Fractal-based Supervised Approach for Dimensionality Reduction of Hyperspectral Images

Vanshika Gupta[a], Sharad Kumar Gupta[b,c,*] and Amba Shetty[d]

[a]Dept. of Civil Engineering, National Institute of Technology Karnataka, Surathkal, India
[b]Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany
[c]Porter School of Environment and Earth Sciences, Tel Aviv University, Tel Aviv, Israel
[d]Dept. of Water Resource and Ocean Engineering, National Institute of Technology Karnataka, Surathkal, India

ARTICLE INFO

Keywords:
Fractal Dimension
Dimensionality Reduction
Intrinsic Dimension
Hyperspectral Imagery
Support Vector Machine

ABSTRACT

Dimensionality reduction is one of the most challenging and crucial issues apart from data mining, security, and scalability, which have retained much traction due to the ever-growing need to analyze the large volumes of data generated daily. Fractal Dimension (FD) has been successfully used to characterize data sets and has found relevant applications in dimension reduction. This paper presents an application of the FD Reduction (FDR) Algorithm on geospatial hyperspectral data, examining its usefulness for data sets with a relatively high embedding dimension. We examine the algorithm at two levels. First is the conventional FDR approach (unsupervised) at the image level. Alternatively, we propose a pixel-level supervised approach for band reduction based on time-series complexity analysis. Techniques for determining an optimal intrinsic dimension for the dataset using these two techniques are examined. We also develop a parallel GPU-based implementation for the unsupervised image-level FDR algorithm, reducing the runtime by nearly 10 times. Furthermore, both approaches use a support vector machine classifier to compare the classification performance of the original and reduced image obtained.

1. Introduction

Technological innovation in hardware and storage space has enabled the collection of increasingly large and complex information, which has helped capture more attributes pertaining to the data and added a higher level of detail to the attributes. However, analyzing and deriving meaningful inferences from it has become more challenging. While we take advantage of the space, we capture data that might be redundant.

Many classification and regression algorithms cannot learn from this redundant information, decreasing the overall performance of the desired task. This behavior called *"Curse of Dimensionality"*, or the *Hughes Phenomenon* (Hughes, 1968), states that with the increase in the embedding dimension, the available data becomes more and more sparse in its own address space (Aggarwal et al., 2001). This results in an exponential increase in the number of data points required for any useful analysis. In most real-world scenarios, including more data points is not easily possible, which directs us toward decreasing the dimensionality of the input data. However, the real-world data may have high dimensionality but could be effectively represented by fewer dimensions due to significant multi-collinearity amongst the attributes.

Hyperspectral remote sensing is one earth observation domain that uses multichannel (high spectral resolution) images acquired within the electromagnetic spectrum's blue and shortwave infrared wavelengths. It provides a higher resolution of the spectral response curve (SRC) to inspect complex surface materials. The image size with hundreds of bands is very large (many gigabytes), posing problems handling the storage and processing time for further analysis. Due to the huge data volume associated with each scene of hyperspectral data, this data type requires more specific attention to the complexity of data receiving, storing, transforming, and processing (Mukherjee et al., 2013).

Hence, it is preferred that the dimension of hyperspectral data be reduced so that both the effect of noise and the curse of dimensionality can be avoided (Mukherjee et al., 2012). Kendall's seminal work in 1961 provided empirical evidence that high-dimensional spaces are predominantly sparse, with data points converging towards the extremities, notably the corners (Kendall, 1961). As a result, high-dimensional hyperspectral data can effectively be represented within a lower-dimensional subspace without significant information loss. In this lower dimensional subspace, the

*Corresponding author
sharadgupta27@gmail.com (S.K. Gupta)
ORCID(s): 0000-0003-4334-1051 (V. Gupta); 0000-0003-3444-1333 (S.K. Gupta); 0000-0002-3633-5254 (A. Shetty)

classifiers can perform well with a much-reduced number of training samples and can produce good classification accuracy. This led to the theory of dimensionality reduction (DR) and a large number of methods had been proposed thereafter to transform the original high dimensional data to another space at a much lower dimension.

DR techniques can be grouped based on the subject or area of focus (Van Der Maaten et al., 2009). From the machine learning or data mining perspective, they can be divided into two categories: *feature selection* and *feature extraction/projection* (Khalid et al., 2014). The former focuses on sub-setting a part of the input set to derive the reduced set (Pudil and Novovičová, 1998). In contrast, the latter is based on a transformed input representation to capture the most relevant information (ping Tian et al., 2013). Another perspective on the DR approaches has also been presented in (Sorzano et al., 2014), where they have been categorized according to the reduction criterion, namely methods based on statistics/information theory, dictionaries, and projection. These techniques range from the most primitive linear techniques like principal component analysis (PCA) (Swain and Banerjee, 2021; Chang et al., 1999; Kambhatla and Leen, 1997) and independent component analysis (ICA) (Wang and Chang, 2006; Lennon et al., 2001) to graph-based techniques such as spectral clustering (Li et al., 2014; Gupta et al., 2018), Laplacian-based embedding (Ghojogh et al., 2023), multi-structure unified discriminative embedding (Luo et al., 2022), discriminative and geometry-preserving adaptive graph embedding (Gou et al., 2023), and collaboration- competition preserving graph embedding (Shah and Du, 2022) etc. to the most recent non-linear (Paul and Chaki, 2019) and neural network-based auto-encoders (Tao et al., 2015; Zabalza et al., 2016).

In this paper, we examine a fractal dimension (FD) based attribute reduction technique, namely the FD reduction (FDR) algorithm introduced by Traina et al. (2000). Fractal geometry, since its introduction by Mandelbrot (Mandelbrot and Van Ness, 1968), has gained popularity in describing various types of naturally occurring structures (Pentland, 1984), which possess self-similarity when viewed at a range of scales (up to a certain limit in a practical setting). This self-similarity can be exact (Figure 1), qualitative, or even statistical (Falconer, 2004). The most popular FD estimation methods in remote sensing include the modified variogram method (Mukherjee et al., 2013), the triangular prism (Clarke, 1986), the blanket method (Peleg et al., 1984), and the adapted Hausdorff distance (Ghosh and Somvanshi, 2008). Moreover, the FD parameter in a single spectral band can be calculated for each pixel (local description) or a square subset of the image (global description) (Aleksandrowicz et al., 2016; Krupiński et al., 2020). Similarly, in the context of hyperspectral data, researchers have used two approaches for FD estimation, i.e., (1) an analysis of spectral signatures or profiles that refer to single pixels (Dong, 2008; Mukherjee et al., 2014), and (2) spectral bands in the whole image or a subset (Qiu et al., 1999). Besides the number of algorithms for FD calculation, there are a few applications where fractal-based analysis has been utilized in several applications in earth observation and remote sensing, such as image compression, dimensionality reduction (DR) (Krupiński et al., 2020), image classification (Su et al., 2019), image segmentation (Karydas, 2020; Coliban et al., 2016), mixture analysis (Patel and Ghosh, 2020), scaling of essential ecosystem variables (Wu et al., 2018), landscapes types classification (Krupiński et al., 2020) etc.

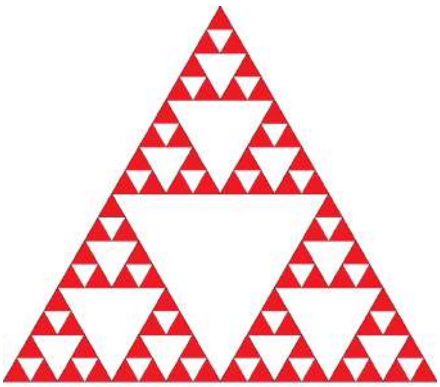


Figure 1: Sierpinski Triangle - an exact spatial fractal

In this research, we have studied a feature selection-based DR approach. In areas such as image analysis and DR, existing studies predominantly build upon two aspects or key properties of fractals, which we have examined in the current research, i.e., (1) as a measure of a shape's roughness and (2) as an indicator of intrinsic dimension. This paper aims to examine, using the FDR algorithm, if the two characteristics mentioned above can be utilized to obtain meaningful results as previously proposed by several studies (Keller et al., 1989; Lam, 1990; Lopes and Betrouni, 2009). Unsupervised FDR (UFDR) methods often fail to capture the intricate variations among different classes, hindering the potential for accurate representation and analysis. By introducing a supervised FDR (SFDR) approach, we aspire to harness the discriminative information present in the dataset's classes, leading to a more meaningful and insightful reduction of dimensions. Moreover, the potential use of FD as a measure of roughness or complexity for time-series fractals has also been utilized using a pixel-level reduction approach of the SRC. A detailed analysis of the FDR is performed on three widely used hyperspectral datasets - *Indian Pines (IP)*, *Pavia University (PU)*, & *University of Houston (UH)*. In this research, we have implemented DR to classify *PU*, *IP* and *UH* hyperspectral images.

For the DR problem, the contributions of this paper are as follows:

- To quickly determine the optimal dimension for the reduced space.
- To develop a novel SFDR method and compare it with the UFDR method for classification accuracy assessment.

2. Theoretical Background

2.1. Concept of Fractals and Fractal Dimension

Within the context of the earth's surface features, every object possesses a characteristic dimension (Mukherjee et al., 2012). Mandelbrot (1982) asserts that when an object exhibits an irregular shape, like a curved line or like the coastline of Britain, its dimension is fractional rather than an integer value. The word "Fractal" was invented by Mandelbrot (1977) to bring together under one heading a large class of objects that have certain structural features in common, although they appear in diverse contexts in astronomy, geography, biology, fluid dynamics, probability theory, and pure mathematics (Xu et al., 1993).

According to Mandelbrot (1982), the term "fractal" comes from the Latin adjective "fractus", which has the same root as "fraction" and "fragment" and means "irregular and fragmented." According to fractal geometry (Mandelbrot, 1977), a straight line has a dimension of one. As the line becomes more and more irregular or curved, the dimension goes on increasing towards two. Thus, the fractal dimension of any curve is intimately related to its shape/ irregularity and can be considered as its characteristic feature (Mukherjee et al., 2012). Fractal structures are characterized by the Hausdorff-Besicovitch dimension (Balka et al., 2015), also known as the "Fractal Dimension", a non-integer dimension that is essentially a measure of "roughness" (Vadrevu, 2023).

In hyperspectral data, a single pixel contains N response values corresponding to N spectral channels. If the pixel is pure, i.e., there is only one ground cover class represented by that pixel, and if the spectral response values of the pixel are plotted against bands, the curve generated resembles the SRC of the ground cover. The SRC of a land cover class exhibits fractal characteristics, evident through its distinct fractal dimension.

2.2. FD Calculation

Most reduction methods need an optimal estimate of the dimensionality of the reduced set, *a.k.a.* intrinsic dimension (ID) of the dataset, in advance. Hence, an ID estimator must precede any DR technique. Various benchmark ID estimation techniques have been highlighted in (Camastra, 2003; Campadelli et al., 2015). One of the ways to estimate the ID is to study the FD of the data. An intuitive way to understand the FD is to visualize the bulk (e.g., length) of an object to vary exponentially with the size of the measuring instrument or the scale of measurement (Theiler, 1990). This exponent is a measure of the FD 1. For a fractal, the bulk (length, area) is a function of the size of the measuring unit. It is considered to theoretically reach an infinite as the size of the measuring unit approaches 0 (see coastline paradox), unlike Euclidean geometrical figures whose size is constant for any measurement scale.

$$Bulk \sim size^D \tag{1}$$

$$D = \lim_{size \to 0} \frac{log(bulk)}{log(size)} \tag{2}$$

Here, D remains constant for a fractal and can be estimated by finding the slope of the $\log(bulk)$ vs. $\log(size)$ graph. However, to translate it into a numerical estimate that can be algorithmically designed for sampled data, various methods/FD estimators, such as box counting, power spectrum, correlation dimension, etc., were proposed to calculate the FD (Theiler, 1990).

These techniques also have certain limitations. Factors such as quantization and sampling frequency (resolution) of the signal (Huang et al., 1994; Lam and Quattrochi, 1992; Theiler, 1990) determine the extent to which the proposed methods converge with the theoretical limit. In theory, the resolution of the data is assumed to be infinite for the proposed techniques to give an accurate estimate of the FD. The increased resolution of the sampled data enhances the estimator's performance (under certain conditions), while quantization may cause a shift in the value of the FD (Huang et al., 1994). Hence, the absolute value of FD obtained from these techniques is a weak descriptor of the true FD. Another limiting aspect reviewed in (Camastra, 2003) is the number of data points essential to arrive at an accurate estimate of D:

$$D < 2\log_{10} N \tag{3}$$

where N is the number of data points, and D is the FD. So, for a data set with nearly 10,000 points, the ID can be at most 8 for the fractal-based methods to give a near-accurate estimate of the ID. Through the results presented in this paper, we examine the applicability of FD in determining the optimal dimensionality of the 200 and 103 dimensional datasets. The proposed technique reduces the generalization of FD from the image level to a pixel level, considering that similar pixels or data points can be grouped to illustrate the same *FDR* behavior. A similar idea is also presented in (Barbará and Chen, 2003), where a change in FD is used to determine clusters of similar data points.

As stated in many studies (Pentland, 1984; Theiler, 1990), FD coupled with *lacunarity* as a second-order statistic can be used to measure a shape's roughness or complexity. Introduced in (Keller et al., 1989), it has shown some potential for utility in segmentation and characterization in geospatial imagery (Lam, 1990; De Cola, 1989; Qiu et al., 1999; Sun et al., 2006) and elsewhere (Lopes and Betrouni, 2009). Although, it cannot be used as a unique descriptor for a shape that is not a true fractal. In this paper, we attempt to leverage this idea of shape complexity to time-series-like data. When such a structure possesses the self-similar scaling properties of a fractal, either exactly or statistically, it is known as a *time-series fractal* (Pilgrim and Taylor, 2018). Various studies characterizing real-time-series data as fractals have successfully been carried out (Evertsz, 1995; Sewell, 2011) for better prediction and analysis models. Here, we treat the SRC as a variable quantity captured at a range of wavelengths and try to analyze its complexity using its correlation dimension. Some studies suggest that SRC at a given pixel in a hyperspectral image has fractal properties, but more research is needed to validate this claim. Now, the spectral response pattern at any pixel location of any hyperspectral image has its characteristic shape, which cannot be defined by any mathematical formula and is therefore considered an irregular curve (Mukherjee et al., 2012). We intend to find the *basis* and the intrinsic dimension (cardinality of the basis) of the set of E N-dimensional vectors, i.e., eliminate all the correlated attributes.

FD technique is used to aid classifiers in identifying important aspects of the data. This study uses correlation dimension to compute FD (Grassberger and Procaccia, 2004). This is because of two reasons. Firstly, we intend to compare how the FDR algorithm in (Traina et al., 2000) compares for datasets of higher dimensionality, particularly hyperspectral imagery. Since correlation dimension has been employed for their study, we use the same FD estimator as different methods can give slightly varying estimates of FD. Secondly, it is computationally simple and robust for higher dimensional data (Mo and Huang, 2010) and hence is more widely utilized.

2.3. Correlation Dimension

The correlation dimension is estimated by computing the slope of the linear part of the correlation integral curve on a log-log scale (Grassberger and Procaccia, 2004). The correlation integral is given by:

$$C(N,r) = \frac{2}{N(N-1)} \sum_{i \neq j} \Theta(r - \|X_i - X_j\|) \tag{4}$$

where N is the number of data points, r is a arbitrarily chosen distance and Θ is the Step Function such that $\Theta(x) = 1$ if $x \geq 0$, else $\Theta(x) = 0$. In simple terms, $N(N-1)/2$ (denominator) is the total number of point-wise distances (excluding distance to self), and the summation (numerator) indicates the number of point-wise distances greater than r, which can vary between a minimum and maximum inter-point distance.

$C(N,r)$ varies exponentially for a fractal as described previously, wherein the $bulk$ can be equated to the $C(N,r)$ and $size$ to r. Therefore, D (FD) is the slope of the linear fit of the $\log(C(N,r))$ vs. $\log(r)$ curve.

2.4. Fractal Dimension Reduction Algorithm

As proposed by (Traina et al., 2000), the FDR is an *unsupervised* iterative approach for attribute selection based on the idea of retaining the intrinsic dimension (equated to the fractal here) of the data on removal of redundant attributes.

It also aligns with the notion of FD as an approximation of the data complexity, as redundant attributes would do little to change the FD. The FDR is an unsupervised iterative method that selects attributes by preserving the data's ID (referred to as the fractal) while removing redundant attributes. It aligns with the concept that FD approximates data complexity, and redundant attributes have little impact on FD. Instead of FD alone, the *change in FD* is used to characterize the data set. Many studies have already proven this idea of FD as a standalone indicator insufficient (Keller et al., 1989). Though it may not perform well with high-dimensional data, the algorithm has the potential as a feature selection tool and can indicate optimal intrinsic dimensions in certain scenarios. This paper uses the term *unsupervised* FDR to refer to this method.

Given a set of N points, with embedding dimensionality E, we can represent the data as (N, E), and the algorithm used can be summed up as follows:

$i \leftarrow 1$
$S \leftarrow (N, E)$
while $i \leq E$ **do**
    $D \leftarrow FD$ of S
    for $j = 1$ **to** $(E - i - 1)$ **do**
        $S' \leftarrow$ Remove j^{th} dimension from S
        $pD_j \leftarrow FD$ of S'
    end for
    $I \leftarrow j$ for which $|D - pD|$ is minimum
    $S \leftarrow$ Remove I^{th} dimension from S
end while

where pD is the *partial FD*. It is the FD of the resulting set obtained after removing one of the attributes from the data. Inside the loop, S' is of the form $(N, E - i)$, i.e., i dimensions have been removed.

2.5. Hyperspectral Data

A hyperspectral image is a remotely sensed geospatial image taken using sensors that can capture surface reflectance at nearly hundreds of wavelengths at a very low bandwidth. RGB images are 3 band images $(H, W, 3)$ captured at 3 wavelengths in the EM spectrum. Beyond RGB comes *multispectral images*, which contain approximately 10 or fewer bands. Similarly, Hyperspectral is of the form (H, W, E) where E is of the order of $100's$. At a particular pixel (h, w), a *spectral response curve* can be obtained as shown in Figure 2. Each pixel represents the overall value of the surface reflectance at a range of wavelengths. Depending on the spatial resolution of the sensor, the reflectance at a pixel can be from a single or mixed land cover.

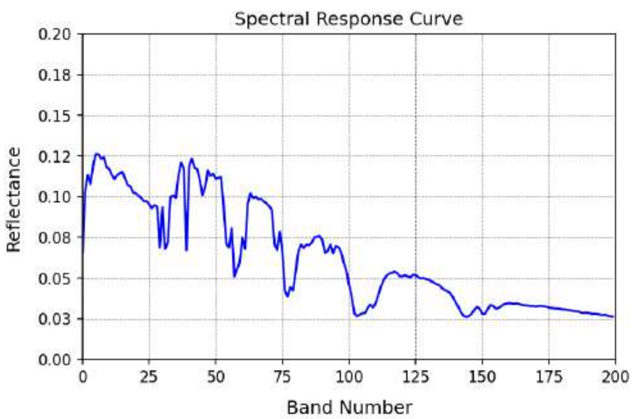


Figure 2: Spectral response at a pixel in a Hyperspectral Image containing 200 bands

2.6. Support Vector Machine Classifier

Support vector machine (SVM) classifier is based on computing an optimal hyperplane that divides the data into two parts such that the two classes lie on either side of the hyperplane. Initially designed for linearly separable data (Boser et al., 1992), kernel tricks have aided the classification of rather complex data distributions applicable to a real setting. The optimization function for SVM maximizes the margin of separation for the two classes, and the points belonging to each class lie on the opposite side of the hyperplane.

Support vectors are the data points closest to the decision boundary or hyperplane and are critical to training. Using Lagrange's multiplier and kernel function (Berwick, 2003), we get the following optimization function:

$$\begin{aligned} L_d &= \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N} a_i a_j y_i y_j K(x_i \cdot x_j) \\ & S.t. \, \forall i, \; 0 < a_i < C, \; \sum_{i=1}^{N} a_i y_i = 0 \end{aligned} \tag{5}$$

$$K(x, y) = (\gamma(x \cdot y) + 1)^p \tag{6}$$

$$K(x, y) = e^{-\gamma \|x-y\|^2} \tag{7}$$

where a_i is the weight to be optimized, y_i is the label for sample x_i and K is the Kernel Function and N is the total number of training samples. Equation 6 and 7 represent the most commonly used kernels: *polynomial* (Poly) and *radial basis function* (RBF). The input parameters for these kernel functions are p and σ, respectively.

A higher penalty parameter C would mean more penalty on miss-classified samples, leading to overfitting and poor generalization, especially when kernel functions are employed. For multi-class, we use the *one-vs-rest* decision approach. As evident from the kernel functions (see eq 6 and 7), the polynomial function fits a polynomial of a specified degree as the decision boundary. In contrast, the RBF kernel can fit more complex distributions and is much more powerful. The n fractal features generated by the proposed methodology have then been used for classification by the SVM classifier using the training pixels, and the accuracy of classification has been obtained using the testing pixels.

3. Datasets

In this research, we have used three widely used hyperspectral datasets, i.e., **Pavia University (PU)** & **Indian Pines (IP)**, & **University of Houston (UH)**, for evaluating the performance of dimensionality reduction using the fractal-based method.

Dataset I

The first dataset used is the IP Hyperspectral image acquired by the AVIRIS sensor across a spectral range 0.4 to 2.5 μm on June 12, 1992, available from the Purdue University Research Repository (PURR)[1]. The image size is 145×145 pixels with a spatial resolution of $20m$, consisting of 200 bands and 16 classes (see Figure 3). We obtained the MATLAB data files of the image from *Grupo de Inteligencia Computacional* website [2].

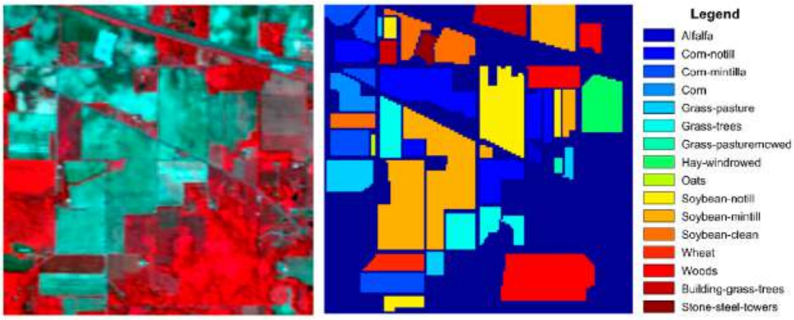
Dataset II

The second dataset is the PU (see Figure 4) Scene captured by the ROSIS sensor at a spectral range of 0.43 to 0.96 μm and resolution of $5nm$ per channel. The image size is 610×340, consisting of 103 bands, 9 classes, and a spatial resolution of $1.3m$. This image is publicly available and was provided by Prof. Paolo Gamba from the Telecommunications and Remote Sensing Laboratory at Pavia University (Italy). We took the MATLAB files from the same source as IP.

These two datasets are widely used in geospatial analysis and concept testing. The PU image has fewer classes and a larger number of samples overall as well as per class. In contrast, the IP image has more classes and a higher level of imbalance coupled with a smaller sample set. The latter will help us evaluate the algorithm for a non-ideal case and test its utility.

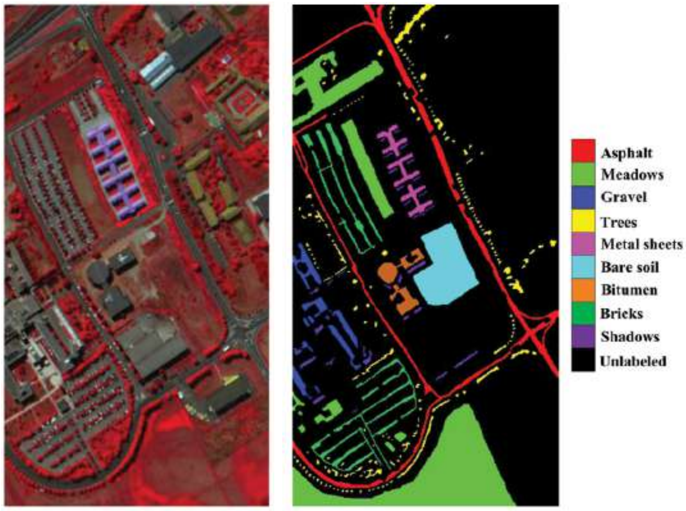
[1] https://purr.purdue.edu/

[2] http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes



(a) (b)

Figure 3: IP hyperspectral image (a) false color composite Image, (b) ground truth



(a) (b)

Figure 4: PU hyperspectral image (a) false color composite Image, (b) ground truth

Dataset III

The third dataset is the Houston 2018 (see Figure 5) scene that was captured by the ITRES CASI 1500 sensor on the University of Houston campus and its vicinity in 2018. This sensor acquires data in a spectral range of 0.38 to 1.05 μm at a spatial resolution of $1m$. The image size is 954×210, consisting of 48 spectral bands and 7 classes. This dataset was originally distributed for the 2018 GRSS Data Fusion Contest [3] (Le Saux et al., 2018) and later provided by Zhang et al. (2023). There are only 22 pixels in the water class, hence, it was merged with background and not used for DR and classification.

4. Methodology

This research is divided into five parts: 1. data processing, 2. fractal dimensionality reduction, 3. estimating the optimal dimension, 4. image classification, and 5. accuracy assessment. The following sections provide details about each part of the developed methodology.

4.1. Data Processing

Data Processing is an integral part of any research on data-driven problems. The data available is usually raw and may not be compatible with the input requirements of the algorithm. It could often be in strings, categorical variables,

[3] https://hyperspectral.ee.uh.edu/?page_id=1075

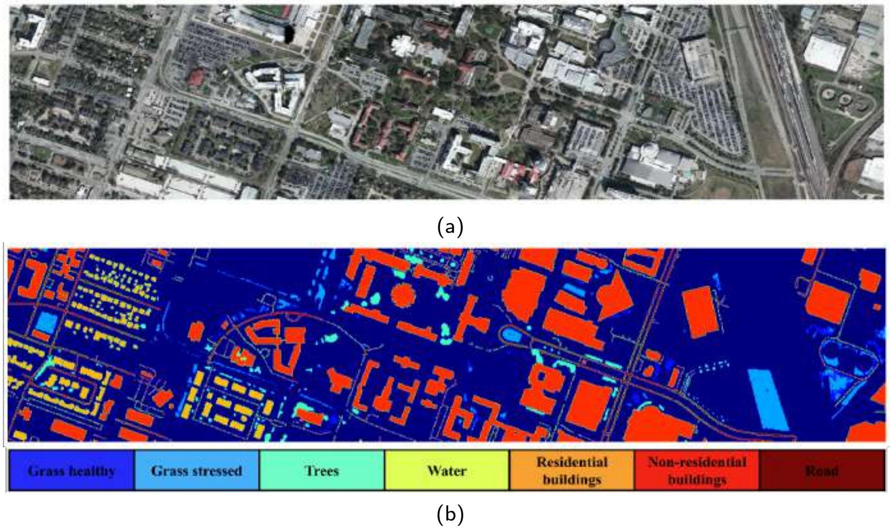


Figure 5: The UH hyperspectral image (a) natural color composite Image (Source: Rasti et al. (2020)), (b) ground truth (Source: Zhang et al. (2023))

or the standard unit format for that particular application. For example, in this study, the remote sensor captures reflectance of the underlying surface, which has a dynamic range of 16 bit, corresponding to integer values from 0 to $65,535$ (Klein et al., 2008). These values can behave abnormally during the classification and, more importantly, with the FD estimation algorithm (Kumaraswamy, 2003). This dependence on data values makes this step critical for our research.

Our primary task is to estimate the image's FD. The original data (.mat file) is $H \times W \times E$, which is first reshaped to a 1D vector of $H * W \times E$. We do not consider unclassified or background classes in any of our implementations. Further, the data is normalized using L^2-norm, each sample row-wise. This step (i.e., normalization) is applied to both supervised (pixel-level) and unsupervised (image-level) algorithms to maintain the characteristics of the data and capture its complexity without changing the nature of the SRC. This property is followed by an L^2-norm.

After applying FDR, we shuffle the original and reduced data with a fixed seed and split it into training and testing sets using a 70:30 ratio. After the split, the same set of normalization consisting of L^2-Norm and feature scaling (0 to 1) is applied to the test and the train data separately. The testing and training samples are normalized separately to not introduce any knowledge from the train to the test set, which can often lead to bias in the test results. The normalization (L^2-norm and feature scaling) step drastically improves the classification performance in SVM.

4.2. Fractal Dimensionality Reduction (FDR)

We have implemented two variants of FDR, i.e., unsupervised and supervised, which are described in the following sections.

4.2.1. Unsupervised Implementation

The unsupervised or conventional FDR (UFDR) approach is followed, as explained in subsection 2.4. Here S is the reshaped and normalized vector ($N \times E$) obtained after applying the procedure as explained in subsection 4.1, where $N = H * W$ is the number of pixels or data points excluding background pixels, and E is the number of bands in the image.

A parallelized algorithm implementation is developed for the procedure, but the algorithm remains unchanged. The large dataset size and high dimensionality make a sequential implementation infeasible. Specifically, we parallelize the inner *for* loop, where the subroutine estimates the *partial FD* vector (pD_j) by temporarily removing attributes one at a time. The parallel pool of processes is closed once the inner loop is complete. The attribute causing a minimum change in FD is then eliminated.

4.2.2. *Supervised Implementation*

This implementation aims to localize the FDR algorithm to a pixel level and analyze its effects on our classification end goal, i.e., compute the FD of the *SRC* at a pixel, treating it as a 1-D time series.

Points are removed iteratively from the SRC to preserve its behavior while minimizing changes to the FD. However, the above-described approach needs to be repeated for all the pixels in the image, thereby processing approximately 10,000 pixels or 1-D wavelength series (based on foreground pixels of the IP image). This process is time-intensive and, hence, infeasible to be pursued.

We propose to follow a SFDR (SFDR) algorithm to handle this issue. We selected one pixel from each class/label representing all the pixels belonging to that class. The wavelengths removed from the SRC of the representative pixel (after applying FDR) are also removed from all the pixels of that class, which means that if we have a pixel (vector at a pixel location) $x_k = (1 \times E)$ with $y_k = 2$, and after an inner loop of FDR, the point at position j gives a minimum change in FD, then $\forall x_i$ for which $y_i = 2$, we remove the point at index j. The proposed approach can be summed up as follows:

$S \leftarrow (1, E)$
for $k = 1$ **to** $num\_classes$ **do**
    $S_k \leftarrow (N_k, E) = \{x_t \mid y_t = k\}$
    $S \leftarrow (1, E)$ - Any one pixel from S_k
    $i \leftarrow 1$
    while $i \leq E$ **do**
        $D \leftarrow FD$ of S
        for $j = 1$ **to** $(E - i - 1)$ **do**
            $S' \leftarrow$ Remove j^{th} dimension from S
            $pD_j \leftarrow FD$ of S'
        end for
        $I \leftarrow j$ for which $|D - pD|$ is minimum
        $S \leftarrow$ Remove I^{th} dimension from S
        $S_k \leftarrow$ Remove I^{th} dimension $\forall\ x_t \in S_k$
    end while
end for

Using this approach, the FDR will run as many times as the number of classes in the dataset. Stopping the result at the correct iteration (optimal dimension) in the *while* loop will yield the reduced set. This algorithm is substantially faster than the conventional FDR due to 2 reasons: (1) FD estimation of a time series is very fast (vector of size $1 \times E$) (2) the algorithm has to be implemented as many times as the number of classes. Hence, it scales well even for a large number of classes. FD estimation of time series with 200 points (considering IP) takes less than 0.01 seconds on average, and ×16 (16 classes for IP) gives an overall 0.16 seconds for one FD estimation. Therefore, 0.16 seconds compared to 15 seconds for the unsupervised implementation on a CPU is considerably faster.

A supervised technique was also introduced in (Mo and Huang, 2010); however, their approach significantly differs from the proposed method. In (Mo and Huang, 2010), attributes are eliminated based on the variables' effect on the dependent or output variable, unlike our proposed method.

4.3. Estimating the Optimal Dimension

We can analyze the results obtained from the unsupervised and supervised approach to estimate the ID of the data. The change in FD after removing the least significant attribute at each iteration is analyzed against the number of iterations completed (or the number of attributes removed). We could examine this graph by searching for the initial point of an increasing trend or by counting the times ΔFD surpasses a set threshold value if a sudden upward trend is not visible.

The value of this count determines the number of significant attributes in the dataset. The threshold is set based on the error allowed (or precision desired) in preserving the original data complexity.

We can plot the partial FD (with the attribute causing the least difference removed) against the number of iterations to gain further understanding. The point at which this plot drops may indicate which attribute removal substantially impacted the FD and suggest the number of significant attributes. Traina et al. (2000) showed that the FD of the original dataset is that point in this graph where the drop occurs. However, since the Embedding dimensionality of our dataset is very high, deflections to this are very likely.

4.4. Image Classification

Once the approximate dimensionality and reduced set containing only significant attributes are obtained, we test the results using SVM. The change in classification accuracy on the reduced set (SFDR and UFDR) is compared to the original image. We perform hyper-parameter tuning for C, *Gamma*, and *Kernel* to find the optimal values. For C, we used the following values: $1, 5, 10, 50, 100$; for *Gamma*: $0.5, 0.1, 0.05, 0.005$, $scale$; for *kernel*: RBF, Poly (Polynomial of degree 3). Setting gamma as *'scale'* implies that the performance of the classifier is invariant to the scale of X, and its value is set as $1/(num\_features * Variance(X))$, where $num\_features$ is the same as the number of dimensions.

4.5. Accuracy Assessment

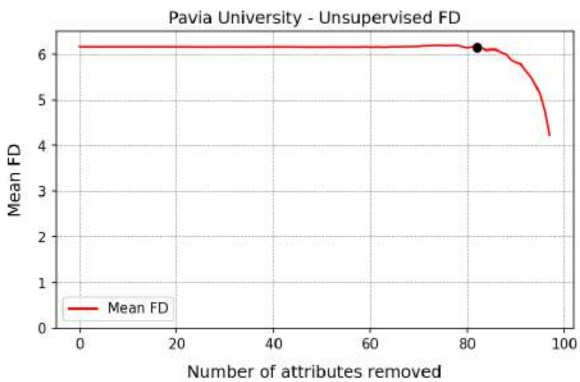
We use accuracy, precision, recall, F1-score, and confusion matrix to test the classifier's performance. While accuracy alone is insufficient for multi-class and imbalanced data, they provide adequate validation for testing the model. *Accuracy* is defined as the percentage of the total number of samples the model correctly classifies. *Precision* measures the ratio of correctly predicted samples to the total number of samples predicted to belong to that class. *Recall*, on the other hand, indicates the proportion of correctly predicted samples to the total number of samples belonging to that class. The model is considered good when precision and recall are close to 1. *F1-score* is the harmonic mean of precision and recall and is ideally close to 1.

5. Results

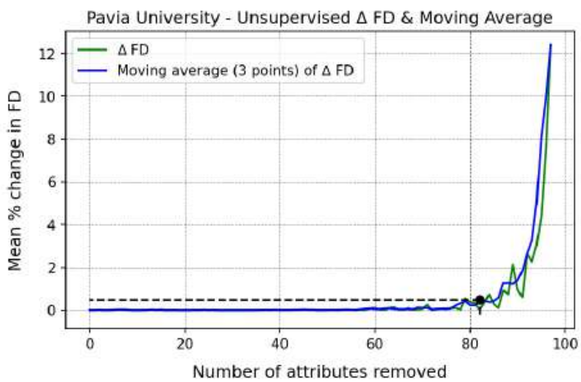
The methodology explained in section 4 was applied to the PU and IP dataset, and the results are explained in this section.

5.1. Unsupervised Approach

The parallel unsupervised version of the algorithm is run on a Windows 10 Pro Intel(R) Xeon(R) Platinum 8160 CPU @2.10 GHz with 24 cores and 48 threads and an NVIDIA Quadro K5200 Graphics card. We use 24 cores to implement the algorithm. The parallelized version was nearly a 10× speed-up as the GPU took approximately 2.5× more time to perform the same FD estimation task. The execution time for computing the FD is noted and is observed to decrease linearly with the number of dimensions, as expected since it has a time complexity of $O(N)$. Since the algorithm is linear, overall speedup compared to sequential implementation is also ~ 10 times. The two graphs discussed in subsection 4.3 for the PU image are shown in Figure 6.



(a)

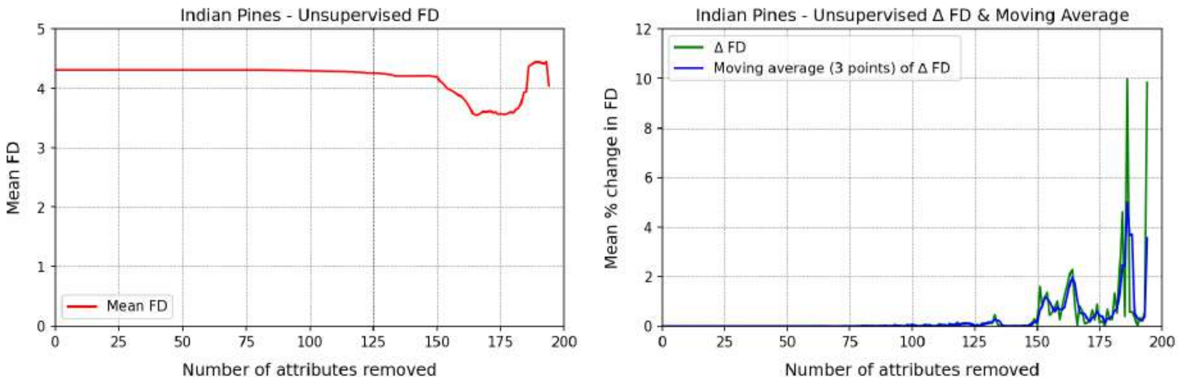


(b)

Figure 6: PU UFDR results (a) FD (b) Δ FD and 3-point moving average of Δ FD as a function of the number of attributes removed

As seen in Figure 6a, removing several attributes does not cause any notable change in the dataset's characteristics. The ΔFD is close to zero in the initial 75% of the curve. The point where the drop occurs (highlighted) is taken at a precautionary distance as we want the precision of selection to be high, or in other words, we can keep some redundant attributes. The important note that we must take here is that although the nature of the graph coincides with that of results shown in Traina et al. (2000), the FD does not act as a guide for the point at which to terminate the algorithm as 6 is too less to represent the intrinsic dimension of such a high dimensional real dataset (Gupta et al., 2018; Golay and Kanevski, 2017; Mukherjee et al., 2012). The insufficiency in the number of data points to concretely establish the FD of the dataset may be attributed as a possible reason (Traina Jr et al., 2010). The change in FD (ΔFD) is plotted in Figure 6 along with its 3-point moving average to perform smoothing and suppress any undulations. We get approximately the same point from the graph, i.e., 19 dimensions. This plot is useful for analyzing when the change in FD starts to increase considerably. However, it cannot give a very accurate estimate as the ΔFD can increase/decrease in a local window of sequential iterations and need not be monotonically non-decreasing in nature (even for ideal datasets). Hence, we only take a 3-point moving average if the trend increases.

The same graphs have been generated for the IP dataset in Figure 7. We can see that although the FD drops at a point, it increases again, which is not the expected behavior of a graph. There can be many possible reasons for this; as explained by PU, the ideal requirement of data points (see Eq 3) is too high compared to the actual number of points available. The deflections to this are more prominent in IP as it has a higher embedding dimension and possibly a higher intrinsic dimension. Hence, the graph obtained cannot be very useful for determining the optimal dimensionality of the data.



(a) (b)

Figure 7: IP UFDR results (a) FD (b) Δ FD and 3-point moving average of Δ FD as a function of the number of attributes removed

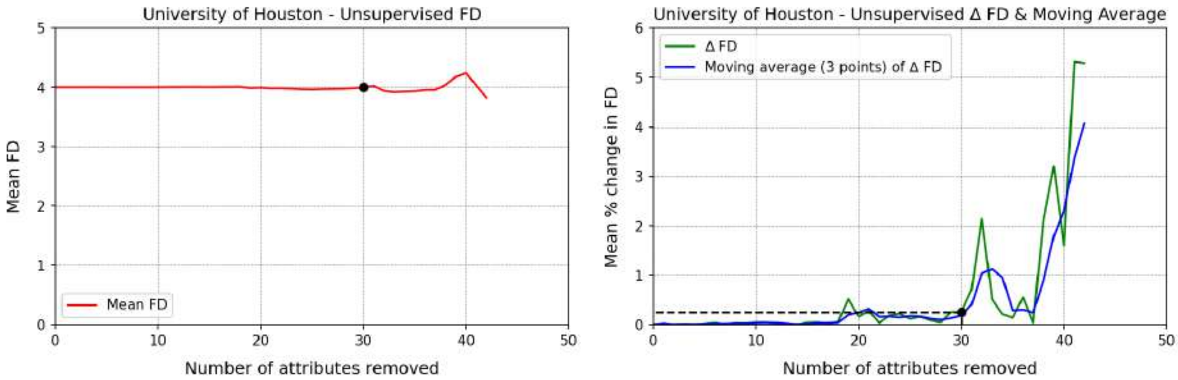
The change in the FD plot, as shown in Figure 7b, also does not give any evidence of a proper elbow. This shows that even if the algorithm proves correctness in reducing to the least dependent attributes, it cannot be relied on statistically, i.e., to estimate the intrinsic dimensionality. Finally, we use a threshold on the allowed percentage change in FD as we wish to eliminate all attributes that do not contribute significantly towards the FD of the dataset. An error of 1% is permitted, and 30 bands crossing this threshold are obtained.

Similarly, for the UH dataset (in Figure 8), the clear transition point cannot be determined using UFDR. One possible reason can be the small number of channels in the dataset. Another possible reason could be the skewness of samples in each class. Even though the number of data points is high, most of them belong to one class.

5.2. Supervised Approach

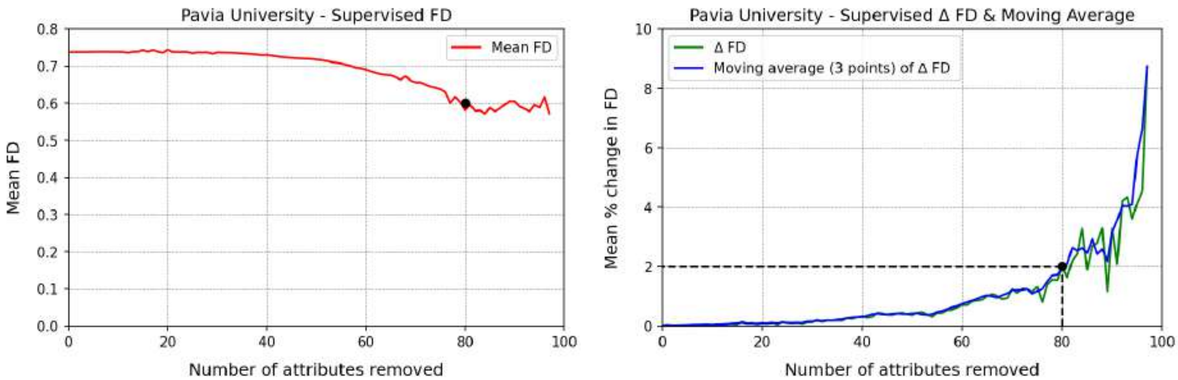
We independently compute each class's FD and percentage ΔFD to derive the same graphs using the SFDR. To satisfy the requirement of having the same dimensions for all classes, we computed the average of these quantities across all classes and then plotted it. As seen in Figure 9a, we do not observe a sudden drop in the FD, and the overall descent is at a smooth gradient. However, we notice an upward trend in Figure 9b. Hence, we take the 3-point moving average and select the point where the graph takes an upfront. Using this approach, we get 22 bands for the reduced set. As mentioned previously, we can also employ a threshold approach here.

Similar results were obtained in the case of the IP image, as shown in Figure 10. Taking a 3-point moving average of the ΔFD, we obtained approximately 30 bands. As stated earlier and from the results, it is clear that this method



(a) (b)

Figure 8: UH UFDR results (a) FD (b) Δ FD and 3-point moving average of Δ FD as a function of the number of attributes removed



(a) (b)

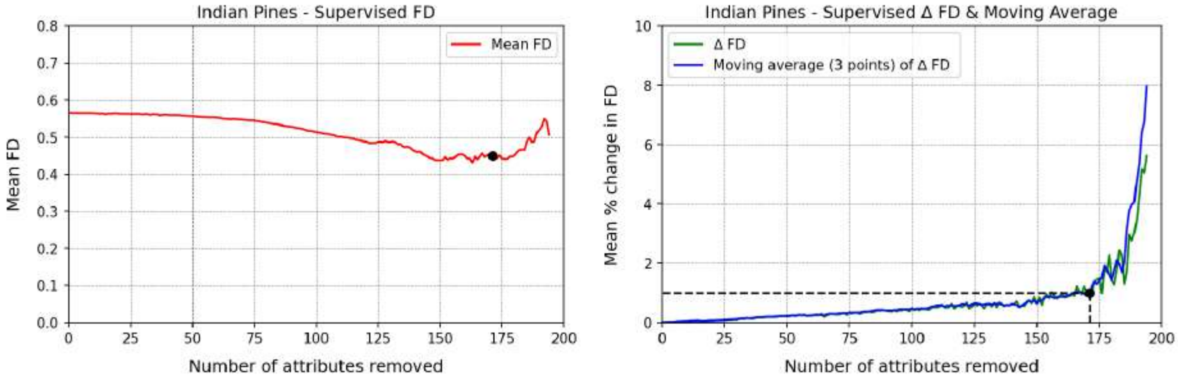
Figure 9: PU SFDR results (a) FD (b) Δ FD and 3-point moving average of Δ FD as a function of the number of attributes removed

is weak in estimating the intrinsic dimensionality of the data. Hence, we can incorporate alternative approaches to estimating this prior (Gupta et al., 2018).

The number of optimal bands chosen for each case has been summarised in Table 1. The estimated number of dimensions for the PU dataset are 19 and 22 using UFDR and SFDR, respectively. The number of dimensions for the IP dataset could not be correctly estimated using UFDR as shown in Figure 7; however, the maximum change in FD is obtained for 30 - 50 dimensions. Hence, we have presented equal FDs for UFDR and SFDR for the IP dataset. For UH datset, the estimated number of dimensions are 18 and 15 using UFDR and SFDR, respectively. It should be noted that the UH dataset contains only 48 bands and the % reduction in dimension is lesser than that of IP and PU datasets. These reduced bands

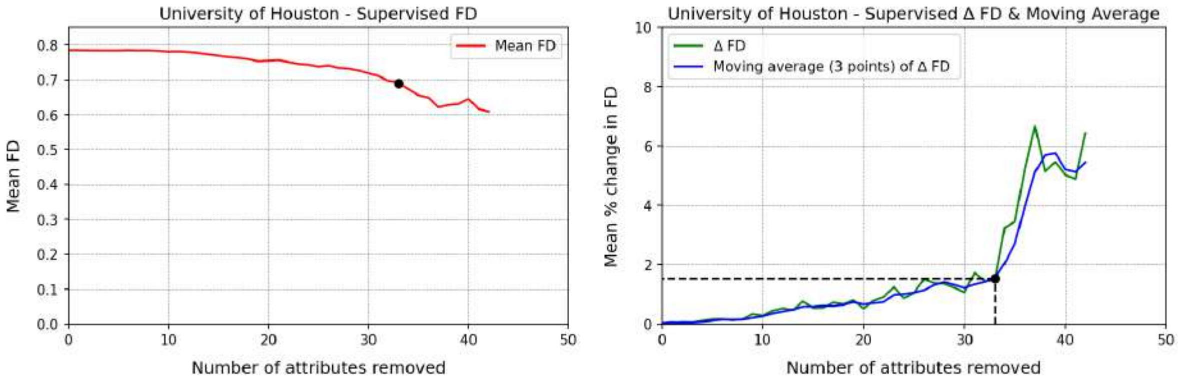
Table 1
Estimated number of dimensions for both datasets

	IP	PU	UH
Unsupervised	30*	19	18
Supervised	30	22	15



(a) (b)

Figure 10: IP SFDR results (a) FD (b) Δ FD and 3-point moving average of Δ FD as a function of the number of attributes removed



(a) (b)

Figure 11: UH SFDR results (a) FD (b) Δ FD and 3-point moving average of Δ FD as a function of the number of attributes removed

The SRC of the reduced and original set obtained using the supervised and unsupervised approaches for PU, IP, and UH datasets (only a few classes) are shown in Figure 12, 13, and 14 respectively. We randomly picked four pixels belonging to different classes and plotted their SRC. For most of the classes, we can see that SFDR-based reduction can preserve the nature of the SRC much better than the UFDR. However, the unsupervised method outperforms the SFDR for class 9 of both PU and IP data. The important point is that the visual similarity of the reduced SRC with the original SRC does not necessarily indicate a better classification. A DR technique that increases the separability of points belonging to different labels can perform better in this respect (similar to LDA, which is a feature projection-based technique).

5.3. Classification

After estimating the number of dimensions and extracting the reduced dataset, we performed SVM classification and compared the original and reduced sets based on the described metrics. The effects of different settings for the hyper-parameters of SVM were studied by exploring variations on a logarithmic scale for each of the four cases described in subsection 4.4. The model is set for convergence until infinite iterations using a tolerance of 0.001. For the PU dataset, there were 29945 and 12831 pixels for training and testing, respectively, whereas, for IP, there were 7176 and 3073, respectively. The data is randomly shuffled using a fixed seed value before splitting between test and train. This data distribution for testing and training in PU after the split is shown in Table 2. A similar split is performed for

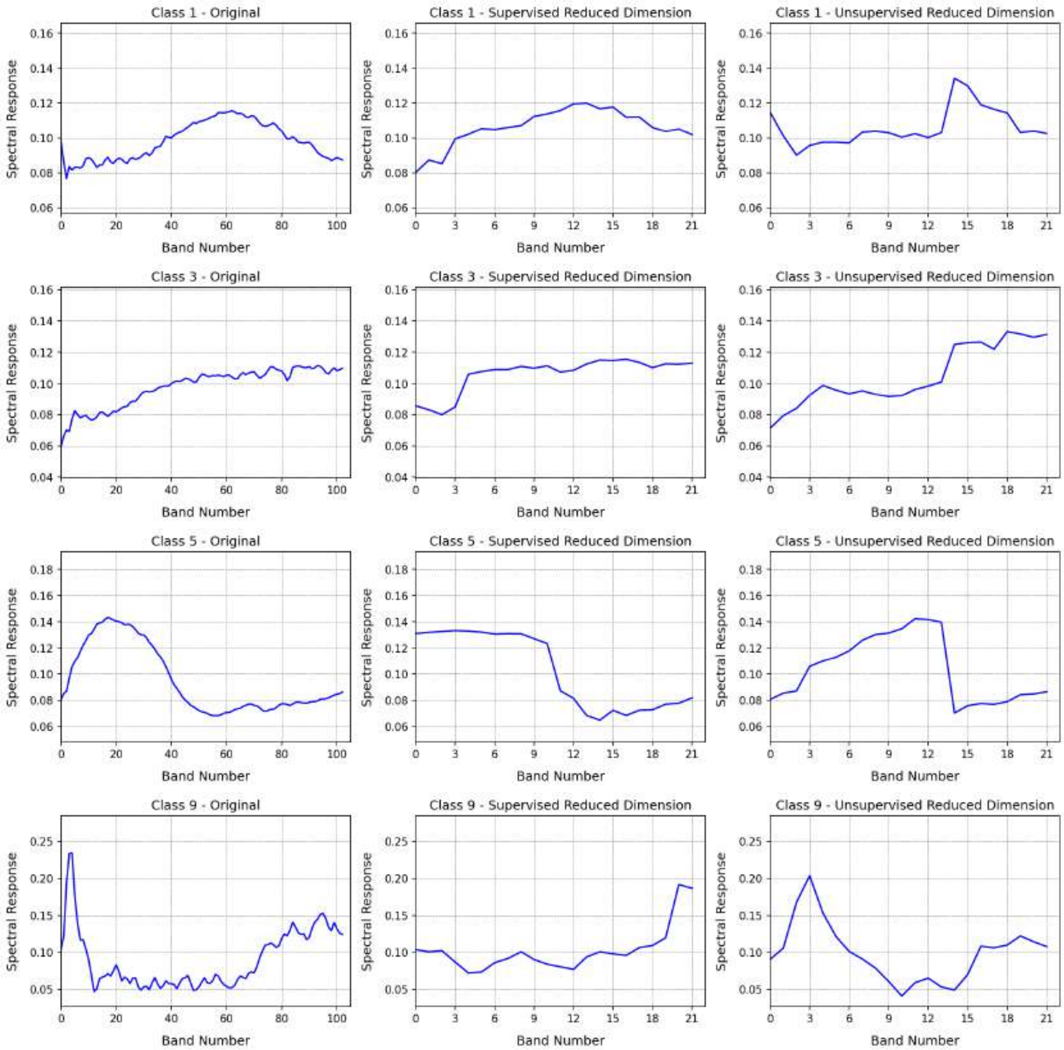


Figure 12: PU - Comparison of Original (left) and Reduced Spectral Response Curve (a) Using SFDR (middle) (b) UFDR (right)

the IP and UH datasets.

After considering all the factors such as overfitting, underfitting, training, and test accuracy, precision, recall, and confusion matrix, parameters, as shown in Table 3, were selected as the best-fit tuning. It is essential to consider all the factors, as the model might exhibit randomness or bias towards a class with more samples. This problem can arise here as the dataset exhibits class imbalance, especially the IP.

The results from the hyper-parameter tuning have been provided as **Supplementary file** containing Tables 1, 2, 3, and 4, 5, and 6, along with classification reports on metrics in Tables 7, 8, 9, and 10). As seen in Table 3, both datasets show a decrease in accuracy in UFDR but a significant increase in accuracy in the supervised approach. The decrease in accuracy is more prominent in the IP image, possibly due to greater imbalance and more classes with fewer samples. The UFDR approach fails to retain the variability of different classes, making it harder for the classifier to construct an optimized decision boundary. Additionally, the lower spatial resolution of IP allows for more mixed pixels, making it difficult to distinguish between the SRC of individual classes. The accuracy of classification is very high on UH image as there are

The confusion matrix for classifying PU image before and after dimensionality reduction is given in Table 4. As can be seen, the model classifies many samples as Class 1. A similar trend can be observed for supervised methods

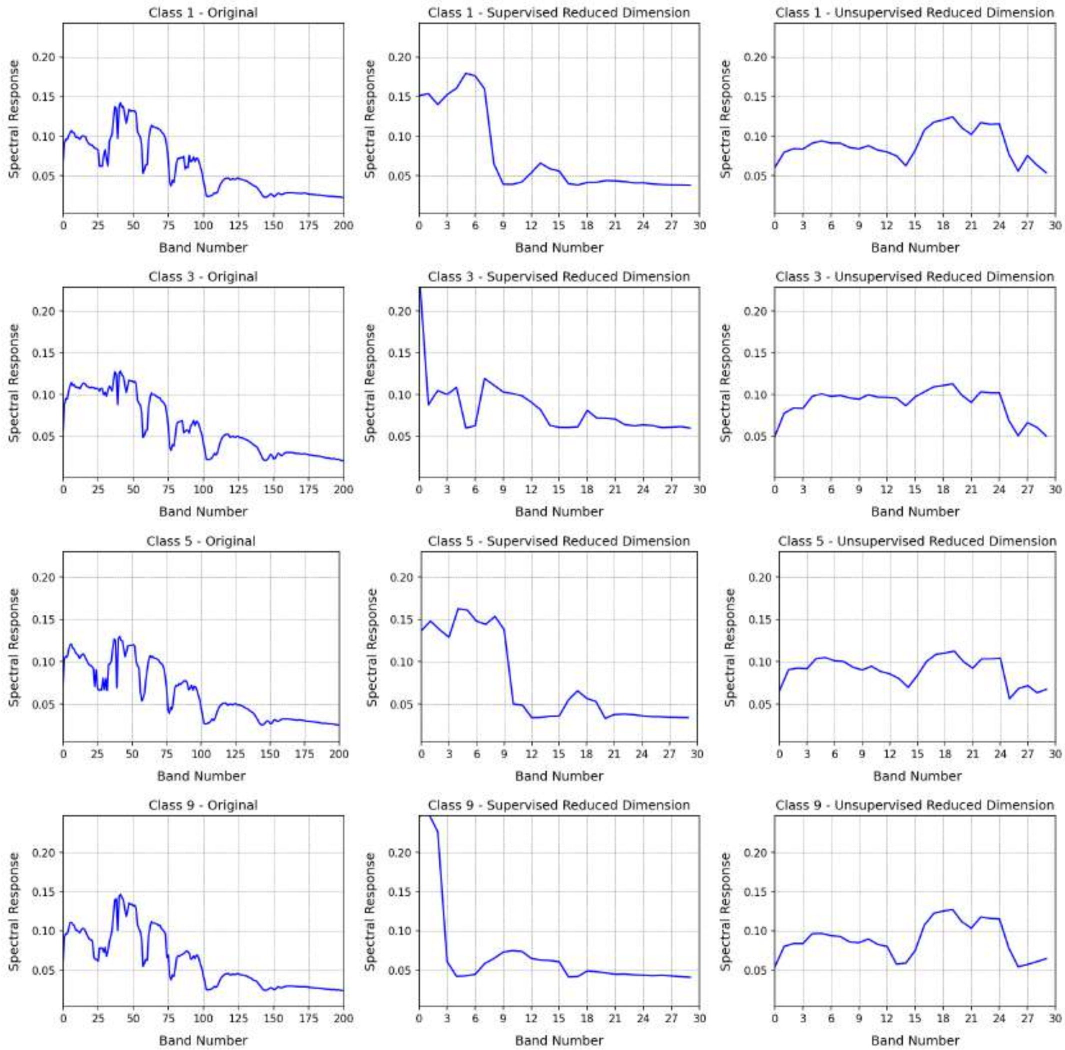


Figure 13: IP - Comparison of Original (left) and Reduced Spectral Response Curve (a) Using SFDR (middle) (b) UFDR (right)

Table 2
Distribution of train-test samples for PU data

Class	# Training	# Testing
1	4665	1966
2	13017	5632
3	1460	639
4	2136	928
5	937	408
6	3542	1487
7	931	399
8	2595	1087
9	662	285

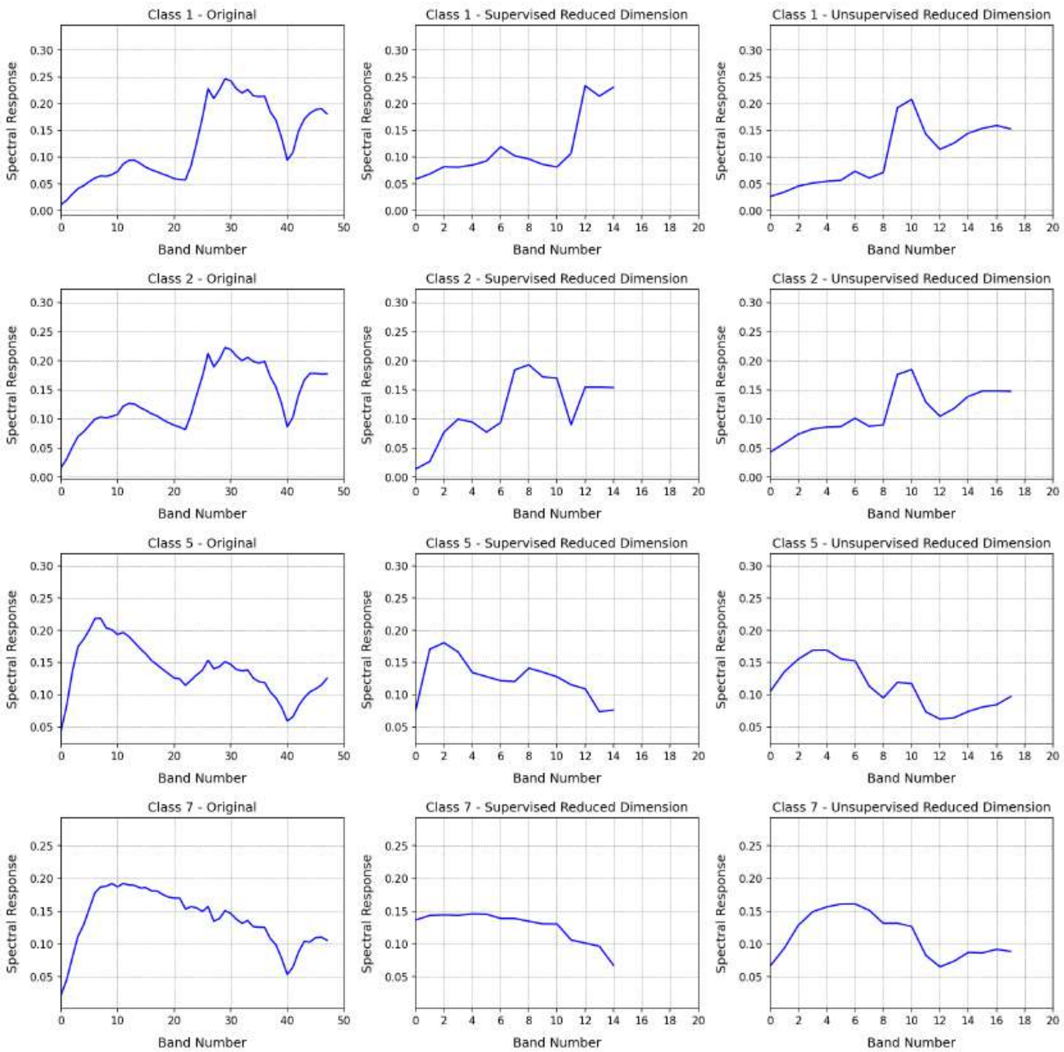


Figure 14: UH - Comparison of Original (left) and Reduced Spectral Response Curve (a) Using SFDR (middle) (b) UFDR (right)

Table 3
Best fit parameters with testing and training accuracy for original and reduced set

Data - Method	Dimension	C	Gamma	Kernel	Original		Reduced	
					Training	Testing	Training	Testing
PU - Unsupervised	19	1	scale	rbf	87.67	85.45	83.69	81.85
IP - Unsupervised	30	50	scale	rbf	95.96	83.24	84.69	68.17
UH - Unsupervised	18	50	scale	poly	95.30	94.40	93.02	92.45
PU - Supervised	22	1	scale	rbf	87.45	86.04	96.87	94.56
IP - Supervised	30	1	0.05	poly	86.38	81.06	94.25	94.43
UH - Supervised	15	50	0.5	poly	95.24	94.44	100.00	100.00

(Table 4(a)) as well as unsupervised methods (Table 4(b)). This indicates that minimal data characteristics are lost even after reducing it to nearly $1/5^{th}$ of the original embedding dimension. SVM's ability to distinguish between classes improves in the SFDR approach.

Table 4
Confusion matrix for PU dataset consisting of 9 classes (1 to 9) (a) supervised for original (left) and reduced data (right) (b) unsupervised for original (left) and reduced data (right)

Actual Label \ Predicted Label	1	2	3	4	5	6	7	8	9
1	1634	5	17	0	0	3	249	58	0
2	45	5441	0	42	0	101	0	3	0
3	93	2	421	0	0	0	2	121	0
4	0	73	0	844	0	11	0	0	0
5	0	0	0	0	408	0	0	0	0
6	26	512	20	1	0	894	5	29	0
7	17	1	3	0	0	0	376	2	0
8	216	4	102	0	0	10	12	743	0
9	4	1	0	0	0	0	0	0	280

Actual Label \ Predicted Label	1	2	3	4	5	6	7	8	9
1	1510	0	0	0	0	0	450	6	0
2	0	5602	0	0	0	30	0	0	0
3	3	1	632	0	0	0	3	0	0
4	1	1	0	925	0	1	0	0	0
5	0	0	0	0	408	0	0	0	0
6	4	24	1	0	0	1448	10	0	0
7	26	2	0	0	0	0	370	1	0
8	5	0	0	1	0	0	116	965	0
9	10	0	0	0	0	0	1	0	274

(a)

Actual Label \ Predicted Label	1	2	3	4	5	6	7	8	9
1	1825	5	12	0	0	10	9	45	0
2	18	5239	0	8	0	447	0	2	0
3	127	1	400	0	0	3	0	117	0
4	0	140	0	744	0	15	0	0	0
5	5	0	0	0	424	1	0	0	0
6	38	270	21	0	0	1177	0	17	0
7	135	0	0	0	0	0	239	1	0
8	289	3	105	0	0	16	0	646	0
9	4	1	1	0	0	0	0	0	271

Actual Label \ Predicted Label	1	2	3	4	5	6	7	8	9
1	1611	10	105	0	0	22	1	157	0
2	4	5645	1	13	0	50	0	1	0
3	30	2	552	0	0	3	0	61	0
4	0	224	0	667	0	8	0	0	0
5	2	1	0	0	424	1	2	0	0
6	27	768	10	0	0	698	1	19	0
7	311	1	4	0	0	2	32	25	0
8	61	7	365	0	0	25	0	601	0
9	3	1	0	0	0	0	0	0	273

(b)

Further, to check if the enhanced classification ability of the supervised approach in the Pavia image is due to more bands, we computed the classification report for a range of reduced dimensions. We found that the test accuracy for PU on the reduced set obtained using UFDR (consisting of 22 bands) was smaller from that of SFDR.

6. Discussion

In this research, a fractal-based method for the DR of hyperspectral data has been proposed. Instead of using all of the bands of the original hyperspectral data for analysis, this method attempts to find a much-reduced dimension by using an algorithm for generating the fractal features. The fractal dimension considers the structure of the SRC, while the corresponding energy accounts for class separation (Mukherjee et al., 2012). Consequently, the reduced-dimensional features encapsulate both class-specific characteristics and inter-class variation.

We compared the proposed SFDR algorithm with various state-of-the-art methods published in the field to establish its advantages. For the sake of fair comparison, we have not compared the proposed algorithm with complex deep learning methods. The methods used for comparison included two-stage subspace projection (Li et al., 2018), joint spectral-spatial fractal method (Su et al., 2019), multiple edge-preserving features and multiple feature learning (Tian et al., 2019), improved spatial-spectral weight manifold embedding (Liu et al., 2020), self-organizing maps (Hidalgo et al., 2021), multi-structure unified discriminative embedding (Luo et al., 2022), spatial-aware collaboration-competitive preserving graph embedding (Shah and Du, 2022). In case of UH dataset, the accuracy of classification has been compared for 6 classes provided in Zhang et al. (2023) except water. As seen from the results presented in Table 5, we can see that the proposed SFDR algorithm has obtained higher overall accuracy and better data reduction compared to the state-of-the-art methods and, to some extent, outperformed the other models.

The proposed method efficiently utilizes the unique FD value associated with each SRC to reduce the dimensionality. It is able to produce appreciable results in terms of classification accuracy for data having high dimensions, having

Table 5
Comparison of classification results for state-of-the-art DR methods with proposed SFDR approach

S. No.	Method	PU Dataset		IP Dataset		UH Dataset	
		Accuracy (%)	Data Reduction (%)	Accuracy (%)	Data Reduction (%)	Accuracy (%)	Data Reduction (%)
1.	Li et al. (2018)	-	-	73.90	**90.00**		
2.	Su et al. (2019)	84.57	65.05	-	-		
3.	Tian et al. (2019)	94.40	63.11	85.22	81.00		
4.	Liu et al. (2020)	86.98	80.85	84.71	**90.00**		
5.	Hidalgo et al. (2021)	81.90	37.80	92.50	77.50		
6.	Luo et al. (2022)	83.00	**88.35**	-	-		
7.	Shah and Du (2022)	-	-	91.57	75.00		
8.	Li et al. (2022)	-	-	93.89	-	**91.82** (6 Classes from best models)	-
9.	Akwensi et al. (2023)	-	-	-	-	**84.52**	37.50
10.	Proposed (SFDR)	**94.56**	78.64	**94.43**	85.00	**100.00 (6 Classes)**	68.75

only pure pixels, and a large number of classes of subtly different spectral response patterns that are otherwise difficult to distinguish from each other. Furthermore, the complexity of the FD algorithm may affect the efficiency of feature analysis. With the rapidly increasing spatial and spectral resolutions in hyperspectral imagery, the efficiency of feature computation and extraction will become more and more critical for complex feature analysis and extraction approaches (Su et al., 2019). As a feature reduction method, FD is also faced with the challenge of condensing information with high accuracy, which affects the accuracy of hyperspectral image classification. Our proposed SFDR approach has demonstrated an effective dimensionality reduction method for hyperspectral imagery, but challenges still exist similar to other feature extraction methods, such as PCA, ICA, and various kernel-based methods, etc. The SFDR algorithm used for FD estimation is simple and effective. It can extract information and reduce the dimension of SRC in the spectral domain. The FD can be used as an additional feature for object segmentation, classification, and recognition in hyperspectral imagery.

7. Conclusion

The application of the FDR examined on hyperspectral images reveals certain limitations in deriving statistical inferences for estimating an optimal intrinsic dimension for the dataset. Qualitatively or in terms of reducing significant attributes, the unsupervised algorithm presents promising results and can be employed for dimensionality reduction. The supervised approach performs better in all datasets (IP, PU and UH) and enhances the classification performance to a great extent. The results show that the FDR analysis is suitable for complex and nonlinear objects in hyperspectral image data reduction and classification. However, a supervised technique can only be utilized in certain settings where the ground truth is available. The rationale behind developing the SFDR approach was that when we apply unsupervised reduction using FDR to the whole image, it tries to generalize the intrinsic dimension. However, each class can have its own intrinsic dimension and may improve the results using the knowledge of different classes.

The supervised version can be further extended to localized unsupervised or semi-supervised versions wherein a local window of some pixels can be considered for carrying out common FDR (instead of one pixel for a class). This local window can be a conventional moving window with some stride or a cluster of points selected based on a similarity metric. Such improvisations can reduce the breadth of data for which we derive a common iterative elimination and might enhance the desired outcome. However, it comes at the cost of increased computation time as the number of iterations for which FDR is run increases.

Code Availability

The code for this work is available at `https://github.com/vansjyo/Fractal_Dimension_MP`

Acknowledgment

The authors would like to thank Dr. Dericks P. Shukla, Associate Professor, IIT Mandi, for providing the computational resources to carry out some parts of the work.

CRediT authorship contribution statement

Vanshika Gupta: Conceptualization of this study, Methodology, Implementation, and Manuscript Preparation. **Sharad Kumar Gupta:** Conceptualization of this study, Methodology, Revision, Review, and Editing. **Amba Shetty:** Conceptualization of this study, Review, and Editing.

References

Aggarwal, C.C., Hinneburg, A., Keim, D.A., 2001. On the surprising behavior of distance metrics in high dimensional space, in: International conference on database theory, Springer. pp. 420–434.

Akwensi, P.H., Kang, Z., Wang, R., 2023. Hyperspectral image-aided LiDAR point cloud labeling via spatio-spectral feature representation learning. International Journal of Applied Earth Observation and Geoinformation 120, 103302. URL: https://doi.org/10.1016/j.jag.2023.103302, doi:10.1016/j.jag.2023.103302.

Aleksandrowicz, S., Wawrzaszek, A., Drzewiecki, W., Krupinski, M., 2016. Change detection using global and local multifractal description. IEEE Geoscience and Remote Sensing Letters 13, 1183–1187. doi:10.1109/LGRS.2016.2574940.

Balka, R., Buczolich, Z., Elekes, M., 2015. A new fractal dimension: The topological Hausdorff dimension. Advances in Mathematics 274, 881–927. URL: https://linkinghub.elsevier.com/retrieve/pii/S0001870815000389, doi:10.1016/j.aim.2015.02.001, arXiv:1108.4292.

Barbará, D., Chen, P., 2003. Using self-similarity to cluster large data sets. Data Mining and Knowledge Discovery 7, 123–152.

Berwick, R., 2003. An idiot's guide to support vector machines (svms). Retrieved on October 21, 2011.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the fifth annual workshop on Computational learning theory, pp. 144–152.

Camastra, F., 2003. Data dimensionality estimation methods: a survey. Pattern recognition 36, 2945–2954.

Campadelli, P., Casiraghi, E., Ceruti, C., Rozza, A., 2015. Intrinsic dimension estimation: Relevant techniques and a benchmark framework. Mathematical Problems in Engineering 2015.

Chang, C.I., Du, Q., Sun, T.L., Althouse, M.L., 1999. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE transactions on geoscience and remote sensing 37, 2631–2641.

Clarke, K.C., 1986. Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. Computers & Geosciences 12, 713–722. doi:10.1016/0098-3004(86)90047-6.

Coliban, R.M., Radoi, A., Ivanovici, M., 2016. A color and multispectral fractal model for forest region identification in satellite images, in: 2016 International Conference on Communications (COMM), IEEE. pp. 381–384.

De Cola, L., 1989. Fractal analysis of a classified landsat scene. Photogrammetric Engineering and Remote Sensing 55, 601–610.

Dong, P., 2008. Fractal signatures for multiscale processing of hyperspectral image data. Advances in Space Research 41, 1733–1743. doi:10.1016/J.ASR.2007.04.090.

Evertsz, C.J., 1995. Fractal geometry of financial time series. Fractals 3, 609–616.

Falconer, K., 2004. Fractal geometry: mathematical foundations and applications. John Wiley & Sons.

Ghojogh, B., Crowley, M., Karray, F., Ghodsi, A., 2023. Laplacian-based dimensionality reduction, in: Elements of Dimensionality Reduction and Manifold Learning. Springer, pp. 249–284.

Ghosh, J.K., Somvanshi, A., 2008. Fractal-based dimensionality reduction of hyperspectral images. Journal of the Indian Society of Remote Sensing 36, 235–241.

Golay, J., Kanevski, M., 2017. Unsupervised feature selection based on the morisita estimator of intrinsic dimension. Knowledge-Based Systems 135, 125–134.

Gou, J., Yuan, X., Xue, Y., Du, L., Yu, J., Xia, S., Zhang, Y., 2023. Discriminative and geometry-preserving adaptive graph embedding for dimensionality reduction. Neural Networks 157, 364–376.

Grassberger, P., Procaccia, I., 2004. Measuring the strangeness of strange attractors, in: The Theory of Chaotic Attractors. Springer, pp. 170–189.

Gupta, V., Gupta, S.K., Shukla, D.P., 2018. Optimal selection of bands for hyperspectral images using spectral clustering, in: International Conference on Recent Trends in Image Processing and Pattern Recognition, Springer. pp. 288–304.

Hidalgo, D.R., Cortés, B.B., Bravo, E.C., 2021. Dimensionality reduction of hyperspectral images of vegetation and crops based on self-organized maps. Information Processing in Agriculture 8, 310–327. doi:10.1016/j.inpa.2020.07.002.

Huang, Q., Lorch, J.R., Dubes, R.C., 1994. Can the fractal dimension of images be measured? Pattern Recognition 27, 339–349.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14, 55–63.

Kambhatla, N., Leen, T.K., 1997. Dimension reduction by local principal component analysis. Neural computation 9, 1493–1516.

Karydas, C.G., 2020. Optimization of multi-scale segmentation of satellite imagery using fractal geometry. International Journal of Remote Sensing 41, 2905–2933.

Keller, J.M., Chen, S., Crownover, R.M., 1989. Texture description and segmentation through fractal geometry. Computer Vision, Graphics, and image processing 45, 150–166.

Kendall, M., 1961. A Course in the Geometry of N Dimensions. Griffin's statistical monographs & courses, Charles Griffin. URL: https://books.google.de/books?id=EakaAAAAMAAJ.

Khalid, S., Khalil, T., Nasreen, S., 2014. A survey of feature selection and feature extraction techniques in machine learning, in: 2014 Science and Information Conference, IEEE. pp. 372–378.

Klein, M.E., Aalderink, B.J., Padoan, R., De Bruin, G., Steemers, T.A., 2008. Quantitative hyperspectral reflectance imaging. Sensors 8, 5576–5618.

Krupiński, M., Wawrzaszek, A., Drzewiecki, W., Jenerowicz, M., Aleksandrowicz, S., 2020. What can multifractal analysis tell us about hyperspectral imagery? Remote Sensing 12, 4077.

Kumaraswamy, K., 2003. Fractal dimension for data mining. Center for Automated Learning and Discovery School of Computer Science Carnegie Mellon University 5000.

Lam, N., 1990. Description and measurement of landsat tm images using fractals. Photogrammetric engineering and remote sensing 56, 187–195.

Lam, N.S.N., Quattrochi, D.A., 1992. On the issues of scale, resolution, and fractal analysis in the mapping sciences. The Professional Geographer 44, 88–98.

Le Saux, B., Yokoya, N., Hansch, R., Prasad, S., 2018. 2018 IEEE GRSS Data Fusion Contest: Multimodal Land Use Classification [Technical Committees]. IEEE Geoscience and Remote Sensing Magazine 6, 52–54. URL: https://ieeexplore.ieee.org/document/8328995/, doi:10.1109/MGRS.2018.2798161.

Lennon, M., Mercier, G., Mouchot, M., Hubert-Moy, L., 2001. Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images, in: IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), IEEE. pp. 2893–2895.

Li, C., Tang, X., Shi, L., Peng, Y., Tang, Y., 2022. A Two-Staged Feature Extraction Method Based on Total Variation for Hyperspectral Images. Remote Sensing 14, 302. URL: https://www.mdpi.com/2072-4292/14/2/302, doi:10.3390/rs14020302.

Li, S., Qiu, J., Yang, X., Liu, H., Wan, D., Zhu, Y., 2014. A novel approach to hyperspectral band selection based on spectral shape similarity analysis and fast branch and bound search. Engineering Applications of Artificial Intelligence 27, 241–250.

Li, X., Zhang, L., You, J., 2018. Hyperspectral image classification based on two-stage subspace projection. Remote Sensing 2018, Vol. 10, Page 1565 10, 1565. doi:10.3390/RS10101565.

Liu, H., Xia, K., Li, T., Ma, J., Owoola, E., 2020. Dimensionality reduction of hyperspectral images based on improved spatial–spectral weight manifold embedding. Sensors 2020, Vol. 20, Page 4413 20, 4413. doi:10.3390/S20164413.

Lopes, R., Betrouni, N., 2009. Fractal and multifractal analysis: a review. Medical image analysis 13, 634–649.

Luo, F., Zou, Z., Liu, J., Lin, Z., 2022. Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding. IEEE Transactions on Geoscience and Remote Sensing 60, 1–16.

Mandelbrot, B.B., 1977. Fractals : form, chance, and dimension. W.H. Freeman San Francisco, San Francisco SE - xvi, 365 pages : illustrations ; 24 cm.

Mandelbrot, B.B., 1982. The fractal geometry of nature. volume 1. WH freeman New York.

Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian Motions, Fractional Noises and Applications. SIAM Review 10, 422–437. doi:10.1137/1010093.

Mo, D., Huang, S.H., 2010. Fractal-based intrinsic dimension estimation and its application in dimensionality reduction. IEEE Transactions on Knowledge and Data Engineering 24, 59–71.

Mukherjee, K., Bhattacharya, A., Ghosh, J.K., Arora, M.K., 2014. Comparative performance of fractal based and conventional methods for dimensionality reduction of hyperspectral data. Optics and Lasers in Engineering 55, 267–274. doi:10.1016/J.OPTLASENG.2013.11.018.

Mukherjee, K., Ghosh, J.K., Mittal, R.C., 2012. Dimensionality reduction of hyperspectral data using spectral fractal feature. Geocarto International 27, 515–531. URL: http://www.tandfonline.com/doi/abs/10.1080/10106049.2011.642411, doi:10.1080/10106049.2011.642411.

Mukherjee, K., Ghosh, J.K., Mittal, R.C., 2013. Variogram fractal dimension based features for hyperspectral data dimensionality reduction. Journal of the Indian Society of Remote Sensing 41, 249–258. doi:10.1007/S12524-012-0225-4/TABLES/4.

Patel, A.K., Ghosh, J.K., 2020. Quantitative analysis of mixed pixels in hyperspectral image using fractal dimension technique. Journal of the Indian Society of Remote Sensing 48, 1237–1244.

Paul, A., Chaki, N., 2019. Dimensionality reduction of hyperspectral images using pooling. Pattern Recognition and Image Analysis 29, 72–78.

Peleg, S., Naor, J., Hartley, R., Avnir, D., 1984. Multiple resolution texture analysis and classification. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6, 518–523. doi:10.1109/TPAMI.1984.4767557.

Pentland, A.P., 1984. Fractal-based description of natural scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6, 661–674.

Pilgrim, I., Taylor, R.P., 2018. Fractal analysis of time-series data sets: Methods and challenges, in: Ouadfeul, S.A. (Ed.), Fractal Analysis. IntechOpen, Rijeka. chapter 2, pp. 05–30.

Pudil, P., Novovičová, J., 1998. Novel methods for feature subset selection with respect to problem knowledge, in: Feature extraction, construction and selection. Springer, pp. 101–116.

Qiu, H.I., Lam, N., Quattrochi, D., Gamon, J., 1999. Fractal characterization of hyperspectral imagery. Photogrammetric Engineering and Remote Sensing 65.

Rasti, B., Hong, D., Hang, R., Ghamisi, P., Kang, X., Chanussot, J., Benediktsson, J.A., 2020. Feature Extraction for Hyperspectral Imagery: The Evolution from Shallow to Deep: Overview and Toolbox. IEEE Geoscience and Remote Sensing Magazine 8, 60–88. doi:10.1109/MGRS.2020.2979764, arXiv:2003.02822.

Sewell, M., 2011. Characterization of financial time series. Rn 11, 01.

Shah, C., Du, Q., 2022. Spatial-aware collaboration–competition preserving graph embedding for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters 19, 1–5.

Sorzano, C.O.S., Vargas, J., Montano, A.P., 2014. A survey of dimensionality reduction techniques. arXiv preprint arXiv:1403.2877 .

Su, J., Li, Y., Hu, Q., 2019. A new spectral-spatial jointed hyperspectral image classification approach based on fractal dimension analysis. Fractals 27, 1950079.

Sun, W., Xu, G., Gong, P., Liang, S., 2006. Fractal analysis of remotely sensed images: A review of methods and applications. International Journal of remote sensing 27, 4963–4990.

Swain, S., Banerjee, A., 2021. Evaluation of dimensionality reduction techniques on hybrid cnn–based hsi classification. Arabian Journal of Geosciences 14, 2806.

Tao, C., Pan, H., Li, Y., Zou, Z., 2015. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. IEEE Geoscience and remote sensing letters 12, 2438–2442.

Theiler, J., 1990. Estimating fractal dimension. JOSA A 7, 1055–1073.

ping Tian, D., et al., 2013. A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering 8, 385–396.

Tian, W., Xu, L., Chen, Z., Shi, A., 2019. Multiple feature learning based on edge-preserving features for hyperspectral image classification. IEEE Access 7, 106861–106872. doi:10.1109/ACCESS.2019.2927786.

Traina, C., Traina, A., Wu, L., Faloutsos, C., 2000. Fast feature selection using fractal dimension. J. Inf. Data Manag. 1, 3–16.

Traina Jr, C., Traina, A., Faloutsos, C., 2010. Fast feature selection using fractal dimension-ten years later. Journal of Information and Data Management 1, 17–17.

Vadrevu, K.P., 2023. Fractal analysis revealed persistent correlations in long-term vegetation fire data in most South and Southeast Asian countries. Environmental Research Communications 5, 011001. URL: https://iopscience.iop.org/article/10.1088/2515-7620/acb041, doi:10.1088/2515-7620/acb041.

Van Der Maaten, L., Postma, E., Van den Herik, J., 2009. Dimensionality reduction: a comparative. J Mach Learn Res 10, 13.

Wang, J., Chang, C.I., 2006. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. IEEE transactions on geoscience and remote sensing 44, 1586–1600.

Wu, L., Liu, X., Qin, Q., Zhao, B., Ma, Y., Liu, M., Jiang, T., 2018. Scaling correction of remotely sensed leaf area index for farmland landscape pattern with multitype spatial heterogeneities using fractal dimension and contextural parameters. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 1472–1481.

Xu, T., Moore, I.D., Gallant, J.C., 1993. Fractals, fractal dimensions and landscapes — a review. Geomorphology 8, 245–262. URL: https://linkinghub.elsevier.com/retrieve/pii/0169555X9390022T, doi:10.1016/0169-555X(93)90022-T.

Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., Du, P., Marshall, S., 2016. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. Neurocomputing 185, 1–10.

Zhang, Y., Li, W., Zhang, M., Qu, Y., Tao, R., Qi, H., 2023. Topological structure and semantic information transfer network for cross-scene hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems 34, 2817–2830. doi:10.1109/TNNLS.2021.3109872.