

CLUSTER-BASED VIRTUAL BAND CONSTRUCTION FOR HYPERSPECTRAL
CLASSIFICATION

by

Ethan Michael Glenn

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Electrical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2026

©COPYRIGHT

by

Ethan Michael Glenn

2026

All Rights Reserved

DEDICATION

I dedicate this thesis to my mother, Amanda Duran, and my grandparents, Michael and Sherry Hanson, for always inspiring me to be the best I can be. To my friends, thank you for your support and belief in me. To my high school teachers, Lefonso Castillo and Steven Salinas, thank you for creating the building blocks for my love of science and math.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Bradley M. Whitaker, for his support and guidance in all things signal processing and machine learning. I would also like to thank the rest of my committee, Dr. Ross K. Snider, for pushing my skills and helping me discover where my interests lay, and Dr. Riley Logan, for his insight into the optics side of hyperspectral image processing. I would like to acknowledge Mahmad Isaq Karankot and Piper Hoxie for their collaboration on this research. I would also like to thank the faculty and staff of the Departments of Electrical and Computer Engineering, Physics, and Science Education at Montana State University, as well as my student colleagues within these departments.

Computational efforts were performed on the Tempest High Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure (RRID:SCR_026229) at Montana State University.

This material is based upon work supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-2242802. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

1. INTRODUCTION	1
Research Questions	3
Contributions	4
Virtual Band Overview	5
Implementation and Reproducibility	6
Thesis Organization	7
2. BACKGROUND	8
Introduction	8
Hyperspectral Imaging Fundamentals	9
Spectral Regions and Physical Interpretation	10
Datasets	12
Spectral Redundancy and Class Separability	14
Challenges in Hyperspectral Classification	15
Dimensionality Reduction in Hyperspectral Imaging	16
Band Selection Methods	18
Clustering-Based Grouping and Motivation for Merging	19
Classifiers	20
Interpretability in Reduced Spectral Representations	21
3. BAND CLUSTERING	23
Introduction	23
Data Preprocessing	24
Band Representation and Similarity	25
Clustering Methods	25
4. WEIGHTED MERGING	28
Introduction	28
Band Merging Algorithms	29
Uniform Merging	29
Entropy-Based Merging	30
Inverse Redundancy Merging	30
PCA-Based Merging	31
Merging Strategy Summary	32

TABLE OF CONTENTS – CONTINUED

5. PIXEL-LEVEL CLASSIFICATION AND RESULTS	33
Introduction	33
Classification and Evaluation.....	34
5-Band Feature Reporting	34
Results.....	35
6. INTERPRETABILITY OF VIRTUAL BANDS	43
Introduction	43
Cluster Weight Analysis.....	43
Practical Implications for Spectral Binning and Sensor Design.....	52
7. THESIS CONCLUSION.....	55
Future Work.....	57
REFERENCES CITED.....	60
APPENDIX: FULL CLASSIFICATION RESULTS.....	67
BOTSWANA RESULTS	68
INDIAN PINES.....	73
PAVIA	78
MONTANA	83

LIST OF TABLES

Table	Page
5.1 Botswana Performance – 5 Bands.....	39
5.2 Indian Pines Performance – 5 Bands	40
5.3 Pavia Performance – 5 Bands.....	41
5.4 Montana Performance – 5 Bands.....	42
A.1 Botswana Performance – 10 Bands.....	68
A.2 Botswana Performance – 20 Bands.....	69
A.3 Botswana Performance – 30 Bands.....	70
A.4 Botswana Performance – 40 Bands.....	71
A.5 Botswana Performance – 50 Bands.....	72
A.6 Indian Pines Performance – 10 Bands.....	73
A.7 Indian Pines Performance – 20 Bands.....	74
A.8 Indian Pines Performance – 30 Bands.....	75
A.9 Indian Pines Performance – 40 Bands.....	76
A.10 Indian Pines Performance – 50 Bands.....	77
A.11 Pavia Performance – 10 Bands.....	78
A.12 Pavia Performance – 20 Bands.....	79
A.13 Pavia Performance – 30 Bands.....	80
A.14 Pavia Performance – 40 Bands.....	81
A.15 Pavia Performance – 50 Bands.....	82
A.16 Montana Performance – 10 Bands	83
A.17 Montana Performance – 20 Bands	83
A.18 Montana Performance – 30 Bands	84
A.19 Montana Performance – 40 Bands	84
A.20 Montana Performance – 50 Bands	85

LIST OF FIGURES

Figure	Page
1.1 Virtual Band Construction Pipeline	6
2.1 Spectral Regions	11
3.1 Virtual Band Clustering Overview.....	23
4.1 Virtual Band Merging Overview.....	28
5.1 Virtual Band Clustering Overview.....	33
5.2 Best RF and KNN Accuracy at $k = 5$	38
5.3 Best RF Accuracy Across All k	38
6.1 Botswana E-FDPC and Entropy Clustering-Merging – 5 Bands	44
6.2 Botswana E-FDPC and Entropy Clustering-Merging – 5 Bands (Zoomed)	46
6.3 Indian Pines UMRMR and PCA Clustering-Merging – 5 Bands	46
6.4 Pavia K-Means and PCA Clustering-Merging – 5 Bands.....	48
6.5 Montana UMRMR and PCA Clustering-Merging – 5 Bands	49
6.6 Montana UMRMR and PCA Clustering-Merging – 5 Bands (Zoomed)	51

NOMENCLATURE

AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BDPC	Band Density Prominence Clustering
E-FDPC	Entropy-based Fast Density Peak Clustering
HSI	Hyperspectral Image
KNN	k -Nearest Neighbors
LDA	Linear Discriminant Analysis
MI	Mutual Information
mRMR	Minimum Redundancy Maximum Relevance
NIR	Near-Infrared
PCA	Principal Component Analysis
RF	Random Forest
ROSIS	Reflective Optics System Imaging Spectrometer
SNNC	Shared Nearest Neighbor Clustering
SWIR	Shortwave Infrared
UMRMR	Unsupervised Max-Relevance, Min-Redundancy
VNIR	Visible and Near-Infrared

ABSTRACT

Hyperspectral image classification is challenged by high spectral dimensionality, inter-band redundancy, and limited labeled training data. This thesis evaluates virtual band construction as a dimensionality reduction strategy in which spectrally similar bands are grouped into clusters and merged into representative features using weighted linear combinations. Four datasets – Botswana, Indian Pines, Pavia University, and a field-collected Montana burn plot – were reduced to five virtual bands using five clustering methods and four merging strategies: uniform, entropy-based, inverse redundancy, and PCA-based. Classification performance was assessed using Random Forest and k-Nearest Neighbor classifiers under an 80/20 train-test split. Merging-based representations matched or exceeded the accuracy of conventional band selection baselines across all datasets, with the most pronounced improvements on spectrally complex scenes such as Indian Pines, where UMRMR clustering with PCA-based merging achieved 0.7455 RF accuracy compared to 0.5684 for selection without merging. Cluster weight analysis showed that merging strategies distribute contributions across neighboring wavelengths rather than isolating single bands, preserving intra-cluster spectral structure while maintaining a direct connection to the original wavelength measurements.

INTRODUCTION

Hyperspectral imaging (HSI) captures images across a large number of narrow spectral bands using a hyperspectral sensor. The number of bands, with wavelengths spanning from ultraviolet to near-infrared, can range from several hundred to thousands. Stacking these images forms a three-dimensional data cube, where two dimensions correspond to spatial coordinates and the third corresponds to wavelength. The resulting spectral signature at each pixel carries detailed information about material composition and condition, enabling pixel-level discrimination that is often not achievable with conventional RGB or multispectral imaging. Applications include environmental monitoring [31, 47], military target detection [26], and noninvasive cancer detection [37].

These advantages come with a practical cost. The high number of spectral features increases computational burden during both training and inference, particularly in experiments that require repeated classifier training across many configurations. Hyperspectral bands are also frequently correlated, especially in contiguous wavelength regions, so many bands provide largely redundant information. At the same time, labeled training samples are often scarce relative to the number of features, which raises the risk of overfitting and reduces generalization to unseen scenes. In field settings, these issues are compounded by sensor noise, illumination variability, and class imbalance, making it difficult to identify a compact spectral representation that holds up across conditions.

Dimensionality reduction addresses these challenges and generally takes one of two forms: feature extraction or band selection. Feature extraction methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) project the data into a lower-dimensional subspace. While effective for classification, the resulting features no

longer correspond to physical wavelengths, which is problematic when the goal includes understanding spectral contributions or when downstream tasks require wavelength-level meaning. Band selection avoids this by retaining a subset of original channels, preserving both discriminative information and spectral interpretability. Recent work has pushed toward more principled selection criteria, including sparse learning methods [45], ranking-based clustering techniques [23], and information-theoretic approaches that account for inter-band redundancy [33]. Survey work [44] reflects growing interest in clustering-based grouping and fusion strategies as a compromise between compact representation and interpretability.

A shared limitation of many selection-based approaches is that they identify the “best” individual bands and discard the rest. This can be a poor strategy when discriminative information is distributed across several correlated bands, or when subtle absorption features span multiple adjacent wavelengths. When clustering is used to organize bands before selection, retaining only one representative per group discards whatever complementary information the remaining cluster members contain. These observations motivate an alternative: rather than selecting a single band from each group, combine all bands within a cluster into a representative feature that retains contributions from the full group.

Prior work along these lines includes split-and-merge approaches with uniform merging [40] and Expectation Maximization with weighted merging [39], both of which show that merging can balance accuracy with computational efficiency. What remains limited, however, is a systematic comparison of virtual band methods against conventional selection approaches across multiple datasets, multiple clustering strategies, and multiple merging rules under a consistent evaluation protocol.

This work addresses that gap by introducing a *virtual band merging framework* in which spectrally similar bands are first grouped into clusters and then merged into representative virtual bands using multiple weighting strategies. Rather than enforcing a single-band representative per cluster, the merging step allows all cluster members to contribute

proportionally to the final representation based on their individual information content or spectral characteristics. Using four hyperspectral datasets, including a Montana burn plot collection [25] and three standard benchmarks, we generate k reduced features and compare performance across five clustering methods and four merging strategies, varying k from 50 down to 5. Results are also compared against selection-style baselines from E-FDPC and UMRMR without merging, representing conventional single-band-per-group behavior. Classification experiments use Random Forest (RF) and k -Nearest Neighbors (kNN) under consistent training protocols, with an 80/20 train-test split and performance measured by accuracy, precision, recall, and F1-score.

Beyond benchmarking, this study examines when merging improves robustness and when traditional selection is sufficient. For the best-performing configuration on each dataset, we report the reduced features along with their cluster membership and normalized weight vectors, providing a direct link between classification performance and spectral interpretation. By evaluating both widely used benchmark scenes and real field-collected data, this work establishes a foundation for using virtual bands in practical hyperspectral classification pipelines where both accuracy and spectral interpretability are required.

Research Questions

To evaluate the capability of virtual band merging as a dimensionality reduction strategy for hyperspectral classification, several key questions must be addressed. Previous work has demonstrated that both uniform and weighted merging can be effective in individual hyperspectral settings, but a systematic comparison across multiple clustering strategies, merging rules, datasets, and reduction targets remains absent from the literature. This research investigates the following questions:

- Can merging-based virtual band representations match or exceed the classification

accuracy of conventional single-band selection approaches across diverse hyperspectral datasets?

- Which combinations of clustering method and merging strategy produce the most consistent performance across different scene types and spectral characteristics?
- How does classification performance scale as the number of virtual bands k decreases from 50 to 5, and does the relative ordering of methods remain stable across reduction targets?
- What do the cluster weight distributions reveal about which spectral regions drive class discrimination, and how does this vary across datasets and merging strategies?
- Can the learned weight distributions be interpreted in physically meaningful terms, and do they offer practical guidance for downstream tasks such as spectral binning and sensor design?

Contributions

To address the primary research questions, the following contributions are made in this thesis:

- A virtual band merging framework that combines five clustering methods with four weighting strategies to produce reduced spectral representations that retain contributions from all cluster members rather than discarding all but one.
- A systematic evaluation of all clustering and merging combinations across four hyperspectral datasets, two classifiers, and six reduction targets ranging from $k = 50$ to $k = 5$, under a consistent and publicly available experimental pipeline.

- A direct comparison between merging-based virtual bands and conventional band selection baselines from E-FDPC and UMRMR, providing a clear empirical account of when and how much merging improves over strict selection.
- A cluster weight analysis for the best-performing configuration on each dataset, connecting classification performance to specific wavelength regions and providing a physically interpretable account of what each virtual band represents.
- A discussion of the practical implications of the weight distributions for spectral binning and sensor design, with particular attention to how data-driven merging recovers physically meaningful spectral boundaries without requiring prior knowledge of the spectral regions.

Virtual Band Overview

The proposed framework reduces hyperspectral dimensionality by combining band clustering with cluster merging to produce a compact set of representative virtual bands. The workflow has three stages: (1) partitioning spectral bands into groups using one of several clustering approaches, (2) merging each group into a single virtual band using a defined weighting strategy, and (3) evaluating classification performance on the resulting reduced representation. Figure 1.1 illustrates this pipeline.

The pipeline converts the original hyperspectral cube into a band-vector representation, partitions the bands into k clusters, and compresses each cluster into a single virtual band. The output is a reduced feature matrix of size $N \times k$, which is passed directly to the classifier. This structure isolates two distinct sources of variation in the final representation: the clustering rule, which determines which bands belong to the same group, and the merging rule, which determines how information is aggregated within each group. Studying these two components jointly, rather than in isolation, is central to the experimental design.

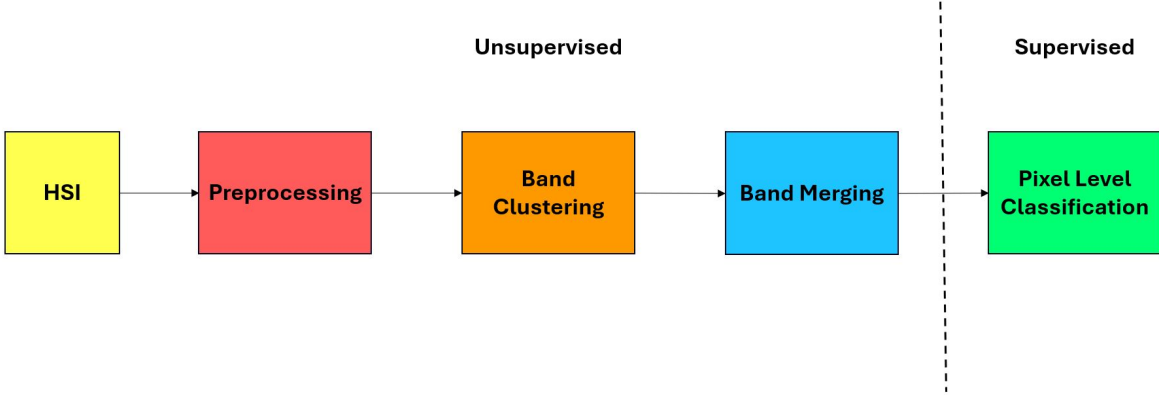


Figure 1.1: Virtual Band Construction Pipeline

Implementation and Reproducibility

All experiments were implemented under a consistent pipeline covering data loading, normalization, band clustering, cluster merging, and classifier training. Classifier hyperparameters were held constant across datasets within each model family, and the same train-test split protocol was applied throughout. The full implementation is publicly available at: github.com/BMW-lab-MSU/hvb-pipeline.

Experiments were conducted across six reduction targets: $k \in \{5, 10, 20, 30, 40, 50\}$ virtual bands. This range spans from aggressive compression, where the full spectrum is reduced to five representative features, to more moderate reduction at $k = 50$, which serves as a reference point for evaluating how much performance changes as dimensionality decreases. Results are reported for $k = 5$ throughout this thesis, as this represents the most demanding reduction target and provides the clearest test of whether merging can preserve classification-relevant spectral structure under extreme compression. Performance at larger values of k was broadly consistent with the $k = 5$ trends, confirming that the relative ordering of clustering-merging combinations is stable across reduction budgets rather than specific to the most aggressive setting.

Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 provides background on hyperspectral imaging fundamentals, spectral regions and their physical interpretation, the datasets used in this study, and the dimensionality reduction methods and classifiers relevant to this work. Chapter 3 describes the band clustering stage of the pipeline, including data preprocessing and the five clustering methods evaluated. Chapter 4 covers the weighted merging stage, deriving and motivating the four merging strategies used to combine cluster members into virtual bands. Chapter 5 presents the pixel-level classification results across all datasets, clustering methods, merging strategies, and reduction targets, with comparison to conventional band selection baselines. Chapter 6 analyzes the interpretability of virtual bands through cluster weight visualizations and discusses the practical implications of the learned weight distributions for spectral binning and sensor design. Chapter 7 concludes the thesis by summarizing the key findings, connecting results back to the research questions, and identifying directions for future work.

BACKGROUND

Introduction

This chapter establishes the technical and scientific context for the virtual band merging framework developed in this thesis. It begins with the physical foundations of hyperspectral imaging, covering how sensors acquire spectral data, how that data is structured, and what the key spectral regions of the electromagnetic spectrum reveal about surface materials. From there it surveys common approaches to hyperspectral dimensionality reduction, including both feature extraction and band selection methods, and reviews clustering-based strategies that organize bands into groups before reduction. The chapter concludes by motivating virtual band construction as a middle ground between these two paradigms, one that retains the physical interpretability of band selection while compressing correlated spectral information rather than discarding it. The datasets used in this study are introduced in the context of the challenges they present.

Raw hyperspectral dimensionality does not translate directly into an equivalent amount of independent information. Adjacent bands sample the same underlying surface properties across wavelengths that are close enough that the material’s spectral response changes only gradually, which produces strong inter-band correlation. This correlation is a property of the scene’s spectral structure rather than of the sensor itself. This correlation supports robust material identification in some contexts, but it also inflates the effective dimensionality of the feature space and can degrade learning algorithms that are sensitive to ill-conditioned inputs.

In many practical pipelines, reducing dimensionality is not simply a computational convenience, it is often a modeling requirement. When the number of spectral features far exceeds the number of labeled training pixels, classifiers risk learning dataset-specific artifacts rather than genuine class-discriminative structure. This problem worsens when classes are

imbalanced or when some classes occupy small spatial regions, since classifiers then have fewer training examples for minority classes and are more likely to learn dataset-specific noise rather than generalizable spectral structure. For this reason, band reduction strategies are commonly evaluated not just on peak accuracy, but on how consistently they perform across datasets and modeling choices.

Hyperspectral Imaging Fundamentals

Hyperspectral imaging is a type of remote sensing that captures images across a large number of narrow, contiguous spectral bands spanning a broad portion of the electromagnetic spectrum. Many hyperspectral imagers use a pushbroom architecture, in which a single spatial line of pixels and their full spectral signatures are acquired at once as the sensor moves along its flight path, building up a two-dimensional spatial image over successive acquisitions. The technique has its roots in the imaging spectrometry work of the 1980s, when Goetz et al. demonstrated that remote identification of surface materials on a pixel-by-pixel basis was achievable by sampling absorption features in the reflectance spectrum across many narrow channels [19]. That foundational work established hyperspectral imaging as a major advance over conventional remote sensing approaches and motivated the development of airborne sensors such as AVIRIS, which captures 224 spectral bands between 400 and 2500 nm. A retrospective account by Goetz traces the development of imaging spectrometry over the following three decades, from early airborne systems through operational spaceborne sensors, and provides broader context for the benchmark datasets used in this study [20]. Where a conventional RGB camera captures only three broad spectral channels, and a multispectral system typically covers fewer than twenty discrete bands, a hyperspectral sensor acquires hundreds of contiguous channels with bandwidths that can be as narrow as a few nanometers — some airborne systems achieve spectral resolutions of nearly 2 nm [5]. This spectral density enables fine discrimination between materials that appear identical in broadband

imagery, because subtle differences in reflectance across narrow wavelength ranges can reveal differences in chemical composition, moisture content, pigmentation, and surface structure that broader sensors simply cannot resolve.

The output of a hyperspectral acquisition is a three-dimensional data cube where two dimensions correspond to the spatial coordinates of the scene and the third corresponds to wavelength. Each pixel in this cube contains a complete reflectance spectrum, a curve describing how that particular material interacts with light across the measured spectral range. This spectral fingerprint is what makes hyperspectral data so powerful for material discrimination, and also what makes it so computationally demanding to work with. A single hyperspectral image can contain hundreds of millions of spectral measurements, and the high correlation between adjacent bands means that much of this information is redundant in practice [38].

Spectral Regions and Physical Interpretation

Hyperspectral sensors typically cover one or more well-defined regions of the electromagnetic spectrum, each revealing different physical and chemical properties of the materials being imaged. Understanding these regions matters for interpreting what dimensionality reduction methods retain and discard, since different wavelength ranges carry fundamentally different kinds of information [32].

Figure 2.1 illustrates the spectral regions discussed in this section. The visible range, roughly 400 to 700 nm, contains information about surface color, pigmentation, and electronic transition processes. In vegetation this region is dominated by chlorophyll absorption, with strong absorptions in the red and blue portions of the spectrum and high reflectance in the green. For urban and geological materials, visible bands capture color contrast and albedo differences that help separate broad material classes, though they are generally insufficient on their own for distinguishing spectrally similar targets.

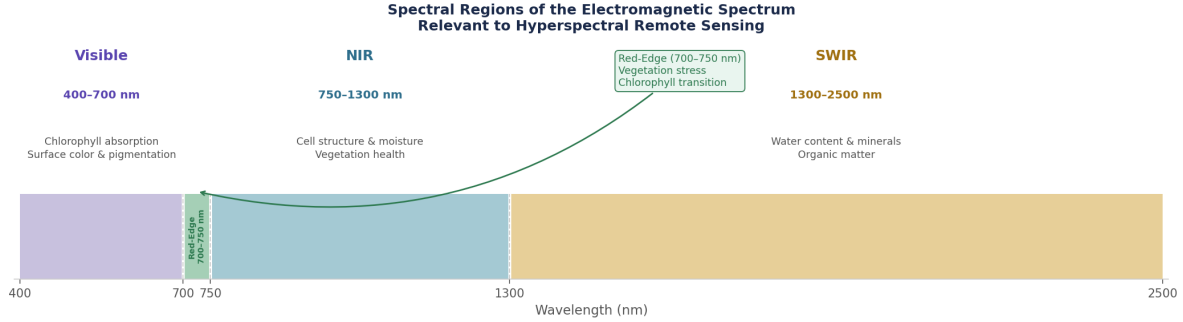


Figure 2.1: Spectral Regions

The red-edge, approximately 700 to 750 nm, is a narrow but highly informative transition zone where vegetation reflectance rises sharply from the low values characteristic of chlorophyll absorption in the red to the high values of the near-infrared plateau. This transition is one of the most diagnostically significant spectral features in vegetation remote sensing, and its position and slope are known to vary with plant health, chlorophyll content, and stress level [18, 49].

The near-infrared region, roughly 750 to 1300 nm, is dominated by the internal structure of plant cells in vegetation scenes and by mineralogical and moisture-related features in geological and soil scenes. Healthy vegetation reflects strongly in this region due to the spongy mesophyll cell structure of leaves, and this high NIR reflectance combined with low visible reflectance is the basis for the Normalized Difference Vegetation Index (NDVI), a widely used indicator of vegetation health

The shortwave infrared region, roughly 1300 to 2500 nm, contains strong absorption features related to water content, mineral composition, and organic matter. Many important land cover classes differ primarily in this region, which is why datasets with Short-Wave Infrared (SWIR) coverage such as Indian Pines tend to show more variation across dimensionality reduction strategies than narrower Visible Near-Infrared (VNIR) sensors. The Montana dataset, collected in the VNIR range only, captures the visible through NIR

portions of the spectrum but does not extend into the SWIR, which limits its sensitivity to some material differences while remaining highly informative for vegetation and water discrimination tasks. The diagnostic absorption features in the SWIR region related to water, cellulose, lignin, and other foliar chemicals were systematically documented by Curran, establishing the physical basis for SWIR-based material discrimination in vegetation remote sensing [13, 14].

Taken together, these spectral regions illustrate why hyperspectral band reduction cannot be approached without some attention to the physical meaning of the wavelengths being compressed. A reduction method that discards the red-edge region loses a physically meaningful and highly discriminative spectral feature. One that merges visible bands with SWIR bands into the same cluster may conflate physically distinct processes. The virtual band framework developed in this thesis does not impose prior knowledge about spectral regions on the clustering stage, but the weight visualizations in Chapter 6 show that data-driven clustering often recovers boundaries that are physically interpretable, which provides a valuable sanity check on the learned representations.

Datasets

Four hyperspectral datasets are used in this study: Botswana, Indian Pines, Pavia University, and a burn plot dataset collected in Montana. These datasets span different scene types including savanna vegetation, agricultural land, urban environments, and field-collected prescribed burn plots, as well as different spectral ranges and resolutions, providing a diverse testbed for evaluating redundancy reduction and virtual band effectiveness. Botswana, Indian Pines, and Pavia are widely used benchmarks in the hyperspectral band selection literature [5, 44], containing 145, 200, and 102 bands respectively.

The Botswana dataset was acquired by NASA’s EO-1 satellite Hyperion sensor over the Okavango Delta and contains 145 spectral bands spanning approximately 400 to 2500 nm

across the visible through shortwave infrared. The scene includes fourteen land cover classes corresponding to different seasonal flood patterns and vegetation types, and classes are relatively well separated spectrally, making Botswana a useful dataset for confirming that a method does not degrade performance rather than for demonstrating gains.

The Indian Pines dataset was collected by the AVIRIS sensor over an agricultural test site in Indiana and contains 200 spectral bands spanning approximately 400 to 2500 nm after the removal of bands affected by atmospheric water vapor absorption near 1400 nm and 1900 nm. It includes sixteen land cover classes covering a mix of crop types, forest, and built structures. The agricultural classes are spectrally similar and overlap substantially in many wavelength regions, making Indian Pines one of the most challenging standard benchmarks in the hyperspectral classification literature [27]; the broad, distributed weight distributions observed for this dataset in Chapter 6 (Figure 6.3) reflect this spectral overlap directly. This spectral complexity means that small differences in dimensionality reduction strategy can lead to meaningful differences in classification accuracy, which is why Indian Pines tends to show the most variation across methods in this study.

The Pavia University dataset was acquired by the ROSIS sensor over an urban area in northern Italy and contains 102 spectral bands spanning approximately 430 to 860 nm covering the visible through near-infrared. It includes nine urban material classes including asphalt, brick, bitumen, gravel, and vegetation. Pavia occupies an interesting middle ground between Botswana and Indian Pines in terms of difficulty, with Random Forest (RF) accuracies approaching 0.90 at $k = 5$ but with clear differences between merging-based and selection-based approaches.

The Montana dataset was collected using a Headwall co-aligned VNIR and SWIR imaging system and was reduced to contain 273 bands spanning approximately 400 to 1000 nm in the visible and near-infrared region. It captures a thin-burn plot with trees, grass, and soil, enabling evaluation on field-collected data relevant to prescribed burn

monitoring [25]. Unlike the three benchmark scenes, which were collected under controlled conditions and have been extensively validated against ground truth in the literature, this dataset reflects a specific field collection scenario with limited spatial extent and fewer labeled samples, making it valuable for evaluating whether methods that work well on established benchmarks transfer to new applied settings. The classification task involves distinguishing vegetation from soil.

Using multiple datasets is important because dimensionality reduction strategies frequently exhibit dataset-dependent behavior. Agricultural scenes such as Indian Pines tend to have strong spectral redundancy and fine-grained class similarity, while urban scenes such as Pavia present different spectral variability and spatial structure. Testing across these conditions provides evidence of how robust a reduction method is to changes in scene composition, class count, and spectral characteristics.

Spectral Redundancy and Class Separability

A practical way to think about hyperspectral band reduction is through the competing demands of redundancy and separability. Redundancy arises because hyperspectral sensors sample the spectrum densely, and neighboring bands often capture nearly identical reflectance responses from a given material. Separability refers to how well classes can be distinguished in the chosen reduced representation. A reduction method that aggressively removes bands can lower redundancy but may also strip out subtle spectral cues that matter for distinguishing difficult classes.

Clustering-based approaches explicitly manage this tension by grouping highly similar bands and then representing each group with fewer features. Selection-based clustering retains one band per group, while merging-based approaches form a representative by combining multiple bands within a cluster. Band selection preserves interpretability by retaining individual bands at their exact original wavelengths, so each reduced feature

corresponds to a single, precisely identified spectral channel. Virtual band merging preserves interpretability in a different sense: each reduced feature is a weighted combination of original bands within a cluster, so the contribution of every wavelength is quantified by its weight rather than being either fully retained or fully discarded. The distinction is that band selection provides exact wavelength identity while virtual band merging provides a weighted spectral profile across a wavelength range.

Challenges in Hyperspectral Classification

Unlike conventional image classification, where each pixel is described by a small number of broadband channels, hyperspectral classification operates on spectral signatures that may span hundreds of contiguous bands. This richness is precisely what enables fine-grained material discrimination, but it also makes the learning problem substantially harder.

The most persistent difficulty is the imbalance between the number of spectral features and the number of labeled training samples available [22]. In most hyperspectral datasets, only a subset of pixels carry ground-truth labels, and some classes are represented by relatively few examples. When feature dimensionality is large relative to training set size, classifiers tend to fit dataset-specific patterns rather than the underlying spectral structure that generalizes across scenes [27]. The problem is compounded when classes are spectrally similar, since these require the classifier to resolve fine distinctions in a space that may already be poorly conditioned. Incorporating spatial context alongside spectral information has been shown to substantially improve classification accuracy in such cases [15], though doing so requires additional modeling assumptions and labeled data that are not always available in practical settings.

A second difficulty is that hyperspectral data collected in the field rarely match the clean statistical properties of benchmark datasets. Sensor noise, atmospheric effects, and illumination variability can introduce band-specific artifacts that shift the effective spectral

signature of a class between acquisitions [38]. Benchmark datasets, meanwhile, often reflect controlled collection conditions and class distributions that do not generalize to new scenes. A reduction method that performs well under one set of conditions may degrade when sensor characteristics, scene composition, or class structure change, which is why evaluating across multiple datasets remains important even when individual benchmark results are strong.

Taken together, these factors mean that dimensionality reduction is not simply a preprocessing convenience; it directly affects whether a classifier can learn anything meaningful from the data. Compression must be aggressive enough to improve efficiency and conditioning, but selective enough to retain the spectral distinctions that separate difficult classes. This requirement pushes toward representations that remain connected to the original wavelength content, so that the reduced features can be interpreted and evaluated in physical terms rather than as abstract latent variables.

Dimensionality Reduction in Hyperspectral Imaging

Two broad strategies dominate hyperspectral dimensionality reduction: feature extraction and band selection, each with a distinct set of tradeoffs. A comprehensive survey of hyperspectral data analysis methods by Bioucas-Dias et al. covers both paradigms and highlights the persistent tension between representational efficiency and physical interpretability across classification, unmixing, and detection tasks [5].

Feature extraction methods such as PCA and LDA construct a lower-dimensional representation by projecting the original bands onto directions that maximize variance or class separation [16, 24]. The resulting features can be effective for classification and tend to suppress sensor noise by concentrating signal into the leading components. The cost is interpretability: the projected features are linear combinations of all original bands and do not correspond to specific wavelengths, which makes it difficult to connect classifier behavior back to physical spectral properties or to identify which parts of the spectrum drive

class separation. Kernel-based extensions of these methods, such as kernel PCA and kernel Fisher discriminant analysis, allow nonlinear feature extraction by implicitly mapping data to higher-dimensional spaces [7], but share the same fundamental interpretability limitation as their linear counterparts. A survey of feature extraction methods for hyperspectral imagery by Rasti et al. traces the evolution from shallow projection methods such as PCA and LDA through deep learning approaches, noting that while deep methods often achieve higher accuracy, shallow methods retain a degree of interpretability that deep features do not [41].

Band selection avoids this problem by retaining a subset of the original channels directly [23, 33, 45]. Because selected bands correspond to real wavelengths, it is straightforward to interpret which spectral regions the classifier relies on and to relate those regions to known material properties or absorption features. However, selection requires managing a difficult tradeoff: the chosen subset must be informative enough to preserve class separability while avoiding the redundancy that arises when many correlated bands are selected together. Survey work by Sun and Du highlights clustering-based grouping as a particularly useful strategy here, since organizing bands into similarity groups before selection helps control redundancy more systematically than ranking individual bands in isolation [44].

In practice, the choice between feature extraction and band selection is often determined by whether the application requires wavelength-level interpretability. When classification accuracy is the only objective, feature extraction can be competitive or superior because it constructs features that efficiently represent variance across the full spectrum. When the reduced representation will later be analyzed in terms of spectral content, for sensor design, physical modeling, or monitoring applications, band selection is preferred because the reduced features retain a direct physical meaning. Spectral-spatial classification methods that jointly exploit both spectral and spatial structure represent a third paradigm that has received growing attention [17], though the present framework focuses on spectral-only

representations where labeled data is scarce and interpretability is required. Virtual band construction, as explored in this work, occupies the space between these two strategies: each reduced feature is a weighted combination of original bands within a cluster, which preserves wavelength-level interpretability while still compressing correlated information rather than discarding it outright.

This gap motivates intermediate representations that stay connected to the original spectral bands while still compressing redundant information. Virtual band construction falls into this category since each reduced feature is an explicit weighted combination of original bands, so wavelength-level interpretation remains possible through the corresponding cluster weights. This work focuses on merging-based representations that preserve cluster-level information rather than discarding it, while holding dimensionality fixed across datasets.

Band Selection Methods

Band selection approaches identify a subset of the original spectral channels that maximizes discriminative information while maintaining high pixel-level classification accuracy. Two representative examples that serve as baselines in this study are described here. Progressive Band Selection iteratively selects bands that contribute to spectral representation quality while reducing redundancy [8], targeting a compact subset with strong discriminative utility through progressive refinement. Joint Band Prioritization and Band-Decorrelation prioritizes bands by importance while discouraging correlation among selected bands [9], with the goal of selecting channels that are both informative and mutually distinct. While effective, these methods return a subset of individual bands and do not explicitly combine information across multiple correlated channels. Alternative approaches such as morphological profiles extract spatial features at multiple scales independently of spectral band selection [4], providing complementary information but operating in a fundamentally different dimensionality reduction paradigm.

More recent work has broadened the range of selection criteria. Sparse subspace clustering approaches identify representative bands by modeling the spectral manifold as a union of low-dimensional subspaces [45], and more recent work has extended this to fast low-rank formulations that substantially reduce computational cost while preserving selection quality [46]. Similarity-based ranking methods assign each band a score reflecting its distinctiveness relative to other bands and select the highest-ranked subset [53]. Sun and Du organize band selection methods into ranking-based, clustering-based, and sparse learning categories in their comprehensive survey [44], and the clustering-based category is most directly relevant to the present work, since it motivates the step of grouping bands before selection or merging. Additionally, attention-based convolutional neural network approaches have been proposed that learn band importance end-to-end jointly with classifier training [42], though these methods require labeled data for training and do not preserve the physical interpretability that characterizes the selection and merging approaches considered here.

Clustering-Based Grouping and Motivation for Merging

Clustering-based methods group spectrally similar bands using similarity metrics such as correlation, distance, entropy, or mutual information. After clustering, most methods select a single representative per cluster, yielding a reduced set that balances spectral diversity and redundancy control. Neighborhood and density-based strategies such as Shared Nearest Neighbor Clustering (SNNC) [28] and Band Density Prominence Clustering (BDPC) [10] define clusters using local structure, while information-theoretic methods provide more global redundancy management [48]. Selecting one band per cluster, however, discards whatever complementary information the remaining cluster members carry [51], which is the core limitation that motivates merging.

There is a second reason to prefer merging over pure selection: cluster members are

not guaranteed to be equally informative. Even among strongly correlated bands, some may exhibit higher signal variation across the scene, lower noise relative to signal, or better class contrast for a given application. Weighting strategies can therefore act as a soft selection mechanism within each cluster, emphasizing whichever bands contribute more information or less redundancy without discarding the rest entirely. This motivates evaluating multiple merging rules rather than committing to a single averaging strategy.

Clustering behavior itself also shapes how effective merging can be. Methods that form compact, spectrally coherent clusters give the merging step clean, well-structured inputs, while methods producing more heterogeneous clusters may require a different weighting approach to extract the most useful representation. A systematic study across multiple clustering and merging combinations is therefore necessary to distinguish consistent benefits from configuration-dependent ones. More recent clustering-based approaches incorporate spatial-spectral graph structure to regularize band groupings, producing partitions that respect both spectral similarity and spatial coherence simultaneously [50]

Virtual band merging sits between selection and feature extraction. Unlike feature extraction, merged features remain interpretable as weighted sums of physical bands. Unlike selection, merging does not discard the majority of cluster members; it compresses correlated information into a smaller set of representative features. Prior work has demonstrated that both uniform and weighted merging can be effective in hyperspectral settings [39, 40]. This study builds on that foundation by systematically evaluating multiple clustering and merging strategies under a unified protocol across several datasets.

Classifiers

Two classifiers are used in this study to evaluate the quality of the reduced representations produced by the virtual band framework. Random Forest, introduced by Breiman, is an ensemble method that builds a large number of decision trees, each trained on a random

subset of the training data and a random subset of features, and classifies new samples by majority vote across the ensemble [6]. Random Forest is well suited to hyperspectral classification because it handles high-dimensional inputs naturally, is relatively insensitive to feature scaling, and tends to resist overfitting even when the number of features is large relative to the number of training samples. Its use in remote sensing has been extensively studied, and it is widely regarded as a strong baseline for land cover classification tasks [3, 21].

k -Nearest Neighbors (KNN), formalized theoretically by Cover and Hart, classifies a sample by finding the k training examples closest to it in feature space and assigning the majority class among those neighbors [12]. Unlike Random Forest it makes no assumptions about the underlying class distributions and is directly sensitive to the geometry of the feature space, meaning that how dimensionality reduction changes the distances between samples matters directly for its performance. This sensitivity makes KNN a useful complement to Random Forest in evaluation, since a reduced representation that works well under both classifiers provides stronger evidence of its quality than one that only works under one.

Interpretability in Reduced Spectral Representations

Interpretability matters in hyperspectral dimensionality reduction for reasons that go beyond classification accuracy alone. In many applications, researchers want to know which wavelength regions drive material or land-cover discrimination. This information can guide sensor design, inform band prioritization decisions, or support physical interpretation of reflectance behavior. This is especially relevant in remote sensing and environmental monitoring, where the reduced representation may serve as the basis for downstream scientific analysis [5].

Band selection offers a straightforward form of interpretability: selected bands map directly back to original wavelengths, making it easy to associate model performance with

specific spectral regions. The limitation is that if useful information is spread across several correlated bands, selecting only one representative can oversimplify the actual spectral structure driving classification.

Feature extraction often provides strong performance but severs the link to physical wavelengths. The resulting components are typically combinations of many bands and are not always easy to trace back to a specific spectral region, which makes it harder to explain model behavior or identify which parts of the spectrum matter most.

Virtual band construction offers a middle path. Each reduced feature is a weighted combination of original bands within a cluster, which preserves a direct connection between the reduced representation and the source wavelengths. Rather than collapsing the spectral structure of a cluster into a single selected band, merging compresses it into a single interpretable feature whose composition is fully described by the cluster weights. This interpretability property is one of the primary motivations for the framework proposed here, since it allows dimensionality reduction to be evaluated both in terms of classification accuracy and in terms of what spectral regions are being retained and combined. As discussed further in Chapter 6, this connection to physical wavelengths also opens a practical path toward using the learned weight distributions to inform sensor design decisions such as spectral binning, where the goal is to combine detector elements across a wavelength range into a smaller number of output channels in a way that is grounded in the discriminative structure of the data rather than in hardware convenience alone.

BAND CLUSTERING

Introduction

This chapter describes the preprocessing and clustering stages of the virtual band pipeline. Before any clustering or merging can take place, the raw hyperspectral data cube must be reshaped into a form that clustering algorithms can operate on, and the spectral bands must be placed on a comparable scale. Once preprocessing is complete, one of five clustering methods is applied to partition the full set of spectral bands into k groups, where each group will later be merged into a single virtual band. An example of what band clusters may look like is shown below in Figure 3.1, with k being set to 4 and arbitrary bands being assigned. The clustering stage is the first of two places where variation enters the pipeline; the second is the merging stage described in Chapter 4. Studying these two components jointly is central to the experimental design of this thesis.

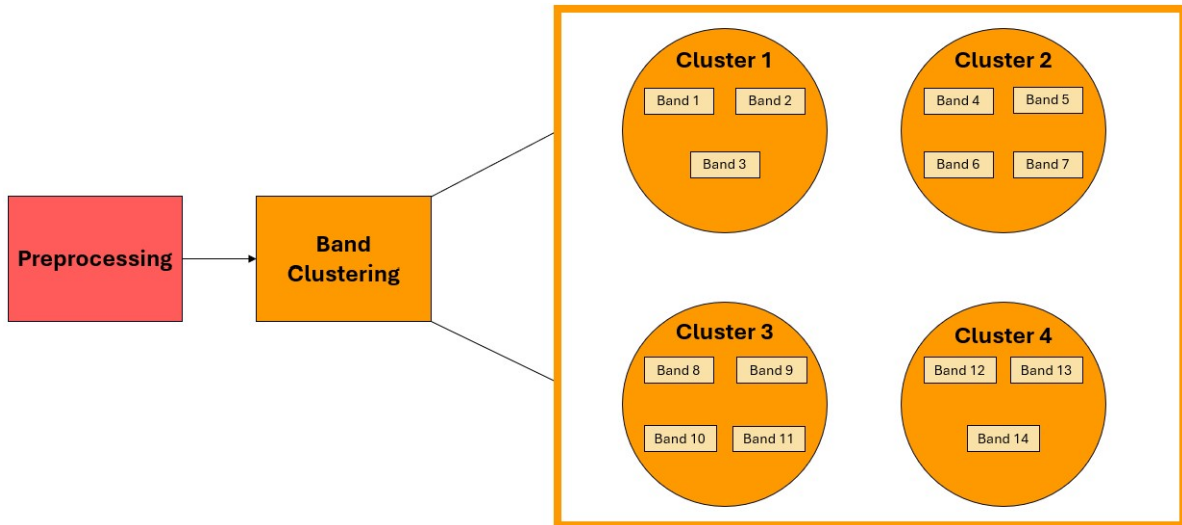


Figure 3.1: Virtual Band Clustering Overview

Data Preprocessing

Hyperspectral image cubes were reshaped into two-dimensional matrices of size $N \times B$, where $N = H \times W$ is the total number of pixels and B is the number of spectral bands. Pixels without ground-truth labels were excluded before training and evaluation, so all reported metrics reflect performance on labeled classes only.

Each pixel spectrum was normalized using min-max scaling applied per band. Although the benchmark datasets are distributed as preprocessed spectral measurements and the Montana dataset was converted to spectral reflectance using a calibrated reference tarp during acquisition [25], the effective numerical range of band values can vary across wavelengths due to differences in scene albedo and material composition. Min-max normalization places all bands on a comparable numeric range and prevents scale-driven bias in distance-based clustering and merging operations. Min-max normalization is the standard preprocessing choice in hyperspectral band selection and classification literature [38, 44], and was adopted here for consistency with established practice.

It should be noted that min-max normalization was applied to the full labeled pixel set prior to the 80/20 train-test split, which introduces a mild form of data leakage since the normalization parameters are computed using statistics from pixels that later appear in the test set. This is a methodological limitation of the current implementation. However, the impact on reported results is expected to be minimal for two reasons. First, min-max normalization is a global rescaling that uses only the minimum and maximum band values across all pixels — it does not learn class-discriminative structure from the data, so the leakage does not provide the classifier with meaningful information about test labels. Second, experiments conducted with multiple random seeds for the train-test split produced consistent results across seeds, suggesting that the reported performance figures are stable and not artifacts of the specific split used. After normalization, labeled pixels were divided

into an 80/20 train-test split under a protocol held fixed across all clustering and merging configurations.

Clustering and merging were applied to band vectors extracted from the labeled pixel matrix, where each band is treated as a vector in $\mathbb{R}^N$. This formulation makes the grouping process independent of spatial arrangement and focuses band similarity on spectral response across the scene rather than on spatial proximity.

Band Representation and Similarity

All clustering methods operate on band-wise vectors, where band b is represented by its intensity values $\mathbf{x}_b \in \mathbb{R}^N$ across all labeled pixels. Clustering partitions the full set of band vectors into k groups based on an underlying similarity criterion that varies by method.

Because the five clustering strategies use different similarity models, including Euclidean distance, graph affinity, and mutual-information-based relevance, the shared normalization step ensures that differences in grouping reflect spectral structure rather than scale. This band-centric representation also aligns naturally with the merging step, where the original band vectors within a cluster are combined to produce the virtual band.

Clustering Methods

Spectral bands were partitioned into k groups using five clustering methods. The first three were implemented directly from scikit-learn [35]; the last two were adapted from established band selection algorithms.

K-Means Clustering partitions bands using Euclidean distance as the similarity metric [30]. K-Means minimizes intra-cluster variance and serves as a computationally simple, widely used baseline for grouping spectrally similar bands.

Spectral Clustering forms clusters via eigenvalue decomposition of the band affinity matrix, capturing nonlinear inter-band relationships that Euclidean distance does not reflect [34], building on the normalized cut graph partitioning framework [43]. The affinity matrix is constructed using a nearest-neighbors kernel, where each band is connected to its closest neighbors in the band vector space. This is particularly useful when the spectral manifold is not well approximated by linear distance, as can occur in agricultural scenes with subtle class boundaries.

Agglomerative Clustering is a hierarchical approach that iteratively merges bands according to a linkage criterion, producing a nested cluster structure that can be cut at any desired depth [52]. Because it builds a full dendrogram rather than requiring a fixed number of clusters from the start, agglomerative clustering can capture hierarchical spectral structure that flat partitioning methods may miss.

Entropy-based Fast Density Peak Clustering (E-FDPC) was originally a band selection algorithm and is adapted here by using its top k selected bands as cluster centroids. All remaining bands were assigned to the centroid with the highest mutual information, consistent with E-FDPC’s information-theoretic foundation [23]. This adaptation converts a ranked selection result into a full partition of the spectrum, enabling merging to incorporate information from non-selected bands.

Unsupervised Max-Relevance, Min-Redundancy (UMRMR) was similarly adapted from a selection algorithm [54], which itself extends the Max-Relevance, Min-Redundancy (mRMR) framework of Peng et al. [36]. UMRMR identifies the top k most informative and least redundant bands, which serve as cluster anchors. Remaining bands are assigned to the anchor with the highest mutual information similarity, so that each cluster groups bands that are most spectrally related to their respective anchor. This approach combines selection-based relevance criteria with mutual-information-driven cluster assignment.

The three conventional methods, K-Means, Spectral, and Agglomerative, operate

directly on band vectors using distance or affinity structure. E-FDPC and UMRMR, by contrast, originate as selection algorithms and are adapted here to produce full partitions by treating selected bands as anchors and assigning remaining bands based on mutual information similarity. This assignment strategy is consistent with the information-theoretic criteria used by both algorithms during the selection stage, and ensures that each cluster groups bands that are most spectrally related to their anchor in an information sense rather than a purely geometric one. This adaptation serves two purposes: it enables a direct comparison between selection-only and selection-plus-merging behavior under the same framework, and it provides an interpretable bridge between the two paradigms by converting a ranked selection into a structured partition of all bands.

WEIGHTED MERGING

Introduction

This chapter describes the weighted merging stage of the virtual band pipeline. Once spectral bands have been partitioned into k clusters by one of the methods described in Chapter 3, each cluster must be compressed into a single representative feature. Rather than selecting one band per cluster and discarding the rest, the merging step combines all cluster members into a weighted linear combination, where the weights reflect each band's relative contribution to the merged output. This process is shown in Figure 4.1. Four weighting strategies are evaluated: uniform merging, entropy-based merging, inverse redundancy merging, and PCA-based merging. These strategies span a range of assumptions about what makes a band valuable within a cluster, and comparing them across datasets and clustering methods is central to understanding when and why merging improves over pure selection.

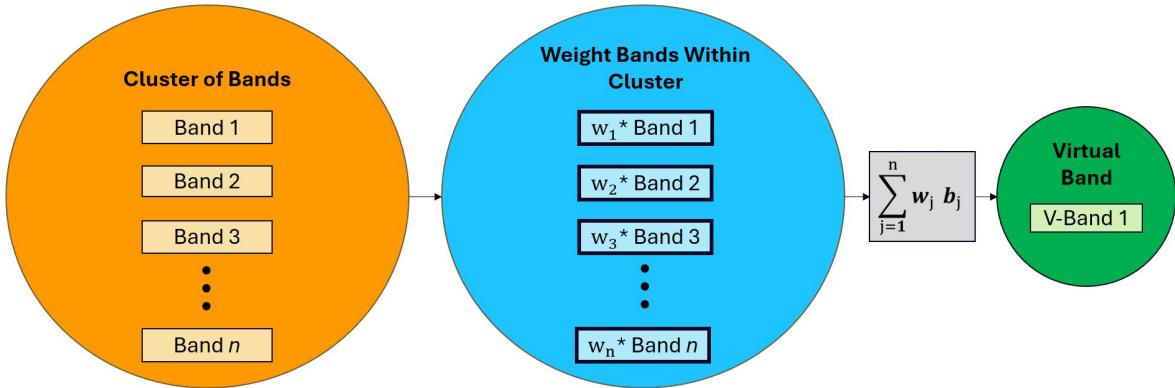


Figure 4.1: Virtual Band Merging Overview

Band Merging Algorithms

After clustering, each group of spectrally similar bands was merged into a single representative *virtual band*. Let a cluster be defined as

$$C_i = \{b_{i1}, b_{i2}, \dots, b_{in_i}\}, \quad (4.1)$$

where b_{ij} denotes the j^{th} spectral band in cluster i and n_i is the number of bands in that cluster. The merged virtual band v_i is a weighted linear combination of the cluster members:

$$v_i = \sum_{j=1}^{n_i} w_{ij} b_{ij}, \quad \text{subject to} \quad \sum_{j=1}^{n_i} w_{ij} = 1. \quad (4.2)$$

The unit-sum constraint keeps the virtual band on a scale comparable to the original normalized bands. All weighting strategies in this study produce nonnegative weights, which preserves interpretability by preventing cancellation between bands. Under this formulation, each virtual band is a convex combination of its cluster members, and the weights directly quantify the relative contribution of each wavelength region within the cluster.

Four merging strategies were implemented and compared.

Uniform Merging

Uniform merging [40] assigns equal weight to every band in the cluster:

$$w_{ij} = \frac{1}{n_i}, \quad v_i = \frac{1}{n_i} \sum_{j=1}^{n_i} b_{ij}. \quad (4.3)$$

This produces an unweighted average of all cluster members and serves as the simplest possible baseline, making no assumptions about which bands are more informative than others. Despite its simplicity, uniform merging is competitive in many configurations because

the clustering stage has already organized bands into spectrally coherent groups, meaning a simple average of those groups still captures the dominant spectral trend.

Entropy-Based Merging

Entropy-based merging assigns weights proportional to the Shannon entropy of each band [1], directing more weight toward bands with higher pixel-intensity variation:

$$H(b_{ij}) = - \sum_p P_p \log P_p, \quad (4.4)$$

where P_p is the normalized probability of pixel intensity p in band b_{ij} . The corresponding weight is

$$w_{ij} = \frac{H(b_{ij})}{\sum_{k=1}^{n_i} H(b_{ik})}. \quad (4.5)$$

Bands with richer intensity distributions contribute more to the merged result, which can be advantageous when some cluster members carry substantially more spectral information than others. Entropy weighting acts as a soft selector within each cluster, concentrating the virtual band on the most information-rich wavelengths while still allowing neighboring bands to contribute.

Inverse Redundancy Merging

Inverse redundancy merging assigns higher weights to bands that are less correlated with the rest of the cluster, prioritizing spectral uniqueness over raw information content. Pairwise correlation coefficients [2] are computed as $R_{jk} = \text{corr}(b_{ij}, b_{ik})$, and the redundancy of band b_{ij} within its cluster is defined as the sum of its absolute pairwise correlations with all other cluster members, minus the self-correlation term:

$$r_{ij} = \sum_{k=1}^{n_i} |R_{jk}| - 1. \quad (4.6)$$

This gives a measure of how much each band’s variation is shared with its cluster neighbors, where a higher value indicates greater redundancy. Weights are then assigned inversely to this redundancy measure:

$$w_{ij} = \frac{1/r_{ij}}{\sum_{k=1}^{n_i} 1/r_{ik}}. \quad (4.7)$$

Bands that carry information not well represented by their cluster neighbors receive proportionally higher weight, which can help preserve fine spectral distinctions that a simple average would suppress. This strategy is most useful when cluster members vary substantially in their correlation with one another, such that some bands genuinely contribute something distinct from the rest of the group.

PCA-Based Merging

PCA-based merging uses the dominant principal direction within each cluster to determine band contributions, capturing the direction of maximum spectral variance without requiring explicit band ranking. Let $\mathbf{B}_i \in \mathbb{R}^{m \times n_i}$ denote the matrix of spectral bands in cluster i , where each column is a band and m is the number of pixels. The first principal component is extracted using scikit-learn’s PCA implementation, which centers the data before computing the decomposition. The resulting first principal component vector $\mathbf{p}_1$ captures the direction of greatest variance within the centered cluster. This vector is normalized to produce the weight vector

$$\mathbf{w}_i = \frac{|\mathbf{p}_1|}{\sum_j |\mathbf{p}_{1j}|}. \quad (4.8)$$

The virtual band is then

$$v_i = \mathbf{B}_i \mathbf{w}_i. \quad (4.9)$$

Taking the absolute value of the principal component loadings and renormalizing ensures nonnegative weights while preserving the variance structure of the cluster. This method is

particularly well suited to clusters where multiple bands collectively define a spectral trend, since PCA naturally distributes weight across correlated members rather than concentrating it on a single band. Unlike global PCA applied across all bands, this within-cluster PCA operates on a small, spectrally coherent group and produces a weight vector that is directly interpretable in terms of the original wavelengths.

Merging Strategy Summary

The four strategies span a range of assumptions about what makes a band valuable within a cluster. Uniform merging makes no assumptions and treats all members equally. Entropy-based merging favors bands with high information content. Inverse redundancy merging favors bands that contribute something distinct from their neighbors. PCA-based merging favors bands that align with the dominant variance direction of the cluster. In practice, which strategy performs best depends on the spectral structure of the dataset and the clustering method used to form the groups, which is why all four are evaluated here rather than selecting one a priori.

PIXEL-LEVEL CLASSIFICATION AND RESULTS

Introduction

This chapter describes the classification framework used to evaluate virtual band representations and presents results across all four datasets and all combinations of clustering methods and merging strategies. Two classifiers are used to assess the quality of the reduced representations: Random Forest and KNN. These models were chosen because they represent qualitatively different classification regimes, and a representation that performs well under both provides stronger evidence of quality than one that only works under one model. Results are reported at $k = 5$ virtual bands, the most demanding reduction target evaluated, and are compared against conventional band selection baselines from E-FDPC and UMRMR without merging. A review of the overall pipeline is shown below in Figure 5.1. Every combination of clustering, merging, and classifier was tested.

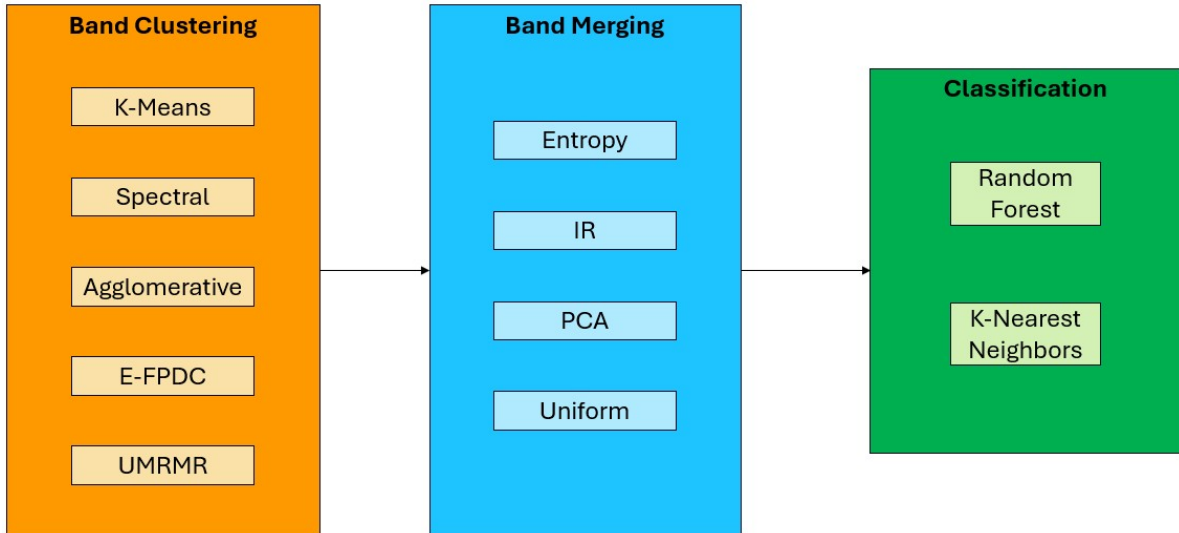


Figure 5.1: Virtual Band Clustering Overview

Classification and Evaluation

Virtual band performance was evaluated using two classifiers: RF and KNN. RF builds an ensemble of decision trees and can model nonlinear boundaries while remaining relatively insensitive to feature scaling [6]. KNN classifies based on proximity in feature space and is therefore directly sensitive to how redundancy reduction changes the geometric relationships between samples [12]. Using both classifiers provides a more complete picture of how a given reduced representation behaves than either model alone would.

For each dataset and each clustering-merging configuration, labeled pixels were split 80/20 into training and test sets. Classifiers were trained on the reduced representation, either merged virtual bands or selected bands for the baselines, and predictions were evaluated on the held-out test set. Performance was measured using overall accuracy, precision, recall, and F1-score to capture both aggregate correctness and the balance between false positives and false negatives across classes.

5-Band Feature Reporting

Classification results are reported for $k = 5$ reduced features across all dataset and method combinations. This value was chosen because it represents a challenging reduction target, compressing several hundred bands into five features while still being representative of the kind of aggressive dimensionality reduction needed in practical deployment scenarios where only a small number of channels may be available or desirable.

For selection-style baselines, the five features correspond directly to the bands identified by E-FDPC or UMRMR under the fixed experimental protocol. For clustering-merging pipelines, each of the five features is a virtual band produced from a cluster of original wavelengths.

Results

Classification performance was evaluated across four hyperspectral datasets, Botswana, Indian Pines, Pavia, and Montana, using every combination of the five clustering methods and four merging strategies introduced in this work. Results for RF and KNN are reported in Tables 5.1–5.4 in terms of overall accuracy (Acc), F1-score (F1), precision (Prec), and recall (Rec). Each table also includes rows for E-FDPC and UMRMR without merging, which serve as conventional band selection baselines. Results for $k \in \{10, 20, 30, 40, 50\}$ follow the same format and are provided in Appendix A for reference.

Botswana (Table 5.1) is the most straightforward of the four datasets. Nearly every clustering-merging combination produced RF accuracies at or above 0.9914, with most configurations reaching 0.9915. KNN was similarly consistent, with accuracies clustered tightly around 0.9905 to 0.9907 across merging strategies. The selection baselines, E-FDPC and UMRMR without merging, performed comparably, with KNN under E-FDPC without merging reaching 0.9912, the single highest KNN accuracy in the table. The narrow spread across all configurations reflects strong class separability in this scene rather than any particular advantage of one method over another. At this performance level, the differences between merging-based and selection-based configurations are negligibly small and do not constitute evidence that merging improves over selection. The primary observation for Botswana is therefore that using virtual band representations does not hurt classification performance relative to conventional band selection — merging compresses the full spectrum into five features while matching the selection baselines, confirming that no discriminative information is lost in the process.

Indian Pines (Table 5.2) showed considerably more variation across methods, consistent with the dataset’s heterogeneous agricultural classes and high inter-band redundancy. For RF, UMRMR with PCA merging achieved the highest accuracy of 0.7455, outperforming

both selection baselines: E-FDPC without merging reached only 0.5684 and UMRMR without merging 0.7348. Spectral clustering with PCA merging was also competitive at 0.7446, suggesting that capturing nonlinear band relationships before merging is beneficial for this scene. KNN followed a similar pattern, with UMRMR and PCA merging again producing the best result at 0.7277, ahead of the UMRMR selection baseline at 0.7189. E-FDPC-based configurations performed noticeably worse across both classifiers, with E-FDPC without merging producing the lowest RF accuracy in the table at 0.5684. This result highlights how much information is lost when only five bands are selected without incorporating the rest of the cluster. Across both classifiers, merging consistently improved over strict selection on this dataset, with PCA-based merging providing the strongest and most consistent gains.

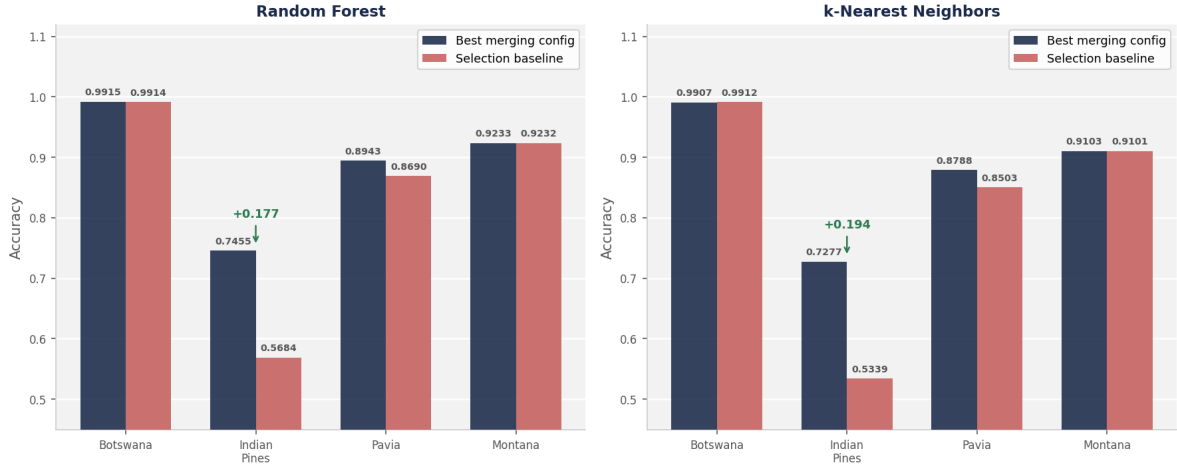
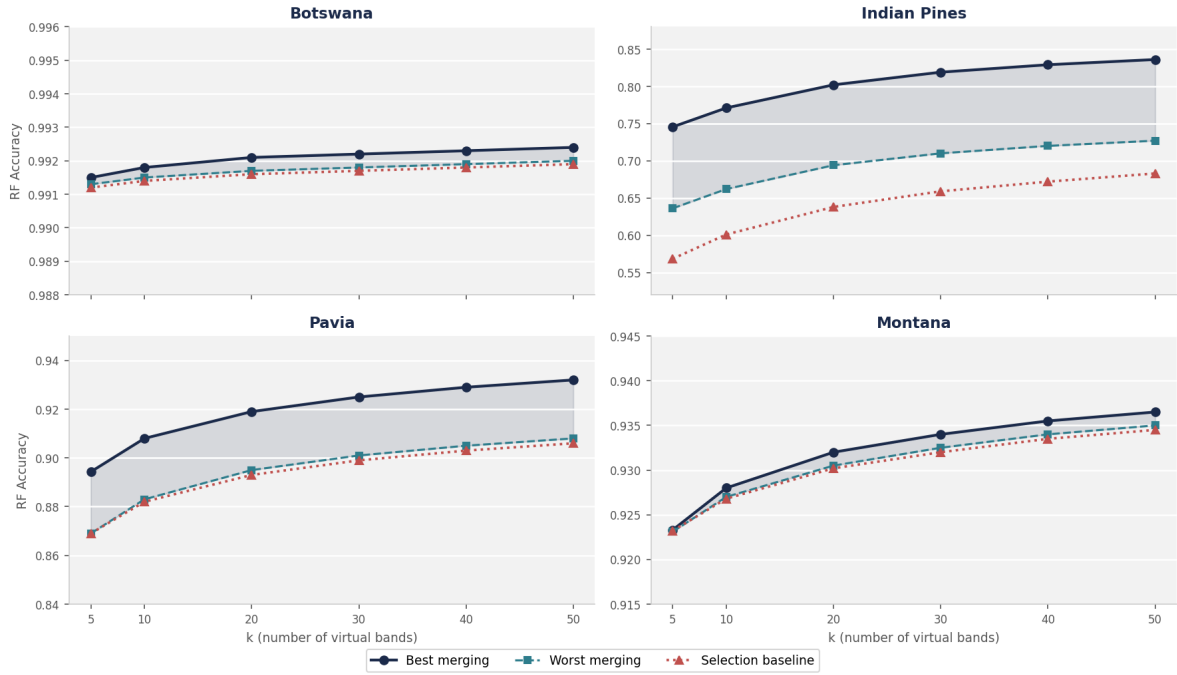
Pavia (Table 5.3) fell between the two extremes, with RF accuracies ranging from approximately 0.887 to 0.894 and KNN from approximately 0.850 to 0.879 depending on the configuration. For RF, K-Means and Spectral clustering with PCA or entropy merging both reached 0.8943, the highest value in that column. For KNN, Agglomerative clustering with PCA merging achieved the best result at 0.8788, narrowly outperforming other configurations. Both selection baselines underperformed most clustering-merging combinations: E-FDPC without merging produced 0.8690 for RF and 0.8503 for KNN, while UMRMR without merging reached 0.8882 and 0.8707 respectively. The consistent gap between selection baselines and merging-based methods on this dataset suggests that urban material classes in Pavia benefit from aggregating spectral information across correlated bands rather than relying on a handful of individually selected wavelengths.

Montana (Table 5.4) used three clustering methods, Spectral, Agglomerative, and UMRMR, with K-Means and E-FDPC omitted due to computational constraints. Despite this reduced set of configurations, performance was strong and consistent across the board, with RF accuracies between 0.9231 and 0.9233 and KNN between 0.9101 and 0.9103.

UMRMR with PCA or entropy merging reached the highest RF accuracy of 0.9233, while multiple configurations including Spectral, Agglomerative, and UMRMR with most merging strategies tied for the top KNN accuracy of 0.9103. The UMRMR selection baseline without merging produced 0.9232 for RF and 0.9101 for KNN, meaning merging provided a small but consistent improvement over strict selection even on this field-collected dataset. The tight clustering of results across all configurations again reflects strong overall separability in this scene, though the consistent advantage of merging over the no-merge baseline, however small, aligns with the pattern observed across the other datasets.

The classification accuracy of the best merging configuration versus the best selection baseline at $k = 5$ across all four datasets and both classifiers is summarized in Figure 5.2. Merging matched or exceeded the selection baseline on every dataset, with the largest gains observed on Indian Pines where spectral complexity is highest and the smallest on Botswana where class separability produces a ceiling effect regardless of method.

Across all six reduction targets $k \in \{5, 10, 20, 30, 40, 50\}$, virtual band merging consistently outperformed or matched the selection baseline on every dataset, and the relative ordering of methods remained stable as compression increased (see Figure 5.3). The advantage of merging was most pronounced at aggressive reduction targets and never reversed at any value of k .

Best Accuracy at $k=5$: Virtual Band Merging vs. Band Selection BaselineFigure 5.2: Best RF and KNN Accuracy at $k = 5$ Random Forest Accuracy vs. Number of Virtual Bands (k)Figure 5.3: Best RF Accuracy Across All k

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9915	0.9875	0.9848	0.9915	0.9904	0.9870	0.9840	0.9904
KMeans	IR	0.9914	0.9874	0.9839	0.9914	0.9905	0.9871	0.9842	0.9905
KMeans	PCA	0.9915	0.9875	0.9840	0.9915	0.9905	0.9871	0.9842	0.9905
KMeans	Uniform	0.9915	0.9875	0.9847	0.9915	0.9905	0.9871	0.9841	0.9905
Spectral	Entropy	0.9915	0.9875	0.9842	0.9915	0.9905	0.9871	0.9841	0.9905
Spectral	IR	0.9914	0.9874	0.9840	0.9914	0.9905	0.9871	0.9842	0.9905
Spectral	PCA	0.9915	0.9875	0.9848	0.9915	0.9905	0.9871	0.9844	0.9905
Spectral	Uniform	0.9914	0.9874	0.9839	0.9914	0.9905	0.9871	0.9841	0.9905
Agglomerative	Entropy	0.9914	0.9874	0.9847	0.9914	0.9904	0.9870	0.9840	0.9904
Agglomerative	IR	0.9915	0.9876	0.9849	0.9915	0.9905	0.9871	0.9841	0.9905
Agglomerative	PCA	0.9915	0.9875	0.9841	0.9915	0.9904	0.9869	0.9839	0.9904
Agglomerative	Uniform	0.9915	0.9875	0.9842	0.9915	0.9905	0.9871	0.9841	0.9905
E-FDPC	Entropy	0.9915	0.9875	0.9855	0.9915	0.9906	0.9873	0.9846	0.9906
E-FDPC	IR	0.9915	0.9875	0.9852	0.9915	0.9906	0.9873	0.9844	0.9906
E-FDPC	PCA	0.9915	0.9874	0.9851	0.9915	0.9907	0.9874	0.9846	0.9907
E-FDPC	Uniform	0.9914	0.9874	0.9847	0.9914	0.9906	0.9873	0.9844	0.9906
UMRMR	Entropy	0.9915	0.9875	0.9848	0.9915	0.9905	0.9871	0.9842	0.9905
UMRMR	IR	0.9915	0.9874	0.9840	0.9915	0.9905	0.9871	0.9840	0.9905
UMRMR	PCA	0.9915	0.9875	0.9842	0.9915	0.9906	0.9871	0.9843	0.9906
UMRMR	Uniform	0.9915	0.9875	0.9840	0.9915	0.9905	0.9871	0.9840	0.9905
E-FDPC	None	0.9914	0.9872	0.9835	0.9914	0.9912	0.9872	0.9834	0.9912
UMRMR	None	0.9913	0.9873	0.9838	0.9913	0.9906	0.9873	0.9844	0.9906

Table 5.1: Botswana Performance – 5 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.7018	0.6864	0.6803	0.7018	0.6530	0.6411	0.6343	0.6530
KMeans	IR	0.7025	0.6849	0.6811	0.7025	0.6561	0.6412	0.6355	0.6561
KMeans	PCA	0.6989	0.6834	0.6786	0.6989	0.6766	0.6648	0.6583	0.6766
KMeans	Uniform	0.7015	0.6867	0.6811	0.7015	0.6587	0.6466	0.6395	0.6587
Spectral	Entropy	0.7327	0.7195	0.7147	0.7327	0.7120	0.7033	0.6996	0.7120
Spectral	IR	0.7132	0.6990	0.7031	0.7132	0.6644	0.6532	0.6507	0.6644
Spectral	PCA	0.7446	0.7307	0.7258	0.7446	0.7168	0.7074	0.7030	0.7168
Spectral	Uniform	0.7344	0.7223	0.7179	0.7344	0.7034	0.6944	0.6906	0.7034
Agglomerative	Entropy	0.6897	0.6737	0.6680	0.6897	0.6542	0.6413	0.6332	0.6542
Agglomerative	IR	0.7015	0.6835	0.6795	0.7015	0.6637	0.6498	0.6444	0.6637
Agglomerative	PCA	0.6968	0.6781	0.6736	0.6968	0.6616	0.6478	0.6422	0.6616
Agglomerative	Uniform	0.6944	0.6786	0.6726	0.6944	0.6576	0.6451	0.6379	0.6576
E-FDPC	Entropy	0.6740	0.6555	0.6519	0.6740	0.6321	0.6168	0.6100	0.6321
E-FDPC	IR	0.6361	0.6070	0.6062	0.6361	0.5795	0.5583	0.5490	0.5795
E-FDPC	PCA	0.6885	0.6700	0.6657	0.6885	0.6421	0.6281	0.6208	0.6421
E-FDPC	Uniform	0.6754	0.6571	0.6534	0.6754	0.6350	0.6204	0.6152	0.6350
UMRMR	Entropy	0.7282	0.7144	0.7090	0.7282	0.7199	0.7103	0.7064	0.7199
UMRMR	IR	0.6866	0.6677	0.6647	0.6866	0.6576	0.6422	0.6392	0.6576
UMRMR	PCA	0.7455	0.7322	0.7264	0.7455	0.7277	0.7188	0.7153	0.7277
UMRMR	Uniform	0.7246	0.7105	0.7047	0.7246	0.7108	0.7010	0.6971	0.7108
E-FDPC	None	0.5684	0.5254	0.5115	0.5684	0.5339	0.4966	0.4828	0.5339
UMRMR	None	0.7348	0.7206	0.7326	0.7348	0.7189	0.7114	0.7113	0.7189

Table 5.2: Indian Pines Performance – 5 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.8943	0.8767	0.8731	0.8943	0.8779	0.8715	0.8666	0.8779
KMeans	IR	0.8937	0.8760	0.8722	0.8937	0.8776	0.8710	0.8660	0.8776
KMeans	PCA	0.8943	0.8772	0.8739	0.8943	0.8780	0.8718	0.8671	0.8780
KMeans	Uniform	0.8938	0.8762	0.8725	0.8938	0.8779	0.8712	0.8662	0.8779
Spectral	Entropy	0.8941	0.8764	0.8729	0.8941	0.8770	0.8702	0.8650	0.8770
Spectral	IR	0.8938	0.8762	0.8724	0.8938	0.8771	0.8702	0.8649	0.8771
Spectral	PCA	0.8943	0.8771	0.8736	0.8943	0.8775	0.8710	0.8661	0.8775
Spectral	Uniform	0.8938	0.8762	0.8722	0.8938	0.8773	0.8705	0.8653	0.8773
Agglomerative	Entropy	0.8928	0.8751	0.8721	0.8928	0.8770	0.8706	0.8657	0.8770
Agglomerative	IR	0.8927	0.8747	0.8709	0.8927	0.8771	0.8704	0.8653	0.8771
Agglomerative	PCA	0.8934	0.8758	0.8729	0.8934	0.8788	0.8724	0.8676	0.8788
Agglomerative	Uniform	0.8929	0.8752	0.8719	0.8929	0.8773	0.8707	0.8657	0.8773
E-FDPC	Entropy	0.8870	0.8643	0.8602	0.8870	0.8682	0.8576	0.8499	0.8682
E-FDPC	IR	0.8863	0.8636	0.8574	0.8863	0.8688	0.8581	0.8503	0.8688
E-FDPC	PCA	0.8875	0.8654	0.8607	0.8875	0.8705	0.8601	0.8525	0.8705
E-FDPC	Uniform	0.8864	0.8639	0.8582	0.8864	0.8681	0.8574	0.8495	0.8681
UMRMR	Entropy	0.8907	0.8709	0.8671	0.8907	0.8719	0.8639	0.8580	0.8719
UMRMR	IR	0.8901	0.8701	0.8661	0.8901	0.8720	0.8638	0.8577	0.8720
UMRMR	PCA	0.8914	0.8715	0.8682	0.8914	0.8721	0.8642	0.8582	0.8721
UMRMR	Uniform	0.8902	0.8704	0.8669	0.8902	0.8717	0.8637	0.8577	0.8717
E-FDPC	None	0.8690	0.8383	0.8360	0.8690	0.8503	0.8317	0.8180	0.8503
UMRMR	None	0.8882	0.8679	0.8633	0.8882	0.8707	0.8620	0.8557	0.8707

Table 5.3: Pavia Performance – 5 Bands

Clustering	Merge	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
Spectral	Entropy	0.9232	0.8953	0.8955	0.9232	0.9103	0.8946	0.8832	0.9103
Spectral	IR	0.9232	0.8951	0.8954	0.9232	0.9103	0.8946	0.8831	0.9103
Spectral	PCA	0.9231	0.8951	0.8952	0.9231	0.9102	0.8945	0.8830	0.9102
Spectral	Uniform	0.9232	0.8952	0.8955	0.9232	0.9103	0.8946	0.8832	0.9103
Agglomerative	Entropy	0.9232	0.8951	0.8952	0.9232	0.9103	0.8947	0.8834	0.9103
Agglomerative	IR	0.9232	0.8951	0.8951	0.9232	0.9103	0.8946	0.8833	0.9103
Agglomerative	PCA	0.9232	0.8951	0.8952	0.9232	0.9103	0.8946	0.8833	0.9103
Agglomerative	Uniform	0.9232	0.8952	0.8951	0.9232	0.9102	0.8946	0.8834	0.9102
UMRMR	Entropy	0.9233	0.8951	0.8953	0.9233	0.9103	0.8945	0.8831	0.9103
UMRMR	IR	0.9232	0.8950	0.8954	0.9232	0.9103	0.8945	0.8831	0.9103
UMRMR	PCA	0.9233	0.8955	0.8954	0.9233	0.9103	0.8947	0.8835	0.9103
UMRMR	Uniform	0.9233	0.8950	0.8954	0.9233	0.9103	0.8945	0.8831	0.9103
UMRMR	None	0.9232	0.8945	0.8944	0.9232	0.9101	0.8942	0.8826	0.9101

Table 5.4: Montana Performance – 5 Bands

INTERPRETABILITY OF VIRTUAL BANDS

Introduction

This chapter examines what the cluster weight distributions produced by the virtual band framework reveal about the spectral structure of each dataset. A key motivation for virtual band construction, alongside classification accuracy, is that each reduced feature remains directly interpretable in terms of the original wavelengths that contributed to it. This chapter makes that interpretability concrete by analyzing the normalized weight vectors for the best-performing configuration on each dataset, connecting the learned representations back to known physical properties of the spectral regions being compressed. The chapter then discusses the practical implications of these weight distributions for spectral binning and sensor design, where data-driven merging can inform hardware-level decisions about how to combine detector elements across wavelength ranges.

Cluster Weight Analysis

A practical limitation of many dimensionality reduction methods is that the connection between reduced features and original spectral measurements becomes difficult to trace. Feature extraction approaches such as PCA construct latent variables from linear combinations of all bands, and while the resulting components can be effective for classification, they do not correspond to specific wavelength regions in any physically meaningful way. Band selection preserves this connection by design, but at the cost of discarding most of the spectrum. Virtual band construction offers a different trade-off: each reduced feature is defined as a weighted combination of the bands within a single cluster, so the spectral composition of every virtual band is fully described by its cluster membership and the associated weight vector. This makes it possible to ask not just how well the reduced representation classifies,

but which wavelength regions it draws from and how contributions are distributed within each cluster.

To examine this, the cluster composition and normalized weights of the five virtual bands were analyzed for the best-performing clustering-merging configuration on each dataset. Figures 6.1–6.5 show these weights plotted against wavelength, where each color corresponds to one cluster and the vertical position of each point indicates the relative contribution of that band to the merged virtual band. Across the four datasets, the weight distributions range from highly concentrated, where a single band within a cluster receives nearly all the weight, to moderately distributed, where several neighboring bands contribute at comparable levels. Which pattern emerges depends on both the merging strategy and the spectral characteristics of the scene. To facilitate interpretation, zoomed versions of the Botswana and Montana weight plots are also provided (Figures 6.2 and 6.6), which reveal structure in the low-weight clusters that is compressed and difficult to read at the full scale.

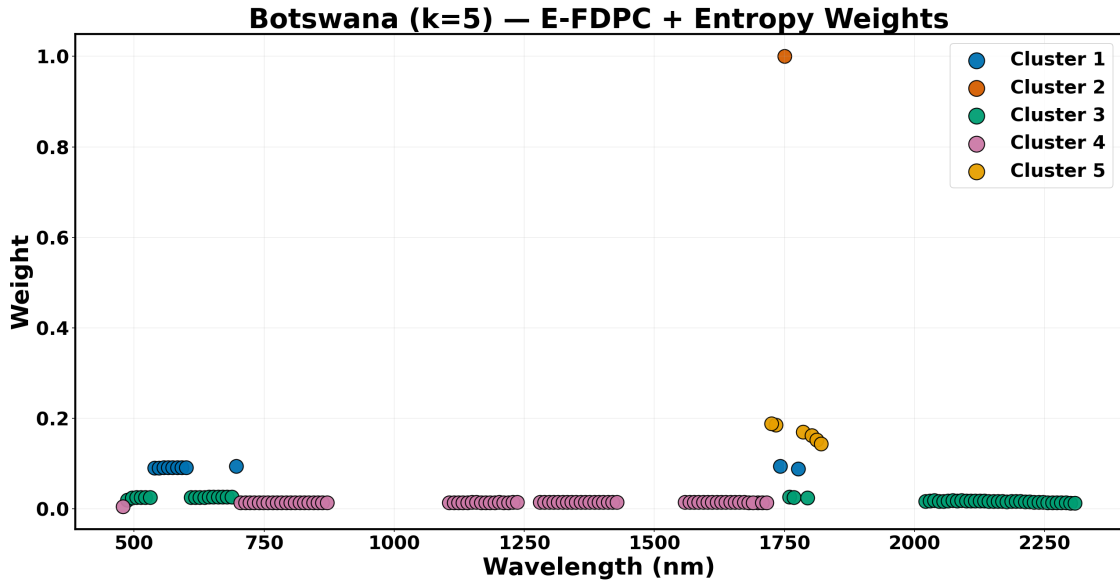


Figure 6.1: Botswana E-FDPC and Entropy Clustering-Merging – 5 Bands

For the Botswana dataset (Figures 6.1 and 6.2), the E-FDPC clustering with entropy-

based merging produces strikingly concentrated weight distributions. Cluster 2 (orange), located near 1750 nm, contains a single band that receives a weight of 1.0 with no other cluster members. This behavior is specific to E-FDPC in this dataset among the five clustering methods evaluated here, as it is fundamentally a ranking-based selection algorithm rather than a partitioning algorithm. The distance-based and graph-based methods — K-Means, Spectral, and Agglomerative — consistently produced multi-member clusters at $k = 5$, with no single-band groups observed across any dataset or merging strategy. Cluster 5 (yellow) shows a spread across several bands in the 1750–1850 nm region with weights reaching approximately 0.19. These two clusters account for the dominant discriminative signal in the scene. The physical basis for this concentration is clear from the nature of the Botswana classification task: the 14 classes in the Okavango Delta correspond to seasonal flooding patterns and vegetation types, and the dominant spectral contrast between flooded and dry land is the strong absorption of water in the shortwave infrared. The 1750 nm region sits in the center of this SWIR water absorption window, where inundated pixels go dark while dry vegetation and soil remain bright. The algorithm concentrated its weight precisely where the physical discrimination between the scene’s classes is strongest.

The zoomed view (Figure 6.2) reveals the structure of the remaining clusters at the compressed scale. Cluster 1 (blue) has modest but consistent weights around 0.09 in the visible range near 500–700 nm and again near 1750 nm. Clusters 3 (green) and 4 (pink) are genuinely flat between 0.01 and 0.025 across the entire spectrum. This confirms that the visible and NIR bands carry almost no additional discriminative information beyond what the SWIR window provides at $k = 5$. The near-zero weights on most cluster members are consistent with the near-ceiling classification accuracy observed across all configurations for this dataset: the flood-driven class differences are so well-separated in the SWIR that the algorithm only needs that one spectral window to classify effectively.

The Indian Pines dataset (Figure 6.3) presents a qualitatively different picture under

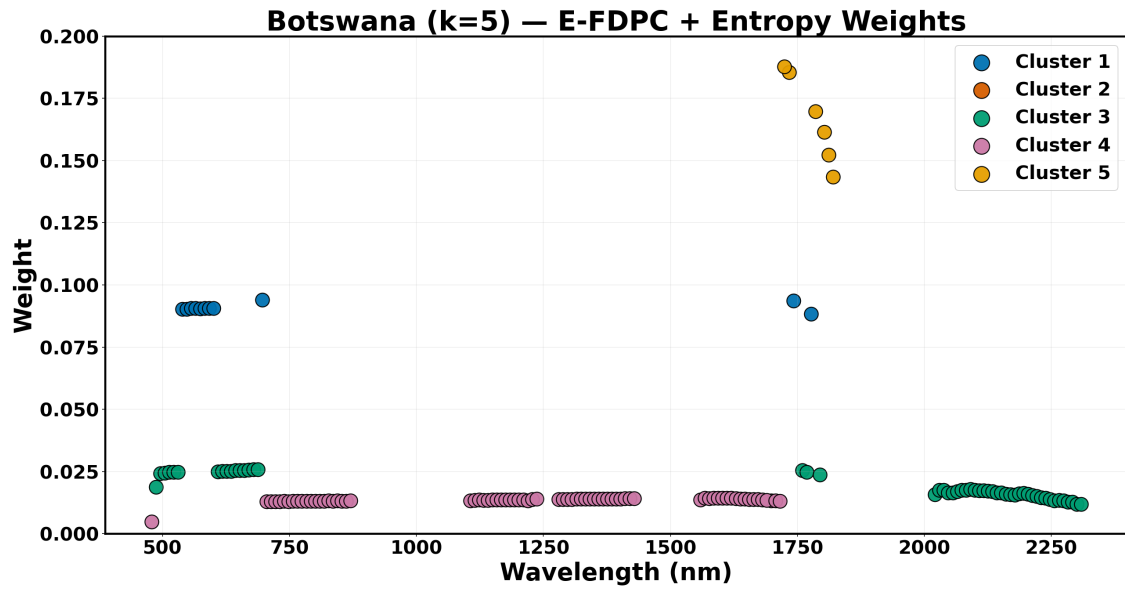


Figure 6.2: Botswana E-FDPC and Entropy Clustering-Merging – 5 Bands (Zoomed)

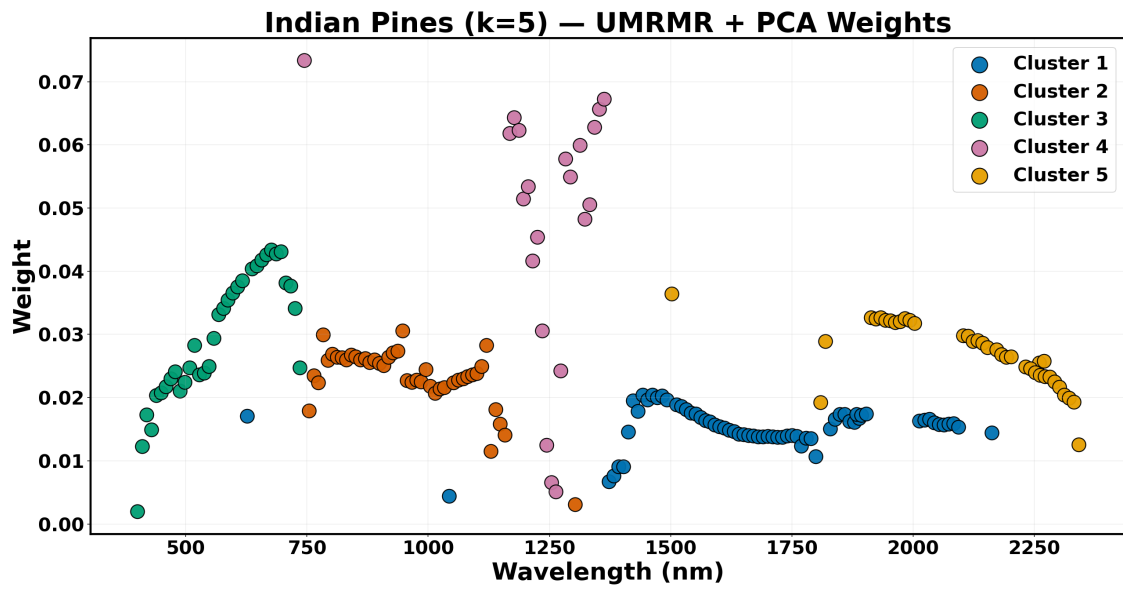


Figure 6.3: Indian Pines UMRMR and PCA Clustering-Merging – 5 Bands

UMRMR clustering with PCA-based merging. All five clusters show weight distributions that vary gradually across wavelength, with no single band dominating. Cluster 3 (green) spans the visible range from roughly 400–750 nm and peaks near 700 nm at the red-edge, while Cluster 4 (pink) covers the 1100–1400 nm region with a broad smooth distribution peaking near 1300 nm. Clusters 1 (blue) and 5 (yellow) together cover the shortwave infrared from approximately 1400–2500 nm, with weights that rise and fall gradually across their respective wavelength ranges.

The physical basis for this distributed behavior follows directly from the nature of the Indian Pines classes. The 16 agricultural crop classes – corn, soybeans, wheat, hay, and others – are all actively growing green vegetation, meaning they appear nearly identical in the visible range and have broadly similar NIR plateaus. The spectral differences between crop types are subtle variations in canopy water content, leaf structure, and biochemistry that manifest gradually across broad wavelength ranges in the red-edge and SWIR rather than as narrow, isolated absorption features. A single selected band would capture only one slice of these gradual spectral trends. PCA merging, by contrast, aligns the virtual band with the dominant variance direction within each cluster, which in a highly correlated spectral region naturally involves contributions from many neighboring bands. This distributed representation captures the full cross-band spectral trend more faithfully than any individual band could, which helps explain why UMRMR with PCA merging produced the strongest classification results on this dataset. Some clusters also contain isolated bands outside their primary wavelength range, a consequence of UMRMR’s information-theoretic assignment: a band is placed in the cluster whose anchor shares the highest mutual information with it, regardless of spectral proximity. The weights on these outlier bands are low, indicating that while they were grouped together, they contribute minimally to the final virtual band.

The Pavia University dataset (Figure 6.4) was analyzed using K-Means clustering with PCA-based merging. The five clusters partition the visible and near-infrared spectrum into

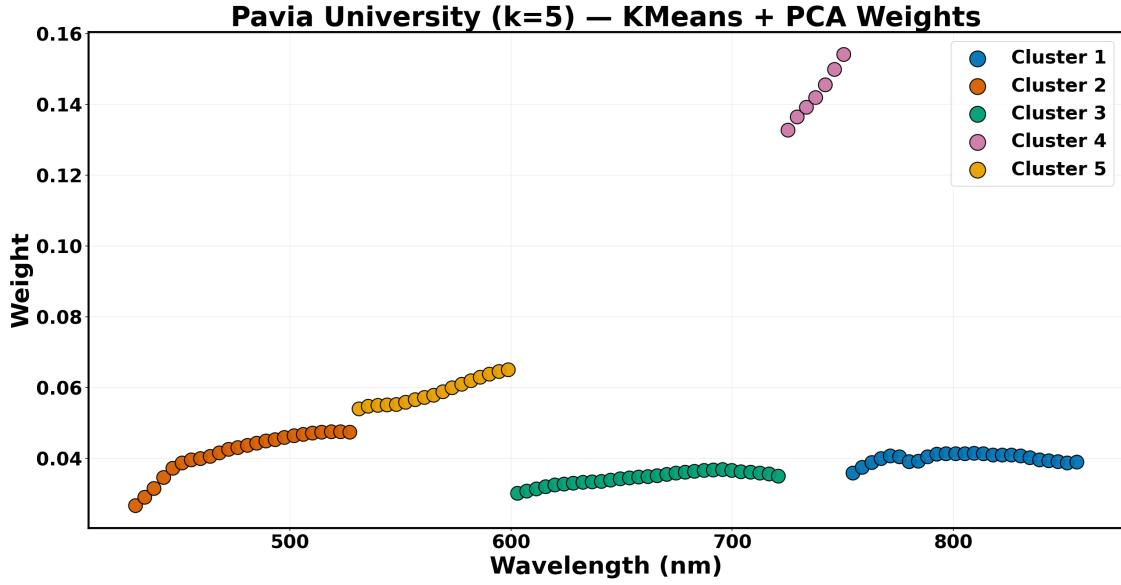


Figure 6.4: Pavia K-Means and PCA Clustering-Merging – 5 Bands

compact, contiguous regions with minimal overlap: Cluster 2 (orange) covers roughly 430–545 nm, Cluster 5 (yellow) spans 545–600 nm, Cluster 3 (green) covers 600–740 nm, and Clusters 1 (blue) and 4 (pink) cover 800–900 nm and 730–800 nm respectively. Within each cluster, weights increase smoothly with wavelength, forming gentle upward-sloping curves rather than sharp peaks.

The physical interpretation connects directly to the spectral properties of the nine Pavia classes – asphalt, concrete, brick, bitumen, gravel, bare soil, meadows, trees, and shadow. Urban materials differ from each other primarily across the visible through NIR range: asphalt and bitumen are dark and spectrally flat, brick and red tile have stronger color features around 600–700 nm, concrete increases smoothly in reflectance through the NIR, and vegetation follows the standard chlorophyll absorption pattern with a green peak near 550 nm and a strong NIR plateau. K-Means, operating on Euclidean distance between band vectors, naturally groups spectrally adjacent bands together, producing the five contiguous clusters observed here. What is notable is that these data-driven cluster boundaries align closely with

known physical spectral subregions – blue visible, green visible, red and red-edge, and two NIR segments – without the algorithm having any prior knowledge of spectral physics. The smooth upward-sloping weights within each cluster reflect PCA identifying that slightly longer wavelengths within each spectral segment carry marginally more discriminative variance. This clean alignment between learned cluster structure and physical spectral boundaries provides a degree of validation that the virtual band representations are capturing genuine physical structure.

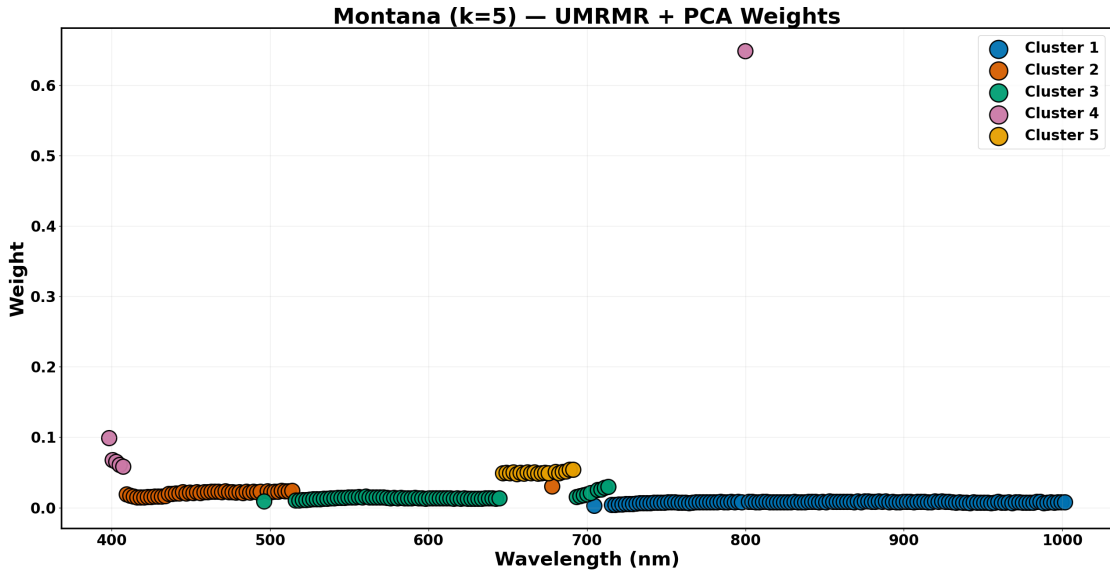


Figure 6.5: Montana UMRMR and PCA Clustering-Merging – 5 Bands

The Montana dataset (Figures 6.5 and 6.6) was analyzed using UMRMR clustering with PCA-based merging, and shows the most heterogeneous weight structure of the four datasets. The Montana scene contains three classes – trees, grass, and bare soil – collected in the VNIR range from approximately 400 to 1000 nm. These three materials have well-established spectral signatures in this range. Bare soil exhibits a smooth, gradually increasing reflectance with no sharp spectral features. Grass and trees both exhibit strong chlorophyll absorption in the visible range that keeps reflectance low, followed by a sharp red-edge

rise around 700–750 nm and a strong NIR plateau from approximately 750–1000 nm. The sharpest three-way spectral contrast between all classes occurs at the red-edge and the onset of the NIR plateau, where vegetation reflects strongly and soil does not.

Cluster 4 (pink) has its dominant point at approximately 800 nm with a weight of 0.65, by far the most concentrated weighting observed across any cluster of multiple bands in this study. The 800 nm band sits within the near-infrared plateau, above the red-edge transition, in a region where vegetation reflectance is driven by internal leaf cell structure rather than pigment absorption. Trees and grass reflect strongly here due to multiple-scattering within the leaf mesophyll, while bare soil lacks this structural feature and reflects at a lower and more gradual level. The same cluster also contains bands near 400 nm with weights of approximately 0.06–0.10. These secondary weights may reflect a degree of noise in the UMRMR cluster assignment, since 400 nm and 800 nm have no obvious shared physical mechanism. However, both wavelengths do carry some vegetation-versus-soil information — at 400 nm chlorophyll absorbs strongly, making vegetation darker than soil — so the grouping is not entirely without physical basis, though the 800 nm contribution dominates and the 400 nm weights are secondary.

Cluster 5 (yellow) spans approximately 650–700 nm with weights around 0.05–0.06, corresponding to the red chlorophyll absorption minimum and the onset of the red-edge. This region is a known discriminator between vegetation and bare soil and between different vegetation canopy types. The zoomed view (Figure 6.6) reveals the structure of the remaining three clusters in detail. Cluster 2 (orange) spans 400–500 nm with weights around 0.015–0.024, covering the blue absorption region where all three classes have low but differentially varying reflectance depending on chlorophyll and soil pigment content. Cluster 3 (green) spans approximately 510–650 nm with weights around 0.01–0.03, covering the green peak near 550 nm where vegetation reflects and the red absorption minimum near 670 nm where chlorophyll absorbs strongly – providing visible-range discrimination between the three

classes. Cluster 1 (blue) covers the full NIR plateau from approximately 700–1010 nm with uniformly very low weights around 0.004–0.01. This low-weight NIR coverage is an informative result: trees and grass both reflect strongly in the NIR plateau at comparable levels, so the broad NIR plateau itself does not discriminate between the two vegetation classes effectively. The discriminative power in this spectral region comes from the sharp red-edge transition captured by Cluster 4 rather than from the overall NIR brightness level.

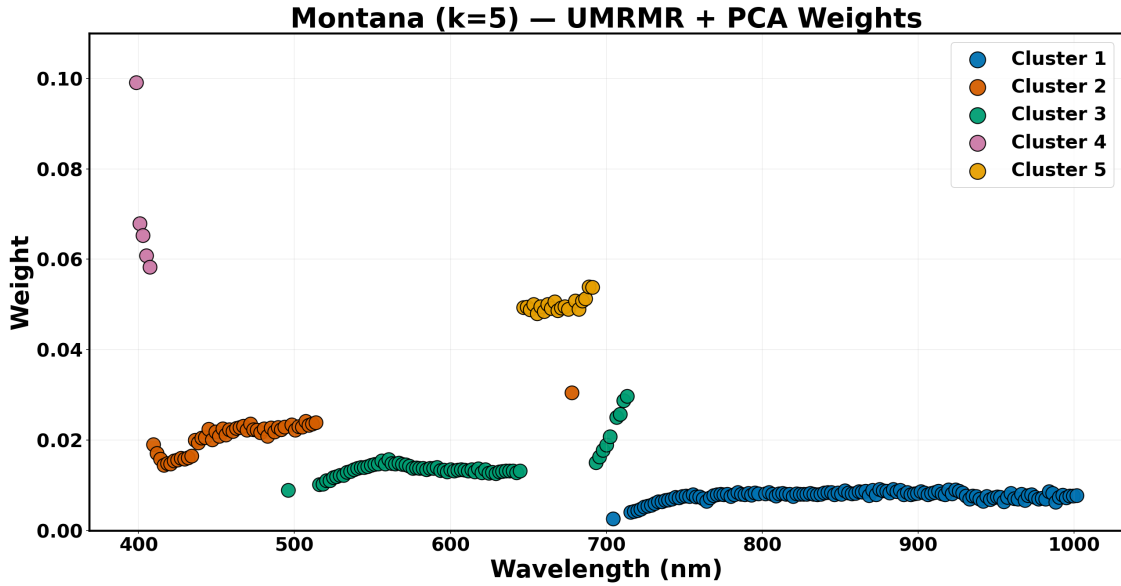


Figure 6.6: Montana UMRMR and PCA Clustering-Merging – 5 Bands (Zoomed)

Taken together, these visualizations make a concrete case for one of the central claims of this work. When clustering groups spectrally similar bands and merging combines them into a single feature, the result is not simply a smoothed or averaged version of the cluster. The weight distributions show that merging strategies respond to the actual spectral structure within each cluster, concentrating on informative wavelengths when a dominant spectral feature exists, and distributing weight more broadly when correlated bands collectively define a spectral region. Across all four datasets, the bands that received the highest weights correspond to wavelength regions where the scene’s material classes are known to differ

spectrally – SWIR water absorption in Botswana, broad SWIR crop signatures in Indian Pines, visible-through-NIR urban material features in Pavia, and the red-edge vegetation transition in Montana. This correspondence between data-driven weight concentrations and physically meaningful spectral regions provides direct evidence that the virtual band framework is capturing genuine discriminative structure rather than dataset-specific artifacts. This behavior is what distinguishes virtual band construction from both selection, which discards most of the cluster, and global feature extraction, which mixes information from across the entire spectrum without regard to spectral locality.

Practical Implications for Spectral Binning and Sensor Design

The cluster weight distributions analyzed in the previous section are interpretable not only in terms of classification performance but also in terms of their practical implications for sensor design. In hyperspectral imaging systems, the number of spectral channels a sensor acquires is a fundamental hardware parameter that determines cost, data throughput, and system complexity. When designing a sensor for a specific application such as vegetation monitoring or urban material mapping, one practical question is how to group the full spectrum into a smaller number of output channels, a process known as spectral binning, in a way that preserves as much class-discriminative information as possible [?].

Conventional approaches to spectral binning typically use fixed, uniformly spaced wavelength intervals or predefined spectral regions based on broad physical knowledge of the scene. The virtual band framework developed in this thesis offers an alternative: the cluster weight distributions produced by data-driven merging provide an empirical basis for deciding which wavelength regions carry redundant information and can be combined without loss, and which carry complementary information that should be preserved in separate output channels. Rather than binning by hardware convenience, this approach bins by discriminative structure learned directly from the data.

The weight analysis across the four datasets in this study reveals several patterns that are relevant to this application. In the Botswana scene, the near-zero weights on most cluster members indicate that large portions of the SWIR spectrum contribute little class-discriminative information at the level of five virtual bands, which suggests that a sensor optimized for this scene could use aggressive binning across those regions without meaningful loss while concentrating detector resources near 1750 nm. In Indian Pines, the broadly distributed weights under PCA merging indicate that agricultural class discrimination depends on gradual spectral trends spanning wide wavelength ranges rather than narrow absorption features, which argues for moderate bin widths that preserve these trends rather than fine spectral resolution that captures individual absorption peaks. The Montana result is perhaps the most actionable: the highly concentrated weight near 800 nm corresponding to the vegetation NIR plateau onset suggests that a sensor designed for vegetation class discrimination in a trees-grass-soil scene should treat this narrow wavelength region as a dedicated output channel rather than binning it with adjacent NIR bands. This is consistent with the well-established importance of the red-edge region in vegetation remote sensing [31, 49] and provides a data-driven confirmation of what physical intuition already suggests.

More broadly, the virtual band framework can be viewed as a tool for translating classification requirements into sensor specifications. Given a set of labeled hyperspectral data from a target scene, the framework produces a set of cluster boundaries and weight distributions that describe which parts of the spectrum are most important for discrimination and how they should be combined. These outputs could directly inform the design of a custom multispectral sensor optimized for that scene, reducing the number of required channels while preserving the spectral structure that drives classification performance. The connection between the data-driven merging results and the physical optics of spectral binning is one of the more practically significant implications of this work, and remains

an active direction for future investigation.

It is also worth noting the relationship between the cluster structures produced by different clustering methods and the established physical boundaries between spectral regions described in Chapter 2. In several configurations, particularly those using distance-based clustering such as K-Means and Agglomerative, the learned cluster boundaries align reasonably well with the physically meaningful transitions between the visible, red-edge, NIR, and SWIR regions. The Pavia K-Means result is a clear example of this, where the five clusters partition the visible and NIR spectrum into segments that correspond closely to known spectral subregions. This alignment was not imposed by the algorithm, which operates on statistical similarity in the data rather than on prior knowledge of spectral physics, but it emerges naturally from the correlation structure of real hyperspectral scenes. When data-driven cluster boundaries agree with physical spectral region boundaries, it provides a degree of validation that the learned representations are capturing genuine physical structure rather than dataset-specific artifacts. When they disagree, as in the non-contiguous UMRMR clusters for Montana, it raises the question of whether the information-theoretic grouping is identifying genuinely complementary cross-region relationships or simply reflecting the idiosyncrasies of the training data, a question that warrants further investigation as the framework is applied to new scenes and sensors.

THESIS CONCLUSION

This thesis investigated cluster-based virtual band construction as a dimensionality reduction strategy for hyperspectral image classification, motivated by a fundamental limitation shared by most band selection approaches: when a single representative band is chosen from a group of correlated channels, the remaining members of that group are discarded entirely, along with whatever complementary spectral information they carry. The work brought together three components — a grounding in hyperspectral dimensionality reduction and spectral interpretability, a systematic study of clustering methods for organizing spectral bands into coherent groups, and a comparative evaluation of weighted merging strategies for combining those groups into representative virtual bands — and examined how these components interact across four diverse datasets and two classifiers.

The background established why dimensionality reduction is not simply a preprocessing convenience in hyperspectral classification but a modeling requirement. When the number of spectral features far exceeds the number of labeled training samples, classifiers risk fitting dataset-specific artifacts rather than the underlying spectral structure that generalizes across scenes. Feature extraction methods such as PCA address this by projecting bands into a compact latent space, but at the cost of wavelength-level interpretability. Band selection preserves interpretability by retaining original channels, but forces a binary keep-or-discard decision that wastes correlated information. Virtual band construction was positioned as a middle ground: by grouping bands into clusters and merging each cluster into a single weighted feature, the framework retains a direct connection to physical wavelengths while still compressing redundant information rather than discarding it outright. This framing connected the motivation for the work to both the classification accuracy goals of the machine learning community and the spectral interpretability requirements of the remote sensing and optics communities.

The clustering chapter and the merging chapter are best understood as two complementary stages of a single pipeline rather than independent contributions. Clustering determines the structure of the problem that merging has to solve: methods that form compact, spectrally coherent groups give the merging step clean, well-structured inputs, while methods that produce heterogeneous or non-contiguous clusters present a different kind of challenge. K-Means and Agglomerative clustering organized bands based on distance and linkage structure, producing the contiguous spectral groupings seen clearly in the Pavia weight visualizations. Spectral clustering captured nonlinear inter-band relationships that Euclidean methods miss, which proved advantageous on Indian Pines where agricultural class boundaries are subtle and spread across broad wavelength ranges. E-FDPC and UMRMR, adapted from selection algorithms by treating their chosen bands as cluster anchors, brought information-theoretic relevance criteria into the grouping stage, producing clusters whose members are related to their anchor in a mutual-information sense rather than a purely geometric one. This diversity in clustering behavior meant that the merging step encountered fundamentally different inputs depending on the method used, which is why evaluating all combinations rather than committing to a single pipeline was central to the experimental design.

Among merging strategies, PCA-based merging was the most consistently strong performer across datasets and classifiers. By aligning each virtual band with the direction of maximum within-cluster variance, it captures the dominant spectral trend of the cluster without requiring an explicit band ranking or threshold. Entropy-based merging directed weight toward bands with richer intensity distributions and performed competitively in several configurations, while inverse redundancy merging provided targeted behavior in cases where cluster members varied substantially in their correlation with one another. Uniform merging, despite making no assumptions about band quality, remained a competitive baseline in many configurations, which reflects the fact that when clustering has already organized

bands into spectrally coherent groups, even a simple average preserves much of the relevant structure. Across all datasets and reduction targets from $k=50$ down to the most aggressive setting of $k=5$, merging-based representations matched or exceeded conventional band selection baselines, with the most pronounced improvements on spectrally complex scenes such as Indian Pines where discriminative information is distributed across many correlated wavelengths.

The interpretability analysis in Chapter 6 provided a concrete basis for understanding why merging improves over selection. The cluster weight visualizations showed that merging strategies respond to the actual spectral structure within each cluster rather than simply averaging its members. On Indian Pines, PCA merging produced broadly distributed weight profiles that tracked gradual spectral trends across wide wavelength ranges, consistent with the distributed nature of agricultural class differences in that scene. On Pavia, K-Means produced clean contiguous clusters whose weights increased smoothly with wavelength, reflecting a consistent spectral trend within each region. On Montana, UMRMR identified a highly concentrated weight near 800 nm corresponding to the red-edge transition in vegetation reflectance, a physically meaningful spectral feature that carries strong discriminative information for distinguishing vegetation from soil. These patterns demonstrate that the reduced representations are not abstract — they can be traced back to specific wavelength regions and interpreted in terms of known spectral physics.

Future Work

This interpretability has practical implications that extend beyond classification accuracy. The cluster weight vectors produced by the merging framework describe, for each virtual band, which original wavelengths contributed and in what proportion. This information could directly inform spectral binning decisions in sensor design, where the goal is to combine detector elements across a wavelength range into a smaller number

of output channels. Rather than choosing bin boundaries arbitrarily or based on sensor convenience, the weight distributions learned from data provide an empirical basis for deciding which wavelength regions carry redundant information and can be combined without loss, and which carry complementary information that should be preserved in separate bins. This connection between data-driven merging and hardware-level design is a direction that warrants further investigation, particularly for applications where custom sensor configurations are feasible and spectral efficiency is a priority.

A related direction concerns the definition and physical interpretation of spectral regions within hyperspectral data. The current framework groups bands based on statistical similarity or information-theoretic criteria, but the resulting clusters can be compared against established spectral region boundaries — visible, near-infrared, red-edge, shortwave infrared — to assess whether data-driven groupings align with physically meaningful divisions of the spectrum. For the Montana burn plot data, the emergence of the red-edge as a dominant feature in the weight analysis suggests that the framework is recovering physically grounded structure without being explicitly told where the red-edge is. Extending this analysis across more scenes and sensor configurations, and comparing data-driven cluster boundaries against established spectral indices and absorption features, would strengthen the connection between the statistical framework developed here and the physical optics of hyperspectral sensing.

Several other directions remain open. The evaluation covered two classifiers chosen for their stability and interpretability as baselines. Extending the evaluation to support vector machines, sparse representation classifiers, or spatial-spectral deep learning architectures would clarify whether the observed merging benefits persist under more expressive models [11, 29]. The four datasets used here span meaningful variation in scene type and spectral resolution but do not cover the full diversity of hyperspectral sensing — longer wavelength ranges, finer spectral sampling, or sensors with different noise characteristics

could reveal edge cases where the framework underperforms. Computational cost is also a practical concern for certain clustering strategies at scale, and future work could explore approximate or GPU-accelerated implementations that make virtual band construction tractable for large scenes or real-time applications. The current interpretability analysis characterizes the importance of individual bands within their cluster but does not address the relative importance of clusters themselves. Extending the framework to quantify how much each virtual band contributes to overall classification performance — for example through permutation importance or Shapley values applied to the virtual band representation — would provide a more complete picture of which spectral regions drive discrimination at both the within-cluster and across-cluster levels, with direct implications for sensor channel prioritization.

Despite these open questions, the results of this thesis establish that cluster-based virtual band construction is an effective and interpretable approach to hyperspectral dimensionality reduction. By allowing all cluster members to contribute proportionally to a merged virtual band rather than discarding all but one, the framework preserves spectral structure that strict selection throws away, without sacrificing the wavelength-level interpretability that feature extraction methods give up. The cluster weights provide a transparent and physically grounded record of what each reduced feature represents, making it possible to connect classification performance back to specific parts of the spectrum — and ultimately, to the sensor design and scientific analysis decisions that depend on knowing which parts of the spectrum matter.

REFERENCES CITED

- [1] *Entropy, Relative Entropy, and Mutual Information*, chapter 2, pages 13–55. John Wiley & Sons, Ltd, 2005. ISBN 9780471748823. doi: 10.1002/047174882X.ch2.
- [2] A. G. Asuero, A. Sayago, and A. G. González. The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry*, 36(1):41–59, 2006. doi: 10.1080/10408340500526766.
- [3] Mariana Belgiu and Lucian Drăguț. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016. ISSN 0924-2716. doi: <https://doi.org/10.1016/j.isprsjprs.2016.01.011>. URL <https://www.sciencedirect.com/science/article/pii/S0924271616000265>.
- [4] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491, 2005. doi: 10.1109/TGRS.2004.842478.
- [5] Jose M. Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, 2013. doi: 10.1109/MGRS.2013.2244672.
- [6] Leo Breiman. Random forests. 45(1):5–32. doi: 10.1023/A:1010933404324.
- [7] G. Camps-Valls and L. Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362, 2005. doi: 10.1109/TGRS.2005.846154.
- [8] Chein-I Chang and Keng-Hao Liu. Progressive band selection of spectral unmixing for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(4):2002–2017, 2014. doi: 10.1109/TGRS.2013.2257604.
- [9] Chein-I Chang, Qian Du, Tzu-Lung Sun, and M.L.G. Althouse. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2631–2641, 1999. doi: 10.1109/36.803411.
- [10] Chein-I Chang, Yi-Mei Kuo, and Kenneth Yeonkong Ma. Band selection via band density prominence clustering for hyperspectral image classification. *Remote Sensing*, 16(6), 2024. ISSN 2072-4292. doi: 10.3390/rs16060942. URL <https://www.mdpi.com/2072-4292/16/6/942>.
- [11] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, 2014. doi: 10.1109/JSTARS.2014.2329330.

- [12] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967. doi: 10.1109/TIT.1967.1053964.
- [13] Paul J Curran. Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3):271–278, 1989. ISSN 0034-4257. doi: [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2). URL <https://www.sciencedirect.com/science/article/pii/0034425789900692>.
- [14] CHRISTOPHER D. ELVIDGE. Visible and near infrared reflectance characteristics of dry plant materials. *International Journal of Remote Sensing*, 11(10):1775–1795, 1990. doi: 10.1080/01431169008955129. URL <https://doi.org/10.1080/01431169008955129>.
- [15] Mathieu Fauvel, Yuliya Tarabalka, Jón Atli Benediktsson, Jocelyn Chanussot, and James C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013. doi: 10.1109/JPROC.2012.2197589.
- [16] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. doi: 10.1111/j.1469-1809.1936.tb02137.x.
- [17] Pedram Ghamisi, Emmanuel Maggiori, Shutao Li, Roberto Souza, Yuliya Tarablaka, Gabriele Moser, Andrea De Giorgi, Leyuan Fang, Yushi Chen, Mingmin Chi, Sebastiano B. Serpico, and Jón Atli Benediktsson. New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geoscience and Remote Sensing Magazine*, 6(3):10–43, 2018. doi: 10.1109/MGRS.2018.2854840.
- [18] Anatoly A. Gitelson and Mark N. Merzlyak. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 148(3):494–500, 1996. ISSN 0176-1617. doi: [https://doi.org/10.1016/S0176-1617\(96\)80284-7](https://doi.org/10.1016/S0176-1617(96)80284-7). URL <https://www.sciencedirect.com/science/article/pii/S0176161796802847>.
- [19] Alexander F. H. Goetz, Gregg Vane, Jerry E. Solomon, and Barrett N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985. doi: 10.1126/science.228.4704.1147. URL <https://www.science.org/doi/abs/10.1126/science.228.4704.1147>.
- [20] Alexander F.H. Goetz. Three decades of hyperspectral remote sensing of the earth: A personal view. *Remote Sensing of Environment*, 113:S5–S16, 2009. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2007.12.014>. URL <https://www.sciencedirect.com/science/article/pii/S003442570900073X>. Imaging Spectroscopy Special Issue.

- [21] J. Ham, Yangchi Chen, M.M. Crawford, and J. Ghosh. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501, 2005. doi: 10.1109/TGRS.2004.842481.
- [22] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, 1968. doi: 10.1109/TIT.1968.1054102.
- [23] Sen Jia, Guihua Tang, Jiasong Zhu, and Qingquan Li. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):88–102, 2016. doi: 10.1109/TGRS.2015.2450759.
- [24] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, NY, 2 edition, 2002. ISBN 978-0-387-95442-4. doi: 10.1007/b98835.
- [25] Mahmad Isaq Karankot, Bradley M. Whitaker, Xiaobing Zhou, and Muhammad Umer Masood. Attention and edge-aware band selection for efficient hyperspectral classification of burned vegetation. In *2025 IEEE 35th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2025. doi: 10.1109/MLSP62443.2025.11204320.
- [26] Chen Ke. Military object detection using multiple information extracted from hyperspectral imagery. In *2017 International Conference on Progress in Informatics and Computing (PIC)*, pages 124–128, 2017. doi: 10.1109/PIC.2017.8359527.
- [27] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. Wiley Series in Remote Sensing and Image Processing. Wiley, 2003. ISBN 9780471420286.
- [28] Qiang Li, Qi Wang, and Xuelong Li. An efficient clustering method for hyperspectral optimal band selection via shared nearest neighbor. *Remote Sensing*, 11(3), 2019. ISSN 2072-4292. doi: 10.3390/rs11030350. URL <https://www.mdpi.com/2072-4292/11/3/350>.
- [29] Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jón Atli Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019. doi: 10.1109/TGRS.2019.2907932.
- [30] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.
- [31] Riley D. Logan, Madison A. Torrey, Rafael Feijó-Lima, Benjamin P. Colman, H. Maurice Valett, and Joseph A. Shaw. Uav-based hyperspectral imaging for river algae pigment estimation. *Remote Sensing*, 15(12), 2023. ISSN 2072-4292. doi: 10.3390/rs15123148. URL <https://www.mdpi.com/2072-4292/15/12/3148>.

- [32] Dimitris G. Manolakis, Ronald B. Lockwood, and Thomas W. Cooley. *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*. Cambridge University Press, 2016.
- [33] Adolfo Martínez-Usó, Filiberto Pla, José Martínez Sotoca, and Pedro García-Sevilla. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4158–4171, 2007. doi: 10.1109/TGRS.2007.904951.
- [34] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf.
- [35] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [36] Hanchuan Peng, Fuhui Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005. doi: 10.1109/TPAMI.2005.159.
- [37] Robert Pike, Guolan Lu, Dongsheng Wang, Zhuo Georgia Chen, and Baowei Fei. A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging. *IEEE Transactions on Biomedical Engineering*, 63(3):653–663, 2016. doi: 10.1109/TBME.2015.2468578.
- [38] Antonio Plaza, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, Mattia Marconcini, James C. Tilton, and Giovanna Trianni. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:S110–S122, 2009. doi: 10.1016/j.rse.2007.07.028.
- [39] Manoharan Prabukumar and Sawant Shrutika. Band clustering using expectation–maximization algorithm and weighted average fusion-based feature extraction for hyperspectral image classification. *Journal of Applied Remote Sensing*, 12(4):046015, 2018. doi: 10.1117/1.JRS.12.046015.
- [40] Shaheera Rashwan and Nicolas Dobigeon. A split-and-merge approach for hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 14(8):1378–1382, 2017. doi: 10.1109/LGRS.2017.2713462.

- [41] Behnood Rasti, Danfeng Hong, Renlong Hang, Pedram Ghamisi, Xudong Kang, Jocelyn Chanussot, and Jon Atli Benediktsson. Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox. *IEEE Geoscience and Remote Sensing Magazine*, 8(4):60–88, 2020. doi: 10.1109/MGRS.2020.2979764.
- [42] Pablo Ribalta Lorenzo, Lukasz Tulczyjew, Michal Marcinkiewicz, and Jakub Nalepa. Hyperspectral band selection using attention-based convolutional neural networks. *IEEE Access*, 8:42384–42403, 2020. doi: 10.1109/ACCESS.2020.2977454.
- [43] Jianbo Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. doi: 10.1109/34.868688.
- [44] Weiwei Sun and Qian Du. Hyperspectral band selection: A review. *IEEE Geoscience and Remote Sensing Magazine*, 7(2):118–139, 2019. doi: 10.1109/MGRS.2019.2911100.
- [45] Weiwei Sun, Liangpei Zhang, Bo Du, Weiyue Li, and Yenming Mark Lai. Band selection using improved sparse subspace clustering for hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2784–2797, 2015. doi: 10.1109/JSTARS.2015.2417156.
- [46] Weiwei Sun, Jiangtao Peng, Gang Yang, and Qian Du. Fast and latent low-rank subspace clustering for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):3906–3915, 2020. doi: 10.1109/TGRS.2019.2959342.
- [47] Ming Yang Teng, Ruby Mehrubeoglu, Scott A. King, Kirk Cammarata, and James Simons. Investigation of epifauna coverage on seagrass blades using spatial and spectral analysis of hyperspectral images. In *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, 2013. doi: 10.1109/WHISPERS.2013.8080658.
- [48] Julius Tschannerl, Jinchang Ren, Peter Yuen, Genyun Sun, Huimin Zhao, Zhijing Yang, Zheng Wang, and Stephen Marshall. Mimr-dgsa: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm. *Information Fusion*, 51:189–200, 2019. ISSN 1566-2535. doi: <https://doi.org/10.1016/j.inffus.2019.02.005>. URL <https://www.sciencedirect.com/science/article/pii/S1566253518300666>.
- [49] Daan van Dijk, Sorosh Shoaie, Thijs van Leeuwen, and Sander Veraverbeke. Spectral signature analysis of false positive burned area detection from agricultural harvests using Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 97:102296, 2021. doi: 10.1016/j.jag.2021.102296.
- [50] Jun Wang, Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, and En Zhu. Graph regularized spatial-spectral subspace clustering for hyperspectral band selection. *Neural Networks*, 153:292–302, 2022. ISSN 0893-6080. doi: <https://doi.org/10.1016/j.neunet.2022.03.005>.

neunet.2022.06.016. URL <https://www.sciencedirect.com/science/article/pii/S0893608022002313>.

- [51] Qi Wang, Fahong Zhang, and Xuelong Li. Optimal clustering framework for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10): 5910–5922, 2018. doi: 10.1109/TGRS.2018.2828161.
- [52] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. ISSN 01621459, 1537274X. URL <http://www.jstor.org/stable/2282967>.
- [53] Buyun Xu, Xihai Li, Weijun Hou, Yiting Wang, and Yiwei Wei. A similarity-based ranking method for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11):9585–9599, 2021. doi: 10.1109/TGRS.2020.3048138.
- [54] Junling Xu. Adapt the mRMR criterion for unsupervised feature selection. In Longbing Cao, Jiang Zhong, and Yong Feng, editors, *Advanced Data Mining and Applications*, pages 111–121. Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-17313-4_11.

APPENDIX

FULL CLASSIFICATION RESULTS

This appendix contains classification results for all clustering-merging configurations at $k \in \{10, 20, 30, 40, 50\}$ for each dataset. Results at $k = 5$ are reported in Chapter 5. Tables are organized by dataset and reduction target.

BOTSWANA RESULTS

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9915	0.9874	0.9844	0.9915	0.9907	0.9874	0.9848	0.9907
KMeans	IR	0.9915	0.9874	0.9844	0.9915	0.9906	0.9874	0.9847	0.9906
KMeans	PCA	0.9915	0.9874	0.9847	0.9915	0.9906	0.9874	0.9847	0.9906
KMeans	Uniform	0.9915	0.9874	0.9847	0.9915	0.9906	0.9874	0.9847	0.9906
Spectral	Entropy	0.9915	0.9875	0.9845	0.9915	0.9904	0.9873	0.9849	0.9904
Spectral	IR	0.9915	0.9875	0.9845	0.9915	0.9904	0.9873	0.9848	0.9904
Spectral	PCA	0.9915	0.9875	0.9841	0.9915	0.9904	0.9873	0.9848	0.9904
Spectral	Uniform	0.9915	0.9874	0.9844	0.9915	0.9904	0.9873	0.9848	0.9904
Agglomerative	Entropy	0.9915	0.9875	0.9850	0.9915	0.9906	0.9875	0.9852	0.9906
Agglomerative	IR	0.9915	0.9874	0.9844	0.9915	0.9906	0.9874	0.9851	0.9906
Agglomerative	PCA	0.9915	0.9875	0.9848	0.9915	0.9906	0.9875	0.9852	0.9906
Agglomerative	Uniform	0.9915	0.9874	0.9839	0.9915	0.9906	0.9874	0.9850	0.9906
E-FDPC	Entropy	0.9915	0.9873	0.9844	0.9915	0.9908	0.9875	0.9848	0.9908
E-FDPC	IR	0.9915	0.9874	0.9846	0.9915	0.9907	0.9874	0.9847	0.9907
E-FDPC	PCA	0.9915	0.9874	0.9845	0.9915	0.9907	0.9875	0.9848	0.9907
E-FDPC	Uniform	0.9915	0.9874	0.9837	0.9915	0.9907	0.9874	0.9847	0.9907
UMRMR	Entropy	0.9914	0.9873	0.9836	0.9914	0.9905	0.9873	0.9847	0.9905
UMRMR	IR	0.9915	0.9875	0.9844	0.9915	0.9906	0.9874	0.9850	0.9906
UMRMR	PCA	0.9915	0.9874	0.9844	0.9915	0.9907	0.9874	0.9850	0.9907
UMRMR	Uniform	0.9915	0.9875	0.9846	0.9915	0.9906	0.9874	0.9849	0.9906
E-FDPC	None	0.9915	0.9873	0.9839	0.9915	0.9911	0.9874	0.9847	0.9911
UMRMR	None	0.9915	0.9875	0.9857	0.9915	0.9906	0.9875	0.9852	0.9906

Table A.1: Botswana Performance – 10 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9915	0.9874	0.9844	0.9915	0.9906	0.9878	0.9859	0.9906
KMeans	IR	0.9916	0.9875	0.9848	0.9916	0.9905	0.9878	0.9859	0.9905
KMeans	PCA	0.9915	0.9875	0.9847	0.9915	0.9906	0.9878	0.9859	0.9906
KMeans	Uniform	0.9915	0.9875	0.9848	0.9915	0.9905	0.9877	0.9858	0.9905
Spectral	Entropy	0.9915	0.9875	0.9846	0.9915	0.9910	0.9880	0.9861	0.9910
Spectral	IR	0.9915	0.9875	0.9850	0.9915	0.9909	0.9880	0.9860	0.9909
Spectral	PCA	0.9915	0.9875	0.9850	0.9915	0.9909	0.9879	0.9860	0.9909
Spectral	Uniform	0.9915	0.9875	0.9854	0.9915	0.9909	0.9880	0.9861	0.9909
Agglomerative	Entropy	0.9916	0.9876	0.9852	0.9916	0.9907	0.9879	0.9861	0.9907
Agglomerative	IR	0.9916	0.9876	0.9853	0.9916	0.9907	0.9879	0.9860	0.9907
Agglomerative	PCA	0.9915	0.9875	0.9851	0.9915	0.9907	0.9879	0.9860	0.9907
Agglomerative	Uniform	0.9915	0.9875	0.9846	0.9915	0.9907	0.9879	0.9860	0.9907
E-FDPC	Entropy	0.9915	0.9874	0.9856	0.9915	0.9909	0.9880	0.9859	0.9909
E-FDPC	IR	0.9915	0.9874	0.9843	0.9915	0.9909	0.9880	0.9860	0.9909
E-FDPC	PCA	0.9915	0.9874	0.9836	0.9915	0.9910	0.9880	0.9860	0.9910
E-FDPC	Uniform	0.9915	0.9874	0.9841	0.9915	0.9909	0.9880	0.9859	0.9909
UMRMR	Entropy	0.9915	0.9874	0.9840	0.9915	0.9906	0.9877	0.9856	0.9906
UMRMR	IR	0.9915	0.9874	0.9842	0.9915	0.9906	0.9877	0.9857	0.9906
UMRMR	PCA	0.9915	0.9874	0.9848	0.9915	0.9905	0.9875	0.9854	0.9905
UMRMR	Uniform	0.9915	0.9874	0.9843	0.9915	0.9906	0.9877	0.9857	0.9906
E-FDPC	None	0.9915	0.9873	0.9846	0.9915	0.9908	0.9872	0.9848	0.9908
UMRMR	None	0.9915	0.9875	0.9846	0.9915	0.9907	0.9880	0.9862	0.9907

Table A.2: Botswana Performance – 20 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9916	0.9875	0.9851	0.9916	0.9907	0.9881	0.9862	0.9907
KMeans	IR	0.9916	0.9876	0.9850	0.9916	0.9907	0.9881	0.9862	0.9907
KMeans	PCA	0.9915	0.9874	0.9844	0.9915	0.9907	0.9881	0.9862	0.9907
KMeans	Uniform	0.9916	0.9876	0.9861	0.9916	0.9908	0.9881	0.9862	0.9908
Spectral	Entropy	0.9915	0.9875	0.9847	0.9915	0.9909	0.9880	0.9864	0.9909
Spectral	IR	0.9916	0.9875	0.9847	0.9916	0.9909	0.9880	0.9864	0.9909
Spectral	PCA	0.9916	0.9875	0.9854	0.9916	0.9910	0.9881	0.9871	0.9910
Spectral	Uniform	0.9916	0.9876	0.9847	0.9916	0.9909	0.9880	0.9864	0.9909
Agglomerative	Entropy	0.9915	0.9874	0.9849	0.9915	0.9909	0.9880	0.9859	0.9909
Agglomerative	IR	0.9916	0.9875	0.9853	0.9916	0.9909	0.9880	0.9860	0.9909
Agglomerative	PCA	0.9915	0.9875	0.9850	0.9915	0.9909	0.9880	0.9860	0.9909
Agglomerative	Uniform	0.9915	0.9875	0.9854	0.9915	0.9909	0.9880	0.9860	0.9909
E-FDPC	Entropy	0.9915	0.9874	0.9849	0.9915	0.9910	0.9882	0.9865	0.9910
E-FDPC	IR	0.9916	0.9875	0.9861	0.9916	0.9910	0.9881	0.9864	0.9910
E-FDPC	PCA	0.9915	0.9875	0.9845	0.9915	0.9910	0.9881	0.9864	0.9910
E-FDPC	Uniform	0.9915	0.9875	0.9856	0.9915	0.9910	0.9881	0.9864	0.9910
UMRMR	Entropy	0.9915	0.9875	0.9847	0.9915	0.9909	0.9882	0.9867	0.9909
UMRMR	IR	0.9915	0.9874	0.9849	0.9915	0.9908	0.9881	0.9865	0.9908
UMRMR	PCA	0.9916	0.9875	0.9849	0.9916	0.9909	0.9882	0.9867	0.9909
UMRMR	Uniform	0.9915	0.9875	0.9855	0.9915	0.9908	0.9881	0.9865	0.9908
E-FDPC	None	0.9915	0.9873	0.9846	0.9915	0.9911	0.9876	0.9855	0.9911
UMRMR	None	0.9916	0.9875	0.9857	0.9916	0.9909	0.9883	0.9863	0.9909

Table A.3: Botswana Performance – 30 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9915	0.9875	0.9853	0.9915	0.9910	0.9882	0.9867	0.9910
KMeans	IR	0.9915	0.9875	0.9850	0.9915	0.9910	0.9883	0.9867	0.9910
KMeans	PCA	0.9916	0.9876	0.9861	0.9916	0.9910	0.9883	0.9868	0.9910
KMeans	Uniform	0.9915	0.9875	0.9848	0.9915	0.9910	0.9883	0.9867	0.9910
Spectral	Entropy	0.9915	0.9874	0.9845	0.9915	0.9910	0.9884	0.9871	0.9910
Spectral	IR	0.9916	0.9875	0.9861	0.9916	0.9911	0.9884	0.9872	0.9911
Spectral	PCA	0.9916	0.9876	0.9854	0.9916	0.9911	0.9885	0.9873	0.9911
Spectral	Uniform	0.9916	0.9875	0.9851	0.9916	0.9911	0.9884	0.9871	0.9911
Agglomerative	Entropy	0.9916	0.9876	0.9864	0.9916	0.9909	0.9884	0.9867	0.9909
Agglomerative	IR	0.9916	0.9876	0.9852	0.9916	0.9909	0.9883	0.9867	0.9909
Agglomerative	PCA	0.9915	0.9875	0.9862	0.9915	0.9909	0.9883	0.9867	0.9909
Agglomerative	Uniform	0.9916	0.9876	0.9860	0.9916	0.9909	0.9883	0.9867	0.9909
E-FDPC	Entropy	0.9915	0.9874	0.9853	0.9915	0.9911	0.9882	0.9864	0.9911
E-FDPC	IR	0.9915	0.9875	0.9847	0.9915	0.9910	0.9881	0.9863	0.9910
E-FDPC	PCA	0.9916	0.9875	0.9848	0.9916	0.9911	0.9883	0.9864	0.9911
E-FDPC	Uniform	0.9915	0.9874	0.9846	0.9915	0.9911	0.9882	0.9863	0.9911
UMRMR	Entropy	0.9916	0.9876	0.9857	0.9916	0.9906	0.9880	0.9862	0.9906
UMRMR	IR	0.9915	0.9875	0.9855	0.9915	0.9907	0.9881	0.9864	0.9907
UMRMR	PCA	0.9916	0.9876	0.9856	0.9916	0.9907	0.9881	0.9863	0.9907
UMRMR	Uniform	0.9916	0.9876	0.9856	0.9916	0.9907	0.9881	0.9864	0.9907
E-FDPC	None	0.9915	0.9873	0.9843	0.9915	0.9913	0.9880	0.9865	0.9913
UMRMR	None	0.9915	0.9875	0.9846	0.9915	0.9909	0.9885	0.9868	0.9909

Table A.4: Botswana Performance – 40 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
K-Means	Entropy	0.9915	0.9875	0.9854	0.9915	0.9909	0.9883	0.9867	0.9909
K-Means	IR	0.9916	0.9877	0.9861	0.9916	0.9909	0.9883	0.9867	0.9909
K-Means	PCA	0.9915	0.9875	0.9855	0.9915	0.9909	0.9883	0.9867	0.9909
K-Means	Uniform	0.9916	0.9876	0.9859	0.9916	0.9909	0.9883	0.9867	0.9909
Spectral	Entropy	0.9915	0.9875	0.9853	0.9915	0.9913	0.9886	0.9877	0.9913
Spectral	IR	0.9915	0.9875	0.9854	0.9915	0.9912	0.9885	0.9875	0.9912
Spectral	PCA	0.9916	0.9875	0.9854	0.9916	0.9913	0.9886	0.9877	0.9913
Spectral	Uniform	0.9915	0.9875	0.9860	0.9915	0.9912	0.9885	0.9875	0.9912
Agglomerative	Entropy	0.9915	0.9875	0.9850	0.9915	0.9910	0.9885	0.9870	0.9910
Agglomerative	IR	0.9916	0.9875	0.9850	0.9916	0.9910	0.9885	0.9870	0.9910
Agglomerative	PCA	0.9916	0.9876	0.9856	0.9916	0.9910	0.9885	0.9870	0.9910
Agglomerative	Uniform	0.9916	0.9875	0.9850	0.9916	0.9910	0.9885	0.9870	0.9910
E-FDPC	Entropy	0.9915	0.9874	0.9839	0.9915	0.9912	0.9883	0.9869	0.9912
E-FDPC	IR	0.9915	0.9875	0.9851	0.9915	0.9911	0.9883	0.9867	0.9911
E-FDPC	PCA	0.9915	0.9875	0.9847	0.9915	0.9910	0.9882	0.9867	0.9910
E-FDPC	Uniform	0.9915	0.9875	0.9845	0.9915	0.9911	0.9883	0.9867	0.9911
UMRMR	Entropy	0.9916	0.9875	0.9856	0.9916	0.9907	0.9883	0.9866	0.9907
UMRMR	IR	0.9915	0.9874	0.9848	0.9915	0.9907	0.9882	0.9865	0.9907
UMRMR	PCA	0.9915	0.9875	0.9848	0.9915	0.9907	0.9883	0.9866	0.9907
UMRMR	Uniform	0.9915	0.9875	0.9855	0.9915	0.9907	0.9883	0.9866	0.9907
E-FDPC	None	0.9915	0.9886	0.9868	0.9915	0.9913	0.9881	0.9868	0.9913
UMRMR	None	0.9915	0.9875	0.9869	0.9915	0.9910	0.9886	0.9869	0.9910

Table A.5: Botswana Performance – 50 Bands

INDIAN PINES

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.7674	0.7545	0.7517	0.7674	0.7122	0.7018	0.6988	0.7122
KMeans	IR	0.7551	0.7404	0.7375	0.7551	0.6832	0.6718	0.6677	0.6832
KMeans	PCA	0.7674	0.7550	0.7514	0.7674	0.7139	0.7042	0.7032	0.7139
KMeans	Uniform	0.7631	0.7503	0.7648	0.7631	0.7132	0.7028	0.6996	0.7132
Spectral	Entropy	0.7534	0.7394	0.7360	0.7534	0.7351	0.7273	0.7230	0.7351
Spectral	IR	0.7572	0.7431	0.7397	0.7572	0.7015	0.6913	0.6881	0.7015
Spectral	PCA	0.7512	0.7370	0.7334	0.7512	0.7341	0.7257	0.7219	0.7341
Spectral	Uniform	0.7567	0.7431	0.7401	0.7567	0.7353	0.7273	0.7222	0.7353
Agglomerative	Entropy	0.7679	0.7549	0.7512	0.7679	0.6939	0.6845	0.6810	0.6939
Agglomerative	IR	0.7574	0.7436	0.7406	0.7574	0.7051	0.6952	0.6913	0.7051
Agglomerative	PCA	0.7627	0.7488	0.7461	0.7627	0.6825	0.6727	0.6679	0.6825
Agglomerative	Uniform	0.7653	0.7516	0.7490	0.7653	0.7011	0.6916	0.6886	0.7011
E-FDPC	Entropy	0.7439	0.7298	0.7251	0.7439	0.6999	0.6905	0.6855	0.6999
E-FDPC	IR	0.7593	0.7459	0.7429	0.7593	0.6882	0.6786	0.6752	0.6882
E-FDPC	PCA	0.7427	0.7280	0.7230	0.7427	0.7177	0.7095	0.7074	0.7177
E-FDPC	Uniform	0.7532	0.7392	0.7530	0.7532	0.7015	0.6922	0.6861	0.7015
UMRMR	Entropy	0.7603	0.7462	0.7421	0.7603	0.7303	0.7215	0.7165	0.7303
UMRMR	IR	0.7477	0.7328	0.7290	0.7477	0.7061	0.6980	0.6957	0.7061
UMRMR	PCA	0.7696	0.7565	0.7515	0.7696	0.7517	0.7443	0.7408	0.7517
UMRMR	Uniform	0.7512	0.7369	0.7338	0.7512	0.7256	0.7167	0.7117	0.7256
E-FDPC	None	0.5988	0.5485	0.5551	0.5988	0.5486	0.5156	0.5017	0.5486
UMRMR	None	0.7717	0.7572	0.7537	0.7717	0.7436	0.7370	0.7346	0.7436

Table A.6: Indian Pines Performance – 10 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.7817	0.7687	0.7644	0.7817	0.6649	0.6523	0.6491	0.6649
KMeans	IR	0.7836	0.7699	0.7671	0.7836	0.6635	0.6497	0.6461	0.6635
KMeans	PCA	0.7857	0.7724	0.7690	0.7857	0.6666	0.6531	0.6497	0.6666
KMeans	Uniform	0.7793	0.7649	0.7627	0.7793	0.6663	0.6531	0.6484	0.6663
Spectral	Entropy	0.7679	0.7543	0.7503	0.7679	0.7470	0.7389	0.7355	0.7470
Spectral	IR	0.7784	0.7647	0.7619	0.7784	0.7068	0.6966	0.6929	0.7068
Spectral	PCA	0.7831	0.7696	0.7664	0.7831	0.7508	0.7435	0.7423	0.7508
Spectral	Uniform	0.7753	0.7620	0.7591	0.7753	0.7465	0.7381	0.7340	0.7465
Agglomerative	Entropy	0.7834	0.7694	0.7669	0.7834	0.6371	0.6202	0.6112	0.6371
Agglomerative	IR	0.7817	0.7665	0.7648	0.7817	0.6312	0.6136	0.6043	0.6312
Agglomerative	PCA	0.7855	0.7712	0.7697	0.7855	0.6402	0.6232	0.6144	0.6402
Agglomerative	Uniform	0.7822	0.7686	0.7662	0.7822	0.6371	0.6199	0.6106	0.6371
E-FDPC	Entropy	0.7320	0.7130	0.7111	0.7320	0.6509	0.6359	0.6269	0.6509
E-FDPC	IR	0.7182	0.6982	0.7001	0.7182	0.6421	0.6256	0.6171	0.6421
E-FDPC	PCA	0.7325	0.7127	0.7142	0.7325	0.6585	0.6444	0.6364	0.6585
E-FDPC	Uniform	0.7322	0.7135	0.7119	0.7322	0.6549	0.6402	0.6316	0.6549
UMRMR	Entropy	0.7738	0.7609	0.7574	0.7738	0.7201	0.7110	0.7066	0.7201
UMRMR	IR	0.7736	0.7598	0.7567	0.7736	0.6923	0.6831	0.6817	0.6923
UMRMR	PCA	0.7765	0.7627	0.7594	0.7765	0.7075	0.6992	0.6974	0.7075
UMRMR	Uniform	0.7788	0.7660	0.7626	0.7788	0.7194	0.7098	0.7044	0.7194
E-FDPC	None	0.6166	0.5658	0.5747	0.6166	0.5748	0.5489	0.5385	0.5748
UMRMR	None	0.7838	0.7706	0.7662	0.7838	0.7479	0.7417	0.7394	0.7479

Table A.7: Indian Pines Performance – 20 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.7876	0.7739	0.7717	0.7876	0.6440	0.6291	0.6221	0.6440
KMeans	IR	0.7791	0.7658	0.7629	0.7791	0.6423	0.6275	0.6207	0.6423
KMeans	PCA	0.7891	0.7750	0.7729	0.7891	0.6438	0.6286	0.6221	0.6438
KMeans	Uniform	0.7836	0.7696	0.7667	0.7836	0.6442	0.6296	0.6235	0.6442
Spectral	Entropy	0.7798	0.7655	0.7640	0.7798	0.7477	0.7402	0.7370	0.7477
Spectral	IR	0.7769	0.7626	0.7605	0.7769	0.7225	0.7134	0.7134	0.7225
Spectral	PCA	0.7845	0.7704	0.7680	0.7845	0.7491	0.7413	0.7367	0.7491
Spectral	Uniform	0.7812	0.7678	0.7650	0.7812	0.7432	0.7350	0.7317	0.7432
Agglomerative	Entropy	0.7969	0.7838	0.7799	0.7969	0.6228	0.6059	0.5976	0.6228
Agglomerative	IR	0.7938	0.7810	0.7777	0.7938	0.6235	0.6066	0.5989	0.6235
Agglomerative	PCA	0.7931	0.7795	0.7768	0.7931	0.6219	0.6049	0.5967	0.6219
Agglomerative	Uniform	0.7922	0.7786	0.7759	0.7922	0.6219	0.6050	0.5973	0.6219
E-FDPC	Entropy	0.7577	0.7424	0.7399	0.7577	0.7146	0.7051	0.7000	0.7146
E-FDPC	IR	0.7298	0.7115	0.7126	0.7298	0.6778	0.6652	0.6596	0.6778
E-FDPC	PCA	0.7560	0.7399	0.7379	0.7560	0.7089	0.6990	0.6923	0.7089
E-FDPC	Uniform	0.7629	0.7479	0.7457	0.7629	0.7118	0.7020	0.6953	0.7118
UMRMR	Entropy	0.7879	0.7750	0.7720	0.7879	0.7149	0.7057	0.6998	0.7149
UMRMR	IR	0.7864	0.7737	0.7702	0.7864	0.6975	0.6873	0.6836	0.6975
UMRMR	PCA	0.7914	0.7791	0.7752	0.7914	0.7042	0.6949	0.6929	0.7042
UMRMR	Uniform	0.7867	0.7739	0.7701	0.7867	0.7172	0.7083	0.7044	0.7172
E-FDPC	None	0.7075	0.6854	0.6901	0.7075	0.6935	0.6814	0.6769	0.6935
UMRMR	None	0.7855	0.7720	0.7688	0.7855	0.7208	0.7103	0.7060	0.7208

Table A.8: Indian Pines Performance – 30 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.7974	0.7843	0.7810	0.7974	0.6302	0.6133	0.6038	0.6302
KMeans	IR	0.7983	0.7852	0.7818	0.7983	0.6288	0.6114	0.6020	0.6288
KMeans	PCA	0.8045	0.7914	0.7893	0.8045	0.6314	0.6144	0.6048	0.6314
KMeans	Uniform	0.7981	0.7849	0.7826	0.7981	0.6293	0.6122	0.6026	0.6293
Spectral	Entropy	0.7926	0.7792	0.7767	0.7926	0.7063	0.6951	0.6910	0.7063
Spectral	IR	0.7891	0.7751	0.7731	0.7891	0.6887	0.6772	0.6738	0.6887
Spectral	PCA	0.7912	0.7775	0.7753	0.7912	0.7070	0.6961	0.6901	0.7070
Spectral	Uniform	0.7883	0.7748	0.7730	0.7883	0.7015	0.6896	0.6846	0.7015
Agglomerative	Entropy	0.7993	0.7859	0.7827	0.7993	0.6300	0.6137	0.6066	0.6300
Agglomerative	IR	0.8024	0.7897	0.7862	0.8024	0.6316	0.6153	0.6083	0.6316
Agglomerative	PCA	0.8024	0.7890	0.7857	0.8024	0.6314	0.6151	0.6080	0.6314
Agglomerative	Uniform	0.7998	0.7870	0.7835	0.7998	0.6307	0.6144	0.6074	0.6307
E-FDPC	Entropy	0.7579	0.7419	0.7419	0.7579	0.6992	0.6873	0.6806	0.6992
E-FDPC	IR	0.7358	0.7176	0.7182	0.7358	0.6654	0.6526	0.6461	0.6654
E-FDPC	PCA	0.7565	0.7401	0.7388	0.7565	0.7054	0.6950	0.6885	0.7054
E-FDPC	Uniform	0.7612	0.7461	0.7442	0.7612	0.6949	0.6827	0.6755	0.6949
UMRMR	Entropy	0.7964	0.7844	0.7801	0.7964	0.7039	0.6944	0.6941	0.7039
UMRMR	IR	0.7974	0.7849	0.7813	0.7974	0.6832	0.6717	0.6702	0.6832
UMRMR	PCA	0.8012	0.7892	0.7859	0.8012	0.6887	0.6773	0.6741	0.6887
UMRMR	Uniform	0.7964	0.7840	0.7804	0.7964	0.6989	0.6893	0.6889	0.6989
E-FDPC	None	0.7161	0.6935	0.7013	0.7161	0.6980	0.6875	0.6822	0.6980
UMRMR	None	0.7950	0.7822	0.7787	0.7950	0.7294	0.7202	0.7183	0.7294

Table A.9: Indian Pines Performance – 40 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
K-Means	Entropy	0.8026	0.7894	0.7865	0.8026	0.6350	0.6199	0.6131	0.6350
K-Means	IR	0.8002	0.7874	0.7840	0.8002	0.6357	0.6206	0.6138	0.6357
K-Means	PCA	0.8067	0.7935	0.7914	0.8067	0.6352	0.6202	0.6135	0.6352
K-Means	Uniform	0.8012	0.7878	0.7845	0.8012	0.6350	0.6198	0.6130	0.6350
Spectral	Entropy	0.7993	0.7855	0.7828	0.7993	0.6656	0.6517	0.6470	0.6656
Spectral	IR	0.8000	0.7868	0.7833	0.8000	0.6654	0.6512	0.6458	0.6654
Spectral	PCA	0.8048	0.7923	0.7890	0.8048	0.6583	0.6410	0.6343	0.6583
Spectral	Uniform	0.8036	0.7909	0.7878	0.8036	0.6637	0.6498	0.6447	0.6637
Agglomerative	Entropy	0.7974	0.7846	0.7820	0.7974	0.6404	0.6253	0.6191	0.6404
Agglomerative	IR	0.8026	0.7897	0.7880	0.8026	0.6404	0.6253	0.6191	0.6404
Agglomerative	PCA	0.7990	0.7858	0.7821	0.7990	0.6404	0.6254	0.6191	0.6404
Agglomerative	Uniform	0.8007	0.7882	0.7844	0.8007	0.6409	0.6257	0.6194	0.6409
E-FDPC	Entropy	0.7769	0.7622	0.7604	0.7769	0.7367	0.7273	0.7212	0.7367
E-FDPC	IR	0.7765	0.7618	0.7612	0.7765	0.7080	0.6985	0.6928	0.7080
E-FDPC	PCA	0.7838	0.7688	0.7666	0.7838	0.7403	0.7307	0.7240	0.7403
E-FDPC	Uniform	0.7753	0.7606	0.7598	0.7753	0.7339	0.7246	0.7195	0.7339
UMRMR	Entropy	0.7986	0.7868	0.7824	0.7986	0.7153	0.7065	0.7053	0.7153
UMRMR	IR	0.7971	0.7848	0.7807	0.7971	0.6987	0.6885	0.6852	0.6987
UMRMR	PCA	0.8012	0.7891	0.7847	0.8012	0.6958	0.6863	0.6826	0.6958
UMRMR	Uniform	0.7964	0.7840	0.7803	0.7964	0.7115	0.7033	0.7021	0.7115
E-FDPC	None	0.7272	0.7079	0.7105	0.7272	0.7082	0.6978	0.6918	0.7082
UMRMR	None	0.7967	0.7839	0.7812	0.7967	0.7175	0.7073	0.7024	0.7175

Table A.10: Indian Pines Performance – 50 Bands

PAVIA

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.8995	0.8823	0.8823	0.8995	0.8781	0.8738	0.8704	0.8781
KMeans	IR	0.8996	0.8823	0.8825	0.8996	0.8781	0.8737	0.8704	0.8781
KMeans	PCA	0.8990	0.8816	0.8805	0.8990	0.8784	0.8741	0.8709	0.8784
KMeans	Uniform	0.8996	0.8822	0.8826	0.8996	0.8783	0.8740	0.8707	0.8783
Spectral	Entropy	0.8996	0.8827	0.8816	0.8996	0.8819	0.8777	0.8744	0.8819
Spectral	IR	0.8999	0.8830	0.8827	0.8999	0.8824	0.8782	0.8750	0.8824
Spectral	PCA	0.8997	0.8830	0.8830	0.8997	0.8825	0.8782	0.8747	0.8825
Spectral	Uniform	0.8997	0.8828	0.8829	0.8997	0.8819	0.8777	0.8744	0.8819
Agglomerative	Entropy	0.8995	0.8820	0.8818	0.8995	0.8755	0.8705	0.8667	0.8755
Agglomerative	IR	0.8996	0.8820	0.8814	0.8996	0.8755	0.8706	0.8668	0.8755
Agglomerative	PCA	0.8992	0.8818	0.8814	0.8992	0.8754	0.8705	0.8666	0.8754
Agglomerative	Uniform	0.8996	0.8822	0.8817	0.8996	0.8755	0.8706	0.8668	0.8755
E-FDPC	Entropy	0.8897	0.8654	0.8626	0.8897	0.8708	0.8612	0.8541	0.8708
E-FDPC	IR	0.8894	0.8645	0.8616	0.8894	0.8704	0.8607	0.8535	0.8704
E-FDPC	PCA	0.8900	0.8659	0.8639	0.8900	0.8718	0.8624	0.8555	0.8718
E-FDPC	Uniform	0.8899	0.8652	0.8634	0.8899	0.8707	0.8611	0.8540	0.8707
UMRMR	Entropy	0.8991	0.8821	0.8816	0.8991	0.8771	0.8726	0.8691	0.8771
UMRMR	IR	0.8991	0.8818	0.8806	0.8991	0.8771	0.8726	0.8690	0.8771
UMRMR	PCA	0.8996	0.8824	0.8821	0.8996	0.8772	0.8726	0.8691	0.8772
UMRMR	Uniform	0.8991	0.8819	0.8814	0.8991	0.8771	0.8726	0.8690	0.8771
E-FDPC	None	0.8745	0.8423	0.8273	0.8745	0.8528	0.8355	0.8226	0.8528
UMRMR	None	0.8962	0.8769	0.8772	0.8962	0.8763	0.8707	0.8664	0.8763

Table A.11: Pavia Performance – 10 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9009	0.8829	0.8850	0.9009	0.8740	0.8693	0.8657	0.8740
KMeans	IR	0.9017	0.8838	0.8867	0.9017	0.8742	0.8695	0.8660	0.8742
KMeans	PCA	0.9012	0.8833	0.8844	0.9012	0.8739	0.8691	0.8656	0.8739
KMeans	Uniform	0.9011	0.8831	0.8846	0.9011	0.8742	0.8694	0.8660	0.8742
Spectral	Entropy	0.9022	0.8849	0.8866	0.9022	0.8763	0.8721	0.8689	0.8763
Spectral	IR	0.9020	0.8844	0.8866	0.9020	0.8764	0.8721	0.8690	0.8764
Spectral	PCA	0.9021	0.8848	0.8867	0.9021	0.8764	0.8722	0.8691	0.8764
Spectral	Uniform	0.9018	0.8843	0.8866	0.9018	0.8764	0.8722	0.8690	0.8764
Agglomerative	Entropy	0.9015	0.8835	0.8854	0.9015	0.8745	0.8702	0.8669	0.8745
Agglomerative	IR	0.9012	0.8832	0.8860	0.9012	0.8747	0.8704	0.8671	0.8747
Agglomerative	PCA	0.9011	0.8831	0.8850	0.9011	0.8747	0.8705	0.8672	0.8747
Agglomerative	Uniform	0.9013	0.8832	0.8851	0.9013	0.8747	0.8704	0.8672	0.8747
E-FDPC	Entropy	0.8933	0.8688	0.8665	0.8933	0.8739	0.8664	0.8608	0.8739
E-FDPC	IR	0.8924	0.8676	0.8654	0.8924	0.8733	0.8659	0.8603	0.8733
E-FDPC	PCA	0.8933	0.8690	0.8689	0.8933	0.8750	0.8675	0.8620	0.8750
E-FDPC	Uniform	0.8925	0.8681	0.8637	0.8925	0.8738	0.8664	0.8609	0.8738
UMRMR	Entropy	0.9014	0.8844	0.8854	0.9014	0.8798	0.8763	0.8736	0.8798
UMRMR	IR	0.9014	0.8845	0.8853	0.9014	0.8794	0.8758	0.8731	0.8794
UMRMR	PCA	0.9017	0.8850	0.8867	0.9017	0.8800	0.8765	0.8738	0.8800
UMRMR	Uniform	0.9013	0.8841	0.8845	0.9013	0.8798	0.8762	0.8736	0.8798
E-FDPC	None	0.8824	0.8524	0.8508	0.8824	0.8597	0.8470	0.8376	0.8597
UMRMR	None	0.8988	0.8796	0.8814	0.8988	0.8791	0.8752	0.8723	0.8791

Table A.12: Pavia Performance – 20 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9013	0.8827	0.8857	0.9013	0.8744	0.8700	0.8668	0.8744
KMeans	IR	0.9013	0.8827	0.8857	0.9013	0.8744	0.8700	0.8668	0.8744
KMeans	PCA	0.9011	0.8826	0.8859	0.9011	0.8744	0.8700	0.8668	0.8744
KMeans	Uniform	0.9017	0.8834	0.8862	0.9017	0.8744	0.8700	0.8668	0.8744
Spectral	Entropy	0.9020	0.8848	0.8866	0.9020	0.8782	0.8747	0.8721	0.8782
Spectral	IR	0.9023	0.8850	0.8872	0.9023	0.8782	0.8748	0.8722	0.8782
Spectral	PCA	0.9017	0.8841	0.8855	0.9017	0.8784	0.8748	0.8721	0.8784
Spectral	Uniform	0.9021	0.8848	0.8861	0.9021	0.8782	0.8748	0.8722	0.8782
Agglomerative	Entropy	0.9014	0.8830	0.8859	0.9014	0.8736	0.8690	0.8656	0.8736
Agglomerative	IR	0.9014	0.8831	0.8864	0.9014	0.8736	0.8690	0.8656	0.8736
Agglomerative	PCA	0.9014	0.8828	0.8855	0.9014	0.8736	0.8690	0.8655	0.8736
Agglomerative	Uniform	0.9014	0.8829	0.8857	0.9014	0.8736	0.8690	0.8656	0.8736
E-FDPC	Entropy	0.8935	0.8687	0.8755	0.8935	0.8741	0.8674	0.8624	0.8741
E-FDPC	IR	0.8936	0.8689	0.8659	0.8936	0.8733	0.8667	0.8617	0.8733
E-FDPC	PCA	0.8944	0.8702	0.8700	0.8944	0.8746	0.8680	0.8630	0.8746
E-FDPC	Uniform	0.8938	0.8692	0.8695	0.8938	0.8743	0.8676	0.8626	0.8743
UMRMR	Entropy	0.9015	0.8841	0.8857	0.9015	0.8770	0.8730	0.8699	0.8770
UMRMR	IR	0.9012	0.8837	0.8845	0.9012	0.8769	0.8728	0.8697	0.8769
UMRMR	PCA	0.9011	0.8835	0.8841	0.9011	0.8771	0.8732	0.8702	0.8771
UMRMR	Uniform	0.9012	0.8836	0.8852	0.9012	0.8769	0.8729	0.8698	0.8769
E-FDPC	None	0.8844	0.8557	0.8516	0.8844	0.8618	0.8507	0.8425	0.8618
UMRMR	None	0.8998	0.8802	0.8831	0.8998	0.8769	0.8730	0.8701	0.8769

Table A.13: Pavia Performance – 30 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
KMeans	Entropy	0.9015	0.8831	0.8862	0.9015	0.8746	0.8702	0.8668	0.8746
KMeans	IR	0.9011	0.8827	0.8857	0.9011	0.8746	0.8702	0.8669	0.8746
KMeans	PCA	0.9009	0.8822	0.8847	0.9009	0.8746	0.8702	0.8669	0.8746
KMeans	Uniform	0.9014	0.8831	0.8856	0.9014	0.8746	0.8702	0.8669	0.8746
Spectral	Entropy	0.9019	0.8844	0.8862	0.9019	0.8776	0.8738	0.8709	0.8776
Spectral	IR	0.9024	0.8851	0.8875	0.9024	0.8776	0.8738	0.8709	0.8776
Spectral	PCA	0.9017	0.8843	0.8864	0.9017	0.8775	0.8737	0.8708	0.8775
Spectral	Uniform	0.9021	0.8848	0.8874	0.9021	0.8775	0.8738	0.8709	0.8775
Agglomerative	Entropy	0.9018	0.8835	0.8865	0.9018	0.8760	0.8721	0.8693	0.8760
Agglomerative	IR	0.9018	0.8834	0.8866	0.9018	0.8760	0.8721	0.8693	0.8760
Agglomerative	PCA	0.9020	0.8837	0.8866	0.9020	0.8761	0.8722	0.8694	0.8761
Agglomerative	Uniform	0.9015	0.8832	0.8851	0.9015	0.8760	0.8722	0.8693	0.8760
E-FDPC	Entropy	0.8955	0.8722	0.8768	0.8955	0.8772	0.8722	0.8685	0.8772
E-FDPC	IR	0.8948	0.8711	0.8720	0.8948	0.8765	0.8714	0.8677	0.8765
E-FDPC	PCA	0.8956	0.8725	0.8751	0.8956	0.8776	0.8724	0.8687	0.8776
E-FDPC	Uniform	0.8952	0.8718	0.8775	0.8952	0.8771	0.8721	0.8684	0.8771
UMRMR	Entropy	0.9023	0.8849	0.8870	0.9023	0.8787	0.8754	0.8729	0.8787
UMRMR	IR	0.9024	0.8851	0.8869	0.9024	0.8789	0.8755	0.8730	0.8789
UMRMR	PCA	0.9023	0.8849	0.8872	0.9023	0.8790	0.8756	0.8730	0.8790
UMRMR	Uniform	0.9021	0.8846	0.8860	0.9021	0.8788	0.8755	0.8730	0.8788
E-FDPC	None	0.8868	0.8600	0.8574	0.8868	0.8657	0.8568	0.8501	0.8657
UMRMR	None	0.9004	0.8812	0.8850	0.9004	0.8787	0.8752	0.8725	0.8787

Table A.14: Pavia Performance – 40 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
K-Means	Entropy	0.9019	0.8835	0.8865	0.9020	0.8759	0.8721	0.8693	0.8759
K-Means	IR	0.9018	0.8835	0.8866	0.9018	0.8759	0.8721	0.8693	0.8759
K-Means	PCA	0.9015	0.8834	0.8856	0.9015	0.8759	0.8721	0.8693	0.8759
K-Means	Uniform	0.9018	0.8835	0.8869	0.9018	0.8759	0.8721	0.8693	0.8759
Spectral	Entropy	0.9014	0.8836	0.8859	0.9014	0.8782	0.8746	0.8719	0.8782
Spectral	IR	0.9019	0.8842	0.8864	0.9019	0.8781	0.8744	0.8717	0.8781
Spectral	PCA	0.9019	0.8841	0.8866	0.9019	0.8781	0.8745	0.8719	0.8781
Spectral	Uniform	0.9015	0.8838	0.8864	0.9015	0.8781	0.8745	0.8718	0.8781
Agglomerative	Entropy	0.9018	0.8833	0.8859	0.9018	0.8757	0.8718	0.8690	0.8757
Agglomerative	IR	0.9018	0.8834	0.8867	0.9018	0.8757	0.8718	0.8690	0.8757
Agglomerative	PCA	0.9017	0.8835	0.8872	0.9017	0.8757	0.8718	0.8690	0.8757
Agglomerative	Uniform	0.9017	0.8832	0.8869	0.9017	0.8757	0.8718	0.8690	0.8757
E-FDPC	Entropy	0.8969	0.8746	0.8792	0.8969	0.8789	0.8744	0.8711	0.8789
E-FDPC	IR	0.8967	0.8745	0.8824	0.8967	0.8788	0.8743	0.8710	0.8788
E-FDPC	PCA	0.8972	0.8753	0.8790	0.8972	0.8793	0.8748	0.8715	0.8793
E-FDPC	Uniform	0.8967	0.8746	0.8787	0.8967	0.8789	0.8744	0.8710	0.8789
UMRMR	Entropy	0.9018	0.8841	0.8858	0.9018	0.8802	0.8770	0.8747	0.8802
UMRMR	IR	0.9022	0.8846	0.8871	0.9022	0.8801	0.8770	0.8747	0.8801
UMRMR	PCA	0.9024	0.8849	0.8879	0.9024	0.8804	0.8772	0.8749	0.8804
UMRMR	Uniform	0.9020	0.8844	0.8868	0.9020	0.8801	0.8770	0.8747	0.8801
E-FDPC	None	0.8909	0.8669	0.8648	0.8909	0.8697	0.8637	0.8592	0.8697
UMRMR	None	0.9008	0.8818	0.8840	0.9008	0.8774	0.8741	0.8716	0.8774

Table A.15: Pavia Performance – 50 Bands

MONTANA

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
Spectral	Entropy	0.9243	0.8957	0.9000	0.9243	0.9111	0.8959	0.8853	0.9111
Spectral	IR	0.9243	0.8955	0.8999	0.9243	0.9110	0.8958	0.8852	0.9110
Spectral	PCA	0.9243	0.8955	0.8999	0.9243	0.9111	0.8959	0.8852	0.9111
Spectral	Uniform	0.9243	0.8957	0.9001	0.9243	0.9110	0.8959	0.8852	0.9110
Agglomerative	Entropy	0.9247	0.8965	0.9013	0.9247	0.9112	0.8961	0.8854	0.9112
Agglomerative	IR	0.9247	0.8965	0.9014	0.9247	0.9113	0.8961	0.8855	0.9113
Agglomerative	PCA	0.9247	0.8966	0.9014	0.9247	0.9112	0.8961	0.8855	0.9112
Agglomerative	Uniform	0.9247	0.8966	0.9014	0.9247	0.9113	0.8961	0.8855	0.9113
UMRMR	Entropy	0.9242	0.8959	0.8995	0.9242	0.9110	0.8960	0.8853	0.9110
UMRMR	IR	0.9242	0.8959	0.8995	0.9242	0.9110	0.8960	0.8854	0.9110
UMRMR	PCA	0.9242	0.8961	0.8994	0.9242	0.9110	0.8959	0.8853	0.9110
UMRMR	Uniform	0.9243	0.8959	0.8996	0.9243	0.9110	0.8960	0.8854	0.9110
UMRMR	None	0.9242	0.8939	0.9010	0.9242	0.9105	0.8947	0.8833	0.9105

Table A.16: Montana Performance – 10 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
Spectral	Entropy	0.9250	0.8964	0.9036	0.9250	0.9122	0.8973	0.8873	0.9122
Spectral	IR	0.9250	0.8962	0.9038	0.9250	0.9122	0.8972	0.8872	0.9122
Spectral	PCA	0.9250	0.8963	0.9036	0.9250	0.9122	0.8973	0.8872	0.9122
Spectral	Uniform	0.9250	0.8964	0.9038	0.9250	0.9122	0.8973	0.8873	0.9122
Agglomerative	Entropy	0.9251	0.8961	0.9044	0.9251	0.9104	0.8946	0.8832	0.9104
Agglomerative	IR	0.9251	0.8962	0.9045	0.9251	0.9104	0.8946	0.8832	0.9104
Agglomerative	PCA	0.9251	0.8962	0.9045	0.9251	0.9104	0.8946	0.8832	0.9104
Agglomerative	Uniform	0.9251	0.8962	0.9045	0.9251	0.9104	0.8946	0.8832	0.9104
UMRMR	Entropy	0.9250	0.8966	0.9033	0.9250	0.9116	0.8969	0.8868	0.9116
UMRMR	IR	0.9250	0.8966	0.9033	0.9250	0.9116	0.8969	0.8868	0.9116
UMRMR	PCA	0.9250	0.8966	0.9033	0.9250	0.9115	0.8969	0.8867	0.9115
UMRMR	Uniform	0.9250	0.8966	0.9034	0.9250	0.9115	0.8969	0.8868	0.9115
UMRMR	None	0.9246	0.8942	0.9036	0.9246	0.9113	0.8955	0.8844	0.9113

Table A.17: Montana Performance – 20 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
Spectral	Entropy	0.9252	0.8964	0.9048	0.9252	0.9128	0.8977	0.8878	0.9128
Spectral	IR	0.9252	0.8964	0.9049	0.9252	0.9127	0.8976	0.8877	0.9127
Spectral	PCA	0.9252	0.8964	0.9048	0.9252	0.9126	0.8976	0.8875	0.9126
Spectral	Uniform	0.9252	0.8964	0.9048	0.9252	0.9127	0.8977	0.8877	0.9127
Agglomerative	Entropy	0.9251	0.8951	0.9054	0.9251	0.9116	0.8955	0.8845	0.9116
Agglomerative	IR	0.9251	0.8950	0.9056	0.9251	0.9116	0.8955	0.8845	0.9116
Agglomerative	PCA	0.9250	0.8951	0.9054	0.9250	0.9116	0.8955	0.8845	0.9116
Agglomerative	Uniform	0.9251	0.8951	0.9054	0.9251	0.9116	0.8955	0.8845	0.9116
UMRMR	Entropy	0.9255	0.8975	0.9055	0.9255	0.9124	0.8976	0.8875	0.9124
UMRMR	IR	0.9255	0.8975	0.9055	0.9255	0.9124	0.8976	0.8875	0.9124
UMRMR	PCA	0.9255	0.8975	0.9055	0.9255	0.9123	0.8974	0.8873	0.9123
UMRMR	Uniform	0.9255	0.8975	0.9055	0.9255	0.9124	0.8975	0.8875	0.9124
UMRMR	None	0.9247	0.8941	0.9043	0.9247	0.9114	0.8956	0.8846	0.9114

Table A.18: Montana Performance – 30 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
Spectral	Entropy	0.9253	0.8963	0.9053	0.9253	0.9124	0.8969	0.8866	0.9124
Spectral	IR	0.9253	0.8963	0.9056	0.9253	0.9124	0.8969	0.8866	0.9124
Spectral	PCA	0.9253	0.8963	0.9054	0.9253	0.9124	0.8969	0.8866	0.9124
Spectral	Uniform	0.9253	0.8963	0.9054	0.9253	0.9124	0.8969	0.8866	0.9124
Agglomerative	Entropy	0.9250	0.8946	0.9059	0.9250	0.9131	0.8970	0.8867	0.9131
Agglomerative	IR	0.9250	0.8945	0.9059	0.9250	0.9131	0.8970	0.8868	0.9131
Agglomerative	PCA	0.9250	0.8946	0.9060	0.9250	0.9131	0.8970	0.8868	0.9131
Agglomerative	Uniform	0.9251	0.8946	0.9060	0.9251	0.9131	0.8970	0.8867	0.9131
UMRMR	Entropy	0.9255	0.8974	0.9058	0.9255	0.9128	0.8979	0.8880	0.9128
UMRMR	IR	0.9255	0.8973	0.9057	0.9255	0.9128	0.8979	0.8880	0.9128
UMRMR	PCA	0.9255	0.8973	0.9057	0.9255	0.9127	0.8978	0.8878	0.9127
UMRMR	Uniform	0.9255	0.8974	0.9057	0.9255	0.9128	0.8979	0.8880	0.9128
UMRMR	None	0.9248	0.8944	0.9047	0.9248	0.9119	0.8961	0.8853	0.9119

Table A.19: Montana Performance – 40 Bands

Clustering	Merging	RF				KNN			
		Acc	F1	Prec	Rec	Acc	F1	Prec	Rec
Spectral	Entropy	0.9252	0.8957	0.9051	0.9252	0.9119	0.8962	0.8856	0.9119
Spectral	IR	0.9252	0.8957	0.9052	0.9252	0.9119	0.8962	0.8856	0.9119
Spectral	PCA	0.9252	0.8957	0.9054	0.9252	0.9119	0.8963	0.8857	0.9119
Spectral	Uniform	0.9252	0.8957	0.9052	0.9252	0.9119	0.8962	0.8856	0.9119
Agglomerative	Entropy	0.9251	0.8949	0.9062	0.9251	0.9146	0.8986	0.8892	0.9146
Agglomerative	IR	0.9251	0.8949	0.9063	0.9251	0.9146	0.8986	0.8892	0.9146
Agglomerative	PCA	0.9251	0.8949	0.9063	0.9251	0.9146	0.8986	0.8892	0.9146
Agglomerative	Uniform	0.9251	0.8948	0.9062	0.9251	0.9146	0.8986	0.8892	0.9146
UMRMR	Entropy	0.9255	0.8972	0.9061	0.9255	0.9125	0.8975	0.8874	0.9125
UMRMR	IR	0.9255	0.8972	0.9060	0.9255	0.9125	0.8975	0.8873	0.9125
UMRMR	PCA	0.9255	0.8972	0.9060	0.9255	0.9124	0.8974	0.8872	0.9124
UMRMR	Uniform	0.9255	0.8971	0.9060	0.9255	0.9125	0.8974	0.8873	0.9125
UMRMR	None	0.9250	0.8951	0.9058	0.9250	0.9125	0.8968	0.8862	0.9125

Table A.20: Montana Performance – 50 Bands