International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Reduction of High Dimensional Data Using Discriminant Analysis Methods

Aigbokhan E. E.[1], Alakiri H. O.[1], Lawal O. N.[1]
[1]Dept. of Computer Technology, Yaba College of Technology, Yaba -Lagos, Nigeria
Email address: [1]enaseddy@gmail.com, [1]halakiri@yahoo.com, [1]kingwales@gmail.com

Abstract— *In recent years, analysis of high dimensional data for several applications such as content based retrieval, speech signals, fMRI scans, electrocardiogram signal analysis, multimedia retrieval, market based applications etc. has become a major problem. To overcome this challenge, dimensionality reduction techniques, which enable high dimensional data to be represented in a low dimensional space have been developed and deployed for varieties of application to fast track the study of the information structure. In this paper, a comparative study of LDA and a KDA among the dimensionality reduction techniques were considered using data samples collected from survey and it was implemented using object oriented programming language (C#). The results reveal that less data components were discovered by LDA across the different dataset tested in comparison with KDA.*

Keywords— *High dimensional data, Dimensionality reduction, sample size, linear and non-linear techniques, LDA, KDA.*

I. INTRODUCTION

Dimensionality decrease is a procedure of removing the basic data from the dataset of high-dimensional information by speaking to them in a more dense type of much lower dimensionality to improve the order precision and to diminish computational complexity. Dimensionality decrease has turned into a practical procedure of giving powerful information portrayal moderately low-dimensional space for some, applications like electrocardiogram flag investigation and substance base recovery [19]. In K.I. [13], "this high dimensional information decrease strategy is vital in several parts, as it guide against the expense of high dimension as well as other undesired properties of high-dimensional spaces by encouraging among others grouping, representation, and pressure of high-dimensional information". True information, for example, discourse signals, advanced photos, or fMRI checks, for the most part has a high dimensionality. To visualize and understands data properly, its high-dimension should be decreased as much as possible to low-dimension of 2-D or 3-D for better interpretation of its content features. In real world scenerio according to [12], "the decreased portrayal ought to have a dimensionality, which can be likens to the inherent data dimension and the innate data dimension is the base number of parameters expected to represent the watched properties of the data".

Customarily, dimensionality decrease was performed utilizing direct methods, for example, Principal Components Analysis (PCA, LDA and so forth). Nonetheless, these procedures don't satisfactorily deal with complex non-straight information. In the most recent decade, countless methods for dimensionality decrease have been proposed. Specifically for genuine information, these nonlinear dimensionality decrease systems may offer favorable position, since true information is probably going to be very nonlinear. Past examinations have appeared nonlinear procedures beat their linear equivalents on difficult counterfeit undertakings. For example, the Swiss roll dataset includes a lot of focuses that lie on a winding like 2-D complex inside a 3-D space. Countless strategies are flawlessly ready to discover this implanting, while linear methods neglect to do as such. As opposed to these triumphs on simulated datasets, effective utilizations of nonlinear dimensionality decrease systems on regular datasets are rare. Past this perception, it isn't obvious to what degree the exhibitions of the different dimensionality decrease methods vary on artificial and characteristic undertakings [16] however this examination is restricted. The aim of this paper is to explore these exhibitions of test datasets. This paper utilizes information tests gathered from study and it was executed utilizing object oriented programming language (C#). The information was fundamentally broke down utilizing Linear dimensional procedures (LDA) against non-Linear dimensional system (KDA).

II. LITERATURE REVIEW

High-dimensional data are hard to understand but easy to interprete and visualize when the original data is fitted in an embedded manifold of 2-D or 3-D inside a higher-dimensional space [5][7][23][18][9]. The problem of high dimensional data reduction is easily stated as follow: Suppose we have K-dimensional Manifold, M embedded in V - dimensional space, with K< V and if there exist differentiable capacity function f that have the rank k, then we can define an embedding capacity function as: $: R_k : \to \Re_v$. Generally $\Re_k$ stands for input-data space and R_v stands for the hidden space. Given a lot of n observations $x = x_1, x_2, \ldots, x_n$ of the v - dimensional arbitrary vector $x = (x_1, x_2, \ldots, x_n)^T$, the dimensionality decrease signifies the estimation of the unknown lower k– dimensional vector, $i = 1,2,\ldots,n$ to such an extent that $x_i = f(y_i) + \varepsilon_i$, with ε indicating the noisy component. The vector k is regularly known as hidden variable (part). The example implies that the VxV test covariance network of the variable are indicated by:

$$\mu = (\mu_1, \mu_2, \ldots, \mu_m)^T \tag{1}$$

and

$$\sum = E = \left\{ (x - \mu)(x - \mu^T) \right\} \tag{2}$$

Aigbokhan E. E., Alakiri H. O., and Lawal O. N., "Reduction of High Dimensional Data Using Discriminant Analysis Methods," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 2, pp. 214-219, 2019.

where *E* stand for the expectation. Assuming that the data are stored in the columns of the *v x n* matrix x, the dimensionality reduction techniques attempt to decompose the data into a *VxK* matrix B where each column of the input data x is a linear (or nonlinear) combination of the elements in B, such that,

$$X = DY \quad (3)$$

where the matrix Y of size *k*x*n* convey in its columns the new *n k* - dimensional data variables (vectors).The linear dimensionality reduction techniques can only retrieve the linear structure of the subspace. In this case, each latent component $\mathbf{y}_i$ is a linear combination of original vector $\mathbf{x}_i$

$$y_{ji} = \sum_{o=1}^{m} a_{jo} x_{oi} \quad (4)$$

With *i* = 1,….,n , j = 1,…….,*k*, and *0* = 1,….,v . In a matrix form this is written:

$$Y = CX \quad (5)$$

Where, A is a weight matrix for linear transformation. Clearly, $C = D^{-1}$. In any case, as a rule, the suspicion that the information space can be spoken to as straight mix of the element subspaces does not generally represent anticipated outcomes from this present reality situations. "For example, converting an object on a uniform context can't epitomized as a linear function of the picture pixels." For example, it is impossible to obtain a linear function representation of image pixels for any object translated on an identical background Hence, suitable subspace can be empitomized increasingly by performing a nonlinear decay of the information space

2.1 Dimensionality Reduction Techniques

Dimensionality decrease diminishes the quantity of factors which improves the execution of the grouping. Handling of the high dimensional information drives the expansion of unpredictability both in execution time and memory utilization [20]. There are number of procedures accessible to lessen the elements of the dataset. Every single method decreases the components of the information dependent on specific criteria. As of late, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are viewed as vital and integral assets of dimensionality decrease for extricating powerful highlights of high-dimensional vectors in information. Contingent upon the information, the decrease strategies are named straight systems and non-linear methods [2][10][14]. In following areas we manage these distinctive procedures. For the most part, there are two kinds of information like direct information and non-linear information.

2.1.1 Linear Dimensionality Reduction Techniques

These data reduction techniques that linearly transform data from high-dimensional space to a lower-dimensional space representation, whose the variance of the original data is maximize. Though, a number of strategies accessible to deal with this sort of linear mapping of data but the two prominent techniques implemented are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) used to map to a linear subspace of lower dimensionality.

a. *PCA:* One of the most seasoned straight dimensionality methods is an outstanding PCA [24][4]. Its straightforwardness and pertinent to vast territory of uses positioned the PCA (otherwise called Karhunen-Loeve change from the correspondence hypothesis) as a standout amongst the most prevalent methodologies, in spite of its weaknesses. PCA is a system that straightly changes the information by finding a lot of p eigenvectors that represents the most extreme information's difference. The principal segment indicates path with the greatest variance. Given the orthogonality ($B^{-1} = B^{T}$) of the principal component (PC), it is possible to obtain the hidden factor y_1 from the original factor when projected onto "weight matrix $y_1 = B^{T}(x_1 - \mu)$ using equation (3) and (5) and this is defined by:

$$b = \frac{arg\max}{\|b\| - 1} \; E\{(b^T x)^2\} \ldots \quad (6)$$

The PCs are contained in the column segments of the matrix B. The second column segment lies in the subspace opposite to the primary, the third column segment obliges the greatest fluctuation bearing in the subspace opposite to the initial two, etc. The symmetrically guideline between the eigen-vectors is the main requirement forced in communicating the eigen-vectors. The eigen-vectors bj (PCs) and the relating eigen esteems λj can be figured by explaining the condition:

$$\sum b_j = \lambda_j b_j, j = 1, \ldots p \quad (7)$$

Or consistently through the characteristic equation given by:

$$|\Sigma| - \lambda I = 0 \quad (8)$$

where I is the identity matrix and the indicates the trademark or coefficient determinant of the matrix. By requesting the eigen-vectors in the request of diving eigen esteems (biggest initial), one can make an arranged symmetrical premise with the first eigen-vector having the course of biggest fluctuation of the information. Thusly, we can discover headings in which the informational index has the most huge measures of vitality. PCA ideally limits remaking mistake under the L2 standard (Euclidean separation) as objective function:

$$\Sigma\|X - BY\|^2 \quad (9)$$

Once the eigen-vectors are revamped by the diminishing estimation of their eigen values, an issue is given by the number p of eigen-vectors (PCs) to be held. The quantity of PCs to be kept depends on the usage and typical method to choose a proper number of PCs is given by registering the overall vitality extent of the fluctuation contained by the main p PCs.

$$\sum_{j=1}^{p} \frac{\lambda_j}{trace(\Sigma)} \quad (10)$$

Next, the quantity of PCs is chosen with the goal that the total vitality is over a specific edge, for example, 95 %. The remainder of the eigenvectors having low vitality are disposed of. A great number of utilizations exist for PCA from various logical fields, for example, Statistics, Computer Vision, Image Processing, Pattern Recognition, Chemistry, Astronomy, and so forth. Despite the fact that the regular PCA is performed in numerous applications, it has a few issues. The primary issue of PCA is that the Mean Square Error (MSE) is commanded with the huge number of mistakes. PCA dependent on L2-standard winds up touchy to anomalies. To defeat this issue PCA dependent on L1-standard is proposed to improves the power [1][15].

Aigbokhan E. E., Alakiri H. O., and Lawal O. N., "Reduction of High Dimensional Data Using Discriminant Analysis Methods," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 2, pp. 214-219, 2019.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

b. *LDA:* This is an organized dimensionality reduction technique. The feature selection in traditional LDA is obtained by exploiting the difference between classes and reducing the distance within classes [6]. For better separation, the high dimensional space is reduced into low dimensional subspace. Given Q distinctive classes, n_c *as* number of samples in class Q, c=1,….,Q and $n = \sum_{c=1}^{Q} n_c$, is the overall amount of "coefficient vectors". Then the average "coefficient vector" for class is $\mu_c = \frac{1}{n_c}\sum_{i=1}^{n_c} x_{cl}$ while overall average "coefficient vector" is $\mu = \frac{1}{n}\sum_{c=1}^{Q}\sum_{i=1}^{n_c} x_{cl}$. Thus, the distribution of samples of the same class about their respective mean is expected to be as small as possible and this visible from the "within-class scatter matrix" given by $S_w = \sum_{l=1}^{v_c}\sum_{c=1}^{Q}(x_d - \mu_c)(x_d - \mu_c)^T$. Also each bunch is shaped from samples of the same class, which must be far away from the other bunch and this made possible through the "between-class scatter matrix" given by $S_b = \sum_{c=1}^{Q}(\mu_s - \mu)(\mu_s - \mu)^T$. By carefully increasing between scatter matrix and decreasing the within scatter matrix, the LDA is able to projects the sample variables at the same time into a subspace. The objective function is defined as:

$$b = \underset{B}{\arg max} \frac{\|B^T S_b B\|}{\|B^T S_w B\|} \tag{11}$$

The PCA is usually apply prior to LDA, whenever the "within-class scatter matrix" Sw is singular for n<<v and this defined by:

$$B_{LDA} = \underset{B}{\arg max} \frac{\|B^T B_{PCA}^T S_b B_{PCA} B\|}{\|B^T B_{PCA}^T S_w B_{PCA} B\|} \tag{12}$$

is calculated, and finally, the optimum LDA projection direction is given by $B_{opt} = B_{LDA}.B_{PCA}$. LDA has been utilized broadly in numerous applications, for example, picture recovery (Swet and Wend, 1996) microarray information characterization [11], speaker acknowledgment [17], and so forth. The LDA [26] generally utilizes the worldwide structure data of the all out preparing tests to decide the straight discriminant vectors and these vectors are altogether worldwide. For a test, the utilization of worldwide straight discriminant vectors to separate highlights from the examples may prompt wrong arrangement, though the utilization of nearby direct discriminant vectors may create right order. At the point when the worldwide information structure isn't totally predictable with the nearby information structure, Local LDA (LLDA) is more dominant than the customary LDA calculations and LLDA can successfully catch the neighborhood structure of tests.

2.1.2 Non-Linear Dimensional Reduction Techniques

These are systems utilized in decreasing complex multi-dimensional nonlinear information. Nonlinear methods beat their straight partners on complex counterfeit assignments [3]. For example, the Swiss roll dataset includes many of focuses that falls within a winding like 2-D complex inside a 3-D space. Countless direct strategies are consummately ready to discover this inserting, though straight systems neglect to do as such. As opposed to these victories on counterfeit datasets, effective uses of nonlinear dimensionality decrease methods on regular datasets are rare.

a. The Kernel PCA (KPCA): The KPCA depends on interpreting the first information xi to a higher-dimensional space. A primary look at the transformation of the first information to a high-dimension space is by all accounts in opposition to the dimensionality decrease standard. Nonetheless, by utilizing a similar investigation with respect to the straight PCA and utilizing the purported "piece trap" [22] the KPCA calculation can be carried out within a low-dimensional space. Recall that, the info space where the first data live is transformed by a nonlinear function onto an hidden space, $f: R^k : \rightarrow \Re^v, v > k$. The viability of this approach stem from the fact that data are easily separated in high-dimensional space compare low-dimensional space and this is justifiable from "statistical pattern recognition field", thus fast-tracking performance in classification process. In view of the above, a new covariance matrix can be obtain by incorporating the map ϕ parameter into the sample covariance matrix equation (2) as:

$$\Sigma_{KPCA} = \mathrm{E}\{\varphi(x-\mu)\varphi(x-\mu^T)\} \tag{13}$$

The eigenvectors $\mathbf{b}_j$ and the corresponding eigen values λ_j can be derived the equation:

$$\Sigma_{kpca} b_j = \lambda_j b_j, j = 1, \dots p \tag{14}$$

Given a K x k matrix O as kernel operator:

$$O_{ij} = \left(\varphi(x_i).\varphi(x_j)\right) \tag{15}$$

Thus, we compute the eigenvectors and eigen values of the dot product matrix O using equation (3) as follows:

$$n\lambda\alpha = O\alpha \tag{16}$$

Hence, nonlinear PCs associated to ϕ will be calculated from normalized eigenvectors and eigen values, leading to a new test variable, which gets projected onto the eigenvectors in order to reduce its dimension:

$$\left(B^k \varphi(x_i)\right) = \sum_{i=1}^{n} \alpha_i^k \left(\varphi(x_i).\varphi(x)\right) \tag{17}$$

b. KDA: So as to make LDA relevant to nonlinearly organized information, part based strategies have been connected to outline input information to a component space by a nonlinear mapping where internal items in the element space can be figured by a bit capacity without knowing the nonlinear mapping unequivocally which makes highlight space frequently turns out to be a lot bigger than that of the first information space, and therefore, the disperse networks become solitary. A non-linear dimension reduction can be achieved from LDA by utilizing a similar kernel theory implemented in KPCA. The KDA also known as "Kernel Fisher Discriminant (KFD)" [21] operates by calculating an objective capacity function for new component space as follows:

$$B^{\varphi} = \underset{B}{\arg max} \frac{\|B^T S_b^{\Phi} B\|}{\|B^T S_w^{\Phi} B\|} \tag{18}$$

Where, ϕ epitomize the nonlinear mapping.
In case of a two-class problem,

Aigbokhan E. E., Alakiri H. O., and Lawal O. N., "Reduction of High Dimensional Data Using Discriminant Analysis Methods," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 2, pp. 214-219, 2019.

$$S_w^{\phi} = \sum_{c=1}^{Q} \sum_{l=1}^{n_c} \left(\varphi(x_{cl}) - \mu_c^{\varphi}\right)\left(\varphi(x_{cl}) - \mu_c^{\varphi}\right)^T \quad (19)$$

And

$$S_w^{\phi} = \sum_{c=1}^{Q} \left(\mu_c^{\phi} - \mu_c^{\varphi}\right)\left(\mu_c^{\phi} - \mu_c^{\varphi}\right)^T \quad (20)$$

where $Q = \{Q_1 Q_2\}$ and $n_c \epsilon \{n_{c1}, n_{c2}\}$;

Then the Fisher's linear discriminant cost function is expressed by:

$$\alpha = \arg max_{\alpha} \frac{\alpha^T M \alpha}{\alpha^T N \alpha} \quad (21)$$

where the notations are defined as:

$$M = (M_1 - M_2)(M_1 - M_2)^T$$

$$(M_i)_Q = {}^{1}\!/_{Q} \sum_{k-1}^{text/Q} k\,(X_Q . X_k^i)$$

$$N = \sum_{Q-1,2} K_Q\,(I - 1_n) K_Q^T,$$

where K_Q represents kernel matrix as a typical application for the class;

The KFD had been used in pattern recognition and has major advantage of KDA is that it can been applied irrespective of singularity of the scatter matrices equally in the original space and hidden space by a nonlinear transformation. The KDA is an effective dimension reduction technique for multi-class problems as proven from the comparison with other techniques of solving the generalized eigenvalue problem shows.

b. *Kernel Independent Component Analysis (KICA):* [8] proposed the KICA as a replacement for the classical ICA whose components are mixed with nonlinear functions but also apply the "kernel Hilbert space" to extract data from sources which have been mixed nonlinearly. In view of this, two different functions: the kernel ICA-KCCA and the ICA-KGV have been defined in this reproducing space and they based on canonical correlations. According to [8] "the Kernel ICA-KCCA reduces the first kernel canonical correlation that depends on the data x_i, i =1,….,n only through the centered Gram matrices for i *iCs.* Also, the Kernel ICA-KGV minimizes the kernel generalized variance. Both functions are connected to a global eigenvector problem $k_k \alpha = \lambda D_k \alpha$, where k is a regularization parameter and $K\kappa$ and $D\kappa$ are block matrices built from the Gram matrices".

III. METHODOLOGY

The goal of this section is to implement suitable algorithms that tried to decrease the component of data by removing irrelevant attributes in the data for this analysis. The research approach used in this study is based on *linear (LDA) and non-linear (KDA) algorithms* for the analysis of two high data dimensional statistical data - *Cigar and Gutierrez-Osuna datasets* obtained from online repositories, specifically the UCI repository. Analyzing and computing all these high dimensional data - *Cigar and Gutierrez-Osuna datasets* is very difficult because, they each consists of many attributes which may not all affect the result of the classification. The two dimensionality reduction algorithms – LDA and KDA simulated into a C# program and was run several times to test the applicability of the algorithms in solving reduction problems and preferences were drawn from result. Though removing of attributes from the data takes less execution time, there may be a loss of data which may affect the classification. Below are the details of our selected datasets used.

TABLE 1: Components Analysis

Data Set	No of variables	LDA No of components	KDA No of components
Cigars	30	2	120
Gutierrez-Osuna	10	2	10

3.1 Experimental Result & Analysis

In this study, the dimensionalities of - *Cigar and Gutierrez-Osuna datasets* were reduced through LDA as well as KDA and the performance of each was measured in terms of Components, Run time and Projection. To assess the performance of LDA and KDA algorithms, the developed C# program was run 10 times on the two datasets of Cigars and Gutierrez-Osuna with various dimensions and the final optimized solutions were recorded along with the processed time as shown in table 2**.**

3.1.1 Execution time analysis

We present below in table 2, the execution time analysis for the Cigars Dataset and Gutierrez-Osuna Dataset

TABLE 2: Result of analysis by execution time

	Cigars Dataset		Gutierrez-Osuna Dataset	
	LDA	**KDA**	**LDA**	**KDA**
SN	**Runtime (mill secs)**	**Runtime (mill secs)**	**Runtime (mill secs)**	**Runtime (mill secs)**
1	6	10	5	6
2	7	11	4	6
3	5	10	4	5
4	7	10	5	9
5	7	11	5	6
6	8	11	5	8
7	8	12	5	6
8	5	9	6	9
9	5	9	6	9
10	6	9	6	9
Average	64	102	51	73

3.1.2 Analysis by Projection

From our result, we discovered that KDA has better projections across the various dataset tested in comparison with LDA. Below are sample projections:

a. Projection on Cigars Dataset based on LDA

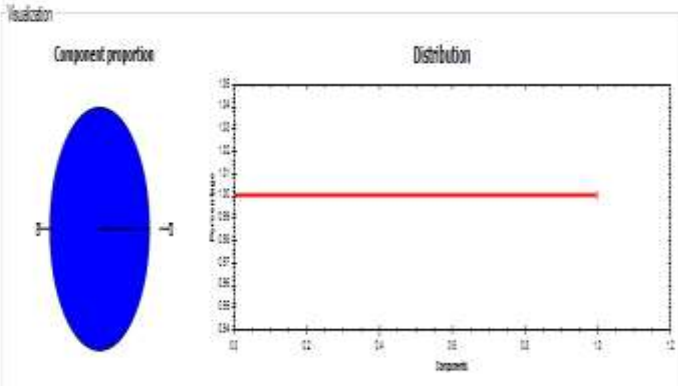

Fig. 1a: Component Visualization of Cigars Dataset based on LDA

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

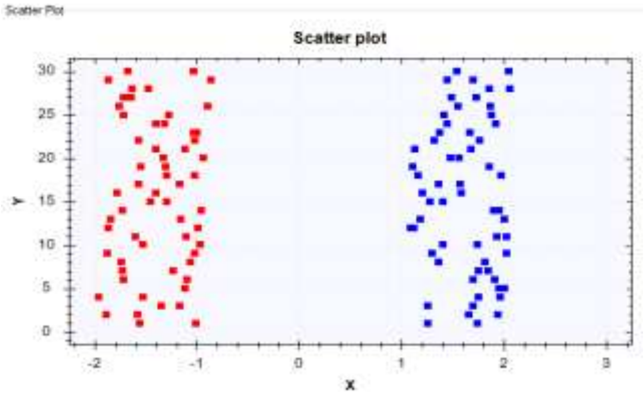

Fig. 1b: Component Scatter Plot of Cigars Dataset based on LDA

Projection on Cigars Dataset based on KDA

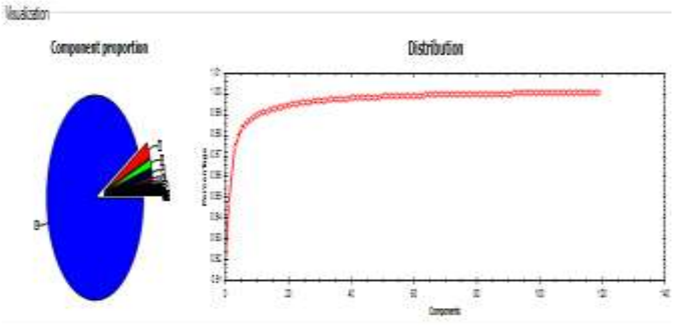

Fig. 2a: Component Visualization of Cigars Dataset based on KDA

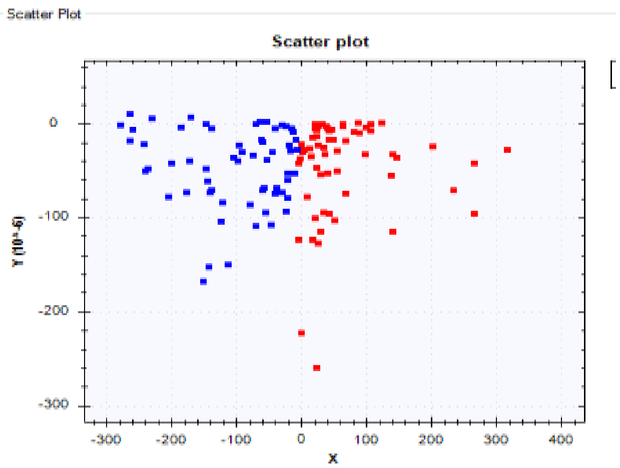

Fig. 2b: Component Scatter Plot of Cigars Dataset based on KDA

b. *Projection on Gutierrez-Osuna Dataset LDA*

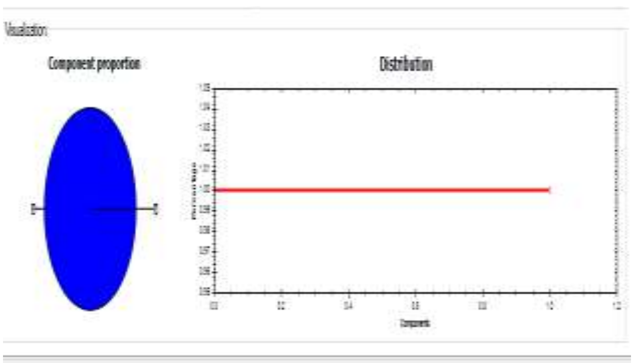

Fig. 3a: Component Visualization of Cigars Dataset based on LDA

Projection on Gutierrez-Osuna Dataset KDA

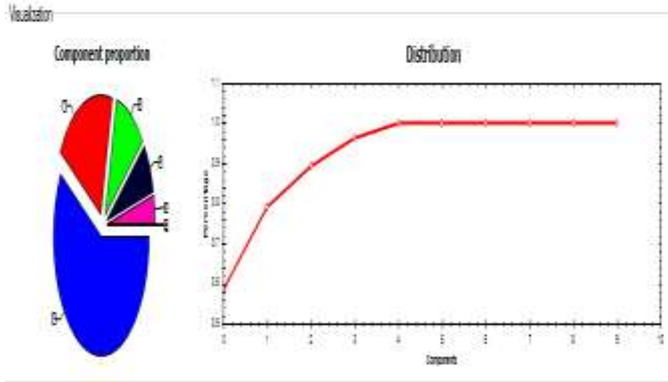

Fig. 3b: Component Visualization of Cigars Dataset based on KDA

IV. Discussion

The result above shown the execution time and projection analysis of *Cigar and Gutierrez-Osuna datasets* whose dimensionalities were reduced via LDA and KDA algorithms. The performance of each algorithm was measured in terms of Components, Run time and Projection. For both the execution time and projection result analysis, it is easily observed that LDA executes faster than KDA and it is important to note that KDA discovers more data points (Eigen values and Eigen vectors) compared to LDA. For instance, LDA discovers 4 data points on Cigar dataset while KDA discovers over 4000 data points. Also, we discovered that less data components were discovered by LDA across the various dataset tested in comparison with KDA. This clearly shows KDA is a more effective dimensionality reduction technique to LDA. Though removal of attributes from the data takes less execution time, there may be a loss of data which may affect the classification as demonstrated by LDA on both datasets. The outcome got from the investigation of that datasets-Cigar and Gutierrez-Osuna demonstrated that KDA can be connected independent of peculiarity of the disperse frameworks both in the first space and in the element space by a nonlinear mapping. It additionally exhibits that KDA is a powerful measurement decrease technique for multi-class issues contrasted with different strategies for taking care of the summed up eigenvalue issue.

V. Conclusion

In this paper, we present the linear and non-linear techniques to reduce the dimensions of the initiall data. From the study, it was understood that nonlinear dimensionality decrease methods (KDA) outflank the conventional direct systems (LDA) on genuine world datasets. Less data components were found by LDA over the different dataset tried despite the fact that they execute quicker in correlation with KDA. This unmistakably demonstrates KDA is increasingly successful dimensionality decrease system, essentially on the grounds that they have better projections over the different dataset tried in examination with LDA. Most genuine world datasets are non-straight in nature, henceforth it is fitting to extricate helpful data through KDA-based system rather than LDA-based procedure. In the event that LDA-based strategies is to connected to nonlinearly organized information, at that point part based techniques will be required to delineate information to a component space by a nonlinear mapping where internal items in the element space can be registered by a piece work without knowing the nonlinear mapping expressly which makes include space regularly turns out to be a lot bigger than that of the first information space, and therefore, the dissipate frameworks become particular. Along these lines, utilizing a similar part hypothesis as in KPCA, the LDA approach can be summed up, prompting nonlinear information dimensionality decrease.

REFERENCE

[1] A. M. Martinez and A. C. Kak *"PCA versus LDA",* IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 228–233, 2001.

[2] Amir Karami, "Taming Wild High Dimensional Text Data with aFuzzy Lash", arXiv:1712.05997v1 [stat.ML], https://arxiv.org/pdf/1712.05997.pdf, 2017.

[3] Andreas Artemiou and Yuexiao Dong, *"Sufficient dimension reduction viaprincipal Lqsupport vector machine"*, Electronic Journal of Statistics Vol. 10 (2016) 783–805, 2016.

[4] Bogaardt Laurens, Goncalves Romulo, Zurita-Milla Raul, and Izquierdo-Verdiguier Emma, *"Dataset Reduction Techniques to Speed Up SVD Analyses on Big Geo-Datasets"*, International Journal of *Geo-Information,* 2018.

[5] Christopher J.C. Burges, *"Dimension Reduction: A Guided Tour",* Article *in* Foundations and Trends® in Machine Learning· DOI: 10.1561/2200000002 Source: DBLP, 2010.

[6] D. L. Swets and J.Weng, *"Using discriminant eigenfeatures for image retrieval"* ,*Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

[7] Daiho Uhm, Sunghae Jun and Seung-Joo Lee, *"A Classification Method Using Data Reduction",* International Journal of Fuzzy Logic and Intelligent Systems, vol. 12, no. 1, pp. 1-5, 2012.

[8] F. R Bach. and M. J. Jordan, "Kernel independent component analysis," *Machine Learning Research*, vol. 3, pp. 1–48,2002.

[9] Guoqing Chao, Yuan Luo and Weiping Ding, "Recent Advances in Supervised Dimension Reduction: A Survey", Machine learning and knowledge extraction, 2019.

[10] Han Shaobo and David B. Dunson, "Supervised Multiscale Dimension Reduction for Spatial Interaction Networks", arXiv:1901.00172v2 [cs.LG], https://arxiv.org/pdf/1901.00172.pdf, 2019.

[11] J. Dudoit, Fridlyand, and T. P. Speed "Comparison of discrimination methods for the classification of tumors using gene expression data", *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 77–87, 2002.

[12] K.Fukunaga, *"Introduction to Statistical Pattern Recognition",* San Diego: Academic Press, 2nd edition edition, 1990.

[13] K.I. Kim, K. Jung, and H.J. Kim *"Face recognition using kernel principal component analysis"*. IEEE Signal Processing Letters, 9(2):40–42, 2002.

[14] Li Didong, Minerva Mukhopadhyay and David B Dunson, "Efficient Manifold Approximation with Spherelets", arXiv:1706.08263v3 [stat.ML], https://arxiv.org/pdf/1706.08263.pdf, 2019.

[15] N. Kwak *"Principal component analysis based on L1-norm maximization*", IEEE Transaction on Pattern Analysis and Machine Intelligence, volume. 30, no. 9, pages.1672 -1680, 2008.

[16] N. Patwari, and A. O Hero III,. "Manifold learning algorithms for localizationin wireless sensor networks" In *Proc. of IEEE Int. Conf.on Acoust. Speech and Signal Processing*, 2004.

[17] Q. Jin and A.Waibel, *"Application of LDA to speaker recognition",* in *Proc. of International Conference on Spoken Language Processing*, vol. 2, pp. 250–253,2000.

[18] R. Rajesh, Agarwal Shikha, Ranjan Prabhat, *"Dimensionality reduction methods classical and recent trends: A survey:"* International Science Press IJCTA, 9(10), 2016, pp. 4801-4808, 2016.

[19] R.Fisher, *"The statistical utilization of multiple measurements"*, Ann.Eugenics, vol. 8, no. 4, pp. 376–386, 1938

[20] S. S. Shylaja, K. N. Balasubramanya Murthy and S. Natarajan, *"Dimensionality Reduction Techniques for Face Recognition"*, Web of Science, Intech Publisher, retrieved from www.intechopen.com, 2011.

[21] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.- R. Müller, *"Fisher discriminant analysis with kernels"*, in *Neural Networks for Signal Processing IX*, pp. 41–48, 1998.

[22] Shawe-Taylor John and Cristianini Nello, "Kernel Methods for Pattern Analysis", Journal of Computer Science, Electronic, Optoelectronic Devices, and Nanotechnology, Engineering, Pattern Recognition and Machine Learning Cambridge University Press, https://doi.org/10.1017/CBO9780511809682, 2004.

[23] Szymon Łukasik, André Moitinho, Piotr A. Kowalski, António Falcão, Rita A. Ribeiro, and Piotr Kulczycki, *"Survey of Survey of Object-Based Data ReductionTechniques in Observational Astronomy",* DeGruyter Open vol 4: Pg 579–587, 2016.

[24] T. Jolliffe, "*Principal Component Analysis"*, (2nd ed.), New York: Springer-Verlag, 2002.

[25] Teh Ying Waha, Samee U. Khan, *"Big Data Reduction Methods: A Survey"*, Data Science and Engineering, Springer ,2016

[26] Zizhu Fan; Yong Xu; Zhang, D, *"Local Linear Discriminant Analysis Framework Using Sample Neighbors"*, IEEE Transactions on Neural Networks, , On page(s):1119 - 1132 Volume: 22 ,2011.

Aigbokhan E. E., Alakiri H. O., and Lawal O. N., "Reduction of High Dimensional Data Using Discriminant Analysis Methods," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 2, pp. 214-219, 2019.