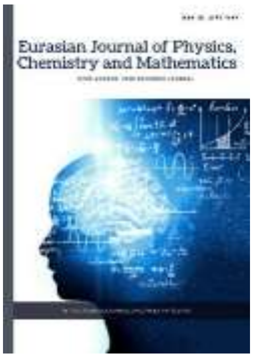

Comparison of some methods of analyzing the three-dimensional spectral image using simulation

Sackineh Shamil Jasim

sackineh.sh@uokerbala.edu.iq
Department of Statistics - College of Administration and Economics -University of Karbala-Iraq

ABSTRACT

In this article, a new method has been proposed in dealing with the spectral image and analyzing it under the principle components method of Three-Dimensional linear-Discriminant-Analysis (3DPCLDA) and the principle components method – Quadratic-Discriminant-Analysis (3DPCQDA) and compared it with the traditional methods, which is the principle components method - linear discriminant analysis. PCLDA) Ordinary and principal components method - Quadratic Discriminant Analysis (PCQDA), Partial Least Squares Method - PLSDA and SVM Method. Using simulation experiments, it was concluded that the three-dimensional methods are better than the traditional methods, as they have achieved the most standards of accuracy and testing.

Keywords: Spectral image, three-dimensional analysis, discriminant analysis, cross-validation, principle components, partial least squares, SVM

1. Introduction:

The multivariate statistical analysis approach's essential component is the description and analysis of phenomena with several dimensions and variables. Discriminant analysis is one of those statistical methods applied to data analysis. A detailed comparison between two or more groups that share a number of characteristics will be made on the basis of several factors. Application of the linearly structured characteristic function, which is composed of the independent variables. Based on numerous characteristics and measurements collected for the examined phenomena variables, it aims to develop a linear or non-linear formula between the study variables to Category a single item into a certain group. Techniques for hyper spectral imaging help in gathering specialized distributed spectral data, where each pixel in the image is made up of a spectrum at a certain wavelength. 3D matrices are made up of the spatial coordinates for these data, which are situated in the x and y axes, and the spectral data, which is positioned in the z axis. The wavelength responses, on the other hand, each represent a two-dimensional (2D) image that is piled on top of the other to generate a three-dimensional object known as a "data cube." In this study, a novel method will be used to investigate the fundamental components of the three-dimensional image within the framework of linear and quadratic discriminant analysis, and its efficacy will be evaluated using some accuracy and sensitivity criteria.

2. Spectral Image

The spectral image, often known as the spectrogram (Specter Gram), is a representation of the spectrum's density and how it changes over time. The spectral falls and sound planning are other names for it. It is employed in a variety of domains, including music, speech processing, sonar and radar, seismology, and others, to recognize severe noises in order to interpret animal screams. The most prevalent type is a geometrical representation with two dimensions, where the horizontal axis denotes time and the vertical axis denotes frequency. The intensity or color at each image point is represented by the third dimension, which displays the amplitude of a specific frequency at a specific time. With the continued use of cutting-edge digital technology in research and its application to all animal sounds, it has been used to study birds. It is very helpful for research on animal frequency modulation. It is helpful in overcoming speech impairments. It is employed in phonetics research and voice synthesis, and spectral pictures are frequently used to make it easier. It can be used to examine a transient signal's test results by a signal processor like a filter to evaluate its performance.. (Auburn, 2022)

3. Linear Discriminant Analysis (LDA)

Finding a linear set of features that separate or distinguishes between two or more Categes of objects or events is done in statistics and other domains using a linear Fisher discriminant function. An expansion of this method is linear discriminant analysis, commonly referred to as normal discriminant analysis (NDA) or discriminant function analysis. The resulting composition may be applied as a linear Categifier or, more frequently, as a dimension-reduction technique before another Categification procedure (which is more commonly used). Analysis of variance, regression analysis, and linear discriminant analysis are all methods that aim to express a certain dependent variable as a linear combination of a number of characteristics or other measures. While discriminant analysis employs continuum independent factors and a categorical dependent variable, analysis of variance uses categorical independent variables and a continuum dependent variable. In comparison to an analysis of variance, linear discriminant analysis and unit probability regression are more similar. For the same reason that they do, the values of continuous independent variables are used to explain a categorical variable. These alternative approaches are used when it is inappropriate to assume that the independent variables have a normal distribution, which is a core tenet of the linear discriminant analysis methodology. Linear discriminant analysis and principal and factor analyses both aim to find linear combinations of variables that best represent the data. The explicit goal of linear discriminant analysis is to define the distinction between different data Categes. As opposed to factor analysis, which adopts the structure of features based on differences rather than similarities, principal component analysis does not account for any variations. The linear discrimination analysis works best when the measurements of the independent variables are comparable for each observation. Discriminant concordance analysis is the comparable method for working with categorical independent variables. multi-layered issues must be handled , linear and quadratic discriminant analysis is a Categifier that may produce linear and quadratic decision surfaces, respectively. It also produces closed solutions and is simple to compute. (Haasdonk &P, ekalska, 2010, 3)

4. Disadvantages of linear and quadratic discriminant analysis in the case of high dimensions

One of the commonly used methods is Fisher's Linear Discriminant Analysis (FLDA) to identify significant linear trends for segregating sample groups rather than PCA principal compounds. However, when the distance of the discriminant is greater than the number of aggregates, then the LDA can find the characteristic trends but in one aggregates. For example, when we have two sets of sample and the need to Categify (separation) between them, the LDA can find only one characteristic trend, Thus, it is possible to lose information important to the distinction in the sample group. Quadratic discriminant analysis (QDA) is a widely used method for Categification and is a general case of linear discriminant analysis

(LDA). When the Categification is in the case of the covariance matrices distinguishing between the categories and in order for the quadratic discriminant analysis to achieve a very high Categification ratio, an accurate estimate must be found about the covariance matrices, in this case everything becomes more difficult in the case of high dimensions since the number of comparable observations has a dimensional property (McLachlan, G. J. , 2004).

5. Three dimension Principal component

Finding the characteristic roots and characteristic vectors of the covariance and covariance matrix for the explanatory variables, or, depending on the type of data, finding the characteristic roots and characteristic vectors of the correlation matrix, forms the basis of the analysis of the main components. . The measurements are different, so we can use the correlation matrix, and the estimations of the basic compounds are among the methods of estimating the regression model that addresses the problem of multilinearity in the data. Principal compound analysis or abbreviated PCA is one of the most successful methods for solving problems that are represented in low dimensions (few dimensions), as Kirby and Sirovich mentioned in their research for more than 20 years that PCA is not only used to reduce the dimensions of samples but also for Categification, analysis and discrimination, noting that PCA illustrates the assumptions of the common variance of all data with the most expressive elements. Martinez, A. M.; Kak, A. C. , 2001)

6. Linear and Quadratic 3D Discriminant Analysis: (Camilo L. et al., 2020)

Similar to other spectroscopic imaging techniques, hyperspectral imaging gathers and analyzes data from the entire electromagnetic spectrum. To locate things, identify materials, or detect processes, hyperspectral imaging seeks to obtain the spectrum of each pixel in a scene image.[1][2] The three main subfields of spectral imaging are as follows: There are three types of scanners: band-series (spectral) scanners, which collect images of a region at various wavelengths; thrust broom and broom-like scanners (spatial scanning), which read images over time; and snapshot imaging. Hyperspectral , which instantly creates an image using a matrix staring gadget. The resulting data are spatially scattered along the x and y axes and spectrally distributed along the z axis. In order to fit the spectral data into a two-dimensional (2D) framework and reconstruct the spatial dimension, this type of data is often analyzed using a multivariate technique. Only a few algorithms, in this case, are able to process a triple set of complete dimensions. Based on an investigation of the key elements of the 3D linear discriminant analysis, new 3D methods are employed here for the discriminant analysis of hyperspectral images: Principle Components Analysis of Discriminant - Linear Discriminant Analysis in Three Dimensions
(3D-PCA-LDA), and the Three-Dimensional Second-Order Discriminant Analysis (Quadratic):-

7. Principle Components Analysis of Discriminant - Quadratic Discriminant Analysis (3D-PCA-QDA).

They are two novel methods for discriminant analysis of multispectral image sets that outperform conventional methods in terms of Categification speed and accuracy. One can collect highly distributed spectral data using high-spectrum imaging techniques, where each image spot (pixel) has a spectrum of a certain wavelength. In a different interpretation, every wavelength response stands for a two-dimensional (2D) image that is piled (accumulated) on top of each other to create a three-dimensional object that is oddly referred to as a "data cube."

Hyperspectral images are Categified using new 3D discrimination analysis techniques termed (3.D-P.C.A-L.D.A) and (3.D-P.C.A-Q.D.A). On the hyperspectral image data set, the following is how the PCA 3D model of major compounds with bilinear locations per pixel position is implemented:

$X_{ij} = T_{ij}P_{ij}^T + E_{ij}$

A temporary matrix with position is called X_{ij} in this scenario. (i and j), samples are represented by the rows, while wave numbers are represented by the columns. T_ij: are the results (degrees) of PCA finding degrees at position i , j , P_{ij}^{T} are the charges or (load process) at position i, j and E_{ij}: represent errors (residues) at position i,j, and the letter T symbolized the shift of the matrix.(Substitution process. Finally, three three-dimensional matrices (D-PCA3) are created of degrees (T), charges (P) and residuals (E). There is a difference in the 3D-PCA model of the linear triad data based on the Tucker3 model ("true 3D-PCA") in which the three-dimensional matrix is decomposed into three charges and a basic matrix, which also differs from the corresponding orthogonal vectors, the "Triple P.C.A". 3.D-P.C.A-L.D.A and .-P.C.A-Q.D.A, linear discriminant analysis (L.D.A) and quadratic discriminant analysis (Q.D.A) are used to Categify and employ them in mean score (3.D-P.C.A), respectively. Thus the 3.D-P.C.A-L.D.A (L_{sk})3 and 3.D-P.C.A-Q.D.A (Q_{sk}) 3 rating scores are calculated as:

$$L_{sk} = (X_s - \bar{X}_k)^T C_{pooled}^{-1}(X_s - \bar{X}_k) - 2Log\ \pi_k \quad \dots (1)$$

$$Q_{sk} = (X_s - \bar{X}_k)^T C_k^{-1}(X_s - \bar{X}_k) + Log|C_k| - 2Log\ \pi_k \quad \dots (2)$$

Whereas, $\boldsymbol{X_s}$: denote the Categ (N × 1) vector where it represents the average score for T of the sample for each main component, S N. $\boldsymbol{\bar{X}_k}$: represented by the Categ vector (N ×1) where it represented by the mean positons for Categ K for each major component N. $\boldsymbol{C_{pooled}}$: represents the pooling covariance matrix. C_k : represents the variance matrix for Categ K. π_k: represents the previous probability of Categ K . $\boldsymbol{C_{pooled}}$ C_k and $\pi_k$are calculated according to the following formula:

$$\boldsymbol{C_{pooled}} = \frac{1}{n}\sum_{k=1}^{k} n_k C_k \quad \dots (3)$$

$$\boldsymbol{C_k} = \frac{1}{nk-1}\sum_{s=1}^{nk}(X_s - \bar{X}_k)(X_s - \bar{X}_k)^T \quad \dots (4)$$

$$\boldsymbol{\pi_k} = \frac{n_k}{n} \quad \dots (5)$$

where the training set's total sample count is n. The number of categories is K. The number of samples in Categ k is known as n k. The PCA-LDA and PCA-QDA detections in Equation 2-1 will be calculated using the unfolded PCA findings, and the calculation process will be the same in both cases..

8. Support Vectors Machines Method

The researchers (Hall, Marron and Park, 1992)) provided the 3D smooth junction (SCV) method with the best smoothest h-independent experimental method (obtained by the program). Then, the researchers introduced the smoothest SCV for multiple variables (Sain, Baggerly, and Scott, 1994), and he initially worked with it using a slightly altered version of the Leave-in-diagonal formula, or LSCV technique, in equation (40-2) to obtain with data samples devoid of duplicate values.

$$Lscv\ (H) = n^{-1}\ (4\pi)^{-d/2}\ |H|^{-1/2} + n^{-2} \sum_{i=1}^{n}\sum_{j=1}^{n} (\ K_{2H} - 2K_H + K_0\ )\ (X_i - X_j\ ) \quad (13)$$

where K0 → is the Dirac delta function

to form (SCV) before smoothing the data differences (Xi – Xj) by K2G, i.e. replacing (Xi – Xj) by winding with K2G (Xi – Xj):

$$Xj:(\ SCV(H) = n^{-1}\ (4\pi)^{-d/2}\ |H|^{-1/2} + n^{-2} \sum_{i=1}^{n}\sum_{j=1}^{n} (\ K_{2H+\ 2G} - 2K_{H\ +\ 2G} + K_{2G})(X_i - X_j\ ) \quad (14)$$

With the exception of step (1), which makes use of the band width matrix (H) obtained using the method (KDA SCV), the alegorithm for this approach is the same as the algorithm for the previous method (Gaonkar, B.& Davatzikos, C. , 2013)

9. PLS - DA . Penal Least Squares Discriminant Analysis Method

The penal least squares method was presented in the context of regression and then it was inherited and used as a method of Categification in the context of discriminant analysis and was then called the discriminant analysis of partial least squares. (Loong et al. 2018):

$$S_x S_y = H^* \quad (15)$$

$$H^* = \frac{1}{(n-1)^2}\sum_{i=1}^{g} n_i^2\left(\overline{X}_i - \overline{X}\right)\left(\overline{X}_i - \overline{X}\right)' \quad (16)$$

$$[corr(a'x, b'y]^{\,2}\ var(ax') = \frac{[cov(a'x, b'y]^{\,2}}{var(b'y)} \quad (17)$$

Suppose that $A = [a_1, a_2, \dots\dots, a_k]$ is a vector of parameters pxk, to get these parameters by:

$$\underset{\substack{a \in R^p \\ b \in R^q \\ a'x \in 0'}}{\operatorname{argmax}}\left[\frac{[cov(a'x, b'y)]^2}{var(b'y)}\right] = [a_{k+1}, b_{k+1}] \quad (18)$$

10. Evaluation and Testing of Methods: (David, 2010) (Camilo L. et al., 2020)

The evaluation of the used traditional and three-dimensional models is done through the following formulas:

Accuracy: It is the total number of correctly Categified samples, taking into account the true negatives and errors.

Sensitivity: This is the percentage of correctly Categified positives.

Specificity: The percentage of negatives Categified correctly.

$$\mathrm{Acc}(\%) = \left(\frac{T.l+T.N}{T.l+F.l+T.N+F.N}\right) \times 100$$

$$\mathrm{Sen}(\%) = \left(\frac{T.l}{T.P+F.N}\right) \times 100$$

$$\mathrm{Spec}(\%) = \left(\frac{T.N}{F.l+T.N}\right) \times 100$$

So, Tl: represents the results that are Categorized as positive. TN: represents the real (realistic) negatives. Fl: results Categorized as negative.

FN: Failed negatives. In addition, the noise arrays contain the correct Categification ratio.

11. Simulation:

Normal distribution data with size n=2500 were generated. The characteristics of the spectral image (Categes) were distinguished in two ways: (1) and (2) Table (1) included the ratio of noise matrices at training and cross-validation groups and testing in the usual and three-dimensional methods used in the analysis of the spectral image, as it was found that the method of the main compounds - The quadratic discriminant analysis is superior to the standard approaches, which are the principal component method and the three-dimensional linear discriminant analysis. The ordinary linear discriminant analysis and the principal component method - the ordinary quadratic discriminant analysis and the method of least squares - the ordinary discriminant analysis and the SVM method, as shown in the tables (2) Which clarified the criteria for the accuracy of the methods used in the simulation, as it showed that the main components method - three-dimensional linear discriminant analysis and the main components method - quadratic discriminant analysis gave the most standard of accuracy, more sensitivity and more specification in the analysis of the three-dimensional spectral image compared to the rest of the traditional methods. As shown in Figure (1), which represents the Categification of the spectral image data, if the form (a) represents the Categification of the data according to the three-dimensional main component method, if the data is Categified from the first Categ (56.68%) and from the second Categ are (88.67).

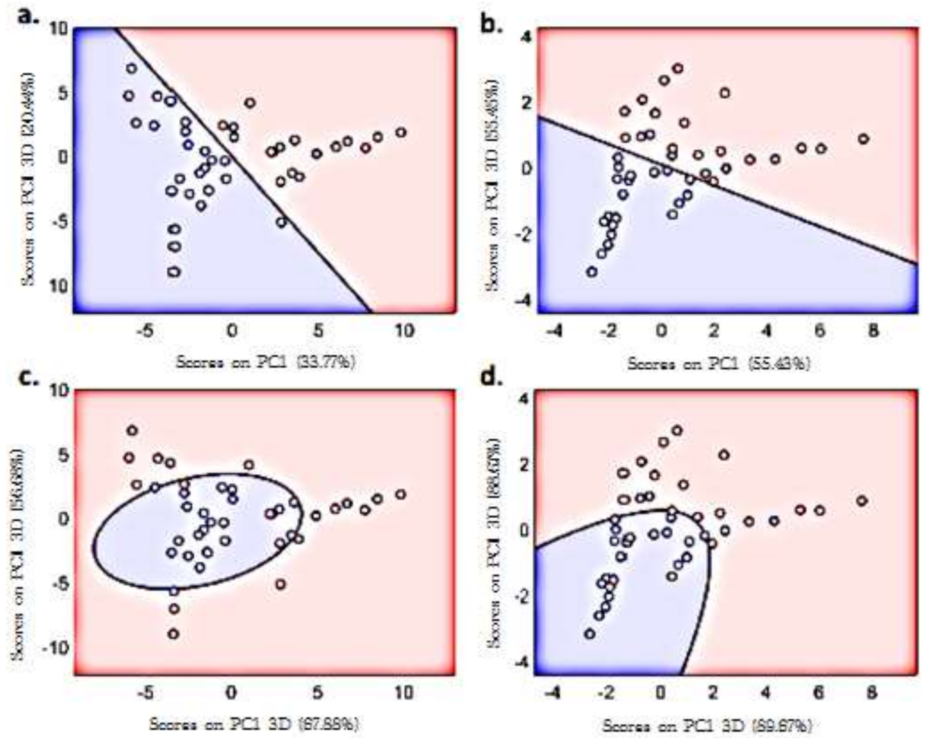

Figure (1) shows the Categification ratios according to the usual and three-dimensional methods

Table (1) Percentage of Noise Matrices at Training and Cross-validation Groups and Testing in Standard and 3D Methods Used in Spectral Image Analysis

Method	Categ	Training		Cross – Validation		Test	
Traditional Methods							
PCA-LDA		Categ 1	Categ 2	Categ 1	Categ 2	Categ 1	Categ 2
	Categ 1	66%	31%	51%	47%	100%	1%
	Categ 2	53%	51%	49%	53%	66%	31%
PCA-QDA							
	Categ 1	73%	30%	72%	30%	88%	21%
	Categ 2	45%	55%	49%	51%	31%	55%
PLS-DA							
	Categ 1	85%	12%	82%	19%	88%	21%
	Categ 2	41%	88%	23%	77%	21%	76%
SVM							
	Categ 1	93%	8%	66%	23%	76%	13%
	Categ 2	1%	99%	31%	66%	12%	78%
3D Methods							
3D-PCA-LDA							
	Categ 1	72%	28%	72%	30%	79%	21%
	Categ 2	23%	77%	25%	77%	10%	90%
3D-PCA-QDA							
	Categ 1	82%	20%	77%	25%	79%	23%
	Categ 2	22%	78%	30%	27%	12%	90%

Table (2) Accuracy standards for the usual and three-dimensional methods used in the analysis of the spectral image

Method	Accuracy	Sensitivity	Specificity
Traditional			
PCA-LDA	64%	31%	100%
PCA-QDA	68%	55%	77%
PLS-DA	71%	68%	77%
SVM	71%	79%	62%
3D			
3D-PCA-LDA	84%	90%	80%
3D-PCA-QDA	84%	90%	80%

References:

1. Auburn, Luke; Rochester Institute of Technology. "Scientists develop spectral imaging techniques to help museums with conservation efforts". Phys.org.
2. Camilo L. M. Morais1*, Panagiotis Giamougiannis2, Rita Grabowska1, Nicholas J. Wood2, Pierre L. Martin-Hirsch2, Francis L. Martin3*, (2022), " A three-dimensional discriminant analysis approach for hyperspectral images", *Analyst*, 2020, DOI: 10.1039/D0AN01328E.
3. David, DE; Lynne, AM; Han, J; Foley, SL (2010). "Evaluation of virulence factor profiling in the characterization of veterinary Escherichia coli isolates". Appl Environ Microbiol. 76 (22): 7509–7513. Bibcode:2010ApEnM..76.7509D. doi:10.1128/aem.00726-10. PMC 2976202. PMID 20889790.
4. Gaonkar, B.; Davatzikos, C. (2013). "Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and Categification".
5. HAASDONK, B. and PEKALSKA, E. (2010): Indefinite kernel Fisher discriminant. In Proc. of ICPR 2008, International Conference on Pattern Recognition.
6. Loong Chuen Lee,*a,b Choong-Yeun Liong **b and Abdul Aziz Jemain b, (2018), " Partial Least Squares-Discriminant Analysis (PLS-DA) for Categification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps", , Analyst, 2018, DOI: 10.1039/C8AN00599K.
7. Martinez, A. M.; Kak, A. C. (2001). "PCA versus LDA" (PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23 (2): 228–233. doi:10.1109/34.908974.
8. McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience. ISBN 978-0-471-69115-0. MR 1190469.