Convolutional autoencoder pan-sharpening method for spectral indices in landsat 8 images

Jessica da Silva Costa[1] - ORCID: 0000-0003-1947-6386

Hideo Araki[1] - ORCID: 0000-0001-7116-6102

[1]Universidade Federal do Paraná, Departamento de Geomática, Curitiba - Paraná, Brasil.
E-mail: jessica.dsilvacosta@gmail.com, haraki@ufpr.br

Abstract:

Pan-sharpening (PS) consists of combining a high spatial resolution (HR) panchromatic image (PAN) and a low spatial resolution (LR) multispectral image (MS) to generate an MS-HR image. However, some PS methods have spectral and spatial distortions that influence subsequent analyses. Thus, this study aimed to develop a PS method based on convolutional autoencoder (CAE) for Landsat 8 images and evaluate its performance in calculating spectral indices. In the PS process, we trained a CAE network and used a multiscale-guided filter. The performance of the proposed method was analyzed using the Kolmogorov-Smirnov (K-S) statistic of the empirical cumulative distribution function (eCDF) between the values of the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Moisture Index (NDMI) of the PS and MS-LR images. The results show that the proposed method is effective for calculating the indices. Therefore, we conclude that it has great potential for preserving the spatial information of the PAN image and the spectral information of the MS-LR image during the PS process for calculating spectral indices.

Keywords: Deep Learning; Remote sensing; Image fusion.

1. Introduction

The pan-sharpening (PS) method, which stands for "panchromatic sharpening", is used to integrate the geometric details of the high spatial resolution (HR) panchromatic image (PAN) and the spectral information of the low spatial resolution (LR) multispectral image (MS) to obtain a high spatial resolution multispectral image (MS-HR) (Vivone et al., 2021; Dadrass Javan et al., 2021). MS-HR images obtained through this method are more suitable for subsequent applications that rely on both spatial and spectral information (Yang et al., 2016), such as thematic mapping (Ouzemou et al., 2018), visual interpretation (used in commercial software products such as Google Earth and Bing Maps), change detection (Maurya et al., 2020), and vegetation index analysis (Beene et al., 2022), among others.

According to the taxonomy proposed by Vivone et al. (2021), PS methods can be grouped into four main categories: 1) Component substitution (CS) (Shettigara, 1992), which consists of replacing the LR component of the MS image with the PAN-HR image to obtain the pan-sharpened MS-HR image – this category includes the Principal Component Analysis (PCA) method (Chavez Jr and Kwarteng, 1989), the Intensity-Hue-Saturation (IHS) method (Tu et al., 2001), the Gram-Schmidt (GS) method (Klonus and Ehlers, 2009), the Brovey transformation, and other arithmetic combinations (Jiang et al., 2012); 2) Multiresolution Analysis (MRA) (Aiazzi et al., 2002), which is based on the spatial decomposition of the PAN-HR image by the wavelet transform or Laplacian pyramids to extract high spatial structures, that are injected into the interpolated MS-LR image to obtain the MS-HR image – it includes the Additive Wavelet Luminance Proportional (AWLP) (Otazu et al., 2005) and the Generalized Laplacian Pyramid (GLP) (Aiazzi et al., 2006) methods; 3) Variational Optimization (VO) (Liu, Xiao and Li, 2018), which is built on the variational theory and the main process is usually based on or converted to the optimization of an energy functional, such as in model-based methods (Liu et al., 2016) and sparse methods (Zhu and Bamler, 2013); 4) Deep learning (DL), such as approaches based on convolutional neural networks (CNNs) designed for image fusion (Dong et al., 2016). The first two classes (CS and MRA) make up the group of classical and traditionally used approaches. The two remaining classes (VO and DL) represent the main emerging lines of research in recent years.

Although several PS methods have been employed and reported in the literature, some common limitations have been identified, including: color distortion in the PS images; limitations in the skills and knowledge of the people applying the method; low number of high-quality results; and difficulty in building relationships between the individual MS bands and the PAN image (Rahaman, Hassan and Ahmed, 2017). PS images are often used as inputs for subsequent analyses, and the quality of these analyses depends on the accuracy and reliability of these products.

In recent years, studies have developed various metrics for performance evaluation, most focused on examining the quality of pan-sharpened images in terms of color preservation, spatial fidelity, and spectral fidelity (Khan, Alparone and Chanussot, 2009; Agudelo-Medina et al., 2019; Vivone, Addesso and Chanussot, 2019). However, specific index performance evaluations promise to optimize the selection of PS methods used for subsequent analyses that depend on obtaining reliable indices (Beene et al., 2022). Spectral indices derived from high-resolution satellite images are a useful source of data for many forestry, agricultural, environmental, and climate studies (Vélez et al., 2023).

Researchers have recently developed DL-based PS methods using Convolutional Autoencoder (CAE) (Azarang, Manoochehri and Kehtarnavaz, 2019; Al Smadi et al., 2021; Al Smadi et al., 2022), a type of convolutional neural network that can be used to generate an output image based on the input image (Goodfellow et al., 2016). These methods have shown great ability to simultaneously preserve the spatial details and the spectral characteristics of pan-sharpened images. However, most of these methods use high spatial resolution satellite images (around 10 meters or less) with four-band MS images and a PAN image covering the visible and near-infrared spectrum (such as QuickBird). On the other hand, methods that use multispectral images of moderate resolution (around 10 meters or more), which have more spectral bands (such as Landsat 8), are not widely explored, despite their capacity to monitor large areas.

Therefore, the aim of this study is to develop a CAE-based pan-sharpening method for Landsat 8 images and evaluate its performance in calculating spectral indices.

2. Methodology

2.1 Pan-sharpening method

2.1.1 Convolutional Autoencoder Network (CAE)

The CAE network developed in this study was trained with unsupervised learning and was composed of an encoding part – which featured three two-dimensional convolution layers and two Max-pooling layers – and a decoding part – which had three two-dimensional convolution layers and two Upsampling layers. Batch Normalization (BN) (Ioffe and Szegedy, 2015) was applied to all layers except the output layer, avoiding model instability and helping the flow of gradients in the networks (Zhu et al., 2020). The Leaky Rectified Linear Unit (Leaky ReLU) activation function (Xu et al., 2015) was used after the encoding convolutions, and the Rectified Linear Unit (ReLU) activation function (Nair and Hinton, 2010) was used after the decoding convolutions. The sigmoid activation function was used for the output layer. The Mean Square Error (MSE) between the reconstructed output data and the input data was used to update the weights. Then, the backpropagation algorithm was used to train the network, which received the spatially reduced PAN-LR image as input and produced an estimated PAN-HR image as output. Figure 1 shows the architecture of the CAE network.
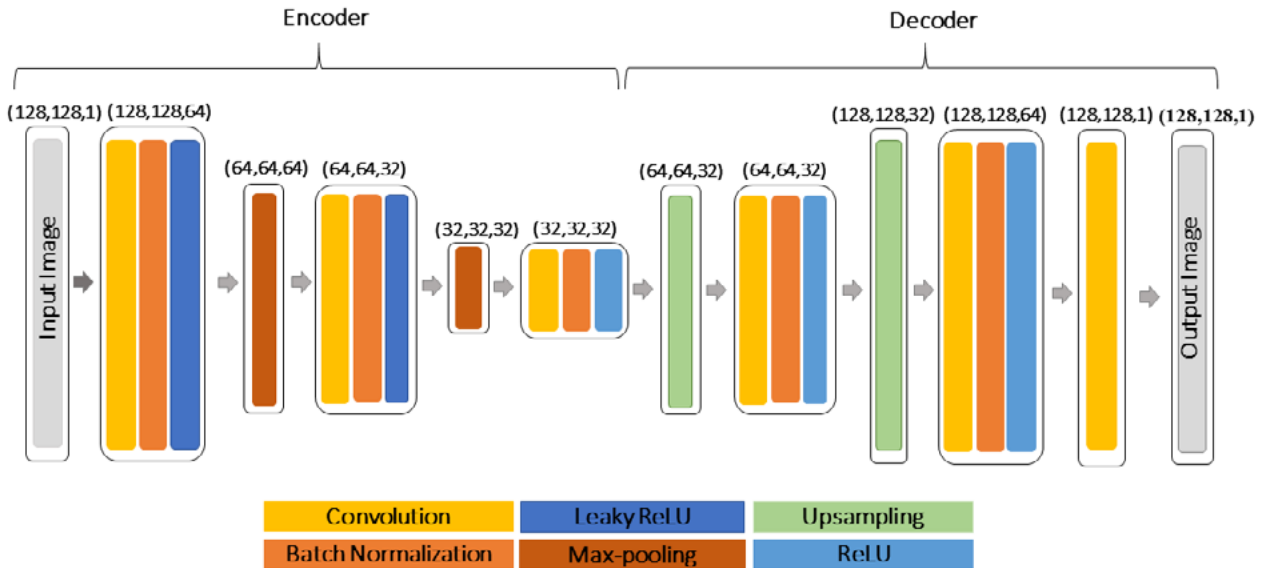

Figure 1: Architecture of the CAE network.

2.1.2 Pan-sharpening convolutional autoencoder method

To develop the proposed PS method, the CAE network was programmed to enhance the spatial details of the intensity (I) component of the resampled MS-LR image, in which the relationship between the PAN image and its degraded form was learned by the network. The trained network is reliable to enhance the corresponding intensity component of the MS image due to its similarity to the PAN image (Al Smadi et al., 2021). Subsequently, spatial details were extracted from the PAN image using a multiscale guided filter method (Yang et al., 2016), in which the

estimated intensity component (EI) was used as the orientation image. Then, the spatial details of the PAN image and an injection gain matrix, used to preserve spectral information, were injected into the resampled MS-LR image, resulting in the PS MS-HR image. Figure 2 shows the methodological flowchart of the proposed PS method.
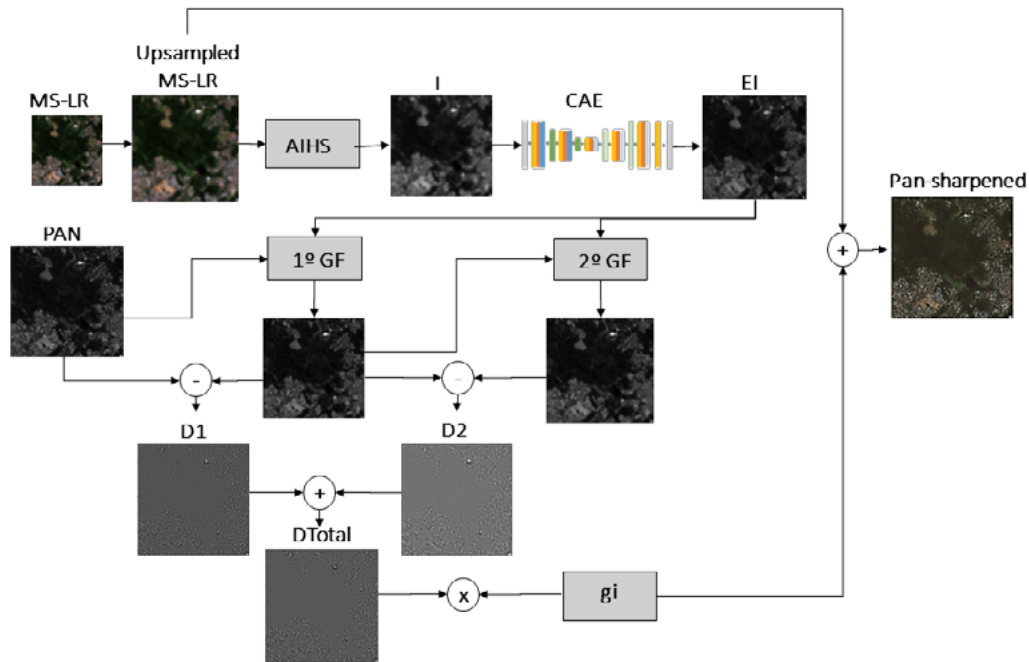

Figure 2: Methodological flowchart of the proposed method.

The Adaptive Intensity-Hue-Saturation method (AIHS) was used to calculate the intensity component (Rahmani et al., 2010). Through Equations 1 and 2, this method reduced the difference between the PAN image and the intensity component (Yang et al., 2016). The original IHS method is based on a color space conversion principle, and it is only suitable for an MS image with exactly three bands. Through some mathematical manipulation, the IHS method can be extended to n bands. The AIHS method has recently been formulated by adaptively adjusting the linear combination coefficients of the MS bands in the spatial detail extraction step (Leung, Liu and Zhang, 2014).

$$I = \sum_{i=1}^{n} \alpha_i M_i \tag{1}$$

$$\alpha_i^* = \arg \min_{\alpha i} \left\| PAN - \sum_{i=1}^{n} \alpha_i M_i \right\|^2 \tag{2}$$

In which αi represents the weight coefficients and n, the number of spectral bands. M_i indicates the ith band of the MS image and PAN corresponds to the panchromatic image.

Derived from a local linear model, the guided filter computes the filtering output by considering the content of a guidance image, which can be the input image itself or another different image. The guided filter can be used as an edge-preserving smoothing operator like the popular bilateral filter (Tomasi and Manduchi, 1998), but it has better behaviors near edges. Moreover, it can transfer the structures of the guidance image to the filtering output, enabling new filtering applications like dehazing and guided feathering (He, Sun and Tang, 2013).

Assuming that the filtering output Z is a linear transformation of the guidance image Y in a local window v_k centered at a pixel k, it can be described by Equation 3.

$$Z_i = a_k Y_i + b_k \quad \forall i \in \nu_k \tag{3}$$

where *Zi* and *Yi* are the ith pixel value of the output and guidance images, respectively, v_k is a window of size (2r+1)x(2r+1). The coefficients a_k and b_k can be estimated by the Equations 4 and 5, respectively.

$$a_k = \frac{\frac{1}{|v|}\sum_{i \in v_k} Y_i X_i - \mu_k \overline{X}_k}{\delta_k^2 + \eta} \tag{4}$$

$$b_k = \overline{X}_k - a_k \mu_k \tag{5}$$

where δ_k^2 and μ_k are the variance and mean of *Y* in v_k, /v/ is the number of pixels in v_k, η is a regularization parameter, and $\overline{X}_k$ is the mean of input image *X* in v_k. All the obtained values of a_k and b_k are first averaged, and then the final guided filter is estimated as Equation 6.

$$Z_i = \overline{a}_i Y_i + \overline{b}_i \tag{6}$$

where $\overline{a}_i = (\frac{1}{|v|}) \sum_{k \in v_i} a_k$ and $\overline{b}_i = \left(\frac{1}{|v|}\right) \sum_{k \in v_i} b_k$.

Equation 7 represents the guided filtering operation in this study.

$$Z = G(X, Y) \tag{7}$$

where *G* represents the guided filter function.

In this method, the EI component is used as an orientation image and the PAN image is used as the input image in the first scale of the guided filter (j=1), as seen in Equation 8.

$$GF(PAN) = G\left(PAN, E_I\right) \tag{8}$$

where *GF* represents the guided filter output.

Afterward, a detail map *(D₁)* to extract the spatial details of the PAN image was obtained by estimating the difference between it and the result of the first scale of the guided filter *GF(PAN)*, as seen in Equation 9. The result of the guided filter is an approximation image of the input image, and the difference between the approximation image and the input image can be considered as the spatial detail information of the input image (Yang et al., 2016).

$$D_1 = PAN - GF(PAN) \tag{9}$$

Then, for the second scale of the guided filter (j=2) was used Equation 10.

$$GF^j(PAN) = G\left(PAN^{j-1}, E_I\right) \tag{10}$$

For the *j*th scale (*j* > 1), *PANⱼ₋₁* is the approximation layer, namely the (*j-1*)th level guided filtered output image.

The difference between *GF(PAN)* and *GFʲ(PAN)* is represented by the spatial detail D_2, as shown in Equation 11.

$$D_2 = GF(PAN) - GF^j(PAN) \tag{11}$$

The total semantic map *(D_{Total})*, shown in Equation 12, is injected into the resampled MS-LR image via the g_i injection gains, which are calculated using Equation 13. Thus, the pan-sharpened MS-HR image is obtained through Equation 14.

$$D_{Total} = D_1 + D_2 \tag{12}$$

$$g_i = \frac{cov(MS_i, E_I)}{var(E_I)} \tag{13}$$

$$Pan - sharpened = M_i + g_i \times D_{Total} \tag{14}$$

2.2 Performance analysis

PS methods' performance is evaluated by a visual analysis, which consists in comparing the spatial and spectral details of the pan-sharpened image with the input PAN and MS images, and by calculating the Empirical Cumulative Distribution Function (eCDF) of the spectral index values, calculated using the original MS images (pre-pan-sharpened) and the PS images. Figure 3 shows the methodological flowchart used to analyze the performance of the PS methods on the spectral indices. All the steps of this methodology were implemented in the Python programming language.
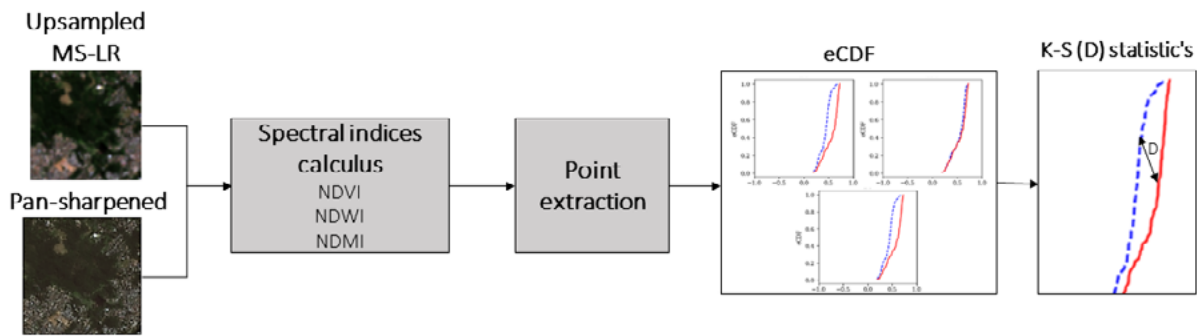

Figure 3: Methodological flowchart for analyzing the performance of pan-sharpening methods.

Equation 15 allows the conversion of digital numbers (DNs) from the MS-LR and pan-sharpened images to top-of-atmosphere (TOA) reflectance. The TOA reflectance was used because the Landsat 8/OLI sensor Collection 2 Level 1 images were utilized in this study.

$$\rho_\lambda = \frac{MpQcal + Ap}{\sin(\theta_{SE})} \tag{15}$$

In which ρ_λ is the TOA reflectance, Mp is the band-specific multiplicative rescaling factor, $Qcal$ are the pixel values (DN), Ap is the band-specific additive rescaling factor, and θ_{SE} is the local sun elevation angle.

The spectral indices used in this analysis are:

1) Normalized Difference Vegetation Index (NDVI)

This index measures green and healthy vegetation. The combination of its normalized difference formulation and the use of chlorophyll's highest absorption and reflectance regions make it useful in a wide range of conditions. The value of this index ranges from -1 to 1. Green vegetation ranges from 0.2 to 0.8 (Rouse et al., 1973). NDVI is calculated through Equation 16.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{16}$$

In which NIR are pixel values in the near-infrared band and Red are pixel values in the visible red band.

2) Normalized Difference Water Index (NDWI)

NDWI aims to delineate and highlight open water features. This method uses reflected near-infrared radiation and visible green light (Green) to increase the presence of such features while eliminating the presence of soil and terrestrial vegetation features. The NDWI value varies from -1 to 1, with 0 being the threshold between water and non-water targets. In other words, if the target is water, then NDWI ≥ 0, and if it is non-water, then NDWI ≤ 0 (Mcfeeters, 1996). The calculation to obtain the NDWI can be seen in Equation 17.

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{17}$$

3) Normalized Difference Moisture Index (NDMI)

NDMI detects moisture levels in vegetation using a combination of near-infrared (NIR) and short-wave infrared (SWIR) spectral bands. The NDMI ranges from -1 to 1, with lower values indicating a low water content in the vegetation and higher values corresponding to a high water content (Gao, 1996). The NDMI is obtained through Equation 18.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \tag{18}$$

In this study, the pixel values of all spectral index (SI) images were extracted using 100 randomly distributed points in each study area. The eCDF was then calculated using Equation 19. Accordingly, we analyzed the discrepancy between the eCDF distributions of the spectral index values extracted from the pan-sharpened image and those of the original MS-LR images. The premise is that an effective PS method will enhance an image whose cumulative distribution pattern of spectral indices is similar or close to the original MS-LR image (Beene et al., 2022).

$$eCDF_{SIj}(SI_j) = \frac{100 \le SI_j}{n} = \frac{1}{n} \sum_{i=1}^{n} 1X_i \le SI_j \tag{19}$$

Where SI is the spectral index value and n is the number of observations.

Furthermore, we used the Kolmogorov-Smirnov (K-S) test (Justel, Peña and Zamar, 1997) to examine whether the eCDF distribution of the pan-sharpened images' spectral indices is statistically different from the eCDF distribution of the pre-pan-sharpened images' spectral indices. The K-S statistic (D) is the absolute maximum distance between the two eCDFs. If the p-value of the K-S test is statistically significant (i.e. $p < 0.05$), we may conclude that the original and pan-sharpened MS images are different and that the method utilized is, therefore, ineffective.

3. Results

3.1 Datasets

The experiment employed PAN images with a spatial resolution of 15 m and MS images with a spatial resolution of 30 m, containing 5 bands (Blue, Green, Red, NIR and SWIR) with a radiometric resolution of 16 bits, from the Landsat 8/OLI sensor (Collection 2, Level 1), located in the municipality of Curitiba, in the state of Paraná, Brazil.

The dataset used to train the CAE network was composed of 208 PAN images of 128x128 pixels for the years 2020, 2021, 2022, and 2023, totaling 832 images. Therefore, 832 PAN-LR images were used as input images, which were spatially degraded at a scale of 2, generated by applying the bilinear interpolation filter, and 832 original PAN-HR images were used, serving as validation images. 558 images were used to train the network, while 274 images and 1000 epochs were used for the test, with a processing time of 3004,536 s (approx. 57 min). The CAE network

training was performed using the Nvidia Tesla T4 GPU (Graphics Processing Units) provided by Google Colab. Figure 4 shows the loss value change curve during training, showing a good fit, with the training and test loss values decreasing to a point of stability.
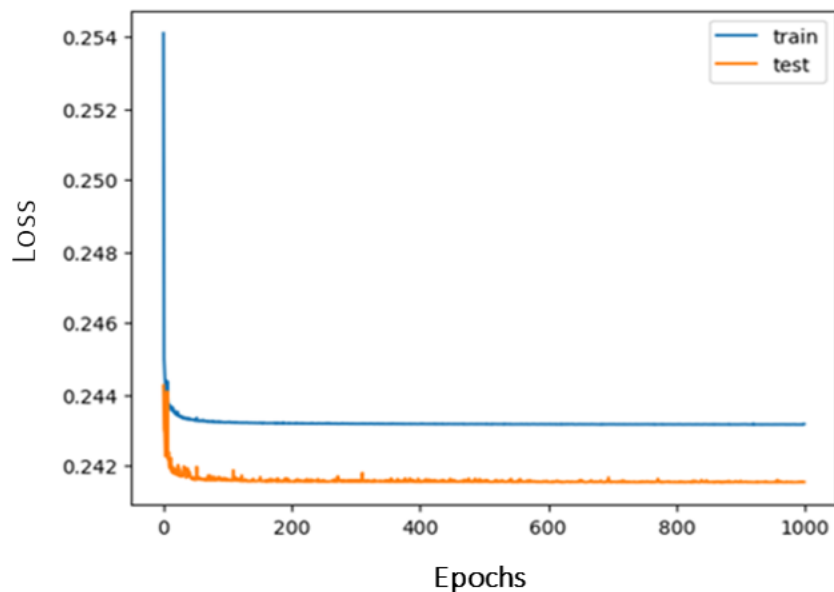

Figure 4: Loss value change curve.

Two datasets of PAN and MS images of 128x128 and 64x64 size, respectively, were selected, one located in a forested area and the other in an urban area. Furthermore, to analyze the performance of the proposed method, the results were compared with a set of traditional PS component replacement methods, including the following methods: Brovey transformation (BT); simple average (SM); ESRI; and IHS, obtained using the ArcGIS software version 10.7. Additionally, the results were also compared with the PS methods Multi-scale guided filter (MSGF) (Yang et al., 2016) and Convolutional autoencoder (CAE) (Azarang, Manoochehri, and Kehtarnavaz, 2019), implemented in the Python programming language.

3.2 Experimental results

3.2.1 Pan-sharpened image results

Figures 5 and 6 show the results of the PS images of a forested area and an urban area, respectively. Visual analysis shows that the proposed method provides good preservation of spatial and spectral information for both areas. In addition, the proposed method performed better in the preservation of the spectral reflectance of the different vegetation areas compared to that obtained by the BT, SM, ESRI, IHS and CAE methods, which showed spectral distortion errors, including changes in vegetation and other land cover characteristics. In addition, the traditional methods showed spectral distortion in urban areas, such as places with buildings, compared to the MS-LR image. Furthermore, it is possible to see that the CAE method presents block artifacts in both areas.
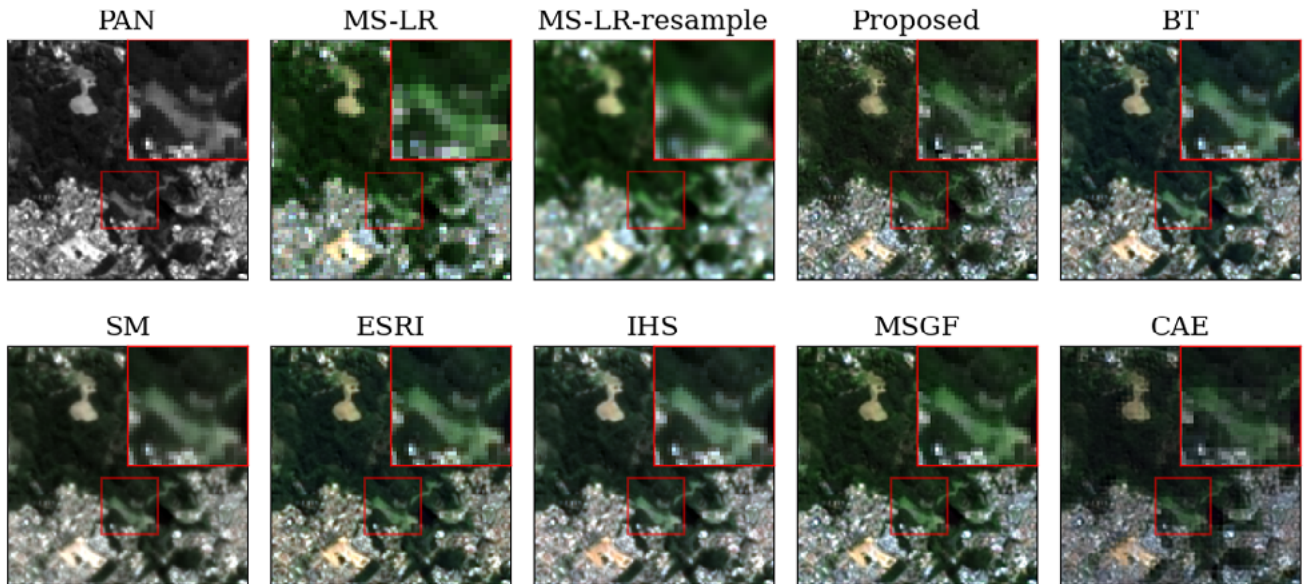
Figure 5: Pan-sharpened results of the "forested area" in true color red, green, and blue (RGB) composition.
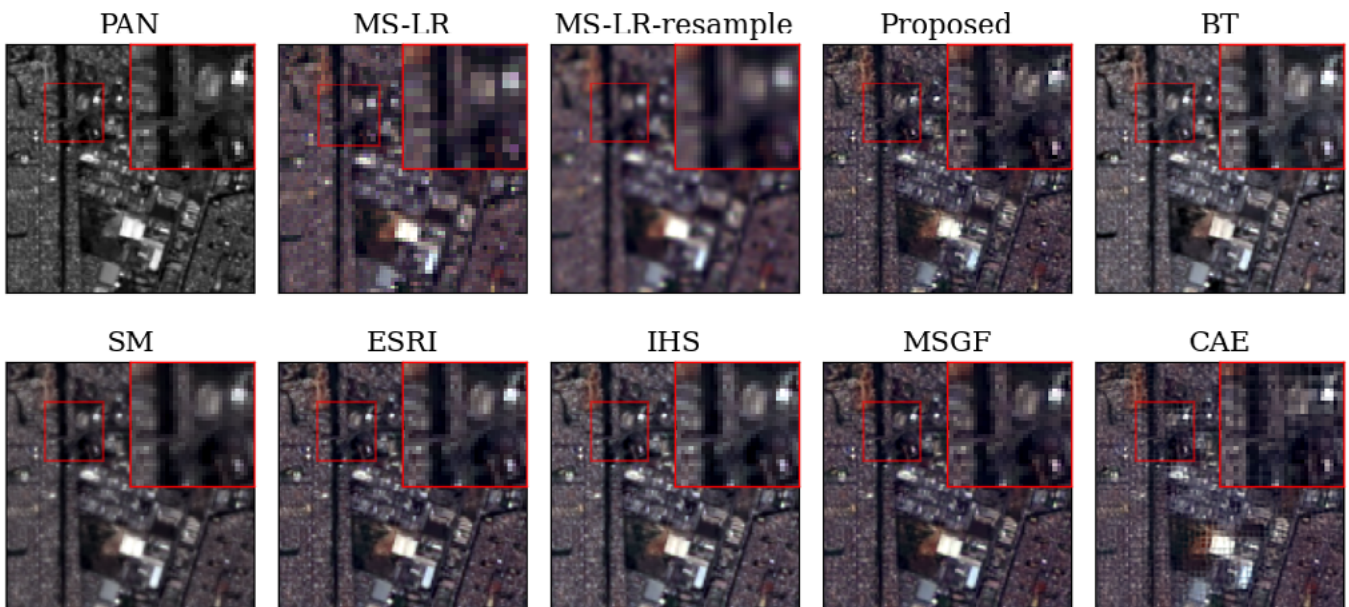

Figure 6: Pan-sharpened results of the "urban area" dataset in true color red, green, and blue (RGB) composition.

3.2.1 NDVI and NDWI index results

Figures 7 and 8 show the distribution of the NDVI and NDWI values of the PS and MS-LR images from the "forested area" and "urban area" datasets, while Figures 9 and 10 show their eCDF distribution. Figures 9 and 10 show that the eCDF distribution of the proposed method's NDVI and NDWI values showed a relatively small discrepancy compared to that obtained from the MS-LR image. Table 1 shows that the proposed method presented the smallest distance (D) between eCDF distributions for the NDVI and NDWI values when compared to those obtained by the other methods for the forested area. The BT method, on the other hand, showed the largest distance (D) between eCDF distributions compared to the other methods. In addition, Table 1 shows that, for the urban area, the proposed method had the smallest distances (D) between eCDF distributions for the NDVI index, while the SM, MSGF and the proposed method had the smallest distances for the NDWI index. The IHS method obtained the greatest distance for both indices. Furthermore, the MSGF, CAE and proposed method showed a

p-value greater than 0.05 for the NDVI and NDWI indices for the two study areas. This indicates that proposed method is effective for calculating these indices in these areas.
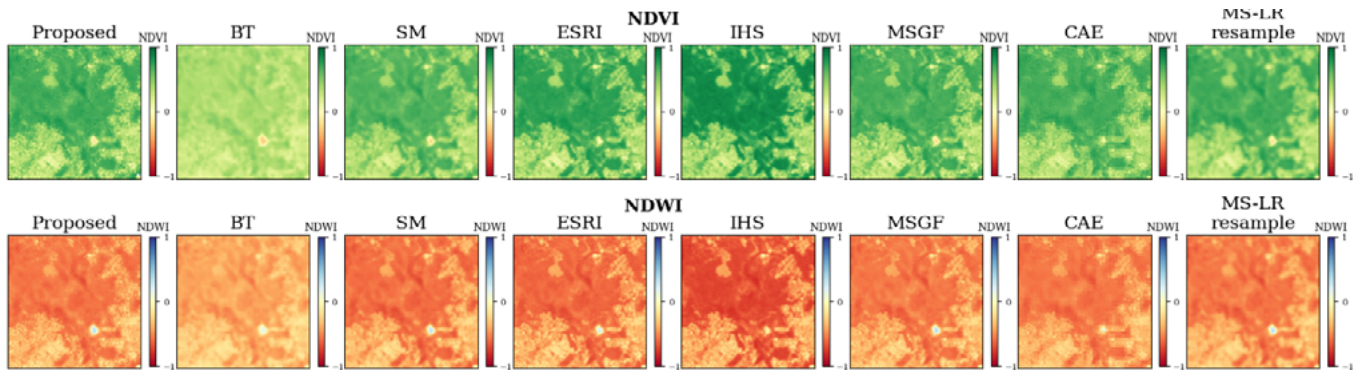

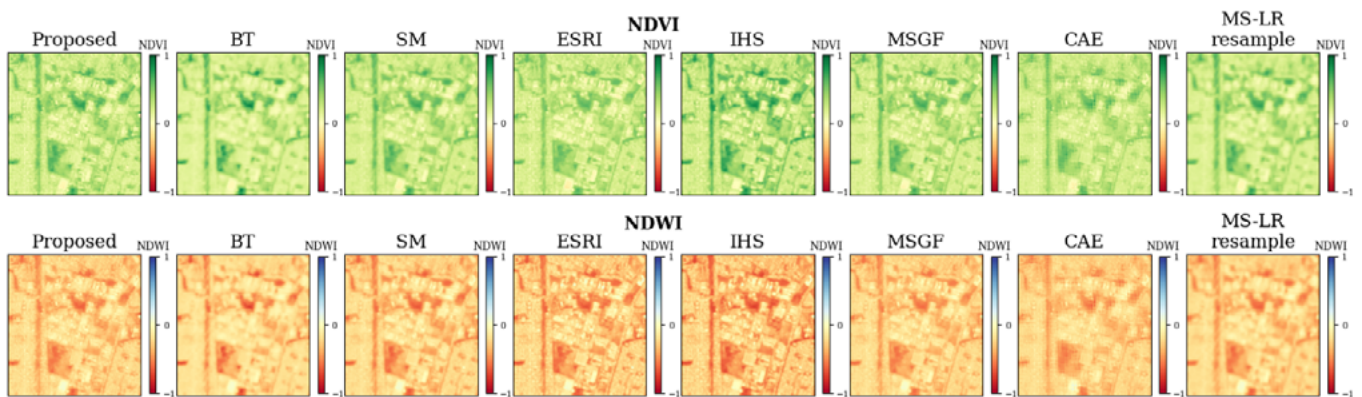**Figure 7:** NDVI and NDWI results for the forested area dataset.

Figure 8: NDVI and NDWI results for the urban area dataset.

Figure 9: Plots of the eCDF of the forested area dataset's NDVI and NDWI values. The red line shows the eCDF of the MS-LR image, while the dashed blue line shows the eCDF of the pan-sharpened image.
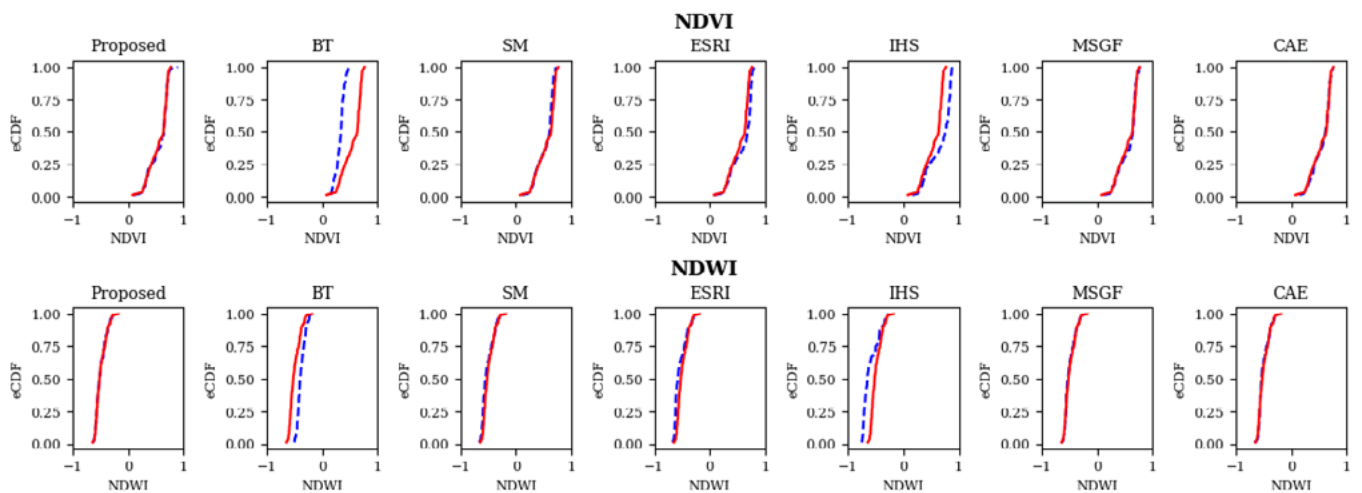
Figure 10: Plots of the eCDF of the urban area dataset's NDVI and NDWI values. The red line shows the eCDF of the MS-LR image, while the dashed blue line shows the eCDF of the pan-sharpened image.
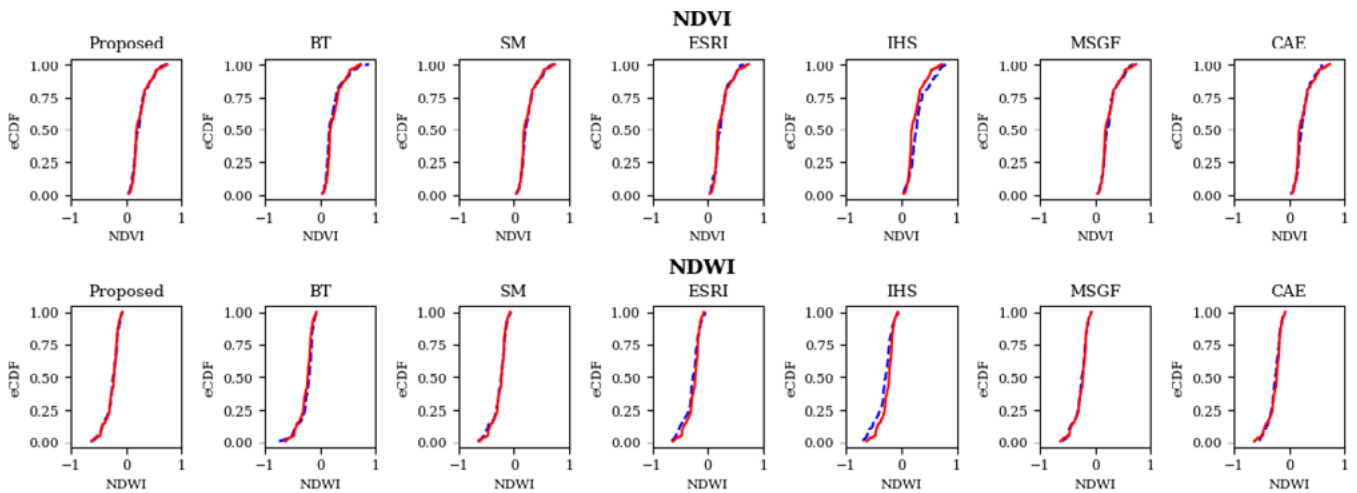
Table 1: Results of the K-S test for the "forested area" and "urban area" datasets for NDVI and NDWI values.

Spectral indices	Pan-sharpening method	Forested Area		Urban Area	
		D	p-value	D	p-value
NDVI	Proposed	0.06	0.9942	0.09	0.8154
	BT	0.69	<0.0001	0.20	0.0363
	SM	0.21	0.0240	0.14	0.2819
	ESRI	0.35	<0.0001	0.14	0.2819
	IHS	0.53	<0.0001	0.29	0.0004
	MSGF	0.08	0.9084	0.13	0.3681
	CAE	0.08	0.9084	0.17	0.1111
NDWI	Proposed	0.05	0.9996	0.10	0.8154
	BT	0.60	<0.0001	0.19	0.0539
	SM	0.19	0.0539	0.07	0.9684
	ESRI	0.35	<0.0001	0.15	0.2111
	IHS	0.58	<0.0001	0.27	0.0012
	MSGF	0.10	0,7020	0.10	0.7020
	CAE	0.10	0,7020	0.12	0.4695

3.2.2 NDMI index results

Figure 11 shows the distribution of the NDMI values of the PS and MS-LR images from the "forested area" and "urban area" datasets, while Figure 12 show their eCDF distribution. The eCDF distributions show that there is a minimal discrepancy between the NDMI values of the PS images compared to the values of the MS-LR image for the two areas. However, Table 2 shows that the proposed method achieved the smallest distance (D) between the eCDF distributions for the two areas, indicating its effectiveness in calculating NDMI.
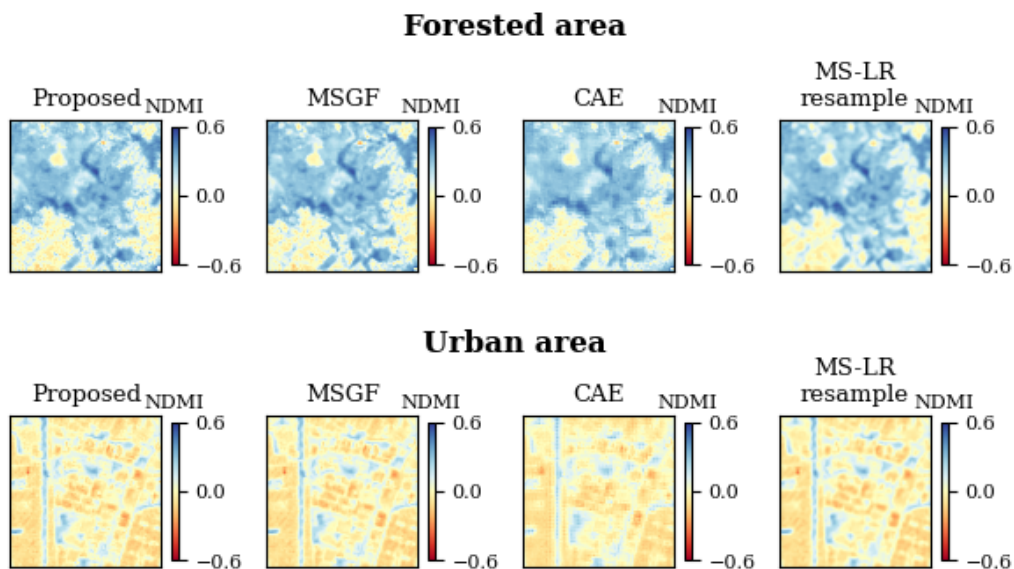
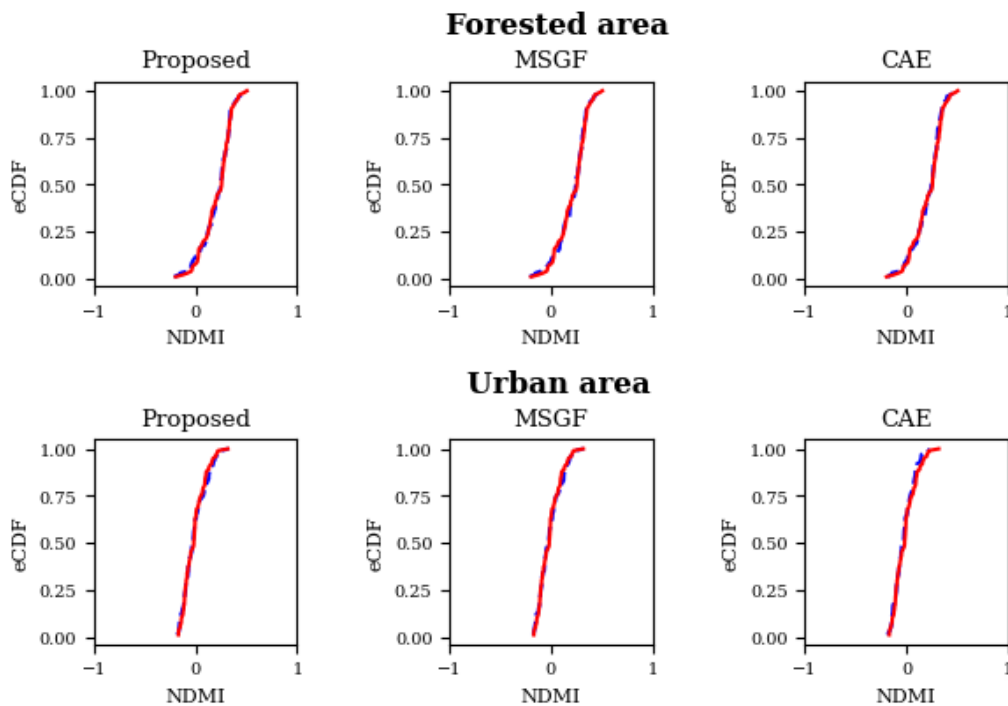**Figure 11:** NDMI results for the forested area and urban area datasets.

Figure 12: Plots of the eCDF of NDMI values for the forested area and urban area datasets. The red line shows the eCDF of the MS-LR image, while the dashed blue line shows the eCDF of the pan-sharpened image.

Table 2: Results of the K-S test for the "forested area" and "urban area" datasets for the NDMI values.

Spectral indices	Pan-sharpening method	Forested Area		Urban Area	
		D	p-value	D	p-value
NDMI	Proposed	0.07	0.9684	0.05	0.9996
	MSGF	0.09	0.8154	0.07	0.9684
	CAE	0.11	0.5830	0.12	0.4695

4. Discussion

According to the studies carried out by Azarang, Manoochehri and Kehtarnavaz (2019), Al Smadi et al. (2021), and Al Smadi et al. (2022), PS methods based on CAE have a great ability to simultaneously preserve the spatial details and spectral characteristics of pan-sharpened images. Pan-sharpened images are used as inputs for subsequent analyses, and the quality of these analyses depends on the quality of the initial products. In this study, we developed a PS method based on CAE for Landsat 8 images and analyzed its potential for calculating the NDVI, NDWI, and NDMI indices. High-resolution spectral indices are a crucial tool for monitoring plant growth and health, assessing the impact of environmental factors on vegetation, and supporting decision-making processes in agriculture and forestry (Vélez et al., 2023).

According to the results found through visual analysis, the proposed method showed spatial characteristics similar to those of the PAN image and managed to maintain spectral reflectance to the low-resolution multispectral image (MS-LR) for the two study areas. The proposed method showed better performance in preserving spectral reflectance in vegetated areas compared to traditional methods, which showed alterations when compared to the MS-LR image. In Landsat 8 images, the PAN band covers the visible spectral regions, within a spectral range of 0.50 to 0.68 mm, resulting in spectral distortion of targets such as vegetation when PS methods are used. Healthy green vegetation is very reflective in the near-infrared, and as the PAN band is generally interpreted as "intensity" in most PS methods, vegetation pixels appear lighter and have a bluish color in real color representations (Borel, Tuttle and Spencer, 2010). In addition, we can analyze the great ability of the proposed method and other analyzed methods to preserve the spectral information of PS images compared to MS-LR images in urban areas. In the studies carried out by Azarang, Manoochehri and Kehtarnavaz (2019) and by Al Smadi et al. (2021), who used a CAE architecture to develop a pan-sharpening method, a better performance in preserving the spectral information for vegetated areas and urban areas was achieved compared to other pan-sharpening methods analyzed in their work. Furthermore, it is possible to see that the CAE method presents block artifacts in both areas, a phenomenon of grid artifact at the edge of the patches used in its processing, which influences the visual quality of the generated pansharpening images.

The results obtained between the traditional analyzed PS approaches depend on the scene used, showing different behaviors for natural and urban settings. On the other hand, the results obtained by the proposed method performed well for both forested and urban areas. Images of urban areas are particularly challenging for the PS process. High-contrast features, such as the edges between the roofs of buildings and a street and details smaller than the spatial resolution, are particularly difficult to render accurately (Vivone et al., 2021). Although vegetated areas have less contrasting spatial characteristics than urban areas, this type of land cover has peculiar features that pose a challenge to the PS process. Vegetated areas are usually textured with patterns that can be regular (e.g. in agricultural fields) or irregular (e.g. in forested regions), leading to potential spatial distortions in the PS results (Vivone et al., 2021).

In addition, the traditional methods analyzed here have a 4-band restriction, and bands in the visible and near-infrared spectral range are often used. Therefore, the traditional methods were not used in this study to calculate the NDMI index, which makes use of the NIR-SWIR (near-infrared and shortwave infrared) combination. However, the MSGF, CAE and proposed methods do not suffer from this band limit, and were managed to obtain a smaller spectral and spatial differences between the images. Moreover, the proposed method demonstrated a smaller distance between the eCDF distributions for the NDVI and NDMI indices for both study areas, and for the NDWI value for the forest area. Thus, it can be used to calculate different spectral indices, which are not limited to the visible and near-infrared spectral bands.

5. Conclusion

In this study, we developed a PS method using a CAE network on Landsat 8 images to be applied in the calculation of spectral indices. The results showed, through statistical analysis, that the proposed method is effective for calculating the NDVI, NDWI and NDMI indices for the two study areas. In addition, visual analysis shows that it achieves greater spectral preservation compared to the original image, compared to other methods. We can therefore conclude that the proposed method has great potential for preserving the spatial information of the PAN image and the spectral information of the original multispectral image during its pan-sharpening process, as well as for calculating spectral indices in forested and urban areas.

For future studies, we recommend applying this method to calculate other spectral indices, such as the Enhanced Vegetation Index (EVI), the Leaf Area Index (LAI), and the Soil Adjusted Vegetation Index (SAVI). We also recommend using this method on other moderate resolution images, such as Sentinel-2 images, which are freely available, have multispectral images with a resolution of 10 m, and do not have a PAN image.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTION

Jessica da Silva Costa: Conceptualization; formal analysis; methodology; validation; writing-original draft; writing - review and editing. **Hideo Araki:** Conceptualization; formal analysis; methodology; validation; writing-original draft; writing – review and editing; supervision; visualization.

REFERENCES

Agudelo-Medina, O. A. et al. 2019. Perceptual quality assessment of pan-sharpened images. *Remote Sens.,* 11(7), pp. 877.

Aiazzi, B. et al. 2002. Context driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Trans. Geosci. Remote Sens.*, 40(10), pp. 2300-12.

Aiazzi, B. et al. 2006. MTF-tailored multiscale fusion of high-resolution MS and pan imagery. *Photogramm. Eng. Remote Sens.*, 72(5), pp. 591–96.

Al Smadi, A. A. et al. 2022. A Pansharpening Based on the Non-Subsampled Contourlet Transform and Convolutional Autoencoder: Application to QuickBird Imagery. *IEEE Access*, 10, pp. 44778-88,

Al Smadi, A. et al. 2021. Pansharpening based on convolutional autoencoder and multi-scale guided filter. *J Image Video Proc*, 25.

Azarang, A, Manoochehri, H. E. and Kehtarnavaz, N. 2019. Convolutional autoencoder-based multispectral image fusion. *IEEE Access*, 7, pp. 35673-83.

Beene, D. et al. 2022. Performance evaluation of multiple pan-sharpening techniques on NDVI: A statistical framework. *Geographies*, 2, pp. 435-452.

Borel, C. C., Tuttle, R, F. and Spencer, C. 2010. Improved panchromatic sharpening of multi-spectral image data, Proc. SPIE 7812, Imaging Spectrometry XV, 78120G.

Chavez Jr. P. S. and Kwarteng, A. W. 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.*, 55(3), pp. 339–48.

Dadrass Javan, F. et al. 2021. A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171, pp. 101-17.

Dong, C. et al. 2016. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.,* 38(2), pp. 295–307.

Gao, Bo-Cai. 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, pp. 257-66.

Goodfellow, I.; Bengio, Y.; Courville, A. 2016. *Deep Learning*. MITPress.

He, K., Sun, J. and Tang, X. 2013. Guided image _ltering. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (6), pp. 1397-1409.

Ioffe, S. and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. Lille: France, 37, pp. 448-56.

Jiang, C. et al. 2012. A practical compressed sensing-based pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.,* 9(4), pp. 629–33.

Justel, A., Peña, D. and Zamar, R. 1997. A multivariate Kolmogorov–Smirnov test of goodness of fit. *Statistics & Probability Letters.*, 35(3), pp. 251–59.

Khan, M. M., Alparone, L. and Chanussot, J. 2009. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE Trans. Geosci. Remote Sens.*, 47(11), pp. 3880–891.

Klonus, S. and Ehlers, M. 2009. Performance of evaluation methods in image fusion. In: International Conference on Information Fusion. Seattle, WA: IEEE, pp. 1409-16.

Leung, Y., Liu, J. and Zhang, J. 2014. An Improved Adaptive Intensity–Hue–Saturation Method for the Fusion of Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 11 (5), pp. 985-89.

Liu, P. 2016. Spatial-Hessian-Feature-Guided variational model for pan-sharpening. *IEEE Tran. Geosci. Remote Sens.* 54 pp. 2235–53.

Liu, P., Xiao, L. and Li, T. 2018. A variational pan-sharpening method based on spatial fractional-order geometry and spectral–spatial low-rank priors. *IEEE Trans. Geosci. Remote Sens.*, 56(3).

Maurya, A. K. et al. 2020. Effect of pansharpening in fusion based change detection of snow cover using convolutional neural networks. *IETE Technical Review*, 37(5), pp. 465-75.

Mcfeeters, S. K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), pp.1425-32.

Nair, V. and Hinton, G. E. 2010. Rectified linear units improve restricted boltzmann machines. In: International conference on machine learning, 27, 2010, Haifa: Israel. pp. 807-14.

Otazu, X. et al. 2005. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.,* 43(10), pp. 2376–85.

Ouzemou, J. E. et al. 2018. Crop type mapping from pansharpened Landsat 8 NDVI data: A case of a highly fragmented and intensive agricultural system. *Remote Sensing Applications: Society and Environment,* 11, pp. 94-103.

Rahaman, K.R.; Hassan, Q.K. and Ahmed, M.R. 2017. Pan-sharpening of Landsat-8 images and its application in calculating vegetation greenness and canopy water contents**.** *ISPRS Int. J. Geo-Inf. 6*(6), pp.168.

Rahmani, S. et al. 2010. An adaptive IHS pan-sharpening method. *IEEE Geosci.Remote Sens. Lett.,* 7 (4), pp.746–50.

Rouse, J., Haas, R., Schell, J; Deering, D. 1973. Monitoring Vegetation Systems in the Great Plains with ERTS. In: *Third ERTS Symposium*, NASA. pp.309-317.

Shettigara, V. K. 1992. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. Photogramm**.** *Eng. Remote Sens*. 58, pp. 561–567.

Tomasi, C. and Manduchi, R. 1998. Bilateral Filtering for Gray and Color Images. *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Bombay, India, 1998, pp. 839-846.

Tu, T. M. et al. 2001. A new look at IHS-like image fusion methods. *Inform. Fusion*, 2(3), pp. 177-86.

Vélez, S., Martínez-Peña, R. and Castrillo, D. 2023. Beyond Vegetation: A review unveiling additional insights into agriculture and forestry through the application of vegetation indices, 6, pp. 421-436.

Vivone, G. et al. 2021. A new benchmark based on recent advances in multispectral pansharpening: revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1), pp. 53-81.

Vivone, G.; Addesso, P.; Chanussot, J. 2019. A combiner-based full resolution quality assessment index for pansharpening. *IEEE Geosci. Remote Sens. Lett***.**, 16(3), pp. 437–41.

Xu, B. et al. 2015. Empirical evaluation of rectified activations in convolutional network.

Yang, Y. et al. 2016. Remote sensing image fusion based on adaptive IHS and multiscale guided filter. *IEEE Access*, 4, pp. 4573-82.

Zhu, D. et al. 2020. Spatial interpolation using conditional generative adversarial neural networks. *International Journal of Geographical Information Science,* 34(4), pp.735-58.

Zhu, X. X. and Bamler, R. 2013. A sparse image fusion algorithm with application to pan-sharpening. *IEEE Trans. Geosci. Remote Sens*., 51(5), pp. 2827–36.