Role of Hybrid Spectral Similarity Measures for Semi-Supervised Fuzzy Classifier

SURBHI SHARMA

March, 2018

SUPERVISORS:

Dr. Anil Kumar

Dr. Valentyn A. Tolpekin

ROLE OF HYBRID SPECTRAL SIMILARITY MEASURES FOR SEMI-SUPERVISED FUZZY CLASSIFIER

SURBHI SHARMA

Enschede, The Netherlands, March, 2018

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfillment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Geoinformatics

SUPERVISORS:

Dr. Anil Kumar

Dr. Valentyn A. Tolpekin

THESIS ASSESSMENT BOARD:

Dr.Ir. R.A. de By (Chair)

Prof. Dr. Ir. A. Stein (ITC Professor)

Prof. S.K. Ghosh (External Examiner, Indian Institute of Technology (IIT), Roorkee)

ABSTRACT

To elucidate the information present in a remote sensing image, classification has helped significantly. With the presence of uncertainties in remotely sensed images, soft classification techniques are often considered. The desideratum for good classification accuracies with minimized efforts has led to the options where a few labeled data could enhance the classification accuracies. In this research, a semi-supervised approach is considered to minimize the efforts in collecting huge labeled training samples which is often strenuous and time-consuming. To handle uncertainties better, a Possibilistic *c* Means classifier is used and considered with different approaches as per the labeled training data. The similarity between two pixel vectors becomes important when it comes to collecting unlabeled data in a semi-supervised approach, or when the classification is performed based on assigning more membership values to similar pixels. The hybrid spectral similarity measures, Spectral Information Divergence with Spectral Angle Measure and Spectral Information Divergence with Spectral Correlation Angle are used to measure the similarity between two pixel vectors due to their proven capabilities for capturing high band-to-band variability for a hyperspectral imagery. Their roles are studied in view of a multispectral imagery. The Possibilistic *c* Means classifier is used with semi-supervised training data using different hybrid measures. Due to the availability of a few labeled training samples, Mean Shift algorithm is employed to refine the training data and to shift the mean for a Possibilistic *c* Means classification algorithm to a higher density region. In addition, the proposed methods relate the bandwidth parameter from Mean Shift algorithm to the bandwidth or resolution parameter of Possibilistic *c* Means classifier with an iterative procedure to capture the class variances. The methods are applied to an input LANDSAT-8 imagery with 30 m spatial resolution. The classification accuracies for the proposed methods are tested in reference to a higher resolution FORMOSAT-2 image (8 m spatial resolution), classified with Support Vector Machine and are reported with the Root Mean Square Error (RMSE) and Fuzzy Error Matrix (FERM). The proposed methods are compared with the conventional methods such as supervised approach and Euclidean distance as a similarity measure.

It is found that the results from a completely supervised approach are comparable with the results of semi-supervised approach when a Possibilistic *c* Means classifier is incorporated with the respective distance measures. In comparison to hybrid measures, Euclidean distance is found to be the best in terms of capturing the high inter-class and intra-class variability when incorporated as a distance measure in a Possibilistic *c* Means classifier and also when measuring the similarity between two pixels with highest overall accuracies and lowest Global Root Mean Square Error. Among hybrid measures, Spectral Information Divergence with Spectral Correlation Angle works best in terms of measuring the similarity between pixel vectors and as a distance measure in a Possibilistic *c* Means classifier. Also, the hybrid measure, Spectral Information Divergence with Spectral Angle Measure works best for Eucalyptus class where intra-class variability is low.

In addition, the classification methods with just a few labeled samples, after shifting the mean to a higher density region using Mean Shift algorithm gives comparable results to both supervised (large labeled data) as well as semi-supervised approaches, with respective distance measures. Also, it is possible to capture variance and achieve higher classification accuracies with a few labeled training samples using Mean Shift algorithm, and a Possibilistic *c* Means classifier.

Keywords: *Possibilistic c* Means*, Semi-supervised learning, Mean Shift algorithm, Spectral similarity measures*

ACKNOWLEDGEMENTS

The journey of this research has left me with a lot of wonderful memories. I am grateful and thankful to every person, conditions, and situations that came into my life and filled my life with experiences. I am grateful to this life for giving me the opportunity of my interest and keeping me motivated throughout this journey.

I would like to draw my sincere regards to both of my supervisors Dr. Anil Kumar and Dr. Valentyn Tolpekin for their constant support, advice and for sharing their enriching experiences throughout this research. I am thankful and grateful to both of them for everything and for helping me become a better person and a researcher.

I would like to thank all of my fellow classmates and all the people at IIRS and ITC who have been a constant source of inspiration, help, and support for the times when I needed the most. The time spent with all of you is unforgettable and would be missed.

Finally, I would like to thank my parents, family, teachers and my dear friends without whom it would have been impossible to complete this research.

Surbhi Sharma

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

A crisp boundary is a subset of vague interpretation by human minds.

- *Anonymous*

1.1. Motivation and Problem Statement

A picture helps to explicate more than a million words. It elucidates complex relationships between objects with their sizes and spatial positions. The field of remote sensing has grown to derive these relationships with ease and has become an indispensable part for obtaining valuable information by means of observing an object, scene or phenomenon without any physical contact (Richards & Jia, 2013). It is one of the ways of obtaining information about the surface of the Earth, atmosphere, and oceans to map remote areas cheaply and efficiently with the help of satellites or aircraft (Campbell, 1996). The ongoing advancements in spatial, spectral, temporal and radiometric resolutions of sensors have increased the applicability of data for mapping and classification of land use land cover, mining, climate change, urban planning, oceanography and many more.

Classification is an image processing technique for translating the reflectance or digital numbers (DN) to thematic information (Richards & Jia, 2013). The objects on images are classified into themes or classes in order to reduce the complexity of information present in remotely sensed images for analysis. It can be unsupervised, supervised, or semi-supervised (Olivier et al., 2006; Richards & Jia, 2013). In unsupervised classification data is divided into a set of spectral clusters or similar group or cluster albeit labels are symbolic and does not represent ground cover or information classes. The clusters or groups are then classified into information classes by the users. On the other hand, supervised classification needs a sufficient amount of training data as a representative of a specific class based on prior knowledge. In case of supervised learning, a collection of pure training or labeled samples, both in terms of quality and quantity is usually expensive, time-consuming and require a lot of manual work (Persello & Bruzzone, 2014). There may be the presence of outliers in training samples or insufficiency of pure training samples.

The semi-supervised approach, on the other hand, is in between unsupervised and supervised learning approaches in which partial supervision is provided (Olivier et al., 2006). It works with a large number of untrained data (unlabeled data) and a minimum number of training samples (labeled data). Based on smoothness, cluster and manifold assumptions, it is a better classification approach than supervised classification to achieve higher accuracies with limited availability of training samples (Persello & Bruzzone, 2014). Semi-supervised learning algorithm uses various models, features, kernel, and similarity functions that help to classify data with least number of labeled training samples that are pure and captures intra-class variability (Olivier et al., 2006).

Often present uncertainties and vagueness in collecting pure samples lead to low classification accuracies, poor analysis, and decision-making. Their existence in remote sensing images makes it difficult to comprehend the veracious nature of information. The uncertainties can be due to vague class definition, mixed pixels, and transition zones or fuzzy boundaries (Lucieer, 2004). The vagueness and imprecision also called fuzziness that arises from complex real-world phenomenon may be dealt with by means of applying Fuzzy Set Theory (Zadeh, 1965).

The fuzzy set as introduced by Zadeh (1965) has a fuzzy membership function with flexible membership values varying between zero and one. This is in comparison to crisp set, a part of well-known and established classical set theory has an advantage of representing vague classes (Enderton, 1977; Zimmermann, 2010). The classical set theory as used in hard classification has sharp and crisp boundaries and is unable to deal with fuzziness and or complexity of real-world phenomena. It assigns a fixed membership value, either zero or one to a pixel. The soft classification technique, on the other hand, is more helpful in classifying the images into a set of themes for further analysis.

Soft classification approaches have been successful in classifying and quantifying uncertainties present in the images due to both vague class definitions as well as mixed pixels (Foody, 1996). As per definition, a mixed pixel is a pixel having multiple classes within itself. Many sub-pixel classification approaches such as Linear Spectral Unmixing have been used to quantify uncertainties due to mixed pixels (Foody & Cox, 1994). However, uncertainties due to poorly defined classes are often dealt with fuzzy classification approaches (Foody, 1996; Lucieer, 2004). In many cases, it has been found that fuzzy classifiers incorporate fuzziness in assigning membership values and are suitable for modeling uncertainties (Stein, 2010).

Fuzzy classifiers can be used to classify remote sensing images and can be based on supervised, unsupervised or semi-supervised learning approaches. Semi-supervised fuzzy classifiers are fuzzy classifiers with a semi-supervised learning approach that works with a limited amount of labeled data and a large number of unlabeled data. The semi-supervised approach can be applied in the pre-classification stage, or it can be integrated with an objective function of the classifier at the time of classification. As a part of pre-classification step, semi-supervision takes into consideration a small amount of pure labeled training samples and then builds upon by increasing the number of unlabeled training samples by following one of the assumptions of semi-supervised learning approaches such as Continuity, Cluster or Manifold (Zhu, 2008). On the other hand, integrating the semi-supervised approach at classification stage needs modification and optimization of the objective function of a classifier with the available labeled training set information (Macario & Carvalho, 2010; Pedrycz & Waletzky, 1997).

In fuzzy classification, representation of similarity of each pixel to a cluster is of key importance. It is often expressed in terms of membership function whose value lies in between zero or one. The values close to one represents higher similarity between a pixel and a cluster and vice versa (Bezdek et al., 1984). Pixels are clustered precisely based on their value of similarity. To enhance the accuracy of classification further, a conventional distance measure (Euclidean distance) may be replaced by a more effective similarity measure that can meticulously evaluate the distance from a pixel to cluster (mean) (Macario & Carvalho, 2015). Therefore, an effective distance measure (statistical distance), or similarity measure can further enhance the classification accuracy. Various spectral similarity measures have proved to work for hyperspectral imagery such as Spectral Angle Measure as introduced by Kruse et al. (1993), Spectral Correlation Measure by van der Meer & Bakker (1997), Spectral Information Divergence by Chang (2000) that calculate the similarity between two pixels based on their spectral information.

Spectral Information Divergence (SID) is a stochastic spectral similarity measure in which spectrum of a pixel vector is modeled as a probability distribution, and spectral similarity is defined by measuring the discrepancy between the probability distribution of two pixels spectra (Chang, 2000). It relies on the concept of Statistical Divergence (Fano, 1961) and Kullback–Leibler divergence (Kullback, 1997). Spectral Angle Measure (SAM) and Spectral Correlation Angle (SCA) are deterministic approaches that calculate spectral angle and spectral correlation between the spectral signatures of two pixel vectors respectively. For smaller distances, the values of Spectral Angle Measure and Euclidean distance are close to each other (Du et al., 2004; van der Meer, 2006). The combination of deterministic and stochastic approaches such as

Spectral Information Divergence with Spectral Angle Measure and Spectral Correlation Angle as used by Du et al. (2004), van der Meer (2006) and Kumar et al. (2011) make two similar spectra even more similar and two dissimilar spectra more distinct. In addition, another parameter defined as Spectral Discriminatory Power, described by Chang (2000) compares the effectiveness of different spectral similarity measures among themselves. It discriminates one pixel from another relative to a reference pixel. These hybrid spectral similarity measures work well with high spectral information that is generally present in hyperspectral images and proved to have a relatively high spectral discriminatory power than individual (non-hybrid) measures (van der Meer, 2006).

Due to the advantages of these hybrid spectral similarity measures in distinctly identifying the spectral similarities in hyperspectral imagery, they can also be used to generate precise spectra or pure samples for multispectral image classification by precisely refining the training set outliers by measuring similarity among them. Work presented by Chauhan & Krishna Mohan (2014) has shown that these hybrid measures were successful in generating precise spectra from hyperspectral field spectra by capturing band to band variability. In case of semi-supervised learning approach, these hybrid measures can be beneficial in identifying the unlabeled data by encapsulating the similar features from the limited labeled training data.

The hybrid spectral similarity measures can be used in a fuzzy classifier as a statistical distance measure which captures the spectral band to band variability of a pixel vector of which the conventional distance and similarity measures such as Euclidean distance are insensitive to. These hybrid spectral similarity measures have been studied with respect to hyperspectral remote sensing images due to their higher spectral resolution. However, hyperspectral images are very much sensitive to spectral variability due to atmospheric effects at the time of data acquisition. On the other hand, multi-spectral images have high within-class variability but are less sensitive to atmospheric variations. Therefore, the study on the behavior of multispectral imagery with these hybrid measures is a part of this research.

Amidst a lot of research on remote sensing digital image classification, this research aims to study the role of these hybrid spectral similarity measures for a semi-supervised fuzzy classifier for classifying multi-spectral remote sensing imagery and to study their behavior in improving the classification accuracies.

1.2. Research Identification

This research addresses the following problems:

1. Uncertainties due to mixed pixels or vague classes.
2. An insufficient number of labeled training samples.

The identification of research lies in the following components:

1. To handle uncertainties, fuzzy classifiers are used.
2. To handles noise more effectively, a Possibilistic *c* Means classifier is used.
3. Semi-supervised learning is incorporated to deal with an insufficient number of labeled samples.
4. Hybrid spectral similarity measures are used since they are more sensitive to spectral variations.

The two major roles of hybrid spectral similarity measures have been identified:

1. To generate unlabeled training data for applying semi-supervised learning approach.
2. Replacing the conventional Euclidean distance with hybrid measures in a semi-supervised Possibilistic fuzzy classifier as a distance based similarity measure.

Therefore, the following research studies the effectiveness of the proposed hybrid spectral similarity measures in solving the aforementioned problems.

1.3. Research Objectives

The main objective of the proposed research is to study the role of hybrid spectral similarity measures for a semi-supervised Possibilistic fuzzy classifier in classifying multispectral imagery.
The sub-objectives proposed to reach the main objectives are as follows:

1. To study the effectiveness of the proposed hybrid spectral similarity measures for multispectral imagery.
2. To develop a precise spectrum by using hybrid spectral similarity measures for a multispectral image.
3. To identify and apply a suitable approach for incorporating semi-supervised learning method in a Possibilistic fuzzy classifier.
4. To develop and optimize an objective function for semi-supervised Possibilistic fuzzy classifier with hybrid spectral similarity measures.
5. To compare the performance of hybrid spectral similarity measure with conventional similarity measures (Euclidean distance).
6. To compare the performance of a semi-supervised approach with the supervised approach.

1.4. Research Questions

The research questions identified to meet the objectives and sub-objectives of the research are as follows:

1. Are the proposed hybrid spectral similarity measures effective in identifying the similarity between pixels for a multi-spectral image?
2. Is semi-supervised approach applied before classification better than optimizing the objective function of the classifier?
3. What will be the effect of hybrid spectral similarity measures on classification accuracy as compared to conventional similarity measures?
4. Is proposed semi-supervised Possibilistic fuzzy classifier with hybrid spectral similarity measures better in dealing with uncertainties than supervised classifier?

1.5. Innovation Aimed At

This research attempts to apply a combination of probabilistic and deterministic approaches as a similarity measure in Possibilistic fuzzy classifiers for classifying a multispectral image. Several studies are based on similarity measures with their applications on the fuzzy classifier, but they are either based on probabilistic or deterministic measures separately. In addition, these hybrid spectral similarity measures have not been applied in identifying unlabeled data as a part of semi-supervised learning for multi-spectral imagery.

1.6. Research Approach

The generalized methodology has been shown in Figure 1.1.

1.7. Thesis Structure

The whole thesis has been organized into seven chapters. **Chapter 1** introduces some basic information on the background of the topic, research identification, objectives and associated research questions with an overview of a generalized methodology. **Chapter 2** briefs about the related literature on various classification techniques as well as the spectral similarity measures. **Chapter 3** discusses the methods adopted and classification approaches. **Chapter 4** documents the study area, dataset and software used for this research. **Chapter 5** shows the results obtained during the research. **Chapter 6** discusses the results obtained. **Chapter 7** concludes the research with answers to the research questions and future recommendations.

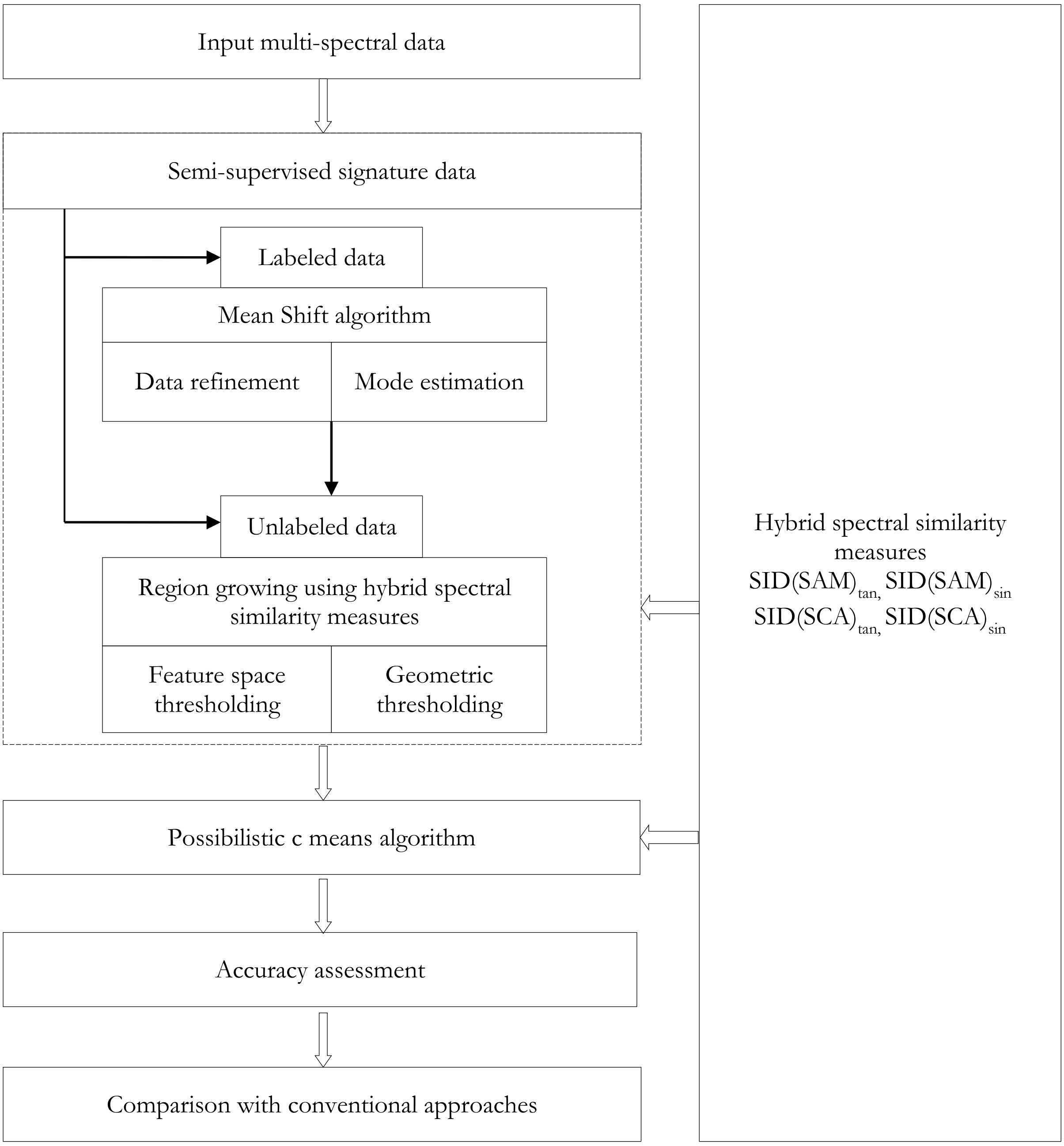


Figure 1.1. Generalized Methodology

2. LITERATURE REVIEW

This chapter presents the survey on the required literature to understand the research and reviewing the possible solutions. **Section 2.1** discusses the background knowledge of the available classification approaches and methods. **Section 2.2** provides the literature on the related works.

2.1. Background Review

2.1.1. Land cover classification techniques

Land cover designates both vegetative and non-vegetative features in a concrete form (Campbell, 1996). The remote sensing images are capable of providing accurate land cover information directly from features or visible evidence available on images. With broadly available literature on various image classification techniques to map land cover, various parametric & non-parametric algorithms, supervised, unsupervised, soft or hard classification techniques are widely used (Campbell, 1996; Jensen & Lulla, 1987). These methods could be used to classify images into spectral or informational classes and to generate thematic maps for further analysis (Richards & Jia, 2013; Tso & Mather, 2009).

Supervised classification technique such as Maximum Likelihood Classification is a parametric approach and requires effective estimations of mean and covariance. In cases where a reliable estimate of covariance matrix is not possible due to limited availability of training data, other non-parametric methods such Minimum Distance to mean classification is an effective approach for classification. But the technique has a drawback of being insensitive to spectral variances (Lillesand & Kiefer, 2015). Despite limitations of classes being normally distributed, Maximum Likelihood Classifier is most commonly used classifier (Davis et al., 1978).

Several other non- parametric classification techniques allow classifying and analyzing data with no pre-assumed distribution. Simplest of the techniques are kNN (nearest Neighbour) classification approach and Look Up tables but are often impractical and inefficient to obtain accurate and reliable results (Richards & Jia, 2013). Advanced techniques such as Artificial Neural Networks (ANN) with its commonly used Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM) are effective in dealing with complex networks by incorporating advanced learning mechanisms (Atkinson & Tatnall, 1997; Mountrakis et al., 2011). However, a large training data set is required for such complex learning procedures to produce generalized outcomes.

In terms of dealing with uncertainties, fuzzy clustering being the most widely used clustering algorithm helps in classifying data with vague definitions (Tso & Mather, 2009). Several sub-pixel classification techniques have also been used for handling mixed pixels.

2.1.2. Clustering and fuzzy classification

Grouping of spectrally similar pixels in a multispectral or a hyperspectral space is often known as Clustering (Richards & Jia, 2013). It is a process in which pixels that are more spectrally similar are put into similar clusters, by means of applying a similarity criterion. It is also known as unsupervised classification approach for classifying data based on the available spectral information as shown in Figure 2.1.

A crisp boundary or a statement is itself a subset of a vague concept (Fisher, 1997). A source deriving some information is often poorly defined and is limited to the perception of an observer. The geographical information, scientific facts, human-interpreted behaviors are themselves falsifiable in order to ameliorate the existing theories and facts.

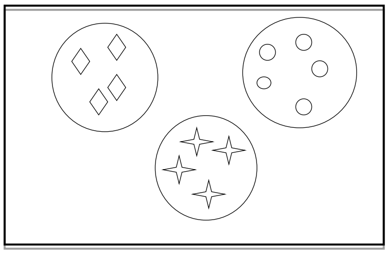


Figure 2.1. Clustering, showing grouping of similar features together. The features are represented by different shapes.

In the field of remote sensing, a few causes for uncertainties in remote sensing comes from the techniques and sources of data acquisition such as sensor resolution, sensor sensitivity, geometric calibration and atmospheric conditions (Lucieer, 2004). Furthermore, information obtained has uncertainties derived from poorly defined boundaries, often represented as fuzzy boundaries or fuzzy transition zones. Presence of mixed pixels and vague classes often results in uncertainties which can restrain the anatomization of remotely sensed image at the time of image classification (Shi et al., 2005).

Fuzzy set theory as proposed by Zadeh (1965) provides a framework for modeling vagueness. A lot of work has successfully substantiated the use of fuzzy approaches in quantifying the uncertainties present in remotely sensed images. A process of visualization of uncertainties as a qualitative approach manifests spatial behaviors and patterns of uncertainties (Cheng, 2002).

Clustering can be categorized as Soft Clustering and Hard Clustering. Soft clustering approach is better where the classification is based on membership values lies in between zero and one and has been applied in various applications to improve classification accuracies (Belohlavek & Vychodil, 2005; Shepard, 2005). Hard clustering approaches are on the other hand inappropriate in accounting reliable information where boundaries are vaguely defined, and pixels are mixed. Out of various fuzzy classification algorithms, Fuzzy c-Means (Bezdek et al., 1984) works well with various land cover classification. It can be implemented by applying both supervised and unsupervised procedures. It is a widely applied soft classification approach. It gives a membership value which is shared among different clusters. It is based on the concept of fuzzy logic and fuzzy set theory (Zadeh, 1965). In the literature, different work on mapping and estimating sub-pixel level information using Fuzzy c Means algorithm is available (Fisher & Pathirana, 1990; Foody, 2000). The Fuzzy c-Means algorithm is however sensitive to noise and outliers (Chawla, 2010).

2.1.3. Possibilistic *c* Means classification

In the domain of pattern recognition and computer vision, clustering methods have been one of the foremost choice for analyzing inherent features in images. Due to limitations of fuzzy clustering algorithms such as Fuzzy *c* Means, a possibilistic framework for clustering has been presented by Krishnapuram & Keller (1993). A possibilistic approach to clustering provides an intuitive way of dealing with class memberships, and the conceptual framework abides the constraint of the degree of compatibility or belongings than of degree of sharing. A possibilistic view of clustering is based on the idea of modeling vagueness by generating membership distribution of a feature vector in a class which is unique for a class and is independent of the distribution of membership in other classes.

The Possibilistic *c* Means algorithm characterize data with partitions that are able to handle noise and outliers more effectively than the conventional algorithms. Having an advantage of flexible hyperplane constraint that can handle noise and outliers, it produces coincident clusters and needs proper initialization values (Krishnapuram & Keller, 1996).

The Possibilistic *c* Means algorithm ameliorate the results of Fuzzy *c* Means algorithm by assigning pixels to more than one cluster by relaxing the hyperplane constraint and producing memberships as per the belongingness of a pixel to a cluster. As described by Krishnapuram & Keller (1993), noise points are assigned very low membership values in all clusters, and the closest point has a greater membership value than the farther ones which helps in obtaining more accurate cluster centers. In case of a single cluster, the Possibilistic *c* Means algorithm tend to outperform the Fuzzy *c* Means algorithm (Krishnapuram & Keller, 1996; Wu & Zhou, 2006).

2.1.4. Spectral similarity measures

Spectral Similarity measures are used to measure the similarity between the spectral signatures of the two pixels. In the literature, most commonly used measure is Euclidean distance. Other commonly used distance-based similarity measures include Manhattan, Chessboard, Cosine, Bray Curtis, Canberra, Mean Absolute Difference, etc. (Dongre, 2016; Mukhopadhaya, 2016) are able to define distance-based similarity between the pixels. Various other spectral similarity measures such as Spectral Angle Measure, Spectral Correlation Angle, and Spectral Information Divergence have been identified and are proved better in finding the similarity between pixels for a hyperspectral image (van der Meer, 2006). As a distance based metric, a similarity measures must follow the following axioms (Goshtasby, 2012).

Distance-Based Similarity Measures follows four distance axioms:

The relation between the distance (D) and Perceived Similarity (S)

Let $a = \{a_1, a_2, \dots a_N\}$ and $b = \{\, b_1,\ b_2, \dots\ b_N\}$ be any two measurements.

- *Equal self-similarity*:

  $D(a,a) = D(b,b)\ \forall a,b$

  $S(a,a) = S(b,b)\ \forall a,b$

- *Minimality:*

  $D(a,b) > D(a,a)\ \forall a \neq b$

  $S(a,b) < S(a,a)\ \forall a \neq b$

- *Symmetry*:

  $D(a,b) = D(b,a)\ \forall a,b$

  $S(a,b) = S(b,a)\ \forall a,b$

- *Triangle Inequality*:

  $D(a,b) + D(b,c)\ \geq D(a,c)\ \forall a,b,c$

If a and b are similar, b and c are similar implies that a and c should also be similar.

2.1.5. Semi-supervised approach to image classification

The semi-supervised approach is a technique in which a classifier takes into consideration only a small number of training data (labeled) all together with a large amount of untrained (unlabeled) data. This technique is effective and efficient as it does not have to deal with the problems related to a collection of a large number of labeled training samples.

In literature, there are many methods available for implementing the semi-supervised learning such as kernels, similarity measures, models, features, etc. (Zhu, 2008). The methods can be transductive or inductive. Some of the methods include co-training, self-training, expectation-maximization with generative mixture models, transductive Support Vector Machines, graph-based methods, etc. (Olivier et al., 2006). The choice of technique or model chosen depends upon clustering process and data characteristics such as similarity and features present in the data.

In case of semi-supervised clustering, a distance-metric is often used to provide some constraints on data to define whether a pixel or data would belong to a particular cluster or not. There are metric based models to detect inconsistency with labeled data (Schuurmans & Southey, 2002). The distance-based learning approach has many applications in semi-supervised, non-linear interpolation and clustering (Orlitsky, 2005).

The semi-supervised learning approach uses one of the following assumptions in order to use the unlabeled training data (Zhu, 2008):

1. *Continuity Assumption:* The likelihood of sharing a label is higher for closer points than the farther points.
2. *Cluster Assumption:* The likelihood of sharing a label is higher for points that are in a set of a common discrete cluster of data.
3. *Manifold Assumption:* The input space or input dimensions could be much larger than the actual data which tend to lie on the manifold of dimensions which are much lower.

2.1.6. Assessment of Accuracy

Accuracy is defined on the basis of how closely the test value is in agreement with the true value. It involves estimating the errors related to the sources. The sources of errors differ with the data, environmental conditions of experiments and measurement techniques. In the field of remote sensing, the different sources of collection of data have different scales of errors and can be accessed qualitatively and quantitatively in different ways by observing the difference in the obtained and true value.

One of the major requirement of assessing the quality of classified products of the remote sensing images comes with the largely available data for interpretation. Accuracy assessment has a major role in assessing the quality of different classification algorithms by quantifying the output of the classified products. It allows to have confidence in the obtained results and serves as an indication whether the objectives of the proposed analysis have been achieved or not. In terms of classification, there are various ways in which accuracies can be computed which include estimation of error matrix or confusion matrix or sometimes also called as contingency matrix. It also gives information about User's Accuracy, Producer's Accuracy, Overall Accuracy and kappa coefficient (Campbell, 1996).

The error matrix has been used successfully in case of hard classified outputs and gives the estimation of misclassification based on the reference data. The reference data can be higher resolution imagery, ground truth information, reference maps for the same area. For soft classified outputs, i.e., where a single pixel belongs to more than one class, the concept of Fuzzy Error Matrix (Binaghi et al., 1999) has been successfully put forth and implemented in accessing the accuracy of remotely sensed images.Various other accuracy assessment techniques involve Entropy, Root Mean Square Error, Sub-pixel Confusion Uncertainty Matrix at the pixel level or class level, and many more (Campbell, 1996; Kandpal, 2016).

2.2. Recent Related Work

In literature, various work has been done in the field of soft classification for various applications to improve classification accuracies (Belohlavek & Vychodil, 2005; Shepard, 2005). Out of various fuzzy classification algorithms, Fuzzy c-Means works well with various land cover classification but is sensitive to noise and outliers (Chawla, 2010). A possibilistic counterpart of Fuzzy c-Means is Possibilistic c-Means that have membership values representing "degree of belongingness" of a pixel instead of "degree of sharing" (Chawla, 2010; Dongre, 2016; Krishnapuram & Keller, 1993). Having an advantage of flexible hyperplane constraint that can handle noise and outliers, it produces coincident clusters and needs proper initialization values (Krishnapuram & Keller, 1996). Various improvements and modifications have been proposed such as Improved Possibilistic c-Means (IPCM) algorithm by Zhang & Leung (2004), Enhanced Possibilistic c-Means (EPCM) by Xie et al. (2008) earlier that proved to work better than both Fuzzy c-Means and Possibilistic c-Means.

Miyamoto et al. (2008) describe various fuzzy clustering algorithms with their modifications based on kernels, similarity and dissimilarity measures. Similarity and dissimilarity measures were incorporated as distance measure in fuzzy classifiers as proposed by Lee et al. (2009). Zwick et al. (1987) demonstrated the measure of similarity among fuzzy concepts. Goshtasby (2012) shows different similarity and dissimilarity measures and their properties. In addition, various different distance, as well as similarity measures such as Euclidean, Mahalanobis, Cosine, Chessboard, were studied with Fuzzy c-Means by Mukhopadhaya (2016), with Possibilistic c-Means, Improved Possibilistic c-Means, Modified Possibilistic c-Mean by Dongre (2016) and with Noise Classifier by Panda (2017).

A discriminatory measure as defined by Kumar et al. (2011) includes hybrid approaches for Spectral Information Divergence (SID) with Spectral Angle Measure (SAM) and Spectral Correlation Angle (SCA) with sin and tan functions that make two similar and dissimilar spectra more similar and dissimilar respectively. Hybrid measures such as Jefferies-Matusita and Spectral Angle Measure has been successful in classifying mangroves using hyperspectral imagery (Padma & Sanjeevi, 2014).

In addition, the semi-supervised approach has been applied in hard and soft classification applications with soft, sparse multimodal regression model (Li et al., 2011). They have been extensively used along with active learning approaches, as shown by Drews et al. (2013). Bouchachia & Pedrycz (2006) and Shanbehzadeh (2013) have shown various improvements of fuzzy classifiers in semi-supervised learning approach with the integration of semi-supervised approach in the objective function of a classifier. Modification of objective function for possibilistic classifiers with semi-supervised approach has also been studied (Liu & Wu, 2013).

3. METHODS AND CLASSIFICATION APPROACHES

This chapter describes the methods and approaches used in this research to achieve the desired objectives. **Section 3.1** presents the conceptual review of various spectral similarity measures. **Section 3.2** discusses the Possibilistic *c* Means classifier. **Section 3.3** provides a review of the algorithms used in this research for Possibilistic *c* Means classifier. **Section 3.4** discusses the semi-supervised approach, and the roles of hybrid measures for a Possibilistic *c* Means classifier. It also discusses the application of mean shift for improving the classification results. This section also discusses the accuracy assessment techniques used in this research. **Section 3.5** discusses the comparison techniques.

3.1. Spectral similarity measures

3.1.1. Euclidean Distance

Euclidean distance (ED) is a well-known spectral similarity measure that calculates the similarity between two spectral signatures by calculating the square root of the squared difference between them (van der Meer, 2006).

$$ED(s,s') = \|s - s'\| = \sqrt{\sum_{j=1}^{L}\left(s_j - s_j'\right)^2} \tag{3.1}$$

In equation (3.1), s and s' are the spectral signature vector of two pixel vectors X and X' and L is a set of wavelengths corresponding to spectral band channel.

3.1.2. Spectral Angle Measure

Spectral Angle Measure (SAM) is a widely used hyperspectral similarity measure that calculates the similarity between two spectral signatures by measuring the angle between them with dimensionality equals to number of spectral bands (Kruse et al., 1993). It calculates angle between two spectral signatures s and s' for pixel vectors X and X'.

$$SAM(s,s') = \cos^{-1}\left(\frac{\sum_{j=1}^{L} s_j s_j'}{\left(\sqrt{\sum_{j=1}^{L} {s_j}^2}\sqrt{\sum_{j=1}^{L} {s_j'}^2}\right)}\right) \tag{3.2}$$

In equation (3.2), L is a set of wavelengths corresponding to spectral band channel. It has a lower bound of 0 and values greater than 1, with a maximum of 1.57, and is insensitive to Albedo and illumination effects.

3.1.3. Spectral Correlation Angle

Spectral Correlation Angle (SCA) as described by van der Meer & Bakker (1997) calculates the Pearsonian Correlation coefficient between the spectral signatures of two-pixel vectors.

$$r(s,s') = \frac{L\sum_{j=1}^{L} s_j s_j' - \sum_{j=1}^{L} s_j \sum_{j=1}^{L} s_j'}{\sqrt{\left[L\sum_{j=1}^{L} s_j^2 - \left(\sum_{j=1}^{L} s_j\right)^2\right]\left[L\sum_{j=1}^{L} s_j'^2 - \left(\sum_{j=1}^{L} s_j'\right)^2\right]}} \tag{3.3a}$$

In equation (3.3a), s and s' are the spectral signature vector of two pixel vectors X and X' and L is a set of wavelengths corresponding to spectral band channel.

The coefficient reflects the extent of the linear relationship between two spectra and has a value in between -1 to 1.

$$SCA(s, s') = \cos^{-1}\left(\frac{r(s,s')+1}{2}\right) \tag{3.3b}$$

It is converted into an angle (in radians) for comparison to other measures and is described by equation (3.3b).

3.1.4. Spectral Information Divergence

Based on the concept of statistical divergence (Fano, 1961) and Kullback–Leibler divergence (Kullback, 1997), Spectral Information Divergence is a stochastic spectral similarity measure in which spectrum of a pixel vector is modeled as a probability distribution. The spectral similarity is defined by measuring the discrepancy between the probability distribution of two pixels spectra (Chang, 2000). It relies on the theory and concept of Spectral Information Measure (SIM) which is an information-theoretic measure (Chang, 2000). As per SIM, the spectral signature histogram is used as the desired probability distribution after normalizing it to unity and each pixel is treated as a random variable. It models spectral band to band variability.

The concept of Spectral Information Measure is explored to understand the derived Spectral Information Divergence. Let X be a pixel vector defined as $(X_1, X_2, \dots., X_L)^T$, where X_i is a pixel in Band B_j, acquired at a wavelength ω_j, where $\{\omega_j\}_{j=1,\dots L}$, where j represents L wavelengths and i represents N pixels. Let $S = (s_1, s_2, \dots., s_L)^T$ be the spectral signature of a pixel vector X, represented as the spectrum of reflectance or radiances.

Assuming, the components of S (reflectance and radiances) as non-negative, we define probability measure in probability space $(\Omega, \Sigma, \mathrm{A})$, where Ω is sample space and defined as $(\omega_1, \omega_2, \dots., \omega_L)$, Σ is an event space which is a set of all subsets of Ω and A is a probability measure whose values lies in range $[0,1]$.

$$A_j = \frac{s_j}{\sum_{l=1}^{L} s_l} \tag{3.4}$$

In equation (3.4), $A = (A_1, A_2, \dots., A_L)^T$ is the desired probability vector for pixel vector X.

$$I_j(X) = -\log(A_j) \tag{3.5}$$

$$H(X) = -\sum_{j=1}^{L} A_j \log(A_j) = \sum_{j=1}^{L} A_j I_j(X) \tag{3.6}$$

The equation (3.5) and equation (3.6) shows the self-information $I_j(X)$ for a particular band B_j and the entropy $H(X)$ of each pixel X.

$$I_j(X') = -\log\left(A'_j\right) \tag{3.7}$$

$$D(X' \parallel X) = \sum_{j=1}^{L} A'_j D_j(X' \parallel X) = \sum_{j=1}^{L} A'_j \left(I_j(X') - I_j(X)\right) = \sum_{j=1}^{L} A'_j \log\left(\frac{A'_j}{A_j}\right) \tag{3.8}$$

Spectral Information Divergence (SID) calculates spectral correlation between two pixel vectors. Let X' be another pixel vector defined as $(X'_1, X'_2, \dots, X'_L)^T$ with spectral the signature $S' = (s'_1, s'_2, \dots s'_L)^T$ and probability vector $A' = (A'_1, A'_2, \dots A'_L)^T$.

$$D(X \parallel X') = \sum_{j=1}^{L} A_j D_j(X \parallel X') = \sum_{j=1}^{L} A_j \left(I_j(X) - I_j(X')\right) = \sum_{j=1}^{L} A_j \log\left(\frac{A_j}{A'_j}\right) \tag{3.9}$$

The equation (3.7) shows the self-information and equation (3.8) shows the relative entropy of X' with respect to X. $D(X' \parallel X)$ is also known as cross-entropy, directed divergence or Kullback-Leibler information measure. The relative entropy of X with respect to X' has been described in equation (3.9)

$$SID(X, X') = D(X \parallel X') + D(X' \parallel X) \tag{3.10}$$

Spectral Information Divergence (SID) as given by Chang (2000) is defined as shown in equation (3.10), where, $D(X \parallel X') \neq D(X' \parallel X)$

Lesser the value of SID, more similar the spectral signatures are to each other.

3.1.5. Hybrid Spectral Similarity Measures

3.1.5.1. Spectral Information Divergence-Spectral Angle Measure

The hybrid approach for Spectral Information Divergence (SID) with Spectral Angle Measure (SAM) has been proposed by Du et al. (2004) as a hyperspectral similarity measure that performs better than ED, SAM and SID if, applied individually. The hybrid approach of Spectral Angle Measure as deterministic approach and Spectral Information Divergence as the stochastic approach multiplies the spectral ability of two signatures that make two similar and dissimilar spectra more similar and dissimilar respectively. The discriminatory ability of the hybrid approach has been proved better than the individuals as the hybrid approach combines the strengths of the two measures (Chang, 2000; Du et al., 2004; Kumar et al., 2011).

Let $S = (S_1, S_2, \dots . S_L)^T$ and $S' = (S'_1, S'_2, \dots . S'_L)^T$ be the spectral signature of two pixel vectors $X = (X_1, X_2, \dots, X_L)^T$ and $X' = (X'_1, X'_2, \dots, X'_L)^T$.

$$SID(SAM_{tan}) = SID(s, s') \times \tan\left(SAM(s, s')\right) \tag{3.11}$$

$$SID(SAM_{sin}) = SID(s, s') \times \sin\big(SAM(s, s')\big) \tag{3.12}$$

Equation (3.11) and equation (3.12) shows the hybrid function for Spectral Information Divergence and Spectral Angle Measure with sin and tan trigonometric functions (Chang, 2000). Here, the cosine function is not used, since it calculates the projection of one spectral signature along the other instead of orthogonal distance, thereby reducing the discriminability.

3.1.5.2. Hybrid Spectral Information Divergence-Spectral Correlation Angle

Similar to the hybrid of Spectral Information Divergence and Spectral Angle Measure, Kumar et al. (2011), proposed a hybrid approach of the stochastic Spectral Information Divergence with the deterministic Spectral Correlation Angle which has the better discriminatory capability for hyperspectral signatures. It minimizes the shading effect as caused when using Spectral Angle Measure and eliminates the negative correlation.

Let $S = (S_1, S_2, \ldots . S_L)^T$ and $S' = (S_1', S_2', \ldots . S_L')^T$ be the spectral signature of two pixel vectors $X = (X_1, X_2, \ldots, X_L)^T$ and $X' = (X_1', X_2', \ldots, X_L')^T$.

$$SID(SCA_{tan}) = SID(s, s') \times \tan\big(SCA(s, s')\big) \tag{3.13}$$

$$SID(SCA_{sin}) = SID(s, s') \times \sin\big(SCA(s, s')\big) \tag{3.14}$$

Equation (3.13) and equation (3.14) shows the hybrid function for Spectral Information Divergence and Spectral Correlation Angle with sin and tan trigonometric functions (Chang, 2000). Here also, the cosine function is not used because of the similar reasons as in $SID - SAM$ hybrid approach.

3.1.6. Spectral Discriminatory Power

To evaluate the effectiveness of different spectral similarity measures among each other, Chang (2000), defines a metric known as relative Spectral Discriminatory Power (SDP). It discriminates the spectral signature of one pixel from another relative to a reference pixel.

Let p and q be the spectral signature of any pair of pixel vectors, D be the spectral signature of a reference pixel d and m is any given spectral similarity measure.

$$SDP^m(p, q, D) = \max_{m>0}\left\{\frac{m(p, D)}{m(q, D)}, \frac{m(q, D)}{m(p, D)}\right\} \tag{3.15}$$

Higher the $SDP^m(p, q, D)$ as described in equation (3.15), the better is the discriminatory power of m with respect to a reference. It is symmetric and bounded below by $1\ i.e. \geq 1$ with equality iff p and q is equal.

3.2. Possibilistic *c* Means classifier

To generate the absolute value of memberships which represents typicality or compatibility with an elastic constraint, the modification to the initial Fuzzy *c* Means algorithm (FCM), equation (3.16a) and equation

(3.16b), have been put forth by Krishnapuram & Keller (1993) known as Possibilistic *c* Means algorithm (PCM).

$$J_{FCM}(v,\mu) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}d_{ij}^{2} \tag{3.16a}$$

Where,

$$\sum\nolimits_{i=1}^{C}\mu_{ij} = 1 \quad \forall j \tag{3.16b}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{3.17a}$$

$$v_i = \frac{\sum_{j=1}^{N}(\mu_{ij})^m x_j}{\sum_{j=1}^{N}(\mu_{ij})^m} \tag{3.17b}$$

Here, μ is membership value of pixel j in cluster i
N is the total number of pixels in an image,
x is the data present in an image, $\{x_1, x_2, \dots . x_N\}$,
d is the distance from pixel j to cluster i
c is the number of clusters
m is the weight parameter that controls the degree of fuzziness, and
v is the center for cluster i

$$\mu_{ij} \in [0,1] \quad \forall\, i,j \tag{3.18}$$

$$0 < \sum_{j=1}^{N}\mu_{ij} < N \quad \forall\, i \tag{3.19}$$

The membership values for the Fuzzy *c* Means algorithm, as described in equation (3.17a), follows the criterion from equation (3.18) and equation (3.19).

$$\mu_{ij} = \frac{1}{1 + d_{ij}} \tag{3.20}$$

$$J_{PCM}(v,\mu) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}d_{ij}^{2} + \sum_{i=1}^{C}\eta_i\sum_{j=1}^{N}(1-\mu_{ij})^m \tag{3.21}$$

A model for membership function as described by Zimmermann & Zysno (1985) conceptualizes a framework of modeling vague clusters with centers as described in equation (3.17b). Following the concept of the model, class memberships of every pixel are defined as shown in equation (3.20) and the objective function of Possibilistic *c* Means algorithm (PCM) is formulated as equation (3.21). In order to avoid the trivial solution, when all memberships are zero, a regularization term or a penalty term is added to the objective function of FCM to have the high memberships of representative feature points.

Also, $max(\mu_{ij}) > 0 \;\; \forall\, j$ relaxing the hyperplane constraint on membership values as compared to FCM in equation (3.16b).

$$\eta_i = \mathrm{K}\frac{\sum_{j=1}^{N}\mu_{ij}^{m} d_{ij}^{2}}{\sum_{j=1}^{N}\mu_{ij}^{m}} \tag{3.22}$$

$$\mu_{ij} = \frac{1}{1+\left(\frac{d_{ij}^{2}}{\eta_i}\right)^{\frac{1}{m-1}}} \tag{3.23}$$

Here η is a positive number known as a resolution parameter or bandwidth of the distribution of possibilistic membership function for each cluster. The distance at which the membership value of a point in a cluster becomes 0.5, is determined by the value of η and is related with the overall shape of a cluster. It is estimated as shown in equation (3.22). The membership values are dependent on the estimation of this parameter as shown in equation (3.23).

Advantages:

Outliers and Noisy data can be handled with Possibilistic *c* Means algorithm, as it assigns lower memberships to the unrepresentative data in all the clusters and in case of supervised classification, PCM works better than FCM for untrained classes (Krishnapuram & Keller, 1993; Wu & Zhou, 2006).

Disadvantages:

PCM is highly sensitive towards initialization values and produces coincident clusters (Zhang & Leung, 2004).

3.3. Algorithmic review for supervised Possibilistic *c* Means classifier

The basic steps for a supervised Possibilistic *c* Means classifier are as follows:

1. Collect the training data with the desired number of classes.
2. Initialize a value of m, the degree of fuzziness.
3. Calculate mean from the training data and initial class membership values.
4. Calculate η and final memberships of every pixel to every class.

The Possibilistic *c* Means classifier can follow either of the ways as described in Figure 3.1(a) and Figure 3.1(b) for estimating the parameter values.

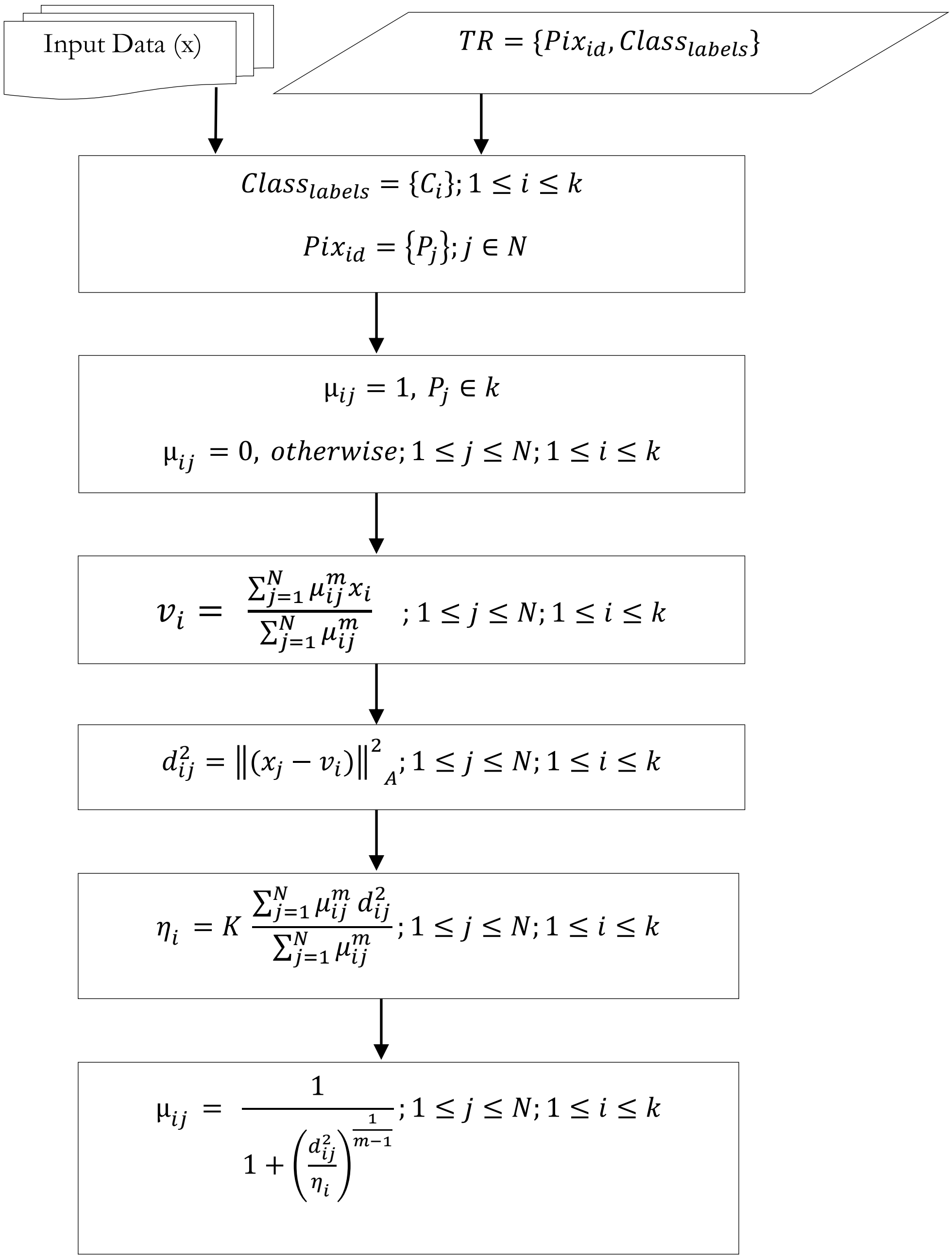
Case 1: When membership values are initialized first.

Initial membership values: obtained from labeled training data

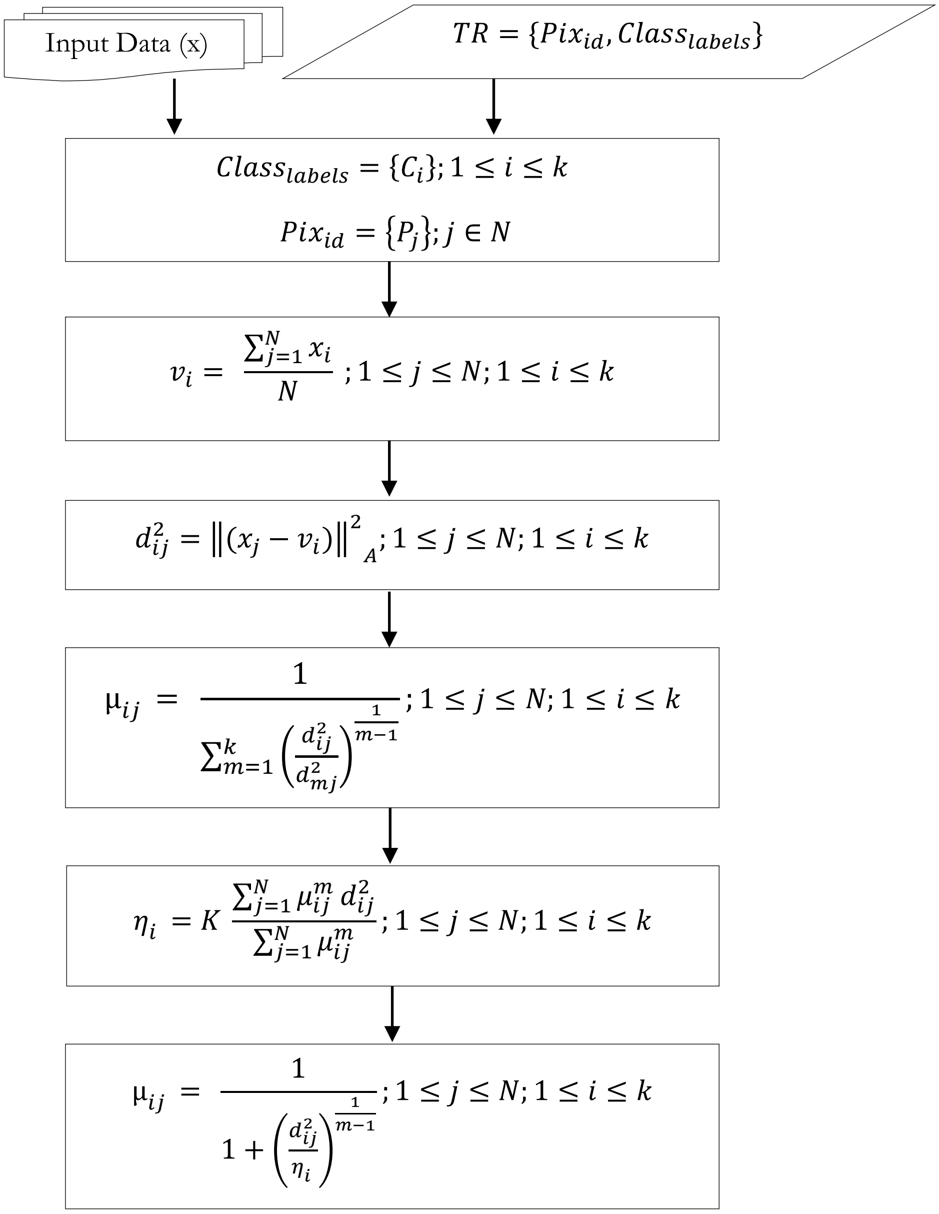
η : Depends only on labeled training data (initial memberships)

The approach has been shown in Figure 3.1(a), and the steps are described as follows:

1. Initialize the class memberships from the labeled training data.
2. Calculate Mean and distance of each pixel to this mean.
3. Calculate η using the initialized class memberships.
4. Calculate final membership values of every pixel to every class.



(a)



(b)

Figure 3.1. Supervised Possibilistic c means approach (a) Case 1 (b) Case 2, where, $A = I$ (Identity Matrix) = Euclidean Norm, $K = 1$, TR is the labeled training data which consists of Pixel Id (P_j) and class labels $(Class_{labels})$, k is the number of clusters, x is the data present in an image, N is the number of pixels in the image, μ_{ij} is the membership value of pixel j in class i, η_i is the bandwidth or resolution parameter, d_{ij}^2 is the squared distance between pixel j and mean value (v) of cluster i, m is the degree of fuzziness with value = 2.

Case 2: When mean value is estimated first.

Initial membership values: obtained from all pixels

η : Depends on all pixel values (estimated initial memberships)

The approach has been shown in Figure 3.1(b), and the steps are described as follows:

STEP 1: Initialize the class mean values from the labeled training data.

STEP 2: Calculate the distance of each pixel to this mean.

STEP 3: Calculate initial membership values using the calculated distance.

STEP 4: Calculate η using the initialized class memberships.

STEP 5: Calculate final membership values of every pixel to every class.

Note: The major difference in both the approaches lies in the estimation of initial class membership values μ and η.

3.4. Semi-supervised Possibilistic *c* Means classifier

3.4.1. The semi-supervised approach in classification

The semi-supervised Possibilistic *c* Means clustering algorithm uses a very few labeled training samples and a large number of unlabeled samples for the calculation of parameters such as mean, memberships and η. The following diagram as represented in Figure 3.2 depicts the generalized semi-supervised approach to classification using a Possibilistic *c* Means classifier.

The semi-supervised Possibilistic *c* Means classifier can follow any of the cases as described in Section 3.3, but in this type of learning approach, the signature data contains both labeled as well as unlabeled data in which labeled data are very few in number.

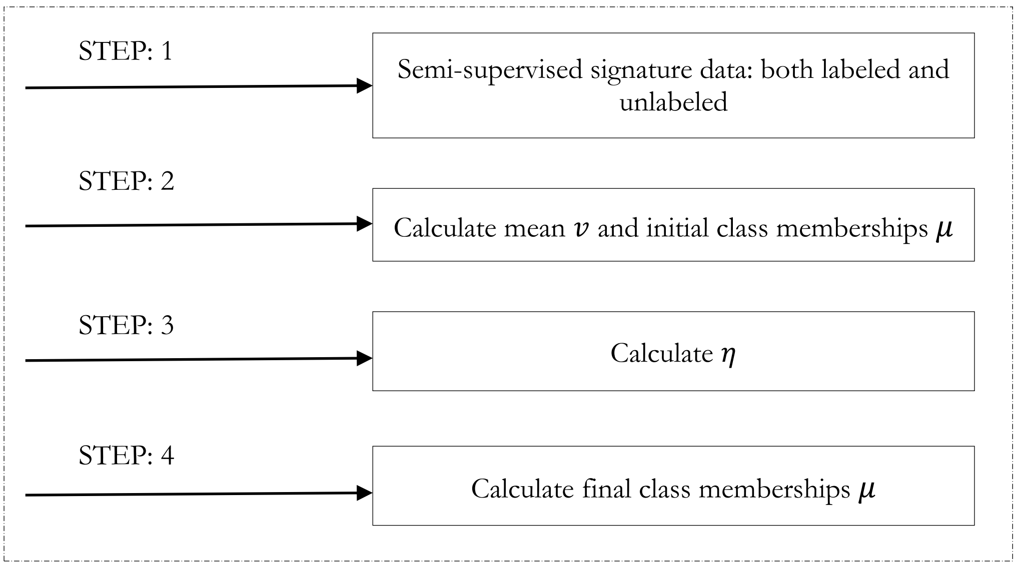


Figure 3.2. Generalized procedure for semi-supervised Possibilistic c means

3.4.2. Semi-supervised signature/training data

The semi-supervised data consists of a small number of labeled training samples and a large number of unlabeled data.

Labeled Data:

The labeled data collected could be limited to a single pixel and could be extended to the $10N$ (where N is the number of bands in a remote sensing image) samples, so as to calculate a complete co-variance matrix as in case of a Maximum Likelihood Classifier (Richards & Jia, 2013). A single pixel may not capture the variability of a class at all, while the collection of huge labeled samples is always time-consuming. In case of a fuzzy classifier, the estimation of full covariance matrix is not required, therefore, the need of labeled samples further reduces. Therefore, only a few labeled data for training the semi-supervised Possibilistic fuzzy classifier are collected. In the literature, several researches have been done based on different number of labeled training dataset, which is usually very little in compared to the unlabeled data, in between ten to forty percent (Triguero et al., 2015). This research uses ten percent of the samples of fully supervised approach, considering $10N$ to be the maximum limit. Therefore, the labeled training data collected as a part of semi-supervised approach consists of eight to ten samples per class.

At training stage, the pixels collected should be as pure as possible and should be representative of a particular class. They are therefore referenced to the available ground data information. Classes can be multi-modal, and this could lead to the collection of poor labeled training data. The labeled data should be refined, and outliers must be removed. They are assessed by using Mean Shift Algorithm (Fukunaga & Hostetler, 1975).

3.4.2.1. Mean Shift Algorithm

The data in the remote sensing images can be considered as a sample of a probability density function. In order to find the maxima or mode of a density function, a non-parametric technique, known as Mean Shift algorithm is used for finding the maxima or modes (typically a local maximum) of a density function in feature space (Fukunaga & Hostetler, 1975). This estimate is iterative and depends upon an initial estimate of mean chosen. The weights from nearby points are calculated using a kernel for re-estimation of the initial point or mean. A Gaussian Kernel is typically chosen due to its several advantages over others (Cheng, 1995; Comaniciu et al., 2002; Fukunaga & Hostetler, 1975). This type of kernel estimation requires a scale parameter or a bandwidth which represents the shape of the clusters.

$$f(x) = \frac{\sum_{x_j \in NB(x)} \exp\left(-\frac{1}{2}\left\|(x - x_j)/\sigma\right\|^2\right).\, x_j}{\sum_{x_j \in NB(x)} \exp\left(-\frac{1}{2}\left\|(x - x_j)/\sigma\right\|^2\right)} \tag{3.24}$$

The modes with Gaussian kernel can be determined using equation (3.24), where $NB(x)$ is the neighborhood of x for which the value after applying kernel is non-zero with x_j neighboring points. The neighborhood is defined by applying an appropriate threshold of a suitable window size using Euclidean distance measure for the mean value x. The Mean Shift as defined by Fukunaga & Hostetler (1975) is given by $f(x) - x$. The Mean Shift Algorithm repeats the whole estimation by replacing x by $f(x)$ until convergence of $f(x)$.

<u>Choice of standard deviation in Mean Shift (Gaussian kernel):</u>

Bandwidth parameter: h or σ (sigma/ Standard deviation) in case of Gaussian kernel, can be related to η (resolution or bandwidth parameter in Possibilistic *c* Means classifier), which also denotes variance and can be used in Mean Shift algorithm for mode estimation. η represents shape of the cluster.

For applying Mean Shift algorithm, a reliable estimate of variance is needed. With just a little labeled training data, a reliable estimate of variance or standard deviation could not be obtained due to the following reasons:

- The labeled data is very little in number and may not capture the entire class variances.
- Also, the parameters such as mean, η and class membership values μ obtained from the semi-supervised labeled data may or may not be correct.

Therefore, an iterative approach in a Possibilistic *c* Means classifier for calculating the parameters (η and class membership values μ) could help to get a reliable estimate described as follows:

(S1) Apply Case 1 or Case 2 as described in Section 3.3 at steps $i, i = 0, \dots lmax$.

(S2) Compute an updated η^{i+1} using equation (3.22).

(S3) Compute an updated μ^{i+1} using equation (3.23).

(S4) Compare η^{i+1} to η^{i}. If $\|\eta^{i+1} - \eta^{i}\| < \epsilon$, Stop. Otherwise return to (S2).

Here, $lmax$ is a very large number taken for iterations and ϵ is chosen to be a very small value for convergence. In this research, iterations tend to converge within $lmax < 1000$ and $\epsilon = 1.0e - 7$.

The value for $\sqrt{\eta}$ estimate is calculated using Possibilistic c Means classifier with Euclidean as a distance measure and is studied with Case 1 and Case 2 as described in Section 3.3. Due to the availability of a few labeled training samples, the iteration are applied. The possible values for the bandwidth parameter per class are shown in Table 3.1.

Table 3.1. Possible values for bandwidth parameter when Possibilistic c Means algorithm is applied using Case 1 and Case 2 and studied with iterations and without iterations.

CLASSES	STANDARD DEVIATION (with labeled training data) 10 pixels per class	$\sqrt{\eta}$ ESTIMATE			
		CASE 1		CASE 2	
		Without Iteration	With Iteration	Without Iteration	With Iteration
Wheat Crop	3822.954	3900.291	7119.043	2334.545	7119.043
Riverine Sand	3005.318	3749.129	4196.145	1897.665	4196.145
Dense Forest	1849.632	1754.757	829.737	1351.201	829.737
Grass Land	2375.56	3376.203	2834.927	2289.842	2834.927
Water	1888.927	1548.571	6436.770	1771.217	6436.770
Eucalyptus	1752.290	625.814	2756.503	1801.006	2756.503

It can be seen that the η from both the cases after iterations (Table 3.1) are almost equal which makes sense as if it were a case of completely supervised procedure (with larger training data), in which both the cases would tend to give similar results without iterations. Therefore, the values chosen for this research to be used as a bandwidth parameter per class in Mean Shift algorithm are shown in the last column of Table 3.1.

Outlier removal using Mean Shift algorithm:
The labeled training data can be assessed by using Mean Shift algorithm, by finding the modes of the classes, taking each labeled data as a starting point. Those points which end up in completely different modes (far from the cluster) are the prominent outliers. These outliers are removed before further processing. Also, if modes are in a cluster but very far from each other could represent a multi-modal class.

Mean Shift algorithm for shifting the 'mean' parameter:
The semi-supervised approach has a very few number of labeled data, which could lead to a poor estimate of the parameters as the training data might not be a complete representation of a particular class due to huge inter-class variance in multi-spectral imagery. For this, Mean Shift algorithm could be further utilized to obtain a mean in a higher density region as clusters for an image can be categorized by separating the regions of different densities.

Unlabeled Data:

The unlabeled data for semi-supervised approach may be obtained by using hybrid spectral similarity measures by following the continuity assumption and cluster assumptions. The spectral similarity measures such as Spectral Information Divergence with Spectral Angle Measure and Spectral Information Divergence with Spectral Correlation Angle are used to measure the similarity between the training samples and included as a part of unlabeled data for further classification.

Refinement of unlabeled data using threshold: To avoid mixed pixels or pixels from different classes to be a part of training data, careful thresholds are applied both in feature space as well as geometric space while collecting the unlabeled data.

Table 3.2. Thresholds in feature space for classes using different spectral similarity measures for region growing

Identified Classes	Spectral Similarity Measures		
	SID-SAM(tan/sin)	SID-SCA(tan/sin)	Euclidean
Dense Forest	7.9×10^{-8}	2.7×10^{-7}	90
Wheat Crop	3.0×10^{-7}	4.0×10^{-7}	215
Riverine Sand	3.0×10^{-7}	7.5×10^{-7}	260
Water	6.0×10^{-7}	1.7×10^{-6}	170
Eucalyptus	2.2×10^{-7}	7.0×10^{-7}	130
Grass land	5.0×10^{-7}	1.5×10^{-6}	260

The following steps are taken:

1. The labeled training data after removing outliers are taken.
2. Mean is calculated using the labeled data and shifted using Mean Shift algorithm.
3. Hybrid spectral similarity measures are applied as a part of region growing in feature space, which calculates the similarity between pixel values to the shifted mean value (as described in Subsection 3.1.5).
4. The similar pixels are chosen by applying an appropriate threshold in feature space (Table 3.2), by following the cluster assumption.
5. The geometric distance (Euclidean distance) is calculated for each of selected pixels with respect to each of the labeled training data.
6. The pixels are then chosen by applying an appropriate threshold in geometric space (Table 3.3). The pixels which are far from the labeled data are discarded, by following the continuity assumption for identifying the unlabeled data in semi-supervised learning.

Following the cluster assumption of semi-supervised learning approach, thresholds in feature space have been set up for approximately 70-80 samples per class as shown in Table 3.2. Table 3.3 presents the thresholds in geometric space (spatial domain) by following the continuity assumption of semi-supervised learning approach to remove any outliers or samples that are far apart spatially from the labeled samples. The thresholds are set from the minimum of the distance obtained between two pixels by applying similarity measure in feature and geometrical domain.

Table 3.3. Thresholds in geometric space for classes calculated after applying Euclidean distance in spatial domain when feature space thresholds are defined using different spectral similarity measures for region growing

Identified Classes	Euclidean Distance when feature space is calculated using		
	SID-SAM(tan/sin)	SID-SCA(tan/sin)	Euclidean
Dense Forest	1000	1000	540
Wheat Crop	800	1000	600
Riverine Sand	700	1000	700
Water	2000	1700	800
Eucalyptus	800	450	200
Grass land	140	110	91

3.4.3. Hybrid spectral similarity measures for semi-supervised Possibilistic fuzzy classifier:

Instead of Euclidean distance as a distance measure, the hybrid spectral similarity measures such as SID-SAM (tan/sin) and SID/SCA (tan/sin) have been used in the objective function of a Possibilistic *c* Means classifier. The approach applied has been shown in Figure 3.3. The algorithmic flowchart is similar to Section 3.3 with different distance-based hybrid spectral similarity measures as described in Section 3.1.5. For the Possibilistic *c* Means classifier with the semi-supervised approach, the mean is estimated from both labeled and unlabeled training data (collected using different spectral similarity measures, Section 3.1).

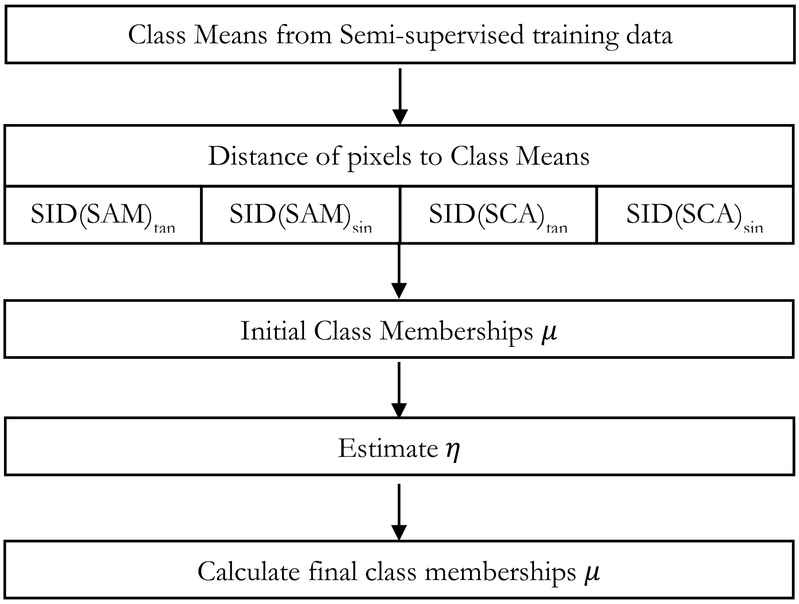


Figure 3.3. Semi-supervised Possibilistic *c* Means classifier with hybrid measures

3.4.4. Classification and Accuracy Assessment:

Classification is performed by using the semi-supervised Possibilistic *c* Means classifier using various hybrid spectral similarity measures, and their accuracies are accessed by using Root Mean Square Error and Fuzzy Error Matrix.

Reference Data Preparation:

The outputs obtained after classifying image is a soft output. To access accuracies, it should be cross-checked with the true value. Therefore, the reference should be as good as possible and should contain minimum errors. To compare the soft outputs, the reference data considered is also a soft output image. A higher resolution image of the Formosat-2 satellite is used as a reference. The finer resolution image is chosen because it defines the features present on the imagery in a much better way than a coarser resolution image. The reference image has been classified using Support Vector Machine (SVM) (Appendix A) in soft classification mode, to get the soft membership values (Mountrakis et al., 2011; Richards & Jia, 2013).

3.4.4.1. Root Mean Square Error

Root Mean Square Error (RMSE) discriminates between the membership values of the classified input image with the membership values of the classified reference image and is calculated using equation (3.25) and equation (3.26). The value for Root Mean Square that is close to zero but always greater than zero represents a lesser deviation in memberships of the two classified images (Kandpal, 2016).

Class wise RMSE:

$$RMSE(\mu_r, \mu_c) = \sqrt{\frac{\sum_{i=1}^{N}(\mu_{rji} - \mu_{cji})^2}{P \times Q}} \tag{3.25}$$

Overall RMSE:

$$RMSE(\mu_r, \mu_c) = \sqrt{\frac{\sum_{j=1}^{c}\sum_{i=1}^{N}(\mu_{rji} - \mu_{cji})^2}{P \times Q}} \tag{3.26}$$

Where N is the number of pixels, c is the number of classes, μ_r is the membership value of classified reference image and μ_c is the membership value of classified input image, μ_{rji} is the membership value of classified reference image for pixel i in class j, μ_{cji} is the membership value of classified input image for pixel i in class j, $P$and Q are the total number of features and feature vectors respectively in the data with dimension of the image as $P \times Q$.

3.4.4.2. Fuzzy Error Matrix

Fuzzy Error Matrix can be used to assess the accuracy of soft classification (Binaghi et al., 1999). It is similar to the concept of the traditional error matrix but is based on soft membership values, i.e., it can take non-negative real numbers instead of non-negative integer values (Stehman et al., 2007). The layout of the concept has been shown in Table 3.4.

Table 3.4. Fuzzy Error Matrix Layout

Soft Classification Input Data	Soft Reference Data				Total Grades
	Class 1	Class 2	…..	Class *c*	
Class 1	$M_{(1,1)}$	$M_{(1,2)}$		$M_{(1,c)}$	C_1
Class 2	$M_{(2,1)}$	$M_{(2,2)}$		$M_{(2,c)}$	C_2
……..	…….	…….		…....	…....
Class *c*	$M_{(c,1)}$	$M_{(c,2)}$		$M_{(c,c)}$	C_c
Total Grades	R_1	R_2		R_c	

Definitions: $M_{(p,q)}$ is the member of the Fuzzy Error Matrix in p class of classified input image and q class of classified reference image

$$M(p,q) = |C_p \cap R_q| = \sum_{x \in X} \min_{x>0}\left(\mu_{C_p}(x), \mu_{R_q}(x)\right) \tag{3.27}$$

In equation (3.27), R_q is a set of membership values for reference image in class q and C_p is a set of membership values for input image in class p, $\mu_{C_p}(x)$ is the membership of test sample x in class p of the input image and $\mu_{\boldsymbol{R_q}}(x)$ is the membership of test sample x in class q of the reference image, where X is the set of sample data. The Fuzzy Error matrix uses min operator to determine the elements of M.

$$OA = \frac{\sum_{j=1}^{c} M_{(j,j)}}{\sum_{j=1}^{c} R_j} \tag{3.28}$$

Overall Accuracy: This is obtained by dividing the sum of correct estimates of membership values (diagonal elements) of a matrix by the total grade of memberships in the reference data, as shown in equation (3.28).

$$PA_j = \frac{M_{(j,j)}}{R_j} \tag{3.29}$$

Producer's Accuracy: It is a measure of error of omission, i.e., how well an image can be classified if the value of the pixel is omitted from the class where it actually belongs. It is obtained by dividing the element of major diagonal of the fuzzy error matrix representing a class by the total grade of memberships in the classified reference image for a class j, as shown in equation (3.29).

$$UA_j = \frac{M_{(j,j)}}{C_j} \tag{3.30}$$

User's Accuracy: It is a measure of error of commission when the value of the pixel is included in a class where it does not belong. It indicates how well a pixel value represents the actual ground information. It is obtained by dividing the element of major diagonal of the fuzzy error matrix representing a class by the total grade of memberships in classified input data for a class j, as shown in equation (3.30).

3.4.5. Spectral Discriminability Power:

In order to check which spectral similarity measure is more effective in differentiating various classes among themselves for a multispectral image, Spectral Discriminatory Power (Section 3.1.6) has been calculated for three similarly looking classes using equation (3.15).

1) Dense Forest
2) Grass Land
3) Eucalyptus

Dense Forest Class is taken as a reference class for which Spectral Discriminatory Power has been calculated.

3.5. Comparison

Comparison of semi-supervised approach is done with supervised Approach. In case of supervised approach, training data is collected manually (70-80 samples) per class (based on the assumption of maximum training data to be $10N$), and the Possibilistic *c* Means classifier is used to classify the input image (Richards & Jia, 2013). The overall accuracies and Root Mean Square errors are compared with the results obtained using semi-supervised approach.

Comparison of hybrid spectral similarity measures is done with the Euclidean Measure. The Euclidean distance is used for increasing the training samples and also as a distance measure in Possibilistic *c* Means classifier. The overall accuracies and Root Mean Square errors are compared with the results obtained by using hybrid measures.

4. RESOURCES AND DATASET USED

This chapter presents the resources and the data used in this research to reach the desired objectives. **Section 4.1** provides the details of the study area and datasets used for achieving the objectives of this research. **Section 4.2** provides the details of the software and other resources used.

4.1. Study area and data used

The study area chosen for this research is located in the Haridwar district of state Uttarakhand in India. The study area has its boundaries connected to Dehradun from the north, Pauri Garhwal from the east and, some districts of Uttar Pradesh from south and west. The central latitude is 29.956° N and longitude is 78.170° E. It is very close to Dehradun and is very diverse in terms of land cover classes.

The main motivation for the choice of this area was that the satellite images of Landsat-8 and Formosat-2 and the field data with six identified classes were already available.

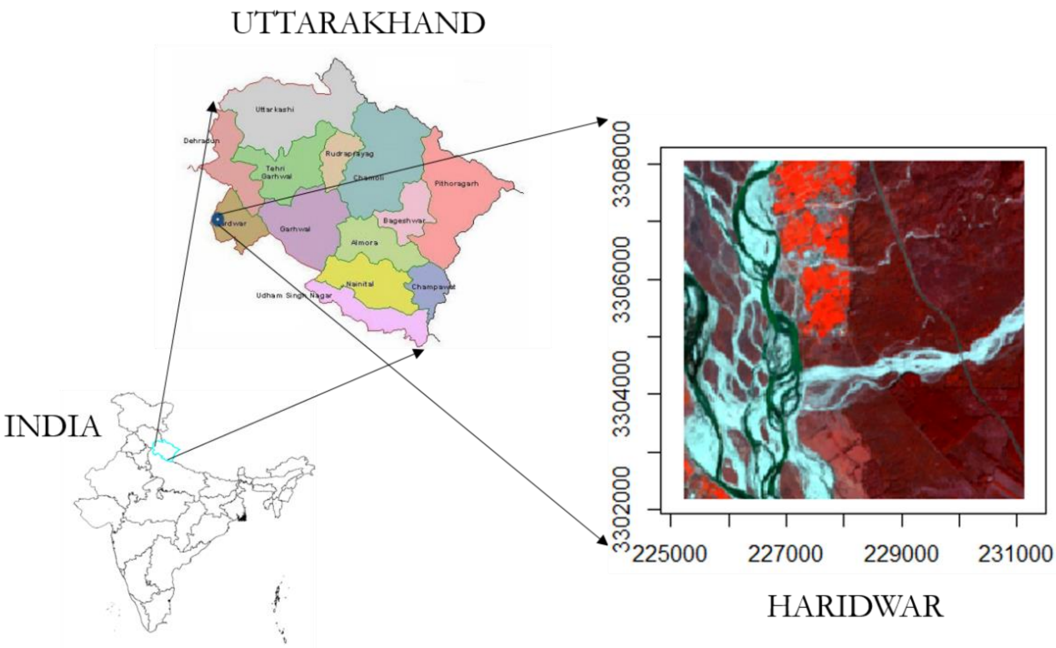


Figure 4.1. Study area used in this research, located in Haridwar district of state Uttarakhand, India.

Table 4.1. Specifications of data used in this research, Landsat-8 and Formosat-2 Satellite Images.

Specifications	LANDSAT-8	FORMOSAT-2
Spatial Resolution	30 m	8 m
Spectral Resolution	8 Bands	4 Bands
Date of Acquisition	Feb 12, 2015	Feb 21, 2015
Scene Size	170 km x 185 km	24 km x 24 km
Return Interval	Every 16 Days	Daily

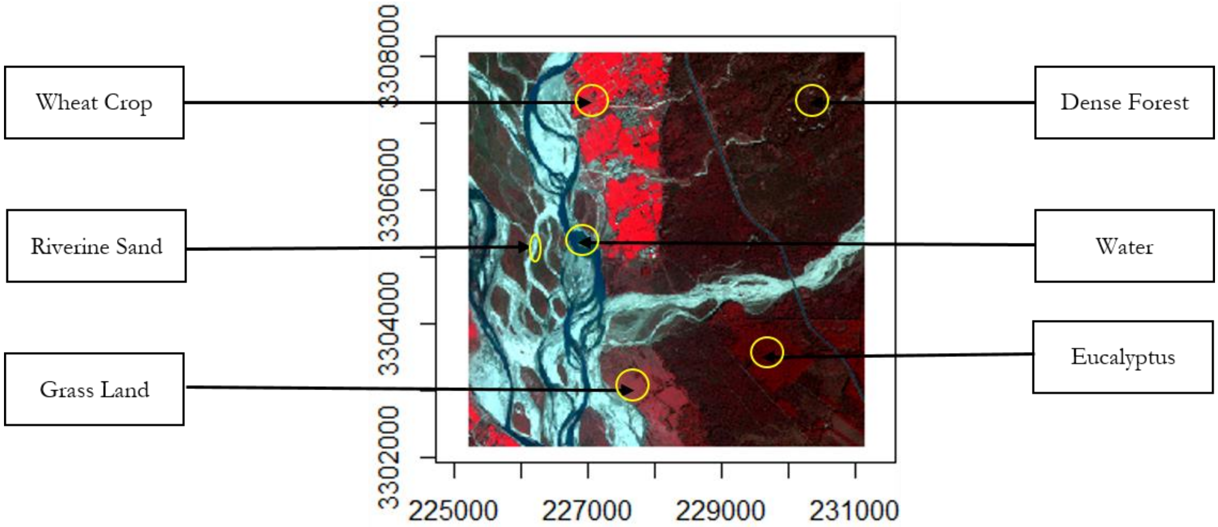


Figure 4.2. Identified classes on the satellite image (Formosat-2) as per the field data

Figure 4.1 shows the study area on the multispectral Landsat-8 satellite imagery and Figure 4.2 shows the identified classes on the satellite image (Formosat-2) as per the field data (Appendix B). The specification of data used in this research is summarized in Table 4.1.

4.2. Software Used

All the methods are applied using R and R Studio (R Development Core Team, 2017). The libraries used for processing are Raster (Hijmans, 2016), Rgdal (Rowlingson, 2017), Gstat (Graler et al., 2017; Pebesma, 2004), Rgl (Adler & Murdoch, 2017), Scatterplot3d (Ligges & Mächler, 2003), e1071 (Meyer et al., 2017). For a collection of training data, generation of maps and region of interest, ArcMap (Esri, 2016) is used, and the work was done on a standard processing workstation.

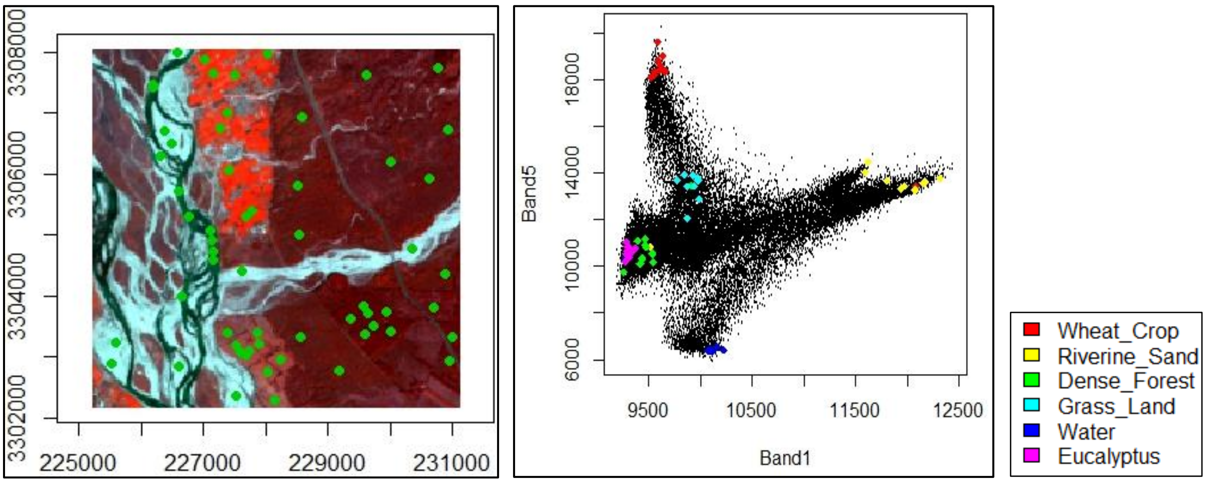
5. RESULTS AND ANALYSIS

This chapter presents the results and analysis done on the input data to achieve the objectives for this research by applying the methods discussed in **Chapter 3**. **Section 5.1** shows the usage of Mean Shift algorithm for a few labeled samples. **Section 5.2** presents the analysis on the choice of Possibilistic *c* Means algorithm. This section also demonstrates the results of Possibilistic *c* Means algorithm using an only limited number of labeled samples using different distance based similarity measures. **Section 5.3** presents the results of semi-supervised Possibilistic *c* Means classifier with hybrid spectral similarity measures. **Section 5.4** presents the results of supervised Possibilistic *c* Means classifier with different distance based similarity measures. **Section 5.5** shows the results of a Spectral Discriminatory Power for different spectral similarity measures. **Section 5.6** compares all the supervised and semi-supervised approaches with hybrid spectral similarity measures and also compares the results with conventional Euclidean distance measure.

5.1. Labeled data

This subsection presents the analysis of the collected training data, part of the semi-supervised labeled data as described in Subsection 3.4.2. They are analyzed and refined before further processing.

As can be seen in Figure 5.1 (a), there are 10 labeled samples collected manually as a part of training data (Section 3.4.2) for each of the classes identified at the time of field visit on the satellite image of Landsat-8 as described in Section 4.1. Even though the geographic locations and training data collection has been cross-referred with the available field data, there are few outliers in the training data which can be seen in Figure 5.1 (b), feature space plots. The spectral values of few labeled pixels denote other classes or are far from the originally identified class. These could be possible outliers due to mixed pixels or could be a case of multi-modal classes. These training samples are further tested for mode analysis using Mean Shift algorithm.



(a) (b) (c)

Figure 5.1. Supervised labeled data collected per class (10 training samples per class) based on the field data on Landsat-8 satellite image (a) Spatial location of the samples collected per class (b) Plot in feature space for labeled samples (d) Legend for feature space plot

5.1.1. Analysis of Mean Shift algorithm on the labeled data

This section describes the Mean Shift analysis on the labeled training data to estimate the correct mean and to analyze the accuracy of the labeled training samples as described in Section 3.4.2.1. The Mean Shift algorithm is applied for estimating the modes of classes if each training data is taken as a starting point. From Figure 5.2, we can see that the modes obtained by each of the labeled training data in each of the classes have values which are either overlapping in some cases, or very close to each other, or very far from the class cluster in the feature space.

From Figure 5.2, it could be seen that the training data in classes Wheat Crop and Water has an almost overlapping mode with one outlier each. The training data from class Eucalyptus results in an almost similar mode with no outlier. The training data from classes Riverine Sand, Dense Forest, and Grass Land has one outlier each with the modes which are comparatively close in the feature space. The overlapping values of modes in a class signify that the training data results in the same mode value after applying Mean Shift algorithm and that the labeled data is a complete representative of a class and the class is unimodal. The different mode values which are closer in feature space signifies multimodal class. The training data for a class that results in modes which are closer in feature space could be considered as a separate class in the spectral domain. The training data that results in closer modes or almost overlapping mode values are treated as a part of the class and labeled data in this research. The training data which results in mode values that are completely different or are far from the clustered values in feature space are treated as potential outliers and are removed from the labeled training data.

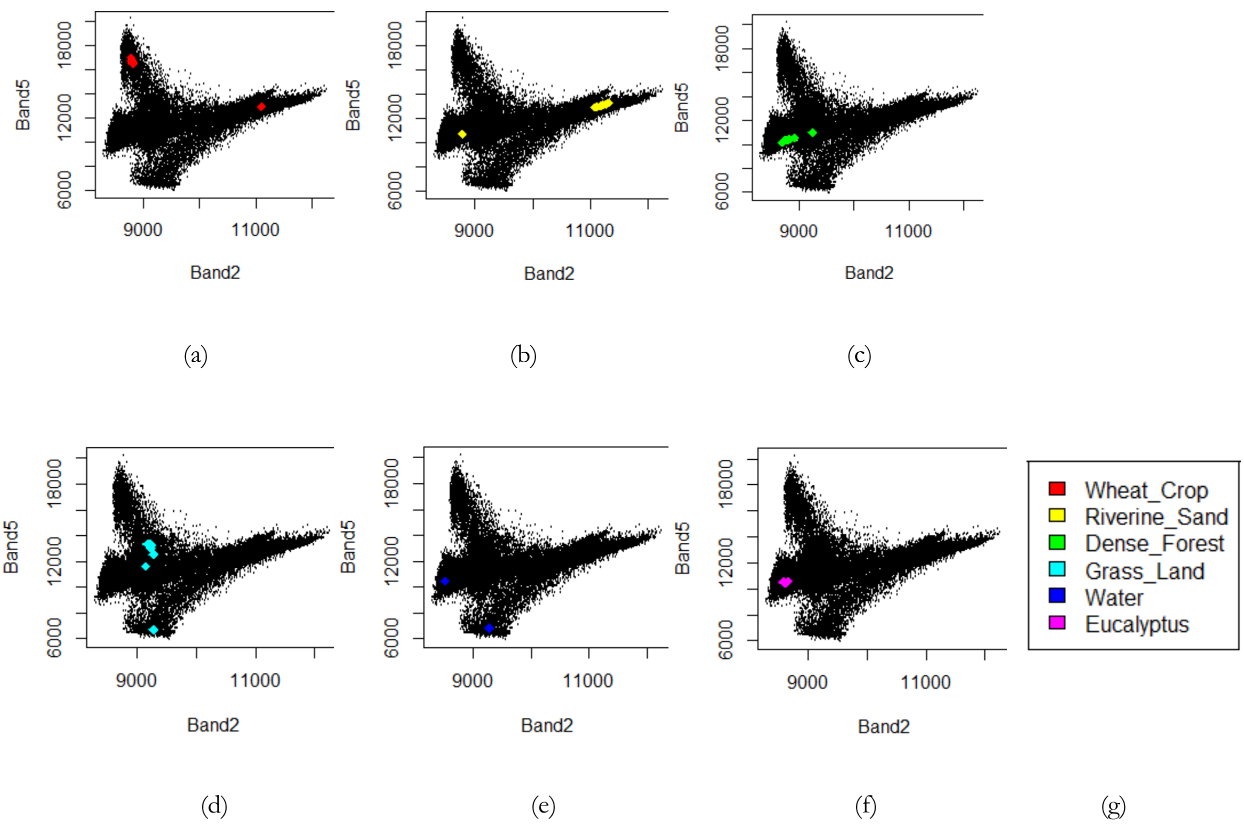


Figure 5.2. Modes estimated using Mean Shift algorithm for the labeled training data per class (a) Wheat Crop (b) Riverine Sand (c) Dense Forest (d) Grass Land (e) Water (f) Eucalyptus (g) Legend

Apart from refining the outliers in the labeled training samples, the Mean Shift algorithm is also used for shifting the mean obtained from the refined labeled samples to a higher density region as described in

Subsection 3.4.2.1. The difference in mean from labeled data (M) and shifted mean using Mean Shift algorithm (MS) per class is shown in Table 5.1 and is estimated using Euclidean distance as described in Section 3.1.1.

Table 5.1. Euclidean distance between mean from labeled data (M) and shifted mean using Mean Shift algorithm (MS) per class

Classes	Euclidean Distance between Mean (M) and Shifted Mean (MS)
Dense Forest	182.767
Wheat Crop	266.523
Riverine Sand	336.572
Water	184.285
Eucalyptus	596.743
Grass land	254.253

The difference in the mean values is highest for Eucalyptus class and lowest for Dense Forest. The higher shift in the Eucalyptus class may be related with the fact that locating pure training samples for Eucalyptus class manually is comparatively difficult as compared to rest of the classes on Landsat -8 satellite imagery, due to the similarity in its spectral values with the neighboring Dense Forest class. In contrast to Eucalyptus class, the collection of pure labeled training data is easiest for the Dense Forest class due to its larger area and distinctive spectral values.

5.2. Possibilistic *c* Means algorithm (with labeled samples)

This section describes the analysis of the algorithm chosen for applying Possibilistic *c* Means classification as discussed in two cases and described in Section 3.2. The choice of one of the case becomes important in case of a few labeled training data, as the initial membership values μ and estimate of η is highly dependent upon the number and quality of training data. In case of semi-supervised approach, quality of the training data could not be completely ensured as compared to completely supervised training data, therefore both the cases for Possibilistic *c* Means classifier have been checked initially with a little number of refined labeled training samples with shifted mean and Euclidean distance as a distance based similarity measure in a Possibilistic *c* Means classifier.

The global Root Mean Square Error is calculated when a Possibilistic *c* Means classifier is applied to little refined labeled samples with the shifted mean for both the cases. The global RMSE value for Case 1 is 0. 560 and that for Case 2 is 0.202. The value for Case 2 is lower than Case 1 which could be due to the reason that the value of initial class memberships μ are derived differently for the two cases and are highly dependent upon the number of labeled data. Since the labeled training data is very small in size, Case 1 is more sensitive towards the estimation of η (variance or the bandwidth parameter or resolution parameter) as η is calculated based on the initial μ. Therefore, Case 2 is preferred for further processing with degree of fuzziness, m=2.

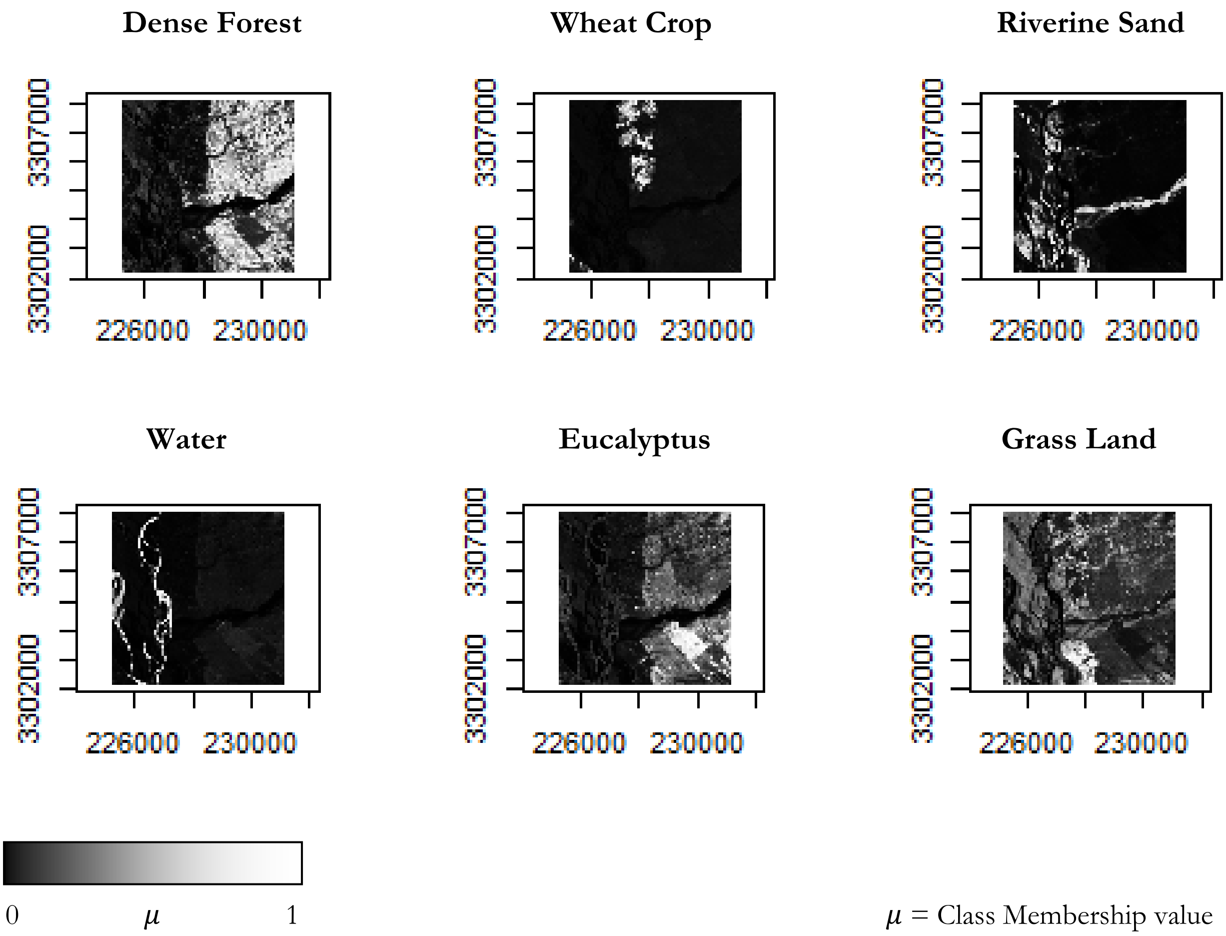
5.2.1. Result of Possibilistic *c* Means algorithm for labeled training data with shifted mean using different distance measures

This subsection presents the results obtained by applying a Possibilistic *c* Means classifier in a supervised mode with only a few available labeled samples. The mean is calculated from the labeled data and shifted using Mean Shift algorithm as presented in Subsection 3.4.2.1. The Possibilistic c Means algorithm is applied as described in Section 3.2 with different distance based similarity measures as defined in Section 3.1, following Case 2 from Section 3.3. The accuracies have been assessed as shown in Table 5.2 by using global Root Mean Square Estimation and overall accuracy of Fuzzy Error Matrix as described in Subsection 3.4.4.2.

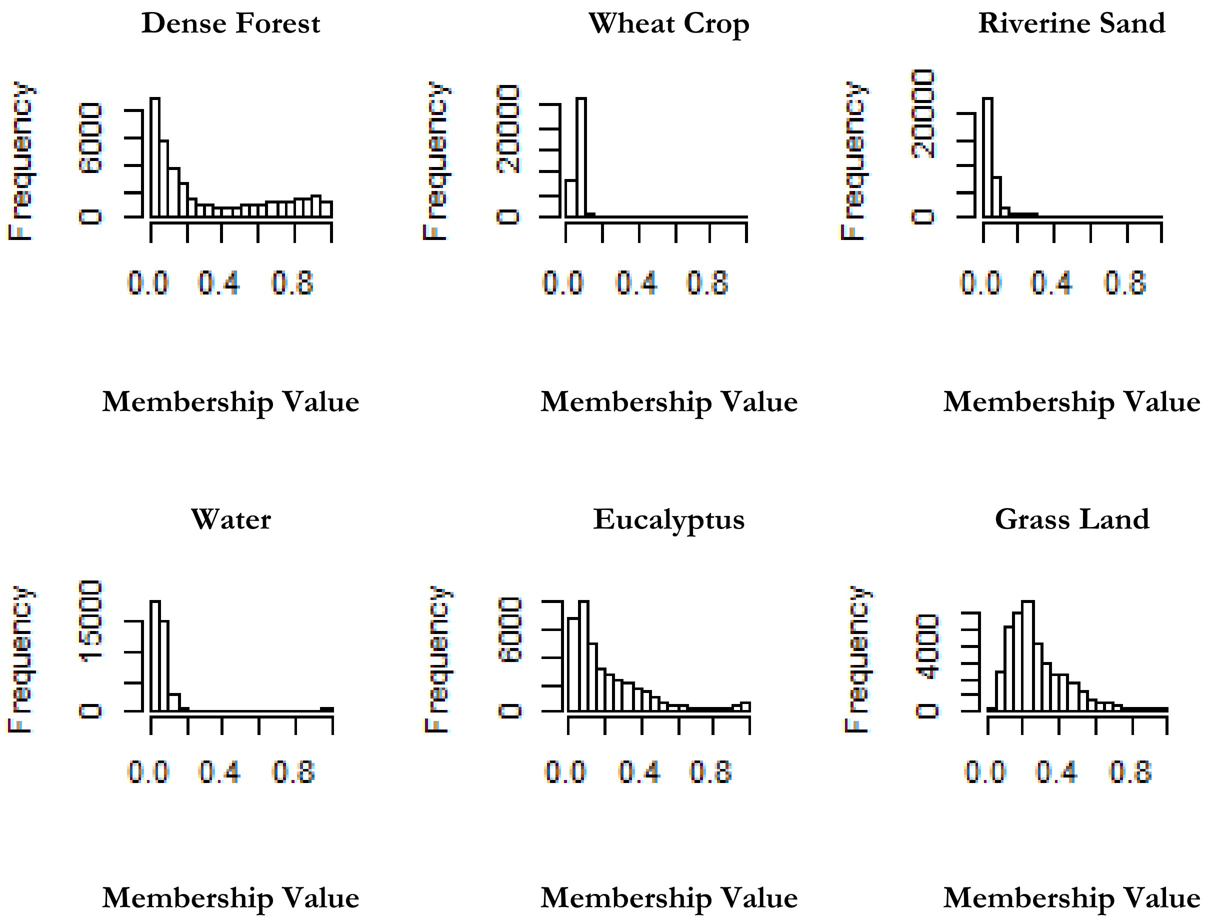
Table 5.2. The result showing global Root Mean Square Error and overall accuracy (in percent) of Fuzzy Error Matrix for a Possibilistic c Means classifier applied on labeled data after shifting mean with different distance based similarity measures.

Spectral Similarity Measures as a distance measure in PCM	Global RMSE for PCM	(FERM) Overall Accuracy of PCM (In percent)
Euclidean	0.202	85.845
SID-SAM-tan	0.359	73.997
SID-SAM-sin	0.358	73.329
SID-SCA-tan	0.330	81.417
SID-SCA-sin	0.323	76.789

FERM: Fuzzy Error Matrix; RMSE: Root Mean Square Estimate; PCM: Possibilistic c Means Classifier



(a)



(b)

Figure 5.3. Fractional Membership values for classes obtained using Possibilistic *c* means algorithm with few labeled samples and shifted mean with Euclidean as a distance measure (a) plot of class memberships (b) histogram of membership values

The results in Table 5.2 shows that the global root mean square value for Possibilistic *c* Means classifier with Euclidean as a distance measure is the lowest, and the overall accuracy of Fuzzy Error Matrix for the same is the highest than the Possibilistic *c* Means algorithm with all other hybrid distance measures. The fractional images as shown in Figure 5.3 represents membership values in each of the class when Possibilistic *c* Means algorithm is applied to the labeled data with shifted mean and Euclidean as a distance measure.

5.3. Result of semi-supervised Possibilistic *c* Means algorithm using different distance measures

This subsection presents the results obtained by applying Possibilistic *c* Means classifier in a semi-supervised mode with the mean estimated from the semi-supervised training samples. The unlabeled samples are obtained using hybrid spectral similarity measures by following the cluster and continuity assumption for which thresholds for different measures. The Possibilistic *c* Means algorithm is applied as described in Section 3.2 with different distance based similarity measures defined in Section 3.1, following Case 2 from Section 3.3. The accuracies have been assessed as shown in Table 5.4 by using global Root Mean Square Error and overall accuracy of Fuzzy Error Matrix Table 5.5 as described in Subsection 4.3.5.

Table 5.3. Euclidean distance between mean from labeled data shifted mean (MS) and Mean from semi-supervised data (semi-mean)

Classes	Euclidean Distance between Mean (MS) and Semi-Mean from (SID-SAM(tan/sin))	Euclidean Distance between Mean (MS) and Semi-Mean from (SID-SCA(tan/sin))	Euclidean Distance between Mean (MS) and Semi-Mean from (ED)
Dense Forest	36.974	38.539	63.814
Wheat Crop	45.669	45.223	63.573
Riverine Sand	86.183	83.883	50.056
Water	97.674	99.289	51.249
Eucalyptus	128.519	134.028	93.344
Grass land	101.823	117.524	75.689

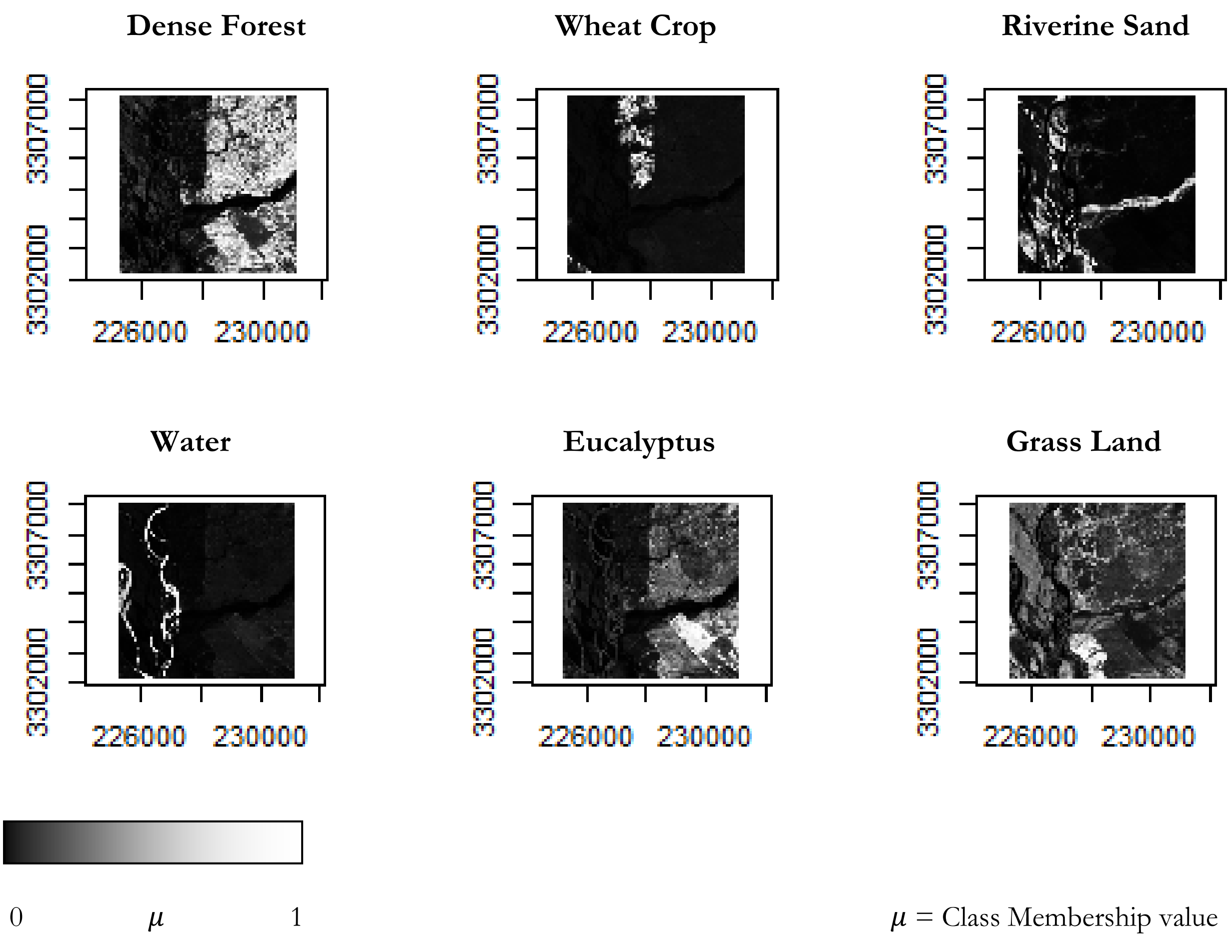
Table 5.4. The result of global Root Mean Square Error obtained by applying Possibilistic *c* Means classifier in a semi-supervised mode. The columns represent the global RMSE obtained when region growing is done using one of the similarity measures. The rows represent Possibilistic *c* Means classifier when one of the similarity measures is used as a distance measure.

Spectral Similarity Measures	Semi-supervised Possibilistic *c* Means (RMSE)		
	SID-SAM(tan/sin)	SID-SCA(tan/sin)	Euclidean
Euclidean	0.201	0.200	0.201
SID-SAM-tan	0.355	0.354	0.353
SID-SAM-sin	0.353	0.352	0.352
SID-SCA-tan	0.327	0.326	0.326
SID-SCA-sin	0.319	0.319	0.319

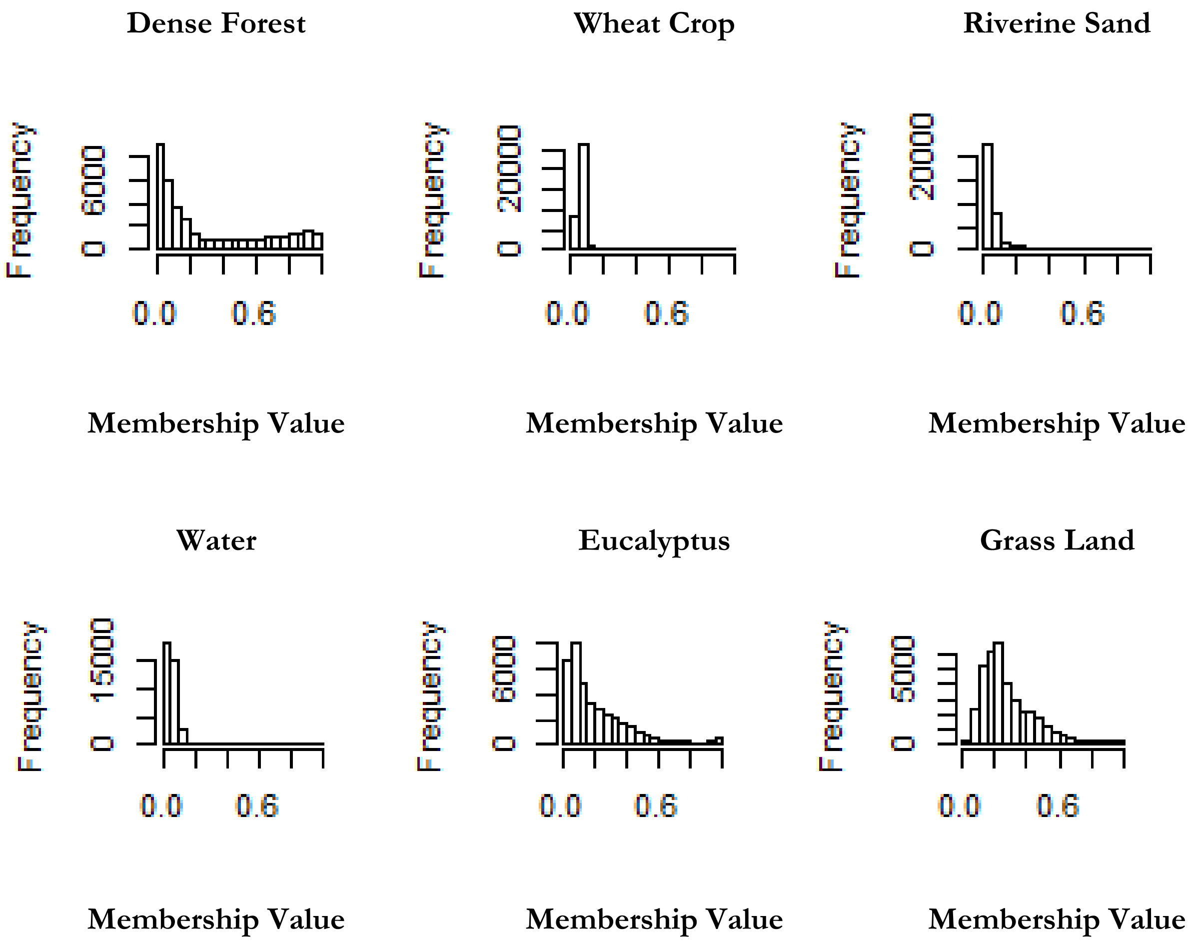
The difference of means obtained from semi-supervised approach applied using different spectral similarity measures and the mean obtained from the labeled training data after applying mean shift are estimated using Euclidean distance and are summarized in Table 5.3. The means estimated vary with the spectral similarity measure used. Higher value or distance represents a higher deviation from the region of high density. For all the spectral similarity measures, the shift in Eucalyptus class is maximum and even higher with hybrid measures. This suggests that the hybrid measures are able to capture the variance in Eucalyptus class. Table 5.4 and Table 5.5 shows that the results from all semi-supervised approach (using different similarity measures) are almost equivalent for all the corresponding distance measure in Possibilistic *c* Means algorithm. Among distance measures, Euclidean has a highest overall accuracy computed from Fuzzy Error Matrix and lowest global Root Mean Square Error. The fractional images showing the membership values in each of the class when semi-supervised Possibilistic *c* Means algorithm is applied using Euclidean as a distance measure is shown in Figure 5.4.

Table 5.5. The result of Overall Accuracy of Fuzzy Error Matrix obtained by applying Possibilistic *c* Means classifier in a semi-supervised mode. The columns represent the overall accuracies obtained when region growing is done using one of the similarity measures. The rows represent Possibilistic *c* Means classifier when one of the similarity measures is used as a distance measure.

Spectral Similarity Measures	Semi-supervised Possibilistic *c* Means (FERM Overall Accuracy in percent)		
	SID-SAM(tan/sin)	SID-SCA(tan/sin)	Euclidean
Euclidean	86.152	86.150	86.309
SID-SAM-tan	72.476	72.531	71.930
SID-SAM-sin	71.759	71.814	71.203
SID-SCA-tan	81.325	81.279	81.293
SID-SCA-sin	76.405	76.352	76.309



(a)



(b)

Figure 5.4. Membership values obtained Possibilistic *c* means algorithm with semi-supervised training data obtained using hybrid Spectral Information Divergence and Spectral Correlation Angle (SID-SCA) with Euclidean as a distance measure (a) plot of class memberships (b) histogram of membership values

5.4. Result of supervised Possibilistic *c* Means algorithm using different distance measures

This section shows the result of a supervised Possibilistic *c* Means algorithm, with a maximum of 80 samples per class collected manually with reference to the field data. The Possibilistic *c* Means algorithm is applied as described in Section 3.2 with different distance based similarity measures as defined in Section 3.1, following Case 2 from Section 3.3. The accuracies have been assessed as shown in Table 5.6 by using global Root Mean Square Error and overall accuracy of Fuzzy Error Matrix as described in Subsection 3.4.4. The results in Table 5.6 shows that the global root mean square value for a Possibilistic *c* Means classifier with Euclidean as a distance measure is the lowest, and the overall accuracy of Fuzzy Error Matrix is highest for the same.

Table 5.6. The result showing global Root Mean Square Error and overall accuracy of Fuzzy Error Matrix for a Possibilistic *c* Means classifier applied in a fully supervised mode with different distance based similarity measures.

Spectral Similarity Measures	Global RMSE	(FERM) Overall Accuracy (in percent)
Euclidean	0.207	86.764
SID-SAM-tan	0.351	73.313
SID-SAM-sin	0.349	72.673
SID-SCA-tan	0.313	76.851
SID-SCA-sin	0.312	72.129

5.5. Spectral Discriminatory Power

This section presents the results of discriminatory power as described in Section 3.4.5 which compares the discriminatory ability of different spectral similarity measures to discriminate between classes with respect to a reference class.

Table 5.7. Spectral Discriminatory Power of different spectral similarity measures.

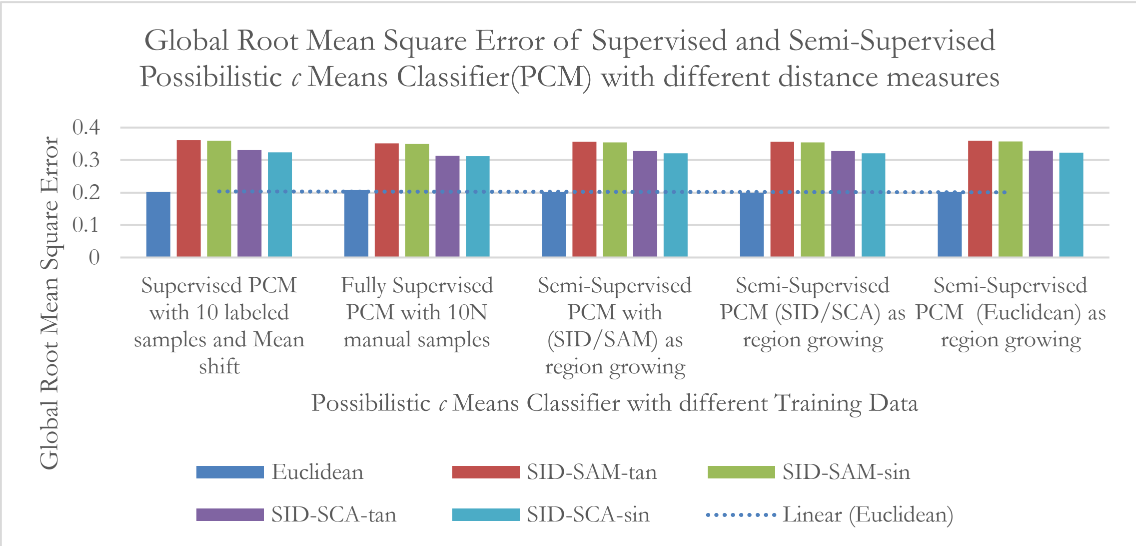
Spectral Similarity Measure	Spectral Discriminatory Power
ED	2.823
SAM	1.153
SCA	1.805
SID	1.343
SID-SAM-tan	1.550
SID-SAM-sin	1.548
SID-SCA-tan	2.523
SID-SCA-sin	2.379

Table 5.7 shows that the Spectral Discriminatory Power of the Euclidean measure is the highest and that of Spectral Angle Measure is the lowest. From Table 5.7, the Spectral Discriminatory Power of the Euclidean measure to discriminate between Forest and Grass Land is almost three times better to distinguish Forest from Eucalyptus. Also, the Spectral Discriminatory Power for hybrid Spectral Information Divergence-Spectral Correlation Angle (SID-SCA-tan) is also close to the Euclidean.

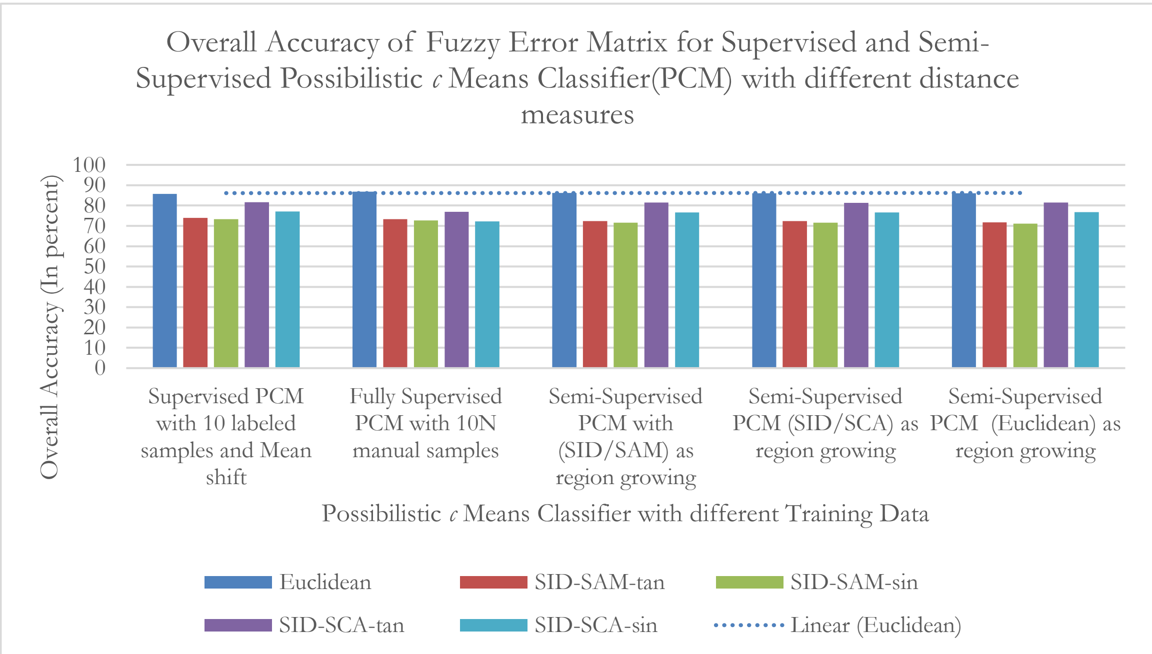
5.6. Comparison of semi-supervised with the supervised approach with different spectral similarity measures

This section compares the results of semi-supervised and supervised approaches in a Possibilistic *c* Means classifier as discussed in Section 5.2.1, Section 5.3 and Section 5.4. Figure 5.5 (a) compares the global Root Mean Square Error, and Figure 5.5 (b) compares the overall accuracies obtained from Fuzzy Error Matrix. This section also shows the trend of results obtained for individual classes.

Figure 5.5 shows that for the corresponding distance measure in Possibilistic *c* Means algorithm, the results of supervised and semi-supervised approaches are almost similar with very slight deviations in the global Root Mean Square values as well as the overall accuracies of Fuzzy Error Matrix. This shows that even a very little number of training data if applied with Mean Shift Algorithm (Section 3.4.2.1), gives a similar estimate of mean and variance of the classes as compared to completely supervised approach. In this research, the bandwidth parameter of the Mean Shift Algorithm has been related to the bandwidth parameter of the Possibilistic *c* Means algorithm, as discussed in Section 3.4.2.1.



(a)



(b)

Figure 5.5. Comparison between supervised and semi-supervised approaches with different spectral similarity measures (a) global Root Mean Square Error (b) overall accuracy of Fuzzy Error Matrix

Figure 5.5 also shows that among all the spectral similarity measures, Euclidean stands out to be the best distance measure when Possibilistic *c* Means algorithm is applied for classifying multispectral imagery irrespective of the supervised or semi-supervised learning approaches in terms of overall accuracies of Fuzzy Error Matrix and global Root Mean Square Error. The results obtained in Section 5.2.1 is also coherent with the results obtained for Possibilistic *c* Means classifier with different spectral similarity measures.

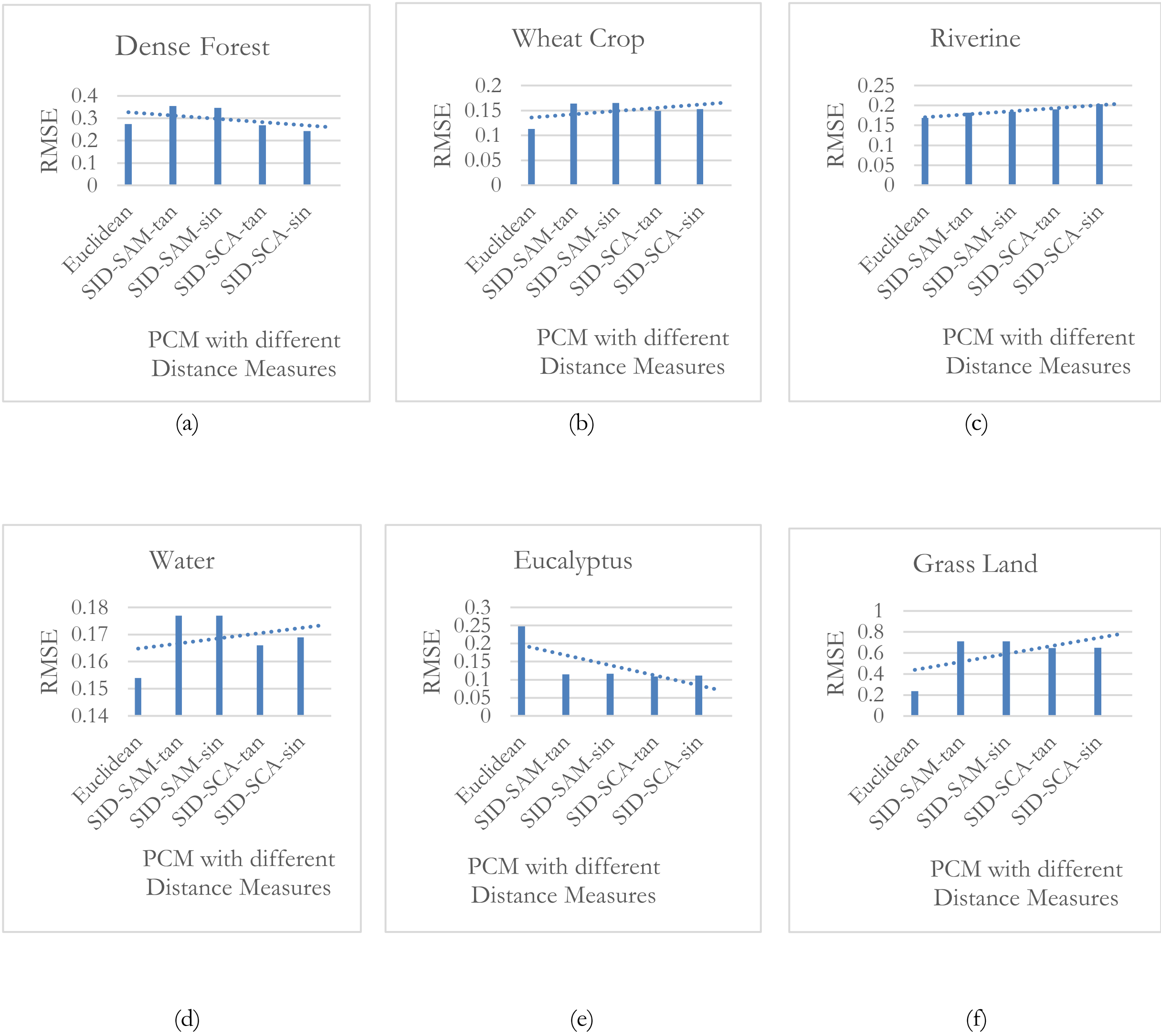


Figure 5.6. Result of trend followed by classes for supervised as well as semi-supervised approach based on Root Mean Square Estimation

Figure 5.6 shows the general trend based on the class-wise implementation of Root Mean Square estimate followed by classes after classifying the multispectral imagery with Possibilistic *c* Means algorithm in case of all supervised and semi-supervised approaches. It shows that the hybrid measure works best in case of Eucalyptus class. For Dense forest class, the hybrid SID-SCA (tan/sin) is analogous to Euclidean. Figure 5.7 shows the membership values obtained for Eucalyptus class when supervised Possibilistic *c* Means algorithm is applied with SID-SAM-tan as a distance based similarity measure and Table 5.8 shows the User's and the Producer's accuracy for Eucalyptus class for the same approach. The User's accuracies for the hybrid measures are higher than for the Euclidean measure, and the Producer's accuracies are lower for hybrid measures than for Euclidean. Figure 5.8 shows the membership values for Eucalyptus class when Euclidean is employed as a distance measure instead of SID-SAM-tan in a Possibilistic *c* Means classifier.

Table 5.8. User's and Producer's Accuracy for Eucalyptus class obtained by applying Fuzzy Error Matrix on Possibilistic c mean classifier with SID-SAM-tan as a distance measure

Spectral Similarity Measures	User's Accuracy (in percent)	Producer's Accuracy (in percent)
Euclidean	57.361	97.981
SID-SAM-tan	91.616	51.893
SID-SAM-sin	92.028	50.457
SID-SCA-tan	82.176	78.073
SID-SCA-sin	86.467	69.224

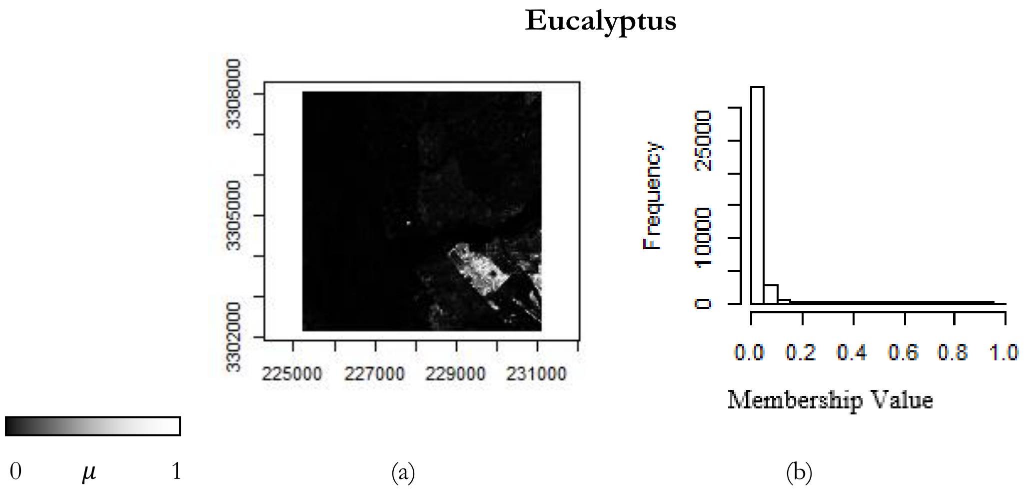


Figure 5.7. Membership values for Eucalyptus class for PCM with SID-SAM-tan as a similarity measure (a) plot of membership values μ (b) histogram of membership values

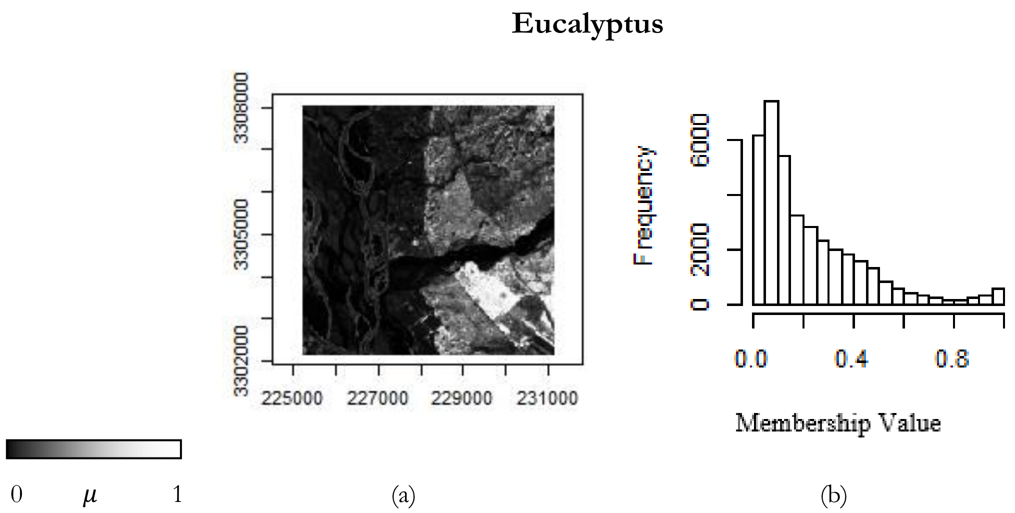


Figure 5.8. Membership values for Eucalyptus class for a Possibilistic *c* Means classifier with Euclidean as a similarity measure (a) plot of membership values μ (b) histogram of membership values

6. DISCUSSION

This chapter discusses the results obtained in Chapter 5 with the objective of analyzing the role of hybrid spectral similarity measures for a semi-supervised fuzzy classifier.

The hybrid spectral similarity measures, Spectral Information Divergence with Spectral Angle Measure and Spectral Information Divergence with Spectral Correlation Angle are used to calculate the spectral similarity between pixels for collecting unlabeled data for semi-supervised approach and as a distance measure in a Possibilistic *c* Means classifier.

As a part of collecting unlabeled samples, it has been observed that the individual hybrid measures and Euclidean measure collected different unlabeled samples and the mean estimated, is also different from them, but the results obtained for all the semi-supervised Possibilistic classifier are analogous to each other. The results suggest that the Possibilistic *c* Means classifier with a particular distance measure is independent of the similarity measure used for increasing the training samples if the number of training data is a representative of a particular class and the mean estimated from them represents the exact mean of a class. This could be seen from the working of Possibilistic *c* Means algorithm as described in Section 3.2, which depends upon the initialization of mean, membership values as well estimation of bandwidth parameter. As long as the training data is able to capture the intra-class variance (related with bandwidth parameter) based on the estimation of mean and initial class memberships, the Possibilistic *c* Means algorithm would give good results.

As a part of distance-based similarity measure in a Possibilistic *c* Means classifier, the Euclidean measure outperforms all the hybrid measures in terms of overall accuracies for classifying multi-spectral images. The observed results suggest that the Euclidean is less sensitive to inter-class and intra-class variations as compared to the hybrid measures. This could also be observed from the classified outputs of classes Eucalyptus and Dense Forest where hybrid measures performed better than the Euclidean or analogous to Euclidean because of the lesser intra-class variability. This means that if there are lesser variations in the classes, the hybrid spectral similarity measure can perform better than Euclidean.

The one of another possible reason could be in terms of measuring band to band spectral variability for measuring the similarity between pixels of multispectral images. The hybrid measures are based on the concept of Spectral Information Measure (Chang, 2000) which calculates the similarity between two pixel vectors by modeling their band to band spectral variability. In case of multi-spectral images, the bands are limited, and the band to band variability is not higher because of lower spectral resolution as compared to hyperspectral images. With the lower spectral resolution, these measures become equivalent to Euclidean or even worse.

It is observed that the results of semi-supervised approach are analogous to a supervised approach where a large training data was collected manually, and Possibilistic *c* Means algorithm was applied using respective distance based similarity measures. The results suggest that the training data collected using spectral similarity measures were comparable to the manual supervised samples in terms of calculating a reliable estimate of class means and variances.

In the other supervised approach, when little labeled samples were collected and Mean Shift algorithm was applied, the results from the Possibilistic *c* Means classifier with different distance based similarity measures were analogous to the other supervised and semi-supervised approaches. This suggests that the estimates of mean and variances for classes obtained from Mean Shift algorithm were comparable to other supervised and semi-supervised training samples. This could be because the Mean Shift algorithm is able to correctly refine the labeled data and shift the mean to a higher density region for classification.

The possible reason of Mean Shift to give better estimates of mean could be related with the usage of the bandwidth parameter that was employed from the iterative procedure of Possibilistic *c* Means algorithm performed using a few labeled sample as described in Section 3.4.2.1. This suggests that it is possible to get a better estimate of variance with a little training data if the variance for a class is related to the bandwidth parameter of Possibilistic *c* Means algorithm obtained iteratively.

To analyze the results and this research, SWOT analysis has been done as shown in Table 6.1.

Table 6.1. SWOT analysis of the research.

	Beneficial to achieving objectives	**Detrimental to achieving objectives**
	STRENGTHS (S)	**WEAKNESSES (W)**
Internal Origin	• The strength of the project lies in the approach and methods which revealed that Mean Shift algorithm could be used for defining mean for a fuzzy classifier. • Even a few labeled training samples can give an estimate of variance if bandwidth parameter is related to variance and estimated iteratively. • The semi-supervised approach is comparable with the supervised approach, therefore, no need for collecting large training data manually. • The methods can help dealing effectively with mixed pixels and vague boundaries even with a few labeled training data using Mean Shift algorithm.	• The hyperspectral measures in measuring similarity in a multispectral imagery were comparatively less effective than conventional measure (Euclidean distance). • The hybrid measures are more complex than conventional measure, still not effective. • The hybrid measures as a distance measures are not able to handle higher class variability of a multi-spectral image. • The fuzzy classifiers with hybrid measures as a distance measure are less effective in dealing with uncertainties due to mixed pixels.
	OPPORTUNITIES (O)	**THREATS (T)**
External Origin	• Methods based on a few labeled samples can be applied in dealing with uncertainties and where a collection of a large training data is not possible. • Methods can be effective in the classification of remote sensing images which has high uncertainties due to fuzziness and mixed pixels without the need of a large labeled data.	• The methods may prove less effective with very high resolution images as provided by Unmanned Aerial Vehicles. • The unavailability of standardized assessment tools for the fuzzy classification techniques questions the applicability of the approaches.

To handle uncertainties due to mixed pixels and vague classes, the study reveals that even a very little amount of labeled training data can help achieving better classification accuracies for a Possibilistic fuzzy classifier if the labeled data is a representative of a class and is able to give a better estimate of class means and variances. Various works on semi-supervised learning including objective function modification of a Possibilistic *c* Means classifier as discussed in Section 2.2 substantiate the use of a little labeled samples with a large unlabeled samples for achieving classification accuracies, but the results from this research with the Mean Shift algorithm eliminates the need for having a large unlabeled data for a Possibilistic *c* Means classifier.

7. CONCLUSION AND RECOMMENDATION

This chapter concludes the research with some recommendation. **Section 7.1** presents the conclusion drawn from this research. **Section 7.2** provides some recommendations for future study.

7.1. Conclusion

The classification of multispectral images is always a challenging task because of the uncertainties present in the image due to fuzzy boundaries and mixed pixels. In order to improve the quality of classification, the choice of methods plays an important role. In this research, the hybrid measures are studied for a multispectral imagery. These hybrid measures are applied in increasing the training data as a part of unlabeled data in semi-supervised learning to get an improved mean from similar samples. These hybrid measures are also used as a distance based similarity measure in a Possibilistic *c* Means classifier applied using semi-supervised training data for improving the classification accuracies by assigning similar pixels to classes. Their roles have been studied in view of multispectral imagery that has higher inter-class and intra-class variance. Due to the uncertainties present in multispectral images, manual collection of pure labeled samples becomes challenging. Therefore, the use of semi-supervised approach is emphasized. For improving the quality of the labeled samples and increasing the knowledge obtained from the labeled training samples, Mean Shift algorithm has been investigated. The bandwidth parameter for the mean shift algorithm is related to the bandwidth parameter of the Possibilistic *c* Means classifier and investigated for obtaining a better initial mean for classification. For improving the classification accuracies, a Possibilistic *c* Means algorithm is chosen and investigated with different techniques. Two different cases for employing Possibilistic *c* Means classifier have been studied and experimented with different training data. The concept of iterations for estimating variance for a mean shift from a fewer labeled data has been studied. The Mean Shift algorithm has been investigated for refining the labeled training data and for shifting the mean from the training data to a higher density region. The Possibilistic *c* Means classifier has been applied to different training datasets using various similarity measures and their classification accuracies have been recorded and observed.

From the experiments and study on using the hybrid spectral similarity measures as a distance based similarity measures in a Possibilistic *c* Means classifier, the investigations shows that they are able to classify multispectral image, by measuring the similarity between pixels but are not better than the conventional Euclidean measure which is easy to use and less complicated than the hybrid measures. As a part of the semi-supervised approach, the study shows that the hybrid measures if used to collect the training samples for a Possibilistic *c* Means classifier gives comparable results to Euclidean measure. The investigations on Mean Shift algorithm revealed that the Possibilistic *c* Means classifier with shifted mean gives comparable accuracies to both completely supervised and semi-supervised approaches, without the need of increasing the size of training data. Among, hybrid measures, SID-SCA measure gives higher overall accuracies from Fuzzy Error Matrix, and lower global Root Mean Square Error than other hybrid measures and SID-SAM measure works better in assigning memberships to Eucalyptus class which has lower intra-class variability.

This study answers the following research questions as proposed:

- **Question:** Are the proposed hybrid spectral similarity measures effective in identifying the similarity between pixels for a multi-spectral image?

Answer: The proposed hybrid spectral similarity measures are not effective in comparison with the conventional Euclidean measure in measuring the similarity between pixels for the multispectral image. This is because of the higher intra-class and inter-class variability which is often inevitable in a multispectral image. The investigation shows that the hybrid measure works better than Euclidean only in cases where intra-class variability is lower.

- **Question:** Is semi-supervised approach applied before classification better than optimizing the objective function of the classifier?

  Answer: Optimization of the objective function for applying semi-supervised approach requires modification of the objective function of a classifier and initializations of memberships that are either one or zero, as presented in Case 1 (Section 3.3), as soft membership values for the increased and unlabeled samples are unavailable initially. With our study on different cases of Possibilistic *c* Means classifier as described in Section 5.2, it is unable to provide better classification accuracies than Case 2. Therefore, not adopted for this research. With modifications and optimizations, it may give results but is not within the scope of this research.

- **Question:** What will be the effect of hybrid spectral similarity measures on classification accuracy as compared to conventional similarity measures?

  Answer: The classification accuracies are assessed based on global Root Mean Square Error and overall accuracies from Fuzzy Error Matrix. The results as described in Section 5.2.1, Section 5.3 and Section 5.4 shows that the overall classification accuracies are lower and the global Root Mean Square Error is higher for a Possibilistic *c* Means classifier with hybrid measures than the conventional Euclidean distance as a distance based similarity measure.

- **Question:** Is proposed semi-supervised Possibilistic fuzzy classifier with hybrid spectral similarity measures better in dealing with uncertainties than supervised classifier?

  Answer: The proposed semi-supervised Possibilistic fuzzy classifier with hybrid measures are equivalent to their supervised counterpart with the respective hybrid similarity measures. But they are not better in comparison to the Euclidean distance as a distance based similarity measure. The semi-supervised approach and supervised approach both tend to give similar results. In addition, the Possibilistic *c* Means classifier works with a very little number of labeled samples, without the need of large unlabeled samples when Mean Shift algorithm is applied for shifting the mean to a higher density region.

This research also answers the additional research questions:

- **Question:** How to get a better estimate of variance with a little training data?

  Answer: This research identifies a method of obtaining an estimate of variance with a little training data with an iterative approach using a Possibilistic *c* Means classifier as described in Section 3.4.2.1.

- **Question:** How mean shift algorithm can help in estimating an effective mode for classification?

  Answer: This estimate of variance (bandwidth parameter of Possibilistic *c* Means classifier, also known as η) if related to the bandwidth parameter of the Gaussian kernel in Mean Shift algorithm and used for shifting the mean of labeled samples, results in the similar classification accuracies as of other supervised and semi-supervised approaches.

7.2. Recommendation

This subsection discusses the recommendations for the future works.

- The optimization of Possibilistic *c* Means in terms of degree of fuzziness can be explored.
- A detailed exploration of the semi-supervised technique with objective function modification can be done.
- A detailed study of the semi-supervised learning assumptions on the choice of a similarity measure can be done.
- The effect of iterations on the modified versions of Possibilistic *c* Means algorithm can be explored.
- A study on hybrid measures for estimating the variance of a class can be done.

LIST OF REFERENCES

Adler, D., & Murdoch, D. (2017). rgl: 3D Visualization Using OpenGL. Retrieved from https://cran.r-project.org/package=rgl

Atkinson, P. M., & Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of Remote Sensing*, *18*(4), 699–709. https://doi.org/10.1080/014311697218700

Belohlavek, R., & Vychodil, V. (2005). Introduction to Fuzzy Sets and Fuzzy Logic. In R. B. Shepard (Ed.), *StudFuzz* (Vol. 186, pp. 1–58). New York, NY: Springer New York. https://doi.org/10.1007/0-387-28098-7_9

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2-3), 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Binaghi, E., Brivio, P. A., Ghezzi, P., & Rampini, A. (1999). A fuzzy set accuracy assessment of soft classification. *Pattern Recognition Letters*, *20*, 935–948.

Bouchachia, A., & Pedrycz, W. (2006). Enhancement of fuzzy clustering by mechanisms of partial supervision. *Fuzzy Sets and Systems*, *157*(13), 1733–1759. https://doi.org/10.1016/j.fss.2006.02.015

Campbell, J. B. (1996). *Introduction to remote sensing. Second edition. Introduction to remote sensing. Second edition.*

Chang, C. I. (2000). An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory*, *46*(5), 1927–1932. https://doi.org/10.1109/18.857802

Chauhan, H., & Krishna Mohan, B. (2014). Effectiveness of Spectral Similarity Measures to Develop Precise Crop Spectra for Hyperspectral Data Analysis. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-8*(December), 83–90. https://doi.org/10.5194/isprsannals-II-8-83-2014

Chawla, S. (2010). *Possibilistic c - means - spatial contextual information based sub - pixel classification approach for multi - spectral data*. University of Twente. Retrieved from http://www.itc.nl/library/papers_2010/msc/gfm/chawla.pdf

Cheng, T. (2002). Fuzzy objects: Their changes and uncertainties. *Photogrammetric Engineering and Remote Sensing*, *68*(1), 41–49.

Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(8), 790–799. https://doi.org/10.1109/34.400568

Comaniciu, D., Meer, P., & Member, S. (2002). Mean Shift : A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(5), 1–18. https://doi.org/10.1109/34.1000236

Davis, S. M., Landgrebe, D. A., Phillips, T. L., Swain, P. H., Hoffer, R. M., Lindenlaub, J. C., & Silva, L. F. (1978). *Remote Sensing: The quantitative approach*. *New York* (Vol. 1).

Dongre, P. K. (2016). *Study of Similarity and Dissimilarity Measures With Possibilistic Based Fuzzy*. Andhra University, Visakhapatnam.

Drews, P., Colares, R. G., Machado, P., de Faria, M., Detoni, A., & Tavano, V. (2013). Microalgae classification using semi-supervised and active learning based on Gaussian mixture models. *Journal of the Brazilian Computer Society*, *19*(4), 411–422. Retrieved from https://journal-bcs.springeropen.com/articles/10.1007/s13173-013-0121-y

Du, Y., Chang, C. I., Ren, H., Chang, C. C., Jensen, J. O., & D'Amico, F. M. (2004). New hyperspectral discrimination measure for spectral characterization. *Optical Engineering*, *43*(8), 1777.

https://doi.org/10.1117/1.1766301

Enderton, H. B. (1977). Elements of Set Theory. *Information Storage and Retrieval*, *58*(303), 279.

Esri. (2016). ArcMap 10.4.1. *ESRI*.

Fano, R. M. (1961). Transmission of Information: A Statistical Theory of Communications. *American Journal of Physics*, *29*(11), 793. https://doi.org/10.1119/1.1937609

Fisher, P. (1997). The pixel: A snare and a delusion. *International Journal of Remote Sensing*, *18*(3), 679–685. https://doi.org/10.1080/014311697219015

Fisher, P. F., & Pathirana, S. (1990). The evaluation of fuzzy membership of land cover classes in the suburban zone. *Remote Sensing of Environment*, *34*(2), 121–132. https://doi.org/10.1016/0034-4257(90)90103-S

Foody, G. M. (2000). Estimation of sub-pixel land cover composition in the presence of untrained classes. *Computers and Geosciences*, *26*(4), 469–478. https://doi.org/10.1016/S0098-3004(99)00125-9

Foody, G. M., & Cox, D. P. (1994). Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *International Journal of Remote Sensing*, *15*(3), 619–631. https://doi.org/10.1080/01431169408954100

Fukunaga, K., & Hostetler, L. D. (1975). The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. *IEEE Transactions on Information Theory*, *21*(1), 32–40. https://doi.org/10.1109/TIT.1975.1055330

Goshtasby, A. A. (2012). Image Registration Principles, Tools and Methods. In *Advances in Computer Vision and Pattern Recognition*. Springer London. https://doi.org/10.1007/978-1-4471-2458-0

Graler, B., Pebesma, E., & Heuvelink, G. (2017). *Encyclopedia of GIS*. (S. Shekhar, H. Xiong, & X. Zhou, Eds.), *Wp* (Vol. 8). Cham: Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-17885-1

Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. Retrieved from https://cran.r-project.org/package=raster

Jensen, J. R., & Lulla, K. (1987). Introductory digital image processing: A remote sensing perspective. *Geocarto International*, *2*(1), 65. https://doi.org/10.1080/10106048709354084

Kandpal, N. (2016). *Non-Linear Separation Of Classes Using Kernel Based Possibilistic c-Means*. University of Twente. Retrieved from http://www.itc.nl/library/papers_2016/msc/gfm/kandpal.pdf

Krishnapuram, R., & Keller, J. M. (1993). A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*, *1*(2), 98–110. https://doi.org/10.1109/91.227387

Krishnapuram, R., & Keller, J. M. (1996). The possibilistic C-means algorithm: Insights and recommendations. *IEEE Transactions on Fuzzy Systems*, *4*(3), 385–393. https://doi.org/10.1109/91.531779

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., & Goetz, A. F. H. (1993). The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*. https://doi.org/10.1016/0034-4257(93)90013-N

Kullback, S. (1997). Information Theory and Statistics. Gloucester, Mass.: Dover.

Kumar, M., Seshasai, M. V. R., Vara Prasad, K. S., Kamala, V., Ramana, K. V., Dwivedi, R. S., & Roy, P. S. (2011). A new hybrid spectral similarity measure for discrimination among Vigna species. *International Journal of Remote Sensing*, *32*(14), 4041–4053. https://doi.org/10.1080/01431161.2010.484431

Lee, S. H., Pedrycz, W., & Sohn, G. (2009). Design of similarity and dissimilarity measures for fuzzy sets on the basis of distance measure. *International Journal of Fuzzy Systems*, *11*(2), 67–72.

Li, J., Bioucas-Dias, J. M., & Plaza, A. (2011). Semi-supervised hyperspectral image classification using a new (soft) sparse multinomial logistic regression model. *2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, *10*(2), 1–4. https://doi.org/10.1109/WHISPERS.2011.6080879

Ligges, U., & Mächler, M. (2003). scatterplot3d - An *R* Package for Visualizing Multivariate Data. *Journal of Statistical Software*, *8*(11), 1–36. https://doi.org/10.18637/jss.v008.i11

Lillesand, T. M., & Kiefer, R. W. (2015). Remote sensing and image interpretation. 7th edition. *Remote Sensing and Image Interpretation.*

Liu, L., & Wu, X.-J. (2013). Semi-Supervised Possibilistic Fuzzy c-Means Clustering Algorithm on Maximized Central Distance. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*. Paris, France: Atlantis Press. https://doi.org/10.2991/iccsee.2013.342

Lucieer, A. (2004). *Uncertainties in Segmentation and their Visualisation*. University of Twente (ITC) and Utrecht University. Retrieved from https://www.itc.nl/library/Papers_2004/phd/lucieer.pdf

Macario, V., & Carvalho, F. D. A. T. De. (2010). A New Approach for Semi-Supervised Clustering Based on Fuzzy C-Means, *970*, 18–23.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. Retrieved from https://cran.r-project.org/package=e1071

Miyamoto, S., Ichihashi, H., & Honda, K. (2008). *Algorithms for Fuzzy Clustering*. (J. Kacprzyk, Ed.), *Methods* (Vol. 229). https://doi.org/10.1007/978-3-540-78737-2

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Mukhopadhaya, S. (2016). *Exploring measures of similarity / dissimilarity for fuzzy classifier : from data quality to distance quality*. University of Twente. Retrieved from http://www.itc.nl/library/papers_2016/msc/gfm/mukhopadhaya.pdf

Olivier, C., Schölkopf, B., & Zien, A. (2006). *Semi-Supervised Learning. Interdisciplinary sciences computational life sciences* (Vol. 1). Cambridge, Massachusetts: The MIT Press.

Orlitsky, A. (2005). Estimating and computing density based distance metrics. *Proceedings of the 22nd International Conference on Machine Learning - ICML '05*, 760–767. https://doi.org/10.1145/1102351.1102447

Padma, S., & Sanjeevi, S. (2014). Jeffries matusita-spectral angle mapper (JM-SAM) spectral matching for species level mapping at Bhitarkanika, Muthupet and Pichavaram mangroves. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *40*(8), 1403–1411. https://doi.org/10.5194/isprsarchives-XL-8-1403-2014

Panda, A. (2017). *Different Measures of Similarity and Dissimilarity for Fuzzy Based Noise Classifier*. University of Twente. Retrieved from https://webapps.itc.utwente.nl/library/2017/msc/gfm/panda.pdf

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, *30*(7), 683–691. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0098300404000676

Pedrycz, W., & Waletzky, J. (1997). Fuzzy clustering with partial supervision. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *27*(5), 787–795.

Persello, C., & Bruzzone, L. (2014). Active and Semisupervised Learning for the Classification of Remote

Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(11), 6937–6956.

R Development Core Team, R. (2017). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing* (Vol. 1). Retrieved from http://www.r-project.org

Richards, J. A., & Jia, X. (2013). *Remote Sensing Digital Image Analysis*. *Journal of Chemical Information and Modeling* (4th ed., Vol. 53). Springer-Verlag Berlin Heidelberg.

Rowlingson, R. B. and T. K. and B. (2017). rgdal: Bindings for the "Geospatial" Data Abstraction Library. Retrieved from https://cran.r-project.org/package=rgdal

Schuurmans, D., & Southey, F. (2002). Metric-based methods for adaptive model selection and regularization. *Machine Learning*, *48*(1-3), 51–84. https://doi.org/10.1023/A:1013947519741

Shanbehzadeh, J. (2013). Improve Semi-Supervised Fuzzy C-means Clustering Based On Feature Weighting, (14), 82–89.

Shepard, R. B. (Ed.). (2005). Introduction to Fuzzy Sets and Logic. In *Quantifying Environmental Impact Assessments Using Fuzzy Logic* (pp. 63–113). New York, NY: Springer New York. https://doi.org/10.1007/0-387-28098-7_9

Shi, W. Z., Ehlers, M., & Molenaar, M. (2005). Uncertainties in integrated remote sensing and GIS. *International Journal of Remote Sensing*, *26*(14), 2911–2915. https://doi.org/10.1080/01431160500197537

Stehman, S. V., Arora, M. K., Kasetkasem, T., & Varshney, P. K. (2007). Estimation of Fuzzy Error Matrix Accuracy Measures Under Stratified Random Sampling. *Photogrammetric Engineering & Remote Sensing*, *73*(2), 165–173. https://doi.org/10.14358/PERS.73.2.165

Stein, A. (2010). Fuzzy Methods in Image Mining. In R. Jeansoulin, O. Papini, H. Prade, & S. Schockaert (Eds.), *Methods for Handling Imperfect Spatial Information* (pp. 243–268). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14755-5_10

Triguero, I., García, S., & Herrera, F. (2015). Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. *Knowledge and Information Systems*, *42*(2), 245–284. https://doi.org/10.1007/s10115-013-0706-y

Tso, B., & Mather, P. M. (2009). *Classification methods for remotely sensed data. Taylor and Francis Group.*

van der Meer, F. (2006). The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, *8*(1), 3–17. https://doi.org/10.1016/j.jag.2005.06.001

van der Meer, F., & Bakker, W. (1997). Cross correlogram spectral matching: application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada. *Remote Sensing of Environment*, *61*(3), 371–382.

Wu, X.-H., & Zhou, J.-J. (2006). An Improved Possibilistic C-Means Algorithm Based on Kernel Methods. In D.-Y. Yeung, J. T. Kwok, A. Fred, F. Roli, & D. de Ridder (Eds.), *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, SSPR 2006 and SPR 2006, Hong Kong, China, August 17-19, 2006. Proceedings* (pp. 783–791). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11815921_86

Xie, Z., Wang, S., & Chung, F. L. (2008). An enhanced possibilistic C-Means clustering algorithm EPCM. *Soft Computing*, *12*(6), 593–611. https://doi.org/10.1007/s00500-007-0231-6

Zadeh, L. A. (1965). Fuzzy Logic. *Information and Control*, *8*, 338–353.

Zhang, J. S., & Leung, Y. W. (2004). Improved possibilistic C-means clustering algorithms. *IEEE Transactions on Fuzzy Systems*, *12*(2), 209–217. https://doi.org/10.1109/TFUZZ.2004.825079

Zhu, X. (2008). Semi-Supervised Learning Literature Survey Contents. *University of Wisconsin-Madison*, 60.

Zimmermann, H. J., & Zysno, P. (1985). Quantifying vagueness in decision models. *European Journal of Operational Research*, *22*(2), 148–158. https://doi.org/10.1016/0377-2217(85)90223-1

Zimmermann, H.-J. (2010). Fuzzy set theory. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 317–332. https://doi.org/10.1002/wics.82

Zwick, R., Carlstein, E., & Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, *1*(2), 221–242. https://doi.org/10.1016/0888-613X(87)90015-6

APPENDIX A

Support Vector Machine (SVM) classified outputs of Formosat-2 satellite image as used for reference, are shown in Figure A-1 and Figure A-2 respectively. The Formosat-2 (8m) image has been resampled to 10m before classifying to compare it with input Landsat-8 image.

The classification model has the following specifications: Classification: C-classification, Kernel: Linear, Cost function: 10, SVM Model Accuracy: 99.513% and 10-fold cross-validation accuracy of 99.352% and prediction accuracy on test data for Hard classification as 99.111%.

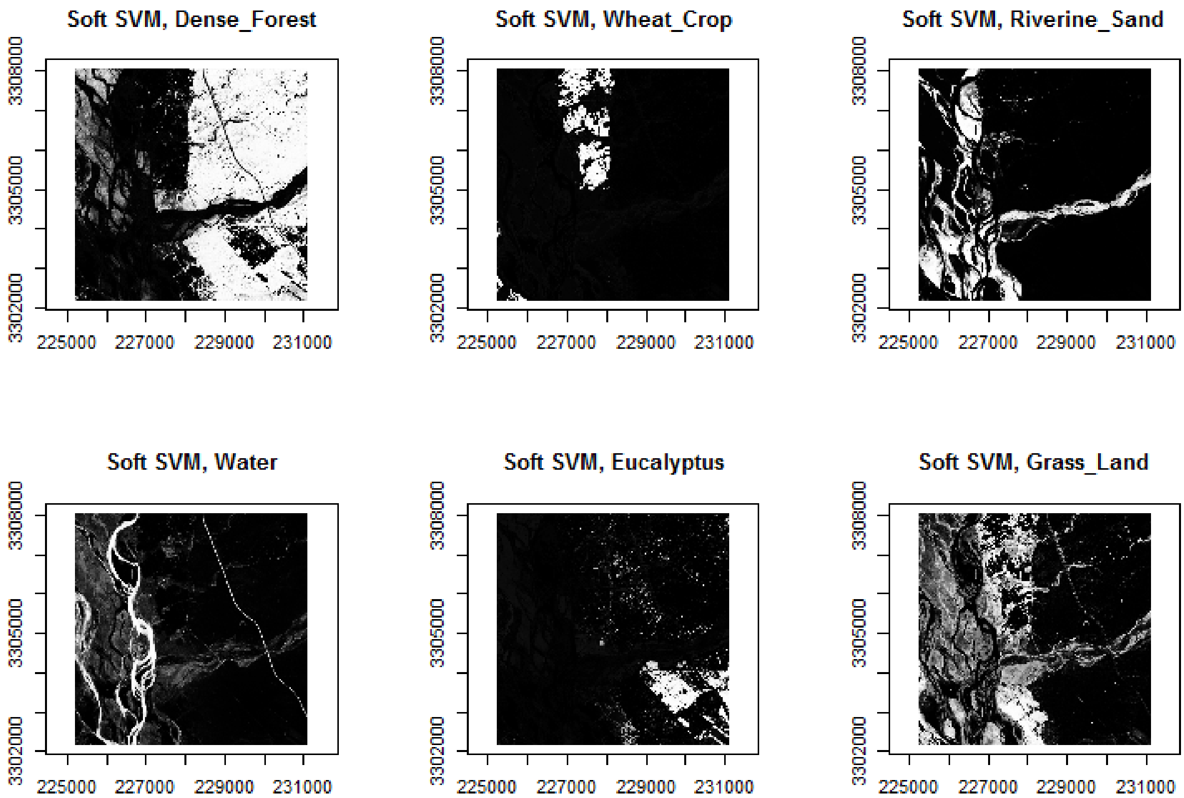


Figure A- 1. Soft Classified Formosat 2 Imagery using Support Vector Machine

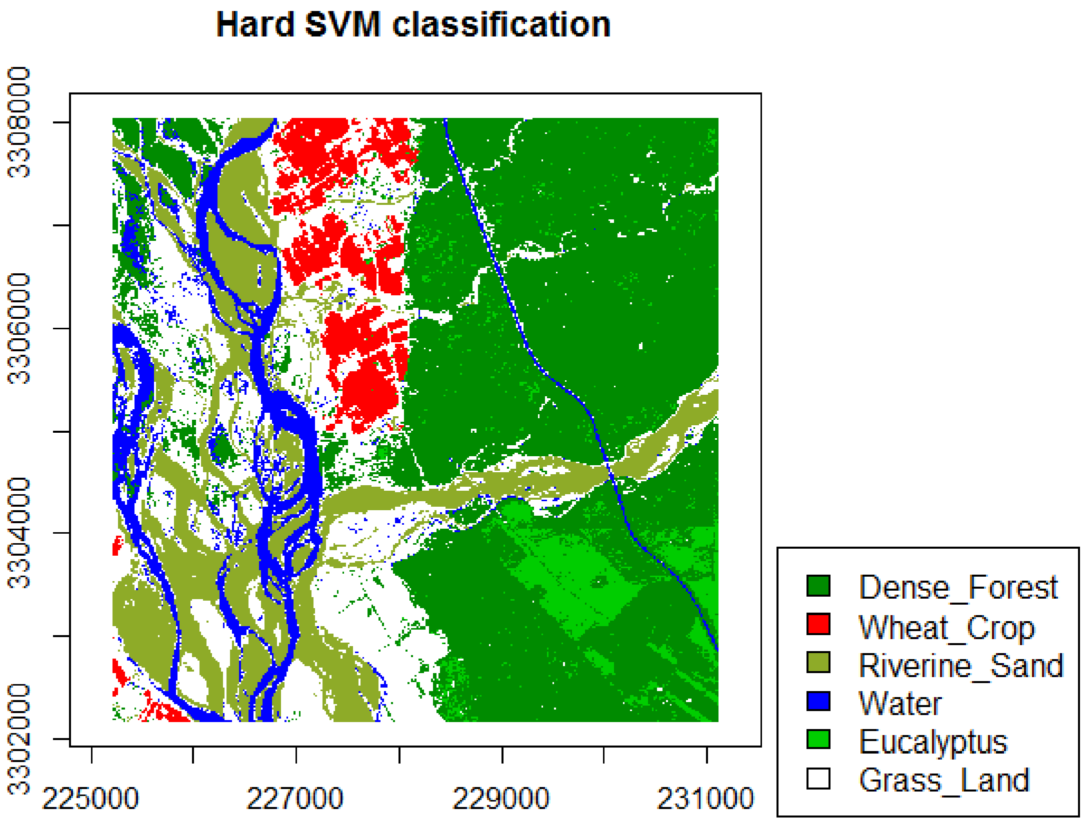


Figure A- 2. Hard Classified output of FORMOSAT-2 Imagery using Support Vector Machine

APPENDIX B

The classes identified on satellite image at the time of actual field visit is shown in Figure B-1.

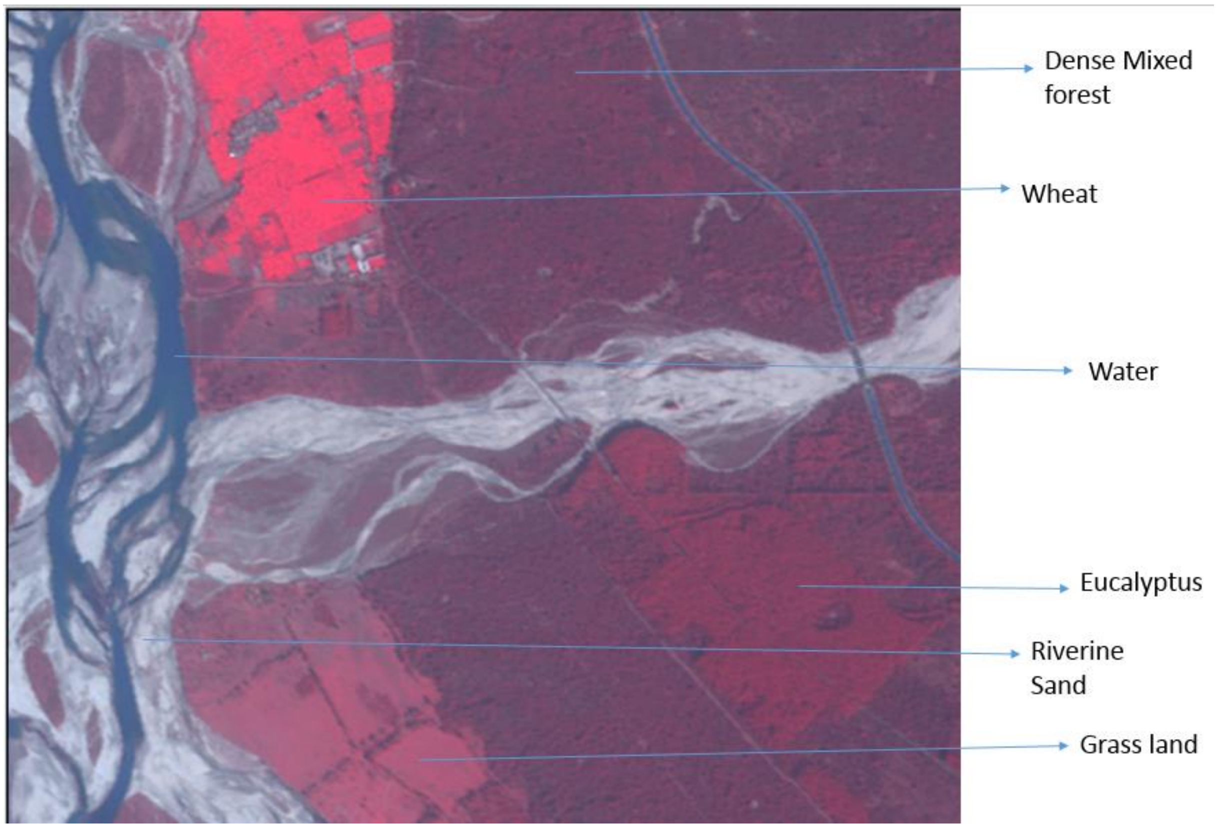


Figure B- 1. Identified Classes at the time of field visit.