iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81828

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Decoding High-Dimensional Data: Linear Dimensionality Reduction Techniques Revisited

Daljeet Kaur Khanduja[1], Surjeet Kaur[2]
[1]Professor, Department of Mathematics, Sinhgad Academy of Engineering, Pune,Maharashtra, India
[2]Associate Professor, Department of Mathematics, SIES College of Arts, Science and Commerce (Empowered Autonomous), Sion (West), Mumbai, 400022, Maharashtra, India

Abstract: In our rapidly digitizing world, data is being collected at a never-before-seen pace from dynamic global sectors like healthcare, manufacturing, sales, IoT devices, the web, smart gadgets, social media, and organizations on a regular basis. The properties of this type of data are high dimensionality, large volume, redundant features, and noise. As such, representing and processing such a large amount of complex, heterogeneous data becomes much more challenging. In many sectors where there is a lot of data with lots of columns or classes, dimensionality reduction techniques are crucial. One essential method for evaluating and understanding high-dimensional data is the application of dimensionality reduction techniques.These methods collect a wealth of interesting data properties, including dynamical structure, correlation between data sets, covariance, and input-output interactions. This paper reviews several types of algorithms used in dimension reduction in order to provide readers with a thorough and lucid summary of the field and to give them a sense of how to assess its increasing importance over the past few years. Since linear dimensionality reduction techniques have straightforward geometric interpretations and generally appealing computing features, they are fundamental to the analysis of high dimensional data. Principal Component Analysis (PCA), Singular Value Decomposition (SVD), linear discriminant analysis(LDA), and Independent Component Analysis (ICA) are the four linear dimensionality reduction approaches that will be examined and compared in this study. The purpose of this study is to examine and contrast the benefits and drawbacks of dimensionality reduction in several widely applied mathematical concepts and techniques.

Keywords: Dimensionality reduction, Principal component analysis, Singular value decomposition, linear discriminant analysis, Independent component analysis.

I. INTRODUCTION

Today, in the digital age, almost every industry—including social media, healthcare, finance, science, user behaviour in social networks, patient data in the medical field, transaction records in the financial market, posts that go viral on social networking sites like Facebook, Twitter, and Instagram, mobile phones, multimedia, etc.—is producing an unprecedented amount of data in real time within fractions of seconds. As a result, data generation is expanding more quickly than it has in the past. High dimensional data is the term used to describe the enormous volume of big data generated at such a fast rate. Since each form of data also contains a variety of traits and attributes, high dimensional data has several dimensions. The data science and artificial intelligence fields are faced with numerous challenges as a result of the rapidly accumulating high-dimensional data[1], but they are also bound to be the source of new theoretical advancements [2]. The unprocessed raw data is frequently dispersed and sparsely distributed while working with high-dimensional data, which gives rise to "The Curse of High Dimensionality." It frequently contains a great deal of superfluous information, including factors that are related or repeated. Although Machine Learning Algorithms (MLAs) are capable of processing massive volumes of data [3], their effectiveness decreases with an increase in the dimensionality of the input. This data is highly heterogeneous, massive, complicated, and multidimensional; it includes structured, unstructured, and semi-structured data types. The various types of data, including text, numbers, XML, photos, videos, audio files, web pages, and more, cause the dimensionality of data to be multiple and extremely complex. Thus, managing such enormous datasets becomes extremely challenging and presents numerous difficulties with regard to data collection, storage, analysis, visualization, and high dimensionality. A challenge with high-dimensional datasets is that not every variable included is "important" for comprehending the phenomenon being studied.In order to manage the processing of high-dimensional data, more potent computing power, more effective algorithms, and more successful data reduction techniques are needed to develop efficient methods for storing and retrieving data while maintaining the important aspects of the data. Alternatively, noise and redundant data in high-dimensional data may affect the quality of prediction and decision-making as well as the accuracy of analysis and modelling [4].

Thus, one of the key areas of focus in the fields of data science and artificial intelligence is the study of effective strategies for handling high-dimensional data[5].Additionally, passing such a large number of input features through ML algorithms impacts performance, slows down processing, and increases model complexity—a phenomenon known as the "curse of dimensionality" [6]. A very effective method called dimensionality reduction can be used to lower the dimensions by transforming high-dimensional data into a meaningful interpretation of decreased dimensionality. When it comes to reporting the observed qualities of the data, the smallest number of parameters required is the fundamental dimensionality of the data. [7].

Dimensionality reduction (DR) is a technique that selects and extracts features from the existing feature space of the data that are pertinent to the application domain[8] and reduces a large number of features or random variables to a small number of features without losing any data[9] [10][11]. Dimensionality reduction can be advantageous for increasing the analysis's accuracy as well as its processing efficiency [12]. Benefits of applying dimensionality reduction methods to a dataset.

(i) Minimize the number of dimensions and data storage capacity.
(ii) The computation time is reduced.
(iii) Data that is redundant, irrelevant, and noisy can be removed.
(iv) It is possible to optimize data quality.
(v) increases accuracy and facilitates the smooth operation of an algorithm.
(vi) make data visualization possible
(vii) It boosts performance and makes classification easier[13][14].

The technique of dimension reduction has been utilized for a considerable amount of time in data visualization and low dimensional feature extraction from larger dimensions [15].

Because dimensionality reduction reduces the high-dimensional features' dimensionality and other undesirable traits, it is crucial in many different contexts [16], [17].

The paper is structured as follows: A basic introduction to linear dimensionality reduction techniques is given in Section 2. Singular Value Decomposition, Principal Component Analysis, Linear Discriminant Analysis, and Independent Component Analysis are covered in Sections 3 through 6, respectively. We examine the benefits and drawbacks of the various dimensionality reduction techniques in Section 7, and the paper is eventually concluded in Section 8.

II. LINEAR DIMENSIONALITY REDUCTION TECHNIQUES

In the fields of statistics and machine learning, linear dimensionality reduction approaches have been developed for over a century. These techniques include denoising or compressing data, identifying significant feature spaces, and showing or examining data structures. Numerous pertinent data features are captured by these methods, including input-output linkages, dynamical structure, margin between data classes, covariance, and correlation across data sets. These methods, with their attractive computational capabilities and simple geometric interpretations, are indispensable instruments for the analysis of noisy and high-dimensional data. When transforming the original high-dimensional data into a low-dimensional linear mapping, these techniques preserve a number of interesting data properties. This study surveys different linear dimensionality reduction methods while offering a thorough analysis of the benefits and drawbacks of each method by the authors whose work is being reviewed. Nonetheless, selecting the right approach for a given set of data is crucial and heavily dependent on the type of data we are working with. Depending on different characteristics, a number of approaches are examined and contrasted, including Principal Component Analysis (PCA)[18], Singular Value Decomposition (SVD)[19],Linear Discriminant Analysis (LDA)[20], and Independent Component Analysis (ICA)[21]. Less computer complexity and time are required with significantly less storage when data dimensions are lower. Less characteristics also assist in preventing overfitting [22]–[25].

III. PRINCIPAL COMPONENT ANALYSIS(PCA)

A common dimensionality reduction method for condensing big data sets of variables into smaller sets while maintaining important patterns and trends is principal component analysis. The method accomplishes this by keeping most of the information in the original set of variables while decreasing its size. Smaller data sets are easier to analyse and visualize, which allows machine learning algorithms to comprehend data points much more quickly and easily because fewer variables need to be processed. In conclusion, in order to retain the best degree of informative content, the fundamental idea behind PCA is to reduce the number of variables in a data collection. The original variables are combined or linearly mixed to form the new variables known as main components. This kind of variable combination leaves the new variables, or primary components, uncorrelated while condensing or compressing most of the information from the original variables into the initial components[26].

A. Literature Survey On PCA

In [27] and [28] the classic texts on PCA in Anglo-Saxon literature are presented. A study of PCA tools and an overview of significant turning points in the method's evolution can be found in [29]. [30] discusses the threefold character of PCA (algebraic, geometrical, and statistical).

A recent assessment of PCA and its most recent developments may be found in [31]. Principal component analysis (PCA), which is based on the variables' covariance matrix, is the most effective linear dimension reduction method in terms of mean-square error [32]. A different approach to using PCA to reduce a dataset's dimension is proposed in [33]. Instead of considering the PCs as the new variables, this method uses the data from the PCs to find important variables in the original dataset. As before, determining the number k of significant variables to retain involves first computing the PCs and then examining the scree plot. After removing the variable with the biggest (absolute value) coefficient from that vector, the eigenvector corresponding to the least significant principal component—the one with the smallest eigenvalue—is taken into consideration This process is repeated until k variables remain. In order to perform dimensionality reduction, principal component analysis [34]–[37], an unsupervised linear approach, embeds the data into a linear subspace of a decreased dimensionality. By using PCA, as much of the variation as is practical is captured in a reduced dimension representation of the original data. This can be done by finding the largest variance in the data and using it to create a linear basis for the data that is reduced in dimensionality and possibly orthogonal. In mathematical words, PCA seeks to maximize as much of the variance as possible by finding a linear mapping P that optimizes the cost function $tr\left(P^T AP\right)$, where A is the zero-mean data sample covariance matrix. In other words, PCA maximizes $\left(P^T AP\right)$ with regard to P, while ensuring that each column v of P has a norm of 1,i.e. $\|v\|^2 = 1$

In [38], another way to think of PCA is as a probabilistic latent variable model. This model employs a linear Gaussian noise model in conjunction with a Gaussian prior over the latent space. A potentially more computationally efficient EM-algorithm is produced by the probabilistic PCA formulation for exceptionally high-dimensional data. Using Gaussian processes, probabilistic PCA can be utilized to find nonlinear mappings between high- and low-dimensional spaces[39]. Minor components, or the eigenvectors corresponding to the smallest eigenvalues, are also included in the linear mapping in a subsequent PCA adjustment because they may be important in classification scenarios[40]. PCA has been effectively used in several fields, such as facial recognition [41] ,coin classification[42], seismological series analysis[43] etc.

In [8],[44] , Principal Component Analysis (PCA), which views dataset A as a collection of points in a multidimensional space as a m x n matrix where the variables are represented by m rows and the observations by n columns.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots. & \dots.. & \dots. & \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \dots.. \\ A_m \end{bmatrix} \qquad \text{Eq(1)}$$

The matrix A must be linearly transformed into the $B_{m\times n}$ matrix using a $Q_{m\times m}$ matrix containing $q_1, q_2, \dots q_m$ as the row vectors and $a_1, a_2, \dots. a_n$ as the column vectors of the A matrix

$$B_{m\times n} = Q_{m\times m}\ A = \begin{bmatrix} q_1 \cdot a_1 & q_1 \cdot a_2 & \dots.. & q_1 \cdot a_n \\ q_2 \cdot a_2 & q_2 \cdot a_2 & \dots\dots & q_2 \cdot a_n \\ \dots\dots\dots & \dots\dots\dots & \dots\dots & \dots\dots. \\ q_m \cdot a_1 & q_m \cdot a_2 & \dots.. & q_m \cdot a_n \end{bmatrix} \qquad \text{Eq(2)}$$

The dot product $q_i \cdot a_j$ shows the projection of the original dataset A onto the columns of Q.

The principal component directions are represented by the rows of Q, which serve as a new basis.

In [44], PCA is a multivariate statistical technique that enhances computational speed and accuracy by utilizing an orthogonal transformation. When examining the associations between a set of variables, PCA aims to describe as much variance as feasible with the fewest available variables and to take the most important information out of the data and present it as a collection of additional orthogonal variables known as main components.

In [45],the most popular unsupervised technique for addressing the issue of data dimensionality called Principal Component Analysis (PCA) is employed. By maximizing the association between the data points, PCA projections are carried out. The low-dimensional space provided by the projections allows for the accurate reflection of all data points without information loss. Below is the PCA equation.

$$\hat{U} = \arg\max Tr\left(U^T Cov(X)U\right) \qquad \text{Eq(3)}$$

$$U^T U = 1 \qquad \text{Eq(4)}$$

where the linear transformation matrix $\hat{U} \in R^{d \times D}$ is represented by $\hat{U}$.

In [46], Apart from the linearity assumptions, principal component analysis (PCA) indicates that a structure of interest is represented by primary components with higher correlated variances, whereas smaller variances are indicative of noise. PCA is reversible using Eigen decomposition techniques because it also assumes that the primary components are in orthogonal form.

A new technique was introduced in [47] to produce lower-dimensional word embeddings by efficiently combining the reduction of PCA-based dimensionality with a previously proposed post-processing algorithm. This approach reduced the dimensionality of the embedding by 50%, resulting in an equivalent or (more often) superior efficiency than the higher dimension embedding, according to empirical evaluations on 12 typical word similarity benchmarks.

In [48] a quick and efficient method for learning the low-dimensional key characteristics of hyperspectral images (HSIs) called super pixels (Super PCA) was proposed . The three data sets demonstrated that the (Super PCA) method was noticeably better at reducing the dimensions based on PCA for classification than the conventional baselines.In [49] Principal Component Analysis (PCA) is a commonly employed unsupervised dimensionality reduction technique that addresses the high dimensionality of time-series data. It is less expensive computationally and uses a linear transformation technique [50]. This method, which is non-iterative, eliminates correlated features and lessens overfitting issues [51]. In [52], the dimensionality reduction technique PCA successfully detect outliers, compress the time-series data dimension, and include more information by using the linear transformation of the Eigen Decomposition It is also possible to use this strategy with non-stationary data. Thus, it is widely believed by researchers that PCA is suitable for decreasing the number of dimensions in time-series datasets [53], [54], [55], [56].

In [57], first, PCA ascertains a dataset's greatest variance. The covariance matrix, which shows the information about the correlation of the features, is then computed using the following formula [58]:

$$Cov(g_1, g_2) = \frac{\sum\left(g_{1i} - \bar{g}_1\right)\left(g_{2i} - \bar{g}_2\right)}{N} \qquad \text{Eq(5)}$$

where, g_1 and g_2 are two random features of the input data set, and $\bar{g}_1$ and $\bar{g}_2$ are the mean of g_1 and g_2 respectively, N is the total number of data points.

Projecting the dataset is a crucial component of PCA. Following a reorientation from the original axis to the new axes indicated by the principal components, the data are then translated into the new subspace.[59].

IV. SINGULAR VALUE DECOMPOSITION (SVD)

One powerful mathematical method in the domains of data analytics and machine learning is singular value decomposition (SVD). SVD is regarded as the computational backbone of numerous data-driven programs and algorithms. SVD can be thought of as a practical application of principal component analysis. It is a linear algebra technique and among the simplest ways to handle complicated data. When using principal component analysis (PCA), SVD is utilized to break down high-dimensional data into statistically significant patterns in lower-dimensional data. More frequently, signal processing, machine learning ,linear algebra, orthogonal decomposition, dynamic mode decomposition, clustering and classifications [60], noise reduction, information extraction, and dimensionality reduction are the many different applications where the matrix factorization technique known as singular value decomposition is employed. With the use of this potent mathematical tool, one can factor a given matrix into three other matrices: a unitary matrix, a diagonal matrix, and its conjugate transpose. By breaking down a matrix into its most basic elements, it enables us to manipulate and comprehend the original matrix's structure and characteristics. Since the SVD makes it possible to accurately describe any matrix and makes it simple to remove less significant data from a matrix in order to create a low-dimensional approximation, it can be thought of as a type of dimensionality reduction.

While minimizing the number of dimensions in the data collection, the goal is to preserve as much information as possible. SVD helps address the curse of dimensionality, a common problem in machine learning where the number of features decreases algorithmic effectiveness In addition to reducing the number of dimensions in algorithms, SVD improves their efficacy by transforming the input data into a more concise and understandable representation.

B. Literature Survey on SVD

In [61], [62], a matrix M is broken down into singular values, left singular vectors, and right singular vectors using the SVD algorithm. The directions of highest variance are represented by the singular vectors, and the singular values indicate the significance of each component.

SVD splits a m x n matrix M (or a data frame with m rows and n columns) into three matrices:

Let $M = U * \sum * V^T$ where

M represents the real m x n matrix that is being decomposed, U is a column orthonormal matrix of size m x r, Σ is a diagonal matrix of size r x r with diagonal entries called the singular values, and V is a row orthonormal matrix of size n x r. The matrix M's rank is denoted by r.

The singular values of the original matrix M make up the diagonal components of Σ. The matrix Σ can be expressed as follows where the order of the singular values is $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r \geq 0$

$$\sum = \begin{bmatrix} \sigma_1 & & & \\ & & .. & \\ & \ldots & & \\ & & & \\ & & \ldots. & \\ & & & \sigma_r \end{bmatrix} \qquad \text{Eq(6)}$$

The primary matrix M is then reconstructed via SVD using the following equation (7) to a lower rank k [63]:

$$Y = U_k * \sum_{k}. * {V_k}^T \qquad \text{Eq(7)}$$

Here, Y being a low dimensional output, the dimensionality of the data can be decreased to k dimensions by choosing only the top k singular values and vectors where U_k, $\mathbf{\Sigma}_k$, V_k^T are truncated versions of U, $\sum$, V^T.This type of SVD is called Truncated SVD[64].The computational cost of SVD can be high despite its many advantages. It performs effectively on a user-defined matrix but poorly in experimental or adaptive processes [65].

For the prediction issue, it is the recommended dimensionality reduction approach [66].

Using Singular Value Decomposition (SVD), Qiao was able to determine the appropriate factorization rank and offer a solid initialization for Non-Negative Matrix Factorization (NMF) methods in 2015 [67]. In [68],since SVD uses a matrix to move data from a high-dimensional representation to a low-dimensional form and provides a suitable representation for any chosen matrix in which a minimal number of significant components may be easily removed, resulting in a reasonably accurate representation with whatever number of dimensions that is required. Using the Discrete Wavelet Transform, Singular Value Decomposition, and Huffman coding, Kumar and Manoj published a compression method for encrypted images in 2017 [69]. In smart distribution systems, De Souza et al.'s [70] lossy data compression method based on SVD was introduced. By using the suggested plan, they were able to efficiently transport data via communication networks while also decreasing the amount of data in smart grids. After reconstructing the data, the authors deduced that SVD is a highly competitive data reduction strategy that preserves the originality of the data. Similarly, to break down smart grid data, the authors in [71] employed a high-order singular value decomposition (HOSVD) technique. Also authors in [72], [73], and [74] illustrated several techniques for diverse domains based on singular value decomposition. In [75], [76], [77], and [78] dimensionality reduction for time-series data is commonly achieved by the unsupervised, linear method of Singular Value Decomposition (SVD) where a real or complex matrix A is broken down into three matrices, each of which retains its useful properties as the principal matrix. This process is a data-driven, fundamental linear algebra-based generalization of the Fourier transform [79].

An ECG signal compression technique was presented by Kumar et al. in 2015 for the massive ambulatory system data. The SVD and wavelet difference reduction techniques serve as the foundation for their algorithm[80].For the purpose of reducing cloud cyber data, Feng et al. (2018) introduced the tensor SVD technique [81]. In order to address the issues associated with storing and calculating huge amounts of historical load curve data, Chen suggested a dimension reduction clustering approach based on the daily load curve of SVD [82]. In addition to improving load curve clustering accuracy, this method reduced running times. A SVD-based feature extraction technique was presented by Kang [83] and is used to categorize and locate induction motor problems. The SVD-based technique has demonstrated good classification accuracy in both noisy and non-noise conditions.In [84] A. Winursito and R. Hidayat examined the accuracy of the SVD and PCA approaches in Mel Frequency Cepstral Coefficients (MFCC), a well-liked feature extraction technique used in speech recognition systems.[85] developed the proper orthogonal decomposition for SVD, which is non-iterative to reduce the impact of outliers on the velocimetry data of particle images. Additionally, a constrained singular value decomposition (SVD) was suggested in [86] in order to address the orthogonal problem of singular value decomposition and sparsity. The imputation approach served as the foundation for the multi-level SVD that was suggested by [87] in order to effectively manage and pre-process datasets gathered from various sources. The technique is found to be effective in a variety of fields, including education, medicine, and the biological sciences. [88]further suggested FFT-PCA/SVD as a more reliable and efficient method for recognizing changeable facial expressions than PCA/SVD. Digital image processing [89], taxonomic classification of biological sequences, bioinformatics, text summarization [90], pattern recognition [91], gene expression data [92], signal processing [93], are just a few domains in which the researchers have applied SVD.

V. LINEAR DISCRIMINANT ANALYSIS (LDA)

A classification and dimensionality reduction method called linear discriminant analysis (LDA) optimizes the differences between various classes in a dataset. .Projecting a dataset with many features onto a less-dimensional space with adequate class-separability is the primary goal of LDA. By reducing the dimensionality of the features, the LDA approach seeks to maximize the ratio of the within-class variance to the between-class variance. hence ensuring optimal class separability. The approach used in LDA is very similar to that of PCA. The process of LDA involves identifying a linear feature combination that describes or differentiates between two or more classes of objects or events. Both LDA and PCA are linear transformation techniques for lowering the dimensionality or number of features. In supervised learning contexts where the classifications of data points are preset, LDA is especially common. LDA employs a "supervised" strategy, whereas PCA is regarded as a "unsupervised" technique that ignores class labels in favour of maximizing dataset variance through the identification of principal components. By computing "linear discriminants," LDA finds the directions that act as axes to optimize the separation between different classes. One benefit of LDA is that it can utilize data from both features to generate a new axis that maximizes the class distance of the variables and minimizes variance. When the goal is to classify data into predefined groups and dimensionality reduction is desired but relevant information required for class discrimination needs to be retained, LDA may be helpful. When it is assumed that every class has a normal distribution, LDA performs effectively. If the task requires dimensionality reduction guided by class labels, LDA is a suitable choice.

A. Literature Survey on LDA

LDA may also be employed as a linear classifier, according to a number of research [94]. As a linear classifier, LDA makes use of a linear feature combination to extract features and reduce dimensions [95], [96], and [97].Face recognition [98], text recognition [99] [100], automatic machine operation diagnosis [101], early disease detection [102], person reidentification [103], hand movement classification [104], motor imagery EEG [105], and ground water redox conditions [106] are some of the applications of LDA. LDA is also effective in a wide range of classification tasks such as document categorization [107] and speech recognition [108].

In [109], LDA is utilized to lower the higher dimensions mathematically. LDA aims to reduce dimensionality while retaining as much of the class discriminating information as feasible [110]. It can be viewed as a pre-processing stage for applications involving pattern categorization and machine learning. The five phases necessary to carry out an LDA are listed in [111]. The authors of [112] employed PCA to eliminate the null values before using LDA to get around the issue of a small sample size. They put forth an innovative plan to implement PCA and LDA together. A robust technique for LDA utilizing Bhattacharyya error bound optimization was proposed by the authors in [113] in a similar scenario. The authors also suggested that sparse LDA is a simple extension of RLDA. Similarly, in [114], authors looked at the recursive problem of orthogonal least square regressions (SLR) and orthogonal LDA.

In [115] , LDA tries to linearly identify a combination of characteristics from two different types of objects or occurrences in order to classify and differentiate them . It determines the equivalent anticipated value or a category label y_i for each given n-dimensional dataset

$X= (x^1,x^2,\ldots x^n)$.

In [116], LDA's strengths lie in its ability to employ the probabilistic algorithm for the model's training in an efficient manner and its consistent rich internal structure. and it has been observed that the classification accuracy of the LDA is insufficient when employed alone. With high-dimensional and small sample size data, the conventional LDA model is susceptible to the singularity issue [117] .Because of this, this technique is frequently used in a hybrid system to solve the singularity problem and provide the best accuracy scores by combining it with other techniques like PCA.

In [118], Given m as members of the data samples and Ci as the class of data i, to map high dimensional data $Y = [x_1, x_2, x_3, \ldots . x_n] \in R^{m\times p}$ into lower dimensional data $Y \in R^{m\times k}$, LDA seeks to discover a linear transformation matrix $W = [w_1, w_2, w_3, \ldots w_n] \in R^{m\times k}$ (k << p) such that $Y = W^T X$

Equation (1) is used to determine the ideal Projection matrix.

$$J(W) = \frac{tr(S_b)}{tr(S_W)} \qquad \text{Eq(8)}$$

Zhang et al. [119] have presented a strategy for capturing local information called Local Intra-class Geometrical Variation Preserving LDA (LIPLDA). Using LIPLDA's high performance as well as enhanced precision for both local and global geometric data were recognized more effectively.

Applying k-Nearest Neighbor (KNN) to local LDA was proposed by Liu et al. [120]. In order to determine how important the sample structural information is, it uses an affinity matrix to assign weights. LDA is less successful when sample sizes are small and data are non-Gaussian. These issues have led to the development of Subclass Discriminant Analysis (SDA) [121] and Mixture Subclass DA (MSDA) [122]. Furthermore, Generalized EDA (GEDA) and Semi-supervised Regularized Discriminant Analysis (SRDA), respectively, were proposed by Ran et al. [123] and Xin et al. [124] to overcome the problem of small sample sizes.

Ye et al. [125] presented the 2D-LDA technique as a means of implicitly resolving the singularity problem and enhancing recognition accuracy. Recent advances indicate that 2D-LDA is an effective matrix-based DR method. However, the singularity problem is outside the scope of 2D-LDA. To solve the singularity problem, Li et al. [126] suggested a Generalized Lp-norm 2D-LDA (G2D-LDA) framework with regularization. By leveraging the kernel functionality, Baudat and Anouar [127] proposed Generalized Discriminant Analysis (GDA) to enhance classification. When features are extracted from the samples using global linear discriminant vectors, the test sample may be incorrectly classified, whereas local linear discriminant vectors may yield accurate classification.

VI. INDEPENDENT COMPONENT ANALYSIS (ICA)

One significant unsupervised method for extracting individual components from high dimensional data and to examine data from several channels is independent component analysis [129][130],[131]. The method entails modelling data derived from a linear combination of separate sources. In order to do a random vector's independent component analysis, a linear transformation that reduces the statistical dependence between the vector's components must be found. ICA is a dimensionality reduction technique that works similarly to PCA in that it starts with a set of features and outputs a separate, potentially useful set. In contrast to PCA, which aims to optimize variance, ICA attempts to extract distinct sources that are mixed within the dataset, based on the assumption that the features are mixtures of independent sources.

Retrieving independent characteristics from observations that are predicated on other specific data is one of ICA's primary goals. By decreasing the dissimilarity information, ICA finds the significant relationship between the data and decorrelates the same data [132]. Whereas PCA, for example, seeks the best data representations, ICA searches for non-Gaussian, statistically independent features.

A. Literature Survey On ICA

In [133], the initial decomposition of the data X using ICA is

$$X \rightarrow A.S \qquad \text{Eq(9)}$$

where the features are as independent as feasible, and the mixing matrix is A, and the basis coefficient is S. In order to produce k dimensions ICA selects the k most important independent elements in a dataset to produce data Y.

$$Y=A_k.S_k \qquad \text{Eq(10)}$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

The elements are attainable in specific order and scales . Knowing very little regarding the mixing procedure or the source signals, ICA focuses on separating the original signals from mixed data. In signal processing, ICA is considered a specific instance of the "blind source separation" problem [134].The ICA formula is.

$(x_1,x_2,\ldots x_m)^T=f(p_1,p_2,\ldots p_k)^T$ Eq(11)

Where f stands for the function of the real m-dimensional vectors and $x\,m$ for the input features. By using unique independent data vectors to capture covariance, ICA reduces the linearity between the data pieces. Thus, the established independence is achieved. A detailed explanation of the steps involved in reducing dimensions utilizing supervised (LDA) and unsupervised (PCA, KPCA, SVD, and ICA) approaches was put forth by Raji Ramachandran et al. [135].

For the purpose of extracting features from images, an enhanced independent component analysis approach based on the basis function's sparsity is employed [136]. This technique provides good sparsity and rapid convergence speed since it does not involve complex optimization of high-order nonlinear comparison functions. By using ICA for face recognition , the literature [137] showed that ICA has a very wide potential for advancement. Wang proposed an independent component analysis method for dimensionality reduction [138], which he named ICA-DR. It measures the statistical independency of the data beyond second-order statistics by using mutual information as a criterion.Mutual information is used to separate these signal sources under their statistically assessed independence [139]. A number of distinct standards have been developed to assess source independence. However, the source of all of them was mutual information for calculating the difference between two unanticipated sources [140].

Analyzing electroencephalographic (or brain-derived) data is an intriguing use case for ICA. In neuroimaging, fMRI, and EEG analysis, ICA is a crucial tool that aids in distinguishing between normal and aberrant signals.

Bayesian detection, blind identification, data analysis, compression, and source localization
are among the fields in which ICA is applied. ICA can extract more relevant components than PCA since it can be used with the independent optimization condition rather than PCA's variance maximization [141]. Additionally, ICA has the capacity to perhaps extract more information from the gathered data [142]. The majority of ICA algorithms try to resolve gradient-descent-based optimization issues, like maximum likelihood estimation [143], mutual information minimization [144], and non-Gaussian of source S [145].

VII. COMPARISON OF LINEAR DIMENSIONALITY TECHNIQUES

PCA can be used when working with datasets that have a lot of characteristics in order to alleviate the effect of curse of dimensionality. PCA is useful in capturing shared information and presenting it with fewer components when features exhibit significant correlation. PCA is useful when high-dimensional data visualization is difficult. Data projection allows for easy visualization into a lower-dimensional domain. PCA is an appropriate method when there is a high degree of linearity in the relationships between variables. Also, it can be used in situations where it is essential to capture global patterns and relationships within the data. However, PCA has the following limitations viz; Since principal components are linear combinations of the original features , hence there may be a loss of interpretability of the original features in the converted space. Also, the linearity of the relationships between variables is assumed by PCA, however this may not always be the case. The sensitivity of PCA to feature scale necessitates the frequent need for normalization. Because PCA aims to capture the maximum variance, which can be influenced by extreme values, outliers can have a major effect on the results.

SVD can be used when the objective is to lessen the data's dimensionality while maintaining its fundamental structure. SVD approximates and compresses huge data sets by preserving only the most significant singular values and vectors. SVD enables a matrix to be represented compactly and facilitates matrix operations computation in an efficient manner. As a result, dealing with huge matrices requires less computing power and storage. SVD can also be used when the effect of noise has to be minimized in the data by using only the most significant singular values . SVD is a suitable choice in ill-conditioned systems to provide numerical stability in solving linear equations. and while addressing missing values or sparse data. SVD is a strong tool for data analysis and dimensionality reduction. It enables us to determine which elements or characteristics of a dataset are most crucial and to examine the connections between them. However, SVD has the following limitations viz; It can be costly to compute the whole SVD for large matrices particularly if the matrix has a high rank since it may require a lot of memory to store the entire matrices U, summation, and V. This may make employing SVD for large-scale data analysis less feasible. SVD is susceptible to missing values in the data, necessitating the use of specific methods to handle them.

Another limitation of SVD is that not all data types may be appropriate for SVD. The data matrix must be dense and well-behaved for SVD to work. Other matrix factorization techniques may be more appropriate for data that is sparse or irregularly organized, in which case this method may not be suitable.

LDA is an effective tool for classification tasks because it is developed to optimize the separation between various classes. Similar to PCA, LDA also has the benefit of taking class information into account while reducing dimensionality. LDA functions well when it is assumed that each class has a normal distribution. LDA is an appropriate option when the task calls for dimensionality reduction under the direction of class labels. However, if there aren't enough samples in each class, LDA might not function properly, since more samples lead to better class parameter estimation. Also, LDA is susceptible to outliers, and their existence may have an impact on the method's performance.

ICA makes it possible to separate mixed signals without knowing anything about the source signals or the mixing mechanism beforehand. It is capable of exposing the data's underlying structures and hidden sources.

ICA relies on the statistical independence assumption, which is frequently more robust than orthogonality and other assumptions. This enables ICA to be used in a variety of real-world situations. Informational features can be extracted by using ICA as a preprocessing step. These characteristics can then be used to enhance the performance and interpretability of machine learning algorithms. However, ICA has the following limitations viz; The underlying sources are assumed to be non-Gaussian by ICA, albeit this may not always be the case. In the event that the underlying sources are Gaussian, ICA may not function. It is assumed by ICA that the sources are blended linearly, albeit this may not always be the case. Nonlinear blending of the sources may render ICA ineffective. When dealing with large datasets, ICA might be expensive to calculate. Because of this, applying ICA to real-world problems may prove difficult.

ICA might run into convergence issues, which could keep it from continuously resolving issues. This can be a problem for complex datasets with plenty of sources. Depending on the particular goals and properties of the data, one can choose between Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA).

Table 1: Conceptual Comparison Of Linear Dimensionality Reduction Techniques

Methods	Objective and Advantages	Limitations	Complexity of computation
PCA	Increase variance while minimizing noise sensitivity.	Confined to projection that is linear	$O(d^2n + n^3)$
SVD	Minimum construction error	Unsuitable for data that is not linear	$O(d^2n + n^3)$
LDA	Optimize the level of class separation	Experience problems with class singularity	$O(d^2n)$, $n > d$; $O(d^3)$, $d > n$
ICA	Increase the degree of statistical independence	High time required for training	O[2di(d +1) n]

VIII. CONCLUSION

The increasing amount of high-dimensional data makes it necessary to employ dimensionality reduction techniques to make the data easier to handle. This paper provides a detailed review of the different linear dimensionality reduction techniques that are used, identifies the various approaches that are used to lower the dimensions in order to increase the accuracy of algorithms, and provides guidance on how to evaluate the field's recent surge in popularity. It is crucial to display the high dimensions of data in lesser dimensions without sacrificing the data's uniqueness. But regardless of the type of data we are dealing with, choosing the appropriate strategy for a particular set of data is essential and greatly influenced by the kind of data we are dealing with. The benefits and drawbacks of each technique are also explained in this study. Many methodologies such as Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) are analysed and compared based on various attributes. Conclusively, all these methods seek to provide the appropriate information while minimizing complexity. Additionally, this work provides a reasonable starting point for comparing various dimensional approaches.

REFERENCES

[1] Tang, W., & Zhong, S. (2007). Pairwise constraints-guided dimensionality reduction. Computational Methods of Feature Selection, 295-312

[2] Donoho, D.L. (2000) High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. Lecture Delivered at the "Mathematical Challenges of the 21st Century" Conference of the American Math. Society, Los Angeles. http://www-stat.stanford.edu/donoho/Lectures/AMS2000/AMS2000.html .

[3] F. Anowar and S. Sadaoui, "Incremental Neural-Network Learning for Big Fraud Data," in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 3551–3557, doi: 10.1109/SMC42975.2020.9283136.

[4] Jain, R., & Xu, W. (2021). RHDSI: a novel dimensionality reduction based algorithm on high dimensional feature selection with interactions. Information Sciences, 574, 590-605.

[5] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik, "Dimensionality Reduction: A Comparative Review," J. Mach. Learn. Res., vol. 10, pp. 1–41, 2009, doi: 10.1080/13506280444000102.

[6] Wei Wei, Member, IEEE, Qin Yue, Kai Feng, Junbiao Cui, and Jiye Liang, Senior Member, IEEE, "Unsupervised Dimensionality Reduction based on fusing multiple clustering results," IEEE Transactions on Knowledge and Data Engineering, 21 September 2021.

[7] K. Fukunaga, Introduction to Statistical Pattern Recognition. San Diego, CA, USA, 1990.

[8] Laurens van der Maaten "An Introduction to Dimensionality Reduction Using Matlab", MICC, Maastricht University.

[9] Verleysen M., François D. (2005) The Curse of Dimensionality in Data Mining and Time Series Prediction. In: Cabestany J., Prieto A., Sandoval F. (eds) Computational Intelligence and Bioinspired Systems. IWANN 2005. Lecture Notes in Computer Science, vol 3512. Springer, Berlin, Heidelberg.

[10] Thippa Reddy Gadekallu, Praveen Kumar Reddy, Kuruva Lakshman, Rajesh Kaluri, "Analysis of Dimensionality Reduction Techniques on Big Data," IEEE Access (Vol: 8) 16 March 2020.

[11] Zebari , R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. Journal of Applied Science and Technology Trends, 1(2), 56 – 70

[12] G.N.Ramadevi and K.Usharani- "Study On Dimensionality Reduction Techniques And Applications" - 2- Vol 04, Special Issue01; 2013 -Publications Of Problems & Application In Engineering Research – Paper

[13] Juvonen, A., Sipola, T., Hämäläinen, T. (2015). Online anomaly detection using dimensionality reduction techniques for HTTP log analysis, Comput. Networks, vol. 91, pp. 46–56, doi: 10.1016/j.comnet.2015.07.019.

[14] Liu, L., Zsu, M. T. (2009). Encyclopedia of Database Systems, Springer Publishing Company

[15] Christopher J. C. Burges "Dimension Reduction: A Guided Tour-Foundations and TrendsR_ in Machine Learning" Vol. 2, No. 4 (2009) 275–365 _c 2010 DOI: 10.1561/2200000002.

[16] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev., vol. 28, no. 1, pp. 39–54, 1998, doi: 10.1109/5326.661089.

[17] F. Salo, A. B. Nassif, and A. Essex, "Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection," Comput. Networks, vol. 148, pp. 164–175, Jan. 2019, Accessed: Sep. 03, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pi i/S1389128618303037.

[18] M. Partridge and C. Rafael, "Fast dimensionality reduction and simple PCA," Intell. Data Anal., vol. 2, no. 3, pp. 292–298, 1998, doi: 10.3233/IDA-1998-2304.

[19] K. V Ravi Kanth, D. Agrawal, A. El Abbadi, and A. Singh, "Dimensionality Reduction for Similarity Searching in Dynamic Databases," Comput. Vis. Image Underst., vol. 75, no. 1, pp. 59–72, 1999, doi: https://doi.org/10.1006/cviu.1999.0762.

[20] P. Switzer, "Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery," J. Int. Assoc. Math. Geol., vol. 12, no. 4, pp. 367–376, 1980, doi: 10.1007/BF01029421.

[21] Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural Netw 13(4–5):411–430

[22] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 Science and Information Conference, 2014, pp. 372–378, doi: 10.1109/SAI.2014.6918213.

[23] A. B. Nassif, M. Azzeh, L. F. Capretz, and D. Ho, "A comparison between decision trees and decision tree forest models for software development effort estimation," in 2013 3rd International Conference on Communications and Information Technology, ICCIT 2013, 2013, pp. 220–224, doi: 10.1109/ICCITechnology.2013.6579553.

[24] M. Azzeh "Analogy-based effort estimation: a new method to discover set of analogies from dataset characteristics," IET Softw., vol. 9, no. 2, pp. 39–50, 2015, doi: 10.1049/ietsen.2013.0165.

[25] M. Azzeh, S. Banitaan, and F. Almasalha, "Pareto efficient multi-objective optimization for local tuning of analogy-based estimation," Neural Comput. Appl., vol. 27, no. 8, pp. 2241– 2265, 2016, doi: 10.1007/s00521-015-2004-y.

[26] M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining", Journal of Computing, Vol. 1, No. 1, pp. 7-11, 2009.

[27] J.E. Jackson. A User's Guide to Principal Components. New York: John Wiley and Sons, 1991.

[28] I. T. Jolliffe. Principal Component Analysis. Sprin, second edition, 2002.

[29] S. Wold. Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems, 2:37–52, 1987.

[30] R. Vidal, Y. Ma, and S. S. Sastry. Generalized Principal Component Analysis. Springer, 2016

[31] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A, page 20150202, 2016.

[32] I.T. Jolliffe. Principal Component Analysis. Springer-Verlag, 1986.

[33] K.V. Mardia, J.T. Kent, and J.M. Bibby. Multivariate Analysis. Probability and Mathematical Statistics. Academic Press, 1995.

[34] Person, K. (1901) On Lines and Planes of Closest Fit to System of Points in Space. Philiosophical Magazine, 2, 559-572. http://dx.doi.org/10.1080/14786440109462720

[35] Jenkins, O.C. and Mataric, M.J. (2002) Deriving Acion and Behavior Primitives from Human Motion Data. International Conference n Robots and Systems, 3, 2551-2556. [5] Jain, A.K. and Dubes, R.C. (1962) Algorithms for Clastering Data. Prentice Hall, Upper Saddle River.

[36] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1995) Multivariate Analysis Probability and Mathematical Statistics. Academic Press, Waltham.

[37] (2002) Francesco Camastra Data Dimensionality Estimation Methods, a Survey INFM-DISI, University of Genova, Genova.

[38] Fukunaga, K. (1982) Intrinsic Dimensionality Extraction, in Classification, Pattern Recognition and Reduction of Dimensionality, Vol. 2 of Handbook of Statistics, North Holland, 347-362.
[39] Roweis, S.T. (1997) EM Algorithms for PCA and SPCA. Advances in Neural Information Processing Systems, 10, 626-632.
[40] Lawrence, N.D. (2005) Probabilistic Non-Linear Proncipal Component Analysis with Gaussian Process Latent Variable Models. Journal of Machine Learning Research, 6, 1783-1816.
[41] Welling, M., Rosen-Zvi, M. and Hinton, G. (2004) Exponential Family Harmoniums with an Application to Information Retrieval. Advances in Neural Information Processing Systems, 17, 1481-1488.
[42] Turk, M.A. and Pentland, A.P. (1991) Face Recognition Using Eigenfaces. Proceedings of the Computer Vision and Pattern Recognition 1991, Maui, 586-591. http://dx.doi.org/10.1109/CVPR.1991.139758
[43] Huber, R., Ramoser, H., Mayer, K., Penz, H. and Rubik, M. (2005) Classification of Coins Using an Eigenspace Approach. Pattern Recognition Letters, 26, 61-75. http://dx.doi.org/10.1016/j.patrec.2004.09.006
[44] Posadas, A.M., Vidal, F., de Miguel, F., Alguacil, G., Pena, J., Ibanez, J.M. and Morales, J. (1993) Spatialtemporal Analysis of a Seismic Series Using the Principal Components Method. Journal of Geophysical Research, 98, 1923-1932. http://dx.doi.org/10.1029/92JB02297
[45] H. Abdi and L. J. Williams, “Principal Component Analysis,” Wiley Interdiscip. Rev. Comput. Stat., vol. 4, no. 2, pp. 433–459, 2010
[46] Arunasakthi, K., &KamatchiPriya, L. (2014). A review on linear and non-linear dimensionality reduction techniques. Machine Learning and Applications: An International Journal (MLAIJ), 1(1), 65-76.
[47] Yan, S., Xu, D., Zhang, B., Zhang, H. J., Yang, Q., & Lin, S. (2006). Graph embedding and extensions: A general framework for dimensionality reduction. IEEE transactions on pattern analysis and machine intelligence, 29(1), 40-51.
[48] Raunak, V. (2017). Simple and Effective Dimensionality Reduction for Word Embeddings, arXiv, no. Nips, pp. 1–6.
[49] Jiang, J., Ma, J., Chen, C., Wang, Z., Cai, Z. and Wang, L. (2018). SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery, IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 4581–4593, doi: 10.1109/TGRS.2018.2828029
[50] N. Verbeeck, R. M. Caprioli, and R. Van de Plas, “Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry,” Mass Spectrometry Reviews, vol. 39, no. 3, pp. 245–291, 2020.
[51] L. J. Al-Omairi, J. Abawajy, M. U. Chowdhury, and T. AlQuraishi, “An empirical analysis of graph-based linear dimensionality reduction techniques,” Concurrency and Computation: Practice and Experience, vol. 33, no. 5, pp. 1–13, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5990
[52] T. Howley, M. G. Madden, M.-L. O’Connell, and A. G. Ryder, “The effect of principal component analysis on machine learning accuracy with high dimensional spectral data,” in Applications and Innovations in Intelligent Systems XIII. Springer London, pp. 209–222. [Online]. Available: https://doi.org/10.1007/1-84628-224-116
[53] D. Cao, Y. Tian, and D. Bai, “Time series clustering method based on principal component analysis,” in Proceedings 5th International Conference on Information Engineering for Mechanics and Materials (ICIMM 2015). Citeseer, 2015, pp. 888–895
[54] F. Alshammri and J. Pan, “Moving dynamic principal component analysis for non-stationary multivariate time series,” Computational Statistics, vol. 36, no. 3, pp. 1–41, Mar. 2021. [Online]. Available: https://doi.org/10.1007/s00180-021- 01081-8
[55] P. Tanisaro and G. Heidemann, “Dimensionality reduction for visualization of time series and trajectories,” in Image Analysis. Springer International Publishing, 2019, pp. 246–257. [Online]. Available: https://doi.org/10.1007/978- 3-030-20205-721
[56] [55] A. Widodo and I. Budi, “Model selection using dimensionality reduction of time series characteristics,” in International Symposium on Forecasting, Seoul, South Korea, 2013, pp. 57–118.
[57] [56] G. Gawde and J. Pawar, “Shape based time series reduction using pca,” in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017, pp. 1–4.
[58] F. Anowar, S. Sadaoui, and B. Selim, “Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne),” Computer Science Review, vol. 40, pp. 1–13, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013721000186.
[59] N. Sharma, “Principal component analysis,” https://heartbeat.fritz.ai/understanding-the-mathematics-behind-principalcomponent-analysis-efd7c9ff0bb3, 2021, last accessed 21 December 2021.
[60] J. Shlens, “A tutorial on principal component analysis,” arXiv preprint arXiv:1404.1100, pp. 1–12, 2014
[61] Y. Jaradat, M. Masoud, I. Jannoud, A. Manasrah, and M. Alia, “A Tutorial on Singular Value Decomposition with Applications on Image Compression and Dimensionality Reduction,” 2021 International Conference on Information Technology (ICIT), 2021, pp. 769-772, doi: 10.1109/ICIT52682.2021.9491732.
[62] P. David. (2015) "Linear Algebra: A Modern Introduction", 4th ed. Cengage Learning
[63] L. J. Sciacca and R. J. Evans, “Multidimensional inverse problems in ultrasonic imaging,” in Multidimensional Systems: Signal Processing and Modeling Techniques, ser. Control and Dynamic Systems, C. Leondes, Ed. Academic Press, 1995, vol. 69, pp. 1–48. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S0090526705800374
[64] S. Brunton and J. Kutz, Singular Value Decomposition (SVD). Cambridge University Press, 02 2019, pp. 3–46.
[65] M. Frank and J. M. Buhmann, “Selecting the rank of truncated SVD by maximum approximation capacity,” in 2011 IEEE International Symposium on Information Theory Proceedings. IEEE, Jul. 2011, pp. 1036–1040. [Online]. Available: https://doi.org/10.1109/isit.2011.6033687
[66] A. Dutta, J. Liang, and X. Li, “A fast and adaptive svd-free algorithm for general weighted low-rank recovery,” arXiv preprint arXiv:2101.00749, pp. 1–27, 2021
[67] F. Anowar, S. Sadaoui, and B. Selim, “Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne),” Computer Science Review, vol. 40, pp. 1–13, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013721000186
[68] Qiao, Hanli. (2015) "New SVD based initialization strategy for non-negative matrix factorization" Pattern Recognition Letters 63: 71-77.
[69] Taufik faudi, bustami yusuf, Munzir, “Singular value Decomposition for dimensionality reduction in unsupervised text learning problems”, ieee conference on Education tech and computer 29th july 2010.
[70] Kumar, Manoj; and Ankita Vaish. (2017) "An efficient encryption-then-compression technique for encrypted images using SVD." Digital Signal Processing 60: 81-89

[71] J. C. S. de Souza, T. M. L. Assis, and B. C. Pal, “Data compression in smart distribution systems via singular value decomposition,” IEEE Transactions on Smart Grid, vol. 8, no. 1, pp. 275–284, 2017. [
[72] D. Kaur, G. S. Aujla, N. Kumar, A. Y. Zomaya, C. Perera, and R. Ranjan, “Tensor-based big data management scheme for dimensionality reduction problem in smart grid systems: Sdn perspective,” IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 10, pp. 1985–1998, Oct 2018.
[73] A. Zhang and D. Xia, “Tensor svd: Statistical and computational limits,” IEEE Transactions on Information Theory, 2018.
[74] L. T. Yang, X. Wang, X. Chen, L. Wang, R. Ranjan, X. Chen, and M. J. Deen, “A multi-order distributed hosvd with its incremental computing for big services in cyber-physical-social systems,” IEEE Transactions on Big Data, 2018.
[75] I. Kajo, N. Kamel, and Y. Ruichek, “Incremental tensor-based completion method for detection of stationary foreground objects,” IEEE Transactions on Circuits and Systems for Video Technology, 2018
[76] F. Fallucchi and F. M. Zanzotto, “Singular value decomposition for feature selection in taxonomy learning,” in Proceedings of the International Conference RANLP-2009. Borovets, Bulgaria: Association for Computational Linguistics, Sep. 2009, pp. 82–87. [Online]. Available: https://www.aclweb.org/anthology/R09-1016
[77] I. Bowala and M. Fernando, “A novel model for Time-Series Data Clustering Based on piecewise SVD and BIRCH for Stock Data Analysis on Hadoop Platform,” Advances in Science, Technology and Engineering Systems Journal, vol. 2, no. 3, pp. 855–864, 2017. [Online]. Available: http://astesj.com/v02/i03/p106/
[78] W. Lin, J. Z. Huang, and T. McElroy, “Time series seasonal adjustment using regularized singular value decomposition,” Journal of Business & Economic Statistics, vol. 38, no. 3, pp. 487–501, feb 2019. [Online]. Available: https://doi.org/10.1080/07350015.2018.1515081
[79] H. Li, C. Lin, X. Wan, and Z. xin Li, “Feature representation and similarity measure based on covariance sequence for multivariate time series,” IEEE Access, vol. 7, pp. 67 018–67 026, 2019.
[80] V. Choqueuse, P. Granjon, A. Belouchrani, F. Auger, and M. Benbouzid, “Monitoring of three-phase signals based on singular-value decomposition,” IEEE Transactions on Smart Grid, vol. PP, pp. 6156–6166, 02 2019.
[81] Kumar, Ranjeet; A. Kumar; and G. K. Singh. (2015) "Electrocardiogram signal compression based on singular value decomposition (SVD) and adaptive scanning wavelet difference reduction (ASWDR) technique" AEU-International Journal of Electronics and Communications 69.12: 1810-1822.
[82] Feng, Jun; et al. (2018) "A Secure Higher-Order Lanczos-Based Orthogonal Tensor SVD for Big Data Reduction." IEEE Transactions on Big Data.
[83] Chen Y et al (2018) Application of singular value decomposition algorithm to dimension-reduced clustering analysis of daily load profiles. Automat Electr Power Syst 42(3):105–111
[84] Kang M, Kim JM (2013) Singular value decomposition-based feature extraction approaches for classifying faults of induction motors. Mech Syst Signal Process 41(1–2):348–356
[85] A. Winursito, R. Hidayat, A. Bejo, and M. N. Y. Utomo, “Feature Data Reduction of MFCC Using PCA and SVD in Speech Recognition System,” 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), 2018, pp. 1-6, doi: 10.1109/ICSCEE.2018.8538414
[86] Higham, J.E., Brevis, W. and Keylock, C.J. (2016) A Rapid Non-Iterative Proper orthogonal Decomposition Based Outlier Detection and Correction for PIV Data. Measurement Science and Technology, 27, Ariticle ID: 125303. https://doi.org/10.1088/0957-0233/27/12/125303
[87] Guillemot, V., Beaton, D., Gloaguen, A., Löfstedt, T., Levine, B., Raymond, N. and Abdi, H. (2019) A Constrained Singular Value Decomposition Method That Integrates Sparsity and Orthogonality. PLoS ONE, 14, e0211463. https://doi.org/10.1371/journal.pone.0211463
[88] Husson, F., Josse, J., Narasimhan, B. and Robin, G. (2019) Imputation of Mixed Data With Multilevel Singular Value Decomposition. Journal of Computational and Graphical Statistics, 28, 552-566. https://doi.org/10.1080/10618600.2019.1585261.
[89] Asiedu, L., Adebanji, A., Oduro, F. and Mettle, F. (2016) Statistical Assessment of PCA/SVD and FFT-PCA/SVD on Variable Facial Expressions. British Journal of Mathematics & Computer Science, 12, 1-23. https://doi.org/10.9734/BJMCS/2016/22141 [29] Ka
[90] Cao, L.J. (2006) Singular Value Decomposition Applied to Digital Image Processing. Arizona State University Polytechnic Campus, Mesa, 1-15.
[91] Santos, A.R., Santos, M.A., Baumbach, J., Mcculloch, J.A., Oliveira, G.C., Silva, A. and Azevedo, V. (2011) A Singular Value Decomposition Approach for Improved Taxonomic Classification of Biological Sequences. BMC Genomics, 12, Artticle No. S11. https://doi.org/10.1186/1471-2164-12-S4-S11
[92] Lassiter, A. (2013) Handwritten Digit Classification and Reconstruction of Marred Images Using Singular Value Decomposition. Virginia Tech.
[93] Meng, C., Zeleznik, O.A., Thallinger, G.G., Kuster, B., Gholami, A.M. and Culhane, A.C. (2016) Dimension Reduction Techniques for the Integrative Analysis of Multi-Omics Data. Briefings in Bioinformatics, 17, 628-641. https://doi.org/10.1093/bib/bbv108
[94] Li, X., Ng, M.K., Ye, Y., Wang, E.K. and Xu, X. (2017) Block Linear Discriminant Analysis for Visual Tensor Objects with Frequency or Time Information. Journal of Visual Communication and Image Representation, 49, 38-46. https://doi.org/10.1016/j.jvcir.2017.08.004
[95] L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee, and Q. M. Gu, “A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine,” Neurocomputing, vol. 55, no. 1, pp. 321–336, 2003, doi: https://doi.org/10.1016/S0925-2312(03)00433- 8.
[96] Tharwat, A., Gaber, T., Ibrahim, A. and Hassanien, A.E. (2017) Linear Discriminant Analysis: A Detailed Tutorial. AI Communications, 30, 169-190. https://doi.org/10.3233/AIC-170729
[97] Fisher, E.M. (1936) Linear Discriminant Analysis. Statistics & Discrete Methods of Data Sciences, 392, 1-5.
[98] C. O. S. Sorzano, J. Vargas, and A. P. Montano, “A survey of dimensionality reduction techniques,” pp. 1–35, 2014.
[99] Park, C.H. and Park, H. (2008) A Comparison of Generalized Linear Discriminant Analysis Algorithms. Pattern Recognition, 41, 1083-1097. https://doi.org/10.1016/j.patcog.2007.07.022
[100] Yu, H. and Yang, J. (2001) A Direct LDA Algorithm for High-Dimensional Data—With Application to Face Recognition. Pattern Recognition, 34, 2067-2070. https://doi.org/10.1016/S0031-3203(00)00162-X
[101] Zhang, T., Fang, B., Tang, Y. Y., Shang, Z. and Xu, B. (2010) Generalized Discriminant Analysis: A Matrix Exponential Approach. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 40, 186-197. https://doi.org/10.1109/TSMCB.2009.2024759
[102] Kedadouche, M., Liu, Z. and Thomas, M. (2018) Bearing Fault Feature Extraction Using Autoregressive Coefficients, Linear Discriminant Analysis and Support Vector Machine under Variable Operating Conditions. Applied Condition Monitoring, 9, 339-352. https://doi.org/10.1007/978-3-319-61927-9_32
[103] Xiong, H., Cheng, W., Hu, W., Bian, J. and Guo, Z. (2017) FWDA: A Fast Wishart Discriminant Analysis with its Application to Electronic Health Records Data Classification. 1-15. http://arxiv.org/abs/1704.07790

[104]Wu, L., Shen, C.H. and van den Hengel, A. (2017) Deep Linear Discriminant Analysis on Fisher Networks: A Hybrid Architecture for Person Re-Identification. Pattern Recognition, 65, 238-250. https://doi.org/10.1016/j.patcog.2016.12.022

[105]Krasoulis, A., Nazarpour, K. and Vijayakumar, S. (2017) Use of Regularized Discriminant Analysis Improves Myoelectric Hand Movement Classification. 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER), Shanghai, 25-28 May 2017, 395-398. https://doi.org/10.1109/NER.2017.8008373

[106]Jusas, V. and Samuvel, S.G. (2019) Classification of Motor Imagery Using a Combination of User-Specific Band and Subject-Specific Band for Brain-Computer Interface. Applied Sciences (Switzerland), 9, 1-17. https://doi.org/10.3390/app9234990 [106] Wilson, S.R., Close, M.E. and Abraham, P. (2018) Applying Linear Discriminant Analysis to Predict Groundwater Redox Conditions Conducive to Denitrification. Journal of Hydrology, 556, 611-624. https://doi.org/10.1016/j.jhydrol.2017.11.045

[107]K. Torkkola. Linear discriminant analysis in document classification. In IEEE TextDM 2001, pages 800–806, 2001.

[108]R. Haeb-Umbach and H. Ney. Linear discriminant analysis for improved large vocabulary continuous speech recognition. In IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992, volume 1, pages 13–16, 1992.

[109]Dimensionality Reduction Eisaku Maeda Tokyo Denki University, Tokyo, Japan (1) © Springer Nature Switzerland AG 2020 K. Ikeuchi, Computer Vision

[110]R. Gutierrez-Osuna, "Dimensionality rereduction-lda." Introduction to Pattern Recognition,Wright State University.

[111]S. Raschka, "Linear discriminant analysis bit by bit." Blog, August 2014

[112]Think India Journal Issn: 0971-1260 Vol-22-Issue-16-August-2019. A Comparative Study on Linear Dimensionality Reduction Techniques For Big Data Analytics Devinder Kaur, Oshin Bhardwaj, Surbhi Sharma,

[113]C. Li, Y. Shao, W. Yin, and M. Liu, "Robust and sparse linear discriminant analysis via an alternating direction method of multipliers," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–12, 2019.

[114]Y. Liu, R. Zhang, F. Nie, X. Li, and C. Ding, "Supervised dimensionality reduction methods via recursive regression," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–11, 2019.

[115]S. Feng and H. Wang, "Comparison of PCA and LDA Dimensionality Reduction Algorithms based on Wine Dataset," 2021 33rd Chinese Control and Decision Conference (CCDC), 2021, pp. 2791-2796, doi: 10.1109/CCDC52312.2021.9602325.

[116]X. Liu, H. Xiong, and N. Shen, "A hybrid model of VSM and LDA for text clustering," 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), 2017, pp. 230- 233, doi: 10.1109/CIAPP.2017.8167213.

[117]S. Ji and J. Ye, "Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection," in IEEE Transactions on Neural Networks, vol. 19, no. 10, pp. 1768-1782, Oct. 2008, doi: 10.1109/TNN.2008.2002078.

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

[119]D. Zhang, Y. Zhao, M. Du, A new supervised dimensionality reduction algorithm using linear discriminant analysis and locality preserving projection, WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS, E-ISSN (2013) 2224–3402.

[120]Z. Liu, G. Liu, J. Pu, X. Wang, H. Wang, Orthogonal sparse linear discriminant analysis, International Journal of Systems Science (2018) 1–11.

[121]M. Zhu, A. M. Martinez, Subclass discriminant analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (8) (2006) 1274–1286.

[122]N. Gkalelis, V. Mezaris, I. Kompatsiaris, Mixture subclass discriminant analysis, IEEE Signal Processing Letters 18 (5) (2011) 319–322.

[123]R. Ran, B. Fang, X. Wu, S. Zhang, A simple and effective generalization of exponential matrix discriminant analysis and its application to face recognition, IEICE TRANSACTIONS on Information and Systems 101 (1) (2018) 265–268.

[124]Y. Xin, Q. Wu, Q. Zhao, Q. Wu, Semi-supervised regularized discriminant analysis for eeg-based bci system, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2017, pp. 516–523.

[125]J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, in: Advances in neural information processing systems, 2005, pp. 1569–1576.

[126]C.-N. Li, Y.-H. Shao, W.-J. Chen, N.-Y. Deng, Generalized two-dimensional linear discriminant analysis with regularization, arXiv preprint arXiv:1801.07426.

[127]G. Baudat, F. Anouar, Generalized discriminant analysis using a kernel approach, Neural computation 12 (10) (2000) 2385–2404

[128]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, July 2011

[129]Tang, Y. and Rose, R. (2008) A Study of Using Locality Preserving Projections for Feature Extraction in Speech Recognition. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, 31 March-4 April 2008, 1569-1572. https://doi.org/10.1109/ICASSP.2008.4517923

[130]Comon P (1994) Independent component analysis, a new concept? Signal Process 36(3):287–314

[131]Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural Netw 13(4–5):411–430

[132]A. Hyvärinen, J. Karhunen, and E. Oja. (2001). Independent Component Analysis. New York: Wiley.

[133]C. O. S. Sorzano, J. Vargas, and A. P. Montano, "A survey of dimensionality reduction techniques," pp. 1–35, 2014.

[134]A. Tharwat, "Independent component analysis: An introduction," Appl. Comput. Informatics, vol. 17, no. 2, pp. 222–249, Jan. 2021, doi: 10.1016/j.aci.2018.08.006.

[135]Raji Ramachandran, Gopika Ravichandran and Aswathi Raveendran, "Evaluation of Dimensionality Reduction Techniques for Big Data," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 23 April 2020

[136]Huang QH, Wang S, Liu Z (2007) Improved algorithm of image feature extraction based on independent component analysis. Opto-Electron Eng 1:121–125

[137]Yuen PC, Lai J-H (2002) Face representation using independent component analysis. Pattern Recogn 35(6):1247–1257

[138]Wang J, Chang C-I (2006) Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. IEEE Trans Geosci Remote Sens 44(6):1586–1600.

[139]Jing Wang and Chein-I Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," in IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 6, pp. 1586-1600, June 2006, doi 10.1109/TGRS.2005.863297.

[140]T. Cover and J. Thomas, Elements of Information Theory. New York: Wiley, 1991.

[141]Scholz, M., Gatzek, S., Sterling, A., Fiehn, O. and Selbig, J. (2004) Metabolite Fingerprinting: Detecting Biological Features by Independent Component Analysis. Bioinformatics, 20, 2447-2454. https://doi.org/10.1093/bioinformatics/bth270

[142]Pochet, N., De Smet, F., Suykens, J.A.K. and De Moor, B.L.R. (2004) Systematic Benchmarking of Microarray Data Classification: Assessing the Role of Non-Linearity and Dimensionality Reduction. Bioinformatics, 20, 3185-3195

[143] Hyvärinen, A. (1999) Sparse Code Shrinkage: Denoising by Nonlinear Maximum Likelihood Estimation. Neural Computation, 11, 1739-176

[144] Amari, S.-I., Cichocki, A. and Yang, H.H. (1996) A New Learning Algorithm for Blind Source Separation. In: Mozer, M., Ed., Advances in Neural Information Processing System, Morgan Kaufmann Publishers, Massachussets, 757-763. https://papers.nips.cc/paper/1115-a-new-learning-algorithm-for-blind-signal-separation.pdf

[145] Hyvärinen, A. and Oja, E. (1997) A Fast Fixed-Point Algorithm for Independent Component Analysis. Neural Computation, 9, 1483-1492. https://doi.org/10.1162/neco.1997.9.7.1483