International Journal of Research and Advanced Development (IJRAD), ISSN: 2581-4451

# Supervised Probabilistic Dimensionality Reduction Techniques in Big Data: A Survey

**Dr. V. Chandrasekaran[1], C. Saravanan[2], A. Logeswaran[3], T. Premamala[4], V. Leela[5] ***

[1]Professor, Department of Electronics & Communication Engineering, Vellalar College of Engineering & Technology, Erode, Tamil Nadu, India. vc4sachin@gmail.com

[2,3,4,5]Assistant Professor, Department of Information Technology, Vellalar College of Engineering & Technology, Erode, Tamil Nadu, India. saravananvanjii@gmail.com, logeswaran18@gmail.com, tpremamala@gmail.com, leelabtechit@gmail.com

Abstract: In current world scenario, storing data, audio, image and video is big challenging task. By reducing the dimensionality of the data from the original dimension to lower dimension leads to good visualization, less computation time and faster execution time. Lot of dimensionality reduction techniques exists and classified into two categories' namely feature selection and feature extraction. Feature selection is removing of irrelevant and redundant data thereby reducing in computation time and increasing accuracy. Feature extraction or projection is mapping higher dimensionality data into lower dimensional data. However, there's no specific review that specialize in the supervised dimension reduction problem. Considering classification or regression as being the most goal of dimension reduction, the aim of this paper is to summarize and organize the present developments within the field into three main classes: PCA-based, Non-negative Matrix Factorization (NMF)-based, and manifold-based supervised dimensionality reduction methods. Moreover, we outline a dozen open problems which will be further explored to advance the event of this subject.

**Keywords:** supervised learning; dimension reduction; representation learning; principal component analysis; Non-negative matrix factorization; manifold learning

## I. INTRODUCTION

Large scale data with higher dimension is a big challenging problem in machine learning. High-dimensional data are quite common in this world. Popular dimension reduction methods belong to the unsupervised learning techniques because there is no label information is employed. The other two traditional machine learning categories are supervised learning and semi-supervised learning, which use all or a neighborhood of the label information. In most real applications, dimensionality reduction is simply an intermediate step towards the ultimate goals, like classification or regression [1-6]. Feature selection or feature extraction methods are first used to category the data into low-dimensional text representation, and then, a classifier helps to make a prediction [7, 8]. Lacking supervision, some important words could also be filtered before training the classifier, which affects the ultimate performance [9]. To manage the problem and to help the growing needs supervised dimensionality reduction techniques are deployed.

The Supervised dimensionality reduction methods are classified into three classes: PCA-based, NMF-based, and manifold-based dimensionality reduction methods. PCA-based and NMF-based methods are linear methods Manifold-based methods are non-linear methods. By studying the label information, we find that there are two main ways: LDA and directly integrating the loss function for classification or regression. To use the loss function directly for classification or regression, the commonly-used loss functions are mainly adopted in Support Vector Machine (SVM), logistic regression, linear regression, polynomial regression, etc. We will elaborate on them within the subsequent sections.

In the recent years many dimensionality reduction techniques where extensively explored, and several reviews [10–17] on dimension reduction where analyzed. We provide a prototype to systematically categorize the methods and helps to list the important open problems which are occurs frequently for the further development of this topic. As feature extraction is very popular when compared to the feature selection, in our paper, we mainly specialize in feature extraction for supervised learning. With reference to feature selection for supervised learning, we refer the reader to [18].

In the other part this paper, we offer a proper definition and the taxonomy of supervised dimensionality reduction in Section 2. In Section 3, we describe supervised dimensionality reduction methods and their three classes in detail. Section 4 projects the real-world applications areas where the supervised dimensionality reduction methods are implemented. In Section 5, several promising future directions are exploration. Finally, conclusion explained in the in Section 6.

## II. RELATED WORKS

### A. Definition and Taxonomy

To obtain an entire picture of the present supervised dimensionality reduction methods, we offer Figure 1 to point out the taxonomy of supervised and semi-supervised dimensionality reduction techniques.

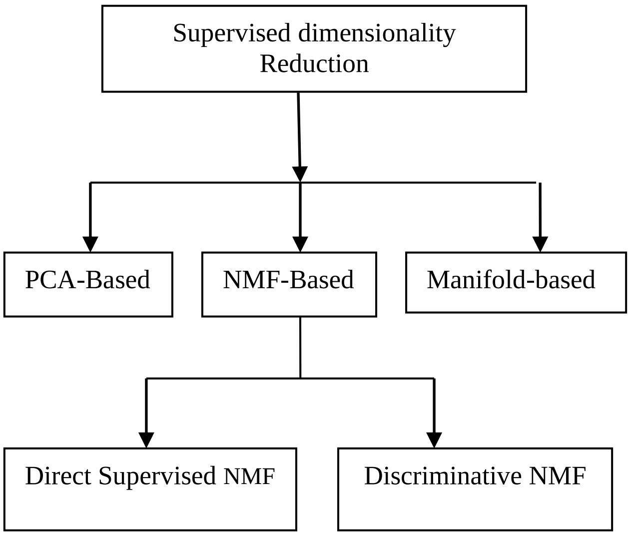


Fig 1. The taxonomy of supervised dimensionality reduction methods.

For simplicity, afterwards, we'll just use supervised dimensionality reduction to incorporate supervised and semi-supervised dimensionality reduction. Supervised dimensionality reduction methods can be divided into three classes: PCA-based, NMF-based, and manifold-based methods.

### B. Supervised Dimensionality Reduction Technique

#### 1) PCA-Based Supervised Dimensionality Reduction:

PCA are often considered as one of very popular dimension reduction technique which helps to learn the orthogonal projection of the original data onto a lower dimensional linear space, known as the principal subspace; such that the variance of the projected data is maximized [19]. PCA is a statistical based approach which helps to transforms a set of correlated variables into linearly uncorrelated variables. Assume that O observations in the data and variables are p, number of principal components is min (O-1, p). The steps involved in principal component analysis is 1) Formation of mean centered data 2) Normalize the data 3) Eigen vector and Eigen value calculation 4) Formation of principal components.

#### 2) NMF-Based Supervised Dimensionality Reduction:

It factors the non-negative into products of two non-negative matrices. Let the matrix be A. The value of A will be WH, where W and H are non-negative matrices. WH is lower approximation to A. Initially random values are given for W and H will be in the form of interactive method, W and H values are generated. In some cases, the algorithm converges lower than k. This is often an oblong matrix. The overall value is the product of W and H which is adequate to A. Non-negative matrix factorization is originally referred to as non-negative rank factorization or matrix factorization. NMF are often classified into two types. They are optimization-based methods and geometry-based methods. NMF may be a non-supervised technique which will be used for future learning. To quantify the approximation error, normally cost functions are used. NMF is that the one among the important methods to approximate the measured data. In document analysis, each document is stored as vector. Each element of vector indicates count of a term appearing that document. In image processing, each vector of matrix represents a picture. A matrix represents a set of images. Each element of the vector represents color of pixel. NMF extracts facial parts from face images. Some of the real time application of NMF is Face and visual perception. Direct supervised NMF and Discriminative NMF are its basic types.

#### 3) Manifold-Based Supervised Dimensionality Reduction:

High-dimensional data points in manifold-Based methods will have a low-dimensional manifold, and therefore the task of manifold learning is to uncover this low-dimensional manifold. Manifold-based dimensionality reduction methods exploit the geometric properties of the manifold on which the information points are speculated to lie. The basic types of manifold-based dimensionality reduction methods include Isomap [52], Locally Linear Embedding

(LLE) [53], and Laplacian Eigenmap (LE) [54]. LLE-Based Supervised dimensionality Reduction

### C. Discussion

Supervised NMF has been successfully applied in computer vision and speech recognition, because NMF features a excellent interpretability due to its non-negativity property. In classification and regression problems PCA-based methods are widely used, but the performance of NMF-based methods is not so competitive in the computer vision and speech recognition fields. Manifold-based methods assume that the information points are located in a low-dimensional manifold or each data that are often represented as the linear combination of its neighbors. Manifold-based methods are normally time consuming because of the inverse of the Laplacian matrix. In summary, based on the performance and usage, the three classes of supervised or semi-supervised methods are listed from top to bottom as PCA-based methods, manifold-based methods, and then, NMF-based methods.

## III. APPLICATION

Supervised dimensionality reduction has been successfully applied to a wide range of applications including computer vision, biomedical informatics, speech recognition, visualization, etc.

### A. Computer Vision

From the inception of NMF [27], it had been effectively applied to face recognition as it has the ability to supply interpretable bases. Face recognition is one of better example application for supervised NMF. Discriminative NMFs [46, 47, 69] are successors of supervised NMF methods at face recognition, and then, many direct NMF methods [35–37, 70] also demonstrated superior performance in this task.

### B. Biomedical Informatics:

In bioinformatics, especially inside the subject of genetics, due to the massive amount of gene markers, it's far tough to become aware of the genuine gene marker had results in a sure ailment directly. As excessive dimension and classification, need to be concurrently tackled, supervised dimensionality reduction becomes the proper preference. Zhang. [74] Proposed a most cancers classification semi-supervised projective NMF technique which is very useful to classes the datasets. Supervised PCA [76] and supervised specific PCA [77,78] was correctly carried out to gene set analysis and genome-wide affiliation analyses respectively. Moreover, supervised probabilistic PCA [26] done thoroughly in gene category. In clinical informatics, with the fast improvement of scientific devices, a spread of functions is gathered in actual packages. How to identify the effective features certainly diseases are difficult and supervised size reduction becomes an honest preference to solve this problem. Chao. [38] Proposed a supervised NMF with the aid of combing NMF which facilitates to enhance the ICU mortality prediction performance. Fuse. [79] Combined NMF and SVM helps to diagnose Alzheimer's disease and obtained an advanced performance. Supervised PCA [20] has been efficiently utilized in DNA micro array information analysis and cancer diagnosis. Finally, we got here to know that the method of knowledge discovery in biomedical informatics is broadly speaking done by using biomedical domain experts. This is broadly speaking due to the excessive complexity of the research domain, which requires deep domain information. At an equivalent time, these domain experts face most important boundaries in managing and analyzing their excessive-dimensional, sophisticated research records. A latest work [80] outlined that ontology-centered information infrastructure for studies project, which actively helps the clinical domain professionals in data acquisition, processing, and exploration, are frequently very beneficial here.

### C. Speech Recognition:

Speech popularity is some other successful application of NMF. Lee [51] Used discriminative NMF to categorize the emotional difference in speech. Weninger. [81] solved the audio supply separation with supervised NMF, even as Nakajima. [82] and Kitamura. [83] Followed supervised NMF for music sign separation. Although there exists a quantity of a success programs in speech recognition, extra attempts can be made inside the future. As we will see that the majority of the present supervised measurement reduction methods are NMF-primarily based, both PCA-based totally and manifold-based totally methods are often investigated and in comparison, with the triumphing methods.

### D. Visualization:

High-dimensional information is hard to elucidate. In ICU mortality prediction problem [38] there are numerous sign functions, and it's hard to interpret them individually way to the excessive dimensionality. As far as we all understand, biomedical experts are increasingly confronted with complex high-dimensional records. Because the number of dimensions is usually very large, one needs to map them to a smaller number of relevant dimensions to be extra amenable to professional analysis. This is because irrelevant, redundant, and conflicting size scan negatively affect the effectiveness and performance of the analytic procedure. This is often additionally the so-referred to as curse of dimensionality problem. To have an effect on this problem, dimensionality reduction can be a possible means, but the possible mappings from high- to low-dimensional areas are ambiguous. Subspace evaluation [84,85] are often wont to are searching for solutions. Since excessive-dimensional information is tough to interpret, a rough picture of the info is pretty helpful; thus, visualization is very critical, and it's also a critical utility of supervised measurement reduction. Barshan. [21] provided a supervised PCA to conduct visualization, even as Vlachos. [56] Gave some other supervised dimensionality reduction approach by using borrowing the LDA idea for visualization. Geng [58] proposed a supervised Isomapto visualize.

## IV. POTENTIAL FUTURE RESEARCH ISSUES

Supervised dimensionality reduction has emerged as a successful technique in much application area during the last two decades, still some challenging problems that need to be tackled in the near future? Below, we unfold some important open problems worth further exploration.

### A. Scalability

For PCA-based methods, the time complexity of covariance matrix computation is O ($D^2$ N), and that of its eigenvalue decomposition is O ($D^3$). Therefore, the complexity of PCA is O ($D^2$ N + $D^3$). In NMF-based methods, due to the additional objective function items some fast solving methods like the projected gradient descent method [31] do not work properly, then the time complexity of its most time-costly part is O (tNDd); t is the iteration numbers it needs to converge. For manifold-based methods, the time complexity of constructing the similarity matrix is O ($N^2$ D), and the frequently-used solving strategy is generalized Eigen value decomposition; the time complexity is O ($D^3$). One of the main focused objective of supervised dimensionality reduction is to solve high-dimensional problems, but when the feature dimension is high when compared to others, the time costs of the existing supervised dimensionality reduction methods are still high, because some specific application which may designed using unsupervised dimensionality reduction methods do not work due to the presence of new objective values or constraints on label information. When dataset is in large pattern like in social networks, there are millions of data points, and the time cost for supervised dimensionality reduction is still unacceptable. Therefore, some specific algorithms directed at supervised dimensionality reduction are urgently in need, especially due to the data explosion in this era.

### B. Missing Values problem

Missing values problems are a common phenomenon in many applications due to a variety of factors like mismatch of results, failure of sensors in computer vision and missing certain laboratory test results over time for some patients in the clinical setting [89]. The existing strategy is imputation with zero, the mean, or the maximum value, or multiple imputations [90]. To overcome problems by missing values, Lee. [34] Introduced an auxiliary matrix to indicate whether the values were missed or not. Obviously, no specific designs are proposed in the supervised dimensionality reduction process. Some methods to handle missing values like the E-M algorithm [91] can be considered to be incorporated into some supervised dimensionality reduction methods.

### C. Heterogeneous Data

Information or Data may contain any form heterogeneous types of features such as numerical, categorical, symbolic, ordinal features, etc. So, it is a very important challenge that how we are going to combine these different types of data together to perform supervised dimensionality reduction for better usage. A normal way to overcome this problem by converting all high dimensional dataset into categorical type. However, many information may be lost during this phase. Additionally, the difference between the continuous values can be categorized into the same category is neglected [95, 96]. Therefore, it is very important to analyze how to exploit the information within mixed data types which is worth exploring in the near future.

## V. CONCLUSION

The field of supervised dimensionality reduction has seen huge growth at an increasing rate. We have mentioned the state-of-the-art research on this review by categorizing it into three fundamental classes: PCA-based, NMF-based, and manifold-based supervised dimensionality reduction methods. To apprehend their characteristics better, we provide an analyzed review to elaborate their benefits and drawbacks. To increase the additional development of this topic, we also list some open upcoming troubles waiting for analytical study in the near future. This review could be beneficial for researchers who need to develop superior supervised dimensionality reduction methods or who are searching for techniques to study low-dimensional representation for certain supervised learning applications.

We accept a truth that supervised dimensionality reduction will continue to stay an active location of look at in the years to come, thanks to an increase in the high-dimensional data and sustained community efforts. In addition, their tighter integration into precise application systems will continuously structure the emerging landscape and provide opportunities for researcher contribution.

## REFERENCES


[1] Ai-Jun, Y.; Xin-Yuan, S. Bayesian variable selection for disease classification using gene expression data. Bioinformatics 2009, 26, 215–222.

[2] Sun, J.; Bi, J.; Kranzler, H.R. Multi-view singular value decomposition for disease subtyping and genetic associations. BMC Genet. 2014, 15, 73.

[3] Luo, Y.; Mao, C.; Yang, Y.; Wang, F.; Ahmad, F.S.; Arnett, D.; Irvin, M.R.; Shah, S.J. Integrating Hypertension Phenotype and Genotype with Hybrid Non-negative Matrix Factorization. Bioinformatics 2018.

[4] Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. Neurocomputing 2018, 300, 70–79.

[5] Sun, S.; Zhang, C. Adaptive feature extraction for EEG signal classification. Med. Biol. Eng. Comput. 2006, 44, 931–935.

[6] Guyon, I.; Elisseeff, A. An introduction to feature extraction. In Feature Extraction; Springer: Berlin/Heidelberg, German, 2006; pp. 1–25.

[7] Rogati, M.; Yang, Y. High-performing feature selection for text classification. In Proceedings of the Eleventh International Conference on Information and Knowledge Management, McLean, VA, USA, 4–9 November 2002; pp. 659–661.

[8] Kim, H.; Howland, P.; Park, H. Dimension reduction in text classification with support vector machines. J. Mach. Learn. Res. 2005, 6, 37–53.

[9] Basu, T.; Murthy, C. Effective text classification by a supervised feature selection approach. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops, Brussels, Belgium, 10 December 2012; pp. 918–925.

[10] Carreira-Perpinán, M.A. A Review of Dimension Reduction Techniques; Technical Report CS-96-09 9; University of Sheffield: Sheffield, UK, 1997, pp. 1–69.

[11] Fodor, I.K. A Survey of Dimension Reduction Techniques; Center for Applied Scientific Computing, Lawrence Livermore National Laboratory: Livermore, CA, USA, 2002, Volume 9,pp. 1–18.

[12] Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: A comparative review. J. Mach. Learn. Res. 2009, 10, 66–71.

[13] Thangavel, K.; Pethalakshmi, A. Dimensionality reduction based on rough set theory: A review. Appl. Soft Comput. 2009, 9, 1–12.

[14] Ma, Y.; Zhu, L. A review on dimension reduction. Int. Stat. Rev. 2013, 81, 134–150.

[15] Blum, M.G.; Nunes, M.A.; Prangle, D.; Sisson, S.A. A comparative review of dimension reduction methods in approximate Bayesian computation. Stat. Sci. 2013, 28, 189–208.

[16] Sorzano, C.O.S.; Vargas, J.; Montano, A.P. A survey of dimensionality reduction techniques. arXiv 2014, arXiv:1403.2877.

[17] Luo, Y.; Ahmad, F.S.; Shah, S.J. Tensor factorization for precision medicine in heart failure with preserved ejection fraction. J. Cardiovasc. Transl. Res. 2017, 10, 305–312.

[18] Tang, J.; Alelyani, S.; Liu, H. A survey of dimensionality reduction techniques. In Data Classification: Algorithms and Applications; CRC Press: Boca Raton, FL, USA, 2015.

[19] Hotelling, H. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 1933, 24, 417.

[20] Bair, E.; Hastie, T.; Paul, D.; Tibshirani, R. Prediction by supervised principal components. J. Am. Stat. Assoc. 2006, 101, 119–137.

[21] Barshan, E.; Ghodsi, A.; Azimifar, Z.; Jahromi, M.Z. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. Pattern Recognit. 2011, 44, 1357–1371.

[22] Gretton, A.; Bousquet, O.; Smola, A.; Schölkopf, B. Measuring statistical dependence with Hilbert-Schmidt norms. In International Conference on Algorithmic Learning Theory; Springer: Berlin/Heidelberg, German, 2005; pp. 63–77.

[23] Fukumizu, K.; Bach, F.R.; Jordan, M.I. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. J. Mach. Learn. Res. 2004, 5, 73–99.

[24] Bin, J.; Ai, F.F.; Liu, N.; Zhang, Z.M.; Liang, Y.Z.; Shu, R.X.; Yang, K. Supervised principal components: A new method for multivariate spectral analysis. J. Chemom. 2013, 27, 457–465.

[25] Roberts, S.; Martin, M.A. Using supervised principal components analysis to assess multiple pollutant effects. Environ. Health Perspect. 2006, 114, 1877. [CrossRef]

[26] Yu, S.; Yu, K.; Tresp, V.; Kriegel, H.P.; Wu, M. Supervised probabilistic principal component analysis. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; pp. 464–473.

[27] Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. Nature 1999, 401, 788.

[28] Dhillon, I.S.; Tropp, J.A. Matrix nearness problems with Bregman divergences. SIAM J. Matrix Anal. Appl. 2007, 29, 1120–1146.

[29] Kong, D.; Ding, C.; Huang, H. Robust nonnegative matrix factorization using l21-norm. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; pp. 673–682.

[30] Lee, D.D.; Seung, H.S. Algorithms for non-negative matrix factorization. In proceedings of the Conference on Neural Information Processing Systems, Denver, CO, USA, 27 November–2 December 2000; pp. 556–562.

[31] Lin, C.J. Projected gradient methods for nonnegative matrix factorization. Neural Comput. 2007, 19, 2756–2779.

[32] Hsieh, C.J.; Dhillon, I.S. Fast coordinate descent methods with variable selection for non-negative matrix factorization. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1064–1072.

[33] Sun, D.L.; Fevotte, C. Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 6201–6205.

[34] Lee, H.; Yoo, J.; Choi, S. Semi-supervised nonnegative matrix factorization. IEEE Signal Process. Lett. 2010, 17, 4–7.

[35] Jing, L.; Zhang, C.; Ng, M.K. SNMFCA: Supervised NMF-based image classification and annotation. IEEE Trans. Image Process. 2012, 21,

4508–4521.

[36] Gupta, M.D.; Xiao, J. Non-negative matrix factorization as a feature selection tool for maximum margin classifiers. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 2841–2848.

[37] Shu, X.; Lu, H.; Tao, L. Joint learning with nonnegative matrix factorization and multinomial logistic regression. In Proceedings of the 2013 International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 3760–3764.

[38] Chao, G.; Mao, C.; Wang, F.; Zhao, Y.; Luo, Y. Supervised Nonnegative Matrix Factorization to Predict ICU Mortality Risk. arXiv 2018, arXiv:1809.10680 .

[39] Luo, Y.; Xin, Y.; Joshi, R.; Celi, L.A.; Szolovits, P. Predicting ICU Mortality Risk by Grouping Temporal Trends from a Multivariate Panel of Physiologic Measurements. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 42–50.

[40] Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Discriminative learned dictionaries for local image analysis. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.

[41] Mairal, J.; Bach, F.; Ponce, J. Task-driven dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. 2012, 34, 791–804.

[42] Zhang, R.; Hu, Z.; Pan, G.; Wang, Y. Robust discriminative non-negative matrix factorization. Neurocomputing 2016, 173, 552–561.

[43] Bisot, V.; Serizel, R.; Essid, S.; Richard, G. Supervised nonnegative matrix factorization for acoustic scene classification. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016, Budapest, Hungary, 3 September 2016.

[44] Sprechmann, P.; Bronstein, A.M.; Sapiro, G. Supervised non-negative matrix factorization for audio source separation. In Excursions in Harmonic Analysis; Springer: Berlin/Heidelberg, Germany, 2015; Volume 4, pp. 407–420.

[45] Wang, Y.; Jia, Y.; Hu, C.; Turk, M. Fisher non-negative matrix factorization for learning local features. In Proceedings of the Sixth Asian Conference on Computer Vision, Jeju, Korea, 27–30 January 2004; pp. 27–30.

[46] Zafeiriou, S.; Tefas, A.; Buciu, I.; Pitas, I. Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification. IEEE Trans. Neural Netw. 2006, 17, 683–695.

[47] Kotsia, I.; Zafeiriou, S.; Pitas, I. A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems. IEEE Trans. Inf. Forensics Secur. 2007, 2, 588–595.

[48] Guan, N.; Tao, D.; Luo, Z.; Yuan, B. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. IEEE Trans. Image Process. 2011, 20, 2030–2048.

[49] Lu, Y.; Lai, Z.; Xu, Y.; Li, X.; Zhang, D.; Yuan, C. Nonnegative discriminant matrix factorization. IEEE Trans. Circuits Syst. Video Technol. 2017, 27, 1392–1405.

[50] Vilamala, A.; Lisboa, P.J.; Ortega-Martorell, S.; Vellido, A. Discriminant Convex Non-negative Matrix Factorization for the classification of human brain tumours. Pattern Recognit. Lett. 2013, 34, 1734–1747.

[51] Lee, S.Y.; Song, H.A.; Amari, S.I. A new discriminant NMF algorithm and its application to the extraction of subtle emotional differences in speech. Cognit. Neurodyn. 2012, 6, 525–535.

[52] Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. Science 2000, 290, 2319–2323.

[53] Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. Science 2000, 290, 2323–2326.

[54] Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 2003, 15, 1373–1396.

[55] Torgerson, W.S. Multidimensional scaling: I. Theory and method. Psychometrika 1952, 17, 401–419.

[56] Vlachos, M.; Domeniconi, C.; Gunopulos, D.; Kollios, G.; Koudas, N. Non-linear dimensionality reduction techniques for classification and visualization. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 645–651.

[57] Ribeiro, B.; Vieira, A.; das Neves, J.C. Supervised Isomap with dissimilarity measures in embedding learning. In Iberoamerican Congress on Pattern Recognition; Springer: Berlin/Heidelberg, German, 2008; pp. 389–396.

[58] Geng, X.; Zhan, D.C.; Zhou, Z.H. Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trans. Syst. Man Cybern. Part B 2005, 35, 1098–1107.

[59] Li, C.G.; Guo, J. Supervised isomap with explicit mapping. In Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06), Beijing, China, 30 August–1 September 2006; pp. 345–348.

[60] Zhang, Y.; Zhang, Z.; Qin, J.; Zhang, L.; Li, B.; Li, F. Semi-supervised local multi-manifold Isomap by linear embedding for feature extraction. Pattern Recognit. 2018, 76, 662–678.

[61] De Ridder, D.; Duin, R.P. Locally Linear Embedding for Classification; Pattern Recognition Group Technical Report PH-2002-01; Delft University of Technology: Delft, The Netherlands, 2002; pp. 1–12.

[62] De Ridder, D.; Kouropteva, O.; Okun, O.; Pietikäinen, M.; Duin, R.P. Supervised locally linear embedding. In Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003; Springer: Berlin/Heidelberg, Germany, 2003; pp. 333–341.

[63] Zhang, S.Q. Enhanced supervised locally linear embedding. Pattern Recognit. Lett. 2009, 30, 1208–1218.

[64] Liu, C.; Zhou, J.; He, K.; Zhu, Y.; Wang, D.; Xia, J. Supervised locally linear embedding in tensor space. In Proceedings of the 2009 Third International Symposium on Intelligent Information Technology Application,NanChang, China, 21–22 November 2009; pp. 31–34.

[65] Raducanu, B.; Dornaika, F. A supervised non-linear dimensionality reduction approach for manifold learning. Pattern Recognit. 2012, 45, 2432–2444.

[66] Zheng,F.; Chen, N.; Li, L. Semi-supervised Laplacian eigenmaps for dimensionality reduction. In Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, China, 30–31 August 2008; pp. 843–849.

[67] Wu, R.; Yu, Y.; Wang, W. Scale: Supervised and cascaded laplacian eigenmaps for visual object recognition based on nearest neighbors. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 25–27 June 2013; pp. 867–874.

[68] Jiang, Q.; Jia, M. Supervised laplacian eigenmaps for machinery fault classification. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, USA, 31 March–2 April 2009; pp. 116–120.

[69] Zhang, X.; Yali, P.; Liu, S.; Wu, J.; Ren, P. A supervised dimensionality reduction method-based sparse representation for face recognition. J.

Mod. Opt. 2017, 64, 799–806.

[70] Chen, W.S.; Zhao, Y.; Pan, B.; Chen, B. Supervised kernel nonnegative matrix factorization for face recognition. Neurocomputing 2016, 205, 165–181. [CrossRef]

[71] Kumar, B. Supervised Dictionary Learning for Action Recognition and Localization. Ph.D. Thesis, Queen Mary University of London, London, UK, 2012.

[72] Santiago-Mozos, R.; Leiva-Murillo, J.M.; Pérez-Cruz, F.; Artes-Rodriguez, A. Supervised-PCA and SVM classifiers for object detection in infrared images. In Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Miami, FL, USA, 22–22 July 2003; pp. 122–127.

[73] Xinfang, X.; Xin, X.; Hao, D.; Han, W.; Luoru, L. A Semi-Supervised Dimension Reduction Method for Polarimetric SAR Image Classification. Acta Opt. Sin. 2018, 4, 045.

[74] Zhang, X.; Guan, N.; Jia, Z.; Qiu, X.; Luo, Z. Semi-supervised projective non-negative matrix factorization for cancer classification. PLoS ONE 2015, 10, 1–20.

[75] Gaujoux, R.; Seoighe, C. Semi-supervised Nonnegative Matrix Factorization for gene expression deconvolution: a case study. Infect. Genet. Evol. 2012, 12, 913–921.

[76] Chen, X.; Wang, L.; Smith, J.D.; Zhang, B. Supervised principal component analysis for gene set enrichment of microarray data with continuous or survival outcomes. Bioinformatics 2008, 24, 2474–2481. [CrossRef]

[77] Lu, M.; Lee, H.S.; Hadley, D.; Huang, J.Z.; Qian, X. Supervised categorical principal component analysis for genome-wide association analyses. BMC Genom. 2014, 15, 1–10.

[78] Lu, M.; Huang, J.Z.; Qian, X. Supervised logistic principal component analysis for pathway-based genome-wide association studies. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, Orlando, FL, USA, 7–10 October 2012; pp. 52–59.

[79] Fuse, T.A.; Jayasignpure, N.D.; Pawar, P.D. NMF-SVM Based CAD Tool for the Diagnosis of Alzheimer's Disease. Int. J. Appl. Innov. Eng. Manag. 2014, 3, 268–274.

[80] Giradi, D.; Holzinger, A. Dimensionality Reduction for Exploratory Data Analysis in Daily Medical Research. In Advanced Data Analytics in Health; Springer: Cham, Switzerland, 2018; pp. 3–20.

[81] Weninger, F.; Roux, J.L.; Hershey, J.R.; Watanabe, S. Discriminative NMF and its application to single-channel source separation. In Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014.

[82] Nakajima, H.; Kitamura, D.; Takamune, N.; Koyama, S.; Saruwatari, H.; Ono, N.; Takahashi, Y.; Kondo, K. Music signal separation using supervised NMF with all-pole-model-based discriminative basis deformation. In Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; pp. 1143–1147.

[83] Kitamura, D.; Saruwatari, H.; Yagi, K.; Shikano, K.; Takahashi, Y.; Kondo, K. Robust music signal separation based on supervised nonnegative matrix factorization with prevention of basis sharing. In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology, Athens, Greece, 12–15 December 2013; pp. 392–397.

[84] Hund, M.; Böhm, D.; Sturm, W.; Sedlmair, M.; Schreck, T.; Ullrich, T.; Keim, D.A.; Majnaric, L.; Holzinger, A. Visual analytics for concept exploration in subspaces of patient groups. Brain Inform. 2016, 3, 233–247.

[85] Sun, S.; Zhang, C. The selective random subspace predictor for traffic flow forecasting. IEEE Trans. Int. Transp. Syst. 2007, 8, 367–373. [CrossRef]

[86] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Twenty-Seventh Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 Demcember 2013; pp. 3111–3119.

[87] Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Dmpirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 Octorber 2014; pp. 1532–1543.

[88] Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv 2018, arXiv:1810.04805.

[89] Luo, Y.; Szolovits, P.; Dighe, A.S.; Baron, J.M. 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. J. Am. Med. Inform. Assoc. 2017, 25, 645–653.

[90] Su, Y.S.; Gelman, A.; Hill, J.; Yajima, M. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. J. Stat. Softw. 2011, 45, 1–31.

[91] Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. 1977, 39, 1–38.

[92] Chao, G.; Sun, S. Consensus and complementarity based maximum entropy discrimination for multi-view classification. Inf. Sci. 2016, 367, 296–310.

[93] Xu, C.; Tao, D.; Xu, C. A survey on multi-view learning. arXiv 2013, arXiv:1304.5634.

[94] Chao, G.; Sun, S. Alternative multiview maximum entropy discrimination. IEEE Trans. Neural Netw. Learn. Syst. 2016, 27, 1445–1456.

[95] Chao, G.; Sun, S.; Bi, J. A survey on multi-view clustering. arXiv 2017, arXiv:1712.06246.

[96] Holzinger, A. From Machine Learning to Explainable AI. In Proceedings of the 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), Kosice, Slovakia, 23–25 August 2018; pp. 55–66.