ENSEMBLE CLASSIFIERS FOR LAND COVER MAPPING

BOLANLE TOLULOPE ABE

Doctor of Philosophy

May 2014

ENSEMBLE CLASSIFIERS FOR LAND COVER MAPPING

BOLANLE TOLULOPE ABE

A thesis submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg

Supervisor: Professor Tshilidzi Marwala

Declaration

I declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Signed this ____26th____ day of ____May__2014___

___

Bolanle Tolulope Abe

Abstract

This study presents experimental investigations on supervised ensemble classification for land cover classification. Despite the arrays of classifiers available in machine learning to create an ensemble, knowing and understanding the correct classifier to use for a particular dataset remains a major challenge. The ensemble method increases classification accuracy by consulting experts taking final decision in the classification process. This study generated various land cover maps, using image classification. This is to authenticate the number of classifiers that should be used for creating an ensemble. The study exploits feature selection techniques to create diversity in ensemble classification. Landsat imagery of Kampala (the capital of Uganda, East Africa), AVIRIS hyperspectral dataset of Indian pine of Indiana and Support Vector Machines were used to carry out the investigation. The research reveals that the superiority of different classification approaches employed depends on the datasets used. In addition, the pre-processing stage and the strategy used during the designing phase of each classifier is very essential. The results obtained from the experiments conducted showed that, there is no significant benefit in using many base classifiers for decision making in ensemble classification. The research outcome also reveals how to design better ensemble using feature selection approach for land cover mapping.

The study also reports the experimental comparison of generalized support vector machines, random forests, C4.5, neural network and bagging classifiers for land cover classification of hyperspectral images. These classifiers are among the state-of-the-art supervised machine learning methods for solving complex pattern recognition problems. The pixel purity index was used to obtain the endmembers from the Indiana pine and Washington DC mall hyperspectral image datasets. Generalized reduced gradient optimization algorithm was used to estimate fractional abundance in the image dataset thereafter obtained numeric values for land cover classification.

The fractional abundance of each pixel was obtained using the spectral signature values of the endmembers and pixel values of class labels. As the results of the experiments, the classifiers show promising results. Using Indiana pine and Washington DC mall hyperspectral datasets, experimental comparison of all the classifiers' performances reveals that random forests outperforms the other classifiers and it is computational effective.

The study makes a positive contribution to the problem of classifying land cover hyperspectral images by exploring the use of generalized reduced gradient method and five supervised classifiers. The accuracy comparison of these classifiers is valuable for decision makers to consider tradeoffs in method accuracy versus complexity. The results from the research has attracted nine publications which include, six international and one local conference papers, one published in Computing Research Repository (CoRR), one Journal paper submitted and one Springer book chapter, Abe *et al*., 2012 obtained a merit award based on the reviewer reports and the score reports of the conference committee members during the conference period.

Dedication

This thesis is dedicated to the Almighty God, the Omnipotent, the Merciful and Omniscience Father. Who reigns and rules in the affairs of men forever and ever at all times. It is also dedicated to my beloved father Late Francis Olufunmilayo Akinmade.

Acknowledgements

I thank the Almighty God, the creator of heaven and earth for His loving kindness over me. I thank you for unlimited love and favor you showed to me and abundant of grace given me for picking me up from hopelessness into royal priesthood. Receive all glory and majesty ascribe to you in my life in Jesus name.

I sincerely declare my unfeigned gratitude to my supervisor Prof. Tshilidzi Marwala for the consistent encouragement and understanding shown to me during the course of this program. All the time you sacrificed to attend to my work painstakingly going through my work and offering useful suggestions. I express my gratitude to my co-supervisor Prof. Anton Van Wyk for his contribution to my work. My thanks also go to Dr. Anthony Gidudu for his contributions towards the success of this research. At this juncture, I want to express my appreciation and gratitude to Prof. Oludayo Olugbara for his unrelenting encouragement and supportive efforts towards the successful completion of this study. Despite his ever-busy schedules found time to offer useful advice and suggestions throughout the course of this study. I equally express my gratitude to Dr. Peter Olubambi, Dr. Olusola Ilemobade, Dr. Jaco Jordan, Dr. Benjamin Akpor and Mr. Modupe Abiodun for their supports and encouragements.

I especially appreciate Ven. (Prof). E.A.O. Laseinde, Mr. &. Mrs Olorunlogbon Adeniyi, Mr. & Mrs. Donald Adebanjo for giving support to my family while I was away in South Africa. My gratitude also goes to Pastor. & Mrs. Babalola Olusegun Akinmade, Mr & Mrs Bayode Akinmade, Mr. & Mrs. Adetunji, Mokesioluwa Adeniyi, my sister-in-law Mrs. Foluke Bosede Olaleye, who overwhelmed me with their moral and spiritual support. God will reward you abundantly.

I am very grateful to my parents late Pa Francis Olufunmilayo Akinmade of blessed memory and my mother Mrs. Marian Olanumo Akinmade who tried despite the

difficult circumstances to bequeath the greatest legacy known as education to me. My sincere appreciation goes to members of Chapel of Faith, Federal University of Technology, Akure, Nigeria, Christian Embassy International, Full Gospel Businessmen's Fellowship International and Eastside Community Church, Pretoria. South Africa, I must equally appreciate my children, Oluwagbemi, Ayooluwadun and Oluwapelumi, who understood when mummy was busy with her research work and needed not to be disturbed. My friends Mrs. Adebukola Adunola, Mrs. Mabel Olanipekun, Mrs. Janet Adegbenro and Dr. Patricia Popoola, thank you for all the supports.

At this point, I thank God for the "Bone of my Bone" Dr. Olusola Taye Abe, my husband, whom God has raised as a pillar of support and encouragement for me in all dimensions. I appreciate deeply, the zeal which you often showed by urging me towards successful completion of this uphill task. I appreciate those times you sat up with me even into the early hours of the morning, showing great interest in the progress of this work. Those times you upheld me in prayers of intercession to our God. God bless you all, Amen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Nomenclature

AMEE: Automatic Morphological Endmember Extraction

ANC: Abundance Non-negativity Constraint

ASC: Abundance Sum-to-one Constraint

AVIRIS: Airborne Visible/Infrared Imaging Spectrometer

BSI: Bhattacharyya Separability Index

CVM: Cross Validation Majority

DSI: Divergence Separability Index

ENVI: Environment for Visualizing Images

GIS: Geographic Information Systems

GPS: Global Positioning Systems

GPUs: Graphics Processing Units

GRG: Generalized Reduced Gradient

HI: Hyperspectral Image

ISDAS (CCRS): Imaging Spectrometer Data Analysis System

ISIS Integrated Software for Imagers and Spectrometers

LARSYS: Laboratory of Applications of Remote Sensing Image Data Processing System

LIBSVM: A Library for Support Vector Machines

LM: Landsat Thematic

LSMA: Linear Spectral Mixture Analysis

MNF: Maximum Noise Fraction

NN: Nearest Neighbour

NSM: No separability measure

PCA: Principal Component Analysis

PPI: Pixel Purity Index

RF: Random Forest

RMS: Root Mean Square

SI: Separability Index

SIPS (CU/CSES) Spectral Image Processing System

SPAM Spectral Analysis Manager

SVM: Support Vector Machines

TDSI: Transformed Divergence Separability Index

WEKA: Waikato Environment for Knowledge Analysis

1A1: One – Against – One

CHAPTER 1

Introduction

1.1 Background of the study

Increasingly earth observation has become a prime source of data in the Geosciences and many related disciplines permitting research into the distant past, the present and into the future. Earth observation is based on the premise that information is available from the electromagnetic energy field arising from the earth's surface or atmosphere (or both) and in particular from the spatial, spectral and temporal variations in that field (Levin, 1999; Kramer, 2002; Sabino, 2005). Through this, the environment can be better monitored, modelled, and consequently, better policy decisions made.

One of the areas of research interest has always been how to relate earth observation output e.g. Aerial photographs and satellite images (remotely sensed imagery) to known features (e.g. Land cover). Land cover refers to the physical surface of the earth, including various combinations of vegetation types, soils, exposed rocks and water bodies as well as anthropogenic elements, such as agriculture and built environments (Mathur, 2004; Udelhoven, 2009; Sánchez *et al*., 2010). A land cover map consists of a set of contiguous map units each of which is labelled according to a land cover class. The main reason for producing land cover maps is to give a clear idea of the stock, state of nature and built resources.

One critical environmental aspect to which satellite images can be applied is land cover mapping using classification algorithms called classifiers. An emerging area of

research interest relating image classification to land cover mapping is ensemble classification (Breiman, 1996; Kittler *et al*., 1998; Opitz, 1999; Giacinto and Roli, 2001; Polikar, 2006; Oza and Tumer, 2008; Marwala, 2009; Gidudu *et al*., 2009a). Whereas previous research (Wacker and Landgrebe, 1972; Roli *et al*., 1997; Pinho *et al*., 2008) has sought to find better classification algorithms, ensemble classification is premised on using a 'consensus' approach to land cover mapping, the 'consensus' being dependent on a collection of base classifiers. Ongoing research in the application of ensemble classification to land cover mapping has focused on the different ways ensembles can be constituted (Chen *et al*., 2007; Chan and Paelinckx, 2008; Udelhoven *et al*., 2009). Some of the common approaches (Breiman, 1996; Opitz, 1999; Pal, 2003; Tsymbal, 2005; Polikar, 2006; Marwala, 2011) have involved constituting ensembles using different classification algorithms, constituting base classifiers from using different training data, or deriving base classifiers using different band combinations (ensemble feature selection) (Gidudu *et al*., 2008b).

Classifying and mapping vegetation is an important technical task for managing natural resources as vegetation provides a base for all living beings and plays an essential role in affecting global climate change, such as influencing terrestrial (carbon dioxide) CO_2 (Xiao *et al*., 2004; Xie *et al*., 2008). But classification accuracy poses serious challenge and this is due to, the design procedure of classifier, choice of training sets from dataset and information conveyed to the algorithm (Oza and Tumer, 2008). Statistical based classifiers have been successfully applied to multispectral data but are not effective for hyperspectral data (Hsu, 2007). The major reason is the fact that the number of spectral bands in hyperspectral data is too large, relative to the training samples. An effective way to solve this problem is

to reduce the dimension of the hyperspectral images. This can be done by extracting a number of salient features from the hyperspectral data (Keshava and Mustard, 2002; Su *et al*., 2008; Sánchez *et al*., 2010).

Furthermore, hyperspectral sensor uses many contiguous bands of high resolution, which covers the visible, near-infrared, and shortwave infrared spectral bands (Adams and Smith, 1986; Vane *et al*., 1993; Lillesand and Kiefer, 1994; Nascimento, 2005). Information obtained from a particular pixel in a given hyperspectral band is a mixture of the energies scattered by the constituent substances in the respective pixel spatial coverage (Adams and Smith, 1986). According to Heinz and Chang (2001), Linear Spectral Mixture Analysis (LSMA) is a commonly accepted approach to mixed-pixel classification in remote sensing to estimate fractional abundance present in the image pixels. This study addresses the concerned issues for a remote sensing hyperspectral data and classification.

1.2 Problem statement

In remote sensing software, there are arrays of classifiers that can be used for image classification. Despite these arrays, knowing and understanding the correct classifier to use for a particular dataset remains a major challenge. Majority of preceding research has centred attention on developing classifiers that are better than existing ones (Steele and Patterson, 2001; Pal, 2007). Ensemble classification is an emerging approach to land cover mapping whereby the final classification output is a result of a ‘consensus’ of classifiers. Ensemble classification has been successfully deployed, but little has been done to systematically analyze the interplay between the ensemble size and the resulting classification accuracy. Hence, to date it has not been ascertained how many base classifiers an ensemble should have, or for a given

ensemble, how many features should the base classifiers have and the dependency of them on the data in question.

An ensemble system should consist of base classifiers which are diverse i.e. Classifiers whose decision boundaries err differently. Nevertheless, it is not established if there is any correlation between classification accuracy and diversity measures. Previous work relating ensemble classification to land cover mapping has focused on investigating how combining different classifiers impacts on classification accuracy (Foody *et al*., 2007), how different types of ensembles can be applied to land cover mapping (Pal, 2007) and also enforcing diversity through bagging for land cover mapping (Steele and Patterson, 2001). Kittler *et al*., (1998) developed a common theoretical framework and revealed that many available algorithms are developed to solve different problems of classification where all the pattern representations are jointly used to make decisions. In this research, the influence of ensemble classification on land cover classification accuracy was investigated.

Hyperspectral Imagery data provide ample spectral information to identify and distinguish spectrally unique materials. Therefore, the classification of the materials and classifier performance over the data is very crucial. While the general procedures (pre-processing and classification) for hyperspectral images and multispectral images are the same, the processing of hyperspectral data remains a challenge. Especially, cost effective and computationally efficient procedures are required to process a large number of image bands (Varshney and Arora 2004; Xie *et al*., 2008, Abe *et al*., 2012).

A major problem with hyperspectral dataset is mixed pixels which are associated with the hyperspectral sensor used during sensing. Spectral mixture analysis provides an efficient mechanism for the interpretation and classification of remotely sensed multidimensional imagery (Plaza *et al*., 2002; Iordache *et al*., 2011). In remote sensing literature, a number of techniques have been developed for unconstrained, partially constrained and fully constrained linear spectral unmixing which can be computationally expensive (Sanchez *et al*., 2010; Iordache *et al*., 2011). For fully constrained linear spectral mixing analysis, two constraints are imposed. First, the sum of the abundance fractions of information present in an image pixel should be one. Secondly, these abundance fractions should be nonnegative. The first constraint can be easily solved while the second has not been fully solved because disparities can be experienced and the solution requires numerical approaches. For this purpose, an experimental comparison of supervised learning classifiers for land cover classification of hyperspectral imagery was investigated.

1.3 Aim and Objectives

The study investigates the influence of ensemble classification approach and spectral mixing problems associated with hyperspectral imagery for land cover classification accuracy.

The objectives are to investigate:

- Diversity through training a given classifier on different features and land cover accuracy
- Interplay between the structure of ensemble and land cover classification accuracy

- Interplay between combination rules of the ensemble and land cover classification accuracy
- Fully constrained spectral unmixing analysis for land cover classification. The process involved;

- Data dimensionality reduction using the Eigenvalues and Maximum Noise Factor (MNF)
- Separate noise from data
- Obtain spectral endmembers and their corresponding spectral signatures
- Obtain the best pure spectral pixels from the dataset using the Pixel Purity Index (PPI)
- Estimate fractional abundance in the dataset thereby obtaining the numeric values for land cover classification

1.4 Contribution to knowledge

Several research studies have been reported in the remote sensing literature on different classification algorithms. The ensemble classification approach has been proven to yield favourable results compared to single systems for a broad range of applications (Polika, 2006). This research revealed that on combining classifiers in its application for land cover mapping, there is no significant benefit in having many base classifiers. In this study, three base classifiers were sufficient to give an accurate result (Gidudu *et al*., 2009a, Abe *et al*., 2010).

Reports in the remote sensing literature on how to quantify diversity in the ensemble classification has focus investigation on finding measures to build diverse ensemble systems (Kuncheva and Whitaker, 2003). This research revealed that diversity

measures do not provide an adequate means upon which to constitute ensembles for land cover mapping (Gidudu *et al*., 2008a).

The major contribution to knowledge of the study is the introduction of Generalized Reduced Gradient (GRG), an optimization technique to estimate fractional abundance in hyperspectral image for land cover classification. From literature, a number of techniques have been developed for unconstrained, partially constrained and fully constrained linear spectral unmixing which can be computationally expensive (Sanchez *et al*., 2010). For instance, quadratic programming has been applied to impose abundance sum-to-one constraint (ASC) and abundance no-negativity constraint (ANC) to obtain fractional abundance, but the disadvantage is that the algorithms are computationally expensive (Boardman, 1995; Settle and Drake, 1993; Heinz and Chang, 2001).

The method used by (Heinz and Chang, 2001) likewise has the limitation of excessive computational complexity as the number of endmember increases. Another approach is the application of the least square method which cannot satisfy the ASC and ANC. If applied, the solved abundance fractions of the material signatures may be negative and their sum within an image pixel may not necessarily be one. Hence, the solutions are generally not optimal in terms of material quantification (Heinz and Chang, 2001; Sanchez, *et al.*, 2011).

Introducing Generalized Reduced Gradient (GRG), an optimization technique with fully constraints algorithm in this study provides solution to the problems of estimating fractional abundance in hyperspectral image. The estimated numeric values obtained was successfully used for land cover classification using various classifiers and the classification accuracy results are remarkably good (Abe *et al.*,

2012). This is important for decision maker to consider tradeoff in accuracy and complexity of methods. In addition, the research permits the analysis of spatial dimension of land cover change and will contribute to the assessment of consequences of land cover change.

Another contribution to knowledge is the ability to successfully apply the GRG algorithm on the Indian pine dataset (Abe *et al.*, 2012). The GRG algorithm has been able to solve the land-cover classification scenario associated with the Indian pine dataset which has been researched for a long time due to the problematic nature of the dataset (Landgrebe, 1998; Plaza *et al*., 2008). The algorithm was successfully applied to Washington DC mall hyperspectral dataset (Abe *et al.*, 2012). The research reveals that the GRG algorithm can be successfully applied to any type of data. Hence, ensemble classifiers improve predictive accuracy.

The research outputs have been able to produce nine publications to its credit. These include five international conference publications, two local conference publications, one published in the Computing Research Repository (CoRR), one submitted for Journal publication and one Book Chapters. Abe *et al*., 2012 obtained a merit award based on the reviewer reports and the score reports of the conference committee members during the conference period.

1.5 Scope of the Thesis

The techniques used in this work are made general and can be used for other applications, other than considered in this thesis. Land cover classification accuracy was investigated throughout this work, using various classification algorithms. According to Congalton and Green (2009), there are two types of map accuracy assessment. They are positional and thematic accuracy assessments. Positional

accuracy involves location of map features accuracy and measures the distance between the spatial feature on a map and its true location on the ground.

On the other hand, thematic accuracy assessment deals with the labels or attributes of the features in the ground truth or reference map. The accuracy assessment that was used for this study was based on thematic. This measures whether the mapped feature labels are different from the true feature label. The process includes; designing of the accuracy assessment sample, data collection for each sample and results' analysis. The study areas used for the research are thematic Landsat imagery of Kampala, the capital of Uganda, Indiana pines and Washington DC mall hyperspectral datasets.

1.6 Thesis Layout

The remaining parts of the thesis are structured as follows:

Chapter 2 presents related theory on remote sensing and applications. It will also include a survey of work done using different classification algorithms as applied to remotely sensed imagery.

Chapter 3 contains different design methodology used for this study. There are six investigations described in the chapter with each investigation carried out in accordance with the objectives of the study.

Chapter 4 shows the results obtained from the investigations carried out in chapter 3. It also contains discussions on the results.

Chapter 5 summarises the major findings of this research and recommendations for further research directions are given.

Appendices: Appendix A presents the Land cover maps obtained using Kampala dataset; Appendix B contains Land cover maps generated using Indiana pine dataset.

List of Publications

From the thesis, the following papers were published:

1. Abe, B. T., Olugbara, O. O. and Marwala, T. (2014). "Experimental comparison of support vector machines with random forest for hyperspectral imaging land cover classification," *Journal of Earth System Science,* Indian Academy of Sciences (Accepted).
2. Abe, B. T., Olugbara, O. O. and Marwala, T. (2014). "Classification of Hyperspectral image using Machine Learning Methods," *IAENG Transactions on Engineering Technologies, Book Chapters of Springer,* Lecture Notes in Electrical Engineering, Volume 247, pp 555-569.
3. Abe, B. T. and Jordaan, J. (2013). "Hyperspectral Image classification based on NMF features selection method," *Proceedings of SPIE,* Vol. 9067, 90670N, SPIE Digital Library.
4. Abe, B. T., Jordaan, J. and Marwala, T. (2013). "Ensemble classifier using GRG algorithm for land cover classification," *Proceedings of SPIE,* Vol. 9067, 90670O, SPIE Digital Library.
5. Abe, B. T., Olugbara, O. O. and Marwala, T. (2012). "Hyperspectral Image Classification using Random Forest and Neural Network," *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering and Computer Science 2012, WCECS 2012,* 24-26 October, San Francisco, USA, pp 522-527.

6. Abe, B., Gidudu, A. and Marwala, T. (2010). “Investigating the effect of ensemble classification on remotely sensed data for land cover mapping,” *In Proceedings of 2009 IEEE International Geoscience & Remote Sensing Symposium Earth observation-Origins to Applications*, Honolulu, USA, 23 –39 July, pp. 2832 – 2835.

7. Gidudu, A., Abe, B. and Marwala, T. (2009b). “Random Ensemble Feature Selection for Land Cover Mapping,” *In Proceedings of 2009 IEEE International Geoscience & Remote Sensing Symposium Earth observation-Origins to Applications*, Cape Town South Africa, 12 – 19 July, pp. 840 – 842.

8. Gidudu, A., Abe, B. and Marwala, T. (2009a). “Investigating the effect of Ensemble Size on Classification Accuracy*”, In Proceedings of the 2nd International Conference on Earth Observation for Global Changes, (EOGC2009)*, Chengdu, China 25 – 29 May, pp. 2195 – 2199.

9. Gidudu, A., Abe, B.T. and Marwala, T. (2008b). “Ensemble feature selection for hyperspectral imagery”, *In Proceedings of the 19th Annual Symposium of the Pattern Recognition Association of South Africa*, Cape Town, South Africa, 27 – 29 November, pp. 27-32.

10. Gidudu, A, Abe, B. and Marwala, T. (2008a). “Land Cover Mapping Using Ensemble Feature Selection Methods”, CoRR abs/0811.2016.

CHAPTER 2

Related theory

This chapter presents a summary of remote sensing and the concept of hyperspectral images. Related theories on feature extraction and classification algorithms used in the study are undertaken.

2.1 Land cover

The dynamic and complex nature of environmental changes poses numerous challenges to human development. With increasing deforestation, industrialization, urbanization, mineral exploration among others, the price has been environmental degradation. Some of the long term consequences of environmental degradation have included: increased poverty, as well as climate change resulting in unexpected prolonged rains and droughts (Xiao *et al*., 2004; Xie *et al*., 2008). Some of the mitigation measures include environmental monitoring, environmental modelling and advocacy about the importance of the environment. Assessing and monitoring the state of the earth's surface is a major requirement for global change research and has resulted into new clarity and better awareness of the earth's dynamic nature (Lambin *et al*., 2001; Jung *et al*., 2006).

Land cover serves as a basic inventory of land resources for all levels of government, environmental agencies, and private industries throughout the world**.** Land Cover is characterized by a large variety of special distinct classes. The diagnosis and evaluation of the spectral separability measure yield the potential for automated identification and mapping of these classes. Land cover mapping has found applications in inventory and baseline land resource mapping, land change and time

series analysis, agricultural monitoring, natural resource monitoring, error assessment and uncertainty management among others (Bruzzone and Cossu, 2002; Wolter, 2005; Randall, 2006; Chen *et al*., 2007; Xie, *et al*., 2008)

2.1.1 Why Land Cover mapping?

Land cover refers to the surface cover of the earth. Land cover mapping constitutes an integral component of the process of managing the land resources and mapped information is the result of analysis of remotely sensed data (Levin, 1999). According to Congalton and Green, (2009), because the earth's resources are scarce, and more people are added to it continually, there will be a shortage of resources and their values. Hence, the need for timely and accurate information dissemination on the type, quantity and degree of resources increase. For this reason, land cover mapping is regarded as important for environmental management and land use policy making. Because of its broad coverage and cost-effectiveness, application of remote sensing to derive land cover information, using either manual interpretation or automated classification has been on the increase. The latter is frequently used as it rapidly reduces the workload of the image interpretation and requires much less expert knowledge (Zhou and Yang, 2009).

2.2 Remote sensing

Remote sensing is the acquisition and analysis of information about the earth from a distance using a computer and sensor through electromagnetic radiation. This started in 1830s with the origination of the camera (Jorgensen, 2004). Figure 2.1 shows the concept of remote sensing. The leap from manual aerial photographic interpretation to 'automatic' classification was inspired by the availability of experimental data in various bands in the mid 1960's as a prelude to the launch of the Earth Resources

Technology Satellite (ERTS – which was later renamed Landsat 1) (Maul and Gordon, 1975). Landsat Thematic (TM), a moderate resolution scanner makes it possible to view the co-occurrence of different materials within the ground instantaneous field of view of urban areas that are characterized by a pattern of very heterogeneous patches. This necessitated the adoption of digital multivariate statistical methods for the extraction of land cover information (Landgrebe, 1997).

The conservation, preservation and sustainable yield of natural resources are increasingly dependent upon remotely sensed data for inventory and monitoring of changes (Xie, et al., 2008). A suite of digital data, such as high resolution satellite images is currently available for this purpose. New technologies such as Image Processing (IP), GPS and GIS are being used to integrate and process these data. Digital image processing is extremely important in fully harnessing the information in high resolution satellite imagery data.

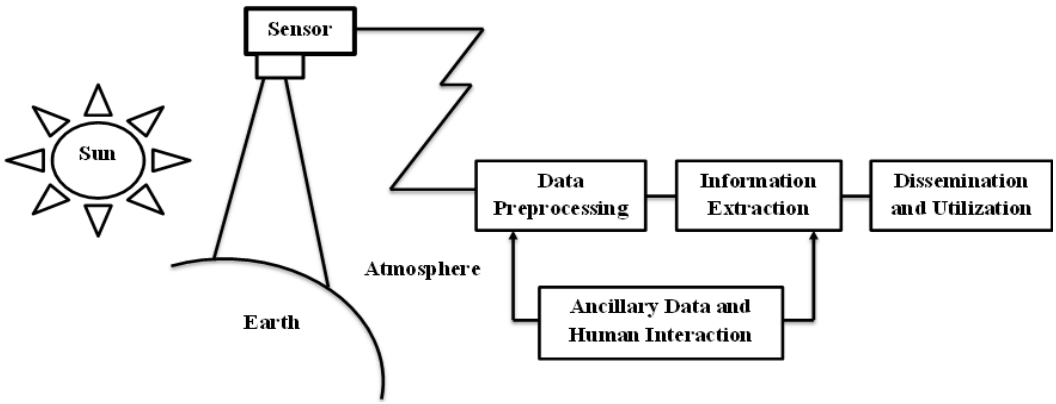


Figure 2.1: Concept of remote sensing

(Landgrebe, 1998).

Remote sensing technology not only can be applied to map vegetation covers over land areas, but also in underwater areas with focus on mapping Submergent Aquatic

Vegetation (SAV), which is regarded as a powerful indicator of environmental conditions in both marine and fresh water ecosystems (Wolter *et al*., 2005; Lathrop *et al*., 2006; Xie *et al.*, 2008).

2.3 Data collection

Remote sensing instruments, measures reflected electromagnetic radiation with the aid of aerial or satellite platform (Figure 2.2). Remotely sensed imageries are obtained using passive or active remote sensors. Passive sensors measure energy that is naturally available through sun ray when available, while active sensors depend on their energy source for illumination. Passive sensor can only be used during the day since that is when the sun is available to illuminate the earth. Active sensors radiate energy directly to the target object to be investigated. The sensor detects and measures the radiation reflected from that target.

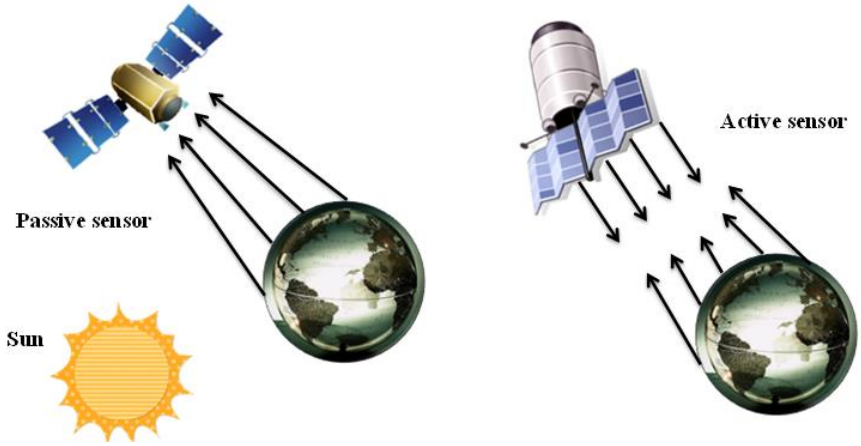


Figure 2.2: Remote sensing sensors
(Levin, 1999).

A remote sensing image is an objective recording of the electromagnetic reaching the sensor. The electromagnetic spectrum (figure 2.3) ranges from shorter wavelengths to the longer wavelengths. Though there are several regions of the spectrum that can be used for remote sensing, the most frequently used is the ultraviolet portion. The ultraviolet portion of the spectrum has the shortest wavelengths of the electromagnetic spectrum.

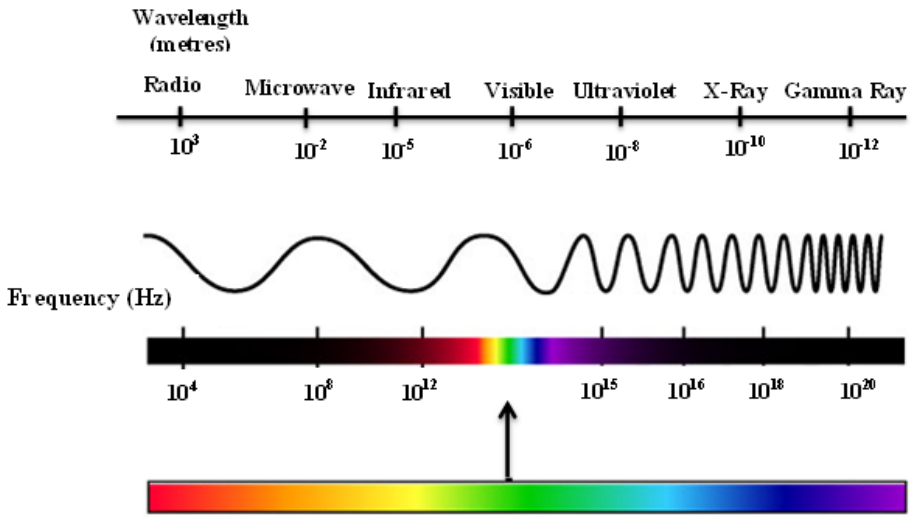


Figure 2.3: Electromagnetic spectrum

(Levin, 1999)

The remote sensing process involves interaction between incident radiation and the targets of interest are as shown in figure 2.4. The radiation used for remote sensing travels through some distance through the atmosphere before reaching the earth to collect the information on the target. The seven stages comprising the process of remote sensing as shown in figure 2.4 are:

- Energy source or illumination (A)

- Radiation and the atmosphere (B)
- Interaction with the target (C)
- Recording of energy by the sensor (D)
- Transmission, Reception and Processing (E)
- Interpretation and analysis (F)
- Application (G)

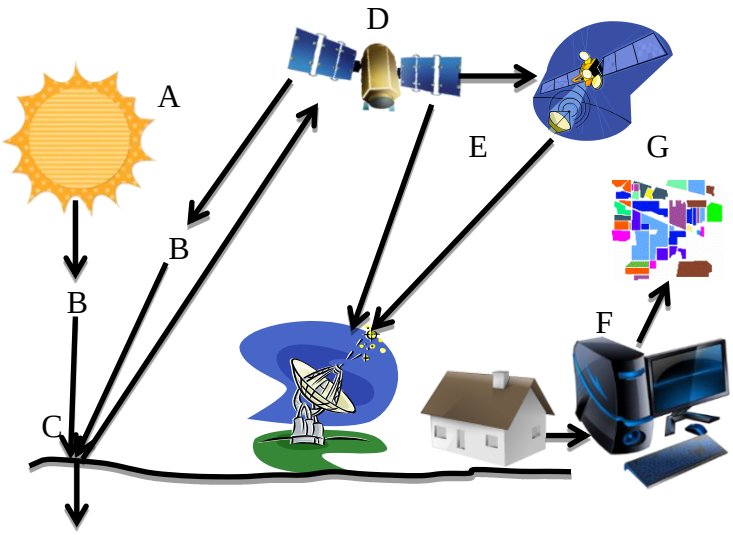


Figure 2.4: Remote Sensing process

(Levin, 1999)

2.4 Image resolutions

Spectral resolution is the ability of a sensor to produce clear or distinguished wavelength interval, known as channels or bands in the electromagnetic spectrum (Fundamental of Remote Sensing). The arrangement of pixels in an image describes the spatial structure while radiometric describes the actual information contained in

the image. The temporal resolution is the length of time taken by the satellite to complete one entire orbit cycle. The image obtained by remote sensors may contain one spectral band called panchromatic image (black and white), multispectral images contains few spectral bands and contiguous spectral bands are hyperspectral images. Different class labels and details in an image can be distinguished when their responses over a distinct wavelength range are compared (Fundamental of Remote Sensing). Figure 2.5 shows the three dimensional hyperspectral data cube and can be treated as a stack of two dimensional spatial images, each corresponding to a particular narrow spectral band.

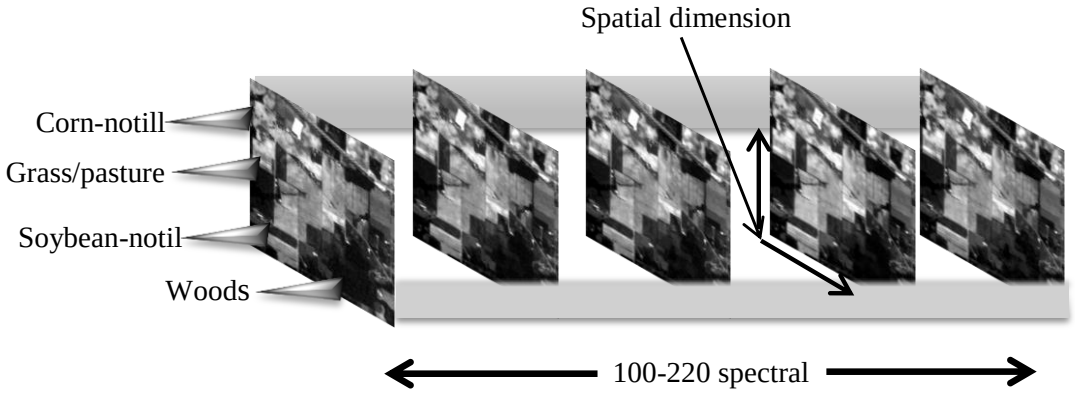


Figure 2.5: Three dimensional hyperspectral data
(Shaw and Burke, 2003)

2.5 Hyperspectral Imagery

Recent developments in sensor technology have resulted in the development of hyperspectral instruments. The instruments are capable of collecting hundreds of images (spectral bands) corresponding to wavelength channels, for the same area of the earth's surface (Green *et al*., 1998; Plaza *et al*., 2003). Each pixel contains in

hyperspectral data cube is linked to spectral signature or fingerprint that uniquely characterize the materials within the pixel (Figure 2.6). Such recognition provides a great advantage for detecting minerals, urban planning and vegetation studies, monitoring and management of environment, security and defense matters among others (Varshney and Arora, 2004; Xie *et al*., 2008). However, accurate classification of remote sensing images is an important task to be able to achieve these advantages (Shaw and Burke, 2003). The existing classifiers are certainly diverse, robust and powerful, it is abundantly clear that much information remains untapped in modern hyperspectral data, awaiting new algorithms and software implementations.

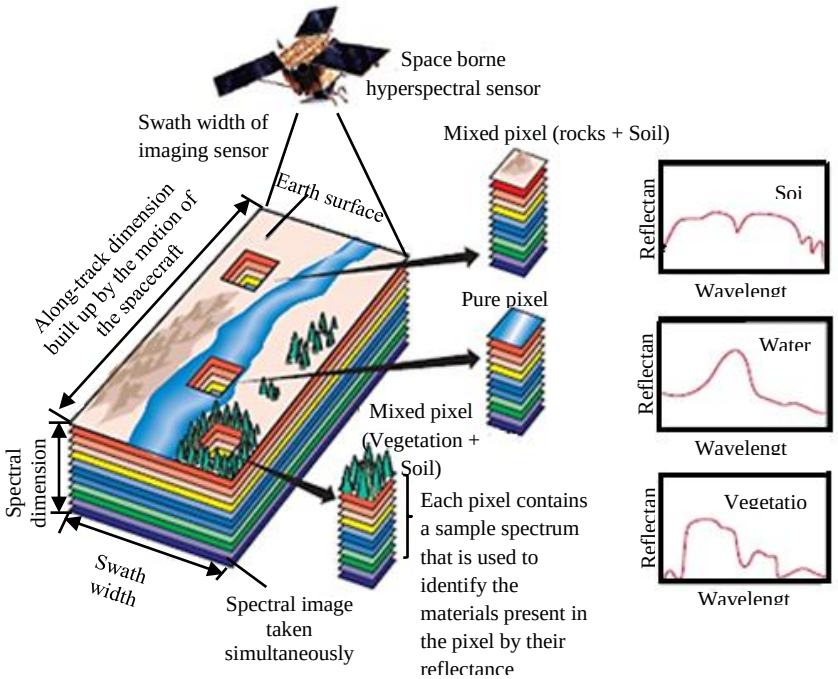


Figure 2.6: Concept of imaging Spectrometer

(Shaw and Burke, 2003)

2.6 Spectral mixture

In hyperspectral imagery, a pixel is usually mixed with a number of materials present in the scene. Spectral mixture analysis has been extensively used in remote sensing for material discrimination, classification and detection. Various spectral mixture techniques have been reported in the remote sensing literature (Plaza *et al*., 2002; Keshava and Mustard, 2002; Plaza *et al*., 2004b; Pinho *et al*., 2008; Zhang and Huang, 2010). However, the spectral signature of a particular pixel contains a mixture of the signatures (fingerprints) of the numerous materials present within the spatial coverage of the target field view by the sensor. This is due to one of the following reasons:

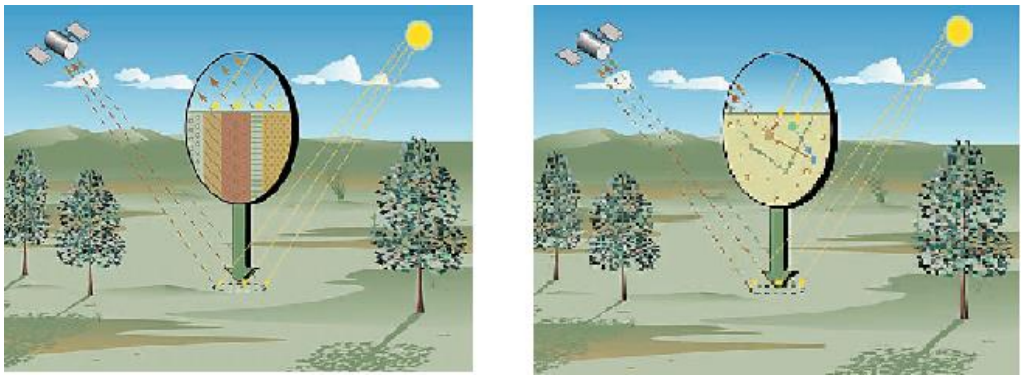
(1) Low spatial resolution of the sensor used that distinct material can jointly occupy one pixel. The outcome spectral measurement contains some composite of the individual spectral. This is common with remote sensing platforms operating at high altitude, covering large area surveillance with low spatial resolution.

(2) Mixed pixel can also occur when unique materials are combined into a homogeneous mixture (Sanchez *et al*., 2010).

Spectral unmixing is the process whereby the measured spectrum of a mixed pixel is broken down into a number of pure spectral components, referred to as endmembers. This is also known as class labels, class types, components or signatures (Gong and Zhang 1999) and a set of corresponding fractional abundance that indicate the amount of each endmember present in the pixel (Plaza *et al*., 2004b; Sanchez *et al*., 2010). In hyperspectral imagery, linear spectral unmixing is a commonly accepted approach to mixed-pixel classification. Distinct substances such as water, tree, bridge, grass among others which are called the endmembers and the fraction in

which they emerge in the mixed pixel is referred to as fractional abundance. The other method used is the nonlinear mixing, where the incident sun ray comes across close mixture that causes multiple bounces.

The disadvantage of using nonlinear method is that the particle size, composition and altered state of the endmembers due to the multiple bounce significantly control the parameters of the solution (Keshava and Mustard, 2002). For experiment 5 of this research, linear unmixing method was adapted to generate fractional abundance in order to determine the degree of abundance in each endmember present in the pixel.



(a) **(b)**

Figure 2.7: Linear and nonlinear mixing

(a) Linear mixing: Radiated sun reflected from the target through a single bounce

(b) Nonlinear mixing: Radiated sun encounters an intimate mixture that stimulates multiple bounce (Keshava and Mustard, 2002).

We chose five algorithms from WEKA software for classification procedure because they are freely available and research has shown that the software has produced good results in remote sensing imagery classification (Garner, 1995; Pinho *et al*., 2008; Moore *et al*., 2009; Nandgaonkar *et al*., 2010). The algorithms used are: C4.5,

Random forest, Support vector machine, neural networks, bagging using REPtree as the base classifier. In the next session, some applications of these classifiers that were used in this research are discussed.

2.7 Image processing and classification

Image processing is any form of signal processing for which the input is an image, such as satellite images, a photograph or video frame. The output of image processing may be either image, a set of characteristics or parameters related to the image. Image processing is of interest because it affords abundant data to be translated into useful information in time (Liu *et al.,* 2009). Numerous sources of imagery are identified by their differences in spectral, spatial, radioactive and temporal characteristics, hence are suitable for different purposes of vegetation mapping. Then, correlations of the vegetation types (communities or species) within the classification system with discernible spectral characteristics of remote sensed imagery have to be identified. These spectral classes of the imagery are eventually translated into the vegetation types in the image interpretation process, which is also called image processing (Xie *et al*., 2008).

The abilities to retrieve information from the data motivates researchers to explore methods of data mining, a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad *et al.*, 1996; Xie *et al.*, 2008; Li, 2011). Increasingly Earth observation has become a prime *source* of data in the Geosciences and many related disciplines permitting research into the distant past, the present and into the future. One of the areas of research interest has always been how to relate earth observation output e.g. Aerial photographs and satellite images (remotely sensed imagery) to known features. Pre-processing of remotely sensed (satellite) images prior to vegetation extraction is important to remove noise

and increase the interpretability of image data. This is true when dealing with time series of imagery or when an area is covered with many images since it is crucial to make these images compatible spatially and spectrally. It is expected that results obtained after image pre-processing should appear as if the image were acquired from the same sensor (Hall et al., 1991; Xie et al., 2008).

Image classification is the process whereby pixels in an image are automatically categorized into land cover classes. It is a fundamental analysis technique for remotely sensed data and involves the categorization of pixels based on their spectral characteristics (Cihlar et al., 1998). Image classification can be categorized into (i) supervised and unsupervised, (ii) Spectral and contextual classifications.

2.7.1 Supervised and unsupervised classification

Supervised classification requires the user to decide which classes exist in the image, and then delineate samples of these classes. These samples (known as training areas) are input into a classification program, which produces a classified image. The choice of the training area is based on the researcher's familiarity with the geographical area and knowledge of the actual land cover types in the image. Unsupervised classification does not require training areas. Actually, it is the opposite of the supervised classification process. Spectral classes are grouped based significantly on the numerical information in the data and then matched by the researcher to the information classes (Bortolot, 1999; Sabino, 2005). Types of supervised classifiers include Minimum – Distance to Means, Neural Networks, and maximum likelihood classifiers, while examples of unsupervised classifiers include K – Mean, Fuzzy C means, and ISODATA among others (Bortolot, 1999; Sabino, 2005).

2.7.2 Spectral and contextual classifications

Spectral techniques are based on the spectral response pattern of a pixel and are divided into two categories, parametric and non-parametric classifiers. Parametric classifiers assume a Gaussian distribution of the data. In supervised parametric classification, a multivariate Gaussian distribution associated with each class is extracted from the training set by estimating the mean vector and the covariance matrix. Parametric classifiers are based on the definition of some discriminate functions based on the parameters of normal distribution. An example of parametric classifiers is a Gaussian maximum likelihood classifier. Non-parametric classifiers imply decision boundaries of arbitrary geometry and needs an iterative process to complete the boundaries. Example of these classifiers includes Nearest Neighbor (1-NN) 3, K-Nearest Neighbor (k-NN) 4 techniques, kernel methods, and classification trees. Contextual classifiers consider the special context of a pixel in the image and are generally applied on remote sensing data when a large variety of spectral responses are observed in the same field. Contextual classifiers have been used successfully in a number of different problems such as coping with segmentation and classification of remotely sensed data (Dwivedi, et al., 2004).

2.8 Classifiers

This section presents a brief discussion on the classifiers used for this research. Supervised classification method was used for all the investigations carried out in the research.

2.8.1 Support Vector Machines (SVMs)

The construction of Support Vector Machines (SVMs) classifier has been described in many literatures (Boser *et al*., 1992; Vapnik, 1999; Witten and Frank, 2005;

Watanachaturaporn & Arora, 2004; Watanachaturaporn *et al*., 2004, Bruzzone *et al*., 2006). The classifier has been proven to have a high generalized ability to solve classification problems and has been extensively used in supervised classification for land cover mapping (Cortes and Vapnik, 1995; Lennon *et al*., 2002; Melgani and Bruzzone, 2004; Gidudu, 2006; Abe *et al.*, 2010). This technique has been used in different application domains, including object detection and text categorization and has outperformed the traditional neural network technique in terms of generalization capacity (Taratinno *et al*., 2006). It has a very interesting property for hyperspectral image analysis in the sense that it does not suffer from Hughes phenomenon and with a small number of training samples it can separate classes easily (Cortes and Vapnik, 1995; Lennon *et al*., 2002; Qi and Huang, 2007). SVM selects the optimal hypothesis as the one yielding the maximum margin of separation between two classes (Vapnik, 1995). Among multi-class SVMs frequently used is one against one (1A1) algorithm, a type used in this research. The method constructs k (k – 1) /2 hyperplanes where each is trained on the data from two classes selected out of k classes. The performance and computation cost of (1A1) has been established, when compared with other SVM methods that no method can compete with one against one in terms of training and the good statistical performance. For a k-class problem, 1A1 algorithm creates all pairwise discriminating hyperplanes, in the sense that (k – 1) hyperplanes are created to separate one class from others. All final borderlines are included in the created hyperplanes. The decision function for class *ij* is defined as (Platt 2004):

$$f^{ij}(X) = \langle \phi(x), w^{ij} \rangle + b^{ij} \tag{2.1}$$

This is obtained by solving this optimization problem (Platt 2004):

$$\min \tfrac{1}{2} \left\| w^{ij} \right\|^2 + C \sum_n \xi_n^{ij}$$

$$\langle \phi(x_n), w^{ij} \rangle + b^{ij} \geq 1 - \xi_n^{ij}, \text{ if } y_n = i \tag{2.2}$$

Subject to:

$$\langle \phi(x_n), w^{ij} + b^{ij} \geq -1 + \xi_n^{ij} \rangle, \text{ if } y_n = j, \tag{2.3}$$

$$\xi_n^{ij} \geq 0$$

for all *n* examples in classes *i* and *j*.

Given that $f_{ij}(X) = -f_{ij}(X), k(k-1)/2$ diverse decision function for a k – class problem exist. The 1A1 method used for class recognition in this research was the "max wins" algorithm. In this method, each classifier casts one vote for its chosen class and the class with majority vote wins. This means (Platt 2004):

$$\text{Class } X = \arg\max_{\mathrm{i}} \sum_{j \neq i=1}^{k} sign\left(f_{ij}\right)(X) \tag{2.4}$$

Where sign (f_{ij}) represents the sign function, meaning that it has the value 1 when f_{ij} is positive otherwise 0. When there is a tie on the decision, each point in the

unclassifiable region is given to the nearest neighbour applying real value decision functions as (Platt 2004):

$$\text{Class } X = \arg\max_{i} \sum_{j \neq i=1}^{k} \left(f_{ij} \right)(X) \tag{2.5}$$

The SVM used in this study experiment is the Gaussian kernel, which is Radial basis function (RBF) is given by (Platt 2004):

$$K\, x_i, x_j = \exp\left(-\gamma \left\| x_i \right) - x_j \right\|^2 \tag{2.6}$$

The network operates in a way that one basis function is centered on all training instances. The predicted outputs are combined linearly by computing the maximum margin hyperplane (Witten and Frank, 2005). The SVM (LIBSVM) used for sequential minimal optimization (Chang and Lin, 2001) in WEKA software was used for the classification.

2.8.2 Neural network

Neural network methods are general classifiers that can handle problems with lots of parameters and can classify objects, even when the distribution of object in n-dimensional parameter space is very complex. Research activities have established that neural networks are capable alternative to numerous conventional methods (Zhang, 2000, Benediktsson et al., 1990; Benediktsson and Swain, 1992, Marwala, 2010). They are data driven, self-adaptive technique that adjust themselves to data under investigation without any explicit specification of the functional or distributional form. They are also universal functional approximations that can approximate any function with arbitrary accuracy (Cybenko, 1989; Hornik, 1991;

Zhang, 2000). In addition, neural networks are nonlinear models that make them flexible in modelling real world complex relationships. Furthermore, they are capable of estimating the posterior capability that provides the basis for creating the classification rule and carrying out statistical analysis (Zhang, 2000; Richard and Lippman, 1991, Marwala, 2010). Various neural networks are available for classification purposes (Lippmann, 1989), but this paper focus on multilayer perceptron (MLP) that uses back propagation to classify instances. The nodes in the network are all sigmoid.

2.8.3 C4.5

The C4.5 for many years was the standard decision tree classification algorithm used in the machine learning and data mining communities. The algorithm creates decision tree classifier to predict membership cases of categorical dependent variable from measures on one or more variable. It uses information gain ration matrix for classification. C4.5 uses the significance of statistic of error to trim branches and uses probability weighting to deal with feature loss during the training period (Richard *et al*., 2006; Wang *et al*., 2008). The algorithm works in a way that each node of the decision tree matches an attribute and individual arc matches a value range of the attribute. The value of the expected attribute is known by the path from the root to individual leaf. The highest attribute is allocated to each node. This aims at associating the attribute to reduce the data entropy to a node (Pinho *et al*., 2008; Silva *et al*., 2008). Moore (2009) claimed that C4.5 is consistent and performs better using large data. Using Remote Sensing (RS) data for classification, Yu and Ai (2009) implemented rough set and C4.5 algorithm. The classifier performs well on that particular data type. Polat and Gunes (2009) conducted investigation using ‘one against all method’ with C4.5. The experiment was conducted with Dermatology,

Image segmentation, Lymphography datasets. They were able to achieve very good accuracy against other algorithms. But nothing was said about time taking by the algorithm and other datasets. Efficiency of an algorithm also depends on the type of data used. Jiang and Yu (2009) suggested a hybrid algorithm based on outlier detection and C4.5. They used imbalance data and make it stable with the application of outlier detection using C4.5 algorithm. The experiment produced good accuracy as compared with other algorithm.

2.9 Ensemble classifier

Ensemble classifiers are essentially a multi-classifier system, implying that their functionality is dependent on a collection of classifiers – an ensemble of classifiers to get an accurate result (Foody *et al*., 2007). In literature, ensemble classification systems go by a variety of names, such as a mixture of experts and a committee of classifiers (Marwala, 2001; Polikar, 2006; Gidudu *et al.*, 2009; Marwala, 2009; Marwala, 2011). Individual predictions of the ensemble members are typically combined by weighted or unweighted voting to classify new data (Foody *et al*., 2007; Qi and Huang, 2007; Marwala, 2011). Research has shown that ensemble generates better classification accuracy results than the individual classifier making up the ensemble (Jimenez *et al*., 1999; Giacinto and Roli, 2001; Polikar, 2006; Pal, 2007). In remote sensing Giacinto and Roli, 1997; and Roli *et al*., (1997) report the application of an ensemble of neural networks and the integration of classification results of different type of classifiers.

The main idea behind ensemble classification is that one is interested in taking advantage of various classifiers at their disposal to come up with a 'consensus' result. This is made possible by the following fundamental reasons (Dietterich, 2009):

(i) Statistical: This problem comes up due to the small amount of training data available as compared to the size of the hypothesis space. By constructing an ensemble out of all of these accurate classifiers, the algorithm can "average" their votes and reduce the risk of choosing the wrong classifier.

(ii) Representational: In most applications of machine learning, the true function cannot be represented by any of the hypotheses in the space.

(iii) Computational: Many learning algorithms work by performing some form of local search that may get stuck in local optima. An ensemble constructed by running the local search from many different starting points may provide a better approximation to the true unknown function.

These three fundamental issues are the three most important ways in which existing learning algorithms fail. Hence, ensemble methods have the promise of reducing (and perhaps even eliminating) these three key shortcomings of standard learning algorithms, than any of the individual classifiers.

The challenge at hand, involves deciding which classifiers to consider and how to combine their results. Polika (2006) affirmed that while the prediction of the ensemble may not be better than the best individual classifier in the ensemble, the method will surely minimize the overall risk of making a particularly poor choice. In another justification, Polika (2006) opine that ensemble method, handling large volume of data set by a single classifier will not be realistic. The best way to handle the data is to partition it into smaller subsets, train different classifiers with different partitions of data, and combine their predictions using an intelligent combination rule. The outcome often proves to be more efficient (Kittler, 1998; Polika; 2006).

From available literatures, Polkar, (2006) recommends that the constituent classifiers in the ensemble have different decision boundaries (that is diverse), because if identical there will be no gain in combining the classifiers (Shipp and Kuncheva, 2002). Diversity in ensemble systems has been more commonly explored by considering different classifiers, training a given classifier on different portions of the data, using a classifier with different parameter specifications and using different features. Diversity in ensemble systems is ensured by selecting base classifiers which err differently, since strategically combining these classifiers can reduce the total error (Tsymbal *et al*., 2005; Parikh and Poikar, 2007; Oza and Tumer, 2008; Marwala, 2011) as illustrated in Figure 2.8.

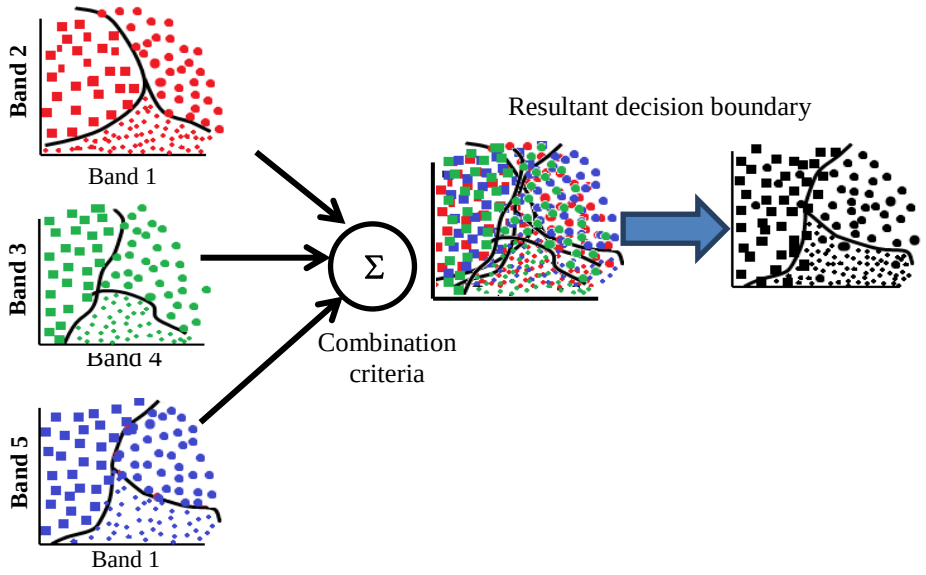


Figure 2.8: Graphical illustration of an Ensemble classifier system (Parikh and Polikar, 2007)

Two methods which have gained prominence in ensemble classification research include bagging or bootstrap aggregating (Breiman, 1996) and Adaboost or reweighting boosting (Freund and Schapire, 1996), which principally involve training a classifier on different training data.

Of equal importance to ensemble classification is how to combine the results of the base classifiers (Foody *et al*., 2007). A number of approaches exist to combine information from multiple classifiers (Giacinto and Roli, 2001; Valentini and Masulli, 2003; Huang and Lees, 2004) such as majority voting (Chan and Paelinckx, 2008), weighted majority voting (Polikar, 2006) or more sophisticated methods like consensus theory (Benediksson and Swain, 1992) and stacking (Džeroski and Zenko, 2004). The majority vote rule was adopted as combination rule for the ensemble method used for this study. Majority vote rule functions on binary value function Δ_{ki}, (Hurber and Dutra, 2000; Kittler *et al*., 1998)

where $\Delta_{ki} = 1$ if

$$p\left(\omega_k \middle| \underset{x_i}{\rightarrow}\right) = \max_{j=1}^{C} P\left(\omega_j \middle| \underset{x_i}{\rightarrow}\right) \text{ and } \Delta_{ki} = 0 \qquad (2.7)$$

Otherwise, class ω_j is assigned by majority voting through (Hurber and Dutra, 2000):

$$\sum_{i=1}^{R} \Delta_{ji} = \max_{k=1}^{m} \sum_{i=1}^{R} \Delta_{ki} \qquad (2.8)$$

For each class ω_k in (2.7), the sum on the right hand side counts the votes obtained for the prediction from the base classifiers. The class that has the highest number of votes is chosen as the consensus (majority) decision.

2.9.1 Random forest

The random forests (RF) ensemble classifier builds several decision trees randomly as proposed (Mountrakis *et al*., 2011) for classification of multisource remote sensing, geographic data and hyperspectral imaging. Various ensemble classification methods have been proposed in recent times and they have been proven to considerably improve classification accuracy. The most famous and widely used ensemble methods are boosting and bagging (Breiman, 2001). The boosting method is based on sample re-weighting technique, but a bagging method uses bootstrapping. Random Forest classifier uses bagging or bootstrap aggregating to yield an ensemble of classification and regression trees. The method works by searching only a random subset of the features for a split at each node to minimize the correlation between the classifiers in the ensemble. The method selects a set of features randomly and creates an algorithm with a bootstrapped sample of the training data (Breiman, 2001). This method provides a potential benefit that it is insensitive to noise or overtraining because resampling is independent of the weighting scheme employed. Additionally, the method is computationally lighter than methods based on boosting and bagging and often produces excellent results (Silva *et al*., 2008).

The random forests algorithm uses Gini Index (Pal, 2005) as a feature selection measure. In this case, the impurity of a feature is measured against the classes. In certain training set T, selecting a pixel at random and allocates it to a class C_i (Pal, 2005), (Rodriguez-Galiano, 2001);

$$\sum \sum_{J=1} \left(f\left(C_i, T\right) / |T| \right) \left(f\left(C_i, T\right) / |T| \right) \tag{2.9}$$

Where $\left(f\left(C_i, T\right) / |T| \right)$ represent the possibility that a selection case belongs to class C_i. Hence, using a certain combination of features, a decision tree grows up to its maximum depth without pruning. For our experiment, 10 trees were constructed. Out of bag error was 0.5471 while considering 187 random features.

2.9.2 Bagging

Bagging (bootstrap aggregating) was the leading widely used technique of selecting sets for ensemble classifiers. The invention of bagging has its root from attempts to reduce the error variance. The algorithm operates by creating new training sets using resampling methods from the original data set n (the number of samples in the original training data) times, randomly with replacement. The sample being chosen will not be removed from the data set in the next draw. Therefore, some of the training samples will be selected more than once while some samples will not be chosen at all in a new set. The algorithm assists classification accuracy by reducing the variance of the classification errors. The classifiers are ensemble using majority vote and vote of each classifier carries the same weight (Breiman, 1996).

2.10 Diversity measure

Different diversity measures have been defined for quantitative assessment of diversity. The main focus of investigation has centered on finding measures which can be used as a basis upon which to build diverse ensemble systems (Masisi *et al*., 2008). In literatures (Kuncheva and Whitaker 2003; Polkar, 2006), there are two categorizations of diversity measures, namely: pair-wise and non-pair-wise diversity

measures. Among pair-wise measures defined between two classifiers are Correlation, Q-Statistic, Disagreement and Double Fault Measures among others. For T classifiers, calculating T (T-1) /2 pair-wise diversity measures and the overall diversity of the ensemble is obtained by averaging these pair-wise measures. According to (Polikar, 2006), for two hypotheses h_i and h_j, the notations

Table 2.1: Hypothesis for diversity

	hj is correct	*hj* is incorrect
hi is correct	a	B
hj is incorrect	c	D

a = fraction of instances that are correctly classified by both classifiers,

b = fraction of instances correctly classified by h_i but incorrectly classified by h_j.

$$a + b + c + d = 1$$

Subsequently, the following pair-wise diversity measures can be defined:

***Correlation*:** Diversity is measured as the correlation between two classifier outputs define as (Polikar, 2006):

$$\rho_{ij} = (ad - bc)/y \quad 0 \leq \rho \leq 1 \tag{2.10}$$

Where y = $\sqrt{[(a+b)(c+d)(a+c)(b+d)]}$

Maximum diversity is obtained for ρ = 0, indicating that the classifiers are uncorrelated.

Q - Statistic is defined as (Polikar, 2006):

$$Q_{ij} = (ad - bc)/(ad + bc) \quad (2.11)$$

Q assumes positive values if the same instances are correctly classified by both classifiers, otherwise, negative values. Maximum diversity is also obtained for Q = 0.

***Disagreement and Double fault Measures*:** The disagreement is the probability that the two classifiers will disagree, whereas the double fault measure is the probability that both classifiers are incorrect. The diversity increases with both the disagreement and the double fault value (Polikar, 2006; Masisi *et al*., 2008).

$$D_{i.j} = b + c, \quad (2.12)$$

$$DF_{i.j} = d. \quad (2.13)$$

The diversity measure for the ensemble is derived by calculating the average of the pair-wise measures of the constituent classifiers (Shipp and Kuncheva, 2002; Tsymbal *et al*., 2005). Non pair-wise diversity measures include: the entropy measure, Kohavi-Wolpert variance and measurement of inter-rater agreement (Kohavi and Wolpert, 1996).

2.11 Feature selection

Hyperspectral data has found applications in many areas due to the number of bands making up the dataset (Varshney and Arora, 2004). However, with increasing number of bands, cost of classification increases exponentially, though accuracy saturates after increase to a certain number of bands. Another problem is the Hughes phenomenon that states that the ratio of the number of pixels with a known class identity (known as instances) and the number of bands must be maintained at or above some minimum value to achieve statistical confidence (Hughes, 1968; Cheg *et*

al., 2006). Reducing the number of bands helps to maintain this law. This is very important at the pre-processing stage of either multispectral or hyperspectral remote sensing data (Cheg *et al*., 2006). Among various searching algorithms are, Random selection, Best-first have been successfully used for searching for an accurate subset (Kohavi, 1995). For ensemble classification method, the features to be included in a given base classifier may be selected using genetic algorithms (Optiz, 1999), exhaustive search methods and a random selection of feature subsets (Ho, 1998). Bhattacharya, Divergence, Transformed divergence and Random search schemes were used as feature selection methods in this study.

2.11.1 Separability index

Bhattacharya distance which measures the similarity of two discrete or continuous probability distributions (Bhattacharya, 1948; Choi and Lee, 2003) approach was used as an evaluation function for exhaustive search. This approach provides the statistical distance between each of the class pairs for each possible subset of features (Bhattacharyya, 1943). It provides the best subset of spectral features to be used for the classification process. The Bhattacharyya distance defined in eq. (2.13) can be written in a *n*-dimensional diagonalized coordinate. Feature selection attempts to select a subset of bands system (Guo *et al*., 2008):

$$b=\frac{1}{8}\left[\mu_2-\mu_1\right]^T\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}\left(\mu_2-\mu_1\right)+\frac{1}{2}\ln\frac{\left|\left(\Sigma_1+\Sigma_2\right)/2\right|}{\left|\Sigma_1\right|^{1/2}\left|\Sigma_2\right|^{1/2}} \tag{2.14}$$

Where μ_1 and $\sum_1$ are the mean vector and covariance matrix of class *i*, respectively. The divergence separability index measures the distance between class means and Euclidean distance. Euclidean distance works using one band at a time.

According to Eastman (2001), divergence is defined as;

$$d_{ij} = 0.5\left\{\left(c_i - c_j\right)\left(c_i^{-1}\right)\right\} + 0.5tr\left\{\left(c_i^{-1} + c_j^{-1}\right)\left(m_i - m_j\right)\left(m_i - m_j\right)^t\right\} \tag{2.15}$$

(tr) represents trace, (-1) means the inverse and (t) depicts the transpose, (*i)* and (*j)* represent subscripts for the two signatures, (*C)* stands for covariance matrix and (*m)* is the vector of means. The transformed divergence can be obtained using the formula (Eastman, 2001):

$$t_{ij} = c(1 - e^a) \tag{2.16}$$

where $a = -d_{ij}/8$, c is a constant multiplier and *d* is the divergence.

2.12 Feature extraction model

Guyon and Elisseeff, (2003) describes feature extraction as the transformation of original features to construct a new feature space. Examples of linear feature extraction models are the Principal Component Analysis (PCA) (Hyvarynën *et al*., 2001), Independent Component Analysis (ICA) (Fukunaga, 1990) and Linear Discriminant Analysis (LDA) (Hyvarynën *et al*., 2001) among others. These algorithms minimize some criterion function like the mean square error (PCA), a class separability criterion (LDA) or an independence criterion (ICA)**.** Various

classification models use the linear features to build a classifier, obtaining improved computational efficiency and accuracy (Grana and d'Anjou, 2004).

Another approach to feature extraction is the linear spectral mixing analysis. A standard technique for spectral mixture analysis is linear spectral unmixing, (Hein and Chang, 2001; Plaza *et al*., 2004b), which assumes that the collected spectra at the spectrometer can be expressed in the form of a linear combination of endmembers weighted by their corresponding abundances.

2.12.1 Linear mixing model

Solving mixed different materials stirred up developing different algorithms to retrieve endmembers and the fractional abundance from mixed pixels. With reference to (Heinz and Chang, 2001; Keshava and Mustard, 2002; Plaza *et al*., 2004b; Sanchez *et al*., 2010; Abe *et al*., 2012), each pixel (vector) in a remotely sensed hyperspectral image (I) having n number of bands is denoted by:

$$X(i,j) = \sum_{z=1}^{p} \Phi_z(i,j).\, e_z + n(i,j) = E\,\Phi\ + n \qquad (2.17)$$

Where, X (i, j) is a vector (L x 1) at a distinct spatial value

$$= [x_1(i,j), x_2(i,j), x_3(i,j), \cdots\cdots\cdots, x_n(i,j)] \in \Re^n$$

$\mathfrak{R}$ = the real numbers in which pixel's response $x_k(i,j)$ at sensor channels

$k = 1, \cdots, n$ is included.

E_z = spectral response of endmember z at pixel $X(i,j)$

p = the total number of endmembers

$n(i,j)$ = noise vector and $\Phi_z(i,j)$ is a scalar value that represents the fractional coverage of the endmember vector e_z at pixel $X(i,j)$

Linear mixing model operates under two constraints of the fractional abundance Φ_z. These are;

(i) The abundance sum to one constraint (Heinz and Chang, 2001; Sanchez *et al*., 2010),

$$\sum_{z=1}^{p} \Phi_z(i,j) = 1. \tag{2.18}$$

(ii) Non-negativity constraint: All abundance should be no negative that is (Heinz and Chang, 2001; Sanchez *et al*., 2010)

$$\Phi_z(i,j) \geq 0, \text{ for all } 1 \leq i \leq E \tag{2.19}$$

This is the way of accounting for the full composition of a mixed pixel. In literatures, for fully constrained linear spectral mixing analysis, the first constraint imposed can be easily solved while the second has not been fully solved because disparities can be experienced and the solution requires numerical approaches (Heinz and Chang, 2001; Sanchez *et al*., 2010).

2.12.2 Linear unmixing procedure

Endmember extraction and fractional abundance can be obtained by using the following three procedures as shown in Figure 2.9. They are dimension reduction, endmember determination and fractional abundance (inversion).

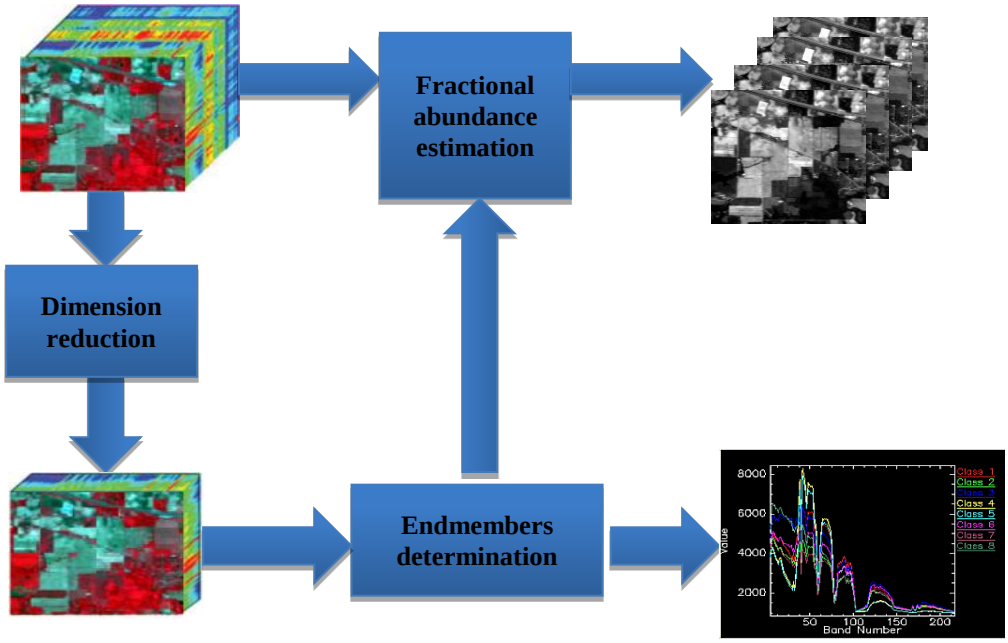


Figure 2.9: Spectral unmixing concepts

(Sanchez *et al*., 2010).

i. Dimension Reduction

Dimension reduction method is aimed at reducing the number of spectral bands in an image. This is not compulsory when dealing with spectral unmixing. The process is carried out to reduce the time used during processing hyperspectral data. Dimension reduction algorithm does not generate an image different from the original image. Instead, it finds a minimum representation of the original image that adequately keeps the original information for successful unmixing in the lower dimension. The algorithms are designed to reduce the error in the procedures action in the lower dimension (Keshava and Mustard, 2002). Among various algorithms developed for dimension reduction are the principal component analysis (PCA), maximum noise fraction (MNF) and the independent component analysis (ICA).

ii. Maximum Noise Fraction

Maximum noise fraction (MNF) transforms and principal component analysis (PCA) are used to determine the dimensionality reduction to reduce computational complexity and also to compact information in transformed components (Green *et al*., 1988; Chaudhry *et al*., 2006). These algorithms are second order statistic-based transforms. PCA has the limitation of not being able to always produce images that gradually reduce image quality with an increasing component number. PCA maximizes the variance while MNF maximizes signal-to-noise ratio, hence the components will show increasing image quality. This is done by carrying out a forward transform, determining the bands that have coherent images (looking at the images and the eigenvalues), and perform an inverse MNF transform using a spectral subset to include only good bands or smoothing the noisy bands before the inverse.

Let a remotely sensed hyperspectral image data set of n-bands with grey levels be given by eq. (2.20) (Green *et al*., 1988):

$$Z_i\,(x, y), \qquad i = 1, \cdots, n \tag{2.20}$$

Where, x, y gives the coordinates of the sample.

$$Z(x, y) = S(x, y) + N(x, y) \tag{2.21}$$

$Z^T(x, y) = \{Z_1(x, y), \cdots, Z_n(x, y)\}$, $S(x, y)$ and $N(x, y)$ are the uncorrelated signal and noise components of $Z(x, y)$ respectively.

$$Cov\{Z(x, y)\} = \sum = \sum_S + \sum_N \tag{2.22}$$

where $\sum_S$ and $\sum_N$ represent the covariance matrices of $S(x,y)$ and $N(x,y)$ respectively.

It is easy to define MNF transform in matrix form (Green *et al*., 1988):

$$Y(x,y) = A^T Z(x,y) \tag{2.23}$$

Where $Y^T(x) = \left(Y_1(x), \cdots, Y_n(x)\right)$ and $A = (a_1, \cdots, a_n)$

A significant difference between MNF transform and PCA is that,

(i) Since MNF depends on the signal-to-noise ratios, it is invariant under scale changes to any band,

(ii) $Z(x,y)$, $S(x,y)$and $N(x,y)$ are orthogonal.

In this study, MNF was adopted for dimensionality reduction in the endmember determination.

iii. Endmember determination

An endmember is defined as a spectrally pure pixel that portrays various mixed pixel in the image (Plaza *et al*., 2004b). For this study, the endmember extraction procedure was used to select the features used for the classification. The method of feature selection involves identifying the most discriminating measurements out of a set of *D* potentially useful measurements, where $d \leq D$. Endmember extraction has been widely used in hyperspectral image analysis due to significantly improved spatial and spectral resolution provided by hyperspectral imaging sensor also known as imaging spectrometry (Chang *et al*., 2006; Chaudhry *et al*., 2006).

Identification of image endmember is a crucial task in hyperspectral data exploitation, especially classification (Martinez *et al*., 2006). When the endmembers have been selected, various methods can be used to construct their spartial distribution, associations and fractional abundances. For real hyperspectral data, various tools (algorithms) developed to execute the task of locating appropriate endmembers include, Pixel Purity Index (PPI), N-FINDR and Automatic Morphological Endmember Extraction (AMEE) (Plaza *et al*., 2004a; Chang *et al*., 2006; Chaudhry *et al*., 2006; Martinez *et al*., 2006).

Pixel Purity Index generates a large number of *n*-dimensional vectors called "skewer" (Boardman *et al*., 1995; Plaza *et al*., 2008), through the dataset. N-FINDR fully automated method locates the set of pixels with the largest possible volume by "inflating" a simplex within the image data (Winter, 1999; Plaza *et al*., 2004b). On the other hand, AMEE uses a morphological method where spectral and spatial information are equally required to derive endmembers. For the purpose of this research, PPI available in Environment for Visualizing Images (ENVI) was used to retrieve endmembers from the data.

iv. Pixel Purity Index

Pixel Purity Index (PPI) developed by Boardman *et al*., (1995) searches through a set of vertices of a convex geometry in a certain dataset that should present pure signatures present in the data (Chaudhry *et al*., 2006). The algorithm has a supervised characteristic nature. To generate the endmembers from the data, first, a "noise whitening" and dimensionality reduction step is performed using MNF transform (Boardman *et al*., 1995; Plaza *et al*., 2004b). Secondly, a Pixel purity score is obtained in the image cube through random producing lines in the *n*-dimensional space containing the MNF-transformed data. The spectral points are projected on the

lines and the points at the extremes of each line are counted. Bright pixels in the PPI image generally are image endmembers. The highest-valued of these pixels are input into the *n*-dimensional Visualizer for the clustering process that develops individual endmember spectra.

v. Fractional abundance

The Generalized Reduced Gradient (GRG) method is a well-known algorithm for the solution of optimization problems with non-linear objective function and constraints (Abadie and Carpentier, 1969; Haggag, 1981). From equation (2.17), $\Phi_z(i,j)$ the fractional abundance can be derived using generalized reduced gradient (GRG) which is defined as (Haggag, 1981; Mouatasim, 2010)

$$\text{Optimize: } y(\mathrm{X}) \tag{2.24}$$

Subject to:

$$f_1(\mathrm{X}) = 0 \qquad for\ i = 1,2,\cdots,m$$

$$\mathrm{X} \in \Omega$$

where y: $\mathbb{R}^n \rightarrow \mathbb{R}$ and $f_1: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are continuously differentiable, and $\Omega \subset \mathbb{R}^n$ is closed and convex set (for example, $\Omega = \Pi_{t=1}^{n}[a_{t,}b_t]$. The problem in the equation (2.17) can be numerically solved by using the generalized reduced gradient method (Plaza *et al*., 2004b; Mouatasim, 2010). This produces a sequence $\{\mathrm{X}^k\}_{k\geq 0}$, where X^0 is an initial possible point. For each k>0, a new possible point X^{k+1} is created from X^k using operator Q_k. The iteration is given by:

$$k \geq 0: \mathrm{X}^{k+1} = Q_k(\mathrm{X}^k) \tag{2.25}$$

We introduce in this thesis an application of the generalized reduced gradient approach to generate fractional abundance from hyperspectral imagery for land cover mapping.

2.13 Summary

A review of relevant reported literature on remote sensing, data collection, image resolutions and processing, and different classification methods were summarized. Feature extraction model base on spectral unmixing was also presented.

A gap in knowledge that is common to reported studies on ensemble is the lack of information on ensemble size and the resulting classification on land cover mapping. Also, it is not established if there is any correlation between classification accuracy and diversity measures.

On spectral unmixing, studies also revealed that, a number of techniques have been developed for unconstrained, partially constrained and fully constrained linear spectral unmixing which can be computationally expensive (Sanchez *et al*. 2010). A major gap in knowledge that is common is, out of the two constraints imposed on fully constrained linear spectral mixing analysis, the second constraint, that is, the abundance fractions of information present in an image pixel should be nonnegative has not been fully solved. This is because disparities can be experienced and the solution requires numerical approaches. For these reasons, this work focuses on investigating the effect of ensemble size and correlation between classification accuracy and diversity measures. The issue on the second constraint imposed on fully constrained linear spectral analysis was also investigated.

CHAPTER 3

Design methodology

3.1 Introduction

The process of extracting information from multispectral and hyperspectral images which are powerful tools in remote sensing are similar but there are procedures to undertake before analyzing hyperspectral datasets. The final performances of the datasets depend mainly on the processing systems (either hardware or software) of the collected data. This section presents various investigations conducted on multispectral and hyperspectral images for land cover classification. All investigations were conducted using supervised classification method in which each classifier is trained with pixels extracted from the region of interest on the selected features (bands) set.

3.2 Research instruments

3.2.1 Software

There are various image processing software that are available among them are; SPAM; ISIS,.Tetracorder, SIPS (CU/CSES); MULTISPEC (Purdue) and ENVI. According to Boardman *et al*., 2006, using the available software on datasets, there is evident that not all information in modern hyperspectral data is tapped. This is because, the performance of each software on a dataset depends on the nature of the dataset been investigated. Therefore, there is need for new algorithms and software implementations

Multispec, a data analysis software system that originated in the LARSYS multispectral image data analysis system (Landgrebe and Biehl, 2001) was used to analyze the hyperspectral image data used. The software is a machine learning tool that has the ability of processing remotely sensed imagery. Among the various functions of the software is the Bhattacharyya seperability index, which was one of the separability index used in the research. The tool was used to separate unwanted bands from the dataset and was also used to combine different bands for the classification process in Matlab (MathWorks Inc., 2010).

Matlab (Matrix laboratory) software is a high level language and interactive environment software used to perform computational intensive task. This software was used for the classification procedure. The algorithm programmed in the software was Support Vector Machine one against one. The predicted results from the classifiers were imported into IDRISI for land cover map generations.

IDRISI, an integrated geographic information system (GIS) and remote sensing software was used for image processing. The software developed by Clark University has the capability of analyzing and displaying digital geospatial information (Eastman, 2006). The software has a comprehensive suite of image processing tools, making it an excellent choice for land cover mapping application with remotely sensed data. The software was used to extract regions of interest from the datasets for classification and was used to generate land cover maps using the results predicted by the classifiers.

Environment for Visualizing Image (ENVI) software capable of extracting information from geospatial imagery was used to process the hyperspectral imagery

data for the purpose of the study. The image processing includes, dimension reduction, noise reduction and extraction of endmembers.

Excel solver, an optimization tool written by Microsoft for use in Microsoft Excel. This was used to obtain the fractional abundance from the spectral signature obtained from the endmembers using Generalized Reduced Gradient (GRG) approach for classification. Solver model in excel has implicit constraints that are required to execute the GRG approach adopted in this study.

WEKA (Waikato Environment for Knowledge Analysis) is a well-known suite of machine learning software written in Java. The software was developed at the University of Waikato, New Zealand. WEKA workbench has a collection of visualization tools and algorithms for remotely sensed imagery data analysis and predictive modelling (Witten and Frank, 2005). The software is freely available and has a comprehensive collection of data preprocessing and model techniques. Some of the classification algorithms available in WEKA were used for classification procedures in this study.

3.2.2 Methods of creating ensembles

Many methods for constructing ensembles have been developed among then are;

(i) Bayesian Voting: This is done by enumerating the Hypotheses. In a Bayesian probabilistic setting, each hypothesis defines a conditional probability distribution. It primarily addresses the statistical component of ensembles. According to Dietterich (2001), the Bayesian approach does not address the computational and representational problems in any significant way

(ii) Manipulating the Training Examples: This involves manipulation of the training examples to generate multiple hypotheses. This technique works especially well for unstable learning algorithms, that is, algorithms whose output classifier undergoes major changes in response to small changes in the training data. Examples are Decision-tree, neural network, and rule learning algorithm. Training set can be manipulated using various methods such as;

- Bagging (as discussed in chapter 2),

- constructing the training sets by leaving out disjoint subsets of the training data

- AdaBoost algorithm developed by Freund and Schapire, (1996): This manipulates the training examples to generate multiple hypotheses.

(iii) Manipulating the Input Features: This method manipulates the set of input features available to the learning algorithm. This technique only works when the input features are highly redundant.

(iv) Manipulating the Output Targets: This manipulates the output values that are given to the learning algorithm. According to Dietterich and Bakiri (1995), the technique improves the performance of both the C4.5 decision tree algorithm and the backpropagation neural network algorithm on a variety of difficult classification problems.

(v) Injecting Randomness: This method injects randomness into the learning algorithm. While this is perhaps the most common way of generating ensembles of neural networks, manipulating the training set may be more effective.

All the methods described were used in the research study to investigate the impact of ensemble over the single classifier in the ensemble. The investigations conducted concludes that injecting randomness performed better as compared with other methods

3.3 Land cover mapping using ensemble features selection methods on Kampala imagery

The focus of this investigation is an ensemble feature selection that entails ensuring diversity through training a given classifier on different features, which in remote sensing would be the different sensor bands (Gidudu *et al*., 2008a). By varying the feature subsets used to generate the ensemble classifier, diversity is ensured since the base classifiers tend to err in different subspaces of the instance space (Tsymbal *et al*., 2005; Oza and Tumer, 2008). Some of the techniques used to select features to be used in ensemble systems include genetic algorithms (Opitz, 1999), exhaustive search methods and a random selection of feature subsets (Ho, 1998; Marwala, 2009)).

3.3.1 Study area

The study area for this research was Kampala, the capital of Uganda (Gidudu, 2006). The optical bands of a 2001 Landsat image (column 171 and row 60) formed the dataset from which ensembles were created and investigated. There were five land cover classes of interest considered: water, built up areas, thick swamps, light swamps and other vegetation. Samples for classification were obtained from the region of interest (Gidudu *et al*., 2008a).

3.3.2 Design procedure

Ten ensembles were created, each with five base classifiers, the number five having been arbitrarily chosen. For each ensemble, the base classifiers were made up of the bands which yielded the best separability indices (best five band combinations in this case). Three separability indices were used, namely, Bhattacharyya distance, divergence and transformed divergence.

For each base classifier and corresponding ensemble, a land cover map was derived using Gaussian Support Vector Machines (Boser *et al*., 1992; Vapnik, 1999; Witten and Frank, 2005, Gidudu *et al*., 2008a). The land cover map for each ensemble was consequently derived through majority voting primarily due to its simplicity (Valentini and Masulli, 2002). Each of the derived land cover maps was compared with ground truth data to ascertain its classification accuracy. In order to determine the diversity of each ensemble, the kappa analysis was used to give the measure of agreement between the constituent base maps and ultimately the overall ensemble diversity. The influence of diversity on land cover classification accuracy for each ensemble was evaluated by comparing the derived land cover classification accuracies with the derived diversity measures.

3.4 Ensemble feature selection for hyperspectral imagery

3.4.1 Data description

The hyperspectral dataset used in this study was sourced from the AVIRIS sensor and represents Indiana's Indian Pines in the United States of America (Landgrebe, 1998). The dataset is as shown in Figure 3.1. It is a freely accessible online dataset which comes with accompanying ground truth data (Figure 3.2). Of the 224 bands, 4 were discarded because they contained zeros and of the remaining 180 bands were

used in this research. The rest of the bands were left out because of being affected by atmospheric distortion (Bazi and Melgani, 2006).

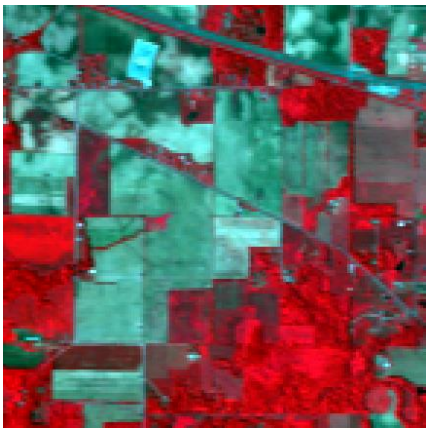


Figure 3.1 Indiana Pine hyperspectral image

(Landgrede, 1998)

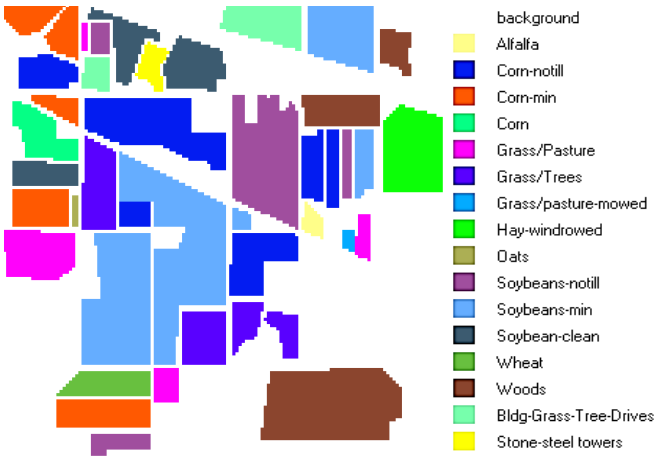


Figure 3.2 Indiana Pine Ground truth & the labels

(Landgrebe, 1998)

The classes of interest are: alfalfa (alf), corn-notill (cnt), corn-minimum till (cmn), corn, grass/pasture (gp), grass/trees (gt), grass/pasture-mowed (gpm), hay-

windrowed (hw), oats, soybeans-notill (snt), soybeans-minimum till (smn), soybean-clean (scl), wheat, woods, building-grass-tree drives (bgtd) and stone-steel towers (sst). These classes were selected in reference to the ground truth data. This data was also used to carry out investigations in experiments 3.5 and 3.6 with the same number of spectral bands and class labels (Gidudu *et al*., 2008b; Abe *et al*., 2010).

3.4.2 Research design

Based on Chen *et al*., (2007)'s categorization, this investigation focuses on the ensemble approach dependent on one learning algorithm (in this case Gaussian SVMs), with diversity being enforced through using different feature (band) combinations. Two ensemble feature selection techniques were used, namely exhaustive search and random selection of feature subsets. The evaluation function of the exhaustive search was the Bhattacharyya Distance separability index (Bhattacharyya, 1943; Gidudu *et al*., 2008b).

The results of the base classifiers in each ensemble were combined using two methods; majority voting and an adaptation of Cross Validation Majority (CVM) also called single best. In CVM, the cross validation method is used as a basis for selecting the best out of the whole ensemble. In this study, this was modified to consider the final classification results of each base classifier instead. For comparison, another ensemble was derived from sequential grouping subsequent bands into 10 base classifiers. That is, bands 1-18 made up the first base classifier, bands 19 – 36 the second base classifier and so on, making a total of 10 base classifiers for all the 180 bands. For each base classifier and corresponding ensemble, classification was carried out in Matlab with the results being imported into Idrisi Andes for data integration and generation of a land cover map. Classification accuracies were then calculated for each derived land cover map, by

making comparisons between the predicted output from the base and ensemble classifiers and the ground truth data. These results were then used as the basis upon which ensemble feature classification and its corresponding effect on land cover mapping was evaluated.

3.5 Investigating the effect of ensemble size on classification `accuracy

This study investigates the interplay between the ensemble size and classification accuracy. The online Indiana Pines hyperspectral dataset was used.

3.5.1 Design procedure

In this research, two aspects of ensemble size were considered. The number of base classifiers per ensemble and the number of features (bands) per base classifier/ensemble. A total of seven (7) ensembles were generated through an exhaustive search process, using the Bhattarcharya distance as separability index (Bhattacharyya, 1943). The seven ensembles had the following distribution of bands per ensemble: 2, 4, 6, 8, 10, 12 and 14. For each ensemble, base classifiers were constituted by considering the band combinations with the best separability index (Gidudu *et al*., 2009a). In each ensemble, the number of base classifiers was cumulatively increased and for each collection of base classifiers corresponding classification accuracy was determined. In all cases, majority voting was adopted as the consensus rule and Gaussian support vector machines (one against one) were used as the base classifier for the classification process.

3.6 Random ensemble features Selection for Land Cover Mapping

This research investigates the effect of ensemble size on classification accuracy of land cover mapping using the hyperspectral AVIRIS data set of the Indian Pine. The research used random ensemble feature selection to investigate the effect of ensemble size on the consequent classification accuracies for land cover mapping (Gidudu *et al*., 2009b).

3.6.1 Design procedure

Two approaches were used to investigate the effect of ensemble size on classification accuracy. In the first case, classification accuracy was related to increasing numbers of features per base classifier and in the second case accuracy was related to increasing numbers of base classifiers per ensemble. A total of seven ensembles were constituted with increasing amount of features per base classifier. The ensemble had two features per base classifier in the sequence 2, 4, 6, 8, 10, 12 and 14. Similarly, for each ensemble predictions were cumulatively added and the ensemble classification consequently derived to give final decision. Image classification was conducted through the use of Support Vector Machines (SVMs) using one against one approach.

3.7 Experimental comparisons of supervised learning classifiers for land cover classification of hyperspectral imagery

This aspect of the study was aimed at identifying a set of reference signatures (also known as endmembers) to model the reflectance for land cover classification.

3.7.1 Data description

The data used for this study was AVIRIS data set of the Indian Pine (AVIRIS Indiana dataset, 1992) test site acquired over a mixed agricultural/forest region in Northwestern Indiana has been researched extensively (Landgrebe, 1998; Plaza *et al*., 2004a; Guo *et al*., 2006; Plaza *et al*., 2009). This dataset is a descriptive of problems associated with hyperspectral image analysis and classification scenario, in the sense that, the primary crops of the area (corn and soybeans) were very early in their growth cycle, with only a 0.05 canopy cover. Distinguishing the differences among the main crops under these conditions can be very difficult (Landgrebe, 1998; Plaza *et al*., 2009). The pixels in the dataset are labelled as belonging to one of 16 classes of vegetation. Not all the pixels are labelled because they correspond to uninteresting regions or were too difficult to label. Noisy channels and water absorption channels were removed (channels 1–3, 104–112, 148–165, and 217–224). About atmospheric correction, the image has been processed to remove path radiance including the light scattered by the interaction between surface and the atmosphere (Camps-Valls and Bruzzone, 2005; Nascimento and Bioucas-Dias, 2005). The remaining 186 channels were used for this study. Each image has size 145 x 145 = 21 025 pixels. The study site has sixteen class labels out of which nine were used for the investigation because the remaining seven class labels has an insufficient number of pixels available (Lennon *et al*., 2002; Melbani and Bruzzone, 2004; Demir and Ertürk, 2007; Qi and Huang, 2007). The land cover classes used were, corn-notill (cnt), corn-minimum till (cmn), grass/pasture (gp), grass/trees (gt), hay-windrowed (hw), soybeans-notill (snt), soybeans-minimum till (smn), soybean-clean (scl), woods and background (Abe *et al*., 2012).

3.7.2 Problem formation

The task of land cover classification can be generally formulated as a linear spectral unmixing problem. The linear spectral unmixing is a sub-pixel classification process that decomposes mixed pixels and determines the combination of fractional abundances. The basic idea behind the linear spectral mixture analysis is that every image pixel is a mixture of different endmember spectral responses. The spectrum recorded by the sensor is a linear combination of endmember spectral responses (Kärdi, 2007). The linear unmixing model can be mathematically expressed as (Sanchez *et al*., 2010; Zhang *et al*., 2011):

$$x(i, j) = \sum_{k=1}^{P} a_k(i, j).e_k + n(i, j) \tag{3.1}$$

The component $x(i, j)$ is a spectral response vector, $a_k(i, j)$ is a scalar value representing fractional abundance of endmember vector e_k at pixel $x(i, j)$, $n(i, j)$ is a vector that denotes the spectral band error and P is the total number of endmember. Equation (3.1) operates under two physical constraints on fractional abundances to account for the full composition of a mixed pixel. These constraints are non-negativity of all fractional abundance values and fractional abundance values must sum to unity (Heinz and Chang 2001; Sanchez *et al*. 2010).

Previous efforts on linear spectral unmixing problem (Sanchez *et al*. 2010; Zhang *et al*. 2011; Iordache *et al*. 2011) have investigated the Least Square (LSU) method (Heinz and Chang, 2001) to estimate a set of fractional abundances as follows.

$$a(i, j) = (e^T e)^{-1} e^T x(i, j) \tag{3.2}$$

In Equation (3.2), e^T is the transpose of the matrix e of endmember spectral responses and e^{-1} represents the inverse matrix of e matrix.

In general, the fractional abundances obtained from Equation (3.2) can only satisfy the sum to unity constraint, but the non-negativity of fractional abundances cannot always be guaranteed. The solutions obtained by the LSU method are therefore generally not optimal in terms of material quantification (Heinz and Chang 2001; Du *et al.* 2008; Sanchez *et al*. 2010). The linear spectral unmixing problem has to be formulated as an optimization problem that minimizes the spectral band error. The reason for the minimization is to obtain optimal fractional abundances that simultaneously satisfy the two changing constraints Δ_1 and Δ_2 with respect to the spectral coordinate (i, j). The following fully constrained linear spectral unmixing optimization problem has to be solved by finding a set of fractional abundances $a(i, j)$ that minimizes the spectral band error $a(i, j)$ in Equation (3.1).

Minimize

$$J(a) = \Delta\{(x(i, j) - a(i, j)e)^T (x(i, j) - a(i, j)e)\} \tag{3.3}$$

Subject to

$$\Delta_1 = \left\{ a(i, j) \mid \sum_{k=1}^{p} a_k(i, j) - 1 = 0 \right\} \tag{3.4}$$

$$\Delta_2 = \{a(i, j) \mid 0 \le a_k(i, j) \le 1\} \tag{3.5}$$

In Equation (3.3), $J(a)$ is the objective function to be minimized, Δ_1 is the abundance sum-to-one constraint and Δ_2 is the abundance nonnegativity constraint with respect to the spatial coordinate (i, j). The PPI algorithm is first applied to extract endmember spectral responses from the hyperspectral image before Equations (3.3) can be solved for optimal fractional abundances using the GRG algorithm (Abadie and Carpentier 1969; Lasdon *et al*. 1974; Su and Lii 1995). Once a set of endmember spectral responses $e = \{e_k\}_{k=1}^{p}$ is determined, the corresponding fractional abundances $a(i, j) = \{a_k(i, j)\}_{k=1}^{p}$ in a specific pixel, vector $x(i,j)$ of the hyperspectral image can be estimated by using the GRG algorithm.

The purpose of the land cover classification is to evaluate the performance of the RF and SVM classifiers per class basis. The GRG algorithm is used to obtain the estimated numeric values of the endmembers' fractional abundance.

3.7.3 Pre-processing procedure

The input hyperspectral image has to be taken through five steps to obtain the desired classification result. Figure 3.3 shows the block diagram of the land cover classification process implemented in this study (Sanchez *et al*., 2010; Abe *et al.*, 2012).

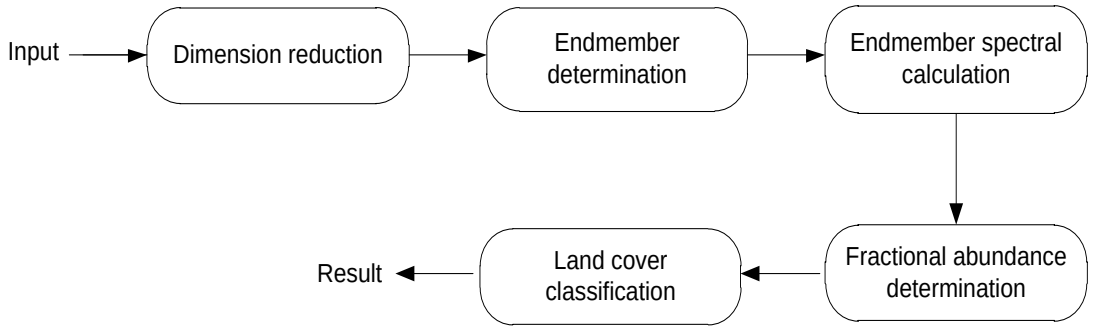


Figure 3.3: Hyperspectral image classification procedure

Image Dimension Reduction

Noise whitening and dimensionality reduction were performed on the dataset using the maximum noise factor (MNF) transform to reduce the number of spectral bands in the hyperspectral image. The endmember spectral responses $e_k, k = 1,2,...,P$ in Equation (3.1) were determined by using PPI algorithm available in Environment for Visualizing Images (ENVI) software.

Endmember Spectral Response Determination

The PPI method efficiently handles hyperspectral images as it provides a convenient and physically motivated decomposition of an image in terms of relatively few components (Theiler *et al*., 2000). The algorithm searches through a set of vertices of a convex geometry in the dataset to present pure signatures in the data. This was accomplished by randomly generating lines in the N-Dimensional space (an ENVI's visualizer that provides an interactive tool for finding endmember spectral responses) containing the MNF transformed data. In an N-Dimensional component space, it is assumed that endmember (P) spectral responses will occur at the vertices of the

hyper-solid or geometric shape bounding the pixel values in that space. The spectral points were projected on the lines and the points at the extremes of each line were selected by drawing a polygon around a few of the extreme data points to create the endmember (P) spectral responses. The highest-value of these pixels were input to the ENVI visualizer for the clustering process that developed the individual endmember (1,2,3,….P) spectral response. The result is not a single map representation as in thematic image classification, but a series of images, each having the size of the original image (Settle and Drake, 1993; Adams *et al*., 1995; Van der Meer and Jia, 2012).

Fractional Abundance Estimation

The GRG optimization algorithm was executed to estimate per pixel fractional abundances by using spectral responses results obtained from PPI algorithm. Using the endmember set produced by the PPI, a set of endmember numeric values were generated as follows:

(i) The first step is to calculate the compute matrix $(e^T e)^{-1} e^T$, where $e = \{e_k\}_{k=1}^{p}$ is formed by the *P* endmember extracted from the PPI. Using excel solver for calculation, the compute matrix is multiplied by all the pixel vectors $x(i, j)$

(ii) The computed matrix calculated in step 1 is multiplied by each pixel $x(i, j)$ from the region of interest, thus obtaining a set of vectors $a_k(i, j)$, each containing the fractional abundances of the *P* endmembers in each pixel.

The new values obtained were used to train and test SVM and RF classifiers for the land cover classification procedure.

Land Cover Classification

The purpose of thematic image classification is to represent the land cover in terms of a number of fixed classes where each image pixel represents a unique endmember which in turn is used to produce a single map representation. Table 3.1 shows the selected region of interest (ROI) from the Indiana pine dataset used for the classification procedure. The new generated numeric values obtained from the GRG algorithm in the two steps mentioned above were used for the classification procedure. The WEKA (Garner 1995) data mining software is selected to build SVM, RF, neural network, C4.5 and bagging classifiers. The five classifiers are experimentally compared for land cover classification.

Table 3.1. Number of pixels extracted from the ROI

Endmembers	Number of Pixels
Corn-notill (cnt)	359
Corn-minimum till (cmn)	305
Grass/pasture (gp)	264
Grass/trees (gt)	339
Hay-windowed (hw)	279
Soybeans-notill (snt)	350
Soybean-clean (scl)	203
Soybeans-minimum till (smn)	425
Woods	400
Background (bg)	300
Total number of pixels	3224

3.7.4 Conclusion

This study aimed to evaluate the performance and a comparison of the supervised classifiers for land cover classification of a heterogeneous area of Indian Pine test site in Northwestern Indiana, USA. Spectral unmixing scheme for the hypespectral image procedure was used for preprocessing the dataset. Environment for visualizing image was used to extract endmember and to generate the spectral values of the region of interest. Generalized Reduced Gradient (GRG), an optimization technique with constraints was used to estimate fractional abundance of each pixel in the region of interest. The computed normalized fractional abundance values obtained were multiplied by all the pixel values extracted from the region of interest for the classification procedure. The experiments have established that ensemble method obtains better accurate classifiers by combining less accurate ones.

3.8 Hyperspectral Image Classification using Random Forests and Neural Networks

This work considers the problem of land cover classification of hyperspectral images by using a linear spectral mixture analysis technique, which is a commonly accepted approach to mixed-pixel classification. The objectives of the study are to (i) identify a collection of spectrally pure constituent spectral, which are referred to as the endmembers (Hein and Chang, 2001; Martinez *et al*., 2006). Thereafter, we express the measured spectrum of each mixed pixel as a linear combination of endmembers weighted by fractional abundances that indicate the proportion of each endmember present in the pixel (Martinez *et al*., 2006; Sanchez *et* al., 2010), (ii) explore Generalized Reduced Gradient (GRG) optimization algorithm to estimate the fractional abundance in the dataset thereby obtaining the numeric values for land cover classification (Abadie and Carpentier, 1969; Lasdo *et al*., 1974; Abe *et al.*,

2012) and (iii) to experimentally compare the performance of random forests and neural network classifiers to examine the suitability of GRG algorithm for solving land cover classification problem. The procedure used in section 3.7 was adopted to execute this experiment.

3.8.1 Data description

Figure 3.8 shows the input airborne hyperspectral image of the Washington, D.C. dataset (Landgrebe, 2003). The sensor used to measure pixel response in 210 bands in the 0.4 to 2.4μm region of the visible and infrared spectrum. It has 1208 scan lines with 307 pixels in each scan line, which is approximately 150 Megabytes. Bands in the 0.9 and 1.4μm region where the atmosphere is opaque have been omitted. The remaining 191 spectral bands are used for this study. The dataset contains seven ground cover types, namely: Roofs, Street, Path, Grass, Trees, Water and Shadow.

Fig 3.8: Hyperspectral image of Washington D. C. Mall
(Landgrebe, 2003).

Accompany the dataset is a copy of the file labelled *dctest.project,* which describes the land cover types used for the experimental procedure.

For this study, the problem formulated in experiment 3.7 was used to ascertain the authentication of our proposed GRG algorithm on the Washington DC mall dataset. The design methodology entails taking the input dataset through four stages to obtain desired classification results. The dataset was taken through the process of dimension reduction after which the endmembers were determined. Thereafter the fractional abundance of the endmembers was estimated and the numeric values obtained were used train the random forests and neural network classifiers for the land cover classification.

Table 3.2 displays the land cover classes and the number of pixels extracted from the original image based on the ROI. The dataset was input into the ENVI visualizer for the clustering process that develops individual endmember spectral. The pixels extraction mechanism enables the image spectral to accurately account for any errors in atmospheric correction.

Table 3.2: Number of pixels extracted from ROI

Classes	Number of Pixels
Roof	724
Paths	211
Water	703
Street	404
Trees	398
Shadow	97
Grass	818
Total	3355

3.8.2 Conclusion

This study explored linear spectral mixture analysis technique to solve the mixed-pixel classification problem associated with land cover classification of hyperspectral images. The study identified a collection of spectrally pure constituent spectral called endmembers in the hyperspectral dataset of the Washington DC mall. Six endmembers were generated from the study. These are used to obtain normalized numeric values of the fractional abundance generated using the spectral signatures of the land cover label. An ensemble method has three fundamental attributes (as discussed in chapter 2), which contributed to successful accuracy performance over the single classifier in the ensemble. These fundamental attributes are lacking in single classifier in the ensemble and this is the reason for the poor performance.

CHAPTER 4

Results and discussions

This section presents the classification results and findings from different experiments carried out in this study as a result of the instruments and methods used on remote sensing data for land cover classification.

4.1 Ensemble classification with Kampala imagery

Tables 4.1 – 4.4 give summaries of the results depicting the ensembles constituted depends on the separability index used, the respective base classifier classification accuracy assessment and the consequent ensemble classification accuracies. The relationship between ensemble classification accuracy and diversity was investigated by determining the correlation between ensemble classification accuracy and the agreement measure which in this case was the Kappa value. This was computed by averaging the in-ensemble pair-wise kappa values of the base classifiers measured against each other. In order to get a better appreciation on the in-ensemble diversity, the variance was also computed for the computed pair-wise kappa values. The calculated results for each separability measure are also presented in Tables 4.1 – 4.4 (Gidudu *et al.*, 2008).

Table 4.1: Separability index classification results using Bhattacharyya distance

Ensemble	B1		B2		B3	
	Bands	Acc	Bands	Acc	Bands	Acc
	5, 6	0.8565	3, 5, 6	0.8718	3, 4, 5, 6	0.9269
	3, 5	0.8496	1, 2, 3	0.3841	2, 3, 5, 6	0.8794
	2, 3	0.5395	4, 5, 6	0.9191	1, 2, 3, 5	0.8599
	4, 5	0.9030	3, 4, 5	0.9201	2, 3, 4, 5	0.9295
	1, 3	0.2220	2, 3, 5	0.8591	1, 2, 3, 4	0.8994
Accuracy	0.8922		0.8940		0.9160	
SB	0.9030		0.9201		0.9295	
Diversity	0.42		0.57		0.70	
Variance	0.03		0.02		0.01	

Table 4.2: Separability index classification results using divergence

Ensemble	D1		D2		D3	
	Bands	Acc	Bands	Acc	Bands	Acc
	2, 4	0.7921	2, 3, 4	0.8964	3, 4, 5, 6	0.9269
	1, 5	0.7283	1, 2, 4	0.8855	2, 4, 5, 6	0.9246
	1, 4	0.7750	4, 5, 6	0.9191	1, 4, 5, 6	0.9211
	3, 4	0.8814	1, 3, 4	0.8574	1, 3, 5, 6	0.8667
	2, 5	0.7569	2, 4, 5	0.9236	2, 3, 5, 6	0.8794
Accuracy	0.8559		0.9007		0.9237	
SB	0.8814		0.9236		0.9269	
Diversity	0.47		0.70		0.72	
Variance	0.05		0.01		0.03	

Table 4.3: Separability index classification results using transformed divergence

Ensemble	T1		T2		T3	
	Bands	Acc	Bands	Acc	Bands	Acc
	2, 4	0.7921	1, 2, 4	0.8855	1, 2, 5, 6	0.8678
	3, 4	0.8599	2, 3, 4	0.8964	1, 3, 5, 6	0.8667
	4, 6	0.9017	4, 5, 6	0.9191	2, 3, 5, 6	0.8794
	1, 4	0.7750	2, 4, 6	0.9175	3, 4, 5, 6	0.9269
	5, 6	0.8565	1, 3, 4	0.8574	1, 4, 5, 6	0.9211
Accuracy	0.8680		0.8982		0.9049	
SB	0.9017		0.9191		0.9269	
Diversity	0.44		0.71		0.65	
Variance	0.04		0.01		0.03	

Table 4.4: Ensemble using no Seperability measure

Bands	Accuracy %
1,2,3,4,5	0.93
1,2,4,5,6,	0.93
1,3,4,5,6	0.88
2,3,4,5,6	0.93
1,2,3,4,6	0.93
Accuracy	0.93
Single Best	0.93
Diversity	0.79
Variance	0.02

Where, Acc. – Accuracy: SB – Single best: SI – Seperability: Diversity – Diversity measure (Agreement): Variance – Diversity measure (variance)

Intuitively, the more diverse the ensemble, the lower the agreement between the classifiers and consequently the lower the consequent kappa values. By extension, the more diverse the ensemble, the bigger the variance between the in-ensemble pair-wise kappa values. The Table also reveals that ensemble accuracy increases as the agreement between the base classifiers increases and as the variance between the base classifier output decreases. In effect, this would ideally imply that the ensemble classification accuracy would increase if there is more agreement between the base classifier outputs. The contradiction this imputes is that, to get higher ensemble classification accuracy there is a need for less diversity among the base classifiers.

4.1.1 Prediction analysis

From the summary of the ensemble classification results Tables 4.1 – 4.4, it can be observed that, for all the ensembles, the ensemble classification accuracy was better than many of the base classifiers, whereas in no case was it better than the best classifiers with the ensemble. It is, however, critical to note, and the possibility is indicated here and reported elsewhere (e.g. Bruzzone and Cossu, 2004), that whereas the ensemble classification may not be more accurate than all of the base classifiers used in its construction (Foody *et al*., 2007), it certainly reduces the risk of making a particularly poor selection (Polikar, 2006). Table 4.1 – 4.4 also shows that across all ensembles, the respective classification accuracy increased as the size of the base classifiers increased.

4.1.2 Binomial test of significance between ensembles

This is further confirmed from Table 4.5 depicting the binomial tests of significance of the between ensemble classification accuracies. In the simple case of determining if there is a difference between two classifications (2 sided test), the null hypothesis

(Ho) that there is no significant difference will be rejected if |Z| > 1.96 (Congalton and Green, 1998). For each separability index used, increasing the number of features in the base classifiers in general significantly increased the ensemble classification accuracy. The ensemble (E) with five features per base classifier was seen to be significantly better than all the other ensembles apart from D3, where the difference was deemed insignificant. From the results, nothing conclusive can be deduced regarding which of the used separability indices is best suited as a basis upon which to build ensembles.

Table 4.5: Binomial Test of significance between ensembles

	B1	B2	B3	D1	D2	D3	T1	T2	T3	E
B1	-									
B2	0.44	-								
B3	6.06	5.62	-							
D1	8.20	8.64	14.22	-						
D2	2.09	1.65	3.97	10.28	-					
D3	8.17	7.74	2.12	6.31	6.09	-				
T1	5.58	6.02	11.61	2.63	7.67	13.71	-			
T2	1.47	1.03	4.59	9.67	0.62	6.71	7.05	-		
T3	3.15	2.71	2.91	11.34	1.06	50.3	8.72	1.68	-	
E	9.65	9.22	3.61	17.76	7.57	1.49	15.17	8.19	6.52	-

Figures 4.1 – 4.3: shows the land cover map of the ensemble classification results.

Appendix A presents other generated maps from the investigation. It can be seen from the map that all the land cover classes are properly labelled.

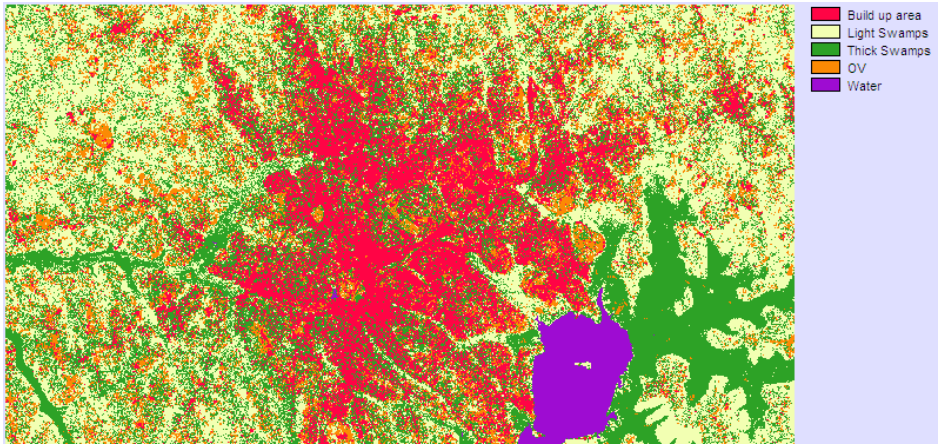


Figure 4.1: Map generated with BSI using 2 bands

Figure 4.2: Map generated with DSI using 2 bands

Figure 4.3: Map generated with TDSI using 2 bands

4.1.3 Conclusion

The results bring to the fore the challenge that comes with including diversity measures in ensemble classification research. Clearly its use in determining diversity for land cover mapping is counter intuitive. The problem may stem from using the classifier output as the basis upon which to measure diversity. Whereas diversity, as defined in ensemble classification research, is premised on having decision boundaries which err differently, using outputs to determine the measure of diversity presupposes that using different decision boundaries would yield different results. In the case of ensemble feature selection, base classifiers from different features certainly result in decision boundaries which err differently (and hence exhibit diversity), however, their final classification outputs are similar as the higher coefficients of correlation depict. Hence, base on the outputs as a measure of diversity clearly gives a poor reflection of how diverse the ensemble is (Gidudu *et al*., 2008a).

In their concluding remarks, Shipp and Kuncheva (2002) posit that the quantification of diversity and its use in determining diversity in ensembles will only be possible when a more precise formulation of the notion of diversity is obtained. Until then different heuristics will have to be employed. Whereas ensemble classification presents a unique approach to land cover mapping, the quantification of diversity and its consequent influence in determining the type of ensembles is clearly still open for research. This is a future work that needs to be addressed.

4.2 Ensemble feature selection for hyperspectral imagery

Table 4.6 shows the results of the different ensembles considered. The classification accuracy is given in terms of the Kappa coefficient of agreement (Cohen, 1960), which is a measure of how well the derived map compares with ground truth data. It ranges from 0 to 1 with 0 implying no agreement between predicted land cover and ground truth, and 1 indicating complete agreement. All the ensembles had ten base classifiers, the figure ten having been arbitrarily chosen. The base classifiers in Ensembles 1, 2, 3 and 4 consisted of 10, 14, 18 and 18 features (bands) respectively, each with different band combinations (feature configurations). Ensembles 1 and 2 were derived from an exhaustive search strategy, with the ten best base classifiers being selected based on their separability indices. Ensemble 3 was constituted by sequentially arranging the 180 bands into ten base classifiers, each with 18 features. On the other hand, all the features constituting the base classifiers in Ensemble 4 were randomly selected. Table 4.6 revealed that in all cases single best had better results than majority voting. It is also observed that in general, results from ensemble 3 were the poorest, while ensemble 4 yielded the best results.

Table 4.6: Ensemble feature selection classification accuracy results

Ensemble	1	2	3	4
BC 1	0.6209	0.6214	0.4591	0.6176
BC 2	0.6134	0.6323	0.4737	0.6531
BC 3	0.6112	0.6264	0.3383	0.6084
BC 4	0.6232	0.6418	0.3937	0.6605
BC 5	0.6128	0.6317	0.4141	0.6276
BC 6	0.6149	0.6323	0.4687	0.6314
BC 7	0.6125	0.6281	0.4885	0.5803
BC 8	0.619	0.6242	0.5288	0.6425
BC 9	0.6202	0.6168	0.4067	0.6151
BC 10	0.6338	80.6435	0.3593	0.5989
MV	0.6212	0.6314	0.4707	0.6482
SB	0.6338	0.6418	0.5288	0.6605

Where: Ens.–Ensemble, BC– Base Classifier, MV–Majority Vote, SB–Single Best

Figures 4.4 to 4.7 are the land cover accuracy maps obtained when compared with the reference map (the ground truth). It can be seen that the maps reflect a mixture of the pure signatures of the various materials found within the spatial extent of the ground instantaneous field view of the imaging instrument. It is obvious that, the mixture of materials affected the accuracy results obtained.

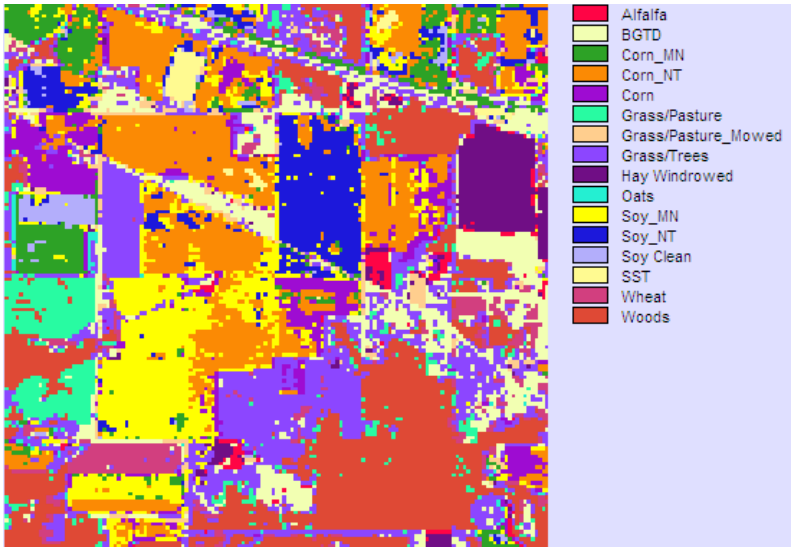


Figure 4.4: Land cover map obtained from an ensemble of 10 bands (Ensemble 1)

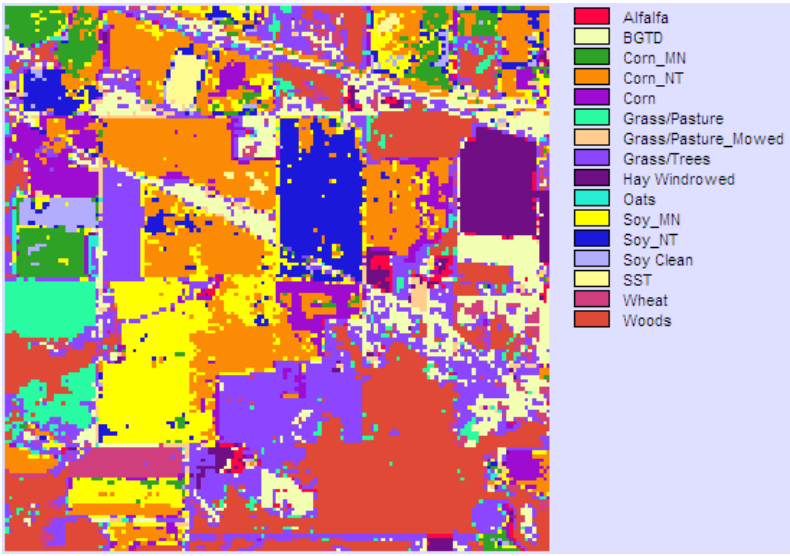


Figure 4.5: Land cover map obtained from an ensemble of 14 bands (Ensemble 2)

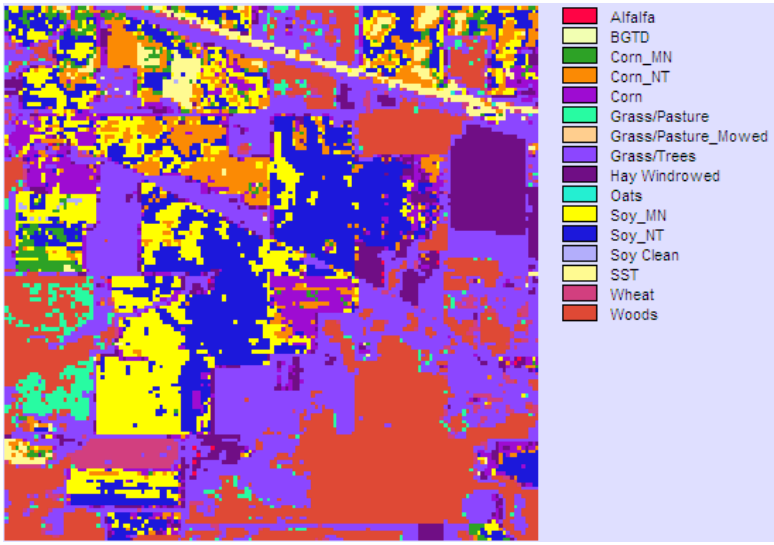


Figure 4.6: Land cover map obtained from an ensemble of 18 bands (Ensemble 3)

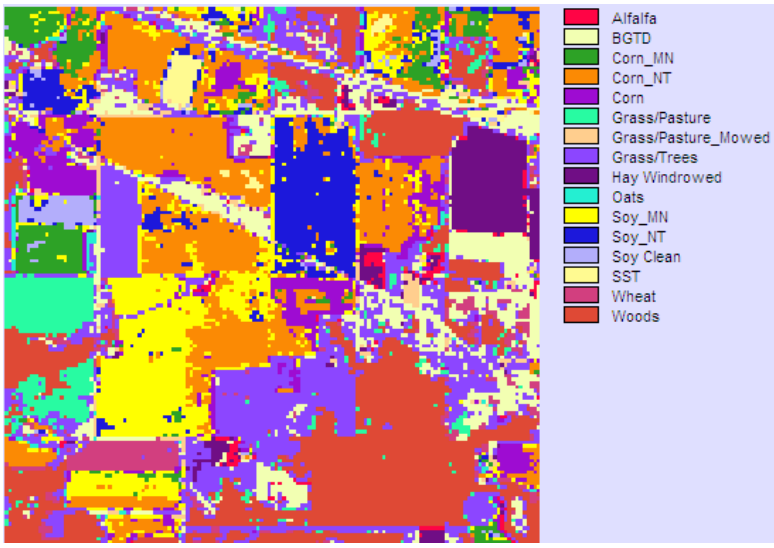


Figure 4.7: Land cover map obtained from an ensemble of 18 bands (Randomly selected, ensemble 4)

4.2.1 Binomial Test of significance

To get a better appreciation of the differences between these results, a binomial test of significance was carried out for each ensemble to ascertain the pairwise difference between majority voting and single best, results of which are illustrated in Table 4.7.

Table 4.7 Binomial Test of Significance between Majority Vote and Single Best

Ensemble	|z|
1	0.99
2	0.82
3	4.43
4	0.99

In the simple case of determining if there is a difference between two classifications (2 sided test), the null hypothesis (Ho) that there is no significant difference will be rejected if $|Z| > 1.96$ (Congalton *et al*., 1983; Rosenfield *et al*., 1986; Congalton *et al*., 1998). In this case, it is only in ensemble 3 that there is a significant difference between majority vote and single best approaches. The same test was carried out to establish if there was any significant difference between the different ensembles, the results of which are shown in Table 4.8 and 4.9. Table 4.8 depicts the pairwise difference between the ensembles based on the majority vote values, while Table 4.9 refers to single best values. In both tables E1, E2, E3 & E4 refer to Ensembles 1, 2, 3 and 4 respectively. Tables 4.8 & 4.9 show that the results of ensemble 4 are significantly better than the results from ensembles 1 & 3. Though the results of ensemble 4 are better than ensemble 2, the difference is insignificant. Results from ensemble 3 are significantly worse than all the results of the ensembles 1, 2 & 4.

Table 4.8: Binomial tests of significance between the different ensembles based on majority vote values

	E 1	E2	E3	E4
E1	0			
E2	0.80	0		
E3	11.65	2.47	0	
E4	2.14	1.33	184	0

Table 4.9: Binomial tests of significance between the different ensembles based on single best values

E1	E2	E3	E4	
E1	0			
E2	0.63	0		
E3	8.15	8.79	0	
E4	2.13	1.49	10.31	0

4.2.2 Conclusion

Of the ensembles considered, evidently the one based on random selection yielded the best classification results. Sequentially selecting bands into base classifiers yielded significantly poorer results. Feature selection resulted in better classification results compared to sequentially selecting the features. However ensemble 2 performed better than ensemble 1. This may have been as a result of using more features in each base classifier. The difference however was not significant.

4.3 Effect of ensemble size on classification accuracy

Table 4.10 shows a tabular representation of how classification accuracy varied as the number of base classifiers and bands per ensemble increased. These results were then used to generate Figure 4.8, in a bid to further illustrate the variation of classification accuracy as the number of base classifiers increased within each ensemble.

Table 4.10: Ensemble size classification Accuracy results

Number of base classifiers	Number of Bands Per Ensemble						
	2	4	6	8	10	12	14
3	0.470	0.566	0.606	0.600	0.617	0.632	0.628
4	0.467	0.561	0.596	0.613	0.620	0.628	0.630
5	0.470	0.564	0.609	0.612	0.617	0.635	0.630
6	0.473	0.563	0.606	0.615	0.617	0.629	0.633
7	0.475	0.562	0.610	0.618	0.617	0.630	0.630
8	0.480	0.561	0.607	0.618	0.619	0.632	0.631
9	0.487	0.562	0.604	0.615	0.620	0.632	0.633
10	0.483	0.563	0.606	0.620	0.621	0.628	0.631

The series in the legend represents the accuracy variation within each ensemble. As can be observed in all ensembles, the almost flat trend is evidence that the classification accuracy varied minimally as the number of base classifiers increases. Each series in Figure 4.8 represents a given ensemble, each with a given number of bands per ensemble. Two of the land cover accuracy maps generated are shown in Figures 4.9 and 4.10. Other generated maps can be seen in appendix B.

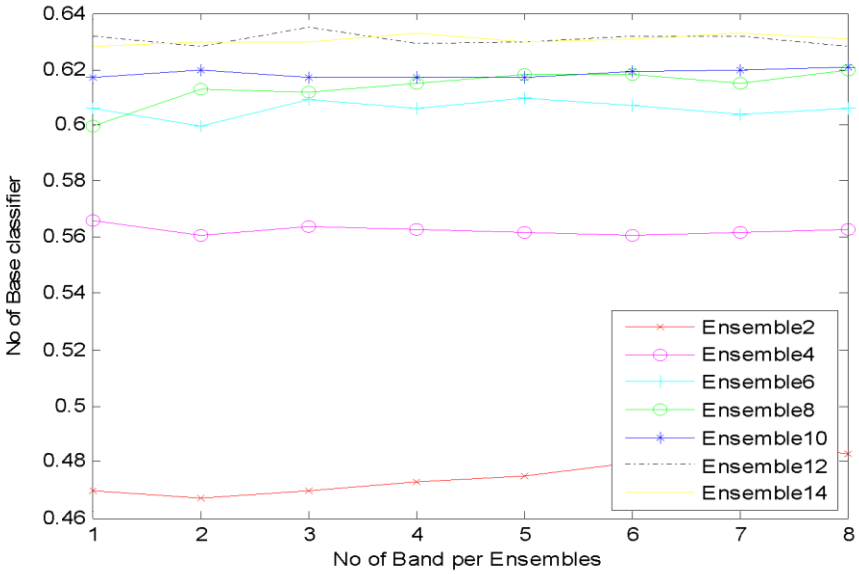


Figure 4.8: Graphical illustration of accuracy variation.

Series: 1, 2, 3, 4, 5, 6, 7 respectively represent bands 2, 4, 6, 8, 10, 12 and 14 per ensemble. General observation from the result reveals that classification accuracy increased as the number of bands per ensemble increased (Gidudu *et al*., 2009b; Abe *et al*., 2010). This increase was most obvious when the number of bands per ensemble increased from two to four. As the number of bands per ensemble increased from four to six, the improvement in accuracy became less pronounced, reducing even further as the number of bands per ensemble increased. A look at ensembles 6 and 7 representing 12 and 14 bands per ensemble are virtually indistinguishable implying that at that stage any additional increase in the number of bands per ensemble, regardless of the number of base classifiers, may not translate into a significant increase in classification accuracy.

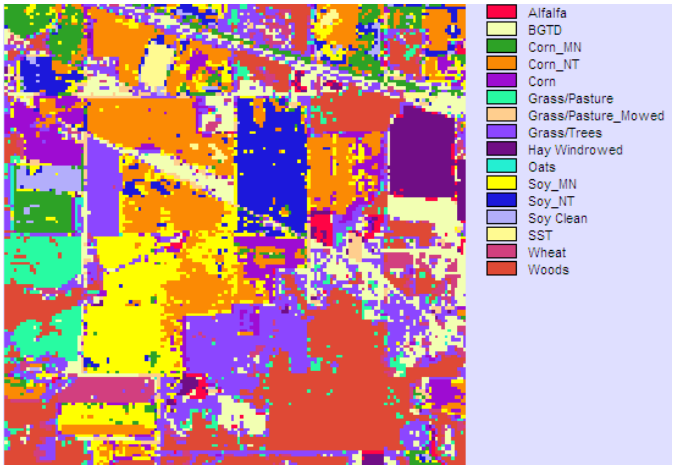


Figure 4.9: Ensemble made up of 4 bands classification Accuracy map

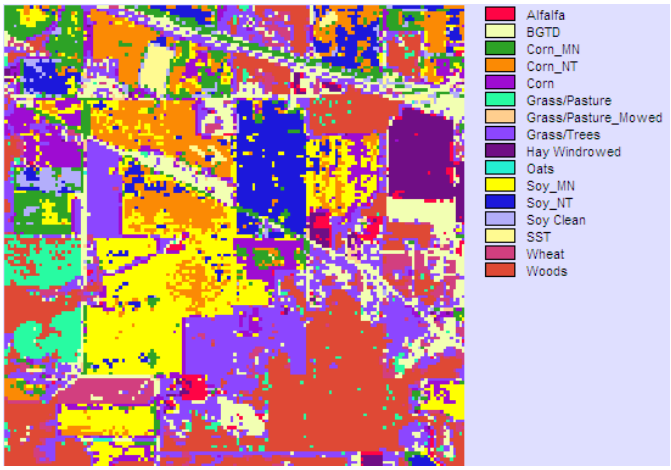


Figure 4.10: Ensemble made up of 6 bands classification Accuracy map

Whereas the results in Table 4.10 and Figure 4.8 depict differences in classification accuracy as the number of bands per ensemble increased, a binomial test of significance was applied to ascertain if these differences were significant.

4.3.1 Binomial Test of Significance

Tables 4.11, 4.12 and 4.13 were derived to illustrate the variation in accuracy between ensembles while maintaining a fixed number of base classifiers. Tables 4.11, 4.12 and 4.13 show the binomial test of significance at the 95% confidence interval, such that significant difference is deducted when Z > 1.96. In the Tables 4.11, 4.12 and 4.13 the number of base classifiers considered was 3, 6 and 10 respectively.

Table 4.11: Between-ensemble Binomial test of significance for 3 base classifiers

	E2	E4	E6	E8	E10	E12
E2	-					
E4	7.36	-				
E6	10.53	3.13	-			
E8	10.04	2.64	0.48	-		
E10	11.41	3.99	0.86	1.34	-	
E12	12.57	5.13	2	2.48	1.14	-
E14	12.28	4.85	1.72	2.2	0.86	0.28

In all the tables the general trend is that, ensembles E2 and E4 (i.e. ensembles with 2 and 4 bands per base classifier) were significantly inferior to the rest of the ensemble classifiers (i.e. E6, E8, E10, E12 and E14). In tandem with Figure 2, Tables 4.11, 4.12 and 4.13 depict that as the number of features per base classifier increase, so does the significance between the classification accuracies decrease. Even as only these three tables are presented here, they are representative of all the other tables.

Table 4.12: Between-ensemble Binomial test of significance for 6 base classifiers

	E2	E4	E6	E8	E10	E12
E2	-					
E4	6.89	-				
E6	10.26	3.33	-			
E8	10.98	4.04	0.7	-		
E10	11.09	4.15	0.82	0.11	-	
E12	12.11	5.14	1.81	1.1	0.99	-
E14	12.37	5.41	2.07	1.36	1.25	0.26

Table 4.13: Between-ensemble Binomial test of significance for 10 base classifiers

	E2	E4	E6	E8	E10	E12
E2	-					
E4	6.07	-				
E6	9.46	3.36	-			
E8	10.55	4.43	1.07	-		
E10	10.67	4.56	1.20	0.13	-	
E12	11.20	5.08	1.72	0.64	0.52	-
E14	11.49	5.36	2.00	0.93	0.80	0.28

Of the series in Figure 4.11, it is only series 1 (E2) which showcases the largest within ensemble variation. Hence, it was used to ascertain if cumulatively increasing the number of base classifiers can result into any statistically significant increase in classification accuracy. Table 4.14 presents within ensemble binomial test of significance for E2 (i.e. where number of features per ensemble are two). From the results it is evident that even with the biggest in-ensemble variation none of the base classifier accuracies were significantly different from the other.

Table 4.14: Within-ensemble Binomial test of significance for 2 features

	E3	E4	E5	E6	E7	E8	E9
E3	-						
E4	0.18	-					
E5	0.02	0.19	-				
E6	0.27	0.45	0.26	-			
E7	0.38	0.56	0.37	0.11	-		
E8	0.85	1.02	0.83	0.57	0.47	-	
E9	1.32	1.5	1.3	1.04	0.94	0.47	-
E10	1.04	1.21	1.02	0.76	0.66	0.19	0.28

4.3.2 Conclusion

In conclusion, the results demonstrate that in the design of ensemble classification systems for land cover mapping, classification accuracy increases as the number of bands per ensemble increase. The results however, show that there is a 'peaking' effect beyond which increasing the number of features per ensemble does not translate into an increase in classification accuracy. It can also be concluded that for ensemble feature classification, in its application for land cover mapping, there isn't any significant benefit in having many base classifiers. From these results, the minimum number of base classifiers (in this case 3) suffices.

4.4 Random Ensemble Feature Selection Classification

Table 4.15 shows the result obtained from the investigation using a random selection approach to classification accuracy for land cover mapping. Results show that in general, there is a significant improvement in classification accuracy as the number of bands per base classifier increased. This could be due to the fact that given the high number of classes, 16 classes in this case more features were needed to

appropriately separate the classes. On the other hand, whereas there were few instances where incrementally adding base classifiers to the ensemble significantly improved the classification accuracy, the general trend is that there is no improvement in accuracy as the number of base classifiers increased (Gidudu et al., 2009b).

Table 4.15: Random selection ensemble accuracy results

Number of Base Classifiers	Number of Bands per Ensemble						
	2	4	6	8	10	12	14
3	0.412	0.438	0.570	0.627	0.623	0.636	0.651
4	0.396	0.467	0.569	0.626	0.633	0.627	0.642
5	0.400	0.442	0.582	0.629	0.625	0.637	0.649
6	0.382	0.483	0.595	0.627	0.639	0.632	0.642
7	0.407	0.497	0.559	0.634	0.632	0.634	0.648
8	0.426	0.523	0.543	0.639	0.641	0.642	0.643
9	0.435	0.521	0.542	0.646	0.645	0.642	0.646
10	0.433	0.540	0.560	0.645	0.641	0.639	0.643

Appendix B presents the land cover accuracy maps generated from the investigation.

4.4.1 Binomial test for significance between ensembles for 3 classifiers

Table 4.16 revealed the results obtained for a binomial test of significance. Looking at E2, E4 and E6 from the table shows a significant increase in classification accuracy. As the ensemble size increases, (E8, E10, E12) there is no significant improvement in classification accuracy. Binomial test of significance was also conducted for 9 base classifiers to ascertain the effect of ensemble size on land cover classification. From Table 4.17, E2, E4 and E6 show significant classification

accuracy, while E8, E10 and E12 show no significant increase in classification accuracy.

Table 4.16: Between-ensemble Binomial test of significance for 3 base classifiers

	E2	E4	E6	E8	E10	E12
E2	-	-	-	-	-	-
E4	2	-	-	-	-	-
E6	12.2	10.1	-	-	-	-
E8	16.8	14.7	4.5	-	-	-
E10	16.5	14.3	4.2	0.3	-	-
E12	17.5	15.4	5.2	0.7	1	-
E14	18.8	16.7	6.4	1.9	2.3	1.3

Table 4.17: Between-ensemble Binomial test of significance for 9 base classifiers

	E2	E4	E6	E8	E10	E12
E2	-	-	-	-	-	-
E4	6.58	-	-	-	-	-
E6	8.19	15.96	-	-	-	-
E8	16.54	9.78	8.16	-	-	-
E10	16.48	9.72	8.1	0.06	-	-
E12	16.22	9.46	7.85	0.31	0.25	-
E14	16.53	9.77	8.15	0.01	0.05	0.3

In conclusion, size in ensemble systems does matter, but only if the number of bands per base classifier is increased. These results principally inform us that in the design of ensemble feature selection classification systems, increasing the number of base classifiers may not necessarily translate into improved ensemble classification accuracies. In which case, the minimum possible number of base classifiers will suffice. Increasing the number of base classifiers evidently doesn't improve classification accuracy (Gidudu et al., 2009b).

4.5 Spectral Unmixing Analysis for land cover classification accuracy

The discussion on the results of our research on the endmember spectral response determination is presented. The session also discusses the performance of the SVM and RF classifiers investigated.

4.5.1 Results of Endmember Spectral Response Determination

Figure 4.11 shows the result obtained, wherein the extreme pixels correspond to endmember spectral responses in each projection that are recorded. The total number of times that each pixel is marked as extreme is noted. A threshold value of (65%) is used to define how many pixels are marked as extreme at the ends of the projected vector.

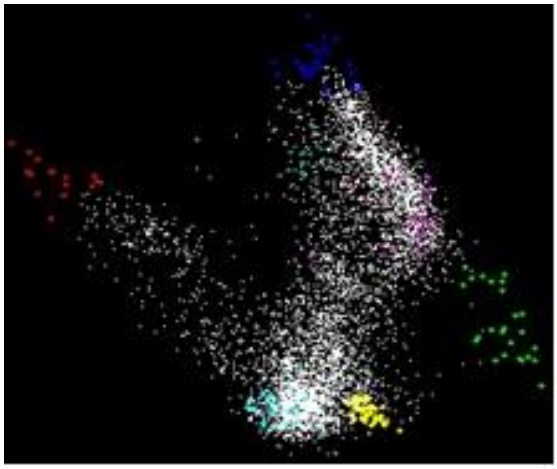


Figure 4.11: Purest pixels occur at the edges of the projected vector

The estimated number of endmember spectral responses and the corresponding spectral signatures were obtained by using ENVI visualizer. At the completion of specified iterations, ten images (*P*) were created in which the value of each pixel

corresponds to the number of times that a pixel was recorded as extreme. The bright pixels in the PPI image are generally the image endmember spectral responses to characterize the vegetation structure. Figure 4.12 shows the generated images and the Root Mean Square (RMS) error of the image.

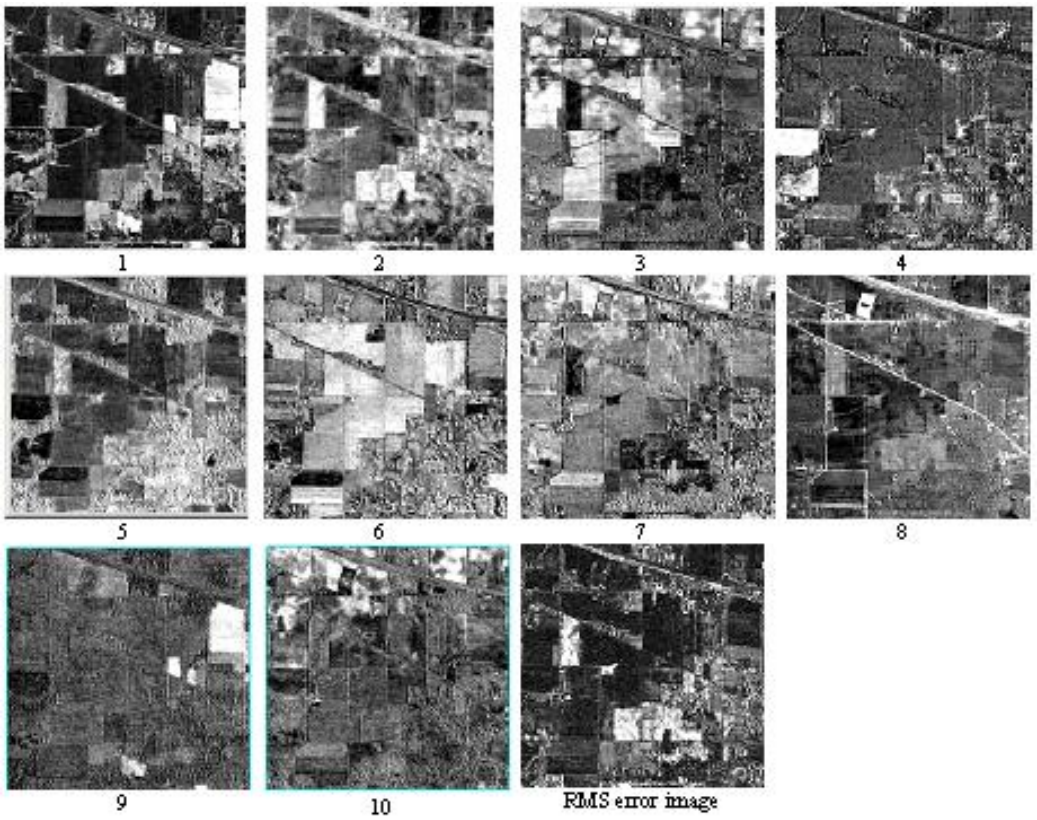


Figure 4.12: Generated images and RMS error from PPI method

4.5.2 Results of Land Cover Classification

Tables 4.18 – 4.22 present the confusion matrixes obtained by each classifier showing the correlation between the class labels. The confusion matrix (also known as error matrix) is a widely accepted method to report the error of raster data and to

assess the classification accuracy of a classifier (Congalton, 1991). The matrix expresses the number of sample units allocated to each land cover type as compared to the reference data. The diagonal of the matrix designates agreement between the reference data and the interpreted land cover types (Congalton, 1991). From Table 4.20 (SVMs), scl, snt, cnt have 100% classification accuracy while others have some of the pixel members misclassified. The remaining classifiers have some of their class labels misclassified.

Table 4.18: Neural Network Confusion Matrix

a	b	c	d	e	f	g	h	i	j	classified as
205	0	0	3	27	0	9	0	20	0	a = gp
0	126	15	19	0	25	0	18	0	0	b = scl
0	32	220	17	0	22	1	58	0	0	c = snt
0	41	77	157	12	71	1	66	0	0	d = smn
2	0	4	2	284	2	7	1	31	6	e = gt
1	70	53	31	16	122	0	66	0	0	f = cnt
1	0	0	1	2	0	275	0	0	0	g = hw
0	30	60	28	4	52	0	131	0	0	h = cmn
107	0	0	0	0	0	0	0	291	2	i = woods
48	0	0	0	84	1	8	0	124	35	j = bg

Table 4.19: Support Vector Machines Confusion Matrix

a	b	c	d	e	f	g	h	i	j	classified as
246	0	0	0	15	0	0	0	1	2	a = gp
0	203	0	0	0	0	0	0	0	0	b = scl
0	0	350	0	0	0	0	0	0	0	c = snt
0	1	0	424	0	0	0	0	0	0	d = smn
0	0	0	0	333	6	0	0	0	0	e = gt
0	0	0	0	0	359	0	0	0	0	f = cnt
0	0	0	0	1	0	278	0	0	0	g = hw
0	0	0	0	0	4	0	301	0	0	h = cmn
8	0	0	0	0	0	0	0	391	1	i = woods
5	0	0	0	7	2	0	0	0	286	j = bg

Table 4.20: Random Forest Confusion Matrix

a	b	c	d	e	f	g	h	i	j	classified as
247	0	0	0	15	0	1	0	1	0	a = gp
0	203	0	0	0	0	0	0	0	0	b = scl
0	0	347	0	0	2	0	1	0	0	c = snt
0	0	3	420	0	1	0	1	0	0	d = smn
1	0	0	0	332	6	0	0	0	0	e = gt
0	1	0	2	0	356	0	0	0	0	f = cnt
0	0	0	0	1	0	278	0	0	0	g = hw
0	0	1	0	0	4	0	300	0	0	h = cmn
2	0	0	0	0	0	0	0	398	0	i = woods
0	0	0	0	9	2	1	0	3	285	j = bg

Table 4.21: C4.5 Confusion Matrix

a	b	c	d	e	f	g	h	i	j	classified as
215	0	0	0	26	0	8	0	10	5	a = gp
0	144	3	20	3	26	0	7	0	0	b = scl
0	12	290	10	0	27	0	11	0	0	c = snt
1	20	39	303	5	24	1	32	0	0	d = smn
1	0	0	1	312	7	0	0	10	8	e = gt
0	29	21	10	5	256	1	37	0	0	f = cnt
2	0	0	0	2	0	275	0	0	0	g = hw
0	10	13	14	2	32	0	234	0	0	h = cmn
59	0	0	0	1	1	0	0	327	12	i = woods
35	1	0	1	67	2	3	0	83	108	j = bg

Table 4.22: Bagging Confusion Matrix

a	b	c	d	e	f	g	h	i	j	classified as
239	0	0	1	7	0	7	0	7	3	a = gp
0	171	9	10	2	9	0	2	0	0	b = scl
0	7	326	2	0	10	0	5	0	0	c = snt
0	7	9	399	3	3	0	4	0	0	d = smn
19	1	0	0	299	6	2	0	6	6	e = gt
0	9	5	7	1	329	1	7	0	0	f = cnt
1	0	0	0	0	0	277	0	0	1	g = hw
0	3	10	6	1	9	0	276	0	0	h = cmn
18	0	0	0	0	0	0	0	374	8	i = woods
26	0	0	0	14	2	4	0	43	211	j = bg

Table 4.23 presents the summary and comparison of the overall accuracy results. It is obvious that not all the instances in each case were correctly classified. All the classifiers performed well with classification accuracy above 70% except the neural network. The low performance of neural network could be that the volume of data used is not enough to train the classifier (Zhang, 2000; Marwala, 2009; Taskin, 2009). However, SVM and RF predictions are tremendous (Abe et al., 2012). Comparing the Kappa statistics values obtained, the SVM has the highest percentage followed by RF. The tables also revealed that there are no unclassified instances during the classification procedure.

Table 4.23: Spectral unmixing classification accuracy results

C	CCI	ICI	UI	KS	MAE	RMSE	RAE (%)	RRSE (%)	A (%)
NN	1846	1378	0	0.5243	0.1049	0.234	58.5418	78.1656	57.26
SVMs	3171	53	0	0.9817	0.0033	0.0573	1.8343	19.1535	98.34
RF	3166	58	0	0.9799	0.0286	0.0875	15.954	29.2364	98.20
C 4.5	2464	760	0	0.7371	0.0627	0.1771	34.9938	59.1559	76.43
B	2901	323	0	0.8882	0.0553	0.1394	30.8356	46.5581	89.98

Where: C – Classifier, CCI – Correctly Classified Instances, ICI–Incorrectly Classified Instances, UI – Unclassified Instances, KS – Kappa Statistic, MAE – Mean Absolute Error, RSE – Root Mean Squared error, RAE– Relative Absolute Error, RRSE – Root Relative Squared Error, A –Accuracy, NN – Neural network, SVM – Support vector machines, RF – Random forest, B – Bagging.

We also looked at the performances of the classifiers against each class label and the outcome is as shown in Table 4.24. The table shows that each classifier performance on each class label is remarkable. Comparing the performance of the each classifier against the individual class label, SVM produces a higher level of classification accuracy per class label as compared to others.

Table 4.24: Overall percentage accuracy results' analysis of different classification schemes performance on the land cover classes

Class label	Random forest	Neural Network	C4.5	SVMs	Bagging
gp	96.00	91.84	93.70	99.04	92.84
bg	92.99	90.73	92.43	99.47	91.07
woods	95.32	87.56	89.24	99.69	87.75
scl	95.07	93.74	94.91	100	93.77
snt	92.15	89.18	92.34	100	89.27
smn	90.23	86.79	91.63	99.97	86.85
gt	93.42	89.52	91.22	99.10	90.11
cnt	91.35	88.90	91.50	100	88.90
hw	97.89	91.38	93.58	99.97	91.97
cmn	92.25	90.57	93.74	99.88	90.60

The entire accuracy assessment procedure is that the error matrix must be a representative of the entire area mapped from the remotely sensed data **(**Congalton, 1988) and is calculated using the formula **(**Story and Congalton, 1986; Congalton, 2005)**:**

$$C = \frac{\sum_{i,j=1}^{N} Z_{(i,j)}}{M} \tag{4.1}$$

Where: C = classification accuracy

Z = all entries in the major diagonal of the matrix

M = total number of sample units in the matrix

If all the non-major diagonal elements of the error matrix are zero, then it means no area in the map has been misclassified and the map accuracy is 100 percent. Otherwise, there are certain percentages of misclassified instances (Congalton *et al.*, 1983; Congalton, 1991; Rodriguez-Galiano *et al.*, 2011). In our experiment from Table 4.23 shows the misclassified instances as follows: NN, 1378 instances = 44.74%, SVMs, 53 instances = 1.66%, RF, 58 instances = 1.8%, C4.5, 760 instances = 23.57%, Bagging, 323 instances = 10.02% respectively. The overall accuracy for correctly classified instances, incorrectly classified instances, unclassified instances and the Kappa statistic are identified from the error matrices (Landgrebe, 1998; Congalton and Green, 2009; Abe *et al*., 2012).

In order to further evaluate the results of the classification accuracy so as to establish which of the classifiers performed better, the results obtained by each classifier per land cover in Figure 4.24 is subjected to Friedman test. Friedman test is a non-parametric statistical test alternative to ANOVA with repeating measures. According to Japkowicz and Shah (2011), the null hypothesis is that all the classifiers perform equally, and rejection of that null hypothesis means that: there exists at least one pair of classifiers with significantly different performances.

Should the null hypothesis be rejected, the Post-hoc test will be conducted. The job is to identify the significantly different pairs of classifiers. Its objective is to identify the specific pair or pairs of factors with a difference in rank sums that is statistically significant, and which may have caused the rejection of the global null hypothesis. The Friedman test is used to investigate if there is a difference between the predictions obtained per land cover class by Random Forest, Support Vector Machines, C4.5 and Bagging using alpha = 0.05. The null hypotheses are;

H_0: there is no difference between the five conditions.

H_1: there is a difference between the five conditions.

To calculate the degree of freedom (df), we used

df = k -1 (where k is the number of classifiers to be compared)

df = 5- 1 = 4 (we used 4 degrees of freedom). Thereafter we find the decision rule. The critical value will be 9.48773, if we calculate Chi square value (χ^2) and it is greater than 9.48773, the hypothesis will be rejected. To calculate Friedman Test the following equation is used:

$$\chi^2 = \frac{12}{nk(k-1)} \sum R^2 - 3n(k+1)$$

Where n represents the number of class labels, k is the number of classifiers and $R$is the classifiers considered. From our calculation, the null hypothesis was rejected because the value obtained is greater than the critical value. The Friedman's test shows that there is a significant difference among the classifiers being tested. Hence, we used Nemenyi test to pinpoint where the difference lies using the formula;

$$q_{yz} = \frac{\bar{R}_y - \bar{R}_z}{\sqrt{\frac{k(k+1)}{6n}}}$$

(n is the number of domains and k, the number of classifiers). From the Friedman test,

$$\bar{R}_A = 23, \bar{R}_B = 49, \bar{R}_C = 27, \bar{R}_D = 10, \bar{R}_E = 41$$

Where,

$\bar{R}_A =$ Random forest, $\bar{R}_B =$Neural Network, $\bar{R}_C =$Support Vector Machines, $\bar{R}_D =$ C4.5 and $\bar{R}_E =$ Bagging

We obtained, $q_{AB} = -36.77$, $q_{AC} = -5.66$, $q_{AD} = 18.38$, $q_{AE} = -25.46$, $q_{BC} = 31.11$, $q_{BD} = 55.15$, $q_{BE} = 11.31$, $q_{CD} = 24.04$, $q_{CE} = -19.80$, $q_{DE} = -42.84$

$q_{\propto} = 9.49$ for $\propto = 0.05$($q_{\propto}$ must be larger than q_{yz} for the hypothesis that y and z perform equally to be rejected).

Therefore, we reject the null hypothesis in the case of RF and NN, RF and C4.5, RF and Bagging, NN and SVMs, NN and C4.5, NN and Bagging, SVMs and Bagging, and C4.5 and Bagging, but not in the case of RF and SVMs.

Kappa statistic (Bishop *et al.*, 1975; Congalton, 1991) a discrete multivariate technique is used to determine (i) if the remotely sensed classification is better than a random classification, (ii) if two or more error matrices are significantly different from each other. The outcome of performing Kappa analysis is a KHAT statistic, which can be viewed as a measure of accuracy or agreement. The KHAT statistic is

calculated using the following formula (Cohen, 1960; Congalton and Mead, 1983; Congalton, 1991, Landgrebe, 1998):

$$\check{K} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}\left(x_{i+} * x_{+i}\right)}{N^2 - \sum_{i=1}^{r}\left(x_{i+} * x_{+i}\right)} \tag{4.2}$$

Where, r = number of rows in the matrix

x_{ii} = no. of observation in row *i* and column *i*

x_{i+} and x_{+i} = the marginal totals of row *i* and column *i* respectively

N = total number of observations

According to (Demir and Ertürk, 2007), the kappa value depicts Kappa coefficient of agreement which is a measure of how well the classifier's prediction compares with the reference (ground truth) data. ,It ranges from – 1 to 1, with negative values meaning agreement worse than expected. Low negative values (0 to – 0.10) generally imply no agreement between the classified land cover and ground truth while 1 indicates complete agreement. Kappa statistic calculated value obtained for both SVM, RF models were very significant taking into account the complexity of the study area and the large number of categories, which indicate that the method used for the image processing and the application of GRG optimization technique is very effective. Evaluating the performances of the classifiers in a land cover classification context are in the sequence, the SVM has the highest classification accuracy, followed by RF, Bagging (ensemble classifiers), followed by C4.5, then

the neural network (Abe et al., 2012). Other results obtained from the experiment are presented in Appendix C.

4.5.3 Conclusion

The classification accuracy results obtained from the unmixing spectral analysis using generalized reduced gradient approach was presented. The results obtained proved that, the generalized reduced gradient approach to image processing is very effective as applied to hyperspectral image. Evaluating the classifiers on classification accuracy bases, generalized Support Vector Machine's performance is the best, followed by Random Forest, Bagging, C4.5 and NN has the least performance. However, using Random Forest instead of Support Vector machine is far less computationally expensive (Abe *et al*., 2012).

4.6 Hyperspectral Image Classification using Random Forests and Neural Networks

This section presents the results and discussion of the experiment conducted on the Washington DC Mall dataset

4.6.1 Endmembers determination

The first experiment performed aimed to obtain endmembers from image dataset using the ENVI software application. The MNF transformation of the input hyperspectral image was performed for dimension reduction. The next stage of the endmember determination is to select a set of endmembers by applying the PPI algorithm on the extracted Region of Interest (ROI) pixels. Figure 4.13 shows this result, wherein the extreme pixels corresponding to the endmembers in each projection are recorded and the total number of times each pixel is marked as

extreme is noted. A threshold value of 1 is used to define how many pixels are marked as extreme at the ends of the projected vector (Abe *et al.*, 2012).

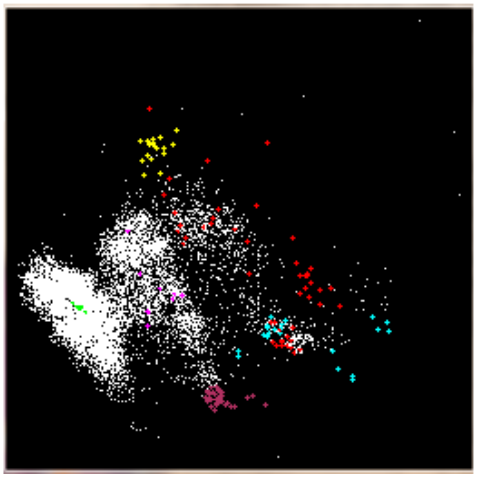


Fig 4.13: Purest pixels occur at the edges of the projected vector

The estimated number of spectral endmembers and their corresponding spectral signatures are obtained using ENVI visualizer. Figure 4.14 shows the generated six fractional endmembers of the image from the PPI method (Abe *et al.*, 2012).

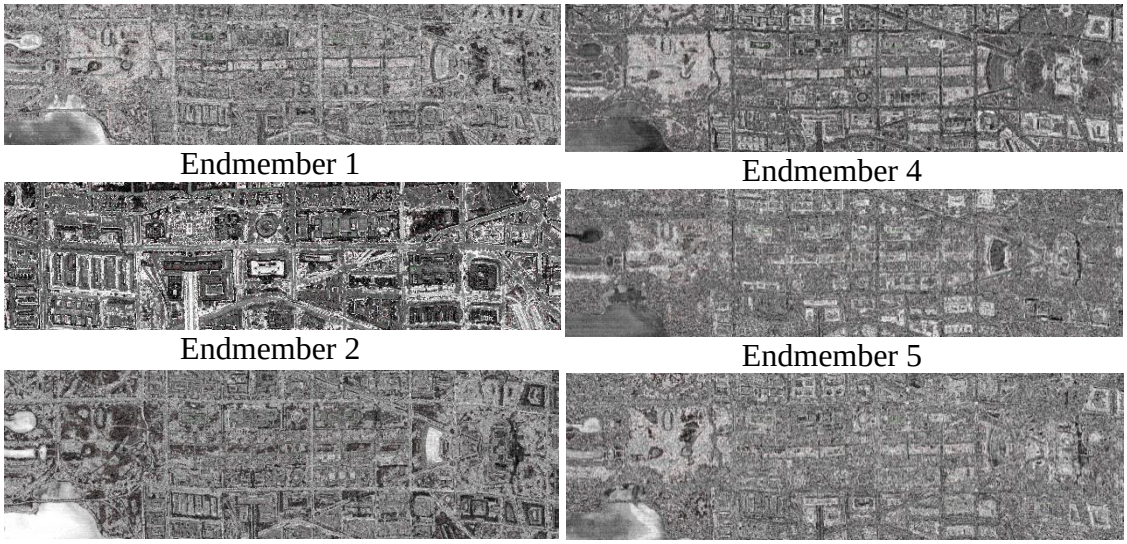


Fig 4.14: Fraction images for each endmember

At the completion of specified iterations, a PPI image is created in which the value of each pixel corresponds to the number of times that a pixel was recorded as extreme. The bright pixels in the PPI image are generally the image endmembers to characterize the land cover structure. This study presents six endmember models to characterize the land cover structure which are: Roofs, Street, Path, Grass, Trees, Water and Shadow. Normalized numerical values of the fractional abundant generated were calculated from the spectral signatures of the land cover label.

4.6.2 Results of Land Cover Classification

RF and NN classifiers are evaluated using the error confusion matrix method, which is a representation of the entire classification result. According to (Benediktsson *et al*., 1990), the error confusion matrix can be used to compute the overall accuracy and the individual class label accuracy. The error confusion matrix is a widely accepted method to report error of raster data and to assess the classification accuracy of a classifier. The matrix expresses the number of sample units allocated

to each land cover type as compared to the reference data. The diagonal of the matrix designates agreement between the reference data and the interpreted land cover types (Congalton, 1991; Abe *et al*., 2012).

Table 4.25 shows the result of the error confusion matrix for the performance of RF classifier. This result shows that roofs, paths, water, streets, trees and grass have 100% classification accuracy because none of their pixel's member is misclassified while the shadow has one of the pixels' members misclassified.

Table 4.25: Random forests error confusion matrix

a	b	c	d	e	f	g	classified as
724	0	0	0	0	0	0	a = Roofs
0	211	0	0	0	0	0	b = Paths
0	0	703	0	0	0	0	c = Water
0	0	0	404	0	0	0	d = Streets
0	0	0	0	398	0	0	e = Trees
0	0	1	0	0	96	0	f = Shadow
0	0	0	0	0	0	818	g = Grass

Table 4.26 records the result of the error confusion matrix for the performance of NN. From the table, it can be observed that roofs and grass are 100% classified while other land cover classes have some of their pixels misclassified.

Table 4.26: Neural networks error confusion matrix

a	b	c	d	e	f	g	classified as
724	0	0	0	0	0	0	a = Roofs
1	210	0	0	0	0	0	b = Paths
1	0	699	0	0	3	0	c = Water
1	0	2	401	0	0	0	d = Streets
0	0	0	0	398	0	0	e = Trees
0	0	17	0	0	80	0	f = Shadow
0	0	0	0	0	0	818	g = Grass

Generally, the two classifiers performed excellently well. Considering the individual class label, RF produces a higher level of classification accuracy per class label as compared to the NN. The entire accuracy assessment procedure is that the error confusion matrix must be a representative of the entire area mapped from the remotely sensed data (Story and Congalton, 1986). The overall accuracy for correctly classified instances, incorrectly classified instances, unclassified instances and the Kappa statistic are identified from the error confusion matrices (Story and Congalton, 1986; Congalton, 1988; Abe *et al.*, 2012).

If all the non-major diagonal elements of the error confusion matrix are zero, then it means no area in the map has been misclassified and the map accuracy is 100 percent. Otherwise, there are certain percentages of misclassified instances (Congalton, 1988). In our experiment, RF as compared to NN has only 1 instance misclassified, while NN has 25 instances misclassified.

The Kappa coefficient of agreement is a measure of how well the accuracy of the classifier compares with the reference or ground truth data (Congalton, 1988). It

ranges from 0 to 1, with 0 implying no agreement between the classified land cover and ground truth and 1 indicates complete agreement. Table 4.27 shows the result of error, Kappa statistics and overall accuracy classification.

Table 4.27: Classification accuracy

C	CCI	ICI	UI	KS	MAE	RMSE	RAE (%)	RRSE (%)	Accuracy (%)
RF	3354	1	0	0.9996	0.0015	0.0176	0.6568	5.1615	99.9702
NN	3330	25	0	0.9909	0.003	0.0379	1.2835	11.1095	99.2548

Where: C– Classifier, CCI – Correctly Classified Instances, ICI–Incorrectly Classified Instances, UI – Unclassified Instances, KS – Kappa Statistic, MAE – Mean Absolute Error, RSE – Root Mean Squared error, RAE – Relative Absolute Error, RRSE – Root Relative Squared Error

The process used for results analysis in section 4.5 was also used in this section. According to this result, there are no unclassified instances during the RF and NN classification procedures and the overall classification accuracies of the classifiers are seen to be comparable. It can be deduced from the predictions that RF outperformed NN. In addition, RF is more computational effective as compared to NN (Abe *et al*., 2012).

Conclusion

This study aimed to establish a performance comparison between RF and NN classifiers for land cover classification. The performance assessment was done, giving overall accuracy and error confusion matrix. Experimental results demonstrate that the generation of RF and NN based land cover classification systems significantly improves overall accuracy. As a result, the classifiers can

significantly contribute to land cover classification system as a source of analysis and increase its accuracy. The comparability and high accuracy performance of RF and NN indicates that the GRG method introduced in this study is effective for solving a linear spectral unmixing problem of land cover classification (Abe *et al*., 2012).

CHAPTER 5

Remark and Conclusions

This chapter highlights the objectives of the study, how they were achieved and the major findings. It also presents the limitation and the future work to be done.

5.1 Remark

This section gives a recap on the study objectives, how they were achieved and the major findings.

- Diversity through training a given classifier on different features and land cover accuracy: This was achieved by using ensemble feature selection to impose diversity in ensembles. The features of the constituent base classifiers for each ensemble were created through an exhaustive search algorithm using different separability indices. For each ensemble, the classification accuracy was derived as well as a diversity measure purported to give a measure of the in-ensemble diversity. The correlation between ensemble classification accuracy and diversity measure was determined to establish the interplay between the two variables. The study was carried out using Indiana pines dataset. The investigation reveals that diversity measures as formulated do not provide an adequate means upon which to constitute ensembles for land cover mapping.
- Interplay between the structure of ensemble and land cover classification accuracy: The study was done by cumulatively increasing the ensemble size (of both the number of base classifiers and bands) and consequently evaluates its effect on the ensemble classification accuracy. The investigations were

carried out using the online Indiana Pines hyperspectral dataset. The main finding of this investigation was that as the number of bands per ensemble increased, so did the classification accuracy significantly improve. On the other hand, increasing the number of base classifiers in the ensemble system portrayed no significant influence on land cover mapping classification accuracy.

- Fully constrained spectral unmixing analysis for land cover classification: This investigation was carried out using Indiana pines and Washington DC Mall hyperspectral image datasets. The pixel purity index algorithm was used to obtain endmember spectral responses from the datasets. The generalized reduced gradient optimization algorithm is thereafter executed on the research datasets to estimate fractional abundances in the hyperspectral images, thereby obtain the numeric values for land cover classification. The Waikato environment for knowledge analysis (WEKA) data mining framework is selected as a tool to carry out the classification process. The classifiers used were; support vector machines, random forests, neural network, Bagging and C4.5. Results show that the performances of all the classifiers are remarkable, but support vector machines ensemble is comparable with that of random forests.
- Interplay between combination rules of the ensemble and land cover classification accuracy: The study explores the ensemble feature selection as a means of ensuring diversity for land cover mapping of Indiana pines hyperpectral data. Two ensemble feature selection techniques were used, namely exhaustive search and random selection of feature subsets. The evaluation function for the exhaustive search was the Bhattacharyya Distance separability index. The results of the base classifiers in each ensemble were

combined using two methods; majority voting and an adaptation of Cross Validation majority (CVM) also called single best. In CVM, cross validation data is used as a basis for selecting the best out of the whole ensemble.

Results show that random selection of features (bands) yielded the best results as compared to building base classifiers depending on search algorithms or as was used in the study, sequentially arranging the features into base classifiers. Of the combination techniques, the single best technique yielded better results than majority vote. However, in most cases the difference between the results was not significant.

5.2 Conclusions

In this research, we investigated the effect of ensemble classification on land cover mapping. This was carried out by looking at the interplay between ensemble system size and classification accuracy for land cover mapping. The work also explores the ensemble feature selection as a means of ensuring diversity for land cover mapping of hyperspectral data. It was observed from the results that;

- There is a ‘peaking’ effect beyond which increasing the number of features per ensemble does not translate into an increase in classification accuracy
- In application of ensemble feature classification for land cover mapping, there is no significant benefit in having many base classifiers. The minimum number of base classifier (in this was 3) should be enough.
- Diversity measures as currently formulated does not provide an adequate means upon which to constitute ensembles for land cover mapping.
- The random selection of bands yielded a good result compared to building base classifiers depending on the bands search algorithm.

Investigation of classification accuracy was also conducted using spatial and spectral signature of AVIRIS hyperspectral imagery data. The numeric values of the estimated fractional abundance obtained from Generalized Reduced Gradient based approach was adequate for land cover classification. The classifiers predictions ascertained that the approach was cost and computational effective with SVM having the best predicted using Indiana pine dataset.

In the same vain, the experimental results demonstrate that the generation of RF and NN based land cover classification systems significantly improves overall accuracy. As a result, the classifiers can significantly contribute to land cover classification system as a source of analysis and increase its accuracy. The comparability and high accuracy performance of RF, SVM and NN indicates that the GRG method introduced in this study is effective for solving a linear spectral unmixing problem of land cover classification.

5.3 Future works

The limitation of this work is the non-availability of different remote sensing data to allow for more experiments. This is due to the high cost involved in procuring or acquiring remote sensing data. The investigator used data available online, which has been researched by various researchers all over the world.

As future work we want to:

- Look into the quantification of diversity measure and its consequent influence in determining the type of ensemble classification for land cover mapping.

- Further investigations will be conducted on various hyperspectral datasets using the Generalized Reduced Gradient base approach to authenticate its efficiency on other datasets.
- Generalized Reduced Gradient algorithm shall be implemented using Matlab.

Bibliography

Adams, J. B., Sabol, D. E., Kapos, V., Filho, R. A., Roberts, D. A., Smith, M. O. and Gillespie, A. R. (1995). “Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon,” *Remote Sensing of Environment*, 52(2) pp. 137-154.

Abadie, J. and Carpentier, J. (1969). “Generalization of the wolfe reduced gradient method in the case of non-linear constraints”, *In: R. Fletcher, Ed., Optimization, Academic Press, London*, pp. 37-47.

Abe, B. T., Olugbara, O. O. and Marwala, T. (2012) “Hyperspectral Image Classification using Random Forest and Neural Network”, *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering and Computer Science 2012, WCECS 2012*, 24-26 October, San Francisco, USA, pp 522-527.

Abe, B., Gidudu, A. and Marwala, T., "Investigating the effects of ensemble classification on remotely sensed data for land cover mapping," *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International* , vol., no., pp.2832-2835, 25-30 July 2010

Bazi, Y., and Melgani, F. (2006). “Toward an optimal SVM classification system for hyperspectral remote sensing images”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3374 – 3385.

Benediktsson, J. A., Swain, P. H., and Ersoy, O. K., (1990). “Neural network approaches versus statistical methods in classification of multisource remote

sensing data", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, pp. 540 – 552.

Benediktsson, J.A. and Swain, P.H. (1992). "Consensus theoretic classification methods", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 22, pp. 688 – 704.

Bhattacharyya, A. (1943). "On a measure of divergence between two statistical populations defined by probability distributions, *Buletin. Calcutta Mathematical Society*", vol. 35, pp 99 – 109.

Boardman, J. W., Kruse, F.A., and Green, R.O. (1995). "Mapping target signatures via partial unmixing of AVIRIS data", *In: Syummaries of the VI JPL Airborn Earth Science Workshop*, Pasadena, CA.

Boardman, J.W.; Biehl, L.L.; Clark, R.N.; Kruse, F.A.; Mazer, A.S.; Torson, J., (2006) "Development and Implementation of Software Systems for Imaging Spectroscopy", *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on* , vol., no., pp.1969 – 1973.

Bortolot, Z. (1999). "Image classification", Online publication http://www.geog.ubc.ca/courses/geog570/notes/classification.html. Cited 9 July 2011.

Breiman, L. (1996). "Bagging predictors", *Machine Learning*, vol. 24, no. 2, pp. 123 – 140.

Breiman, L. (2001). "Random forests, *Machine Learning*", vol. 45, no.1, pp.5–32.

Bruzzone, L. and Cossu, R. (2002). “A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 1984–1996.

Chan, J., C., and Paelinckx, D. (2008). “Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery”, *Remote Sensing of Environment*, vol. 112, pp. 2999 – 3011.

Chang, C.-C. and Lin, C.-J. (2001). LIBSVM: “A Library for Support VectorMachines”, Online publication *http://www.csie.ntu.edu.tw/~cjlin/libsvm*. Cited 8 June 2011.

Chang, C-I, Wu, C-C., Liu, W-m. and Ouyang, Y-C. (2006). “A new growing method for simplex-based endmember extraction algorithm”, *IEEE Trans. Geosci. & Remote Sensing*, vol. 44, no. 10, pp. 2804-2819.

Chaudhry, F., Wu, C., Liu, W., Chang, C-I and Plaza, A. (2006). “Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery”, In *Recent Advances in Hyperspectral Signal and Image Processing*”, C.-I Chang, Ed. Trivandrum, India: Research Signpost, vol. 3, pp. 31-61.

Chen, Y., Crawford, M. and Ghosh, J. (2007). “Knowledge based stacking of hyperspectral data for land cover classification”, *In Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, pp. 316-322.

Cheng, Q., Varshney, P. K., Arora, M.K (2006). “Logistic regression for feature selection and soft classification of remote sensing data, *IEEE Geoscience and Remote Sensing Letters*”, vol. 3, no. 4, pp. 491 - 494.

Cihlar, J., Xiao, Q., Chen, J., Beaubien, J., Fung, K. and Latifovic, R. (1998). “Classification by progressive generalization: A new automated methodology for remote sensing multichannel data”, *International Journal of Remote Sensing*, vol.19, no. 14, pp. 2685 – 2704.

Cohen, J. (1960). “A coefficient of agreement for nominal scales”, *Educational and Psychological Measurement*, vol.20, pp 37 – 46.

Congalton, R. G., Oderwald, R. G., and Mead, R. A. (1983). “Assessing landsat classification accuracy using discrete multivariate analysis statistical techniques”, *Photogrammetric Engineering and Remote Sensing*, vol.49, pp. 1671 – 1678.

Congalton, R. (1988). “A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data”, *Photogrammetric Engineering and Remote Sensing*, vol. 54, no.5, pp. 593-600.

Congalton, R.G. (1991). “A review of assessing the accuracy of classifications of remotely sensed data”, *Remote Sensing of Environment*, vol. 37, no.1, 35-46.

Congalton, R. G., and Green, K. (1998). “Assessing the accuracy of remotely Sensed data: Principles and practices”, (*Boca Raton, Florida: Lewis Publishers*).

Congalton, R.G, Green, K. (2009). “Assessing the accuracy of remotely sensed data: principles and practices”, *2nd Edition*, CRC Press, Taylor & Francis Group, Boca Raton, London, New CRC Press, Taylor & Francis Group, Boca Raton.

Cortes, C. and Vapnik, V. (1995). “Support vector networks”, *Machine Learning*, vol. 20, no.3, pp. 273–297.

Cybenko, G. (1989). “Approximation by superpositions of a sigmoidal function”, *Math. Contr. Signals Syst*, vol. 2, pp. 303–314.

Demir, B. and Ertürk, S. (2007). “Hyperspectral image classification using relevance vector machines”, *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 586-590

Dietterich, T. G., and Bakiri, G. (1995). “Solving multiclass learning problems via error-correcting output codes”, *Journal of Artificial Intelligence Research*, vol. 2, pp. 263 – 286.

Dietterich, T. G. (2001). “Ensemble methods in machine learning”, *In J. Kittler and F. Roli, editors, Multiple Classifier Systems*, LNCS vol. 1857, Springer, pp. 1-15.

Dwivedi, R.S., Kandrika, S. and Ramana, K.V. (2004). “Comparison of classifiers of remote-sensing data for land use/land cover mapping”, *Current Science,* vol. 86, no. 2, pp. 328-335.

Džeroski, S. and Zenko, B. (2004). “Is combining classifiers with stacking better than selecting the best one?”, *Machine Learning*, vol. 54, no. 3, pp. 255 – 273, (Hingham, MA, USA: Kluwer Academic Publishers)

Eastman, J.R., (2006) IDRISI Andes (Worcester, MA: Clark University).

Fayyad, U.M., Shapiro, G.P., Smyth, P., and Uthurusamy, R. (Eds.), (1996). Advances in knowledge discovery and data mining, AAAI/MIT Press.

Feller, W. (1968). “The Strong Law of Large Numbers”, *In An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed. New York: Wiley, pp. 243-245.

Foody, G.M., Boyd, D.S. and Sanchez-Hernandez, C. (2007). “Mapping a specific class with an ensemble of classifiers”, *International Journal of Remote Sensing*, vol. 28, no. 8, pp 1733 – 1746.

Freund, Y., and Schapire, R. (1996). “Experiments with a new boosting algorithm”, *In Proceedings of the 13th International Conference on Machine Learning*, Bari, Italy, pp. 148 – 156, (Morgan Kaufmann).

Garner, S. R., (1995). “WEKA: The Waikato environment for knowledge analysis,” *Proceedings of the NewZealand Computer Science Research Students Conference*, pp. 57–64.

Giacinto, G. and Roli, F. (1997). “Ensembles of neural networks for soft classification of remote sensing images”, In Proceedings of the European

Symposium on Intelligent Techniques, European Network for Fuzzy Logic and Uncertainty Modelling in Information Technology", Bari, Italy, pp. 166-170.

Giacinto, G. and Roli, F., (2001). "Design of effective neural network ensembles for image classification purposes", *Image and Vision Computing*, vol.19, pp. 699 – 707.

Gidudu, A. (2006). "Land cover mapping through optimizing remote sensing data for SVM classification", *PhD Thesis, University of Cape Town*, South Africa.

Gidudu, A, Abe, B and Marwala, T. (2008a). "Land Cover Mapping Using Ensemble Feature Selection Methods", CoRR abs/0811.2016**.**

Gidudu, A., Abe, B.T. and Marwala, T. (2008b). "Ensemble feature selection for hyperspectral imagery", *In Proceedings of the 19th Annual Symposium of the Pattern Recognition Association of South Africa*, Cape Town, South Africa, 27th – 29th November, pp. 27-32.

Gidudu, A., Abe, B. and Marwala, T. (2009a). "Investigating the effect of Ensemble Size on Classification Accuracy", *In Proceedings of the 2nd International Conference on Earth Observation for Global Changes*, (EOGC2009), Chengdu, China 25th – 29th May, pp. 2195 – 2199.

Gidudu, A., Abe, B. and Marwala, T. (2009b). "Random Ensemble Feature Selection for Land Cover Mapping", *In Proceedings of 2009 IEEE International Geoscience & Remote Sensing Symposium Earth observation-Origins to Applications*, Cape Town South Africa, 12th – 19th July, pp. 840 – 842.

Gong, P. and Zhang, A. (1999). “Noise effect on linear spectral unmixing”, *Geographic Information Sciences*, 5(1) pp. 52–57.

Green, A.A., Berman, M., Switzer, P. and Craig, M.D. (1988). “A transformation for ordering multispectral data in terms of image quality with implications for noise removal”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65-74.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M. S., Olah, M. R. and Williams, O. (1998). “Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)”, *Remote Sens. Environ*., vol. 65, no. 3, pp. 227–248.

Guo, B., Gunn, S.R., Damper, R.I. and Nelson, J.D.B. (2006). “Band selection for hyperspectral image classification using mutual information”, *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 4, pp. 522-526.

Guo, B., Gunn, S. R., Damper, R. I. and Nelson, J. D. B. (2008). “Customizing kernel functions for SVM – based hyperspectral image classification”, *IEEE Transactions on Image Processing*, vol. 17, no.4, pp. 622 – 629.

Guyon, I. and Elisseeff, A. (2003). “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182.

Haggag, A. A. (1981). “A variant of the generalized reduced gradient algorithm for nonlinear programming and its applications”, *European Journal of Operational Research*, vol. 7, no. 2, pp. 161–168.

Hall, F. G., Strebel, D. E., Nickeson, J. E., et al. (1991). “Radiometric rectification: toward a common radiometric response among multidate, multisensor images”, *Remote Sensing of Environment*, vol. 35, pp. 11–27.

Heinz, D.C. and Chang, C. (2001). “Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery”, *IEEE Transactions Geoscience & Remote Sensing*, vol. 39, no. 3, pp. 529-545.

Ho, T. K. (1998). “The random subspace method for constructing decision forests”, *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832 – 844.

Hornik, K. (1991). “Approximation capabilities of multilayer feedforward networks”, *Neural Networks*, vol. 4, pp. 251–257.

Hsu, Pai-Hui, (2007). “Feature extraction of hyperspectral images using wavelet and matching pursuit”, *In ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 78 – 92.

Huang, Z and Lees, B.G., (2004). “Combining non-parametric models for multisource predictive forest mapping”, *Photogrammetric Engineering and Remote Sensing*, vol.70, pp 415 – 425.

Huber, R. and Dutra, L.V. (2000). “Huber, R. and Dutra, L.V., (2000). “Classifier combination and feature selection for land-cover mapping from high-resolution airborne dual-band SAR data”, *Proc World Multiconf Systemics, Cybernetics and Informatics*, vol. 5, pp.370-375.

Hughes, G. F. (1968). “On the mean accuracy of statistical pattern recognition”, *IEEE Transactions on Information Theory*, vol. IT-14, no. 1, pp. 55–63.

Hugo Carrão1, H., Capão1, L., Bação, F. and Caetano1, M (2006). “Meris based land cover classification with self-organizing maps: Preliminary results”, *In Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover*, Center for Remote Sensing of Land Surfaces, Bonn, pp. 16 – 24.

Iordache, M.-D., Bioucas-Dias, J. M. and A., Plaza, (2011). “Sparse unmixing of hyperspectral data”, *Geoscience and Remote Sensing, IEEE Transactions*, vol. 49, no. 6, pp. 2014-2039.

Jiang, S. and Yu, W. (2009). “A Combination Classification Algorithm Based on Outlier detection and C4. 5,” *Springer Publications*.

Jimenez, L.O., Morales-Morell, A. and Creus, A. (1999). “Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and, neural networks”, *IEEE Transactions on Geoscience and Remote Sensing*, vol.37, no.3, part 1, pp. 1360 – 1366.

Jorgensen, J. (2004). “Remote sensing”, Online publication, *http://maps.unomaha.edu/Peterson/GIS/notes/RS2.htm*, Cited 9 July 2011.

Jung, M., Churkina, G., Henkel K, *et al.* (2006). “Exploiting synergies of global land cover products for carbon cycle modeling”, *Remote Sensing Environment*, vol. 101, pp. 534–53.

Keuchela, J., Naumanna, S., Heilera, M., and Siegmund, A. (2003). “Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data”, *Remote Sensing of Environment*, vol. 86, pp. 530–541.

Keshava, N. and Mustard, J. F. (2002). “Spectral unmixing”, *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44-57.

Kittler, J., Hatef, M., Duin, R.P. and Mates, J. (1998). “On combining classifiers”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no.3, pp. 226-239.

Kohavi, R. (1995). “A study of cross-validation and bootstrap for accuracy estimation and model selection,” In C. S. Mellish (Ed.), *Proceedings of the 14th International Joint Conference on Artificial Intelligence* pp. 1137–1143. San Francisco: Morgan Kaufmann

Kohavi, R and Wolpert, D. (1996). “Bias plus variance decomposition for zero-one loss functions”, In: L. Saitta (ed.): *Machine Learning: Proc. 13th International Conference*. pp.275 – 283, Morgan Kaufmann.

Kramer J. H. (2002). “Observation of the earth and its environment: Survey of missions and sensors”, *(4th Edition)*. (Berlin: Springer).

Kuncheva, L., and Whitaker, C., J. (2003). “Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy”, *Machine Learning*, vol. 51, pp. 181 – 207 (Hingham, MA, USA: Kluwer Academic Publishers).

Lacar, F.M. (2001). “Use of hyperspectral imagery for mapping grape varieties in the BarossaValley, South Australia”, *Geoscience and Remote Sensing Symposium (IGARSS,)01) – IEEE International*, vol. 6, pp. 2875 – 2877.

Lambin, E.F., Turner II, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X.B., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skanes, H., Stone, G.D., Svedin, U., Veldkamp, A., Vogel, C. and Xu, J.C., (2001). “The causes of land-use and land-cover change: moving beyond the myths”, *Global Environmental Change*, vol. 11, no. 4, pp. 261–269.

Landgrebe, D. (1997). “The evolution of landsat data analysis”, *Photogrammetric Engineering and Remote Sensing*, vol. 63, pp. 859 – 867.

Landgrebe, D. (1998). “Multispectral data analysis: A signal theory perspective”, *School of Electr. Comput. Eng.*, Purdue Univ., West Lafayette, IN.

Landgrebe, D. and Biehl, L. (2001). “An introduction to multispec”, Available online, *http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/*, Cited 10 July 2008.

Landgrebe, D. A. (2003). “Signal Theory Methods in Multispectral Remote Sensing”, *John Wiley and Sons, Inc.*, 111 River Street, Hoboken, NJ,

Landgrebe, D. (2005). “Multispectral land sensing: Where from, where to?” *IEEE Transactions on Geoscience and Remote Sensing*, vol.43, pp. 433 - 440.

Lathrop, R. G., Montesano, P., Haag, S., (2006). “A multi-scale segmentation approach to mapping seagrass habitats using airborne digital camera imagery”, *Photogramm Eng Remote Sens*, vol. 72, pp. 665–75.

Lennon, M., Mercier, G. and Hubert-Moy, L. (2002). “Classification of hyperspectral images with nonlinear filtering and support vector machines*”, In Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, Toronto, ON, Canada, vol. 3, pp. 1670 – 1672.

Levin N. (1999). “Fundamentals of remote sensing”, a book compiled using Internet resources, articles and personal knowledge, 225 pages.

Lippmann, R. P. (1989). “Pattern classification using neural networks”, *IEEE Commun. Mag.*, pp. 47–64.

Li, J. (2011) "Remote sensing image information mining with HPC cluster and DryadLINQ," in *Proceedings of the 49th ACM Southeast Conference ACM SE’11*, 2011, pp. 227 – 232.

Liu, H., Fan, Y., Deng, X. and Ji, S. (2009). "Parallel Processing Architecture of Remotely Sensed Image Processing System Based on Cluster", *Image and Signal Processing, 2009, CISP '09. 2nd International Congress on*, pp.1-4.

Martinez, P.J., Perez, R. M., Plaza, A., Aguilar, P.L., Cantero, M.C. and Plaza, J. (2006). “Endmember extraction algorithms for hyperspectral images”, *Annals of Geophysics*, vol. 49, no. 1, pp. 93-101.

Marwala, T. (2009). “Computational intelligence for missing data imputation, estimation, and management: Knowledge optimization techniques“, *Information Science Reference,* ISBN: 978-1-60566-336-4. Foreword by Dr Fulufhelo V. Nelwamondo, Harvard University.

Marwala, T. (2010). “Finite-element-model Updating Using Particle-swarm Optimization. Finite-element-model Updating Using Computional Intelligence Techniques*,” Springer London:* 67-84.

Marwala, T. and Lagazio, M. (2011). “Militarized conflict modeling using computational intelligence techniques“, *Springer-Verlag*, London, UK ISBN: 978-0857297891. Foreword by Professor Bruce Russett (Yale University, USA).

Marwala, T. (2012). “Condition monitoring using computational intelligence methods: Applications in mechanical and electrical systems“, *Springer-Verlagi*, London, UK ISBN-10: 1447123794, ISBN-13: 978-1447123798, Foreword by Professor Arthur Mutambara

Masisi, L., Nelwamondo, F.V. and Marwala, T., (2008). “The effect of structural diversity of an ensemble of classifiers on classification accuracy”, *International Association of Science and Technology for Development: Modeling and Simulation* (IASTED), pp. 135-140.

Mathur, A. and Foody, G. M. (2004). “Land cover classification by support vector machine: Towards efficient training”, *In Proceeding IEEE International Geoscience and Remote Sensing Symposium*, pp. 742 – 744.

MATLAB version 7.10.0. Natick, Massachusetts: The MathWorks Inc., 2010.

Maul, G. A. and Gordon, H. R. (1975). “On the use of LANDSAT 1 in optical oceanography,” *Rem. Sens. of Env*., vol 4, pp. 95 – 128

Melgani, F. and Bruzzone, L. (2004). “Classification of hyperspectral remote sensing images with support vector machines”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778-1790.

Moore, S. A., D’Addario, D. M., Kurinskas J. and Weiss, G. M., (2009). “Are Decision Trees Always Greener on the Open (Source) Side of the Fence?” *Proceedings of DMIN*, pp. 185–188

Mouatasim, A. E. (2010). “Two-phase generalized reduced gradient method for constrained global optimization”, *Journal of Applied Mathematics*, vol. 2010, Article ID 976529, 19 pages. doi:10.1155/2010/976529

Mountrakis, G., Im, J. and Ogole, C. (2011). “Support vector machines in remote sensing: a review”, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol, 66, pp. 247–259.

Nandgaonkar, S.; Jagtap, R.; Anarase, P.; Khadake, B.; Betale, A. (2010). "Image mining of textual images using low-level image features," Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on , vol.9, no., pp.588 – 592.

Natural Resources Cananda, \Tutorial: Fundamentals of Remote Sensing, *http://www.ccrs.nrcan.gc.ca/resource/tutor/fundam/chapter1/01_e.php* Cited 9 september 2007.

Opitz, D. (1999). "Feature Selection for Ensembles", In Proceedings of the 16th National Conference on Artificial/ Intelligence (AAAI), pp. 379 – 384.

Oza, C. N. and Tumer, K. (2008). "Classifier ensemble: Select real – world applications", *Information Fusion*, vol. 9, pp. 4 – 20.

Pal, M. (2003). "Random forests for land cover classification", *In Proceedings Geoscience and Remote Sensing Symposium*, vol. 6, pp. 3510 – 3512.

Pal, M. (2005). "Random forest classifier for remote sensing classification", *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217 – 222.

Pal, M. (2007). "Ensemble learning with decision tree for remote sensing classification", *In Proceedings of World Academy of Science, Engineering and Technology*, vol. 26, pp. 1307 – 6884.

Parikh, D., and Polikar, R. (2007). "An ensemble-based incremental learning approach to data fusion", *IEEE Transactions on Systems, Man, And Cybernetics—Part B: Cybernetics*, vol. 37, no. 2, pp. 437-450.

Platt, R. V., and A. F. H. Goetz. 2004. A comparison of AVIRIS and synthetic Landsat data for land use classification at the urban fringe", *PhotogrammetricEngineering and Remote Sensing* 70, pp. 813–19.

Plaza, A., Martinez, P., Perez, R. and Plaza, J. (2002). “Spatial/spectral endmember extraction by multidimensional morphological operations”, *IEEE Trans. Geosci. & Remote Sensing*, vol. 40, pp. 2025-2041.

Plaza, A., Martínez, P., Plaza, J. and Pérez, R. (2003). “Spectral analysis of hyperspectral image data”, *Advances in Technique for Analysis of Remotely Sensed Data, IEEE Workshop*, pp. 298 – 307.

Plaza, A., Martinez, P., Perez, R.M. and Plaza, J. (2004a). “A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profile, *Pattern Recognition*”, vol. 37, no.6, pp. 1097 – 1116.

Plaza, A., Martinez, P., Perez, R. and Plaza, J. (2004b). “A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data”, *IEEE Trans. Geosci. & Remote Sensing*, vol. 42, no. 3, pp. 650-663.

Plaza, A., Plaza, J. and Cristo, A. (2008). “Morphological feature extraction and spectral unmixing of hyperspectral images”, *IEEE International Symposium on Signal Processing and Information Technology*, pp. 12-17.

Pinho, C. M. D., Silva, F. C., Fonseca, L. M. C. and Monteiro, A. M. V. (2008). “Intra-urban land cover classification from high-resolution images using the C4.5 algorithm”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37. Part B7. pp. 695-699.

Polat, K. and Güne, S. A., (2009). “Novel hybrid intelligent method based on C4. 5 decision tree classifier and one against-all approach for multi-class

classification problems," *Expert Systems with Applications*, vol. 36, pp. 1587-1592

Polikar, R. (2006). "Ensemble based systems in decision making", *IEEE Circuits and Systems Magazine*, third quarter, pp. 21 – 45.

Qi, H. and Huang, M. (2007). "Research on SVM ensemble and its application to remote sensing classification", *In of Proceeding International Conference on Intelligent Systems and Knowledge Engineering*, pp. 271 – 278.

Randall, B. S. (2006). "Introduction to hyperspectral imaging", *MicroImages, Inc*.,

Rosenfield, G.H, and Fitzpatrick-Lins, K. (1986). "A coefficient of agreement as a measure of hematic classification accuracy", *Photogrammetric Engineering and Remote Sensing*, vol. 52, pp. 223 – 227

Richard, O. D., Perter E. H. and David G S. (2006). "Pattern classification [M]", *Beijing,: Citic Publishing House & China Machine Press, CITIC Publishing House*, pp. 24 - 28.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. (2011). "An assessment of the effectiveness of a random forest classifier for land cover classification", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol., 67, pp. 93-104.

Roli, F., Giacinto, G. and Vernazza, G. (1997). "Comparison and combination of statistical and neural networks algorithms for remote-sensing image classification", *Neurocomputation in Remote Sensing Data Analysis, Austin, J.*,

Kanellopoulos, I., Roli, F. and Wilkinson G. (Eds.), Berlin: Springer-Verlag, pp. 117-124.

Richards, J. A., and Jia, X. (1999). "Remote sensing digital image analysis: An introduction", (3rd Edition). (Berlin: Springer).

Richard, M. D. and Lippmann, R. (1991). "Neural network classifiers estimate Bayesian a posteriori probabilities", *Neural Comput.,vol.* 3, pp. 461–483.

Sabino (2005). "Fundamentals of remote sensing", *Canadian Center for Remote Sensing*, pp. 148 & 161

Sanchez, S., Martin, G., Plaza, A. and Chang, C. (2010). "GPU implementation of fully constrained linear spectral unmixing for remotely sensed hyperspectral data exploitation", *In Proceedings SPIE Satellite Data Compression, Communications, and Processing VI*, vol. 7810, pp. 78100G-1 – 78100G-11.

Settle, J. J. and Drake, N. A. (1993). "Linear mixing and the estimation of ground cover proportions," *International Journal of Remote Sensing*, 14(6) pp.1159-1177.

Shaw G.A., Burke H.K. (2003). "Spectral imaging for remote sensing", *Lincoln Laboratory Journal*, vol.14, no.1., pp. 3 – 28.

Shipp, C., A. and Kuncheva, L. (2002). "Relationship between combination methods and measures of diversity in combining classifiers", *Information Fusion*, vol. 3, pp. 135 – 148.

Silva, M.P.S., Câmara, G., Escada, M. I. S. and Souza, R. C. M. (2008). “Remote sensing image mining: Detecting agents of land use change in tropical forest areas”, *International Jornal of Remote Sensing*, vol. 29, no. 16, pp. 4803 – 4822.

Steele, B.M. and Patterson D.A. (2001). “Land cover mapping using combination and ensemble classifiers”, *Proceedings of the 33rd symposium on the Interface Foundation of North America*, pp. 236-247.

Story, M. and Congalton, R.G. (1986). “Accuracy assessment: a user’s perspective”, *Photogrammetric Engineering and Remote Sensing*, 52, pp. 397-399.

Tarantino, C., D'Addabbo, A., Castellana, L., Blonda, P., Pasquariello, G., Ancona, N. and Satalino, G. (2006). “Neural network ensemble and support vector machine classifiers: an Application to remote sensed data”, *Fuzzy Logic and Applications Lecture Notes in Artificial Intelligence*, vol. 2955, pp. 317-323, ISBN: 3-540-31019-3, Springer-Verlag.

Theiler J, Lavenier D, Harvey N, Perkins S, and Szymanski, J. (2000). “Using blocks of skewers for faster computation of pixel purity index”, *Proceedings of the SPIE International Conference on Optical Science and Technology*, 4132 pp. 61-71.

Tsymbal, A., Pechenizkiy, M., and Cunningham, P. (2005). “Diversity in search strategies for ensemble feature selection”, *Information Fusion*, vol. 6, no. 1, pp. 83 – 98.

Yu, M. and Ai, T. H., (2009). “Study of RS data classification based on rough sets and C4. 5 algorithm,” *In Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*

Valentini, G., and Masulli, F. (2002). “Ensembles of learning machines”, *In: Neural Nets WIRN Vietri, Lecture Notes in Computer Sciences*, edited by; Tagliaferri, R and Marinaro, M., vol. 2486, pp 3 – 19.

Van der Meer, F. D. and Jia, X. (2012). “Collinearity and orthogonality of endmembers in linear spectral unmixing,” *International Journal of Applied Earth Observation and Geoinformation*, 18(0), pp. 491-503.

Vapnik, V. (1995). “The nature of statistical learning theory”, Springer-Verlag, New York.

Vapnik, V. (1999). “The nature of statistical learning theory”, second edition. New York: Springer-Verlag.

Varshney, P. K. and Arora, M. K. (2004). “Advanced image processing techniques for remotely sensed hyperspectral data”, New York: Springer-Verlag.

Udelhoven, T., Waske, B., Linden, S. and Heitz, S. (2009). “Land-cover classification of hypertemporal data using ensemble systems”, *IEEE Transaction on Geoscience & Remote Sensing*, pp. III-1012 – III -1015.

Wacker, A. G., and Landgrebe, D. (1972). “Minimum distance classification in remote sensing”, *LARS Technical Reports. Paper 25.* Online publication. *http://docs.lib.purdue.edu/larstech/25* Cited 11 July 2011.

Wang, C., Du, Z., Liu, Z. and Liu, Y. (2008). “Study on decision tree land cover classification based on MODIS data”, *In International Workshop on Earth Observation and Remote Sensing Applications*, pp. 1 – 6.

Winter, M.E. (1999). “N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data”, *In Proceedings SPIE Imaging Spectrometry V*, vol. 3753, pp. 266-275.

Witten, I. H. and Frank, E. (2005). “Data Mining: Practical machine learning tools and techniques”, 2nd Edition. Morgan Kaufmann, San Francisco, pp. 189 – 199, 316 – 319, 337 – 423.

Wolter, P. T., Johnston, C. A. and Niemi, G. (2005). “Mapping submergent aquatic vegetation in the US Great Lakes using Quickbird satellite data”, *International Journal of Remote Sensing*, vol. 26, pp. 5255–74.

Xiao, X. M, Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore III, B., Ojima, D. (2004). “Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data”, *Remote Sens. Environ.*, vol. 91, pp. 256–70.

Xie, Y., Sha, Z. and Yu, M., (2008). “Remote sensing imagery in vegetation mapping: a review”, *Journal of Plant Ecology*, vol. 1, pp. 9–23.

Zhang, G.P. (2000). “Neural networks for classification: A survey”, *IEEE Transactions on Systems, Man, and Cybernetics-part C: Applications and Reviews*, vol. 30, no. 4, pp. 451 – 462.

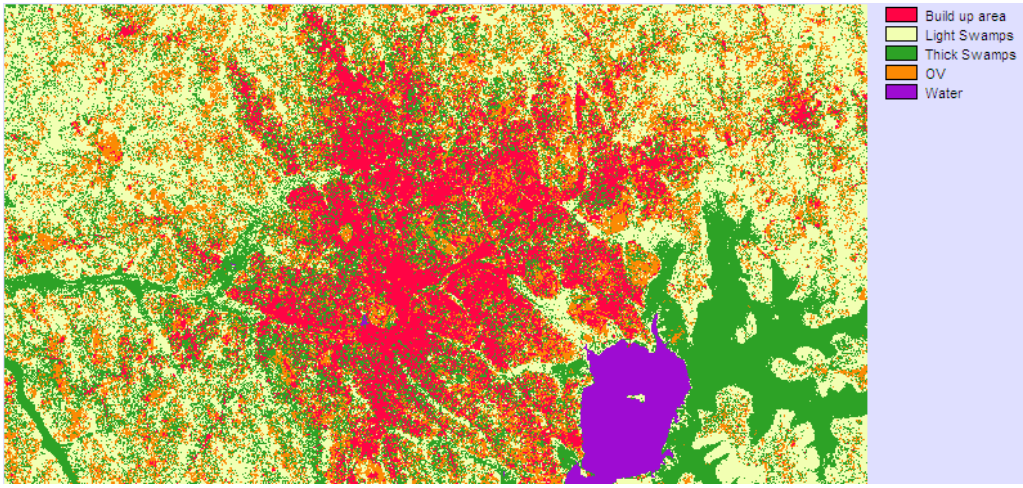
Zhang, L. and Huang, X. (2010). “Object-oriented subspace analysis for airborne hyperspectral remote sensing imagery”, *Neurocomputing*, vol. 73, pp. 927-936.

Zhou, L and Yang, X. (2009). “A preliminary study of three training methods for land cover classification by artificial neural networks”, *In Urban Remote Sensing Joint Event*, pp. 1 – 5.

Appendix A

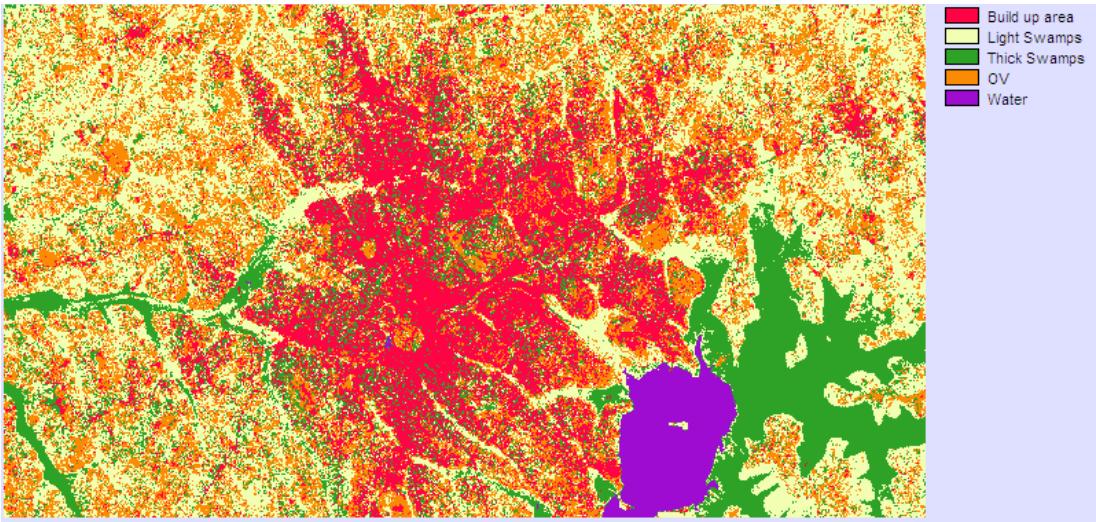
Land cover classification accuracy maps for Kampala dataset

Bhattacharyya separability index



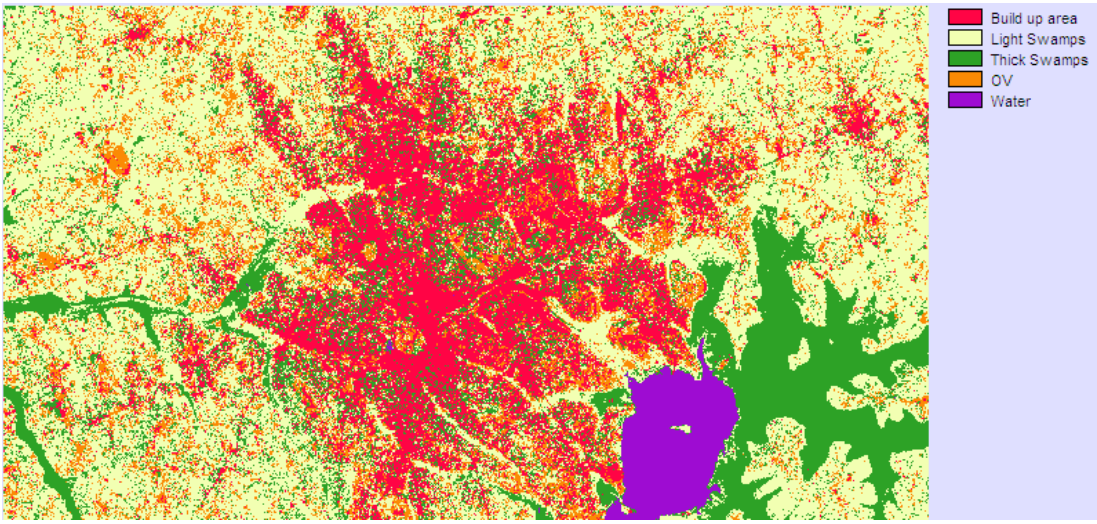
Map generated from accuracy result of Ensemble using BSI with 3 bands

Map generated from accuracy result of Ensemble using BSI with 4 bands

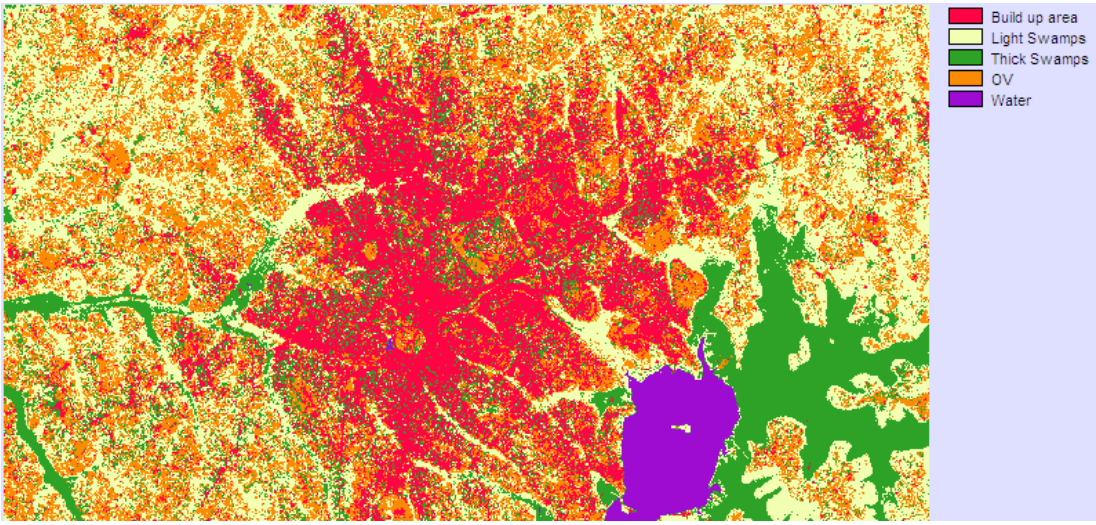


Map generated from accuracy result of Ensemble using BSI with 5 bands

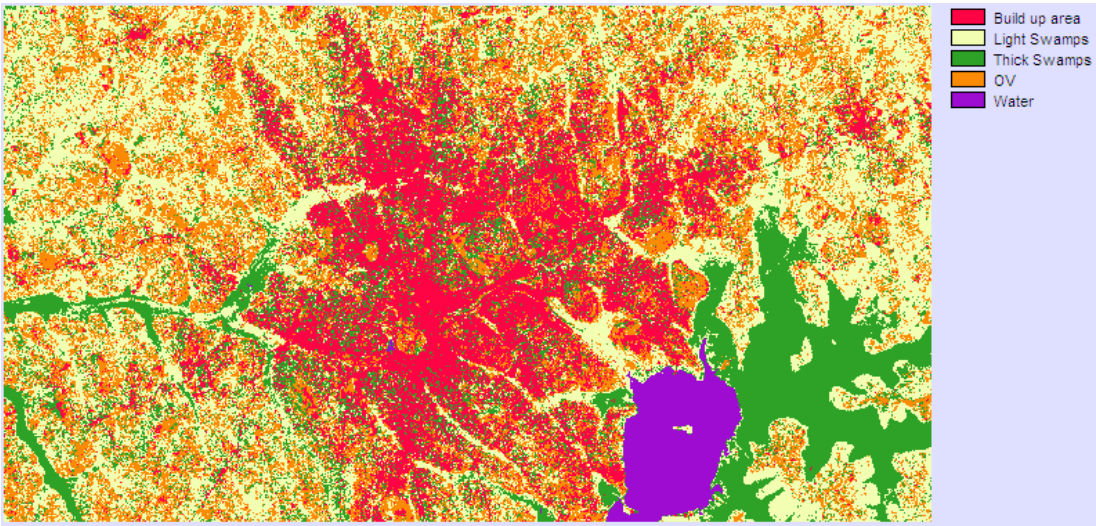
Divergence separability index



Map generated from accuracy result of Ensemble using DSI with 3 bands



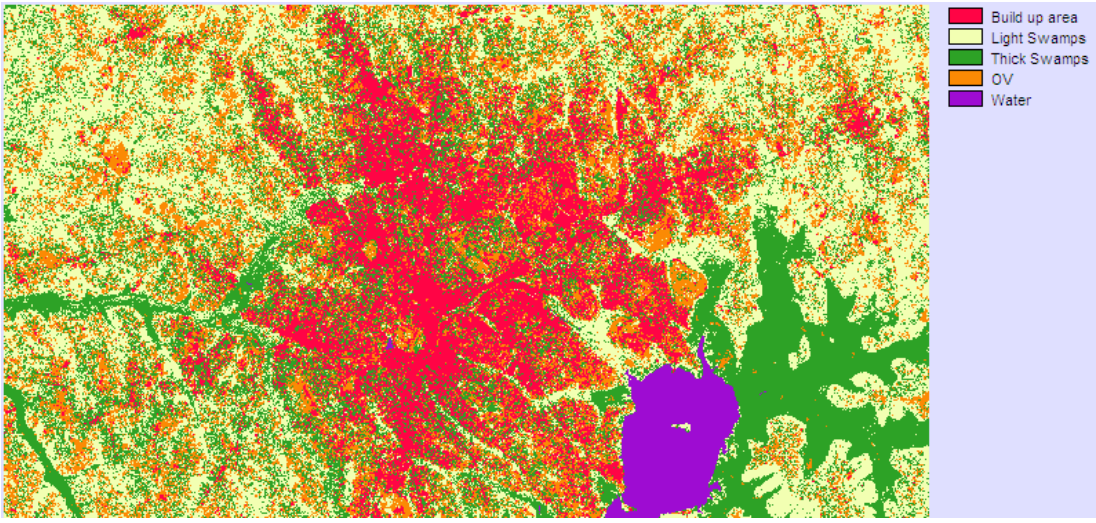
Map generated from accuracy result of Ensemble using DSI with 4 bands



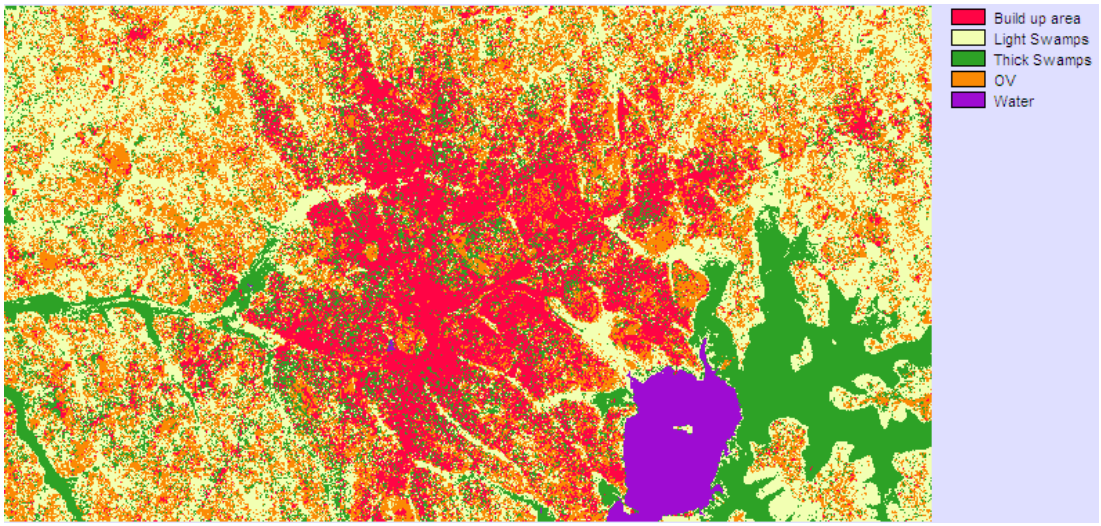
Map generated from accuracy result of Ensemble using DSI with 5 bands

Transformed divergence separability index

Map generated from accuracy result of Ensemble using TDSI with 3 bands

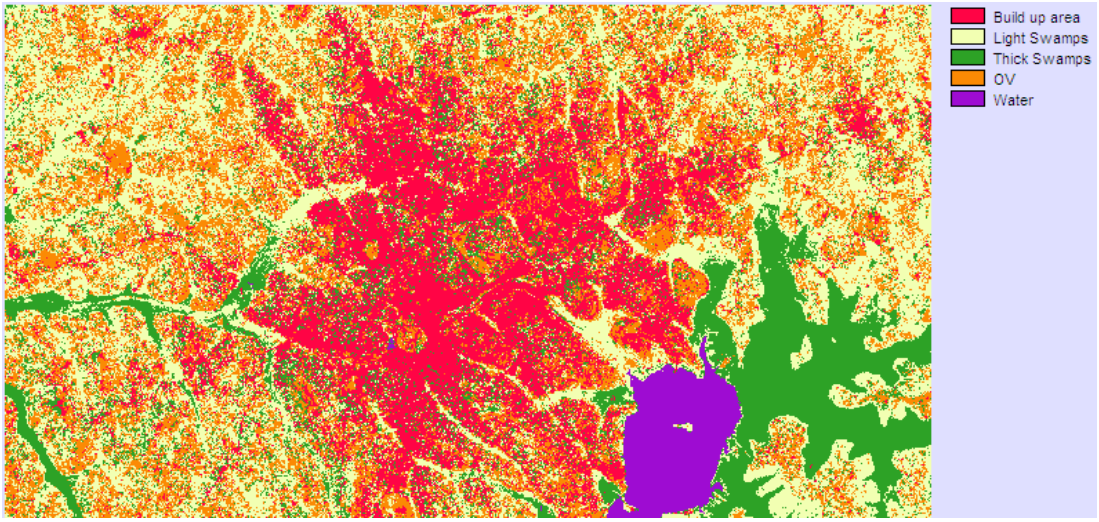


Map generated from accuracy result of Ensemble using TDSI with 4 bands



Map generated from accuracy result of Ensemble using TDSI with 5 bands

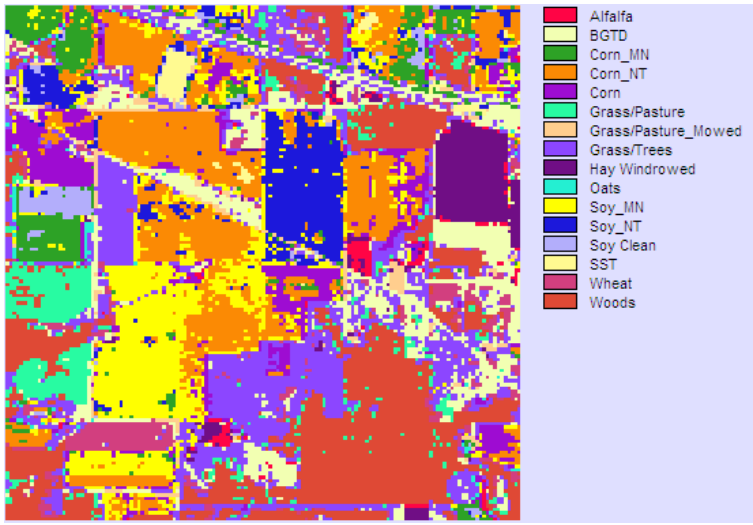
No seperability measure



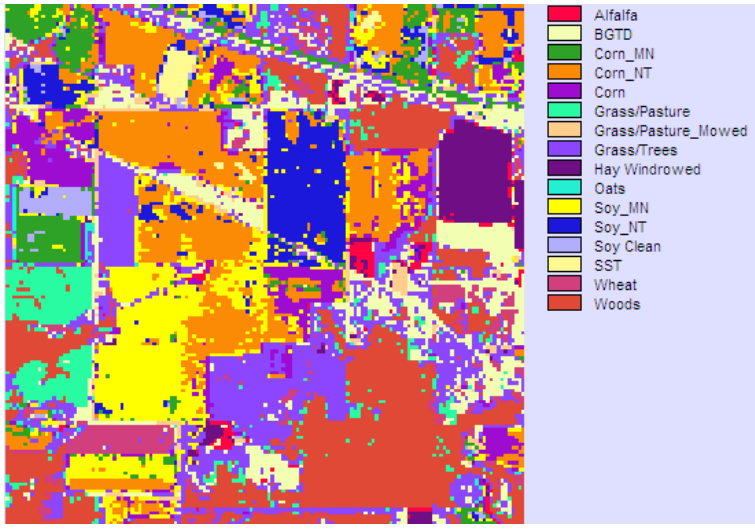
Map generated from accuracy result of Ensemble using NSM with 6 bands

Appendix B

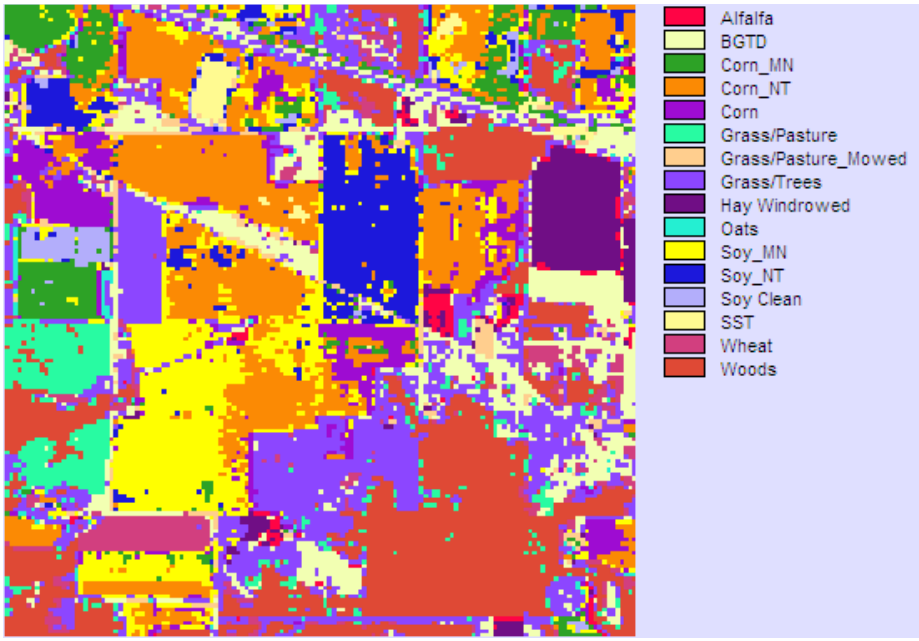
Indiana land cover accuracy maps on effect of ensemble size on Classification



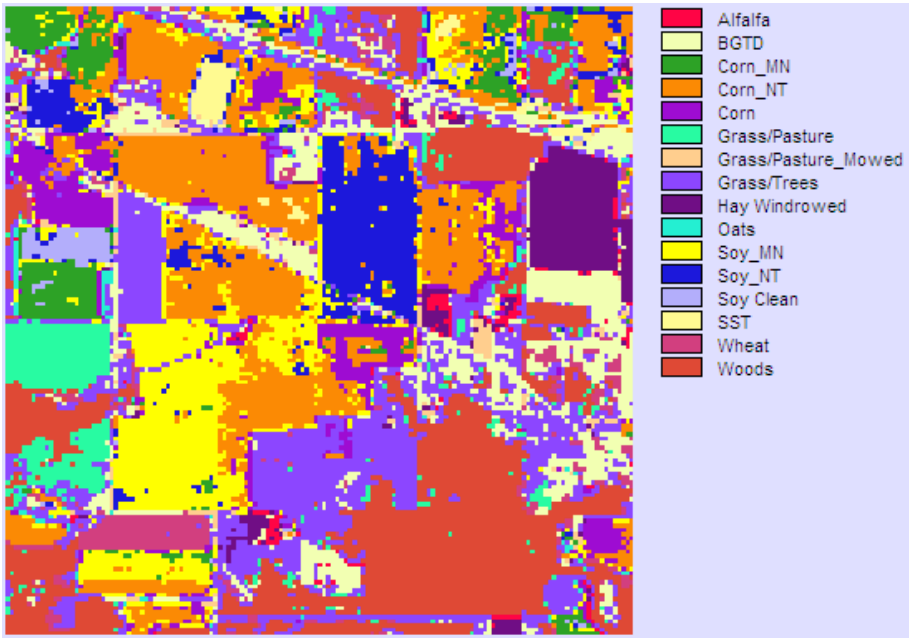
Ensemble made up of 8 bands classification Accuracy map



Ensemble made up of 10 bands classification Accuracy map



Ensemble made up of 12 bands classification Accuracy map



Ensemble made up of 14 bands classification Accuracy map