Review of Dimension Reduction Methods

Salifu Nanga[1*], Ahmed Tijani Bawah[2], Benjamin Ansah Acquaye[3], Mac-Issaka Billa[4], Francis Delali Baeta[5], Nii Afotey Odai[6], Samuel Kwaku Obeng[7], Ampem Darko Nsiah[8]

[1]Department of Basic Science, School of Basic and Biomedical Science, University of Health and Allied Sciences, Ho, Ghana
[2]Department of Medical Laboratory Science, School of Allied Health Sciences, University of Health and Allied Sciences, Ho, Ghana
[3]SANA Research and Consult, Accra, Ghana
[4]University of Management Studies, Weija-Accra, Ghana
[5]Department of Mathematics and Statistics, Ho Technical University, Ho, Ghana
[6]Health Information Unit, Tema General Hospital, Tema, Ghana
[7]Department of Data Quality Monitoring, Ghana Statistical Service, Accra, Ghana
[8]Criminal Investigation Department, Ghana Police Service, Accra, Ghana
Email: *snanga@uhas.edu.gh, ahmed024gh@yahoo.com, acquayeansa@gmail.com, macbilla@ymail.com, baetadelali@gmail.com, afotey777@gmail.com, samobeng2002@yahoo.com, ampemn@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 (ROBCA) 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-

gorithm is that, it is computationally expensive. The fast and orthogonal version of LPP, called FOLPP was proposed by [94] to address the challenge of OLPP. 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 (LL-RDA) 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

[16]. 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 "interestingness" 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 whiles 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 is 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 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 $\sqrt{n}$ 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

[1] Shamsolmoali, P., Kumar Jain, D., Zareapoor, M., Yang, J. and Afshar Alam, M. (2019) High-Dimensional Multimedia Classification Using Deep CNN and Extended Residual Units. *Multimedia Tools and Applications*, **78**, 23867-23882. https://doi.org/10.1007/s11042-018-6146-7

[2] Brodinová, Š., Filzmoser, P., Ortner, T., Breiteneder, C. and Rohm, M. (2019) Robust and Sparse K-Means Clustering for High-Dimensional Data. *Advances in Data Analysis and Classification*, **13**, 905-932. https://doi.org/10.1007/s11634-019-00356-9

[3] De La Torre, F. and Black, M.J. (2001) Robust Principal Component Analysis for Computer Vision. *Proceedings of the IEEE International Conference on Computer*

Vision, **1**, 362-369.

[4] Vidal, Rene, Ma, Y. and Sastry, S. (2005) *Generalized Principal Component Analysis* (*GPCA*), **27**, 1945-1959. https://doi.org/10.1109/TPAMI.2005.244

[5] Tang, B., Heywood, M.I. and Shepherd, M. (2005) Comparing and Combining Dimension Reduction Techniques for Efficient Text Clustering. In: HillolKargupta, J.S., Chandrika, K. and Arnold, G., Eds., *Proceeding of SIAM International Workshop on Feature Selection for Data Mining*, 17-26.

[6] 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

[7] Ji, C., Li, Y., Qiu, W., Jin, Y., Xu, Y., Awada, U. and Qu, W. (2012) Big Data Processing: Big Challenges. *Journal of Interconnection Networks*, **13**, 1-19. https://doi.org/10.1142/S0219265912500090

[8] Pearson, K. (1901) LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London*, *Edinburgh*, *and Dublin Philosophical Magazine and Journal of Science*, **2**, 559-572. https://doi.org/10.1080/14786440109462720

[9] Hotelling, H. (1933) Analysis of a Complex of Statistical Variables into Principal Components. *The Journal of Educational Psychology*, **24**, 417-441, 498-520. https://doi.org/10.1037/h0070888

[10] Spearman, C. (1904) General Intelligence. *The American Journal of Psychology*, **15**, 201-292. https://www.jstor.org/stable/1412107?origin=crossref&seq=1#metadata_info_tab_contents https://doi.org/10.2307/1412107

[11] Webb, A.R. and Copsey, K.D. (2011) Statistical Pattern Recognition: Third Edition. John Wiley & Sons, Hoboken. https://doi.org/10.1002/9781119952954

[12] Pratihar, D.K. (2011) Non-Linear Dimensionality Reduction Techniques. In: Wang J., Ed., *Encyclopedia of Data Warehousing and Mining*, *Second Edition*, IGI Global, Pennsylvania, 1416-1424. https://doi.org/10.4018/978-1-60566-010-3.ch219

[13] Modarresi, K. (2015) Unsupervised Feature Extraction Using Singular Value Decomposition 2 Feature Extraction and Dimensional Reduction for Modern Data. *Procedia Computer Science*, **51**, 2417-2425. https://doi.org/10.1016/j.procs.2015.05.424

[14] Deerwester, S., Furnas, G.W., Landauer, T.K. and Harshman, R. (1990) Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, **41**, 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

[15] He, X.F. and Niyogi, P. (2003) Locality Preserving Projections. *Neural Information Processing Systems*, **16**, 153-160.

[16] Comon, P. (1994) Independent Component Analysis, a New Concept? *Signal Processing*, **115**, 27-37.

[17] 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

[18] Espezua, S., Villanueva, E., Maciel, C.D. and Carvalho, A. (2015) Neurocomputing a Projection Pursuit framework for Supervised Dimension Reduction of High Dimensional Small Sample Datasets. *Neurocomputing*, **149**, 767-776.

https://doi.org/10.1016/j.neucom.2014.07.057

[19] Schölkopf, B., Smola, A. and Müller, K.R. (1997) Kernel Principal Component Analysis. In: Gerstner, W., Germond, A., Hasler, M. and Nicoud, J.D., Eds., *Artificial Neural Networks—ICANN*97, Springer, Berlin, 583-588. https://doi.org/10.1007/BFb0020217

[20] Meulman, J.J. (1992) The Integration of Multidimensional Scaling and Multivariate Analysis with Optimal Transformations. *Psychometrika*, **57**, 539-565. https://doi.org/10.1007/BF02294419

[21] Lee, J.A., Lendasse, A. and Verleysen, M. (2004) Nonlinear Projection with Curvilinear Distances: Isomap versus Curvilinear Distance Analysis. *Neurocomputing*, **57**, 49-76. https://doi.org/10.1016/j.neucom.2004.01.007

[22] Kohonen, T. (1998) The Self-Organizing Map. *Neurocomputing*, **21**, 1-6. https://doi.org/10.1016/S0925-2312(98)00030-7

[23] Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J. and Torkkola, K. (1996) LVQ-PAK: The Learning Vector Quantization Program Package. Helsinki University of Technology, Espoo. http://cis.legacy.ics.tkk.fi/research/papers/lvq_tr96.ps.Z

[24] McInnes, L., Healy, J. and Melville, J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. http://arxiv.org/abs/1802.03426

[25] Hinton, G. and Roweis, S. (2003) Stochastic Neighbor Embedding. In: Becker, S., Thrun, S. and Obermayer, K., Eds., *Advances in Neural Information Processing Systems*, MIT Press, Massachussets, Vol. 15, 857-864.

[26] Ahmadkhani, S. and Adibi, P. (2016) Face Recognition Using Supervised Probabilistic Principal Component Analysis Mixture Model in Dimensionality Reduction without Loss Framework. *IET Computer Vision*, **10**, 193-201. https://doi.org/10.1049/iet-cvi.2014.0434

[27] Jollife, I.T. and Cadima, J. (2016) Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A*: *Mathematical*, *Physical and Engineering Sciences*, **374**, 1-16. https://doi.org/10.1098/rsta.2015.0202

[28] 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] Kambhatla, N. and Leen, T.K. (1997) Dimension Reduction by Local Principal Component Analysis. *Neural Computation*, **9**, 1493-1516. https://doi.org/10.1162/neco.1997.9.7.1493

[30] Seheult, A.H., Green, P.J., Rousseeuw, P.J. and Leroy, A.M. (1989) Robust Regression and Outlier Detection. *Journal of the Royal Statistical Society*: *Series A* (*Statistics in Society*), **152**, 133-134. https://doi.org/10.2307/2982847

[31] Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T., Cohen, K.L. and Cohen, K.L. (1999) Robust Principal Component Analysis for Functional Data. *Test*, **8**, 1-73. https://doi.org/10.1007/BF02595862

[32] Serneels, S. and Verdonck, T. (2008) Principal Component Analysis for Data Containing Outliers and Missing Elements. *Computational Statistics and Data Analysis*, **52**, 1712-1727. https://doi.org/10.1016/j.csda.2007.05.024

[33] Hubert, M., Rousseeuw, P.J. and Vanden Branden, K. (2005) ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics*, **47**, 64-79. https://doi.org/10.1198/004017004000000563

[34] Wang, T. and Gu, I.Y.H. (2007) Object Tracking Using Incremental 2D-PCA Learning and Ml Estimation. 2007 *IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP* 07, Honolulu, 15-20 April 2007, 933-936. https://doi.org/10.1109/ICASSP.2007.366062

[35] Lu, H., Member, S., Plataniotis, K.N.K., Member, S. and Venetsanopoulos, A.N. (2008) MPCA : Multilinear Principal Component Analysis of Tensor Objects. *IEEE Transactions on Neural Networks*, **19**, 18-39. https://doi.org/10.1109/TNN.2007.901277

[36] Zou, H., Hastie, T. and Tibshirani, R. (2006) Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, **15**, 265-286. https://doi.org/10.1198/106186006X113430

[37] Journ, M. and Richt, P. (2010) Generalized Power Method for Sparse Principal Component Analysis. *Journal of Machine Learning Research*, **11**, 517-553.

[38] Kumar, N., Singh, S. and Kumar, A. (2018) Random Permutation Principal Component Analysis for Cancelable Biometric Recognition. *Applied Intelligence*, **48**, 2824-2836. https://doi.org/10.1007/s10489-017-1117-7

[39] Barshan, E., Ghodsi, A., Azimifar, Z. and Zolghadri Jahromi, M. (2011) Supervised Principal Component Analysis: Visualization, Classification and Regression on Subspaces and Submanifolds. *Pattern Recognition*, **44**, 1357-1371. https://doi.org/10.1016/j.patcog.2010.12.015

[40] Uysal, A.K. (2018) On Two-Stage Feature Selection Methods for Text Classification. *IEEE Access*, **6**, 43233-43251. https://doi.org/10.1109/ACCESS.2018.2863547

[41] Karami, A. (2019) Application of Fuzzy Clustering for Text Data Dimensionality Reduction. *International Journal of Knowledge Engineering and Data Mining*, **6**, 289. https://doi.org/10.1504/IJKEDM.2019.102487

[42] Jégou, H. and Chum, O. (2012) Negative Evidences and Co-Occurences in Image Retrieval: The Benefit of PCA and Whitening. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y. and Schmid, C., Eds., *Lecture Notes in Computer Science*, Springer, Berlin, 774-787. https://doi.org/10.1007/978-3-642-33709-3_55

[43] Chamundeeswari, V.V., Singh, D. and Singh, K. (2009) An Analysis of Texture Measures in PCA-Based Unsupervised Classification of SAR Images. *IEEE Geoscience and Remote Sensing Letters*, **6**, 214-218. https://doi.org/10.1109/LGRS.2008.2009954

[44] Rajitha, G. and Raju, K.U. (2018) PCA-ICA Based Acoustic Ambient Extraction. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, **3**, 51-59.

[45] Dong, S., Hu, R., Tu, W., Zheng, X., Jiang, J. and Wang, S. (2012) Enhanced Principal Component Using Polar Coordinate PCA for stereo Audio Coding. 2012 *IEEE International Conference on Multimedia and Expo*, Melbourne, 9-13 July 2012, 628-633. https://doi.org/10.1109/ICME.2012.22

[46] Bouwmans, T., Javed, S., Zhang, H., Lin, Z. and Otazo, R. (2018) On the Applications of Robust PCA in Image and Video Processing. *Proceedings of the IEEE*, **106**, 1427-1457. https://doi.org/10.1109/JPROC.2018.2853589

[47] Arunnehru, J. and Geetha, M.K. (2013) Motion Intensity Code for Action Recognition in Video Using PCA and SVM. In: Prasath, R. and Kathirvalavakumar, T., Eds., *Lecture Notes in Computer Science*, Springer, Cham, 70-81. https://doi.org/10.1007/978-3-319-03844-5_8

[48] Li, W., Han, M. and Feng, S. (2018) Multivariate Chaotic Time Series Prediction:

Broad Learning System Based on Sparse PCA. In: Gedeon, T., Wong, K.W., Lee, M. and Xu, C.Z., Eds., *International Conference on Neural Information Processing*, Springer International Publishing, Berlin, 56-66.

[49] Divya, S. and Padmavathi, G. (2014) A Novel Method For detection of Internet Worm Malcodes Using Principal Component Analysis and Multiclass Support Vector Machine. *International Journal of Security and Its Applications*, **8**, 391-402. https://doi.org/10.14257/ijsia.2014.8.5.34

[50] Shin, H., Jeong, H., Park, J., Hong, S. and Choi, Y. (2018) Correlation between Cancerous Exosomes and Protein Markers Based on Surface-Enhanced Raman Spectroscopy (SERS) and Principal Component Analysis (PCA). *ACS Sensors*, **3**, 2637-2643. https://doi.org/10.1021/acssensors.8b01047

[51] Stewart, G.W. (1993) On the Early History of the Singular Value Decomposition. *Bulletin of the Geological Society of America*, **76**, 287-306. https://doi.org/10.1137/1035134

[52] Golub, G.H. and Reinsch, C. (1971) Singular Value Decomposition and Least Squares Solutions. *Numerische Mathematik*, **14**, 403-420. https://doi.org/10.1007/BF02163027

[53] Cao, L.J. (2006) Singular Value Decomposition Applied to Digital Image Processing. Arizona State University Polytechnic Campus, Mesa, 1-15.

[54] 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

[55] Lassiter, A. (2013) Handwritten Digit Classification and Reconstruction of Marred Images Using Singular Value Decomposition. Virginia Tech.

[56] 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

[57] 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

[58] Silvério-Machado, R., Couto, B.R.G.M. and Dos Santos, M.A. (2015) Retrieval of Enterobacteriaceae Drug Targets Using Singular Value Decomposition. *Bioinformatics*, **31**, 1267-1273. https://doi.org/10.1093/bioinformatics/btu792

[59] 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

[60] 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

[61] 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

[62] Murty, M.R., Murthy, J.V. and Reddy, P.V.G.D. (2011) Text Document Classifica-

tion basedon Least Square Support Vector Machines with Singular Value Decomposition. *International Journal of Computer Applications*, **27**, 21-26. https://doi.org/10.5120/3312-4540

[63] Williams, T.P. and Gong, J. (2014) Predicting Construction Cost Overruns Using Text Mining, Numerical Data and Ensemble Classifiers. *Automation in Construction*, **43**, 23-29. https://doi.org/10.1016/j.autcon.2014.02.014

[64] Singh, G. and Goel, N. (2016) Entropy Based Image Watermarking Using Discrete Wavelet Transform and Singular Value Decomposition. 2016 *3rd International Conference on Computing for Sustainable Global Development* (*INDIACom*), New Delhi, 16-18 March 2016, 2700-2704.

[65] Moghaddasi, Z., Jalab, H.A. and Noor, R.M. (2019) Image Splicing Forgery Detection Based on Low-Dimensional Singular Value Decomposition of Discrete Cosine Transform Coefficients. *Neural Computing and Applications*, **31**, 7867-7877. https://doi.org/10.1007/s00521-018-3586-y

[66] Bhat, K.V., Das, A.K. and Lee, J.H. (2019) A Mean Quantization Watermarking Scheme for Audio Signals Using Singular-Value Decomposition. *IEEE Access*, **7**, 157480-157488. https://doi.org/10.1109/ACCESS.2019.2949691

[67] Allwinnaldo, Budiman, G., Novamizanti, L., Alief, R.N. and Ansori, M.R.R. (2019) QIM-Based Audio Watermarking Using Polar-Based Singular Value in DCT Domain. 2019 *4th International Conference on Information Technology, Information Systems and Electrical Engineering* (*ICITISEE*), Yogyakarta, 20-21 November 2019, 216-221. https://doi.org/10.1109/ICITISEE48480.2019.9003921

[68] Chetverikov, D. and Axt, A. (2010) Approximation-Free Running SVD and Its Application to Motion Detection. *Pattern Recognition Letters*, **31**, 891-897. https://doi.org/10.1016/j.patrec.2009.12.031

[69] Gürsun, G., Crovella, M. and Matta, I. (2011) Describing and Forecasting Video Access Patterns. 2011 *Proceedings IEEE INFOCOM*, Shanghai, 10-15 April 2011, 16-20. https://doi.org/10.1109/INFCOM.2011.5934965

[70] Xie, Y., Wulamu, A., Wang, Y. and Liu, Z. (2014) Implementation of Time Series Data Clustering Based on SVD for Stock Data Analysis on Hadoop Platform. 2014 *9th IEEE Conference on Industrial Electronics and Applications*, Hangzhou, 9-11 June 2014, 2007-2010. https://doi.org/10.1109/ICIEA.2014.6931498

[71] Potha, N. and Maragoudakis, M. (2015) Cyberbullying Detection Using Time Series Modeling. 2014 *IEEE International Conference on Data Mining Workshop*, Shenzhen, 14 December 2014, 373-382. https://doi.org/10.1109/ICDMW.2014.170

[72] Das, L., Das, J.K. and Nanda, S. (2018) Advanced Protein Coding Region Prediction Applying Robust SVD Algorithm. 2017 *2nd International Conference on Man and Machine Interfacing* (*MAMI*), Bhubaneswar, 21-23 December 2017, 1-6. https://doi.org/10.1109/MAMI.2017.8307887

[73] Zear, A., Singh, A.K. and Kumar, P. (2018) Multiple Watermarking for Healthcare Applications. *Journal of Intelligent Systems*, **27**, 5-18. https://doi.org/10.1515/jisys-2016-0036

[74] Mirzal, A. (2013) The Limitation of the SVD for Latent Semantic Indexing. 2013 *IEEE International Conference on Control System, Computing and Engineering*, Penang, 29 November-1 December 2013, 413-416. https://doi.org/10.1109/ICCSCE.2013.6720000

[75] Dumais, S.T. (2004) Latent Semantic Analysis. *Annual Review of Information Science and Technology*, **38**, 188-230. https://doi.org/10.1002/aris.1440380105

[76] Landauer, T.K., Foltz, P.W. and Laham, D. (1998) An Introduction to Latent Semantic Analysis. *Discourse Processes*, **25**, 259-284.
https://doi.org/10.1080/01638539809545028

[77] Kintsch, W. (2001) Predication. *Cognitive Science*, **25**, 173-202.
https://doi.org/10.1207/s15516709cog2502_1

[78] Howard, M.W. and Kahana, M.J. (2002) When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language*, **46**, 85-98.
https://doi.org/10.1006/jmla.2001.2798

[79] Sehra, S.S., Singh, J. and Rai, H.S. (2017) Using Latent Semantic Analysis to Identify Research Trends in OpenStreetMap. *ISPRS International Journal of Geo-Information*, **6**, 1-31. https://doi.org/10.3390/ijgi6070195

[80] Evangelopoulos, N.E. (2013) Latent Semantic Analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, **4**, 683-692. https://doi.org/10.1002/wcs.1254

[81] Abedi, V., Yeasin, M. and Zand, R. (2014) Empirical Study Using Network of Semantically Related Associations in Bridging the Knowledge Gap. *Journal of Translational Medicine*, **12**, 1-6. https://doi.org/10.1186/s12967-014-0324-9

[82] Hofmann, T. (2001) Unsupervised Learning by Probabilistic LATENT Semantic Analysis. *Machine Learning*, **42**, 177-196. https://doi.org/10.1023/A:1007617005950

[83] Si, L. and Jin, R. (2005) Adjusting Mixture Weights of Gaussian Mixture Model via Regularized Probabilistic Latent Semantic Analysis. In: Ho, T.B., Cheung, D. and Liu, H., Eds., *Lecture Notes in Computer Science*, Springer, Berlin, 622-631.
https://doi.org/10.1007/11430919_72

[84] Zhai, C. and Geigle, C. (2018) A Tutorial on Probabilistic Topic Models for Text Data Retrieval and Analysis. 41*st International ACM SIGIR Conference on Research and Development in Information Retrieval*, June 2018, 1395-1397.
https://doi.org/10.1145/3209978.3210189

[85] Saquicela, V., Baculima, F., Orellana, G., Piedra, N., Orellana, M. and Espinoza, M. (2018) Similarity Detection among Academic Contents through Semantic Technologies and Text Mining. *CEUR Workshop Proceedings*, **2096**, 1-12.

[86] Uysal, A.K. and Gunal, S. (2014) Text classification Using Genetic Algorithm Oriented Latent Semantic Features. *Expert Systems with Applications*, **41**, 5938-5947.
https://doi.org/10.1016/j.eswa.2014.03.041

[87] Zhou, D., Yang, D., Zhang, X., Huang, S. and Feng, S. (2019) Discriminative Probabilistic Latent Semantic Analysis with Application to Single Sample Face Recognition. *Neural Processing Letters*, **49**, 1273-1298.
https://doi.org/10.1007/s11063-018-9852-2

[88] Ozsoy, M.G., Cicekli, I. and Alpaslan, F.N. (2010) Text Summarization of Turkish Texts Using Latent Semantic Analysis. *Coling* 2010, 23*rd International Conference on Computational Linguistics*, *Proceedings of the Conference*, Beijing, 23-27 August 2010, 869-876.

[89] Kao, A. and Poteet, S.R. (2007) Natural Language Processing and Text Mining. Springer, Berlin. https://doi.org/10.1007/978-1-84628-754-1

[90] Zhu, L. and Zhu, S. (2007) Face Recognition Based on Orthogonal Discriminant Locality Preserving Projections. *Neurocomputing*, **70**, 1543-1546.
https://doi.org/10.1016/j.neucom.2006.12.004

[91] He, X.F. (2004) Incremental Semi-Supervised Subspace Learning for Image Retrieval. *ACM Multimedia* 2004—*Proceedings of the* 12*th ACM International Conference on Multimedia*, October 2004, 2-8. https://doi.org/10.1145/1027527.1027530

[92] Wong, W.K. and Zhao, H.T. (2012) Supervised Optimal Locality Preserving Projection. *Pattern Recognition*, **45**, 186-197. https://doi.org/10.1016/j.patcog.2011.05.014

[93] Cai, D., He, X., Han, J. and Zhang, H.J. (2006) Orthogonal Laplacianfaces for Face Recognition. *IEEE Transactions on Image Processing*, **15**, 3608-3614. https://doi.org/10.1109/TIP.2006.881945

[94] Wang, Y., Huang, S., Liu, Z., Wang, H. and Liu, D. (2016) Locality Preserving Projection Based on Endmember Extraction for Hyperspectral Image Dimensionality Reduction and Target Detection. *Applied Spectroscopy*, **70**, 1573-1581. https://doi.org/10.1177/0003702816665992

[95] Wu, Y. and Gu, R.M. (2006) A New Subspace Analysis Approach Based on Laplacianfaces. In: King, I., Wang, J., Chan, L.W. and Wang, D., Eds., *Lecture Notes in Computer Science*, Springer, Berlin, 253-259. https://doi.org/10.1007/11893257_28

[96] Huang, S., Yang, D., Yang, F., Ge, Y., Zhang, X. and Lu, J. (2013) Face Recognition via Globality-Locality Preserving Projections. http://arxiv.org/abs/1311.1279

[97] Huang, P., Li, T., Shu, Z., Gao, G., Yang, G. and Qian, C. (2018) Locality-Regularized Linear Regression Discriminant Analysis for Feature Extraction. *Information Sciences*, **429**, 164-176. https://doi.org/10.1016/j.ins.2017.11.001

[98] Zhong, F., Zhang, J. and Li, D. (2014) Discriminant Locality Preserving Projections Based on L1-Norm Maximization. *IEEE Transactions on Neural Networks and Learning Systems*, **25**, 2065-2074. https://doi.org/10.1109/TNNLS.2014.2303798

[99] Turan, C., Lam, K.-M. and He, X. (2018) Soft Locality Preserving Map (SLPM) for Facial Expression Recognition. arXiv:1801.03754.

[100] Wang, B., Hu, Y., Gao, J., Sun, Y., Chen, H., Ali, M. and Yin, B. (2017) Locality Preserving Projections for Grassmann Manifold. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main Track*, Australia, 19-25 August 2017, 2893-2900. https://doi.org/10.24963/ijcai.2017/403

[101] Chen, S., Zhao, H., Kong, M. and Luo, B. (2007) 2D-LPP : A Two-Dimensional Extension of Locality Preserving Projections. *Neurocomputing*, **70**, 912-921. https://doi.org/10.1016/j.neucom.2006.10.032

[102] Lin, C., Jiang, J., Zhao, X., Pang, M. and Ma, Y. (2015) Supervised Kernel Optimized Locality Preserving Projection with Its Application to Face Recognition and Palm Biometrics. *Mathematical Problems in Engineering*, **2015**, 13-15. https://doi.org/10.1155/2015/421671

[103] Ji, Z., Pang, Y., He, Y. and Zhang, H. (2014) Semi-Supervised LPP Algorithms for Learning-To-Rank-Based Visual Search Reranking. *Information Sciences*, **302**, 83-93. https://doi.org/10.1016/j.ins.2014.10.037

[104] Hays, P., Ptucha, R. and Melton, R. (2013) Mobile Device to Cloud Co-Processing of Asl Finger Spelling to Text Conversion. 2013 *IEEE Western New York Image Processing Workshop* (*WNYIPW*), Rochester, 22-22 November 2013, 39-43. https://doi.org/10.1109/WNYIPW.2013.6890987

[105] Wang, Z.Q., Liu, Y. and Sun, X. (2008) An Efficient Web Document Classification Algorithm Based on LPP and SVM. 2008 *Chinese Conference on Pattern Recognition*, Beijing, 22-24 October 2008, 437-440. https://doi.org/10.1109/CCPR.2008.91

[106] Fernandes, S.L. and Bala, G.J. (2013) A Comparative Study on ICA and LPP Based Face Recognition under Varying Illuminations and Facial Expressions. 2013 *International Conference on Signal Processing, Image Processing & Pattern Recognition*, Coimbatore, 7-8 February 2013, 122-126. https://doi.org/10.1109/ICSIPR.2013.6497972

[107] Lu, J., Zhao, Y., Xue, Y. and Hu, J. (2008) Palmprint Recognition via Locality Preserving Projections and Extreme Learning Machine Neural Network. *International Conference on Signal Processing Proceedings*, ICSP, **2096**, 2096-2099.

[108] Baniya, B.K. and Lee, J. (2016) Importance of Audio Feature Reduction in Automatic Music Genre Classification. *Multimedia Tools and Applications*, **75**, 3013-3026. https://doi.org/10.1007/s11042-014-2418-z

[109] Huang, T.S., Nijholt, A., Pantic, M. and Pentland, A. (2007) LNAI4451—Artifical Intelligence for Human Computing. Springer, Berlin. https://link.springer.com/content/pdf/10.1007/978-3-540-72348-6.pdf#page=86

[110] Lu, K., Ding, Z., Zhao, J. and Wu, Y. (2010) Video-Based Face Recognition. *Proceedings—2010 3rd International Congress on Image and Signal Processing*, CISP, **1**, 232-235. https://doi.org/10.1109/CISP.2010.5646345

[111] Weng, X. and Shen, J. (2008) Classification of Multivariate Time Series Using Locality Preserving Projections. *Knowledge-Based Systems*, **21**, 581-587. https://doi.org/10.1016/j.knosys.2008.03.027

[112] Guo, Z.Q., Wang, H.Q. and Liu, Q. (2013) Financial Time Series Forecasting Using LPP and SVM Optimized by PSO. *Soft Computing*, **17**, 805-818. https://doi.org/10.1007/s00500-012-0953-y

[113] Zheng, H., Wang, R., Yin, J., Li, Y., Lu, H. and Xu, M. (2020) A New Intelligent Fault Identification Method Based on Transfer Locality Preserving Projection for Actual Diagnosis Scenario of Rotating Machinery. *Mechanical Systems and Signal Processing*, **135**, Article ID: 106344. https://doi.org/10.1016/j.ymssp.2019.106344

[114] 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

[115] 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. https://doi.org/10.1093/bioinformatics/bth383

[116] 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

[117] 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

[118] Hyvärinen, A. (1999) Sparse Code Shrinkage: Denoising by Nonlinear Maximum Likelihood Estimation. *Neural Computation*, **11**, 1739-1768. https://doi.org/10.1162/089976699300016214

[119] Hyvärinen, A. (1999) Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, **10**, 626-634. https://doi.org/10.1109/72.761722

[120] He, X.S., He, F. and He, A.-L. (2018) Super-Gaussian BSS Using Fast-ICA with Chebyshev-Pade Approximant. *Circuits, Systems, and Signal Processing*, **37**, 305-341. https://doi.org/10.1007/s00034-017-0554-1

[121] Akkalkotkar, A. and Brown, K.S. (2017) An Algorithm for Separation of Mixed Sparse and Gaussian Sources. *PLoS ONE*, **12**, e0175775.

https://doi.org/10.1371/journal.pone.0175775

[122] Yang, Z., LaConte, S., Weng, X. and Hu, X. (2008) Ranking and Averaging Independent Component Analysis by Reproducibility (RAICAR) *Human Brain Mapping*, **29**, 711-725. https://doi.org/10.1002/hbm.20432

[123] Radüntz, T., Scouten, J., Hochmuth, O. and Meffert, B. (2017) Automated EEG Artifact Elimination by Applying Machine Learning Algorithms to ICA-Based Features. *Journal of Neural Engineering*, **14**, Article ID: 046004. https://doi.org/10.1088/1741-2552/aa69d1

[124] Rahmanishamsi, J., Dolati, A. and Aghabozorgi, M.R. (2018) A Copula Based ICA Algorithm and Its Application to Time Series Clustering. *Journal of Classification*, **35**, 230-249. https://doi.org/10.1007/s00357-018-9258-x

[125] Glasser, M.F., Coalson, T.S., Bijsterbosch, J.D., Harrison, S.J., Harms, M.P., Anticevic, A. and Smith, S.M. (2018) Using temporal ICA to Selectively Remove Global Noise While Preserving Global Signal in Functional MRI Data. *NeuroImage*, **181**, 692-717. https://doi.org/10.1016/j.neuroimage.2018.04.076

[126] Ince, H. and Trafalis, T.B. (2017) A Hybrid Forecasting Model for Stock Market Prediction. *Economic Computation and Economic Cybernetics Studies and Research*, **51**, 263-280.

[127] Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L. and Smith, S.M. (2014) Automatic Denoising of Functional MRI Data: Combining independent Component Analysis and Hierarchical Fusion of Classifiers. *NeuroImage*, **90**, 449-468. https://doi.org/10.1016/j.neuroimage.2013.11.046

[128] Abrahamsen, N. and Rigollet, P. (2018) Sparse Gaussian ICA. 1-27. http://arxiv.org/abs/1804.00408

[129] Ablin, P., Cardoso, J.F. and Gramfort, A. (2018) Faster ICA under Orthogonal Constraint. 2018 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Calgary, 15-20 April 2018, 4464-4468. https://doi.org/10.1109/ICASSP.2018.8461662

[130] Bingham, E., Kuusisto, J. and Lagus, K. (2002) ICA and SOM in Text Document Analysis. *SIGIR Forum* (*ACM Special Interest Group on Information Retrieval*), August 2002, 361-362. https://doi.org/10.1145/564376.564444

[131] Sevillano, X., Alías, F. and Socoró, J.C. (2004) Reliability in ICA-Based Text Classification. In: Puntonet, C.G. and Prieto, A., Eds., *Lecture Notes in Computer Science*, Springer, Berlin, 1213-1220. https://doi.org/10.1007/978-3-540-30110-3_153

[132] Kim, J., Choi, J., Yi, J. and Turk, M. (2005) Effective Representation Using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 1977-1981. https://doi.org/10.1109/TPAMI.2005.242

[133] Le, T.H. (2011) Applying Artificial Neural Networks for Face Recognition. *Advances in Artificial Neural Systems*, **2011**, 1-16. https://doi.org/10.1155/2011/673016

[134] Saimurugan, M. and Ramachandran, K.I. (2015) A Comparative Study of Sound and Vibration Signals in Detection of Rotating Machine Faults Using Support Vector Machine and Independent Component Analysis. *International Journal of Data Analysis Techniques and Strategies*, **7**, 188-204. https://doi.org/10.1504/IJDATS.2014.062458

[135] Agurto, C., Barriga, S., Burge, M. and Soliz, P. (2015) Characterization of Diabetic

Peripheral Neuropathy in Infrared Video Sequences Using Independent Component Analysis. 2015 *IEEE 25th International Workshop on Machine Learning for Signal Processing* (*MLSP*), Boston, 17-20 September 2015, 1-6. https://doi.org/10.1109/MLSP.2015.7324362

[136] Lu, C.J., Lee, T.S. and Chiu, C.C. (2009) Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression. *Decision Support Systems*, **47**, 115-125. https://doi.org/10.1016/j.dss.2009.02.001

[137] Grigoryan, H. (2016) A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). *Database Systems Journal*, **7**, 12-21.

[138] Mueller, A., Candrian, G., Kropotov, J.D., Ponomarev, V.A. and Baschera, G.M. (2010) Classification of ADHD Patients on the Basis of Independent ERP Components Using a Machine Learning System. *Nonlinear Biomedical Physics*, **4**, 1-12. https://doi.org/10.1186/1753-4631-4-S1-S1

[139] Welsh, R.C., Jelsone-Swain, L.M. and Foerster, B.R. (2013) The Utility of Independent Component Analysis and Machine Learning in the Identification of the Amyotrophic Lateral Sclerosis Diseased Brain. *Frontiers in Human Neuroscience*, **7**, 1-9. https://doi.org/10.3389/fnhum.2013.00251

[140] Fisher, E.M. (1936) Linear Discriminant Analysis. *Statistics & Discrete Methods of Data Sciences*, **392**, 1-5.

[141] Pan, F., Song, G., Gan, X. and Gu, Q. (2014) Consistent Feature Selection and Its Application to Face Recognition. *Journal of Intelligent Information Systems*, **43**, 307-321. https://doi.org/10.1007/s10844-014-0324-5

[142] Sharma, A. and Paliwal, K.K. (2012) A New Perspective to Null Linear Discriminant Analysis Method and Its Fast Implementation Using Random Matrix Multiplication with Scatter Matrices. *Pattern Recognition*, **45**, 2205-2213. https://doi.org/10.1016/j.patcog.2011.11.018

[143] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 711-720. https://doi.org/10.1109/34.598228

[144] Wang, H.X., *et al.* (1999) Fisher Discriminant Analysis with Kernels. *IEEE Transactions on Cybernetics*, **44**, 828-842. https://doi.org/10.1109/TCYB.2013.2273355

[145] Yang, W. and Wu, H. (2014) Regularized Complete Linear Discriminant Analysis. *Neurocomputing*, **137**, 185-191. https://doi.org/10.1016/j.neucom.2013.08.048

[146] 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

[147] Paliwal, K.K. and Sharma, A. (2012) Improved Pseudoinverse Linear Discriminant Analysis Method for Dimensionality Reduction. *International Journal of Pattern Recognition and Artificial Intelligence*, **26**, 1-9. https://doi.org/10.1142/S0218001412500024

[148] Ye, J. and Xiong, T. (2006) Computational and Theoretical Analysis of Null Space and Orthogonal Linear Discriminant Analysis. *Journal of Machine Learning Research*, **7**, 1183-1204. https://doi.org/10.1145/1143844.1143979

[149] Ran, R., Fang, B., Wu, X. and Zhang, S. (2018) A Simple and Effective Generalization of exponential Matrix Discriminant Analysis and Its Application to Face Rec-

ognition. *IEICE Transactions on Information and Systems*, **E101D**, 265-268. https://doi.org/10.1587/transinf.2017EDL8198

[150] Dai, G. and Qian, Y. (2004) Face Recognition Using Novel LDA-Based Algorithms. *Frontiers in Artificial Intelligence and Applications*, **110**, 455-459.

[151] Wang, S., Lu, J., Gu, X., Du, H. and Yang, J. (2016) Semi-Supervised Linear Discriminant Analysis for Dimension Reduction and Classification. *Pattern Recognition*, **57**, 179-189. https://doi.org/10.1016/j.patcog.2016.02.019

[152] 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

[153] 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

[154] 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

[155] 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

[156] 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

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

[158] 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

[159] 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

[160] Wang, Z.Q. and Qian, X. (2008) Text Categorization Based on LDA and SVM. 2008 *International Conference on Computer Science and Software Engineering*, Wuhan, 12-14 December 2008, 674-677. https://doi.org/10.1109/CSSE.2008.571

[161] Suhas, S.S. (2012) Face Recognition Using Principal Component Analysis and Linear Discriminant Analysis on Holistic Approach in Facial Images Database. *IOSR Journal of Engineering*, **2**, 15-23. https://doi.org/10.9790/3021-021241523

[162] Wang, Z., Ruan, Q. and An, G. (2012) Facial expression Recognition Based on Tensor Local Linear Discriminant Analysis. 2012 *IEEE* 11*th International Conference on Signal Processing*, Beijing, 21-25 October 2012, 1226-1229. https://doi.org/10.1109/ICoSP.2012.6491797

[163] Subasi, A. and Gursoy, M.I. (2010) EEG Signal Classification Using PCA, ICA, LDA and Support Vector Machines. *Expert Systems with Applications*, **37**, 8659-8666. https://doi.org/10.1016/j.eswa.2010.06.065

[164] Umapathy, K., Krishnan, S. and Jimaa, S. (2005) Multigroup Classification of Audio

Signals Using Time-Frequency Parameters. *IEEE Transactions on Multimedia*, **7**, 308-315. https://doi.org/10.1109/TMM.2005.843363

[165] Sharma, Y., Pl, T., Hammerla, N., Mellor, S., Mcnaney, R., Olivier, P. and Essa, I. (2014) Automated Surgical Osats Prediction from Videos. 2014 *IEEE* 11*th International Symposium on Biomedical Imaging* (*ISBI*), Beijing, 29 April-2 May 2014, 461-464. https://doi.org/10.1109/ISBI.2014.6867908

[166] Siddiqi, M.H., Ali, R., Rana, M.S., Hong, E.K., Kim, E.S. and Lee, S. (2014) Video-Based Human Activity Recognition Using Multilevel Wavelet Decomposition and Stepwise Linear Discriminant Analysis. *Sensors* (*Switzerland*), **14**, 6370-6392. https://doi.org/10.3390/s140406370

[167] Maharaj, E.A. and Alonso, A.M. (2014) Discriminant Analysis of Multivariate Time Series: Application to Diagnosis Based on ECG Signals. *Computational Statistics and Data Analysis*, **70**, 67-87. https://doi.org/10.1016/j.csda.2013.09.006

[168] Sakarya, U. and Demïrpolat, C. (2018) SAR Image Time-Series Analysis Framework Using Morphological Operators and Global and Local Information-Based Linear Discriminant Analysis. *Turkish Journal of Electrical Engineering and Computer Sciences*, **26**, 2958-2966. https://doi.org/10.3906/elk-1712-339

[169] Varatharajan, R., Manogaran, G. and Priyan, M.K. (2018) A Big Data Classification Approach Using LDA with an Enhanced SVM Method for ECG Signals in Cloud Computing. *Multimedia Tools and Applications*, **77**, 10195-10215. https://doi.org/10.1007/s11042-017-5318-1

[170] Holz, M. (2016) The Creative Suburb: A Study in Creative Reflective Practice. *International Journal of Sustainable Development and Planning*, **11**, 1-14. https://doi.org/10.2495/SDP-V11-N1-1-14

[171] Friedman, J.H. and Tukey, J.W. (1974) A Projection Pursuit Algorithm for Exploratory Data Analysis. *IEEE Transactions on Computers*, **C-23**, 881-890. https://doi.org/10.1109/T-C.1974.224051

[172] Yuan, H. (2016) A Study on Wide-Area Measurement-Based Approaches for Power System Voltage Stability. Doctoral Dissertations, University of Tennessee, Knoxville. http://trace.tennessee.edu/utk_graddiss/4119

[173] Okeke, B., Uchenna, J. and Nkiruka, E. (2015) Discriminant Analysis by Projection Pursuit Discriminant Analysis by Projection Pursuit. *Global Journal of Science Frontier Research*: *F Mathematics and Decision Sciences*, **15**, 220-226.

[174] Peña, D. and Prieto, F.J. (2001) Cluster Identification Using Projections. *Journal of the American Statistical Association*, **96**, 1433-1445. https://doi.org/10.1198/016214501753382345

[175] Aladjem, M., Israeli-ran, I. and Bortman, M. (2018) Analysis Density Estimation. *IEEE Transactions on Neural Networks and Learning Systems*, **29**, 1-14.

[176] Ren, Y., Liu, H., Yao, X. and Liu, M. (2007) Prediction of Ozone Tropospheric Degradation Rate Constants by Projection Pursuit Regression. *Analytica Chimica Acta*, **589**, 150-158. https://doi.org/10.1016/j.aca.2007.02.058

[177] Chen, S.W., Lin, S.H., Liao, L. De, Lai, H.Y., Pei, Y.C., Kuo, T.S. and Tsang, S. (2011) Quantification and Recognition of Parkinsonian Gait from Monocular Video Imaging Using Kernel-Based Principal Component Analysis. *BioMedical Engineering Online*, **10**, 1-21. https://doi.org/10.1186/1475-925X-10-99

[178] Jones, M.C. and Sibson, R. (1987) What Is Projection Pursuit? *Journal of the Royal Statistical Society*: *Series A* (*General*), **150**, 1-37. https://doi.org/10.2307/2981662

[179] Nason, G.P. (1992) Design and Choice of Projection Indices. *Journal of the Royal*

Statistical Society: *Series B*, **63**, 551-567. https://doi.org/10.1111/1467-9868.00298

[180] Guo, Q., Wu, W., Questier, F., Massart, D.L., Boucon, C. and De Jong, S. (2000) Sequential Projection Pursuit Using Genetic Algorithms for Data Mining of Analytical Data. *Analytical Chemistry*, **72**, 2846-2855. https://doi.org/10.1021/ac0000123

[181] Lee, E.K., Cook, D., Klinke, S. and Lumley, T. (2005) Projection Pursuit for Exploratory Supervised Classification. *Journal of Computational and Graphical Statistics*, **14**, 831-846. https://doi.org/10.1198/106186005X77702

[182] Berro, A., Larabi Marie-Sainte, S. and Ruiz-Gazen, A. (2010) Genetic Algorithms and Particle Swarm Optimization for Exploratory Projection Pursuit. *Annals of Mathematics and Artificial Intelligence*, **60**, 153-178. https://doi.org/10.1007/s10472-010-9211-0

[183] Intrator, N. (1993) Combining Exploratory Projection Pursuit and Projection Pursuit Regression with Application to Neural Networks. *Neural Computation*, **5**, 443-455. https://doi.org/10.1162/neco.1993.5.3.443

[184] Jimenez, L. and Landgrebe, D. (1994) High Dimensional Feature Reduction via Projection Pursuit. *International Geoscience and Remote Sensing Symposium* (*IGARSS*), **2**, 1145-1147. https://doi.org/10.1109/IGARSS.1994.399367

[185] Galeano, P., Peña, D. and Tsay, R.S. (2006) Outlier Detection in Multivariate Time Series by Projection Pursuit. *Journal of the American Statistical Association*, **101**, 654-669. https://doi.org/10.1198/016214505000001131

[186] Makhotkina, L.Y., Khristoliubova, V.I. and Khannanova Fakhrutdinova, L.R. (2016) Design of Special Purposes Products Made of Nanomodified Leather. *Mathematics Education*, **11**, 1495-1503.

[187] Lee, Y.D., Cook, D., Park, J.W. and Lee, E.K. (2013) PPtree: Projection Pursuit Classification Tree. *Electronic Journal of Statistics*, **7**, 1369-1386. https://doi.org/10.1214/13-EJS810

[188] Zhang, H., Wang, C. and Fan, W. (2015) A Projection Pursuit Dynamic Cluster Model Based on a Memetic Algorithm. *Tsinghua Science and Technology*, **20**, 661-671. https://doi.org/10.1109/TST.2015.7350018

[189] da Silva, N., Cook, D. and Lee, E.-K. (2018) A Projection Pursuit Forest Algorithm for Supervised Classification. *Journal of Computational and Graphical Statistics*, 1-13. http://arxiv.org/abs/1807.07207 https://doi.org/10.1080/10618600.2020.1870480

[190] Barcaru, A. (2019) Supervised Projection Pursuit—A Dimensionality Reduction Technique Optimized for Probabilistic Classification. *Chemometrics and Intelligent Laboratory Systems*, **194**, Article ID: 103867. https://doi.org/10.1016/j.chemolab.2019.103867

[191] Hofmeyr, D. and Pavlidis, N. (2015) Semi-Supervised Spectral Connectivity Projection Pursuit. 2015 *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference* (*PRASA-RobMech*), Port Elizabeth, 26-27 November 2015, 201-206. https://doi.org/10.1109/RoboMech.2015.7359523

[192] Geng, X.Q. and Ma, Z. (2014) PPC: A Novel Approach of Chinese Text Mining Based on Projection Pursuit Model. 2014 *IEEE Workshop on Advanced Research and Technology in Industry Applications* (*WARTIA*), Ottawa, 29-30 September 2014, 950-952. https://doi.org/10.1109/WARTIA.2014.6976431

[193] Trizna, D.B., Bachmann, C., Sletten, M., Allan, N., Toporkov, J. and Harris, R. (2001) Projection Pursuit Classification of Multiband Polarimetric SAR Land Im-

ages. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 2380-2386. https://doi.org/10.1109/36.964974

[194] Malpica, J.A., Rejas, J.G. and Alonso, M.C. (2008) A projection Pursuit Algorithm for Anomaly Detection in Hyperspectral Imagery. *Pattern Recognition*, **41**, 3313-3327. https://doi.org/10.1016/j.patcog.2008.04.014

[195] Van Der Maaten, L.J.P., Postma, E.O. and Van Den Herik, H.J. (2009) Dimensionality Reduction: A Comparative Review. *Journal of Machine Learning Research*, **10**, 1-41.

[196] Osadchy, R. (2011) Kernel PCA—Unsupervised Learning 2011. PPT Presentation, 26. http://www.cs.haifa.ac.il/~rita/uml_course/lectures/KPCA.pdf

[197] Rosipal, R. (2001) An Expectation-Maximization Approach to Nonlinear. *Neural Computation*, **510**, 505-510. https://doi.org/10.1162/089976601300014439

[198] Qiao, Z., Wang, Z., Zhang, C., Yuan, S., Zhu, Y. and Wang, J. (2012) PVAm-PIP/PS Composite Membrane with High Performance for CO_2/N_2 Separation. *AIChE Journal*, **59**, 215-228.

[199] Liu, X.F. and Yang, C. (2009) Greedy Kernel PCA for Training Data Reduction and Nonlinear Feature Extraction in Classification. *MIPPR* 2009: *Automatic Target Recognition and Image Analysis*, **7495**, Article ID: 749530. https://doi.org/10.1117/12.832806

[200] Washizawa, Y. (2009) Subset Kernel Principal Component Analysis. 2009 *IEEE International Workshop on Machine Learning for Signal Processing*, Grenoble, 1-4 September 2009, 1-6. https://doi.org/10.1109/MLSP.2009.5306221

[201] Debruyne, M. and Verdonck, T. (2010) Robust Kernel Principal Component Analysis and Classification. *Advances in Data Analysis and Classification*, **4**, 151-167. https://doi.org/10.1007/s11634-010-0068-1

[202] Chen, J., Wang, G. and Giannakis, G.B. (2019) Nonlinear Dimensionality Reduction for Discriminative Analytics of Multiple Datasets. *IEEE Transactions on Signal Processing*, **67**, 740-752. https://doi.org/10.1109/TSP.2018.2885478

[203] Kim, K.I., Franz, M.O. and Sch, B. (2014) Image Modeling Based on Kernel Principal Component Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, 1-17.

[204] Pasquariello, L. and Kantor, J. (2006) Pattern Recognition. *Artforum International*, **44**, 326-323.

[205] Leitner, C., Pernkopf, F. and Kubin, G. (2011) Kernel PCA for Speech Enhancement. *NTERSPEECH* 2011, 12*th Annual Conference of the International Speech Communication Association*, Florence, 27-31 August 2011, 1221-1224.

[206] Leitner, C. and Pernkopf, F. (2011) The Pre-Image Problem and Kernel PCA for Speech Enhancement. In: Travieso-González, C.M. and Alonso-Hernández, J.B., Eds., *Advances in Nonlinear Speech Processing*, Lecture Notes in Computer Science, Vol. 7015, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-25020-0_26

[207] Fei, C. and Chongzhao, H. (2007) Time Series Forecasting Based on Wavelet KPCA and Support Vector Machine. 2007 *IEEE International Conference on Automation and Logistics*, Jinan, 18-21 August 2007, 1487-1491.

[208] Ni, J., Ma, H. and Ren, L. (2012) A Time-Series Forecasting Approach Based on KPCA-LSSVM for Lake Water Pollution. 2012 9*th International Conference on Fuzzy Systems and Knowledge Discover*y, Chongqing, 29-31 May 2012, 1044-1048.

https://doi.org/10.1109/FSKD.2012.6234207

[209] Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M. and Kimmel, R. (2008) Analysis of Two-Dimensional Non-Rigid Shapes. *International Journal of Computer Vision*, **78**, 67-88. https://doi.org/10.1007/s11263-007-0078-4

[210] Wang, H.X. (2014) Scaling for Open-Analysis. *IEEE Intelligent Systems*, **14**, 44-52.

[211] Blouvshtein, L. and Cohen-Or, D. (2018) Outlier Detection for Robust Multi-dimensional Scaling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**, 2273-2279. https://doi.org/10.1109/TPAMI.2018.2851513

[212] Ma, Y., He, K., Hopcroft, J. and Shi, P. (2018) Neighbourhood-Preserving Dimension Reduction via Localised Multidimensional Scaling. *Theoretical Computer Science*, **734**, 58-71. https://doi.org/10.1016/j.tcs.2017.09.021

[213] Chen, L. and Buja, A. (2009) Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Drawing, and Proximity Analysis. *Journal of the American Statistical Association*, **104**, 209-219.
https://doi.org/10.1198/jasa.2009.0111

[214] Haj Mohamed, H., Belaid, S., Naanaa, W. and Ben Romdhane, L. (2018) Local Commute-Time Guided MDS for 3D Non-Rigid Object Retrieval. *Applied Intelligence*, **48**, 2873-2883. https://doi.org/10.1007/s10489-017-1114-x

[215] Peterfreund, E. and Gavish, M. (2018) Multidimensional Scaling of Noisy High Dimensional Data. http://arxiv.org/abs/1801.10229

[216] Hanley, A.W., Baker, A.K., Hanley, R.T. and Garland, E.L. (2018) The Shape of Self-Extension: Mapping the Extended Self with Multidimensional Scaling. *Personality and Individual Differences*, **126**, 25-32.
https://doi.org/10.1016/j.paid.2017.12.013

[217] Wagner, D. (2008) Lecture Notes in Computer Science: Preface. In: Berthold, M.R., Ed., *Lecture Notes in Computer Science* (*Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), Springer Berlin Heidelberg, Ljubljana, Slovenia, 304-315.

[218] Poelmans, J., Van Hulle, M.M., Viaene, S., Elzinga, P. and Dedene, G. (2011) Text Mining with Emergent Self Organizing Maps and Multi-Dimensional Scaling: A Comparative Study on Domestic Violence. *Applied Soft Computing Journal*, **11**, 3870-3876. https://doi.org/10.1016/j.asoc.2011.02.026

[219] Tang, Z., Huang, Z., Zhang, X. and Lao, H. (2017) Robust Image Hashing with Multidimensional Scaling. *Signal Processing*, **137**, 240-250.
https://doi.org/10.1016/j.sigpro.2017.02.008

[220] Yang, F., Yang, W., Gao, R. and Liao, Q. (2018) Discriminative Multidimensional Scaling for Low-Resolution Face Recognition. *IEEE Signal Processing Letters*, **25**, 388-392. https://doi.org/10.1109/LSP.2017.2746658

[221] Williams, D. and Brookes, T. (2010) Perceptually-Motivated Audio Morphing: Warmth. *Journal of the Audio Engineering Society*, **1**, 216-226.

[222] Rafailidis, D., Nanopoulos, A. and Manolopoulos, Y. (2011) Nonlinear Dimensionality Reduction for Efficient and Effective Audio Similarity Searching. *Multimedia Tools and Applications*, **51**, 881-895.
https://doi.org/10.1007/s11042-009-0420-7

[223] Ho, C.C., MacDorman, K.F. and Pramono, Z.A.D.D. (2008) Human Emotion and the Uncanny Valley: A GLM, MDS, and Isomap Analysis of Robot Video Ratings. *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction* (*HRI*), Amsterdam, 12-15 March 2008, 169-176.

[224] Tenreiro MacHado, J., Duarte, F.B. and Monteiro Duarte, G. (2011) Analysis of Financial Data Series Using Fractional Fourier Transform and Multidimensional Scaling. *Nonlinear Dynamics*, **65**, 235-245.
https://doi.org/10.1007/s11071-010-9885-1

[225] He, J., Shang, P. and Xiong, H. (2018) Multidimensional Scaling Analysis of Financial Time Series Based on Modified Cross-Sample Entropy Methods. *Physica A*: *Statistical Mechanics and Its Applications*, **500**, 210-221.
https://doi.org/10.1016/j.physa.2018.02.105

[226] Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R.R. and McNally, K.L. (2015) SNP-Seek Database of SNPs Derived from 3000 Rice Genomes. *Nucleic Acids Research*, **43**, D1023-D1027.
https://doi.org/10.1093/nar/gku1039

[227] Tenenbaum, J.B., De Silva, V. and Langford, J.C. (2000) A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, **290**, 2319-2323.
https://doi.org/10.1126/science.290.5500.2319

[228] Alizadeh Sani, Z., Shalbaf, A., Behnam, H. and Shalbaf, R. (2014) Automatic Computation of Left Ventricular Volume Changes over a Cardiac Cycle from Echocardiography Images by Nonlinear Dimensionality Reduction. *Journal of Digital Imaging*, **28**, 91-98. https://doi.org/10.1007/s10278-014-9722-z

[229] Yang, M., Rajasegarar, S., Rao, A.S., Leckie, C. and Palaniswami, M. (2016) Anomalous Behavior Detection in Crowded Scenes Using Clustering and Spatio-Temporal Features. *IFIP Advances in Information and Communication Technology*, **486**, 132-141. https://doi.org/10.1007/978-3-319-48390-0_14

[230] Najafi, A., Joudaki, A. and Fatemizadeh, E. (2016) Nonlinear Dimensionality Reduction via Path-Based Isometric Mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 1452-1464.
https://doi.org/10.1109/TPAMI.2015.2487981

[231] Shi, H., Yin, B., Bao, Y. and Lei, Y. (2016) A Novel Landmark Point Selection Method for L-ISOMAP. 2016 12*th IEEE International Conference on Control and Automation*, *ICCA*, Kathmandu, 1-3 July 2016, 621-625.
https://doi.org/10.1109/ICCA.2016.7505347

[232] Liu, Y., Xia, C., Fan, Z., Wu, R., Chen, X. and Liu, Z. (2018) Implementation of Fractal Dimension and Self-Organizing Map to Detect Toxic Effects of Toluene on Movement Tracks of Daphnia magna. *Journal of Toxicology*, **2018**, Article ID: 2637209. https://doi.org/10.1155/2018/2637209

[233] Mousavi Nezhad, M., Gironacci, E., Rezania, M. and Khalili, N. (2018) Stochastic modelling of Crack Propagation in Materials with Random Properties Using Isometric Mapping for Dimensionality Reduction of Nonlinear Data Sets. *International Journal for Numerical Methods in Engineering*, **113**, 656-680.
https://doi.org/10.1002/nme.5630

[234] Dawson, K., Rodriguez, R.L. and Malyj, W. (2005) Samle Phenotype Clusters in High-Density Oligonucleotide Microarray Data Sets Are Revealed Using Isomap, a Nonlinear Algorithm. *BMC Bioinformatics*, **6**, Article No. 195.
https://doi.org/10.1186/1471-2105-6-195

[235] Geng, X., Zhan, D. and Zhou, Z. (2005) Supervised Nonlinear Dimensionality Reduction for Visualization and Classification. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, *Part B* (*Cybernetics*), **35**, 1098-1107.
https://doi.org/10.1109/TSMCB.2005.850151

[236] Huang, D.S., Jo, K.H., Lee, H.H., Kang, H.J. and Bevilacqua, V. (2009) Emerging

Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence. In: Siekmann, J. and Wahlster, W., Eds., *5th International Conference on Intelligent Computing*, *ICIC* 2009, Ulsan, South Korea, 16-19 September 2009, Vol. 5755.

[237] Sekiya, T. (2010) Analysis of Computer Science Related Curriculum on LDA and Isomap. *Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education*, Ankara, Turkey, June 2010, 48-52. https://doi.org/10.1145/1822090.1822106

[238] Lukui, S., Jun, Z., Enhai, L. and Pilian, H. (2007) Text Classification Based on Nonlinear Dimensionality Reduction Techniques and Support Vector Machines. *Third International Conference on Natural Computation* (*ICNC* 2007), Haikou, 24-27 August 2007, 674-677.

[239] Sun, W., Halevy, A., Benedetto, J.J., Czaja, W., Liu, C., Wu, H. and Li, W. (2014) UL-Isomap Based Nonlinear Dimensionality Reduction for Hyperspectral Imagery Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **89**, 25-36. https://doi.org/10.1016/j.isprsjprs.2013.12.003

[240] Weychan, R., Marciniak, T., Stankiewicz, A. and Dabrowski, A. (2015) Real Time Recognition of Speakers from Internet Audio Stream. *Foundations of Computing and Decision Sciences*, **40**, 223-233. https://doi.org/10.1515/fcds-2015-0014

[241] Reyes, A.K., Caicedo, J.C. and Camargo, J.E. (2017) Identifying Colombian Bird Species from Audio Recordings. In: Beltrán-Castañón, C., Nyström, I. and Famili, F., Eds., *Lecture Notes in Computer Science*, Springer, Cham, 274-281. https://doi.org/10.1007/978-3-319-52277-7_34

[242] Menaria, S. (2015) Manifold Feature Extraction of Video Based on ISOMAP. *International Journal of Engineering Science and Technology*, **7**, 169-172.

[243] Nie, X., Liu, J., Sun, J. and Zhao, H. (2011) Key-Frame Based Robust Video Hashing Using Isometric Feature Mapping. *Journal of Computational Information Systems*, **7**, 2112-2119.

[244] Roweis, S.T. and Saul, L.K. (2000) Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* (*New York, N.Y.*), **290**, 2323-2326. https://doi.org/10.1126/science.290.5500.2323

[245] Liu, X., Tosun, D., Weiner, M.W. and Schuff, N. (2013) Locally Linear Embedding (LLE) for Mri Based Alzheimer's Disease Classification. *NeuroImage*, **83**, 148-157. https://doi.org/10.1016/j.neuroimage.2013.06.033

[246] Rudovic, O., Nicolaou, M.A. and Pavlovic, V. (2017) Machine Learning Methods for Social Signal Processing. Cambridge University Press, Cambridge, 234-254. https://doi.org/10.1017/9781316676202.018

[247] Pan, Y., Ge, S.S. and Mamun, A.A. (2009) Weighted Locally Linear Embedding for Dimension Reduction. *Pattern Recognition*, **42**, 798-811. https://doi.org/10.1016/j.patcog.2008.08.024

[248] Park, J.H., Zhang, Z., Zha, H. and Kasturi, R. (2004) Local Smoothing for Manifold Learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **2**, 1-8.

[249] Karbauskaite, R., Kurasova, O. and Dzemyda, G. (2007) Selection of the Number of Neighbours of Each Data Point in the Locally Linear Embedding. *Information Technology and Control*, **36**, 359-364.

[250] Kouropteva, O., Okun, O. and Pietikäinen, M. (2005) Incremental Locally Linear Embedding Algorithm. *Lecture Notes in Computer Science*, **3540**, 521-530.

https://doi.org/10.1007/11499145_53

[251] Zhang, Z.Y. and Wang, J. (2007) MLLE: Modified Locally Linear Embedding Using Multiple Weights. *Advances in Neural Information Processing Systems*, **19**, 1593-1600.

[252] Yuan, Q., Dai, G. and Zhang, Y. (2017) A Novel Multi-Objective Evolutionary Algorithm Based on LLE Manifold Learning. *Engineering with Computers*, **33**, 293-305. https://doi.org/10.1007/s00366-016-0473-y

[253] Ma, L., Crawford, M.M. and Tian, J. (2010) ANOMALY detection for Hyperspectral Images Based on Robust Locally Linear Embedding. *Journal of Infrared, Millimeter, and Terahertz Waves*, **31**, 753-762. https://doi.org/10.1007/s10762-010-9630-3

[254] Wang, X., Zheng, Y., Zhao, Z. and Wang, J. (2015) Bearing Fault Diagnosis Based on Statistical Locally Linear Embedding. *Sensors* (*Switzerland*), **15**, 16225-16247. https://doi.org/10.3390/s150716225

[255] Nichols, J.M., Bucholtz, F. and Nousain, B. (2011) Automated, Rapid Classification of Signals Using Locally Linear Embedding. *Expert Systems with Applications*, **38**, 13472-13474. https://doi.org/10.1016/j.eswa.2011.04.146

[256] Embedding, D., Nie, X., Liu, J., Member, S., Sun, J. and Liu, W. (2011) Robust Video Hashing Based on. *Signal Processing*, **18**, 307-310. https://doi.org/10.1109/LSP.2011.2126020

[257] Alain, M., Guillemot, C., Thoreau, D. and Guillotel, P. (2015) Inter-Prediction Methods Based on Linear Embedding for Video Compression. *Signal Processing: Image Communication*, **37**, 47-57. https://doi.org/10.1016/j.image.2015.07.011

[258] Kohonen, T. (1990) The Self-Organizing Map. *Proceedings of the IEEE*, **78**, 1464-1480. https://doi.org/10.1109/5.58325

[259] De la Hoz, E., De La Hoz, E., Ortiz, A., Ortega, J. and Prieto, B. (2015) PCA Filtering and Probabilistic SOM for Network Intrusion Detection. *Neurocomputing*, **164**, 71-81. https://doi.org/10.1016/j.neucom.2014.09.083

[260] Merényi, E. and Taylor, J. (2017) SOM-Empowered Graph Segmentation for Fast Automatic Clustering of Large and Complex Data. 2017 12*th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization* (*WSOM*), Nancy, 28-30 June 2017, 1-9. https://doi.org/10.1109/WSOM.2017.8020004

[261] Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V. and Saarela, A. (2000) Self Organization of a Massive Document Collection. *IEEE Transactions on Neural Networks*, **11**, 574-585. https://doi.org/10.1109/72.846729

[262] Mohan, P. and Patil, K.K. (2018) Weather and Crop Prediction Using Modified Self Organizing Map for Mysore Region. *International Journal of Intelligent Engineering and Systems*, **11**, 192-199. https://doi.org/10.22266/ijies2018.0430.21

[263] Isa, D., Kallimani, V.P. and Lee, L.H. (2009) Using the Self Organizing Map for Clustering of Text Documents. *Expert Systems with Applications*, **36**, 9584-9591. https://doi.org/10.1016/j.eswa.2008.07.082

[264] Yang, H.C., Lee, C.H. and Hsiao, H.W. (2015) Incorporating Self-Organizing Map with Text Mining Techniques for Text Hierarchy Generation. *Applied Soft Computing Journal*, **34**, 251-259. https://doi.org/10.1016/j.asoc.2015.05.005

[265] Sacha, D., Kraus, M., Bernard, J., Behrisch, M., Schreck, T., Asano, Y. and Keim, D.A. (2018) SOMFlow: Guided Exploratory Cluster Analysis with Self-Organizing Maps and Analytic Provenance. *IEEE Transactions on Visualization and Computer Graphics*, **24**, 120-130. https://doi.org/10.1109/TVCG.2017.2744805

[266] Alruily, M., Ayesh, A. and Al-Marghilani, A. (2010) Using Self Organizing Map to Cluster Arabic Crime Documents. *Proceedings of the International Multiconference on Computer Science and Information Technology*, **5**, 357-363. https://doi.org/10.1109/IMCSIT.2010.5679616

[267] Yang, H.C. and Lee, C.H. (2000) Automatic Category Generation for Text Documents by Self-Organizing Maps. *Proceedings of the International Joint Conference on Neural Networks*, **3**, 581-586.

[268] Sampath, R. and Saradha, A. (2015) Alzheimer's Disease Image Segmentation with Self-Organizing Map Network. *Journal of Software*, **10**, 670-680. https://doi.org/10.17706/jsw.10.6.670-680

[269] Majumder, A., Behera, L. and Subramanian, V.K. (2014) Emotion Recognition from Geometric Facial Features Using Self-Organizing Map. *Pattern Recognition*, **47**, 1282-1293. https://doi.org/10.1016/j.patcog.2013.10.010

[270] Germen, E., Başaran, M. and Fidan, M. (2014) Sound Based Induction Motor Fault Diagnosis Using Kohonen Self-Organizing Map. *Mechanical Systems and Signal Processing*, **46**, 45-58. https://doi.org/10.1016/j.ymssp.2013.12.002

[271] Toiviainen, P. (2007) Visualization of Tonal Content in the Symbolic and Audio Domains. *Computing in Musicology*, **35**, 187-199.

[272] Du, Y., Yuan, C., Li, B., Hu, W. and Maybank, S. (2017) Spatio-Temporal Self-Organizing Map Deep Network for Dynamic Object Detection from Videos. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 4245-4254. https://doi.org/10.1109/CVPR.2017.452

[273] Maddalena, L. and Petrosino, A. (2008) A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. *IEEE Transactions on Image Processing*, **17**, 1168-1177. https://doi.org/10.1109/TIP.2008.924285

[274] Hammer, B., Micheli, A., Neubauer, N., Sperduti, A. and Strickert, M. (2005) Self-Organizing Maps for Time Series. *Proceedings of WSOM*, 115-122.

[275] Hagenbuchner, M., Sperduti, A. and Tsoi, A.C. (2003) A Self-Organizing Map for Adaptive Processing of Structured Data. *IEEE Transactions on Neural Networks*, **14**, 491-505. https://doi.org/10.1109/TNN.2003.810735

[276] Aiolli, F., Da San Martino, G., Sperduti, A. and Hagenbuchner, M. (2007) "Kernelized" Self-Organizing Maps for Structured Data. University of Wollongong, Australia, 19-24.

[277] Kohonen, T. (1995) Learning Vector Quantization. In: *Self-Organizing Maps*, Springer, Berlin, 175-189. https://doi.org/10.1007/978-3-642-97610-0_6

[278] Nova, D. and Estévez, P.A. (2014) A Review of Learning Vector Quantization Classifiers. *Neural Computing and Applications*, **25**, 511-524. https://doi.org/10.1007/s00521-013-1535-3

[279] Yousefi, J. and Hamilton-Wright, A. (2014) Characterizing EMG Data Using Machine-Learning Tools. *Computers in Biology and Medicine*, **51**, 1-13. https://doi.org/10.1016/j.compbiomed.2014.04.018

[280] Sen, O. (2002) Application of LVQ Neural Networks Combined with Genetic Algorithm in Power Quality Signals Classification. *International Conference on Power System Technology*, , Kunming, China, 13-17 October 2002, 491-495.

[281] Fitria, D., Ma'sum, M.A., Imah, E.M. and Gunawan, A.A. (2014) Automatic Arrhythmias Detection Using Various Types of Artificial. *Journal of Computer Science and Information*, **2**, 90-100. https://doi.org/10.21609/jiki.v7i2.262

[282] Hammer, B., Hofmann, D., Schleif, F.M. and Zhu, X. (2014) Learning Vector Quantization for (Dis-)Similarities. *Neurocomputing*, **131**, 43-51. https://doi.org/10.1016/j.neucom.2013.05.054

[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 8*th International Conference on Biomedical Engineering and Informatics* (*BMEI*), Shenyang, 14-16 October 2015, 742-746. https://doi.org/10.1109/BMEI.2015.7401601

[284] Kaden, M., Nebel, D. and Villmann, T. (2016) Adaptive Dissimilarity Weighting for Prototype-Based Classification Optimizing Mixtures of Dissimilarities. *ESANN* 2016 *Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, 27-29 April 2016, 135-140.

[285] Hofmann, D., *et al.* (2013) Efficient Approximations of Robust Soft Learning Vector Quantization for Non-Vectorial Data. *Neurocomputing*, **147**, 96-106.

[286] Schleif, F. (2017) Small Sets of Random Fourier Features by Kernelized Matrix LVQ. 2017 12*th International Workshop on Self-Organizing Maps and Learning Vector Quantization*, Clustering and Data Visualization (WSOM), Nancy, 28-30 June 2017. https://doi.org/10.1109/WSOM.2017.8020026

[287] Martín-Valdivia, M.T., Ureña-López, L.A. and García-Vega, M. (2007) The Learning Vector Quantization Algorithm Applied to Automatic Text Classification Tasks. *Neural Networks*, **20**, 748-756. https://doi.org/10.1016/j.neunet.2006.12.005

[288] Kumar, P.V.N. and Ganesamoorthy, R. (2016) Recognition of Sign and Text Using LVQ and SVM. *International Research Journal of Advanced Engineering and Science*, **1**, 118-125.

[289] Orhan, E., Onursal, C., Serdar, B.M. and Feyzullah, T. (2016) A Comparative Study on Parkinson's Disease Diagnosis Using Neural Networks and Artificial Immune System. *Journal of Medical Imaging and Health Informatics*, **6**, 264-268. https://doi.org/10.1166/jmihi.2016.1606

[290] González, R., Barrientos, A., Toapanta, M. and Del Cerro, J. (2017) Aplicación de las Máquinas de Soporte Vectorial (SVM) al diagnóstico clínico de la Enfermedad de Párkinson y el Temblor Esencial. *RIAI—Revista Iberoamericana de Automatica e Informatica Industrial*, **14**, 394-405. https://doi.org/10.1016/j.riai.2017.07.005

[291] Manurung, D.B., Dirgantoro, B. and Setianingsih, C. (2019) Speaker Recognition for Digital Forensic Audio Analysis Using Learning Vector Quantization Method. 2018 *IEEE International Conference on Internet of Things and Intelligence System* (*IOTAIS*), Bali, 1-3 November 2018, 222-226. https://doi.org/10.1109/IOTAIS.2018.8600852

[292] Carletti, V., Foggia, P., Percannella, G., Saggese, A., Strisciuglio, N. and Vento, M. (2013) Audio Surveillance Using a Bag of Aural Words Classifier. 2013 10*th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Krakow, 27-30 August 2013, 81-86. https://doi.org/10.1109/AVSS.2013.6636620

[293] Mochamad, H., Loy, H.C. and Aoki, T. (2004) LVQ-Based Video Object Segmentation through Combination of Spatial and Color Features. *IEEE Region 10 Annual International Conference*, Chiang Mai, 21-24 November 2004, 473-281.

[294] Rahman, A.Y., Sumpeno, S. and Purnomo, M.H. (2017) Video Minor Stroke Extraction Using Learning Vector Quantization. 2017 5*th International Conference on Information and Communication Technology* (*ICoIC7*), Melaka, 17-19 May 2017, 1-7. https://doi.org/10.1109/ICoICT.2017.8074687

[295] Poulos, M. and Papavlasopoulos, S. (2013) Automatic Stationary Detection of Time

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

[296] Dilli, R., Argou, A., Reiser, R. and Yamin, A. (2018) Fuzzy Information Processing (Vol. 831). Springer International Publishing, Berlin.

[297] Wang, L., Zhao, X., Pei, J. and Tang, G. (2016) Transformer Fault Diagnosis Using Continuous Sparse Autoencoder. *SpringerPlus*, **5**, Article No. 446. https://doi.org/10.1186/s40064-016-2107-7

[298] van der Maaten, L. and Hinton, G. (2008) Visualizing Data Using t-SNE. *Machine Learning*, **219**, 187-202.

[299] Xie, B., Mu, Y., Tao, D. and Huang, K. (2011) M-SNE: Multiview Stochastic Neighbor Embedding. *IEEE Transactions on Systems*, *Man, and Cybernetics*, *Part B*: *Cybernetics*, **41**, 1088-1096. https://doi.org/10.1109/TSMCB.2011.2106208

[300] Schubert, E., Spitz, A., Weiler, M., Geiß, J. and Gertz, M. (2017) Semantic Word Clouds with Background Corpus Normalization and T-Distributed Stochastic Neighbor Embedding. http://arxiv.org/abs/1708.03569

[301] Ghassemi, M.M., Mark, R.G. and Nemati, S. (2015) A Visualization of Evolving Clinical Sentiment Using Vector representations of Clinical Notes. *Computing in Cardiology*, **42**, 629-632. https://doi.org/10.1109/CIC.2015.7410989

[302] Zhang, J., Chen, L., Zhuo, L., Liang, X. and Li, J. (2018) An Efficient Hyperspectral Image Retrieval Method: Deep Spectral-Spatial Feature Extraction with DCGAN and Dimensionality Reduction Using t-SNE-Based NM Hashing. *Remote Sensing*, **10**, 271. https://doi.org/10.3390/rs10020271

[303] Zhong, Z., Li, J., Ma, L., Jiang, H. and Zhao, H. (2017) Deep Residual Networks for Hyperspectral Image Classification. 2017 *IEEE International Geoscience and Remote Sensing Symposium* (*IGARSS*), Fort Worth, 23-28 July 2017, 1824-1827. https://doi.org/10.1109/IGARSS.2017.8127330

[304] Carr, C.J. and Zukowski, Z. (2019) Curating Generative Raw Audio Music with D.O.M.E. *CEUR Workshop Proceedings*, **2327**, 3-6.

[305] Wong, K.Y. and Chung, F.L. (2019) Visualizing Time Series Data with Temporal Matching Based t-SNE. 2019 *International Joint Conference on Neural Networks* (*IJCNN*), Budapest, 14-19 July 2019, 1-8. https://doi.org/10.1109/IJCNN.2019.8851847

[306] Khan, S., Taverna, F., Rohlenova, K., Treps, L., Geldhof, V., De Rooij, L. and Carmeliet, P. (2019) EndoDB: A Database of Endothelial Cell Transcriptomics Data. *Nucleic Acids Research*, **47**, D736-D744. https://doi.org/10.1093/nar/gky997

[307] Beaulieu-Jones, B.K., Orzechowski, P. and Moore, J.H. (2018) Mapping Patient Trajectories Using Longitudinal Extraction and Deep Learning in the MIMIC-III Critical Care Database. *Biocomputing*, **2018**, 123-132. https://doi.org/10.1142/9789813235533_0012

[308] Spivak, D.I. (2009) Metric Realization of Fuzzy Simplicial Sets. *Self Published Notes*, 1-4. http://math.mit.edu/~dspivak/files/metric_realization.pdf

[309] Smets, T., Verbeeck, N., Claesen, M., Asperger, A., Griffioen, G., Tousseyn, T. and De Moor, B. (2019) Evaluation of Distance Metrics and Spatial Autocorrelation in Uniform Manifold Approximation and Projection Applied to Mass Spectrometry Imaging Data. *Analytical Chemistry*, **91**, 5706-5714. https://doi.org/10.1021/acs.analchem.8b05827

[310] Becht, E., *et al.* (2018) Evaluation of UMAP as an Alternative to t-SNE for Sin-

gle-Cell Data. *Nature Biotechnology*, **53**, 1689-1699. https://doi.org/10.1101/298430

[311] Smets, T., Waelkens, E. and De Moor, B. (2020) Prioritization of M/Z-Values in Mass Spectrometry Imaging Profiles Obtained Using Uniform Manifold Approximation and Projection for Dimensionality Reduction. *Analytical Chemistry*, **92**, 5240-5248. https://doi.org/10.1021/acs.analchem.9b05764

[312] Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R. and Rosen, M.S. (2018) Image Reconstruction by Domain-Transform Manifold Learning. *Nature*, **555**, 487-492. https://doi.org/10.1038/nature25988

[313] Verma, P. and Salisbury, K. (2020) Unsupervised Learning of Audio Perception for Robotics Applications: Learning to Project Data to T-SNE/UMAP Space. 1-5. http://arxiv.org/abs/2002.04076

[314] Ninevski, D. and Leary, P.O. (2020) Detection of Derivative Discontinuities. In: Berthold, M., Feelders, A. and Krempl, G., Eds., *Advances in Intelligent Data Analysis XVIII. IDA* 2020, Springer, Cham, 366-378. https://doi.org/10.1007/978-3-030-44584-3_29

[315] Maan, H., Mbareche, H., Raphenya, A.R., Banerjee, A., Nasir, J.A., Kozak, R.A. and Wang, B. (2020) Genotyping SARS-CoV-2 through an Interactive Web Application. *The Lancet Digital Health*, **2**, e340-e341. https://doi.org/10.1016/S2589-7500(20)30140-0

[316] Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G. and Newell, E.W. (2019) Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nature Biotechnology*, **37**, 38-47. https://doi.org/10.1038/nbt.4314

[317] Diaz-Papkovich, A., Anderson-Trocmé, L., Ben-Eghan, C. and Gravel, S. (2019) UMAP Reveals Cryptic Population Structure and Phenotype Heterogeneity in Large Genomic Cohorts. *PLoS Genetics*, **15**, e1008432. https://doi.org/10.1371/journal.pgen.1008432

[318] Li, K.C. (1991) Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association*, **86**, 316-327. https://doi.org/10.1080/01621459.1991.10475035

[319] Dennis Cook, R. (2000) Save: A Method for Dimension Reduction and Graphics in Regression. *Communications in Statistics—Theory and Methods*, **29**, 2109-2121. https://doi.org/10.1080/03610920008832598

[320] Li, K.C. (1992) On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma. *Journal of the American Statistical Association*, **87**, 1025-1039. https://doi.org/10.1080/01621459.1992.10476258

[321] Cook, R.D. (1998) Principal Hessian Directions Revisited. *Journal of the American Statistical Association*, **93**, 84-94. https://doi.org/10.1080/01621459.1998.10474090

[322] Xia, Y., Tong, H., Li, W.K. and Zhu, L.-X. (2002) An Adaptive Estimation of Dimension Reduction Space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 363-410. https://doi.org/10.1111/1467-9868.03411

[323] Li, B., Zha, H. and Chiaromonte, F. (2005) Contour Regression: A General Approach to Dimension Reduction. *Annals of Statistics*, **33**, 1580-1616. https://doi.org/10.1214/009053605000000192

[324] Cook, R.D. and Ni, L. (2005) Sufficient Dimension Reduction via Inverse Regression: A Minimum Discrepancy Approach. *Journal of the American Statistical Association*, **100**, 410-428. https://doi.org/10.1198/016214504000001501

[325] Li, B. and Wang, S. (2007) On Directional Regression for Dimension Reduction.

Journal of the American Statistical Association, **102**, 997-1008.
https://doi.org/10.1198/016214507000000536

[326] Zhu, Y. and Zeng, P. (2006) Fourier Methods for Estimating the Central Subspace and the Central Mean Subspace in Regression. *Journal of the American Statistical Association*, **101**, 1638-1651. https://doi.org/10.1198/016214506000000140

[327] Zeng, P. and Zhu, Y. (2010) An Integral Transform Method for Estimating the Central Mean and Central Subspaces. *Journal of Multivariate Analysis*, **101**, 271-290. https://doi.org/10.1016/j.jmva.2009.08.004

[328] Wang, H.Sh. and Xia, Y. (2008) Sliced Regression for Dimension Reduction. *Journal of the American Statistical Association*, **103**, 811-821.
https://doi.org/10.1198/016214508000000418

[329] Yin, J., Geng, Z., Li, R. and Wang, H. (2010) Nonparametric Covariance Model. *Statistica Sinica*, **20**, 469-479.

[330] Yin, X. and Li, B. (2011) Sufficient Dimension Reduction Based on an Ensemble of Minimum Average Variance Estimators. *Annals of Statistics*, **39**, 3392-3416.
https://doi.org/10.1214/11-AOS950

[331] Boulesteix, A.L. and Strimmer, K. (2004) Partial Least Squares: A Versatile Tool for the Analysis of High-Dimensional Genomic Data. *Briefings in Bioinformatics*, **8**, 32-44. https://doi.org/10.1093/bib/bbl016

[332] Li, L., Cook, R.D. and Tsai, C.L. (2007) Partial Inverse Regression. *Biometrika*, **94**, 615-625. https://doi.org/10.1093/biomet/asm043

[333] Yamada, M., Niu, G., Takagi, J. and Sugiyama, M. (2011) Computationally Efficient Sufficient Dimension Reduction via Squared-Loss Mutual Information. *Journal of Machine Learning Research*, **20**, 247-262.

[334] Fukumizu, K., Bach, F.R. and Jordan, M.I. (2009) Kernel Dimension Reduction in Regression. *Annals of Statistics*, **37**, 1871-1905. https://doi.org/10.1214/08-AOS637

[335] Cook, R.D. and Li, L. (2009) Dimension Reduction in Regressions with Exponential Family predictors. *Journal of Computational and Graphical Statistics*, **49**, 774-791.
https://doi.org/10.1198/jcgs.2009.08005

[336] Forzani, L. and Su, Z. (2017) Envelopes for Elliptical Multivariate Linear Regression. *Statistica Sinica*, **15**, 20-25.

[337] Cook, R.D. and Forzani, L. (2009) Likelihood-Based Sufficient Dimension Reduction. *Journal of the American Statistical Association*, **104**, 197-208.
https://doi.org/10.1198/jasa.2009.0106

[338] Zhang, J. and Chen, X. (2019) Robust Sufficient Dimension Reduction via Ball Covariance. *Computational Statistics and Data Analysis*, **140**, 144-154.
https://doi.org/10.1016/j.csda.2019.06.004

[339] Mai, Q. and Zou, H. (2015) Nonparametric Variable Transformation in Sufficient Dimension Reduction. *Technometrics*, **57**, 1-10.
https://doi.org/10.1080/00401706.2014.901254

[340] Gather, U., Hilker, T. and Becker, C. (2002) A Note on Outlier Sensitivity of Sliced Inverse Regression. *Statistics*, **36**, 271-281. https://doi.org/10.1080/02331880213194

[341] Li, Y. and Zhu, L.X. (2007) Asymptotics for Sliced Average Variance Estimation. *Annals of Statistics*, **35**, 41-69. https://doi.org/10.1214/009053606000001091