MULTIPLE CLASSIFIERS AND DIMENSIONALITY REDUCTION METHODS FOR HYPERSPECTRAL IMAGE CLASSIFICATION

A thesis submitted
in partial fulfillment for the degree of

Doctor of Philosophy

by

BHARATH BHUSHAN D

DEPARTMENT OF EARTH AND SPACE SCIENCES

INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY
THIRUVANANTHAPURAM - 695547

JANUARY 2015

CERTIFICATE

This is to certify that the thesis titled **Multiple Classifiers and Dimensionality Reduction Methods for Hyperspectral Image Classification**, submitted by **Mr. Bharath Bhushan D**, to the Indian Institute of Space Science and Technology, Thiruvananthapuram, for the award of the degree of **Doctor of Philosophy**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Rama Rao Nidamanuri
Supervisor
Department of Earth and Space Sciences

Thiruvananthapuram
January 2015

Counter signature of HOD with seal

DECLARATION

I declare that this thesis titled **Multiple Classifiers and Dimensionality Reduction Methods for Hyperspectral Image Classification** submitted in fulfillment of the Degree of Doctor of Philosophy is a record of original work carried out by me under the supervision of **Dr. Rama Rao Nidamanuri**, and has not formed the basis for the award of any degree, diploma, associateship, fellowship or other titles in this or any other Institution or University of higher learning. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever the findings of others have been cited.

Bharath Bhushan D
SC11D007

Place: Thiruvananthapuram
January 2015

ACKNOWLEDGEMENTS

First and foremost, I express my sincere thanks and profound gratitude to my supervisor Dr. Rama Rao Nidamanuri for his continuous support throughout my Ph.D studies and research; for his patience, motivation, enthusiasm, immense knowledge and insightful discussions. His guidance helped me throughout my research and also in writing of this thesis.

I thank my Doctoral committee members: Prof. A Chandrasekar, Dr. B S Daga Sagar, Dr. S K Sasamal, Dr. L Gnanappazham, and Dr. Deepak Mishra for their encouragement and insightful comments. I extend my thanks to Dr. R Krishnan, former Dean Academics, IIST for his valuable suggestions during the course of work. I would like to thank Director, IIST, Dr. K. S. Dasgupta and Dean R & D, IIST for providing me this opportunity and support throughout my Ph.D. I like to express my appreciation to the entire staff and lab members of the department, library and administrative staff at IIST for their timely help.

I am grateful to Dr. K. P. Soman and Dr. M. Sabarimalai Manikandan, my master supervisors, for their inspiration and motivation to pursue doctoral research. I also take this opportunity to thank Dr. Yuliya Tarabalka for her scientific discussions in the last stage of my research work.

I thank all my friends for their never ending love, support and excitement. Last but not the least, I would like to thank my parents, brother and sisters for their continuous support they had given throughout my educational career and life.

Bharath Bhushan D

ABSTRACT

The aim of this thesis is to develop efficient classification methodologies based on a multiple classifier system which minimize the classifier and data dependence and offer acceptable classification accuracy across various hyperspectral images and application domains. Hyperspectral remote sensing has been emerging as a reliable data source for the mapping and monitoring of various land surface features and processes. Hyperspectral data are interesting and challenging. Traditional supervised image classification techniques that use all available spectral bands often fail on hyperspectral data due to the curse of dimensionality that comes along. Dimensionality reduction methods coupled with appropriate classifiers could mitigate the dimensionality problem. But, identification of an appropriate dimensionality reduction method and classifier is subjective, based on analysts' prior knowledge and the performance is method and data specific.

A multiple classifier system provides a conceptual framework to combine the relative advantages of several classifiers to enhance reliability and accuracy of classification. Having diversity in the performance of classifiers is a pre-condition for the success of multiple classifier system. In principle, multiple transformations of the same hyperspectral image by different dimensionality reduction methods lead to differential classification performances. The potential of dimensionality reduction methods to create diversity in the multiple classifier system is not well understood. As the multiple classifier system involves parallel application of a diverse group of classifiers, understanding the impact of dimensionality reduction methods on the overall classification performance is important for developing optimal classification methodologies for land cover mapping.

The aim of this thesis is twofold. In the first part, there are two objectives. The first objective is to study the impact of different dimensionality reduction methods on the classification performance of a multiple classifier system. A multiple classifier system designed with five dimensionality reduction methods and seven classifiers has been used for a series of classification experiments on five different hyperspectral images (acquired at different sites) for land cover mapping. The change in classification accuracy for various combinations of dimensionality reduction methods and classifiers has been tracked across various information classes and land cover settings. Results indicate substantial change in the performance of the multiple classifier system with different dimensionality reduction methods (*peer reviewed journal manuscript on this aspect: Impact of dimensionality reduction methods on the classification performance of the multiple classifier system for hyperspectral image classification, International Journal of Remote Sensing*). The second objective is to assess the relationship between information class, classifier and dimensionality reduction method within the multiple classifier system framework. The multiple classifier system's architecture has been manoeuvred to compute magnitudes and patterns in the per-class classification accuracy of each information class for all the possible combinations of classifiers and dimensionality reduction methods within and across the different hyperspectral images. The results indicate the existence of empirical relationships across different hyperspectral images, wherein different information classes prefer different combinations of classifiers and

dimensionality reduction methods (*peer-reviewed journal manuscript on this aspect: Assessment of the impact of dimensionality reduction methods on information classes and classifiers for hyperspectral image classification by multiple classifier system, Advances in Space Research*).

The first part of the aim of thesis highlights the necessity of introducing dynamism in the multiple classifier system for the selection of classifiers and dimensionality reduction methods for effective hyperspectral image classification. Consequently, in the second part, the objective is to develop a novel classification technique which selects classifiers and dimensionality reduction methods according to input data dynamics. In particular, we propose two novel modifications to the functional architecture of the multiple classifier system. The first contribution is an algorithmic extension of the multiple classifier system, we name it as dynamic classifier system. The proposed dynamic classifier system pairs up optimal combinations of classifiers and dimensionality reduction methods, specific to the hyperspectral image, and performs image classification based only on the identified combinations (*peer-reviewed journal manuscript on this aspect: Dynamic linear classifier system for hyperspectral image classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*). The second contribution is the development of a new dynamic classifier selection approach based on extreme learning machine regression which selects a subset of optimal classifiers relative to each input pixel by exploiting the local information content of the pixel. Further, the spatial contextual information is incorporated in the proposed dynamic classifier selection approach to develop a new spectral-spatial classification model to exploit high spatial resolution of modern airborne hyperspectral images (*peer-reviewed journal manuscript on this aspect: Dynamic ensemble selection approach for hyperspectral image classification with joint spectral and spatial information, IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*).

The objectives of this thesis make a significant contribution to the current knowledge about the application of multiple classifier system for hyperspectral image classification and present a novel multiple classifier system based dynamic classification framework, which offers optimal classification performance across different hyperspectral images, land cover settings, and information classes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AA	Average Accuracy
ACE	Adaptive Coherence Estimator
CF	Classifier Fusion
CV	Correlation of Variation
DCS	Dynamic Classifier Selection
DCS	Dynamic Classifier System
DCS-Beta-LSR	Dynamic Classifier Selection-Beta-Least Square Regression
DCS-ELM	Dynamic Classifier Selection-Extreme Learning Machine
DCS-LA	Dynamic Classifier Selection-Local Accuracy
DCS-MLA	Dynamic Classifier Selection-Modified Local Accuracy
DES	Dynamic Ensemble Selection
DES-Beta-LSR	Dynamic Ensemble Selection-Beta-Least Square Regression
DES-ELM	Dynamic Ensemble Selection-Extreme Learning Machine
DM	Disagreement Measure
DWT-DR	Discrete Wavelet Transform Based Dimensionality Reduction
FCC	False Color Composite
ICA	Independent Component Analysis
KC	Kappa Coefficient
KPCA	Kernel Principal Component Analysis
KWM	Kohavi-Wolpert Variance Measure
LDC	Linear Discriminant Classifier
LRC	Logistic Regression Classifier
MF	Matched Filter
MLC	Maximum-Likelihood Classifier
MNF	Minimum Noise Fraction
MRF	Markov Random Field
MV	Majority Voting
NBC	Naive Bayes Classifier

NED	Normalized Euclidean distance
OA	Overall Accuracy
OBS	Optimal Band Selection
OSP	Orthogonal Subspace Projection
PA	Producer Accuracy
PCA	Principal Component Analysis
QCD	Quartile Coefficient of Dispersion
RBF	Radial basis function
RMD	Relative mean difference
RRC	Random Reference Classifier
RSM	Random Subspace Method
SAM	Spectral Angle Mapper
SB	Single Best Classifier
SSM	Spectral Similarity Measure
SVM	Support Vector Machine
TCIMF	Target Constraint Interference Minimized Filter
UA	User Accuracy

CHAPTER 1

INTRODUCTION

1.1 Hyperspectral Imaging

For the last decade hyperspectral remote sensing has been an active area of research and development in remote sensing. Recent advances in detecting technologies made it possible to extend the collection of hyperspectral images to the thermal infrared region of electromagnetic spectrum (Johnson et al., 2011; Schlerf et al., 2012; Johnson et al., 2014; Santini et al., 2014; Jie-lin et al., 2014). Hyperspectral remote sensing is related to multispectral remote sensing. The multispectral remote sensing acquires image data in several discrete spectral bands. Hyperspectral remote sensing collects image data simultaneously in hundreds of narrow (typically in 10 nm) and continuous spectral bands in the entire optical wavelength region of the electromagnetic spectrum ($0.4 - 2.5\mu m$). These measurements make it possible to derive a continuous spectrum for each image pixel as shown in the Figure 1.1. After adjustments for sensor, atmospheric, and terrain effects are applied, these image spectra can be compared with field or laboratory reflectance spectra in order to recognize and map surface materials such as particular types of vegetation or diagnostic minerals associated with ore deposits. Originally developed for mining and geology, hyperspectral remote sensing is now used in a wide array of applications and the technology is continually becoming more available to the public and remote sensing community.

1.1.1 Applications

Thanks to the abundant spectral information inherent in hyperspectral images, hyperspectral images become important data sources for a wide range of applications, as widespread as ecology and surveillance, as well as historical manuscript research, such as the imaging of the Archimedes Palimpsest (Liang, 2011; Rapantzikos and Balas, 2005; Salerno et al., 2006).

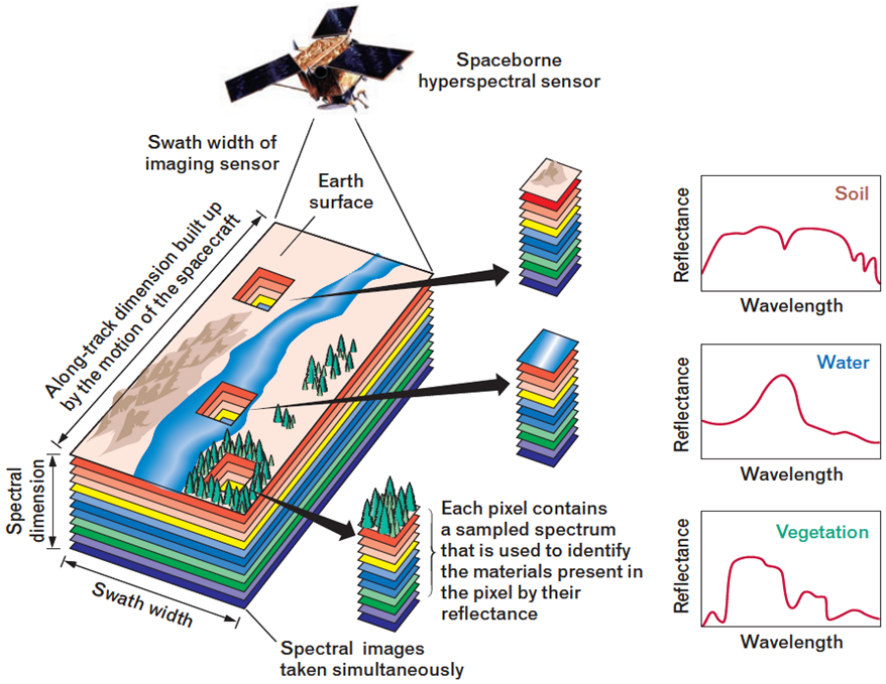


Figure 1.1: The concept of hyperspectral imaging and respective spectral signatures of the different materials (soil, water and vegetation), *Image courtesy: www.markelowitz.com/Hyperspectral.html*.

The following are some of the application areas where hyperspectral remote sensing is being applied actively.

- Monitoring and management of environment: Hyperspectral images can be used to study the nature of the environment and track the changes over the time. In particular, it is useful to monitor water bodies, harmful algal blooms, human settlement in urban areas, forest management, air pollution, CO_2 emissions and in detection of oil spills (Andrew and Ustin, 2008; Ghiyamat and Shafri, 2010; Borengasser et al., 2007; Sanchez et al., 2003).

- Vegetation mapping and biophysical characterization: Hyperspectral images can be used in vegetation mapping applications at the community and species level, and in characterizing the biophysical properties of the vegetation cover such as bio-mass, leaf area index, chlorophyll content, etc. These properties are important in understanding the dynamics of ecosystems (Darvishzadeh et al., 2008; Wu et al., 2010a; Thenkabail et al., 2011, 2013).

- Precision agriculture: Hyperspectral remote sensing can be effectively used in precision agriculture to monitor the health of the crop, characterize the physical

and chemical properties of the soil, detection of nutrient deficiencies, identification of insect and weed infestation, and early detection of crop stress (Boggs et al., 2003; Cetin et al., 2005; Thenkabail et al., 2011).

- Mineralogy: Hyperspectral images can be used to identify many valuable minerals that are exposed or weathered in the residual soil and to retrieve the surface compositional information for mineral exploration purposes (Sabins, 1999; Cloutis, 1996; Van Der Meer et al., 2012).

- Military and defence: Hyperspectral imaging can used in military surveillance to detect and track the military objects that are invisible to the naked eye (Briottet et al., 2006; Manolakis et al., 2003).

- Detection of hazardous materials: Hyperspectral technology detects and localize the presence of the chemical plumes and other hazardous compounds in the atmosphere and helps in deciding the timely pre-caution measures (Hirsch and Agassi, 2007; Gurram and Kwon, 2010; Dao et al., 2012; Manolakis et al., 2014).

Further, the technological evolution of optical sensors has led to development of numerous airborne hyperspectral sensors with different spectral, spatial and operational range specifications, thus enabling to explore new applications. Table 1.1 lists the available hyperspectral sensors, future planned missions and their spectral properties (Dalponte et al., 2009; Staenz and Held, 2012).

1.1.2 Challenges

Hyperspectral images contain a wealth of spectral data, but effective interpretation and processing methods are crucial in translating the rich spectral data into more useful information. The very nature of a hyperspectral image poses many challenges to exploit hyperspectral images effectively. Some of the most common limiting factors are given below.

- High dimensionality: extremely large size of data causes storage and memory problems; demands fast computing solutions that can accelerate the information exploitation of hyperspectral images for many real time applications such as target detection, change detection, and food quality inspection, etc.

Table 1.1: List of recent and future planned hyperspectral sensors and their spectral properties.

Sensor name	Manufacturer	Platform	Number of Spectral Bands	Spectral Resolution nm	Spectral range μm
Hyperion on EO-1	NASA Goddard Space Flight Center	satellite	220	10	0.4-2.5
CHRIS Proba	ESA	satellite	63	1.25	0.415-1.05
HySI	ISRO	satellite	64	10	0.4-0.950
HJ-1A	Chine CAST	satellite	110-128	5	0.45-0.950
HICO	NASA/ONR	satellite	128	5.7	0.353-1.08
AVIRIS	NASA Jet Propulsion Lab	aerial	224	10	0.4-2.5
HYDICE	Naval Research Lab	aerial	210	7.6	0.4-2.5
PROBE-1	Earth Search Sciences Inc.	aerial	128	12	0.4-2.45
CASI 550	ITRES Research Limited	aerial	288	1.9	0.4-1
CASI 1500	ITRES Research Limited	aerial	288	2.5	0.4-1.05
SASI 600	ITRES Research Limited	aerial	100	15	0.95-2.45
TASI 600	ITRES Research Limited	aerial	64	250	8- 11.5
HyMap	Integrated Spectronics	aerial	125	17	0.4-2.5
ROSIS	DLR	aerial	84	7.6	0.4-0.85
EPS-H	GER Corporation	aerial	133	0.67	0.43-12.5
EPS-A	GER Corporation	aerial	31	23	0.43-12.5
DAIS 7915	GER Corporation	aerial	79	15	0.43-12.3
AISA Eagle	Spectral Imaging	aerial	244	2.3	0.4-0.97
AISA Eaglet	Spectral Imaging	aerial	200	-	0.4-1.0
AISA Hawk	Spectral Imaging	aerial	320	8.5	0.97-2.45
AISA Dual	Spectral Imaging	aerial	500	2.9	0.4-2.45
MIVIS	Daedalus	aerial	102	20	0.43-12.7
AVNIR	OKSI	aerial	60	10	0.43-1.03
EnMAP	DLR	future	218	5/10	0.42-2.45
PRISMA	ASI	future	237	12	0.4-2.5

- Redundant information: hyperspectral images being spectrally oversampled, correlation among neighbouring spectral bands is high and this correction often leads to sampling problems in the analyses of hyperspectral images.

- Noisy bands: hyperspectral images are often affected by various types of noise such as sensor noise, photon noise, calibration errors which reduce the precision of the subsequent processing (Kerekes and Baum, 2005; Guo et al., 2013).

- Limited labelled samples: small ratio between the dimensionality of hyperspectral image and available reference samples, popularly known as Hughes phenomena or curse of dimensionality.

These factors make the information exploitation of hyperspectral image a very challenging task. The methods developed for analysing multispectral images can be naturally extended to interpret the hyperspectral data. However, the afore mentioned limiting factors, especially the case of a limited number of labelled samples impede the effectiveness of data exploration methods. Thus development of new hyperspectral data processing methods is necessary to take full advantage of hyperspectral data.

1.2 Motivation, Objectives, and Contribution of the thesis

Supervised classification is an analysts' method of choice to translate remote sensing image data into meaningful labelled pixel information. The supervised classification is the process of assigning class labels to each of the pixels in image by learning a hypothesis function from the training samples. The effectiveness of supervised classification depends upon its ability to trade upon the available training samples, data dimensionality and information classes. In the last decade, significant research efforts have been made in improving hyperspectral image classification performance by developing 1) dimensionality reduction methods for overcoming the curse of dimensionality (Lennon et al., 2001; Kumar et al., 2001; Kaewpijit et al., 2003; Manolakis and Marden, 2004; Hsu, 2007; Amato et al., 2009; Bor-Chen Kuo et al., 2009; Dalla Mura et al., 2011; Yin et al., 2012; Imani and Ghassemian, 2014); and 2) non-parametric and knowledge

based classifiers which are not sensitive to some of limiting factors of hyperspectral images (Camps-Valls and Bruzzone, 2005; Plaza et al., 2009; Chen et al., 2013; Li et al., 2013; Camps-Valls et al., 2014; Kuo et al., 2014). These studies tackle some of the limiting factors of hyperspectral image analysis and also report good classification accuracy. The general conclusions of these studies highlight that the available techniques are inadequate to derive substantial benefits of the spectrally rich hyperspectral image.

Nevertheless, with plethora of classifiers and dimensionality reduction methods available, the selection of a classifier and/or a dimensionality reduction method which offers optimal performance for a given hyperspectral image and/or application domain is a subjective task and is still an active area of research (Wolpert, 1996; Fabio et al., 1997; Chen and Qian, 2008). In general, the most appropriate classifier, and dimensionality reduction method are identified before hand by analyst's prior knowledge or on a heuristic basis. This makes the obtained results subjective and the procedure expert-dependent.

A multiple classifier system (MCS), an advanced pattern recognition technique is a flexible way to merge complimentary advantages of several algorithms into a single framework. MCS consists of several classifiers called as base classifiers. Each of the classifiers in the MCS produces decision function values for a given input and these decision function values are combined in a suitable manner to obtain the final decision. Benediktsson and Kanellopoulos (1999) explored the potential of MCS for hyperspectral image classification and have shown that it alleviates the small sample size problem. Since then, several studies have appeared in the literature on the classification of hyperspectral data using MCS framework and show improved classification accuracy (Petrakos et al., 2001; Ham et al., 2005; Doan and Foody, 2007; Kuo et al., 2011; Kwon and Rauss, 2011; Du et al., 2012b; Xia et al., 2014). The appropriate design of MCS which maintains diversity among the base classifiers is necessary to improve the classification accuracy. The design of MCS for hyperspectral image classification in the literature can be categorized into different categories: 1) the divide and conquer approach 2) the multi-source and multi-temporal approach 3) the training samples manipulation approach 4) the spectral and spatial features approach, and 5) the supervised and unsupervised approach. These approaches consider different ways of introducing complimentary information in the design of MCS to obtain higher classification accuracies.

In the divide and conquer approach, the diversity in the MCS is created by partitioning a hyperspectral image into several smaller subsets and each of these subsets are given as the input data source to the MCS (Benediktsson and Kanellopoulos, 1999; Gidudu et al., 2009; Kalluri et al., 2010; Kwon and Rauss, 2011; Prasad et al., 2012). On the other hand, the multi-source and multi-temporal approach introduces diversity in the MCS by considering the hyperspectral images acquired on different dates and different sensors (Landgrebe, 1999; Prasad et al., 2008; Udelhoven et al., 2009; Du et al., 2012b; Dos Santos et al., 2012). Apart from these approaches, the diversity in the MCS can also be created by using training samples manipulation methods such as bagging, boosting, random forest and their variants (Fernandez-Redondo et al., 2004; Chan et al., 2012). This approach selects a different subset of training samples to train the base classifiers in the MCS. Kumar et al. (2002) introduced hierarchical based approach for creating diversity in the MCS, in that multiple binary classifiers are used in the design of the MCS (Kumar et al., 2002; Ham et al., 2005; Jun and Ghosh, 2009; Rajan et al., 2006). Apart from the spectral information, the high spatial content of the hyperspectral image has also been exploited to create complimentary information in the design of MCS by deriving spatial information features and spectral feature (Dalla Mura et al., 2011; Benediktsson et al., 2004; Huang and Zhang, 2013; Wang et al., 2009). Recently, other than the supervised classifiers, different segmentation techniques and unsupervised classification techniques are used in the MCS design and to create diversity among the base classifiers (Yang et al., 2010a; Tarabalka et al., 2010; Alajlan et al., 2012, 2013). These studies prove that MCS is a competitive learning method to offer substantial improvement in classification accuracy. Nonetheless, these available techniques are again data, application specific and are limited by generalization. In particular, the choice of classifiers in the MCS has significant impact on the classification performance.

Dimensionality reduction methods enable the creation of multiple transformations of same hyperspectral image with different mathematical and statistical properties, thus leading to have complimentary information among different dimensionality reduction methods. The potential of creating diversity by deploying multiple dimensionality reduction methods in the MCS is not understood. In particular, creating differential performances among the classifiers in the MCS against different dimensionality reduction methods is not studied yet. Further as mentioned earlier, the performance of the MCS

is subjective to the classifier set which constitutes the MCS. The proper selection of the classifiers and dimensionality reduction methods for the data at hand is necessary to achieve generalization performance across different hyperspectral images. Our extensive literature survey reveals that the current methods lack the capability to select classifiers dynamically based on the input data dynamics.

In addition to classifiers combination methods, the final decision of MCS can also be obtained by classifiers selection methods. In the classifiers combination methods, the decision function values of all the classifiers are used; where as in the classifiers selection methods, the decision function values of certain classifiers are only used to obtain final classification results. Most of the studies in MCS for hyperspectral image classification are mainly focused on the classifiers combination, while little or no attention has paid to the classifiers selection methods. In particular, dynamic selection of subset of classifiers relative to each image pixel has not been studied. This approach has the potential to relax the rigid diversity and accuracy constraints of the MCS and to account for local variations within the image.

Moreover, in recent years several studies addressed the incorporation of spatial information in the classification system (Fauvel et al., 2008; Chen et al., 2011; Zhang et al., 2012; Fauvel et al., 2013; Gurram and Kwon, 2013; Chen et al., 2014). In particular Markov random field (MRF) regularization has gained enormous attention. However, these studies are performed with a single classifier; and the potential of MRF model in the MCS framework is least understood. Hence the general aim of this thesis is to address the aforementioned limitations in the MCS framework and develop a dynamic MCS classification framework for effective hyperspectral image classification.

Objectives

The overall aim of this thesis is to study the impact of dimensionality reduction methods on the classifiers and information class relationships in the MCS framework, and develop a novel MCS based dynamic classification technique which adaptively selects optimal pairs of classifiers and dimensionality reduction methods and relative to the information content in hyperspectral images. Enhancing the MCS classification framework to a functionally dynamic classification framework, adaptive to the hyperspectral

image at hand, would be a valuable methodological tool for objective and quantifiable decisions, choices, and optimal performance in hyperspectral image classification across different images. The main contributions and hence objectives of this thesis can be summarized as:

1. an empirical analysis on the impact of dimensionality reduction methods on the classification performance of multiple classifier system for hyperspectral image classification

2. an assessment of classifier and dimensionality reduction method relationship, and information class, classifier and dimensionality reduction method relationship in the framework of multiple classifier system for hyperspectral image classification

3. a novel system for automatically selecting the classifiers relative to each dimensionality reduction method based on the input data dynamics for hyperspectral image classification

4. to propose a new dynamic classifier selection approach for hyperspectral image classification which selects best subset of classifiers relative to each pixel based local distribution of the pixel, and

5. a novel spectral-spatial classification model to exploit the spectral and spatial information of the hyperspectral image

In the next subsections, the main objectives and novelties are briefly discussed.

1. Impact of dimensionality reduction methods on the classification performance of MCS

The information exploitation of the hyperspectral image is complex from the theoretical and computational point of view. Dimensionality reduction is an effective preprocessing technique to overcome the Hughes phenomena and to reduce the computational load for further analysis. Whilst several dimensionality reduction methods are available in the literature, the choice of dimensionality reduction method and its compatibility with the classifier selected is crucial for achieving acceptable classification

results. Despite many studies available on suitability of dimensionality reduction methods for the classical supervised classification methods which uses only a single classifier, the application and hence its compatibility with base classifiers and impact on the overall performance within the MCS framework is not reported. In this thesis, we present an empirical study aimed at understanding the impact of dimensionality reduction methods on the performance of MCS for hyperspectral image classification. In particular, we analyse the behaviour of the classification accuracy of the MCS on different dimensionality reduction methods and different combination functions. Further, the sensitivity of the individual base classifiers is also analysed with different dimensionality reduction methods. The outcome of the study has important implications on the current understanding of the MCS on selection of different variants of input data sources and their combined classification performance.

2. An assessment of the existence of relationship between classifier and dimensionality reduction method and information class, classifier and dimensionality reduction methods

The availability of a plethora of classifiers and dimensionality reduction methods makes the identification of optimal choice of classifier and dimensionality reduction method a challenging task. Limited studies are available in comparing the performance of several dimensionality reductions and several classifiers across different hyperspectral datasets. Therefore in this thesis, we assess the relationship between classifiers and dimensionality reduction methods as well as class, classifier and dimensionality reduction methods for hyperspectral image classification. In particular, we conducted experiments to analyse whether i) there exists a optimal pair of classifier and dimensionality reduction method across multiple hyperspectral images; ii) there exists an information class dependent classifiers and dimensionality reduction methods; iii) the observed the relationships are scale dependent; iv) the deployment of multiple dimensionality reduction can increase the classification accuracy of the MCS. Further, the classifiers which are above the threshold are combined and compared with the state-of-the-art-methods.

3. Dynamic classifier system for hyperspectral image classification

Having differential performance among the classifiers is a necessity in image classification by multiple classifier system. The major limitation is that without addressing the input data dynamics, merely inclusion of classifiers from divergent groups may not necessarily yield the results expected from the MCS. In this thesis, we present a novel system named the dynamic classifier system to acquire the capability to dynamically identify optimal pairs of classifiers and dimensionality reduction methods from the pool of classifiers to simplify, and reduce both data and application dependence of MCS. The diversity and performance measures are used as the criteria in pairing the optimal classifiers and dimensionality reduction methods. Apart from this, the impact of different non-trainable and trainable combination functions on the performance of the MCS and on the proposed dynamic classification system is also analysed. The experimental results of the proposed system are compared with the state-of-the-art approaches.

4. Dynamic classifier selection approaches for hyperspectral image classification

To increase the performance of MCS, the optimal subset of classifiers are selected from the pool of the classifiers, as implemented in the dynamic classifier system. However, the selection of classifiers in this method is independent of the location of image pixel in the feature space; hence all the classifiers take part in classifying each image pixel. On the other hand, the optimal subset of classifiers varies for different spatial locations in the image. Therefore, the performance of MCS can be improved by selecting the best subset of classifiers dynamically relative to each image pixel, known as dynamic classifier/ensemble selection (DCS/DES). In this thesis, we explore the potential of the dynamic classifier selection approaches available in the non-imaging signal processing arena and propose a new dynamic classifier selection approach for hyperspectral image classification.

5. Spectral-spatial classification model for hyperspectral image classification

In addition to high spectral information content of the hyperspectral image, recent hyperspectral sensors also provide rich spatial information. Most of the studies in the literature deal with incorporation of the spatial contextual information for a single clas-

sifier. However, limited studies are available in incorporating the spatial contextual information in the MCS. In this thesis, we develop a new spectral-spatial classification model for hyperspectral image classification. The spectral information is extracted from the dynamic classifier selection approach and the spatial information is derived from the Markov Random Field framework. The proposed novel model can be considered as a unified framework to take the full advantages of hyperspectral image.

1.3 Structure of Thesis

The remainder of this thesis has been organized in such a way that the specific objectives listed above are presented in different chapters. Each chapter is made self-contained and includes introduction, materials and methods, results and analysis, discussion and conclusions.

- Chapter 2 presents an overall view of multiple classifier system and theoretical foundations of classifiers and dimensionality reduction methods which are used in designing the MCS in the later chapters.
- Chapter 3 addresses the impact of different dimensionality reduction methods on the classification performance of MCS. The analysis is carried out to understand the sensitiveness of the classifiers in the MCS with changing dimensionality reduction method.
- Chapter 4 assesses the existence of empirical relationship between classifier and dimensionality reduction method, and class, classifier and dimensionality reduction methods for hyperspectral image classification. Further, in order to understand whether the observed relationship is sensitive to the change in spectral and spatial resolution of the image, the experiments were also performed on the spectrally and spatially downscaled synthetic hyperspectral images generated in this thesis.
- Chapter 5 proposes an extended version of the multiple classifier system we named as dynamic classifier system to select the optimal classifiers relative to each input dimensionality reductions to form an effective MCS. The diversity and classifier's performance measure are considered as the criteria in selecting

the classifier in the MCS. Further, the impact of different non-trainable and trainable combination functions is also studied in the framework of dynamic classifier system.

- Chapter 6 explores the potential of dynamic classifier selection approaches for hyperspectral image classification. We propose a new dynamic classifier selection approach based on extreme learning machine regression to offer better results in terms accuracy and computational complexity. Another contribution of this chapter is introducing a new spectral-spatial classification model to incorporate the spatial contextual information in the hyperspectral image classification system.

- Chapter 7 presents the summary and future directions of this thesis.

CHAPTER 2

THEORETICAL BACKGROUND

Prelude: In this chapter, the background of multiple classifier system, reasons for its increasing relevance within the context of pattern recognition and image classification, and the combination functions used in it are presented. Further, mathematical foundations of the dimensionality reduction methods, and classifiers which are used within the framework of multiple classifier system in the later chapters of this thesis are briefly described.

2.1 Multiple Classifier System

2.1.1 An overview of multiple classifier system

Multiple classifier system (MCS) is a classification framework that enables the combination of the decision function values of several individual classifiers to obtain higher classification accuracies. The classifiers forming the MCS are called base classifiers. Figure 2.1 shows general schematic diagram of a MCS. The MCS has been pursued as an important topic of research in machine learning although with different names in literature such as combination of multiple classifiers, classifier ensemble, mixtures of experts, divide and conquer classifiers, and composite classifier system (Dasarathy and Sheela, 1979; Jacobs et al., 1991; Xu et al., 1992; Hull and Srihari, 1994; Kuncheva et al., 2001; Prasad and Bruce, 2011).

Kanal (1974) gave strong motivation in the late 70's in his review work that there is no single model which works well for all the pattern recognition problems, rather it is better to use a bag of tools to solve problems more effectively. The idea of multiple classifiers originated in 70's and the significant developments in this field were achieved during 90's. Hansen and Salamon (1990) demonstrated the potential of combing similarly configured neural networks to improve predictive performance. Since then, the MCS has been studied extensively in theoretical and empirical studies

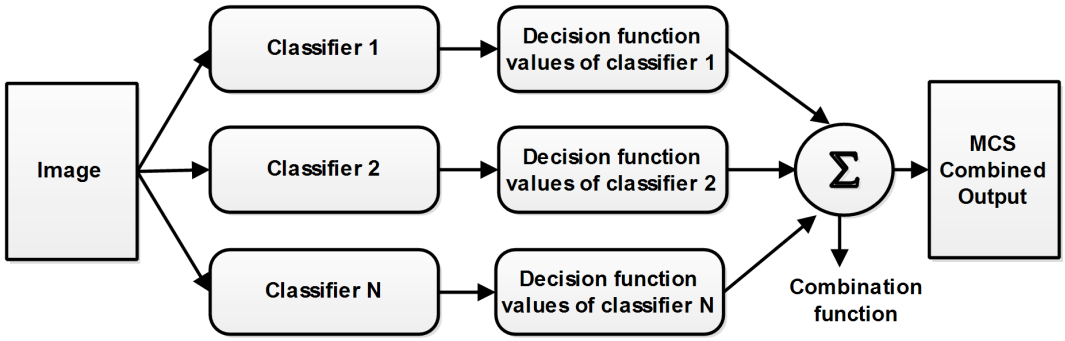


Figure 2.1: Flow chart of multiple classifier system.

(Schapire, 1990; Jacobs et al., 1991; Xu et al., 1992; Hull and Srihari, 1994; Lam and Suen, 1995). Schapire (1990) introduced the 'boosting' algorithm and showed that a stronger classifier can be generated by combining several weak classifiers. Breiman (1996) introduced the 'bagging algorithm'. These two algorithms have laid the foundations for the famous *Adaboost*, and *Random forest* classifier and their variants which have shown great performance in many pattern recognition applications (Freund and Schapire, 1996; Breiman, 2001; García-Pedrajas et al., 2007; Galar et al., 2012). The benefits of several individual classifiers are merged by aggregating the individual classifiers decision function values using combination functions. Designing a combination function is the main component in the MCS. Kittler (1998) developed a common theoretical framework for combining classifiers and showed that performance can be improved by using simple combination functions (Lam and Suen, 1995, 1997; Alexandre et al., 2001; Fumera and Roli, 2005; Mi and Huo, 2011). However, the development of an effective combination function alone is not sufficient to increase the classification accuracy. In addition there should be difference in the opinions among the classifiers forming the MCS, generally called as diversity. In other words, classifiers should commit different types of errors. Diversity is one of the crucial factors that determines the success of MCS. It is argued that diversity among the weak classifiers could be the main reason for the outstanding performance of bagging and boosting algorithms (Kuncheva et al., 2002; García-Pedrajas et al., 2007). Several studies have attempted to study the relationship between the level of diversity and accuracy from the MCS (Brown et al., 2005; Kuncheva et al., 2002; Shipp and Kuncheva, 2002; Windeatt, 2005). Kuncheva and Whitaker (2003) show that the pair of classifiers should have negative dependence and an empirical relationship is evident between MCS accuracy and some of the di-

versity measures (but not always). Thus, diversity among the classifiers is an essential component in the MCS. The methods of introducing and creating diversity among the base classifiers are indeed an active area of research in recent years.

2.1.2 Reasons for MCS

Dietterich (2000) states three reasons why MCS is better than a single classifier. These three reasons are: *statistical, computational and representational*.

- *Statistical*: In classification problems, the choice of classifier for the problem at hand is decided based on the training performance of the classifiers. However, there is a possibility that the chosen classifier may not yield good generalization performance on the test dataset. For instance, two different classifiers with the same training accuracy may have different testing accuracies. This problem particularly arises when the available labelled samples are limited, which is often the case with hyperspectral images. The data complexity of hyperspectral image and inadequate training samples makes it difficult to characterize land cover class information, thus ending up with poor generalization performance. In such a scenario, a combination of several classifiers reduces the risk of selecting a poorly performing classifier and may also lead to good generalization performance.

- *Computational*: In general, classifiers' training process starts from a random initialization point and ends closer to the optimal point in the classifier space. The different initialization points of same algorithm leads to different optimal points in the classifier space. Further, the local search of some algorithms may get stuck in the local optima. These problems can be dealt with by running the local search with different initializations and combining all of them will yield a better approximation than the individual classifiers. The computational arguments particularly important if the classifiers' performances are very sensitive to change of their parameters and input datasets.

- *Representational*: In most of the cases, the classifier space does not contain an optimal classifier for a given problem at hand. In such cases, combining several classifiers may lead to a classifier with an optimal decision boundary. For example, the combination of several linear classifiers can approximate the non-linear

decision boundary of the optimal classifier. In case of hyperspectral images, the spectral signatures vary across different parts of the image. This introduces the complexity in finding the optimal decision boundary, and hence the combination of classifiers is a suitable approach for such problems.

2.1.3 Combination functions

The combination function is a mathematical function/scheme which combines the intermediate decision function values of different classifiers in the MCS to produce a final classified image. There are two types of combination functions: non-trainable combination function, and trainable combination function.

The non-trainable combination function does not require training once the classifier's produce decision function values. On the other hand, the trainable combination function requires training for estimating weight of the combination function (Kuncheva, 2004). The non-trainable combination functions used are: majority voting, average rule, maximum rule, minimum rule, median rule and product rule. The trainable combination function can be formulated as the stack generalization problem, where one more learner is used on the classifiers' decision function value in order to get final classified results.

Let $\mathbf{x} \in R^n$ be the sample and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be the set of class labels, and let $\Psi = \{\psi_1, \ldots, \psi_L\}$ be the set of classifiers in the ensemble or MCS and each of the classifier ψ_l produces the c decision function vectors $[d_{l,1}(\mathbf{x}), \ldots, d_{l,c}(\mathbf{x})]$. Then, the L classifiers' outputs for a given $\mathbf{x}$ can be organized as a decision profile ($\mathbf{DP}$) matrix as shown in Figure 2.2.

The combination function can be formulated as a mapping a function $F : R^{LC} \rightarrow R^C$ that maps the decision profile matrix to set of class labels, such that $\boldsymbol{\mu} = F(DP)$, where $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_c\}$. The class label of $\mathbf{x}$ is assigned as the maximum index value of $\boldsymbol{\mu}$.

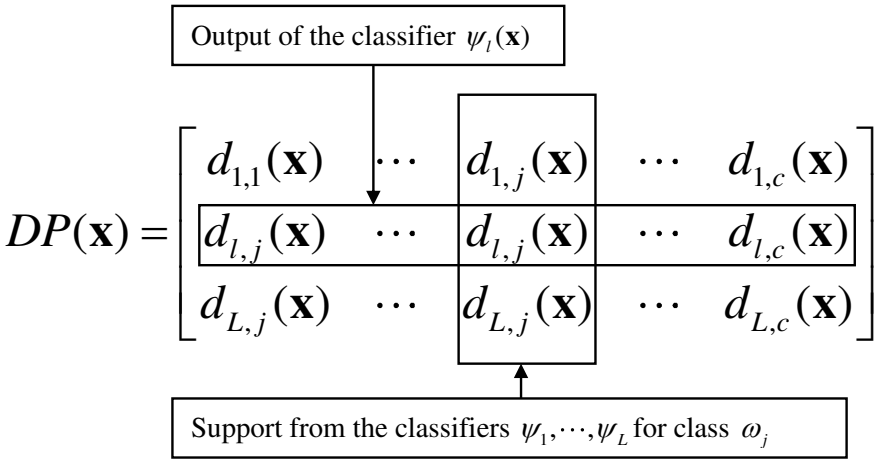


Figure 2.2: Decision profile matrix of x from the L classifiers.

2.1.3.1 Non-trainable combination function

- Majority voting (MV)

  MV is a simple and robust combination function used in MCS. The number of classifiers that agree on each class label is counted. Then the test pixel is assigned to the class which has the highest number of votes. Let us assume that the classifiers produce c binary decision function vectors $[y_{l,1}, y_{l,2}, \ldots, y_{l,c}] \in \{0, 1\}$, where $y_{l,j} = 1$ if classifier ψ_l labels $\mathbf{x}$ in the class ω_j, 0 otherwise. Then the MV is given as

$$\mu_j(\mathbf{x}) = \sum_{i=1}^{L} y_{i,j}(\mathbf{x}), \quad j = 1, 2, \ldots, c \tag{2.1}$$

- Average rule

$$\mu_j(\mathbf{x}) = \frac{1}{L} \sum_{i=1}^{L} d_{i,j}(\mathbf{x}), \quad j = 1, 2, \ldots, c \tag{2.2}$$

- Maximum rule

$$\mu_j(\mathbf{x}) = \max_i \{d_{i,j}(\mathbf{x})\}, \quad j = 1, 2, \ldots, c \tag{2.3}$$

- Minimum rule

$$\mu_j(\mathbf{x}) = \min_i \{d_{i,j}(\mathbf{x})\}, \quad j = 1, 2, \ldots, c \tag{2.4}$$

- Median rule

$$\mu_j(\mathbf{x}) = \operatorname{median}_i \{d_{i,j}(\mathbf{x})\}, \quad j = 1, 2, \ldots, c \tag{2.5}$$

- Product rule

$$\mu_j(\mathbf{x}) = \prod_{i=1}^{L} d_{i,j}(\mathbf{x}), \quad j = 1, 2, \ldots, c \tag{2.6}$$

2.1.3.2 Trainable combination function

The trainable combination function can be modelled as the stack generalization approach. The idea is to apply the learner after stacking the output of the base classifiers (Wolpert, 1992; Duin, 2002; Jin Chen et al., 2009). The assumption is that the classifiers' intermediate output could also provide the discriminant information and the learner (or combiner) catches these information. The trainable combination function can be formulated as

$$F(\mathbf{s}) = \mathbf{W}\mathbf{s} + b \tag{2.7}$$

where F is the combination function (learner), $\mathbf{s}$ is the stacked output of all the classifiers. The aim is to find the weight matrix $\mathbf{W} \in R^{N \times MN}$ and $b \in R^N$ using the $\{(\mathbf{s_i}, \omega_i)\}_{i=1}^{I}$. In other words, the base classifiers' outputs are linearly weighted and combined to get enhanced classification results.

2.2 Dimensionality reduction methods

Advances in imaging technologies have led to data overload in remote sensing. While the high dimensional remote sensing dataset provides abundant information to explore and critically examine the physical and chemical properties of materials, it poses challenges in processing and analyzing the data. Previous studies have demonstrated that high dimensional data spaces are mostly empty, and the data tends to lie in lower dimensional subspace (Jimenez and Landgrebe, 1998). Hence representing hyperspectral image in a lower dimensional space is essential for the betterment of hyperspectral image classification. The application of dimensionality reduction methods on hyperspectral images can

- cure the high dimensionality problem,
- reduce computational load in storage and processing time,

- translate redundant features into a non-redundant features (for instance, independent, orthogonal) depending on the nature of the dimensionality reduction methods,
- reduce the effect of noise in the data.

Dimensionality reduction methods transform the high dimensional dataset into a lower dimensional representation with some degree of loss of information, however small it may be. Let $\mathbf{A} \in R^{M \times N}$ be the dataset, where M is number of pixels in each band, and N is the number of dimensions (bands). Dimensionality reduction methods transform the N dimensional dataset $\mathbf{A} \in R^{M \times N}$ into K dimensional dataset $\mathbf{Y} \in R^{M \times K}$ where $K << N$ by suitable transformation matrix $\mathbf{W} \in R^{N \times K}$ as

$$\mathbf{Y} = \mathbf{AW}. \tag{2.8}$$

In general K is called as intrinsic dimension of the dataset, however in practice the intrinsic dimension is not known. The general aim of dimensionality reduction method is to find the transformation matrix $\mathbf{W}$ by optimizing criteria function. A brief description of the most commonly used dimensionality reduction methods which are used in this thesis is provided below.

2.2.1 Principal component analysis

Principal component analysis (PCA) translates the correlated features into a set of orthogonal features by seeking a new set of directions which maximize the variance of data. The basis of the transformation function is obtained by solving the following eigenvalue problem

$$\text{cov}(\mathbf{A})\mathbf{W} = \mathbf{\Lambda W} \tag{2.9}$$

where $\text{cov}(\mathbf{A})$ is the covariance matrix of the data, $\mathbf{W}$ is the eigenvectors and $\mathbf{\Lambda}$ is the eigenvalues of the covariance matrix. Since the covariance matrix is real and symmetric, the resulting eigenvectors are orthogonal. The eigenvalues represent the variance of the data and eigenvectors represent direction of spread of the data. The eigenvectors matrix $\mathbf{W}$ is used as the basis for PCA transformation and the lower di-

mensional representations ($\mathbf{Y}$) of the dataset $\mathbf{A}$ is obtained as $\mathbf{Y} = \mathbf{AW}$. The eigenvector corresponding to the highest eigenvalue is known as the first principal direction and the data projected on this axis represent the maximum variance of the original data, called as first principal component. The number of lower dimensions required is decided based on the eigenvalue threshold.

2.2.2 Minimum noise fraction

Minimum noise fraction (MNF) reduces the dimension of hyperspectral image by segregating noise in the data. The image pixel vector in MNF is modelled as

$$\mathbf{x_i} = \mathbf{s_i} + \mathbf{n_i}$$

where $\mathbf{x_i}$ is the ith hyperspectral image pixel having N bands, $\mathbf{s_i}$ and $\mathbf{n_i}$ are the signal and noise components respectively. The MNF transformation is essentially a two cascaded principal component analysis. The basis of MNF transformation is obtained by maximizing the noise fraction $\frac{\text{var}(\mathbf{n_i})}{\text{var}(\mathbf{x_i})}$ in the projected components.

$$\mathbf{Y} = \mathbf{W}^T\mathbf{A} \tag{2.10}$$

$$\mathbf{Y} = \mathbf{U}^T\mathbf{V}^T\mathbf{A} \tag{2.11}$$

The first transformation is based on the estimated noise co-variance matrix ($\mathbf{C_N}$)[1], which decorrelates and rescales noise in the data such that the noise has unit variance and no band to band correlation.

$$\mathbf{V}^T\mathbf{C_N}\mathbf{V} = \mathbf{I}$$

and the transformation matrix $\mathbf{V}$ can be obtained by $\mathbf{V} = \mathbf{B}\mathbf{\Lambda}^{-1/2}$ with $\mathbf{B}$ being the matrix that diagonalizes $\mathbf{C_N}$ and $\mathbf{\Lambda}$ being diagonal matrix

$$\mathbf{B}^T\mathbf{C_N}\mathbf{B} = \mathbf{\Lambda}$$

[1]The covariance of the noise is in general not known for hyperspectral image. However, a reasonable estimate of noise can be obtained through maximum autocorrelation factor by exploiting the correlation in the spatial neighbourhood of the image pixels.

The second transformation is the standard PCA of the noise whitened data. The covariance matrix of noise whitened image ($\tilde{\mathbf{C}_\mathbf{A}}$) can be obtained by

$$\tilde{\mathbf{C}_\mathbf{Z}} = \mathbf{D}^T \mathbf{C}_\mathbf{A} \mathbf{D}$$

and from the diagonalization or eigenvalue decomposition of matrix $\tilde{\mathbf{C}_\mathbf{Z}}$, we have

$$\mathbf{U}^T \tilde{\mathbf{C}_\mathbf{A}} \mathbf{U} = \mathbf{\Delta} \tag{2.12}$$

where $\mathbf{U}, \mathbf{\Delta}$ are the corresponding eigenvectors and eigenvalues of the matrix $\tilde{\mathbf{C}_\mathbf{A}}$. Thus the final MNF transformed image ($\mathbf{Y}$) can be obtained as

$$\mathbf{Y} = \mathbf{W}^T \mathbf{A}. \tag{2.13}$$

This projection results in high signal to noise ratio bands which are ordered according to noise fraction in the transformed domain. The inherent dimensionality of the image is determined by examining the eigenvalues as similar to PCA. Further, MNF transform can also be used to remove the noise from the original bands by performing inverse transform using a spectral subset of the data, which includes only the high SNR bands.

2.2.3 Independent component analysis

Independent component analysis (ICA) was originally developed for unmixing a linear combination of signals in the signal processing community. However, the same formulation can also be used for reducing the dimension of high dimensional hyperspectral data. The key of idea of ICA is to find a maximum statistically independent basis by minimizing the statistical dependency using higher order statistics. The central limit theorem states that if the Gaussian like observations are broken into a set of non-Gaussian mixtures, then the individual signals will be independent. Let $\mathbf{x}$ be the observed signal

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2.14}$$

where $\mathbf{A}$ is the scalar mixing matrix and $\mathbf{s}$ is the vector of source signals. The

ICA assumes that the components of $\mathbf{s}$ are statistically independent and they have non-Gaussian distribution. Thus the maximization of non-Gaussianity leads to independent components. The kurtosis (fourth order moment) can be used as the measure for non-Gaussianity. Hence, the objective of ICA is to find a matrix $\mathbf{W} = \mathbf{A}^{-1}$ that maximizes the non-Gaussianity in the projected components $\mathbf{y}$. The Fast ICA algorithm was employed to obtain the transformation matrix $\mathbf{W}$ (Hyvärinen, 1999; Hyvärinen and Oja, 2000).

$$\mathbf{y} = \mathbf{W}\mathbf{x} \Rightarrow \mathbf{y} = \mathbf{W}\mathbf{A}\mathbf{s} \tag{2.15}$$

2.2.4 Discrete wavelet transform based dimensionality reduction method

Discrete wavelet transform is a powerful tool used in many image processing applications such as image compression, feature extraction, registration, fusion and image segmentation. The inherent multi-resolution property makes wavelet transform an excellent tool for feature extraction from hyperspectral images. The basis of wavelet transform is fixed unlike PCA, MNF, and ICA transformations. The wavelet transform decomposes a signal into different levels of decomposition using scaling ($\phi(t)$) and wavelet ($\psi(t)$) basis functions.

$$\phi_{(j,k)}(t) = 2^{j/2} \sum_{k} h(k)\phi(2^j t - k) \tag{2.16}$$

$$\psi_{(j,k)}(t) = 2^{j/2} \sum_{k} g(k)\phi(2^j t - k) \tag{2.17}$$

where $h(k), g(k)$ are the low pass, high pass filter coefficients respectively, and j is the level of decomposition. The filter coefficients are derived using the properties of scaling and wavelet functions. The forward wavelet transform projects the signal into the scaling and wavelet basis functions, thus resulting in an approximation, and detailed wavelet coefficients. The forward wavelet transform consists of filtering and down-sampling. The inverse wavelet transform is performed to reconstruct the decomposed signal back into the original signal and it consists of up-sampling and filtering.

The discrete wavelet transform (DWT) can be used to reduce the dimensionality of hyperspectral image in two stages (Kaewpijit et al., 2003). The first stage determines

the level of decomposition and the second stage reduces dimensionality of hyperspectral image. Let $\{\mathbf{x}_i, y_i\}_{i=1}^{m}$ be a set of training samples and its corresponding c class labels and let $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_c]$ be the mean of training samples belonging to each class. The steps followed can be summarized as

- step 1: Each of the class mean vectors is decomposed into the approximation and detailed coefficients using a forward wavelet transform.
- step 2: The inverse wavelet transform is computed only with approximation coefficients by discarding the detailed coefficients to reconstruct the original signal.
- step 3: The level (j) of decomposition for a given class mean pixel is decided based on the acceptable correlation threshold with the original class mean pixel.
- step 4: The required level of decomposition (L) for reducing the dimension of hyperspectral image is determined by the lowest level of decomposition in step 3 for all the class mean vectors.
- step 5: Using the computed level of decomposition in step 4, all the hyperspectral image pixels are decomposed into level L with low-pass filter coefficients. If the original image contains N bands, the number of bands in the dimensionality reduced image is $N/2^L$

2.2.5 Kernel principal component analysis

Kernel principal component analysis (KPCA) is the non-linear version of PCA and is capable of capturing the higher-order statistics to better represent information in original data (Scholkopf and Smola, 2001). The input data points are mapped into a higher dimensional feature space by kernel mapping function, and the normal PCA is performed in the higher dimensional feature space. Let $\mathbf{x}_i$ be the set of input data points, and $\phi(\mathbf{x}_i)$ be the kernel mapping function. The computation of PCA in the high dimensional feature space has a high computationally cost. Using the kernel trick, it is possible to performs all the calculations in the input data space itself (Scholkopf and Smola, 2001). The kernel function reduces the dot product in the feature space to a

function in the input space as

$$\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) \tag{2.18}$$

The kernel PCA is performed by solving the following eigenvalue problem

$$\lambda\alpha = \mathbf{K}\alpha, \quad \text{subject to } \|\alpha\|_2 = \frac{1}{\lambda} \tag{2.19}$$

where $\mathbf{K}$ is the kernel matrix, the polynomial and Gaussian function are most used as the kernel functions. After solving the eigenvalue problem, the projection is performed as

$$\mathbf{Y} = \alpha\mathbf{K} \tag{2.20}$$

2.2.6 Optimal band selection

The optimal band selection reduces dimensionality of hyperspectral image by selecting the best subset of bands based on physical properties of the objects as described in the literature. This method enables incorporating the experts' domain knowledge on the available land cover information classes in the classification system. The selection of bands (wavelengths) depends upon number of classes, nature of classes and underlying image itself.

2.3 Classifiers

Let $\{\mathbf{x}_i, y_i\}_{i=1}^N$ be the pair of training samples and corresponding class labels. The classification problem can be defined as estimating a function $f : \boldsymbol{\chi} \rightarrow \Omega$ that maps the input space $\boldsymbol{\chi} \in R^n$ to the set of class labels $\Omega = \{\omega_1, \ldots, \omega_c\}$. The function f is called as classifier or learner. If $\mathbf{x}$ is the unknown pixel (test pixel), then labelling is performed as $\mathbf{x} \in \omega_j$, such that $j = \arg\max f(\mathbf{x})$

Image classifiers can be categorized into two groups: parametric classifiers and non-parametric classifiers. The former assumes that the data follows the statistical distribution, where as the later doesn't need to have the data following any statistical distribution. This section describes the classifiers used in this thesis. Based on the nature of

classifiers, the classifiers can be further categorized into three groups: spectral matching methods, covariance based methods and subspace based methods.

2.3.1 Spectral matching methods

The spectral matching methods consist of normalized Euclidean distance classifier (NED), spectral angle mapper (SAM), spectral similarity measure (SSM).

2.3.1.1 Normalized Euclidean distance

NED is the common geometric distance matching algorithm, which computes Euclidean distance between reference pixel and unknown pixel (Robila and Gershman, 2005). The unknown pixel is classified to a class which is at a minimum distance to the mean spectra of the classes. The Euclidean distance is given by

$$T_{NED_i}(\mathbf{x}) = \frac{1}{N}\left((\mathbf{m}_i - \mathbf{x})^T(\mathbf{m}_i - \mathbf{x})\right), \quad i = 1, 2, \ldots, c \tag{2.21}$$

where N is the number of elements in the image pixel vector, $\mathbf{m_i}$ is the desired ith class vector (mean class vector) and $\mathbf{x}$ is the unknown pixel vector. The unknown pixel is assigned to the class, which has minimum distance measure.

2.3.1.2 Spectral angle mapper

Spectral angle mapper (SAM) is one of the leading image classification and spectral library search methods used in hyperspectral image due to its simplicity and insensitive to illumination changes. SAM measures cosine angle between the reference pixel and unknown pixel (Kruse et al., 1993). Unknown pixels are assigned to the class, which has minimum angle measure. It is defined mathematically as

$$T_{SAM_i}(\mathbf{x}) = \cos^{-1}\left(\frac{\mathbf{m}_i^T\mathbf{x}}{\|\mathbf{m}_i\|\|\mathbf{x}\|}\right), \quad i = 1, 2, \ldots, c \tag{2.22}$$

2.3.1.3 Spectral similarity measure

The spectral similarity measure (SSM) is a hybrid approach which combines the relative merits of both the spectral angle mapper and normalized Euclidean distance classifier (Granahan and Sweet, 2001). The SSM constitutes two parts: in the first part Euclidean distance is computed between reference pixel and unknown pixel; the second part measures similarity in shape between the reference and unknown pixel using correlation measure.

$$T_{SSM_i}(\mathbf{x}) = \left(\frac{1}{N} \sum_{j=1}^{N} \left(m_i(j) - x(j)\right)^2 + \left(1 - \left[\frac{\frac{1}{N-1} \sum_{j=1}^{N} \left(m_i(j) - \mu_{m_i}\right)\left(x(j) - \mu_x\right)}{\sigma_{m_i}\sigma_x} \right]^2 \right)^2 \right)^{1/2} \tag{2.23}$$

where N is the number of the spectral bands; μ_{m_i} and μ_x the mean of the reference pixel $\mathbf{m}_i$ and the unknown image pixel $\mathbf{x}$ respectively; σ_{m_i}, σ_x the standard deviation of $\mathbf{m}_i, \mathbf{x}$ respectively.

2.3.2 Covariance based methods

The covariance based methods consist of matched filter (MF), and adaptive coherence estimator (ACE).

2.3.2.1 Matched filter

Pixel labelling in supervised classification can be considered as a target detection problem (Harsanyi and Chang, 1994; Du et al., 2003) formulated as a binary hypothesis with two competing hypotheses: target absent (background) or target present. Matched Filter (MF) can be derived from the Likelihood ratio detectors, which assumes that the target and background classes follow the multivariate normal distributions with different mean and covariance (Manolakis et al., 2003). This formulation thus leads to quadratic detector by Neyman-Pearson detector and becomes linear detector if both hypotheses

have same covariance ($\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_b$).

$$T_{MF}(\mathbf{x}) = (\boldsymbol{\mu}_t - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}^{-1}\mathbf{x} \tag{2.24}$$

However, in practice it is quite difficult to satisfy this assumption and these quantities are estimated from the available data using maximum likelihood estimates. Finally, the MF can be described as

$$T_{MF_i}(\mathbf{x}) = \frac{(\mathbf{m}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{(\mathbf{m}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{m}_i - \boldsymbol{\mu})}, \quad i = 1, 2, \ldots, c \tag{2.25}$$

where $\mathbf{m}_i$ is the desired class mean vector, $\mathbf{x}$ is the input pixel vector, $\boldsymbol{\mu}$ is the background mean vector and $\boldsymbol{\Sigma}$ is the background covariance matrix.

2.3.2.2 Adaptive coherence estimation

In hyperspectral data, the background statistics are unknown and the variability of background can be described either by subspace model (structured background) or a statistical distribution (unstructured background). The unstructured background models the target pixel as a subspace model and includes additive noise in the background. The Generalized Likelihood Ratio (GLR) is the unstructured background model which has same covariance for both hypotheses. If we model the covariance structure of both hypotheses with different variance, GLR approach leads to the ACE (Manolakis et al., 2003).

$$T_{ACE_i}(\mathbf{x}) = \frac{\left(\mathbf{m}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{x}\right)^2}{\left(\mathbf{m}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{m}_i\right)\left(\mathbf{x}^T \boldsymbol{\Sigma}^{-1}\mathbf{x}\right)}, \quad i = 1, 2, \ldots, c \tag{2.26}$$

where $\boldsymbol{m}_i$ is the desired class mean vector, $\boldsymbol{\Sigma}$ is the back ground covariance matrix, $\mathbf{x}$ is the input pixel vector. When the whitening transform is applied to the data, then the ACE is equal to the cosine angle between the test pixel and the target subspace into whitened space.

2.3.3 Subspace based methods

The subspace based methods consist of orthogonal subspace projection (OSP), and target constrained minimized interference filter (TCIMF).

2.3.3.1 Orthogonal subspace projection

The orthogonal subspace projection (OSP) approach has received considerable interest in hyperspectral target detection applications. The OSP models hyperspectral pixel vector as a linear combination of a set of finite class signatures present in the image (Harsanyi and Chang, 1994). Let $\mathbf{x}$ be the pixel vector of size $N \times 1$, then

$$\mathbf{x} = \mathbf{M}\boldsymbol{\alpha} + n \tag{2.27}$$

where $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_c]$ is a $N \times c$ signature matrix with c classes, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_c]$ is a $c \times 1$ coefficient vector. The OSP model divides the signature matrix $\mathbf{M}$ into two parts, desired signature of interest $\mathbf{d} = \mathbf{m}_c$ and the undesired signature matrix $\mathbf{U} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{c-1}]$ as

$$\mathbf{x} = \mathbf{d}\alpha_d + \mathbf{U}\boldsymbol{\alpha}_{\boldsymbol{u}} + n \tag{2.28}$$

where α_c is the coefficient of the desired target signature. The goal of OSP is to find an orthogonal complement projector that eliminates the undesired signature matrix (background matrix) by projecting the data onto a subspace orthogonal to the undesired signatures in $\mathbf{U}$. The orthogonal subspace projector can be derived as $\mathbf{P}\mathbf{U}^{\perp} = \mathbf{I} - \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$. The OSP classifier projector which maximizes the SNR is derived as

$$T_{osp_i}(\mathbf{x}) = \mathbf{d}^T\mathbf{P}\mathbf{U}^{\perp}\mathbf{x}, \quad i = 1, 2, \ldots, c \tag{2.29}$$

2.3.3.2 Target constraint interference minimized filter

Target constrained interference minimized filter (TCIMF) is the extension of the constrained energy minimization (CEM) algorithm. TCIMF assumes that the image pixel vector is made up of three separate sources such as desired class signatures $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{nd}]$, undesired class signatures $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{nu}]$ and interference (Ren and Chang, 2000). TCIMF detects multiple target (desired) signal sources, annihilates undesired signal source while suppressing the interference caused by the other signal sources in single operation (Chang, 2005). It is designed by the FIR filter that passes the desired class signatures $\mathbf{D}$ using a $nd \times 1$ unity constraint vector, while

annihilating the undesired class signatures in $\mathbf{U}$ using $nu \times 1$ zero constraint vector. The filter coefficients of TCIMF are derived by solving the constrained optimization problem (Chang, 2005).

$$T_{TCIMF_i}(\mathbf{x}) = (\mathbf{w}_{TCIMF_i})^T \mathbf{x}, \quad i = 1, 2, \ldots, c \tag{2.30}$$

where $\mathbf{w}_{TCIMF_i} = \mathbf{R}^{-1}[\mathbf{DU}]\left([\mathbf{DU}]^T\mathbf{R}^{-1}[\mathbf{DU}]\right)^{-1}\begin{bmatrix} 1_{nd\times 1} \\ 0_{nu\times 1} \end{bmatrix}$, $\mathbf{R}$ is the background correlation or covariance matrix, $\mathbf{x}$ is the pixel vector, $\mathbf{U}$ is the undesired class signature matrix and $\mathbf{D}$ is the desired class signature matrix.

2.3.4 Support vector machine

Support vector machine is a non-parametric binary classification method which has been successfully applied to numerous pattern recognition applications. Let $\{\mathbf{x_i}, y_i\}$ be the training samples ($\mathbf{x_i}$) and the corresponding class labels ($y_i \in \{+1, -1\}$). The objective of the SVM is to find an optimal hyperplane which separates the two classes while maximizing the margin between the two classes. The hyperplane is determined by using few training samples and these samples are called as support vectors. In noisy cases, the two classes may not be linearly separable and in such scenarios the slack variables ξ_i are introduced to allow for the overlapping training samples. The optimal hyperplane is obtained by solving the following optimization problem

$$\begin{aligned} \min_{\mathbf{w},b,\xi} L(\mathbf{w}, b) &= \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i \\ \text{subject to } y_i\left(\mathbf{w}^T\mathbf{x_i} + b\right) &\geq 1 - \xi_i, \forall i = 1, \ldots, n \\ \xi_i &\geq 0 \end{aligned} \tag{2.31}$$

where $\mathbf{w}$ is the normal vector of the separating hyperplane, b is the bias of the separating hyperplane from the origin, and $C > 0$ is a constant that determines the tradeoff between minimizing the training error and maximizing the margin. The optimal C value has to be determined for the good generalization performance and it is data dependent. The above optimization problem 2.31 can be reformulated through a Lagrangian multipliers $\{\alpha_i, i = 1, \ldots, n\}$ as a dual optimization leading to a Quadratic Programming (QP) solution (Scholkopf and Smola, 2001). Then the decision on the test sample $\mathbf{x}$ is

evaluated as

$$f(\mathbf{x}) = \text{sign}\left(\sum_i y_i \alpha_i \mathbf{x}^T \mathbf{x_i} + b\right) \tag{2.32}$$

When the data samples are not linearly separable, the samples are mapped into a higher dimensional feature space using a mapping function, in this case it is called as non-linear SVM. In general, the classes are linearly separable in the higher dimensional feature space, hence the equation 2.31 can be formulated in this space. The kernel trick is used to implicit represent the transformation using the kernel function as

$$K(\mathbf{x}, \mathbf{x_i}) = \Phi(\mathbf{x})^T \Phi(\mathbf{x_i}) = \exp\left(-\frac{1}{2}\frac{\|\mathbf{x} - \mathbf{x_i}\|^2}{\sigma^2}\right) \tag{2.33}$$

where Φ is the function that transforms the input data to a higher dimensional feature space. The equation 2.33 is the popularly used kernel called as Gaussian radial basis function (RBF) kernel (Scholkopf and Smola, 2001; Gustavo Camps-Valls, 2009). In this case, the decision on the test sample $\mathbf{x}$ is evaluated as

$$f(\mathbf{x}) = \text{sign}\left(\sum_i y_i \alpha_i K\left(\mathbf{x}^T \mathbf{x_i}\right) + b\right) \tag{2.34}$$

The Lagrange multipliers α_i weights each training sample according to its importance in determining the hyperplane. The training samples associated with nonzero weights are termed support vectors.

2.4 Chapter Conclusions

The multiple classifier system (MCS) is an emerging methodological framework for combining the relative performances of several independent classifiers to produce a single classified output which is better compared to any of the classifiers forming the MCS. This chapter presented a background of the MCS and the reasons for the evolution of multiple classifier system as an important framework in pattern recognition with promising scope for hyperspectral image classification. A brief mathematical description of the combination functions, classifiers, and dimensionality reduction methods which are used to design the multiple classifier system in this thesis are presented in this chapter.

CHAPTER 3

IMPACT OF DIMENSIONALITY REDUCTION METHODS ON THE PERFORMANCE OF MULTIPLE CLASSIFIER SYSTEM FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Prelude: This chapter presents our empirical study on understanding the impact of dimensionality reduction methods on the classification performance of multiple classifier system (MCS). Transformed image components from the five widely used dimensionality reduction methods are considered as input data sources and seven linear classifiers are considered as the base classifiers in the MCS design. Two statistical dispersion measures are used to analyze the sensitivity of the base classifiers individually and also as combined in the MCS framework relative to different dimensionality reduction methods. Classification experiments are performed on five multi-sensor airborne hyperspectral images covering diverse land cover categories.

3.1 Introduction

Hyperspectral imagery provides rich spectral information to discriminate large number of materials as different materials respond differently in different spectral channels of the electromagnetic spectrum. Supervised image classification has been the method of choice for the discrimination and mapping of a range of land cover categories using multispectral and hyperspectral remote sensing data. The accuracy of image classification depends upon the chosen classifier's ability to trade upon the relationships between available training samples and information classes. While providing sufficient number

This chapter will be published in *International Journal of Remote Sensing,* with the title: " Impact of Dimensionality Reduction Methods on the Performance of the Multiple Classifier System for Hyperspectral Image Classification". Authors: Bharath Bhushan Damaodaran, Rama Rao Nidamanuri

of training samples is not a major problem with multispectral image, it is a major challenge with hyperspectral image because of its huge dimensionality and this is known as Hughes phenomena (Jimenez and Landgrebe, 1998; Landgrebe, 2002). This phenomena arises when the ratio between the number of features to the number of training samples is very small, resulting in inaccurate estimation of parameters of the classifiers and thus leads to the poor generalization capability of the classifier.

To circumvent this problem, a number of dimensionality reduction methods and non-parametric classification schemes are available in the literature (Bruce et al., 2002; Wang and Chang, 2006; Camps-Valls et al., 2004; Camps-Valls and Bruzzone, 2005; Mendenhall and Merenyi, 2008; Zhong and Zhang, 2012). These classifiers are widely differentiable by their performance, application specifications and level of vulnerability to high dimensionality. Dimensionality reduction methods enables the creation of multiple transformations of the same imagery with different mathematical and statistical properties, thus leading to have complimentary information among different dimensionality reduction methods. The performance of the classifiers is also variable based on the information content of the dimensionality reduction methods. It has further been established that there is no ideal classifier that can offer optimal performance across applications, sites and data-specific conditions of image classifications (Kanal, 1974; Fabio et al., 1997; Wolpert and Macready, 1997). Hence, the identification of classifiers and dimensionality reduction methods optimal to the specific application at hand has become a key step in image classification. In spite of image classification goals being precise and objective, there is always a high risk of the selected method being suboptimal. To minimise this enormous subjectivity, the ensemble classification approach, popularly known as multiple classifier system (MCS) has been proposed as a general image classification paradigm that offers acceptable classification results under various application domains (Kittler, 1998; Dietterich, 2000; Kuncheva, 2004; Rokach, 2010; Ceamanos et al., 2010).

The MCS consists of a set of image classification algorithms, called as base classifiers within the MCS terminology; each base classifier produces its own decision function value for labelling an unknown pixel. Then, the decision values of all the base classifiers are combined in an intelligent way to produce a final classified output, whose accuracy and precision is better than any of the classifiers used. The MCS, in principle, seems to be the preferred approach over the typical supervised classification approach

for handling hyperspectral images because of the condition that different spectral signatures lead to different levels of image property generalization. The different generalization levels of classifiers diminish the risk of picking up a bad classifier (Dietterich, 2000). Employing an effective combination scheme and creation of diversity among the chosen base classifiers are the major factors that contribute to successful application of MCS for hyperspectral image classification. A large body of literature is available reporting on the development of theoretical and implementation strategies for creating and capturing diversity, and sophisticated combination schemes (Benediktsson and Kanellopoulos, 1999; Gabrys and Ruta, 2006; Fauvel et al., 2006; Prasad and Bruce, 2008b; H et al., 2010; Ceamanos et al., 2010; Yang et al., 2010b).

Incorporating the dimensionality reduction method as an integral part in the design of the MCS architecture assumes significance especially when the classifiers are parametric, and complex. Akin to the classifier selection, dimensionality reduction is data and application specific. The inclusion of dimensionality reduction methods in the design of the MCS has been studied in literature (Benediktsson and Kanellopoulos, 1999; Prasad et al., 2008; Kalluri et al., 2010) and the performance is validated under only one particular dimensionality reduction method. The different dimensionality reduction methods has the capability to provide different levels of data transformation and generalization, and influence classifier's performance significantly. Therefore understanding these important properties of various dimensionality reduction methods on the classification performance of the MCS is necessary for practical applications. Our extensive literature survey reveals that there is limited or no literature available addressing the potential impact of the differential data generalization by various dimensionality reduction methods on the performance of MCS for hyperspectral image classification.

The objective of this chapter is to assess the impact of various dimensionality reduction methods on the performance of a MCS designed for classifying airborne hyperspectral images for discriminating various land cover categories. In particular, we provide an empirical analysis with different dimensionality reduction methods on the fixed design of MCS to study the performance sensitivity of the MCS to the different input data sources to the MCS.

This chapter is organized into five sections. The next section describes the datasets used in our analysis while the section 3.3 presents the dimensionality reduction meth-

ods, classifiers, and methodological framework used to investigate the goals of this chapter. Section 3.4 describes and discusses the experimental results obtained. Finally, the last section presents conclusions of this chapter.

3.2 Datasets

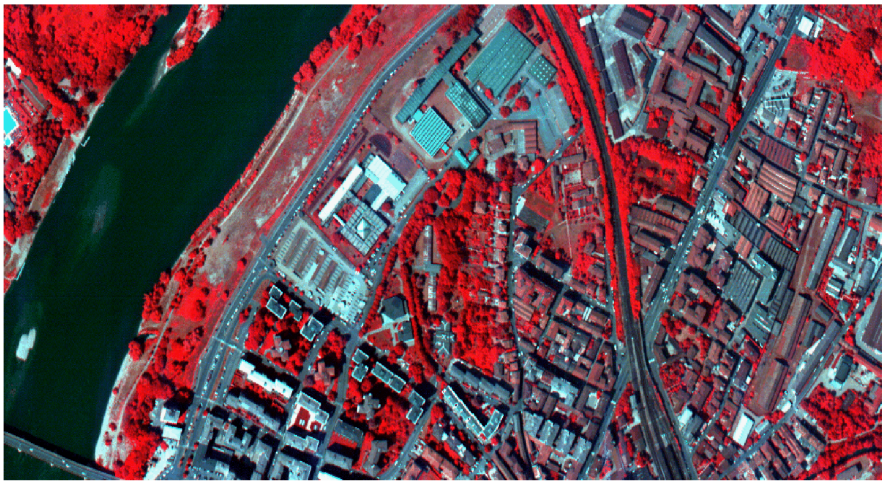
We used five different sources of airborne hyperspectral images (one image each from HyMAP, ProSpecTIR, HYDICE and two images from ROSIS airborne hyperspectral imaging system) covering several land cover categories and sites. As there are two images acquired by the ROSIS sensor for two different sites, we appended name of the location to the ROSIS image for its readily identification. False colour composites of the image are shown in Figures 3.1, 3.2.

HyMAP image: The HyMAP hyperspectral image was acquired over the Dedelow research station of the Leibnitz-Centre for Agricultural Landscape Research (ZALF), Germany on 9 May, 1999 (Nidamanuri and Zbell, 2011). The predominant land use categories in the study site were agricultural crops namely winter barley, winter rape, winter wheat and winter rye, built up, and grass. The image has a spatial resolution of 5 m and 128 spectral bands in the spectral range 0.40 to 2.48 μm. A subset of the image acquired was used in this study.

ROSIS-University image: The next hyperspectral image used in our experiment was acquired on 8 July, 2002 over the University of Pavia, Italy, by ROSIS airborne hyperspectral sensor in the framework of the HySens Project managed by DLR (German Aerospace Agency) (Fauvel et al., 2009). The image has 103 spectral bands in the spectral range 0.43 to 0.86 μm with spatial resolution of 1.3 m. This image consists of ten land cover classes namely, trees, asphalt, meadow, gravel, metal sheet, bare soil, bitumen, bricks, shadows and built up.

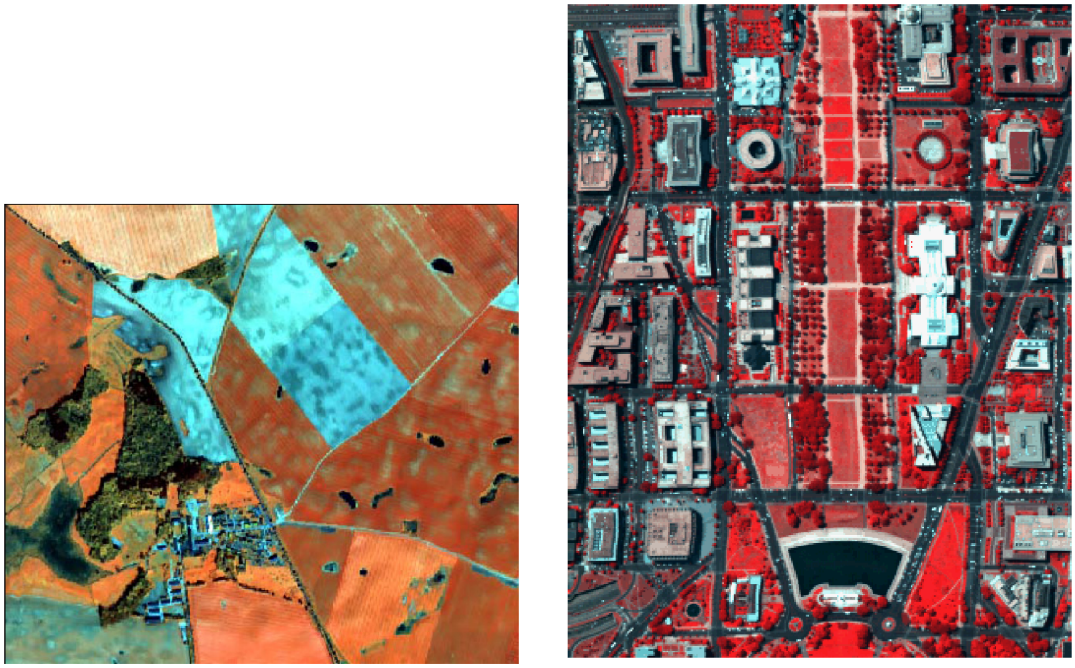
ProSpecTIR image: The ProSpecTIR airborne hyperspectral image was acquired over the City of Reno, USA on 13 September, 2006. This image consists of 356 spectral bands in the spectral range 0.39 to 2.45 μm with spatial resolution of 1 m. The dominant land use categories in the image are trees, water, bare soil, asphalt, built up, shadows, and vehicles.

(a) (b)



(c)

Figure 3.1: False color composites of the hyperspectral images used (a) ROSIS-University (b) ProSpecTIR (c) ROSIS-City of Pavia.



(a) (b)

Figure 3.2: False color composites of the hyperspectral images used (a) HyMAP (b) HYDICE.

ROSIS-City of Pavia image: ROSIS airborne hyperspectral sensor was acquired on 8 July, 2002 over the City of Pavia, Italy in the framework of the HySens Project managed by DLR (German Aerospace Agency) (Fauvel et al., 2009). The image has 102 spectral bands in the spectral range 0.43 to 0.86 μm with spatial resolution of 1.3 m. This image consists of ten land cover classes namely, water, trees, asphalt, meadow, self building blocks, tiles, bare soil, bitumen, shadows and built up.

HYDICE image: This airborne hyperspectral image was acquired over a mall in Washington DC by the HYDICE hyperspectral image sensor on 23 August, 1995. A total of 191 bands were collected in the spectral range 0.4 to 2.4 μm. The spatial resolution of the image is 2 m. This image consists of seven land cover classes namely, water, road, grass, trees, roof, path, and shadow.

The water absorption bands were removed in all the hyperspectral images and the number of bands mentioned above are the total number of bands available after removal of water the absorption bands.

3.3 Methodology

In order to evaluate the impact of various dimensionality reduction methods on the performance of MCS, we applied a diverse group of dimensionality reduction methods namely, PCA, MNF, ICA, DWT-DR and OBS. These methods are different from one another in their capability to retain spectral integrity, noise segregation, and rescaling of transformed data relative to original data representation. These dimensionality reduction methods are given as the input data source to the MCS. An MCS was created using seven linear classifiers as the base classifiers. Diversity among the classifiers was maintained by selecting the classifiers which belong to various groups of classifiers. These classifiers can be categorized broadly into three categories namely, spectral matching based methods, covariance modelling and subspace modelling based methods. The spectral matching based methods consist of Normalized Euclidean distance (NED) (Robila and Gershman, 2005), Spectral angle mapper (SAM) (Kruse et al., 1993), and Spectral similarity measure (SSM) (Granahan and Sweet, 2001). The covariance modelling methods consist of Matched filter (MF), and Adaptive coherence estimation (ACE) (Manolakis et al., 2003). The subspace modelling methods consist of Orthogonal subspace projection (OSP) (Harsanyi and Chang, 1994), and Target constraint interference minimized filter (TCIMF) (Ren and Chang, 2000).

In the following we briefly recall about the dimensionality reduction methods and classifiers used in this chapter. Detailed description of the below methods are presented in chapter 2.

3.3.1 Dimensionality reduction methods

PCA is an efficient linear transformation method extensively used for dimensionality reduction of multispectral and hyperspectral imagery. Eigenvectors of covariance matrix are the basis for PCA transformation. The required number of components of PCA transformation is decided based on the resulting eigenvalues. MNF has been especially used for dimensionality reduction of hyperspectral image because of its ability to increase the signal to noise ratio in the extracted components. It is essentially a two cascaded principal component analysis. The first transformation is based on the estimated noise covariance of image which results in unit noise variance and no band to

band correlation (Green et al., 1988; Amato et al., 2009). Then PCA is applied on the whitened data. When a target is characterized by relatively smaller land cover extend, it may not able to be captured by second-order statistics (variance) and may be buried and discarded in lower order components as noise (Cheriyadat and Bruce, 2003; Wang and Chang, 2006). ICA captures both the second and higher-order statistics. The hyperspectral image is projected on the basis that is maximally statistically independent and retains the signal components of targets relatively with lesser area extent (Villa et al., 2009). ICA was applied on the hyperspectral images and the numbers of components were decided based on the eigenvalues. As it is established that class separable information could be buried in the lower order components, we considered lower order components up to a certain number of components for both PCA and ICA (10 components for all the hyperspectral images) in this experiment. However for the MNF, eigenvalue threshold differs from image to image and it was decided based on the number of components which produce optimal classification accuracy[1]. Thus for the MNF, we selected 10 components for the ROSIS images and 15 components each for the ProSpecTIR and HYDICE images.

The statistical dimensionality reduction methods PCA, MNF, and ICA project data on to a new coordinate system by maximizing statistical measures. In contrast to that, the information content in hyperspectral image may not always coincide with such projections (Bruce et al., 2002). It has been shown that the DWT-DR produce better or comparable classification accuracy with PCA (Kaewpijit et al., 2003). Each pixel in the hyperspectral image was decomposed by the wavelet transform using the Daubechies filter (Bruce et al., 2002). The high frequency information (outliers) were discarded and the approximation coefficients were reconstructed using inverse discrete wavelet transform. The required level (L) of decomposition was achieved based on the acceptable correlation between the decomposed signal and the original signal. A correlation coefficient of 0.98 was chosen as the threshold thus resulting in a 3 level (L) decomposition. Then all the pixels in hyperspectral image were decomposed into level L and the approximation coefficients were used for further analysis.

Optimal band section is another dimensionality reduction method in which the dimensionality of the image is reduced manually by selecting a subset of bands based on

[1] training samples classification accuracy.

expert prior knowledge on the land cover classes. Based on literature and the nature of land cover categories found in the images, we identified 10 spectral bands (560 nm, 630 nm, 671 nm, 681 nm, 763 nm, 824 nm, 844 nm, 1410 nm 1800 nm, and 2200 nm) for the HyMAP image (Herold et al., 2003; Thenkabail et al., 2004; Rao et al., 2007), 15 bands for the ROSIS-University and ROSIS-City of Pavia images, (430 nm, 446 nm, 474 nm, 538 nm, 560 nm, 580 nm, 630 nm, 671 nm, 686 nm, 726 nm, 763 nm, 782 nm, 806 nm, 824 nm, 838 nm) and 15 bands for the ProSpecTIR and HYDICE images (400 nm, 560 nm, 630 nm, 671 nm, 681 nm, 763 nm, 844 nm, 1106 nm, 1410 nm, 1560 nm, 1800 nm, 2180 nm, 2330 nm) (Herold et al., 2003; Thenkabail et al., 2004; Rao et al., 2007; Marpu et al., 2009) for classification by the MCS.

3.3.2 Base classifiers

The performance of MCS depends on the base classifiers' ability to commit the complimentary errors. Diversity among the base classifiers was maintained by selecting the classifiers which belong to various groups of classifiers namely, spectral matching, covariance modelling and subspace modelling methods.

The spectral matching based methods consist of normalized Euclidean distance classifier (NED), spectral angle mapper (SAM), and spectral similarity measure (SSM). These classifiers differ in their ability to exploit spectral information in the hyperspectral image. For instance, the NED classifier captures spectral brightness, whereas the SAM is insensitive the illumination changes. On the other hand, the SSM captures both spectral brightness and spectral angle differences.

The covariance modelling methods consist of matched filter (MF) and adaptive coherence estimator (ACE). The covariance based methods are modelled as the binary classification problem with the hypothesis, whether the desired class is present or not. The MF is modelled as the noise free model, whereas ACE models accounts for the background information in the desired class target with different variance (Manolakis and Shaw, 2002). When the whitening transformation is applied then the MF uses the distance threshold and the ACE uses the angle threshold. The covariance matrix of the MF and ACE are computed from the training samples of all the classes.

The subspace modelling methods consist of orthogonal subspace projection (OSP)

and target constraint minimized interference minimized filter (TCIMF). The subspace based model methods assume that the image pixel is a linear combination of finite set of class signatures present in the image. The OSP projects image pixel into a subspace orthogonal to undesired class signatures and then performs the matching with the desired class signature. TCIMF models both the desired and undesired class signatures as the subspace model, and the finds the linear filter to suppress the undesired class signatures and interferences.

3.3.3 MCS formultation

The schematic diagram of the methodology adopted is shown in Figure 3.3. In order to assess the influence of dimensionality reduction methods on the MCS from a classification performance perspective, the hyperspectral image were transformed to the lower dimensional subspaces by five dimensionality reduction methods and each transformed image was given as the input data source to the MCS. This resulted in five different

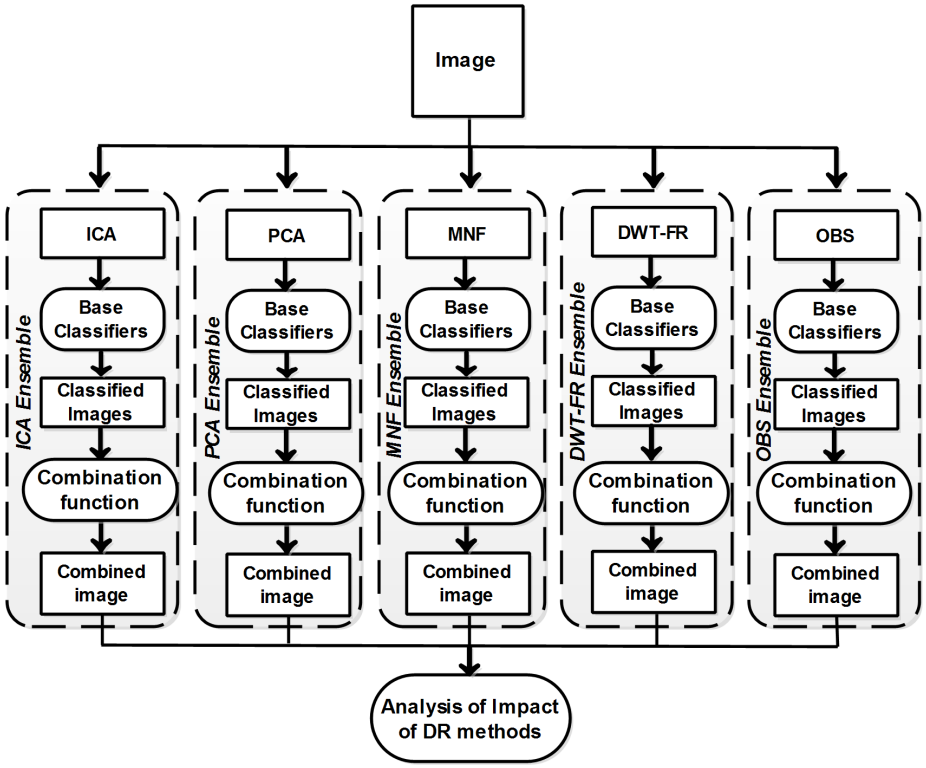


Figure 3.3: Flow chart of the experimental design to study the impact of dimensionality reduction methods on the MCS.

MCS or five ensembles. Each we name on the basis of the dimensionality reduc-

tion methods used in their creation: ICA-ensemble, PCA-ensemble, MNF-ensemble, DWTDR-ensemble, and OBS-ensemble. The classifiers in each ensemble produce decision function values independently. These intermediate decision function values were combined by various combination schemes to produce the final classified image. The combination schemes consist of six non-trainable combination functions such as majority voting rule, maximum rule, minimum rule, sum rule, product rule and median rule (Kittler, 1998; Fumera and Roli, 2005). In addition, the MCS was allowed to produce classified image for each pair of the classifier-dimensionality reduction method in the MCS. These independent classified images were considered for benchmarking the performance of the combined MCS output classified images. Statistical dispersion measures, namely Correlation of variation (CV), Relative mean difference (RMD), Quartile coefficient of dispersion (QCD) are used to assess the variability of the MCS with different dimensionality reduction methods.

The classification performance of the MCS and base classifiers was assessed using independent testing samples created from the ground truth reference map. Since the considered classifiers in the MCS are linear, the classifiers training involves in computing mean class signature for each class and common covariance matrix. The overall accuracy (OA), kappa coefficient (KC), producer accuracy (PA), and user accuracy (UA) were used as the evaluation measures in the experiment.

3.4 Experimental Results and Analysis

In order to achieve the goals of this chapter, we have conducted two set of experiments. In the first experiment, we analyzed (1) the sensitivity of individual classifiers to different dimensionality reduction methods, and (2) the effect of dimensionality reduction methods on the base classifiers' overall performance. In the second experiment, we assessed the impact of different dimensionality reduction methods on the classification performance of the MCS.

3.4.1 Analysis of base classifiers relative to different dimensionality reduction methods

Figure 3.4 shows the behaviour of the classification accuracies with different dimensionality reduction methods for each of the classifiers considered in the MCS. From the Figure 3.4, we can derive some inference on the effect of different dimensionality reduction methods on different classifiers. First of all, the classification accuracies are variable across different classifiers and different dimensionality reduction methods. This phenomenon highlights the sensitivity of the classifiers to different input data sources. The level of variability of the classification accuracies is a function of the dimensionality reduction method, classifier and hyperspectral images. For example, with PCA dimensionality reduction method there exists 35%, 61%, 29%, 63% and 23% differences between maximum and minimum classification accuracy for five hyperspectral images. Similar differences are also evident with other dimensionality reduction methods. Further it is interesting to see that for each of the hyperspectral images the highest classification accuracies are obtained with different pairs of classifiers and dimensionality reduction methods. However among the classifiers, NED (HyMAP, ROSIS-City of Pavia), SSM (ROSIS-University, ProSpecTIR) has repeated twice as the best classifier and SAM (HYDICE) occurred once as the best classifier. Similarly, among the dimensionality reduction methods, MNF has resulted as the best dimensionality reduction method with three images. This observation shows the domination of a particular category of classifiers and dimensionality reduction methods.

When the classification results are analyzed within each image, different dimensionality reduction methods prefers different classifiers to obtain the best classification results. For example, with HyMAP image, NED is resulted as the best classifier for ICA and MNF; MF for PCA; ACE for DWTDR and SSM for OBS dimensionality reduction methods. Moreover, there exist significant accuracy differences among the best classifiers relative to each dimensionality reduction method for a given hyperspectral image. This indicates the relative importance of the input data source and classifier relationship for better classification performance. Further, this also highlights that the maximum achievable classification accuracy for each of the hyperspectral image is limited due to its overall information content characteristics.

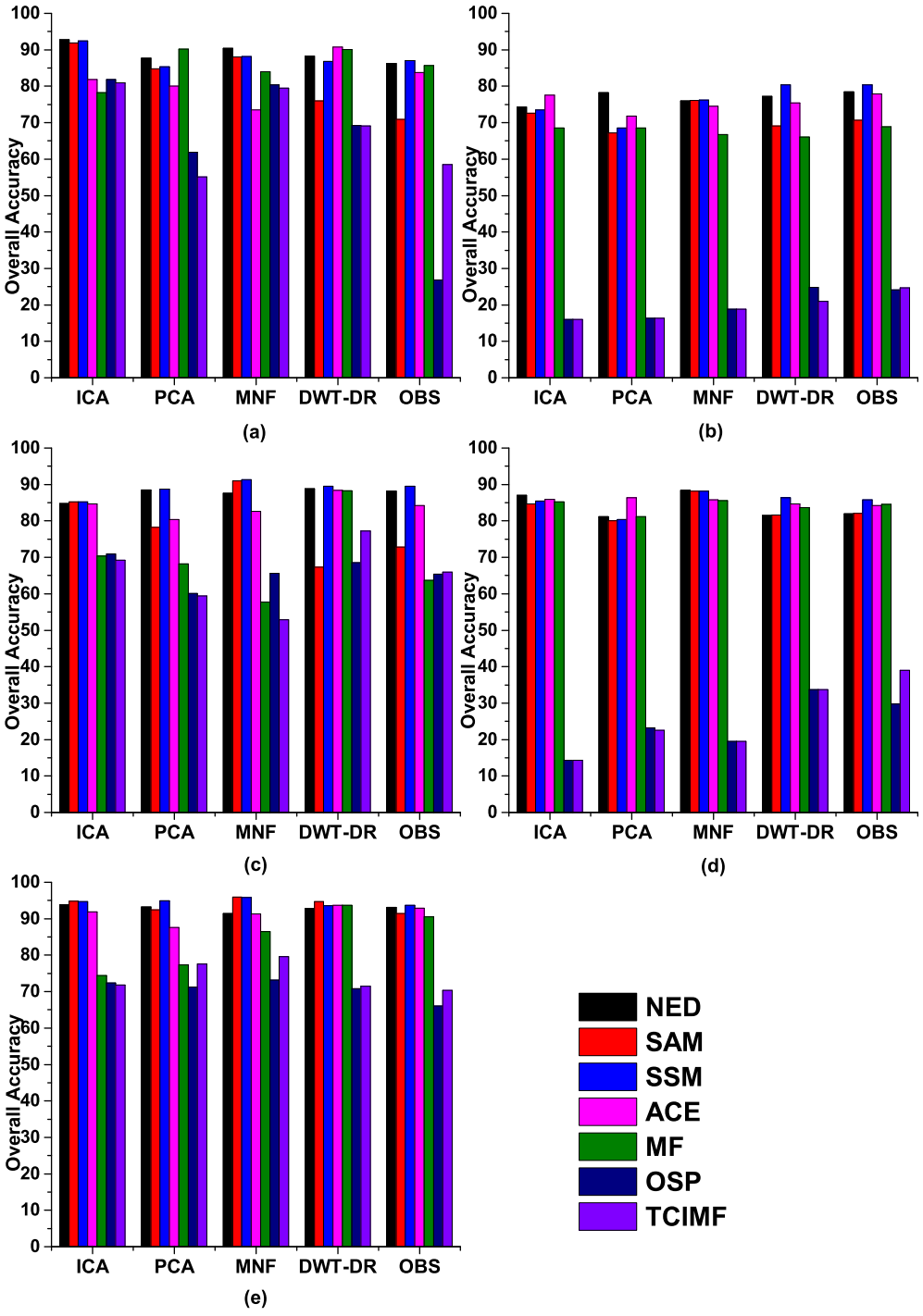


Figure 3.4: Overall accuracy of classifiers relative to different dimensionality reduction methods of (a) HyMAP (b) ROSIS-University (c) ProSpecTIR (d) ROSIS-City of Pavia and (e) HYDICE hyperspectral images.

3.4.1.1 Assessing the effect of dimensionality reduction methods on the overall base classifiers performance and individual classifiers performance

In order to do this, we conducted statistical dispersion analysis based on the overall accuracy of the classifiers. In particular, we analyzed (a) which dimensionality reduction are most and least sensitive to the base classifiers of the MCS and (b) which individual classifier is most and least sensitive to change in dimensionality reduction methods. The statistical dispersion measures correlation of variation (CV), relative mean difference (RMD), quartile coefficient of dispersion (QCD) are used to measure the spread and variability of the data. The statistical dispersion is zero when all the values are same and increases as the data becomes diverse. Hence it is an effective measure to analysis the impact of the dimensionality reductions on the classifiers performance. In this experiment, we have computed CV, RMD, and QCD to analysis the impact of dimensionality reduction methods. Apart from this, the average accuracy was also calculated relative to each dimensionality reduction method and relative to each classifier to characterize the overall average performance of the dimensionality reduction methods and classifiers. The average accuracy relative to each dimensionality reduction methods was computed by averaging the overall accuracy of all the classifiers (base classifiers) in the MCS. Similarly, the average accuracy relative to each classifier was computed by averaging each classifier's accuracy resulting from all the dimensionality reduction methods.

Sensitivity analysis of dimensionality reduction methods:

Different dimensionality reduction methods were given as input to the base classifiers of the MCS and the statistical dispersion measure were computed based on performance of the base classifiers. If all the classifiers perform in a similar manner with respect to a particular dimensionality reduction method, then statistical dispersion measures will be minimum.

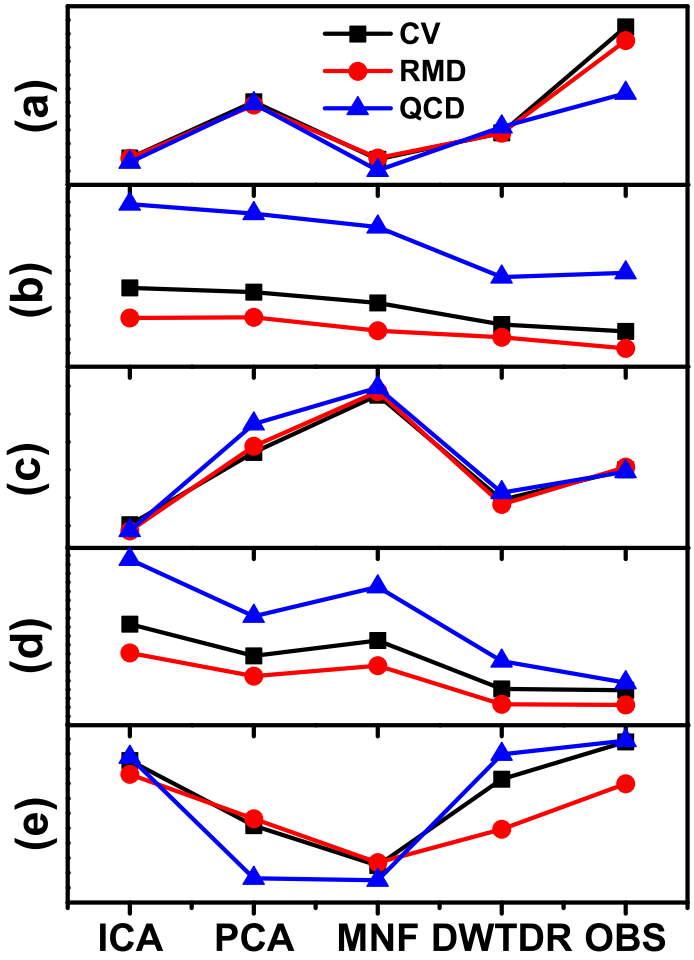


Figure 3.5: Statistical dispersion measures (CV, RMD, and QCD) calculated from the base classifiers performance relative to each dimensionality reduction methods in the MCS for (a) HyMAP (b) ROSIS-University (c) ProSpecTIR (d) ROSIS-City of Pavia and (e) HYDICE hyperspectral images.

Figure 3.5 shows behaviour of the three statistical dispersion measures relative to each dimensionality reduction methods for different sets of classifiers. From the Figure 3.5, we can infer that different dimensionality reduction methods have different levels of impact on the classifiers performance; and the variability of dimensionality reduction method differs from image to image. For the HyMAP, and HYDICE images, the base classifiers are most sensitive with OBS dimensionaltiy reduction method; for the ROSIS University, ROSIS City of Pavia images the classifiers are most sensitive with ICA dimensionality reduction method and MNF for the ProSpecTIR image. It is noticeable

that the impact of the PCA dimensionality reduction method is moderate for all the hyperspectral images; and the similar type of dimensionality reduction methods are resulted as the most and least sensitive dimensionality reduction methods for the similar type of hyperspectral images[2]. For certain hyperspectral images, the high statistical dispersion values were observed compared to other images; this is because of the worst performance of certain classifiers. The average accuracy of the base classifier shown in Figure 3.6 better characterizes the dimensionality reduction method to be used on the hyperspectral images irrespective of the classifiers.

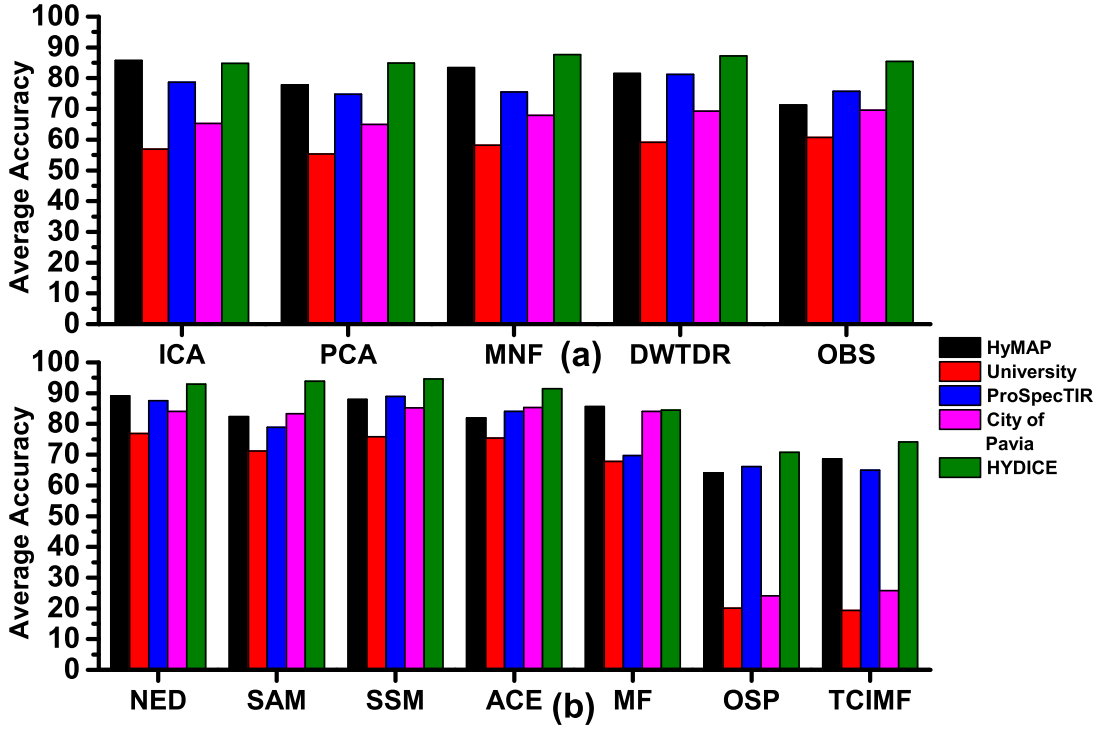


Figure 3.6: (a) Average accuracy of dimensionality reduction methods: The overall accuracies of all the classifiers are averaged relative to each dimensionality reduction methods. (b) Average accuracy of classifiers: overall accuracies of a particular classifier was averaged over all the dimensionality reduction methods.

As observed with statistical dispersion measures, the best average accuracies are obtained with different dimensionality reduction methods for different images. For example, higher average accuracies are obtained with OBS dimensionality reduction method for ROSIS sensor datasets, and ICA, DWTDR, MNF for the HyMAP, ProSpecTIR and HYDICE hyperspectral images respectively. From the Figure 3.5 and 3.6 we can infer

[2]Images containing similar types of land cover classes.

that similar type of dimensionality reduction methods are resulted as the best choice for the ROSIS University and City of Pavia images. This indicates that the information extraction capability of the dimensionality reduction methods is a function of underlying land cover information and sensors' specifications.

Sensitivity analysis of Classifiers:

Figure 3.7 shows the statistical dispersion measures relative to each classifier across different dimensionality reduction methods for the five hyperspectral images. The average accuracy of the classifiers is shown in Figure 3.6 (b). Figure 3.7 reveals that a high statistical dispersion measure is obtained with the subspace modelling classifiers for most of the images. This is due to the poor performance of subspace modelling classifiers for certain dimensionality reduction methods (see Figure 3.4).

Apart from the subspace modelling classifiers SAM, SSM, and NED are more sensitive to different dimensionality reduction methods for HyMAP, ROSIS University, and ROSIS City of Pavia hyperspectral images respectively. For the remaining two images MF is resulted as the most sensitive classifier to different dimensionality reduction methods. There are a few classifiers which are least sensitive to the different dimensionality reduction methods for the four hyperspectral images (NED, SSM with HyMAP, ProSpecTIR, and HYDICE images and NED, MF with ROSIS-University image). For the ROSIS-City of Pavia image, ACE is resulted as the least sensitive classifier. From the Figure 3.6 (b), it is interesting to observe that the classifiers which are least sensitive to different dimensionality reduction methods have offered higher average accuracies. For example, higher average accuracy is obtained with NED classifier for the HyMAP, ROSIS-University image and the similar conclusion holds for the remaining hyperspectral images as well.

3.4.2 Impact of dimensionality reduction methods on the performance of the MCS

The base classifiers in the MCS were combined relative to each dimensionality reduction method, thus resulting with the five ensembles for each image and it is named as ICA ensemble, PCA ensemble, MNF ensemble, DWTDR ensemble and OBS ensemble.

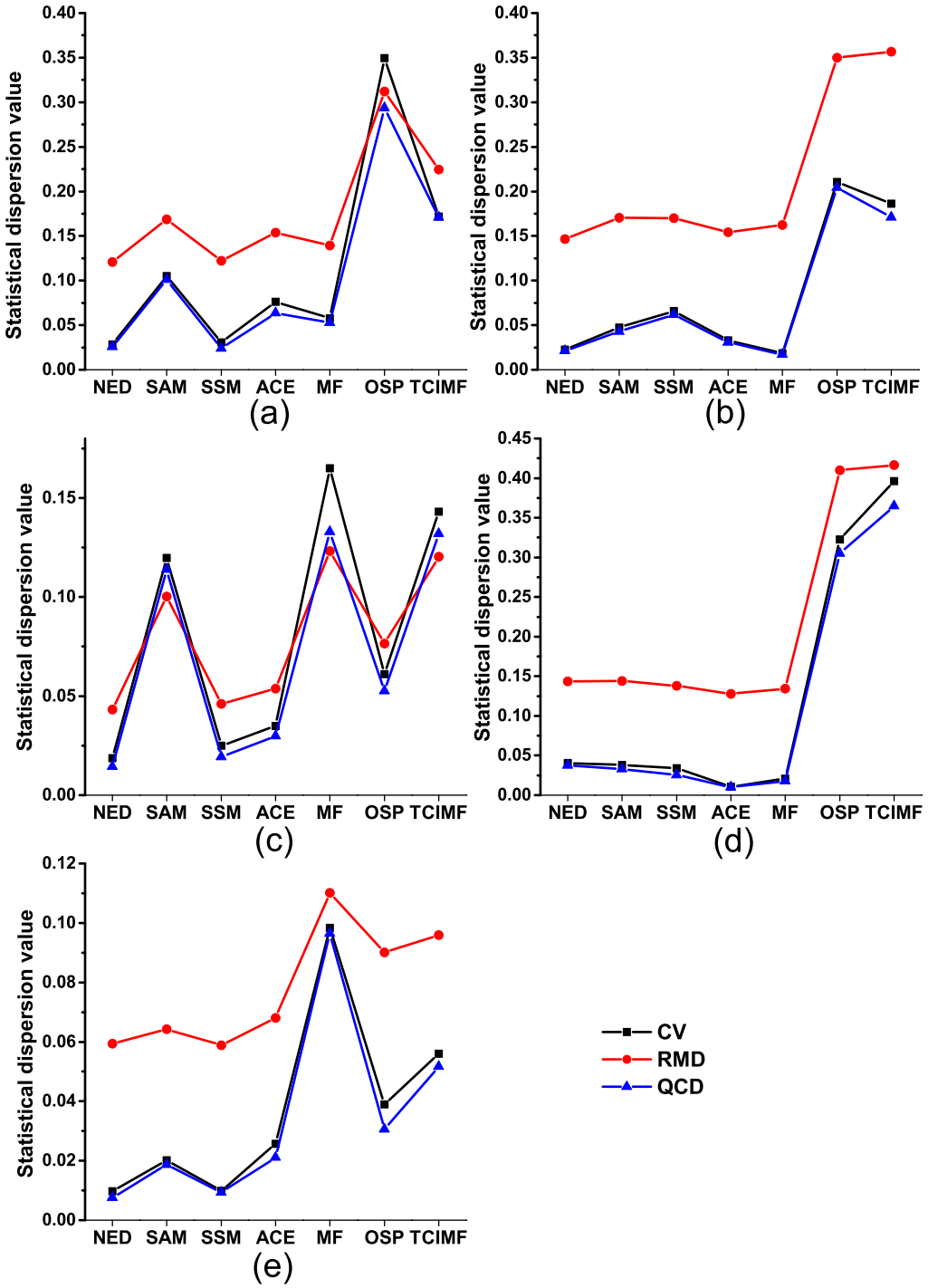


Figure 3.7: Statistical dispersion measures (CV, RMD, and QCD) of the individual classifiers over all the dimensionality reduction methods (a) HyMAP (b) ROSIS-University (c) ProSpecTIR (d) ROSIS-City of Pavia and (e) HYDICE.

Figure 3.8 shows the classification accuracies resulted from the combination functions of the MCS. It can be observed that there is a significant to marginal increase in the classification accuracy over the single best classifier with respect to each dimensionality reduction method ensemble. The magnitudes of the accuracy improvement achieved by the MCS are variable across different dimensionality reduction method ensembles within each hyperspectral image. At this point it is worth to note that our aim is to analyze the impact of dimensionality reduction method on the performance (magnitude) improvement of the MCS over the single best classifier (relative to each dimensionality reduction method). So we focus our analysis on the magnitude of the improvement of MCS over the single best classifier and not on the overall classification accuracy[3]. There are upto 4-6% differences in accuracy improvement among the dimensionality reduction method ensembles for the HyMAP, ROSIS-University, HYDICE, ProSpecTIR hyperspectral images. These accuracy differences highlight the impact of the different dimensionality reduction methods (different input data sources to the MCS). As mentioned earlier, the different dimensionality reduction methods characterize the information content of original hyperspectral image based on different principle. Therefore understanding the proper choice of the input data sources is necessary for the hyperspectral image classification in the MCS framework. It is also important to note that the performance of the MCS is not improved with all the dimensionality reduction methods. There are one or two dimensionality reduction method ensembles for each of the hyperspectral image which decrease the classification accuracy compared to the single best classifier. This might be due to the classifiers forming the MCS not committing different types of errors with certain dimensionality reduction method.

The MCS has offered about 3-5%, 1-4%, 2%, 1-2.5%, and 1-3% magnitude improvement when compared to the single best classifier for the HyMAP, University, ProSpecTIR, City of Pavia, and HYDICE hyperspectral images respectively. Further, different dimensionality reduction methods offered best accuracy improvement for the different hyperspectral images. For example, the highest improvement is obtained with OBS for the HyMAP image, MNF for the ROSIS University and City of Pavia image, and DWTDR for the ProSpecTIR and HYDICE hyperspectral images. Thus, the mag-

[3]For example with the HyMAP image, the overall accuracy of 91.5% is obtained with OBS ensemble and 93.4% is obtained with DWTDR ensemble, but OBS ensemble has offered 5.3% improvement and DWTDR has offered 3.3% improvement compared to the single best classifier.

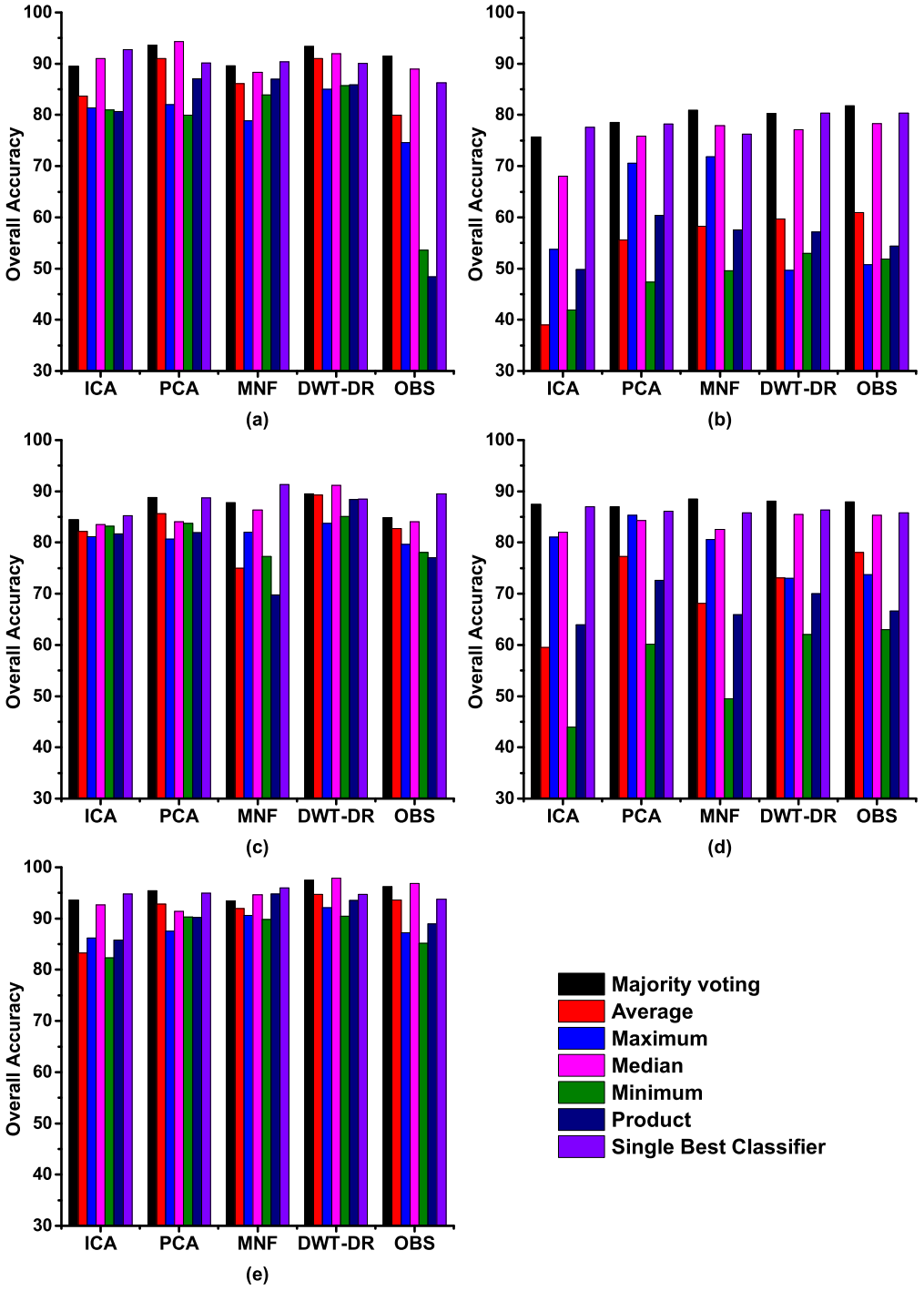


Figure 3.8: Overall accuracy of the different combination functions relative to each dimensionality reduction methods (a) HyMAP (b) ROSIS-University (c) ProSpecTIR (d) ROSIS-City of Pavia (e) HYDICE.

nitude of improvement and the performance of the dimensionality reduction methods are variable across different images. This indicates that the performance of the MCS is dynamic in nature, and the information exploitation capability of different dimensionality reduction methods is also dynamic in nature. The MCS adapting to this dynamic relationship of classifiers and dimensionality reduction methods with respect to each image may offer stable and improved performance.

3.4.2.1 Impact of combination function on the performance of the MCS

The classifiers' decision function values were combined using six non-trainable combination functions. From the Figure 3.8, it is evident that the type of combination function used has a significant bearing on the performance of the MCS. The majority voting rule improved accuracy across the hyperspectral images. However, for certain dimensionality reduction methods, the median rule improved accuracy, which is marginally higher than the improvement observed with the majority voting. For instance, in the HyMAP, ProSpecTIR image median rule has offered high accuracy improvement for one dimensionality reduction method ensemble and in the HYDICE image it is better with three dimensionality reduction method ensembles. The remaining combination functions (average, maximum, minimum, and product) provided insignificant improvement or decrement in terms of classification accuracy when compared with the single best classifier. This might be because of the presence of worse performing classifiers in the MCS. In other words, the performance of the maximum, minimum, average and product rule is sensitive to one or more worst performing classifiers' decision values whereas the majority voting and median rule are insensitive to the outliers or inaccurate classifiers' decision values. The statistical dispersion measures of the dimensionality reduction methods shown in Figure 3.5 also support this inference. The combination functions which decreased classification accuracy substantially relative to each dimensionality reduction method ensemble exhibit higher values of dispersion measures.

3.4.3 Discussion

The plethora of image classifiers available in the literature are widely varying by their performance being data and task specific. There is no single classifier which offers op-

timal performance over a range of classification tasks. Moreover, there has been great diversity observed in the type and distribution of errors made by classifiers. Recent developments in pattern recognition techniques as envisaged by MCS try to use diversity in classifiers' performance by combining the better aspects of various classifiers for classification results which are ought to be convincing and reliable. Dimensionality reduction is a common pre-requisite for any type of hyperspectral image classification task including the ensemble of classifiers popularly called as MCS. A great body of research is directed at developing of better classifier ensembles and combination function schemes (Petrakos et al., 2001; Fernandez-Redondo et al., 2004; Doan and Foody, 2007; Du et al., 2012b). The overall aim of this study was to examine the change in the classification performance of the MCS while applying various dimensionality reduction methods independently and simultaneously. Whether the diversity in the data variance captured by various dimensionality reduction methods are complimentary to each other and, if it is so, can the same be used for enhancing classification performance by MCS is an important aspect of image classification. The knowledge gained from this study can be used to determine which dimensionality reduction method has the most and least impact on the performance of the set of classifiers being considered and which classifier is affected most and least by the changing the dimensionality reduction method for different hyperspectral images.

First we analysed the impact of dimensionality reduction methods on the variability of the base classifiers' performance. The analysis of statistical dispersion measures indicates that the base classifiers' performance is sensitive to changing dimensionality reduction methods or different input data sources within each image. For different hyperspectral images, different dimensionality reduction methods are resulted as the most and least sensitive dimensionality reduction methods, indicating the advantage of dimensionality reduction methods to create the variability among the classifiers in the MCS. Similar observation also holds when the average accuracy is used to interpret the combined performance of the base classifiers relative to each dimensionality reduction method.

Analyses of the impact of the dimensionality reduction methods relative to each classifier's performance indicate that different classifiers have different levels of variability in the performance within and across the images. This observation indicates the need of using image specific methods for betterment of the hyperspectral image classi-

fication. When the performance of the single best classifier is analysed, it interesting to see that the optimal classifier for each dimensionality reduction method is different and their classification accuracies are also variable. With this dynamic nature of the performance of the dimensionality reduction methods and classifiers, it is necessary to have a prior knowledge on the relation between the input data sources and classifier to obtain the optimal classification accuracy. However obtaining this data specific knowledge is tedious task and this obtained knowledge may not be reliable.

From the above analysis, it can be concluded that the different dimensionality reduction methods have the capability to characterize the image information in different perspectives and create variability in the performance among different classifiers. With this differential information of dimensionality reduction methods, it necessary to understand the impact of the dimensionality reduction methods on the performance of the MCS for the hyperspectral image classification. The classification results obtained by combining the classifiers' decision function values point out significant differences in the magnitude of accuracy increase with changing dimensionality reduction method. There are 3-5% accuracy differences over the single best classifier when the percentage of accuracy improvement is considered and it is about 2-4% accuracy differences when the best overall classification accuracy is considered among the dimensionality reduction methods. The observed significant increase in the classification accuracy of hyperspectral images supports the theoretical advantages of using MCS for hyperspectral image classification. However, the accuracy differences are highly subjective to the dimensionality reduction methods and hyperspectral images. Contrary to the improvement in accuracy for some images, there is a decrease in accuracy for some other images and again the magnitude of decrease in accuracy depends upon the dimensionality reduction method and the hyperspectral image. This observation indicates that selecting a single dimensionality reduction method arbitrarily for the MCS is not an optimal choice as it is uncertain whether the performance of the MCS will be improved or not. This work also presented an objective way of determining the optimal dimensionality reduction method which can be used with the MCS to get accurate results for the specific hyperspectral image. The dimensionality reduction with moderate statistical dispersion value can be considered as the optimal dimensionality reduction for the MCS.

While different classifiers and different dimensionality reduction methods are considered in this experiment, it is necessary to analysis the dynamics of pairs of classifiers

and dimensionality reduction methods with reference to the information classes in the hyperspectral image. Further the potential of applying multiple dimensionality reduction methods within a single MCS for achieving consistent performance of the MCS needs to be explored. In next chapter we address the aforementioned limitations of this study for effective hyperspectral image classification using MCS.

3.5 Chapter Conclusions

In this chapter the experimental analysis on the impact of dimensionality reduction methods on the performance of MCS is presented for hyperspectral image classification. The results indicate that efficient classification of hyperspectral images by MCS requires an understanding of the suitability of various dimensionality reduction methods on the ability of the classifiers to create diversity in the results. In particular, our analysis points out significant variability in the performance of the base classifiers with different dimensionality reduction methods leading to substantial changes in the accuracy improvement of the MCS. This study also suggests that dimensionality reduction methods can be used to create diversity among the base classifiers. Not all the classifiers and dimensionality reduction methods are able to exploit the information available with the hyperspectral imagery. In summary we can conclude that

1. choice of dimensionality reduction methods for creating variability in the base classifiers (MCS) performance has to be done according to the underlying nature of hyperspectral images
2. selection of the optimal classifier for different dimensionality reduction methods requires understanding of the classifiers' performance according to the underlying hyperspectral image
3. the performance improvement of the MCS is adaptive to the dimensionality reduction method and the hyperspectral images.

Finally we provided an intuitive way of determining the optimal dimensionality reduction method for the MCS classification based on the statistical dispersion measures. Examination of the overall accuracy of image classifications relative to the apparent classifier-dimensionality reduction method relationships suggests the need for adaptive classifiers and dimensionality reduction methods in the MCS for efficient hyperspectral image classification.

CHAPTER 4

ASSESMENT OF THE RELATIONSHIP BEWTEEN INFORMATION CLASS, CLASSIFIER AND DIMENSIONALITY REDUCTION METHOD FOR HYPERSPECTRAL IMAGE CLASSIFICATION BY MULTIPLE CLASSIFIER SYSTEM

Prelude: This chapter presents the experimental analysis of the relationship between classifiers and dimensionality reduction methods at overall image level as well as at land cover category level for hyperspectral image classification in the framework of multiple classifier system. In order to understand whether the observed relationship is variable with respect to resolution of the image, the classification experiments were also carried out on the synthetic hyperspectral images generated from the original high resolution hyperspectral images used in this study. The potential of deploying multiple dimensionality reduction methods in a single MCS architecture was analysed by combining the better performing classifiers in the MCS. Further the MCS results were compared with the support vector machine based classification.

4.1 Introduction

Dimensionality reduction is a common pre-processing step in the supervised classification of hyperspectral images for various applications such as land cover mapping (Melesse et al., 2007; Ghiyamat and Shafri, 2010). Data generalization schema of various dimensionality reduction methods differs by their ability to retain spectral integrity

This chapter is published in *Advances in Space Research,* Vol. 53, pp. 1720-1734, June 2014, with the title: " Assessment of the impact of dimensionality reduction methods on information classes and classifiers for hyperspectral image classification by multiple classifier system", . Authors: Bharath Bhushan Damaodaran, Rama Rao Nidamanuri

and residual spectral information required for discrimination of materials. Often, critical information needed for class separation in a hyperspectral image is lost as noise by the application of dimensionality reduction methods (Cheriyadat and Bruce, 2003; Farrell and Mersereau, 2005; Prasad and Bruce, 2007, 2008a). Thus, the use of dimensionality reduction methods without knowledge on the types of land covers available in image may lead to poor results. The performance of the classifiers is variable across the different images and applications, and there is no single best classifier which can be applied across different images and land cover categories (Fabio et al., 1997). Hence the identification of the classifier which is optimal to the application and data at hand is thus a recurring task in every image classification task and it is a very important problem. Numerous studies are available in the literature dealing with selection and comparison of classifiers for various applications of multispectral and hyperspectral image (Smits et al., 1999; Rogan et al., 2002; Erbek et al., 2004; Joshi et al., 2006; Zalazar, 2006; Lu et al., 2008; Al-Ahmadi and Hames, 2009; Brenning, 2010; Moran, 2010; Lu et al., 2011; Szuster et al., 2011; Srivastava et al., 2012; Lu et al., 2012; Xu et al., 2014). Similarly, a number of studies have reported on the selection of dimension reduction methods for various land cover classification scenario (Barker, 1997; Deogun et al., 1998; Dutra, 1999; Linders, 2000; Bruzzone and Serpico, 2000; Colás et al., 2001; Miao et al., 2007; Lu et al., 2007; Lu and Weng, 2007; Chen and Qian, 2008; Clemmensen et al., 2010; Duro et al., 2012). For suggesting optimal classifiers and dimensionality reduction methods for land cover classification, however, most of the studies have used a single classifier for comparing the performance of various dimensionality reduction methods or a single dimensionality reduction method for comparing the performance of various classifiers. Having the theoretical framework to combine the differential performances of various dimensionality reduction methods and classifiers enhances the robustness and reliability of hyperspectral image classification.

Multiple classifier system (MCS) provides the conceptual framework to incorporate various input data sources and classifiers in the classification process (Fernandez-Redondo et al., 2004; Doan and Foody, 2007; Pal, 2008; Ceamanos et al., 2010; Yang et al., 2010b). The apparent one-to-one relationship between classifier and dimensionality reduction method found in multispectral image classification is seldom evident in hyperspectral image classification. Our extensive literature survey reveals the lack of understanding on the suitability of classifiers and dimension reduction methods for

hyperspectral image classification by MCS for a range of land cover categories. The objective of this research is to assess the impact of the relationship between classifiers and dimensionality reduction methods as well as among the information class, classifier, and dimensionality reduction method on the hyperspectral image classification for land cover classification by MCS. The understanding gained from this research is valuable for (a) identification of adaptable classifiers for a given dimensionality reduction method, and (b) identification of the information class dependent sets of classifiers and dimensionality reduction methods for hyperspectral image classification by MCS.

Apart from that, the spectral and spatial resolutions have a vital role in remote sensing applications. The advancement in the sensor technology enables to capture the images in different spectral and spatial resolutions. Several studies have been conducted to understand the optimal spectral and spatial requirements for hyperspectral remote sensing applications like leaf area index (LAI), forest fire analysis, classification etc (Lee, 2001; Yanchen Bo et al., 2005; Sprintsin et al., 2007; Dalponte et al., 2009; Matheson and Dennison, 2012; Thorp et al., 2013). These studies conclude that the image's spatial and spectral resolution has to be selected based on underlying applications, and for certain applications the medium resolution image is sufficient. Therefore, it is necessary to understand whether the observations made with the high resolution image can be directly transferred to the medium or low resolution images. In particular, whether the empirical relationship observed between classifiers and dimensionality reduction methods, and between class, classifier and dimensionality reduction methods is similar for both high resolution (original) image and medium resolution image needs to be evaluated. If the observations are valid then it will minimize the time consuming task to identify the optimal classifiers and dimensionality reduction methods. In order to address this, the high resolution images or original hyperspectral image were downscaled to medium resolution images in both spatial and spectral domain to generate synthetic hyperspectral images. Then classification experiments were performed on the synthetic hyperspectral images.

The novel contributions of this chapter are summarized as i): establishes the existence of empirical relationship between classifier and dimensionality reduction method as well as information class, classifier and dimensionality methods ii): addresses the potential of deploying multiple dimensionality reduction methods in the MCS to offer enhanced classification performance iii): analysis of empirical relationship between land

cover classes, classifiers and dimensionality reduction methods from the classification accuracy perspective for the spatially and spectrally downscaled synthetic hyperspectral images.

The rest of the chapter is organized into different sections as follows: Section 4.2 presents the materials and methods. In the section 4.3 the results and analyses obtained from the original hyperspectral images are presented. In the sections 4.4, and 4.5 the results obtained from spatially and spectrally downscaled synthetic hyperspectral images are presented. The results are discussed in the section 4.6 and the chapter conclusions are presented in the section 4.7.

4.2 Methodology

We generated a MCS using five dimensionality reduction methods and seven classifiers. The dimensionality reduction methods (PCA, ICA, MNF, DWT-DR and OBS) were used to generate different input data sources to the MCS. These methods have been widely used in hyperspectral image classification proving their effectiveness in many hyperspectral remote sensing applications. In order to investigate the choice of optimal classifier for each dimensionality reduction method, the seven classifiers are selected from three different categories namely spectral matching methods, covariance based methods and subspace based methods. The motivation for the selection of these classifiers is 1) the classifiers have a linear decision boundary, 2) low complexity in parameter estimation and fast training time, and 3) insensitive to small sample size problem. The classifiers with above properties are desirable for any application tasks, and the outcome of this study further adds significance in choosing optimal classifier without pre-image classification exercise. In the following we describe briefly the classifiers and dimensionality reduction methods used.

4.2.1 Selection of dimensionality reduction methods

Dimensionality reduction is an essential pre-processing technique to reduce dimensionality of hyperspectral image in order to minimize the correlation between successive bands of hyperspectral image. The resulting components of PCA and MNF transfor-

mation are orthogonal, where as ICA results in independent components. In addition to orthogonal components, MNF has good signal to noise ratio components. In order to minimize the loss of information, the number of components (10 components for all the hyperspectral images considered) which represent data variance of 99% are considered for further processing with PCA and ICA method. For the MNF method, number of components (10 components for the ROSIS image and 15 components each for the ProSpecTIR and HYDICE images) which produce optimal classification accuracy[1] are selected for further processing. The DWT-DR is the mathematical transformation method, where the basis is fixed for all the hyperspectral images unlike the statistical transformation method (PCA, ICA and MNF). The selection of components in DWT-DR depends upon the level of decomposition. An correlation coefficient of 0.98 was chosen as the threshold between the reconstructed signal (after discarding the high frequency components) and the original signal. This resulted in three level decomposition for all the hyperspectral images.

OBS is another dimensionality reduction method in which the dimensionality of the image is reduced manually by selecting a subset of bands based on expert's prior knowledge on land cover classes. Based on literature and the nature of land cover categories found in the images, we identified 10 spectral bands (560 nm, 630 nm, 671 nm, 681 nm, 763 nm, 824 nm, 844 nm, 1410 nm 1800 nm, and 2200 nm) for the HyMAP image (Herold et al., 2003; Thenkabail et al., 2004; Rao et al., 2007), 15 bands for the ROSIS-University and ROSIS-City of Pavia images, (430 nm, 446 nm, 474 nm, 538 nm, 560 nm, 580 nm, 630 nm, 671 nm, 686 nm, 726 nm, 763 nm, 782 nm, 806 nm, 824 nm, 838 nm) and 15 bands for the ProSpecTIR and HYDICE images (400 nm, 560 nm, 630 nm, 671 nm, 681 nm, 763 nm, 844 nm, 1106 nm, 1410 nm, 1560 nm, 1800 nm, 2180 nm, 2330 nm) (Herold et al., 2003; Thenkabail et al., 2004; Rao et al., 2007; Marpu et al., 2009) for classification by the MCS.

4.2.2 Selection of classifiers

The spectral matching based methods consist of normalized Euclidean distance classifier (NED), spectral angle mapper (SAM), and spectral similarity measure (SSM). These classifiers differ in their ability to exploit spectral information in the hyperspec-

[1]The assessment is based on training classification accuracy

tral image. For instance, the NED classifier captures spectral brightness, whereas the SAM is insensitive to the illumination changes. On the other hand, the SSM captures both spectral brightness and spectral angle differences.

The covariance modelling methods consists of matched filter (MF) and adaptive coherence estimator (ACE). The covariance based methods are modelled as the binary classification problem with the hypothesis, whether the desired class is present or not. The MF is modelled as the noise free model, whereas ACE models accounts for the background information in the desired class target with different variance. When the whitening transformation is applied then the MF uses the distance threshold and the ACE uses the angle threshold. The covariance matrix of the MF and ACE were computed from the training samples of all the classes.

The subspace modelling methods consist of orthogonal subspace projection (OSP) and target constraint minimized interference minimized filter (TCIMF). The subspace based model methods assumes that the image pixel is a linear combination of finite set of class signatures present in the image. The OSP projects the image pixel into a subspace orthogonal to undesired class signatures and then performs the matching with the desired class signature. TCIMF models both the desired and undesired class signatures as the subspace model, and the finds the linear filter to suppress the undesired class signatures and interferences.

4.2.3 Generation of synthetic hyperspectral images

The synthetic hyperspectral images were generated by downscaling the original high resolution hyperspectral images in both spatial and spectral domain to evaluate whether the observed best set of classifiers and dimensionality reduction methods varies when the resolution of the image is different. The spatial, and spectral characteristics of the HyMAP hyperspectral image[2] were used to generate the spatially and spectrally downscaled hyperspectral images. The cubic convolution interpolation method was used to downscale the spatial resolution of the hyperspectral images. Thus, all the spatially downscaled synthetic hyperspectral images have $5\,m$ spatial resolution. In

[2] Among all the hyperspectral images considered in this thesis, HyMAP image was the relatively low spatial and spectral resolution hyperspectral image. So HyMAP image is used as the base reference image for the medium or low spatial resolution.

order to extract training and testing samples, the corresponding ground truth map was also downscaled to $5\,m$ spatial resolution.

The spectrally downscaled synthetic hyperspectral images were generated by re-sampling the spectral bands of the original high resolution hyperspectral image based on the wavelength information of the HyMAP image. The ROSIS University and ROSIS City of Pavia hyperspectral images were not considered in this experiment, since the spectral range ($0.43 - 0.838\,\mu m$) of these two images is less than the spectral range of the HyMAP image ($0.403 - 2.48\,\mu m$). The remaining two images (ProSpecTIR, HYDICE) were downscaled about $20\,nm$ spectral resolution, and this process resulted in 128 bands for both the images.

4.2.4 Experimental design of the MCS

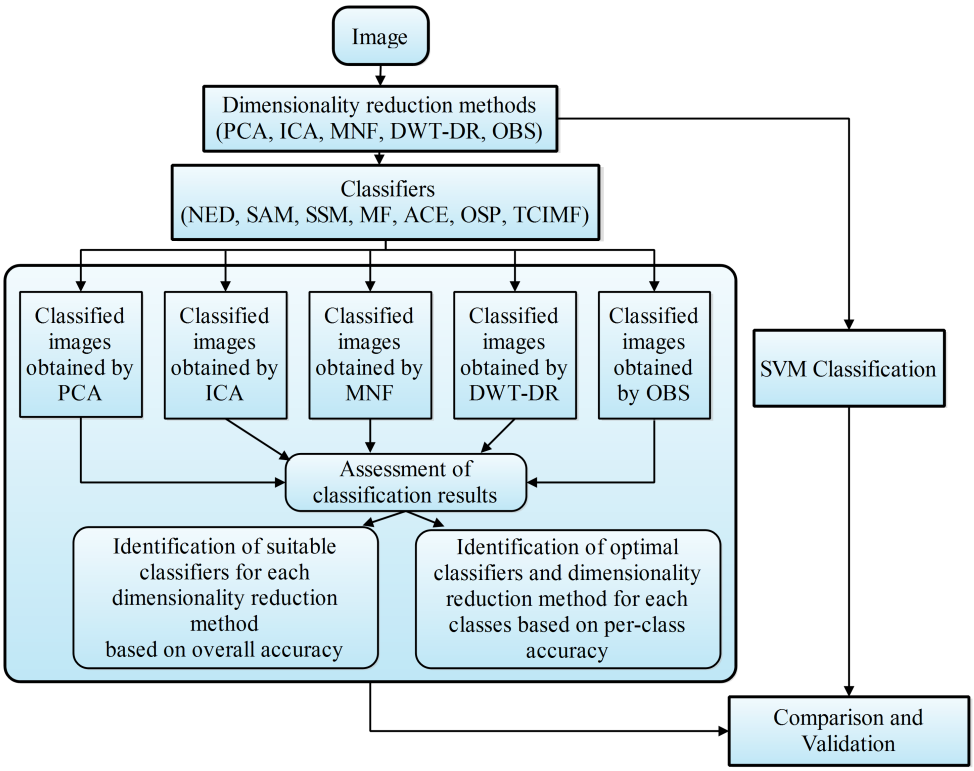


Figure 4.1: Scheme followed to assess the relationship between class, classifier and dimensionality reduction method for hyperspectral image classification.

In order to achieve the goals of this chapter two different processing schemes were adopted. The first experimental scheme was performed on the original hyperspectral

images and the second scheme was on the synthetic hyperspectral images. The outputs of the dimensionality reduction methods were given as input for supervised classification of hyperspectral images by the MCS. The sequence of steps followed for the entire classification tasks with original hyperspectral images is shown in Figure 4.1.

The MCS was programmed to generate all the intermediate classification results as separate outputs and perform classification accuracy assessment based on the testing pixels. The sets of classifiers and dimensionality reduction methods which dominate the labelling of various land cover categories in each image and subsequently the best sets of classifier(s) and dimensionality reduction method(s) for each land cover category were identified based on the overall and per-class classification accuracy, checking for repetition of the observed best sets of classifiers and dimensionality reduction methods. In addition, the pairs of classifiers and dimensionality reduction methods which offered accuracy above the acceptable threshold were combined further in the MCS to further improve the accuracy of hyperspectral image classification. The sequence of the steps followed to analysis the scale dependency of the hyperspectral images are shown in Figure 4.2.

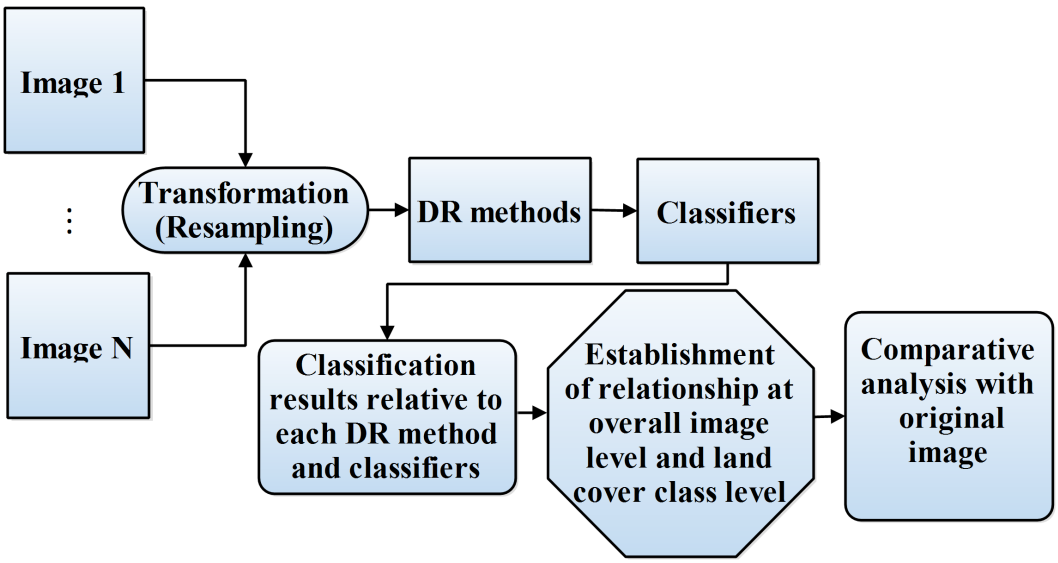


Figure 4.2: Sequence of steps followed to assess the relationship between class, classifier and dimensionality reduction method for the classification of synthetic hyperspectral images.

For all the image classification experiments, the training samples required for classification were extracted from the image itself. The classified results were validated using independent testing samples drawn from the image. The significance of classification

accuracy enhancement of the MCS was assessed by two tailed Kappa Z-test at 95% confidence interval (Russell G. Congalton, 2008). If the computed value of Z exceeds 1.96, classification results of the MCS are significant.

4.2.5 Image classification using SVM

In order to assess the comparative performance of MCS relative to the state-of-the-art methods, all the five hyperspectral images were classified by support vector machines (SVM) using the same reference data samples. The Radial basis function (RBF) was used as the kernel function in the SVM. All the SVM based image classifications were done using the LIBSVM toolbox (Chang, Chih-Chung and Lin, 2011). Because the SVM can classify high dimensionality data even without dimensionality reduction, we performed image classification experiments with and without applying dimensionality reduction and the overall accuracy estimates were compared with the MCS.

4.2.6 Datasets

We used five different sources of airborne hyperspectral images (one each from the sensors HyMAP, ProSpecTIR and HYDICE and two images from the sensor ROSIS) covering multiple sites and multiple land cover classes. False color composite of the hyperspectral images are shown in Figure 3.1, 3.2. For detailed description of hyperspectral images see Chapter 3.

4.3 Results and Analysis for the Original Hyperspectral Images

In this section, we examine the results of the class, classifier and dimensionality reduction method relationship for hyperspectral image classification with the original hyperspectral image. Further, the results of assessing the utility of deploying multiple dimensionality reduction methods in a single MCS architecture are presented. Finally, the results of the MCS are compared with SVM classification method to asses the potential of combining the simple classifiers in the MCS.

4.3.1 Examination of the relationship between classifiers and dimensionality reduction methods

The numerous classified images generated by the MCS were analyzed for identifying the set of classifiers and dimensionality reduction methods which exhibit relatively higher classification performance. Except with the ROSIS University image, for all the other images a minimum threshold of 85% overall accuracy was taken as the benchmark for considering the classification performance as 'acceptable' (Foody, 2002; Russell G. Congalton, 2008). For the ROSIS University image, the best classification accuracy obtained is 80%. This is similar to the best accuracy reported by Fauvel et al. (2012) who used the same image for urban land cover classification by several methods which are various combinations of SVM kernel based spatial-spectral approaches. Hence the minimum threshold of overall accuracy for the ROSIS University image was pegged at 75%.

4.3.1.1 HyMAP image classification

Results of the classification of the HyMAP image classification with each of the dimensionality reduction methods are shown in Table 4.1. From Table 4.1, it can be observed that for the ICA and MNF dimensionality reduction methods, out of the seven classifiers NED, SAM, and SSM offered acceptable overall accuracies. The overall accuracies of the remaining four classifiers are considerably lower than the threshold. For the PCA and OBS dimensionality reduction methods NED, SSM and MF classifiers offered acceptable classification accuracy. As observed with the ICA dimensionality reduction method, the overall accuracy obtained for remaining four classifiers with the OBS and three classifiers with the PCA is below the threshold. For the DWT-DR dimensionality reduction method NED, ACE, MF, and SSM offered acceptable overall accuracy. For the HyMAP image, it is apparent that only two (NED and SSM) out of the seven classifiers offered acceptable classification accuracies with all the dimensionality reduction methods. There were considerable differences in the overall accuracies produced by classifiers with respect to each dimensionality reduction method. For example, with the PCA dimensionality reduction method, overall classification accuracies of the TCIMF and MF are 55.10% and 90.17% respectively, indicating a difference of about 35%. The

classified image produced by the MCS after applying the majority voting rule on the decision function values of the better performing classifiers and dimensionality reduction methods is shown Figure 4.3. The overall classification accuracy of the classified image is 97.02%, an increase of 4.3% over the best performing classifier in Table 4.1. This increase is significant as evident from the two tailed Z-test at 95% confidence interval (computed Z-score is 8.53).

Table 4.1: Overall accuracy (OA) and Kappa coefficient (KC) for HyMAP image classification for various classifiers relative to each dimensionality reduction (DR) method.

DR method		**Classifiers**						
		NED	SAM	SSM	ACE	MF	OSP	TCIMF
ICA	OA (%)	92.73	91.87	92.42	81.86	78.19	81.9	80.93
	KC	0.9147	0.9047	0.911	0.7854	0.7458	0.786	0.7746
PCA	OA (%)	87.74	84.76	85.34	80.04	90.17	61.87	55.1
	KC	0.8571	0.8222	0.829	0.7707	0.8852	0.5566	0.478
MNF	OA (%)	90.37	88.08	88.13	73.54	83.99	80.43	79.5
	KC	0.8871	0.8613	0.8617	0.6984	0.8141	0.7747	0.7642
DWT-DR	OA (%)	88.28	76.0247	86.85	90.79	90.06	69.25	69.1
	KC	0.8635	0.717	0.8453	0.8916	0.8845	0.6379	0.6352
OBS	OA (%)	86.23	70.95	87.08	83.72	85.73	26.79	58.46
	KC	0.8393	0.6579	0.8489	0.8094	0.8324	0.1719	0.5098

4.3.1.2 ROSIS University image classification

The overall accuracy of the ROSIS University classified image for each dimensionality reduction method is shown in Table 4.2. Out of the seven classifiers, the number of classifiers which offered the overall accuracy above the threshold is three each for the MNF, DWT-DR, and OBS and one each for the ICA and PCA. The three classifiers which performed better for the DWT-DR and OBS are identical. The subspace modelling classifiers exhibit poor discrimination across the dimensionality reduction methods. As observed with the HyMAP classified image, the variation in the overall accuracy between various combinations of the classifiers and dimensionality reduction methods is drastic. Figure 4.4 (a) shows the classified image produced by the MCS after applying the majority voting rule on the decision function values of the better performing classifiers and dimensionality reduction methods. The overall classification

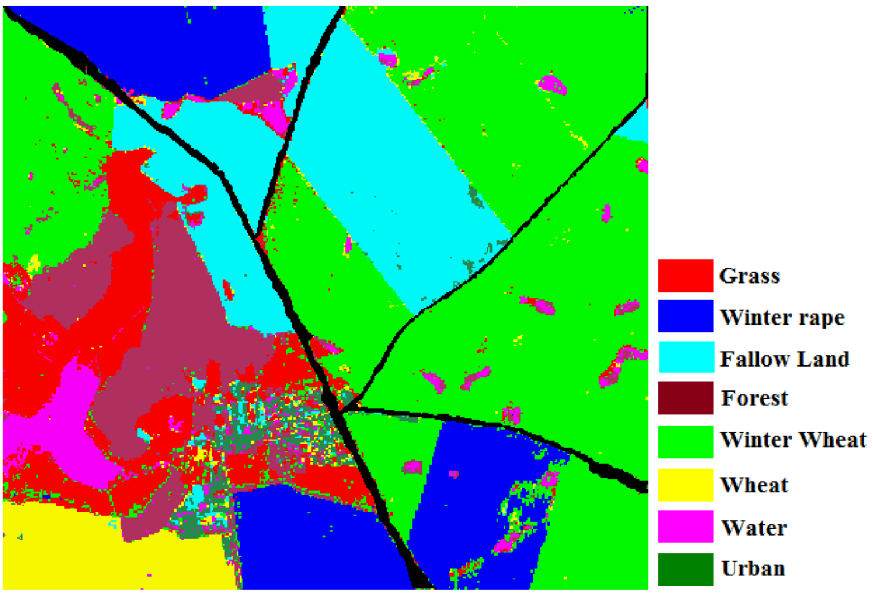


Figure 4.3: Classified image produced by combining the classifiers which are above acceptable threshold of HyMAP image.

accuracy of this classified image is 84.6%, which is 4.22% over the best individual. This apparent increase in the overall accuracy is significant at 95% confidence interval (computed Z-score is 7.79).

Table 4.2: Overall accuracy (OA) and Kappa coefficient (KC) for ROSIS University image classification for various classifiers relative to each dimensionality reduction (DR) method.

DR method		**Classifier**						
		NED	SAM	SSM	ACE	MF	OSP	TCIMF
ICA	OA (%)	74.31	72.58	73.47	77.55	68.59	16	16
	KC	0.709	0.6899	0.6999	0.7451	0.6344	0.0038	0.0038
PCA	OA (%)	78.23	67.24	68.56	71.8	68.57	16.36	16.35
	KC	0.7521	0.6318	0.6461	0.6838	0.6342	0.0089	0.0077
MNF	OA (%)	75.97	76.05	76.25	74.48	66.78	18.77	18.78
	KC	0.73	0.7322	0.7342	0.7139	0.6124	0.0293	0.0295
DWT-DR	OA (%)	77.28	69.15	80.33	75.39	66.1	24.85	20.97
	KC	0.7409	0.6501	0.7756	0.7203	0.6045	0.1274	0.0855
OBS	OA (%)	78.49	70.63	80.38	77.93	68.89	24.13	24.74
	KC	0.7548	0.6671	0.7762	0.7495	0.6348	0.1184	0.1239

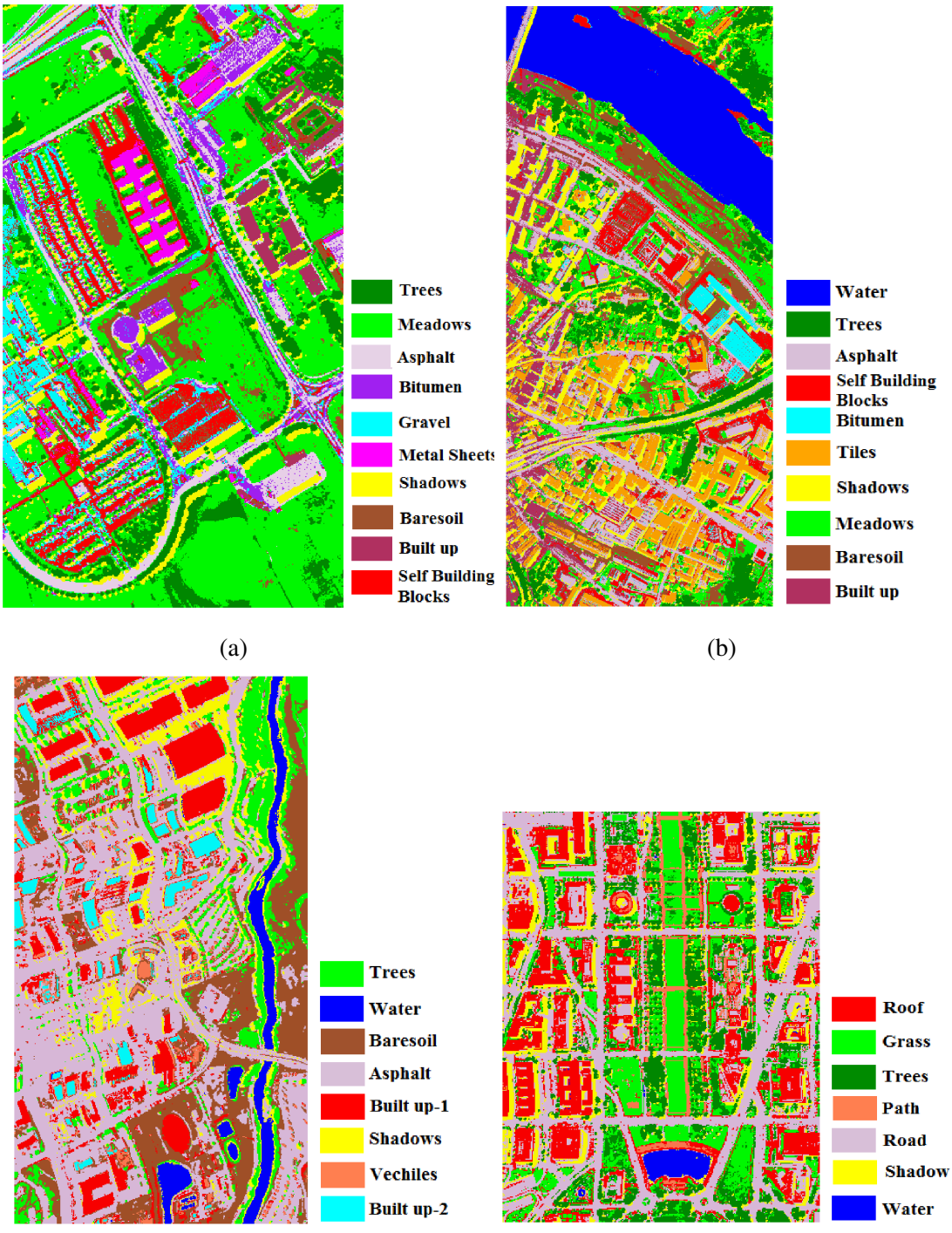


Figure 4.4: Classified images produced by combining the classifiers which are above acceptable threshold (a) ROSIS University (b) ROSIS City of Pavia (c) ProSpecTIR and (d) HYDICE.

4.3.1.3 ProSpecTIR image classification

Table 4.3 shows the overall accuracy of the ProSpecTIR image relative to each dimensionality reduction method. Compared to the few numbers of the classifiers which offered acceptable overall accuracy for the ROSIS University image, there are multiple classifiers which offered acceptable overall accuracy for each dimensionality reduction method. These classifiers are SAM, and SSM for ICA; NED, and SSM for PCA, and OBS; NED, SAM, and SSM for MNF; and NED, SSM, ACE, and MF for DWT-DR dimensionality reduction methods. Compared to the results obtained with ROSIS University image, the overall accuracies are consistently higher. Classification by the MCS using the majority voting as combination function with the better performing classifiers increased the overall accuracy by 2.18% (significant with computed Z-value is 3.55), to 93.5%. The corresponding classified image is shown in Figure 4.4 (c).

Table 4.3: Overall accuracy (OA) and Kappa coefficient (KC) for ProSpecTIR image classification for various classifiers relative to each dimensionality reduction (DR) method.

DR method		**Classifier**						
		NED	SAM	SSM	ACE	MF	OSP	TCIMF
ICA	OA (%)	84.81	85.18	85.28	84.66	70.33	70.91	69.26
	KC	0.8215	0.8261	0.8272	0.8198	0.659	0.6613	0.6421
PCA	OA (%)	88.5	78.2	88.74	80.37	68.22	60.12	59.46
	KC	0.8657	0.7475	0.8685	0.7734	0.6307	0.5325	0.5236
MNF	OA (%)	87.56	91.01	91.32	82.62	57.74	65.66	52.9
	KC	0.8548	0.8956	0.8993	0.7977	0.5215	0.6054	0.4591
DWT-DR	OA (%)	88.89	67.35	89.52	88.34	88.32	68.51	77.23
	KC	0.8703	0.6199	0.8776	0.8629	0.8632	0.6337	0.733
OBS	OA (%)	88.19	72.82	89.49	84.21	63.67	65.39	65.92
	KC	0.8621	0.6799	0.8773	0.8153	0.575	0.5983	0.6051

4.3.1.4 ROSIS City of Pavia image classification

The overall accuracy of the classifiers obtained with each dimensionality reduction method for ROSIS City of Pavia image is shown in Table 4.4. Though this image and the ROSIS University image were acquired by the same sensor and for the same geographical region with several common land cover categories, there are considerable

variations in the better performing classifiers for each dimensionality reduction methods. Only with ICA and MNF dimensionality reduction methods, there are multiple classifiers with accuracy exceeding the threshold limit. Table 4.4 indicates a reduced performance of the classifiers relative to threshold with remaining dimensionality reduction methods. For example, in Table 4.2 there are four classifiers for the DWT-DR, where as there is only one classifier with accuracy exceeding the threshold for DWT-DR in Table 4.4. As observed with ROSIS University image, subspace modelling classifiers exhibit poor classification results. Further, classification of the ROSIS City of Pavia image by the MCS (see Figure 4.4 (b)) with the better performing classifiers and dimensionality reduction methods indicate the overall accuracy to 89.7%. This is a marginal improvement (1.3%) over the best performing classifier in the Table 4.4, but the accuracy improvement is statistically significant (computed Z-value is 3.29) at 95% confidence interval.

Table 4.4: Overall accuracy (OA) and Kappa coefficient (KC) for ROSIS City of Pavia image classification for various classifiers relative to each dimensionality reduction (DR) method.

DR method		**Classifier**						
		NED	SAM	SSM	ACE	MF	OSP	TCIMF
ICA	OA (%)	87	84.64	85.48	85.93	85.18	14.34	14.33
	KC	0.8591	0.8247	0.8343	0.8392	0.8298	0.0236	0.0234
PCA	OA (%)	81.2	80.04	80.42	86.31	81.17	23.23	22.51
	KC	0.7842	0.7713	0.7756	0.8434	0.7846	0.1473	0.1486
MNF	OA (%)	88.4	88.13	88.14	85.77	85.57	19.51	19.51
	KC	0.8672	0.8643	0.8644	0.8375	0.8343	0.0979	0.0979
DWT-DR	OA (%)	81.54	81.64	86.32	84.64	83.64	33.71	33.75
	KC	0.7879	0.7897	0.8427	0.8247	0.811	0.2518	0.2492
OBS	OA (%)	82.01	82.11	85.81	84.16	84.55	29.82	38.97
	KC	0.7932	0.7951	0.8369	0.8193	0.8226	0.21	0.307

4.3.1.5 HYDICE image classification

The overall classification accuracy of the HYDICE image using the various classifiers relative to each dimensionality reduction is shown in Table 4.5. From Table 4.5, it is obvious that most of the classifiers performed well with the HYDICE image. The NED, SAM, SSM, and ACE classifiers offered acceptable classification accuracy for all the

dimensionality reduction methods. The classified image produced by the MCS after applying the majority voting rule on the decision function values of the better performing classifiers and dimensionality reduction methods is shown Figure 4.4. The overall classification accuracy of this classified image is 98.91%, which is 3.89% (significant with computed Z=9.54) higher compared to the best classifier in the Table 4.5.

Table 4.5: Overall accuracy (OA) and Kappa coefficient (KC) for ROSIS City of Pavia image classification for various classifiers relative to each dimensionality reduction (DR) method.

DR method		**Classifier**						
		NED	SAM	SSM	ACE	MF	OSP	TCIMF
ICA	OA (%)	93.81	94.82	94.69	91.86	74.45	72.35	71.83
	KC	0.9263	0.9383	0.9367	0.9032	0.6935	0.6734	0.6675
PCA	OA (%)	93.24	92.42	94.98	87.61	77.35	71.26	77.62
	KC	0.9196	0.91	0.9403	0.8527	0.7283	0.6615	0.7361
MNF	OA (%)	91.41	95.05	94.86	91.29	86.48	73.22	79.62
	KC	0.8977	0.9423	0.9384	0.8965	0.8377	0.6834	0.758
DWT-DR	OA (%)	92.82	94.73	93.58	93.72	93.66	70.8	71.45
	KC	0.9147	0.9373	0.9237	0.925	0.9243	0.6572	0.6642
OBS	OA (%)	93.12	91.41	93.74	92.9	90.53	66.12	70.34
	KC	0.9183	0.8985	0.9256	0.9151	0.8865	0.6021	0.6498

4.3.1.6 Suitability of dimensionality reduction method and classifier

Comparison of the Tables 4.1 through 4.5 reveals that only small subsets of classifiers (with members ranging from one to four) produce acceptable classification results for multiple hyperspectral images. For example, two classifiers (NED and SSM) produced acceptable overall accuracy for three different hyperspectral images (HyMAP, ProSpecTIR, and HYDICE images). Table 4.6 shows the list of classifiers which produced high classification results. The possibility for a classifier to be a member of this subset (of classifiers which produce high classification results) appears to depend upon the dimensionality reduction method. Further, within this subset of classifiers, the maximum overall accuracy achieved is a function of specific hyperspectral image and dimensionality reduction method. Based on the apparent changes in the pairing of classifiers and dimensionality reduction methods which produce high classification results (see Table 4.6), it can be inferred that there is no single classifier or dimensionality reduc-

tion method which is optimal across hyperspectral images. It is particularly important when different hyperspectral images represent different land cover and environmental settings. However, it can be observed that (see Table 4.6) for a given dimensionality reduction method, there are few classifiers (e.g. SSM classifier for MNF dimensionality reduction method) which can produce acceptable classification results for multiple hyperspectral images. This observation indicates the presence of empirical relationships between classifier and dimensionality reduction method. The observed consistent increase in the overall accuracy when the hyperspectral images were classified by the MCS with multiple dimensionality reduction methods indicates the complementary nature of dimensionality reduction methods.

Table 4.6: List of classifiers which offered better classification accuracy relative to each dimensionality reduction (DR) method for original high resolution hyperspectral images (numbers (%) indicate overall accuracy).

DR method	**Hyperspectral image**				
	HyMAP	ROSIS University	ProSpecTIR	ROSIS City of Pavia	HYDICE
ICA	NED (92.5), SAM(91.8), SSM (92.4)	ACE (77)	NED (84.8), SAM(85.1), SSM (85.2), ACE (84.6)	NED (87)	NED (93.8), SAM (94.8), SSM (94.7)
PCA	MF (90.2)	NED (78.2)	NED (88.5), SSM (88.7)	ACE (86.3)	NED (93.2), SAM(92.4), SSM (94.9)
MNF	NED (90.3), SAM (88.1), SSM (88.1)	NED (75.9), SAM (76),	SSM (91.3), SAM (91)	NED (88.4), SAM (88.1), SSM (88.1)	SAM (95.05), SSM (94.8)
DWT-DR	ACE (90.7), MF (90)	SSM (80.3)	NED (88.8), SSM (89.5), ACE (88.3), MF (88.3)	SSM (86.3), ACE(84.6)	SAM (94.7), SSM (93.5), ACE (93.7), MF (93.6)
OBS	NED (86.2), SSM (87)	SSM (80.4)	NED (88.2), SSM (89.5)	SSM (85.8), ACE (84.2), MF (84.5)	NED (93.1), SSM,(93.7), ACE (92.9)

4.3.2 Examination of the relationship between information class, classifier and dimensionality reduction method

Results presented in Tables 4.1 through 4.6 indicate the influence of dimensionality reduction methods and classifiers on the accuracy achieved in hyperspectral image classification at overall image level. However, it is important to understand the dynamics of dimensionality reduction method and classifier relationships at the information class level. In order to assess the premise that there exist more than one set of classifier and dimensionality reduction method which may offer better classification accuracy for a given land use category, the per-class accuracy estimates obtained from each of the images were analyzed for identifying the classifiers and dimensionality reduction methods which were dominant in the labelling of pixels of each land use category. For the land use categories which are found in all the five images, the classifiers and dimensionality reduction methods which produced better classification results are listed in Table 4.7.

For each land cover category, there are up to three classifiers which offer better classification accuracy. Examination of the different land covers and classifiers across the hyperspectral images reveals that different land covers within a single hyperspectral image are best classified by different classifiers. For example, for the HyMAP image, fallow land and bare soil indicate acceptable performance by any combination of four classifiers (NED/SAM/SSM/OSP) and two dimensionality reduction methods (ICA/MNF); winter wheat and winter rape, however could only be classified by either ACE/ MF with DWT-DR or NED with PCA. Similar relationships are found for the other images as well. This indicates the presence of empirical information class-classifier relationship, which again depends upon the specific image. Adding to that, the sets of dimensionality reduction methods which resulted in better classification performance for different land cover categories are comparable to the sets of classifiers which produced better classification accuracy (see Table 4.7). Further, Table 4.7 indicates the presence of several pairs of classifier, and dimensionality reduction methods which are optimal to similar land cover categories found in multiple hyperspectral images. For example, land cover category ‘grass’ found in four images (HyMAP, ROSIS-University, ROSIS-City of Pavia, and HYDICE images) was classified with SAM/SSM-MNF classifie-dimensionality reduction method combination in three images (HyMAP, ROSIS-City of Pavia, and HYDICE images).

Table 4.7: List of classifiers and dimensionality reduction methods (DR) which offered better classification accuracy relative to different original high resolution hyperspectral images and land cover categories (only the classifiers which performed well with multiple dimensionality reduction methods are listed; the per class accuracy (in%) of the land cover classes relative to classifier and dimensionality reduction method are indicated in brackets).

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
Water	NED(95.1)	ICA			NED (99.5)	ICA	NED (98.3)	ICA	NED (99.4)	ICA
					NED (99.8)	MNF	NED (100)	MNF	NED(99.6)	MNF
Fallow land, Baresoil	NED(99.8) SAM(99.7) SSM(99.3)	ICA	NED(69.4),SAM(68.5) SSM (69.6)	MNF	NED(98.6) SSM(100)	PCA	SAM	PCA		
	NED(98.4) SAM(100) OSP(100)	MNF			NED(100),SSM(100) NED(99.5),SSM(99.5)	DWT-DR OBS	MF(59.8)	DWT-DR		
Grass, Meadows	SAM(100) SSM(100)	MNF	SAM(97.6),SSM(97.8) ACE(98.8)	ICA			SAM(95.6) SSM(95.8)	MNF	SAM(99.6), SSM(99.8)	PCA
			SAM(96.5),SSM(97.9) ACE(96.1)	DWT-DR					SAM(99.3), SSM(98.1)	MNF
			SAM(97.8),SSM(97.6)	OBS					SAM(99.7), SSM(100)	DWT-DR
			ACE(97.4)						SAM(99.9), SSM(99.9)	OBS
Trees			SAM(99.5),SSM(97.1)	DWT-DR	SAM(98.6) SSM(98.6)	PCA	NED(97.7) SAM(99.2) SSM(99.1)	DWT-DR	NED(97.4),SAM(95.8) SSM(96.0)	PCA
					SAM(100) SSM(99.7) SSM(99.7)	MNF	NED(98.3) SAM(99.1) SSM(99.5)	OBS	SAM(97.02),SSM(97.02)	MNF
									NED(98.6),SAM(98.4)	DWT-DR

Continued on next page

Table 4.7 – continued from previous page

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
									SSM(98.1)	
			SAM(99.6), SSM(97.4)	OBS	SAM(99.7), SSM(99.4)	DWTDR	NED(98.1)	PCA	SAM(99.8), SSM(99.2)	OBS
					SAM(99.7), SSM(99.2)	OBS				
Shadows			NED(99.7)				NED(99.6)	ICA	NED(95.2 SSM(96.5)	PCA
			NED(99.5)	ICA			NED(98.6) SSM(98.6)	PCA	NED(95.9) SSM(95.9)	DWT-DR
			NED(99.4) SSM(97.4)	DWT-DR	SAM(88.9) SSM(88.1) ACE(89.8)	ICA	NED(99.1) SSM(98.8)	DWT-DR	NED(95.7) SAM(95.4) SSM(97.6)	
			NED(99.4) SSM(99.2)	OBS			NED(97.9) SSM(98.9)	OBS		
Road, Asphalt			MF(84.7)	ICA	NED(96.1) SSM(95.0)	PCA	NED(82.8) MF(92.4)	ICA	NED(98.4), SSM(97.9)	PCA
			MF(84.7)	PCA	ACE(99.6)	DWT-DR	MF(91.8)	PCA	NED(98.4) SSM(98.6)	DWTDR
			MF(89.2)	MNF			NED(84.8) MF(92.5)	MNF	NED(98.4) SSM(98.6)	OBS
Bitumen			ACE(89.9)	PCA			SAM(83.1) ACE(84.5)	MNF		
			SSM(82.7)	DWT-DR			SAM(85.5) ACE(83.8)	DWT-DR		
			NED(81.1) SSM(83.6)	OBS			SAM(86.0) ACE(84.2)	OBS		
Self			MF(90.1)	ICA			NED(84.6)	MNF		

Continued on next page

Table 4.7 – continued from previous page

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
building blocks			MF(91.1) MF(91.2)	PCA MNF			SAM(88.1) SSM(87.9)			
Built up, Urban	ACE(78.5)	DWT-DR	ACE(95.6)	ICA	ACE(100) OSP(99.3)	ICA	MF(81.3)	ICA	MF(98.2), ACE(97.4)	OBS
Roof					ACE(99.5) OSP(98.9)	PCA	MF(80.8)	MNF		
			ACE(94.2)	OBS	ACE(98.9) OSP(99.1)	OBS			MF (99.4)	MNF
Winter wheat	NED (95.8)	PCA								
Winter rape	ACE (98.1) MF(100)	DWT-DR								
Metal sheets			NED(97.8),SAM(97.8)	PCA			ACE(98.2)	ICA		
Tiles			SSM(97.8),ACE(98.8)				MF(98.8)	PCA		
			NED(98.5) SSM(98.0)	DWT-DR			ACE(97.1)	MNF		
			ACE(97.7)				MF(98.9)	DWT-DR		
			NED(98.8) SSM(98.7),ACE(98.8)	OBS			ACE(97.4) MF(97.9)	OBS		

This indicates the existence of class-classifier-dimensionality reduction method relationship and the diversity of information generalization by various dimensionality reduction methods thus leading to data (or application) specific nature of dimensionality reduction methods.

4.3.3 Comparative analysis with SVM based image classification

Table 4.8 shows the best overall accuracies obtained from the MCS and SVM methods. The overall accuracy obtained by the MCS after combining the outputs of best combination of classifiers and dimensionality reduction methods indicates consistently higher accuracies. As evident from Table 4.8, except for the ROSIS-City of Pavia image for which both SVM and MCS show similar accuracy, MCS indicate about 5% higher accuracy over the SVM. However, it can be observed that the overall accuracy from the SVM meets the acceptable threshold for all the images indicating its strong generalization capability.

Table 4.8: Overall accuracy (in %) obtained by the MCS and SVM methods. For SVM image classification, the overall accuracy with and without dimensionality reduction is included (acronyms in the bracket indicate the dimensionality reduction method).

Hyperspectral image	**MCS**	**SVM**	
		With dimensionality reduction	Without dimensionality reduction
HyMAP	97.02	91.93 (ICA)	90.52
ROSIS-University	84.6	80.21 (MNF)	79.3
ProSpecTIR	93.5	88.25 (MNF)	88.6
ROSIS-City of Pavia	89.7	89.96 (MNF)	86.4
HYDICE	98.91	92.59 (PCA)	90.3

Apart from the overall accuracy estimates, computational complexity represented by time of computation of the SVM was compared with the MCS (Figure. 4.5). It is interesting to observe that the computation time of MCS is better than or comparable to the SVM. The apparent matching of the computational time of MCS with the SVM can be justified by the fact that the classifiers used in the MCS are linear with advantage of low complexity parameters estimation (as calculation of only mean and common vari-

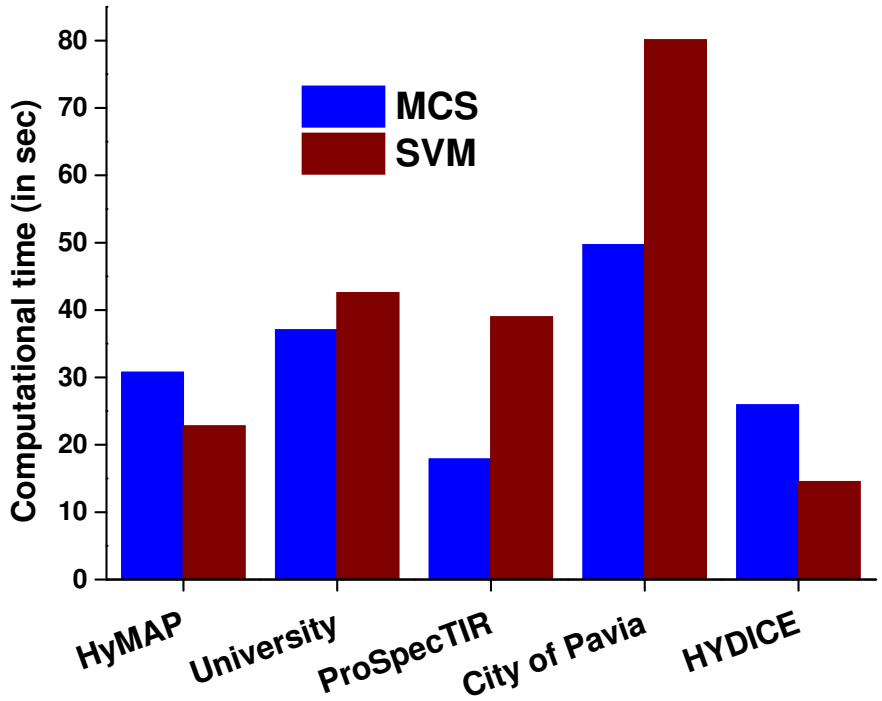


Figure 4.5: Computational time (in sec) of the MCS and SVM classification methods (time was calculated after dimensionality reduction method and includes both training and testing time; experiments were performed on a desktop computer Intel i3 processor, 3.2 GHz, 3 GB RAM and 64 bit operating system).

ance matrix is required for all the classifiers). Image classification by SVM involves mapping input data points to a higher dimensional space and solving $\frac{C(1-C)}{2}$ numbers of quadratic optimization problems (C is number of classes) thus requiring relatively more computation time. However, this apparent computational matching of MCS is case specific and depends upon the composition of classifiers in the MCS. Dimensionality reduction on the SVM performance is not effective; the difference in overall accuracies obtained with and without dimensionality reduction is found marginal across the images.

In this section, we analyzed the existence relationship between class, classifiers and dimensionality reduction methods of the original high resolution hyperspectral image. In the next section, we analyze the existence of relationship between class, classifier and dimensionality reduction method of the synthetic generated hyperspectral images.

4.4 Results and Analysis for the Spatially Downscaled Synthetic Hyperspectral Images

In this section, we analyze whether the established class, classifier, and dimensionality reduction method relationship (Table 4.6, and 4.7) is sensitive to change in the spatial resolution of the hyperspectral image. The better performing classifiers and dimensionality reduction methods are only presented in this section, since it is sufficient to confirm the validity of the class, classifier, and dimensionality reduction method relationship observed in the previous section. Further, the MCS classification results are presented to highlight the potential of deploying multiple dimensionality reduction methods. In the following the observed class, classifier, and dimensionality reduction method relationship with the spatially downscaled hyperspectral images are presented and compared with the original high resolution hyperspectral images.

4.4.1 Examination of classifier and dimensionality reduction method relationship

Table 4.9 shows the better performing classifiers relative to each dimensionality reduction method for all the five spatially downscaled hyperspectral images. As observed with original hyperspectral image (see Table 4.6), only a subset of classifiers ranging from one to four classifiers resulted in high classification accuracies for multiple images. This indicates the importance of pairing classifiers relative to each dimensionality reduction method when the hyperspectral images represent different land cover settings. From Table 4.9, it can be inferred that there exists a classifier which produce better results with a particular dimensionality reduction method across different spatially downscaled hyperspectral images. For example, SSM classifier with MNF and OBS method, ACE classifier with DWT-DR method produces better classification results with multiple images. This observation indicates the existence of the empirical relationship between classifier and dimensionality reduction method as observed with the original hyperspectral image. Further when the maximum overall accuracy is considered, as expected, the classification accuracies are degraded compared to the original hyperspectral image with each of the dimensionality reduction methods. In contrary,

the increase in classification accuracy has been observed with certain dimensionality reduction method for few hyperspectral images. For example, the classification accuracy has been increased for ROSIS University and City of Pavia hyperspectral images with ICA and OBS dimensionality reduction methods. Further, it is interesting that observed maximum overall accuracy of the DWT-DR method is similar with all the original and spatially coarse synthetic hyperspectral images. This indicates the insensitive nature of the DWT-DR to the spatial scale of the image.

Table 4.9: List of classifiers which offered acceptable classification accuracy relative to each dimensionality reduction (DR) method for the spatially downscaled hyperspectral images (numbers (%) indicate overall accuracy).

DR Method	Hyperspectral image				
	HyMAP	ROSIS University	ProSpecTIR	ROSIS City of Pavia	HYDICE
ICA	NED (92.50), SAM (91.8), SSM (92.4), MF (90.2)	SAM (84.84), SSM (84.84), ACE (85.15)	SAM (69.33), SSM (69.33), ACE (71.17)	NED (87.57), SAM (87.92), SSM (88.15), ACE (88.04)	NED (92), SAM (90.61), SSM (90.49), ACE (90.01)
PCA	MF (90.20)	MF (75.04)	NED (84.66)	NED (82.70), SSM (80.14), MF (81.07)	NED (89.53), SSM (90.98)
MNF	NED (90.30), SAM (88.10), SSM (88.10)	NED (58.81), ACE (59.42), MF (59.72)	NED (76.07), SAM (74.23), SSM (74.85)	NED (79.21), SAM(78.28), SSM (78.28), ACE (79.56)	NED (82.31) SAM (82.43), SSM (82.43)
DWT-DR	ACE (90.7), MF (90)	ACE (82.39)	ACE (88.34), MF (89.57)	SSM (84.55), ACE (86.41),	NED (92.18), SAM (91.34), SSM (91.94), ACE (92.54), MF (93.86)
OBS	NED (86.20) SSM (87)	SSM (84.27), ACE (84.69)	NED (84.05), SSM (84.66), ACE (82.21)	ACE (88.15)	NED (90.61), SSM (91.58), ACE (89.41), MF (91.46)

With the change in resolution of the image, it is necessary to analyze whether there is any change in the observed classifier and dimensionality reduction method relationships with the original hyperspectral images. A comparative analysis of Table 4.6 and Table 4.9, shows an interesting observation that similar type of classifiers are paired as the optimal classifiers relative to each dimensionality reduction method across differ-

ent images, as the case with original hyperspectral images (for e.g, SSM classifier with MNF, OBS method and ACE classifier with DWT-DR method). From these observations, we can conclude that: (i) the inferences of the original high resolution hyperspectral images can be directly applied to the spatially downscaled synthetic images (ii) the classifiers' performance is mainly influenced by the underlying information content of the hyperspectral image.

4.4.2 Examination of the class, classifier and dimensionality reduction method relationship

Results of the Table 4.9 indicate the existence of the empirical preference of classifier and dimensionality reduction method for the spatially downscaled hyperspectral image as observed with the original hyperspectral image. It is also important to analysis the influence of the dimensionality reduction method and classifier relationships at the land cover class information level. Table 4.10 shows the resulting better performing classifier and dimensionality reduction method pair with all the land cover classes for the spatially downscaled hyperspectral images. Since the aim of this experiment is to assess whether there exists a similar kind of relationship between class, classifier and dimensionality reduction method with original high resolution image, only the better performing classifier and dimensionality reduction method pairs are listed in Table 4.10. The per-class accuracy of the land cover classes are indicated in brackets. From Table 4.7 and Table 4.10, we can derive some interesting inferences. First of all, there are three to four classifiers which offer better classification accuracy for most of land cover classes as is the case with original high resolution images. Secondly, the per-class classification accuracy has been increased, decreased, and neither increased nor decreased when compared to original scale image. However, this is significantly dependent upon the type of land cover classes.

Table 4.10: List of classifiers and dimensionality reduction methods (DR) which offered better classification accuracy relative to different spatially downscaled synthetic hyperspectral images and land cover categories (only the classifiers which performed well with multiple dimensionality reduction methods are listed; the per class accuracy (in%) of the land cover classes relative to classifier and dimensionality reduction method are indicated in brackets).

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
Water	NED(95.1)	ICA			NED(100), SAM(100)	PCA	NED(100), SAM(100)	All	NED(98), SAM(98)	All
						DWT-DR	SSM(100)	DR	SSM(98)	DR
Fallow land	NED(99.8)	ICA	ACE(81.63)	PCA	NED(75)	PCA	SAM(53.33)	PCA		
Bare soil	SAM(99.7)	ICA	ACE(81.63)	PCA	MF(75)		SSM(53.33)			
	SSM(99.3)	ICA								
	NED	MNF	MF(71.43)	DWT-DR	NED(70)	DWT-DR				
	SAM				SSM(70)					
	SSM				TCIMF(80)					
	OSP				TCIMF(80)					
			ACE(85.71),MF(85.71)	OBS	NED(70)	OBS				
					SSM(70)					
Grass, Meadows	SAM(100)	MNF	SAM(97.59),SSM(97.59)	ICA			SSM(96.15)	PCA	NED(98.89)	PCA
Meadows	SSM(100)		ACE(98.8),MF(98.8)	ICA			SAM(96.15)		SSM(95)	
			MF(98.8)	PCA					NED(98.89),SSM(98.89)	
									ACE(96.11),MF(97.22)	DWT-DR
			SAM(95.18)	DWT-DR						
			SAM(96.39),MF(95.18)	OBS					NED(99.44),SSM(98.33)	OBS
Trees			NED(98.7)	PCA	SAM(94.44)	DWT-DR	NED(99.28)	PCA	NED(96.67),SAM(96.67)	DWT-DR
					ACE(94.44)		SSM(90.65)		SSM(96.67),ACE(97.78)	
					MF(94.44)				MF(98.89)	
			NED(98.71), SAM(100)	DWT-DR	NED(83.33)	OBS	NED(99.28)	DWT-DR	NED(96.67),SSM(97.78)	OBS

Continued on next page

Table 4.10 – continued from previous page

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
			SSM (98.7)		SAM(94.44) SSM(83.33) ACE(83.33)		SAM(100) SSM(100) SSM(100)			
			NED(98.71), SAM(100) SSM(100)	OBS			NED(99.28) SAM(100) SSM(100)	OBS		
Shadows			SAM(86.67),SSM(88.89) ACE(86.67)	ICA	ACE(100) MF(100)	DWT-DR	SAM(100),SSM(100)	PCA	ACE(88.37),MF(91.47)	DWT-DR
			NED(86.67)	PCA			NED(94.74)	OBS	ACE (85.27), MF (82.95)	OBS
			NED(86.67),SSM(82.22), ACE(84.44)	DWT-DR						
			NED(86.67),SSM(84.44) ACE(88.89)	OBS						
Road Asphalt			SAM(82.28),SSM(82.28) ACE(82.99),MF(81.07)	ICA	NED(84.38) SSM(84.38)	PCA	SAM(74.82),SSM(74.82 ACE(74.1),MF(93.53)	ICA	NED(95.96),SAM(95.96) SSM(95.96),ACE(93.94)	ICA
			MF(81.01)	PCA	MF(84.38)	MNF	MF(86.33)	MNF	ACE(82.83),MF(87.88)	DWT-DR
			ACE(80.38)	OBS	NED(84.38)	OBS	SSM(73.38),MF(89.93)	DWT-DR	MF(97.98)	OBS
					SSM(87.5) MF(90.63)	DWT-DR	SSM(73.38),ACE(78.42 MF(89.2)	OBS		
Bitumen			SAM(95.74),SSM(95.74) ACE(95.74)	ICA			NED(88.24),SAM(89.92) SSM(89.92),ACE(89.92)	ICA		
							NED(89.08),SAM(89.92) SSM(89.92),ACE(89.92)	MNF		
			ACE(93.62)	DWT-DR			SAM(89.08),SSM(84.03) ACE(89.92)	DWT-DR		
			ACE(91.49)	OBS			SAM(89.08),SSM(84.03)	OBS		

Continued on next page

Table 4.10 – continued from previous page

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
							ACE(89.92),MF(89.19)			
Self			MF(94.12)	ICA			SSM(100)	PCA		
Building			MF(95.59)	PCA						
Blocks			SSM(85.89),MF(91.18)	DWT-DR			NED(96.67),SAM(96.67) SSM(96.67),ACE(96.67)	MNF		
			SSM(85.29),MF(94.12)	OBS						
Built up	ACE(78.5)	DWT-DR	SAM(87.5),SSM(87.5)	ICA	NED(100)	ICA	NED(84.62),SAM(82.05)	ICA	NED(95.63)	ICA
Urban			ACE(87.5),MF(70.83)	ICA	SAM(100)		SSM(82.05),ACE(82.05)	ICA	NED(95.63)	ICA
Roof			ACE(87.5),MF(70.83)	ICA	SSM(100)		SSM(82.05),ACE(82.05)	ICA	NED(95.63)	ICA
			ACE(87.5),MF(70.83)	ICA	ACE(100)		SSM(82.05),ACE(82.05)	ICA	NED(95.63)	ICA
			NED(70.83),ACE(91.67) MF(70.83)	PCA	ACE(100),OSP(100) TCIMF(100)	PCA	OSP(82.05),TCIMF(82.05)	PCA	SSM(96.72)	PCA
			NED(91.67),ACE(100) MF(100)	MNF	NED(100) ACE(100) MF(100)	MNF				
			SAM(91.67),ACE(83.33)	DWT-DR	MF(100) OSP(100) TCIMF(100)	DWT-DR			SAM(98.91),SSM(95.63) ACE(93.44),MF(96.17)	DWT-DR DWT-DR
			ACE(83.33),MF(83.33)	OBS	NED(100) SSM(100) ACE(100) MF(100) OSP(100)	OBS	ACE(87.17),MF(82.05)	OBS	SAM(100),SSM(94.54) MF(93.44)	OBS
Metal Sheets			SAM(100),SSM(100)	ICA			NED(95.83),SAM(96.89))	ICA		
Tiles			ACE(100)				SSM(96.89),ACE(96.89) MF(96.89)			

Continued on next page

Table 4.10 – continued from previous page

Classes	HyMAP		ROSIS University		ProSpecTIR		ROSIS City of Pavia		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR	Classifiers	DR
			SAM(100),SSM(100) ACE(100)	MNF						
			ACE(100)	PCA						
			SAM(100),ACE(100)	DWT-DR			ACE(95.83),MF(97.91)	DWT-DR		
			SAM(100),ACE(100)	OBS			ACE(95.83),MF(95.83)	OBS		

It is interesting to note that classification accuracy of the urban class materials (for e.g, bitumen, self building blocks, metal sheets etc.,) increased while the classification accuracy of the vegetation classes (grass, meadows, trees) remain unchanged. This indicates that the classes which have high intra-class variability could be classified better with the spatially downscaled synthetic hyperspectral images. However, the validity of this observation depends upon the image and the spatial area extent of the objects in the image.

Further, the examination of the better performing classifier and dimensionality reduction method pair relative to each land cover category indicates the presence of the similar type of classifier and dimensionality reduction method pairs across multiple images. For example, the land cover category 'trees' found in four images (expect HyMAP image) was classified better with NED/SAM/SSM-DWTDR, OBS in all the four hyperspectral images, and the land cover category 'road/asphalt' found in four images was classified optimally with ACE-ICA combination in three hyperspectral images (ROSIS University, ROSIS City of Pavia, and HYDICE image). This indicates the existence of semi-empirical preference of classes for certain classifier and dimensionality reduction methods. When the resulted optimal classifer-dimensionality reduction method pairs are compared with original high resolution image, similar optimal pairs of classifier and dimensionality reduction method are observed for certain classes (for e.g, water, trees, road, asphalt). However, for the 'grass' land cover category the optimal classifiers are similar but the dimensionality reduction method is different for two hyperspectral images; for the 'shadow' region different pairs of classifier and dimensionality reduction method are observed. This difference in relationship for the 'grass' land cover class could be due the sub-optimal performance of the classifiers with MNF dimensionality reduction method. These observations indicate that the suitable combination of classifier and dimensionality reduction methods for land cover classes is a function of underlying information content and is insensitive to changes in the spatial resolution of the image.

4.4.3 MCS classification results

The classifiers which offer classification accuracy above the acceptable threshold for the spatially downscaled hyperspectral image are combined using majority voting to test

the possibility for the further enhancement in the accuracy. The classification results are shown in Figure 4.6 and 4.7. There is about 2.5%, 3.02%, and 1.2% increase in the classification accuracy when compared to the single best classifier. Despite relatively lower in magnitude, these increases in the classification accuracy are statistically significant. The decrease in classification accuracy is statistically insignificant (ProSpecTIR image). Further, it can be observed that higher classification accuracies are achieved for the ROSIS University and ROSIS City of Pavia hyperspectral image when compared to the MCS results of the original high resolution images. This might be due to the effect of decrease in intra-class variability of the hyperspectral image when the image is spatially coarser, and as a result the neighbouring pixels tend to be smooth. The results conclude that the optimal resolution of the image depends upon the underlying land cover classes, and the better classification results could be obtained with medium resolution (spatially coarser) image for the urban class images.

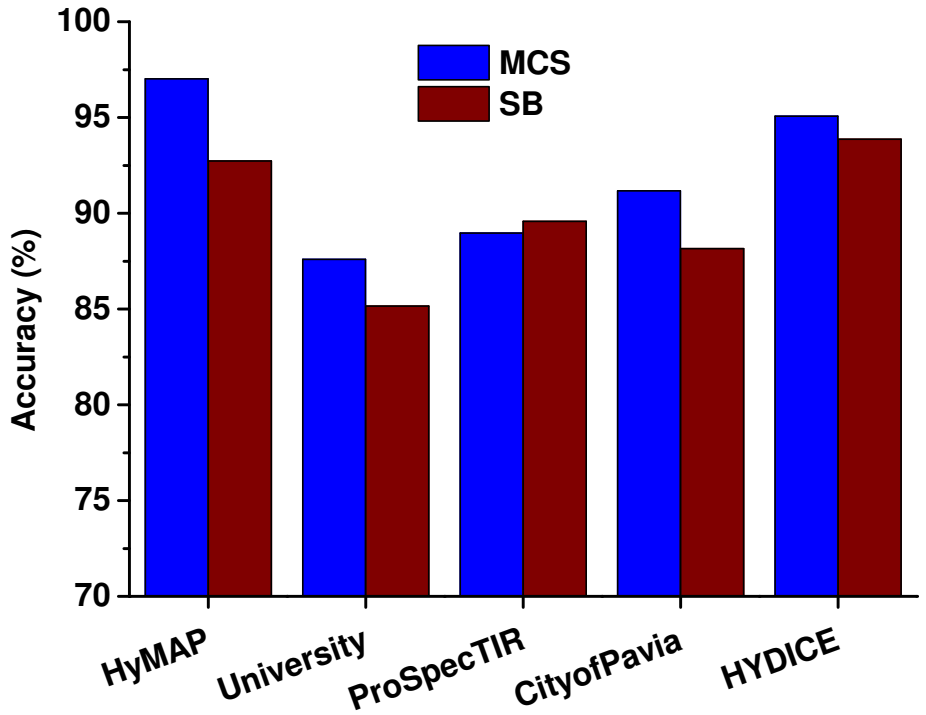
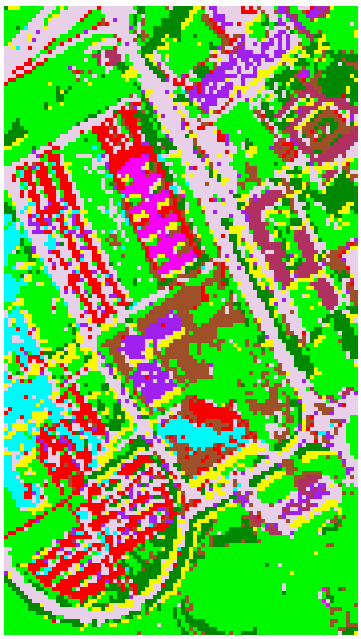
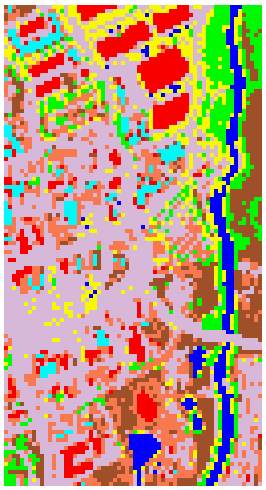


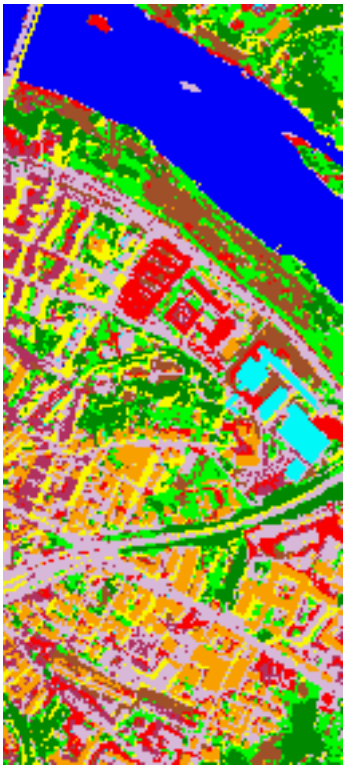
Figure 4.6: Overall accuracy of the MCS based classification after combing the classifiers which are above the acceptable threshold for the spatial downscaled hyperspectral images. Overall accuracy of the single best classifier (SB) is included for the reference.



(a)



(b)



(c)

(d)

Figure 4.7: Classified images produced by the MCS by combining the decision function values of the classifiers and dimensionality reduction methods which meet the acceptable threshold criterion for the spatially downscaled hyperspectral images (a) ROSIS University (b) ProSpecTIR (c) ROSIS City of Pavia (d) HYDICE.

4.5 Results and Analysis for the Spectrally Downscaled Synthetic Hyperspectral Images

In this section, we analyze whether the established class, classifier, and dimensionality reduction method relationship is sensitive to the changes in spectral resolution of the hyperspectral images. The experimental analysis was done only for the relatively better performing classifiers and dimensionality reduction methods. Further, the results are compared with the results obtained from the original hyperspectral images and the synthetic images generated by spatially downscaling the hyperspectral images. Finally, the better performing classifiers were combined in the MCS framework.

4.5.1 Examination of the classifier and dimensionality reduction method relationship

The list of better performing classifiers relative to each dimensionality reduction method for the three hyperspectral images are shown in Table 4.11. There are multiple classifiers which offer better accuracy for different images relative to different dimensionality reduction methods. Table 4.11, reveals that the existence of a set of classifiers which offers better classification performance across multiple images for a given dimensionality reduction method. This indicates that the inferences derived from the original hyperspectral image classification can be extended to the spectrally downscaled images.

When compared with the results obtained from the original hyperspectral images for each dimensionality reduction method, it can be observed that there is no significant difference in the maximum overall classification accuracy. In fact, a marginal improvement has been observed with the spectrally downscaled images. For example, for the ProSpecTIR image with the MNF method 91.7%, and 91.3% of classification accuracy is observed respectively for the spectrally downscaled image and original high resolution image; and 95.8%, and 95.1% classification accuracy observed for HYDICE image. Similar trend has been observed for other dimensionality reduction methods and other images as well. This is an important observation for the hyperspectral image classification where the inclusion of large number of spectral bands in classification is always a question to deal with.

Table 4.11: List of classifiers which offered better classification accuracy relative to each dimensionality reduction (DR) method for the spectrally downscaled hyperspectral images (numbers (%) indicate overall accuracy).

DR	Hyperspectral Image		
method	HyMAP	ProSpecTIR	HYDICE
ICA	NED (92.5), SAM (91.8,) SSM (92.4) , MF (90.2)	NED (85.4), SAM (86.5), SSM (86.5), ACE (86.5)	NED (93.6), SAM (94.1), SSM (94.1), ACE (94.5)
PCA	MF (90.2)	NED(87.4), SSM (86.7)	NED (93.8), SAM (95.3), SSM (94.9)
MNF	NED (90.3), SAM (88.1), SSM (88.1)	SAM (91.7), SSM (91.7)	SAM (94.2), SSM (94.2), MF (95.8)
DWTDR	ACE (90.7), MF (90)	NED (88.4), SSM (89.7), ACE (87.9), MF (89)	SSM (95.3), ACE (93.9), MF (93.1)
OBS	NED (86.2), SSM (87)	NED (88), SSM (90),	NED (92.9), SSM (93.4), ACE (93.2)

4.5.2 Examination of class, classifier and dimensionality reduction method relationship

Table 4.12 shows the list of better performing classifiers and dimensionality reduction method pairs relative to each land cover class across different hyperspectral images. From the comparison of the Tables 4.7 and 4.12, it can be inferred that land cover classes exhibit empirical preferences for certain pairs of classifiers and dimensionality reduction methods across the hyperspectral images. The resulting optimal classifier and dimensionality reduction method pairs are comparatively similar to the original high resolution images. Further, there is no significant difference in the classification accuracy of the land cover classes as compared to the original high resolution image. These observations indicates that the suitability of classifier and dimensionality reduction methods for the land cover classes are insensitive to the change in spectral resolution of the hyperspectral image.

Table 4.12: List of classifiers and dimensionality reduction methods (DR) which offered better classification accuracy relative to different spectrally downscaled synthetic hyperspectral images and land cover categories (only the classifiers which performed well with multiple dimensionality reduction methods are listed; the per class accuracy (in%) of the land cover classes relative to classifier and dimensionality reduction method are indicated in brackets).

Classes	HyMAP		ProSpecTIR		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR
Water	NED(95.1)	ICA	NED(99),SSM(99)	All DR	NED(99), SAM(99), SSM(99)	All DR
Bare soil	NED(99.8), SAM(99.7), SSM(99.3)	ICA	SSM(99.09)			
Fallowland	NED(100), SAM(100), SSM(100)	MNF	NED(98.6), SSM(99.09)	DWT-DR		
	OSP(100)		NED(99.5), SSM(99.3)	OBS		
Grass	SAM(100), SSM(100)	MNF			NED(99.7),SSM(99.8)	PCA
					NED(98.6),SAM(99.4),SSM(99.4)	DWT-DR
					NED(99.9)	OBS
Trees			NED(98.6), SAM(98.1), SSM(98.1)	PCA	SAM(97.2), SSM(97.2), ACE(98.8)	MNF
			SAM(100), SSM(100)	MNF	NED(97.4), SSM(98.8)	DWT-DR
			NED(98.6), SAM(99.7), SSM(99.4)	DWT-DR	NED(97.8), SSM(98.6)	OBS
			NED(98.4),SAM(99.4),SSM(99.4)	OBS		
Shadows			SAM(88.1), SSM(88.1),ACE(88.1)	ICA	NED(97.7),SAM(95.7), SSM(95.8)	ICA
			NED(93.2),SAM(93.2),SSM(93.2)	DWT-DR	NED(95.9),SAM(96.0),SSM(96.8)	PCA
			OSP(97.1),TCIMF(97.0)			
			ACE(93.2)	OBS	NED(99.3),SAM(99.6),SSM(99.6)	MNF
					MF(99.0)	
					NED(95.6), SAM(95.13), SSM(95.3)	OBS
Road			NED(84.2),SAM(81.2),SSM(84.1)	PCA	NED(100),SAM(100),SSM(100)	MNF
Asphalt			NED(88.7),SAM(84.0),SSM(84.1)	MNF	ACE(100),MF(99.7)	MNF

Continued on next page

Table 4.12 – continued from previous page

Classes	HyMAP		ProSpecTIR		HYDICE	
	Classifiers	DR	Classifiers	DR	Classifiers	DR
			NED(84.7),SSM(86.3)	OBS		
Built up	ACE(78.5)	DWT-DR	ACE(99.2), OSP(99.8), TCIMF(99.3)	PCA, ICA	SAM(93.6)	PCA
Urban			ACE(100)	MNF, DWT-DR	MF(92.2)	MNF
			NED(100),SSM(100),OSP(99.1)	OBS	ACE(92.1),MF(92.2)	DWT-DR
					ACE(92.1)	OBS

4.5.3 MCS classification results

Figure 4.8 shows the classification accuracies obtained after combining the classifiers whose accuracy exceeds the acceptable threshold for the three hyperspectral images. The classified images produced by the MCS after combining the better performing classifiers are shown in Figure 4.9. As observed with spatially downscaled image, the classification results are improved by 4.2%, and 2.4% for the ProSpecTIR and HYDICE hyperspectral images respectively. The MCS results of the spectrally downscaled images exhibit 2% increase when compared with the MCS results of spatially downscaled images and original high resolution images. This confirms the notion that large number of narrow spectral bands are not necessary for optimal classification results. Adding to that, the spectral downscaling of the hyperspectral image also reduces the time and memory complexity of the dimensionality reduction methods and the classifiers. This observation concludes that the choice of the optimal spectral resolution has to be decided based on the underling land cover classes. However, more experiments need to be conducted to assess the effect of different spectral resolutions on different land cover images.

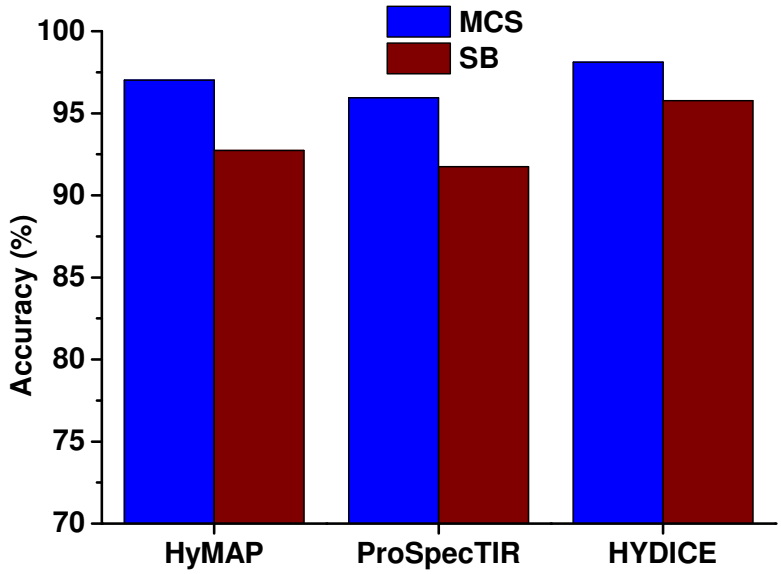
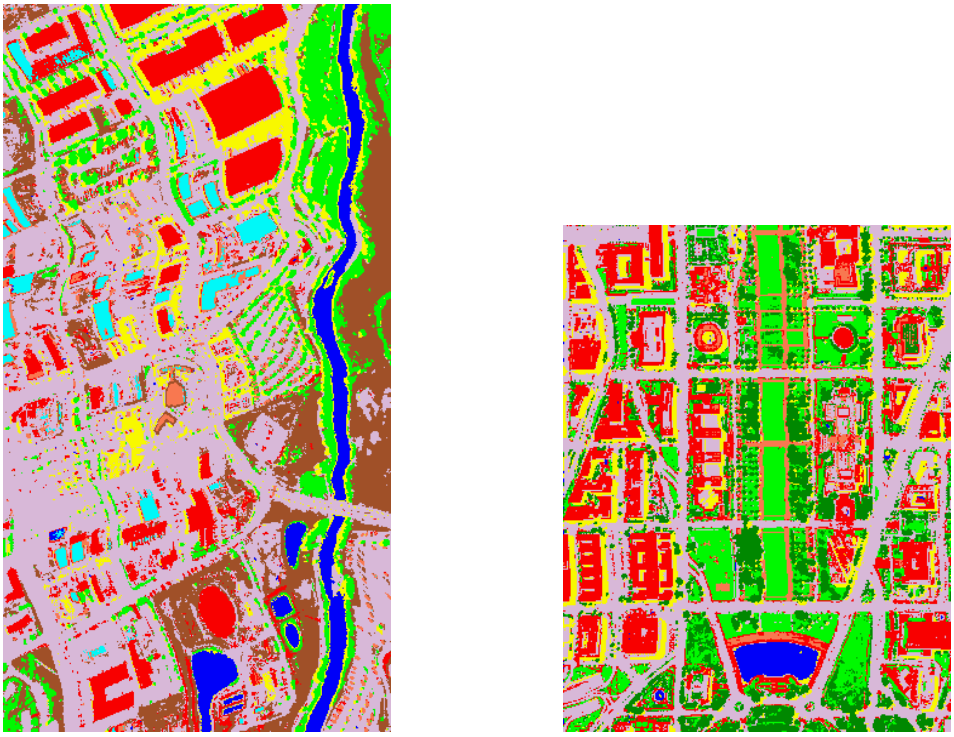


Figure 4.8: Overall accuracy from the MCS based classification after combing the classifiers which offer accuracy above the acceptable threshold for the spectrally downscaled hyperspectral images. Overall accuracy of the single best classifier (SB) is included for reference.



(a) (b)

Figure 4.9: Classified image produced by the MCS by combining the decision function values of the classifiers and dimensionality reduction methods which offered accuracy above the acceptable threshold for the spectrally downscaled hyperspectral images (a) ProSpecTIR (b) HYDICE.

4.6 Discussion

Identification of optimal classifier and dimensionality reduction method for a given hyperspectral image is an important pre-processing step in hyperspectral image classification. Generally, this is accomplished by heuristic methods such as pre-image classification, class spectral separability analyses etc., which often lead to the identified classifier and dimensionality reduction method being site and image specific. Parallel to the developments in acquisition of high quality hyperspectral images, a number of image classifiers and dimensionality reduction methods have been reported with varying levels of capabilities to handle hyperspectral image. The suitability of various classifiers and dimensionality reduction methods for land use and land cover classification has been reported reasonably well (Stathakis and Vasilakos, 2006; Oki et al., 2006; Xu and Gong, 2007; Chen and Peter Ho, 2008). However, the prospect of combining the relative merits of several classifiers from the perspective of suitability of dimensionality reduction methods as made possible by MCS is not understood. In particular, the impact of the si-

multaneous application of multiple classifiers and dimensionality reduction methods on the labelling of various land cover categories in a hyperspectral image is not reported. We conducted this study to address this important research aspect in the pathway to answer three fundamental questions: (1) is there any classifier(s) which offers acceptable classification accuracy across dimensionality reduction methods and data sources (as characterized by information classes)?, (2) given a set of information classes as characterized by data sources and site of data acquisition, what is the appropriate classifier and dimensionality reduction method combination?, (3) if there are multiple sets of classifiers and dimensionality reduction methods and if they are information class dependent, what could be the viable approach to pick up the best performing combination? We have classified multi-source airborne hyperspectral images acquired over five different sites covering a range of land cover categories by MCS.

The variability of classification performance of the classifiers and dimensionality reduction methods across the images has been examined based on the ability to achieve acceptable classification accuracy by the combined deployment of one or more classifiers and dimensionality reduction methods at the whole image and land cover category level. Results indicate that the performance of classifier(s) for labelling various land use categories depends considerably on the dimensionality reduction method and the land use category itself. At the image level, there is an evidence of the existence of classifiers which can offer higher classification accuracy for multiple sites and images (see Table 4.6). This apparent classifier-image preference is again depends upon the dimensionality reduction method. Further, it has been observed that the per-class classification accuracy is classifier sensitive. For example, land use categories such as fallow land, bare soil are best classified across multiple images by spectral matching based classifiers such as NED and SSM. For most of the urban classes, the covariance based methods (ACE and MF) seem perform well.

Amongst the entire set of classifiers considered, only a subset of classifiers offered acceptable classification accuracy. Adding to that, classifiers forming this subset are image and dimensionality reduction method dependent. Thus, it is clear that for betterment of hyperspectral image classification by MCS, the possibility of the existence of a limited number of combinations of classifiers and dimensionality reduction methods should not be overlooked. This is evident from the apparent increase in classification accuracy when the intermediate decision function values of subsets of classifiers and

dimensionality reduction methods are combined and classified by majority voting. The classification accuracy improvement offered by the MCS is statistically significant for all the five hyperspectral images. However, the magnitude of the performance improvement of MCS varies with different images. Further, the accuracy and reliability of land cover classification of a hyperspectral image can be enhanced by exploiting the information class dependent combinations of classifiers and dimensionality reduction methods by the dynamic selection of classifiers and dimensionality reduction methods.

Moreover, the choice of the optimal spatial and spectral resolutions has a significant impact on the accuracy of hyperspectral remote sensing applications (Yanchen Bo et al., 2005; Sprintsin et al., 2007; Thorp et al., 2013). Hence, the examination of the information class, classifier and dimensionality reduction method relationship was also conducted on the spatially and spectrally downscaled synthetic hyperspectral images. The results show that at an overall image level, similar type of optimal classifiers are evident with each dimensionality reduction method; at land cover class level, similar optimal pairs of classifier and dimensionality reduction method are also evident for the land cover classes as observed with the original hyperspectral images. Thus, the inferences derived from the hyperspectral image of a particular resolution can be directly applied to another resolution of the hyperspectral image.

The decrease in accuracy with second order statistics based dimensionality reduction method could be due to the fact that the second order statistics based dimensionality reduction method buries the objects with smaller spatial area extend as the noise in the lower order components, which is the case with the spatially downscaled images. The DWT-DR method offers similar classification performance with both the spatially downscaled and original hyperspectral images. This could be due to the smoothing effect of the DWT-DR method. This apparent stable behaviour of DWT-DR might be important for many remote sensing applications especially when the spatial extent of image is very large. Where as, with the spectrally downscaled hyperspectral images, the classification accuracies are comparable with the original hyperspectral image for all the dimensionality reduction methods.

The observed improvement in the performance of the MCS which included multiple dimensionality reduction methods is very useful for the classification of planetary datasets. There is a growing interest for classification of planetary hyperspectral images

to automatically identify the presence of various minerals and other geological materials of interest (Moussaoui et al., 2008; Hueso et al., 2010; Themelis et al., 2012). As it is challenging to check the performance of the classifiers, given the minimal reference data samples, it is imperative to deploy several classifiers and dimensionality reduction methods in parallel and combine the results in the MCS architecture as demonstrated in this study. Since the available reference samples are very few, the potential of applying machine learning algorithms for classification of planetary hyperspectral images is limited. The classifiers considered in this study have the inherent ability to perform image classification even with limited number of samples, as they require computing means of the information classes and common co-variance matrix.

The classification results obtained from the MCS designed in this study are fairly better than the results from the SVM while showing equal computational complexity. Our results suggest that the MCS with simple linear classifiers can be considered as the alternative image classification strategy to obtain optimal classification accuracy. Despite reasonably large datasets and methods used in this study, caution must be exercised while generalizing conclusions of this study because of the limitations such as choice of classifiers and dimensionality reduction methods, and within class variability of land use categories across sites.

4.7 Chapter Conclusions

In this chapter, we assessed the existence of the empirical relationships between land cover categories, classifiers and dimensionality reduction methods for hyperspectral image classification. The experimental analyses state that the choice of dimensionality reduction method and classifier significantly influences the classification accuracy of hyperspectral image. Different land cover categories within the same image prefer different combinations of classifier and dimensionality reduction method. The empirical relationships existed between land cover categories, classifiers and dimensionality reduction methods are vital for reliable classification of hyperspectral image by adaptive selection of classifiers and dimensionality reduction methods within the MCS architecture. The classification results and computational time complexity obtained from the MCS are compared against the popular SVM classifier. The results of spatially

and spectrally downscaled hyperspectral images reveal that the empirical relationship between class, classifier and dimensionality reduction methods is insensitive to the resolution of the hyperspectral image and is only depend upon the nature of information content in the hyperspectral image. Further the negative impact of varying classifier-dimensionality reduction method on the hyperspectral image classification can be mitigated by integrating multiple dimensionality reduction methods in the MCS architecture.

CHAPTER 5

DYNAMIC CLASSIFIER SYSTEM FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Prelude: Studies presented in the chapters 3 and 4 emphasize the need of improving the MCS architecture so as to be functionally 'dynamic' in the classifiers and dimensionality reduction methods selection for effective hyperspectral image classification. This chapter presents our algorithmic development, named as dynamic classifier system, aimed at introducing dynamism in the MCS architecture for selecting the optimal pairs of classifiers and dimensionality reduction methods relative to the input data dynamics. Further, the impact of combination schemes on the dynamic classifier system is analysed and the results are compared with the state-of-the-art methods.

5.1 Introduction

Dimensionality reduction methods mitigate the Hughes phenomena and make possible the application of classical statistical supervised classifiers on hyperspectral image (Jimenez and Landgrebe, 1998). A number of dimensionality reduction methods and classifiers are proposed in the literature for hyperspectral image classification. However, the performance of a dimensionality reduction method depends upon the nature of information classes and the classifier used (Chen and Qian, 2008; Bakos and Gamba, 2009). Thus, the classification accuracy depends upon the subjective choice of dimensionality reduction methods and classifiers. Identifying an optimal pair of dimensionality reduction method and classifier is a tedious task; given the numerous possibilities in classifiers, dimensionality reduction methods, and the nature and distributions of information classes. Numerous studies have addressed the suitability of dimensionality

This chapter is published in the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* Vol. 7, pp. 2080-2093, June 2014, with the title: "Dynamic Linear Classifier System for Hyperspectral Image Classification for Land Cover Mapping". Authors: Bharath Bhushan Damaodaran, Rama Rao Nidamanuri

reduction methods and classifiers for different applications and have suggested particular classifier and/or dimensionality reduction method appropriate for that particular application (Bakos and Gamba, 2009; Wu et al., 2010b; Wang and Chang, 2006; Srivastava et al., 2012). A common observation of these studies is that the selection of dimensionality reduction methods for the application at hand has to be done in relation with the classifier to be used. Most of the studies report optimal classifiers for specific applications and conclude that the classifier selection is a function of dimensionality reduction method. This context-specific knowledge may not be applicable across different hyperspectral images and applications. A multiple classifier system (MCS), an advanced pattern recognition technique, is emerging (Ceamanos et al., 2010; Du et al., 2012b; Samiappan et al., 2013) as an alternative paradigm for image classification to avoid the need of determining optimal classifier for each and every image analysis task a priori. An MCS permits simultaneous application of several classifiers on the input data, and the intermediate outputs of all the classifiers are combined for producing the final classified image. Most of the MCS architectures reported in the literature are designed to provide different input data sources to same classifier or same input data source to different classifiers (Ceamanos et al., 2010; Waske et al., 2007; Yang et al., 2010b). When the same input data are given to different classifiers, there could be an overlap in the decision boundaries of the classifiers (Yan and Shaker, 2011). Moreover, recently, it has been reported that there are no significant accuracy differences among most of the commonly used classifiers at the overall image level, but significant differences in perclass accuracy for some information classes (Bakos and Gamba, 2009; Szuster et al., 2011). In other words, certain type of information classes prefer certain type of classifiers and dimensionality reduction methods for better discrimination. This indicates the importance of selecting classifiers and dimensionality reduction methods in relation to information content of hyperspectral image.

Maintaining diversity in the classifiers and choosing appropriate combination scheme are vital to the functioning of the MCS. A major limitation of the MCS is that, without addressing the input data dynamics, merely inclusion of the classifiers of divergent groups may not necessarily yield the results expected from the MCS; instead end up producing results which are inferior to that of classical supervised classification (Yi et al., 2006; Zhou et al., 2002). Most of the literature on MCS deal with innovations in combination function schemes and creating diversity using different simple classi-

fiers (Yang et al., 2010b; Santos et al., 2013; Gokaraju et al., 2012; Alajlan et al., 2013; Senaras et al., 2013; Pal, 2008; Kumar et al., 2002; Udelhoven et al., 2009; Jun and Ghosh, 2009). The information representation characteristics of the dimensionality reduction methods in the MCS has been least studied. Prasad and Bruce (2008b) divided hyperspectral images into different subspaces and in each subspace linear discriminant analysis dimensionality reduction method was used to reduce the dimension of the subspace. Several studies have used the dimensionality reduction methods in the MCS framework. However their investigations were limited to a particular dimensionality reduction methods (Prasad et al., 2008, 2012; Kalluri et al., 2010; West et al., 2009; Lee et al., 2009). Further, our literature review revealed that the potential of creating diversity in the MCS based on differential performances of various classifiers against different dimensionality reduction methods has not been addressed. The creation of diversity in the classification process by deploying multiple dimensionality reduction methods is, in principle, possible as each different dimensionality reduction method generates data variance characteristic to its formulation and the distinct way of transformation. Thus, the extension of the MCS framework to let it acquire the capability to dynamically identify the optimal pairs of dimensionality reduction methods and classifiers simplifies image classification and reduces the data and application dependence of the MCS.

The objective of this chapter is to develop an algorithm, we label as dynamic classifier system (DCS), to dynamically select optimal pairs of classifiers and dimensionality reduction methods from a pool of classifiers and dimensionality reduction methods for hyperspectral image classification for land cover mapping. Further, the comparative performance of various trainable and non-trainable combination functions for combining the decision function values of the DCS has been assessed. The proposed DCS method is aimed at reducing the extensive human expert involvement in the selection of classifiers and dimensionality reduction methods in supervised image classification for various land cover mapping scenarios. The proposed DCS was designed with five dimensionality reduction methods and seven classifiers and was implemented for the classification of five multi-site multi-sensor airborne hyperspectral images for the discrimination of a range of land cover classes.

The organization of this chapter is as follows: Section 5.2 describes the methods used in this chapter. The proposed architecture of the system is presented in Section 5.3

and the experimental results are analyzed and discussed in Section 5.4. Finally, in the Section 5.5 we present conclusions drawn from this chapter.

5.2 Methodology

In this section, we provide brief description of the dimensionality reduction methods used for generating different variants of the hyperspectral image and the set of classifiers that constitutes the MCS. Further, in order to quantify the diversity existed between the classifiers, two diversity measures are also described in this section.

5.2.1 Dimensionality reduction method

The popular dimensionality reduction methods-principal component analysis (PCA), independent component analysis (ICA), minimum noise fraction (MNF), kernel principal component analysis (KPCA, and discrete wavelet transform-based dimensionality reduction (DWT-DR) method-were used to transform the hyperspectral images in to a low dimensional space. The PCA, MNF, ICA, and KPCA are based on statistical transformations which project data on to a new coordinate system by maximizing or minimizing certain statistical measures. The PCA and MNF maximize the second-order statistics (variance) in the projected components, where as ICA maximize the higher-order-statistics in the projected components. The KPCA is an extension of the PCA in which the input data points are transformed to higher-dimensional space (called as feature space) by a nonlinear transformation. In the feature space, the KPCA performs similar to the linear PCA. The KPCA has the advantage of capturing the higher-order statistics and provides better separability of the classes over the linear PCA .

The wavelet transform is an effective tool extensively used in signal and image processing. Each pixel in the hyperspectral image was decomposed by the wavelet transform using Daubechies filter (DB6). The outlier in the image was discarded and the approximation coefficients were reconstructed using inverse discrete wavelet transform. The number of reduced dimension of hyperspectral image depends on the level of decomposition. In this work, we generated series of wavelet transformed hyperspectral image with different level of decompositions.

5.2.2 Classifiers used in MCS

A set of simple classifiers with linear decision boundary, which can be broadly categorized into spectral matching methods, probabilistic methods, and subspace modeling methods, were selected for designing the MCS. The advantage of these methods is its fast training performance, since it requires only calculating mean of each class and common covariance matrix of the training samples. We give below a brief description of the classifiers selected.

5.2.2.1 Spectral matching methods

The spectral matching-based methods consist of minimum distance classifier (MDC) (Robila and Gershman, 2005) and spectral similarity measure(Granahan and Sweet, 2001) (SSM). The MDC labels an unknown pixel based on the minimum distance criterion. The SSM labels the pixels based on spectral angle and spectral brightness differences between the image and reference pixels. These classifiers label the unknown pixel based on the minimum decision function value.

5.2.2.2 Probabilistic methods

The probabilistic methods consist of linear discriminant classifier (LDC), logistic regression classifier (LRC), and naive Bayes classifier (NBC). The LDC can be derived from the maximum-likelihood classifier (MLC). An MLC assumes that the classes are normally distributed with different mean and covariance. The MLC becomes LDC, under the condition that all the classes have equal co-variance (Duda et al., 2000). The NBC assumes that features of input data are linearly independent and are normally distributed. It estimates the parameters (mean and variance) along each feature and finds the likelihood for each of the features in the input data. Since the features are independent, the posterior probabilities are estimated by taking the product of the likelihood of each of the features and prior probability (Hastie et al., 2001). The LRC is a binary classifier which linearly weights the input data features, and the weights are obtained by maximizing the log-likelihood function through maximum-likelihood estimate. Then, the sigmoid function is applied over the weighted sum of the input features. For multiclass problem, the logistic regression is implemented by one strategy versus rest strategy

(Hastie et al., 2001; Cheng et al., 2006). This group of classifiers labels unknown image pixel based on the maximum posterior probability values.

5.2.2.3 Subspace modelling methods

The subspace modelling methods consist of orthogonal subspace projection (OSP) and target-constrained interference minimized filter (TCIMF). These methods work based on nullifying the undesired and interfering class members. The goal of OSP is to find an orthogonal complement projector for each of the desired class (one at time) which projects the unknown pixel into a subspace orthogonal to the undesired class mean matrix. The unknown pixel is assigned to the class which has maximum decision value. The TCIMF assumes that the hyperspectral pixel vector is made up of three separate sources such as desired pixel vector, undesired pixel vectors, and interference. Similar to OSP, the TCIMF also splits the class mean vectors into desired class and undesired class mean vectors. It is designed by the finite impulse response (FIR) filter that passes the desired class mean vector, while annihilating the undesired class mean vectors. The weight vector of FIR filter for each of the desired class is computed by minimizing the energy of the filter (Ren and Chang, 2000). These methods label unknown image pixel based on the maximum decision function value.

5.2.3 Diversity measurement

Diversity among the classifiers set is important when constructing an MCS. However, quantitative description of the diversity is not straight forward and there is no accepted formal definition. Further, there is no direct relationship between the diversity and accuracy of the MCS (Kuncheva and Whitaker, 2003). To quantify the diversity introduced by deploying hyperspectral image after applying the dimensionality reduction methods, we adopted two statistical diversity measures: disagreement measure (DM) (Skalak, 1996) and Kohavi-Wolpert variance measure (KWM) (Kohavi and Wolpert, 1996) which belong to pairwise and nonpairwise categories of diversity measures, respectively. The DM is the ratio between the number of samples on which one classifier is correct and other is incorrect to the total number of samples. Suppose there are L

classifiers, then the averaged pairwise DM is given by

$$DM = \frac{2}{L(L-1)} \sum_{i=1}^{L} \sum_{j=i+1}^{L} \frac{N_{10}^{i} + N_{01}^{j}}{N} \tag{5.1}$$

where N_{10}^{i} is the number of samples correctly classified by ith classifier and incorrectly classified by the jth classifier, N_{01}^{j} is the number of samples correctly classified by jth classifier and incorrectly classified by the ith classifier, and N is the total number of samples. Similarly, the Kohavi-Wolpert measure is given by

$$KWM = \frac{1}{NL^2} \sum_{i=1}^{N} l_i \left(L - l_i\right) \tag{5.2}$$

where N is the total number of samples and l_i is the number of classifiers which correctly classify the ith sample. The diversity increases with increasing values of DM and KWM. In principle, any one of the diversity measures can be used for quantifying the diversity. However, a combination of these two categories of measures provides complimentary characteristics of diversity as both of them can be linearly related. Hence, the total diversity measure (DA) of the set of classifiers in the MCS can be calculated from 5.1 and 5.2 as

$$DA = DM + KWM \tag{5.3}$$

5.3 Approach for Introducing Dynamism in the Selection of Classifiers in MCS

In a typical MCS application, all the classifiers forming the classifiers set are applied on the input data irrespective of the inherent data dynamics with reference to the classifiers set. This structural constraint of the MCS may lead to poor classification performance and end up being no better than the classification performance achieved in a typical supervised image classification. We redesigned the MCS framework to automatically select and execute pairs of classifiers and dimensionality reduction methods which are compatible and offer optimal results.

Let $C = \{C_1, C_2, \ldots, C_L\}$ be the set of classifiers, $D = \{D_1, D_2, \ldots, D_M\}$ be the set of dimensionality reduction methods, n is the number of test pixels drawn from

the classified image, c is the number of categories in the classified image, then the error matrix $\mathbf{E}$ can be expressed as the distribution of n test pixels into c^2 cells. Let E_{ij} denote the number of test pixels classified into category j $(j = 1, \ldots, c)$ in the classified image of category i $(i = 1, \ldots, c)$ in the reference pixels, and let $u_i = \sum_{j=1}^{c} E_{ij}$ and $v_j = \sum_{i=1}^{c} E_{ij}$ are the row and column sums of the error matrix $\mathbf{E}$ respectively. Then, the overall accuracy can be computed as $O = \frac{\sum_{i=1}^{c} E_{ii}}{n}$. Similarly, producer's accuracy (P) and user's accuracy (U) can be computed as $P_i = \frac{E_{ii}}{v_i}$ and $U_j = \frac{E_{jj}}{u_j}$, respectively. The kappa coefficient is calculated as

$$K = \frac{n \sum_{i=1}^{c} E_{ii} - \sum_{i=1}^{c} u_i v_i}{n^2 - \sum_{i=1}^{c} u_i v_i}.$$

These intermediate accuracy estimates were used for estimating the optimal dimension of the dimensionality reduction methods and for pairing up the classifiers and dimensionality reduction methods by classifying the training pixels by all the classifiers and dimensionality reduction methods. Apart from this approach, the optimal dimensionality of dimensionality reduction methods can also be estimated by using a class separability measure which has computational advantage. Hence, the proposed DCS has also been implemented to obtain the optimal dimensionality of hyperspectral image, which is independent of the classifier by a class separability measure, Jeffreys-Matusita (JM) distance measure extended for multiclass category, and is equivalent to Bhattacharya bound (Bruzzone et al., 1995).

$$J_{Bh} = \sum_{i=1}^{c} \sum_{j>1}^{c} \sqrt{p(\omega_i)p(\omega_j)} J_{ij}^2 \tag{5.4}$$

where $J_{ij} = \sqrt{2\left(1 - e^{-b_{ij}}\right)}$, b_{ij} is the Bhattacharyya distance[1]. As the JM distance measure is a monotonic function, the flattening point of the curve is selected as the optimal dimension of the dimensionality reduction methods. The performance of the proposed DCS with these two approaches was compared. Figure 5.1 depicts the schematic outline of the proposed DCS with three stages and the algorithmic development of the DCS is shown in algorithm 1. Stage I involves 1) constructing the pool of dimensionality reduction methods and classifiers and 2) finding the optimal dimension of

[1] $b_{ij} = -\log\left\{\int_x \sqrt{p(\mathbf{x}/\omega_i)p(\mathbf{x}/\omega_j)}dx\right\}$, where $p(\mathbf{x}/\omega_i)$, and$p(\mathbf{x}/\omega_j)$ are the conditional density functions for the vector $\mathbf{x}$, given the classes ω_i, ω_j respectively.

Algorithm 1 Algorithmic representation of the proposed DCS.

Input:
I : Hyperspectral Image
C : Classifier set as $C = \{C_1, C_2, \ldots, C_L\}$
D : be the set of dimensionality reduction methods as $D = \{D_1, D_2, \ldots, D_M\}$
X : be the training set, which contains the location index of the training samples in the image
Y : be the testing set, which contains the location index of the testing samples in the image
c : total available classes in the image, classes are represented as $\omega_1, \omega_2, \ldots, \omega_c$
threshold : accuracy difference (in %)

Output:
S : selected classifier for each dimensionality reduction method

procedure DCS ALGORITHM
  Let $S = \{\}$ be a null set
  for $i = 1 \rightarrow M$ **do**
    Apply the i^{th} dimensionality reduction method (D_i) on the hyperspectral image I
    Estimate the optimal dimension of the dimensionality reduction method
      (a): Using the classification accuracy of training samples for each classifier with holdout strategy
      (b): Using Extended JM distance class seperability measure of training samples
    for $j = 1 \rightarrow L$ **do**
      Train and test the classifier C_j using holdout strategy of training samples
      Compute the error (confusion) matrix E, using the testing samples (remaining part of training samples)
      Calculate: Overall accuracy $O_{ij} = \frac{\sum_{m=1}^{c} E_{mm}}{n} \times 100$, and kappa coefficient

$$K_{ij} = \frac{n \sum_{l=1}^{c} E_{ll} - \sum_{k=1}^{c} u_k v_k}{n^2 - \sum_{k=1}^{c} u_k v_k}$$

      where E_{lk} denote number of test pixels classified into category k in the classified image of category k; u_k and v_k are the row and column sum of the error matrix E respectively.
    end for
    if $i = 1$ **then**
      $kd = \arg\max_j O_{ij}, j = 1 \ldots, L$
    else
      select the classifier based on diversity analysis
      $\{k1, k2, \ldots, kp\} = \arg\max_j (O_{ij} - O_{ij'}) \leq treshold > 0, j, j' = 1, 2, \ldots, L$
      $d = \arg\max_s \text{DA}\left(S \cup \{C_{ks}\}\right), s = 1, \ldots, p$ where $d \in \{k1, \ldots, kp\}$, DA is a function which provides diversity value when the classifier C_{ks} is added in the ensemble S
    end if
  end for
end procedure

DCS Combination:

- Train the selected classifiers/dimensionality reduction method pair in S using entire training sample
- Perform classification using selected classifier/dimensionality reduction method pair
- Combine the resulting decision values of the classifier/dimensionality reduction pair using combination functions

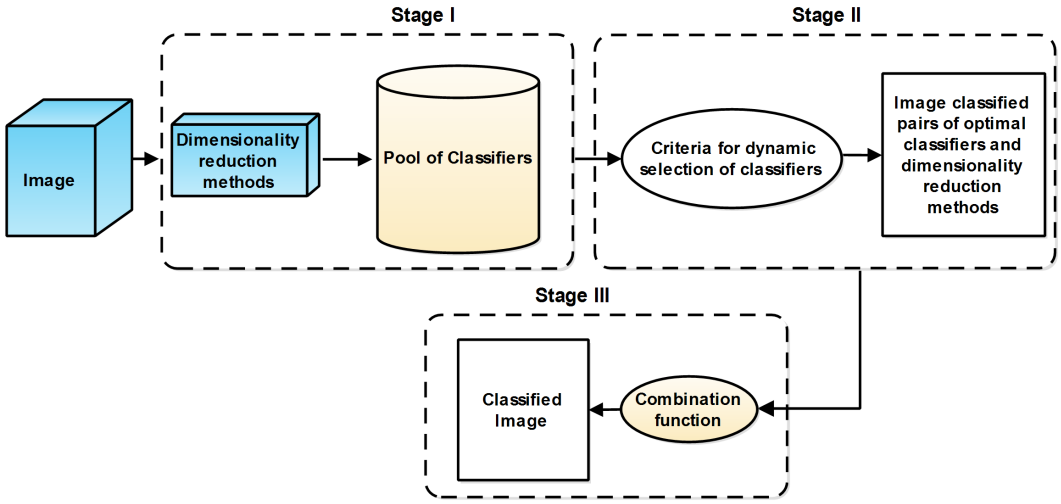


Figure 5.1: Schematic outline of the proposed DCS. In Stage I, the image classifications with all the classifiers relative to each dimensionality reduction method are performed. In Stage II, the DCS identifies pairs of optimal classifier and dimensionality reduction method. In Stage III, the output from the pairs of classifiers and dimensionality reduction method are combined to obtain final classified image.

dimensionality reduction methods relative to each classifier. In stage II, the MCS was programmed to select an optimal classifier relative to each dimensionality reduction method based on the classification accuracy estimates of training samples and the class separability measure while considering the diversity score. By this stage, the anticipated advantage is to give the MCS the ability to select only a subset of classifiers and corresponding dimensionality reduction methods for further processing for producing final classification results. In the present case, the MCS identifies five pairs of classifier and dimensionality reduction method which are accurate and diverse for further processing. The relative merits of identified classifier relative to each dimensionality reduction method are combined together in stage III to produce final classified image. We label this modified version of the MCS in which the classifiers are selected adaptive to specific image data at hand as "DCS".

Irrespective of the sophistication and suitability of the classifiers and corresponding dimensionality reduction methods, the image classification results are ought to be influenced by the combination function used (Du et al., 2012b; Yan and Shaker, 2011). It is, therefore, desirable to have an idea of the impact of various combination functions on the proposed DCS.

5.3.1 Evaluation of classifier combination functions

As is the case with MCS, the performance of the proposed DCS depends upon the type of combination function adopted for combining the intermediate classified images. It is, therefore, necessary to assess the impact of various combination functions on the performance of the proposed DCS. As the classifiers considered in the MCS are heterogeneous, the classifiers' decision values have to be transformed into a common scale before combining the classifiers. Let $\mathbf{d} = [d_1, d_2, \ldots, d_c]$ be the decision values of the classifier for c classes, then decision values $d_i, i = 1, 2, \ldots, c$ of the classifier can be transformed as follows

$$d_i = \begin{cases} \exp\left(-\frac{d_i}{\left(\sum_i d_i^2\right)^{1/2}}\right), & \text{for minimum decision value based classifiers} \\ \exp\left(\frac{d_i}{\left(\sum_i d_i^2\right)^{1/2}}\right), & \text{Otherwise} \end{cases}$$

After the above transformation, the selected classifiers relative to each dimensionality reduction method were combined by six nontrainable combination functions and two trainable combination functions. The nontrainable combination functions are majority voting (MV), maximum rule (max), minimum rule (min), product rule (prod), average rule (avg), and median rule (med). These combination functions have been extensively used in MCS due to their simplicity and robustness (Kittler, 1998). In the trainable combination functions, we used supervised classifiers namely, support vector machines (SVM) and MLC. The decision function values of the classifiers in the DCS are stacked together as a 1-D vector and fed as the input to the trainable combination function (Ceamanos et al., 2010; Thoonen et al., 2012). If there are L classifiers and c number of classes, then $d_{ik}, i = 1, 2, \ldots, L, k = 1, 2, \ldots, c$, is the decision function value of ith classifier and kth class. Then, the decision function values of L classifiers are stacked together as $\mathbf{t} = [d_{1k}, d_{2k}, \ldots, d_{Lk}], k = 1, 2, \ldots, c$, where element in $\mathbf{t}$ is a vector consisting of classifiers decision values for c classes.

5.3.2 Image classification using MCS and SVM

In order to assess the comparative performance of the proposed DCS with the state-of-the-art methods, all the five hyperspectral images were classified by the MCS and SVM using the same reference data samples. Because the SVM and some of the classifiers in the MCS can classify high dimensionality data even without dimensionality reduction, we performed the image classification experiments with and without dimensionality reduction for all the combination functions. The SVM classification is performed with RBF kernel using LIBSVM (Chang, Chih-Chung and Lin, 2011). Grid search method has been used to determine the optimal parameters of the SVM. The best classification accuracies were retained for comparison with the DCS.

5.3.3 Validation of the results

All the classified images were validated by cross-validation (Steele, 2005) to obtain robust estimates of accuracies of the classification experiments. All the reference data were split into training (90%) and testing (10%) samples by 10-fold cross-validation and then performed classification of each image 10 times for the possible 10 splits. For each classified image, an error matrix, overall accuracy, and kappa coefficient were calculated. The accuracy estimates thus obtained were averaged to obtain representative accuracy and error estimates. The number of reference data samples used is shown in Table 5.1. Further, the statistical significance of the differences among the different classification results was assessed by Z-test (Russell G. Congalton, 2008) and McNemar test (Chi-squared test) (Foody, 2004) at 95% of confidence interval.

5.4 Results and Analysis

5.4.1 Hyperspectral datasets

Experiments were performed on five different hyperspectral images (HyMAP, ROSIS University, ROSIS City of Pavia, ProSpecTIR,and HYDICE) which represent diverse land cover categories in agriculture, urban, and mixed environmental settings.

Table 5.1: Number of reference data samples of the five hyperspectral images.

HyMAP		**ROSIS University**		**ProSpecTIR**		**ROSIS City of Pavia**		**HYDICE**	
Class	Reference samples	Class	Reference samples	Class	Reference samples	Class	Reference samples	Class	Reference samples
Grass	1101	Trees	1363	Trees	982	Water	1800	Roof	1788
Winter wheat	1907	Asphalt	1839	Water	1354	Trees	1812	Grass	1512
Winter rape	1735	Bitumen	1101	Bare soil	1162	Asphalt	1700	Trees	1336
Wheat	1764	Gravel	1325	Asphalt	1239	Bitumen	1698	Path	793
Barren Land	2046	Metal sheets	862	Built up-I	1300	Tiles	1500	Road	1464
Water	889	Shadow	956	Shadow	670	Shadow	1616	Shadow	1065
Forest	1222	Meadows	1297	Vehicles	638	Meadows	995	Water	1014
Urban	590	Bare soil	1244	Built up -II	1316	Bare soil	908		
		Built up	705			Built up	978		
		Self building	1423			Self building	837		
Total	11254	Total	12115	Total	8661	Total	13844	Total	8972

5.4.2 Dynamic selection of classifiers relative to dimensionality reduction methods by the proposed DCS

For each dimensionality reduction method, one optimal classifier was identified from the pool of classifiers by 1) DCS constructed with the accuracy estimates obtained from the classification of training samples [Table 5.2 (a)] and 2) DCS constructed with class separability measure [Table 5.2 (b)].

Table 5.2: Identified pairs of optimal classifier and dimensionality reduction (DR) method by the proposed DCS and the corresponding best classification accuracy (OA: overall accuracy, KC: kappa coefficient); Estimated optimal dimension of the dimensionality reduction method is in brackets.

(a) Optimal dimension of DR method estimated based on training samples classification

Hyperspectral image	Selected classifiers relative to dimensionality reduction methods					best pair	
	ICA	PCA	MNF	DWT	KPCA	OA	KC
HyMAP	NBC (9)	LRC (14)	OSP (22)	LDC (5)	LRC (31)	93.25	0.915
ROSIS University	LRC (12)	LDC (14)	NBC (29)	MDC (2)	LRC (26)	82.78	0.802
ProSpecTIR	LRC (8)	LDC (17)	SSM(17)	MDC (5)	NBC (9)	90.31	0.887
ROSIS City of Pavia	LRC (7)	LDC (14)	LRC (9)	LRC (3)	LDC (31)	90.13	0.886
HYDICE	LRC (21)	NBC (10)	LDC (21)	SSM (4)	TCIMF(18)	91.73	0.898

(b) Optimal dimension of DR method estimated based on class separability measure

Hyperspectral image	Selected classifiers relative to dimensionality reduction methods					best pair	
image	ICA	PCA	MNF	DWT	KPCA	OA	KC
HyMAP	LRC (10)	LDC (12)	MDC (9)	LRC (5)	NBC (10)	92.48	0.907
ROSIS University	LRC (12)	MDC (14)	LRC (19)	LDC (3)	NBC (16)	81.23	0.784
ProSpecTIR	LRC (9)	MDC (9)	NBC (8)	NBC (7)	SSM (10)	90.21	0.885
ROSIS City of Pavia	LRC (12)	NBC (12)	MDC (13)	LDC (3)	LDC (14)	89.19	0.876
HYDICE	LDC (7)	SSM (8)	LRC (14)	NBC (3)	TCIMF (13)	91.73	0.898

The estimated optimal dimension of the dimensionality reduction methods is indicated in brackets. It can be observed that the DCS identified different classifiers for different dimensionality reduction methods. The comparison of classifiers identified

from the pool of classifiers reveals that only one classifier is repeated twice for different hyperspectral images. Further, an examination for repetition of the optimal classifiers for multiple hyperspectral images which contain similar land covers indicates significant variations with respect to dimensionality reduction methods.

This indicates that the optimal classifier identified from a particular hyperspectral image need not necessarily be applicable to another similar type of hyperspectral image, thus emphasizing the need for adaptive selection of classifiers. The comparison between Table 5.2 (a) and (b) indicates that the classification accuracies of the best classifier and dimensionality reduction pair are not highly variable across different hyperspectral images despite the occurrence of different classifiers as optimal. Frequency of the selected classifiers by the DCS relative to different dimensionality reduction methods and hyperspectral images indicates that some classifiers are most selected, whereas few classifiers are least selected. These observations highlight the potential of DCS to avoid influence of unsuitable classifiers in image classification.

5.4.3 Impact of the combination function on the classification performance of the proposed DCS

In principle, the classified image obtained from any one of the pairs of classifiers and dimensionality reduction methods identified by the DCS could be the final classified image for land cover mapping. However, running the DCS unrestricted leads to performing classification using all the pairs of optimal classifiers and dimensionality reduction methods for the same hyperspectral image; this results in multiple classified images with variations in the classification accuracy, however, marginal. Within the framework of the proposed DCS, we assessed the possibility of further enhancement of classification accuracy by combining the classified images obtained from the multiple pairs of classifiers and dimensionality reduction methods using several trainable and non-trainable combination functions. The overall accuracies and kappa coefficient are presented in Table 5.3 (for DCS with optimal dimensionality estimated by training samples classification) and Table 5.4 (for DCS with optimal dimensionality estimated by class separability measure) and the corresponding classified images with best combination functions in Figures 5.2, 5.3, 5.4, 5.5, 5.7, and 5.6.

5.4.3.1 DCS with non-trainable combination function

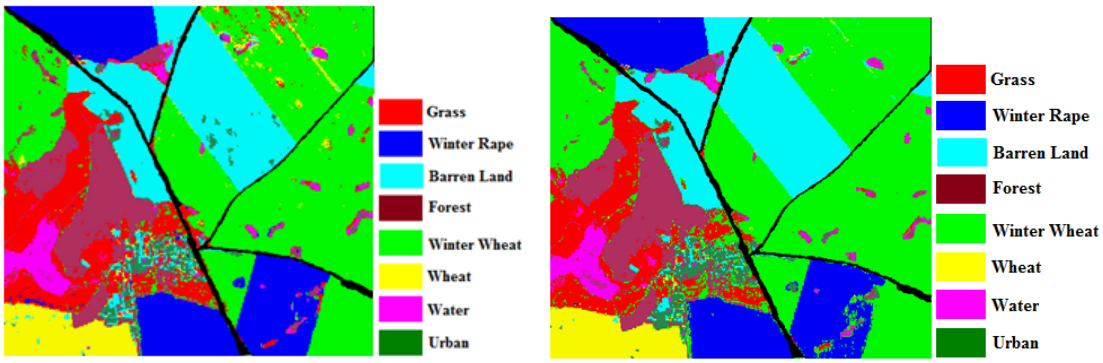
The performance of proposed DCS changed considerably with different nontrainable combination functions. When compared with the best classifier and dimensionality reduction pair (see Table 5.2), there has been marginal to significant increase in the overall accuracy. Table 5.3 shows the classification accuracy obtained by the DCS combination function, when the optimal dimension of the dimensionality reduction methods are determined by the classification accuracy of the training samples. The classification accuracy increased by 4.09% and 5.22% with average rule and majority voting rule for HyMAP and HYDICE hyperspectral images, respectively. For ROSIS University, ProSpecTIR, and ROSIS City of Pavia hyperspectral images, the classification accuracy improved by 5.46%, 5.42%, and 3.59% with product rule, respectively.

Table 5.3: Classification results from DCS (optimal dimension of the dimensionality reduction method was estimated based on training samples classification): Overall accuracy (OA, in %) and Kappa coefficient (KC) of the classified image obtained by combination of the optimal pairs of classifiers and dimensionality reduction methods for various combination functions.

Hyperspectral Image		Non-trainable combination scheme						Trainable combination scheme	
		MV	Avg	Max	Min	Med	Prod	MLC	SVM
HyMAP	OA	96.33	97.34	92.15	93.92	96.14	97.28	97.81	99.85
	KC	*0.958*	*0.969*	*0.903*	*0.932*	*0.953*	*0.964*	*0.972*	*0.998*
ROSIS University	OA	86.65	88.16	84.91	81.94	85.91	88.24	89.45	92.37
	KC	*0.848*	*0.863*	*0.829*	*0.792*	*0.839*	*0.865*	*0.877*	*0.909*
ProSpecTIR	OA	95	95.72	91.28	90.99	92.83	95.73	95.63	96.74
	KC	0.942	0.95	0.898	0.891	0.916	0.95	*0.949*	*0.962*
ROSIS City of Pavia	OA	93.41	93.61	92.32	90.02	93.72	93.64	93.11	95.78
	KC	*0.924*	*0.926*	*0.912*	*0.885*	*0.928*	*0.927*	*0.919*	*0.951*
HYDICE	OA	96.95	94.09	92.26	96.52	96.43	94.43	98.52	96.72
	KC	*0.963*	*0.927*	*0.905*	*0.957*	*0.956*	*0.931*	*0.976*	*0.959*

Table 5.4: Classification results from DCS (optimal dimension of dimensionality reduction method was estimated based on class separability measure): Overall accuracy (OA, in %) and Kappa coefficient (KC) of the classified image obtained by combination of the optimal pairs of classifiers and dimensionality reduction methods for various combination functions.

Hyperspectral Image		Non-trainable combination scheme						Trainable combination scheme	
		MV	Avg	Max	Min	Med	Prod	MLC	SVM
HyMAP	OA	95.47	96.26	92.53	95.51	95.39	96.23	97.79	98.5
	KC	*0.944*	*0.952*	*0.907*	*0.946*	*0.939*	*0.957*	*0.969*	*0.982*
ROSIS University	OA	84.77	85.47	83.97	80.36	82.12	85.39	88.3	90.56
	KC	*0.825*	*0.833*	*0.818*	*0.774*	*0.793*	*0.832*	*0.865*	*0.892*
ProSpecTIR	OA	93.4	94.96	92.14	93.39	94.96	94.91	95.52	97.28
	KC	*0.923*	*0.941*	*0.908*	*0.923*	*0.938*	*0.94*	*0.949*	*0.968*
ROSIS City of Pavia	OA	92.53	92.63	91.39	89.67	91.54	92.74	93.04	93.79
	KC	*0.914*	*0.915*	*0.901*	*0.881*	*0.903*	*0.916*	*0.92*	*0.928*
HYDICE	OA	97.58	93.83	91.8	92.51	96.14	94.11	97.99	97.61
	KC	*0.969*	*0.925*	*0.898*	*0.91*	*0.951*	*0.928*	*0.975*	*0.971*



(a) (b)

Figure 5.2: DCS-based classified images of HyMAP data with best nontrainable combination function (a) the optimal dimension of the dimensionality reduction methods was estimated based on training samples classification (b) the optimal dimension of the dimensionality reduction methods was estimated based on class separability measure.

The classification accuracy of the combination functions, when the optimal dimension of the dimensionality reduction methods was estimated by the class separability measure, is shown in Table 5.4. The magnitude of increase in classification accuracy is similar to Table 5.3 for ROSIS City of Pavia and HYDICE hyperspectral images.

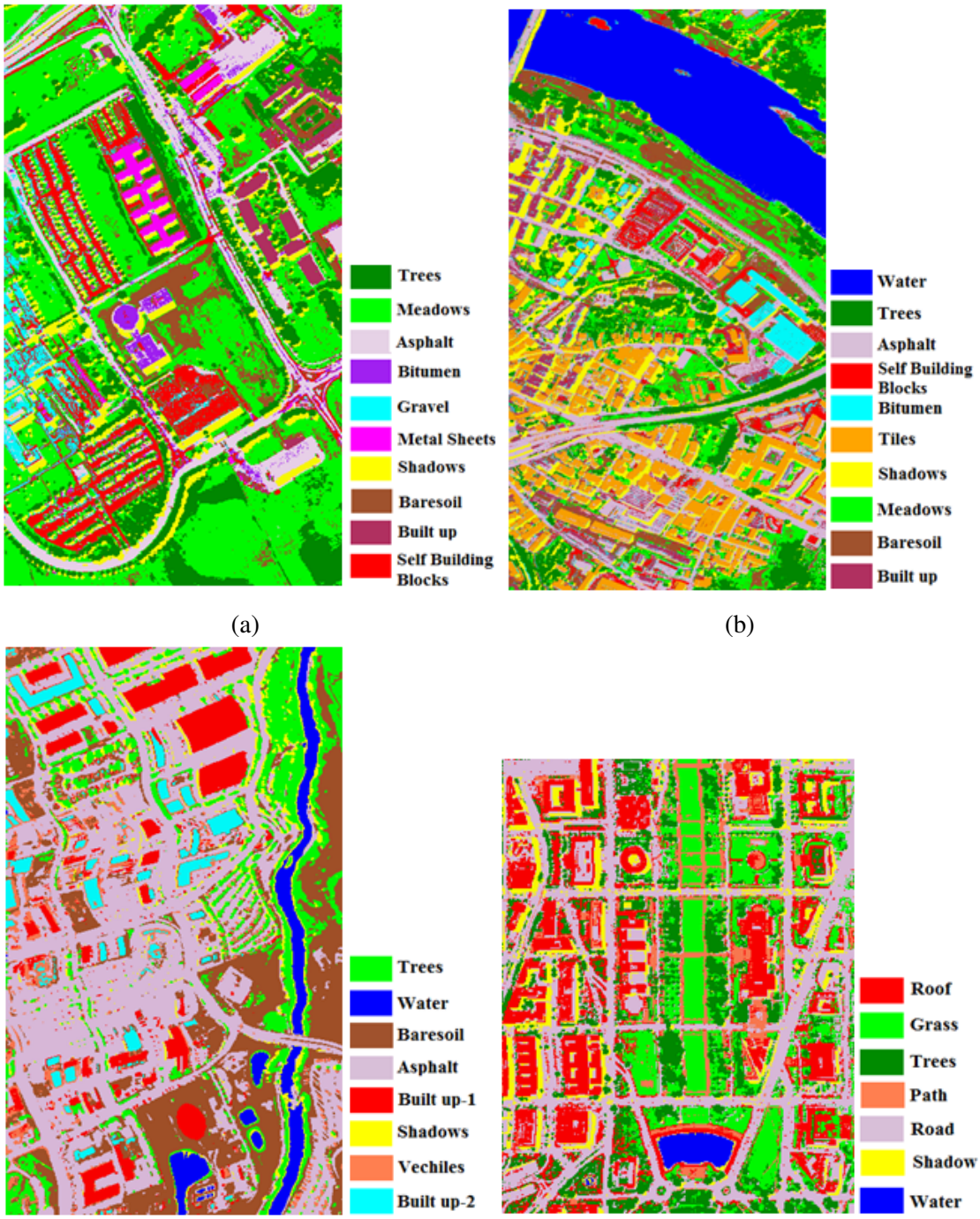


Figure 5.3: DCS-based classified images (best nontrainable combination function; the optimal dimension of the dimensionality reduction methods was estimated based on training samples classification): (a) ROSIS University, (b) ROSIS City of Pavia, (c) ProSpecTIR, and (d) HYDICE

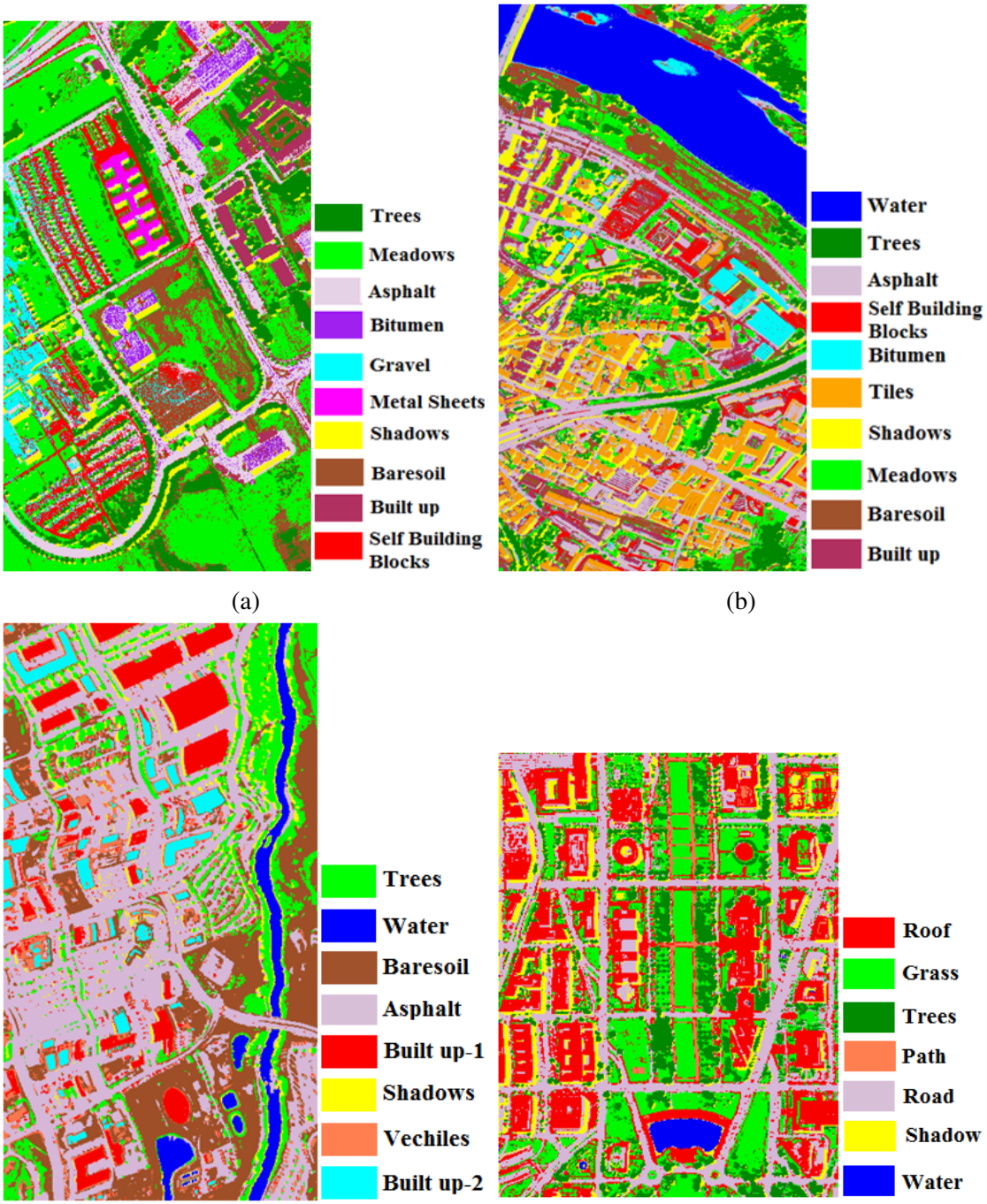


Figure 5.4: DCS-based classified images (best nontrainable combination function; the optimal dimension of the dimensionality reduction methods was estimated based on class separability measure): (a) ROSIS University, (b) ROSIS City of Pavia,(c) ProSpecTIR, and (d) HYDICE.

For remaining hyperspectral images, the increase in classification accuracy is less than one order of magnitude over the best classifier and dimensionality reduction pair. From Tables 5.3 and 5.4, it can also be inferred that there is more than one nontrainable combination functions which offer better classification accuracy with all the five hyperspectral images. Moreover, the product and average rule resulted in similar classification accuracy for all the five hyperspectral images. The DCS-based classification using maximum and minimum as the combination function resulted in accuracy either decreased or no change to the best classifier and dimensionality reduction pair for some of the hyperspectral images. This observation suggests the significance of the adopting appropriate combination function to fully exploit the potential of DCS.

5.4.3.2 DCS with trainable combination function

When the intermediate classified images in the DCS were combined with trainable combination functions, the classification accuracy increased further when compared with the results obtained from the non-trainable combination functions. From Table 5.3, the classification accuracy of the DCS (when the optimal dimension of the dimensionality reduction methods was estimated based on classification accuracy of training samples) with the SVM combination scheme is increased by 6.60%, 9.59%, 6.43%, 5.65%, and 4.99% for the HyMAP, ROSIS University, ProSpecTIR, ROSIS City of Pavia, and HYDICE hyperspectral images, respectively. With the MLC combination scheme, the classification accuracy has improved up to 6% for ROSIS University and HYDICE images and 2-4% for the remaining hyperspectral images.

The combination of classifiers selected by the DCS (when the optimal dimension of the dimensionality reduction methods was estimated based on class separability) in Table 5.2 (b) resulted in 3-9% improvement with the SVM combination scheme and 3-6% of improvement with the MLC combination scheme. Compared to Table 5.3, there is a similar magnitude of improvement in the classification accuracy for the three images and negligible magnitude difference (1%) for the remaining images. However, the classification accuracies resulted in Table 5.3 are higher for the four hyperspectral images compared to Table 5.4. Contrary to this, Table 5.4 results in higher classification accuracy for ProSpecTIR hyperspectral image.

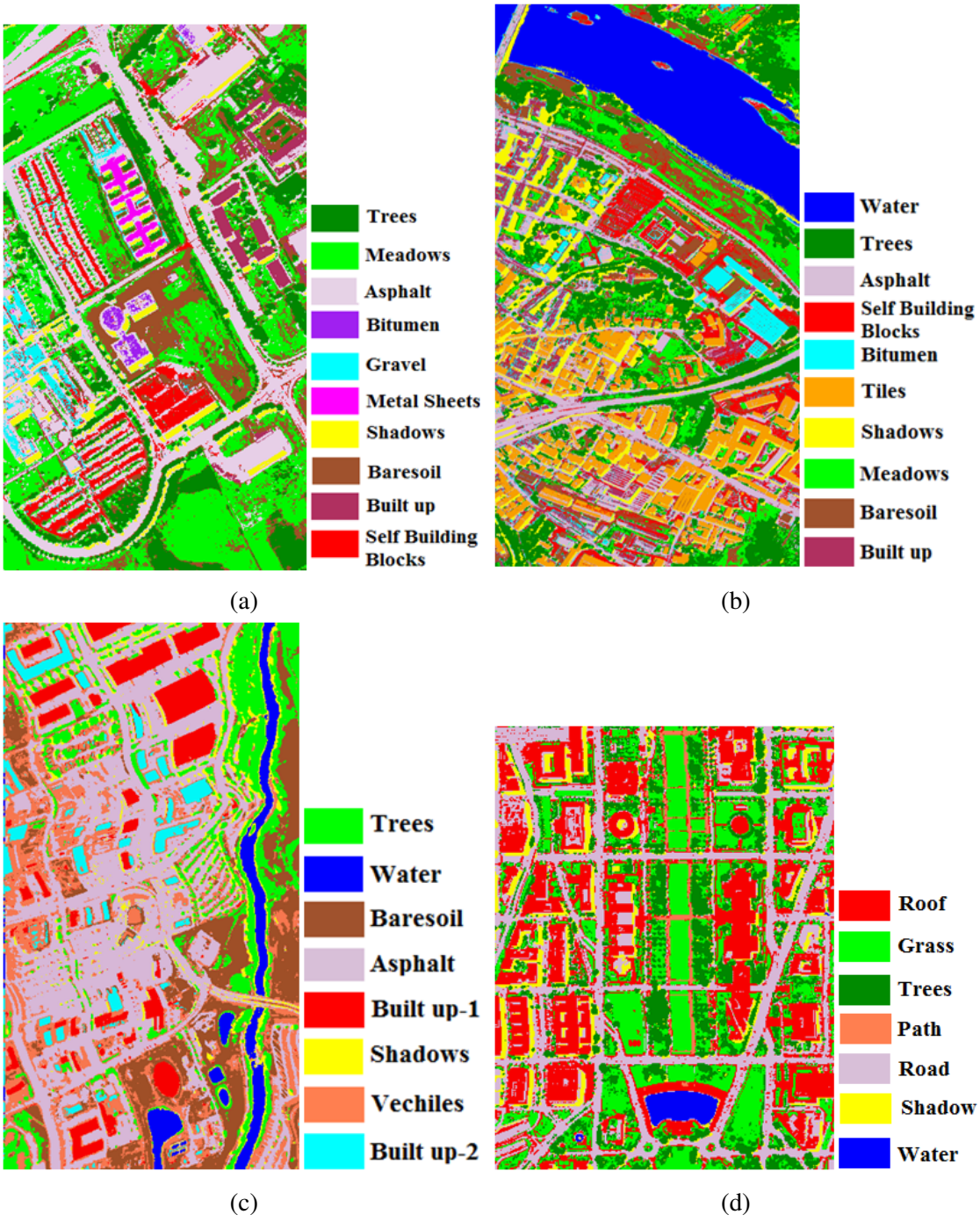
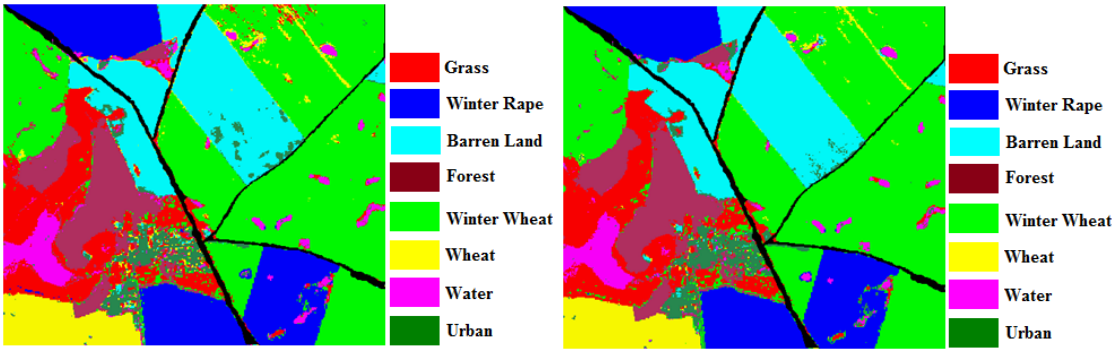


Figure 5.5: DCS-based classified images (best trainable combination function; the optimal dimension of the dimensionality reduction methods was estimated based on training samples classification): (a) ROSIS University, (b) ROSIS City of Pavia,(c) ProSpecTIR, and (d) HYDICE.



(a) (b)

Figure 5.6: DCS-based classified images of HyMAP data with best trainable combination function (a) the optimal dimension of the dimensionality reduction methods was estimated based on training samples classification (b) the optimal dimension of the dimensionality reduction methods was estimated based on class separability measure.

Table 5.5: Statistical significance test (Z-test and McNemar test (Chi-squared test)) between DCS and single best classifier and dimensionality reduction method pair for all the combination functions (optimal dimension of dimensionality reduction method was estimated based on training samples classification). The cases with statistically significant are highlighted in bold.

Kappa variance statistical significance test (Z-test)

Hyperspectral Image	Non-trainable combination function						Trainable combination function	
	MV	Avg	Max	Min	Med	Prod	MLC	SVM
HyMAP	**11.18**	**16.03**	**3.95**	**2.44**	**9.94**	**15.87**	**17.17**	**23.09**
ROSIS University	**8.27**	**7.72**	**3.2**	0.12	**3.84**	**7.79**	**14.19**	**17.36**
ProSpecTIR	**5.21**	**7.09**	1.18	0.54	**3.19**	**7.14**	**6.84**	**10.01**
ROSIS City of Pavia	**6.39**	**7.04**	**5.91**	0.4	**7.55**	**7.15**	**5.26**	**9.58**
HYDICE	**17.06**	**5.4**	0.12	**13.03**	**12.74**	**6.21**	**18.87**	**13.12**

McNemar test (Chi-squared test)

Hyperspectral Image	Non-trainable combination function						Trainable combination function	
	MV	Avg	Max	Min	Med	Prod	MLC	SVM
HyMAP	**206.54**	**374.56**	21.17	**10.72**	**119.14**	**375.3**	**385.68**	**504**
ROSIS University	**129.89**	**86.42**	**10.09**	0.19	**15.56**	**89.65**	**297.99**	**379.7**
ProSpecTIR	**49.85**	**111.34**	**7.59**	0.52	**9.09**	**111.43**	**109.92**	**175.9**
ROSIS City of Pavia	**17.72**	**18.18**	**10.03**	1.23	**20.62**	**18.36**	**14.85**	**96.25**
HYDICE	**497.46**	**69.76**	0.16	**376.49**	**279.62**	**86.56**	**556.44**	**285.8**

Table 5.6: Statistical significance test (Z-test and McNemar test (Chi-squared test)) between DCS and single best classifier and dimensionality reduction method pair for all the combination functions (optimal dimension of dimensionality reduction method was estimated based on class separability measure). The cases with statistically significant are highlighted in bold.

Kappa variance statistical significance test (Z-test)								
Hyperspectral Image	Non-trainable combination function						Trainable combination function	
	MV	Avg	Max	Min	Med	Prod	MLC	SVM
HyMAP	**10.91**	**14.5**	0.16	**11.1**	**10.6**	**14.37**	**18.38**	**21.21**
ROSIS University	**5.71**	**6.1**	**2.55**	1.43	1.45	**5.94**	**14.59**	**16.95**
ProSpecTIR	**3.86**	**6.87**	**2.18**	**3.54**	**6.93**	**6.27**	**7.96**	**12.01**
ROSIS City of Pavia	**5.93**	**6.24**	**4.88**	0.87	**5.09**	**7.54**	**8.23**	**9.05**
HYDICE	**22.35**	**8.56**	0.73	1.82	**16.4**	**9.45**	**19.78**	**7.84**
McNemar test (Chi-squared test)								
Hyperspectral Image	Non-trainable combination function						Trainable combination function	
	MV	Avg	Max	Min	Med	Prod	MLC	SVM
HyMAP	**218.34**	**339.03**	2.29	**157**	**136**	**320.57**	**394.51**	**483.56**
ROSIS University	**109.89**	**76.42**	**9.72**	0.29	0.42	**89.65**	**301.29**	**363.41**
ProSpecTIR	**37.65**	**107.42**	**12.6**	**36.3**	**109**	**105.63**	**128.17**	**197.57**
ROSIS City of Pavia	**31.45**	**28.32**	**20.3**	1.98	**23**	**33.75**	**38.08**	**89.23**
HYDICE	**622.58**	**67.18**	1.92	3.09	**286**	**82.95**	**501.01**	**59.23**

The results of the statistical significance test are shown in Tables 5.5-5.7. Under the condition $Z > 1.96$ and $\chi^2 > 3.841$, the difference in accuracy is regarded as statistically significant at 95% confidence interval. The statistically significant cases are highlighted in bold. As evident from Table 5.5, and 5.6, the accuracy differences between the single best classifier/dimensionality reduction pair and the DCS are significant for most of the combination functions. A significant increase in the classification accuracy has been observed with all the five hyperspectral images. However, the accuracy differences amongst the non-trainable combination functions are marginal to moderate. When compared to non-trainable combination functions, the accuracy improvements by DCS with trainable combination function are consistently higher by magnitude and

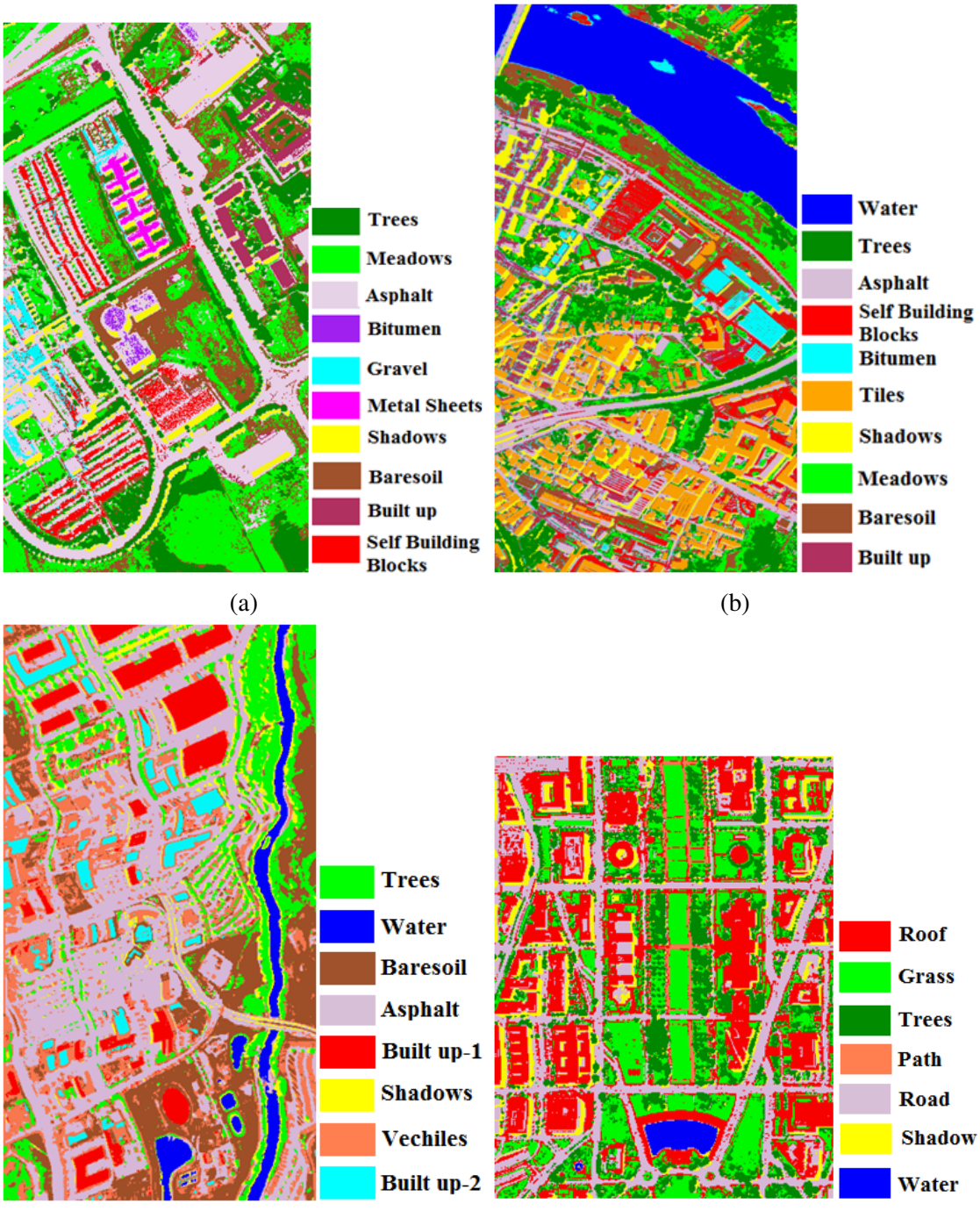


Figure 5.7: DCS-based classified images (best trainable combination function; the optimal dimension of the dimensionality reduction methods was estimated based on class separability measure): (a) ROSIS University, (b) ROSIS City of Pavia,(c) ProSpecTIR, and (d) HYDICE.

Table 5.7: Statistical significance test (Z-test and McNemar test (Chi-squared test)) between the best non trainable combination function and the trainable combination function of the DCS. The cases with statistically significant are highlighted in bold.

(a) Optimal dimension of DR method estimated based on training samples classification

Hyperspectral image	Kappa variance statistical significance test (Z-test)		McNemar test (Chi-squared test)	
	MLC	SVM	MLC	SVM
HyMAP	1.62	**15.02**	3.05	**234.46**
ROSIS University	**6.45**	**9.65**	**65.74**	**223.57**
ProSpecTIR	0.46	**2.95**	0.95	**18.09**
ROSIS City of Pavia	**1.99**	**10.5**	**11.93**	**158.66**
HYDICE	**6.95**	1.71	**72.34**	2.86

(b) Optimal dimension of DR method estimated based on class separability measure

Hyperspectral image	Kappa variance statistical significance test (Z-test)		McNemar test (Chi-squared test))	
	MLC	SVM	MLC	SVM
HyMAP	**7.27**	**11.38**	**98.88**	**159.62**
ROSIS University	**5.64**	**11.07**	**51.67**	**235.5**
ProSpecTIR	1.75	**6.69**	3.04	**81.33**
ROSIS City of Pavia	1.01	**3.36**	2.98	**40.24**
HYDICE	1.17	0.21	2.21	0.78

are statistically significant for four of the five hyperspectral images (see Table 5.7).

5.4.4 Comparison of classification performance of the proposed DCS with MCS and SVM

Table 5.8 shows the overall accuracy estimates obtained from SVM, MCS, and DCS for all the five images. For reference, the overall accuracy obtained with the best classifier/dimensionality reduction pair is also included. Table 5.8 (a) shows the accuracies offered by the DCS (the optimal dimension of the dimensionality reduction methods was estimated based on the training samples classification) when combining the selected classifier and dimensionality reduction pair in Table 5.2 (a).

Table 5.8: Classification accuracy (%) obtained with MCS (with best combination function), SVM (with and without dimensionality reduction (DR) method), the proposed DCS (with the best combination function), and the single best classifier/dimensionaliy reduction pair.

Hyperspectral image	DCS	MCS		SVM		Best pair
		With DR	Without DR	Without DR	With DR	
(a) Optimal dimension of DR method estimated based on training samples classification						
HyMAP	99.85	96.65	87.79	87.43	95.91	93.25
ROSIS University	92.37	85.21	82.44	78.87	87.98	82.78
ProSpecTIR	96.74	94.32	92.86	91.12	94.84	90.31
ROSIS City of Pavia	95.78	91.26	87.17	87.71	92.27	90.13
HYDICE	98.52	95.73	94.48	92.33	95.66	91.73
(b) Optimal dimension of DR method estimated based on class separability measure						
HyMAP	98.5	96.38	87.79	87.43	95.91	92.48
ROSIS University	90.5	84.33	82.44	78.87	87.98	81.23
ProSpecTIR	97.23	93.4	92.86	91.12	94.84	90.21
ROSIS City of Pavia	93.79	89.67	87.17	87.71	92.27	89.19
HYDICE	97.99	96.25	94.48	92.33	95.66	91.73

The DCS offered consistently higher accuracies when compared with the MCS or SVM methods, even though with relatively lower margin for HYDICE image. The DCS shows 3.2%, 7.16%, 2.42%, 4.52%, and 2.79% increase when compared with MCS and 3.94%, 4.39%, 1.9%, 3.51%, and 2.86% increase when compared with SVM for the HyMAP, ROSIS University, ProSpecTIR, ROSIS City of Pavia, and HYDICE image, respectively. When the pair of classifiers and dimensionality reduction method resulted in Table 5.2 (b) are combined, the DCS offers about 4-6% enhancement in classification accuracy over the MCS and about 2% improvement over the SVM classification. Further, the drastic enhancement in classification accuracy of DCS, an increase of about 11-13%, can be observed when compared with SVM and MCS classification without dimensionality reduction method in Table 5.8 (a) and (b). The comparison between Table 5.8 (a) and (b) shows that there is about 2% accuracy difference for HyMAP, ROSIS University, and ROSIS City of Pavia hyperspectral images and for the remaining images, the accuracies are comparable. Table 5.8 indicates that SVM's performance is comparable or better than the MCS when applied with dimensionality reduction: out

of the five images, SVM offered marginally higher accuracy for three images; whereas the MCS indicates marginal increase for two images. However, the performance of SVM and MCS varied significantly while implementing with and without applying dimensionality reduction. These observations conclude that DCS is a good candidate to produce diverse and adaptive classifiers in the MCS for hyperspectral image classification.

5.4.5 Computational complexity analysis

Table 5.9 shows the computational complexity analysis of the DCS and MCS. The MCS combines all the classifiers relative to each dimensionality reduction method, whereas the DCS combines only the identified optimal classifier relative to each DR method.

It is interesting to observe that the computation time of DCS is better than the MCS with single dimensionality reduction method and with all the five dimensionality reduction methods. This may be due to the computational time difference of the SVM combination function in MCS and DCS. In the MCS, all the classifiers are engaged in the combination function, which gets further complicated by the use of the SVM combination function. In the DCS, only a subset of the classifiers is engaged in the combination function, thereby compensating the computational complexity introduced by the dynamic selection criteria. The computation time mentioned in Table 5.9 is calculated for 10 runs and it includes both training and testing time. The experiments are performed with a typical desktop computer (4 GB RAM, Intel i-5 processor @3.20 GHz, 64-bit operating system), and it has been observed that the computational time for the selection of the optimal classifiers and DR methods in the DCS is $1s$. The computation time of DCS with optimal dimensionality estimated by training samples classification and class separability measures differed considerably with the class separability measure showing relatively better performance. However, the apparent computational comparisons are variable by the number of information classes in the hyperspectral image.

Table 5.9: Computation time (CPU time in sec) taken by the MCS with all the dimensionality reduction methods (DR), MCS with single dimensionality reduction method (DR), and the proposed DCS.

(a) Optimal dimension of DR method estimated based on training samples classification

Hyperspectral Image	Optimal dimension estimation	MCS with all the DR methods	MCS with single DR method	DCS
HyMAP	894.22	1928.56	1396.86	1376.24
ROSIS University	942.14	2073.49	1521.76	1502.54
ProSpecTIR	5537.69	6198.19	5885.29	5876.89
ROSIS City of Pavia	1325.73	2759.23	2105.53	2034.13
HYDICE	1207.68	1950.78	1559.28	1542.74

(b) Optimal dimension of DR method estimated based on class-separability measure

HyMAP	34.78	899.52	470.08	448.25
ROSIS University	45.96	1292.36	742.66	706.76
ProSpecTIR	272.94	1034.44	683.04	669.34
ROSIS City of Pavia	35.89	1421.89	752.89	714.19
HYDICE	51.11	1065.02	674.8	656.12

5.4.6 Diversity creation in the MCS with multiple dimensionality reduction methods

The statistical diversity measures, DM and KWM, were applied on the MCS to quantify the diversity existed in the MCS with and without dimensionality reduction methods. Table 5.10 shows the computed DM and KWM values for the various hyperspectral images. The DM and KWM values for each hyperspectral image without dimensionality reduction are the diversity existed in the classifiers considered for designing the MCS. As seen in Table 5.10, the diversity estimate increased considerably after the dimensionality reduction methods were applied. Further, diversity in MCS is considerably higher when the optimal dimension of the dimensionality reduction methods was determined by class separability measure [see Table 5.10 (b)]. The change in diversity values across different images indicates the inherent differences in the data acquired from different sources and environmental conditions.

The analysis of the diversity measures shows that there is a positive relationship between diversity and classification accuracy. For example, the DCS offered highest

Table 5.10: Diversity estimates of the MCS (with and without dimensionality reduction (DR) methods) and DCS for the hyperspectral images considered.

Hyperspectral image	Diversity measures of MCS				Diversity measure of DCS	
	DM		KWM		DM	KWM
	without	with DR	without DR	with DR		
(a) Optimal dimension of DR method estimated based on training samples classification						
HyMAP	0.1833	0.2051	0.0855	0.1008	0.112	0.0448
ROSIS University	0.3001	0.3621	0.1629	0.1886	0.1946	0.0778
ProSpecTIR	0.1523	0.1709	0.0563	0.0843	0.0822	0.0329
ROSIS City of Pavia	0.2307	0.2791	0.1263	0.1426	0.07973	0.03189
HYDICE	0.1696	0.2003	0.0838	0.0937	0.1228	0.0491
(b) Optimal dimension of DR method estimated based on class separability measure						
HyMAP	0.1833	0.2875	0.0855	0.1413	0.0937	0.0375
ROSIS University	0.3001	0.3798	0.1629	0.1867	0.1808	0.0723
ProSpecTIR	0.1523	0.3041	0.0563	0.1495	0.0859	0.0344
ROSIS City of Pavia	0.2307	0.3014	0.1263	0.1481	0.082	0.0328
HYDICE	0.1696	0.2873	0.0838	0.1412	0.1308	0.0523

increase in the accuracy for ROSIS University image for which the diversity measures show highest magnitudes.

5.5 Discussion

The classical hyperspectral image classification techniques are exposed with new challenges in terms of data dimensionality, limited training samples, and to cater to the needs of wider application domains (Jia et al., 2012; Braun et al., 2012; Hasanlou and Samadzadegan, 2012; Bioucas-Dias et al., 2013). Consequently, a number of dimensionality reduction methods and classifiers have been developed. However, there is no classifier or dimensionality reduction method which is optimal across hyperspectral data sources and application domains. In general, the appropriate classifier and dimensionality reduction method are identified before hand by analyst's prior knowledge or on heuristic basis, thus making the results being subjective and the procedure being expert dependent. There has been an increasing interest in the development of methods

for efficient classification of hyperspectral image. An MCS is viewed as one of the effective methodologies to improve hyperspectral image classification performance (Du et al., 2012b; Samiappan et al., 2013; Thoonen et al., 2012; Yang et al., 2010a). During the last two decades, MCS has developed significantly by theories and empirical studies and has been used widely in various applications of pattern recognition. The necessity of generating multiple transformations of image and selection of appropriate classifier from a pool of classifiers for hyperspectral image classification can be easily handled by the MCS architecture. However, the success of MCS depends upon the diversity of classifiers' performance (Ceamanos et al., 2010; Yang et al., 2010b). In principle, different variants of a hyperspectral image can be generated by using different dimensionality reduction methods and the set of all these images can then be exploited for diversity requirement in the MCS. Our results indicate that the diversity in performance of the classifiers in the MCS can be generated by applying multiple dimensionality reduction methods on the input data (see Table 5.10).

Identifying classifiers adaptive to different hyperspectral images avoids the negative impact of the presence of bad classifiers, thereby increasing the classification performance of the MCS. In this chapter, we have proposed and demonstrated experimentally an algorithmic extension of the MCS framework, named as DCS, to automatically identify classifiers optimal to the type of hyperspectral image and the distribution of land cover classes. The proposed DCS pairs up optimal classifiers and dimensionality reduction method and classifies hyperspectral image within the MCS framework using only the selected set of classifiers. The algorithm thus picks up a classifier which is optimal at the overall image level, relative to a dimensionality reduction method. The proposed DCS algorithm has been tested with experiments on five multi-site airborne hyperspectral images of different land cover and environmental settings. The classification performance of the DCS has been compared with MCS and SVM.

Compared to the accuracy estimates obtained with the MCS, SVM, and the typical supervised classification using the best classifier/dimensionality reduction pair, there has been 4-10% increase in the overall classification accuracy. This is an important improvement, given the expectation of higher accuracy estimates from the hyperspectral images and the importance of increase at the higher end of accuracy line. However, it can be observed that for some hyperspectral images (e.g., HYDICE), this increase is marginal by overall accuracy. This apparent increase in the classification accuracy

indicates functional relevance of the optimal pairs of classifiers and dimensionality reduction methods and the prospects of creating diversity in the classifiers performance by feeding data from different dimensionality reduction methods in the MCS by the proposed DCS. Interestingly, the SVM has shown relatively better performance over the MCS when applied with dimensionality reduction. However, its performance is comparable with that of the best classifier/dimensionality reduction pair when applied without dimensionality reduction. Apart from the appropriate classifiers and dimensionality reduction methods, the type of combination function used has significant bearing on the performance of the DCS, akin to the MCS. We observed the overall classification accuracy of all the five hyperspectral images increased by marginal to significant magnitudes when the classifiers are combined using trainable combination functions. Since the classifiers and the dimensionality reduction methods are heterogeneous, the resulting decision function values of the classifiers contain extreme differences in magnitude and direction with reference to the decision function values. Our observation of better overall accuracies for the different hyperspectral images indicates the superior capability of trainable combination functions (SVM) over non-trainable combination functions.

As the performance of the classifier significantly depends on the optimal dimension of the dimensionality reduction methods, we used two approaches: accuracy estimates from training samples classification and extended JM distance class separability measure for estimating the optimal dimension. Although there has been only marginal difference in the classification accuracy, the computational time is significantly lower for the optimal dimensionality estimation using the class separability measure. The proposed DCS algorithm could be used with any combination of classifiers, dimensionality reduction methods and hyperspectral data. Depending upon the complexity of the information classes and the combination function, the DCS algorithm can offer classification performance better than the MCS and SVM. The DCS shows comparatively equal or better computational performance to the MCS, because the computational complexity of the SVM combination function is higher with the increasing number of the classifiers.

5.6 Chapter Conclusions

In this chapter, we proposed and implemented a modified version of the MCS framework, named as dynamic classifier system, to automatically select and execute pairs of optimal classifiers and dimensionality reduction methods for the hyperspectral image classification. We implemented the proposed approach by classifying five different airborne hyperspectral images. The major conclusions of this chapter are:

- the proposed dynamic classifier system is effective in providing diverse and accurate classifiers for MCS based classification
- the improvement in the classification accuracy of the proposed system outperforms the state-of-the-art methods considered
- deployment of multiple dimensionality reduction methods may be considered as alternative to the current practice of depending upon diversity created by differential error estimates of the classifiers in MCS.

CHAPTER 6

DYNAMIC CLASSIFIER SELECTION APPROACHES FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Prelude: Continuing our work on the algorithmic developments for incorporating dynamism in the multiple classifier system presented in the Chapter 5, this chapter presents one more improved and innovative algorithmic development for dynamic classifier selection for hyperspectral image classification. The approach presented in this chapter is different to that presented in the Chapter 5 in the sense that the present approach also incorporates spatial contextual information apart from the spectral information. To incorporate spatial contextual information, extreme learning machine based dynamic classifier selection approach is regularized with Markov Random Field model. The experimental results are compared with the state-of-the-art methods.

6.1 Introduction

Multiple classifier system (MCS) has been recently explored to improve the performance of hyperspectral image classification by combining the predictions of multiple classifiers (Benediktsson and Kanellopoulos, 1999; Santos et al., 2013; Samiappan et al., 2013), thereby reducing the dependence on the performance of a single classifier. For the MCS to perform better than the single best classifier, the classifiers used in the MCS construction have to be diverse, because combining similar classification results may not improve accuracy (Kuncheva and Whitaker, 2003).

This chapter will be published in *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing,* with the title: " Dynamic Classifier Selection Approach for Hyperspectral Image Classification with Joint Spectral and Spatial Information". Authors: Bharath Bhushan Damaodaran, Rama Rao Nidamanuri, Yuliya Tarabalka

Diversity in the MCS can be created explicitly and implicitly. Explicitly, the diversity in the MCS is created by defining a diversity measure and optimizing it. Implicitly, diversity can be introduced by selecting a subset of features (Tin Kam Ho, 1998; Yang et al., 2010b; Ceamanos et al., 2010), training samples manipulation (Fernandez-Redondo et al., 2004; Chan et al., 2012), selecting classifiers from different categories and different feature extraction methods (Bakos and Gamba, 2011; Damodaran and Nidamanuri, 2014). However, the diversity constraint alone does not guarantee that the MCS always performs better. The possibility of inaccurate base classifiers and the incompatible combinations of the classifiers may instead end up the MCS with the suboptimal performance. An ensemble pruning approach has been proposed to select reasonably accurate base classifiers (Zhang and Zhu, 2011; Yi et al., 2006; Zhou et al., 2002). In this method, instead of combining all the available base classifiers in the MCS, a subset of classifiers is selected based on the criteria like diversity measures, performance measures for decision fusion. This approach has been expanded by proposing a unified framework (Damodaran and Nidamanuri, 2014) which consists of both diversity creation (implicit and explicit) and performance measures of base classifiers, as well as selecting the classifiers with non-zero weights by sparse optimization methods (Gurram and Kwon, 2013) to form an effective MCS. However, the selection of classifiers in this method is independent of the location of the image pixel in the feature space; hence all the classifiers take part in classifying each image pixel. On the other hand, the optimal subset of classifiers varies for different spatial locations in the image. Therefore, the performance of the MCS can be improved by selecting the best subset of classifiers dynamically relative to each image pixel, known as dynamic classifier/ensemble selection (DCS/DES) (Kirchhoff and Bilmes, 1999; Singh et al., 2008; Didaci et al., 2005).

The success of DCS depends on accurate estimation of classifiers competence for a given image pixel. The classifiers forming a dynamic subset are chosen based on estimating the accuracy (competence) of the each base classifier in a local region around the image pixel, and the selection is based on the highest accuracy criterion. The classifier competence can be computed by local accuracy (LA) estimation methods (Didaci et al., 2005; Woods et al., 1997; Smits, 2002) and probabilistic model based methods (Woloszynski and Kurzynski, 2011). Recently, Du et al. (2012a) have studied the capability of DCS for hyperspectral image classification using local accuracy estimation. However, their study is limited to two optimal classifiers as the base classifiers. When

the number of base classifiers increases and in the presence of inaccurate classifiers the performance of DCS based on LA is uncertain. Both the LA estimation methods and a probabilistic model approach compute the distance between the test pixel and the training (validation) samples. Therefore, this results in computational expensive for large scale problems. The regression based probabilistic model approach reduces the computational factor at the cost of the accuracy (Woloszynski and Kurzynski, 2011). Hence it is highly desirable to develop an accurate and computationally efficient DCS methodology for hyperspectral image classification.

Apart from spectral content, airborne hyperspectral sensors also provide rich spatial information (Tarabalka et al., 2009; Lixia Yang et al., 2014; Fauvel et al., 2012). Markov random field (MRF) model is a powerful method for modelling the spatial contextual information, which assumes that the neighbouring pixels are likely to belong to the same class (Fauvel et al., 2013; Tarabalka et al., 2010; Ghamisi et al., 2014). Most of the state-of-the-art studies deal with MRF regularization for a single classifier, yielding significant improvement in the classification performance (Tarabalka et al., 2010; Zhang et al., 2011; Bai et al., 2013; Aghighi et al., 2014; Li, 2014; Li et al., 2014). There are few studies which tested the application of an MRF model to MCS-based image classification (Junshi Xia, Peijun Du, 2013; Khodadadzadeh et al., 2014). Furthermore, DCS emerges as a strong candidate to capture the spectral information of a hyperspectral image, since it overcomes the structural limitation of a single classifier, as well as classifier combination or classifier fusion methods. Hence, the spectral information-based DCS enhanced with rich spatial information can be a sound classification methodology for hyperspectral image classification.

The main contributions of this chapter are the following: We propose an extended version of the probabilistic model DCS based on extreme learning machine (ELM) approach. The new DCS-ELM method is sensitive to local variations in an image during the ensemble classifier selection stage. Further, we propose a new unified framework to exploit both spectral and spatial information for hyperspectral image classification.

The flow-chart of the spectral-spatial DCS method is shown in Figure 6.1. The proposed framework extracts the spectral information by using the DCS framework and the spatial information by applying the MRF model. Experiments have been conducted with two multi-site airborne hyperspectral images, and results showed that proposed

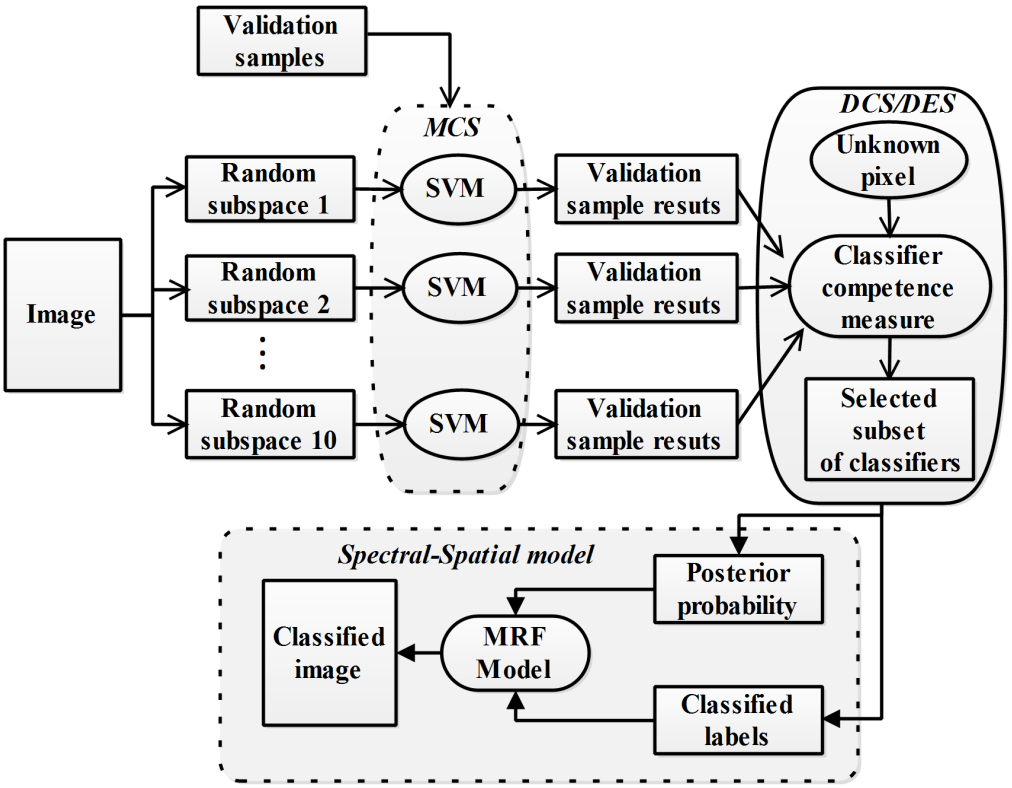


Figure 6.1: Flowchart of the dynamic classifier selection with joint spectral and spatial approach.

spectral-spatial DCS method yielded the improved classification accuracies when compared to the previously proposed techniques.

The remainder of the chapter is organized as follows. In the Section 6.2, the dynamic classifier selection approaches are introduced for hyperspectral image classification. Then, the proposed dynamic classifier approach with joint spectral and spatial information is described. The experimental results of dynamic classifier approaches are presented in the Section 6.3 and it is discussed in Section 6.4. In the last section, the conclusions are drawn.

6.2 Methodology

6.2.1 Multiple classifier system

Let $\Psi = \{\psi_1, \psi_2, \ldots, \psi_L\}$ be the base classifiers forming a MCS, and each classifier $\psi_l, l = 1, 2, \ldots, L$ be a function $\psi_l : \boldsymbol{\chi} \rightarrow \Omega$ from an input space $\boldsymbol{\chi} \subseteq R^n$ to a set

of class labels $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ (M is the number of classes). For any given $\mathbf{x} \in \chi$, classifier ψ_l produces a vector of decision values $\mathbf{d} = [d_{l1}, d_{l2}, \ldots, d_{lM}]$ and $\mathbf{x}$ is assigned to the class which has the maximum probability (decision) value. The base classifiers have to commit different types of errors in their predications on different parts of the input space, so that the MCS produces more accurate results when compared to individual classifiers. The random subspace method (RSM) is a popular ensemble generation technique to generate multiple input data sources from a single input data, thus creating diversity among the classifiers in a MCS (Ceamanos et al., 2010).

The RSM partitions the hyperspectral image bands into L subsets and each subset contains $\frac{P}{L}$ number of bands, where P denotes the number of bands in the original hyperspectral image. Each input data source generated from the RSM is returned as the input to the learning algorithm ψ. Support vector machines (SVM) have gained interest in hyperspectral image classification due to their ability to deal effectively with high-dimensional data and small training sets (Plaza et al., 2009; Camps-Valls and Bruzzone, 2005; Pal and Mather, 2005). The performance of SVM varies across different input data sources, thus introducing diversity in the MCS. Apart from SVM, RSM also has the capability to mitigate the small sample size problem and offers good classification accuracies in the heterogeneous environment. The SVM coupled with the output from the RSM were used as base classifiers in the MCS. The concept of combining all the base classifiers available in the MCS to obtain the final classified image, known as classifier combination or classifier fusion (CF) method were extensively explored. Therefore, here we focus on the dynamic classifier selection (DCS) approach, which dynamically selects a subset of classifiers for a given image pixel. In the following section, we describe the various dynamic classifier approaches that we propose to apply for hyperspectral image classification.

6.2.2 Dynamic classifier selection (DCS) approaches

The basic idea of the DCS is to find the classifier with the highest probability of being correct for a given unseen sample. The selection of correct classifiers and hence the success of the DCS depends upon the estimation of the classifiers competence for a given sample. The classifier competence measure is estimated from validation samples (different from training samples and testing samples). Apart from the training and testing

samples, validation samples are also generated for estimating the classifier competence in dynamic classifier selection. Let $\mathbf{V} = \{(\mathbf{v}_1, j_1), (\mathbf{v}_2, j_2), \ldots, (\mathbf{v}_N, j_N)\}$, be the validation set containing pairs of validation samples and their corresponding class labels. A brief description of different methods used to estimate the classifier competence is given below.

6.2.2.1 Dynamic classifier selection by local accuracy estimate (DCS-LA)

The DCS-LA estimates accuracy of each classifier in a local surrounding region of the image pixel to be classified in the validation samples, and selects the classifier that exhibits higher local accuracy. Let $\mathbf{x}$ be an image pixel to be classified and let us consider k-nearest neighbours of $\mathbf{x}$ in the validation set, denoted as $N(\mathbf{x}) \in \mathbf{V}$. Woods et al. (1997) introduced two strategies to measure the local accuracy: overall local accuracy and local class accuracy. In this work we used the second strategy, since it achieves better results both (Woods et al., 1997) in our experiments.

Now let $\mathbf{x}$ be the image pixel to be classified and without loss of generality we assume that the classifier ψ_l assigns class label ω_m to the image pixel $\mathbf{x}$ (i.e., $\psi_l(\mathbf{x}) = \omega_m$). Then the local accuracy (LA) of classifier ψ_l (LA is known as the classifier competence measure) is denoted by

$$LA(\psi_l, \mathbf{x}) = \frac{N_m}{\sum_{i=1}^{M} N_{im}}, N(\mathbf{x}) \in \mathbf{V} \mid \psi_l(\mathbf{v}_j) = \omega_m, j = 1, 2, \ldots, k \tag{6.1}$$

where N_m is the number of correctly classified samples by the classifier ψ_l to the class ω_m in the neighborhood $N(\mathbf{x})$, and $\sum_{i=1}^{M} N_{im}$ is the number of the k-nearest samples of $\mathbf{x}$ in $\mathbf{V}$ that have been assigned to the class ω_m by the classifier ψ_l. The classifier which exhibits highest local accuracy is selected as the adaptive classifier for image pixel $\mathbf{x}$ and the classifier selection is performed as

$$l = \arg\max_{i} LA(\psi_i, \mathbf{x}), i = 1, 2, \ldots, L \tag{6.2}$$

where l is the index of the classifier which provides highest local accuracy for image pixel $\mathbf{x}$.

However, in this approach all the neighbouring samples are given equal significance

and the classifiers probability values have not been considered. The validation samples which are closer to the image pixel may have more impact than the samples which are farther away. The posterior probability based DCS has been proposed to incorporate the classifiers probability values and weights based on the distance to the neighbouring samples to improve the estimation of local accuracy (Didaci et al., 2005). The posterior local accuracy (PLA) is estimated as follows

$$PLA(\psi_l, \mathbf{x}) = \frac{\sum_{\mathbf{v}_j \in \omega_m} P\left(\omega_m \mid \mathbf{v}_j, \psi_l\right) w_j}{\sum_{i=1}^{M} \sum_{\mathbf{v}_j \in \omega_i} P\left(\omega_m \mid \mathbf{v}_j, \psi_l\right) w_j}, N(\mathbf{x}) \in \mathbf{V} \mid \psi_l(\mathbf{v}_j) = \omega_m \quad (6.3)$$

where $\mathbf{v}_j \in N(\mathbf{x})$, $P\left(\omega_m \mid \mathbf{v}_j, \psi_l\right)$ is the posterior probability value of the validation sample $\mathbf{v}_j$ assigned to the class ω_m by the classifier ψ_l and $w_j = 1/d_j$, d_j is the Euclidean distance between the image pixels $\mathbf{x}$ and $\mathbf{v}_j$. The classifier which has the maximum PLA is selected for classifying the image pixel $\mathbf{x}$ similar to 6.2. In order to select a subset of T classifiers, the classifier competence values (PLA, LA) are arranged in descending order and the first T classifiers are selected. The classification process is then performed by using weighted Bayesian average methods (Kittler, 1998) and is called dynamic ensemble selection (DES):

$$P\left(\frac{\omega_i}{\mathbf{x}}\right) = \sum_{t=1}^{T} \eta_t p_t \left(\frac{\omega_i}{\mathbf{x}}\right), i = 1, 2, \ldots, M \quad (6.4)$$

The class label is obtained as

$$\mathbf{x} \in \omega_m, m = \arg\max_i P\left(\frac{\omega_i}{\mathbf{x}}\right)$$

where η_t is the weight of the classifier ψ_t (for instance, it is obtained as $\eta_t = PLA(\psi_t, \mathbf{x})$), and $p_t\left(\frac{\omega_i}{\mathbf{x}}\right)$ is the resulting posterior probability of class ω_i for a classifier ψ_t. For more details please see (Didaci et al., 2005; Woods et al., 1997).

6.2.2.2 Dynamic classifier selection with modified local accuracy (DCS-MLA)

This approach is similar to DCS-LA, except that the local accuracy is estimated using weighted nearest neighbours of the image pixel $\mathbf{x}$. Motivated by the performance of the distance weighted k-NN classifier, Smits (2002) used generalized Dudani's weighting scheme for scaling distance with the user specified parameter and sth nearest neighbour

for scaling the distances as

$$w_r(\mathbf{x}) = \begin{cases} \frac{d_k - d_r + \alpha(d_s - d_1)}{(1+\alpha)(d_s - d_1)}, & d_s \neq d_1, \alpha = 0, s = 3k \\ 1 & \text{otherwise} \end{cases} \tag{6.5}$$

where d_k is the distance between kth sample and the image pixel $\mathbf{x}$, and d_r is the distance between the rth sample of the kth nearest neighbour and the image pixel $\mathbf{x}$. Then the modified local accuracy (MLA) is estimated as

$$MLA\left(\psi_l, \mathbf{x}\right) = \frac{1}{k} \sum_{j \in N(\mathbf{x}) | \psi_l(\mathbf{x}) = \omega_m} w_j \tag{6.6}$$

The classifier which exhibits maximum local accuracy is chosen as the adaptive classifier to classify the image pixel and if a subset of classifiers are selected then classification is performed similar to 6.4 with $\eta_t = MLA\left(\psi_t, \mathbf{x}\right)$. For more details about the DCS-MLA, please refer to (Smits, 2002). Furthermore, we modified the 6.6 by incorporating classifiers posterior probability values of the neighbouring samples for better local accuracy estimation:

$$MPLA\left(\psi_l, \mathbf{x}\right) = \frac{\sum_{\mathbf{v}_j \in \omega_m} P\left(\omega_m \mid \mathbf{v}_j, \psi_l\right) w_j}{\sum_{i=1}^{M} \sum_{\mathbf{v}_j \in \omega_i} P\left(\omega_m \mid \mathbf{v}_j, \psi_l\right) w_j}, N(\mathbf{x}) \in \mathbf{V} \mid \psi_l\left(\mathbf{v}_j\right) = \omega_m \tag{6.7}$$

where w_j is the weight obtained from 6.5 and $P\left(\omega_m \mid \mathbf{v}_j, \psi_l\right)$ is the posterior probability value of the validation sample $\mathbf{v}_j$ assigned to the class ω_m.

6.2.2.3 DCS-Beta probabilistic model

The third employed method for estimating the classifier competence is based on the beta probabilistic model (Woloszynski and Kurzynski, 2011). The classifier competence is modelled as the probability of correct classification of a random reference classifier (RRC). The random reference classifier produces a randomised vector of class supports such that its expected value is equal to the vector of class supports produced by the classifier ψ_l for each of the samples $\mathbf{v}_j, j = 1, 2, \ldots, N$ in the validation set. The RRC depends on the beta probability distribution with the parameters $\alpha_m, \beta_m, m = 1, \ldots, M$. The parameters α_m, β_m are derived from the vector of class supports produced by the classifier ψ_l.

Let ω_j be the original class label of the sample $\mathbf{v}_j \in \mathbf{V}$, and the classifier ψ_l produces a vector of class supports as $[d_1(\mathbf{v}_j), d_2(\mathbf{v}_j), \ldots, d_M(\mathbf{v}_j)]$. The estimation of classifier competence can be summarized as: (1) estimate the parameters of beta distribution as $\alpha_m = Md_m(\mathbf{v}_j)$, $\beta_m = M[1 - d_m(\mathbf{v}_j)]$ (2) construct the RRC and compute its conditional probability of correct classification as

$$P_c\left(RRC \mid \mathbf{v}_j\right) = \int_0^u b\left(u, \alpha_{mj}\left(\mathbf{v}_j\right), \beta_{mj}\left(\mathbf{v}_j\right)\right) \left[\prod_{m=1, m\neq\omega_j}^{M} B\left(u, \alpha_m\left(\mathbf{v}_j\right), \beta_m\left(\mathbf{v}_j\right)\right)\right] du \tag{6.8}$$

where $b\left(u, \alpha_{mj}(\mathbf{v}_j), \beta_{mj}(\mathbf{v}_j)\right)$ is the beta probability distribution and $B\left(u, \alpha_m(\mathbf{v}_j), \beta_m(\mathbf{v}_j)\right) = \int_0^u b\left(w, \alpha_m(\mathbf{v}_j), \beta_m(\mathbf{v}_j)\right) dw$ is the beta cumulative distribution function. The classifier competence (C) for each validation sample is returned as

$$C\left(\psi_l, \mathbf{v}_j\right) = P_c\left(RRC \mid \mathbf{v}_j\right), j = 1, 2, \ldots, N;\ l = 1, 2, \ldots, L \tag{6.9}$$

The classifier competence is computed for all the validation samples relative to each classifier, which essentially indicates which classifier is most suited for the validation samples. In order to choose the optimal classifier for a given image pixel, the classifier competence set is generalized to the entire feature space by the weighted sum of the competence set as follows:

$$c\left(\psi_l, \mathbf{x}\right) = \frac{\sum_{j=1}^{N} C\left(\psi_l, \mathbf{x}\right) \exp\left(-dist\left(\mathbf{x}, \mathbf{v}_j\right)^2\right)}{\sum_{j=1}^{N} \exp\left(-dist\left(\mathbf{x}, \mathbf{v}_j\right)^2\right)} \tag{6.10}$$

where $dist$ is the Euclidean distance between the image pixel $\mathbf{x}$ and the validation samples. The most competing classifier is selected as similar to 6.2 to classify the image pixel $\mathbf{x}$ and if the subset of competent classifiers is selected, then it is classified by using 6.4. The criterion 6.10 is known as potential model (DCS-Beta Potential). This method eliminates the necessity of finding nearest neighbours for each image pixel $\mathbf{x}$; instead it weights the validation samples which are closer to $\mathbf{x}$ with high weights and the validation samples which are farther away with low weights. However, it is required to compute N distances for each image pixel, yielding high computational complexity, especially when the size of the image is large.

In order to reduce the computational complexity, the classifier selection problem can

be formulated as the regression (classification) problem. Let us consider L classifiers as L classes, the objective is to learn a function that selects a classifier for each of the image pixels. In other words, the classifier selector is a function $f : \mathbf{V} \rightarrow C$, that maps from the validation data set to the competence set of validation samples.

Let $\{(\mathbf{v}_1, C(\psi_l, \mathbf{v}_1)), \ldots, (\mathbf{v}_j, C(\psi_l, \mathbf{v}_j))\}, l = 1, \ldots, L$ be the pairs of a validation sample and its corresponding classifier competence value of the classifier ψ_l. For simplicity, let $\mathbf{C}(\psi_l) = [C(\psi_l, \mathbf{v}_1), C(\psi_l, \mathbf{v}_2), \ldots, C(\psi_l, \mathbf{v}_N)]$, now

$$f(\mathbf{V}; \boldsymbol{\beta}_l) = \boldsymbol{\beta}_l^t \mathbf{V} \Rightarrow \boldsymbol{\beta}_l^t \mathbf{V} = \mathbf{C}(\psi_l) \tag{6.11}$$

where $\boldsymbol{\beta}_l$ is the parameter to be estimated for the classifier ψ_l. The classifier competence of the image pixel $\mathbf{x}$ can be obtained by

$$c(\psi_l, \mathbf{x}) = \boldsymbol{\beta}_l^t \phi(\mathbf{x}) \tag{6.12}$$

and the parameter $\boldsymbol{\beta}_l$ can be obtained by pseudo inverse as below

$$\boldsymbol{\beta}_l = \left(\Phi^t \Phi\right)^{-1} \Phi^t \mathbf{C}(\psi_l) \tag{6.13}$$

where $\Phi = [\phi(\mathbf{v}_1), \ldots, \phi(\mathbf{v}_n)]$ and $\phi(\mathbf{v})$ is the polynomial transformation of the sample $\mathbf{v}$ as $\sum_{i=0}^{r} \mathbf{v}^i, r = 2, 3, 5$. It is found that the transformation with r=3 has offered better results and we adopted the same. The classifier which has the maximum competence value in 6.12 is selected for classifying the image pixel $\mathbf{x}$. We call this method as the DCS-Beta Least square regression (DCS-Beta-LSR). When a subset of classifiers is considered, then the classification is performed by applying 6.4 and we call it as DES-Beta-LSR.

6.2.3 Proposed dynamic classifier selection approach with joint spectral and spatial information

6.2.3.1 Dynamic classifier selection based extreme learning machine regression

The performance of the DCS-Beta-LSR depends on the feature transformation of the validation and input samples. It is shown that the DCS-Beta-LSR has suboptimal performance compared to the DCS-Beta-Potential model (Woloszynski and Kurzynski, 2011). Hence it would be beneficial to have regression model which is independent of feature transformation and results in better performance. Recently extreme learning machine (ELM) has demonstrated its superior capability to offer better generalization ability and performance at extremely fast learning speed for classification and regression problems (Huang et al., 2012). In this paper, we propose ELM as an alternative regression model for the dynamic classifier selection, labelled as DCS-ELM. The result of DES-ELM is then regularized with MRF to incorporate spatial contextual information.

The ELM based regression can be modelled as

$$\sum_{i=1}^{R} \beta_i g_i(w_i, \mathbf{v}_j) = C(\psi_l, \mathbf{v}_j) \Rightarrow \mathbf{h}(\mathbf{v}_j)\boldsymbol{\beta} = C(\psi_l, \mathbf{v}_j), j = 1, 2, \ldots, N \tag{6.14}$$

where R is number of the hidden nodes, $\mathbf{h}(\mathbf{v}_j) = [g_1(w_1, \mathbf{v}_j), \ldots, g_R(w_1, \mathbf{v}_j)]$ is the output row vector of hidden layer for input $\mathbf{v}_j$, $g_i(w_i, \mathbf{v}_j)$ is the output of the ith hidden node and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_R]^t$ is the output weight between the hidden layer nodes and the output nodes. For all j, 6.14 can represented as

$$\mathbf{H}\boldsymbol{\beta}_l = \mathbf{C}(\psi_l) \tag{6.15}$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{v}_1) \\ \vdots \\ \mathbf{h}(\mathbf{v}_N) \end{bmatrix}$ and the parameter $\boldsymbol{\beta}_l$ can be obtained as

$$\boldsymbol{\beta}_l = (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{C}(\psi_l) \tag{6.16}$$

The classifier competence for a image pixel $\mathbf{x}$ is computed as

$$c(\psi_l, \mathbf{x}) = (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{C}(\psi_l)\mathbf{h}(\mathbf{x}) \tag{6.17}$$

The classifier competence are arranged in the descending order and first T classifiers are selected as the adaptive classifier for image pixel $\mathbf{x}$. Then the classification process is performed by computing the weighted Bayesian average 6.4. The class label is obtained as $\mathbf{x} \in \omega_m, m = \arg\max_i P\left(\frac{\omega_i}{\mathbf{x}}\right)$, where η_t is the weight of the classifier ψ_t (it is obtained as $\eta_t = c(\psi_t, \mathbf{x})$), and $p_t\left(\frac{\omega_i}{\mathbf{x}}\right)$ is the resulting posterior probability of class ω_i of the classifier ψ_t.

6.2.3.2 DCS-ELM with MRF regularization

Furthermore, the spatial contextual information is incorporated into the DES-ELM classification by using the MRF-based regularization model. In the MRF framework, the classification task is formulated as an energy minimization problem on the graph of image pixels. The energy to optimize is computed as a sum of spectral and spatial energy terms and assumes that a pixel belonging to a specific class tends to have neighbouring pixels belonging to the same class. The MRF model can be written as

$$\hat{\omega} = \arg\min_{\omega} \left(-\sum_{i \in S} \log P\left(\frac{\omega_i}{\mathbf{x}_i}\right) + \gamma \sum_{j \in N(\mathbf{x}_i)} (1 - \delta(\omega_i, \omega_j)) \right) \tag{6.18}$$

where $\delta(.)$ is the Kronecker function ($\delta(\omega_i, \omega_j) = 1$ for $\omega_i = \omega_j$; $\delta(\omega_i, \omega_j) = 0$ for $\omega_i \neq \omega_j$), $N(\mathbf{x}_i)$ is the neighbouring pixels of $\mathbf{x}_i$, $\hat{\omega}$ is the resulting class labels from the MRF regularization, S is the set of all image pixels and γ is a positive constant parameter that controls the importance of spatial smoothing. The first term $P\left(\frac{\omega_i}{\mathbf{x}_i}\right)$ characterizes the spectral information and it is derived from the DES-ELM by employing Equation 6.4. The second term is expressed by using a Potts model, which penalizes spatial transitions among neighbouring pixels with different class labels, thus resulting classification map of connected regions associated with similar land cover class (Moser et al., 2013). This MRF regularization can be solved using graph cut methods (Boykov et al., 2001).

6.3 Experimental Results

6.3.1 Hyperspectral image description

To study the potential of DCS/DES for hyperspectral image classification, we adopted two benchmark hyperspectral images with different land cover settings (one in urban area and one in agricultural area) captured by two different sensors (ROSIS, AVIRIS).

ROSIS University: The first hyperspectral dataset was collected over the University of Pavia, Italy by the ROSIS airborne hyperspectral sensor in the framework of HySens project managed by DLR (German national aerospace agency). The ROSIS sensor collects images in 115 spectral bands in the spectral range 0.43 to 0.86 μm with spatial resolution 1.3 m/pixel. After the removal of the noisy bands, 103 bands are selected for the experiments. The image contains 610 x 340 pixels with nine classes of interest. Figure 6.2 shows a false color composite (FCC) image and its corresponding ground truth map. The available reference samples in each class are shown Table 6.1.

AVIRIS Indian pines: The second hyperspectral image used was collected by the AVIRIS sensor over the Indian pines site in North Western Indiana. The AVIRIS sensor collects images in 220 spectral bands in the spectral range 0.43 to 0.86 μm at 20 m spatial resolution. Twenty water absorption bands were removed, and 200 bands were used for experiments. This image contains 145 x 145 pixels with sixteen classes of interest. Figure 6.5 shows the FCC image and its corresponding ground truth map. The available reference samples in each class are shown Table 6.2.

Table 6.1: Number of reference samples considered for the experiment of University image.

Class name	Training	Validation	Testing
1. Asphalt	100	100	6431
2. Meadows	100	100	18449
3. Gravel	100	100	1899
4. Trees	100	100	2864
5. Metal sheets	100	100	1145
6. Bare soil	100	100	4829
7. Bitumen	100	100	1130
8. Bricks	100	100	3482
9. Shadows	100	100	747
Total	900	900	40976

Table 6.2: Number of reference samples considered for the Indian pines image.

Class name	Training	Validation	Testing	Class name	Training	Validation	Testing
1. Alfalfa	12	11	23	9. Oats	5	5	10
2. Corn-no till	100	100	1228	10. Soybeans-no till	100	100	772
3. Corn-min till	100	100	630	11. Soybeans-min till	100	100	2255
4. Corn	60	59	118	12. Soybeans-clean till	100	100	393
5. Hay-windowed	100	100	283	13. Wheat	53	52	102
6. Grass/Trees	100	100	530	14. Woods	100	100	1065
7. Grass/pasture-mowed	7	7	14	15.Bldg-grass-trees-drives	100	100	186
8. Grass/pasture	100	100	278	16. Stone-steel towers	24	23	46
Total					1158	1158	7933

6.3.2 Design of experiments

From the available ground truth samples, we randomly selected 100 samples for training, 100 samples for validation and the remaining samples were used for testing. If the total number of available reference samples was lower than 300 samples per class, then 25% of samples were selected for training, another 25% of samples for validation and remaining samples were used as the testing samples (see Tables 6.1 and 6.2). The experimental results were assessed by overall accuracy (OA), average accuracy (AA), and producer accuracy (PA). In order to avoid bias induced by random sampling of the training and validation samples, ten independent Monte Carlo runs are performed and the accuracies (OA, AA, PA) are averaged over the ten runs.

In each of the RSM, multiclass pair-wise probabilistic SVM classification with the Gaussian radial basis function (RBF) kernel was performed (Chang, Chih-Chung and Lin, 2011). The SVM parameters in all our experiments were automatically tuned with $C = 2^{\alpha}, \alpha = \{-5, -4, \cdots, 15\}$ and $gamma = 2^{\beta}, \beta = \{-15, -13, \cdots, 3\}$ (C is the cost function and $gamma$ through five-fold cross validation strategy of the training samples. In the DCS using local accuracy estimation based methods (DCS-LA and DCS-MLA) the classifier competence are estimated based on both strategies 6.1 and 6.3 and the best results are retained. The performance of the DCS-LA and DCS-MLA approaches depends upon the k-nearest neighbours of the test sample in the validation

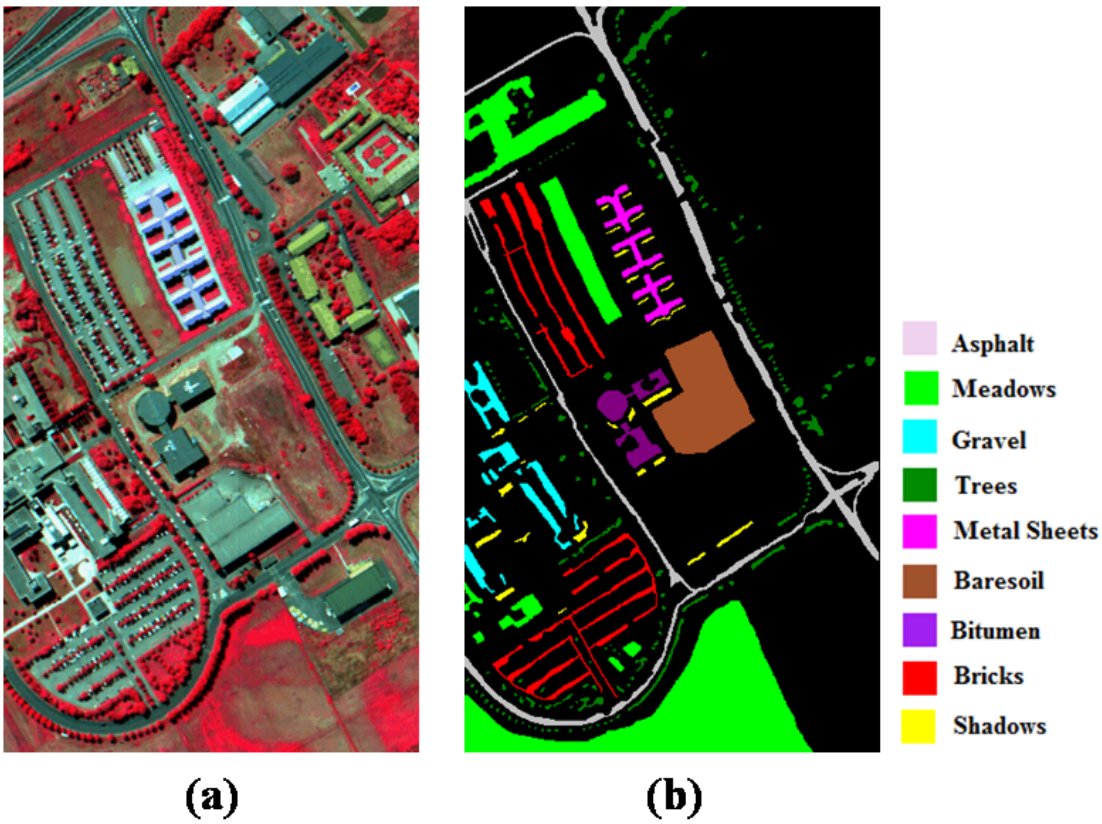


Figure 6.2: (a) False color composite of the ROSIS University image (R: 0.8340 μm G: 0.6500 μm B: 0.5500 μm), (b) Ground truth image and its corresponding class labels.

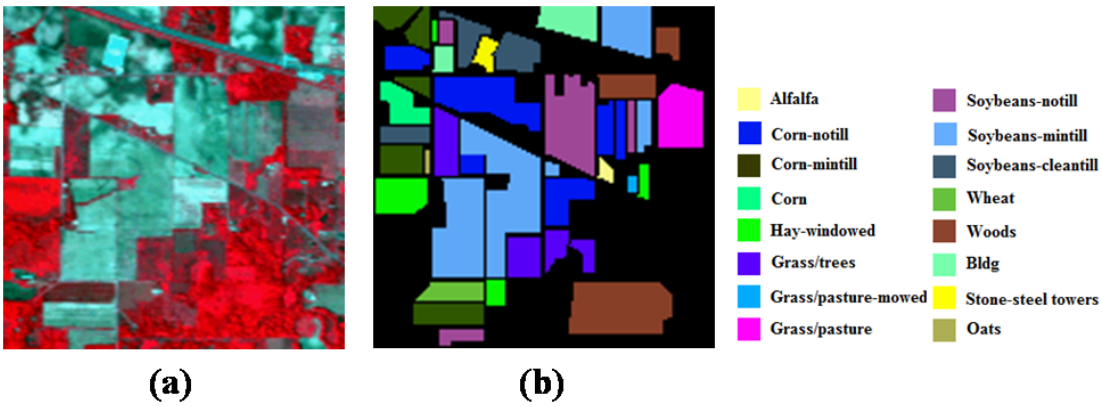


Figure 6.3: (a) False color composite of the AVIRIS Indian pines image (R: 0.8314 μm G: 0.6566 μm B: 0.5574 μm), (b) Ground truth image and its corresponding class labels.

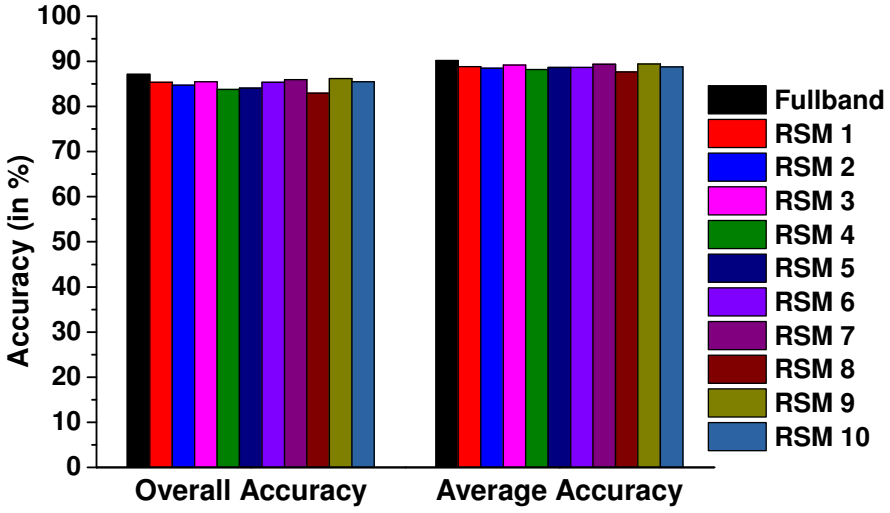
data set. Hence, the value of k was varied from 3 to 25 and only the best classification accuracies were retained. When more than one classifier was selected, the performance of the DES varies with the number of classifiers (T) included in 6.4. Hence, in the experiment we varied the number of classifiers from 2 to 7 and only the best accuracy is reported. However in most of the Monte Carlo runs, the optimal results are obtained with four and five classifiers.

Further, the classifier combination or classifier fusion (CF) method using Bayesian average method is adapted to combine all the base classifiers in the MCS (Kittler, 1998) to compare with the DCS approaches. The results of the proposed spectral-spatial DES (DES-ELM+MRF) were compared with the results of the full-band SVM classification, single best classifier, and CF. In order to have fair comparison we also performed MRF regularization with full band SVM, single best classifier and CF.

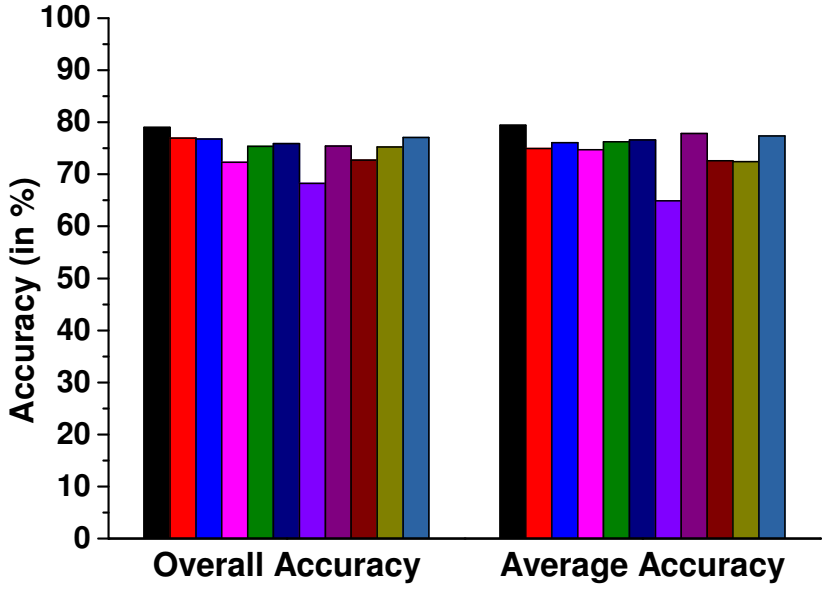
6.3.3 Classification results of RSM

Figure 6.4 shows the classification accuracies of SVM classification relative to each random subspace. The results show that there is a considerable variability among the base classifiers in the MCS in terms of overall accuracy and class-specific accuracies, thus indicating the suitability of RSM for forming the MCS.

The variability of classifiers accuracy is less significant for the University image. There is a 2% accuracy difference between the maximum and minimum overall classification accuracy in the MCS; whereas it is about 8% difference in the AVIRIS hyperspectral image. Similar observation also holds good with the average accuracy. In addition, the maximum classification accuracy (SB classifier) obtained with the MCS is comparable with the full band SVM classification accuracy, showing the inherent nature of RSM to sample uncorrelated bands in each subspace and introduces diversity in the MCS. Apart from sampling the uncorrelated bands, it also provides informative bands for classification.



(a)



(b)

Figure 6.4: Overall accuracy (OA) and Average accuracy (AA) of the SVM classifier relative to each random subspace and Full band hyperspectral image (a) ROSIS University (b) AVIRIS.

6.3.4 Classification results of the DCS and DES

In this section, the classification results of the different DCS/DES approaches and the proposed DES-ELM+MRF are presented. Tables 6.3 and 6.4 summarize the overall accuracies, average accuracies and class-specific accuracies of the dynamic classifier/ensemble selection for both hyperspectral images. DCS indicates that only the most competent classifier is selected for each image pixel; whereas DES indicates that a subset of classifiers was selected for each image pixel. In the case of the DCS, DCS-LA and DCS-ELM methods yield a marginal increase of overall classification accuracies. The remaining methods resulted in 2% decrease of classification accuracies when compared to the SB classifier for both images. This observation highlights the need for selecting an adaptive subset of classifiers for each image pixel, instead of selecting only the most competent classifier. This is analogous to the case of selecting multiple classifiers instead of one classifier to avoid the risk of the suboptimal performance.

Table 6.3: Producer accuracy (PA), overall accuracy (OA), and average accuracy of the dynamic classifier selection and dynamic ensemble selection methods of University image.

Class name	Dynamic classifier selection (DCS)						Dynamic ensemble selection (DES)					
	SB	LA	MLA	Beta-Pot	Beta-LSR	ELM	LA	MLA	Beta-Pot	Beta-LSR	ELM	ELM+MRF
Asphalt	77.66	77.09	79.53	78	78.65	78.31	78.39	80.5	79.51	79.97	79.94	97.17
Meadows	86.66	87.05	84.19	81.72	82.24	85.34	88.75	89.22	88.96	69.47	89.78	98.92
Gravel	82.47	86.91	81.6	83.22	82.09	83.4	87.04	86.22	86.29	86.12	86.6	97.3
Trees	94.54	96.31	91.43	92.12	94.14	94.78	96.04	94.79	95.17	78.72	95.51	95.1
Metal Sheets	99.48	99.66	99.48	99.55	99.59	99.62	99.62	99.57	99.58	99.48	99.6	99.84
Bare soil	87.48	89.26	85.93	84.23	86.33	88.26	91.29	90.64	90.22	88.26	91.83	99.85
Bitumen	93.12	94.7	91.85	92.8	92.63	93.43	94.87	94.11	94.38	94.27	94.4	98.38
bricks	83.28	79.39	81.38	83.18	81.84	82.02	84.56	85.4	85.75	85.23	85.87	98.59
Shadow	99.93	99.95	99.88	99.79	99.87	99.88	99.95	99.92	99.85	99.93	99.96	99.96
AA	89.4	90.04	88.36	88.29	88.6	89.45	91.17	91.15	91.08	86.83	91.5	98.35
OA	86.19	86.53	84.74	83.48	83.96	86.25	88.18	88.56	88.38	78.16	89.82	98.41

When a subset of classifiers is selected for each image pixel and combined by the weighted Bayesian average method, the classification accuracy improved significantly with all the DES approaches (except the DES-LSR method for the University image).

There is a significant improvement in classification accuracy (about 5-6 percentage points) for the Indian pines image, and a moderate improvement (about 2-3 percentage points) for the University image. Among the DES approaches, the DES-ELM achieved the highest accuracy for the University image and the DES-LA, DES-Potential and DES-ELM achieved the highest accuracies for the Indian pine image. The least performance could be observed with DES-LSR method. This is mainly due to its inefficiency in estimating the classifier competence of a test pixel. However, it is computationally efficient compared to the local accuracy estimation based methods.

Table 6.4: Producer accuracy (PA), overall accuracy (OA), and average accuracy of the dynamic classifier selection and dynamic ensemble selection methods of the Indian pines image.

Class name	Dynamic classifier selection (DCS)						Dynamic ensemble selection (DES)					
	SB	LA	MLA	Beta-Pot	Beta-LSR	ELM	LA	MLA	Beta-Pot	Beta-LSR	ELM	ELM+MRF
Alfalfa	62.17	59.13	61.74	63.48	51.3	48.7	56.96	55.65	67.39	50.87	58.7	81.74
Corn-no till	70.2	75.29	67.55	70.98	65.58	69.41	77.5	76.47	77.6	75.52	78.12	81.74
Corn-min till	74.06	75.46	70.33	71.51	71.89	72.89	78.22	78.33	79.52	79.38	79.49	94.05
Corn	79.92	85.17	76.95	74.75	76.44	80.76	88.47	86.69	86.86	86.53	87.37	99.07
Hay windowed	93.22	94.95	90.21	91.31	90.35	92.08	94.35	93.5	93.99	93.22	93.85	97.03
Grass/Trees	90.81	97.62	91.17	92.49	92.08	94.94	97.38	96.28	96.62	92.94	96.94	99.55
Grass/pasture-mow	62.86	63.57	47.14	56.43	57.14	50.71	52.14	46.43	65.71	57.86	62.86	82.14
Grass/pasture	99.1	99.39	98.24	98.67	98.96	99.17	99.53	99.5	99.42	99.17	99.46	100
Oats	32	5	2	29	44	39	10	12	39	36	35	38
Soybeans-no till	79.29	89	76.72	75.69	77.95	80.97	89.84	87.09	87.64	86.99	87.33	98.34
Soybeans-min till	67.01	61.66	62.16	64.85	65.46	65.31	72.36	72.79	73.54	73.43	70.36	91.17
Soybeans-clean till	80.84	85.06	77.89	81.93	78.98	83.66	91.12	90.03	90.59	89.59	88.8	98.37
wheat	97.64	99.22	98.43	98.14	97.35	98.04	99.41	99.41	99.51	99.02	99.31	99.51
Woods	87.5	92.07	87.29	87.17	87.88	89.49	93.46	92.66	92.92	91.46	93.16	94.19
Bldy	69.14	69.62	71.67	73.01	69.68	69.95	79.78	81.18	81.88	73.49	76.72	99.68
Stone-steel towers	92.39	96.74	87.17	87.39	89.35	93.26	95	92.61	92.17	93.7	93.7	98.91
AA	77.38	78.06	72.92	76.05	75.9	76.77	79.72	78.79	82.77	79.95	81.32	90.84
OA	77.06	78.82	74.37	76.03	75.45	77.16	83.23	82.65	83.4	82.18	83.2	93.12

Apart from the overall classification accuracy, the per-class accuracy and average

class accuracy has also been improved (see Table 6.3 and 6.4). There is about 7-8% improvement in per-class accuracy for most of the classes in the Indians pines image while it is moderate with the University image. This observation supports the need of adopting the adaptive classifiers based on local pixel information for enhanced classification performance. However the poor per-class accuracy is observed with the classes oats, and alfalfa. This is because the classifier fails to characterize the class information due to the presence of the insufficient number of training samples. Furthermore, our proposed DES-ELM approach has outperformed other DES methods both in terms of accuracies and computational time.

From the above observations, we can conclude that DES-ELM better characterizes the spectral information and provides reliable probability estimates and class labels when compared to the other considered methods. The inclusion of spatial contextual information in DES-ELM by the MRF model further significantly increases classification performance. In this case (DES-ELM+MRF), the overall and average classification accuracies are improved by 12-15% and by 9-12% over the SB classifier, respectively. When compared to its earlier version (DES-ELM), about 9% enhancement in the classification accuracy is observed. Furthermore, the class-specific accuracies exceed 95% for medium and large spatial structures, and are less than 95% for small spatial structures (for example, trees, alfalfa, oats). The lower per-class classification accuracy of oats, alfalfa might be due to insufficient number of training samples. The classification maps of the SB, DES-ELM and DES-ELM+MRF are shown in Figures 6.5 and 6.6. Visual inspection of Figures 6.5 (a), 6.5 (b), 6.6 (a), and 6.6 (b) reveal that DES produced smoother classification maps than the SB classifier. Figures 6.5 (c) and 6.6 (c) confirm a significant increase in classification accuracies and highlight the potential of the MRF model to produce a smooth classification map with spatially connected regions.

6.3.5 Comparative performance of the spectral-spatial DCS with classifier fusion and full band SVM

The accuracy of spectral-spatial DCS (DES-ELM+MRF) classification is compared with the classifier fusion (CF) method and full-band SVM. When compared with the full-band SVM, DES-ELM+MRF yields improvement of the overall accuracy by 11 and 14 percentage points for the University and the Indian Pines datasets, respectively.

Figure 6.5: Classified images of the University image (a) Single best (SB) classifier (b) DES-ELM (c) DES-ELM-MRF.

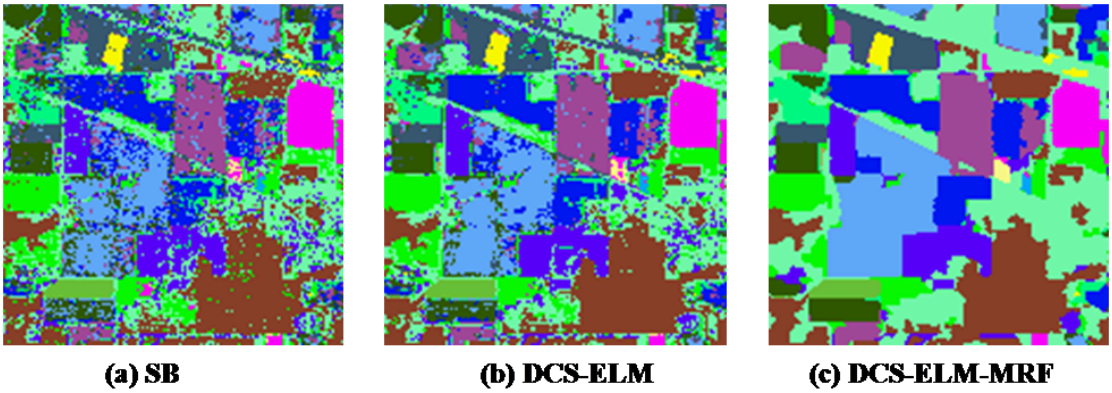


Figure 6.6: Classified images of the Indian pines image (a) Single best (SB) classifier (b) DES-ELM (c) DES-ELM-MRF.

When compared with CF, it yields about 9% improvement for both images. In order to have a fair comparison, we also performed MRF regularization with the full-band SVM, CF and SB classifier and the results are shown in Table 6.5. It can be seen that the accuracy of the DES-ELM+MRF classification outperforms SVM+MRF, SB+MRF and CF+MRF methods by 3% for the University image and 1.5% for the Indian pines image, respectively. The proposed DCS-ELM+MRF approach also yields the highest class-specific accuracies for most classes. This observation highlights the potential of merging the advantages of the two different approaches into a unified framework.

Table 6.5: Overall accuracy (OA), Average accuracy (AA) of the pixel wise classification methods (full band SVM, SB, and CF) and the full band SVM+MRF, SB+MRF, CF+MRF methods

Image		Full band SVM	SB	CF	Full band SVM+MRF	SB+ MRF	CF+ MRF	DES-ELM+ MRF
University	OA	87.12	86.19	88.65	95.41	94.73	96.47	98.41
	AA	90.16	89.4	91.31	95.51	95.55	96.14	98.35
AVIRIS	OA	79.04	77.06	83.22	90.91	90.5	91.5	93.12
	AA	79.43	77.38	81.34	85.86	84.74	86.56	90.84

Table 6.6: Kappa statistical significance test of different pixel wise classification methods and spatial contextual methods of University and Indian pines image. The results are considered as significant at 95% confidence interval if the tabulated value $|Z| > 1.96$.

(a) Kappa Statistical Significance Test of University Image

Methods	SB	DES-ELM	CF	Full band SVM	SB+ MRF	DES-ELM+ MRF	CF+MRF	Full band+ MRF
SB	0	15.09	10.52	7.59	58.71	67.67	64.09	57.41
DCS-ELM		0	6.58	8.3	37.34	42.73	32.95	35.97
CF			0	6.91	38.89	46.25	34.48	37.54
Fullband-SVM				0	40.75	65.96	58.25	51.48
SB+MRF					0	16.36	7.23	8.97
DCS-ELM+ MRF						0	8.34	15.22
CF+MRF							0	10.23

(b) Kappa Statistical Significance Test of Indian Pines Image

SB	0	9.6	9.62	3.16	25.24	28.74	26.93	26.36
DCS-ELM		0	0.23	6.65	15.73	19.29	17.44	16.87
CF			0	6.67	15.7	19.26	17.4	16.84
Fullband-SVM				0	22.29	25.8	23.97	23.41
SB+MRF					0	5.69	3.98	3.78
DCS-ELM+ MRF						0	2.98	3.94
CF+MRF							0	2.14

In order to examine the statistical significance of the results, we have conducted two-tail kappa statistical significance test at 95% confidence interval and the results are shown in Table 6.6. The results are statistically significant at 95% confidence interval, if the tabulated value $\mid Z \mid > 1.96$. As evident from Table 6.6, the accuracy differences of our proposed DES-ELM are statistically significant when compared with SB classifier, CF and full band SVM. However, there is no significant difference between DES-ELM and CF for the Indian pines image. The high statistical significance values are observed when the spatial contextual information is incorporated with the DES approach. This

observation confirms the advantage of the spatial contextual models to obtain accurate classification image over the pixel wise classification methods. Furthermore, it is interesting to see that accuracy improvement offered by our proposed DES-ELM+MRF approach is statistically significant when compared with other spatial contextual models (SB+MRF, CF+MRF and full band SVM+MRF).

6.4 Discussion

MCS has evolved as a promising approach for hyperspectral image classification. The classifiers in the MCS are combined in two ways by classifier fusion and classifier selection. Many studies have demonstrated that combining multiple classifiers (for instance, Bayesian average) has the potential to deliver significant performance for hyperspectral image classification (Samiappan et al., 2013; Thoonen et al., 2012). However, the classifiers forming the MCS have to be diverse in order to get enhanced performance; otherwise it may end up with suboptimal performance. It is understood that along with the diversity constraint, the classifiers forming the MCS also should be accurate enough to enrich the performance of the MCS. This requirement is often met by developing methodologies to select the classifiers by considering both diversity and performance measure. However, most often all the selected classifiers take part in the decision making and do not account for local class diversity and distribution variations within the image. Dynamic classifier selection is an alternative way of combining multiple classifiers in the MCS which selects a classifier (or a subset of classifiers) relative to each image pixel (Woloszynski and Kurzynski, 2011). Most of the previous studies using MCS for hyperspectral image classification are mainly focused on the classifier combination or classifier fusion, while little or no attention has been paid to the classifier selection mechanism. In this chapter, we explored the potential of the dynamic classifier selection (DCS) by proposing new DCS methods for hyperspectral image classification. Estimating classifier competence is the central part of the DCS algorithm. We explored both the local accuracy-based methods and a probabilistic model-based method (DCS-Beta Potential and DCS-Beta LSR) to estimate the classifier competence. The local accuracy-based methods and DCS-Beta potential computes the distance between each image (test) pixel and the whole set of validation samples, resulting in a computational burden. On the other hand, DCS-Beta LSR finds a function which maps the validation

data samples to the classifier competence and reduces the computational burden at the cost of accuracy. Hence, it will be beneficial to have an effective framework which reduces the computational burden without reducing the accuracy. We proposed the extreme learning machine (ELM) based regression framework, which estimates the function mapping validation samples to the classifier competence measure, thus reducing the computational burden without degrading the accuracy. Furthermore, ignoring the spatial correlation among the neighbouring pixels yields poor classification. Our proposed MRF-based framework combines both the spectral information from the DES-ELM and spatial contextual information, resulting in accurate and smooth classification maps.

We introduced diversity in the MCS by partitioning the hyperspectral image bands randomly using RSM with SVM as the base classifier. The classification results indicate that the best accuracy with the RSM is comparable to the full band SVM, highlighting significance of the RSM in providing uncorrelated as well as optimal bands for classification. The result of DCS establishes that selecting one best classifier for each image pixel is not an optimal choice and it could end up with the accuracy no better or less than SB for hyperspectral image classification. For example except DCS-LA and DCS-ELM, the remaining methods result in accuracy about 2% lower than the SB classifier for the University image. However, it has shown in (Du et al., 2012a) that DCS fared better for hyperspectral image classification. When the subset of classifiers is selected, DES offers 2-6% increase in overall classification accuracy. When compared to other considered methods, both DES-ELM and DES-LSR are computationally efficient for large scale image classification, but the latter method reduces the classification accuracy (Woloszynski and Kurzynski, 2011). Our proposed DES-ELM method has the inherent ability to transform the input data samples by hidden neuron nodes and it offers good computational and classification accuracy performance, as evident from the results. The dual advantage of the ELM method is that the performance is independent of its parameters (parameter free) and a variety of transformation functions can be used. However, in this study the best results are obtained with RBF-based transformation in the hidden neurons nodes (Huang et al., 2012). Further, ELM-based regression method shows better performance in both cases when one and a subset of classifiers is selected for each image pixel.

Compared to the single classifier, DES-ELM provides reliable probability estimates by alleviating the limitation of the single classifiers and the classifier fusion. The incor-

poration of the spatial contextual information shows remarkable performance of about 9% in OA over the pixel wise classification of DES-ELM. Compared to the pixel-based classification methods (full band SVM, SB, and CF), there is 9-16% increase in overall accuracy. Further, it is interesting to observe that DCS-ELM+MRF also have the potential to offer significant results in terms of OA and AA. It is noted that there is no significant accuracy difference between CF and DES-ELM for the Indian pines image, despite a significant accuracy difference between CF+MRF and DES-ELM+MRF. This indicates the superior capability of DES-ELM to better characterize the spectral information and provide reliable probability estimate to be used with MRF regularization than CF, and SB classifier. In addition, the experiments are performed with few training samples per class (around 5% of total reference samples for the University image and around 20% of total reference samples for the Indian pines image).

6.5 Chapter Conclusions

In this chapter, we explored the potential of dynamic classifier selection to exploit rich spectral information of hyperspectral images. Experimental results show that the DES has the potential to offer accurate classification results and can be considered alternative to the classifier fusion methods in the MCS. Furthermore, to improve both classification and computational performance, we proposed a new DCS method based on ELM regression and a new spectral-spatial classification framework to incorporate both spectral and spatial information for hyperspectral image classification. The proposed classification model shows impressive performance when compared to the state-of-the-art methods considered.

CHAPTER 7

SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS

Prelude: This chapter summarizes the overall observations and conclusions of the studies presented in the previous chapters, and important contributions of this thesis to the state-of-the art in hyperspectral image processing and analyses. Further, recommendations and directions of future research in this high impact area of remote sensing research are presented in this chapter.

This thesis has investigated some important issues involved in the application of a novel pattern recognition approach *multiple classifier system* for efficient hyperspectral image classification. In pursuance of the objectives of this thesis, we developed and implemented novel algorithmic schemes to make the multiple classifier system a reliable approach for hyperspectral image classification.

The main contributions of this thesis to the literature are: 1) an empirical analysis on the impact of dimensionality reduction methods on the classification performance of the multiple classifier system, 2) establishment of the existence of empirical relationships between classifier and dimensionality reduction method and between class, classifier and dimensionality reduction method within the framework of multiple classifier system, 3) demonstrated that transformed components from multiple dimensionality reduction methods can be used to sustain diversity in the functioning of base classifiers in the multiple classifier system, 4) a novel classifier system for automatically selecting input hyperspectral image adaptive classifiers and dimensionality reduction methods, 5) explored the potential of dynamic classifier selection approaches for hyperspectral image classification and developed two new dynamic classifier selection approaches for hyperspectral image classification. Of the two new approaches proposed, the first approach uses only spectral information, whereas the second approach is based on the spectral-spatial classification model to incorporate the spatial contextual information. These contributions will be valuable in devising efficient and generic methodologies

for the analyses of various sources of hyperspectral images for applications in environmental monitoring, and the natural resource management.

The summary of thesis is presented below chapter wise.

- Chapter 3: The impact of different dimensionality reduction methods on the performance of the multiple classifier system was studied. This study was conducted to understand the potential of using multiple dimensionality reduction methods to create the essential performance variability required among the base classifiers in the MCS and to understand how this variability influences the performance of MCS. The empirical results show that there is a significant variability in the performance of the individual base classifiers in the MCS and the level of variability is a function of the dimensionality reduction method. This indicates the significance of adopting dimensionality reduction methods to create differential performance in the classifiers for the MCS classification. Further, the magnitude of classification improvement of the MCS has a significant bearing on the change in dimensionality reduction methods. The random choice of dimensionality reduction method could adversely affect the performance of MCS. Hence the choice-specific dimensionality reduction method has to be selected based on the information content of the hyperspectral image for exploiting the anticipated benefits of MCS. This study suggests the importance of domain adaptive knowledge and understanding the suitability of the dimensionality reduction methods in relation with the underlying hyperspectral image for successful hyperspectral image analysis.

- Chapter 4: With plethora of classifiers and dimensionality reduction methods available in literature, we assessed the relationship between classifier and dimensionality reduction method as well as information class, classifier and dimensionality reduction method for hyperspectral image classification. The results indicate that at overall image level, there are some empirical relationships indicating preferred pairs of classifiers and dimensionality reduction methods across the different hyperspectral images. At the land cover category level, different classes are best classified by different classifiers within each image and the existence of different preferred combinations of class, classifier, and dimensionality reduction method. These observations are found stable even at different spatial and

spectral resolutions as per our experiments on the synthetic hyperspectral images generated using original hyperspectral images. This remarkable observation puts forward the concept of information class specific methodologies for the analyses of hyperspectral images.

- Chapter 5: A novel classification approach named as dynamic classifier system was developed. This approach, functions within the MCS framework, dynamically identifies optimal pairs of classifiers and dimensionality reduction methods based on the input data dynamics and executes classification only with the identified pairs. The required optimal dimension of the dimensionality reduction methods is optimized by the classification accuracy and class separability of training samples. The identified pairs of classifiers and dimensionality reduction methods are combined with different non-trainable and trainable combination functions, and both of them showed significant increase in classification accuracy over the single best classifier. However, the magnitude of improvement is higher with trainable combination function. Moreover, the classification results are compared with that of MCS and SVM. The experimental results on five different hyperspectral images confirm the robustness of the proposed system to significantly increase the classification accuracy over the typical MCS and SVM.

- Chapter 6: We further improved our dynamic classifier system approach presented in the Chapter 5 to dynamic selection of classifier for identifying the best subset of classifiers relative to each image pixel. As part of this, we implemented some of the algorithms used in the pattern recognition methods (local accuracy based dynamic classifier selection and probabilistic based dynamic classifier selection) for testing their suitability for dynamic selection of classifiers with reference to image pixel for hyperspectral image classification context. Experimental results show significant improvement in the classification accuracy. However these approaches demand high computational time since it requires computing the distance between each image pixel and validation sample. To make the dynamic classifier selection accurate and computationally fast, we modelled the classifier selection problem as the classification problem and proposed a new dynamic classifier selection approach based on extreme learning machine regression framework. The proposed approach offers convincing results when compared to single best classifier and also to other dynamic classifier selection approaches. Fur-

ther, we proposed a new spectral-spatial classification model to incorporate spatial contextual information in the hyperspectral image classification framework. Classification results carried out on two different hyperspectral images demonstrate that the proposed spectral-spatial classification model yields a significant increase in accuracy when compared to the state-of-the-art approaches.

Recommendations and future research

The following are some important directions for future research to further enhance the scientific utility of the studies presented in this thesis. Our future works will include some of the following.

- The dimensionality reduction methods used to create diversity in this thesis can be further modified to work as feature selection methods. We believe that deployment of multiple feature selection methods with different complimentary characteristics could improve the classification performance of the MCS.
- Evaluation of the proposed dynamic classifier system to automatically select and combine the classifiers and dimensionality reduction methods based on each land cover class type. Moreover, the dynamic classifier system can be extended to incorporate spatial contextual information.
- Another interesting direction is optimizing the weights of the spectral-spatial classification model based on shape of the objects present in the image. This may improve the classification performance of small spatial structures. Moreover, devising a classifier competence strategy for hyperspectral image classification could also be of interest. Indeed, very limited studies are reported in this arena.
- The extension and investigation of the proposed techniques on the planetary hyperspectral images as well as on the other hyperspectral sensors should also be of interest.

REFERENCES

1. Aghighi, H., Trinder, J., Tarabalka, Y., and Lim, S. (2014). Dynamic Block-Based Parameter Estimation for MRF Classification of High-Resolution Images. *IEEE Geoscience and Remote Sensing Letters*, *11*, 1687–1691.

2. Al-Ahmadi, F. S., and Hames, A. S. (2009). Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Sattelite Images for Some Remote Arid Areas, Kingdom of Saudi Arabia. *Earth Science*, *20*, 167–191.

3. Alajlan, N., Bazi, Y., AlHichri, H. S., and Melgani, F. (2013). Using OWA Fusion Operators for the Classification of Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*, 602–614.

4. Alajlan, N., Bazi, Y., Melgani, F., and Yager, R. R. (2012). Fusion of supervised and unsupervised learning for improved classification of hyperspectral images. *Information Sciences*, *217*, 39–55.

5. Alexandre, L. A., Campilho, A. C., and Kamel, M. (2001). On combining classifiers using sum and product rules. *Pattern Recognition Letters*, *22*, 1283–1289.

6. Amato, U., Cavalli, R. M., Palombo, A., Pignatti, S., and Santini, F. (2009). Experimental Approach to the Selection of the Components in the Minimum Noise Fraction. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 153–160.

7. Andrew, M., and Ustin, S. (2008). The role of environmental context in mapping invasive plants with hyperspectral image data. *Remote Sensing of Environment*, *112*, 4301–4317.

8. Bai, J., Xiang, S., and Pan, C. (2013). A Graph-Based Classification Method for Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 803–817.

9. Bakos, K. L., and Gamba, P. (2009). Potential of hyperspectral remote sensing for vegetation mapping in high mountain ecosystems. In *Proc. 6th EarSEL SIG IS Workshop*. Tel Aviv, Israel.

10. Bakos, K. L., and Gamba, P. (2011). Combining Hyperspectral Data Processing Chains for Robust Mapping Using Hierarchical Trees and Class Memberships. *IEEE Geoscience and Remote Sensing Letters*, *8*, 968–972.

11. Barker, A. L. (1997). *Selection of distance metrics and feature subsets for k-nearest neighbor classifiers*. Phd thesis University of Virginia.

12. Benediktsson, J., and Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *37*, 1367–1377.

13. Benediktsson, J., Palmason, J., Sveinsson, J., and Chanussot, J. (2004). Decision level fusion in classification of hyperspectral data from urban areas. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS '04* (pp. 73–76).

14. Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J. (2013). Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geoscience and Remote Sensing Magazine*, *1*, 6–36.

15. Boggs, J. L., Tsegaye, T. D., Coleman, T. L., Reddy, K. C., and Fahsi, A. (2003). Relationship Between Hyperspectral Reflectance, Soil Nitrate-Nitrogen, Cotton Leaf Chlorophyll, and Cotton Yield: A Step Toward Precision Agriculture. *Journal of Sustainable Agriculture*, *22*, 5–16.

16. Bor-Chen Kuo, Cheng-Hsuan Li, and Jinn-Min Yang (2009). Kernel Nonparametric Weighted Feature Extraction for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 1139–1155.

17. Borengasser, M., Hungate, W. S., and Watkins, R. (2007). *Hyperspectral Remote Sensing: Principles and Applications*. CRC Press.

18. Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 1222–1239.

19. Braun, A. C., Weidner, U., and Hinz, S. (2012). Classification in High-Dimensional Feature Spaces-Assessment Using SVM, IVM and RVM With Focus on Simulated EnMAP Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*, 436–443.

20. Breiman, L. (1996). Bagging Predictors. *Machine Learning*, *24*, 123–140.

21. Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

22. Brenning, A. (2010). Land Cover Classification by Multisource Remote Sensing: Comparing Classifiers for Spatial Data. Studies in Classification, Data Analysis, and Knowledge Organization (pp. 435–443). Springer Berlin Heidelberg.

23. Briottet, X., Boucher, Y., Dimmeler, A., Malaplate, A., Cini, A., Diani, M., Bekman, H., Schwering, P., Skauli, T., Kasen, I., Renhorn, I., Klasén, L., Gilmore, M., and Oxford, D. (2006). Military applications of hyperspectral imagery. In W. R. Watkins, and D. Clement (Eds.), *Proc. SPIE 6239, Targets and Backgrounds XII: Characterization and Representation* (pp. 62390B–62390B–8).

24. Brown, G., Wyatt, J., Harris, R., and Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Information Fusion*, *6*, 5–20.

25. Bruce, L. M., Koger, C. H., and Li, J. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 2331–2338.

26. Bruzzone, L., Roli, F., and Serpico, S. (1995). An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection. *IEEE Transactions on Geoscience and Remote Sensing*, *33*, 1318–1321.

27. Bruzzone, L., and Serpico, S. B. (2000). A technique for feature selection in multiclass problems. *International Journal of Remote Sensing*, *21*, 549–563.

28. Camps-Valls, G., and Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *43*, 1351–1362.

29. Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J., Martin-Guerrero, J., Soria-Olivas, E., Alonso-Chorda, L., and Moreno, J. (2004). Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 1530–1542.

30. Camps-Valls, G., Tuia, D., Bruzzone, L., and Atli Benediktsson, J. (2014). Advances in Hyperspectral Image Classification: Earth Monitoring with Statistical Learning Methods. *IEEE Signal Processing Magazine*, *31*, 45–54.

31. Ceamanos, X., Waske, B., Benediktsson, J. A., Chanussot, J., Fauvel, M., and Sveinsson, J. R. (2010). A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, *1*, 293–307.

32. Cetin, H., Pafford, J., and Mueller, T. (2005). Precision agriculture using hyperspectral remote sensing and GIS. In *Proceedings of 2nd International Conference on Recent Advances in Space Technologies, 2005. RAST 2005* (pp. 70–77).

33. Chan, J. C.-W., Beckers, P., Spanhove, T., and Borre, J. V. (2012). An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery. *International Journal of Applied Earth Observation and Geoinformation*, *18*, 13–22.

34. Chang, C.-I. (2005). Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *43*, 502–518.

35. Chang, Chih-Chung and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27:1—-27:27.

36. Chen, C. H., and Peter Ho, P.-G. (2008). Statistical pattern recognition in remote sensing. *Pattern Recognition*, *41*, 2731–2741.

37. Chen, G., and Qian, S.-E. (2008). Evaluation and comparison of dimensionality reduction methods and band selection. *Canadian Journal of Remote Sensing*, *34*, 26–36.

38. Chen, Y., Nasrabadi, N. M., and Tran, T. D. (2011). Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*, *49*, 3973–3985.

39. Chen, Y., Nasrabadi, N. M., and Tran, T. D. (2013). Hyperspectral Image Classification via Kernel Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 217–231.

40. Chen, Y., Zhao, X., and Lin, Z. (2014). Optimizing Subspace SVM Ensemble for Hyperspectral Imagery Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*, 1295–1305.

41. Cheng, Q., Varshney, P., and Arora, M. (2006). Logistic Regression for Feature Selection and Soft Classification of Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, *3*, 491–494.

42. Cheriyadat, A., and Bruce, L. M. (2003). Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. *IEEE International Geoscience and Remote Sensing Symposium, 2003*, *6*, 3420–3422.

43. Clemmensen, L., Hansen, M., and Ersbø ll, B. (2010). A comparison of dimension reduction methods with application to multi-spectral images of sand used in concrete. *Machine Vision and Applications*, *21*, 959–968.

44. Cloutis, E. (1996). Review Article Hyperspectral geological remote sensing: evaluation of analytical techniques. *International Journal of Remote Sensing*, *17*, 2215–2242.

45. Colás, R. M., Roig, V. J. T., and Garreta, J. S. S. (2001). A comparative study of Dimensionality reduction Methods for Image Classification. In *9th Spanish Symposium on Pattern Recognition and Image Analysis* (pp. 325–330).

46. Dalla Mura, M., Villa, A., Benediktsson, J. A., Chanussot, J., and Bruzzone, L. (2011). Classification of Hyperspectral Images by Using Extended Morphological Attribute Profiles and Independent Component Analysis. *IEEE Geoscience and Remote Sensing Letters*, *8*, 542–546.

47. Dalponte, M., Bruzzone, L., Vescovo, L., and Gianelle, D. (2009). The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, *113*, 2345–2355.

48. Damodaran, B. B., and Nidamanuri, R. R. (2014). Dynamic Linear Classifier System for Hyperspectral Image Classification for Land Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*, 2080–2093.

49. Dao, M., Nguyen, D., Tran, T., and Chin, S. (2012). Chemical plume detection in hyperspectral imagery via joint sparse representation. In *MILCOM 2012 - 2012 IEEE Military Communications Conference* (pp. 1–5).

50. Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., and Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*, 409–426.

51. Dasarathy, B., and Sheela, B. (1979). A composite classifier system design: Concepts and methodology. *Proceedings of the IEEE*, *67*, 708–713.

52. Deogun, J. S., Choubey, S. K., Raghavan, V. V., and Sever, H. (1998). Feature selection and effective classifiers. *Journal of the American Society for Information Science*, *49*, 423–434.

53. Didaci, L., Giacinto, G., Roli, F., and Marcialis, G. (2005). A study on the performances of dynamic classifier selection based on local accuracy estimation. *Pattern Recognition*, *38*, 2188–2191.

54. Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In *Multiple Classifier Systems Lecture Notes in Computer Science* (pp. 1–15). Springer Berlin Heidelberg volume 1857 of *Lecture Notes in Computer Science*.

55. Doan, H. T. X., and Foody, G. M. (2007). Increasing soft classification accuracy through the use of an ensemble of classifiers. *International Journal of Remote Sensing*, *28*, 4609–4623.

56. Dos Santos, J. A., Gosselin, P.-H., Philipp-Foliguet, S., da S. Torres, R., and Falao, A. X. (2012). Multiscale Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 3764–3775.

57. Du, P., Xia, J., Chanussot, J., and He, X. (2012a). Hyperspectral remote sensing image classification based on the integration of support vector machine and random forest. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 174–177).

58. Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., and Liu, S. (2012b). Multiple classifier system for remote sensing image classification: a review. *Sensors*, *12*, 4764–92.

59. Du, Q., Ren, H., and Chang, C.-I. (2003). A comparative study for orthogonal subspace projection and constrained energy minimization. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1525–1529.

60. Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience.

61. Duin, R. (2002). The combining classifier: to train or not to train? In *16th International Conference on Pattern Recognition, 2002* (pp. 765–770).

62. Duro, D. C., Franklin, S. E., and Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, *118*, 259–272.

63. Dutra, L. V. (1999). Feature extraction and selection for ERS-1/2 InSAR classification. *International Journal of Remote Sensing*, *20*, 993–1016.

64. Erbek, F. S., Özkan, C., and Taberner, M. (2004). Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing*, *25*, 1733–1748.

65. Fabio, R., Giacinto, G., and Vernazza, G. (1997). Comparison and Combination of Statistical and Neural Network Algorithms for Remote-Sensing. In *Neurocomputation in Remote Sensing Data Analysis, Advances in Spatial Science Series* (pp. 117–124). Springer-Verlag.

66. Farrell, M., and Mersereau, R. (2005). On the Impact of PCA Dimension Reduction for Hyperspectral Detection of Difficult Targets. *IEEE Geoscience and Remote Sensing Letters*, *2*, 192–195.

67. Fauvel, M., Benediktsson, J. A., Chanussot, J., and Sveinsson, J. R. (2008). Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing*, *46*, 3804–3814.

68. Fauvel, M., Chanussot, J., and Benediktsson, J. (2012). A spatialâĂŞspectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition*, *45*, 381–392.

69. Fauvel, M., Chanussot, J., and Benediktsson, J. A. (2006). Decision Fusion for the Classification of Urban Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2828–2838.

70. Fauvel, M., Chanussot, J., and Benediktsson, J. A. (2009). Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data over Urban Areas. *EURASIP Journal on Advances in Signal Processing*, *2009*.

71. Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., and Tilton, J. C. (2013). Advances in Spectral-Spatial Classification of Hyperspectral Images. *Proceedings of the IEEE*, *101*, 652–675.

72. Fernandez-Redondo, M., Hernandez-Espinosa, C., and Torres-Sospedra, J. (2004). Hyperspectral image classification by ensembles of multilayer feedforward networks. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)* (pp. 1145–1149). IEEE volume 2.

73. Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*, 185–201.

74. Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing Remote Sensing*, *70*, 627–633.

75. Freund, Y., and Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning (ICML 1996)* (pp. 148–156).

76. Fumera, G., and Roli, F. (2005). A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE transactions on pattern analysis and machine intelligence*, *27*, 942–56.

77. Gabrys, B., and Ruta, D. (2006). Genetic algorithms in classifier fusion. *Applied Soft Computing*, *6*, 337–347.

78. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F. (2012). A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*, 463–484.

79. García-Pedrajas, N., García-Osorio, C., and Fyfe, C. (2007). Nonlinear Boosting Projections for Ensemble Construction. *The Journal of Machine Learning Research*, *8*, 1–33.

80. Ghamisi, P., Benediktsson, J. A., and Ulfarsson, M. O. (2014). Spectral-Spatial Classification of Hyperspectral Images Based on Hidden Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing*, *52*, 2565–2574.

81. Ghiyamat, A., and Shafri, H. Z. M. (2010). A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing*, *31*, 1837–1856.

82. Gidudu, A., Bolanle, A. T., and Marwala, T. (2009). Random ensemble feature selection for land cover mapping. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 840–842).

83. Gokaraju, B., Durbha, S. S., King, R. L., and Younan, N. H. (2012). Ensemble Methodology Using Multistage Learning for Improved Detection of Harmful Algal Blooms. *IEEE Geoscience and Remote Sensing Letters*, *9*, 827–831.

84. Granahan, J. C., and Sweet, J. N. (2001). An evaluation of atmospheric correction techniques using the spectral similarity scale. *IEEE International Geoscience and Remote Sensing Symposium, 2001*, (pp. 2022–2024).

85. Green, A., Berman, M., Switzer, P., and Craig, M. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, *26*, 65–74.

86. Guo, X., Huang, X., Zhang, L., and Zhang, L. (2013). Hyperspectral image noise reduction based on rank-1 tensor decomposition. *ISPRS Journal of Photogrammetry and Remote Sensing*, *83*, 50–63.

87. Gurram, P., and Kwon, H. (2010). A full diagonal bandwidth gaussian kernel SVM based ensemble learning for hyperspectral chemical plume detection. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 2804–2807).

88. Gurram, P., and Kwon, H. (2013). Sparse Kernel-Based Ensemble Learning With Fully Optimized Kernel Parameters for Hyperspectral Classification Problems. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 787–802.

89. Gustavo Camps-Valls, L. B. (2009). *Kernel Methods for Remote Sensing Data Analysis*. John Wiley & Sons, Ltd.

90. H, Y., Du, Q., and Ma, B. (2010). Decision Fusion on Supervised and Unsupervised Classifiers for Hyperspectral Imagery. *IEEE Geoscience and Remote Sensing Letters*, *7*, 875–879.

91. Ham, J., Crawford, M., and Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *43*, 492–501.

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

93. Harsanyi, J. C., and Chang, C.-I. (1994). Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing,*, *32*, 779–785.

94. Hasanlou, M., and Samadzadegan, F. (2012). Comparative Study of Intrinsic Dimensionality Estimation and Dimension Reduction Techniques on Hyperspectral Images Using K-NN Classifier. *IEEE Geoscience and Remote Sensing Letters*, *9*, 1046–1050.

95. Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations*. New York: Springer-Verlag.

96. Herold, M., Gardner, M. E., and Roberts, D. A. (2003). Spectral resolution requirements for mapping urban areas. *IEEE Transactions onGeoscience and Remote Sensing*, *41*, 1907–1919.

97. Hirsch, E., and Agassi, E. (2007). Detection of gaseous plumes in IR hyperspectral images using hierarchical clustering. *Applied optics*, *46*, 6368–74.

98. Hsu, P.-H. (2007). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*, 78–92.

99. Huang, G.-B., Zhou, H., Ding, X., and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics*, *42*, 513–29.

100. Huang, X., and Zhang, L. (2013). An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 257–272.

101. Hueso, R., Legarreta, J., Rojas, J., Peralta, J., Pérez-Hoyos, S., del Río-Gaztelurrutia, T., and Sánchez-Lavega, A. (2010). The Planetary Laboratory for Image Analysis (PLIA). *Advances in Space Research*, *46*, 1120–1138.

102. Hull, J., and Srihari, S. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*, 66–75.

103. Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on neural networks*, *10*, 626–34.

104. Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, *13*, 411–430.

105. Imani, M., and Ghassemian, H. (2014). Band Clustering-Based Feature Extraction for Classification of Hyperspectral Images Using Limited Training Samples. *IEEE Geoscience and Remote Sensing Letters*, *11*, 1325–1329.

106. Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive Mixtures of Local Experts. *Neural Computation*, *3*, 79–87.

107. Jia, S., Ji, Z., Qian, Y., and Shen, L. (2012). Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*, 531–543.

108. Jie-lin, Z., Jun-hu, W., Mi, Z., Yan-ju, H., and Ding, W. (2014). Aerial visible-thermal infrared hyperspectral feature extraction technology and its application to object identification. *IOP Conference Series: Earth and Environmental Science*, *17*, 012184.

109. Jimenez, L., and Landgrebe, D. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, *28*, 39–54.

110. Jin Chen, Cheng Wang, and Runsheng Wang (2009). Using Stacked Generalization to Combine SVMs in Magnitude and Shape Feature Spaces for Classification of Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 2193–2205.

111. Johnson, W. R., Hook, S. J., Mouroulis, P., Wilson, D. W., Gunapala, S. D., Realmuto, V., Lamborn, A., Paine, C., Mumolo, J. M., and Eng, B. T. (2011). HyTES: Thermal imaging spectrometer development. In *2011 Aerospace Conference* (pp. 1–8). IEEE.

112. Johnson, W. R., Hulley, G., and Hook, S. J. (2014). Remote gas plume sensing and imaging with NASAâĂŹs Hyperspectral Thermal Emission Spectrometer (HyTES). In M. A. Druy, and R. A. Crocombe (Eds.), *SPIE Sensing Technology + Applications* (p. 91010V). International Society for Optics and Photonics.

113. Joshi, C., Leeuw, J. D., Skidmore, A. K., van Duren, I. C., and van Oosten, H. (2006). Remotely sensed estimation of forest canopy density: A comparison of the performance of four methods. *International Journal of Applied Earth Observation and Geoinformation*, *8*, 84–95.

114. Jun, G., and Ghosh, J. (2009). Multi-class Boosting with Class Hierarchies. In J. A. Benediktsson, J. Kittler, and F. Roli (Eds.), *Multiple Classifier Systems, Lecture Notes in Computer Science Volume 5519* (pp. 32–41). Springer Berlin Heidelberg.

115. Junshi Xia, Peijun Du, X. H. (2013). MRF-Based Multiple Classifier System for Hyperspectral Remote Sensing Image Classification. In Z.-H. Zhou, F. Roli, and J. Kittler (Eds.), *Multiple Classifier Systems, Lecture Notes in Computer Science, Volume 7872* (pp. 343–351). Springer Berlin Heidelberg.

116. Kaewpijit, S., Le Moigne, J., and El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 863–871.

117. Kalluri, H. R., Prasad, S., and Bruce, L. M. (2010). Decision-Level Fusion of Spectral Reflectance and Derivative Information for Robust Hyperspectral Land Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *48*, 4047–4058.

118. Kanal, L. (1974). Patterns in pattern recognition: 1968-1974. *IEEE Transactions on Information Theory*, *20*, 697–722.

119. Kerekes, J., and Baum, J. (2005). Full-spectrum spectral imaging system analytical model. *IEEE Transactions on Geoscience and Remote Sensing*, *43*, 571–580.

120. Khodadadzadeh, M., Li, J., Plaza, A., Ghassemian, H., Bioucas-Dias, J. M., and Li, X. (2014). Spectral-Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization. *IEEE Transactions on Geoscience and Remote Sensing*, *PP*, 1–17.

121. Kirchhoff, K., and Bilmes, J. (1999). Dynamic classifier combination in hybrid speech recognition systems using utterance-level confidence values. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 693–696).

122. Kittler, J. (1998). Combining classifiers: A theoretical framework. *Pattern Analysis and Applications*, *1*, 18–27.

123. Kohavi, R., and Wolpert, D. H. (1996). Bias Plus Variance Decomposition for Zero-One Loss Functions. In *Machine Learning: Proceedings of the Thirteenth International* (pp. 275–283).

124. Kruse, F., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., and Goetz, A. (1993). The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, *44*, 145–163.

125. Kumar, S., Ghosh, J., and Crawford, M. (2001). Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 1368–1379.

126. Kumar, S., Ghosh, J., and Crawford, M. M. (2002). Hierarchical Fusion of Multiple Classifiers for Hyperspectral Data Analysis. *Pattern Analysis & Applications*, *5*, 210–220.

127. Kuncheva, L., Skurichina, M., and Duin, R. (2002). An experimental study on diversity for bagging and boosting with linear classifiers. *Information Fusion*, *3*, 245–258.

128. Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*.

129. Kuncheva, L. I., Bezdek, J. C., and Duin, R. P. (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, *34*, 299–314.

130. Kuncheva, L. I., and Whitaker, C. J. (2003). Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning*, *51*, 181–207.

131. Kuo, B.-C., Chen, I.-L., Li, C.-H., and Hung, C.-C. (2011). Combining ensemble technique of support vector machines with the optimal kernel method for hyperspectral image classification. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 3903–3906).

132. Kuo, B.-C., Ho, H.-H., Li, C.-H., Hung, C.-C., and Taur, J.-S. (2014). A Kernel-Based Feature Selection Method for SVM With RBF Kernel for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*, 317–326.

133. Kwon, H., and Rauss, P. (2011). Feature-based ensemble learning for hyperspectral chemical plume detection. *International Journal of Remote Sensing*, *32*, 6631–6652.

134. Lam, L., and Suen, C. Y. (1995). Optimal combinations of pattern classifiers. *Pattern Recognition Letters*, *16*, 945–954.

135. Lam, L., and Suen, S. Y. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, *27*, 553–568.

136. Landgrebe, D. (1999). Decision fusion approach for multitemporal classification. *IEEE Transactions on Geoscience and Remote Sensing*, *37*, 1227–1233.

137. Landgrebe, D. (2002). Hyperspectral image data analysis. *IEEE Signal Processing Magazine*, *19*, 17–28.

138. Lee, L. (2001). Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2657–2663.

139. Lee, M., Prasad, S., Bruce, L., West, T., Reynolds, D., Irby, T., and Kalluri, H. (2009). Sensitivity of hyperspectral classification algorithms to training sample size. In *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (pp. 1–4).

140. Lennon, M., Mercier, G., Mouchot, M., and Hubert-Moy, L. (2001). Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 2893–2895).

141. Li, J. (2014). Bayesian Classification of Hyperspectral Imagery Based on Probabilistic Sparse Representation and Markov Random Field. *IEEE Geoscience and Remote Sensing Letters*, *11*, 823–827.

142. Li, J., Marpu, P. R., Plaza, A., Bioucas-Dias, J. M., and Benediktsson, J. A. (2013). Generalized Composite Kernel Framework for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 4816–4829.

143. Li, W., Prasad, S., and Fowler, J. E. (2014). Hyperspectral Image Classification Using Gaussian Mixture Models and Markov Random Fields. *IEEE Geoscience and Remote Sensing Letters*, *11*, 153–157.

144. Liang, H. (2011). Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Applied Physics A*, *106*, 309–323.

145. Linders, J. (2000). Comparison of three different methods to select feature for discriminating forest cover types using SAR imagery. *International Journal of Remote Sensing*, *21*, 2089–2099.

146. Lixia Yang, Shuyuan Yang, Penglei Jin, and Rui Zhang (2014). Semi-Supervised Hyperspectral Image Classification Using Spatio-Spectral Laplacian Support Vector Machine. *IEEE Geoscience and Remote Sensing Letters*, *11*, 651–655.

147. Lu, D., Batistella, M., Evaristo, d. M., and Moran, E. (2008). A Comparative Study of Landsat TM and SPOT HRG Images for Vegetation Classification in the Brazilian Amazon. *Photogrammetric Engineering and Remote Sensing*, *74*, 311–321.

148. Lu, D., Li, G., Moran, E., Dutra, L., and Batistella, M. (2011). A Comparison of Multisensor Integration Methods for Land Cover Classification in the Brazilian Amazon. *GIScience & Remote Sensing*, *48*, 345–370.

149. Lu, D., Li, G., Moran, E., Freitas, C. C., Dutra, L. D., and SantAnna, S. J. (2012). A comparison of maximum likelihood classifier and object-based method based on multiple sensor datasets for land -use/cover classification in the Brazilian Amazon. In *4th International Conference on Geographic Object-Based Image Analysis,(GEOBIA), Brazil*.

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

151. Lu, S., Oki, K., Shimizu, Y., and Omasa, K. (2007). Comparison between several feature extraction/classification methods for mapping complicated agricultural land use patches using airborne hyperspectral data. *International Journal of Remote Sensing*, *28*, 963–984.

152. Manolakis, D., Golowich, S., and DiPietro, R. (2014). Long-Wave Infrared Hyperspectral Remote Sensing of Chemical Clouds: A focus on signal processing approaches. *IEEE Signal Processing Magazine*, *31*, 120–141.

153. Manolakis, D., Marden, D., and Shaw, G. A. (2003). Hyperspectral Image Processing for Automatic Target Detection Applications. *Lincoln Laboratory Journal*, *14*, 79–115.

154. Manolakis, D., and Shaw, G. (2002). Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, *19*, 29–43.

155. Manolakis, D. G., and Marden, D. B. (2004). Dimensionality reduction of hyperspectral imaging data using local principal components transforms. In S. S. Shen, and P. E. Lewis (Eds.), *Proc. SPIE 5425, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X, 393* (pp. 393–401).

156. Marpu, P. R., Gamba, P., and Niemeyer, I. (2009). Hyperspectral data classification using an ensemble of class-dependent neural networks. In *First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009* (pp. 1–4).

157. Matheson, D. S., and Dennison, P. E. (2012). Evaluating the effects of spatial resolution on hyperspectral fire detection and temperature retrieval. *Remote Sensing of Environment*, *124*, 780–792.

158. Melesse, A. M., Weng, Q., Thenkabail, P. S., and Senay, G. B. (2007). Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. *Sensors*, *7*, 3209–3241.

159. Mendenhall, M. J., and Merenyi, E. (2008). Relevance-based feature extraction for hyperspectral images. *IEEE transactions on neural networks*, *19*, 658–72.

160. Mi, A., and Huo, Z. (2011). Experimental Comparison of Six Fixed Classifier Fusion Rules. *Procedia Engineering*, *23*, 429–433.

161. Miao, X., Gong, P., Pu, R., Carruthers, R. I., and Heaton, J. S. (2007). Applying class-based feature extraction approaches for supervised classification of hyperspectral imagery. *Canadian Journal of Remote Sensing*, *33*, 162–175.

162. Moran, E. F. (2010). Land Cover Classification in a Complex Urban-Rural Landscape with Quickbird Imagery. *Photogrammetric Engineering and Remote Sensing*, *76*, 1159–1168.

163. Moser, G., Serpico, S. B., and Benediktsson, J. A. (2013). Land-Cover Mapping by Markov Modeling of Spatial-Contextual Information in Very-High-Resolution Remote Sensing Images. *Proceedings of the IEEE*, *101*, 631–651.

164. Moussaoui, S., Hauksdóttir, H., Schmidt, F., Jutten, C., Chanussot, J., Brie, D., Douté, S., and Benediktsson, J. (2008). On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation. *Neurocomputing*, *71*, 2194–2208.

165. Nidamanuri, R. R., and Zbell, B. (2011). Normalized Spectral Similarity Score (NS3) as an Efficient Spectral Library Searching Method for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*, 226–240.

166. Oki, K., Shan, L., Saruwatari, T., Suhama, T., and Omasa, K. (2006). Evaluation of supervised classification algorithms for identifying crops using airborne hyperspectral data. *International Journal of Remote Sensing*, *27*, 1993–2002.

167. Pal, M. (2008). Ensemble of support vector machines for land cover classification. *International Journal of Remote Sensing*, *29*, 3043–3049.

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

169. Petrakos, M., Atli Benediktsson, J., and Kanellopoulos, I. (2001). The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2539–2546.

170. Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., and Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, *113*, S110–S122.

171. Prasad, S., and Bruce, L. (2008a). Limitations of Principal Components Analysis for Hyperspectral Target Recognition. *IEEE Geoscience and Remote Sensing Letters*, *5*, 625–629.

172. Prasad, S., and Bruce, L. (2011). A divide-and-conquer paradigm for hyperspectral classification and target recognition. In S. Prasad, L. M. Bruce, and J. Chanussot (Eds.), *Optical Remote Sensing* (pp. 99–122). Springer Berlin Heidelberg.

173. Prasad, S., and Bruce, L. M. (2007). Limitations of subspace LDA in hyperspectral target recognition applications. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 4049–4052).

174. Prasad, S., and Bruce, L. M. (2008b). Decision Fusion With Confidence-Based Weight Assignment for Hyperspectral Target Recognition. *IEEE Transactions on Geoscience and Remote Sensing*, *46*, 1448–1456.

175. Prasad, S., Bruce, L. M., and Kalluri, H. (2008). A Robust Multi-Classifier Decision Fusion Framework for Hyperspectral, Multi-Temporal Classification. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. II–273–II–276).

176. Prasad, S., Li, W., Fowler, J. E., and Bruce, L. M. (2012). Information Fusion in the Redundant-Wavelet-Transform Domain for Noise-Robust Hyperspectral Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 3474–3486.

177. Rajan, S., Ghosh, J., and Crawford, M. (2006). Exploiting Class Hierarchies for Knowledge Transfer in Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 3408–3417.

178. Rao, N. R., Garg, P. K., and Ghosh, S. K. (2007). Development of an agricultural crops spectral library and classification of crops at cultivar level using hyperspectral data. *Precision Agriculture*, *8*, 173–185.

179. Rapantzikos, K., and Balas, C. (2005). Hyperspectral imaging: potential in non-destructive analysis of palimpsests. In *IEEE International Conference on Image Processing* (pp. II–618).

180. Ren, H., and Chang, C.-I. (2000). A target-constrained interference-minimized filter for subpixel target detection in hyperspectral imagery. In *IEEE International Geoscience and Remote Sensing Symposium.* (pp. 1545–1547).

181. Robila, S., and Gershman, A. (2005). Spectral matching accuracy in processing hyperspectral data. In *International Symposium on Signals, Circuits and Systems* (pp. 163–166).

182. Rogan, J., Franklin, J., and Roberts, D. A. (2002). A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, *80*, 143–156.

183. Rokach, L. (2010). *Pattern Classification Using Ensemble Methods*. World Scientific Publishing Co., Inc.

184. Russell G. Congalton, K. G. (2008). *Assessing the Accuracy of Remotely Sensed Data*. CRC Press.

185. Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore Geology Reviews*, *14*, 157–183.

186. Salerno, E., Tonazzini, A., and Bedini, L. (2006). Digital image analysis to enhance underwritten text in the Archimedes palimpsest. *International Journal of Document Analysis and Recognition (IJDAR)*, *9*, 79–87.

187. Samiappan, S., Prasad, S., and Bruce, L. M. (2013). Non-Uniform Random Feature Selection and Kernel Density Scoring With SVM Based Ensemble Classification for Hyperspectral Image Analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (pp. 1–9).

188. Sanchez, G., Roper, W. E., and Gomez, R. B. (2003). Detection and monitoring of oil spills using hyperspectral imagery. In N. L. Faust, and W. E. Roper (Eds.), *Proc. SPIE 5097, Geo-Spatial and Temporal Image and Data Exploitation III, 233* (pp. 233–240).

189. Santini, F., Palombo, A., Dekker, R., Pignatti, S., Pascucci, S., and Schwering, P. (2014). Advanced Anomalous Pixel Correction Algorithms for Hyperspectral Thermal Infrared Data: The TASI-600 Case Study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *PP*, 1–12.

190. Santos, A. B., Araujo, A. d. A., and Menotti, D. (2013). Combining Multiple Classification Methods for Hyperspectral Data Interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (pp. 1–10).

191. Schapire, R. E. (1990). The Strength of Weak Learnability. *Machine Learning*, *5*, 197–227.

192. Schlerf, M., Rock, G., Lagueux, P., Ronellenfitsch, F., Gerhards, M., Hoffmann, L., and Udelhoven, T. (2012). A Hyperspectral Thermal Infrared Imaging Instrument for Natural Resources Applications. *Remote Sensing*, *4*, 3995–4009.

193. Scholkopf, B., and Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.

194. Senaras, C., Ozay, M., and Yarman Vural, F. T. (2013). Building Detection With Decision Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*, 1295–1304.

195. Shipp, C. A., and Kuncheva, L. I. (2002). Relationships between combination methods and measures of diversity in combining classifiers. *Information Fusion*, *3*, 135–148.

196. Singh, R., Vatsa, M., and Noore, A. (2008). Multiclass mv-granular soft support vector machine: A case study in dynamic classifier selection for multispectral face recognition. In *19th International Conference on Pattern Recognition* (pp. 1–4).

197. Skalak, D. B. (1996). The Sources of Increased Accuracy for Two Proposed Boosting Algorithms. In *In Proc. American Association for Arti Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop* (pp. 120—-125).

198. Smits, P. (2002). Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 801–813.

199. Smits, P. C., Dellepiane, S. G., and Schowengerdt, R. A. (1999). Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. *International Journal of Remote Sensing*, *20*, 1461–1486.

200. Sprintsin, M., Karnieli, A., Berliner, P., Rotenberg, E., Yakir, D., and Cohen, S. (2007). The effect of spatial resolution on the accuracy of leaf area index estimation for a forest planted in the desert transition zone. *Remote Sensing of Environment*, *109*, 416–428.

201. Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., and Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, *50*, 1250–1265.

202. Staenz, K., and Held, A. (2012). Summary of current and future terrestrial civilian hyperspectral spaceborne systems. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 123–126).

203. Stathakis, D., and Vasilakos, A. (2006). Comparison of computational intelligence based classification techniques for remotely sensed optical image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2305–2318.

204. Steele, B. (2005). Maximum posterior probability estimators of map accuracy. *Remote Sensing of Environment*, *99*, 254–270.

205. Szuster, B. W., Chen, Q., and Borger, M. (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, *31*, 525–532.

206. Tarabalka, Y., Benediktsson, J., and Chanussot, J. (2009). Spectral-Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques. *IEEE Transactions on Geoscience and Remote Sensing*, *47*, 2973–2987.

207. Tarabalka, Y., Fauvel, M., Chanussot, J., and Benediktsson, J. A. (2010). SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters*, *7*, 736–740.

208. Themelis, K. E., Schmidt, F., Sykioti, O., Rontogiannis, A. A., Koutroumbas, K. D., and Daglis, I. A. (2012). On the unmixing of MEx/OMEGA hyperspectral data. *Planetary and Space Science*, *68*, 34–41.

209. Thenkabail, P. S., Enclona, E. A., Ashton, M. S., and Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, *91*, 354–376.

210. Thenkabail, P. S., Lyon, J. G., and Alfredo Huete (2011). *Hyperspectral Remote Sensing of Vegetation*. CRC Press.

211. Thenkabail, P. S., Mariotto, I., Gumma, M. K., Middleton, E. M., Landis, D. R., and Huemmrich, K. F. (2013). Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Twoband Vegetation Indices (HVIs) for Biophysical Characterization and Discrimination of Crop Types Using Field Reflectance and Hyperion/EO-1 Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*, 427–439.

212. Thoonen, G., Mahmood, Z., Peeters, S., and Scheunders, P. (2012). Multisource Classification of Color and Hyperspectral Images Using Color Attribute Profiles and Composite Decision Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*, 510–521.

213. Thorp, K., French, A., and Rango, A. (2013). Effect of image spatial and spectral characteristics on mapping semi-arid rangeland vegetation using multiple endmember spectral mixture analysis (MESMA). *Remote Sensing of Environment*, *132*, 120–130.

214. Tin Kam Ho (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 832–844.

215. Udelhoven, T., van der Linden, S., Waske, B., Stellmes, M., and Hoffmann, L. (2009). Hypertemporal Classification of Large Areas Using Decision Fusion. *IEEE Geoscience and Remote Sensing Letters*, *6*, 592–596.

216. Van Der Meer, F. D., Van Der Werff, H. M., van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., van der Meijde, M., Carranza, E. J. M., de Smeth, J. B., and Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, *14*, 112–128.

217. Villa, A., Chanussot, J., Jutten, C., Benediktsson, J. A., and Moussaoui, S. (2009). On the use of ICA for hyperspectral image analysis. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. IV–97–IV–100).

218. Wang, J., and Chang, C.-i. (2006). Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1586–1600.

219. Wang, X.-L., Waske, B., and Benediktsson, J. A. (2009). Ensemble methods for spectral-spatial classification of urban hyperspectral data. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. IV–944–IV–947).

220. Waske, B., Member, S., and Benediktsson, J. A. (2007). Fusion of Support Vector Machines for Classification of Multisensor Data. *IEEE Transactions on Geoscience and Remote Sensing*, *45*, 3858–3866.

221. West, T., Prasad, S., Bruce, L. M., and Reynolds, D. (2009). Utilization of local and global hyperspectral features via wavelet packets and multiclassifiers for robust target recognition. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. III–825–III–828).

222. Windeatt, T. (2005). Diversity measures for multiple classifier system analysis and design. *Information Fusion*, *6*, 21–36.

223. Woloszynski, T., and Kurzynski, M. (2011). A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition*, *44*, 2656–2668.

224. Wolpert, D. (1996). The lack of priori distinctions between learning algorithms. *Neural Computation*, *8*, 1341–1390.

225. Wolpert, D., and Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*, 67–82.

226. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*, 241–259.

227. Woods, K., Kegelmeyer, W., and Bowyer, K. (1997). Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 405–410.

228. Wu, C., Han, X., Niu, Z., and Dong, J. (2010a). An evaluation of EO-1 hyperspectral Hyperion data for chlorophyll content and leaf area index estimation. *International Journal of Remote Sensing*, *31*, 1079–1086.

229. Wu, J. C., Chang, C.-P., and Tsuei, G.-C. (2010b). Comparison of feature extraction methods in dimensionality reduction. *Canadian Journal of Remote Sensing*, *36*, 645–649.

230. Xia, J., Du, P., He, X., and Chanussot, J. (2014). Hyperspectral Remote Sensing Image Classification Based on Rotation Forest. *IEEE Geoscience and Remote Sensing Letters*, *11*, 239–243.

231. Xu, B., and Gong, P. (2007). Land-use/Land-cover Classification with Multispectral and Hyperspectral EO-1 Data. *Photogrammetric Engineering & Remote Sensing*, *73*, 955–965.

232. Xu, L., Krzyzak, A., and Suen, C. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*, 418–435.

233. Xu, L., Li, J., and Brenning, A. (2014). A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery. *Remote Sensing of Environment*, *141*, 14–23.

234. Yan, W. Y., and Shaker, A. (2011). The effects of combining classifiers with the same training statistics using Bayesian decision rules. *International Journal of Remote Sensing*, *32*, 3729–3745.

235. Yanchen Bo, Jinfeng Wang, and Xiang Li (2005). Exploring the scale effect in land cover mapping from remotely sensed data: the statistical separability-based method. In *IEEE International Geoscience and Remote Sensing Symposium* (pp. 3884–3887).

236. Yang, H., Du, Q., and Ma, B. (2010a). Decision Fusion on Supervised and Unsupervised Classifiers for Hyperspectral Imagery. *IEEE Geoscience and Remote Sensing Letters*, *7*, 875–879.

237. Yang, J.-M., Kuo, B.-C., Yu, P.-T., and Chuang, C.-H. (2010b). A Dynamic Subspace Method for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *48*, 2840–2853.

238. Yi, Z., Samuel, B., and Street, W. N. (2006). Ensemble pruning via semi-definite programming. *Journal of Machine Learning Research*, *7*, 1315–1338.

239. Yin, J., Wang, Y., and Hu, J. (2012). A New Dimensionality Reduction Algorithm for Hyperspectral Image Using Evolutionary Strategy. *IEEE Transactions on Industrial Informatics*, *8*, 935–943.

240. Zalazar, L. (2006). *Comparison of classification approaches for land cover mapping in the Wielkoposka region*. Phd thesis.

241. Zhang, B., Li, S., Jia, X., Gao, L., and Peng, M. (2011). Adaptive Markov Random Field Approach for Classification of Hyperspectral Imagery. *IEEE Geoscience and Remote Sensing Letters*, *8*, 973–977.

242. Zhang, L., Zhang, L., Tao, D., and Huang, X. (2012). On Combining Multiple Features for Hyperspectral Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 879–893.

243. Zhang, R., and Zhu, D. (2011). Study of land cover classification based on knowledge rules using high-resolution remote sensing images. *Expert Systems with Applications*, *38*, 3647–3652.

244. Zhong, Y., and Zhang, L. (2012). An Adaptive Artificial Immune Network for Supervised Classification of Multi-/Hyperspectral Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 894–909.

245. Zhou, Z.-H., Wu, J., and Tang, W. (2002). Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, *137*, 239–263.

LIST OF PAPERS BASED ON THESIS

Papers in Refereed International Journals

1. **Damodaran, B.B.**, and Nidamanuri, R.R., (2014). Assessment of the impact of dimensionality reduction methods on information classes and classifiers for hyperspectral image classification by multiple classifier system. ***Advances in Space Research***, 53, pp. 1720-1734.

2. **Damodaran, B.B.**, and Nidamanuri, R.R., (2014). Dynamic linear classifier system for hyperspectral image classification for land cover mapping. ***IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing***, 7, pp. 2080-2093.

3. **Damodaran, B.B.**, Nidamanuri, R.R. and Tarabalka, Y., (2014). Dynamic Classifier Selection Approach for Hyperspectral Image Classification with Joint Spectral and Spatial Information, ***IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing***.

4. **Damodaran, B.B.**, and Nidamanuri, R.R. Impact of dimensionality reduction methods on the classification performance of multiple classifier system for hyperspectral image classification, ***International Journal of Remote Sensing***.

5. **Damodaran, B.B.**, and Nidamanuri, R.R. Assessment of impact of spectral and spatial resolution on the class, classifier and dimensionality reduction method relationship for hyperspectral image classification by multiple classifier system, ***GIScience and Remote Sensing***.

Presentations in Conferences

1. **Damodaran, B.B.**, Nidamanuri, R.R., (2014). Dynamic classifier selection approaches for hyperspectral image classification, Accepted in *SPIE Remote sens-*

ing 2014, Amsterdam, Netherlands, Sep 22-25.

2. **Bhushan, D.B.**, and Nidamanuri, R.R., (2013). Dynamic classifier system for hyperspectral image classification, *IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, Melbourne, July 21-23 (won the IEEE Geoscience and Remote Sensing Society Travel Grant of $ 1500).

3. **Damodaran, B.B.**, Nidamanuri, R.R., (2012). Assessment of relationship between information classes, classifiers and dimensionality reduction methods for hyperspectral image classification by multiple classifier system, *National Symposium on Space Technology for Food and Environmental Security and Annual Convention of Indian Society of Remote Sensing*, New Delhi, Dec 5-7.

4. **Damodaran, B.B.**, Nidamanuri, R.R., (2012). Impact of feature reduction methods on target detection methods for hyperspectral image classification by multiple classifier system, *39th COSPAR Scientific Assembly*, Mysore, July 14-22.

5. **Damodaran, B.B.**, Nidamanuri, R.R., (2011). Multiple classifier system for hyperspectral image classification, *National Symposium on Empowering Rural India through Space Technology and Annual Convention of Indian Society of Remote Sensing*, Bhopal, November 8-11 (won the best poster presentation award).