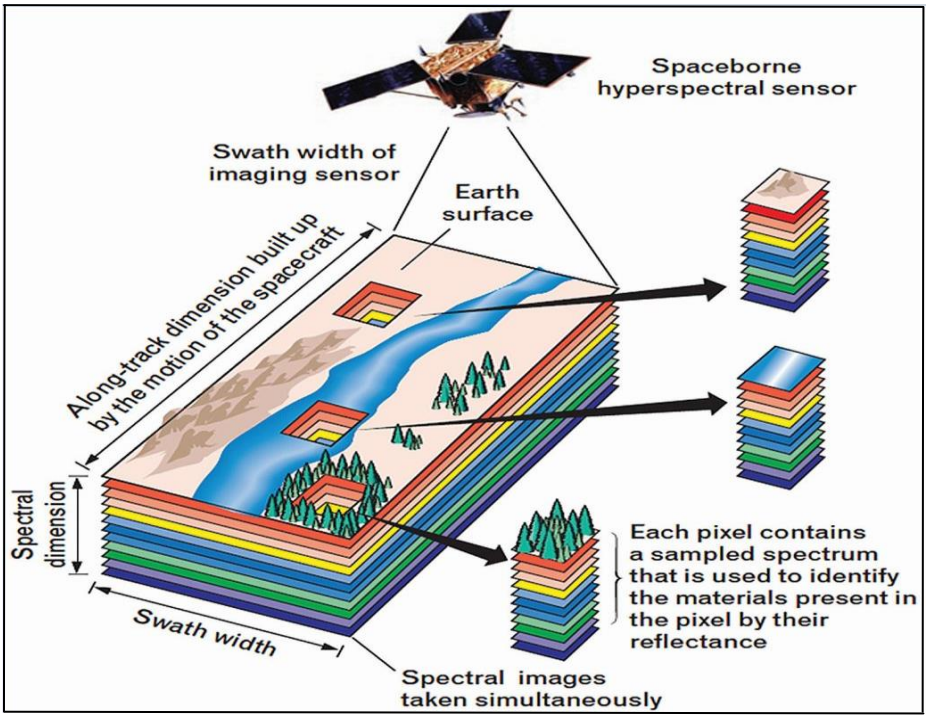
HyMap – Airborne, Number of bands: 100 to 200, Spectral range: Visible to thermal infrared.

1.4 Hyperspectral Image Acquisitions:

The process elements of the hyperspectral data acquisition process are energy source, propagation of energy through the atmosphere, interaction of energy among the objects of the Earth's surface, retransmission of reflected energy through the atmosphere, sensing of energy using air-borne/ space-borne sensors, resulting in the generation of sensor data in pictorial form. In brief, sensors are used to record variations in the way features of earth surfaces reflect an electromagnetic energy [1].

All hyperspectral sensors are treated as imaging spectrometers, which means that they include a prism or diffraction grating that can break the incoming radiation into discrete wavelengths, which are then dispersed separately to detectors. Like other remote sensing sensors, hyperspectral sensors are also of two types: air-borne and space-borne. Signals measure by an air-borne hyperspectral sensor is generally digitized and then recorded on board the aircraft, and retrieved once the aircraft lands. Signals measured by a satellite

hyperspectral sensor is digitized and then electronically transmitted to a ground receiving station. The recorded images from these sensors suffer from various degree of distortion of which radiometric and geometric distortions are primary ones. When light travels from the Sun to the Earth surface and then is reflected to an air-borne or space-borne sensor, the intervening atmosphere often scatters some light. Therefore, the light received by the sensor may be more or less than that caused by reflectance from the surface alone. This is called radiometric distortion. The geometric distortion refers to shape and scale distortions which occur mainly due to perspective projection and instrumentation. Atmospheric correction attempts to minimize these effects on image spectra and is an indispensable step before conducting quantitative image analysis [1]. Figure 1.4 shows the detailed acquisition process.



A

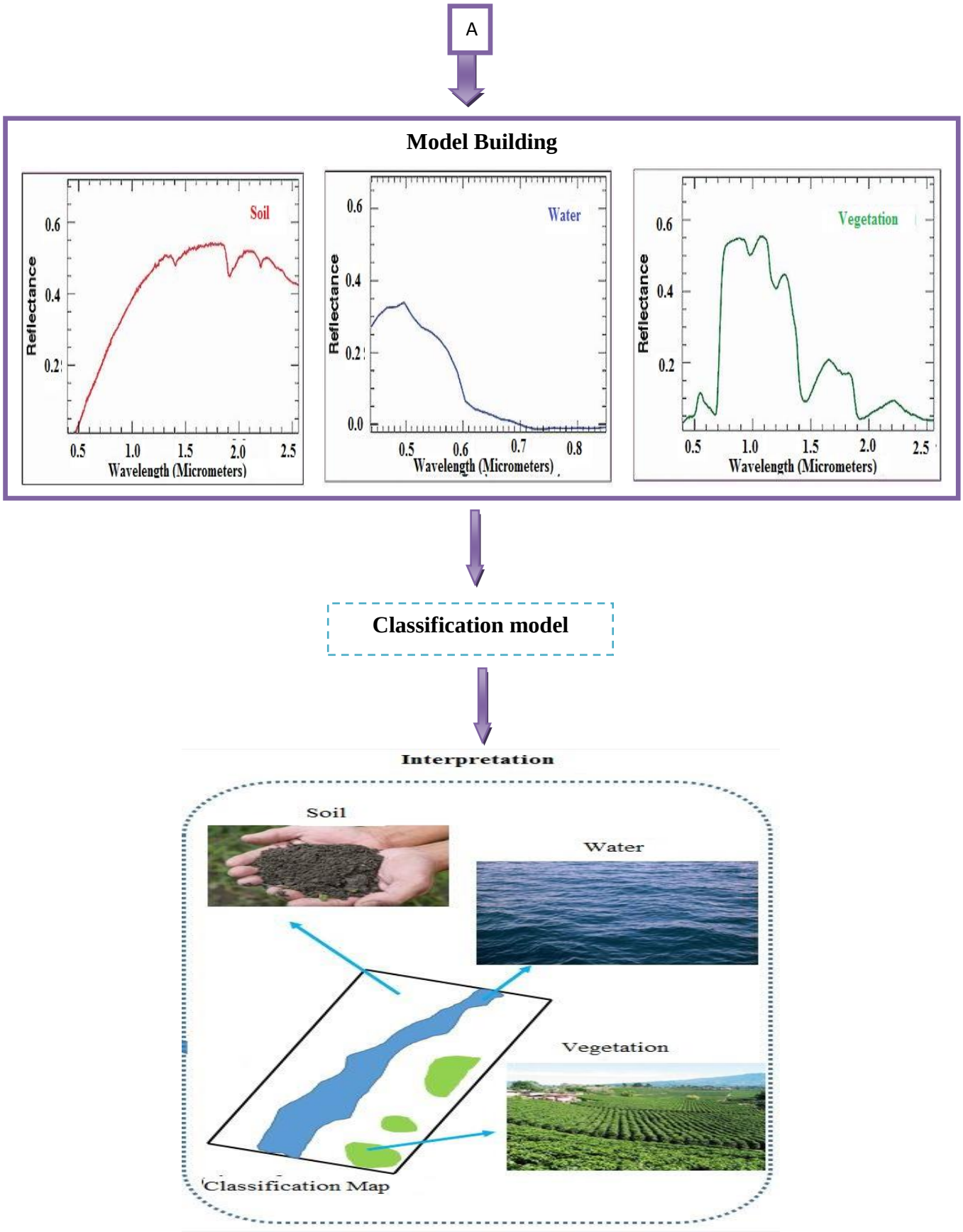


Fig.1.4: Hyperspectral Image Acquisition Process

1.5 Characteristics of Hyperspectral Images:

- Hyperspectral image is represented as a data cube with spatial information in (x, y) coordinate and spectral information in the z-direction.
- Each pixel is represented as a spectral signature which is the response (or the reflectance value) of different wavelengths within the range. These spectral signatures are helpful to identify and distinguish among spatially similar looking objects.
- A sensor that measures a few band images may be considered as hyperspectral if these images were captured in narrow spectral resolution (i.e. ~10 nm wide) and they are contiguous, means not separated by non-measured wavelength bands.
- Hyperspectral images are spectrally over-determined. There is high correlation between adjacent image bands.
- Hyperspectral images provide more accurate and detailed information than any other type of remotely sensed data. Hyperspectral datasets are generally composed of about 100-200 band images with narrow bandwidths (~ 10 nm).
- The size of hyperspectral images is very high which makes it costly for storage and transmission purpose.

1.6 Applications of Hyperspectral Images:

Hyperspectral images contain lots of information which is useful for distinguishing different objects/elements. Several applications of hyperspectral images are there due to its measurement capability with many narrow and contiguous bands. Followings are some of them.

In **remote sensing technology** it is very important to distinguish features of earth surfaces, each features have different spectrum band. Multispectral satellite can capture image up to few bands (10-100); for example, Land sat 7 have 8 bands. But Hyperspectral imaging satellite can capture earth surface in more than 200 bands which helps scientist to differentiate objects that were not possible in multispectral imaging because of spectral resolution [4].

Land use or land cover mapping is the area where hyperspectral images are used efficiently. The term “land-cover” relates to the type of area present on the surface on the earth i.e., cultivation field, lakes, trees, concrete or highways are examples of land cover

types. The term “land use” relates to human activity or economic function associated with a specific land, i.e., land use could be described as urban use, residential use etc. Hyperspectral images are useful in discriminating land cover types due to different spectral signatures of various lands [1, 2].

In the area of **atmospheric characterization and climate research,** certain regions of electromagnetic spectrum are sensitive to different atmospheric constituents. For example, water absorption is prominent at 820nm, 940nm, 1130 nm, 1380nm and 1880 nm, while CO2 absorption occurs at 1600nm and O2 absorption occurs at 760nm and 1270nm. Using this information, estimation of constituents of atmospheric data can be derived using hyperspectral images. The images also used for monitoring natural hazards such as forest fire, volcanic activity, weather forecasting and so on [1].

In **monitoring of coastal environments,** hyperspectral data have been used to characterize and map coral reefs and submerged aquatic vegetation, including sea grasses and water quality. At the land-sea interface, applications include characterization of aquatic vegetation, salt marshes and mangroves. Monitoring of coral reefs using remote sensing data has provided a cost effective and time effective tool for reef surveys, change detection and management [4].

Snow and ice hydrology is the study of cryosphere i.e., the frozen water part of the earth system; for example, study of frozen water surrounding of the Antarctica and the Arctic. Hyperspectral imaging techniques have been used to retrieve information relating to physical and chemical properties of snow including grain size, fractional cover, impurities and liquid water content within snow [1].

Geological and soil mapping is another application of hyperspectral image analysis. Understanding soil suitability for various land use activity is essential to prevent environmental deterioration associated with misuse of land. Geologic mapping involves identification of land forms, rock types and rock structures. Mineral resource exploration is an important type of geologic mapping activity. Hyperspectral image is a preliminary step for most field-based geologic mapping especially in remote regions [1].

Hyperspectral sensors are well suited for **vegetation mapping.** Knowledge of variation in the spectral reflectance of vegetation is a fundamental thing to understand the information content of spectra derived from hyperspectral sensors. Key descriptive elements of the vegetation spectral reflectance signatures include the green peak, the chlorophyll well, the red-edge and water absorption features. Discrimination of species is only achievable where a combination of leaf chemistry, structure and moisture content culminate to form a unique

spectral signature [1, 3].

Hyperspectral images are useful in **military application** also. Military personnel have used hyperspectral images for detection military vehicles under partial vegetation canopy, making differences between camouflage and actual vegetation and many other military target detection objectives [1].

In **Seed viability study** by using the hyperspectral image and plotting the reflectance spectrum one can conclude that whether those seeds are viable or not. Seed might be looking same through naked eyes but its viability will be traced down by the hyperspectral image [3].

In the area of **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 [4].

In **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 [4].

Hyperspectral imaging is widely used in the **food sector**. It is used in different disciplines 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 [5].

In **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 powder [4, 5].

In **Medical Diagnose domain,** early disease detection and disease prevention are very important for the healthy body. Hyperspectral imaging technology can be used to early detect the various types of cancer or retinal diseases [4, 5].

In the area of **Forensic Sciences,** hyperspectral imaging technology can differentiate fine spectral resolution, which makes suitable in the forensic laboratory. It can be used in different ways: questioned document analysis, arson investigation, bloodstain visualization, fiber comparison, gun powder residue, visualization, duct tape examination, fingerprint enhancement, and TLC plate visualization. For example, hyperspectral imaging technology can differentiate between dark marks and bloodstains. This type of differentiating is quite important while solving the crime case [4, 5].

In the domain of **Thin Films** as hyperspectral imaging can distinguish tiny object, it has been widely used for the quality control of the Thin Film manufacturing process [3].

In the area of **Oil and GAS,** hyperspectral imaging technology is widely used in the exploration of Oil and Gas. Now because of hyperspectral imaging sensors it is possible to detect onshore oil seeps [22]. Hyperspectral imaging systems aboard aircraft and spacecraft can detect hydrocarbons and played an important role in the response to the deep-water Horizon oil spill [22]. As trained experts sometimes make mistake on marine or biological phenomena for oil spills when looking at only photographs in the visual spectrum, hyperspectral imaging is particularly useful. Many systems analyze collected data in real-time and automatically identify potential locations of oil spills and seeps. However, because oil spills evolve rapidly, hyperspectral data from aircraft and spacecraft can become obsolete in less than a day, so it is necessary for systems to collect data often and to process it quickly.

In **Hazardous Waste Monitoring,** hyperspectral imaging systems find a broad range of uses regarding hazardous materials management. Images of mine waste can be quickly analyzed to search for minerals that produce acids which pollute rivers. Heavy metals, such as cadmium, lead, and arsenic, are also monitored using hyperspectral imaging. Interestingly, hyperspectral data can also be used to characterize vegetation, which is then used to infer the mineral composition of the underlying soil [23].

Thermal imaging has also been successfully used to locate burial sites of hazardous materials from weapons development projects by distinguishing between disturbed and undisturbed soil. Additionally, underground fires in mines and landfills have also been detected and monitored by hyperspectral imaging systems. On the other hand, hyperspectral imaging has also produced maps of minerals in areas such as the Rockies, which is useful for resource exploration [23].

1.7 Dimensionality Reduction in Hyperspectral Images:

Performance of a classifier depends on the interrelationship among sample sizes, number of features and classifier's complexity. The minimum number of training patterns required for training of a classifier is an exponential function of the number of features [28]. This phenomenon is known as "curse of dimensionality" [29]. In 1968, Hughes conducted a statistical analysis, showing how the accuracy of a classifier depends on the number of training samples. So, the "curse of dimensionality" is also known as "Hughes phenomena"

[30]. Classification accuracy of any classifier does not degrade as the number of features increases, as long as the availability of the number of training samples is arbitrarily large and can represent the underlying densities well. But surprisingly, it has been often observed that more features may degrade the performance of a classifier if the number of training samples is small relative to the number of features. However, hyperspectral imaging sensors acquire a three dimensional image cube (the third dimension specifies the spectral band) as hyperspectral image which usually contains large amount of information to identify and distinguish spectrally unique materials. In this regard, each pixel of the image may be treated as a pattern whose number of attributes/ features is equal to the number of bands present in the image. Moreover, neighboring features are strongly correlated, i.e., information present in hyperspectral image is redundant in nature. So, increasing the number of features may not always increase the discriminating capabilities for classification purpose. Also, high dimensionality (as noticed in hyperspectral image) may lead to "Hughes" phenomena. So reduction in dimensionality is an important step where the aim is to discard redundant features, preserve crucial information, make it less time consuming for classification, as well as, avoid the "Hughes phenomena" [30].

The main two approaches of dimensionality reduction are feature selection and feature extraction. In brief, feature selection is the process of selecting a subset of features from the original set of features, whereas feature extraction is a method of transforming the original set of features into a lower dimensional space. The main two advantages of feature selection and feature extraction are to improve classification accuracy by avoiding "curse of dimensionality" and to reduce the computational cost for classification/ clustering of data [29].

1.7.1 Feature Selection: In feature selection method, a subset of features is selected that are not noisy, redundant or irrelevant for classification from the original set of features to preserve crucial information and reduce redundancy among information [1]. Feature selection methods preserve the original physical meaning of features. Depending on the availability of labeled patterns, feature selection techniques can be categorized as supervised and unsupervised. Supervised feature selection methods use class label information to perform selection of features, whereas, unsupervised feature selection is used when no labeled pattern is available. Feature selection methods consist of two steps: subset generation and subset evaluation. Subset generation is a search process where each point in the search space specifies a possible candidate subset. Several methods exist in literature for finding out such

subsets. In subset evaluation step, the ‘goodness” of a subset of features is measured using some objective functions. We will discuss on those methods in our next chapter (literature survey).

1.7.2 Feature Extraction: Transforming the original set of features into a reduced set of features, which preserves the class separability as much as possible in the reduced space, is called feature extraction. The extracted features lose the meaning of the original features but each of the original features may contribute to make a transformed feature. Also reconstruction of the original features can be done from transformed features and vice-versa. The process of combining features (i.e., how to form the transformation matrix in most cases) is the key issue here. Depending on the way of selecting transformation matrix, feature extraction methods can be grouped into two categories: supervised and unsupervised. If a priori information about class label is available, then supervised methods [1] are chosen. Otherwise, when the class label information is unknown, unsupervised method is a way to perform feature extraction [1]. We will discuss on the existing methods in our next chapter (literature survey).

1.8 Scope of the thesis:

Feature Selection/extraction has become the focus of much research in areas of application for which datasets with hundreds or thousands of variables/features are available. To reduce high dimensionality of such data, there are methods which involve only either feature selection or feature extraction. These methods provide reduced feature subset with only either original features or transformed features. Sometimes it may be useful to have both original and transformed features in the reduced feature set. Can we have a method which provides reduced feature subset containing original as well as transformed features? Can we have a method which incorporates both feature selection and feature extraction to get the best of two? The goal of this work is to develop a method which combines both selected and extracted features and produce combined reduced features set to reduce high dimensionality.

1.9 Organisation of the thesis:

In Chapter 2 Literature survey on different dimension reduction techniques and algorithms are discussed. In Chapter 3 proposed methodology is given. Dataset descriptions, results and their analysis of the work are detailed in chapter 4 and conclusion is described in Chapter 5.

Chapter 2

Literature Survey

So many feature selection and feature extraction methods exist in literature for dimension reduction. In this chapter we will discuss the existing techniques and algorithms for feature selection and feature extraction for dimension reduction of hyperspectral images.

2.1 Feature Selection Techniques, Algorithms & Related Works

Feature selection is one approach to identify relevant features for dimension reduction. Various studies show that features can be removed without performance deterioration. In general, feature selection is a search problem according to some evaluation criterion [12]. As mentioned earlier, feature selection methods consist of two steps: feature subset generation and feature subset evaluation.

2.1.1 Feature subsets generation:

Subset of features is obtained using some optimization technique. One intuitive way is to generate subsets of features by exhaustive search but it is not used due to its higher complexity for real time data. Also Branch and Bound [15] algorithm may produce optimal solution for monotonic feature sets, but the algorithm faces a nesting effect for non-monotonic set of features. A few suboptimal methods are sequential forward selection (SFS), sequential backward selection (SBS), generalized sequential forward selection (GSFS), generalized sequential backward selection (GSBS), sequential floating forward search (SFFS), sequential floating backward search (SFBS) [1, 28, 31 and 32]. If start with an empty set and gradually add one feature at a time, a scheme called sequential forward selection can be adopted; if start with a full set and remove one feature at a time, the scheme is called sequential backward selection.

However, optimal search techniques [15] are not applicable for subset generation in hyperspectral images due to their high complexity. Various suboptimal feature selection

techniques in hyperspectral images use sequential search techniques in subset generation where features are added/ removed sequentially [1, 32 and 33]. In [33], SFS algorithm is used as subset selection algorithm and in spite of using classification accuracy, due to higher computational complexity, class spectral signatures are used in the objective function. Su. et. al. [35] proposed a feature selection algorithm by using SFS as searching technique and the minimum estimated abundance covariance (MEAC) as objective function. These methods have a tendency to get stuck to local optima. They do not perform better in case of hyperspectral images due to the existence of strong correlation among the features [34, 35]. Some techniques for subset generation follow stochastic search methods, e.g., evolutionary based or swarm based algorithms. Stochastic method incorporates randomness into the search procedure to escape from local optima. Evolutionary algorithms [35] are a broad class of stochastic optimization algorithms, which require only the objective function information.

2.1.2 Feature subsets evaluation:

In feature subset evaluation step, each selected subset is evaluated by an evaluation criterion. An optimal subset is always relative to a certain evaluation criterion (i.e., an optimal subset chosen using one evaluation criterion may not be the same as that using another evaluation criterion). Evaluation criteria can be broadly categorized into two groups based on their dependence on the learning algorithm applied on the selected feature subset: filter model and wrapper model. The filter model relies on general characteristics of the data and tries to evaluate the goodness of a feature or feature subset without the involvement of a learning algorithm (for example classification algorithm). Here, the criterion function evaluates features subsets by using various separability measures or their information content. Some of the independent criteria are distance measure, information measure, dependency measure. Separability among classes is measured by inter-class distance (as an example, by Euclidean distance). Separability of classes can also be measured probabilistically where the main aim is to optimize a discriminating function, such as Mahalanobis distance [32], Bhattacharya distance [32], Cheronoff distance [32] etc.. On the other hand, information content based measurements, basically, calculate some statistical properties of features, like, correlation, mutual information, divergence etc.. The wrapper model, on the contrary, requires one predetermined classification algorithm and uses its performance as evaluation criterion. It searches for features which better suit the classification algorithm in terms of performance [1]. For supervised model, the primary goal of classification is to maximize predictive accuracy; while for unsupervised model, there exist a number of heuristic criteria for

estimating the quality of clustering results, such as cluster compactness, scatter separability, and maximum likelihood [12]. Wrapper model tends to be computationally more expensive than filter model [1]. Filter methods are faster than wrapper methods, because it depends on some type of estimation of the importance of individual features or subset of features. On the other hand, wrapper methods are more accurate as the importance of feature subsets is measured using a classification algorithm [36]. In [36] the goodness of subset of features is measured in wrapper approach in subset evaluation step. Various stochastic optimization based feature selection techniques for hyperspectral images are found in literature e.g., genetic algorithm (GA), particle swarm optimization (PSO) [37], ant colony optimization (ACO) [37], differential evolution [37] etc.. Genetic algorithm based feature selection methods for hyperspectral images use binary encoding where fitness value of each chromosome is calculated either by filter method or by wrapper method [1]. In [37] guaranteed convergence binary PSO (GCBPSO) with wrapper model is used. In [14] Datta et. al. have used Self-adaptive differential evolution (SADE) for searching feature subset and Fuzzy kNN classifier is used to calculate classification accuracy which is used as the evaluation criteria. For removing the redundant features, ReliefF method as a feature estimating technique is used.

Existing methods for selecting features of hyperspectral images in unsupervised framework [38] are categorized in various ways. One of the ways is ranking based method [38]. Finding a set of distinctive and informative features, depending on some statistical criteria (e.g., Kullback-Leibler divergence), is the basic idea behind ranking based feature selection methods. Some ranking based methods for hyperspectral images, present in the literature, are information divergence (ID) based method [aloke 45], maximum variance based principle component analysis (MVPCA) [46] etc.. In [34], linear prediction using representative pixels is calculated to select features in an unsupervised way. Assigning weightage to features using pair-wise separability criterion and removal of redundant features using phase correlation are the key concepts in [48]. Li et. al. [37] proposed a method to select features those better preserves the clustering/discriminant information of the data via group sparsity in a linear regression model. In [17, 18, 19] ReliefF algorithm is used for ranking the features according to their importance and preliminary filter high-dimensional features from the feature set.

2.2 Feature Extraction Techniques, Algorithms & Related Works:

Feature extraction is a process that extracts a set of new features from the original features through some functional mapping [1]. Assuming there are D features (or attributes) $A_1, A_2, \dots, A_D$, after feature extraction, we have another set of new features $B_1, B_2, \dots, B_d$, where $B_i = F_i(A_1, A_2, \dots, A_D)$, and F_i is a mapping function. Intensive search is generally required in finding good transformations. The goal of feature extraction is to search for a minimum set of new features via some transformation. There are mainly two ways of mapping/ transforming original features to new features, linear and nonlinear mapping. Transformation can be done depending on the number of required new features and also by ensuring that the true nature of data remains same after transformation.

Two most popular feature extraction techniques are principle component analysis (PCA) and linear discriminant analysis (LDA). PCA [13] transforms the original data into the transformed space where the maximum variance of data is within the first few orthogonal directions. In this method eigenvalues and corresponding eigenvectors are calculated over the covariance matrix of the data. Fisher's LDA [19, 28] is a feature extraction technique which transforms the original set of data into a set of data with lower dimensionality where the class discriminating information is maximized. The main aim of this transformation is to maximize the ratio of between-class variance to within-class variance.

Supervised feature extraction methods for hyperspectral images are divided into different categories. First one is based on clustering or grouping of features. Others follow the transformation of features based on Rayleigh criterion [19]. Prototype space (PS) feature extraction with independent component (PSFE-IC), clustering based feature extraction (CBFE), best-based feature extraction [39], piecewise constant function approximation based feature extraction spectral channels extraction [1] etc. are example of this category. In PSFE-IC method, each band is represented using the reflection properties of classes. Then, Fuzzy C-means clustering operation is performed over features in prototype space to cluster the channels [1]. CBFE is a supervised feature extraction method which, associated with each feature, considers a vector containing the mean values of all the classes of the corresponding feature, and then clustering operation is followed to group the similar features in one cluster. The number of clusters is equal to the number of extracted features [1]. The best-based feature extraction algorithm [40] combines subsets of adjacent features into a smaller number

of features retaining the original spectral information either by top-down or bottom-up approach. The top-down algorithm recursively partitions the feature into two (not necessarily equal) sets of features and then replaces each final set of features by its mean value. The bottom-up algorithm builds an agglomerative tree by merging highly correlated adjacent features and projecting them onto their Fishers direction, yielding high discrimination among classes. In [39], Jensen and Solberg divides the spectral curves into contiguous regions by piecewise constant function approximation. The extracted constants are then used as new features, which are simple average of contiguous spectral features allowing straightforward interpretation.

The other methods of supervised feature extraction are based on the maximization of the ratio of between-class variance to an average within-class variance. Fisher's LDA [13], discriminant analysis feature extraction (DAFE), modified Fisher's linear discriminant analysis (MFLDA) [28] are some well-known methods for supervised feature extraction which maximize the Rayleigh coefficient. One of the popular unsupervised feature extraction techniques rely on principle component analysis (PCA). This method transforms the original data into another space, where maximum variance of data is an orthogonal direction [14]. PCA is an orthogonal basis transformation with the advantage that the first few principle components preserve more variance than that is preserved by any other orthogonal directions [1]. Genetic algorithm based selective PCA (GA-PCA), segmented principal component transformation (SPCT), PCA based extraction method [1] are examples of this type of techniques. Segmented principle component transformation (SPCT) [6] partitions the complete set of bands into several highly correlated subgroups and then performs PCA based transformation on each subgroup.

A supervised neural network approach is Multilayer perceptron [12] having a single hidden layer can be used to extract new features. Here each neuron can be seen as a feature extractor that measures the sensitivity of inputs and internal weights between input and hidden layer, and then transforms the inputs using a nonlinear function. The idea is to use the hidden units as newly extracted features. The predictive accuracy is estimated and used as the performance measure/ the feature extraction criteria. The transformation from input units to hidden units is non-linear. In [20] feature extraction by multilayer perceptron is presented for manipulating the internal representation of input images. In a deep neural network, each layer represents a description of the input image. In a recurrent network, the representation is given by the sequence of internal states computed by the network.

Combination of selected and extracted features of hyperspectral images [11, 12] is done in various ways. In [11], Jia et. al. have proposed a method where discrete wavelet transform (DWT) and affinity propagation (AP) are used for feature extraction and feature selection respectively. DWT and AP are combined together to accomplish the dimension reduction task. DWT-based feature extraction is applied to original hyperspectral data to acquire the wavelet coefficients on different scales and AP-based feature selection is utilized to choose the exemplars from the extracted features. K-nearest neighbor (KNN) classification algorithm is adopted to examine the efficiency of the chosen exemplars. In [12], Sreevani & C.A Murthy has proposed a simultaneous feature selection and feature extraction method. In the first stage correlation coefficient is used for elimination of features from all possible pairs of features and thus remaining features are selected. In the second stage for each pair of random features, first principal component of the pair is taken to form reduced feature set and the pair is discarded.

Chapter 3

Proposed Methodology

Dimension reduction of hyperspectral images has been done using feature extraction and feature selection methods independently as we have seen in the previous chapter (Chapter 3). In this chapter an approach on the combination of feature selection and feature extraction and few strategies of reduction of combined features set to get reduced combined features set is given. To do that an unsupervised correlation based feature selection [18], a filter based feature selection [24], a sequential feature selection [1, 26] and a supervised maximum margin based feature extraction algorithm [19] is used. At first using the extraction algorithm [19] a reduced extracted feature set and using the selection algorithm [18] a reduced feature subset is obtained. Then both the feature sets are combined and finally obtained reduced combined feature set using three different ways. In one strategy, combination has been considered in an exhaustive manner while in other two approaches ReliefF [24] and SBS [1, 26, and 27] feature selection algorithm is used. It has been observed that the reduced combined features set improve classification accuracy as compared to using them in an isolated manner (either feature selection or feature extraction). Using the above mentioned feature selection [18] and feature extraction [19] algorithm we have obtained certain sets of reduced features. The number of extracted and selected features lies in a range and the range depends on the image under consideration.

Let's say d and e are the number of selected and extracted reduced features respectively where $1 \leq d, e \leq r$ (r varies from dataset to dataset). In the exhaustive method, for each extracted features set (say e), d number of features are combined where, d ranges from 1 to 10. That means for each extracted features set there are 10 combined features sets are

obtained. While reporting results, only those combined features set are considered which provides higher classification accuracy as compared to the set of extracted features of same cardinality. In other two approaches, while executing ReliefF [24] and SBS [1, 26 and 27] for reduction of combined feature sets, r_1 number of selected and r_1 number of extracted features are taken into consideration (where value of r_1 may vary from images to images). For the reference purpose all the algorithms that we have used in this work are described in the subsequent sections. After that the combination strategy and the reduction process on the combined features set is discussed in details.

3.1 Feature Selection Algorithm (Unsupervised Correlation based Feature Selection)

To do feature selection, we have used the same algorithm as developed in [18].The algorithm is an unsupervised method where the ground truth value for an instance of the dataset is not needed. Correlation and discriminating capability are the two methods that are considered to eliminate one band iteratively, where Correlation is a mutual relationship or connection between two or more things (here two) and discriminating capability is a property of measuring a device or a test to categorize objects according to their measured values.

However, initially a pair of band images is selected which has high correlation as compared to all other pairs. Then the less discriminant band image between these two is removed. The relationship between two band images is calculated by partitioning each band image into smaller blocks and then averaging the correlation coefficient of the blocks of the two images. One band image is eliminated in each iteration until desired number of band images is left. So clearly there are three steps involved to do the job: In the first step each band image is partitioned into smaller blocks. In the next step, two bands which have the highest correlation between them are searched using partitioned band image correlation. In the third step, less discriminant band, in terms of capacitory discrimination, is removed from these two highly correlated bands. The 2nd and 3rd steps are repeated until the desired number of bands is left. The detailed activity of these three steps is discussed below.

Step 1: ***Each band image is partitioned into smaller blocks***.

Let B_k be a band image of size $M \times N$, where M is the number of row pixels and $N$is the number of column pixels.

M is divided into n_r parts and N is divided into n_c parts.

So band B_k is partitioned into n_b number of blocks: $n_b = n_r \times n_c$.

Now let us assume that the $(i,j)^{th}$ block of the k^{th} band image B_k be represented as:

$$W_k(i,j), \text{where } i = 1,2,\dots,n_r,$$
$$j = 1,2,\dots,n_c \text{ and}$$
$$k = 1,2,\ \dots,D$$

Then the size of each block $= M_r \times N_c$ where $M_r = \frac{M}{n_r}$ and $N_c = \frac{N}{n_c}$.

Each block contains pixels of band image B_k for all $x = 1,2,\dots,M_r$ and $y = 1,2,\dots,N_c$ as follows:

$$W_k(i,j) = B_k(iM_r + x, jN_c + y)$$

Each block is of square shaped when $M_r = N_c$.

Step 2: ***Two highly correlated band images are selected.***

Let the k^{th} and the $(k+1)^{th}$ band images each of size $M \times N$ be B_k *and* B_{k+1} and μ_k & μ_{k+1} be mean of them.

The correlation coefficient between these two bands is calculated by partitioning the images into smaller blocks and the average of the block-wise correlation coefficient between them is calculated as follows.

$$CC_{k,k+1}{}^{(i,j)} = Correlation\_Coefficient\ (W_k{}^{(i,j)},\ W_{k+1}{}^{(i,j)})$$

$$CC_{k,k+1}{}^{(i,j)} = \frac{\sum_{x=1}^{M_r}\sum_{y=1}^{N_c} diff_k^{i,j}(x,y) \times diff_{k+1}^{i,j}(x,y)}{\sqrt{(\sum_{x=1}^{M_r}\sum_{y=1}^{N_c} squ\_diff_k^{i,j}(x,y))(\sum_{x=1}^{M_r}\sum_{y=1}^{N_c} squ\_diff_{k+1}^{i,j}(x,y))}}$$

$$where, diff_k{}^{i,j}(x,y) = (B_k(iM_r + x, jM_c + y) - \mu_k^{i,j}).$$

$$diff\ {}_{k+1}^{i,j}(x,y) = (B_{k+1}(iM_r + x, jM_c + y) - \mu_{k+1}^{i,j}).$$

$$squ\_diff_k^{i,j}(x,y) = (B_k(iM_r + x, jM_c + y) - \mu_k^{i,j})^2.$$

$$squ\_diff_{k+1}^{i,j}(x,y) = (B_{k+1}(iM_r + x, jM_c + y) - \mu_{k+1}^{i,j})^2.$$

Here, $\mu_k^{i,j}$ and $\mu_{k+1}^{i,j}$ are mean of block (i, j) of band images B_k and B_{k+1} respectively and these are defined as:

$$\mu_k^{i,j} = \frac{1}{M_r \times N_c} \sum_{x=1}^{M_r} \sum_{y=1}^{N_c} B_k(iM_r + x, jM_c + y)$$

$$\mu_{k+1}^{i,j} = \frac{1}{M_r \times N_c} \sum_{x=1}^{M_r} \sum_{y=1}^{N_c} B_{k+1}(iM_r + x, jM_c + y)$$

Now considering the block-wise correlation coefficient, denoted as $\Gamma_{k,k+1}$ between band images B_k and B_{k+1} is computed as follows:

$$\Gamma_{k,k+1} = \frac{1}{n_r \times n_c} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} CC_{k,k+1}^{i,j}$$

Only magnitude of correlation coefficient is taken into consideration. So the range of CC is between 0 and 1. If the value is close to 1 then the two images are highly correlated. For each band k, correlation with its neighboring band k+1 is calculated and stored in a list. From the list the highly correlated band image pairs will be selected for the next step.

Step3: ***Eliminate less discriminate band image using capacitory discrimination (CD)***

Discriminating capability of a band image is measured by the deviation of the normalized probability distribution of a band image from the corresponding Gaussian probability distribution with the same mean and variance. The deviation is measured by capacitory discrimination of the band image with its corresponding Gaussian image.

Let us say, the probability distribution for the k^{th} band image is P_k that calculates the normalized histogram of the image and Gaussian distribution Q_k with the mean μ and standard deviation σ is $\frac{1}{\sqrt{2\pi} \times \sigma 2} \times e^{-(x-\mu)2}$; where x is a pixel of a band image.

The point probabilities corresponding to P_k and Q_k are denoted as p_{ki} and q_{ki} respectively.

Now the KL-Divergence (Kullback-Leibler divergence) between P_k & Q_k is calculated as follows:

$$KL - Divergence\ (P_k \,||\, Q_k) = \sum_i p_{ki} \times log\left(\frac{p_{ki}}{q_{ki}}\right)$$

If the mean probability distribution between P_k and Q_k is G_k $then$ G_k is calculated as:

$G_k = \frac{1}{2}(P_k + Q_k)$ and point probability $g_{ki} = \frac{1}{2}(p_{ki} + q_{ki})$.

Capacitory discrimination (CD) of two probability distributions P_k & Q_k is:

$$CD(P_k, Q_k) = \sum_i p_{ki} \times log\left(\frac{p_{ki}}{g_{ki}}\right) + \sum_i q_{ki} \times log\left(\frac{q_{ki}}{g_{ki}}\right) \text{ where } CD >= 0$$

If CD is close to 0 then P_k indicates less deviation from Q_k i.e., P_k has less discriminatory capability and vice versa.

Step 2 & Step 3 are iterated until the desired number of band images is left in the set. The pseudo-code of the steps for this method is given below.

$Feature\_selection(D, d)$

$Start\ with\ the\ hyperspectral\ image\ of\ D\ features: data_D$

$partioning\ of\ the\ image\ into\ K\ number\ of\ blocks\ of\ each\ block\ size\ b1 \times b2$

$Initialize\ set\ sel_{bands} with\ D\ features\ number: sel_{bands} = D$

$Until\ sel_{bands}\ is\ equal\ to\ d$

$For\ i\ in\ range\ of\ 1\ to\ sel_{bands}$

$For\ each\ block\ b\ in\ the\ range\ of\ 1\ to\ K$

$Calculate\ the\ average\ correlation\ coefficient\ of\ each\ block\ for\ two\ consecutive\ bands, B_k\ and\ B_{k+1}, : Avg_{CC} = CC(B_k, B_{k+1})/K$

$End\ for$

$End\ for$

$Choose\ pair\ of\ bands\ from\ sel_{bands}\ which\ has\ the\ highest\ Avg_{CC}\ value$

$For\ the\ pair\ of\ bands, B_k\ and\ B_{k+1}, compute\ the\ capacitory\ discriminatory\ values\ CD_k\ and\ CD_{k+1}, respectively$

$If\ value\ of\ CD_k > value\ of\ CD_{k+1}$

$Eliminates\ the\ band\ B_{k+1}\ from\ sel_{bands}\ set$

$End\ if$

3.2 Feature Extraction Algorithm (supervised maximum margin based feature extraction)

For the feature extraction purpose we have implemented a supervised algorithm using partitioned maximum margin criterion (MMC) which was proposed in [19]. As the correlations between neighboring spectral bands of hyperspectral data are very high due to material spectral correlation, this property is considered during extraction of features. So the hyperspectral bands are partitioned into a number of groups of contiguous bands to use the locality (correlation) property of hyperspectral data, then a MMC [20] based transformation is used on each group to achieve the maximum discrimination among different classes. The method is supervised in nature as prior information about the class label of data is required to calculate the inter-class and intra-class scatter matrices. The MMC uses difference of

between-class (inter-class) scatter and within-class (intra-class) scatter. MMC maximizes the margin between classes.

The entire algorithm works using two steps: first calculation of correlation of neighboring bands and making groups of adjacent bands to use the local property of hyperspectral data and second linear transformation i.e., transformation of group of bands is performed to achieve the maximum discrimination among different classes using maximum margin criterion based transformation. The detailed activity of these steps is discussed below.

Step 1: ***Partitioning of Hyperspectral bands***

D number of original bands of the hyperspectral images is partitioned into K subgroups (K number of contiguous intervals). Highly correlated bands are in the same subgroup.

Let $I_1, I_2, \dots, I_k$ be the number of bands in the $1^{st}, 2^{nd}, \dots, k^{th}$ subgroups respectively. The partition follows the principle that each band is inside one block entirely. Let Γ be the correlation matrix of size $D \times D$. Each element of Γ is γ_{ij}, where γ_{ij} represents the correlation between band images B_i *and* B_j.

Let the size of each band image be $M \times N$. The correlation coefficient between B_i and B_j is computed as:

$$\gamma_{ij} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} |B_i(x,y) - \mu_i| \, |B_j(x,y) - \mu_j|}{\sqrt{\left(\sum_{x=1}^{M}\sum_{y=1}^{N}[B_i(x,y) - \mu_i]^2\right)\left(\sum_{x=1}^{M}\sum_{y=1}^{N}[B_j(x,y) - \mu_j]^2\right)}}$$

where μ_i and μ_j are the mean of band images B_i *and* B_j respectively.

$|B_i(x,y) - \mu_i|$ measures the difference between the reflectance values of pixels (x, y) from the mean value of the image.

It has been observed that the correlation between neighboring spectral bands are generally higher than for bands further apart. Partitioning is performed based on the results obtained by considering correlations whose absolute value exceeds a given threshold. Each value of the correlation matrix is compared with a threshold. If the magnitude is greater than the threshold value (denoted by Θ) then replace the value by 1 otherwise by 0. The value of Θ has been determined depending on the values of average correlation (μ_{corr}) and standard deviation (σ_{corr}) of correlation matrix Γ as

$$\theta = \mu_{corr} + \sigma_{corr}$$

where $\mu_{corr} = \frac{1}{D^2}\sum_{i=1}^{D}\sum_{j=1}^{D}\gamma_{ij}$ and $\sigma_{corr} = sqrt\left(\frac{1}{D^2}\sum_{i=1}^{D}\sum_{j=1}^{D}(\gamma_{ij} - \mu_{corr})^2\right)$

Thus the image of the threshold correlation matrix will be a binary image (0 value for black and 1 value for white) with the square blocks (highly correlated region) of white color in diagonal direction. These square blocks of white color are treated as a subgroup or partition of bands. Image of the correlation matrix of the Indian pines hyperspectral data set is shown in figure 3.1 (a) & (b) where the white patches indicates higher correlation. Now, maximum margin criterion-based transformation is conducted on each subgroup of data.

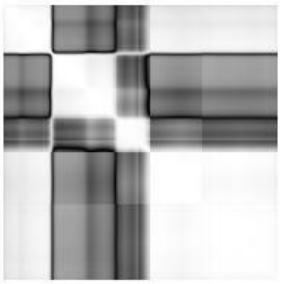


Fig. 3.1 (a): Grayscale image

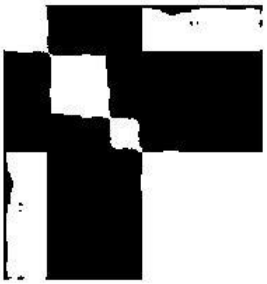


Fig. 3.1 (b): Binary image

Step 2: ***Extraction of features using maximum margin criterion (MMC)***

Let $s_i \in \Re^D, (i = 1,2, \dots, n)$ be D-dimensional patterns/pixels and $c_i \epsilon \{1,2, \dots, c\}$ be the associated class labels, where n and c denote the total number of samples and classes respectively. The main aim is to keep similarity or dissimilarity information intact as much as possible after transforming x_i from $\Re^D$ to $\Re^d$, where $d<<D$. The characteristic of a good feature extractor is to maximize the distances between classes after the transformation of features. Hence the feature extraction criterion is defined as: Maximization of $J = tr(S_b - S_w)$ where S_b and S_w are denoted as the between class and within class scatter matrix respectively. Function tr calculates the trace of a matrix.

This method transforms sample s into t where $s \in \Re^D$ and $t \in \Re^d$; $d<<D$ with a transformation matrix W i.e., $t = W.s$. The objective of the transformation matrix, $W \in \Re^{D \times d}$, is to maximize

$$j(W) = tr(W^T(S_b - S_w)W); \text{ where } W^T \text{ is the transpose of } W \qquad (1)$$

where $j(W)$ is maximized when W is composed of the first d largest eigenvectors of $(S_b - S_w)$. So, the optimal projection axes $w_1, w_2, \dots, w_d$ can be selected as the orthonormal eigenvectors corresponding to the first d largest eigenvalues denoted as, $\lambda_1, \lambda_2, \dots, \lambda_d$, i.e., $(S_b - S_w)w_j = \lambda_j w_j$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$.

There are K blocks and each block has I_k number of features, where $k = 1,2, \dots, K$. The MMC-based transformation is applied on each of this blocks. Let the linear mapping matrix for k^{th} block be $W_k \in \Re^{I_k \times d_k}$, which transforms a dataset of I_k dimension to d_k dimensions. The desired number of features is $d = \sum_{k=1}^{K} d_k$. So the aim is now to optimize the equation (1) for each block, i.e., to form W_k for k^{th} block so as to optimize MMC.

To determine the value of d_k (i.e., how many eigenvectors will be selected from each block), the eigenvectors with their corresponding eigenvalues from each group are considered at first. Then the ratio of eigenvalues with overall eigenvalues of that block is calculated for each eigenvector, i.e., for each eigenvector w_i^k , the corresponding ratio of eigenvalue Λ_i^k is calculated by $\Lambda_i^k = \frac{\lambda_i^k}{\sum_{j=1}^{I_k} \lambda_j^k}$

The eigenvectors (w_i^k) are rearranged or sorted in descending order, depending on the values of ratio of eigenvalues (Λ_i^k). The pseudo-code of the steps of this method is given below.

$Feature\_extraction(D, d)$

$Start\ with\ the\ hyperspectral\ image\ of\ D\ features: data_D$

$Calculate\ the\ correlation\ coefficient\ of\ each\ feature, i, with\ every\ other\ features$

$$T_{ij} = \sum_{i=1}^{D}\sum_{j=1}^{D} Corr(i,j)$$

$Calculate\ the\ average\ of\ the\ correlation\ matrix\ T_{ij}\ as: avg_{corr} = T_{ij}/(D \times D)$

$Calculate\ the\ standard\ deviation\ of\ the\ correlation\ matrix, sig_{corr}$

$Calculate\ the\ threshold\ using\ the\ avg\_corr\ and\ sig\_corr, T_{theta}$

$Using\ T_{theta}\ determine\ the\ binary\ correlation\ matrix, as, T_{ij}^{binary}$

$Find\ the\ number\ of\ blocks\ from\ the\ matrix\ T_{ij}^{binary}, as, K$

$For\ b\ in\ range\ of\ 1\ to\ K$

$Calculate\ the\ covariance\ matrix\ from\ the\ correlation\ matrix, as, T_{ij}^{cov}$

$Calculate\ eigenvalues\ and\ corresponding\ eigenvectors\ of\ each\ block, as,$

$$Eigval_b\ and\quad Eigvec_b$$

$End\ for$

$Select\ d\ number\ of\ eigenvalues\ such\ that, Eigval_1 > Eigval_2 > \cdots > Eigval_d$

$Calculate\ transformation\ matrix\ W, containing\ Eigvec_1, Eigvec_2, \ldots, Eigvec_d$

$$(eigenvectors\ of\ corresponding\ eigenvalues)$$

$Transformed\ features, D_{trans} = W \times data_D$

As in this work, two feature selection algorithms (ReliefF [24] and Sequential backward selection (SBS) [26]) are used to obtain the reduced combined features a brief description of these algorithms is given in the next two sections.

3.3 Feature Selection Algorithm: ReliefF

ReliefF algorithm [24] is one of the most successful filtering based feature selection methods. The algorithm calculates a feature score for each feature and ranks them accordingly. The top scoring features (i.e., having higher ranks) are selected as the important ones. The key idea of the algorithm is to estimate the quality of the features depending on how well their values help to classify the patterns. When this algorithm is used in the field of hyperspectral images, image bands are represented as the features.

Suppose a dataset with n instances and D features having C number of different classes. The algorithm selects a pattern, $R_i (i\epsilon 1,2,..,n)$, and searches for p number of nearest neighbors from the class to which it belongs, and this set of p patterns is termed as nearest hits (denoted as H_j, where, $j = 1,2,..,p$). The algorithm also searches for p number of nearest neighbors from each of the classes except the one to which it belong, and this set of p patterns is called nearest misses for class label C (denoted as $M_j(C)$, where, $j = 1,2,\dots,p$). The algorithm updates the feature quality estimation (the feature score value) $W(f)$ for all features, f, depending on their values for pattern R_i, hits H_j, and misses $M_j(C)$. If the pattern R_i and hits H_j have different values for a feature f, then the corresponding feature separates the two patterns even if they belong to the same class, and it is not desirable. As a result the value of $W(f)$will decrease. On the contrary, if the patterns R_i and M_j have different values for the attribute f, then the said feature can well discriminate the said two patterns, thereby increasing the value of $W(f)$ [24]. Such increments/ decrements will be weighted with a priori probability of the class under consideration.

The pseudo-code of the algorithm is given in Algorithm 3.

Algorithm 3:

$ReliefF\_Algorithm(D)$

1. $Initialize\ all\ feature\ weights\ as\ 0: W(f) = 0; where\ 1 \leq f \leq D$
2. $for\ i = 1\ to\ n\ do$
3. $\quad select\ a\ pattern\ R_i.$
4. $\quad find\ p\ nearest\ hits\ H_j.$
5. $\quad for\ each\ class\ c(\in C) \neq class(R_i) do$
6. $\quad\quad from\ class\ c\ find\ p\ nearest\ misses\ M_j(c).$
7. $\quad end\ for$
8. $\quad for\ each\ feature, f = 1\ to\ D\ do$
9. $$W(f) = W(f) - \sum_{j=1}^{p} \frac{diff(f, R_i, H_j)}{n.p} + \sum_{c \neq class(R_i)} \left[\frac{P(c)}{1 - P(class(R_i))} \sum_{j=1}^{p} \frac{diff(f, R_i, M_j(c))}{n.p} \right]$$
10. $end\ for$

Here, in this algorithm, a function $\left(diff(f, R_i, R_j)\right)$ is used that denotes the difference between values of the features f for two patterns R_i and R_j . The function $diff(f, R_i, R_j)$ is defined as:

$$diff(f, R_i, R_j) = \frac{|value(f,\ R_i)\ - value(f,\ R_j)|}{\max(f) - \min(f)}$$

The user-defined parameter, p controls the locality of the estimates and ensures greater robustness of the algorithm.

3.4 Sequential Backward Feature Selection Algorithm (SBS)

Sequential backward feature selection algorithm [1, 26] is a greedy search algorithm that is used to reduce an initial *D*-dimensional feature space to a *d*-dimensional feature subspace where $d \ll D$. It is a top down approach which starts with the complete set of features, F and one feature is discarded at a time until $(D - d)$ features have been deleted. The criterion of discarding a feature is to maximize the objective function J, of the subset of features. Mathematically, this method can be defined as follows. Let $F = \{f_1, f_2, \dots, f_D\}$ be the given set of *D* features from where $(D - d)$ features has to be deleted to obtain subset of d features. Let A_{D-k} be the set of $D - k$ features at step k. Let, a feature $f^0 \in A_{D-k}$ be chosen such that $J(A_{D-k} - \{f^0\}) \geq J(A_{D-k-1} - \{f\}),\ \forall f \in A_{D-k}$. Then $A_{D-k-1} = A_{D-k} - \{f^0\}$. The algorithm starts with $A_D = F$ and the loop is repeated $(D - d)$ times. At each stage of the algorithm, the feature to be removed from the current feature set is determined by investigating statistical dependence of the features in the set. The pseudo code of the algorithm is given in algorithm 4.

Algorithm 4:

Input: Set of features $F = \{f_1, f_2, \dots, f_D\}$, size of feature set D, and size of target feature set d.

Output: Suboptimal feature subset A_d of size d.

$SBS\_Algorithm(F, d)$

1. $A_D \leftarrow F$
2. $for\ i = 1\ to\ (D - d)\ do$
3. $\quad select\ a\ feature\ f^0$
4. $\quad f^0 \in A_{D-i}\ such\ that\ J(A_{D-i} - \{f^0\}) \geq J\big(A_{D-(i-1)} - \{f\}\big),\ \forall f \in A_{D-i}$
5. $\quad A_{D-(i-1)} = A_{D-i} - \{f^0\}$
6. $end\ for$
7. $return\ A_d$

Sometimes the number of features to be deleted by the algorithm is not specified when the desired number of reduced feature space is not known. In that scenario, the algorithm continues discarding the features until the objective/ criteria function J satisfies the criteria. Such functions should be non-monotonic in nature. Depending on the criteria either it should be a maximization objective function or a minimization objective function. For the minimization objective function the algorithm continues to eliminate a feature that results either decrease in the objective function or remains same. The opposite is happened if maximization objective function is used. The pseudo-code with the minimization objective function of the algorithm is given in algorithm 5.

Algorithm 5:

Input: Set of features $F = \{f_1, f_2, \dots, f_D\}$, size of feature set D.

Output: A_{D-k} is the final reduced feature subset.

$SBS\_Algorithm(F)$

1. $Start\ with\ the\ full\ set\ A_0 = F;\ k = 0.$
2. $Remove\ the\ feature\ f^0\ such\ that\ f^0 = \text{argmin}_{f^0 \epsilon A_k} J(A_k - f^0)$
3. $Update\ A_{k+1} = A_k - f^0\ ; k = k + 1$
4. $Go\ to\ step\ 2\ until\ the\ value\ of\ the\ objective\ function\ remains\ same\ or\ decreases$

However all these algorithms discussed in section 3.1, 3.2, 3.3 and 3.4 are used in the proposed work to obtain reduced combined features set. For this purpose, three different proposed strategies/ approaches are discussed in detail in the subsequent sections.

3.5 Combinations of Extracted and Selected Features: Exhaustive Approach

In this approach, reduced combined features set is obtained by combining the selected features (obtained by using algorithm 1 in section 3.1) and the extracted/ transformed features (obtained by using algorithm 2 in section 3.2) in an exhaustive manner. Let $E = \{e_1, e_2, \dots, e_r\}$ be the set of new transformed (extracted) features and $S = \{s_1, s_2, \dots, s_t\}$ be the set of features, selected (described earlier in this chapter) from the original features set (having D number of features) where $r, t << D$. In the proposed work, for each value of the set of extracted features, E, each value of the selected features set, S, is combined together to generate combined features set. So, for r numbers of extracted features set total $r \times t$ numbers of combined features set is generated. Let say, CF be the set of new combined features set, which is defined as,

$$\text{CF} = \{(e_1 + s_1), (e_1 + s_2), \dots (e_1 + s_t), (e_2 + s_1), (e_2 + s_2), \dots, (e_2 + s_t), \dots, (e_r + s_1), (e_r + s_2), \dots, (e_r + s_t)\}$$

or

$$\text{CF} = \sum_{i=1}^{r} \sum_{j=1}^{t} (e_i + s_j)$$

where the value of r and t lie in a range and the ranges depend on the hyperspectral image under consideration. Now with each of these combined features set, a classifier is used for the pixel classification accuracy of the hyperspectral image. It has been noticed that higher classification accuracy is obtained with these combined features set. While reporting results, we have considered only those combined features set which provides higher classification accuracy as compare to the extracted (or selected) features set of same cardinality. But the disadvantage of this method is that it is time consuming, as the combination is exhaustive in nature. If the value of t is high, then the number of possible combinations increases very much. The step by step process of the proposed method is shown in figure 3.2.

$Start\ with\ the\ hyperspectral\ image\ of\ D\ features: data_D$

$For\ i\ in\ range\ of\ 1\ to\ r$

$Find\ extracted\ features\ set, e_i, using\ Feature\_extraction(data_D, i)$

$Calculate\ classification\ accuracy\ using\ a\ classifier\ with\ data_{e_i}$

$For\ j\ in\ range\ of\ 1\ to\ t$

$Find\ selected\ features\ set, s_j, using\ Feature\_selection(data_D, j)$

$Make\ combined\ dataset\ having\ combined\ features\ set: data_{(e_i+s_j)}$

$Calculate\ classification\ accuracy\ using\ a\ classifier\ with\ data_{(e_i+s_j)}$

$End\ for$

$End\ for$

Fig. 3.2: Process of exhaustive feature combination method

Here, in this method, $data_D$ is the hyperspectral image having D number of bands (features). $Feature\_selection(data_D, j)$ and $Feature\_extraction(data_D, i)$ are two methods with the desired number of features j and i, discussed in section 3.1 and 3.2, respectively. Because of the exhaustive nature, the method is time intensive. Hence, another method is proposed to combine the two different types of features (extracted and selected) and obtained combined reduced features set using two different features selection methods. For this purpose, another two approaches are given in the following two sections.

3.6 Dimension Reduction of Combined Set of Extracted and Selected Features using Sequential Backward Selection (SBS) Algorithm

In this approach, the combination of features is done by a heuristic method. Let $EH = \{e_1, e_2, \dots, e_h\}$ be the set of new transformed (extracted) features and $SH = \{s_1, s_2, \dots, s_h\}$ be the set of features, selected from the original features set (having D number of features) where $h << D$. In the proposed work, combination of features has been done by taking every

element of the set of extracted features (EH) as e_i and every element of the set of selected features (SH) as s_i, where $1 \leq i \leq h$. Thus, the set of combined features is defined as:

$$CH = \sum_{i=1}^{h} (e_i + s_i)$$

where the value of h differs from hyperspectral images to images. Now a selection method called sequential backward feature selection (SBS) is used for elimination of redundant or insignificant features on every element of CH and thus, h different reduced combined features sets are generated. In the SBS method, the elimination of features continues until any further improvement in classification accuracy is not yield. Let say, l_i be the number of features that are eliminated by SBS method for combined features set CH_i. Then the final reduced combined features set is defined as:

$$CF_{SBS} = \sum_{i=1}^{h} (CH_i - l_i)$$

Now, a classifier is used for the pixel classification of the hyperspectral image with each of the element of the set CF_{SBS} (which contains h different reduced combined features set). By eliminating the insignificant features, SBS method improves the classification accuracy. It has been seen that the final reduced combined features provides higher classification accuracy as compare to the set of extracted (or selected) feature of same cardinality. While reporting results, only those reduced combined features set are considered which provides higher classification accuracy as compare to the extracted (or selected) features set of same cardinality

The step by step process of the proposed method is shown in figure 3.3 as below:

$Start\ with\ the\ hyperspectral\ image\ of\ D\ features: data_D$
$For\ i\ in\ range\ of\ 1\ to\ h$
    $Find\ extracted\ features\ set, e_i, using\ Feature\_extraction(data_D, i)$
    $Calculate\ classification\ accuracy\ using\ a\ classifier\ with\ data_{e_i}$
    $Find\ selected\ features\ set, s_i, using\ Feature\_selection(data_D, i)$
    $Combine\ the\ extracted\ and\ selected\ features: CH_i = (e_i + s_i)$
    $Apply\ Algorithm\ 5\ (SBS\ method)\ on\ CH_i : CF_{SBS}(i) = SBS\_Algorithm(CH_i)$
    $Calculate\ classification\ accuracy\ using\ a\ classifier\ with\ data_{CF_{SBS}(i)}$
$End\ for$

Fig. 3.3: Process of dimension reduction of combined features using SBS method

In the above mentioned method, $data_D$ is the hyperspectral image having D number of bands (features). Similarly, $data_{e_i}$ and $data_{CF_{SBS}(i)}$ are the hyperspectral image having e_i and $CF_{SBS}(i)$ number of bands/ features respectively. $Feature\_selection(data_D, i)$ and $Feature\_extraction(data_D, i)$ are two methods with the desired number of features i, discussed in section 3.1 and 3.2, respectively. CH_i is the i^{th} combined features set and $CF_{SBS}(i)$ is the i^{th} reduced combined set obtained, after SBS method is applied on CH_i. Algorithm 5 is the SBS algorithm discussed in section 3.4. In this case, in SBS method, mis-classification accuracy of the classifier has been used as the objective/ criterion function.

Furthermore, the similar process has been done where another feature selection algorithm is used for the purpose of reduction of combined features. The detailed procedure is discussed in the next section.

3.7 Dimension Reduction of Combined Set of Extracted and Selected Features using ReliefF Algorithm

In this approach, the combination of extracted and selected features has been done as the similar way that has been done in section 3.6 and thus, obtained the set of combined features set $CH,$ which contains h different sets of combined features set. The combined features se CH is defined as:

$$CH = \sum_{i=1}^{h} (e_i + s_i)$$

where e_i and s_i be the number of extracted and the number of selected features for the i^{th} combined features set CH_i. Now a feature selection method called ReliefF is used on each of the combined features set CH_i , where $1 \leq i \leq h$ (the value of h varies from different hyperspectral images to images). The algorithm calculates the feature score for every feature present in the features set, CH_i. The calculation of feature score is done as the similar way as that has been discussed in section 3.3. These feature score is considered as the weight value of a feature and ranking of the feature is done based on its weight value. Larger weighted feature gets higher rank and smaller weighted feature gets lower rank. For each set of combined features CH_i, several sets of reduced combined features have been obtained, depending on the desired number of features (d). The reduced combined features set, $CF_{ReliefF}(i)$, contains d number of features $(f_1, f_2, \dots, f_d) \epsilon\, CH_i$, such that $\text{rank}(f_1) \geq \text{rank}(f_2), \dots, \geq \text{rank}(f_d)$. Therefore, the i^{th} reduced combined features set, $CF_{ReliefF}(i)$, can be defined as:

$$CF_{ReliefF}(i) = \sum_{d=1}^{size(CF_i)} \sum_{j=1}^{d} \{f_j\}$$

where the range of the desired number of reduced features (i.e., d) can be varied between 1 to size (CF) (i.e., size of the combined features set) to get different sets of reduced combined features. It has been noticed that, the classification accuracy with the reduced & combined features, obtained by this method, yield higher accuracy than the accuracy with set of

extracted (or selected) features of same cardinality. While reporting results, we have considered only those combined features set from each of the reduced combined features set which provides comparatively higher classification accuracy.

The step by step process of the above said proposed method is shown in figure 3.4 as below:

$Start\ with\ the\ hyperspectral\ image\ of\ D\ features: data_D$

$For\ i\ in\ range\ of\ 1\ to\ h$

$Find\ extracted\ features\ set, e_i, using\ Feature\_extraction(data_D, i)$

$Calculate\ classification\ accuracy\ using\ a\ classifier\ with\ data_{e_i}$

$Find\ selected\ features\ set, s_i, using\ Feature\_selection(data_D, i)$

$Combine\ the\ extracted\ and\ selected\ features: CH_i = (e_i + s_i)$

$Apply\ Algorithm\ 3\ (ReliefF\ method)\ on\ CH_i : CF_{ReliefF}(i) = ReliefF\_Algorithm(CH_i)$

$For\ d\ in\ range\ of\ 1\ to\ size(CH_i)$

$Find\ the\ features\ set, containing\ d\ numbers\ of\ top\ ranked\ features,\ CF^i_{ReliefF}(d),$

$such\ that, (rank(f1) \geq rank(f2) \geq \cdots, \geq rank(fd) \in CF_{ReliefF}(i)$

$Calculate\ classification\ accuracy\ using\ a\ classifier\ with\ data_{CF^i_{ReliefF}(d)}$

$End\ for$

$End\ for$

Fig. 3.4: Pseudo-code of the process of dimension reduction of combined features using ReliefF method

In the above mentioned method, $data_D$ is the hyperspectral image having D number of bands (features). Similarly, $data_{e_i}$ and $data_{CF^i_{ReliefF}(d)}$ are the hyperspectral image having e_i and $CF^i_{ReliefF}(d)$ number of bands/ features respectively. $Feature\_selection(data_D, i)$ and $Feature\_extraction(data_D, i)$ are two methods with the desired number of features i, discussed in section 3.1 and 3.2, respectively. Algorithm 3 is the ReliefF algorithm discussed in section 3.4.

Chapter 4

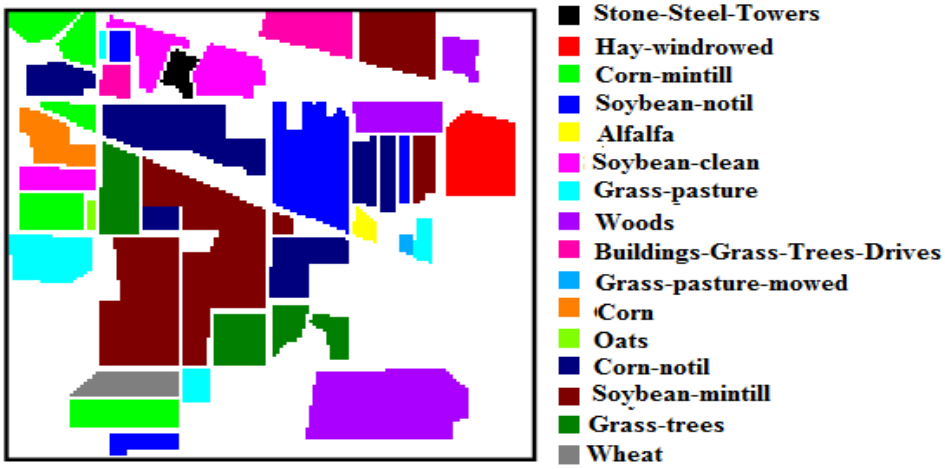
Results and Analysis

The experiment of the proposed methods has been made on three hyperspectral images, namely, Indian pines, Kennedy-space center (KSC) and Botswana. The description of each of these images is given in the next section. The detailed results and its analysis are given in tabular as well as in graphical representation form.

4.1 Datasets Used

Indian Pines – The AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) instrument is used for acquisition of the data over agricultural land of northwest Indiana's pine test site in the early growing season of 1992. Spectral range is 400-2500 nm; spectral resolution is about 10 nm, 220 spectral bands. The size of the image is 145X145 pixels and spatial resolution is 20m. 20 water absorption bands (104-108, 150-163 and 220) and 15 noisy bands (1-3, 103, 109-112, 148-149, 164-165, and 217-219) were removed, resulting in a total of 185 bands. There are 16 classes in this image. We have used 185 images in tiff format for 185 bands [16].

(a)

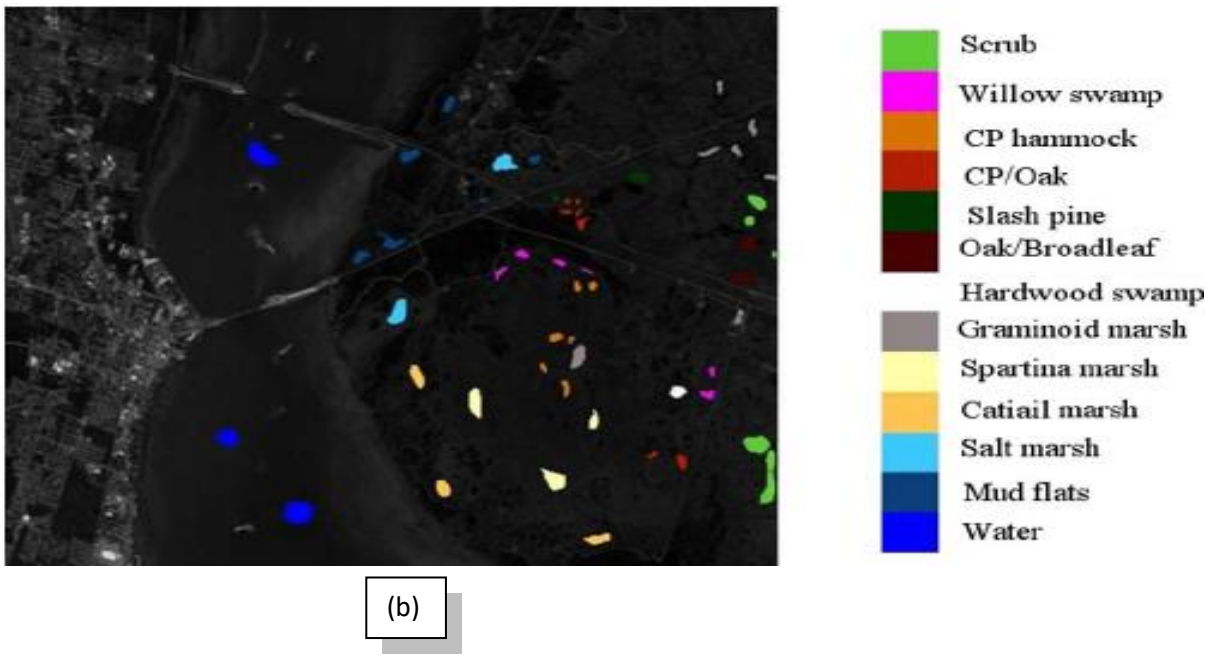


(b)

Fig 4.1: Indian pines image: (a) single band image (band 20) and (b) ground truth image

KSC – NASA AVIRIS sensors acquired data over Kennedy Space Centre on 23 March 1996. AVIRIS acquired data in 224 bands of 10 nm width from an altitude of approximately 20 km and with a spatial resolution of 18 m. The size of the image is 512X614 pixels. After removing water absorption and low signal-to-noise (SNR) bands (1–4, 102–116, 151–172, and 218–224), 176 bands were used for analysis. There are 13 land cover classes. We have used 176 images in tiff format for 176 bands [16].

(a)



(b)

Fig 4.2: KSC image: (a) single band image (band 20) and (b) ground truth image

Botswana – The NASA Earth Observing 1 (EO-1) satellite acquired a sequence of data over the Okavango Delta, Botswana, in 2001-2004. The Hyperion sensor on EO-1 acquired data at 30 m pixel resolution over a 7.7 × 44 km surface area in 242 bands from the 400–2500 nm portion of the spectrum in 10 nm windows. Uncalibrated and noisy bands (such as water

absorption bands) were removed, and the remaining 145 bands were included as candidate features (10–55, 82–97, 102–119, 134–164, 187–220). These data were acquired on 31 May 2001 and consist of observations from 14 identified classes representing land cover types in seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the Delta. We have used 145 images in tiff format for 145 bands [16].

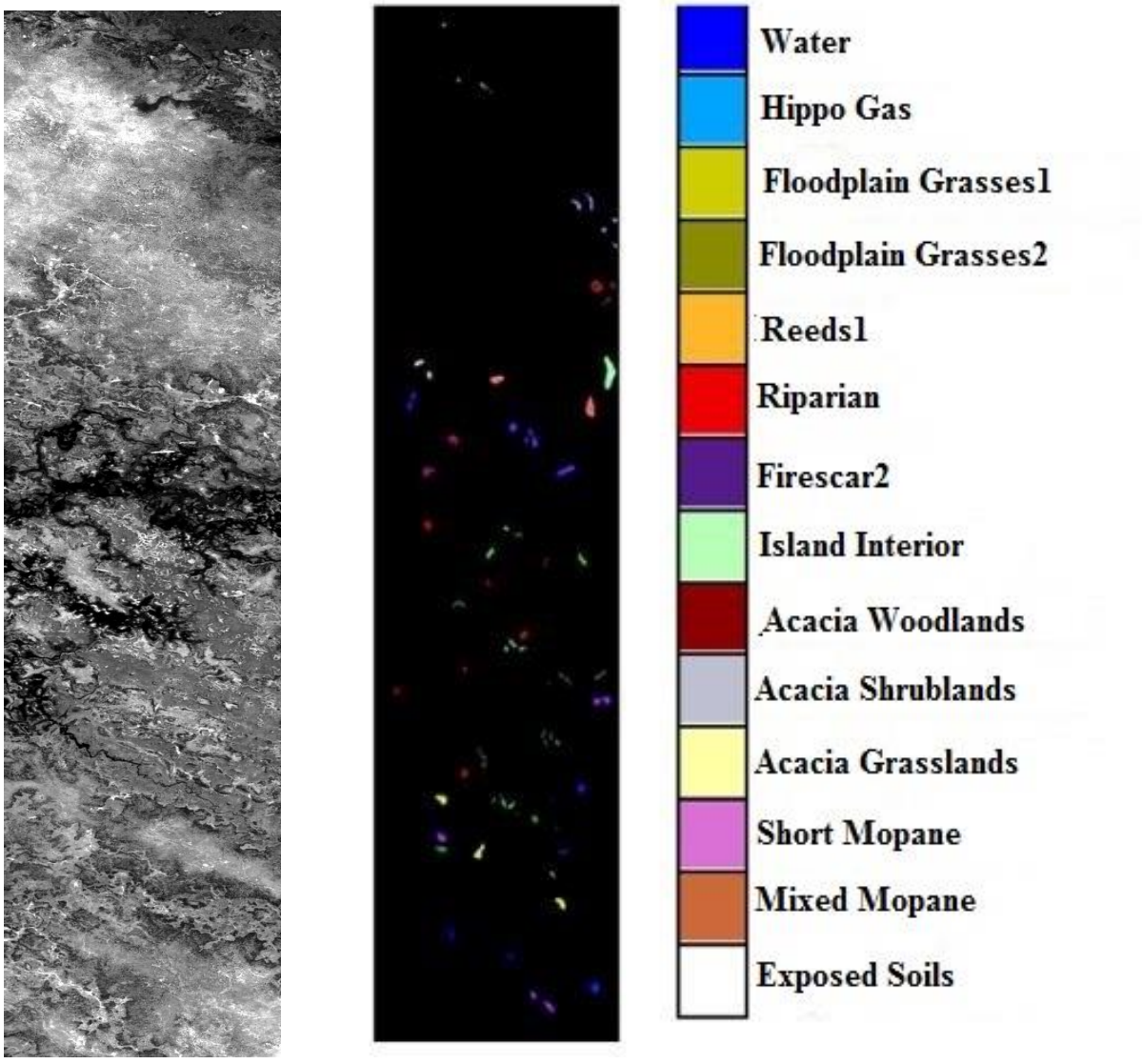


Fig 4.3: Botswana image: (a) single band image (band 20) and (b) ground truth image

4.2 Parameters Details

Following are the parameters and their corresponding values that have been taken into consideration while the experiment has been done with three hyperspectral images.

In Indian pines image:

In feature correlation based partitioning selection algorithm (described in section 3.1) – image block size: 29×29.

In maximum margin based feature extraction algorithm (described in section 3.2) – number of blocks for each image: 4; each block contains band numbers as follows: 1-33, 34-77, 78-100, 101-185

Classifier for classification – In all the proposed methods, Support Vector Machine (SVM) classifier (Multi-class, one to one, with RBF kernel function and 10 fold cross validation (see [25] for further details) is used.

In KSC image:

In feature correlation based partitioning selection algorithm (described in section 3.1) – image block size: 32×25

In maximum margin based feature extraction algorithm (described in section 3.2) – number of blocks for each image: 4; each block contains band numbers as follows: 1-37, 38-86, 87-104, 105-176

Classifier for classification – The same classifier with the same specification as used for Indian pines image is used.

In Botswana image:

In feature correlation based partitioning selection algorithm (described in section 3.1) – image block size: 36×16

In maximum margin based feature extraction algorithm (described in section 3.2) – number of blocks for each image: 6; each block contains band numbers as follows: 1-40, 41-52, 53-69, 70-87, 88-125, 126-145

Classifier for classification – The same classifier with the same specifications as used for Indian pines and KSC images is used.

4.3 Results and its Analysis

4.3.1 Results of Indian pines Images

# of bands	Accuracy using Extracted Features		# of bands (Extracted + Selected)	Accuracy using Combined Features	
	Average (%)	Best (%)		Average (%)	Best (%)
14	0.9011	0.9016	13+1	0.9022	0.9029
15	0.9034	0.9044	14+1	0.9041	0.9056
16	0.8999	0.9015	12+4	0.9043	0.9076
17	0.9012	0.9024	12+5	0.9047	0.9056
18	0.8953	0.8978	14+4	0.9061	0.9086
19	0.8952	0.8971	14+5	0.9060	0.9071
20	0.8952	0.8954	14+6	0.9054	0.9055
21	0.8973	0.8991	15+6	0.9080	0.9091
22	0.8917	0.8926	15+7	0.9059	0.9060
23	0.8926	0.8938	17+6	0.9040	0.9054
24	0.8936	0.8940	17+7	0.9052	0.9070
25	0.8908	0.8909	21+4	0.9007	0.9028
26	0.8883	0.8892	21+5	0.9012	0.9016
27	0.8894	0.8914	21+6	0.9014	0.9018
28	0.8895	0.8912	21+7	0.9012	0.9018
29	0.8899	0.8906	22+7	0.8979	0.8982
30	0.8932	0.8957	23+7	0.8995	0.9002
31	0.8916	0.8930	24+7	0.8979	0.8981
32	0.8959	0.8961	25+7	0.8964	0.8979
33	0.8961	0.8976	30+3	0.8962	0.8976
34	0.8959	0.8967	33+1	0.8983	0.8996
35	0.9006	0.9017	17+18	0.9082	0.9101
36	0.8999	0.9010	17+19	0.9071	0.9075
37	0.9016	0.9019	17+20	0.9089	0.9101
38	0.9042	0.9050	17+21	0.9084	0.9090
39	0.9074	0.9091	17+22	0.9082	0.9097
40	0.9070	0.9079	17+23	0.9114	0.9121

Table 4.1: Classification accuracy obtained using extracted and a combination of extracted and selected features (combination is done using exhaustive method) on Indian pines image

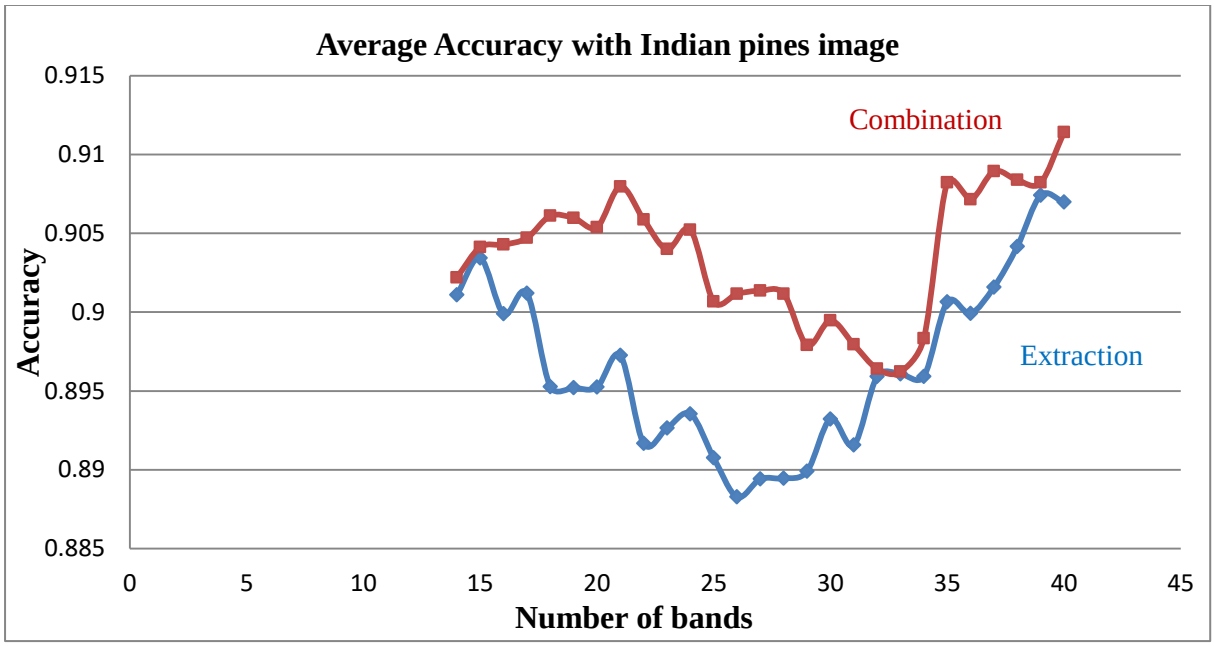


Fig. 4.4: Variation of average accuracy values with number of bands obtained using extracted and reduced combined set of features (combination is done using exhaustive method) on Indian pines image

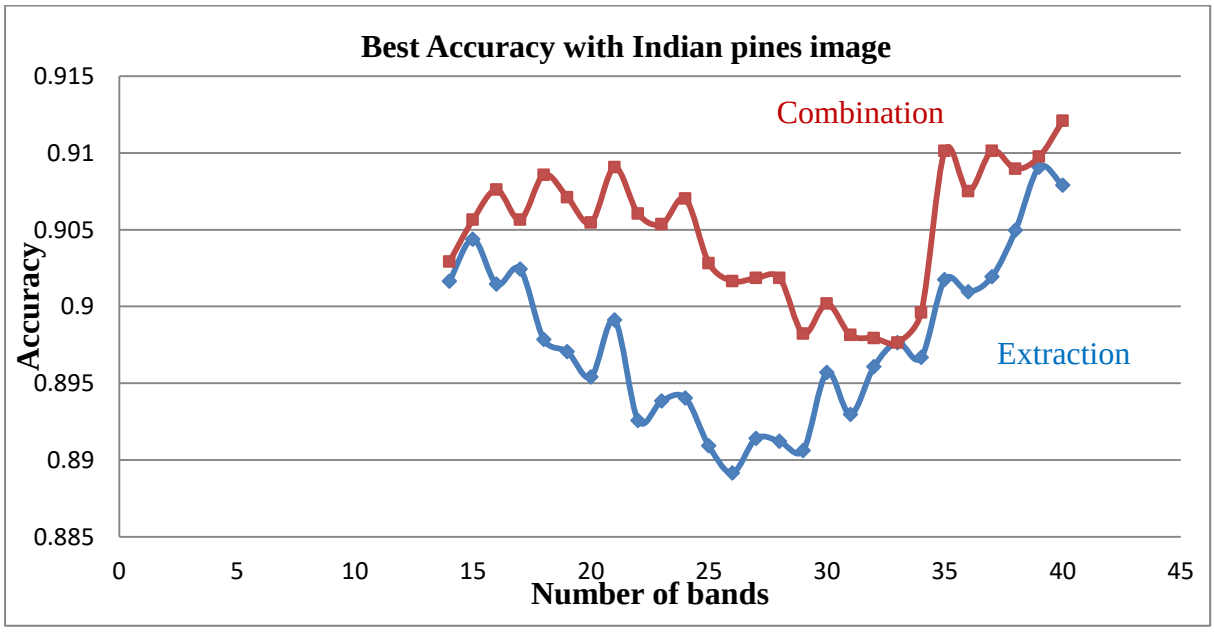


Fig. 4.5: Variation of best accuracy values with number of bands obtained using extracted and reduced combined set of features (combination is done using exhaustive method) on Indian pines image

# of bands	Accuracy using Extracted Features	# of bands (Extraction + Selection)	Accuracy using Extracted Features
8	0.8794	6+2	0.8937
10	0.8965	8+2	0.9093
14	0.9011	9+5	0.9060
16	0.8999	11+5	0.9039
19	0.8952	11+8	0.9070
25	0.8908	12+13	0.9071
27	0.8894	14+13	0.8965
30	0.8932	14+16	0.9076
31	0.8916	15+16	0.9070
33	0.8961	17+16	0.9135

Table 4.2: Classification accuracy obtained using extracted features and reduced combined set of features (reduction is done using SBS method) on Indian pines image

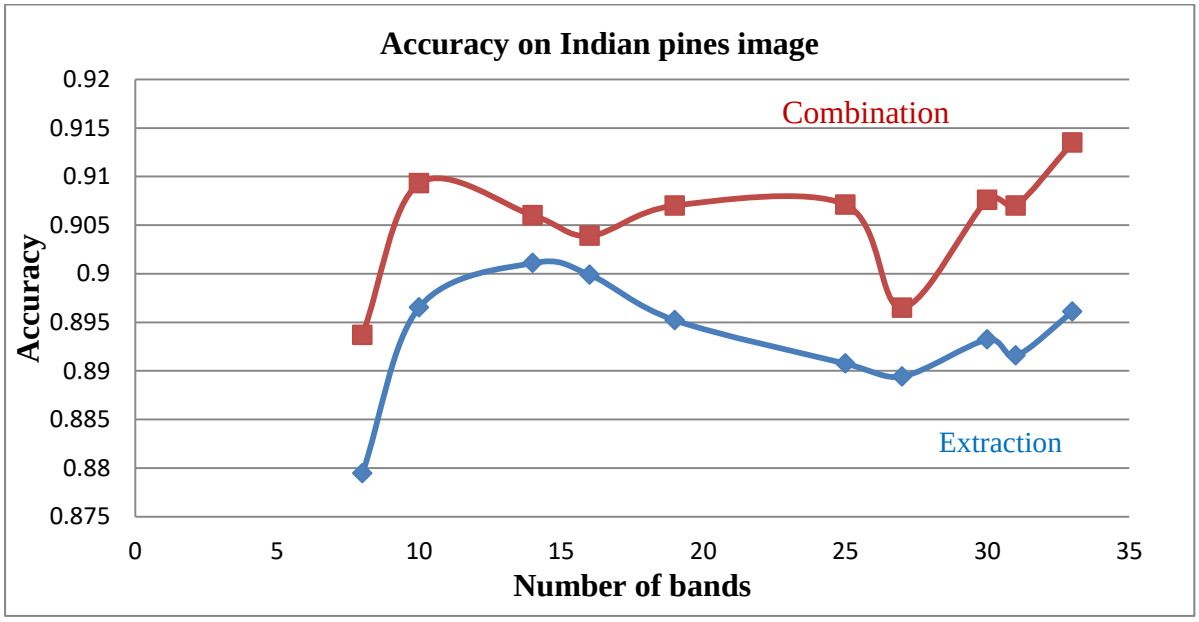


Fig. 4.6: Variation of accuracy values with number of bands obtained using extracted features and reduced combined set of features (reduction is done using SBS method) on Indian pines image

# of bands	Accuracy using Extracted Features	# of bands (Extraction + Selection)	Accuracy using Extracted Features
7	0.8456	5+2	0.8516
22	0.8917	10+12	0.8972
23	0.8926	11+12	0.9008
24	0.8936	12+12	0.9000
25	0.8908	13+12	0.9012
26	0.8883	14+12	0.9038
27	0.8894	14+13	0.9014
28	0.8895	14+14	0.9020
29	0.8899	14+15	0.9029
30	0.8932	14+16	0.9041
31	0.8916	15+16	0.9053
32	0.8959	16+16	0.9080
33	0.8961	16+17	0.9074
34	0.8959	16+18	0.9085
35	0.9006	16+19	0.9098
36	0.8999	16+20	0.9092
37	0.9016	17+20	0.9083
38	0.9042	18+20	0.9107
39	0.9074	19+20	0.9083
40	0.9070	20+20	0.9078

Table 4.3: Classification accuracy obtained using extracted features and reduced combined set of features (reduction is done using ReliefF method) on Indian pines image

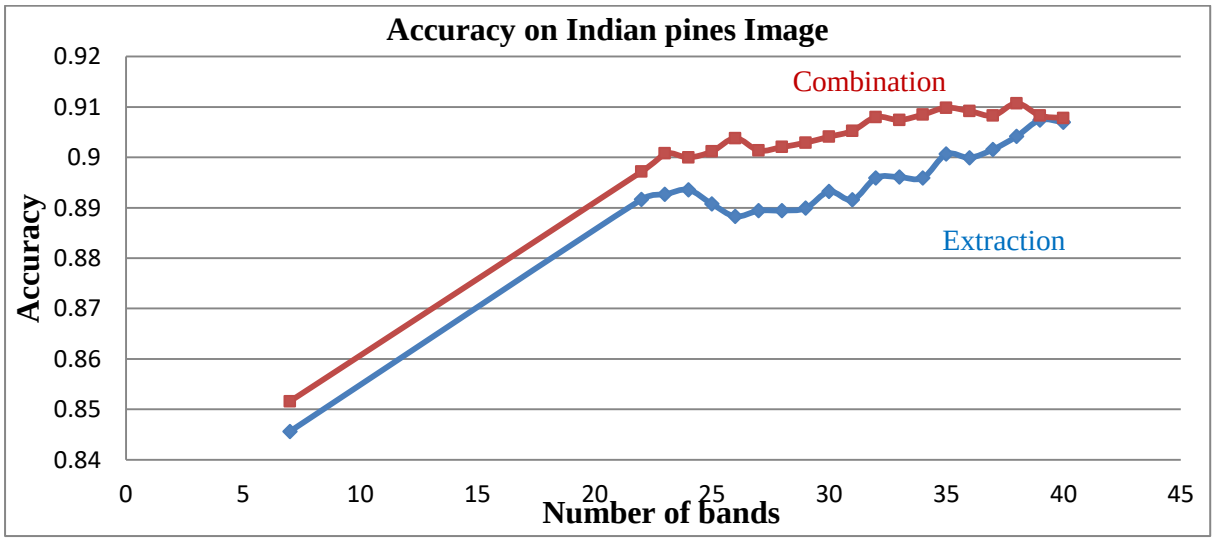


Fig. 4.7: Variation of accuracy values with number of bands obtained using extracted features and reduced combined set of features (reduction is done using ReliefF method) on Indian pines image

Analysis of results:

The pixel classification accuracy for the Indian pines image dataset with extracted features and with reduced combined features is shown in Table 4.1. The results are obtained by using the approach described in Section 3.5, where the number of original features (D) is 185, the range of extracted features varies from 4 to 40 (the value of r) and the range of selected features varies from 1 to 10 (the value of t). However, in this table, we have only provided the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. Both average and the best accuracies (over 6 simulations) are shown in Table 4.1. It is to be noted that the classification accuracy using reduced combined (extracted and selected) features is higher for all the cases when r lies in 14-40. Variation of average and best accuracy values with number of bands obtained using Indian pines dataset are shown in Figures 4.4 and 4.5, respectively. These figures also establish the superiority of using reduced combined features as compared to extracted features.

Table 4.2 shows the classification accuracy obtained using extracted features and reduced combined set of features, where the reduction on combined features is done using SBS method described in Section 3.6, where the range of extracted and selected features varies from 4 to 20 (the value of h). However, in this table, we have only provided the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. Variation of accuracy values with number of bands obtained by using SBS method for reduced combined features, are shown in Figures 4.6. From the figures 4.4 and 4.6, it has been found that this method provides higher classification accuracy as compare to the exhaustive method for this hyperspectral image. The figure also establishes the superiority of using reduced combined features as compared to extracted features. Similarly, Table 4.3 shows the classification accuracy obtained using extracted features and reduced combined set of features, where the reduction on combined features is done using ReliefF method described in Section 3.7, where the range of extracted and selected features varies from 4 to the size to combined features set. The table contains only the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. Variation of accuracy values with number of bands obtained by using ReliefF method for reduced combined features, are shown in Figures 4.7. From the figures 4.4 and 4.7, it is noticed that higher classification accuracy is obtained by this method as compare to the exhaustive method. The figure also establishes the superiority of using reduced combined features as compared to extracted features.

4.3.2 Results of KSC Images

# of bands	Accuracy using Extracted Features		# of bands (Extracted + Selected)	Accuracy using Combined Features	
	Average (%)	Best (%)		Average (%)	Best (%)
5	0.8429	0.8446	(4+1)	0.8883	0.8893
6	0.8593	0.8601	(4+2)	0.8854	0.8874
7	0.8769	0.8778	(6+1)	0.8916	0.8933
8	0.9028	0.9058	(7+1)	0.9096	0.9115
9	0.9042	0.9071	(8+1)	0.9114	0.9136
10	0.9040	0.9058	(8+2)	0.9127	0.9159
11	0.9022	0.9025	(8+3)	0.9133	0.9148
12	0.9099	0.9115	(8+4)	0.9135	0.9163
13	0.9136	0.9142	(12+1)	0.9184	0.9204
14	0.9118	0.9140	(13+1)	0.9188	0.9211
15	0.9141	0.9156	(14+1)	0.9190	0.9204
16	0.9108	0.9121	(12+4)	0.9197	0.9202
17	0.9161	0.9171	(12+5)	0.9193	0.9206
18	0.9184	0.9209	(12+6)	0.9204	0.9221
19	0.9185	0.9204	(12+7)	0.9210	0.9248
20	0.9195	0.9202	(15+5)	0.9217	0.9221

Table 4.4: Classification accuracy obtained using extracted and reduced combined of extracted and selected features (combination is done using exhaustive method) on KSC image

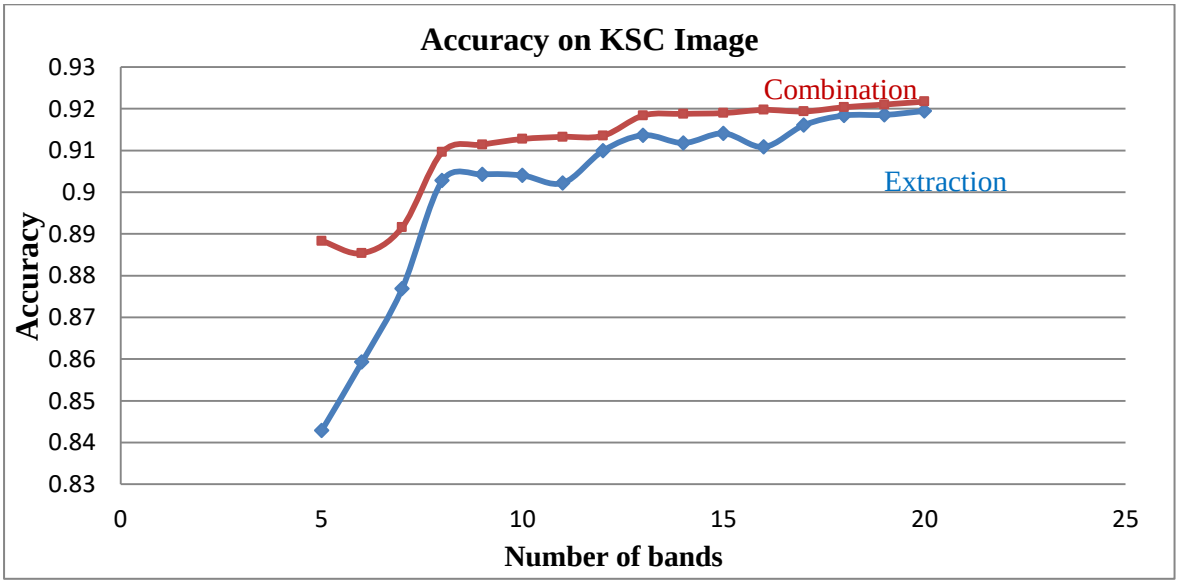


Fig. 4.8: Variation of average accuracy values with number of bands obtained using extracted and reduced combined set of features (combination is done using exhaustive method) on KSC image

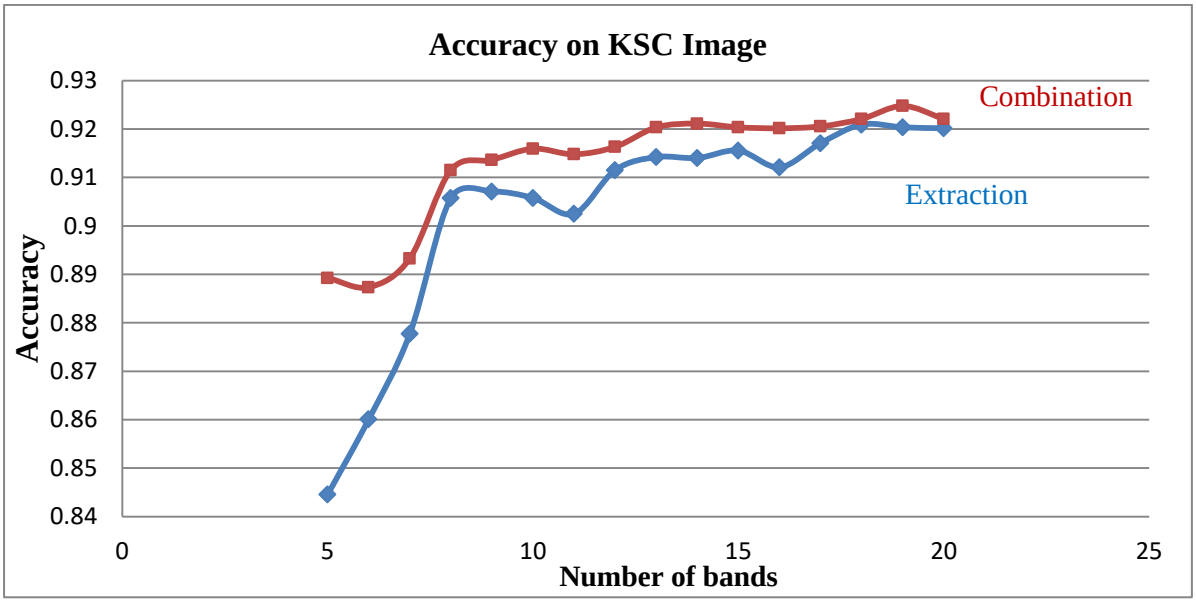


Fig. 4.9: Variation of best accuracy values with number of bands obtained using extracted and reduced combined set of features (combination is done using exhaustive method) on KSC image

# of bands	Accuracy using Extracted Features	# of bands (Extraction + Selection)	Accuracy using Extracted Features
6	0.8593	3+3	0.8987
9	0.9042	5+4	0.9156
12	0.9099	6+6	0.9158
14	0.9118	7+7	0.9179
20	0.9195	10+10	0.9246
22	0.9297	13+9	0.9298
27	0.9244	19+7	0.9244
33	0.9241	18+15	0.9261
37	0.9220	19+18	0.9277

Table 4.5: Classification accuracy obtained using extracted features and reduced combined set of features (reduction is done using SBS method) on KSC image

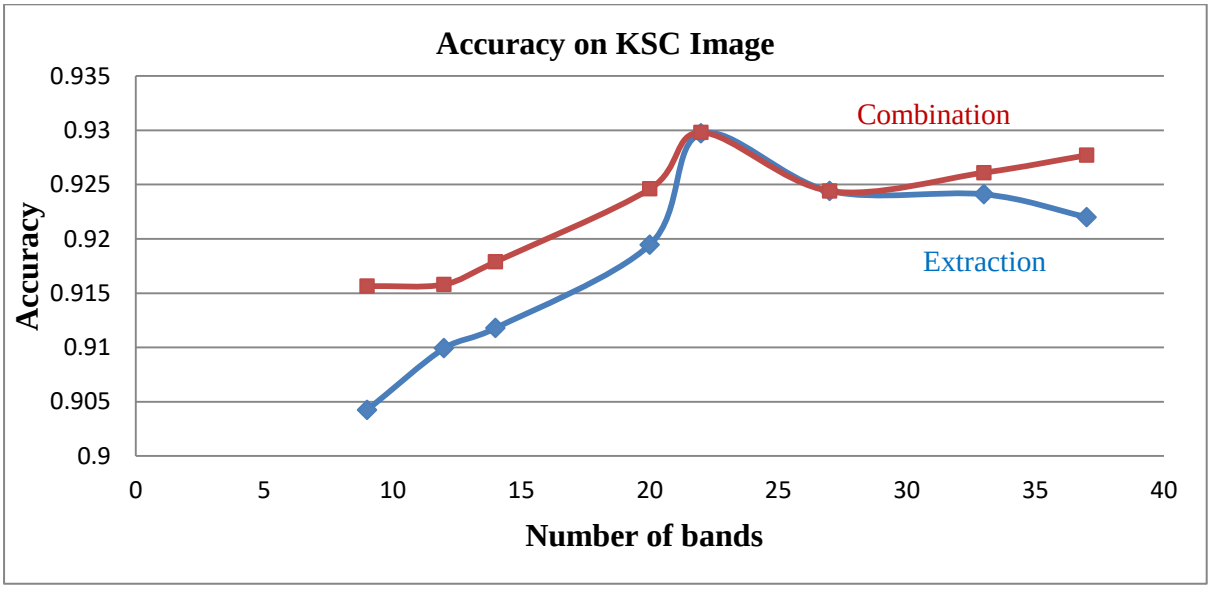


Fig. 4.10: Variation of accuracy values with number of bands obtained using extracted features and reduced combined set of features (reduction is done using SBS method) on KSC image

# of bands	Accuracy using Extracted Features	# of bands (Extraction + Selection)	Accuracy using Extracted Features
6	0.8593	3+3	0.8925
8	0.9028	4+4	0.9086
10	0.9040	5+5	0.9111
12	0.9099	7+5	0.9200
14	0.9118	10+4	0.9151
16	0.9108	10+6	0.9213
18	0.9184	11+7	0.9209
20	0.9195	13+7	0.9221

Table 4.6: Classification accuracy obtained using extracted features and reduced combined set of features (reduction is done using ReliefF method) on KSC image

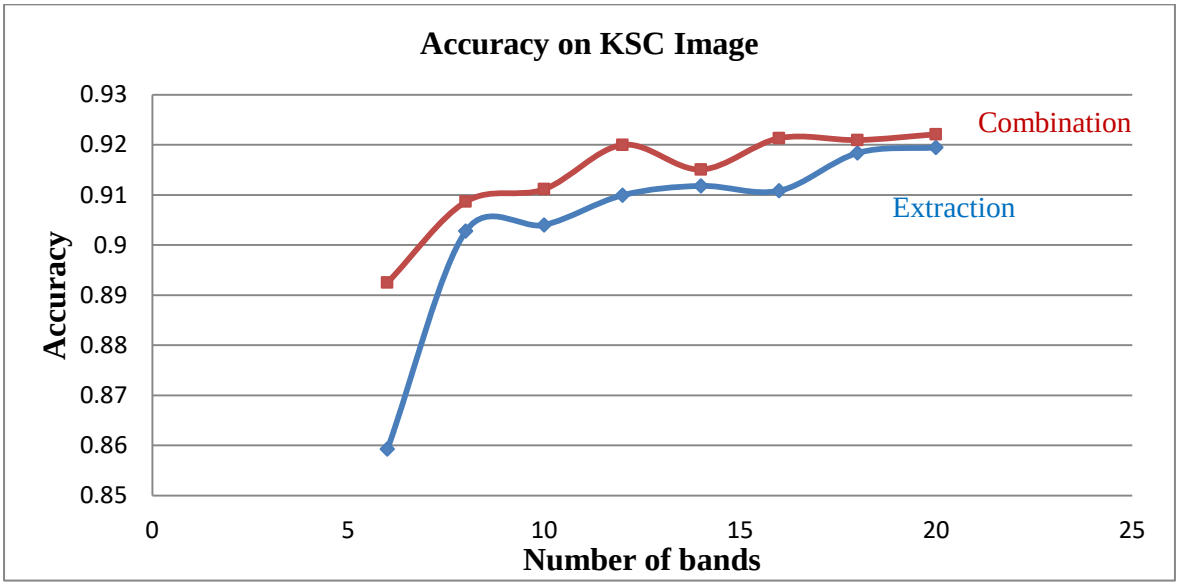


Fig. 4.11: Variation of accuracy values with number of bands obtained using extracted features and reduced combined set of features (reduction is done using ReliefF method) on KSC image

Analysis of results:

The pixel classification accuracy for the Indian pines image dataset with extracted features and with reduced combined features is shown in Table 4.4. The results are obtained by using the approach described in Section 3.5, where the number of original features (D) is 176, the range of extracted features varies from 4 to 40 (the value of r) and the range of selected features varies from 1 to 10 (the value of t). However, in this table, we have only provided the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. Both average and the best accuracies (over 6 simulations) are shown in Table 4.4. It is to be noted that the classification accuracy using reduced combined (extracted and selected) features is higher for all the cases when r lies in 5-20. Variation of average and best accuracy values with number of bands obtained using Indian pines dataset are shown in Figures 4.8 and 4.9, respectively. These figures also establish the superiority of using reduced combined features as compared to extracted features.

Table 4.5 shows the classification accuracy obtained using extracted features and reduced combined set of features, where the reduction on combined features is done using SBS method described in Section 3.6, where the range of extracted and selected features varies from 4 to 20 (the value of h). Variation of accuracy values with number of bands obtained by using SBS method for reduced combined features, are shown in Figures 4.10. The figure also establishes the superiority of using reduced combined features as compared to extracted features. Similarly, Table 4.6 shows the classification accuracy obtained using extracted features and reduced combined set of features, where the reduction on combined features is done using ReliefF method described in Section 3.7, where the range of extracted and selected features varies from 4 to the size to combined features set. The table contains only the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. Variation of accuracy values with number of bands obtained by using ReliefF method for reduced combined features, are shown in Figures 4.11.

4.3.3 Results of Botswana images

# of bands	Accuracy using Extracted Features		# of bands (Extracted + Selected)	Accuracy using Combined Features	
	Average	Best		Average	Best
7	0.9146	0.9163	6+1	0.9162	0.9178
8	0.9161	0.9206	7+1	0.9178	0.9212
9	0.9127	0.9156	7+2	0.9192	0.9215
10	0.9166	0.9178	8+2	0.9196	0.9212
11	0.9121	0.9144	4+7	0.9257	0.9280
12	0.9278	0.9301	5+7	0.9312	0.9326
13	0.9268	0.9268	5+8	0.9330	0.9338
14	0.9252	0.9252	6+8	0.9333	0.9366
15	0.9274	0.9301	12+3	0.9347	0.9400
16	0.9263	0.9289	12+4	0.9358	0.9372
17	0.9275	0.9295	12+5	0.9384	0.9412
18	0.9317	0.9341	12+6	0.9389	0.9418
19	0.9342	0.9342	12+7	0.9413	0.9437
20	0.9337	0.9363	12+8	0.9411	0.9440
21	0.9375	0.9393	12+9	0.9404	0.9424
22	0.9392	0.9412	12+10	0.9397	0.9412
23	0.9388	0.9409	21+2	0.9403	0.9421
24	0.9398	0.9421	21+3	0.9416	0.9424
25	0.9399	0.9412	21+4	0.9434	0.9449
26	0.9390	0.9415	23+3	0.9435	0.9443
27	0.9365	0.9378	22+5	0.9435	0.9452
28	0.9391	0.9409	23+5	0.9449	0.9477
29	0.9378	0.9418	24+5	0.9447	0.9461
30	0.9395	0.9415	24+6	0.9442	0.9455
31	0.9412	0.9452	24+7	0.9454	0.9477
32	0.9427	0.9427	22+10	0.9467	0.9474
33	0.9434	0.9461	32+1	0.9450	0.9474
34	0.9452	0.9464	32+2	0.9454	0.9470
35	0.9449	0.9461	33+2	0.9458	0.9477
36	0.9455	0.9467	33+3	0.9472	0.9495
37	0.9456	0.9489	33+4	0.9494	0.9507
38	0.9475	0.9486	32+6	0.9479	0.9514
39	0.9490	0.9501	33+6	0.9500	0.9517

Table 4.7: Classification accuracy obtained using extracted and reduced combined of extracted and selected features (combination is done using exhaustive method) on Botswana image

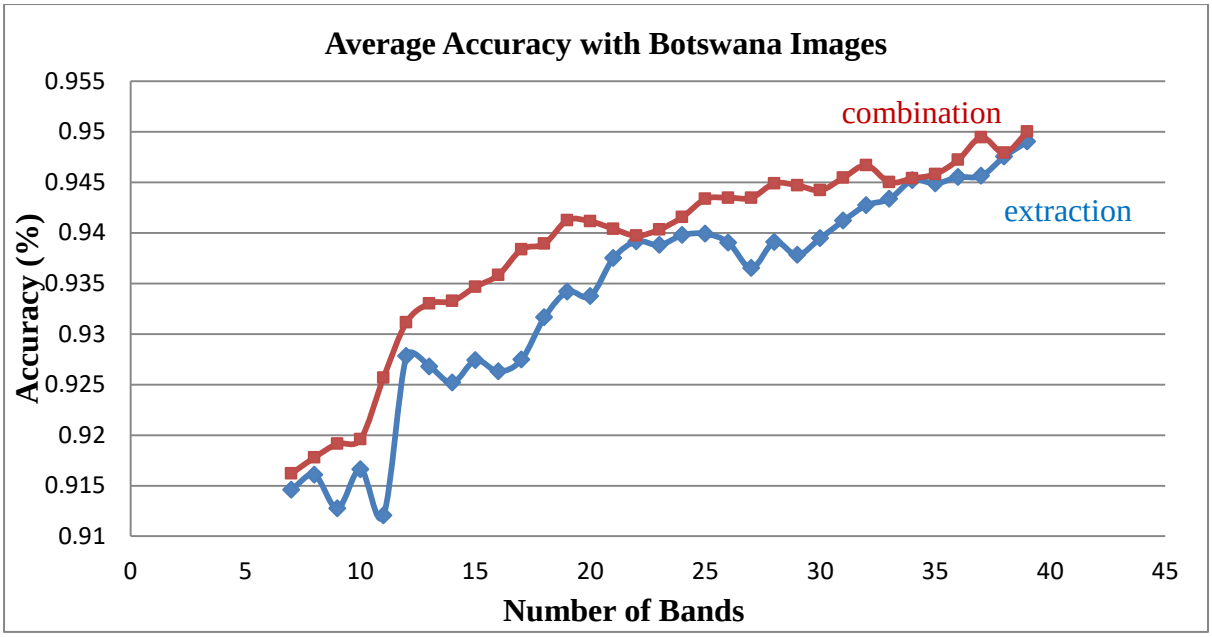


Fig. 4.12: Variation of average accuracy values with number of bands obtained using extracted and reduced combined set of features (combination is done using exhaustive method) on Botswana image

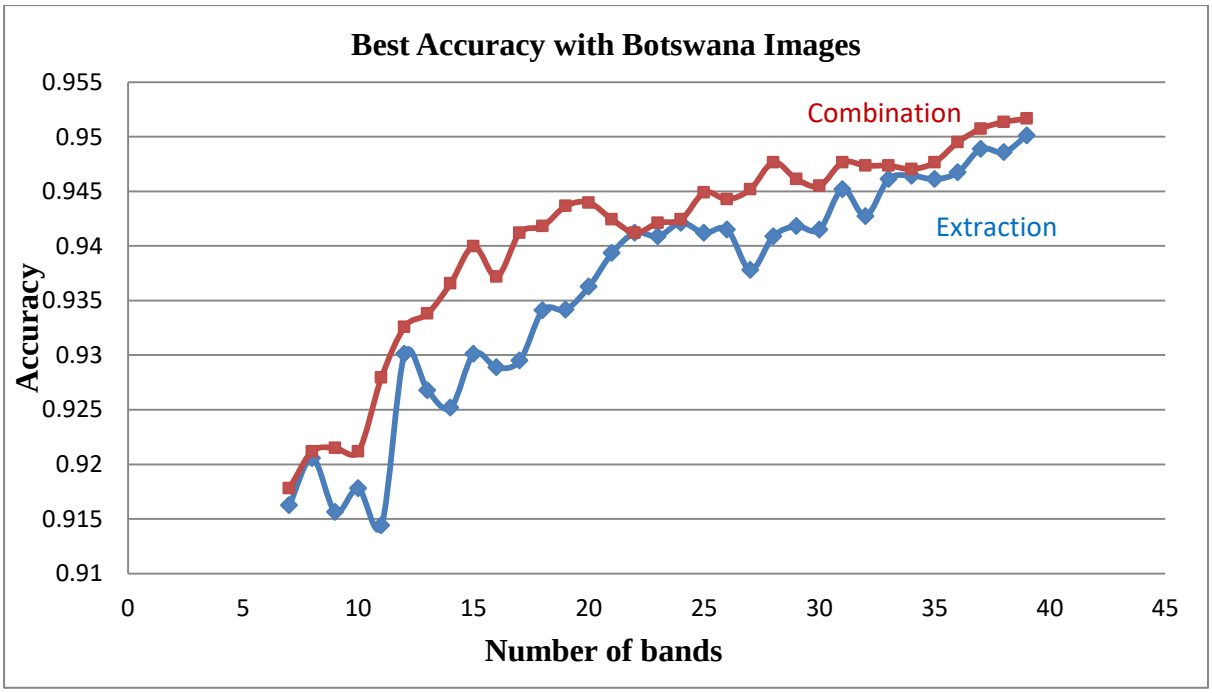


Fig. 4.13: Variation of best accuracy values with number of bands obtained using extracted and reduced combined set of features (combination is done using exhaustive method) on Botswana image

# of bands	Accuracy using Extracted Features	# of bands (Extracted + Selected)	Accuracy using Combined Features
9	0.9127	4+5	0.9193
10	0.9166	5+5	0.9357
11	0.9121	6+5	0.9276
13	0.9268	7+6	0.9307
15	0.9274	8+7	0.9360
17	0.9275	8+9	0.9378
18	0.9317	10+8	0.9477
19	0.9342	11+8	0.9449
20	0.9337	11+9	0.9375
21	0.9375	9+12	0.9495
23	0.9388	11+12	0.9464
24	0.9398	12+12	0.9483
29	0.9378	15+14	0.9437
30	0.9395	14+16	0.9535
31	0.9412	15+16	0.9569
32	0.9427	15+17	0.9483
37	0.9456	19+18	0.9563
38	0.9475	18+20	0.9581
41	0.9536	22+19	0.9544
43	0.9515	21+22	0.9532

Table 4.8: Classification accuracy obtained using extracted features and reduced combined set of features (reduction is done using SBS method) on Botswana image

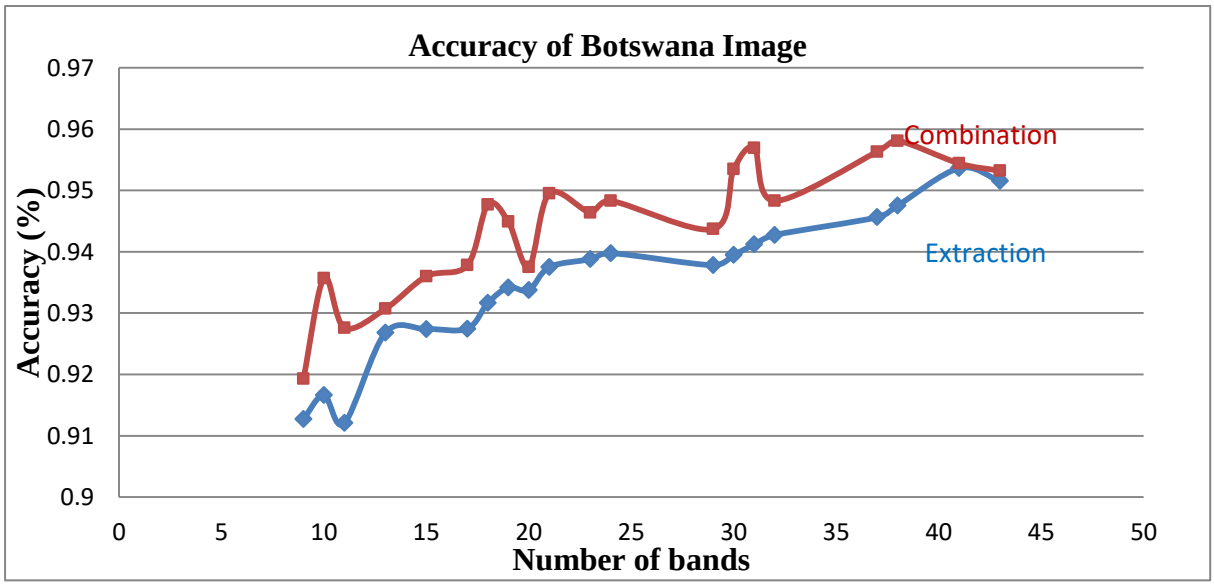


Fig. 4.14: Variation of accuracy values with number of bands obtained using extracted and reduced combined set of features (reduction is done using SBS method) on Botswana image.

# of bands	Accuracy using Extracted Features	# of bands (Extraction + Selection)	Accuracy using Extracted Features
10	0.9166	5+5	0.9193
14	0.9252	7+7	0.9313
16	0.9263	8+8	0.9332
18	0.9317	10+8	0.9326
24	0.9398	12+12	0.9415
26	0.9390	13+13	0.9461
28	0.9391	14+14	0.9446
30	0.9395	15+15	0.9446
32	0.9427	14+18	0.9452
34	0.9452	17+17	0.9483
36	0.9455	17+19	0.9507
38	0.9475	18+20	0.9483
42	0.9503	20+22	0.9510

Table 4.9: Classification accuracy obtained using extracted features and reduced combined set of features (reduction is done using ReliefF method) on Botswana image.

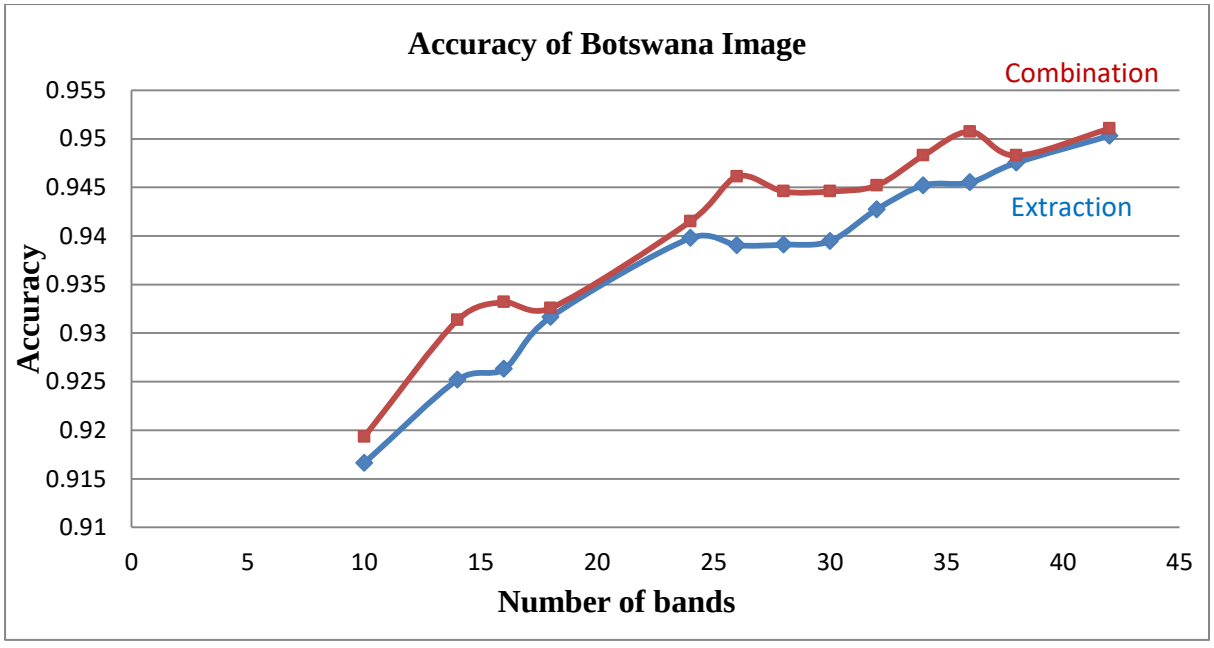


Fig. 4.15: Variation of accuracy values with number of bands obtained using extracted and reduced combined set of features (reduction is done using ReliefF method) on Botswana image

Analysis of results:

The pixel classification accuracy for the Botswana image dataset with extracted features and with reduced combined features is shown in Table 4.7. The results are obtained by using the approach described in Section 3.5, where the same configuration is used that has been used in Indian pines image (described in section 4.3.1). However, in this table, we have only provided the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. Both average and the best accuracies (over 6 simulations) are shown in Table 4.7. It is to be noted that the classification accuracy using reduced combined (extracted and selected) features is higher for all the cases when r lies in 7-39. Variation of average and best accuracy values with number of bands obtained using Indian pines dataset are shown in Figures 4.12 and 4.13, respectively. From the figure, it is noticed that, higher the number of bands, better is the classification accuracy. It is also seen that, if we further increase the number of bands, the classification accuracy either remains same or decreases. Moreover, if we increase the number of bands, computation time increases. These figures also establish the superiority of using reduced combined features as compared to extracted features.

The results of pixel classification accuracy of the Botswana image dataset with extracted features and reduced combined features is shown in table 4.8. In this case, the dimension reduction has been done on combined features set using the SBS method, which is discussed in section 3.6. Here, the range of extracted and selected features set is from 4 to 25 (the value of h). However, in this table, we have only provided the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. It is to be noted that the classification accuracy using reduced combined (extracted and selected) features is higher for most of the cases when r lies in 9-43. Variation of average and best accuracy values with number of bands obtained using Botswana image are shown in Figures 4.14. This figure also establishes the superiority of using reduced combined features as compared to extracted features.

The results shown in table 4.9 are the pixel classification accuracy values with extracted

features and reduced combined features, where the reduction of combined features is done using ReliefF algorithm, which has been discussed in section 3.7. Here, the range of extracted and selected features set is same as the SBS method. However, in this table, we have only provided the results which give higher classification accuracy using the reduced combined features as compared to those obtained using the extracted features of same size. It is to be noted that the classification accuracy using reduced combined (extracted and selected) features is higher for most of the cases when r lies in 10-42. Variation of average and best accuracy values with number of bands obtained using Botswana image are shown in Figures 4.15. This figure also establishes the superiority of using reduced combined features as compared to extracted features.

Chapter 5

Conclusion

In this thesis, we have demonstrated an approach for combination of extracted and selected features and three strategies for the dimension reduction of these combined features for hyperspectral images. For the combination purposes, two different heuristic methods are incorporated, one is the combination of selected and extracted features set of different cardinalities and another one is the combination of both types of features set having same cardinality. Thereafter for dimension reduction purpose of these combined features, one suboptimal method and one filter based method for feature selection is used. When dimension reduction is done on these combined sets, higher classification accuracy is obtained as compare to the extracted features set of same size. Combination of features is an active area of research. This work is an investigation on the effectiveness of combining the original features with the transformed features to achieve better classification accuracy.

5.1 Limitations of the work

The exhaustive approach for combining extracted and selected features is computation intensive as seen, though provides higher accuracy. Also SBS method consumes large amount of time if the cardinality of the combined feature set is high. As a suboptimal method is used for the reduction of combined features, it may not always provide the optimal reduced features from the combined features set.

5.2 Future Scopes of the work

There are lots of future scopes of this work that can be investigated further. The reduction of combined features can be done using an optimal feature selection technique, like, branch and bound to get the optimal sets of reduced combined features. The combination of both types of features can be done in an optimal way rather than heuristically. Also some other different algorithms or methods can be incorporated for extraction of features as well as for selection of subset of features from the original features set.

References:

[1] A.Dutta, “Dimensionality Reduction in Hyperspectral Images”, Ph.D thesis, Department of Computer Science & Engineering, Jadavpur University, 2017.

[2] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, “Remote Sensing and Image Interpretation”, 6th ed. New Delhi, India: Wiley, 2014.

[3] J. B. Campbell and R. H. Wynne, “Introduction to Remote Sensing”, 5th ed. New York, USA: Guilford Press, 2011.

[4] Jenson and R. John, “Remote Sensing of the Environment: An Earth Resource Perspective”, Person Prentice Hall, 2007.

[5] P. K. Varshney and M. K. Arora, “Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data”, 2nd ed. Berlin, Germany: Springer-Verlag, 2004.

[6] J. A. Richards and X. Jia, “Remote Sensing Digital Image Analysis: An Introduction”, 1st ed. New York, USA: Springer, 1999.

[7] D. Landgrebe, “Hyperspectral Image Data Analysis”, IEEE Signal Processing Magazine, pp. 17-28, 2002.

[8] C. I. Chang, “Hyperspectral Imaging: Techniques for Spectral Detection and Classification”, 1st ed. New York, USA: Kluwar Academic/ Plenum Publisher, 2003.

[9] M. T. Eismann, “Hyperspectral Remote Sensing”, 1st ed. Washington, USA: SPIE Press, 2012.

[10] Remote sensing: url- https://en.wikipedia.org/wiki/Remote_sensing

[11] S.Jia, Y.Qian, J.Li, W.Liu & Z.Ji, “Feature extraction and selection hybrid algorithm for hyperspectral imagery classification”, IEEE International Geoscience and Remote Sensing Symposium, July 2010.

[12] Sreevani and C. A. Murthy, “Simultaneus Feature Selection and Feature Extraction for Pattern Classification”, Masters’ Thesis, ISI Kolkata, 2009.

[13] S. Theodoridis, K. Koutroumbas, “Pattern Recognition”, 2nd ed. San Diego, USA: Elsevier Academic Press, 2003.

[14] A. Dutta, S. Ghosh and A. Ghosh,“Wrapper Based Feature Selection in Hyperspectral Image Data Using Self-adaptive Differential Evolution”, 2011 International Conference on Image Information Processing, 2011.

[15] Narendra and Fukunaga, “A Branch and Bound algorithm for Feature Subset Selection”, IEEE Transactions on Computers, pp. 917-922, 1977.

[16] Hyperspectral remote sensing scenes: url- http://www.ehu.eus/ccwintco/index.php/

[17] A. L. Blum and P. Langley, “Selection of Relevant Features and Examples in Machine Learning”, Artificial Intelligence, pp. 245-271, 1997

[18] A. Datta, S. Ghosh and A. Ghosh, “Band Elimination of Hyperspectral Imagery Using Correlation of Partitioned Band Images,” International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 412-417, 2013.

[19] A. Datta, S. Ghosh and A. Ghosh, “Supervised Feature Extraction of Hyperspectral Images Using Partitioned Maximum Margin Criterion,” IEEE Geoscience and Remote Sensing Letters”, pp. 82-86, 2017.

[20] H. R. Li, T. Jiang, and K. Zhang, “Efficient and robust feature extraction by maximum margin criterion,” IEEE Transaction Neural Network, pp. 157–165, Feb. 2006.

[21] K. Kira and L. A. Rendell, “The feature selection problem: traditional methods and a new algorithm,” in AAAI-92. Tenth National Conference on Artificial Intelligence , Menlo Park, CA, USA, pp. 129-34, 1992.

[22] I. Leifer *et. al.*, "State of the Art Satellite and Airborne Marine Oil Spill Remote Sensing: Application to the BP Deepwater Horizon Oil Spill," Remote Sens. Environ. **124**, 185 (2012).

[23] T. Slonecker *et. al.*, "Visible and Infrared Remote Imaging of Hazardous Waste: A Review," Remote Sens. **2**, 2474 (2010).

[24] M. R. Sikonja and I. Kononenko, “Theoretical and Empirical Analysis of ReliefF and RReliefF”, Machine Learning 53, pp.23-69, 2003.

[25] SVM Classification: url - https://in.mathworks.com/help/stats/support-vector-machine-classification.html

[26] Sequential Feature Selection: url - https://in.mathworks.com/help/stats/sequentialfs.html

[27] A. L. Blum and P. Langley, “Selection of Relevant Features and Examples in Machine Learning”, Artificial Intelligence, pp. 245-271, 1997.

[28] A. K. Jain, R. P.W. Duin, and J. Mao, “Statistical pattern recognition: A review”, IEEE Transactions on Pattern Analysis and Machine Intellegence, pp.4-37, 2000

[29] R. Bellman, “Adaptive Control Processe: A guided Tour”, 1st ed. London, UK: Princeton University Prress, 1961.

[30] G. F. Hughes, “On the Mean Accuracy of Statistical Pattern Recognition”, IEEE Transactions on Information Theory, pp. 55-63, 1968.

[31] K. Fukunaga, “Introduction to Statistical Pattern Recognition”, 2nd ed. San Diego, CA, USA: Academic Press, 1990.

[32] P. A. Devijver and J. Kittler, “Pattern Recognition: A statistical Approach”, 1st ed. New Delhi, India: Prentice-Hall Iternational, 1982.

[33] H. Yang, Q. Du, H.Su and Y. Sheng, “An efficient method for supervised hyperspectral band selection”, IEEE Geoscience and Remote Sensing Letters, pp. 138-142, 2011

[34] H. Su, Q. Du and P. Du, “Hyperspectral Image Visualization Using Band Selection”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 2647-2658, 2014.

[35] A. P. Englebrecht, “Computational Intellegence: An Introduction”, 2nd ed. West Sussex, England: John Wiley and Sons, 2007.

[36] H. Liu and L. Yu, “Towards integrating feature Selection Algorithms for Classification and Clustering ”, IEEE Transactions on Pattern Analysis and Machine Intellegence, pp. 153-189, 1997.

[37] H. A. Firpi and E. Goodman, “Swarmed Feature Selection”, in Proceedings of the 33rd Applied Imagery Pattern Recognition workshop, pp. 112-118, 2004.

[38] Q. Du and H. Yang, “Similarity-based unsupervised band selection for Hyperspectral image Analysis”, IEEE Geoscience and Remote Sensing Letters, pp.564-568. 2008.

[39] S. Kumar, J. Ghosh and N. M. Crawford, “Best-bases feature extraction algorithms for Classification of Hyperspectral data”, IEEE Transactions on Geoscience and Remote Sensing, pp.1368-1379, 2001.