QUALITY ASSESSMENT OF DIMENSIONALITY REDUCTION TECHNIQUES ON HYPERSPECTRAL DATA: A NEURAL NETWORK BASED APPROACH

Deepa C[1, *], Amba Shetty[2], Narasimhadhan A V[3]

[1] Department of Applied Mechanics and Hydraulics, 177am003.deepa@nitk.edu.in
[2] Department of Applied Mechanics and Hydraulics, amba_shetty@yahoo.co.in
[3] Department of Electronics and Communication Engineering, dhan257@gmail.com

National Institute of Technology Karnataka, Surathkal, Mangalore-575025, India

Commission III, WG III/4

KEYWORDS: Hyperspectral remote sensing, Dimensionality reduction, Autoencoders, Coranking matrix.

ABSTRACT

Dimensionality reduction of hyperspectral images plays a vital role in remote sensing data analysis. The rapid advances in hyperspectral remote sensing has brought in a lot of opportunities to researchers to come up with advanced algorithms to analyse such voluminous data to better explore earth surface features. Modern machine learning algorithms can be applied to explore the underlying structure of high dimensional hyperspectral data and reduce the redundant information through feature extraction techniques. Limited studies have been carried out on dimensionality reduction for mineral exploration. The current study mainly focuses on the application of autoencoders for dimensionality reduction and provides a qualitative (visual) analysis of the obtained representations. The performance of autoencoders are investigated on Cuprite scene. Coranking matrix is used as evaluation criteria. From the obtained results it is evident that, deep autoencoders provide better results compared to single layer autoencoders. An increase in the number of hidden layers provides a better embedding. The neighborhood size $K \geq 40$ of deep autoencoders provides a better transformation compared to autoencoders which shows an improved embedding only after $K \geq 80$.

1. INTRODUCTION

Recent advances in sensor technology has led to an increased availability of hyperspectral data with sufficiently high spectral and spatial resolutions. Each pixel contains a detailed representation of various materials on ground in the form of reflectance curve based on its absorption features in different portions of the electromagnetic spectrum. This detailed spectral information increases the possibility of more accurately discriminating materials of interest (Chutia, et al., 2015). The burst of informative content conveyed by hyperspectral data permits an improved understanding of different land coverage on earth surface. In spite of that, it introduces a series of challenges that needs to be addressed, such as the computational complexity and resources required to process such voluminous data. The main difficulty in analysing such high-dimensional data sets is that the number of observations required to estimate functions at a certain level of accuracy grows exponentially with the dimension. This problem, often referred to as the curse of dimensionality, has led to various techniques that attempts to reduce the dimensionality of the original hyperspectral data (Daniela, 2014).

Although, the hyperspectral data are both voluminous and multidimensional, nowadays with the availability of advanced computing systems that possess high speed processors and enormous storage power, data volume is no longer a constraint. The problem lies in the data redundancy that needs to be removed to obtain the bands that convey maximum information.

Much of the data does not add to the inherent information content for a particular application, even though it often helps in discovering that information; it contains redundancies. The data recorded by hyperspectral sensors often have substantial overlap of information content over the bands of data recorded for a given pixel. But not all of the data is required for a particular application. Therefore, dimensionality reduction plays a vital role in hyperspectral data analysis.

The high-dimensional nature of hyperspectral imagery imposes certain challenges to perform classification and conventional algorithms do not adapt well. When there is a limited number of reference samples to train a classification system, as the number of dimensions increases, the accuracy of the classification tends to drop. This is because the reliable estimation of statistical class parameters becomes more and more difficult as dimensionality increases (Hughes, 1968). In designing classifiers, the goal is to improve the accuracy of predictions. A vast number of classification techniques for hyperspectral imagery have been presented in the literature, which share the goal of attenuating the Hughes effect and accurately identifying the classes. The kernel methods such as Support Vector Machines (SVMs) have become very popular for hyperspectral image analysis, proving to be extremely well suited to classify high dimensional data when a limited number of training samples are available (Valls, 2005). To further boost up classification accuracies, ensemble classification systems have been investigated for hyperspectral image classification. These approaches combine multiple learning algorithms to improve the predictive accuracy. The use of Random Forest framework was investigated for classification of hyperspectral data (Ham et al., 2005). Aiming to classify high-dimensional data, we have to take into account that the Hughes phenomenon might have a negative impact on the classification results. To address this general issue regarding the classification of high-dimensional

---

* Corresponding author

data, it has been proposed to introduce techniques for dimensionality reduction (Sumithra et al., 2015).

Dimensionality reduction can also be used to visualize the interiors of deep neural networks where the high dimensionality comes from the large number of weights used in a neural network and convergence can be visualized by means of DR (Han et al., 2017). The difficulty in applying these techniques is that each method is designed to maintain certain aspects of the original data and therefore may be appropriate for one task and inappropriate for another. Most of the methods also have parameters to tune and follow different assumptions. The quality of the outcome may strongly depend on their tuning, which adds additional complexity to the process. Choosing an inadequate method may imply that much of the underlying structure remains undiscovered. Depending on the chosen parameters, even a single method can lead to vastly diverse results. Moreover, many nonlinear techniques do not arrive at a unique solution due to random aspects of the algorithm. Instead, they can produce different results in every run, corresponding to different local optima of the objective.

In this paper, systematic analysis of autoencoders and deep autoencoders for dimensionality reduction of hyperspectral data is carried out. Autoencoders provide an unsupervised methodology for generating meaningful latent representations of original data. Such approaches can be utilized in two ways. First, the generated latent representations of the data can be exploited for further analysis. Second, the architectures can be deployed on other heterogeneous data sets. This can be extremely helpful especially for mineral exploration, since there is no ground truth available for better analysis. Also, the obtained results can be used as valuable input for further in-depth analysis of the concerned data. Though various techniques have been proposed in the past, the real challenge is to select an appropriate algorithm for the task at hand. Limited studies have been carried out for dimensionality reduction of hyperspectral data for mineral exploration. Hence the main objective of the current study is to verify the suitability of neural network based approaches on hyperspectral data.

2. DIMENSIONALITY REDUCTION METHODS

Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data (Fukanaga, 1990). Dimensionality reduction is important in many domains, since it mitigates the curse of dimensionality and other undesired properties of high-dimensional spaces (Jimmenez, 1997). As a result, dimensionality reduction aids classification, visualization, and compression of high-dimensional data.

Dimensionality reduction techniques can be classified into two types namely feature extraction and band selection. Traditionally, dimensionality reduction was performed using popular linear techniques such as Principal Components Analysis (PCA) (Pearson, 1901), factor analysis (Spearman, 1904), and classical scaling (Turk, 1991). However, these linear techniques cannot adequately handle complex nonlinear relationships inherent in the data. In the past few years, a large number of nonlinear techniques for dimensionality reduction have been proposed to overcome the insufficiencies of traditional linear techniques. Also, many algorithms were

developed to perform embedding for manifold based datasets namely, Isomap (Tenenbaum et al., 2000), Local Linear Embedding (LLE) (Roweis and Saul, 2000), Laplacian Eigenmaps (Belkin and Niyogi, 2003), Local Tangent Space Alignment (LTSA) (Zhang and Zha, 2004), Hessian Eigenmaps (Donoho and Grimes, 2004), Diffusion Maps (Coifman and Lafon, 2006) and Semi definite Embedding (SDE) (Weinberger and Saul, 2006).This group of algorithms attempts to effectively generate a subspace, resulting in a model which maximizes the distance between different classes.

Band selection aims to select a subset of bands from the original spectral bands that can well represent the actual data. The simplest suboptimal search strategy employs sequential forward selection and sequential backward selection techniques (Webb et al., 2011) which achieves the best subset of features based on adding a set of prefixed features to the current one. Also, with rapid evolvement in the field of soft computing, genetic algorithms have also received considerable attention for feature selection (Zhang, et al., 2012).

Recently, many deep learning approaches have been proposed for analysing hyperspectral data (Wei et al., 2015). Typically, they rely on extracting prominent features using deep convolutional neural networks for analysis and classification. However, these methods refer to supervised learning, and require many labelled observations in order to perform well. In contrast, unsupervised approaches learn representations by identifying patterns in the data and extracting meaningful knowledge while overcoming data complexities. Particular variants of deep learning networks, referred to as autoencoders, have demonstrated good performance for unsupervised representation learning (Bengio et al., 2013). The advantage of unsupervised learning is that there is no need to specify classes or a target variable for the data under observation.. Instead the chosen algorithm arranges the input data. For example, arranged into clusters or into a lower dimensional representation. In contrast to a supervised problem, there is no natural way to directly measure the quality of any output or to compare two methods by an objective measure like modeling efficiency or classification error.

Autoencoders learn a compressed representation of the input data by reconstructing it on the output of the network (Suvash et al., 2015). This compressed representation captures the structure of the data and therefore allows for more accurate analysis (Belkin and Niyogi, 2003). Autoencoders have been deployed on a variety of tasks across different data types such as dimensionality reduction, data denoising, compression, and data generation. Hence in this context, autoencoders can be utilised effectively for dimensionality reduction of hyperspectral data.

3. NEURAL NETWORK BASED DIMENSIONALITY REDUCTION

In the past few years, neural network based approaches for classifying hyperspectral data received a lot of attention (Merenyi, 2005). Neural network models have an advantage over statistical methods in that they are distribution free and thus no prior knowledge about the statistical distribution of classes is needed. In a neural network, a set of weighted sums and nonlinearities describe the function that classifies the input features. The training procedure involves finding the appropriate weights, which is done iteratively.

3.1 Autoencoders: Generally, an autoencoder consists of two networks, an encoder and a decoder, which broadly perform the following tasks:
Encoder: Maps the high dimensional input data into a latent variable embedding which has lower dimensions than the input.
Decoder: Attempts to reconstruct the input data from the embedding.
An autoencoder neural network is an unsupervised machine learning algorithm that applies back propagation, setting the target values to be equal to the inputs. A simple autoencoder will have single hidden layer between the input and output. Autoencoders can represent both linear and non-linear transformation in encoding. The network is trained to minimize the mean squared error between the input and the output of the network. In order to allow the autoencoder to learn a nonlinear mapping between the high dimensional and low dimensional data representation, sigmoid activation functions are used.

The main challenge when designing an autoencoder is its sensitivity to the input data. Several variants have been proposed since autoencoders were first introduced. These variants mainly aim to address shortcomings such as improved generalization and modification to sequence input models. Some significant examples include the denoising, sparse and variational autoencoders.

3.2 Deep autoencoders: A deep autoencoder will have multiple hidden layers. An increase in the number of hidden layers permits the network to be used to solve complex problems. But as the number of hidden layers are increased, the errors propagated back to the earlier layers are drastically reduced. This means that the weights in hidden layers close to the output layer are updated normally, whereas weights in hidden layers close to the input layer are updated minimally or not at all. Generally, this problem prevented the training of very deep neural networks and was referred to as the vanishing gradient problem. This problem can be reduced considerably by the process of pre-training (Diehao Kong et al., 2019). Features learned by pre-training a deep autoencoder structure produce spectral features that outperform conventional feature extraction methods.

4. EXPERIMENT AND RESULTS

In this section, a qualitative and comparative analysis of autoencoders and deep autoencoders for dimensionality reduction of hyperspectral data has been carried out.

4.1 Dataset: To demonstrate the role of autoencoders in the dimensionality reduction of hyperspectral data, a detailed analysis is carried out on the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral data. AVIRIS, flown by NASA/Jet Propulsion Laboratory (JPL) is a 224-channel imaging spectrometer with approximately 10 nm spectral resolution covering the 0.4 –2.5 μm spectral range. (Kruse et al., 2003). The 0.4 to 2.5 μm spectral range provides abundant information about many important earth surface minerals. The AVIRIS data used in the study was captured from Earth Resource-2 aircraft on August 8, 2011, at Cuprite, Nevada, USA.

4.2 Quality measures: It is often too hard for researchers to judge the quality of the resulting embedding by visual inspection. Also, it cannot be compared against ground truth due to high dimensional nature of data. Therefore, formal measures play a vital role in judging the quality of a given data

embedding. Several quality measures have been proposed in the past to serve the purpose (Lee et al., 2009). A quality measure based on coranking matrix is used to evaluate the performance of autoencoders on hyperspectral data (Leuks et al., 2011).

The coranking matrix is a way to capture the changes in ordinal distance. The column wise distances in a distance matrix are replaced by their ranks. The comparison of the ranks in the high and low dimensional spaces is carried out in a systematic way. In a perfect dimensionality reduction, the matrix will only have non zero entries in the diagonal, if most of the non-zero entries are in the lower triangle, then the process of dimensionality reduction collapsed far away points onto each other; if most of the non-zero entries are in the upper triangle, then it is understood that close points are torn apart.

Rank errors and concepts such as neighbourhood intrusions and extrusions can be associated with different blocks of the coranking matrix. The model is pre-trained using stacked denoising autoencoders. They are designed using a greedy layer wise strategy. Pre-training is based on the assumption that it is easier to train a shallow network instead of a deep network, which also reduces generalization error. Deep neural networks can easily jump out of local minima with the help of pre-training (Diehao Kong et al., 2019).

The high dimensional hyperspectral dataset is represented by, $Y = \{y_1, y_2, \ldots, y_N\} \in R^H$ and low dimensional dataset $X = \{x_1, x_2, \ldots, x_N\} \in R^L$. Let δ_{ij} be the distance from y_i to y_j in R^H and d_{ij} be the distance from x_i to x_j in R^L. The rank of y_j with respect to y_i in R^H is given by,

$$\delta_{ij} = |\{k \mid \delta_{ik} < \delta_{ij} \text{ or } (\delta_{ik} = \delta_{ij} \text{ and } 1 \le k < j \le N)\}| \quad (1)$$

Similarly, the rank of x_j with respect to x_i in low dimensional space is,

$$r_{ij} = |\{k \mid d_{ik} < d_{ij} \text{ or } (d_{ik} = d_{ij} \text{ and } 1 \le k < j \le N)\}| \quad (2)$$

The differences $R_{ij} = r_{ij} - \rho_{ij}$ are the rank errors. The coranking matrix C is the histogram of all rank errors and is given by,

$$C_{kl} = |\{(i,j) \mid \rho_{ij} = k \text{ and } r_{ij} = l\}| \quad (3)$$

Pairs of points which change their rank between the original data and its projection are considered as errors. They result in non-zero off-diagonal entries in the coranking matrix. A point x_j with $\rho_{ij} > r_{ij}$ is called intrusion and $\rho_{ij} < r_{ij}$ is called extrusion. The un-weighted sum of C is expressed as a quality,

$$Q_{NX}(K) = \frac{1}{KN} \sum_{k=1}^{K} \sum_{l=1}^{K} C_{kl} \quad (4)$$

where K defines the neighborhood points.

To display the quality of embedding, a curve of $Q_{NX}(K)$ is plotted for fixed range of K. A single parameter K is replaced by the pair (K_s, K_t), where K_s determines the region of interest and K_t is the size of tolerated rank errors which results in a new quality measure $Q_{ND}(K_s, K_t)$ (Bassam Mokbel et al., 2013). For better visualisation, the quality $Q_{ND}(K_s, K_t)$ is parameterized by two values. Hence, rather than a single curve, the results are now represented by a surface. The full quality surface can easily be displayed as a colored matrix, (K_s, K_t) is assigned a color value according to $Q_{ND}(K_s, K_t)$.

Figure 1 and 2 depicts the quality measure Q_{NX} (K) for autoencoder and deep autoencoder respectively. Figure 3 and 4 depicts the quality measure Q_{ND} (K_s, K_t). As shown in the figures 1and 2, the curve steadily rises to the maximum.For fixed neighbourhood range K, the value of Q_{NX} (K) is almost close to 1 beyond K ≥ 80 in autoencoders and K ≥ 40 in deep autoencoders. A perfect embedding provides a Q value of 1. From this it can be inferred that, the global errors are relatively minimum. Figure 3 and 4 provides a better insight through visual interpretation, wherein the white portion indicates perfect embedding with the quality index value 1. It can be observed that, the errors originating in the smaller regions (small K_s) are rather small, errors only occurs for $K_t < K_s$, which implies that the absolute size of rank errors increases.

In ideal case, off diagonal entries in a coranking matrix should be zero. But, it is not so because of intrusions and extrusions induced by rank errors. Autoencoders provides appropriate embedding for K ≥ 80 whereas deep autoencoders provides a good embedding from K ≥ 40. This is due to the fact that, increase in the number of hidden layers allows the network to learn more complex features inherent in the data which in turn results in a better embedding.
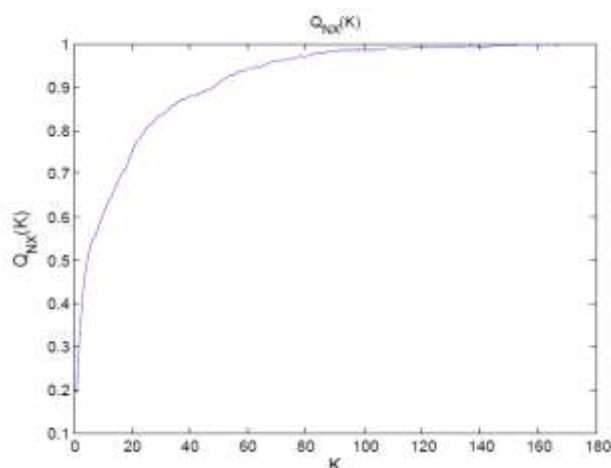

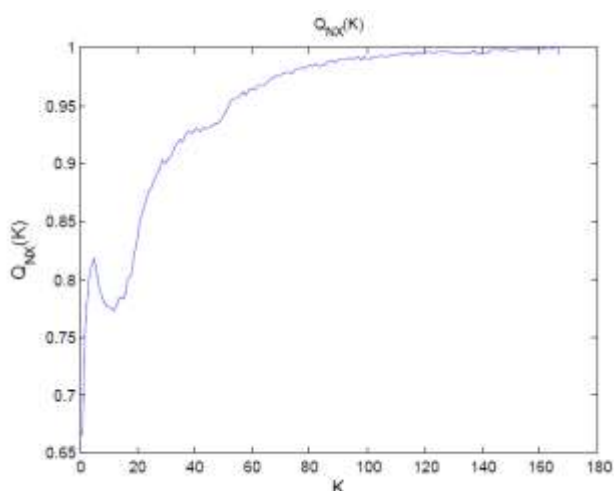Figure 1. Plot of Q_{NX} (K) versus K for autoencoder

Figure 2. Plot of Q_{NX} (K) versus K for deep autoencoder
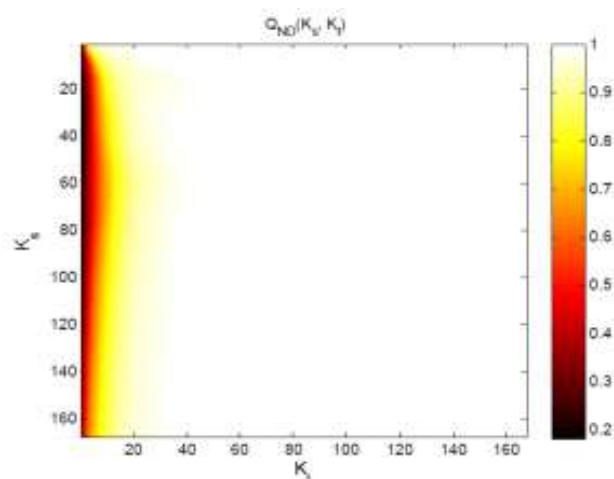

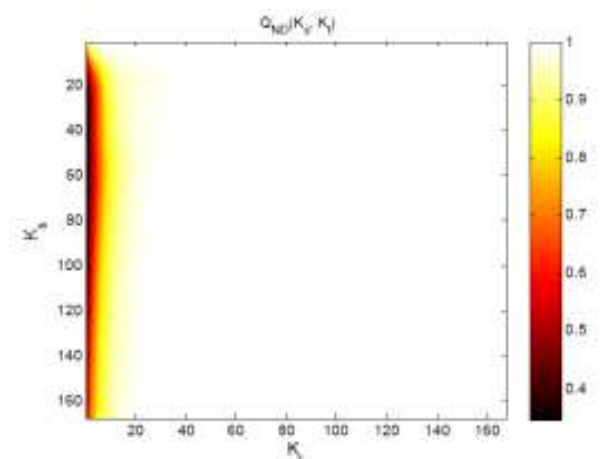Figure 3. Q_{ND} quality measure for autoencoder

Figure 4. Q_{ND} quality measure for deep autoencoder

In summary, experimental evaluation is used to verify that the neural network based approaches produce valid representations and can be applied for dimensionality reduction of hyperspectral data. Also, visual inspection of the learned representations of the whole data set are also obtained from autoencoders and deep autoencoders. The obtained results clearly reveal that autoencoders based approaches are really capable of reconstructing the original data from latent space representation.

5. CONCLUSIONS

Dimensionality reduction techniques play a significant role in hyperspectral data analysis. In particular for mineral exploration limited studies have been carried out to overcome the problem of curse of dimensionality. Although there are many new methods to reduce dimensionality, their assessment and comparison still remains open. In this study, application of autoencoders for dimensionality reduction of hyperspectral data is investigated and evaluated under coranking framework. The studied approaches have several distinguishing properties. First, they are able to produce representations that capture the intrinsic relationships between the data variables and therefore allow for more accurate analysis. Second, they are capable of

reducing the dimensionality of the input data without much loss of quality or performance. Consequently, from the results obtained it can be concluded that, nonlinear techniques outperform the traditional linear techniques. In particular, deep autoencoders with three hidden layers perform well compared to simple autoencoders. For a fixed neighbourhood range, deep autoencoders provides better transformation beyond K ≥ 40 and autoencoders as well beyond K ≥ 80. Hence autoencoders can be considered as a better choice for dimensionality reduction of hyperspectral data since it does not require labelled data for training. On the other hand, increase in the number of hidden layers adds to the computational complexity and reduces the generalisation capability of the network. Restricted Boltzman machine can be further included in training deep autoencoders to further improve the quality of embedding.

REFERENCES

Bassam Mokbel., Wouter Lueks., AndrejGisbrecht., BarbaraHammer., (2013). Visualizing the quality of dimensionality reduction. *Neurocomputing*, 112, 109–123.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computing,* 15(6), 1373-1396.

Bengio, Y., Courville, A., and Vincent, P. 2013. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 35, 1798–1828. doi: 10.1109/TPAMI.2013.50.

Chein-I Chang and Su Wang, 2006. Constrained Band Selection for Hyperspectral Imagery.*IEEE Transactions on Geoscience and Remote Sensing,* 44(6), 1575-1585.

Chutia, D., Bhattacharyya D.K., Sarma K.K., Kalita R., Sudhakar S., 2015. Hyperspectral remote sensing classifications: a perspective survey. *Trans. GIS* 20(4), 463–490.

Coifman, R.R., Lafon, S., 2006. Diffusion maps. *Applied and Computational Harmonic Analysis,* 21(1), 5-30.

Daniela Baran, Nicolae Apostolescu, 2014. A Virtual Dimensionality Method for Hyperspectral Imagery. *Procedia Engineering,* 100, 450-455.

Diehao Kong and Xuefeng Yan, 2019. Novel Model Based on Stacked Autoencoders with Sample Wise Strategy for Fault Diagnosis. *Mathematical Problems in Engineering,* Article ID 8985657, 1-10.

Donoho, D. L., Grimes, C., 2004. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of National Academy of Science, U.S.A.,* 100(10), 5591-5596.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition, *Academic Press,* San Diego, California.

Goel, S.O. Prasher , R.M. Patel , J.A. Landry , R.B. Bonnell , A. A. Viau, 2003. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computer and Electronics in Agriculture,* 39(2), 67-93.

Goetz, Vane, B. Rock, 1985. Imaging Spectrometry for Earth Remote Sensing. *Science,* 228(4704), 1147-1153.

Ham, J., Chen, Y., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*. 43(3), 492–501.

Han, Y, J. Kim, and K. Lee., (2017). Deep Convolutional Neural Networks for Predominant Instrument Recognition in Polyphonic Music. *IEEE-ACM Transactions on Audio, Speech and Language Processing,* 25(1), 208–221.

Hughes, 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information theory,* 14(1), 55-63.

Jimenez, L.O., D.A. Landgrebe, 1998. Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics,* Part C (Applications and Reviews), 28(1), 39 - 54 doi:10.1109/5326.661089.

Kruse,F.A., Boardman, J.W., and Huntington, J.F, 2003.Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), 1388–1400, doi:10.1109/TGRS.2003.812908

Lee, John A., Michel Verleysen, 2009. Quality assessment of Dimensionality Reduction: Rank based Criteria. *Neurocomputing,* 72(7), 1431-1433.

Lennon, G. Mercier, M.C. Mouchot, L. Hubert-Moy, 2001. Independent Component Analysis as a tool for the dimensionality reduction and the representation of hyperspectral images. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium,* 2893-2895.

Lueks, W., B. Mokbel, M. Biehl, and B. Hammer., 2011. How to Evaluate Dimensionality Reduction? - Improving the Co-ranking Matrix. arXiv: 1110.3917.

Liangpei Zhang, Bo Du, 2012.Recent advances in hyperspectral image processing. *Geo-spatial Information Science*, 15(3), 143–156.

Mathieu Fauvel, Jon Atli Benediktsson, Jocelyn Chanussot and Johannes R. Sveinsson, 2008. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing,* 46(11), 3804-3814.

Merenyi, E., 2005. Intelligent understanding of hyperspectral images through self-organizing neural maps. *Proceedings of the 2nd International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2005), Orlando, FL, USA,* 30-35.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philiosophical Magazine,* 2, 559–572.

Rodarmel and Jie Shan, 2002.Principal Component Analysis for Hyperspectral image Classification. *Surveying and Land Information Systems*, 62(2), 115-123.

Roweis,S.T.,L.K.Saul.,2000.Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290(5500), 2323-2326.

Spearman, C., 1904. General intelligence objectively determined and measured. *American Journal of Psychology*, 15, 206–221.

Sumithra, Subu, 2015.A Review of various linear and nonlinear dimensionality reduction techniques. *International Journal of Computer Science and Information Technologies*, 6(3), 2354-2360.

Suvash Sedhain, Aditya Krishna Menon, Sanner, and Lexing Xie., 2015. Autorec: Autoencoders meet collaborative filtering. *Proceedings of the 24th International Conference on World Wide Web*, 111–112.

Tenenbaum, J. B., V. de Silva, and J. C. Langford, 2000. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319-2323.

Turk, M.A., Pentland., 1991. Face recognition using eigenfaces. *Proceedings of the Computer Vision and Pattern Recognition,* 586–591.

Vaddi, Prabukumar, 2017. Comparative study of Feature extraction techniques for Hyperspectral remote sensing Classification: A survey. *Proceedings of International Conference on Intelligent Computing and Control systems*, 543-548.

Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing.* 43(6), 1351–1362.

Vane, Goetz, 1988. Terrestrial Imaging Spectroscopy. *Remote Sensing of Environment*, 24(1), 1-29.

Webb, A.R., Ccpsey, K.D., Cawley, G., 2011. Statistical Pattern Recognition, Wiley: Chichester.

Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li., 2015. Deep convolutional neural networks for hyperspectral image classification.*Sensors,* Article ID 258619, 12:10–1155.

Weinberger,K.Q.,L.K. Saul.,2006. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision,* 70(1), 77-90.

Zebin Wu, Yonglong Li, Antonio Plaza, Jun Li, 2016. Parallel and Distributed Dimensionality reduction of hyperspectral data on cloud computing architectures. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 9(6), 2270-2278.

Zena, Duncan, 2015. A review of feature selection and feature extraction methods applied on Microarray data. *Review Article in Advances in Bioinformatics*, 1-13.

Zhang, Z.Y., H.Y. Zha., 2004. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal of Scientific Computing*, 26(1), 313-338.