**DIMENSIONALITY REDUCTION WITH WEIGHTED FUZZY C-MEANS FOR HYPERSPECTRAL IMAGE CLASSIFICATION**

A Thesis Presented to

The Faculty of Department of Computer Science

By

Michael Wong

In Partial Fulfillment

of Requirements for the Degree

Master of Science, Computer Science

Kennesaw State University

May 2020

**DIMENSIONALITY REDUCTION WITH WEIGHTED FUZZY C-MEANS FOR HYPERSPECTRAL IMAGE CLASSIFICATION**

Approved:

Dr. Chih-Cheng Hung – Advisor

Dr. Coskun Cetinkaya – Department Chair

Dr. Jon Preston - Dean

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Kennesaw State University, I agree that the university library shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish, this thesis may be granted by the professor under whose direction it was written, or, in his absence, by the dean of the appropriate school when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from or publication of, this thesis which involves potential financial gain will not be allowed without written permission.

Michael Wong

**Notice To Borrowers**

Unpublished theses deposited in the Library of Kennesaw State University must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this thesis is:

Michael Wong

1100 S Marietta PKWY, Marietta, GA 30060

The director of this thesis is:

Dr. Chih-Cheng Hung

1100 S Marietta PKWY, Marietta, GA 30060

Users of this thesis not regularly enrolled as students at Kennesaw State University are required to attest acceptance of the preceding stipulations by signing below. Libraries borrowing this thesis for the use of their patrons are required to see that each user records here the information requested.

# ACKNOWLEDGEMENTS

I would like to thank my parents for supporting me throughout the course of achieving my education. Without their support, none of this would have been possible. Secondly, I want to express gratitude to my advisor Dr. Chih-Cheng Hung, I would not have become the person I am today if it was not for him. Meeting different people influences you and changes you in different ways, in this case, Dr. Hung has directly impacted the direction of my future through shaping me as a student, guidance towards becoming a good researcher, and has instilled principals in me that will carry on for the remainder of my lifetime. Through patience, faith, and education Dr. Hung has taught and motivated. I am forever grateful for having such a good advisor. Thirdly, I would like to thank my significant other Dawn, my sister Jenny, my network of friends, and supportive family members that believed in and encouraged me to strive for and achieve my goals. They were there for me when times were difficult and bolstered my confidence. Lastly, I would like to thank my thesis committee members Dr. Edward Jung and Dr. Zhiling Long for their time and critical feedback.

# Publication Produced

Published:

1. Mallapragada, S., Wong, M., & Hung, C. (2018). “Dimensionality Reduction of Hyperspectral Images for Classification” (pp. 153-159). International Information Institute. *(link)*
2. Abeysinghe, W., Wong, M., Hung, C., & Bechikh, S. (2019). “Multi-Objective Evolutionary Algorithm for Image Segmentation”. IEEE SoutheastCon. *(link)*
3. Abeysinghe, W., Wong, M., Hung, C., & Bechikh, S. (2019). “A Constrained Box Algorithm for Imbalanced Data in Satellite Images.” IEEE IGARSS. *(link)*
4. Wong, M., Abeysinghe, W., Hung, C. (2019). “A Massive Self-Organizing Map for Hyperspectral Image Classification.” IEEE Whispers.

Accepted:

1. Wong, M., Hung, C. (2020). “Dimensionality Reduction with Weighted K-Means for Hyperspectral Image Classification.” IEEE IGARSS.

# ABSTRACT

K-Means and Fuzzy C-Means clustering are well-known algorithms for unsupervised learning. Commonly used for clustering, image segmentation, and classification; they are quick and practical algorithms for partitioning a dataset into k classes. However, when presented with high-dimensional and sparse data, the conventionally used Euclidean metric fails to perform well in these conditions. Affected by Hughes's phenomena, dissimilar objects may appear to be identical, resulting in poor classification performance and low-quality clusters. To solve the disadvantages of K-Means and Fuzzy C-Means, a new fuzzy clustering algorithm that incorporates feature weighting along with an alternate distance metric is introduced. The experimental results show that high-quality clusters were found and yielded a higher classification performance with hyperspectral data.

**Table of Contents**

[ACKNOWLEDGEMENTS 5](#_Toc40797554)

[Publication Produced 6](#_Toc40797555)

[ABSTRACT 7](#_Toc40797556)

[LIST OF TABLES 10](#_Toc40797557)

[LIST OF FIGURES 11](#_Toc40797558)

[CHAPTER I 12](#_Toc40797559)

[1 INTRODUCTION 12](#_Toc40797560)

[1.1 Hyperspectral imaging 12](#_Toc40797561)

[1.2 Problem Statement 13](#_Toc40797562)

[1.3 Contribution 15](#_Toc40797563)

[1.4 Thesis Organization 16](#_Toc40797564)

[CHAPTER II 17](#_Toc40797565)

[2 CLUSTERING METHODS 17](#_Toc40797566)

[2.1 Clustering 17](#_Toc40797567)

[2.2 K-Means Clustering 17](#_Toc40797568)

[2.3 Fuzzy C-Means Clustering 19](#_Toc40797569)

[2.4 Feature Weighted Clustering 21](#_Toc40797570)

[CHAPTER III 25](#_Toc40797571)

[3 RELATED WORKS 25](#_Toc40797572)

[CHAPTER IV 27](#_Toc40797573)

[4 DIMENSIONALITY REDUCTION WITH ROBUSTLY WEIGHTED FUZZY C-MEANS 27](#_Toc40797574)

[4.1 Robustly Weighted Fuzzy C-Means Algorithm 27](#_Toc40797575)

[4.2 Datasets and Descriptions 31](#_Toc40797576)

[4.3 Experimental Results 38](#_Toc40797577)

[4.3.1 Pattern Datasets 39](#_Toc40797578)

[4.3.2 Noisy Images 42](#_Toc40797579)

[4.3.3 Hyperspectral Images 43](#_Toc40797580)

[4.4 Discussion: Performance, Advantages, and Disadvantages 46](#_Toc40797581)

[CHAPTER V 48](#_Toc40797582)

[5 CONCLUSION AND FUTURE WORK 48](#_Toc40797583)

[REFERENCES 50](#_Toc40797584)

# LIST OF TABLES

[Table 4‑1 Initial and final weights after feature learning with PCA reduced dataset 30](#_Toc40797633)

[Table 4‑2 Iris dataset description 34](#_Toc40797634)

[Table 4‑3 Wine dataset description 35](#_Toc40797635)

[The hyperspectral datasets used are the Salinas-A and a cropped patch from the Pavia University (PaviaU) dataset [28]. The ground truth image for the Salinas-A dataset can be seen in Fig 4-4. The visualization of the Salinas-A dataset is shown in Fig 4‑5 and the dataset description is described in Table 4‑4. 35](#_Toc40797636)

[Table 4‑5 Salinas-A dataset description 36](#_Toc40797637)

[Table 4‑6 Comparison of different feature weighted clustering algorithms. 40](#_Toc40797638)

[Table 4‑7 50 Experimental Results for Iris Dataset 42](#_Toc40797639)

[Table 4‑8 50 Experimental Results for the Wine Dataset with PCA reduced to 4 dimensions 43](#_Toc40797640)

[Table 4‑9 Results of 10 Experiments on Salinas-A dataset 45](#_Toc40797641)

# LIST OF FIGURES

[Figure 1‑1 Differences between various types of imaging such as RGB color images, MSI, and HSI [6]. 14](#_Toc40797616)

[Figure 1‑2 KNN example [11] 15](#_Toc40797617)

[Upon convergence, by minimizing equation 1, ideally each cluster is densely compact and well separated from each other as shown in Fig 2‑1 below: 20](#_Toc40797618)

[Figure 2‑2 K-Means clustering example [14] 20](#_Toc40797619)

[Figure 2‑3 Feature weights example 22](#_Toc40797620)

[Figure 4‑1 PCA 15 extracted bands from Salinas-A dataset 30](#_Toc40797621)

[Figure 4‑2 Visualization of iris dataset with t-SNE 33](#_Toc40797622)

[Figure 4‑3 Visualization of wine dataset with t-SNE 34](#_Toc40797623)

[Figure 4‑4 Ground truth image of Salinas-A hyperspectral image 35](#_Toc40797624)

[Figure 4‑5 Visualization of Salinas A dataset with t-SNE 36](#_Toc40797625)

[For the PaviaU dataset, a small patch of the hyperspectral image was cropped and only 8 out of the 9 classes were considered. The ground truth image for PaviaU can be see in Fig 4‑6. The visualization of the PaviaU dataset is shown in Fig 4-8 and the dataset description is described in Table 4-7. 37](#_Toc40797626)

[Figure 4‑7 Ground truth of cropped PaviaU hyperspectral image 37](#_Toc40797627)

[Figure 4‑8 Dataset description of the reduced PaviaU dataset 37](#_Toc40797628)

[Figure 4‑9 Visualization of the cropped PaviaU dataset with t-SNE 38](#_Toc40797629)

[Figure 4‑10 Image segmentation results for color mountain image with 5 classes corrupted 44](#_Toc40797630)

[Figure 4‑11 Hyperspectral image segmentation results of Salinas-A dataset 45](#_Toc40797631)

[Figure 4‑12 Hyperspectral image segmentation results of cropped PaviaU dataset 46](#_Toc40797632)

# CHAPTER I

# INTRODUCTION

## Hyperspectral imaging

Hyperspectral imaging is a type of imaging that is able to capture a greater range of the electromagnetic (EM) spectrum ranging from 380 nm to 2500 nm. Each pixel captured is a culmination of features or specific spectral bands that represents the reflectance of a particular part of the EM spectrum in order to form a hyperspectral cube. The hyperspectral cube is generated from a set of images with each spectral band representing information collected from different spectral signatures [1, 2]. Hyperspectral imaging (HSI) is very useful as it is able to collect information that traditional imaging techniques cannot acquire, such as Multispectral Imaging (MSI) and regular color images. Compared to MSI and color images, HSI is composed of hundreds to thousands of narrower spectral bands [3], whereas MSI is only capable of capturing 3 to 10 wider spectral bands and 3 bands for color images [4] as shown in Fig 1.1-1 below. Thus, HSI allows for the collection of spectral bands from a much greater spectral range spanning from the visible light spectrum extending into the ultraviolet and infrared spectrum. Although HSI has been around for some time, its applications have mostly been catered for remote sensing purposes through satellite imaging such as the Airborne/Visible / Infrared Imaging Spectrometer (AVIRIS) hyperspectral sensor [5]. As the number of applications continue increase for HSI, expanding into other areas such as medical diagnosis, agriculture, food quality

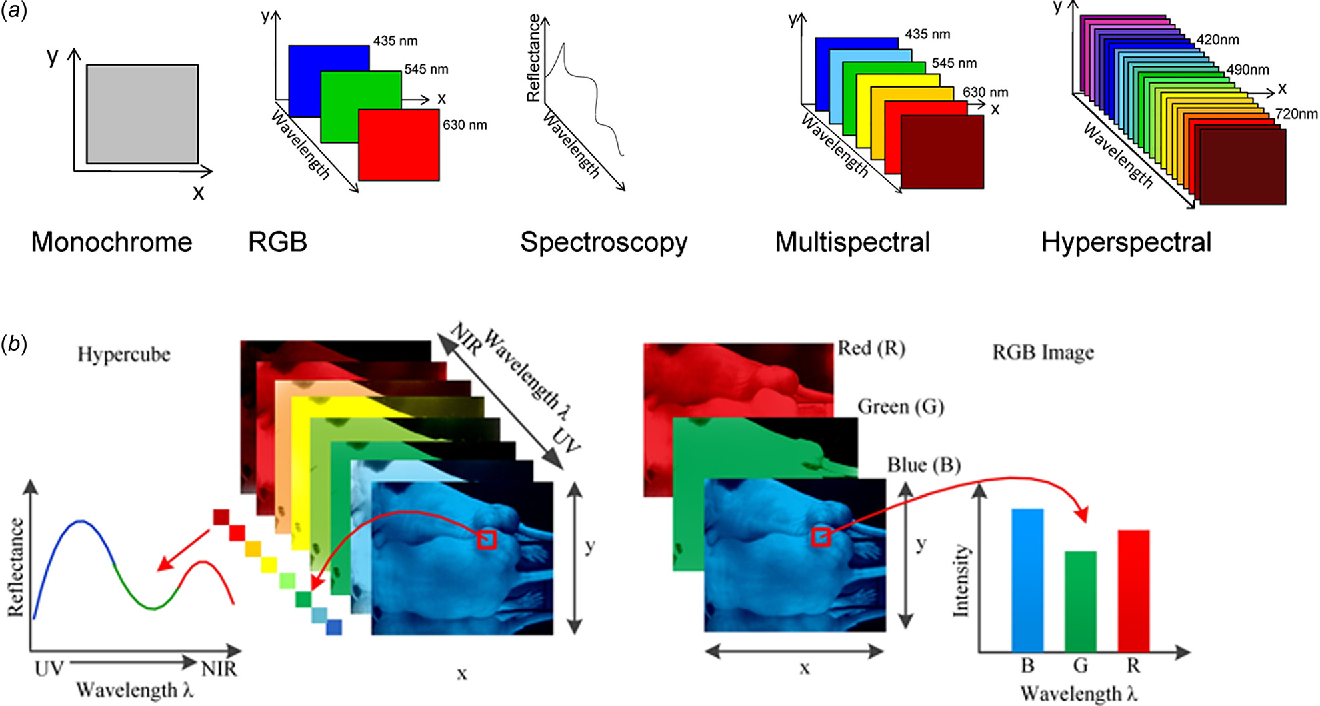


Figure 1‑1 Differences between various types of imaging such as RGB color images, MSI, and HSI [6].

verification, national defense, forensic sciences, and environmental monitoring [2], it is becoming increasingly important to be able to harness this type of data.

## Problem Statement

As discussed in section 1.1, HSI is significantly richer and more abundant in pixel information when compared to traditional imaging techniques. However, the difficulty of processing and deriving useful knowledge from such data is a challenge. Considering the high-dimensionality of HSI’s, with each spectral band representing a feature/dimension, most traditional pattern recognition and machine learning algorithms do not perform well in extracting useful knowledge from such data. Data in high-dimensional settings makes it difficult for solving issues such as classification problems and image segmentation. Traditional algorithms alone are inaccurate or ineffective due to the curse of dimensionality, or otherwise known as Hughes Phenomenon, where relevant pieces of data are masked by the innate properties of redundancy, sparseness, and noise that comes with HSI. The large complex dimensions of hyperspectral data contain empty or redundant bands making objects that are similar appear dissimilar and objects that are dissimilar appear alike. For instance, similarity metrics such as the Euclidean distance metric does not work well. Algorithms such as K-Nearest Neighbors (KNN) that rely heavily on Euclidean distance may achieve poor classification accuracy due to the curse of dimensionality. For example, in the KNN illustration shown in Fig. 1.2-1, the unknown class would normally be classified as Class B, however in a high dimensional setting it might be misclassified as Class A as the data points that are far away may appear to be the same distance as the datapoints that are much closer.

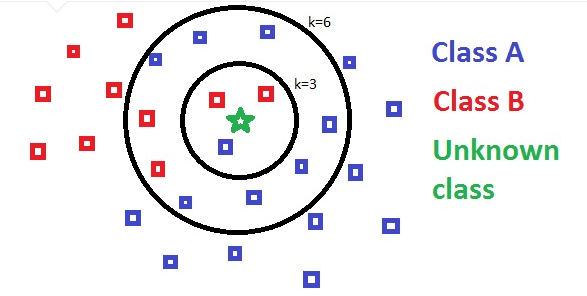


Figure 1‑2 KNN example [11]

In addition, the empty and redundant bands attributes to feature irrelevance issues when working with HSI. Hence, we would like to have a machine learning algorithm or methodology that is able to perform accurate hyperspectral image classification and segmentation. Currently, supervised learning methods such as convolutional neural networks (CNN) are able to perform significantly well with hyperspectral imaging achieving almost near perfect classification accuracy [8]. However, it is not realistic nor feasibly to use supervised learning methods such as CNN with hyperspectral imaging as these techniques requires labels and an abundant amount of training data. In the case of hyperspectral imaging, one image is captured and labeling is possible but is extremely expensive and time consuming as each pixel must be hand labeled. Another issue lies with data generation of HIS, taking a new hyperspectral image within the same location might yield a drastically different data distribution. In this case, unsupervised learning methods are desirable as labels are not required and only the single hyperspectral image is needed. Unsupervised learning methods such as clustering is able to derive useful knowledge from the hyperspectral data without the disadvantages of supervised learning approaches. Therefore, this research aims to discover a new unsupervised learning approach that works well with hyperspectral imaging.

## Contribution

In this thesis, an unsupervised learning approach is proposed which involves dimensionality reduction with a feature weighted Fuzzy C-means (FCM) named Robustly Feature Weighted Fuzzy C-means (RFWFCM). Unsupervised learning methods such as clustering algorithms are commonly used for image segmentation of regular color images, however there are still a limited number of improvements made for clustering hyperspectral images [9, 10]. With our proposed approach, accurate classification of hyperspectral imaging was attained. With the framework of RFWFCM, a hyperspectral image can be precisely segmented without the need of large amounts of data and labels. This is a more realistic approach as deriving labels for every hyperspectral image collected is not feasible and is better for the limited amount of hyperspectral data available. RFWFCM incorporates a distance metric that is proposed to work well with noisy environments, dimensionality reduction, and feature weighting to combat feature irrelevance. The experimental results will show that this unsupervised learning framework is superior to algorithms with feature weighting alone and produces surpassing hyperspectral image segmentation results. Several cluster validity indexes were also used to ensure that the images were clustered well. Overall, this framework can be applied to the applications mentioned in section 1.1 for the analysis of hyperspectral data and future applications as hyperspectral imaging becomes more commonly used.

## Thesis Organization

The remainder of the thesis is organized as follows:

* Chapter 2: Introduction to clustering methods
* Chapter 3: Discussion of related works to FCM feature weighted clustering
* Chapter 4: Dimensionality with Robustly Weighted Fuzzy C-Means
* Chapter 5: Conclusion and Future Works
* References

# CHAPTER II

# CLUSTERING METHODS

## Clustering

Clustering is an unsupervised learning method that aims to partition a dataset into K classes or groups with the goal of achieving well-defined clusters where each cluster is densely grouped with minimal overlap. K random datapoints are initially assigned as cluster centers and datapoints are iteratively assigned to each cluster through the Euclidean distance metric until the sum of distances of each data point to the nearest respective centroid is minimized. Some of the most common clustering techniques includes partitioning methods, hierarchical clustering, fuzzy clustering, density-based clustering, and model-based clustering [12]. Generally clustering is separated into two categories, hard clustering and soft clustering. Hard clustering refers to the assignment of datapoints, where one datapoint can only be assigned to one cluster at a time. On the other hand, soft clustering is a probabilistic approach where each datapoint can belong to multiple clusters at once through association with a fuzzy membership matrix and the sum of cluster probabilities (belongingness) equal to one. In this thesis, the primary focus will be on fuzzy clustering methods.

## K-Means Clustering

K-Means is one of the most traditionally used clustering algorithms as it is valued for its simplistic implementation and fast results [13]. However, the initialization of the cluster centers largely affects the results of the clustering which in consequence requires multiple runs to determine the best result. Given a set of data points with dimensions n x m, where n is the number of datapoints and m is the number of features, the vector X can be partitioned into K separate classes/clusters through minimizing the objective function the equation below:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

The membership variable [0,1]. The similarity metric Euclidean distance is defined below in equation 2:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

The equation to compute the new center vectors at each iteration is defined in equation 3:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Equation 1 can be solved through iteratively repeating the following steps until convergence.

**Step 1**: Randomly select K data points as initial cluster centers

**Step 2**: Assign datapoints to each cluster according to the minimum Euclidean distance between each datapoint and centroids. Update labels matrix.

**Step 3**: Compute the new centroids vector Z’.

**Step 4**: Check termination condition, if Z’ = Z or max iterations reached terminate loop, otherwise repeat steps 2-3.

Upon convergence, by minimizing equation 1, ideally each cluster is densely compact and well separated from each other as shown in Fig 2‑1 below:

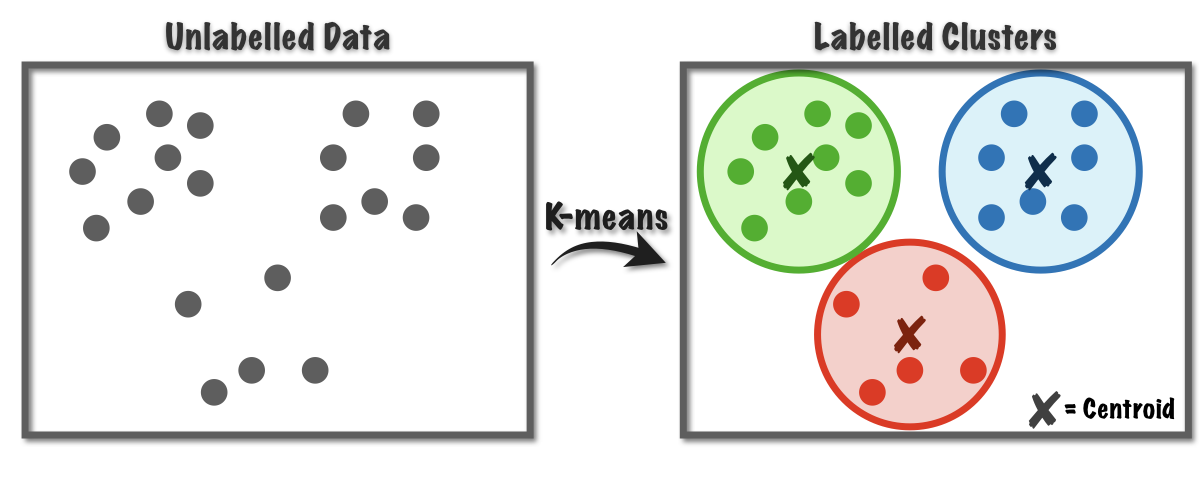


Figure 2‑2 K-Means clustering example [14]

## Fuzzy C-Means Clustering

Fuzzy C-Means (FCM) clustering is a soft clustering probabilistic approach proposed by Bezdek et al. [15] which aimed to alleviate some of the problems of K-Means while producing better clustering results. Bezdek et al., introduced a fuzzy membership matrix that is used to assign a probability of belongingness from a datapoint to each cluster center. This is known as soft clustering where each data point can belong to multiple clusters at a time, whereas with hard clustering a data point can be assigned to only one cluster at a time. In addition, a fuzzifier parameter was introduced to address issues with datasets that contains more noise or have a data distribution that is more overlapped. A fuzzifier value m that is set to 1 is equivalent to K-Means, a crisp clustering, while a value set closer to 3 or higher leads to a fuzzier clustering solution. The FCM algorithm was able to discover clusters that are better separated and more dense that the solutions found by K-Means. From K-Means, an additional membership matrix is added, the update centroids equation is modified to include the membership values, and a fuzzifier parameter is incorporated. The FCM objective function can be seen below in equation 4:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Equation 5 below is used to update the membership matrix:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Equation 6 below is used to compute the new centroids:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Equation 4 can be minimized by iteratively repeating the steps below until convergence:

**Step 1**: Randomly select K data points as initial cluster centers

**Step 2**: Random initialize the membership matrix U where

**Step 3**: Assign datapoints to each cluster according to the minimum Euclidean by updating the membership matrix U with equation 5.

**Step 4**: Compute the new centroids vector Z’ with equation 6.

**Step 5**: Check termination condition, if U’ = U or max iterations reached terminate loop, otherwise repeat steps 3-4.

## Feature Weighted Clustering

The concept of feature weighting was introduced to clustering to combat noise and feature irrelevance when working with datasets that are complex and high in dimensionality [16, 17]. Highly overlapped datasets are an issue for K-Means clustering and FCM determined to address the problems with overlapped data by introducing a clustering a method that utilized fuzzy logic. Huang et al. [16], and Wang et al. [17] found different weighting techniques for clustering and both arrived at the conclusion that some features are irrelevant and should hold little to no weight in the clustering process. Huang et al. found that learning feature weights with K-Means can help guide the search process for optimal cluster centers as clusters were found when relevant features were paired together. Wang et al. discovered the use of weights with FCM allowed for better clustering results as determined with various performance metrics to show its superiority. From these publications, it is clear that feature weighting can make clustering algorithms such as K-Means and FCM robust to noise while discerning important features from irrelevant ones. Once the weights are derived, features with smaller weights can be removed to form a subset of features that will optimally yield well-defined clusters. The process of removing features after deriving the weights can also be thought of as a feature selection method. Fig 2.4-1 shows an example of feature weight learning with the initial weights being and the derived final feature weights . From Fig 2.4-1, we can also see that band 4 has the lowest weight and can be completely removed for a subset consisting of bands 1, 2, 3, and 5.

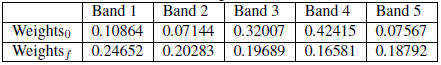


Figure 2‑3 Feature weights example

It is important that clustering was introduced first in this thesis and then followed by feature weighting as these concepts are used to form the proposed algorithm and methodology in this thesis. The feature weight learning concepts considered are from Huang et al. [16] and Wang et al. [17]. Huang et al. introduced an automatic variable feature weighting algorithm that was used with K-Means by adding a weight vector to be learned, where and . The higher the weight represents higher feature importance. The objective function of K-Means is modified as shown in equation (7) to include the weights.

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

and

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

In equation 8, h is the number of features where the distance ≠ 0. The beta value β is a hyperparameter where β > 1 and β ≤ 0. In the case that β = 1, the objective function in equation 7 becomes traditional K-Means. Beta values β ranging from [-10, 10] is the typical range for tuning the weights, these values have to be tested for each dataset and it is not clear if an optimal β value can be determined. The algorithm modifies K-Means by including a randomly initialized weight vector in the beginning steps along with an update feature weights phase to obtain .

Wang et al. developed another feature learning method that learns the optimal feature weights prior to clustering. This feature weight learning algorithm works by determining the similarity between data samples. The similarity measure used for this algorithm was chosen for its simplicity which is shown in equation 9 below:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

Ideally, data sets are uniformly distributed in [0, 1], however most real datasets do not contain these properties. Therefore, to adjust the mean of the distribution of , a positive parameter value β > 0 is desired. Considering that 0.5 is the mean of the uniform distribution, a beta value β = 0.5 is a good value such that:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

Solving equation 10 will give us the beta value β. Then we can evaluate the loss function in equation 11 by minimization with the optimization technique called gradient descent as shown in equation 12.

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

and

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

Once the weights are derived, a weighted Euclidean distance can be used to replace the Euclidean distance commonly used in clustering. Below is the modified objective function of FCM in equation 13:

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

Where is the weighted Euclidean distance as shown in equation 7.

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

Others have further expanded these feature weighting concepts, and these works will be briefly explored in the proceeding chapter.

# CHAPTER III

# RELATED WORKS

In this chapter will briefly go over extensions of FCM feature weighting algorithms that have been proposed. Additionally, we will dive into a distance metric that was proposed to replace Euclidean distance which is determined to be more robust that the Euclidean norm. First, we will discuss the distance metric of interest as it is used along with several of the proposed featured weighted algorithms. This distance metric is also used to derive the new algorithm that will be discussed in Chapter 4.

Wu et al. [18] proposed Alternative C-Means clustering algorithms that utilized a new distance metric that is found to be more robust than the Euclidean norm when used with FCM and K-Means. With the alternative distance metrics, new clustering algorithms were derived which were named alternative FCM (AFCM) and alternative K-Means (AKM). The alternative distance metric aimed to solve the disadvantages of Euclidean distance where clustering methods that used the Euclidean norm encountered issues with different cluster sizes, cluster shape, and inaccuracy due to noise [18]. In essence, the proposed alternative distance metric is more robust to noise and performs invariantly with different cluster sizes. The algorithms for the alternative distance metrics applied to K-Means and FCM can be found in [18].

The automated feature weighting algorithm for K-Means proposed by Huang et al. was extended to an FCM version by Nazari et al. [18] where it was shown to have improvements over FCM type algorithms. However, a disadvantage of learning feature weights with this method is the uncertainty of how to determine an optimal beta value β. Beta values in the range of [-10, 10] would have to be tested for each fuzziness of FCM [1.5, ) to find the best possible clustering. The optimal value β varies with each dataset and can be time consuming and tedious to identify the best clustering results. Li and Yu [19] proposed a similar feature weighted algorithm that is nearly identical to Nazari et al. that also demonstrates a performance increase in FCM type clustering algorithms. It is clear that the algorithm proposed by Li and Yu also has the same disadvantages as the one proposed by Nazari et al. as they both use the same weighting mechanism.

In another publication, the alternative distance metrics from [20] were used to expand weighted FCM algorithms due to their robustness. Xiang et al. [21] introduced an algorithm named Alternative weighted FCM (AWFCM), in which a weight was computed through the normalization of the cluster centroids and was combined with the alternative distance metric. The weighting approach by Xiang et al. combined with the alternative distance metrics showed improvements upon the algorithms AFCM and AKM proposed by Wu et al. [18].

Although many of these algorithms were proposed, there are not many clustering algorithms and methods developed mainly for working with hyperspectral imaging and high-dimensional data. Many of these feature weighting algorithms and distance metrics appears to be good solutions for tackling the properties of hyperspectral imaging. Hung et al. [22] introduced a feature weighting algorithm named a new weighed FCM (NW-FCM) that was used with a feature extraction technique called non-parametric feature extraction (NWFE) by Kuo et al. [23] to reduce the dimensions of hyperspectral data and perform clustering. It can be seen that the area of clustering with hyperspectral imaging has not been fully explored and there is room for improvement.

# CHAPTER IV

# DIMENSIONALITY REDUCTION WITH ROBUSTLY WEIGHTED FUZZY C-MEANS

## Robustly Weighted Fuzzy C-Means Algorithm

In this thesis a new dimensionality reduction with robustly weighted FCM (RWFCM) is proposed to work with hyperspectral datasets. This algorithm takes inspiration from Wu et al. [18] by considering the alternative distance metric proposed as it is robust to noise and the performance is invariant when working with various cluster sizes. In addition to using the alternative distance metric compared to the traditional Euclidean norm, feature weight learning for FCM is used to discern relevant features to address the feature irrelevance issue associated with hyperspectral data. As discussed in chapter 3, many feature learning algorithms have been developed over the years for FCM and is tested mostly on pattern datasets. However, the algorithms proposed have not been tested on more complex data such as hyperspectral data. Xiang et al. [21] has proposed a similar algorithm to the algorithm presented in this thesis, but uses a single weight through computing the normalization of the cluster centers. Comparatively, the method proposed in this thesis uses a weight vector that is learned from sampling datapoints from the entire dataset through the use of a similarity measure between datapoints. Also, this feature weight vector is only computed once prior to clustering which makes it computationally less expensive compared to [21] where the weights are computed each iteration.

When working with hyperspectral data, it is common to reduce the high-dimensionality of a hyperspectral image to a more reasonable set. Pre-processing hyperspectral data is a necessary step as there is an abundance of redundant and empty features. This can be done through common feature selection algorithms were a subset of features with the most information is selected from the original data to form a new subset. Another technique is referred to as dimensionality reduction where the dataset is transformed from a high-dimensional space into a new lower dimensional feature space. As part of our clustering method, principal component analysis (PCA) is used for feature extraction of hyperspectral images to transform the hyperspectral images and reduce the dimensions. Although there are many different feature extraction and decomposition methods, PCA was used for dimensionality reduction as it is one of the most powerful and commonly used for remote sensing data. Srivatsa et al. [24] also showed that the accuracy differences for clustering after preprocessing with various feature extraction methods were negligible.

For hyperspectral image segmentation the proposed dimensionality reduction with RWFCM first reduced with PCA for feature extraction. It was discovered that even after pre-processing with PCA, noise still exists within the reduced subset. Fig 4-1 below shows fifteen extracted bands by using PCA. Naturally PCA organizes the extracted features in a descending fashion with the eigen vectors that hold the most weight located in the first feature vector. However, the most relevant features are not used to guide the search process for well-defined cluster centers and noise may still be present. For example, Table 4-1 shows the results of feature learning after processing with PCA. The weights for the first five bands were shown and were ordered in a descending behavior, however some higher principal component features were not as informative as some lower order features.

Table 4‑1 Initial and final weights after feature learning with PCA reduced dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 |
| **Initial Weights** | 0.10864 | 0.07144 | 0.32007 | 0.42415 | 0.07567 |
| **Final Weights** | 0.94652 | 0.89283 | 0.80689 | **0.56581** | 0.78792 |

It can be seen that in band 4 from Table 1 and the corresponding image in Fig. 4-1, even though PCA orders each extracted feature by the most relevant, is more informative than band 5 from Table 1 and Fig 4-1. It can also be observed that band 12 in Fig 4-1 has a lot less noise and more information when compared to band 11 in Fig 4-1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A picture containing knife  Description automatically generated | A close up of a piano  Description automatically generated | A picture containing saw, circuit, knife, drawing  Description automatically generated | A picture containing knife  Description automatically generated | A close up of a logo  Description automatically generated |
| Principal Component 1 | Principal Component 2 | Principal Component 3 | Principal Component 4 | Principal Component 5 |
| A close up of a logo  Description automatically generated | A close up of a logo  Description automatically generated | A picture containing drawing  Description automatically generated | A group of people  Description automatically generated | A group of people  Description automatically generated |
| Principal Component 6 | Principal Component 7 | Principal Component 8 | Principal Component 9 | Principal Component 10 |
| A group of people  Description automatically generated | A close up of a logo  Description automatically generated | A close up of a persons face  Description automatically generated | A group of people  Description automatically generated | A crowd of people  Description automatically generated |
| Principal Component 11 | Principal Component 12 | Principal Component 13 | Principal Component 14 | Principal Component 15 |

Figure 4‑1 PCA 15 extracted bands from Salinas-A dataset

With the dimensionality reduction RWFCM method, the aim is to find an algorithm that is robust to noise while using feature learning to combat feature irrelevance to achieve good performance when working with hyperspectral imaging. The new objective function for RWFCM can be seen in equation 14 below:

|  |  |  |
| --- | --- | --- |
|  |  | (14) |
|  |  |  |
|  |  |  |

The feature weights are learned by using the method proposed by Wang et al. [17] and the alternative distance metric from Wu et al. [18]. The new update equation for the membership function can be found in equation 15.

|  |  |  |
| --- | --- | --- |
|  |  | (15) |

Below the update equation for updating the center vectors is shown equation 16.

|  |  |  |
| --- | --- | --- |
|  |  | (16) |

The new objective function in equation 14 can be minimized by solving U and Z with equations 15 and 16. The steps for the RWFCM algorithm are shown below:

**Preprocessing:** Use Principal Component Analysis for dimensionality reduction and feature extraction

**Step 1:** Find the optimal Beta value and derive the Weights (**w**)

**Step 2:** Used **w** in new distance metric to update the membership matrix (**U**)

**Step 3:** Compute new centroids matrix (**Z**) with new distance metric

**Step 4:** Check for convergence by termination condition

**Step 4:** Repeat steps 2-3 until Max iterations have been reached or Termination conditions are met

**Termination condition:** if (U’-U < ) and (Z’-Z < ) then Stop

Experimental results will show that the new methodology works extremely well with feature weighting and in some cases the RWFCM only performs well but only marginally.

## Datasets and Descriptions

For the experimental results, various types of datasets were used to compare the performance of the proposed approach and different weighting algorithms. This section will describe each dataset in detail along with visualization of each dataset. For visualization, the well-known t-distributed Stochastic Neighbor Embedding (t-SNE) is used. Each dataset is reduced to two components and then visualized in a graph. Currently, t-SNE is one of the most popular technique for visualizing high-dimensional data. The algorithm t-SNE is able to represent high-dimensional data accurately in a lower dimensional space by maintaining points that are close in high dimension in a lower dimensional space as well. T-SNE is able to represent data in a non-linear and local approach [25].

For pattern datasets the iris pattern dataset and the wine dataset were used [26, 27]. Both datasets are standard benchmark datasets commonly used in machine learning. Fig 4-2 below shows the visualization of the iris dataset and Table 4-2 shows the dataset description.

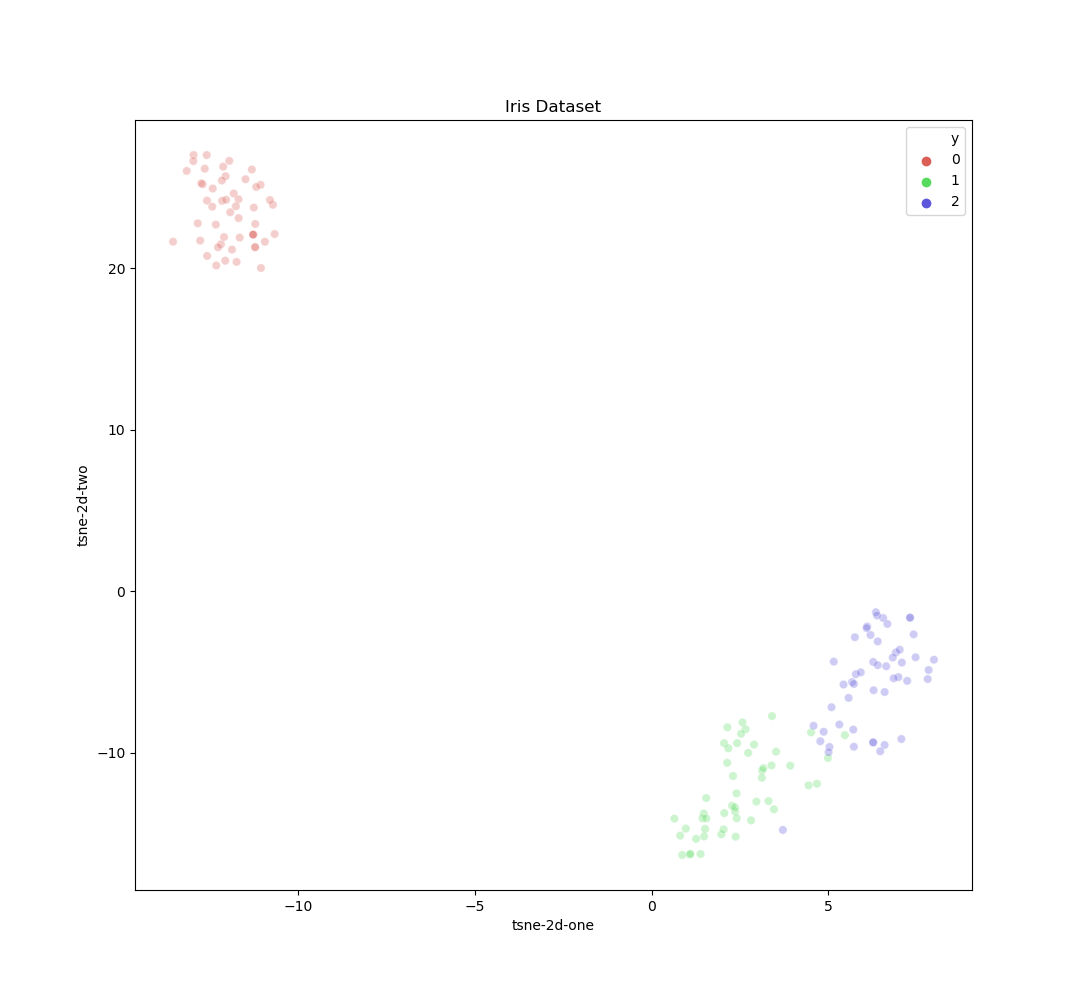


Figure 4‑2 Visualization of iris dataset with t-SNE

Table 4‑2 Iris dataset description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset Name | Features (4) | No. Classes | Data Type | No. Datapoints |
| Iris | Sepal Length (cm) | 3 | Real | 150 |
| Sepal Width (cm) |
| Petal Length (cm) |
| Petal Width (cm) |

The visualization of the wine dataset can be seen in Fig 4-3 and the dataset description can be seen in Table 4-3.

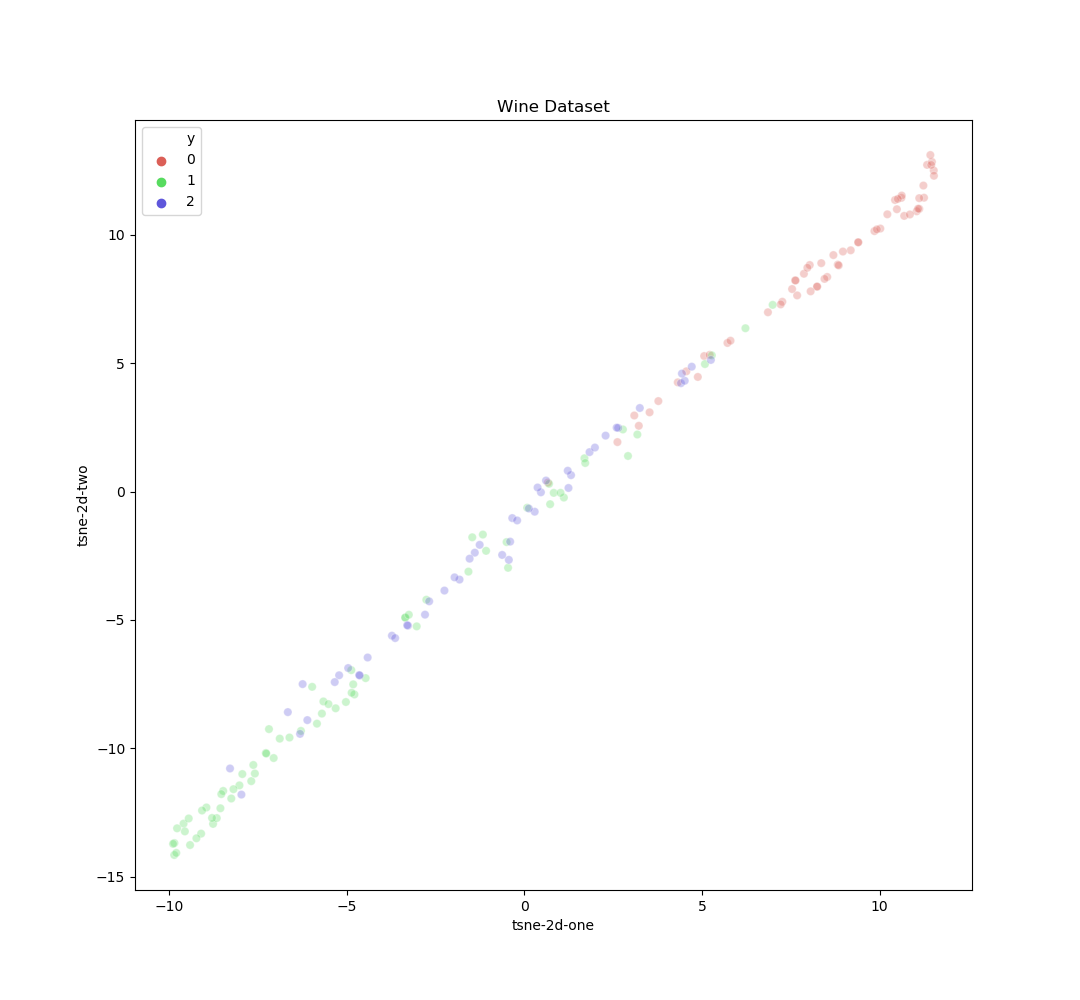


Figure 4‑3 Visualization of wine dataset with t-SNE

Table 4‑3 Wine dataset description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset Name | Features (13) | No. Classes | Data Type | No. Datapoints |
| Wine | Alcohol | 3 | Real | 178 |
| Malic acid |
| Ash |
| Alcalinity of ash |
| Magnesium |
| Total phenols |
| Flavanoids |
| Nonflavanoid phenols |
| Proanthocyanins |
| Color intensity |
| Hue |
| OD280/OD315 of diluted wines |
| Proline |

The hyperspectral datasets used are the Salinas-A and a cropped patch from the Pavia University (PaviaU) dataset [28]. The ground truth image for the Salinas-A dataset can be seen in Fig 4-4. The visualization of the Salinas-A dataset is shown in Fig 4‑5 and the dataset description is described in Table 4‑4.

A picture containing drawing

Description automatically generated

Figure 4‑4 Ground truth image of Salinas-A hyperspectral image

Table 4‑5 Salinas-A dataset description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset Name | Classes | No. Bands | Data Type | No. Datapoints |
| Salinas-A | Broccoli green weeds 1 | 204 (Corrected) | Pixel (Real) | 86\*83 |
| Corn senesced green weeds |
| Lettuce romaine 4wk |
| Lettuce romaine 5wk |
| Lettuce romaine 6wk |
| Lettuce romaine 7wk |

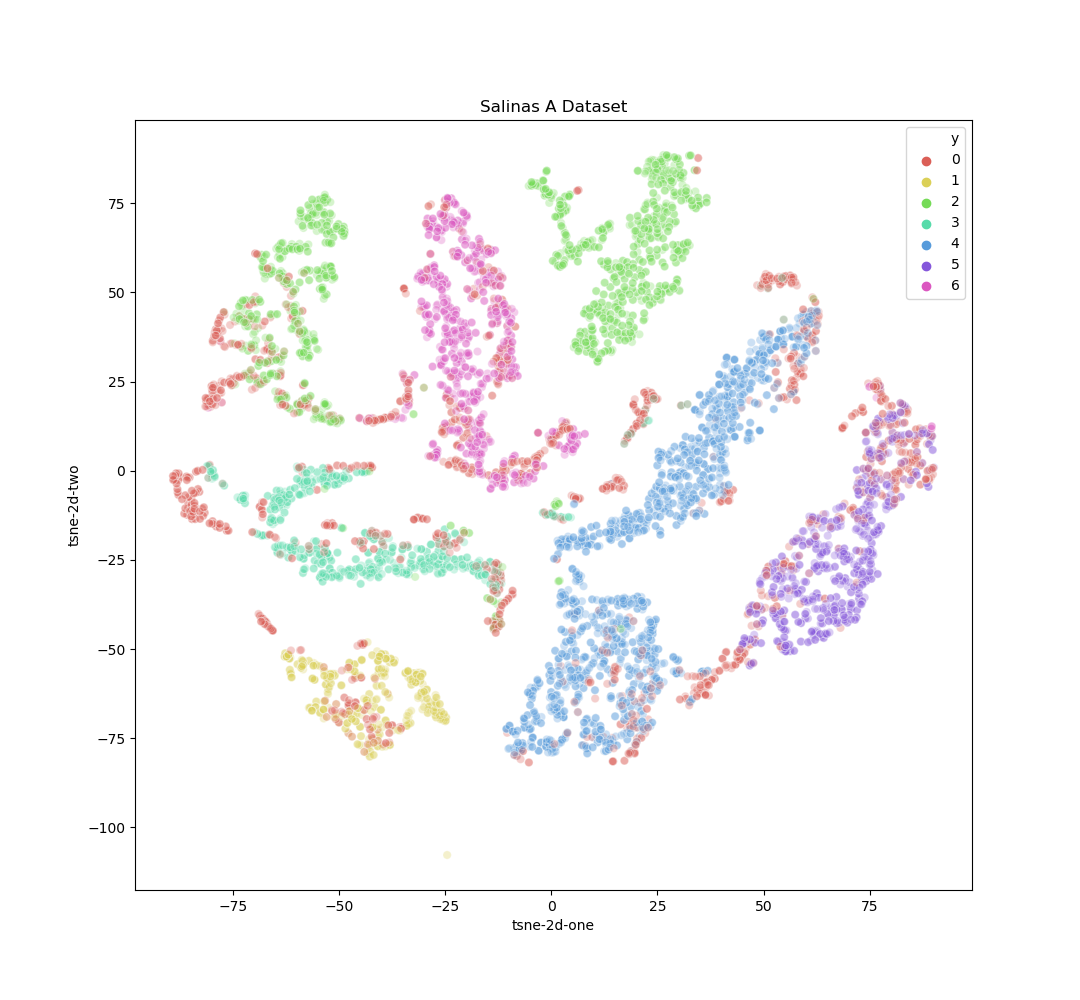


Figure 4‑5 Visualization of Salinas A dataset with t-SNE

For the PaviaU dataset, a small patch of the hyperspectral image was cropped and only 8 out of the 9 classes were considered. The ground truth image for PaviaU can be see in Fig 4‑6. The visualization of the PaviaU dataset is shown in Fig 4-8 and the dataset description is described in Table 4-7.

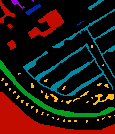


Figure 4‑7 Ground truth of cropped PaviaU hyperspectral image

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset Name | Classes | No. Bands | Data Type | No. Datapoints |
| PaviaU | Asphalt | 200 (Corrected) | Pixel (Real) | 115\*134 |
| Meadows |
| Gravel |
| Trees |
| Painted metal sheets |
| Bare Soil |
| Bitumen |
| Self-Blocking Bricks |
| Shadows |

Figure 4‑8 Dataset description of the reduced PaviaU dataset

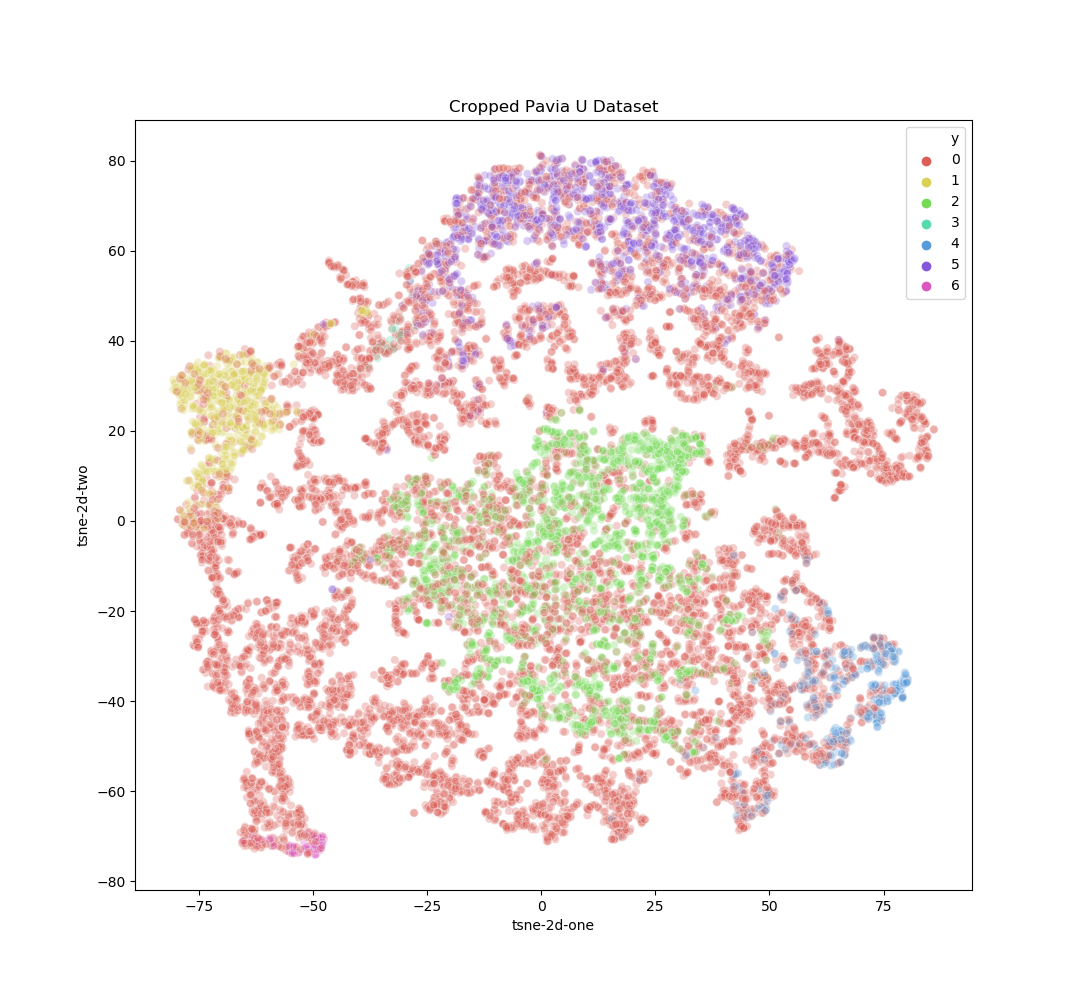


Figure 4‑9 Visualization of the cropped PaviaU dataset with t-SNE

## Experimental Results

The weighted FCM algorithms were all tested and evaluated with several performance metrics. Those performance metrics included min accuracy, max accuracy, average accuracy, Xie-Beni index (xb), Davies-Bouldin index (db), Dunn index, and the fuzzy partition coefficient. For the hyperspectral images an additional evaluation method, commonly used for remote sensing data, called the kappa coefficient was used.

In order to determine well-defined clusters or identify high-quality clusters, two criteria must be met. Clusters must be as compact as possible with little variance within each cluster while maintaining minimal cluster overlap, meaning well separated. There are several cluster validity index algorithms used to determine the quality of the clustering achieved by a clustering algorithm. A popular cluster validity index is the Xie-Beni index which measures the ratio the inter-cluster separation and the within-cluster scatter to determine the quality of the clustering. A smaller xb value denotes better cluster quality [29]. The Davies-Bouldin index evaluates the optimal number of clusters by measuring the scatter within each cluster and the separation between the cluster centers. A smaller db value can determine the best number of classes, however it has the disadvantage that a good db value does not mean good quality clusters were found [29]. The Dunn index is an internal evaluation scheme were the results are based on the clustered data, it is also used to determine the compactness of each cluster as well as measuring the separation between each [29]. The fuzzy partition coefficient (FPC) is also another cluster evaluation in which it determines how well the clustering was performed [30]. A value closer towards one denotes that the dataset was well clustered.

For hyperspectral imaging, the kappa coefficient was used in addition to the other cluster evaluation metrics. The kappa coefficient is commonly used in remote sensing as it is a statistical measure that measures the level inter-rater reliability or agreement between two raters. It may often give better insight to the performance of an algorithm in addition to the overall accuracy.

### Pattern Datasets

For the pattern datasets, each were ran on 50 separate experiments and the various performance metrics mentioned in section 4.3 were saved and averaged. When comparing the RWFCM algorithm to other feature weighting algorithms, it was found that it performed the best with the iris dataset except when it is compared to the weighted FCM (WFCM) alone. Even though the accuracies are equivalent, the RWFCM achieved a higher FPC than the WFCM which means a higher clustering quality was achieved. Table 4-6 below shows the performance comparisons between each of the feature weighted algorithms and algorithms that used the alternative distance metrics. Only the iris dataset was used for comparison between other proposed FCM extensions as it was the most common dataset used across all algorithms for evaluation.

Table 4‑6 Comparison of different feature weighted clustering algorithms.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Dataset** | **Average Accuracy** | **Average FPC** |
| K-Means [20] | Iris | 0.89333 | - |
| FCM [20] | 0.89333 | - |
| Alternative K-means (AKM) [20] | 0.89333 | - |
| Alternative Fuzzy C-Means (AFCM) [20] | 0.9267 | - |
| Robust Local Feature Weighted K-Means (RLFWKM) – Beta [32] | 0.92 | - |
| Attribute Weighted Fuzzy C-Means (AWFCM) – Beta [18] | 0.96 | - |
| Alternative Weighted Fuzzy C-Means (AWFCM) – single weight [21] | 0.94 | - |
| Feature Weight Learning-FCM (WFCM) – vector weight [17] |  | 0.94667 | 0.86 |
| Robustly Weighted Fuzzy C-Means (RWFCM) |  | **0.94667** | **0.94116** |

From Table 4-6, the attributed weighted FCM was able to achieve the highest average accuracy, however the large number of beta parameters that needs to be tested at each fuzziness makes the algorithm unfeasible and unrealistic in real life applications. The weighting algorithm used by WFCM and RWFCM can derive optimal feature weights without having to test beta values between [-10, 10].

A subset of the most relevant fuzzy feature weighting techniques was chosen for a more thorough comparison with RWFCM. Table 4-7 below shows the results averaged from 50 experimental results. For the iris dataset, the WFCM performed better than the RWFCM in terms of accuracy and clustering results, however this is a smaller and simpler dataset. In addition, dimensionality reduction was used other experiments as proposed in the beginning of chapter 4, however that method did not yield good results. Fig 4-2 shows the visualization the iris dataset, and it is evident that the dataset is already well separated. Pre-processing with PCA may interfere with and produce a data distribution that is not as optimal.

Table 4‑7 50 Experimental Results for Iris Dataset

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Fuzzifier m** | **Std Dev** | **Max**  **Acc** | **Avg**  **Acc** | **Xb** | **Db** | **Dunn** | **FPC** |
| Fuzzy C-Means (FCM) | 1.5 | 0 | 0.89333 | 0.89333 | 0.15469 | 0.72188 | 0.00901 | 0.97084 |
|  | 2.0 | 0.04441 | 0.66666 | 0.89333 | 0.17649 | 0.72865 | 0.01244 | 0.92081 |
|  | 2.5 | 0 | 0.89333 | 0.89333 | 0.10471 | 0.70426 | 0.01677 | 0.85699 |
|  | 3.0 | 0 | 0.89333 | 0.89333 | 0.07773 | 0.70183 | 0.0208 | 0.78813 |
|  |  |  |  |  |  |  |  |  |
| Alternative Fuzzy C-Means (AFCM) | 1.5 | 0.85803 | 0.89333 | 0.84813 | 0.41372 | 0.96967 | 0.0126 | 0.83723 |
|  | 2.0 | 0.08815 | 0.90667 | 0.86559 | 0.12791 | 0.89953 | 0.02344 | 0.61773 |
|  | 2.5 | 0.10625 | 0.92 | 0.84453 | 0.03871 | 1.11759 | 0.02443 | 0.37495 |
|  | 3.0 | 0.10495 | 0.92667 | 0.868 | 0.00963 | 0.94155 | 0.02953 | 0.2346 |
|  |  |  |  |  |  |  |  |  |
| Weighted Fuzzy C- Means (WFCM) | 1.5 | 1.11\*10^-16 | 0.94667 | 0.94667 | 0.17984 | 0.77781 | 0.01993 | 0.97076 |
|  | 2.0 | 1.11\*10^-16 | 0.93333 | 0.93333 | 0.14961 | 0.74899 | 0.01858 | 0.93556 |
|  | 2.5 | 1.11\*10^-16 | 0.93333 | 0.93333 | 0.12596 | 0.73918 | 0.01813 | 0.90235 |
|  | 3.0 | 1.11\*10^-16 | 0.93333 | 0.93333 | 0.10589 | 0.73757 | 0.01814 | 0.86403 |
|  |  |  |  |  |  |  |  |  |
| Robustly Weighted Fuzzy C-Means (RWFCM) | 1.5 | 0 | 0.94667 | 0.94667 | 0.19396 | 0.82765 | 0.01894 | 0.94116 |
|  | 2.0 | 0 | 0.94667 | 0.94667 | 0.13364 | 0.82282 | 0.01853 | 0.83276 |
|  | 2.5 | 0.11184 | 0.94667 | 0.86933 | 0.24277 | 1.22283 | 0.02128 | 0.63889 |
|  | 3.0 | 0.08 | 0.93333 | 0.90667 | 0.84186 | 0.97412 | 0.01524 | 0.56705 |

The results for the wine dataset with 50 experimental runs are shown in Table 4-8 below. The dataset was reduced from 13 to 4 dimensions. The dataset was tested at different dimensions reduced with PCA for each algorithm and four features yielded the best results. RWFCM was able to achieve a slightly better max accuracy when compared with dimensionality with FWCM, but dimensionality reduction with FWCM achieved a higher average accuracy. In terms of cluster quality, RWFCM displayed better clustering results when examining with xb and the Dunn index.

Table 4‑8 50 Experimental Results for the Wine Dataset with PCA reduced to 4 dimensions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Fuzzifier m** | **Std Dev** | **Max**  **Acc** | **Avg**  **Acc** | **Xb** | **Dunn** | **FPC** |
| Fuzzy C-Means (FCM) | 1.5 | 0 | 0.8764 | 0.8764 | 0.526 | 0.00229 | 0.87153 |
|  | 2.0 | 0 | 0.88202 | 0.88202 | 0.34831 | 0.01813 | 0.73549 |
|  | 2.5 | 0 | 0.89325 | 0.89325 | 0.21685 | 0.00399 | 0.61733 |
|  | 3.0 | 0 | 0.89887 | 0.89887 | 0.12973 | 0.00359 | 0.5345 |
|  |  |  |  |  |  |  |  |
| Alternative Fuzzy C-Means (AFCM) | 1.5 | 0.095505 | 0.89325 | 0.79775 | 0.45253 | 0.02166 | 0.4873 |
|  | 2.0 | 0.038691 | 0.89325 | 0.74011 | 0.1186 | 0.01029 | 0.22157 |
|  | 2.5 | 1.11\*10^-16 | 0.7191 | 0.7191 | 0.01743 | 0.0083 | 0.10414 |
|  | 3.0 | 0.092026 | 0.89325 | 0.78943 | 0.00233 | 0.0196 | 0 |
|  |  |  |  |  |  |  |  |
| Weighted Fuzzy C- Means (WFCM) | 1.5 | 0 | 0.89887 | 0.89887 | 0.52151 | 0.01236 | 0.87226 |
|  | 2.0 | 0 | 0.89887 | 0.89887 | 0.35065 | 0.00767 | 0.73747 |
|  | 2.5 | 0 | 0.89325 | 0.89325 | 0.21905 | 0.01245 | 0.61996 |
|  | 3.0 | 0 | 0.88764 | 0.88764 | 0.13142 | 0.00297 | 0.53717 |
|  |  |  |  |  |  |  |  |
| Robustly Weighted Fuzzy C-Means (RWFCM) | 1.5 | 0.09202 | 0.89325 | 0.78943 | 0.39693 | 0.0196 | 0.5098 |
|  | 2.0 | 1.11 \* 10^-16 | 0.89887 | 0.7337 | 0.11191 | 0.00991 | 0.23238 |
|  | 2.5 | 0.02595 | 0.90449 | 0.7228 | 0.01817 | 0.02069 | 0.10785 |
|  | 3.0 | 1.11 \* 10^-16 | 0.71348 | 0.713483 | 0.00253 | 0.0196 | 0.5098 |

### Noisy Images

As the alternative distance function was proposed to be a robust distance metric for noise and cluster size variance, experiments were also conducted on noisy images. Fig 4-10 below shows a regular color image with five separate classes and corrupted with noise to compare the performance of RWFCM compared to other FCM methods.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Ground Truth  Mountain Image | Mountain image with 50% noise added | **RFWCM**  Xb Index – 6.87363 | **WFCM**  Xb Index – 1.12329 |

Figure 4‑10 Image segmentation results for color mountain image with 5 classes corrupted

After segmenting the color image with fifty percent noise added to the color mountain image, WFCM achieved better quality clusters when compared to WFCM with using the alternative distance metric. The xb index shows better cluster compactness and separation as WFCM achieved a much lower value of 1.12329 compared to RFWCM.

### Hyperspectral Images

The primary goal of this thesis was to find an unsupervised learning method that was able to classify and segment hyperspectral images well. For the experiments, the Salinas-A dataset and a cropped patch from the PaviaU dataset was used. The kappa coefficient and average accuracy was only obtained for the Salinas-A dataset as there was some issues with accurate class correction post clustering with the cropped PaviaU dataset. Majority voting for class rectification did not work well with the PaviaU dataset, it is speculated that this is due to a higher overlap between data samples of different classes. With PCA and WFCM Table 4-9 shows that it was able to achieve superior results when compared the PCA with RWFCM. PCA and WFCM was able to achieve an overall average accuracy of 92.10%, a max accuracy of 97.99%, an average kappa coefficient of 89.06%, and a max kappa coefficient of 97.28%. The hyperspectral image segmentation results can be found in Fig 4-11.

Table 4‑9 Results of 10 Experiments on Salinas-A dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Fuzzifier m** | **Std Dev** | **Max Acc** | **Avg Acc** | **Dunn** | **Kappa max** | **Kappa avg** |
| Fuzzy C-Means (FCM) | 1.5 | 0.05417 | 0.97157 | 0.90345 | 0.00253 | 0.96154 | 0.8675 |
|  | 2.0 | 0.30467 | 0.8835 | 0.85607 | 0.00168 | 0.84274 | 0.80498 |
|  | 2.5 | 0.03437 | 0.7399 | 0.69753 | 0.00159 | 0.64625 | 0.58056 |
|  | 3.0 | 0.02429 | 0.71054 | 0.66828 | 0.00119 | 0.5924 | 0.52809 |
|  |  |  |  |  |  |  |  |
| Alternative Fuzzy C-Means (AFCM) | 1.5 | 0.14899 | 0.65183 | 0.43552 | 0.00252 | 0.52951 | 0.17195 |
|  | 2.0 | 0.0268 | 0.75351 | 0.66025 | 0.00203 | 0.65502 | 0.55567 |
|  | 2.5 | 0.02882 | 0.73373 | 0.69016 | 0.00214 | 0.65614 | 0.58596 |
|  | 3.0 | 0.11007 | 0.79936 | 0.66288 | 0.00189 | 0.73146 | 0.55231 |
|  |  |  |  |  |  |  |  |
| Weighted Fuzzy C- Means (WFCM) | 1.5 | 0.0587 | 0.97943 | 0.92101 | 0.00591 | 0.97214 | 0.89096 |
|  | 2.0 | 0.048669 | 0.97999 | 0.91138 | 0.0018 | 0.97289 | 0.87823 |
|  | 2.5 | 0.07525 | 0.97774 | 0.90308 | 0.00214 | 0.96987 | 0.86616 |
|  | 3.0 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Robust Fuzzy Weighted C-Means (RFWCM) | 1.5 | 0.06308 | 0.90613 | 0.81729 | 0.00107 | 0.87162 | 0.75188 |
|  | 2.0 | 0.06012 | 0.90501 | 0.86174 | 0.00246 | 0.87006 | 0.8127 |
|  | 2.5 | 0.08425 | 0.89622 | 0.81828 | 0.00138 | 0.8586 | 0.752 |
|  | 3.0 | 0.06909 | 0.90332 | 0.85609 | 0.00167 | 0.86765 | 0.80401 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A picture containing light, drawing  Description automatically generated |  |  |  |  |
| **Ground Truth** | **PCA+FCM**  Overall Acc - 0.97157  Kappa - 0.96154 | **PCA+AFCM**  Overall Acc - 0.79936  Kappa - 0.73146 | **PCA+WFCM**  Overall Acc - 0.97999  Kappa - 0.97289 | **PCA+RFWFCM**  Overall Acc – 0.90613  Kappa - 0.87162 |

Figure 4‑11 Hyperspectral image segmentation results of Salinas-A dataset

For the PaviaU dataset, a small cropped patch was taken from the original PaviaU dataset for experimentation. The reason for cropping the image is due to the large size of the image and to reduce the computation time of the algorithms. Fig 4-12 below shows the image segmentation results along with the Dunn index that was used to measure the clustering quality of each algorithm.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Ground Truth** | **PCA+ FCM**  Dunn Index – 0.00184 | **PCA+ AFCM**  Dunn Index – 0.00186 |
|  |  |  |
| **PCA+ WFCM**  Dunn Index – 0.00319 | **PCA+ RFWCM**  Dunn Index – 0.00375 |  |

Figure 4‑12 Hyperspectral image segmentation results of cropped PaviaU dataset

For the Dunn index, a higher value represents better cluster quality. From Fig 4-12, PCA and RWFCM achieved the highest Dunn index, hence it acquired the best clustering result. However, PCA with WFCM performed almost as well as PCA with RWFCM.

## Discussion: Performance, Advantages, and Disadvantages

The proposed method of feature weighting post dimensionality reduction yielded great results. From the experimental results shown in section 4.1-4.3, better clustering results were achieved and the cluster validity indices indicated that well-defined clusters were found. The performance metrics xb, db, Dunn, and FPC verified that cluster centers for dimensionality reduction with feature weighting obtained clusters that had a lower intra-cluster variance and higher inter-cluster separation. Feature weighting with [17] works well but is computationally expensive with a time complexity of where c is associated to the number of features and n is the number of samples within the dataset. Hence, the run time of the algorithm scales with the number of features and the number of samples. In order to mitigate the computational expense, sampling from the dataset to learn the feature weights was used to significantly improve the feature weight learning time.

The RWFCM with PCA and WFCM with PCA works fairly well with smaller dataset sizes and lower number of classes. From section 4.1-4.3, hyperspectral image segmentation with the proposed methodology works extremely well and produced well segmented images. The algorithm is also more robust to noise when compared to traditional algorithms as Fig 4-1 shows that noise may still be present after pre-processing with PCA. Pre-processing with PCA removes some noise, however noise may still be present in the reduced subset. In addition, although the most important features are extracted with PCA, the informative features do not directly influence the clustering process. Adding feature weights helped guide the clustering process by giving weights to the features that have the most importance in the search process for optimal cluster centers.

As discussed earlier in this section, the time complexity is an issue as the larger the hyperspectral image and larger number of features significantly increases the computation time. RWFCM with PCA and WFCM with PCA works well with smaller hyperspectral images, but does not perform as well when the number of classes and image resolution is increased.

# CHAPTER V

# CONCLUSION AND FUTURE WORK

In this thesis an unsupervised learning method was introduced along with a new proposed algorithm named RWFCM. The goal was to discover an unsupervised learning method that works well with hyperspectral image classification. As clustering have been explored, there have been limited advancements with fuzzy clustering hyperspectral imaging. Supervised learning techniques are able to achieve tremendously high accuracy; however, it is impractical to obtain labels for all hyperspectral images acquired and generating a large dataset is infeasible. Thus, unsupervised learning methods are desirable as they do not require labels for learning and inference.

Dimensionality reduction with RWFCM and WFCM was proposed to work with hyperspectral imaging. The Fig 4-1 demonstrates that even though PCA removes noise from a dataset, noise can still be present. In this case, feature weighting and the alternative distance metric was considered to combat the issue of noise and feature irrelevance upon pre-preprocessing of a dataset. The experimental results show that feature weighting following dimensionality reduction significantly improves hyperspectral image classification performance. The extracted features found with feature weighting helped guide the search process for well-defined cluster centers. Although the addition of the alternative distance metric along with feature weighting did not show a significant improvement, it was found that dimensionality reduction with feature weighting techniques works well.

The future direction of this work includes further exploration of these clustering methods with hyperspectral data. Currently, this algorithm works well with hyperspectral images with smaller resolutions. Ideally, we would like to develop an unsupervised algorithm that is able to work well with large resolution hyperspectral images with a high number of classes.

# REFERENCES

[1] Hyperspectral Imaging. (n.d.). Retrieved from <https://www.sciencedirect.com/topics/medicine-and-dentistry/hyperspectral-imaging>

[2] Imaging, T. (n.d.). Applications. Retrieved from <https://www.techimaging.com/applications/hyperspectral-imaging>

[3] Prasad S Thenkabail, Murali Krishna Gumma, Pardhasaradhi Teluguntla, and Irshad A Mohammed, “Hyperspectral remote sensing of vegetation and agricultural crops,” Photogrammetric Engineering & Remote Sensing (PE&RS), vol. 80, no. 8, pp. 697–723, 2014.

[4] Mota, E. C. O., Maple, D., Cliff, Chris, Mahmoodullah, K.T., V., … Clarence, K. (2020, March 6). Multispectral vs Hyperspectral Imagery Explained. Retrieved from <https://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/>

[5] AVIRIS - Airborne Visible / Infrared Imaging Spectrometer. (n.d.). Retrieved from https://aviris.jpl.nasa.gov/

[6] Mehta, N., Shaik, S., Devireddy, R., & Gartia, M. R. (2018). Single-cell analysis using hyperspectral imaging modalities. Journal of biomechanical engineering, 140(2).

[7] Anand, Anand, Anand, & Anand. (2018, September 22).  k-Nearest Neighbour  – KNN. Retrieved from <https://bigishere.wordpress.com/2018/09/22/k-nearest-neighbour-knn/>

[8] Ghamisi, P., Chen, Y., & Zhu, X. X. (2016). A self-improving convolution neural network for the classification of hyperspectral data. IEEE Geoscience and Remote Sensing Letters, 13(10), 1537-1541.

[9] Zhang, H., Zhai, H., Zhang, L., & Li, P. (2016). Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 54(6), 3672-3684.

[10] Wang, Q., Zhang, F., & Li, X. (2018). Optimal clustering framework for hyperspectral band selection. IEEE Transactions on Geoscience and Remote Sensing, 56(10), 5910-5922.

[11] Sanjay.M. (2018, November 2). KNN using scikit-learn. Retrieved from https://towardsdatascience.com/knn-using-scikit-learn-c6bed765be75

[12] Ahmed, J. (2019, December 25). 5 Amazing Types of Clustering Methods You Should Know. Retrieved from <https://www.datanovia.com/en/blog/types-of-clustering-methods-overview-and-quick-start-r-code/>

[13] Lloyd, S. (1982). Least squares quantization in PCM. IEEE transactions on information theory, 28(2), 129-137.

[14] Jeffares, A. (2019, November 19). K-means: A Complete Introduction. Retrieved from <https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c>

[15] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3), 191-203.

[16] Joshua Zhexue Huang, Michael K Ng, Hongqiang Rong, and Zichen Li, “Automated variable weighting in kmeans type clustering,” IEEE Transactions on Pattern Analysis & Machine Intelligence, , no. 5, pp. 657–668, 2005.

[17] Wang, X., Wang, Y., & Wang, L. (2004). Improving fuzzy c-means clustering based on feature-weight learning. Pattern recognition letters, 25(10), 1123-1132.

[18] Nazari, M., Shanbehzadeh, J., & Sarrafzadeh, H. (2013). Fuzzy c-means based on automated variable feature weighting.

[19] Li, C., & Yu, J. (2006, July). A novel fuzzy c-means clustering algorithm. In International Conference on Rough Sets and Knowledge Technology (pp. 510-515). Springer, Berlin, Heidelberg.

[20] Wu, K. L., & Yang, M. S. (2002). Alternative c-means clustering algorithms. Pattern recognition, 35(10), 2267-2278.

[21] Xiang, W., Rui, G., Jizhong, L., Xiaoying, G., Lina, W., Wei, L., ... & Ke, Z. (2008, December). A Novel Alternative Weighted Fuzzy C-means Algorithm and Cluster Validity Analysis. In 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application (Vol. 2, pp. 130-134). IEEE.

[22] Hung, C. C., Kulkarni, S., & Kuo, B. C. (2010). A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification. IEEE Journal of Selected Topics in Signal Processing, 5(3), 543-553.

[23] B. C. Kuo and D. A. Landgrebe, “Nonparametric weighted feature extraction for classification,” IEEE Trans. Geosci. Remote Sens., vol. 42, no. 5, pp. 1096–1105, May 2004.

[24] Mallapragada, S., Wong, M., & Hung, C. (2018). “Dimensionality Reduction of Hyperspectral Images for Classification” (pp. 153-159). International Information Institute.

[25] Folkman, Tyler. “Why You Are Using t-SNE Wrong.” Medium, Towards Data Science, 5 Apr. 2020, towardsdatascience.com/why-you-are-using-t-sne-wrong-502412aab0c0.

[26] UCI Machine Learning Repository: Iris Data Set, archive.ics.uci.edu/ml/datasets/iris.

[27] UCI Machine Learning Repository: Wine Data Set, archive.ics.uci.edu/ml/datasets/wine.

[28] “Hyperspectral Remote Sensing Scenes.” Hyperspectral Remote Sensing Scenes - Grupo De Inteligencia Computacional (GIC), [www.ehu.eus/ccwintco/index.php/Hyperspectral\_Remote\_Sensing\_Scenes](http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

[29] Li, H., Zhang, S., Ding, X., Zhang, C., & Dale, P. (2016). Performance evaluation of cluster validity indices (CVIs) on multi/hyperspectral remote sensing datasets. Remote Sensing, 8(4), 295.

[30] Trauwaert, E. (1988). On the meaning of Dunn's partition coefficient for fuzzy clusters. Fuzzy sets and systems, 25(2), 217-242.

[31] Stephanie. “Cohen's Kappa Statistic.” Statistics How To, 12 Oct. 2017, [www.statisticshowto.com/cohens-kappa-statistic/.\](http://www.statisticshowto.com/cohens-kappa-statistic/.\)

[32] Zhi, X. B., Fan, J. L., & Zhao, F. (2014). Robust local feature weighting hard c-means clustering algorithm. Neurocomputing, 134, 20-29.