DIMENSION REDUCTION METHODS FOR HYPERSPECTRAL IMAGERY

ONUR HALİLOĞLU

MAY 2022

ÇANKAYA UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF ELECTRONIC AND COMMUNICATION ENGINEERING

PH.D THESIS IN

ELECTRONIC AND COMMUNICATION ENGINEERING

DIMENSION REDUCTION METHODS FOR HYPERSPECTRAL IMAGERY

ONUR HALİLOĞLU

MAY 2022

ABSTRACT

DIMENSION REDUCTION METHODS FOR HYPERSPECTRAL IMAGERY

HALİLOĞLU, Onur

PhD in Electronic and Communication Engineering

Supervisor: Assoc. Prof. Dr. Orhan GAZİ

May 2022, 88 Pages

Hyperspectral Images has huge dimensions of data compared to single band or multispectral band images. This results from the fact that it contains hundreds of spectral bands with a high spectral resolution. Therefore, hyperspectral data processing, storing, and transmitting are critical issues to deal with. Additionally, it is a fact that required sample size for training a specific classification method increases exponentially with increasing number of spectral bands. In order to handle these problems, either the training data size has to be enlarged or dimensionality of hyperspectral images has to be reduced with some dimension reduction techniques. In this thesis, supervised and unsupervised dimension reduction methods are investigated, and some new methods are proposed. The proposed methods aim to reduce the dimensionality of the hyperspectral data before classification while preserving the classification accuracy as much as possible and to achieve reduced dimension with a low computational complexity.

Keywords: Hyperspectral image processing, dominant sets, band selection, dimension reduction, feature selection.

ÖZ

Hiperspektral Görüntüde Boyut İndirgeme Yöntemleri

HALİLOĞLU, Onur
Elektronik ve Haberleşme Mühendisliği Doktora

Tez Yöneticisi: Doç. Dr. Orhan GAZİ
Mayıs 2022, 88 Sayfa

Hiperspektral Görüntüler, tek bant ve çok bantlı görüntülere kıyasla çok büyük boyutlara sahiptir. Bu durum, hiperspektral görüntülerin yüksek çözünürlüklü yüzlerce spektral bant içermesinden kaynaklanmaktadır. Bu nedenle, hiperspektral veri işleme, depolama ve iletme üzerinde uğraşılması gereken önemli konulardır. Ayrıca spektral bantların sayısının artmasıyla belirli bir sınıflandırma yöntemini eğitmek için gerekli örnek boyutunun katlanarak (eksponensiyel olarak) arttığı bir gerçektir. Bu sorunlarla başa çıkmak için ya eğitim veri boyutu genişletilmeli ya da hiperspektral görüntülerin boyut büyüklüğü bazı boyut indirgeme teknikleriyle azaltılmalıdır. Bu tez çalışmasında eğitimli ve eğitimsiz boyut indirgeme yöntemleri incelenmektedir ve bazı yeni yöntemler sunulmaktadır. Sunulan yöntemler, sınıflandırma doğruluğunu mümkün olduğu kadar muhafaza ederek hiperspektral verinin boyutunu azaltmayı ve düşük hesaplama karmaşıklığı ile indirgenmiş boyuta ulaşmayı hedeflemektedir.

Anahtar Kelimeler: Hiperspektral görüntü işleme, baskın kümeler, band seçimi, boyut indirgeme, öznitelik seçimi.

ACKNOWLEDGEMENTS

I would like to thank to my supervisor Assoc. Prof. Dr. Orhan Gazi for his support, contributions, and valuable guidance.

I also would like to thank to my previous supervisors Prof. Dr. Behçet Uğur Töreyin and Assoc. Prof. Dr. Ufuk Sakarya for their guidance on hyperspectral image processing.

I wish to express my deep thanks to my parents Nurten and Cemal Haliloğlu for their significant support during my educational and professional career.

TABLE OF CONTENTS

STATEMENT OF NON-PLAGIARISM ..... **iii**
ABSTRACT ..... **iv**
ÖZ ..... **v**
ACKNOWLEDGEMENTS ..... **vi**
TABLE OF CONTENTS ..... Hata! Yer işareti tanımlanmamış.
LIST OF FIGURES ..... **ixx**
LIST OF ABBREVIATIONS ..... **xi**
CHAPTER I: INTRODUCTION **1**
1.1 STATEMENT OF THE PROBLEM 5
1.2 MOTIVATION AND OBJECTIVES OF THE STUDY 6
1.3 SUMMARY 6
CHAPTER II: REVIEW OF RELEVANT LITERATURE **8**
2.1 FEATURE SELECTION 9
2.1.1 Unsupervised Band Selection 10
2.1.2 Supervised Band Selection 13
2.2 FEATURE EXTRACTION 13
2.2.1 PCA-based Feature Extraction 14
2.2.2 Wavelet-based Feature Extraction 15
2.2.3 EMD-based Feature Extraction 17
CHAPTER III: PERFORMED WORK **19**
3.1 1D EMD 19
3.2 SUB-BAND DECOMPOSITION BASED SUPERVISED FEATURE EXTRACTION 22
3.3 HOW TO SELECT TRANSFORMATION TYPE, FILTER TYPE, AND WAVELET LEVEL? 29
3.4 AGRICULTURAL DATA CHARACTERISTICS 36

3.5 DOMINANT SETS BASED FEATURE SELECTION METHOD IN HYPERSPECTRAL IMAGERY 36
3.5.1. Dominant Set 43
3.5.2 Dominant Sets Based Feature Selection Method (DSbFSM) 44
3.5.2.1 First Stage of DSbFSM 44
3.5.2.2 Second Stage of DSM 47
3.5.3 Other Feature Selection Methods 49
3.5.4 Performance Evaluation and Results 50
3.5.5. The Importance of the Proposed Method in the Literature 55
3.6. DOMINANT SETS METHOD PROPOSED AS PREPROCESSING BEFORE FEATURE SELECTION 55
3.6.1 DSM utilized to increase classification performance of SFS (DSM-SFS) 49
3.7 DOMINANT SET BASED FILTERED FEATURE SELECTION METHOD (DSFM) 61
CHAPTER IV: CONCLUSION 66
REFERENCES 69

LIST OF FIGURES

Figure 1.1:	Panchromatic and Hyperspectral Images …….....	1
Figure 1.2:	Crop precision Classification Using PHI Image, Japan……..	3
Figure 1.3:	Mineral Identification with Hyperspectral Imagery……...	3
Figure 1.4:	Camouflaged vehicle detection in hyperspectral imagery…….....	4
Figure 2.1:	Processing chain of EMD study in…….....	19
Figure 3.1:	Algorithm of Proposed 1D EMD Feature Extraction Method……	22
Figure 3.2:	Sub-band Decomposition Based Supervised Feature Extraction Algorithm……	25
Figure 3.3:	Linear mapping of Classification Accuracies to Find out the Weights of Subbands……	29
Figure 3.4:	Comparative Results……..	31
Figure 3.5:	Wavelet (Daubechies 4) Low Pass Filter Output of SALINAS Data…	34
Figure 3.6:	Wavelet (Daubechies 4) High Pass Filter Output of SALINAS Data…	34
Figure 3.7:	Performance of Proposed Method and PCA Using Haar, Daubechies4, Daubechies10, Symmlet4, Symmlet10, and Coiflet4 Wavelets…….	36
Figure 3.8:	Performed Wavelet Packet Tree Structure……..	37
Figure 3.9:	Octave Band Decomposition Tree Structure……..	38
Figure 3.10:	DSbFSM Algorithm……	45
Figure 3.11:	Salinas A ground truth including 6 classes and unclassified regions….	55
Figure 3.12:	Ground truth of Pavia selected partition including 6 classes and unclassified regions…….	56
Figure 3.13:	Classification Accuracy vs Number of SVM training set samples performances of DSbFSM, CBS, and SFS for Salinas Data……..	58

Figure 3.14: Classification Accuracy vs Number of SVM training set samples performances of DSbFSM, CBS, and SFS for Pavia Data……. 59
Figure 3.15: F metric values of Spectral Bands for 'Brocoli Green Weeds' Class in Salinas Data…… 62
Figure 3.16: Classification Accuracy vs Number of SVM training set samples performances of DSM-SFS, CBS, and SFS for Pavia Data…… 64
Figure 3.17: Classification Accuracy vs Number of SVM training set samples performances of DSM-SFS, CBS, and SFS for Indian Pines Data……. 65
Figure 3.18: Classification Accuracy vs Number of SVM training set samples performances of DSM-SFS, CBS, and SFS for Salinas Data……. 66
Figure 3.19: Classification Accuracy vs Number of SVM training set samples performances of DSFM, CBS and SFS for Pavia Data…….. 69
Figure 3.20: Classification Accuracy vs Number of SVM training set samples performances of DSFM, CBS and SFS for Indian Pines Data……... 70

LIST OF ABBREVIATIONS

CBS	:Correlated Band Selection
DSbFSM	:Dominant Sets based Feature Selection Method
DSFM	:Dominant Set Based Filtered Feature Selection Method
DSM	:Dominant Sets Method
DWT	:Discrete Wavelet Transform
EMD	:Empirical Mode Decomposition
GPU	:Graphics Processing Unit
IMF	:Intrinsic Mode Function
NATO	:North Atlantic Treaty Organization
NFL	:No Free Lunch
PCA	:Principal Component Analysis
PSO	:Partical Swarm Optimization
RGB	:Red Green Blue
SFS	:Sequential Forward Selection
SNR	:Signal to Noise Ratio
STFT	:Short Time Fourier Transform
SVM	:Support Vector Machine
SWIR	:Short Wave InfraRed
TIR	:Thermal InfraRed
VNIR	:Visible Near InfraRed
VSWIR	:Visible Short Wave InfraRed
WT	:Wavelet Transform

CHAPTER 1
INTRODUCTION

Hyperspectral imaging started with the development of a two-dimensional imager that captures spectral data with a large number of bands in the mid-1980s [1]. Hyperspectral imaging is based on the fact that all materials reflect, absorb, and emit electromagnetic energy at specific wavelengths according to their molecular decomposition. In hyperspectral imaging, the data is gathered in hundreds of contiguous narrow spectral bands with high spectral resolution and the evaluation of responses in each band defines the spectral resolution for the material. Thus, distinct materials have distinct spectral signatures.

In order to determine what the hyperspectral image data looks like and how it differs from panchromatic and multispectral images, Figure 1.1 can be beneficial. Single band image has two dimensions whereas hyperspectral imager has two spatial dimensions and one additive spectral dimension. Panchromatic image has a single, wide-spectrum band. On the other hand, hyperspectral image has many contiguous narrow spectral bands.

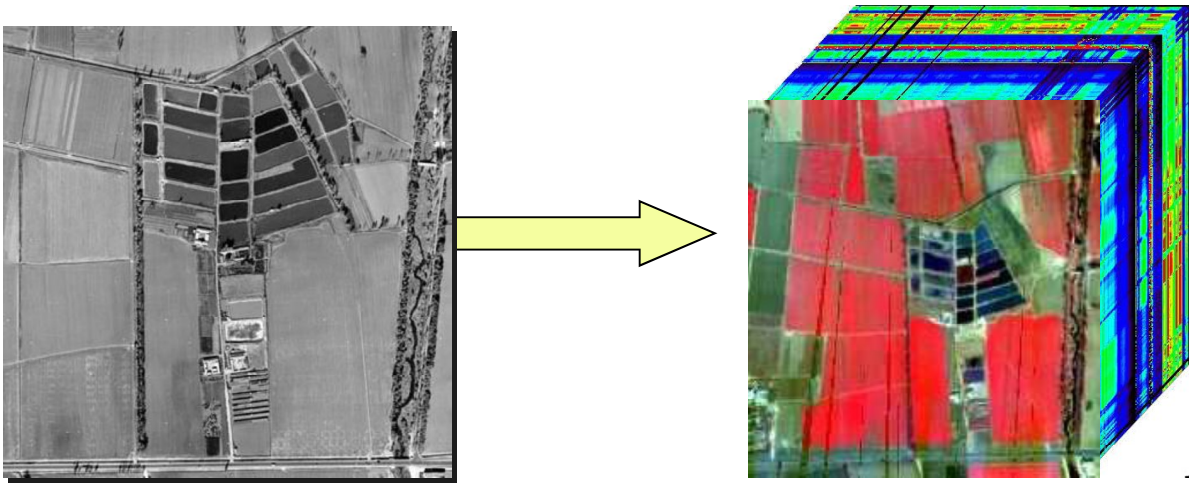

Figure 1.1: Panchromatic and Hyperspectral Images

With the advance in technology, the interest to hyperspectral imaging has been increased since its very high spectral resolution leads to provide material identification. Hyperspectral image classification is realized to discriminate materials or regions that have distinct spectral signatures. In some cases, panchromatic and multi-spectral imaging cannot have sufficient spectral resolution. Therefore, as a high spectral resolution imaging, hyperspectral remote sensing has been used over a wide range of applications that has grown especially for the last 30 years. Some of these applications can be mentioned as:

- Material Identification
- Agriculture
- Forestry
- Geology
- Ecological Monitoring
- Disaster Monitoring
- Camouflage and Concealment
- City Planning and Real Estate
- Biological and Chemical Detection
- Health Care

In Figure 1.2 [2], an example of the agriculture application of hyperspectral image classification is shown. The observation land located in Japan is classified with respect to nine distinct agriculture products.

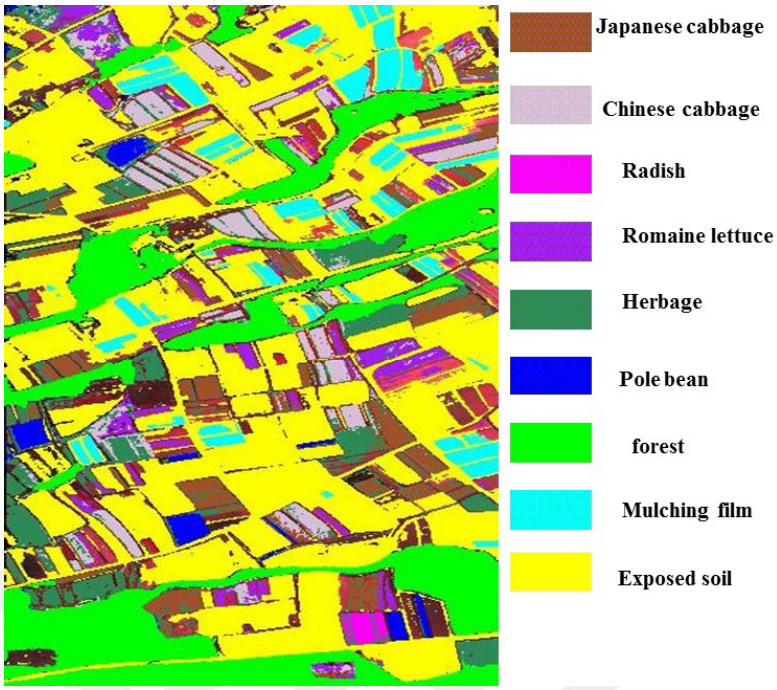

Figure 1.2: Crop precision Classification Using PHI Image, Japan [2]

As an example of hyperspectral remote sensing application in mining and oil industry, Figure 1.3 [2] indicates mineral structure observation on the surface of land.

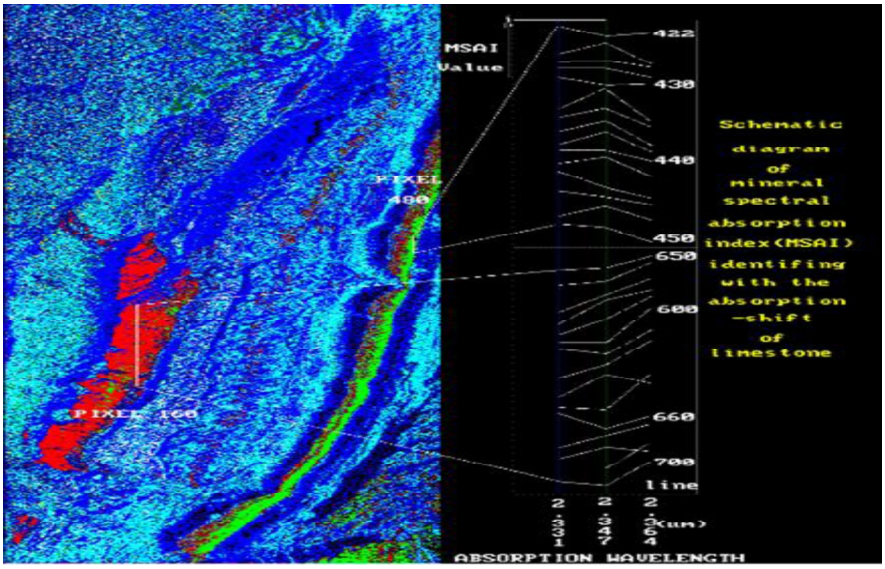

Figure 1.3: Mineral Identification with Hyperspectral Imagery [2]

Hyperspectral images can also be used for military applications as shown in Figure 1.4 [2]. A camouflaged vehicle cannot be detected from panchromatic and

multispectral images whereas it can be detected from hyperspectral image constructed by some of the valuable bands over hundreds of bands.

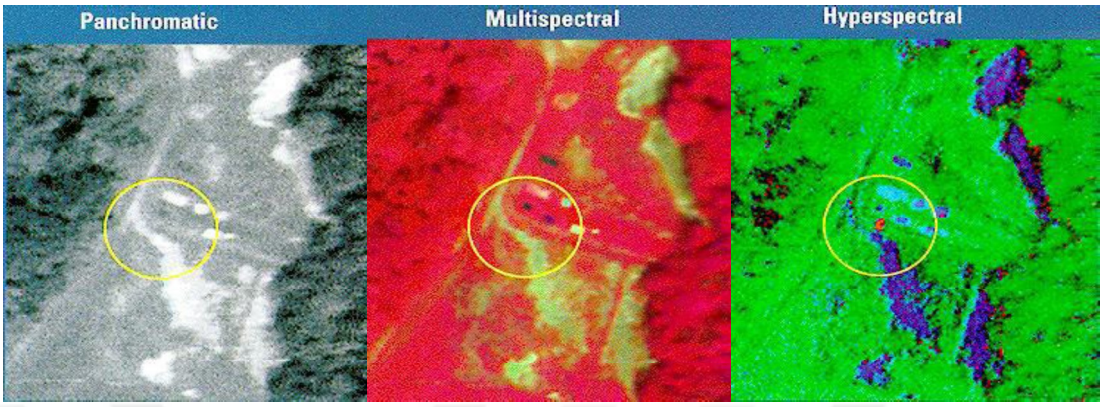

Figure 1.4: Camouflaged Vehicle Detection in Hyperspectral Imagery [2]

There are two widely used hyperspectral data sources: airplane and satellite hyperspectral sensors. There are many hyperspectral sensors working on aerial platforms such as AVIRIS, CASI, HyperCAM, Hymap etc. [3,4] Although aerial imagers have higher ground sampling distances with respect to the satellite imagers, their swath widths are lower, and one can image all parts of the world by the satellite imager without any limitation. Therefore, satellite imagers have crucial advantages over aerial imagers.

There are five main current hyperspectral satellite sensors on orbit announced officially: MODIS, CHRIS, HYPERION, ENMAP, and PRISMA. The first three of these sensors are out of their design lifetime. Therefore, these imagers tend to indicate performance decrease. MODIS is an American imager which has 36 spectral bands in the 0.4-14.4 um spectral region with a 250m ground sampling distance [5,6]. CHRIS is the light-weighted 14 kg imager working in the VNIR region with 11nm spectral resolution and 17m spatial resolution [6]. On the other hand, HYPERION can take images both in the VNIR region and also in the SWIR region. Its spectral resolution is 10nm and its spatial resolution is 30m [7]. [8] ENMAP, a German hyperspectral sensor has launched in 2022. ENMAP operates nearly the same spectral region as PRISMA (420nm - 2450nm). The mean spectral resolution is 6.5nm for VNIR and 10nm for SWIR with a 30m spatial resolution. The objective of the mission is to provide high quality images for agriculture, forestry, coastal zones, and inland waters.

resolution. The objective of the mission is to provide high quality images for agriculture, forestry, coastal zones, and inland waters.

[9] PRISMA the European hyperspectral imager program started to be developed in 2007. The interest area of the imager is the whole European and Mediterranean region. The application areas of the sensor are vegetation, environmental protection, and the climate change. Its orbital altitude will be 700km and the design lifetime will be 5 years. The imager is operating in visible, near infrared, and short-wave infrared regions with 10nm spectral resolution. The mass of the imager is below 90 kg. It has more than 210 contiguous bands. The satellite was launched in 2019.

There is a well-known hyperspectral satellite under development called HyspIRI [10]. MODIS will be substituted with HyspIRI. Its application areas are ecosystem monitoring, vegetation and natural disasters monitoring such as volcanoes, wildfires, and drought. HyspIRI has two instruments one for VNIR and SWIR (VSWIR) spectrums and the other for mid and thermal infrared (TIR) spectrums. It VSWIR instrument has 220 bands with 10nm spectral resolution whereas its TIR instrument has 8 bands. The ground sampling distance of the sensor is 60m at 700km orbital altitude.

1.1 STATEMENT OF THE PROBLEM

The area covered by the single hyperspectral imager is very large compared to the size of the several materials within the swath of the single pixel. Therefore, the response from the single pixel is a mixture of several materials. Moreover, the spectral signature of the same material can be varied due to atmospheric conditions, noise in the imager, and the material mixture effect. Especially, mixed material effect cannot be matched with a single well-defined material. Supervised classification methods require a-priori information such as training data. However, the cost of forming a training data is high. Therefore, unsupervised classification algorithms are required since they do not have dependence on prior information about spectral signatures.

Hyperspectral Imagery has much richer information than multispectral and single-band imagery since it usually includes hundreds of contiguous bands. Therefore, hyperspectral images include rich and fine spectral information. On the other hand, huge data volume in hyperspectral imagery results in data storage and data transmission problems. Processing time of this huge data should also be considered. Additionally, it is

a fact that required sample size for training a specific classification method increases exponentially with increasing number of spectral bands. This is known as curse of dimensionality [11]. Curse of dimensionality leads to Hughes phenomenon [12]. If the required training data size for a specific classifier is insufficient, then the estimation of the parameters used to calculate classification accuracy becomes inaccurate and unreliable. To handle this problem, either the training data size has to be enlarged or dimensionality of hyperspectral images has to be reduced with some dimension reduction techniques. The former, enlarging the training data size, is usually not possible for hundreds of spectral bands in practice. Therefore, dimension reduction is a key issue while dealing with hyperspectral image processing.

The problem that will be studied throughout this thesis is that "Can novel dimension reduction methods for hyperspectral satellite imagery be devised that outperform state-of-the-art techniques with a low computational complexity? Another problem to be investigated is that "Is there any way to reduce the dimensionality of hyperspectral imagery data without much degradation in classification performance?"

1.2 MOTIVATION AND OBJECTIVES OF THE STUDY

Today, hyperspectral imaging is very popular not only due to its several advantages over panchromatic or multi-spectral imaging but also because of its virginity. Application areas of hyperspectral imaging increase day by day. NATO leads a series of meetings about hyperspectral image classification in the last years [13]. This shows us that, there is much to explore in hyperspectral imaging and a great contribution to the literature can be provided with the proposed study.

The main objective of the study is as follows.

- Proposing new dimension reduction methods for hyperspectral imagery that reduces the data as a preprocessing method before classification while maintaining the classification performance as much as possible

1.3 SUMMARY

The rest of the proposal is organized as follows. Chapter 1 presents the introduction, hyperspectral image applications, hyperspectral data sources, the statement of the problem, motivation of the study, the objectives of the study, and organization of the

study. Chapter 2 is a review of relevant literature. Chapter 3 presents the methodology to be used in the study. Chapter 4 presents the summary of the thesis study.

CHAPTER 2

REVIEW OF RELEVANT LITERATURE

Hyperspectral classification problems can be summarized as extreme number of spectral bands, the cost of correct classification, data quality, and the variety of the spectral signature in the spatial axis.

In literature, Kernel based methods are widely used to convert the data in the nonlinear feature space into the high dimension data in the linear feature space. Support Vector Machine [14], Kernel Fischer Discriminant [15], Relevance Vector Machine [16] are the major Kernel based classification methods.

The hyperspectral image classification can be divided into two categories, namely, supervised and unsupervised, based on the dependence on apriori information.

Supervised classification depends on the apriori information about spectral signatures or any pre-classification result. In [17], a new method for classification of hyperspectral images for extracting cartographic object is developed by using the combination of spatial and spectral information. Empirical mode decomposition technique is also used for combining spatial and spectral information to reduce the varieties in the class and increase the varieties between different classes [18]. This method separates the data into local frequency components.

Unlike supervised methods, unsupervised methods do not require apriori information or training data. In [19], independent component analysis method is proposed for hyperspectral image analysis. Moreover, in [20] K average method is used to reduce the effect of spectral signature variance in the same class and to increase the success rate. K-average method is applied to linearly separable hyperspectral data. In this study, graph-based methods can be applied to non-linearly separable hyperspectral data.

It is stated in the introduction section that dimension reduction techniques are applied before classification to reduce the data size and to reduce the dependence of huge number of training samples. In case of hyperspectral images, dimension reduction can be realized in two ways; feature selection and feature extraction. In feature selection techniques, a subset from the original input data which is optimum for the decided criteria is selected. On the other hand, in feature extraction methods, input data is transformed to a different space where a set of new features are produced without sacrificing the discriminative ability much.

2.1 FEATURE SELECTION

Band selection methods, called as feature selection methods, work with a subset of the original bands without changing the physical meaning.

Since noise variations in redundant data and unrelated spectral information to the classified feature lead to deterioration in classification, contrary to what is believed, the accuracy of classification do not have to increase as the number of hyperspectral bands increases. Therefore, band selection is important for the success of hyperspectral classification.

Some crucial aspects in terms of operating spectral range should be considered. Water absorption bands 1400-1900nm are unused and chlorophyll absorption bands i.e., 450nm in blue and 650nm in red must be considered during calibration because of their low reflectance [21].

Some of the most popular band selection metrics can be described as:

- *Information entropy metric* calculates the entropy measure of each band. When the entropy value is high, it means that amount of information in the data is large.
- *First Spectral Derivative metric* evaluates the bandwidth variable as a function of added information and looks for the first spectral derivative of each band. Bands that have similar metric can be eliminated and bands that differ a lot based on metrics should be preserved.
- *Second Spectral Derivative metric* evaluates the bandwidth variable as a function of added information and looks for the first spectral derivative of each band. The larger the deviation from a linear model, the higher the information value of the band.

- *Spectral Ratio metric* calculates the band ratios with respect to the average value of ratios.
- *Correlation metric* calculates the correlation of all bands as pairs and selects the most uncorrelated bands.

Band selection methods can be divided into two main categories: unsupervised band selection and supervised band selection methods.

2.1.1 Unsupervised Band Selection

To provide computational efficiency by using a subset of original hundreds of bands, unsupervised band selection methods can be used when the related object information is unknown. This is because of the fact that unsupervised band selection methods provide reliable solutions regardless of the types of objects to be classified.

Unsupervised band selection methods are computationally efficient. Moreover, they do not need registration, geo-referencing, and much pre-processing.

Some unsupervised band selection studies in the literature are explained below. In [22], a clustering type band selection approach based on statistical and geometrical characteristics of hyperspectral images. Firstly, hyperspectral bands are clustered based on three metrics such as city block metric, Euclidian distance, and cosine distance. Then, the band among the clustered band group having the maximum mean deviation is selected since the selected band possesses the maximum mutual information with respect to other bands within the cluster. In order to define the minimum number of bands that can provide the sufficient information for classification, virtual dimensionality concept is used. The main idea is to provide not only minimum differentiation in the clusters and but also maximum differentiation among separate clusters. A well-known 224 band AVIRIS Cuprite image with a 10nm spectral resolution is used for the proposed method. Water absorption and low SNR bands including 1-3,105-115, 150-170 were extracted. After band removal, 189 bands are used in experiments. By using virtual dimensionality concept, the minimum number of bands to be selected is considered as 22 for this application. It is sated that mean average deviation method based on city block metric preserves the sufficient information for band selection while reducing the data to a considerable level with fast implementation.

In [23], N-FINDR technique is applied to determine which pixels to use before linear prediction method used for band selection. After selecting several pixels from the whole pixels, it is not possible to apply data whitening. However, it is possible to use the whitened versions of the selected pixels. The number of pixels to be selected is determined as one less than the number of bands to be selected for feasible least squares solution. GPU based parallel processing is performed only for pixel selection since spatial dimension is larger than the spectral dimension. That is why it was not applied to band selection part. AVIRIS Cuprite image is used for the proposed method. Water absorption and low SNR bands were removed. After band removal, 189 bands are used in experiments. Constrained linear discriminant classifier method is applied for classification. Data whitening is applied to 189 bands before band selection, but the selected bands are the original bands instead of the whitened ones. N-FINDR algorithm mainly tries to select some pixels from the whole set which can build up a simplex with the maximum volume. The proposed method reduces computational complexity by using N-FINDR pixel selection method and GPU based parallel processing while maintaining classification performance.

In [24], a further improvement based on a collaborative sparse model is developed over linear prediction band selection method. The faster NFINDR+LP method requires sparse linear regression coefficients. Removed bands share same non-removed bands set. This sparse model can also be used for defining the minimum number of selected bands. AVIRIS Cuprite image is used for the proposed method. Water absorption and low SNR bands are removed. After band removal, 189 bands are used in experiments by applying constrained linear discriminant classifier method for classification. Band similarity is calculated jointly instead of pair wisely. Firstly, the best two dissimilar bands are selected and new dissimilar band with the selected bands is found next. This process continues until the predefined number of bands is reached. The similarity between a band and multiple bands is computed by using linear prediction method. However, linear prediction method is computationally complex while considering all the pixels. N-FINDR method can be performed for reducing the number of pixels to be used in band selection. In the first part of this study N-FINDR+LP method is performed for pre-band selection and in the second part linear sparse regression algorithm is used to improve the classification

accuracy with the same number of bands. The proposed method results in an improvement in the classification accuracy.

In [25], a soil electrical conductivity ground measurements data set in agriculture domain is selected as an application area. According to no-free-lunch (NFL) theorem [26], particular supervised method cannot be the best performing method in all data sets. That is why, all combinations of seven unsupervised band selection and three supervised classification methods are performed to find the best performing methods pair that maximizes classification accuracy and minimizes computational complexity. Unsupervised methods can be listed as information entropy, first and second spectral derivative, spatial contrast, spectral ratio, correlation, and principal component analysis. On the other hand, supervised methods can be listed as regression, regression tree and instance based supervised methods. As input for the band selection and classification methods, images of a 120 channel, prism grading push broom sensor were used. The images were obtained from an aerial camera in the range of 1200 to 4000m. Since 400-900nm band range is related with agricultural applications, the sensor used in this study works in the spectral range of 471nm to 828nm. Spectral resolution of the imager is 3nm and the spatial resolution of the imager is 1m. Radiometric calibration including sensor noise and illumination, geometric calibration, and geo-registration are applied to the images before processing. Atmospheric corrections are not considered. It is surprisingly achieved that only at most 4-8 bands including 7 bands in the red spectrum, 4 bands in the near infrared spectrum, and 2 bands close to the border of blue and green spectra are sufficient for the electrical conductivity.

In [27], the aim of the study is to select the most distinctive and informative bands with unsupervised band selection algorithms based on band similarity measurement. The experimental results of the study indicate that similarity based algorithms achieves better success with respect to widely used techniques in terms of information conservation and class separability. Usually, similarity metrics are calculated for each pair of bands, however in this study they are calculated jointly instead of pair wisely. To determine jointly selected bands among large number of bands with less computational time sequential forward search algorithm is utilized. The algorithm firstly finds the best two band combination and increases the number of bands until it reaches the desired number of bands. The similarity-based band selection method does not require predetermined

number of bands. When all the pixels in a band are used for band selection methods, the size of the matrices used in band selection methods becomes very large. Additionally, it is observed that using a small part of the whole pixels does not affect the classification performance much since high spatial correlation exists in each band image. Therefore, 10% or 1% of the pixels is used in band selection methods. linear prediction method is mathematically identical to orthogonal subspace projection method. However, computational complexity is higher in the orthogonal subspace projection method. Therefore, linear prediction method is presented in this study. Compared to classification with the original data, classification with the whitened data has better performance. Linear prediction based band selection method has the better classification accuracy among the other methods in both supervised and unsupervised classification techniques.

2.1.2 Supervised Band Selection

Supervised band selection methods require a training data. Firstly, hyperspectral data is calibrated and this calibrated data is registered with ground measurements to form the training data. Moreover, it is required to have a greater number of examples in a training set than the number of bands in hyperspectral data.

In [27], a supervised band-selection algorithm is proposed by only taking the known class signatures into account. Since the original bands and band combinations are not taken into account, the processing time is much faster than traditional methods. The results indicate that classification performance of the proposed method is better than the traditional methods.

In [28], the proposed method is a particle swarm optimization (PSO) based supervised band selection algorithm. Class signatures which are known previously are used instead of original bands or class training samples. In order to get rid of the high PSO computation time, parallel processes are implemented by GPUs that can lead to higher speeds.

2.2 FEATURE EXTRACTION

Feature extraction methods usually change the physical meaning of the original data since they are dealing with combinations of the original data. Therefore, unlike band

selection, feature extraction can lead to deterioration of the critical data during data transformation.

Reduced number of features may result in a decreased clustering power and classification accuracy. Therefore, preserving the significant information is crucial in dimension reduction methods.

As in feature selection, there are supervised and unsupervised feature extraction algorithms. PCA-based, wavelet-based and empirical mode decomposition (EMD) based feature extraction techniques are investigated among dimension reduction methods.

2.2.1 PCA-based Feature Extraction

PCA is a linear transformation of the original signal into a feature vector by means of an orthogonal set of basis vectors. These vectors are the covariance matrix eigenvectors that minimize the mean square error. Thus, Principal Component transformation accounts for the covariance matrix. It can be explained that PCA converts the correlated variables to uncorrelated one via linear transformation such that first principal component has the largest variance. Then, most probably the first few components include most of the information about class discrimination. That is why; first few of the components of PCA are usually used to reduce the size of the data to be processed.

Although PCA is used as a common dimensionality reduction technique in hyperspectral imaging, it is not very successful in case of low interclass variation.

The only constraint for the PCA basis vectors is orthonormality and it does not account for localization information of bands. Besides, a few of the principal components are considered in most of the implementations. Therefore, PCA used for dimension reduction does not include all the information data. However, PCA-based feature extraction methods can benefit from the PCA advantages while compensating its disadvantageous with the help of other techniques.

In [29], it is investigated that PCA applied on HYDICE and AVIRIS hyperspectral data sets reduces the original data size while maintaining much of the information. In [30], PCA feature extraction algorithm demonstrates a minimal negative effect on hyperspectral target detection. In [31], the hyperspectral data is divided into several parts and hierarchical PCA algorithm is applied to each part of the data then few principal components are combined to be classified. Results of [31] indicate that the method

realizes a remarkable reduction in data size without much degradation in classification accuracy.

2.2.2 Wavelet-based Feature Extraction

Unlike PCA, Wavelet Decomposition coefficients contain all the info about the original signal. Moreover, wavelet-based feature extraction methods give local information about spectral variation of a hyperspectral signal in separate bands and at each scale. Therefore, while using Wavelet Transform (WT) it is possible to analyze data at different scales and resolutions since WT decomposes the original signal into a series of shifted and scaled versions of the determined mother wavelet.

Wavelet transform coefficients can be defined as a large library of orthonormal basis including various distinct time frequency-atoms. Wavelet Transform contains both the variations in amplitude and spectral info. Wavelet decomposition has two stages: filtering and down sampling.

Low pass filter outputs of the wavelet transform are called wavelet approximation coefficients whereas high pass filter outputs of the wavelet transform are called wavelet detail coefficients. Low pass filter outputs represent the large scale information and the high pass filter outputs represent the fine scale information.

Wavelet Transformation was used for image classification in photogrammetry and remote sensing areas [32]-[33]. Wavelet decomposition is applied to hyperspectral data several times in the literature.

In [34], Discrete Wavelet Transform (DWT) is performed for feature extraction and extracted features are evaluated by automated maximum likelihood classifier for their classification accuracy in two agricultural fields. Simplest DWT filter Haar which has 2 coefficients is used for DWT. 1D DWT is applied to spectral domain. Individual wavelet coefficients are evaluated with respect to their class discriminating capability and subset of the best discriminating coefficients are used as features. Subset of 1D best wavelet coefficients is used and it is examined that the technique performs better than PCA. The results indicate that larger scale wavelet coefficients are more useful and utilizing these coefficients gives better results than observing discrete spectral bands.

In [35], 1D Automatic Discrete Wavelet Transform is applied in the spectral domain for dimension reduction. This technique is compared with PCA and it is observed

that classification accuracy of the former is better than the latter. Automatic decomposition level is determined by calculating the correlation between the original data and the decomposed sub-band in the specified decomposition level on the training data. After that, the determined decomposition level is applied to the whole data. As a result, 1D Wavelet Decomposition applied for dimension reduction has higher classification rate than PCA while preserving its simplicity advantage over PCA.

In [36], matching pursuit method which uses a greedy strategy to find an adaptive basis iteratively from wavelet packets dictionary is proposed as a feature extraction technique for hyperspectral imagery. Moreover, other dimension reduction techniques based on wavelet transform are implemented and compared with the matching pursuit method. In this work, training data is used to find the effective and usable features for classification. Both linear wavelet feature extraction which only uses approximate coefficients of wavelet transform and nonlinear wavelet feature extraction which accounts for the approximate and detail wavelet transform coefficients based on entropy of the energy distribution or discriminant measure are implemented. In the Matching pursuit procedure, selected features are determined from the wavelet dictionary one by one in order to provide the best fitting with the original signal. At the beginning of the method, matching pursuit projects the original signal on a vector such that the vector is determined to minimize the residue. Then the residue is projected on a series of orthogonal vectors (wavelets) iteratively. Finally, M vectors that best fits with the original signal and have minimum overall residue are selected as features for the hyperspectral data. When compared with linear and nonlinear wavelet feature extraction methods implemented in this study, Matching Pursuit technique that uses projection vectors from a wavelet dictionary gives better results in terms of classification accuracy.

In [37,38], best wavelet packet basis which is an adaptive wavelet decomposition algorithm is performed as feature extraction method for hyperspectral imaging. Best wavelet packet basis is determined by minimizing the cost function based on between-class distances of the training data. It is declared that best wavelet packet basis method preserves the classification accuracy while reducing the computational complexity.

2.2.3 EMD-based Feature Extraction

As in DWT, the intrinsic information of the data can be decomposed into several components in Empirical Mode Decomposition Method. These several components are called as intrinsic mode functions (IMFs). Each IMFs resemble to a band pass filter banks [39]. First IMF is the one having the center in the origin, so it is a high pass filter. Therefore, IMFs can be described as oscillatory modes at narrow range of frequencies. It is observed that lower order IMFs indicates the spatial structure of the data more with respect to higher orders. The number of IMFs is lower than the logarithm base 2 of the signal. Thus, expressing the data with empirical mode decompositions is reducing the dimensionality. The number of IMFs depends on both the length of the signal and deviations from the average of the signal. One advantage of EMD is that it can be proposed for nonlinear and/or non-stationary signals.

EMD is suitable for nonlinear and nonstationary signals since it is adaptive and it gives info about the local time characteristics of the data. Fourier transform assumes the system is linear and stationary however it is not practically correct. In wavelet transform, if the sampling frequency and decomposition level are constant, frequency localization is fixed. On the other hand, EMD has adaptive frequency localization based on local signal properties. Therefore, EMD has many advantages over Fourier transform and Wavelet transform.

As a result of the advantages of EMD in signal processing it is used as a feature extraction method in hyperspectral imaging.

In [40], Empirical Mode Decomposition Mode is proposed as a preliminary dimension reduction technique in order to decrease the dimensionality before further feature extraction and classification. This technique is performed on agricultural crop area in which discrimination between classes is really hard. AVIRIS [41] data is used for the proposed method. Data whitening is applied before EMD in order to remove the correlations among spectral channels. Thus, EMD can result in higher performance since the features become independent. The processing chain of the study is given below in Figure 2.1. After the data is whitened, the mean operation is performed on the training samples belonged to separate classes. Therefore, one signal with the size of number of bands is obtained for each separate class. EMD is performed on these class's mean signals. Then the original hyperspectral data is transformed based on the IMFs of the

class's mean signals. Data is dimensionally reduced since it is expressed in terms of IMFs of separate class's means. Finally, dimension reduction method DAFE is applied on the pre-reduced data. It is dedicated that the preliminary EMD dimension reduction method improves the performance of the further feature extraction methods regardless of training data size. The amount of performance increases reduces with increasing training data size. Therefore, when sufficient training data is acquired, there is no need for preliminary dimensionality reduction.

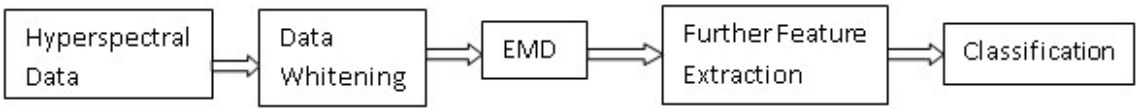

Figure 2.1: Processing chain of EMD study in [40]

In [42], Empirical Mode Decomposition of hyperspectral data is proposed to increase the performance of classification accuracy while using SVM classification. 2D EMD is applied to each spectral band. Two EMD approaches are mentioned, one is summing the first two IMFs and the second is combining the first two IMFs with respect to composite kernel logic. Indian Pine and DC Mall Data are used for the proposed methods. 191 of the DC Mall data over 210 bands and 200 of the Indian Pine data over 220 bands are used due to the removal of water absorption and noisy bands. It is stated that 2D EMD based techniques improve the SVM classification accuracy in hyperspectral data.

CHAPTER 3

PERFORMED WORK

To investigate dimension reduction methods before dealing with hyperspectral classification methods, literature survey is realized. First of all, feature selection methods are investigated, and a wavelet-based band selection method is tried to be implemented. However, the results and the implementation were not very successful. Then, feature extraction methods are investigated most of which are summarized in the previous section. After realizing a literature survey on dimension reduction techniques and hyperspectral imaging, a survey paper is published [43] in RAST 2013 which is added in the Appendix section.

After the survey, the studies were concentrated on feature extraction methods and two feature extraction methods are implemented. The implemented methods are 1D EMD and supervised sub-band decomposition based feature extraction methods which are described in the following parts. The wavelet-based one-dimensional sub-band decomposition applied before principal component analysis (PCA) is proposed to increase the classification accuracy [44].

In this study, band selection methods are investigated besides feature extraction methods. Novel approaches based on dominant sets are presented. Dominant set technique is first applied to hyperspectral data [45]. Dominant sets-based dimension reduction method is applied on hyperspectral data as a band selection technique. Moreover, a distinct dominants sets based study is implemented as a pre-band selection method [83].

3.1 1D EMD

2D EMD dimension reduction techniques applied to the spatial domain of the hyperspectral data are used in [42],[46], [47]. In [40] and [46] EMD is applied as a

preliminary feature extraction method for further dimension reduction while maintaining the classification accuracy. It is estimated that applying 1D EMD to the spectral domain of the hyperspectral data as a preliminary dimensionality reduction, the performance of further feature extraction method can be increased. In order to achieve this, the preliminary dimension reduction method has to maintain the valuable data during reduction. Therefore, it is decided to perform 1D EMD before PCA to increase the classification performance and to reduce the data size more.

In this work, AVIRIS Indian Pine image [41] is used as an input to the 1D EMD feature extraction method. Indian Pine data has 145x145 pixels with 224 spectral bands. First of all noisy and water absorption bands are eliminated so the spectral band number is reduced to 200. 1D EMD is applied to every spectral band i.e., every 145x145 pixel distinct spectral image. A list of intrinsic mode functions (IMFs) is formed and only sum of the first three of the IMFs which are believed to maintain most of original information are contributed to the remaining part of the study. Then, PCA is applied to each IMF sums. Only the first 8 of the components of PCA are taken into account. Finally, the output of PCA is subjected to SVM classification. For SVM classification LIBSVM [48] and for EMD implementation the code in [49] is used in this study. Ground truth of Indian pine data for supervised SVM classification is also available. The unclassified parts and small classes that contain small amount of samples are eliminated from the ground truth in this work. 100 training samples are defined randomly for each nine classes. This training data set is utilized for SVM to find out the SVM classification parameters. Classification is repeated five times and the average of the classification accuracies are accepted as a performance parameter. The algorithm of the work is indicated in Figure **3.1**3.1. To evaluate the performance of the algorithm, the results are compared with the results of direct PCA applied before SVM classification.

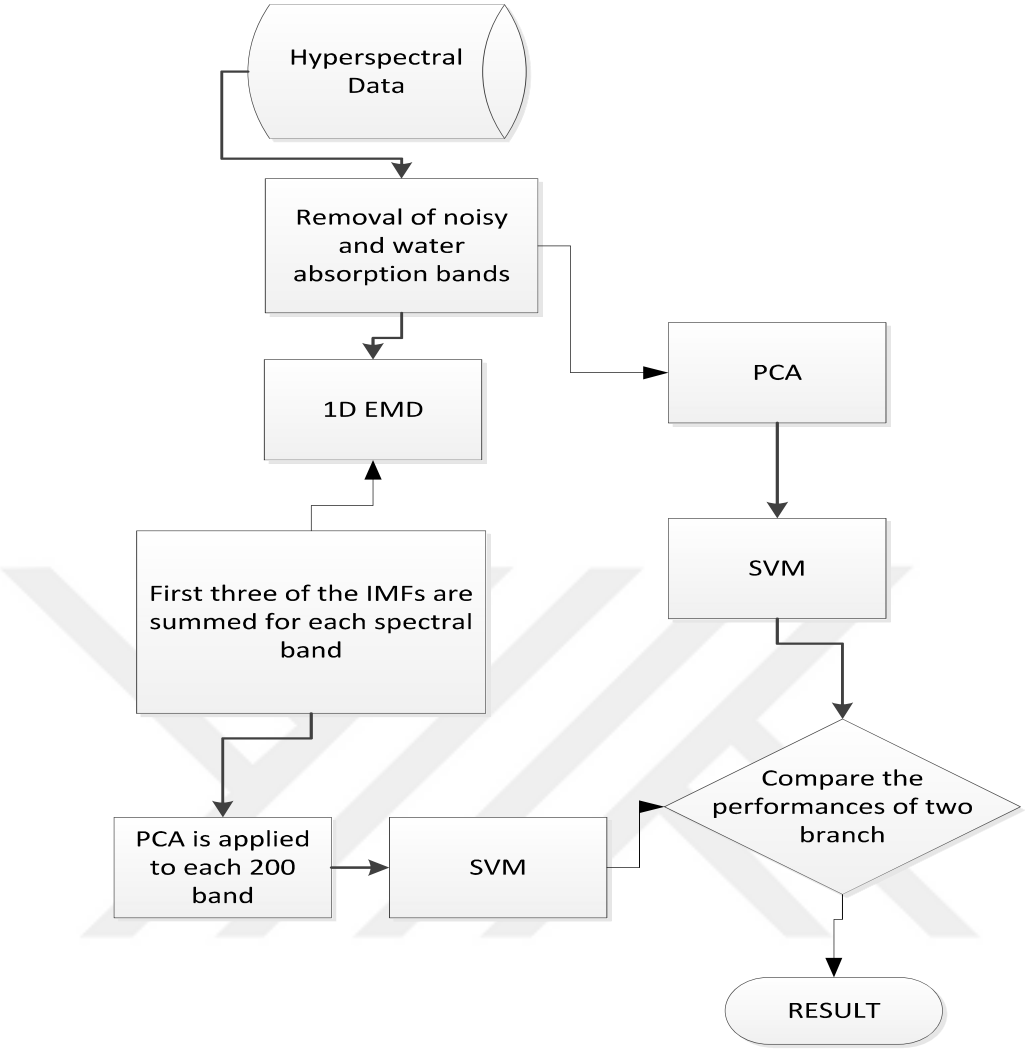

Figure 3.1: Algorithm of Proposed 1D EMD Feature Extraction Method

Results of 1D EMD applied to the spectral domain of hyperspectral data before PCA indicates that classification accuracy is decreased by 1D EMD. The classification accuracy of the direct PCA branch followed by SVM classification is about 0,80 whereas the classification accuracy of 1D EMD branch in which sum of first three IMFs are taken into account is about 0.58. Moreover, when a smaller number of IMFs are taken into account, the classification performance increases. For instance, if only first IMFs are considered then the classification accuracy increases to 0.68. In [42] it was stated that 1D EMD deteriorates the classification performance but the reason of it was not explained.

EMD approach determines the maximum and minimum values and constructs two separate functions for determined maximum points and minimum points. The mean function is calculated between maximum point function and minimum point function. Calculated mean function is subtracted from the overall signal. The algorithm works iteratively on residue signal to minimize the final residue under the defined threshold. Therefore, it can be expressed that IMFs can continue the variations of the original signal with significant changes in the mean value only. If all the hyperspectral data or an image selected from the hyperspectral data for a distinct band is subjected to EMD as in 2D EMD, the change in the mean value is not very important since the variations in the image are protected and the pixel amplitude values are somehow normalized with respect to each other. However, in 1D EMD the situation is different. In 1D EMD applied to every spectral band, each pixel is expressed by IMFs and the mean loss can be different for every pixel. Therefore, the amplitudes of spectral response of the pixels are deviated with respect to each other. This can result in the classification accuracy decrease because both the variations and the amplitude in the spectral response are important in case of classification is hyperspectral imaging.

3.2 SUB-BAND DECOMPOSITION BASED SUPERVISED FEATURE EXTRACTION

As mentioned in the previous section, the studies about wavelet transform (WT) [34]-[36] emphasize that wavelet decomposition applied for dimension reduction has higher classification rate than PCA while preserving its simplicity advantage over PCA. However, PCA can reduce the data much more. That is why; wavelet is combined with PCA in [50]. In the study, PCA is applied to the hyperspectral image such that the selected PCA basis elements are not only orthogonal to each other but wavelets as well. It is observed that data dimension is reduced with this technique and further improvement on classification accuracy is obtained when compared with PCA method. Therefore, combining PCA with wavelet decomposition somehow can improve the PCA performance while preserving efficient data reduction of PCA algorithm.

In [44], We propose a method that first decomposes the hyperspectral signal into its wavelet sub-bands and that reconstructs a new signal from these sub-bands based on the determined weights. Then, PCA is implemented to the reconstructed signal and

support vector machine (SVM) method is utilized as a classifier of hyperspectral data. The algorithm of the study is given in Figure **3.2**3.2.

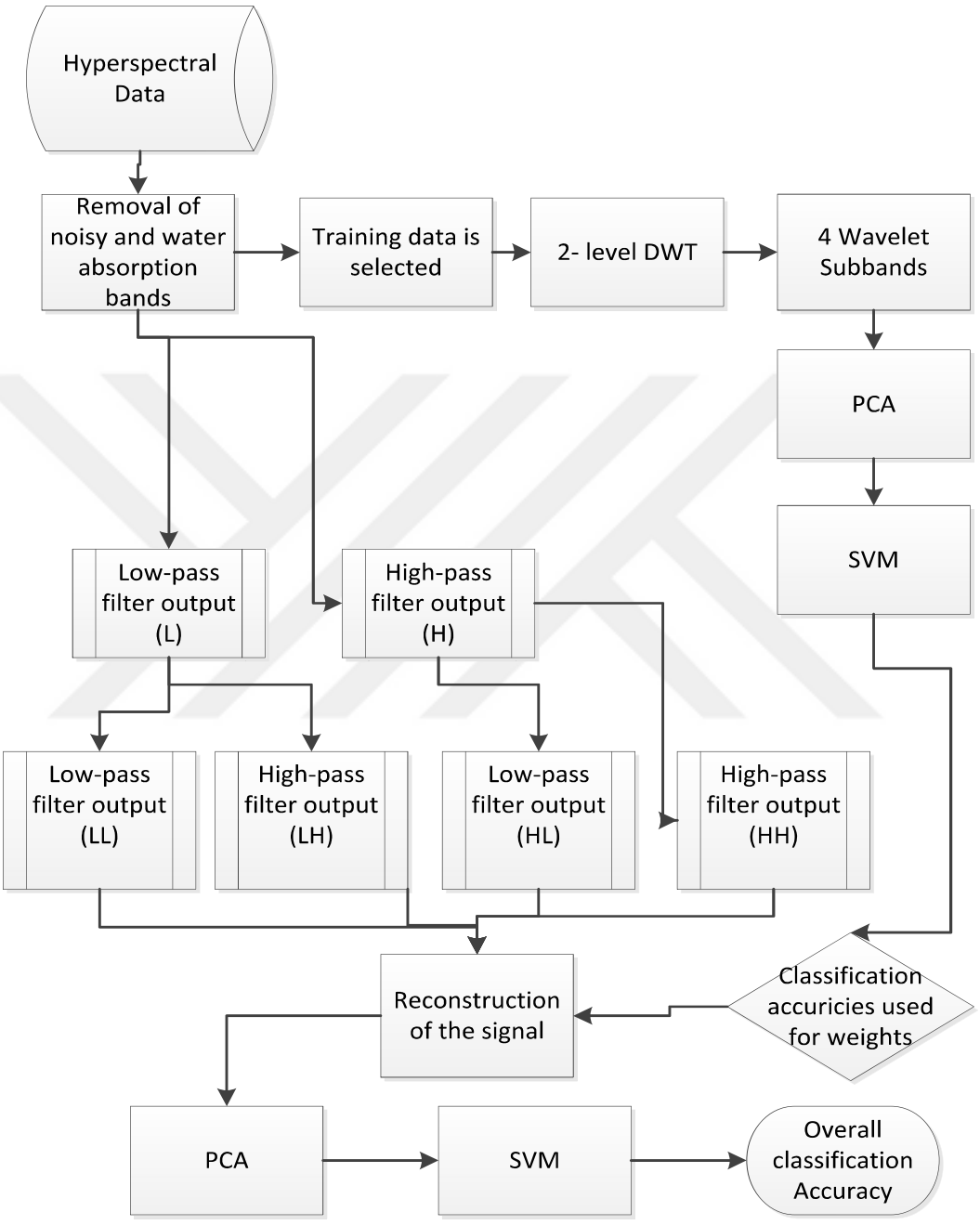

Figure 3.2: Sub-band Decomposition Based Supervised Feature Extraction Algorithm

Wavelet decomposition has two stages: filtering and down sampling. For filtering purposes 2-level Daubechies 4 wavelet [51] transform is applied to Indian Pine AVIRIS

data in this study. Indian Pine test data has 224 bands and active in the spectra between 400nm and 2500nm. After removing water absorption and low SNR bands 200 bands are used in the experiments. After removal of noisy and water absorption bands, 200-band 145x145 pixel hyperspectral data is formed for experiments. The ground truth of the test image is also available. The non-classified parts and classes that have small number of samples are eliminated from the ground truth. As a result, the hyperspectral data is evaluated over nine discriminative classes.

Let the hyperspectral image be denoted by x_{mn} and every $x_{mn} \in X$ for $1 \leq m, n \leq 145$ where m represents the row number and n represents the column number. Therefore, x_{mn} is the spectral response of the pixel at the m^{th} row and n^{th} column that has signal in 200 distinct bands such that

$$x_{mn} = x_{1,mn}\ x_{2,mn}\ x_{3,mn} \dots \dots\ x_{200,mn} \tag{3.1}$$

Indeed, x_{mn} is an array whose elements are responses from each distinct spectral band for the specific pixel at the m^{th} row and n^{th} column.

DWT decomposes the signal into two parts. Low pass filter outputs of the wavelet transform are called wavelet approximation coefficients whereas high pass filter outputs of the wavelet transform are called wavelet detail coefficients. Low pass filter outputs represent the large scale information and the high pass filter outputs represent the fine scale information.

The outputs of the first level Daubechies 4 filter applied to each x_{mn} pixel can be represented as

$$L_{mn}[p] = \sum_{n=-\infty}^{\infty} h[n-2p]x_{mn}[n] \tag{3.2}$$

$$H_{mn}[p] = \sum_{n=-\infty}^{\infty} g[n-2p]x_{mn}[n] \tag{3.3}$$

where h and g represent the low pass and high pass filters respectively, L_{mn} represents the approximation coefficients or low pass filter output, and H_{mn} represents the detail coefficients or high pass filter output.

The outputs of the second level Daubechies 4 filter applied to first level wavelet transform outputs are

$$LL_{mn}[p] = \sum_{n=-\infty}^{\infty} h[n-2p]L_{mn}[n] \tag{3.4}$$

$$LH_{mn}[p] = \sum_{n=-\infty}^{\infty} g[n-2p]L_{mn}[n] \tag{3.5}$$

$$HL_{mn}[p] = \sum_{n=-\infty}^{\infty} h[n-2p]H_{mn}[n] \tag{3.6}$$

$$HH_{mn}[p] = \sum_{n=-\infty}^{\infty} g[n-2p]H_{mn}[n] \tag{3.7}$$

Randomly selected training data from each class is applied to 2-level DWT and this forms four sub-branches LL, LH, HL and HH. 100 samples are selected randomly for each class, so the training data size is defined as 900×200 where 900 represents the pixel numbers located in nine classes and 200 represents the number of bands. Therefore, m and n are selected randomly to define 100 samples for each class provided that all the 100 selected samples are within the same class.

Then PCA is performed on each sub-branch and only few of the PCA components are contributed to the classification of wavelet sub-bands. In the experiments, the effect of the number of PCA components on the proposed algorithm is investigated. In order to achieve this, Monte Carlo simulations are run for different PCA dimensions ranging from 4 to 30. It is thought that if the PCA dimension is below 4 then the selected principal components can be insufficient to represent the information of the hyperspectral data. Also, it is thought that if the PCA dimension is above 30 then the data size cannot be reduced much.

Suppose that PCA dimension is k where $4 \leq k \leq 30$, then the outputs of PCA applied to each four sub-branch of the two level WT becomes:

$PCA\_LL_{mn} = [wLL_{mn,1}\ wLL_{mn,2}\ \ldots\ wLL_{mn,k}]$, PCA output of LL_{mn} (3.8)

$PCA\_LH_{mn} = [wLH_{mn,1}\ wLH_{mn,2}\ \ldots\ wLH_{mn,k}]$, PCA output of LH_{mn} (3.9)

$PCA\_HL_{mn} = [wHL_{mn,1}\ wHL_{mn,2}\ \ldots\ wHL_{mn,k}]$, PCA output of HL_{mn} (3.10)

$PCA\_HH_{mn} = [wHH_{mn,1}\ wHH_{mn,2}\ \ldots\ wHH_{mn,k}]$, PCA output of HH_{mn} (3.11)

where $wLL_{mn,j}$ is the jth component of PCA for $1 \leq j \leq k$ for the LL_{mn} sub-branch.

As a classification method SVM is utilized. After the utilization of SVM, four classification accuracies for each sub-band appear.

$\text{acc}_{\text{LL}} =$ SVM (PCA_LL_{mn}), classification accuracy of LL sub-branch (3.12)

$\text{acc}_{\text{LH}} =$ SVM (PCA_LH_{mn}), classification accuracy of LH sub-branch (3.13)

$\text{acc}_{\text{HL}} =$ SVM (PCA_HL_{mn}), classification accuracy of HL sub-branch (3.14)

$\text{acc}_{\text{HH}} =$ SVM (PCA_HH_{mn}), classification accuracy of HH sub-branch (3.15)

SVM $(.)$ is the supervised support vector machine classification methods that outputs the classification accuracy by considering the ground truth. Cross validation is applied to training samples for SVM training before calculating classification accuracies of each wavelet sub-band.

These four classification accuracies (acc_{LL}, acc_{LH}, acc_{HL}, acc_{HH}) are used to determine the weight of each wavelet sub-band. In order to calculate the weights, linear mapping method shown in Figure 3.33.3 is performed. Linear mapping is formed based on two reference points max_par and min_par. Maximum parameter is the reference classification accuracy point whose weight is determined as 2 whereas the minimum parameter is the reference classification accuracy point whose weight is determined as 1. As it is known, if it is not preferred to change the weights of the outputs of the DWT then all the weights of sub-branches should be taken as 1. However, in this study it is preferred that the sub-branch that has the largest classification accuracy dominates the reconstructed signal from the four sub-branches. Moreover, it is required to take all the sub-branches into reconstruction process. Consequently, max_{par} is taken as 0.80 and min_{par} is taken as 0.50. This means that if the classification accuracy of the sub-band is

0.80 its weight is 2 and if the classification accuracy of the sub-band is 0.50 its weight is 1.

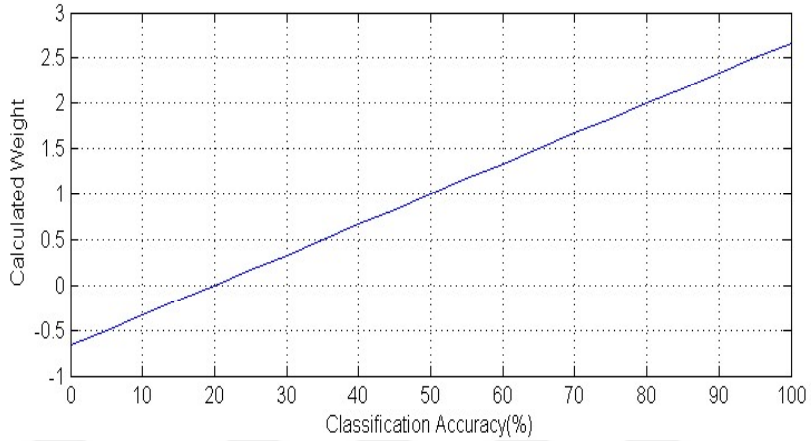

Figure 3.3: Linear mapping of Classification Accuracies to Find out the Weights of Subbands

The weights of the four sub-bands are calculated as follows:

$$\text{weight}_{\text{HH}} = slope * acc_{LL} + \text{const, weight of HH subband} \quad (3.16)$$

$$\text{weight}_{\text{LH}} = slope * acc_{LH} + \text{const, weight of LH subband} \quad (3.17)$$

$$\text{weight}_{\text{HL}} = slope * acc_{LL} + \text{const, weight of HL subband} \quad (3.18)$$

$$\text{weight}_{\text{HH}} = slope * acc_{LL} + \text{const, weight of HH subband} \quad (3.19)$$

where slope equals to $1/(max_{par} - min_{par})$ and const equals to $2 - slope * max_{par}$.

After the calculations of the weights of four sub-bands from the training data, the 2-level DWT is performed on the overall hyperspectral data eliminated from noisy and water absorption bands. The four sub-band outputs are LL_{mn}, LH_{mn}, HL_{mn}, and HH_{mn} as described above and m, n includes all values between 1 and 145 since it is worked on all data instead of a training data. Four sub-bands of the wavelet decomposition are combined to form a new feature signal based on the calculated weights. Therefore, weighted sub-bands are subjected to 2-level Inverse DWT. First level outputs of Inverse DWT are

$$\text{Data\_L} = \text{IDWT}(\text{weight}_{\text{LL}} * LL_{mn}, \text{weight}_{\text{LH}} * LH_{mn}), \text{ low branch output} \quad (3.20)$$

$$\text{Data\_H} = \text{IDWT}(\text{weight}_{\text{HL}} * HL_{mn}, \text{weight}_{\text{HH}} * HH_{mn}, \text{high branch output} \quad (3.21)$$

where $\text{IDWT}(\text{low}, \text{high})$ is the Daubechies 4 type inverse DWT operation with low pass filter output as low and high pass filter output as high.

With the second $IDWT$ operation, the reconstructed signal becomes

$$\text{Reconstructed\_signal} = \text{IDWT}(\text{Data\_L}, \text{Data\_H}), \text{reconstructed signal} \quad (3.22)$$

Since the signal is reconstructed proportional to the classification accuracy of each wavelet sub-band, it emphasizes the significant coefficients that have the most class discriminative properties. The reconstructed signal is subjected to PCA and same number of components used for training data is used for SVM classification method.

In the experiments, the effect of the number of PCA components on the proposed algorithm is investigated. To achieve this, Monte Carlo simulations are run for different PCA dimensions ranging from 4 to 30. In order to evaluate the performance of the proposed method, the results of the method are compared with the results of PCA algorithm. Figure 3.4 indicates the classification accuracies of the proposed wavelet based sub-band decomposition method and PCA with respect to the PCA dimensions. It is investigated that the classification accuracies of PCA and the proposed method are nearly the same for lower PCA dimensions. However, for higher PCA dimensions, the classification performance of the Proposed Method is better than that of PCA.

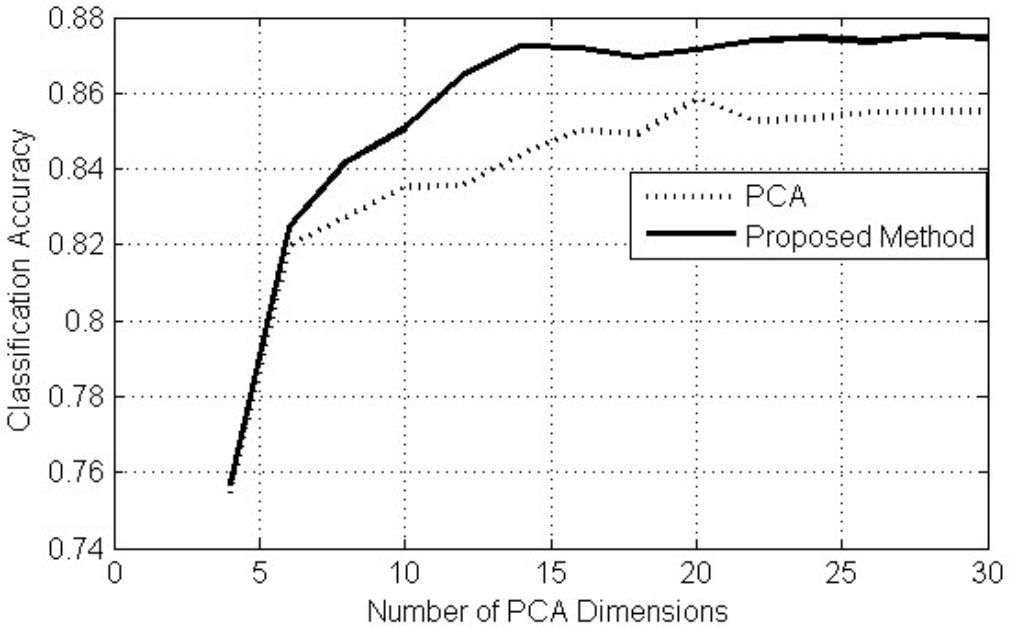

Figure 3.4: Comparative Results

PCA provides remarkable reduction in hyperspectral imagery. However, PCA is computationally complex with respect to DWT. Unlike DWT, it cannot eliminate arbitrary anomalies in a band. Wavelet Transform contains both the variations in amplitude and spectral info. PCA is optimal when the joint distribution of the noise and the clutter is Gaussian. However, this is not the case in practice. It is often hard to distinguish between the clutter coming from the background and the target signal.

In this study, wavelet-based sub-band decomposition method applied before PCA to increase the classification accuracy. It is investigated that classification performance of the supervised method that is proposed here is better than PCA especially for higher PCA dimensions larger than 6 while maintaining the reduction efficiency of PCA on hyperspectral imagery.

3.3 HOW TO SELECT TRANSFORMATION TYPE, FILTER TYPE, AND WAVELET LEVEL?

Most of the signals are represented as time domain signals in practice. This means that the measured value by the signal is a function of time. However, for most applications the distinct or valuable information can be recognized from the frequency representation of the signal. That is why; the frequency domain of the signal is required to be represented.

Fourier transform is a popular way of achieving the signal information in frequency domain. Fourier transform gives information about how much of each frequency exists in the signal. On the other hand, there are other transformations to convert time domain signal to a frequency domain one.

Fourier transformed signal has no time information. In other words, Fourier transform can give information about which frequency components exist in the signal but cannot give any information about at what times these frequency components exist. If the frequency content of the signal does not change over time, i.e., the signal is stationary, or if it is not required to know at what times the frequency components exist, Fourier transform is very useful. Since Fourier transform is lack of the time information of frequency components in the signal and it cannot be used for non-stationary signals, a window-based function called short time Fourier transform (STFT) is developed. STFT takes part of a non-stationary signal which is assumed to be stationary and process each of these partitions separately.

STFT gives information about time and frequency of the signal with some resolutions. Indeed, it is not possible to know the time instant of a specific spectral component. In other words, high resolution in time and frequency domain cannot be achieved at the same time which is called Heisenberg's uncertainty principle [52]. On the other hand, the time interval of the specified spectral components can be found. STFT presents a fixed resolution for time and frequency. This limits the time information of time spectral components. What is needed is a variable resolution. That is why; wavelet transform is preferred to be used widely in signal processing. STFT provides a fixed resolution, whereas Wavelet Transform provides a variable resolution.

The agricultural hyperspectral data is non-stationary. Therefore, in this study wavelet transform is selected to investigate signal characteristics and wavelet-based method is performed to increase classification accuracies of hyperspectral data.

Wavelet transform is becoming a common tool to analyze localized variations in time and frequency domain. Wavelet transform has been used for a wide variety of applications such as geophysics [53], compression, denoising, feature extraction [34-37], etc.

Wavelets can be mainly categorized into two: non orthogonal and orthogonal wavelets. Non orthogonal wavelet analysis is highly redundant especially at large scales.

Moreover, the filter banks are correlated. Therefore, non-orthogonal wavelet transform is beneficial when the wavelet transformed amplitude of the signal is expected to be smooth and continuous. On the other hand, orthogonal wavelet transform forms a wavelet spectrum consisting of discrete wavelet blocks. This structure provides the most compact representation of the signal. In hyperspectral image processing, one aim is to reduce the correlated and redundant data. That is why, dimension reduction methods are applied to the hyperspectral data. In consequence of this, it is preferred to use an orthogonal wavelet transform in order to reduce the redundant information. Additionally, hyperspectral agricultural data may have sharp peaks instead of having a smooth amplitude when wavelet transform is applied on it. In the Figures 3.5 and 3.6, wavelet low pass and high pass filter outputs of the SALINAS data exhibit sharp peaks instead of smooth curves.

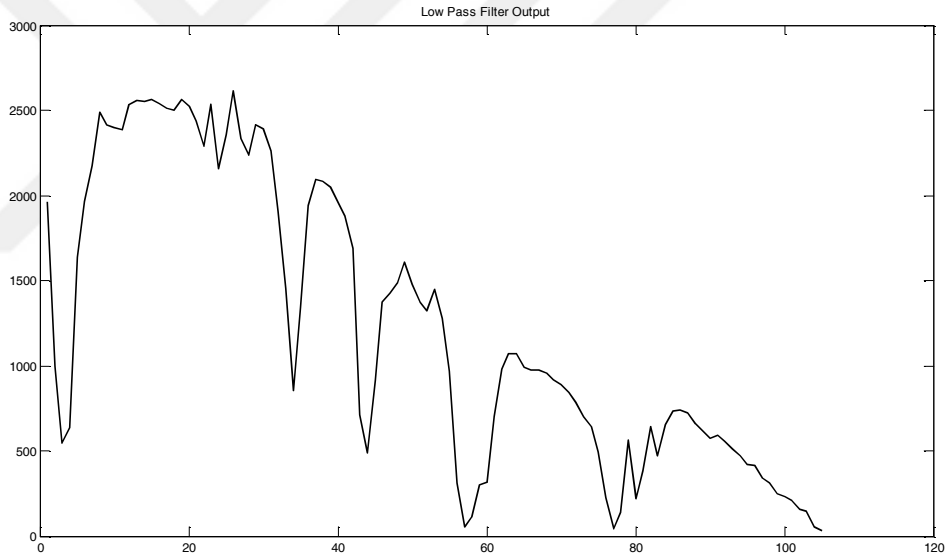

Figure 3.5: Wavelet (Daubechies 4) Low Pass Filter Output of SALINAS Data

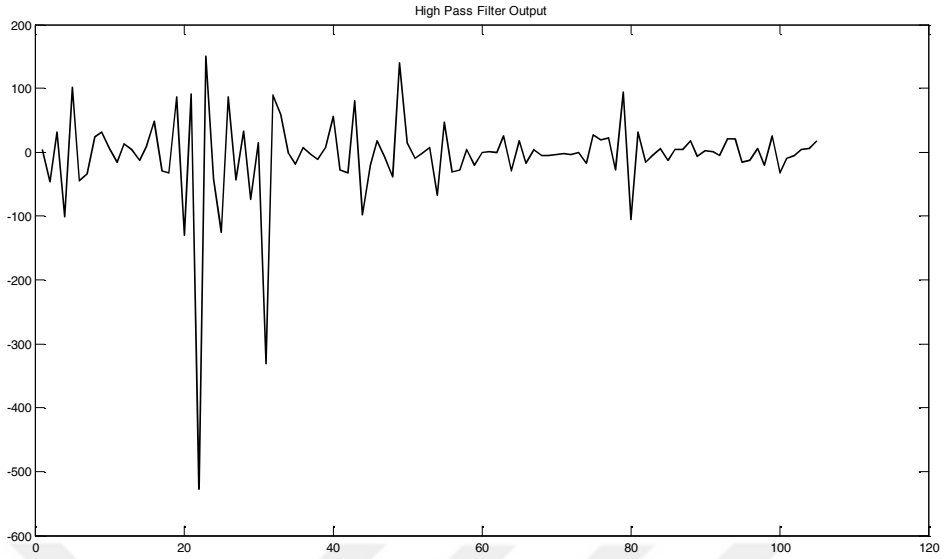

Figure 35.6: Wavelet (Daubechies 4) High Pass Filter Output of SALINAS Data

Based on the information about orthogonal wavelets and the figures [10, 11], it is decided to use orthogonal wavelets. Firstly, Daubechies 4 type wavelet is determined to be performed. Daubechies wavelets indicate a collection of orthogonal mother wavelets. [54] For affixed number of vanishing moments, Daubechies wavelets have a minimum size support length. Support length is defined by width of the window function. The lower the support size of the mother wavelet is the less number of high amplitude coefficients occur. Moreover, wavelets having a compact support length can be computed with finite impulse response conjugate mirror filters. Compactly support wavelet has a narrow window function. The Daubechies mother wavelets are not symmetric.

There are other orthogonal wavelets such as Haar, Symmlet, Coiflet, and etc. Haar is the simplest one and it is memory efficient, so it is fast. However, it has only one vanishing moment. Therefore, it does not fit well in order to approximate smooth functions. Symmlets are more symmetric than Daubechies whereas Coiflets are more symmetric than Symmlets.

Any orthogonal wavelet can be applied to the performed method. However, it is a misunderstanding that any wavelet is suitable for any signal and any applications. It is crucial to select the correct wavelet type for a high performance application result.

Crucial properties of a wavelet function can be expressed as compact support, symmetry, orthogonality, and degree of smoothness. The shape of the wavelet function is

also important must be taken into account based on the characteristics of the hyperspectral data. For instance, if the data has sharp jumps and steps in time domain Haar wavelet function can be used. On the other hand, when the data is smoothly varied in time domain then a smoother wavelet function such as Daubechies can be used.

Considering the best wavelet type for a specific application is a research topic in the literature. Jain Brothers investigated the performance of five wavelet types for image compression [55]. In [56], a recent study on image compression, wavelet functions were compared with respect to some quality metrics such as minimum square error. [57] is about a wavelet-based forecasting method for WiMAX traffic time series. The aim of this paper is to propose a strategy for the selection of wavelet type. The results of the studies were determined by the data and the performed methods. There is no generic solution. Our aim in this study is to increase the classification accuracy of the hyperspectral classes with respect to the SVM classification method.

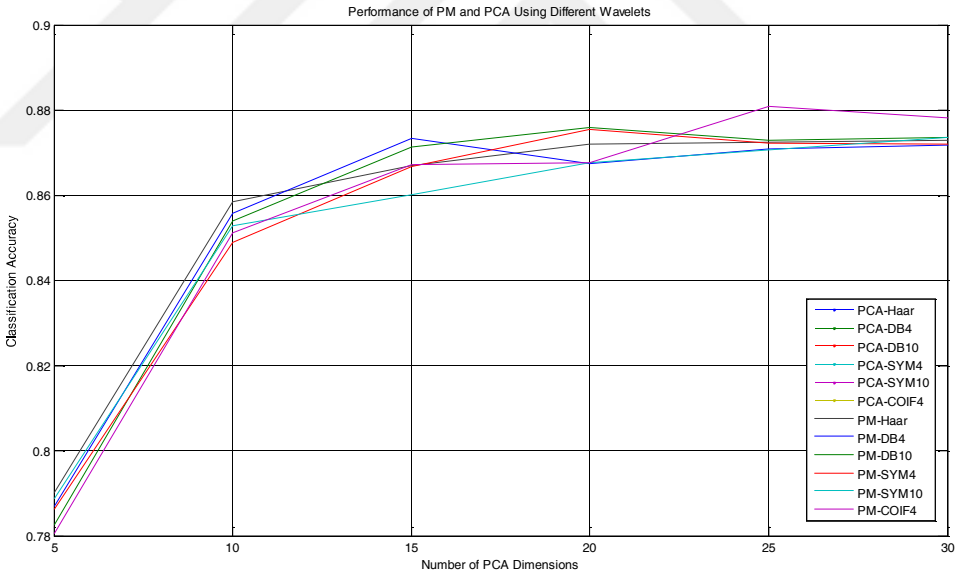

Figure 3.7: Performance of Proposed Method and PCA Using Haar, Daubechies4, Daubechies10, Symmlet4, Symmlet10, and Coiflet4 Wavelets

Figure 3.7 indicates that the wavelet based proposed method described in the section above outperforms PCA method for different PCA dimensions in terms of classification accuracies. When the wavelet performances are compared for the proposed

method, it can be observed that there is not much difference among the wavelet types for Indian Pine data. Daubechies 4 wavelet which is used in most of the proposed work is one of the best performance wavelets. In the proposed work, wavelet packet structure shown in the Figure 3.8 is performed. The level order of the wavelet packet used in the work is 2 and all the subbands are used with some determined weights based on the classification accuracy performances of each subband. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and high pass filters, respectively. Besides the wavelet packet tree used here, there can be many other wavelet tree structures. In wavelet applications usually the approximation component of the decomposition is decomposed iteratively whereas the detail components are not used or used without iterative decomposition. Octave band decomposition [58], seen in Figure 3.8 can be an example of this approach. However, if the important information is in higher frequency components, it requires further decomposition even in detail components.

Once the wavelet tree order is determined; there are many options for the wavelet packets that are going to be used. However, most probably many of them have a poor match for the application. To determine wavelet packet tree structure and wavelet packet tree order is a common problem in literature that has been investigated. It is possible to find a good match between the wavelet packet tree and the application data with an effective search algorithm.

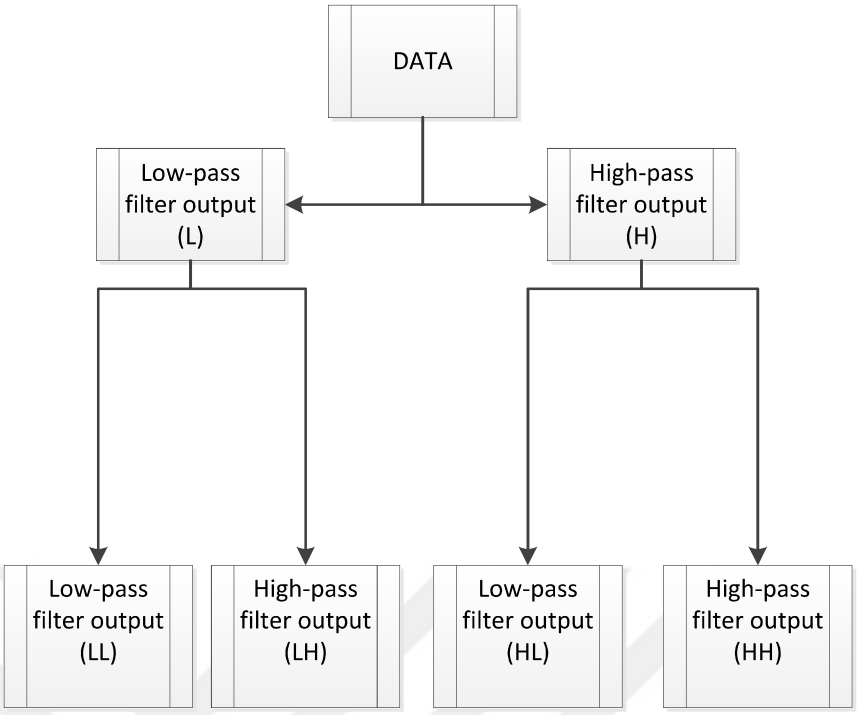

Figure 3.8: Performed Wavelet Packet Tree Structure

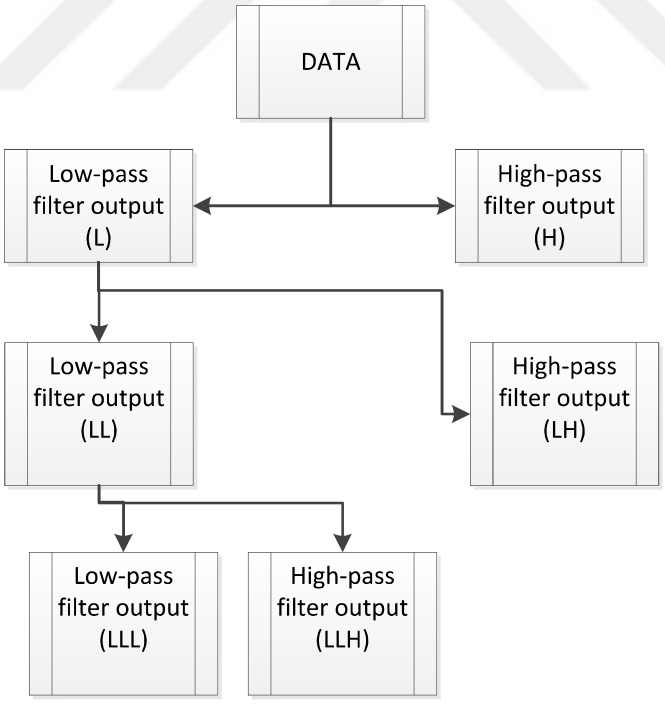

Figure 3.9: Octave Band Decomposition Tree Structure

In [59], wavelet families and packets were evaluated for image texture classification. In [60], to find out best wavelet subbands and best wavelet basis entropy based algorithm is implemented. In this study, entropy is used as a decision criterion for selecting which subbands to use. Moreover, in [61, 62] the decomposition of a signal is decided with respect to an entropy-based criterion. Entropy is a common concept mainly in signal processing since it gives information about the content of the signal.

It is not easy to determine the wavelet packet level. In [62], it was stated that based on the experimental results 2- or 3 levels could be chosen consistently for speech signals. In the proposed work 2 level wavelet decomposition was performed. However, it is possible to determine wavelet orders and wavelet basis with respect to the entropy criteria.

3.4 AGRICULTURAL DATA CHARACTERISTICS

Agriculture authorities in the world are required to know amount of crop yield every year so as to feed the earth human population. In our country, ministry of agriculture provide subsidy payments based on crop distribution and crop species that farmers being grown up. Therefore, agricultural land use information is crucial, and it has to be updated every year. This information has been gathered by communication with farmers and site visit of agriculture inspectors. However, this method is really time consuming, expensive, and insufficient. On the other hand, hyperspectral remote sensing can be time efficient, cost effective and more reliable in order to achieve agricultural land use information.

For agricultural applications, there are two main remote sensing methods: multispectral and hyperspectral. Both methods require phenological information since crops and vegetation can indicate different spectral characteristics at different seasons. In [63], vegetation classification problem due to seasonal variation in the spectral characteristics is emphasized. Therefore, it is beneficial to have the phenological information gathered from time series remote sensing data. Indeed, multitemporal data can increase the vegetation separability.

The agriculture species reflects valuable signals in the spectrum covering the visible, near infrared and the short-wave infrared regions i.e., $400nm$ to $2500nm$. That is why; well-known remote sensing satellites for agricultural purposes either fully or partially cover the defined spectrum. Scientists and engineers have worked to discover whether multispectral or hyperspectral methods would be more appropriate for crop

discrimination. Thenkabail, Lyon, and Huet stated in [64] that in order to discriminate among the biochemical materials involved in plant structures such as the starches, proteins, lignin, chlorophylls a and b, total chlorophyll, carotenoids, anthocyanin, nitrogen, water, and cellulose requires 3 to 5nm spectral bandwidths. Moreover, if multispectral sources are used, different vegetation species may have similar spectral signatures. On the other hand, hyperspectral sources provide detailed spectral signatures for distinct agricultural products so that classification can be realized with higher classification accuracies.

In [65], Landsat ETM bands are compared with Hyperion data. The latter gives higher spectral information and results in better classification results for distinct agricultural crops. In this study, the main aim is to find out optimal hyperspectral vegetation indices and hyperspectral narrow bands for agricultural crops. The study is realized based on (EO-1) Hyperion scenes and surface hyperspectral data for the eight leading worldwide crops wheat, corn, rice, barley, soybeans, pulses, cotton, and alfalfa). A field spectro-radiometer working in ($350-2500nm$) with 1-nm spectral sampling was used. In order to explore the best vegetation indices and narrow bands for agricultural products, temporal resolution is required in remote sensing applications. Therefore, in this study defined agricultural areas were investigated for 6 distinct plant growth stages: early vegetative, mid vegetative, flowering, tillering, critical, and senescing. 33 optimal narrow bands are determined to characterize, monitor, and classify these 8 crops. The narrow bands include 8 bands in the far short wave infrared ($1945-2230nm$), six in the near short-wave infrared ($1440-1770nm$), three in the near-infrared ($1095-1250nm$), four in the near-infrared ($750-1050nm$), three in the red-edge ($700-745nm$), two in the red ($645-700nm$), four in the green ($510-575nm$), and three in the blue ($400-495nm$).

Beside [65], there are other studies on agricultural crop classification such as [4, 66]. In [4], winter rape is automatically mapped and discriminated relative to its coexisting crops by using hyperspectral images with the existence of a product specific characteristic spectral signature. Hymap images with a $10-20nm$ spectral sampling were used for this study. ASD FieldSpec JR spectroradiometer was used for field spectral measurements. To reduce the effect of illumination difference spectral measurements were collected during 11:00–13:00 local time. The spectral data calibration was realized

with a Barium Sulphate white reference panel. Spectral information for the crops in the spectral library was resampled based on Hymap specifications to compare the Hymap hyperspectral images with the spectral signatures of the spectral library.

Use of pre-knowledge or training data for spectral information is hard to deal with when the pre-knowledge is obtained by different sensor under different atmospheric conditions and when the geometry of the sensor with respect to the imaging area is different. Therefore, in most of the studies the training data and the targeted data were collected in almost the same time from the same locations. However, in [4] there is 6 years of time difference between field measurements and hyperspectral images. In addition to this, there is $100km$ distance between the imaged area and the measurement field. Therefore, this study is more generic and independent to temporal and spatial changes compared to other studies.

When limited number of crops is to be classified, it is focused on finding the optimal wavebands and spectral samplings with respect to the distinct valuable wavebands. The aim of [4] is to determine the optimal waveband centers and spectral samplings for agricultural crop characteristics. Mainly 6 crops are investigated in Syria location: barley, wheat, lentil, cumin, chickpea, and etc. Narrowband data from $1.43nm$ hand held spectrometer manufactured by analytical spectral devices, broadband data from LANDSAT 5TM, and ground-truth data are used to realize this study. All measurements with spectroradiometer were taken under bright clear-sky conditions. The determination of the optimal waveband centers and widths are compared with the previous study of Thenkabail in 2000 [5]. The 10 of the band centers of the 12 optimal wavebands of this study are the same as those in the other study within $7nm$ margin. These results can be used to design an optimal hyperspectral imager for the six crops or to reduce the data used for the discrimination of the six crops.

It is shown in [3,4,5] that few spectral bands among a huge number of contiguous narrow spectral bands of hyperspectral imagery are sufficient for specific applications including limited types of products. For instance, calculation of greenness vegetation indices does not require the full spectrum data. Additionally, increasing the number of used bands for classification increases the required training data size. However, it is beneficial to keep in mind that the bands that are redundant for one application may be

valuable in other applications. Therefore, contiguous full spectrum data will always be required.

In [4] and [5] crucial bands for agriculture are provided. Pixel heterogeneity, mixed pixels, spectral similarity, and crop pattern variability are problems of crop discrimination. Some methods give inaccurate results because of these problems. To handle with this issue, many studies have been performed. [6] is one of these studies. [6] is an example of successful crop classification application. This study indicates that object-based image analysis can be a solution to the problems of spectral variability and mixed pixel.

Some studies indicated that some crops do not have unique spectral signatures even in high resolution spectral data. Therefore, it can be declared that some crops have unique spectral signatures whereas some others do not. Based on this information, it is beneficial to continue spending efforts of building spectral libraries for automated hyperspectral image discrimination.

3.5. DOMINANT SETS BASED FEATURE SELECTION METHOD IN HYPERSPECTRAL IMAGERY

There have been many studies on hyperspectral image processing. In order to make a valuable contribution on this area, it is crucial to investigate the studies of recent advances in hyperspectral image processing [67-70]. These studies indicate that last studies on hyperspectral image processing have concentrated on combining the spatial and spectral information. Using spatial information in hyperspectral applications is crucial since without the use of spatial information, hyperspectral images are assumed to be unordered lists of spectral measurements without particular spatial arrangement.

Spatial and spectral information combination can be utilized both in feature extraction methods and feature selection methods. Unlike feature extraction, feature selection cannot lead to deterioration of the critical data during data transformation since a subset from the original input data which is optimum for the decided criteria is selected. Moreover, usually a set of bands among the whole spectra is sufficient for a specific application. Band selection methods [25, 27] are the most popular feature selection methods that select the most valuable bands and remove the redundant bands based on a determined criterion.

The hyperspectral band selection technique can be ranking based [71-72] or clustering based [73-75]. Ranking based methods are utilized for detecting the most informative and distinguishing bands. However, there is a drawback that ranking based methods can select redundant bands since correlation among the selected bands is not considered. Unlike ranking based methods, clustering based methods take correlation among the bands into account. These methods firstly cluster the bands based on a determined criteria and then select one band from each cluster which represents the cluster. Therefore, it is decided to propose a clustering based band selection method that takes the spatial relation among the pixels into account.

Graph-based representations in pattern recognition have become state of the art approaches for problem solving specifically clustering problems. There are advantages in using graphs instead of using feature vectors for object representation [76]. Shi and Malik [77] introduced a normalized cuts method which was based on a generalized eigenvalue problem to obtain graph partitions according to normalized cuts criterion. On the other hand, Pavan and Pelillo [78] proposed a dominant sets method (DSM) in which a graph-theoretic definition of a cluster was introduced. DSM is also related with the maximum clique problem in graphs. Pavan and Pelillo [78] demonstrated the relation between DSM and maximum clique when weighted undirected graph is induced to unweighted undirected graph. The main idea of these methods is to get a graph-based representation of a problem by mapping the graph elements with vertices and mapping the relationship of the graph elements with weighted edges. Therefore, the problem is simply induced to solve a graph-based optimization problem. In this paper, we propose novel graph-based representations in hyperspectral band selection.

Hyperspectral image classification is an important application in agriculture, forestry, geology, ecological monitoring, and disaster monitoring [43]. According to application area, materials or elements under consideration can vary. A set of selected bands can be useful in one application domain; on the other hand, the same set may not have any use in another application domain. In real life scenario, a few numbers of distinct classes are required to be discriminated for a specific application with a limited number of training data. The main challenge is to detect the predetermined classes among a larger number of distinct classes. To achieve this, a framework is required to obtain an optimized set of bands that has not only inter-class discriminative information, but also

intra-class coherence for a specific application. DSM has the advantage of graph-based representation simplification and aims to increase intra-cluster heterogeneity while decreasing interclass heterogeneity. Therefore, it is a good idea to combine DSM approach with band selection in order to handle the main challenge in the real-life scenario. In this paper, we propose a novel framework based on dominant sets approach i.e., dominant sets based feature selection method (DSbFSM) in order to overcome the above mentioned problem.

DSbFSM is a graph-based clustering band selection method so that it aims to cluster the bands by dominant sets approach and selects the most valuable band from each cluster. In clustering band selection methods, clustering can be realized based on different criteria. In [73] mutual information is selected to cluster the spectral bands whereas mean average deviation criterion is used in [74] for clustering the bands. However, rather than finding the most informative bands among the whole spectra, it is crucial to find the most distinctive bands for predetermined classes. Therefore, DSbFSM computes the discrimination performances of the bands for determined classes in a specific application in order to utilize the result as a band clustering criterion. To evaluate the discrimination performances of each band, it is meaningful to use a clustering method. Therefore, the classification performance of each band is again evaluated by dominant sets approach.

The DSbFSM has two stages. In the first stage [45], the pixels in each band are clustered into groups. These clusters are matched with known classes by using the ground truth and the clustering performance of each band is evaluated for each class. Therefore, the first stage puts forward to performances of each band for each class. The second stage uses these results in order to cluster the bands of hyperspectral image. The most representative band in each cluster is selected. As a result, a set of bands that contains most discriminative information according to training data and has the least redundancy is selected. Shortly, the first stage is the performance evaluation of bands with respect to defined classes whereas the second stage is the band selection process by using the outputs of the first stage. In both stages, dominant sets based clustering approach is performed. In the former, DSM is utilized to cluster the pixels. On the other hand, in the latter DSM is utilized to cluster the bands.

The DSbFSM algorithm is shown in Figure 3.103.10

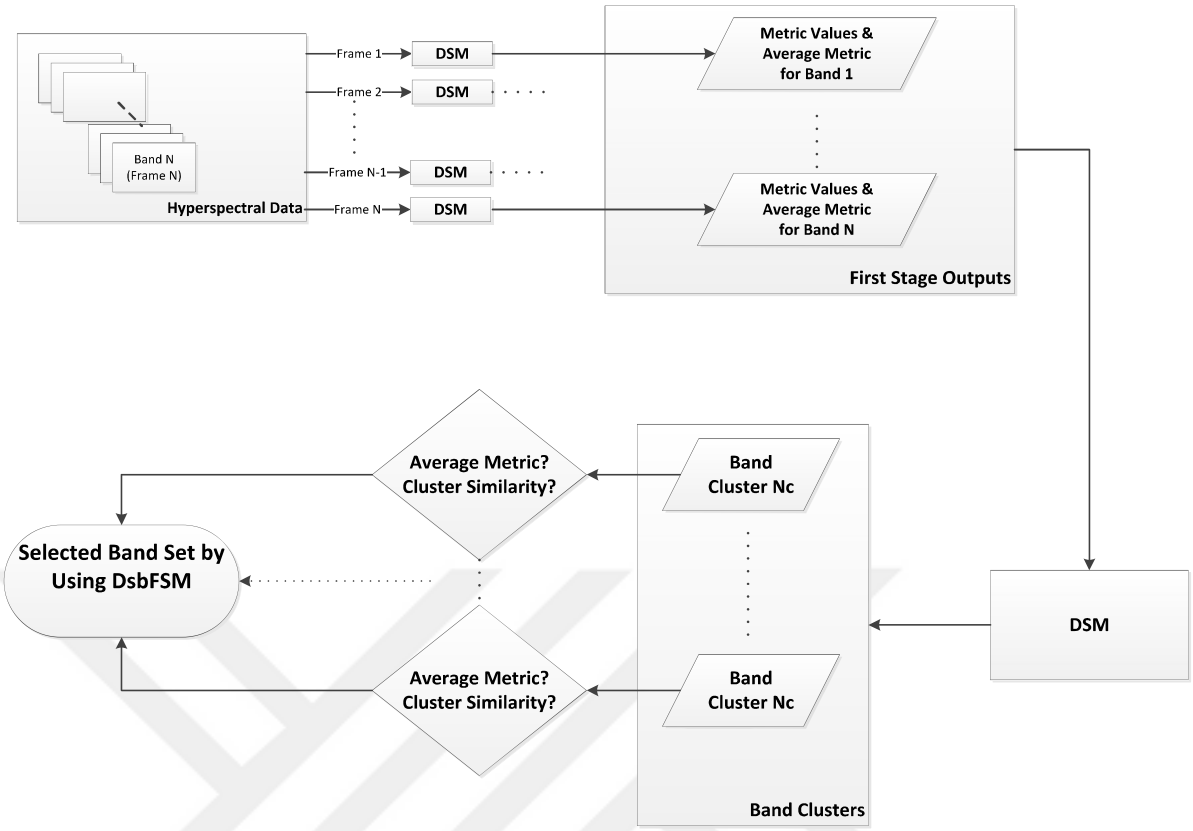

Figure 3.10: DSbFSM Algorithm

The first and second stage of the proposed method will be explained in detail in the following sections. Briefly, the main steps of the proposed method are as follows:

1. DSM is applied at each frame that belongs to a specific band for clustering.
2. Metric values of each frame are evaluated for each class. As a result of the process, a frame has metric values for each class and average metric value for the overall classes. The average metric is the mean value of the metric values for each class.
3. These metric values formed for each frame are utilized for finding out the similarity among frames in order to cluster the bands of the hyperspectral image by using the DSM.
4. The band that represents its cluster properties most i.e. the band that has high interclass similarity and high average metric value is selected from each band cluster. The selected bands form the set of bands determined by the proposed feature selection method.

Both stages of the proposed scheme are based on the dominant sets method. Fundamentals of the dominant sets method are provided in the sequel.

3.5.1. Dominant Set

Clustering problems occur in many areas of engineering such as image segmentation, computer vision, data mining and etc. Pavan and Pelillo [78] proposed a graph theoretic approach that combined the concepts of cluster and dominant set of vertices. This dominant sets method aims to increase the inter-cluster homogeneity while decreasing the intra-cluster homogeneity. This method is used as a movie scene detection approach in [79].

In this framework, the first step is to find out the similarities between the graph vertex pairs to form the similarity matrix $W(i,j)$ which is a $N\ x\ N$ nonnegative symmetric matrix and N is the number of vertices and i , j are the corresponding vertices. $W(i,j)$ is represented by a weighted undirected edge $edg(i,j)$ between each vertex i and j. A graph $G(V,E,W)$ is constructed by a set of vertices $V=\{1,\dots,N\}$, and a set of edges $E=\{(i,j)|W(i,j)>0\}$. For each $W(i,i)=0$, there are no self-loops in the graph.

Let $S \subseteq V$ be a nonempty subset of vertices and $i \in S$. $M_S(i)$ is defined to be the average weighted degree of i with respect to S:

$$M_S(i)=\frac{1}{|S|}\sum_{j\in S} W(i,j) \tag{3.23}$$

The equation (3.23) mentions the relation of $i \in S$ with the other vertices in S. In addition to this, it is required to form the relation between the vertices in S and j provided that $j \notin S$. Therefore, then function φ_S is defined to measure the similarity between vertices j and i, with respect to the average similarity between vertex i and its neighbors in S.

$$\varphi_S(i,j)=W(i,j)-M_S(i) \tag{3.24}$$

The total weight of subset S is defined as:

$$A(S)=\sum_{i\in S} a_S(i) \tag{3.25}$$

where $a_S(i)$ is weight of i $(i \in S)$ with respect to S and it is represented as below.

$$a_S(i) = \begin{cases} 1, & if\,|S| = 1, \\ \sum_{j \in S\{i\}} \varphi_{S\{i\}}(j,i)\, a_{S\{i\}}(j), & otherwise. \end{cases} \tag{3.26}$$

$a_S(i)$ defines the overall similarity between vertex i and the vertices of $S\backslash\{i\}$ with respect to the overall similarity among the vertices in $S\backslash\{i\}$. A non-empty subset of vertices $S \subseteq V$ such that $A(U) > 0$ for any non-empty $U \subseteq S$, is said to be dominant, if:

1. $a_S(i) > 0$, for all $i \in S$,
2. $a_{S\cup\{i\}}(i) < 0$, for all $i \notin S$.

The resulting two statements mentioned above correspond to the cluster conditions that are related to inter-cluster homogeneity and intra-cluster inhomogeneity respectively. The dominant sets found based on the two criteria can be classified as clusters. In practice, direct implementation of this framework is not feasible because of its computational complexity. To solve this problem, a continuous optimization technique, known as replicator dynamics can be used [77].

$$x_i(t+1) = x_i(t)\frac{(Wx)_i}{x(t)^T W x(t)} \tag{3.27}$$

where t is the iteration index and T indicates the transpose of the vector. The iteration is terminated if and only if the difference between two consecutive iterations of x is below the predetermined threshold. Once the dominant set is detected, the remaining vertices can be subjected to the same equation in order to find new dominant sets i.e. clusters. By this way, clustering can be realized for the data.

3.5.2 Dominant Sets Based Feature Selection Method (DSbFSM)

This feature selection method proposes a selected set of bands for discriminating the classes under concern. In order to achieve this, a small part of the hyperspectral image that contains sufficient samples for the interested classes is selected. The DSbFSM will be performed on the selected part of the whole image. Therefore, it is aimed to find out a general frame work that can define required bands for classification without performing on the whole data set. Indeed, the selected set of bands can be used to discriminate the classes under concern even for another region that contains not only the same classes but also contains other distinct classes.

3.5.2.1 First Stage of DSbFSM

In the proposed approach, the first step is to evaluate the performance of bands with respect to defined classes based on a particular metric [45]. In this study, F-measure metric [79] is used to quantify the classification performance of bands. In order to evaluate the discrimination performance of bands, pixels in each frame that belongs to a specific band are clustered by DSM. Then, the clustered pixels are classified by the ground truth. The cluster is appointed to the class that has highest number of member pixels in the cluster. As a result, each band frame is classified. Classification performance is evaluated using F-measure metric [79].

Let H_f be the $L \times M \times N$ hyperspectral image i.e., the hyperspectral cube has L bands, M pixels in the image frame row and N pixels in image frame column. Let H be the partition of H_f including sufficient samples of the classes under concern. That is, H is also a hyperspectral image having L bands and K number of pixels where $K < MxN$. Therefore, H can be expressed as:

$$H = \{H_l\}_{l=1}^{L} \tag{3.28}$$

where H_l is the l^{th} band of the hyperspectral image and it is a K pixel image frame.

In this stage, each frame that belongs to a specific band i.e. H_l is subjected to DSM to form the clusters. The first step is to form the similarity matrices for each band. Let W_l is the KxK similarity matrix for the l^{th} band and $W_l(i,j)$ is the element of the similarity matrix that represents the similarity between the i^{th} and j^{th} pixels, and $1 \leq i,j \leq K$.

In order to form the similarity matrix, it is required to have a similarity measurement function. It is determined that this function is related with both spatial and spectral information. Therefore, $W_l(i,j)$ is defined as:

$$W_l(i,j) = e^{-C_d {d_{ij}}^2} e^{-C_s |r_l(i) - r_l(j)|} \tag{3.29}$$

where C_d and C_s are spatial and spectral information parameters respectively. Once these parameters are determined they are used for every $W_l(i,j)$. $r_l(i)$ is the reflectance value of i^{th} pixel for the l^{th} band. Indeed, $r_l(i)$ is an element in H_l . The spatial distance in the image frame between pixels i and pixel j is represented as d_{ij} .

After forming the similarity matrix, the dominant set is iteratively calculated by using (3.30). The threshold Th defines the number of iterations. The remaining part of the first stage of the proposed method is given in the following steps:

Step1. The initial value of x_l is a K size array consisting of all ones. It means that initially all pixels of the l^{th} band are in the dominant set.

Step2. Calculate the dominant set by using (3.30):

$$x_l(t+1) = x_l(t)\frac{(Wx)_l}{x(t)^T W x(t)}$$

Step 3. If $|x_l(t+1) - x_l(t)| > Th$ then $x_l(t) = x_l(t+1)$ and go to *Step 2*. Otherwise go to *Step 4*.

Step 4. The dominant set is finalized. The dominant set elements constitute the cluster. This cluster is assigned to the proper class by using the ground truth. The maximum number of cluster pixels that are affiliated to the same class defines the class of the cluster. Therefore, the pixels in the dominant set are classified.

Step 5. The classified pixels are extracted from the image and the similarity matrix elements related to these pixels are eliminated from the similarity matrix W_l.

Step 6. If there are unclassified pixels, form x_l array with a size equal to the number of unclassified pixels consisting of all ones and go to *Step 2*. Otherwise go to *Step 7.*

After performing the iterative algorithm above, all the pixels are classified. The next step is to evaluate the performance of bands for each class with a metric. For the evaluation process, it is important to succeed in both clustering the pixels that belongs to the same class in the same cluster and preventing from mismatch of the cluster pixels with incorrect classes. Therefore, F-measure metric [79] is selected to evaluate the performances since it takes not only the detection but also the false alarm into account. It measures the quality of the detected clusters and ranging from 0 to 1. $F = 1$ shows a perfect result.

F is defined as follows.

$$F = \frac{1}{Z}\sum_{C_i \in GT} |C_i| \max_{c_j \in DT}\{f(C_i, C_j)\} \tag{3.31}$$

where

$$f(C_i, C_j) = \frac{2 \times \mathrm{Re}(C_i, C_j) \times \mathrm{Pr}(C_i, C_j)}{\mathrm{Re}(C_i, C_j) + \mathrm{Pr}(C_i, C_j)}, \tag{3.32}$$

GT is the ground truth and DT is the dominant set.

$$Z = \sum_{C_i \in GT} |C_i| \tag{3.33}$$

Recall (Re) and precision (Pr) functions are defined as follows:

$$\mathrm{Re}(C_i, C_j) = \frac{|C_i \cap C_j|}{|C_i|} \tag{3.34}$$

$$\mathrm{Pr}(C_i, C_j) = \frac{|C_i \cap C_j|}{|C_j|} \tag{3.35}$$

Step 7. As described in Step 4, F-metric is calculated after the clusters are assigned to the proper class. Calculate the F-metric of the l^{th} band for each class by using the equation:

$$F_l(i) = \frac{1}{Z} \sum_{c_i \in GT, c_j \in DT} |C_i| \times f(C_i, C_j) \tag{3.36}$$

for $1 \leq i \leq P$ where P is the number of classes in the ground truth.

Step 8. F_l is l^{th} band performance for each class and each element of this array $F_l(i)$ refers to a distinct class in the ground truth. Besides this, an overall metric is calculated for the l^{th} band called $F_{l,tot}$. $F_{l,tot}$ is the mean value of F_l .

All the steps given above are applied to each band of the hyperspectral cube. Therefore, the performances of the hyperspectral bands are evaluated separately. The first stage of the algorithm was published as a part of this study [45].

3.5.2.2 Second Stage of DSM

In the first stage of the proposed method, performance evaluation of bands with respect to defined classes is realized. In the second stage, the aim is to find out the set of bands that mostly discriminates the given classes by using the information coming from the first stage.

Since the output information of the first stage is the input of the second stage, we have to start with F_l s for $1 \leq l \leq L$. As mentioned before F_l is the F-metric values of the l^{th} band for each class. Actually, $F_{l,tot}$ that can be defined as the overall performance of the l^{th} band was also calculated for each band in the former stage. However, it is not meaningful to select the bands that have largest $F_{l,tot}$ values. Due to the fact that the band

which has low overall performance can possess high performance for a distinct band, it is better to use F_l s that includes distinct performances for each class.

The main aim of the band selection is to eliminate redundancy. Therefore, it is required to know the redundant bands. Band clustering is a good solution to define the redundant bands. F_l is behaved as a signal having P samples and dominant sets technique is applied among all F_l s for $1 \leq l \leq L$. To apply dominant set technique, similarity measurement matrix W_B is formed. W_B is a LxL symmetric matix and its elements can be defined as:

$$W_B(i,j) = e^{-C_B(1-corr(F_i,Fj))} e^{-C_V \frac{1}{P} sum(abs(F_i,F_j))} \tag{3.37}$$

where C_B is the band correlation parameter and $corr(f,g)$ is the correlation of signals f and g . Additionally, C_V is the mean square distance parameter and $abs(.)$ is the absolute value function. At first glance, it seems that correlation function is sufficient to form the similarity matrix. However, the distance between band pairs is also crucial to distinguish the bands. The first exponential term of the similarity matrix is related with the correlation between band pairs whereas the second exponential term is related with mean square distance term in the literature that defines the distance between the band pairs. Besides the correlation function, the distance between band pairs is also crucial to distinguish the bands. Let A be a R dimensional signal and B is another signal that is equal to c times A where c is a constant between 0 and 1. $corr(A,A)$ and $corr(a,B)$ are both equal to 1 unless B is different from A. In order to differentiate between F_ls such as the signals A and B, mean square distance term shall be used while calculating the similarity matrix.

Using (5), the dominant sets of bands are calculated iteratively. The iterations finalize provided that the distance between two successive iterations is smaller than the threshold T_B. When the iterations finalize, a dominant set is found. This dominant set refers to a band cluster consisting of similar and/or redundant bands. The remaining bands are subjected to the dominant sets algorithm described above until no unclustered band exists.

At this stage, there are N_C numbers of band clusters. Therefore, N_C bands have to be selected from each cluster to determine the set of bands formed by the proposed feature

selection method. The selected band within the cluster has to satisfy the below statements:

1. The selected band has high $F_{l,tot}$
2. The selected band has highly correlated with other bands within the cluster

To satisfy the two statements above, band selection criteria is defined as:

$$\eta_{l,z} = F_{l,tot} \times \sum_{j \in D_z} W_B(l,j) \tag{3.38}$$

where D_z is the band cluster for $1 \leq z \leq N_C$ and $\eta_{l,z}$ is the band selection parameter of the l^{th} band which is located in the z^{th} cluster. The band which has the maximum $\eta_{l,z}$ value is selected from its cluster. The selected bands form the output of the proposed feature selection method.

3.5.3 Other Feature Selection Methods

In order to evaluate the performance of the proposed method, two other feature selection methods are implemented: Correlated band selection (CBS) method and sequential forward selection (SFS) method.

CBS is a simple method that realizes the band selection by considering the correlation among bands. First of all, correlation parameters between band pairs are calculated. Band pair correlation is found by using the MATLAB function $corr\ (X,Y)$ where X and Y refers to image frames for the first and second bands. The most uncorrelated band pair that has the lowest $abs(corr(X,Y))$ value is chosen as the first two members of the selected band set. Then the correlation pair values of the current selected bands are summed to find the next band. The most uncorrelated band with the current bands is selected as a new member. The last step is performed iteratively until the required number of bands is achieved.

In SFS method, features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion. Criterion is selected as the classification accuracy of SVM. SFS does not propose the optimum set of bands, but it is a suboptimum method. The band (the frame of the band) that has the highest classification accuracy is selected as the first member of the band set. In order to achieve this, SVM is applied to all band frames. That is, if the band number is N then SVM is applied N times. Next step is to find out the next band that increases the classification accuracy provided

that it will be used with the current bands of the selected band set. Therefore, for the next step SVM is performed $N-1$ times. The last step is run iteratively until the classification accuracy does not increase any more. Since SVM is performed iteratively in SFS, its computational complexity is high.

3.5.4 Performance Evaluation and Results

In this study Salinas and Pavia scenes are used [80]. As described earlier, DSbFSM is performed on a small partition of the whole hyperspectral image that contains sufficient samples for the interested classes.

Salinas A data includes six classes DSbFSM is performed to find out valuable hyperspectral bands for the six classes. Salinas A data is the partition of the Salinas data. Salinas A is approximately 5% of Salinas data when removing the unclassified regions. Salinas data includes 10 more classes but these class samples are signed as unclassified since they are not under concern. The ground truth of Salinas A is shown in Figure.

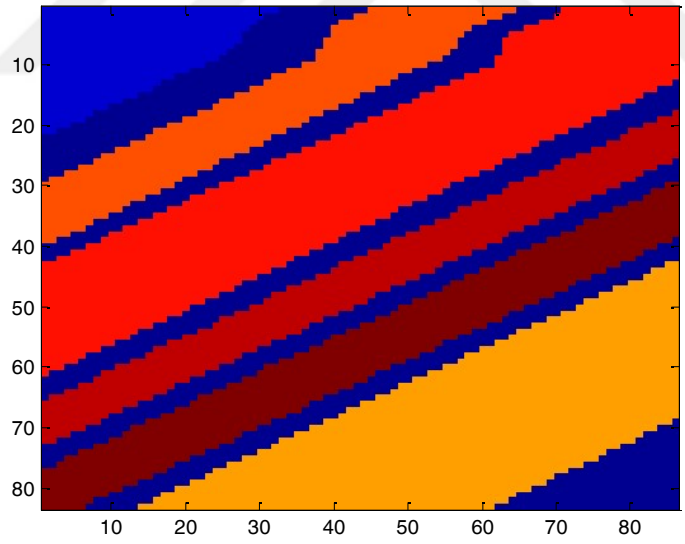

Figure 3.11: Salinas A ground truth including 6 classes and unclassified regions

A partition includes trees, bitumen, tiles, meadows, bare soil, and water classes are selected from the Pavia data. This partition is approximately 0.3% of the full Pavia data when removing the unclassified regions. Selected part of Pavia including four classes is shown in Figure 3.12.

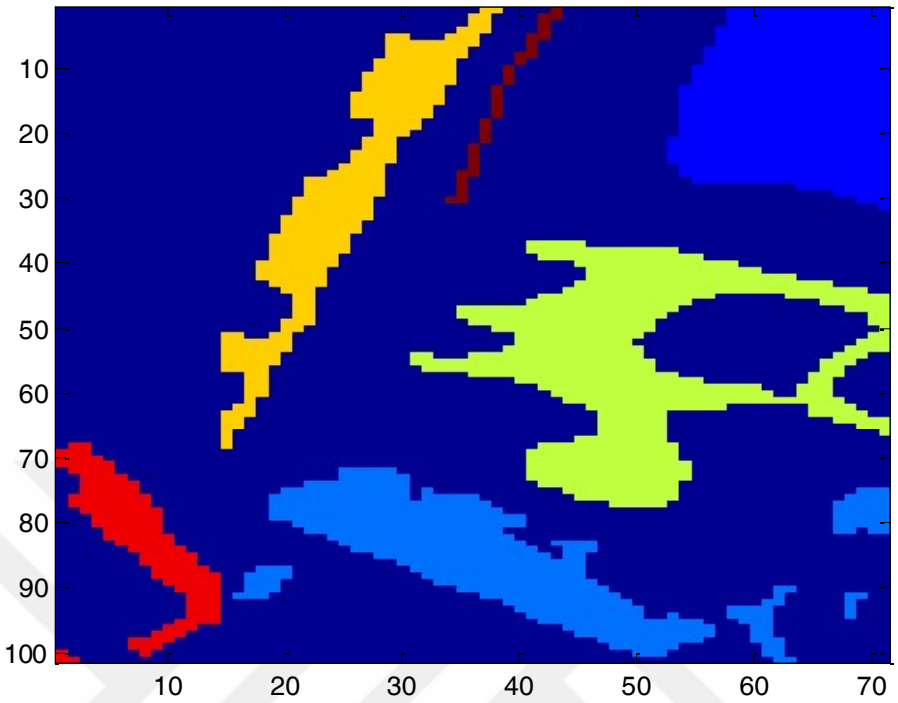

Figure 3.12: Ground truth of Pavia selected partition including 6 classes and unclassified regions

DSbFSM is performed on the three distinct hyperspectral data as mentioned in the above part. While implementing the DSbFSM, there are four parameters to determine. Two of them are for the first stage C_d and C_s. These parameters are found iteratively by using the red edge band in Pavia data as 0.3 and 0.01 for the spatial and spectral parameters respectively. For Salinas data spatial parameter remains constant and spectral parameter is multiplied by the ratio of the radiance mean values between the performed data and the Pavia data. The remaining two parameters C_B and C_V are for the second stage and taken as 200 and 30 for all hyperspectral data.

While performing the DSbFSM, only the samples of the classes under concern are taken into account. Once DSbFSM is performed on small partition of the whole hyperspectral data and valuable bands are selected, a new hyperspectral image that includes all the samples but contains only the selected bands is formed. SVM is applied to this new hyperspectral data in order to find out the performance of DSbFSM in terms of overall classification accuracy for the classes under concern. In order to evaluate the

performance of DSbFSM, correlation band selection (CBS) and sequential forward selection (SFS) [81] band selection methods are implemented. Results are compared in the section below. SFS is a suboptimum band selection method that is used in many studies recently. CBS is a simple method that realizes the band selection by considering the correlation among bands. Most uncorrelated bands are selected by CBS. For fair evaluation, number of selected bands of CBS is equalized to the number of selected bands of DSbFSM. On the other hand, since SFS is an iterative suboptimum method the selected bands of SFS are accepted as is. Classification accuracy [82] vs number of training set performances are analyzed for three band selection methods. Number of training set value defines the number of training samples selected for each class. SVM is performed for different number of training sets and classification accuracy is calculated for 20 times for the same training set number to get the average value. Training set for SVM is randomly selected and it includes samples for each class. While comparing the performances of the three band selection methods with SVM, training set for SVM is randomly selected and the same training set is used for all band selection methods.

DSbFSM, CBS, and SFS performances for the Salinas data are shown in Figure 6. It is observed that DSbFSM performance is slightly larger than that of SFS performance. CBS performance is lower than the others especially for small numbers of training set.

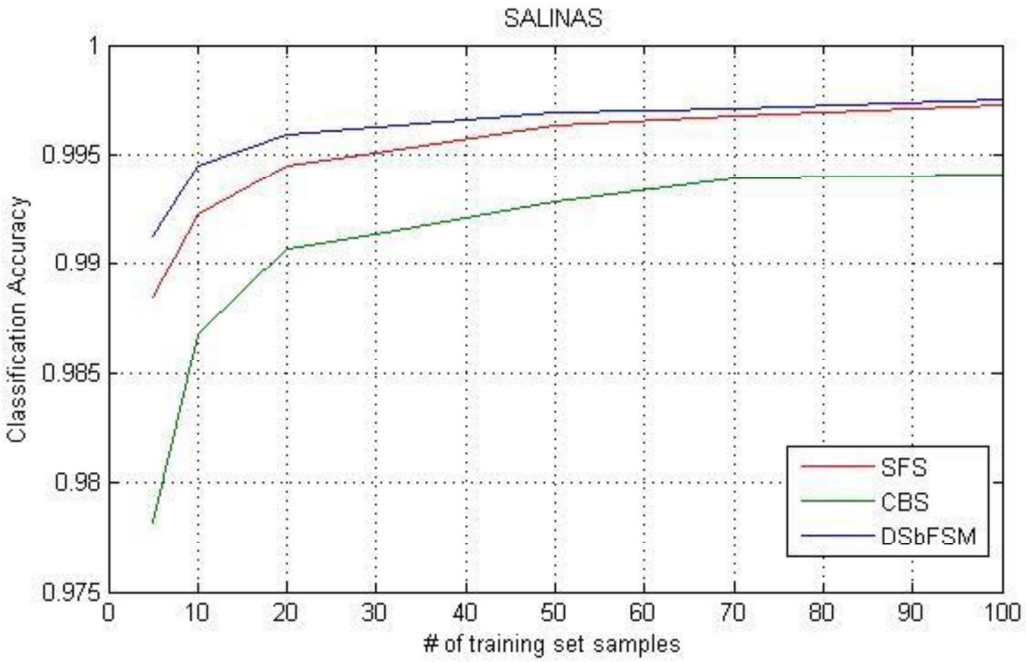

Figure 6: Classification Accuracy vs Number of SVM training set samples performances of DSbFSM, CBS, and SFS for Salinas Data

DSbFSM, CBS, and SFS performances for the Pavia data are shown in Figure 3.143.14. It is observed that DSbFSM performance is greater than SFS and CBS performances. The size of the partition the Pavia data used to realize band selection is really small compared to whole Pavia data. It is investigated that DSbFSM performs well even the size of the training data for band selection is very low.

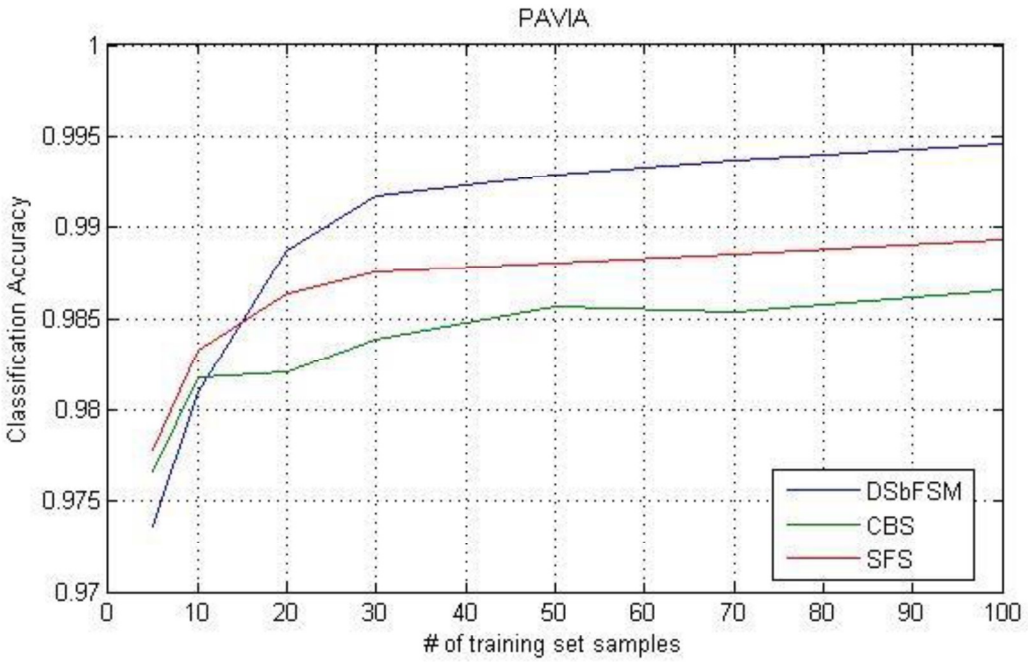

Figure 3.14: Classification Accuracy vs Number of SVM training set samples performances of DSbFSM, CBS, and SFS for Pavia Data

DSbFSM proposed a general set of selected bands for the classes under concern with high classification accuracies. Moreover, CBS and SFS is time consuming since they perform iterative steps to find out the selected bands. In addition to this, the classification method SVM is run for each iteration when SFS is performed.

The proposed feature selection method DSbFSM performs on a small size of the entire data that includes samples from classes under concern. The performance evaluation of DSbFSM indicates that the proposed set of selected bands by DSbFSM can be used to discriminate the classes under concern even for another region that contains not only the same classes but also contains other distinct classes. Therefore, DSbFSM is a general framework that can define required bands for classification. When compared with the other feature selection methods mentioned in this study such as CBS and SFS, DSbFSM has higher classification performance. Moreover, DSbFSM is time efficient since it performs on small size of data, and it does not require to calculate classification accuracy iteratively in order to realize feature selection.

3.5.5. The Importance of the Proposed Method in the Literature

The dominant set framework has been applied to many applications such as image segmentation and machine learning. However, in this study dominant set framework is firstly applied to the hyperspectral image. Moreover, dominant set framework is used twice in the proposed method. The first is utilized to cluster the scenes for each spectral band as in the image segmentation applications. The second is utilized to cluster the spectral bands to find out the redundant bands. Therefore, it can be declared that dominant set has not been used for clustering the bands until this study.

Usually, in hyperspectral image processing applications a group of classes is under concern. Therefore, rather than using the full spectra a set of few bands selected by a feature selection method is sufficient. It is obvious that the bands in the set can alter based on the application. With a small training data including the classes under consideration, the DSbFSM method proposes a general solution to discriminate these classes whether the discrimination is realized on other hyperspectral data including other classes. As a result, the proposed DSbFSM is a novel framework that has discrimination capability inter classes and coherence capability intra classes.

3.6. DOMINANT SETS METHOD PROPOSED AS PREPROCESSING BEFORE FEATURE SELECTION

In Section 3.5 above, dominant sets based feature selection method is proposed. This method is a two stage method and DSM is performed in both stages. In the first stage, DSM is applied to find the clustering performances of each spectral band based on the F-metric values. In the second stage, spectral bands are clustered based on their clustering performances by DSM. The results indicate that DSM is an effective method for feature selection in hyperspectral imagery. However, this two-stage method is analytically complex and DSM is applied at each stage. Therefore, it is determined to simplify the two stage method. The idea is to realize feature selection by using the outputs of the first stage of DSbFSM.

3.6.1 DSM utilized to increase classification performance of SFS (DSM-SFS)

Sequential forward selection (SFS) feature selection technique is a sub-optimum method as described in the above sections. SFS works on the whole spectral bands set.

The classification performance of SFS can be increased by eliminating the spectral bands that decreases the classification accuracy. Moreover, decreased number of spectral bands results in a low computational complexity of SFS.

A DSM based preprocessing method is proposed to increase the classification performance of the sub-optimum feature selection method SFS for hyperspectral images [83]. First of all, as in the first stage of DSbFSM, the clustering performances of each spectral band are calculated by applying DSM to each spectral band. In order to evaluate performances F-metric values are formed. Therefore, the eight steps mentioned in *Section 3.5.2.1* are performed to calculate F-metric of the l^{th} band for each class by using the equation:

$$F_l(i) = \frac{1}{Z} \sum_{c_i \in GT, c_j \in DT} |C_i| \times f(C_i, C_j) \tag{3.39}$$

for $1 \leq i \leq P$ where P is the number of classes in the ground truth.

The aim is to eliminate some of the spectral bands that do not increase the classification accuracy before performing SFS. Therefore, redundancy and useless spectral bands for the defined classes will be omitted. Moreover, classification accuracy is expected to be increased since some spectral bands have negative effect on the overall classification performance.

The F-metric values are utilized by DSM based pre-feature selection method to eliminate the redundant and useless spectral bands. Figure 3.153.15 indicates F-metric values of spectral bands for 'Brocoli Green Weeds' class in Salinas data. It is observed that there are some spiky values within local windows. The spectral bands that have that kind of spiky subdued F-metric values may decrease the overall classification performance.

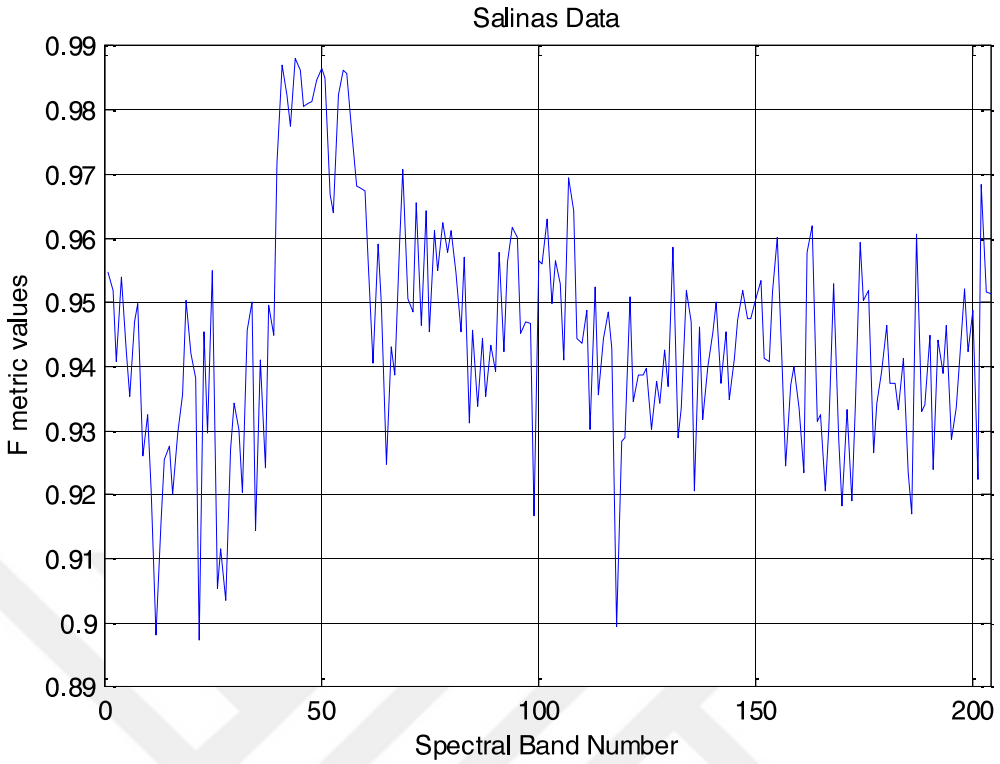

Figure 3.15: F-metric values of Spectral Bands for 'Brocoli Green Weeds' Class in Salinas Data

In this proposed method, unsharp masking filter is applied to F-metric values for each class. Unsharp masking filter is a local filter that derives an unsharp and less blurry signal by creating a mask of the original signal [84]. Generally, unsharp filter can be described as below:

$$y[n] = h_{unsh}[n].x[n] \tag{3.40}$$

where $x[n]$ is the original signal, $y[n]$ is the enhanced signal, and h_{unsh} is the unsharp masking filter. In this study, h_{unsh} is described as:

$$h_{unsh}[n] = \begin{cases} 1, \ if \ x[n] > \left(\frac{1}{L_u}\sum_{m=-\frac{L_u-1}{2}}^{\frac{L_u-1}{2}} x[n+m]\right) + std_{L_u}(x[n]) \\ 0, \qquad otherwise \end{cases} \tag{3.41}$$

where L_u is the length of the unsharp masking filter and std_{L_u} is the local standard deviation within L_u length. Actually, a local threshold is defined for each sample of the signal based on its neighbor samples. This threshold equals to the sum of mean and standard deviation values within the filter length. This unsharp masking filter is applied to F-metric values of spectral bands for each class. Therefore, this process removes the sharp and noisy signals and forms the most valuable spectral bands for each class. Then

the valuable bands for each class are merged to form the spectral bands set to be given as an input to SFS.

The remaining part of the preprocessing DSM based method after the eight steps in *Section 3.5.2.1* is given in the following steps:

Step1. Perform the unsharp masking filter on F-Metric values:

$$F_{l,new}(i) = h_{unsh} \times F_l(i), \tag{3.42}$$

for $1 \leq i \leq P$ where P is the number of classes under concern.

Step2. For $F_{l,new}(i)$, eliminate the spectral bands that have zero F-metric values for every class under concern. Remaining spectral bands forms the selected set of bands.

Step3. SFS is applied to selected set of bands.

In this study, the length of the unsharp masking filter L_u is taken as 5. SVM is utilized to evaluate the classification performance of the proposed feature selection method in this section. The classification performance of the method is compared with that of SFS and CBS. In order to realize feature selection, the proposed method, SFS, and CBS work on a small set of data. The same datasets in *Section 3.5.2.1* are utilized for feature selection. The proposed method is referred as DSM-SFS.

DSM-SFS, SFS, and CBS performances for the Pavia data are shown in Figure **3.16**3.16. It is observed that DSM-SFS performance is slightly larger than that of SFS. CBS performance is lower than the others.

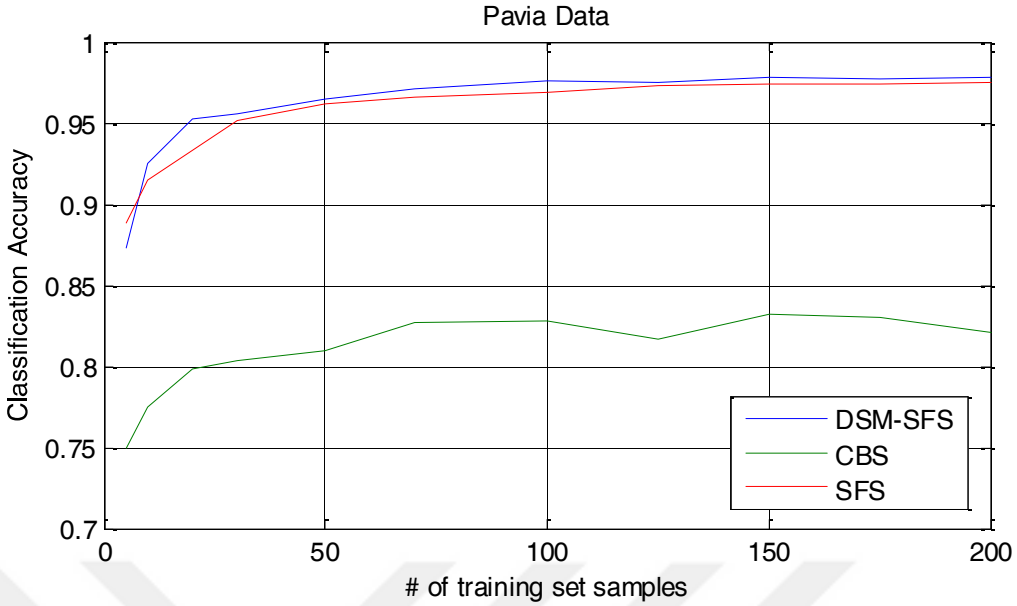

Figure 3.16: Classification Accuracy vs Number of SVM training set samples performances of DSM-SFS, CBS, and SFS for Pavia Data

DSM-SFS, SFS, and CBS performances for the Indian Pines data are shown in Figure 3.177. It is observed that DSM-SFS performance and SFS performance are almost the same for small number of training set samples. On the other hand, DSM-SFS performance is larger than that of SFS when the number of training set samples increases. CBS performance is lower than the others.

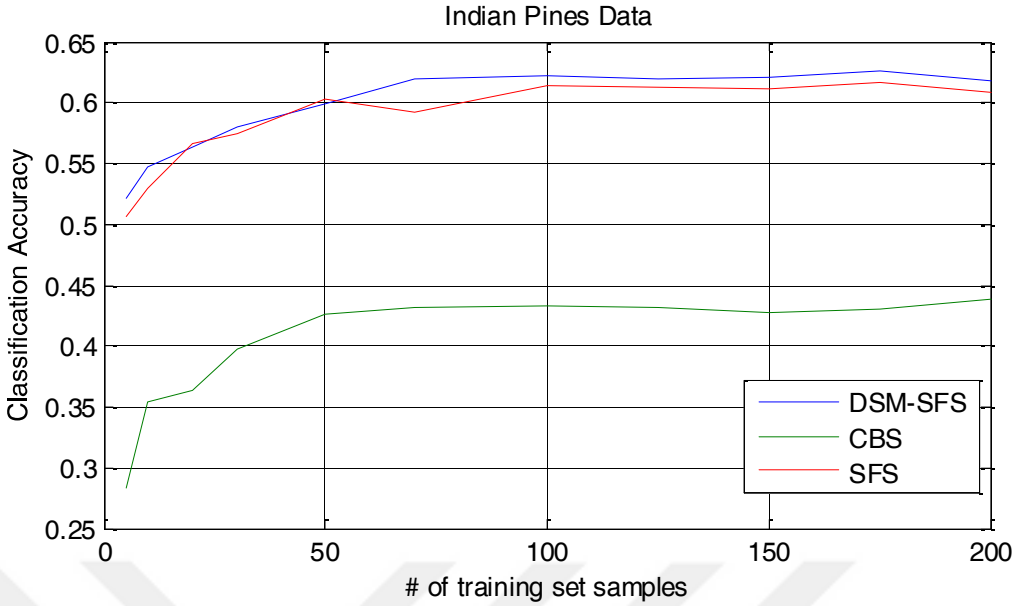

Figure 3.17: Classification Accuracy vs Number of SVM training set samples performances of DSM-SFS, CBS, and SFS for Indian Pines Data

DSM-SFS, SFS, and CBS performances for the Salinas data are shown in Figure 3.183.18. It is observed that DSM-SFS performance is almost the same as SFS performance. CBS performance is lower than the others.

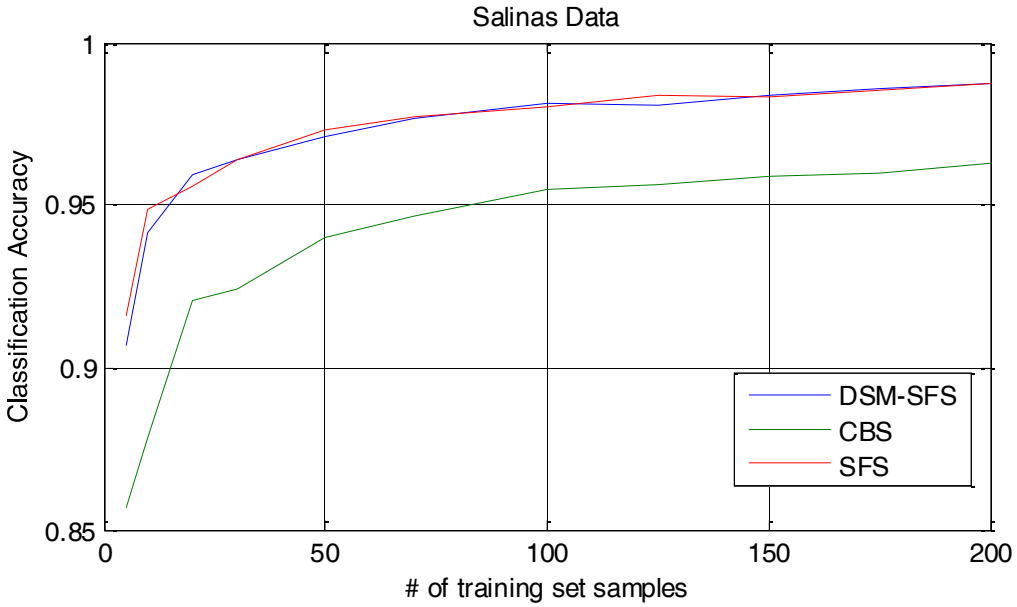

Figure 3.18: Classification Accuracy vs Number of SVM training set samples performances of DSM-SFS, CBS, and SFS for Salinas Data

To conclude the results of DSM-SFS method, it can be observed that applying DSM based feature selection method as a preprocessing increases the performance of a sub-optimum feature selection method SFS.

The proposed feature selection method DSM-SFS performs on a small size of the entire data that includes samples from classes under concern. The performance evaluation of DSM-SFS indicates that the proposed set of selected bands by DSM-SFS can be used to discriminate the classes under concern even for another region that contains not only the same classes but also contains other distinct classes. Therefore, DSM-SFS is a general framework that can define required bands for classification. When compared to the other feature selection methods mentioned in this study such as CBS and SFS, DSM-SFS has higher classification performance. Moreover, DSM-SFS is time efficient since it performs on small size of data, and it does not require to calculate classification accuracy iteratively in order to realize feature selection.

SFS works on the whole spectral bands set. In the first step of DSM-SFS the spectral bands are clustered by dominant sets approach and some of the spectral bands that do not increase the classification accuracy are eliminated before performing SFS. The classification performance of DSM-SFS is greater than SFS based on our Monte Carlo simulations mentioned above. Therefore, eliminating the spectral bands that decreases the classification accuracy before implementing SFS can increase the classification performance of SFS.

DSM and unsharp masking filter eliminate some useless bands to decrease the computational complexity of the SFS method. Therefore, DSM-SFS can be preferred instead of SFS since it increases the classification performance slightly and it decreases number of iterations of SFS.

3.7 DOMINANT SET BASED FILTERED FEATURE SELECTION METHOD (DSFM)

In Section 3.6.1, clustering performances of the spectral bands for each class obtained by DSM are utilized to increase the SFS classification performance. In this section, Dominant Set Based Filtering Method (DSFM) is proposed as a feature selection method that utilizes the clustering performances of the spectral bands. DSFM is not a preprocessing method; it is a complete band selection method by itself.

The aim is to select most valuable spectral bands for each class and to select most invaluable spectral bands that decrease the performance. It is possible that most valuable bands for one class can lead to degradation in the classification of others. This can lead to a decrease in the overall classification performance. Therefore, DSFM subtracts the most invaluable band set from the most valuable band set. The resulting spectral band set forms the selected features of our proposed feature selection method: DSFM.

The remaining part of the DSFM after the eight steps in *Section 3.5.2.1* is given in the following steps:

Step1. Perform the high unsharp masking filter H_{us} on F-Metric values:

$$F_{l,H}(i) = \times F_l(i) \tag{3.43}$$

for $1 \leq i \leq P$ where P is the number of classes under concern.

$$H_{us}[n] = \begin{cases} 1, & if\ F_{l,i}[n] > \left(\frac{1}{L_u} \sum_{m=-\frac{L_u-1}{2}}^{\frac{L_u-1}{2}} F_{l,i}[n+m] \right) \\ 0, & otherwise \end{cases} \tag{3.44}$$

where $F_{l,i} = F_l(i)$, L_u is the filter length of H_{us}

Step2. For $F_{l,H}(i)$, eliminate the spectral bands that have zero F-metric values for every class under concern.

Step3. Define a high threshold $Th_H(i) = mean\left(F_{l,H}(i)\right) + std\left(F_{l,H}(i)\right), 1 \leq i \leq P$ for each class. Select valuable spectral bands that exceeds the threshold $Th_H(i)$ for each class and combine these bands to form the most valuable spectral band set S_{val}.

Step4. Perform the low unsharp masking filter L_{us} on F-metric values:
$F_{l,L}(i) = L_{us} \times F_l(i)$, for $1 \leq i \leq P$ where P is the number of classes under concern.

$$L_{us}[n] = \begin{cases} 1, & if\ F_{l,i}[n] < \left(\frac{1}{L_u} \sum_{m=-\frac{L_u-1}{2}}^{\frac{L_u-1}{2}} F_{l,i}[n+m] \right) \\ 0, & otherwise \end{cases} \tag{3.45}$$

where $F_{l,i} = F_l(i)$, L_u is the filter length of L_{us}

Step5. For $F_{l,L}(i)$, eliminate the spectral bands that have zero F-Metric values for every class under concern.

Step6. Define a low threshold $Th_L(i) = mean\left(F_{l,L}(i)\right) - std\left(F_{l,L}(i)\right), 1 \leq i \leq P$ for each class. Select invaluable spectral bands that does not exceed the threshold $Th_L(i)$ for each class and combine these bands to form the most invaluable spectral band set S_{inval}.

Step7. Form the selected features of DSFM by removing S_{inval} band members from S_{val}.

In this study, the length of the unsharp masking filter L_u is taken as 5. SVM is utilized to evaluate the classification performance of the proposed feature selection method in this section. The classification performance of the DSFM is compared with that of SFS and CBS. In order to realize feature selection, the proposed method, SFS, and CBS work on a small set of data. The same datasets in *Section 3.5.2.1* are utilized for feature selection.

DSFM, SFS, and CBS performances for the Pavia data are shown in Figure 3.193.19. It is observed that DSFM performance is very close that of SFS. For higher training samples DSFM is slightly larger than SFS. This may result from the fact that SFS and DSFM almost reach the highest classification accuracy (> 0.95). CBS performance is lower than the others.

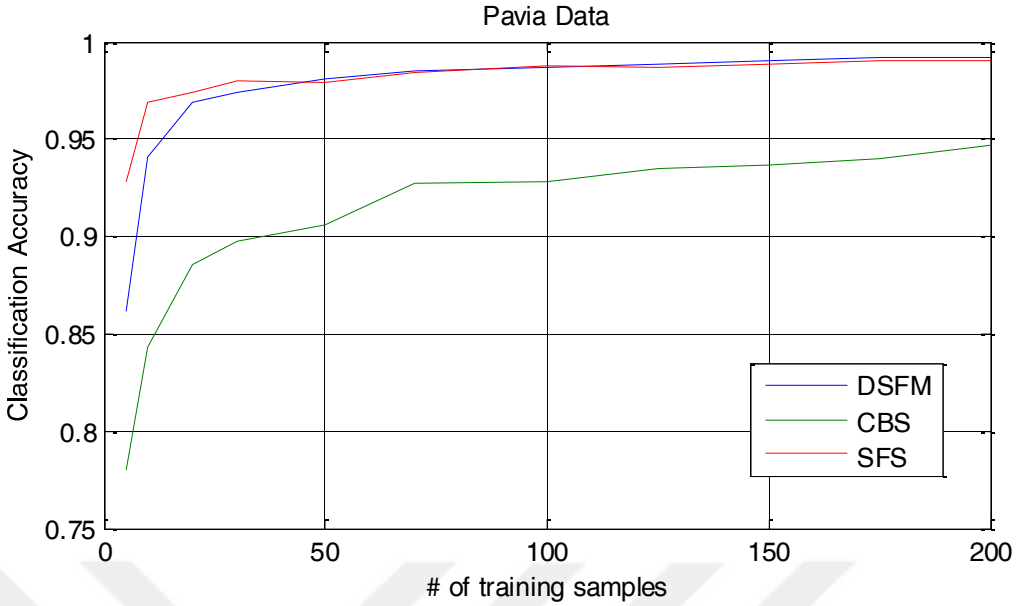

Figure 3.19: Classification Accuracy vs Number of SVM training set samples performances of DSFM, CBS and SFS for Pavia Data

DSFM, SFS, and CBS performances for the Indian Pines data are shown in Figure **3.20**3.20. It is observed that DSFM performance larger than that of SFS and CBS especially for larger numbers of training samples.

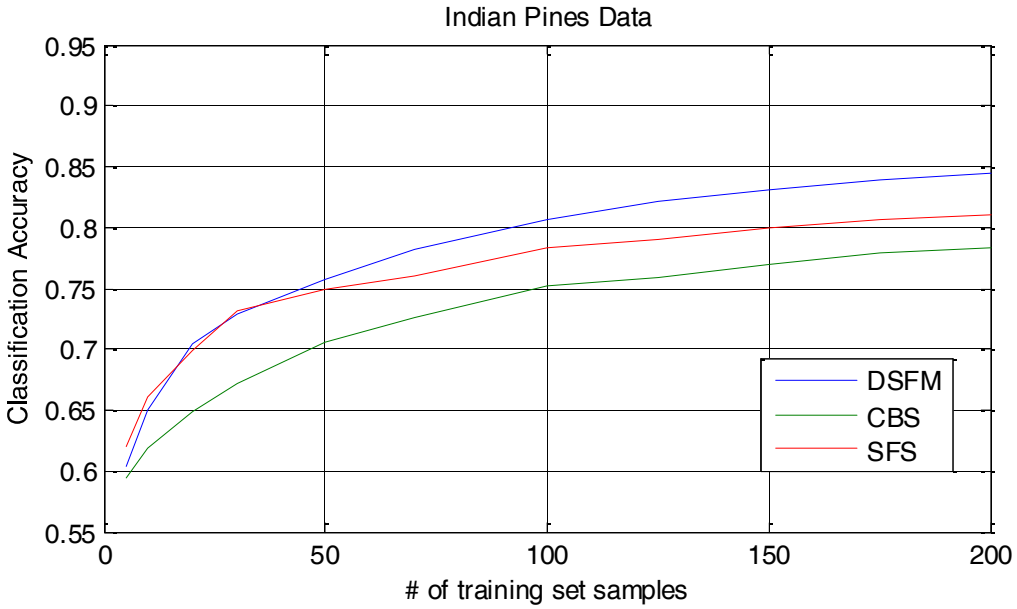

Figure 3.20: Classification Accuracy vs Number of SVM training set samples performances of DSFM, CBS and SFS for Indian Pines Data

It can be observed from the above Figures 3.19 and 3.20 that DSFM has higher classification performance over SFS and CBS. It is learned experimentally that selected spectral bands of feature selection methods DSFM, SFS, and CBS have increased classification performance with SVM classification than that of the whole hyperspectral data set. DSFM eliminates some useless bands to decrease the computational complexity. Therefore, DSFM can be preferred instead of sub optimum feature selection method SFS since it increases the classification performance, and it gets rid of the time-consuming iterations of SFS.

CHAPTER 4

CONCLUSION

Hyperspectral image contains variety of information both in the spatial domain and spectral domain. It has hundreds of narrow contiguous bands. The diversity of the hyperspectral data is huge compared to panchromatic and RGB data. This diversity leads to higher performances for clustering, classification, and detection processes. However, hyperspectral imagery has some drawbacks such as huge data volume and redundancy. Huge data volume requires large processing time, bigger data storage capacity, and higher transmission rates. In order to handle this huge data volume, it is required to have high-tech hardware and higher budgets. It is not always possible to provide the requirements of huge data volume. Therefore, dimension reduction techniques are applied to reduce the data size.

Dimension reduction techniques can be divided into two main groups: feature extraction and feature selection. In feature extraction methods, input data is transformed to a different space where a new set of features are innovated in order to increase the discriminative ability. On the other hand, in feature selection techniques, a subset of the original input data is selected based on a set of criteria. Unlike feature extraction, feature selection does not alter data during data transformation.

In this study, dimension reduction techniques are applied on hyperspectral data. Supervised and unsupervised dimension reduction methods are investigated, and some new methods are proposed. Firstly, 1D EMD is applied as a preliminary feature extraction method before PCA for further dimension reduction while maintaining the classification accuracy. Additionally, wavelet-based dimension reduction technique is applied as a preliminary feature extraction method before PCA. In the proposed method, one-dimensional sub-band decomposition is applied before principal component analysis (PCA) to increase the classification accuracy [44]. It is experimentally shown

that the classification performance of the proposed method is better than the comparative method (PCA) for higher PCA dimensions.

In this study, “dominant sets” technique is first applied to hyperspectral image processing as a clustering method. As a first step, a clustering method based on dominants sets technique is proposed [45]. This method is clustering each spectral band based on “dominant sets” technique and it evaluates the clustering performance of each band. The proposed method is time efficient since it works on a small set of training data instead of the whole hyperspectral data. Then a general dominant sets dimension reduction framework is described in Section 3.5.2. This method is called dominant sets based feature selection method (DSbFSM). DSbFSM is based on finding “dominant sets” in hyperspectral data, so that spectral bands are clustered. From each cluster, the band that reflects the cluster behavior the most is selected to form the most valuable band set in the spectra for a specific application. Results indicates that the proposed framework performs better than the state-of-the-art feature selection methods in terms of classification accuracy.

A dominant sets based pre-band selection framework is also presented to reduce the data size and to reduce the complexity of a well-known band selection method in hyperspectral imagery: sequential forward selection (SFS) [83]. The aim of the method (DSM-SFS) is to reduce the computational complexity of SFS by applying a dominant sets based pre band selection method. Besides reducing the computational complexity of SFS method, results on Pavia and Indian Pines datasets show that the proposed pre-feature selection method performs slightly better than the state-of-the-art feature selection methods in terms of classification accuracy. Therefore, DSM-SFS can be preferred instead of SFS since it increases the classification performance slightly and it decreases number of iterations of SFS.

Finally, dominant set based filtering method (DSFM) is performed on hyperspectral data. This method subtracts the most invaluable band set from the most valuable band set and these sets are selected according to clustering performance of bands based on dominant sets. DSFM performs better than the sub optimum feature selection method SFS and it gets rid of the time consuming iterations of SFS.

All the dimension reduction techniques applied in this thesis aim to reduce the dimensionality of the hyperspectral data before classification while preserving the

classification accuracy as much as possible and to achieve reduced dimension with a low computational complexity. The results of this thesis indicate that general frameworks that achieve reduced data with higher classification performances are proposed. It is evaluated that the mentioned frameworks contribute to the hyperspectral image processing literature.

REFERENCES

[1] Goetz Alexander Vane Gregg Solomon Jerry and Rock Barrett (1985) "*Imaging spectrometry for earth remote sensing*", Science, vol. 228, no. 4704, pp. 1147–1153.

[2] Lifu Zhang (2012), *APSCO Training Course on Demonstration of Remote Sensing Data Usage for Earthquake Monitoring and Evaluation*, Beijing China

[3] Varshney Pramod and Arora Manoj (2004), *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer.

[4] Vorovencii Iosif (2009), "*The Hyperspectral Sensors used in Satellite and Aerial Remote Sensing*", Bulletin of the Transilvania University of Braşov, Vol. 2 (51) - Series II.

[5] Thenkabail Prasad and Smith Ronald (2000), "*Hyperspectral vegetation indices and their relationships with agricultural crop characteristics*", Elsevier Remote sensing of Environment, Vol. 71, Iss. 2, pp. 158-182.

[6] http://modis.gsfc.nasa.gov/

[7] Liao Liang (2000) "*Radiometric Performance Characterization of the Hyperion Imaging Spectrometer Instrument*", Proc. Optical Science and Technology Symposium, Earth Observing Systems V, SPIE 1435.

[8] Kaufmann Hermann Segl Karl Guanter Luis Hofer Stephan Mueller Andreas Richter Rudolf Bach Heike Hostert Patrick and Chlebek Christian "Environmental Mapping and Analysis Program (EnMAP) - Recent Advances and Status", IEEE Geoscience and Remote Sensing Symposium, IGARSS 2008, vol. 4, pp. 109-112, 2008.

[9] Galeazzi Claudio Sacchetti Andrea Cisbani Andrea and Babini Gianni (2008), "*The Prisma Program*", IEEE Geoscience and Remote Sensing Symposium, IGARSS 2008, vol. 4, pp. 105–108.

[10] Chien Stieve Silverman Dorothy and Davies Ashley (2009) "*Onboard Science Processing Concepts for the HyspIRI Mission*", IEEE Intelligent Systems, vol. 24, no. 6, pp. 12–19.

[11] Bellman Richard (1961), "*Adaptive control process: A guided tour*" Princeton University Press, 255 pp.

[12] Hughes Gordon (1968), "*On the mean accuracy of statistical pattern recognizers*," IEEE Trans. On Information Theory, vol. 14, no. 1, pp. 55–63.

[13] Phenomenology and Exploitation of Thermal Hyperspectral Sensing (SET-190), NATO,RTO,source: http://www.cso.nato.int/ACTIVITY_META.asp?ACT=2233

[14] Gualtieri Anthony and Cromp Robert (1998), "*Support Vector Machines for Hyperspectral Remote Sensing Classification*", SPIE 27th Applied Imagery Pattern Recognition Workshop, pp. 221–232.

[15] Dundar Murat and Landgrebe David (2004), "*A Cost-Effective Semisupervised Classifier Approach with Kernels*", IEEE Transactions on Geoscience and Remote Sensing, vol. 42, pp. 264–270.

[16] Demir Begüm ve Ertürk Sarp (2007), "*Hyperspectral Image Classification Using Relevance Vector Machines*", IEEE Geoscience and Remote Sensing Letters, vol. 4, pp. 586-590.

[17] Borghys Dirk Shimoni Michal and Perneel Christiaan (2005), "*Supervised Classification of Hyperspectral Images Using a Combination of Spectral And Spatial Information*", SPIE Image and Signal Processing for Remote Sensing XI, vol. 5982.

[18] Demir Begüm (2010),"*Hiperspektral Görüntülerin Yüksek Doğruluklu Sınıflandırılması*" [Hyperspectral Image Classification with High Accuracy], Kocaeli University, PhD Thesis, 194 pp.

[19] Chiang Shao-Shan Chang Chein-I and Ginsberg Irwing (2000) "*Unsupervised Hyperspectral Image Analysis Using Independent Component Analysis*", IEEE International Geoscience and Remote Sensing Symposium, vol. 7, pp. 3136-3138.

[20] Du Quian (2009), "*Decision Fusion for Classifying Hyperspectral Imagery with High Spatial Resolution*", SPIE Electronic Imaging & Signal Processing.

[21] Campbell Barbara (1996), "*Introduction to Remote Sensing*", The Guilford Press, Second Edition, New York.

[22] Haq Ikram-Ul and Xu Xia (2008), "*A new approach to band clustering and selection for hyperspectral imagery*", IEEE 9th International Conference on Signal Processing ICSP, pp. 1198-1202.

[23] Yang Hsiuhan and Du Quian (2011), "*Fast Band Selection for Hyperspectral Imagery*", IEEE 17th International Conference on Parallel and Distributed Systems (ICPADS), pp. 1048-1051.

[24] Du Quian Bioucas-Dias Jose and Plaza Antonio (2012), "*Hyperspectral band selection using a collaborative sparse model*", IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3054-3057.

[25] Bajcsy Peter and Groves Peter (2004), "*Methodology for Hyperspectral Band Selection*", Photogrammetric Engineering and Remote Sensing Journal, vol. 70, pp. 793-802.

[26] Du Quian Richard Hart Peter and Stork David (2001), "*Pattern Classification*", Second Edition, Wiley-Interscience.

[27] Yang Hsiuhan Du Quian and Sheng Yehua (2011), "A*n Efficient Method for Supervised Hyperspectral Band Selection* ", IEEE Geoscience and Remote Sensing Letters, vol. 8, Iss. 1, pp.138-142.

[28] Wei Wei and Du Quian (2012), "*Fast supervised hyperspectral band selection using graphics processing unit* ", Journal of Applied Reomte Sensing, vol. 6.

[29] Rodarmel Craig and Shan Jie (2002), "*Principal Component Analysis for Hyperspectral Image Classification*" Surveying and Land Information Systems, vol. 62, no. 2, pp. 115–122.

[30] Farrell Ruth and Mersereau Russell (2005), "*On the Impact of PCA Dimension Reduction for Hyperspectral Detection of Difficult Targets*" IEEE Trans. On Geoscience and Remote Sensing, vol. 2, no. 2, pp. 192–195.

[31] Agarwal Abhishek El-Ghazawi Tarek El-Askary Hesham and Le-Moigne Jacqueline (2007), "*Efficient Hierarchical-PCA Dimension Reduction for Hyperspectral Imagery*," 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 353–356.

[32] Zhu Congshan and Yang Xiaodong (1998) "*Study of remote sensing image texture analysis and classification using wavelet*" International Journal of Remote Sensing, vol. 19, no. 16, pp. 3197–3203.

[33 Pu Ruiliang and Gong Peng (2004) "*Wavelet Transformation applied to EO-1 hyperspectral data for forest LAI and crown closure mapping*" ELSEVIER Remote Sensing of Environment, vol. 31, pp. 212–224.

[34] Bruce Lori Koger Cliff and Li Jiang (2002), "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction ", IEEE Transactions on Geoscience and Remote Sen, vol. 40, Iss. 10.

[35] Salehi Bahram and Zouj Mohammad Javad Valadan (2004), "*Wavelet Spectral Analysis for Automatic Reduction of Hyperspectral Imagery*", http://www.gisdevelopment.net/technology/ip/mi04013.htm

[36] Hsu Pai-Hui (2007), "*Feature extraction of hyperspectral images using wavelet and matching pursuit*," ELSEVIER ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, no. 2, pp. 78–92.

[37] Hsu Pai-Hui and Y. Tseng, (2002), "*Feature extraction of hyperspectral data using the Best Wavelet Packet Basis*," IEEE IGARSS 2002, vol. 3, pp. 1167–1669.

[38] Bruce Lori Koger Cliff and Li Jiang (2002), "*Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction*," IEEE Transactions on Geoscience And Remote Sensing, vol. 40, no. 10, pp. 2331–2338.

[39] A. Linderhed, (2004), "*Adaptive Image Compression with Wavelet Packets and Empirical Mode Decomposition,*" published by LTAB.

[40] K. L. Wu and P. F. Hsieh, (2005), "*Empirical mode decomposition for dimensionality reduction of hyperspectral data*," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2, pp. 1241-1244.

[41] AVIRIS NW Indiana's Indian Pines 1992 data set [Online]. Source:ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.tif.zip (data), ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/ThyFiles.zip (ground truth), Acces Date: 07 May 2012.

[42] Demir Begüm and Erturk Sarp (2010), "*Empirical Mode Decomposition of Hyperspectral Images for Support Vector Machine Classification*," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, Iss. 11, pp. 4071-4084.

[43] Teke Mustafa Deveci Hüsne Seda, Haliloğlu Onur Gurbuz Sevgi, and Sakarya Ufuk (2013), "*Hyperspectral Remote Sensing Applications in Agriculture*," 6th International conference on Recent Advances in Space Technologies.

[44] Haliloğlu Onur Sakarya Ufuk and Töreyin Behçet Uğur (2013), "*Sub-band decomposition based supervised feature extraction for hyperspectral image classification*," MUSCLE International Workshop on Computational Intelligence for Multimedia Understanding.

[45] Haliloğlu Onur Sakarya Ufuk and Töreyin Behçet Uğur (2015), "Evaluation of clustering performance of hyperspectral bands," IEEE 2015 23rd Signal Processing and Communications Applications Conference (SIU).

[46] Gormus Esra Tunç Canagarajah Nishan and Achim Alin Marian (2011), "*Dimensionality Reduction of Hyperspectral Images with Wavelet Based Empirical Mode Decomposition*," 18th IEEE International Conference on Image Processing (ICIP), pp. 1709-1711.

[47] Gormus Esra Tunç Canagarajah Nishan and Achim Alin Marian (2012), "*Dimensionality Reduction Of Hyperspectral Images Using Empirical Mode*

Decomposition And Wavelets," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, Iss. 6, pp. 1821-1830.

[48] Chang Chih-Chung and Lin Chih-Jen (2011), "*LIBSVM: a library for support vector machines*", ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 2, pp. 1-27.

[49] Kopsinis Yannis, http://perso.ens-lyon.fr/patrick.flandrin/emd.html, 2008

[50] Bosch Edward Howard and Lin Jeng Eng (2003), "*Wavelet-based dimension reduction for hyperspectral image classification*," Proc. SPIE 5093, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, vol. 5093, pp. 57–69.

[51] Daubechies Ingrıd (1992), "*Ten Lectures on Wavelets,*" Society for Industrial and Applied Mathematics.

[52] Heisenberg Werner (1927), "*Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik,*" Zeitschrift für Physik, vol. 43, pp. 172-198.

[53] Kumar Praveen (1997), "*Wavelet Analysis for Geophysical Applications*," Reviews Of GEOPHYSICS, vol. 35, no. 4, pp. 385-412.

[54] Mallat Stephane (2009), "*A Wavelet Tour Of Signal Processing*," Elsevier Inc., Third Edition.

[55] Jain Yogendra Kumar and Shah Dharambhai, (2007), "*Performance Analysis and Comparison of Wavelet Families Using for Image Compression*," International Journal of Soft Computing, vol. 2, issue 1, pp. 161-171.

[56] Patel Akash Patel Ravi Mangal and Jain Suman (2013), "*Performance Analysis of Wavelet Analysis*," International Journal of Innovative Technology, vol. 2, issue 12.

[57] Stolojescu Crisan Railean Ion and Isar Alexandru (2010), "*Comparison of Wavelet Families with Application to Wimax Traffic Forecasting*," 12th International Conference on Optimization of Electrical and Electronic Equipment, pp. 932-937.

[58] Vetterli Martin Kovacevic Jelena and Goyal Vivek (2010), "*Fourier and Wavelet Signal Processing,*" Noncommercial Book (http://zolotarev.fd.cvut.cz/static/mni/Fourier_Wavelet.pdf).

[59] Avcı Engin (2008), "*Comparison of Wavelet Families for texture classification by using wavelet packet entropy adaptive network Based Fuzzy inference system*," Journal of Applied Soft Computing, vol. 8, iss. 1, pp. 225-231.

[60] Coifman Raphael Wickerhauser Miaden Victor (2013), "*Entropy based algorithms for best basis selection*," International Journal of Innovative Technology, vol. 2, issue 12.

[61] Amiri Gholamreza Ghodrati Asadi Amir (2009), "*Comparison of different Methods of Wavelet and Wavelet packet transform in Processing ground Motion records*," International Journal of Civil Engineering, vol. 7, no. 4.

[62] Coifman Ronald Meyer Yves Quake Steven and Wickerhauser Victor (1992), "*Signal Processing and Compression with Wavelet Packets*,".

[63] Dennison Philip and Roberts Dar (2003), "*The Effects of Vegetation Phenology On Endmember Selection And Species Mapping In Southern California Chaparral*," Remote Sensing of Environment, vol. 87, pp. 295-309.

[64] Thenkabail Prasad Lyon John and Huete Alfredo (2011), "*Advances in Hyperspectral Remote Sensing Of Vegetation And Agricultural Crops*," Hyperspectral Remote Sensing of Vegetation Book, Chapter 2, CRC Press.

[65] Thenkabail Prasad Mariotto Isabella Gumma Murali Krishna Landis David and Huemmrich Karl (2013), "*Selection of Hyperspectral Narrowbands (HNBs) And Composition Of Hyperspectral Two-band Vegetation Indices (HVIs) For Biophysical Characterization And Discrimination Of Crop Types Using Field Reflectance And Hyperion/EO-1 Data*," IEEE Journal Of Selected Topics In Applied Earth Observations and Remote Sensing, vol.6, issue 2, pp. 427-439, CRC Press.

[66] Nidamanuri Rama Rao and Zbell Bernd (2012), "*Existence of Characteristic Spectral Signatures For Agricultural Crops –Potential For Automated Crop Mapping By Hyperspectral Imaging,*" Geocarto International, vol. 27, no. 2.

[67] Camps-Valls Gustau Tuia Devis Bruzzone Lorenzo and Benediktsson John (2014), " *Advances in Hyperspectral Image Classification: Earth Monitoring with Statistical Learning Methods* ", IEEE Signal Processing Magazine, vol. 31, Issue: 1, pp. 45-54.

[68] Plaza Antonio Benediktsson John Boardman Joseph Brazile Jason Bruzzone Lorenzo Camps-Valls and Gustau Chanussot Joselyn (2009), "*Recent advances in techniques for hyperspectral image processing*," ELSEVIER Remote Sensing of Environment, vol. 113, pp. 110-122.

[69] Goetz Alexander (2009), "*Three decades of hyperspectral remote sensing of the Earth: A personal view*," ELSEVIER Remote Sensing of Environment, vol. 113, pp. 5-16.

[70] Quesada-Barriuso Pablo Arguello Francisco Heras Dora Blanco (2014), " *Spectral–Spatial Classification of Hyperspectral Images Using Wavelets and*

Extended Morphological Profiles", Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, vol. 7, issue: 4.

[71] Chang Chih-Chung and Wang Shengjia (2006), "*Constrained band selection for hyperspectral imagery*," Geoscience and Remote Sensing, IEEE Transactions on, vol 44, no. 6, pp. 1575-1585, June.

[72] Du Quian ve Yang Hsiuhan, (2008), "*Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis*," Geoscience and Remote Sensing Letters, IEEE, vol 5, no. 4, pp. 564-568.

[73] Martinez-Uso Adolfo Pla Filiberto Sotoca Jose and Garcia-Sevilla Pedro (2007), "*Clustering-Based Hyperspectral Band Selection Using Information Measures*," Geoscience and Remote Sensing, IEEE Transactions on, vol. 45, no. 12, pp. 4158-4171, Dec.

[74] Haq Ihsan and Xu Xia (2008), "*A new approach to band clustering and selection for hyperspectral imagery,*" Signal Processing, 2008. ICSP 2008. 9th International Conference.

[75] Datta Aloke Ghosh Sanjay and Ghosh Aniruddha (2012), "*Clustering based band selection for hyperspectral images,*" Communications, Devices and Intelligent Systems (CODIS), 2012 International Conference.

[76] Bunke Horst (2008), "*IAM Graph Database Repository for Graph Based Pattern Recognition and Machine Learning*," Structural, Syntactic, and Statistical Pattern Recognition, vol. 5342, N. da Vitoria Lobo, T. Kasparis, F. Roli, J. Kwok, M. Georgiopoulos, G. Anagnostopoulos ve M. Loog, Dü, Springer Berlin Heidelberg, pp. 287-297.

[77] Shi Jing-jing (2000), "*Normalized cuts and image segmentation,*" Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8, pp. 888-905.

[78] Pavan Massimiliano and Pelillo Marcello (2007), "*Dominant Sets and Pairwise Clustering*," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol 29, no. 1, pp. 167-172.

[79] Sakarya Ufuk Telatar Ziya and Alatan Aydın (2012), "*Dominant sets based movie scene detection,*" ELSEVIER 29, vol. 92, issue: 1, pp. 107-119.

[80] "*Hyperspectral Remote Sensing Scenes*," 2014. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[81] Kittler Josef (1986), "*Feature selection and extraction*," Handbook of pattern recognition and image processing, pp. 59-83.

[82] Foody Giles (2002), "*Status of land cover classification accuracy assessment*," Remote Sensing of Environment, vol. 80, no. 1, pp. 185-201.

[83] Haliloğlu Onur ve Gazi Orhan (2022), "*Dominant Sets Based Pre-Feature Selection Method for Hyperspectral Data*," Cankaya University Journal of Science and Engineering, vol. 19, Iss. 1, pp. 51-61.

[84] Ramponi Giovanni Strobel Norbert and Mitra Sanjit (1996), "*Nonlinear unsharp masking methods for image contrast enhancement*," J. Electron. Imag., vol. 5, pp. 353–366.