

NON-LINEAR AUTOENCODER BASED ALGORITHM FOR DIMENSIONALITY REDUCTION OF AIRBORNE HYPERSPECTRAL DATA

Swati Priya^{1*}, Ranendu Ghosh¹, B.K Bhattacharya²

¹Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar - swtpr.04@gmail.com

²Space Applications Centre (SAC), Ahmedabad

Commission III, WG III/10

KEY WORDS: Hyperspectral data, AVIRIS-NG, Dimensionality reduction, Non-linear, Autoencoder

ABSTRACT:

Hyperspectral remote sensing is an advanced remote sensing technology that enhances the ability of accurate classification due to presence of narrow contiguous bands. The large number of continuous bands present in hyperspectral data introduces the problem of computational complexity due to presence of redundant information. There is a need for dimensionality reduction to enhance the ability of users for better characterization of features. Due to presence of high spectral correlation in the hyperspectral datasets, optimum de-correlation technique is required which transforms the hyperspectral data to lower dimensions without compromising with the desirable information present in the data. In this paper, focus has been to reduce the spectral dimensionality problem. So, this research aimed to develop computationally efficient non-linear autoencoder algorithm taking the advantage of non-linear properties of hyperspectral data. The proposed algorithm was applied on airborne hyperspectral image of Airborne Visible Infrared Imaging Spectrometer - Next Generation (AVIRIS-NG) over Anand region of Gujarat and the performance of the algorithm was evaluated. The Signal-to-Noise Ratio (SNR) increased from 22.78 dB to 48.48 dB with increase in number of nodes in bottleneck layer for reconstruction of image. Spectral distortion was also measured using Spectral Angle Mapper Algorithm (SAM), which reduced from 0.38 to 0.05 with increase in number of nodes in bottleneck layer up to 10. So, this algorithm was able to give good reconstruction of original image from the nodes present in the bottleneck layer.

1. INTRODUCTION

Recent advances in hyperspectral remote sensing have improved the scope of hyperspectral data application in various fields like agriculture, mineral mapping and stress detection etc. These data carry huge spatial as well as spectral information corresponding to hundreds of narrow continuous bands. Large number of bands present makes it difficult to process data. Dimensionality reduction is the transformation of the information from higher dimensions to low dimension (Plaza et al., 2005). It preserves certain properties of the original data necessary to classify and account for the data. The research carried out in this field shows an increase in the data in an effective way. Therefore, dimensionality reduction is an extremely important step while working with these come in classification accuracy as dimensionality is reduced (Burgers et al., 2009). Mostly in the case of hyperspectral data, PCA (Principal Component Analysis) (Agrawal et al 2011) has been used by researchers which projects the data based on their variance in orthogonal subspace and provides output in the form of uncorrelated principal component. But PCA concentrate on global features present in the image and ignore some of the local features resulting in loss of information present. (Cheriyadat et al 2011). Other conventional dimensionality reduction methods based on orthogonal transformation like LDA (Linear Discriminate Analysis) (Li et al., 2011) MNF (Minimum Noise Fraction) (Harris et al., 2005) and ICA (Independent Component Analysis) (Villa et al. 2011), have been proposed by many researchers for dimensionality reduction. Researches carried out on these techniques establish that LDA outperforms PCA and MNF (Baranowski et al. 2012).

Many researchers have also used wavelet transform for dimensionality reduction of hyperspectral data as it preserves peaks and valleys found in spectra (Bruce et al. 2002 and Kaewpijit et al., 2003). But the major drawbacks in these methods are that they only consider linear properties present in the datasets and ignoring the non-linear properties of data (Sakurda et al., 2012). Linear transformation based method for dimensionality reduction is not suitable for hyperspectral data which has non-linear properties (Chen et al. 2016). Non-Linear properties are present in hyperspectral image due to heterogeneity in pixels and multiple scattering of photons. Therefore, non-linear approach has been used in this research to reduce the dimensionality of hyperspectral data. An attempt has been made to reduce the dimensionality of hyperspectral data using non-linear autoencoder network consisting of three hidden layers with a bottleneck layer whose dimensions is less than input and output layer and a middle bottleneck layer which contains maximum information to give a compact representation of reconstructed output layer and a sigmoid activation function.

2. DATASET USED

The dataset used in this study is acquired by AVIRIS-NG which is a whiskbroom sensor having around 425 contiguous narrow spectral bands in the spectral range of 380-2500 nm with high spectral resolution of about 5 nm. AVIRIS –NG level L2 data which is radiometrically and geometrically corrected datasets over the Anand region of Gujarat has been used in this study. Out of 425 bands, water vapor absorption bands and noisy bands have been removed making the total number of bands to 370 which are used for further analysis. A small subset of 467 X

* Corresponding author

440 pixels of the image is used to evaluate the performance of the proposed dimensionality reduction algorithm.

3. METHODOLOGY

The autoencoder (AE) is a conventional feed-forward neural network based on unsupervised learning algorithm. It is a specific type of neural network that represents a set of data by encoding, which compresses the input data into a lower-dimensional code and then reconstruct the output from this representation. The code is a compact compression of the input. In autoencoder topology, there are three hidden layer, an input layer, a bottleneck layer of lower dimension and a reconstructed output layer. To take into account non-linear properties of the data sigmoid (non-linear) activation function has been used as shown in equation (1)

$$\phi = \frac{1}{(1+e^{-x})} \quad (1)$$

The training of autoencoder network is performed by considering pixels from the image where each wavelength band corresponds to one input layer in an automatic unsupervised manner without having *a priori* knowledge. The training part of the network performs identity mapping where the input layer has to be equal to the output layer. The sum of square error as shown in equation (2) should be minimized while training phase using SCG (Standard Conjugate Mechanism) (Jeyachandran et al., 2011) to get a good reconstruction of the image.

$$E = \frac{1}{NB} \sum_{n=1}^N \sum_{k=1}^B (X_{nk} - X'_{nk})^2 \quad (2)$$

Where,

E is mean squared error; N and B are spatial and spectral dimensions of the hyperspectral image.

As shown in Figure 1, auto encoder network includes an internal bottleneck layer as well as coding and decoding layer which is used to reconstruct the image with reduced dimensions. Various numbers of nodes were tried in coding and decoding layers but keeping the number of nodes in coding and decoding layer to 50 resulting in the best performance of the model which was in accordance with the researches done in this field (Liccardi et al., 2012). The nodes in the internal bottleneck layer are varied to get the exact reconstruction of the image.

To take into consideration non-linear properties of hyperspectral data, autoencoder network was activated with non-linear activation nodes. To compare the proposed algorithm with conventional PCA the autoencoder network was used without coding and decoding layer and using linear activation function (Sakurda et al., 2012). It will give us output with a linear combination of input nodes. The autoencoder network consists of functional mapping and de-mapping of original signal successively. In mapping part, the original image is encoded to a lower dimensional subspace using the non-linear activations functions in the central hidden layer (bottleneck) as given in the equation (3). After training, the mapping function is used to extract the nodes in bottleneck layer having maximum information.

$$G = S(X) \quad (3)$$

In a similar way, second half - part of *i.e.* de-mapping part of the network defines an arbitrary functional mapping that projects the original image from the lower dimensional space back to the original space using only the bottleneck layers as shown in the equation (4). The number of nodes in bottleneck layer was varied from 1 to 10 to optimize the number of nodes that results in reduction the number of feature by decorrelation in the spectral feature space which should contain information to get the best reconstruction of the original image. Then the output image was reconstructed from the optimum nodes with maximum information of the bottleneck layer.

$$X' = T(G) \quad (4)$$

Autoencoder Architecture

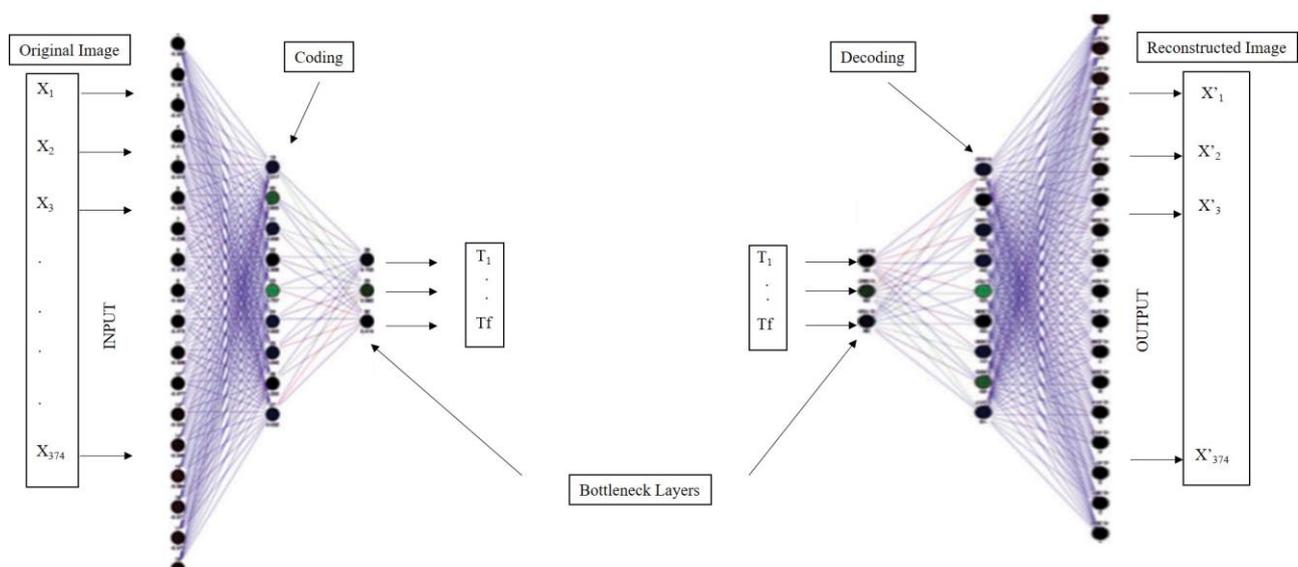


Figure 1. Autoencoder Architecture

The performance of the proposed non-linear autoencoder based network and conventional PCA for dimensionality reduction of hyperspectral data was evaluated in terms of distortion or change in the original data. In this study, Signal-to-Noise Ratio (SNR) was used to evaluate the rate distortion. (Fowler, 2007) The SNR represented in dB (decibels) is used to measure the quality of reconstructed image. Higher SNR shows that quality of reconstructed image has been improved due to removal of noise. SNR is computed using the equation (5).

$$SNR = 10 \log_{10} \frac{\sigma^2}{MSE} \quad (5)$$

where σ^2 = variance of the original image
 MSE = means square error between the original and the compressed image.

Spectral similarity between the original and reconstructed image was evaluated using spectral matching algorithm of Spectral Angle Mapper (SAM) (Feng et al 2015) that computes the similarity in terms of spectral angle:

$$SAM(X, X') = \cos^{-1} \left[\frac{X \cdot X'}{\|X\| \|X'\|} \right] \quad (6)$$

Where X and X' are original and reconstructed image respectively.

A smaller angle represents more similarity and less distortion between the original and reconstructed image.

4. RESULTS AND DISCUSSION

Training of the autoencoder architecture is done using about sixty percent of the randomly selected pixels of the hyperspectral datasets and the error as given in equation (2) is reduced till the error stops reducing stepwise and starts converging. The plot in the Figure 2 shows the model loss in terms of training and validation loss.

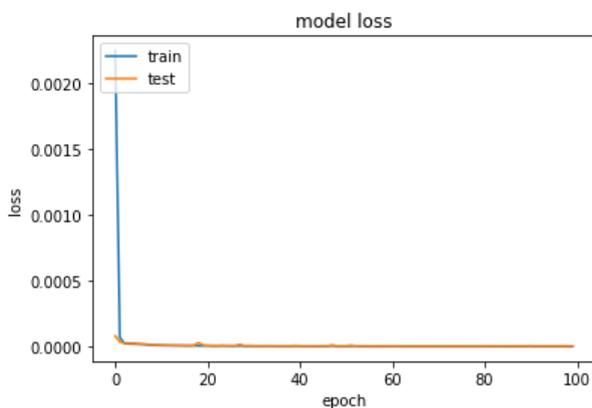


Figure 2. Model Loss in terms of train and test

Figure 2 shows that training and validation loss is synchronized. During the training phase, there is no gap between training and validation loss and the validation loss is decreasing not increasing. This shows that model is not over fitting and the models generalization capability is good. In the autoencoder architecture to compare with traditional PCA linear activation function was used along with non-linear activation function as

in case of proposed algorithm. The performance of the applied algorithm of dimensionality reduction is evaluated using two aspects First aspect is SNR which measures the rate distortion present in the reconstructed image. The second aspect is measured in terms of SAM which is a measure of spectral distortion by evaluating the similarity between the original and reconstructed image obtained by number of nodes in the bottleneck layer of the non-linear autoencoder architecture.

4.1 SNR

The SNR (signal to noise ratio) was calculated using equation (5) between the original and reconstructed image with varying the layer of nodes. It showed an improvement from 22.74 dB to 48.48 dB when varying number of nodes used in bottleneck layer for reconstruction of the image in case of nonlinear as seen in Table 1.

No. of nodes in bottleneck	SNR (non-linear) (dB)	SAM (non-linear) (θ)	SNR (linear) (dB)	SAM (linear) (θ)
1	22.74	0.378	26.19	0.38
2	22.75	0.378	38.11	0.33
3	22.73	0.378	41.99	0.14
4	22.75	0.378	44.45	0.11
5	38.42	0.309	44.71	0.06
6	43.36	0.159	44.10	0.06
7	44.26	0.151	44.42	0.06
8	46.59	0.056	44.21	0.05
9	47.22	0.054	44.09	0.05
10	48.48	0.047	44.73	0.05

Table 1. SAM and SNR values with varying number of nodes in bottleneck layers

The SNR was also computed for linear PCA which shows saturation at 44 dB after 4 neurons in the bottleneck layer seen in Figure 3 This algorithm was also able to remove the presence of striping as shown in Figure 9 and Figure 10 which is not possible in case of traditional PCA.

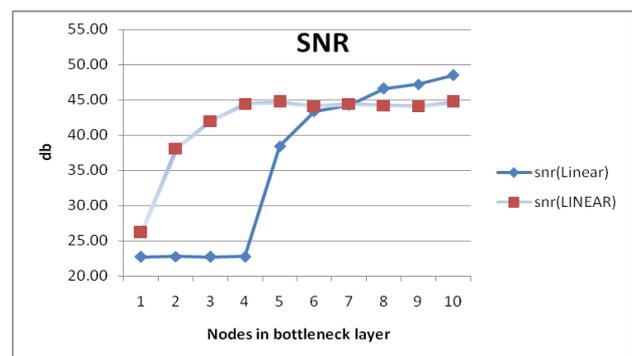


Figure 3. Distortion in terms of SNR

4.2 SAM

The spectral matching was done using SAM algorithm to measure the spectral distortion present in the spectral profile of original and reconstructed image as given in the equation (6). As seen in the Figure 4, SAM was also reduced from 0.38 to 0.05 but in case of linear activation function, it showed saturation after six nodes in bottleneck layer.

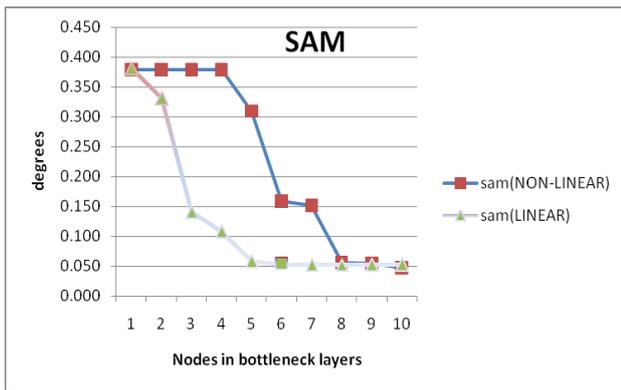


Figure 4. Spectral distortion in terms of SAM

Therefore, best trade-off between the SNR and spectral distortion was obtained by retaining 10 nodes in the bottle neck layer. Reconstruction of the spectral profile was found to be more accurate in case of use of ten nodes in bottleneck layer as compared to use of one node in bottleneck layer for reconstruction of the image as seen in Figure 5 and. Figure 6

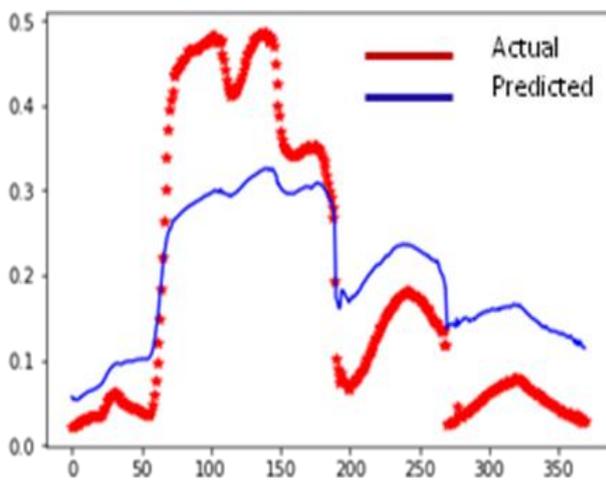


Figure 5. Spectral profile of original and reconstructed image in case of reconstruction using one node in bottleneck layer

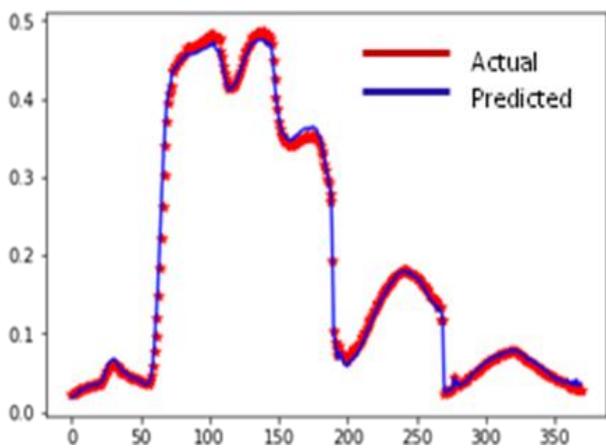


Figure 6. Spectral profile of original and reconstructed image in case of reconstruction using ten nodes in bottleneck layer

So, the proposed algorithm was found better in performance as compared to conventional PCA. The propose algorithm shows

an improvement as it is able to directly compress the image without loss of information and taking into consideration both linear and non-linear properties of the data which may not be possible in case of traditional PCA (Khan et al., 2012).



Figure 7. Original False Color Composite(FCC) image



Figure 8. Reconstructed False Color Composite(FCC) image using 10 nodes in bottleneck layer

As seen in Figure 7 and Figure 8, this algorithm was able to give good reconstruction of the original image using only ten nodes in bottleneck layer. The main advantage of using this network is that once trained it can be used for dimensionality reduction of the entire image acquired by the same sensor. This algorithm was also able to remove the striping effect present in the original image as shown in Figure 9 and Figure 10. Therefore, this algorithm can effectively remove noise along with dimensionality reduction and was able to improve the quality of the image. So this algorithm is able to map linear and nonlinear spectral correlations with a good reconstruction of spectral shape of pixel.

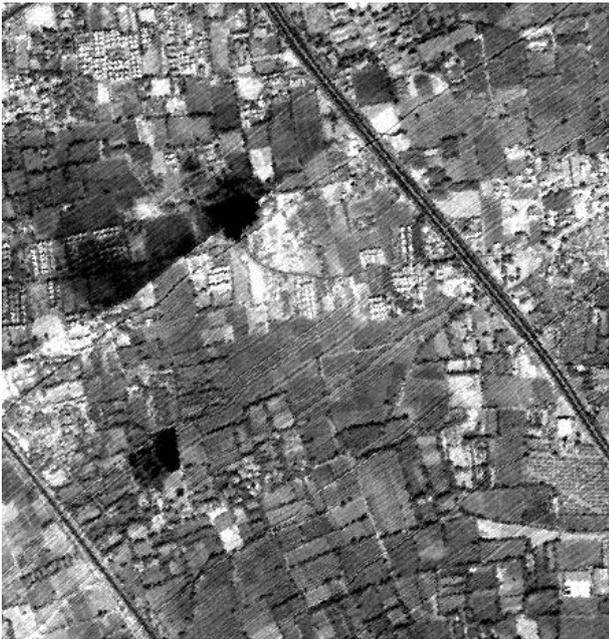


Figure 9. Band 190 (SWIR) of original image with striping effect



Figure 10. Band 190(SWIR) of reconstructed image without striping effect

5. CONCLUSION

The proposed algorithm, non-linear autoencoder was found appropriate for dimensionality reduction of hyperspectral data by preserving the desirable information. The best reconstruction of the original image was obtained by using only ten nodes in the bottleneck layer. So, ten nodes can be utilized for band reduction which is optimum for reconstruction of the original image. In addition to this, the algorithm was also helpful to remove the striping effect which was not possible using the conventional dimensionality reduction algorithm like PCA. This method can remove linear as well as non-linear noise

component present in the image. by taking into consideration non-linear properties of data. The main advantage of using this network is it can be used for dimensionality reduction of the entire image acquired by the same sensor once trained. Therefore, this algorithm can effectively remove noise along with dimensionality reduction and was able to improve the quality of the image. One of the advantage over conventional PCA is that this algorithm can re-project the data into original space by using de-mapping function which is not possible in case of PCA where the data gets transformed and cannot be projected back to original space. It is able to preserve the spectral features of the hyperspectral data which can be used for further classification process based on spectral analysis which is not possible in case of conventional linear dimensionality reduction methods.

ACKNOWLEDGEMENT

Authors are very grateful to K.S Dasgupta Director DAICT for his encouragement and support during the period of investigation. The authors are also thankful to geospatial lab members of DA-IICT, Gandhinagar for their support and also Space Applications Centre-ISRO for providing opportunity to carry out this work and guidance during the study.

REFERENCES

- Agarwal, A., El-Ghazawi, T., El-Askary, H. and Le-Moigne, J., 2007, December. Efficient hierarchical-PCA dimension reduction for hyperspectral imagery. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pp. 353-356, IEEE.
- Bruce, L.M., Koger, C.H. and Li, J., 2002. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on geoscience and remote sensing*, 40(10), pp. 2331-2338.
- Cheriyadat, A. and Bruce, L.M., 2003, July. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International Vol. 6*, pp. 3420-3422, IEEE.
- Feng, S., 2015. Wavelet-Based Non-Homogeneous Hidden Markov Chain Model for Hyperspectral Signature Classification.
- Fowler, J.E. and Rucker, J.T., 2007. Hyperspectral data exploitation: theory and applications. *ch. 3D Wavelet-Based Compression of Hyperspectral Imager*, pp. 379-407.
- Harris, J.R., Rogge, D., Hitchcock, R., Ijewliw, O. and Wright, D., 2005. Mapping lithology in Canada's Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering. *Canadian Journal of Earth Sciences*, 42(12), pp. 2173-2193.
- Harsanyi, J.C. and Chang, C.I., 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Transactions on geoscience and remote sensing*, 32(4), pp.779-785.

- Jeyachandran, C. and Rajaram, M., 2011, December. Comparative performance analysis of various training. In *Trends in Information Sciences and Computing (TISC), 2011 3rd International Conference on*, pp. 5-10, IEEE.
- Kaewpijit, S., Le Moigne, J. and El-Ghazawi, T., 2003. Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE transactions on Geoscience and Remote Sensing*, 41(4), pp. 863-871.
- Li, W., Prasad, S., Fowler, J.E. and Bruce, L.M., 2012. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4), pp. 1185-1198.
- Licciardi, G.A., Khan, M.M., Chanussot, J., Montanvert, A., Condat, L. and Jutten, C., 2012. Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA band reduction. *EURASIP Journal on Advances in Signal processing*, 2012(1), pp. 207.
- Plaza, A., Martinez, P., Plaza, J. and Perez, R., 2005. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Transactions on Geoscience and remote sensing*, 43(3), pp. 466-479.
- Sakurada, M. and Yairi, T., 2014, December. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, pp. 4, ACM.