Review of Dimension Reduction Methods

Salifu Nanga1*, Ahmed Tijani Bawah2, Benjamin Ansah Acquaye3, Mac-Issaka Billa4, Francis Delali Baeta5, Nii Afotey Odai6, Samuel Kwaku Obeng7, Ampem Darko Nsiah8

1Department of Basic Science, School of Basic and Biomedical Science, University of Health and Allied Sciences, Ho, Ghana
2Department of Medical Laboratory Science, School of Allied Health Sciences, University of Health and Allied Sciences, Ho, Ghana
3SANA Research and Consult, Accra, Ghana
4University of Management Studies, Weija-Accra, Ghana
5Department of Mathematics and Statistics, Ho Technical University, Ho, Ghana
6Health Information Unit, Tema General Hospital, Tema, Ghana
7Department of Data Quality Monitoring, Ghana Statistical Service, Accra, Ghana
8Criminal Investigation Department, Ghana Police Service, Accra, Ghana

Email: *snanga@uhass.edu.gh, ahmed024gh@yahoo.com, acquayeans@gmail.com, macbilla@gmail.com, baetadelali@gmail.com, afotey777@gmail.com, samobeng2002@yahoo.com, ampeamn@yahoo.com

Abstract

Purpose: This study sought to review the characteristics, strengths, weaknesses variants, applications areas and data types applied on the various Dimension Reduction techniques. Methodology: The most commonly used databases employed to search for the papers were ScienceDirect, Scopus, Google Scholar, IEEE Xplore and Mendeley. An integrative review was used for the study where 341 papers were reviewed. Results: The linear techniques considered were Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Singular Value Decomposition (SVD), Latent Semantic Analysis (LSA), Locality Preserving Projections (LPP), Independent Component Analysis (ICA) and Project Pursuit (PP). The non-linear techniques which were developed to work with applications that have complex non-linear structures considered were Kernel Principal Component Analysis (KPCA), Multi-dimensional Scaling (MDS), Isomap, Locally Linear Embedding (LLE), Self-Organizing Map (SOM), Latent Vector Quantization (LVQ), t-Stochastic neighbor embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP). DR techniques can further be categorized into supervised, unsupervised and more recently semi-supervised learning methods. The supervised versions are the LDA and LVQ. All the other techniques are unsupervised. Supervised variants of PCA, LPP, KPCA and MDS have been developed. Supervised and semi-supervised variants of PP and t-SNE have also been developed and a semi supervised version of the LDA has been developed. Conclusion: The various application areas, strengths, weaknesses and variants of the DR tech-
niques were explored. The different data types that have been applied on the various DR techniques were also explored.

**Keywords**

Dimension Reduction, Machine Learning, Linear Dimension Reduction Techniques, Non-Linear Reduction Techniques

---

### 1. Introduction

The world in recent times has seen huge amount of data being churned out in different areas of application, resulting in an exponential growth in the complexity, heterogeneity, dimensionality and the size of data [1]. Areas such as education, medicine, web, social media and business are inundated with huge amount of data in this era of Information Communication and Technology (ICT) [2]. There is a continuous evolvement of data in different forms such as digital images [3]; videos [4]; text [5] and speech signals [6].

The existing classical statistical methodologies that have been relied on were developed from an era where the collection of data was not easy as it is now and the magnitude of datasets was much smaller. Therefore, there is a challenge of analyzing these large and sophisticated data sets which require a more sophisticated statistical and computational way of analyzing such data. As a result, the area of machine learning has evolved rapidly to help address this problem. It applies artificial intelligence and automatic learning of data. There is a focus on computer programs to access data and use it to learn for themselves.

In machine learning modelling, high dimensionality of data may raise issues for the accuracy of classification, pattern recognition, and visualization [7]. Computations in high dimensional spaces can become difficult due to the complexity of data which could lead to what is referred to as the curse of dimensionality and might lead to overfitting [7]. Dimension reduction is a terminology used when data with vast dimensions is reduced into lesser dimensions but ensures that it concisely conveys similar information. Dimension reduction techniques are used to typically solve machine learning problems during the stage of preprocessing to obtain better features for a classification or regression task. Dimension reduction algorithms have gained a lot of interest over the past few years. Before applying ML models, Dimension Reduction techniques provide a robust and also an efficient way to reduce the number of dimensions. Some techniques might be appropriate for some type of data but may not be appropriate for other types of data. As well, some DRTs are limited in application areas and constrained in scope. Dimensionality Reduction (DR) can be performed through feature selection and feature extraction. For feature selection, only a few related covariates are selected from the available covariates. All others are considered redundant and deemed not to have real explanatory effect. Feature extrac-
tion assumes that the dependent variable has a relationship with only a few linear combinations of many covariates. All covariates could have explanatory effect but this effect could be represented in a few linear combinations. Dimensionality reduction as a result facilitates classification, visualization, and the compression of high-dimensional data.

This study sought to investigate variants, applications areas, strengths and weaknesses of various DR techniques. The study also looked at the different data types that have been cited in literature to have been used for the different dimension reduction techniques considered in the paper. The data types considered for this paper are text, image, audio, video, time series and structured data. The paper is organized into five (5) main sections. It includes an Introduction, Discriminative features of Dimension Reduction Techniques, Dimension Reduction Techniques, Overview of Sufficient Dimension Reduction and Conclusion.

2. Methodology

Papers related to dimension reduction techniques were considered relevant for the review and based on the keywords “Dimension Reduction, Machine Learning, Linear Dimension Reduction Techniques, Non-linear Reduction Techniques”. The databases employed to search for the papers were ScienceDirect, Scopus, Google Scholar, IEEE Xplore and Mendeley. An integrative review was used for the study where 314 papers were reviewed. Review results were presented and summarized to include; the characteristics of various dimension reduction techniques, their strengths and weaknesses, application areas and variants of DR techniques addressing some identified limitations of the classical dimension reduction techniques.

3. Results

The area of dimension reduction has always been viewed broadly as important to statistical concept that can effectively reduce the dimensions whiles preserving the most important information. Principal component analysis was one of the earliest dimension reduction techniques which emerged as a general method for the reduction of multivariate observations in the early 20th century by [8] and was later independently developed by [9] and Factor analysis which was also consequently developed by [10] which are all linear techniques.

Dimension reduction can be classified into two main categories: linear and non-linear methods. For linear methods, a significant low-dimensional space is proposed to be discovered in data input with space that is high-dimensional, where the embedded data in the input space has a linear structure for linear reduction methods [11]. Also, techniques that are Non-linear were also developed to work with applications that have complex non-linear structures [12]. Other linear dimension reduction techniques considered for this review include Singular Value Decomposition (SVD) [13], Latent semantic analysis [14], Locality
Preserving Projections (LPP) [15], Independent Component Analysis (ICA) [16], Linear Discriminant Analysis (LDA) [17] and Projection pursuit [18]. The non-linear techniques include Kernel Principal Component Analysis (KPCA) [19], Multidimensional Scaling (MDS) [20], Isomap, Locally linear embedding [21], Self-Organizing Map (SOM) [22], Learning Vector Quantization (LVQ) [23], Uniform manifold approximation and projection (UMAP) [24] and the T-Stochastic neighbor embedding [25].

3.1. Linear Dimension Reduction Techniques

3.1.1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is one of the oldest and most widely used techniques which are an unsupervised linear dimension reduction technique. The idea of PCA is the reduction of the dimensions of a dataset, while it preserves as much variability as possible. Preserving as much variability as possible refers to the discovery of new variables that are linear functions of those in the original dataset. These linear functions maximize the variance and are also uncorrelated with each other [26]. Literature on PCA dates back from [8] and also [9] who coined the term principal components. PCA does not require any distributional assumptions as a descriptive tool. It is therefore an adaptive exploratory method that can be used different types of data [27].

The application areas of Principal Component Analysis (PCA) include machine learning, image and speech processing, computer vision, text mining, visualization, biometrics, robotic sensor data and facial recognition [28].

The identification of Principal Components (PCs) which is a set of uncorrelated features is the main aim of PCA. The largest amount of variance in data set is held by the first PC and in that order. Although it is a robust dimension reduction technique it has some limitations. The PCA transformation, despite its widespread use, relies on second-order statistics. The principal components can be highly statistically dependent though uncorrelated and this can lead to PCA failing to find the most compact description of the data. PCA geometrically models the data as a hyperplane embedded in a space that is ambient and requires a larger dimensional representation than would be found by a non-linear technique if the data components have non-linear dependencies. This has prompted the development of non-linear alternatives to PCA [29]. PCA methods also fail to account for outliers which are common in realistic training sets because they employ least squares estimation techniques [30].

There are many extensions of PCA to help address some of its challenges and also to improve efficiency. Robust PCA (RPCA) which was proposed by [31] was developed to analyze corneal images data. RPCA was found to be robust and could work with different kinds of image data. RPCA was found to also work well when outliers are present in image data [3]. Serneels and Verdonck [32] proposed an Expectation Robust PCA (ERPCA) and experimental results revealed that it was suitable for data of different sizes and also robust in dimension reduction when outliers are present.
An extension of PCA, the Local PCA (LPCA) was introduced by [29] and experimental results revealed that LPCA performed better than the classical PCA for image and speech data. Another extension of PCA, the Robust PCA (ROBPCA) which was also proposed by [33] was based on the DR technique, Projection Pursuit (PP) in applying a robust scatter matrix. Experimental results revealed that ROBPCA was more accurate and was also computationally faster than the traditional PCA. The Generalized PCA (GPCA), another extension of PCA proposed by [4] was developed with the main aim of dealing with data of high dimensions with the number of subspaces unknown. Different data sets including the 3D motion segmentation, clustering faces and temporal video segmentation applied GPCA and experimental results revealed that is was efficient [4]. Incremental 2D-PCA was proposed by [34] for videos particularly for tracking of moving objects. Multi-linear PCA (MPCA) was also proposed by [35] and it worked better that PCA and 2D-PCA in facial recognition. The Sparse PCA (SPCA) was also developed to manage sparsity of gene expression data [36]. [37] proposed the Generalized Power Sparse PCA (GP-SPCA) which was developed to overcome the curse of dimensionality issue of Dimension Reduction. [38] introduced the Random Permutation PCA (RP-PCA) and RP-2D-PCA. They were efficient in recognition of images in a biometric system. Bishop in 1999 proposed the Bayesian PCA which used maximum likelihood for latent variable model that is generative. Results revealed that through the Bayesian inference it is able to effectively reduce dimensions in latent space.

Although PCA is unsupervised in nature, a supervised PCA was proposed by [39] which is uniquely effective for classification and regression problems with high dimensional input data. The various data types applied on PCA from literature search are Text [40] [41], Image [42] [43], Audio [44] [45], Video [46] [47], Times series [48] and Structured data [49] [50].

3.1.2. Singular Value Decomposition (SVD)
One of the unsupervised Linear Dimension Reduction Technique (LDRT) is the Singular Value Decomposition (SVD) technique. The SVD is seen to be closely related to PCA and can be used in the computation of metric equations and problems in the form of data reduction [13]. Five mathematicians are credited with playing significant roles resulting in the existence of the SVD and development of theory. These mathematicians are Eugenio Beltrami, Camille Jordan, James Joseph Sylvester, Erhard Schmidt, and Hermann Weyl [51]. [52] however, was the one who has been credited with put finishing touches to the algorithm. SVD have been used in different areas by researchers. These include the area of digital image processing [53], taxonomic classification of biological sequences [54], pattern recognition [55], gene expression data [56], signal processing [57], Natural Language Processing (NLP), bio-informatics [54], and text summarization [54]. SVD is developed specifically for matrix decomposition and can be applied to any real-world matrix.

One drawback of SVD is that it is expensive computationally. It can however
be improved when random sampling is applied. SVD is also sensitive to non-linearities and outliers in data [58].

The non-iterative proper orthogonal decomposition for SVD was proposed by [59] to remove the influence of outliers in particle image velocimetry measurements. Also, a constrained SVD was proposed to work with sparsity and orthogonal issue of Singular value decomposition [60]. The multi-level SVD proposed by [61] was based on imputation method for efficient management and pre-processing of datasets collected from different sources. Fields such as life sciences, medical and education are some of the areas in which the technique is found to be useful. [28] also proposed FFT-PCA/SVD as a comparatively consistent and efficient than PCA/SVD algorithm in variable facial expressions recognition. Optimal dimension reduction is the main objective of SVD. The various data types applied on SVD from literature search are Text [62] [63], Image [64] [65], Audio [66] [67], video [68] [69], Text [62] [63], Times series [70] [109] [71] and Structured data [72] [73].

3.1.3. Latent Semantic Analysis (LSA)

An unsupervised LDR mapping technique, Latent Semantic Analysis (LSA) was designed specifically for text data and is developed on computations from PCA or SVD. LSA which was introduced by [14] is a DR technique introduced for improving the performance of the retrieval of an information retrieval system. This is done by grouping into same clusters related documents such that each document indexes the same words or almost the same words and relatively unrelated documents different words [74]. LSA is a technique that is vector based that is used to make comparisons and as well represent HD corpus text data into one of lower dimensions [5] [75]. LSA is premised on the theory of meaning which is engineered by psychology professor Thomas Landauer. He posited that meaning is constructed through the continuous experience with language [76].

The cognitive functions of LSA include the learning and understanding of the meaning of words [77] especially by students, episodic memory [78], discourse coherence [79], semantic memory [80], and the comprehension of metaphors [77]. LSA is able to produce measures of word-word, passage-passage, word-word relationships. LSA can also handle Synonymy problems to some extent depending on the nature of the dataset [75].

LSA has some limitations although it is seen to be an effective DR tool for text documents. It captures partially, the multiple meanings of a word (polysemy). This is because each word that occurs is treated as having the same meaning due to the word being represented as a single point in space. An example is the word “chair” occurring in a document that contains “The Chair of the Board” and also in a separate document containing “the chair maker” are considered the same. This behavior results in the representation of vectors as an average of all the different meanings of the words in the corpus, which may make it difficult for comparison purposes [14]. The effect of this limitation is however lessened due to the fact that words have a predominant sense throughout a corpus (i.e. not all
meanings are equally likely). Another drawback of LSI is the bag of words Model (BOW), which refers to texts being represented in an unordered collection of words. Multi-gram dictionary can be used however to address this limitation. It is used to find direct and indirect association as well as higher-order co-occurrences among terms [81]. Another limitation of LSA is that, it is unable to recover the intended optimal semantic factors. There has been some extensions to the LSA over the years which includes the technique introduced by [82], the Probabilistic LSA (PLSA). PLSA is effective for retrieval of information, ML, Natural Language Processing (NLP), and other related areas. Experimental results have revealed that the probabilistic method was substantially and consistently better than the standard LSA when different categories of linguistic data collections and text documents were accessed through indexing of documents automatically. [83] also proposed a Regularized Probabilistic LSA (RP-LSA) model to help in adjusting the model flexibility of the classical LSA and also to avoid over fitting issues. Experimental results have revealed that the RP-LSA reduces response and computational time [84]. The hk-LSA [85] was also introduced for the reduction of text documents dimensions. [86] introduced a Genetic Algorithm which was based on Latent Semantic Features (GALSFs) to improve text classification and experimental results revealed that GALSF outperformed the LSI. [87] introduced the Discriminative PLSA (DPLSA) which was proposed for facial recognition. DPLSA was successful in facial recognition based on single training sample [87]. The data type applied for LSA from literature search is Text data [88] [89].

3.1.4. Locality Preserving Projections (LPP)

Locality Preserving Projections (LPP) which was proposed by [15] is an unsupervised linear dimensionality reduction algorithm. They are projective maps that solve problems that are variational in nature and preserve optimally the neighborhood structure of the data set [15]. Because LPP is a classical linear method that also projects data along the usage of maximum variance, it is viewed as an alternative to PCA. LPP shares some of the properties of non-linear methods such as the Locally Linear Embedding or Laplacian Eigenmaps in terms of data representation [15].

There are a number of interesting perspectives to LPP. The objective criterion forms the classical linear techniques is minimized for the maps designed. LPP is seen as an appropriate alternative to PCA in pattern recognition, information retrieval and exploratory data analysis [15]. LPP has different application areas such as face recognition [90], image retrieval [91], image and video classification [15], pattern recognition [15], automatic speech recognition [6], and computer vision [92].

A drawback of LPP is that it has difficulties for reconstruction because the projection matrix in LPP is not orthogonal. As a result the orthogonal LPP (OLPP) was proposed by [93] such that projection matrix that is orthogonal can be obtained through a step by step procedure. The challenge with the OLPP al-
The fast and orthogonal version of LPP, called FOLPP was proposed by [94] to address the challenge of OLP. The algorithm minimizes simultaneously the locality and as well maximizes the globality under the orthogonal constraint.

There has been extensions to the LPP over time and these include the discriminant LPP (DLPP) which was proposed to remove noise which also a limitation of LPP from image data [95] as well as the uncorrelated DLPP (UDLPP) which was proposed to enhance recognition performance [96]. The parametric regularized LPP (PRLPP) algorithm was also proposed to overcome or mitigate the small sample size (SSS) problem. [97] also introduced a Locality-Regularized Linear Regression Discriminant Analysis (LLRDA) based on LL Regression Classification (LLRC) [97]. The Discriminant Locality preserving projections (DLPP) proposed by [98] is founded on the maximization of L1-norm for better pattern recognition performance. The algorithm was efficient when outliers are present and it also resolves small sample size issues which are some of the limitations of LPP. Another extension by [99] was the Soft Locality Preserving Map (SLPM) technique. It effectively reduces the feature vector dimensions. [100] introduced a Grassmann manifold (GLPP) which was based on the LPP. Results from experiments revealed that GLPP was effective for image/video classification. LPP has a singularity matrix issue and as a result 2D image vectors cannot be implemented. As a result, a 2D-LPP was proposed by [101]. 2D-LPP is able to save local information and helps in the detection of an intrinsic manifold structure of images which enhances recognition of images by using images of 2D matrices instead of 1D vector.

There are supervised versions of LPP which includes the Supervised Kernel LPP (SKLPP) proposed by [102] to enhance the accuracy of face recognition. An enhanced supervised locality preserving projections (ESLPP) was introduced by [93] for facial recognition. Cai also proposed a semi-supervised LPP (SSLPP) and experimental results revealed that the SSLPP technique improved LPP by the incorporation of the relevance degree information [103]. The data types applied on LPP from literature search are Text [104] [105], Image [106] [107], Audio [108] [109], Video, [110], Times series [111] [112] and Structured data [113].

### 3.1.5. Independent Component Analysis (ICA)

Independent Component Analysis (ICA) which was initially proposed by [6] is an unsupervised LDR statistical signal processing technique which is extensively used for the exploration of multi-channel data. The technique involves the modelling of data that is a linear mixture of independent source. Independent component analysis of a random vector involves the searching of a transformation that is linear resulting in the minimization of the statistical dependence between its components. ICA as a concept, may be seen as an extension of PCA, which only imposes independence up to the second order and as a result, defines directions that are orthogonal [16]. ICA has applications in blind identification, Bayesian detection, data analysis and compression and localization of sources.
In comparison to PCA, ICA has the ability to provide more components that are meaningful and could be extracted by the independent optimization condition instead of the maximization of variance in PCA [114]. ICA is also able to extract potentially more information from the data collected [115]. Apart from reducing the risk of overfitting, ICA allows for data reconstruction in the original space [115].

The major issue with ICA algorithms however, has to do with its stochasticity. Most ICA algorithms attempt to solve problems involving gradient-descent-based optimization such as maximization such as the non-Gaussianity of source S [116], mutual information minimization [117], and maximum likelihood estimation [118]. Also, in the case of high-dimensional signal space as in non-targeted data, the curse of dimensionality makes it more complicated. Consequently, it is not likely that local minima that are obtained from an algorithm run will be the global minima desired and therefore they are to be interpreted with great caution [119]. The Fast fixed point ICA, FastICA was suggested by [119] for the separation of linearity mixed source signals and complex values and has been employed for feature extraction as well as Blend Source Separation (BSS). BSS has many applications such as remote sensing, biomedical, finance, communication, signal processing and many others [120]. The mixed ICA/PCA was proposed by [121] through Reproducibility Stability approach which utilizes estimation through an iterative method to rank different sources which is utilized in the determination of dimensions of non-Gaussian subspaces from mixture of data. Another extension of ICA, Functions of Ranking and Averaging ICA by Reproducibility (RAICAR) was introduced by [122] to tackle the challenges spatial ICA face for functional Magnetic Reasoning Imaging (fMRI). When the signal mixture contains both Gaussian and non-Gaussian sources, Gaussian sources cannot be recovered by ICA and influence the estimate of non-Gaussian sources. The Mixed ICA/PCA via Reproducibility Stability (MIPReSt) was proposed by [121] to separate features of Gaussian and non-gaussian sources. The IICA-based feature extraction method was also proposed by [123] for automatic EEG artifact elimination. [124] also introduced the Capola ICA (CICA) which is based on measure of dependence of Hoeffding for time series data. [125] proposed the temporal ICA (tICA) to separate global noise signals when capturing fMRI data. [126] introduced a mixed method by combining techniques of ICA as well as kernel methods in the prediction of variations in the stock market. A hybrid of hierarchical clustering and ICA called ICAclust were combined so it could ignore issues like the normality of data as well as small temporal observations which is a feature of classical clustering [126]. Experimental results revealed that ICAclust performed better than traditional k-mean clustering [126] for temporal gene expression data. Other extensions of ICA included Probabilistic ICA (PICA) for fMRI [127], Sparse Gaussian ICA (SGICA) [128], Faster ICA under orthogonal constraint [129] and the Super Gaussian BSS via Fast-ICA with the approximation of Chebyshev Pade [120]. Types of data applied using ICA from literature...
search are Text data [130] [131], Image [132] [133], Audio/signals [123] [127] [134], Video [135], Times series [136] [137], and Structured data [138] [139].

3.1.6. Linear Discriminant Analysis (LDA)
Linear Discriminant Analysis (LDA) is a well-known and widely used supervised LDRT invented by [140], who used it successfully for the classification of flowers in his 1936 paper, “The use of multiple measurements in taxonomic problems”. LDA uses the linear combination of features as a linear classifier for the extraction of features and dimension reduction [17] [140]. It maximizes the ratio of the between-class variance to the within-class variance, thereby guaranteeing maximum class separability through its transformation of features into a lower dimensional space [141]. An advantage of LDA is that it is able to use information from both the features to create a new axis which in turn minimizes the variance and maximizes the class distance of the variables.

Although the LDA is one of the most well-used data reduction techniques, it has a number of limitations. If the dimensions are much higher than the number of samples in the data matrix, LDA is unable to find the lower dimensional space resulting in the within-class matrix becoming singular. This is known as the small sample problem (SSS). Different approaches have been proposed to solve this problem. The first approach proposed was to remove the null space of within-class matrix as was reported by [142]. The second approach utilizes the conversion from an intermediate subspace for example PCA to a within-class matrix to a full-rank matrix [143]. In the second approach, if linearity problem exists, that is if different classes are non-linearly separable, the LDA is unable to discriminate between these classes. Kernel functions can used as reported in [144] as a solution to the problem. The third approach which is a well-known one, is to apply the regularization problem in solving singular linear systems [143].

Different extension of LDA has been proposed to solve the SSS problem. This includes the regularized LDA (RLDA) [145], Direct LDA (DLDA) [146], PCA + LDA [147], Null LDA [148], Generalized EDA (GEDA) [149], kernel DLDA (KDLDA) [150] and PCA + LDA [147]. A semi supervised variant of LDA was proposed by [151] with its main objective of combining both labeled and unlabeled data for training LDA and to allow for the situation where the labeled data are few. Experimental results revealed that it performed better than the classical LDA.

Application of LDA includes facial recognition [152], text recognition [146] [152] [153], automatic diagnosis of machine operations [154], early detection of diseases [155], person reidentification [156], hand movement classification [157], motor imagery EEG [158], and ground water redox conditions [159]. The data types applied on LDA in literature search are Text [152] [160], Images [161] [162], Audio [163] [164], Video [165] [166], Time series [167] [168] and Structured data [169] [170].

3.1.7. Project Pursuit (PP)
Projection Pursuit (PP) proposed by [171] is an unsupervised non-parametric
LDRT. The idea originated from Kruskal in 1969. PP has been used widely for data exploration analysis. It is a technique that is able to find low dimensional linear projections and discovers patterns that are interesting for analysis [172]. A measure of “interestngness” is employed to this end, which is known as projection pursuit index (PP index). One key advantage of PP is its ability to fit different pattern recognition tasks flexibly, depending on the PP index used. Some areas of PP application are classification [173], clustering analysis [174], density estimation [175] and regression analysis [176]. One other advantage of PP, is its ability to mount new examples in the projection space after construction because of the out-of-sample mapping capability of PP.

Although Projection pursuit (PP) is unsupervised learning technique, it has successfully been applied in several domains for supervised analyses as well [177]. Many projection pursuit indices have been consequently developed to define interesting projections. Because most low-dimensional projections are approximately normal, a number of the projection pursuit indices that have been proposed are focused on non-normality. For example, the Legendre index [171], the Hermite index, the natural Hermite index and the entropy index and the moment index [178].

A limitation of PP is its high computational difficulty in finding optimal projection spaces for such cases. Notable PP optimization methods are the gradient techniques (Liu, 1988), the Newton-Raphson method [179], genetic algorithm [180], simulated annealing [181], and also the particle swarm optimization [182].

An extension of PP, the Project Pursuit regression was introduced by [183] to address the complexity issue and also to reduce computation cost of PP technique. Another extension of PP introduced by [184] was the Exploratory PP (EPP). Its objective is to combine an assemblage of data analytic techniques for low dimensional representation. [185] also developed a learning technique for outlier detection and this learning technique was based on PP. Random Projection (RP) was proposed by [130] for image and text data. Comparative tests revealed that in comparison with other techniques, RP was computationally less expensive and as well not affected by the curse of dimensionality [186]. [187] also introduced PP algorithm that was tree-based for the classification purposes with its key strength being its ability to find correlation between features. Also, with the interpretation of results, it provides 1D visualization of group differences. [18] also introduced an extension of PP, purposefully for the reduction of HDD with small sample size and was referred to as the PP framework. [188] introduced the Projection Pursuits Dynamic Cluster (PPDC) to address issues of HDD and non-linearity. [189] also proposed the Projection Pursuits Random Forest (PPRF) technique to solve problems of classification. Experimental results revealed that PPRF was more efficient than Random Forest (RF) when there was a separation of classes applying linear combination of features or when there is an increase in correlation between features. [190] proposed a supervised projection pursuit (SuPP) based on Jensen-Shannon divergence capable...
of working with missing data as well as large variable-to-sample ratio. When SuPP was combined with Naïve Bayes it performed better than compared to PCA and LDA on Iris data. [191] proposed a projection pursuit method based on semi-supervised spectral connectivity. Experimental results revealed that it was competitive in terms of classification accuracy using benchmark data sets. Semi-supervised variants of PP have also been developed [151]. Types of data from literature search to have been applied on PP are Text [192] and Image [193] [194].

3.2. Non-Linear Dimensionality Reduction Techniques

3.2.1. Kernel Principal Component Analysis (KPCA)

Kernel Principal Component Analysis (KPCA) is a Non-linear Dimension Reduction Technique (NLDRT) which was introduced by [19]. It is an extension of traditional PCA that works with High Dimension (HD) feature space employing the kernel method. The difference between KPCA and PCA is that, there is an eigen vector computation of kernel matrix with KPCA whiles PCA calculates the covariance matrix [195]. Also, non-linear principal components can be extracted with less computation power with KPCA. For data having non-linear manifolds, KPCA offers good encoding [196]. With KPCA, there is a non-linear transformation of the input data from the original input space to kernel for each data. A kernel matrix K is then formed from the inner product of the new feature. PCA is consequently applied the centralized K in the estimation of the covariance matrix of the new feature vectors [197]. Some extensively used kernels include Gaussian, Polynomial, and Hyperbolic tangent and Radial.

A drawback of the KPCA is that the cost of computation could be extremely high which could lead to attendant numerical problems of diagonalizing large matrices [197]. To overcome these drawbacks, Rosipal and Girolami proposed an EM algorithm for KPCA [197], which is an expectation-maximization approach for performing kernel principal component analysis and experimental results showed that it an efficient method computationally, especially for large number of data points. One drawback of this approach however is that it needs to still store the N × N kernel matrix, which limits its applicability in many large dataset problems.

The Block Adaptive KPCA (BAKPCA) was developed by [198] to add non-iteratively and dynamically new blocks and to remove old blocks of data. It is efficient in signal processing and also monitoring of processes. Greedy KPCA was also proposed by [199] to improve the performance of SVM classifier. Results showed that the greedy kernel PCA can significantly reduce complexity while it retains classification accuracy. Greedy KPCA was however found to be unsuitable for denoising. The Subset KPCA (SKPCA) was also introduced by [200] to reduce complexities in computations of KPCA for Dimension reduction as well as classification. The Robust KPCA has also been proposed by [201] to deal with outliers and to improve accuracy for protein classification. [202] introduced the discriminative PCA (dPCA) for discriminative analysis of multiple
datasets and has been applied in areas such as health data, sensor data, and facial images. Supervised Kernel Construction for Unsupervised PCA on Face Recognition was also proposed. Experimental results revealed that Supervised Kernel Construction for Unsupervised PCA (SK-PCA) performed better than KPCA with RBF kernel (RBF-PCA) using ORL and FERET databases. The types of data cited in literature that KPCA has been applied on and performed well are Image [203] [204], Audio [205] [206], Video [177] and time series data [207] [208].

3.2.2. Multidimensional Scaling (MDS)

Multidimensional Scaling (MDS) introduced by Kruskal and Wish in 1978 is an unsupervised NLDRT. The main objective of MDS is to preserve a measure of similarity or dissimilarity between pairs of data points. Multidimensional scaling is one of the techniques of dimensional reduction that has the ability to convert multidimensional data into a lower dimensional space while it keeps the intrinsic information. One main objective of MDS is to display graphically a set of given data making results much easier to understand and easy interpretability of complex structural data. Although there are a number of dimension reduction techniques, MDS has become much popular because of its simplicity as well as the various areas of application and has established itself as a standard tool for statisticians and researchers in general. In analysis involving MDS, spatial maps of objects are found given the similarity and dissimilarity of information that exists between available objects [209].

In MDS analysis, the data are embedded typically into a 2 or 3 dimensional map such that given the similarity or dissimilarity, information is matched closely to distances between points [210]. Objects of interest such as items, attributes, stimuli, respondents, etc. correspond to points such that those that are near to each other are similar empirically, and those that are far apart are seen to be different. MDS and factor analysis are seen to be similar but the advantage MDS has over factor analysis is the fact that MDS does not depend on the rigid assumptions of linearity and normality [210]. The only significant assumption of MDS is that the number of dimensions should be one less than the number of points which implies that three variables should at least be entered in the model and also at least two dimensions must be specified [209]. MDS has been applied in exploratory data analysis visualization and multivariate analysis. A limitation of MDS is that it is sensitive to outliers. An outlier detection mechanism was proposed by [211] using Robust MDS (RMDS) and based on geometric reasoning. Another limitation of MDS is that it suffers from increase in noise levels. This is as a result of the fact that MDS is dependent on the noise levels and number of dimensions. Extension of MDS has been proposed over time. The localized MDS which is a neighbor preserving DR algorithm was proposed by [212] to create data that is low dimensional and has a latent manifold structure. [213] also introduced a Local MDS (LMDS) which uses local information to construct a global structure and has been applied for graph drawing as well as proximity analysis. [214] in another study, introduced LMDS purposefully for
non-rigid 3D retrieval of shapes. Another variant of MDS known as the MDS+ was proposed by [215] to act uniquely as a shrinkage function that is asymptotically optimal. MDS+ is able to overcome the external estimation issue for embedding dimensions and also computes the optimal number of lower dimensions into which the dataset can be embedded. The MDS-T was proposed by [216] for the analysis of psychological data. A Partially Supervised Metric Multi-dimensional Scaling Algorithm for Textual Data Visualization has also been developed [217]. The various data types applied on MDS in literature search are Text [210] [218], Image [219] [220], Audio [221] [222], Video [223], Times series [224] [225] and structured data [226].

### 3.2.3. ISOMAP

Another popular unsupervised NLDRT whose objective is to intrinsically find structures of data from a non-linear manifold is the ISOMAP. The algorithm which was proposed by [227] attempts to extract parameterizations for data sets into a low dimensional space to form a high dimensional space such that there is a preservation of the pairwise geodesic distances so that nearby points are far in high dimensional space map to nearby points that are far in low dimensional space. The distinguishing feature of ISOMAP is its ability to get a lower dimensional representation of data, whiles the geodesic distance is preserved [227]; [228]. ISOMAP combines the major characteristics of PCA and MDS in terms of computational efficiency, asymptotic convergence guarantees and global optimality with the flexibility to an extensive class of non-linear manifolds. The ISOMAP approach basically builds on the traditional MDS but the distinguishing property is that is seeks to preserve the intrinsic geometry of the data which is captured in the geodesic manifold distance between the pairs of data points [227]. ISOMAP has been efficient when used in detecting irregularities from real time video analytics [229]. A path based ISOMAP was proposed by [230] for the enhancement of memory and as well as time complexities. Geodesic path is used in this approach to find the low dimension embedding. Some of the drawbacks of ISOMAP are that, it is computationally expensive and performs poorly when manifold is not well sampled and contains holes [230]. The Landmark Isomap (L-Isomap) was presented by [231] to enhance the Isomap scalability.

ISOMAP has successfully been applied on the condition of urban road traffic [232], speech summarization and crack identification in materials [233], facial recognition [234]. Supervised versions of Isomap have been proposed. This includes Supervised Isomap for classification [235], Supervised Isomap with dissimilarity measures in embedding learning [236] and Supervised Isomap for plant leaf image classification [237]. The data types applied on Isomap from literature search are Text [237] [238], Images [239], Audio [240] [241], Video [242] [243].

### 3.2.4. Locally Linear Embedding (LLE)

Locally Linear Embedding (LLE) which is an unsupervised NLDRT and intro-
duced by [244] aims to preserve only local properties of data. LLE as a learning algorithm involves the computation of low-dimensional neighborhood preserving embedding of inputs that are of high dimensions in nature. LLE has the ability to learn the global structure of non-linear manifolds like those from images of faces or documents of text by exploiting the local symmetries of linear reconstructions. LLE has been applied successfully in a wide range of applications which includes face recognition and remote sensing [177]. More recently, LLE has more recently been used in MRI which includes functional MRI [245], shape analysis of the hippocampus in AD, diffusion tensor imaging, breast lesion segmentation, feature fusion and image classification [246].

LLE is popular among researchers because of its ability to deal with large data sets of high dimensional data and its non-iterative way of finding embedding. LLE however has some drawbacks which include sensitivity to noise, the inability to deal with novel data and the inevitable ill-conditioned Eigen problems. Another drawback is that, LLE as an unsupervised technique which assumes that all data reside in a continuous manifold but this is not the case for problems of multiple class classification. Some efforts have recently been made to develop extensions of the classical LLE. [247] proposed the weighted locally linear embedding (WLLE) for dimension reduction. This was to discover the intrinsic structures of data, such as global distributions neighborhood relationships, and clustering. One major advantage of WLLE is to optimize the intrinsic structure process discovery by avoiding unreasonable neighbor searching and also at the same time is able to adapt to novel data. Simulated experiments revealed that the WLLE performed better in dimension reduction and manifold learning than the classical LLE and was more robust to changes in parameter. [248] proposed Local Smoothing for Manifold Learning purposely for outlier detection and noise reduction. Experimental examples with image datasets revealed that manifold learning methods in combination with weighted local linear smoothing give more accurate results. [249] proposed a dimensional reduction technique that was non-linear and computes a low-dimensional and preserving of neighborhood embedding of high dimensional data. Other extensions of LLE include the Hessian Locally Linear Embedding (HLLE) proposed by [250] which is constructed based on Incremental LLE for dynamically adding new data and also preserves significant features of the original data while whiles performing DR.. The Modified LLE (MLLE) was proposed by [251] using multiple weights. A Multiple Manifold LLE proposed by [252] is an approach that allows for learning multiple manifolds for multiple classes and is efficient in classification and objects recognition.

A Supervised version of LLE was proposed by [250] for plant classification based on images of leaves. A semi supervised version of LLE was also proposed for classification of leave images [246]. The types of data applied on LLE from literature search are image [245] [253], Audio [254] [255] and Video [256] [257].

3.2.5. Self-Organizing Map
Self-Organizing Map (SOM) is a cognitive learning unsupervised NLDRT which
was introduced by [258]. SOM is an architecture that was suggested for Artificial Neural Networks. One of the properties of SOM is that it can create effectively spatially organized internal representations of many input signals of features and their abstractions. As a result, from the self-organizing process of SOM, it is able to identify semantic relationships in sentences. SOM has performed particularly well in pattern recognition tasks involving signals that are very noisy. These maps of SOM have been used successfully in speech recognition [258]. SOM is also seen as a very good tool in exploratory phase of data mining [258]. SOM has the ability to reduce complex problems down to data mappings that easily be interpreted. SOM are also capable of handling different types of problems while providing an interactive and useful summary of the data. As well, SOMs are capable of clustering large and complex data sets. SOM however has some drawbacks. It requires data that is sufficient and necessary in order to develop meaningful clusters. Also, the weight of vectors should be based on the successful grouping of the data and distinguishing inputs. Scanty data or extraneous data in the weights may add randomness to the groupings. Another drawback of SOM is that, obtaining a perfect mapping is difficult in cases where groupings are unique within the map. Application areas of SOM include intrusion detection [259], noise removal from spectral images [260], massive documents automatic organization [261] and also weather and crop production rate prediction [262].

Some extensions of SOM include the Community SOM (CSOM) with the specialty of enhancing the overall learning process of SOM. The hybrid approach of SOM was also proposed by [263] for prediction of huge volume of text documents based on the combination of probability distribution and SOM with the Naive Bayes. Experimental results revealed that it achieved better classification accuracy. A text mining novel algorithm approach of SOM was also developed by [264] to enhance the performance of SOM. [262] also proposed a correntropy based technique which was used in place of Mean Square Error (MSE) and used by SOM to enhance the efficiency of SOM in the presence of outliers. [265] also introduced a multistage Visual Analytical (VA) method with SOM flow. The algorithms were to iteratively refine clusters to help in time series data analysis. SOM is suitable for all kinds of data which includes Text [266] [267], Image [268] [269], Audio [270] [271], Video [272] [273], Time series [274] and Structured data [275] [276].

3.2.6. Learning Vector Quantization (LVQ)

Learning Vector Quantization (LVQ) is a competitive based neural network supervised NLDRT which is similar to SOM. The technique which was introduced by Kohonen in 1995 is a technique specifically for statistical pattern recognition with the aim of learning prototypes representing class regions. Voronoi partitions are yielded when the class regions are defined by hyperplanes between prototypes. Several variants of LVQ have been developed by Teuvo Kohonen since the late 1980s [277]. LVQ techniques are similar to SOM in the sense that
all output nodes compete and the winning node is selected according to its similarity to the input pattern presented. Unlike SOM, LVQ updates only the winning neuron and as a result, the output feature space is not topologically ordered. LVQ is mostly applied to find the feature map after analysis on training data is performed using SOM. Unsupervised learning can also be carried out on LVQ for purposes of clustering. LVQ can also be trained without labels by unsupervised learning for clustering purposes [278]. An advantage of LVQ classifiers is that they are intuitive and simple to understand which is an advantage it has over SVMs. Although SVM is considered to be robust, LVQ has shown to be a valuable alternative. LVQ classifiers are also able to deal with multi-class problems. LVQ as a result has been applied in different areas which includes its classification accuracy [279]. LVQ have been applied in speech recognition and control pattern recognition. LVQ however has two major limitations which are slow convergence and unstable behavior. The problem of convergence has been solved using the Genetic algorithm introduced by [280] which increased the classification performance rate prior to power quality disturbances. The LVQ family consists LVQ1, LVQ2, and there are improved versions namely LVQ2.1, LVQ3, OLVQ1, OLVQ3, Multipass LVQ, and HLVQ algorithms. There has been other extension of LVQ which includes the LVQ based artificial neural network classifiers proposed by [281]. The algorithm was developed for different kinds of methods for signal processing to help in the recognition and classification of arrhythmia from the ECG signals. LVQ in combination with Gabor filter was successfully applied to recognize different facial expressions. Different variants of LVQ were also proposed by [282] to help improve accuracy of classification for different kinds of data.

[283] combined PCA and LVQ for classification for strategies of mobile learning employed by college students. Also, the dissimilarities based Generalized LVQ (GLVQ) was proposed by [284] to help in the enhancement of classification accuracy. The Kernel based RSLVQ which was proposed by [285] used the general gram matrix to handle complex non vector data. A hybrid approach of LVQ proposed by [286] was the random Fourier features extraction for an LVQ matrix for the provision of smaller and discriminative feature sets. The suitable data types from literature search for LVQ are Text [287] [288], Image [289] [290], Audio [291] [292], Video [293] [294], Time series [295] [296] and Structured data [289] [297].

3.2.7. t-Stochastic Neighbor Embedding (t-SNE)

t-Stochastic Neighbor Embedding (t-SNE) is an unsupervised NLDRT which was introduced by [298]. The technique is a variation of the Stochastic Neighbor Embedding introduced by [25] whose main objective is the construction of probability distributions from pairwise distances such that larger distances correspond to smaller probabilities and vice versa. T-SNE is the most commonly used learning method in single-cell analysis. T-SNE however has some limita-
tions which includes slow computation time, its inability to meaningfully represent very large datasets and loss of large scale information [299]. A multi-view Stochastic Neighbor Embedding (mSNE) was proposed by [299] and experimental results revealed that it was effective for scene recognition as well as data visualization [299]. The suitable data types for t-SNE are text [300] [301], Image [302] [303], Audio [241] [304], video [217], Time series [305], and Structured data [306] [307].

3.2.8. Uniform Manifold Approximation and Projection (UMAP)
Uniform manifold approximation and projection (UMAP) is an unsupervised NLDRT proposed by [24]. It was constructed based on a theoretical framework in Riemannian geometry and algebraic topology. [24] credits their work on the mathematical work on the mathematical foundations of Laplacian Eigen maps of Belkin and Niyogi. UMAP explores the issue of uniform data distribution on manifolds through the combination of the work of David Spivak [308] and the Riemannian geometry. UMAP at a high level uses the approximations of the local manifold and then patches their local fuzzy simplicial representation of sets to construct a topological representation of high dimensional data. A similar process can be used to construct an equivalent topological representation given some low dimension representation of data. The data representation is then optimized in the low dimensional space to minimize cross entropy between the two topological representations. UMAP is seen to compete well with t-SNE which is currently a robust technique for visualization quality in DR. UMAP also preserves more of the global structure with a better run time performance than t-SNE [24]. Also, the topological foundations of UMAP enable it to significantly scale larger data set than are feasible for t-SNE. UMAP also does not have computational restrictions on embedding dimension hence making it viable for dimension reduction. UMAP is similar to t-SNE but probably has a higher processing speed and better visualization. The main disadvantage of UMAP is the fact that it is a relatively new technique and therefore lacks maturity.

UMAP algorithm was compared to PCA, t-SNE using MSI data sets acquired from pancreas and human lymphoma samples. Results from the study revealed that that UMAP is competitive with t-SNE in terms of visualization and it is also well-suited for the dimensionality reduction of large (>100,000 pixels) MSI data sets. The runtime also markedly reduced by fourfold in comparison with the state of art t-SNE [309]. UMAP was also evaluated as an alternative to t-SNE for single-cell data [310]. The data types applied on UMAP from literature search are Image [309] [311] [312], Audio [313], Video [314] [315] and Structured [24]; [316] [317].

3.3. Overview of Sufficient Dimension Reduction
Sufficient dimension reduction (SDR) is a feature extraction class of methods for classification as well as regression. Its main purpose is to reduce the size of data set with a lot of dimensions to just few features of importance with the potential
of establishing important relationship between variables through improved visualization. Sufficient dimension in recent times has undergone significant development. This could partly be because of increase in demand for methodologies that are able to effectively work with high-dimensional data in the era of big data.

Some of the earliest methods of SDR, include the seminal sliced inverse regression (SIR) by [318], the sliced average variance estimation (SAVE) by Cook and Weisberg [319], principal Hessian direction (PHD) [320] [321], minimum average variance estimation (MAVE) [322], simple contour regression (SCR) [323], the inverse regression (IR) by [324] and also the directional regression (DR) by [325]. Other methods include, the simple contour regression (SCR) [323], Fourier transform method proposed by [326] and [327], sliced regression [328], the Kullback-Leibler based approach which was also proposed by [329] and the ensemble method [330]. There is also the partial least square (PLS) [331] [332], sufficient component analysis (SCA) [333], kernel dimension reduction (KDR) [334]. Other methods include, but are not limited to, the method proposed by [335] for exponential family predictors and the methods suggested by [336] with exponential family inverse predictors and the likelihood based dimension reduction method which was proposed by [337]. The limitation of most of the SDR techniques however, is that they require linearity condition which includes SIR and SAVE [338] or the constant variance condition [320] [321] or even both to hold for some techniques, which is practically difficult to verify. Also, although it is well know that inverse regression methods are easy to compute relatively and also practically useful, many of them fail in estimating the central subspace exhaustively by Cook in 1998 [328]. For example, the PHD is known to detect only patterns that are non-linear and estimates directions in only the central subspace [339]. On the other hand, SIR [318], Slicing regression and IR may not perform well if the relationship of the regression is highly symmetric [321]. [340] pointed out that SIR is also very sensitive to outliers, and at some extreme situations, the estimators provide very wrong efficient dimension reduction directions simply orthogonal to the true dimension reduction directions [340]. [341] also pointed out that, SAVE cannot be √n consistent and that it is not consistent when each slice contains a fixed number of data points that do not depend on n, where n is the sample size [341].

4. Conclusions

The area of Dimension reduction is becoming very relevant in different application areas such as healthcare, economics, environment, social science, agriculture, and many more because of the sheer amount of data being generated in the era of big data. Big data is a phenomenon that was not anticipated by the scientists who contributed to groundbreaking mathematical and statistical models that are still relevant till date. The earliest Dimension reduction techniques were the linear PCA and the linear LDA. Although robust they have their limitations. As a result, variants of these techniques such as the LPCA, RPCA, ROBPCA,
GPCA etc. in the case of PCA have been proposed to address these limitations. Variants of LDA also include RLDA, DLDA, Null LDA, PCA + LDA, kernel DLDA etc. Other linear dimension reduction techniques such as the SVD, LSI, PP, ICA and LPP have been developed with their own unique strengths. One limitation of linear dimension reduction techniques is their inability to perform well when the data has non-linear structures. Non-linear Dimension techniques have consequently been proposed to address this limitation. The KPCA for example is the non-linear version of PCA. Other non-linear techniques include the MDS, ISOMAP, LLE, SOM, LVQ, t-SNE and UMAP. The aim of PCA is the preservation of variance; SVD is optimal dimension reduction; LSI/LVQ is classification accuracy; LPP, KPCA, MDS, LLE and Isomap are the extraction of manifolds; SOM looks at prediction accuracy and t-SNE and UMAP is the preservation of neighborhood. Sufficient dimension reduction (SDR) techniques are being explored recently, with Li proposing the first technique, the seminal sliced inverse regression.

The area of a proper fusion between the dimension reduction techniques and statistics should be explored for further research. Also, most of the dimension reduction techniques reviewed are unsupervised learning techniques. Further research should be carried out on classical supervised dimension reduction techniques as well as semi-supervised techniques. Further research should also be carried out to illustrate practical implementation of DR techniques using example data.

**Author's Contribution**

The idea was developed by SN. Literature was reviewed by all authors. All authors contributed to manuscript writing and approved the final manuscript.

**Funding**

The study attracted no funding.

**Conflicts of Interest**

The authors declare no conflicts of interest regarding the publication of this paper.

**References**


Vision, 1, 362-369.


https://doi.org/10.1080/01638539809545028

https://doi.org/10.1207/s15516709cog2502_1

https://doi.org/10.1016/j.jmla.2001.2798

https://doi.org/10.3390/ijgi6070195

https://doi.org/10.1002/wcs.1254

https://doi.org/10.1186/s12967-014-0324-9

https://doi.org/10.1023/A:1007617005950

https://doi.org/10.1007/11430919_72

https://doi.org/10.1145/3209978.3210189


https://doi.org/10.1016/j.eswa.2014.03.041

https://doi.org/10.1007/s11063-018-9852-2


https://doi.org/10.1007/978-1-84628-754-1

https://doi.org/10.1016/j.neucom.2006.12.004

https://doi.org/10.1145/1027527.1027530


https://doi.org/10.1109/MLSP.2015.7324362


[283] Hu, S., Gu, Y. and Jiang, H. (2016) Study of Classification Model for College Students’ M-Learning Strategies Based on PCA-LVQ Neural Network. 2015 8th *International Conference on Biomedical Engineering and Informatics (BMEI)*, Shenyang, 14-16 October 2015, 742-746. [https://doi.org/10.1109/BMEI.2015.7401601](https://doi.org/10.1109/BMEI.2015.7401601)


Series Using Auto-Correlation Coefficients and LVQ—Neural Network. *JISA* 2013, Piraeus, 10-12 July 2013, 1–6. https://doi.org/10.1109/JISA.2013.6623678


S. Nanga et al.