Middle East Journal of Agriculture Research
Volume: 11 | Issue: 02| April – June| 2022

EISSN: 2706-7955 ISSN: 2077-4605
DOI: 10.36632/mejar/2022.11.2.42
Journal homepage: www.curresweb.com
Pages: 630-637

Development of a PCA-based land use/land cover classification utilizing Sentinel-2 time series

Makar R. S.[1], Shahin Sahar A.[1], El-Nazer Mostafa [2] and Wheida Ali[2]

[1]Soils and Water Use Department, Agricultural and Biological Research Institute, National Research Centre, Dokki, Cairo, Egypt.
[2]Theoretical Physics Department, Physics Research Institute, National Research Centre, Dokki, Cairo, Egypt.

Received: 25 May 2022 **Accepted:** 15 June 2022 **Published:** 25 June 2022

ABSTRACT

Land use/ land cover mapping and characterization is required for resource management and planning. In this aspect, remote sensing methods can be employed to classify the land use/ land cover classes over selected areas in an effective and economical manner compared to traditional surveys. In this research an improved supervised classification scheme for Sentinel-2 images classification of a selected area in El-Beheira governorate was developed. Field survey was carried out to collect ground truth data for from December 2020 to March 2021. The supervised classification was preformed after applying various principle component analyses (PCAs) on the used sequential Sentinel-2 images within the winter season. The results revealed that utilizing the proposed image classification technique, an overall accuracy of 86.8% could be achieved for the produced Land Use/Land cover map. The agricultural area covered about 89% of the studied area and was occupied by seven crops. Wheat and Egyptian clover were the major crops and covered about 67% of the studied area while green beans, potato and citrus covered about 21%.

Keywords: principle component analysis, land use, land cover, image classification, Sentinel-2

1. Introduction

Land cover refers to the observed biophysical cover of the earth's surface (Satir & Berberoglu, 2012), it describes the natural and anthropogenic features which can be observed on the Earth's surface (Fonji and Taff, 2014). Conversely, land use refers to act on the land, influenced by economic, cultural, political, historical, and land-tenure factors (Desai, 2012).

Accurate and reliable information about land use and land cover is essential for monitoring and change detection. It's also useful for updating the geographical information about the world (Deb and Nathr, 2012). Utilizing remote sensing for monitoring land use/ land cover (LU/LC) has various advantages compared to traditional surveys such as the potential to produce fast, accurate and up to date inventory of large areas.

Various image processing and pattern recognition techniques are used to produce LU/LC maps. Nevertheless, the large datasets, which constitute these images, are often difficult to interpret (Jolliffe and Cadima, 2016). Principal components analysis (PCA) is a technique applied to both multispectral and hyperspectral remotely sensed data and commonly used for handling quantitative data. PCA can provide more accurate remote sensing image classification, speeds the processing, and easily handles a large number of image data due to its capability of reducing data dimensionality and complexity (Tripathy *et al.*, 2013). It is one of the widely used techniques for dimensionality reduction of large datasets and at the same time, it minimizes information loss and preserves as much variability as possible (Jolliffe and Cadima, 2016). This reduction is achieved by

Corresponding Author: R. S. Makar, Soils and Water Use Department, Agricultural and Biological Research Institute, National Research Centre, El Buhouth St., Dokki, Cairo, Egypt Postal Code 12622. Email: randa_sgmm@yahoo.com, ORCiD number: 0000-0002-1445-0075

transforming the data into a new set of variables called the principal components (PCs). These PCs are uncorrelated, and ordered in a manner that the first few contain most of the variation present in all of the original variables (Naik, 2018). The images produced from PCA are then subjected to classification either supervised (Abdu, 2019) or unsupervised (Abedini *et al.*, 2012). The main advantage of applying PCA prior to image classification is mostly to enhance the used images and reduce the correlation between the used bands (Dadon *et al.*, 2019). The aim of the current research is to suggest a simple, accurate and repeatable PCA based technique for LU/LC mapping utilizing Sentinel-2 time series especially for complex agricultural landscapes.

2. Materials and Methods

2.1. Study area

The study area is located about 45 km south of Alexandria City and considered as part of El-Beheira Governorate. The area is bounded by latitudes 30°41'55" and 30°53'20" N and longitudes 29°58'40" and 30°11'20" E. It extends to cover a total acreage of about 42,000 feddans. It is surrounded by three irrigation canals; Nubariyah Canal to the east, Al-Hidayah Canal to the north and An-Nasr Canal to the south, while Cairo-Alexandria Desert Road outlined its western boundary (Fig. 1) .

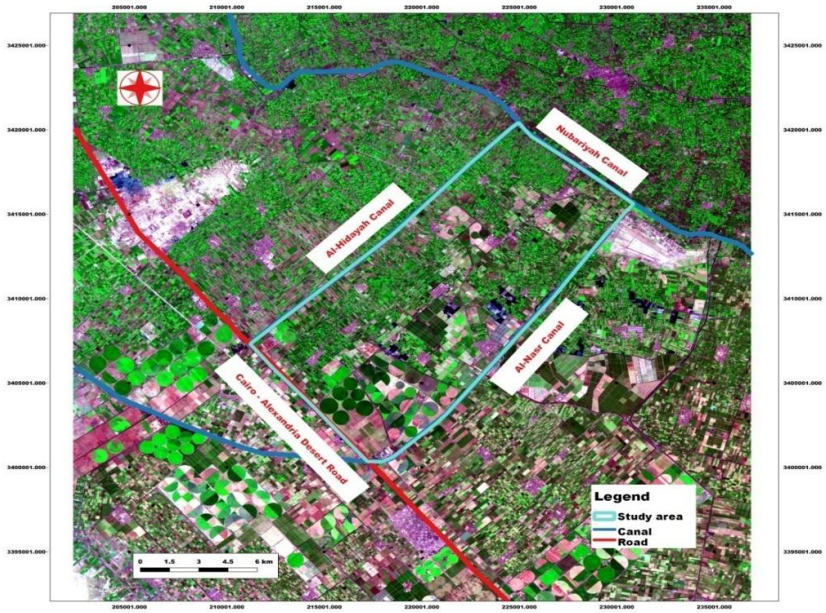

Fig 1: Location map of the study area

2.2. Remote sensing data collection and pre-processing

Sentinel-2 (S2) data were used in this research and four images were collected on 18/10/2020, 27/12/2020, 17/03/2021 and 6/04/2021. Fieldwork was carried out from December, 2020 to March, 2021. Sixty observations were collected and geo-referenced using GPS utility. These images were selected cloud free and covered most of the winter growing season, which extended from October 2020 to May 2021. S2 was developed by the European Space Agency (ESA) and provides high spatial, spectral and temporal resolution images of the earth. The satellites provide a set of 13 spectral bands, four bands at 10m, six at 20m and three at 60m spatial resolution (Vajsová and Aastrand 2015). In this study, only nine S2 bands were used and their characteristics are shown in Table 1 (Li *et al.*, 2021). The S2 data were available free of charge at https://earthexplorer.usgs.gov. They were available in Geographic Markup Language JPEG2000 (GMLJP2) format and were radiometrically and geometrically corrected and projected into Universal Transverse Mercator (UTM) and World Geodetic System 1984 (WGS 84) with datum-zone 36 N. The S2 data were pre-

processed using the Semi-Automatic Classification Plugin (SCP) of the QGIS software. They were imported, stacked into one multiband image for each date, resampled into 10m and subsetted into the geographical extent of the study area. Furthermore, the matching bands of each image for each date were stacked resulting nine images each including four bands representing the four dates.

Table 1: Sentinel-2 Multispectral Imager band characteristics

Band Number	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
2	490	65	10
3	560	35	10
4	665	30	10
5	705	15	20
6	740	15	20
7	783	20	20
8	842	115	10
11	1610	90	20
12	2190	180	20

The Principal component analysis

The principal component analysis was adopted in this study. PCA is considered a powerful tool for data analysis and pattern recognition, which is often used in signal and image processing (Gonzales and Woods, 2002). It allows redundant data to be compacted into fewer bands, by which the dimensionality of the data is reduced. The bands of PCA will be non-correlated and independent, and are ordered in terms of the amount of variance. The first two or three components will carry most of the real information of the original data set (Santiesteban, 2003).

Image classification and accuracy assessment

Supervised classification utilizing the maximum likelihood classification (MLC) is the most common supervised classification method used for analyzing satellite image data. It makes use of a discriminate function to assign pixel to the class with the highest likelihood (Ahmad and Quegan, 2012). MLC requires the selection of training samples representing the different “spectral classes" to adequately represent each class. These training data are used to estimate the parameters of a probability density function for each spectral class (Bolstad and Lillesand, 1991). In the present research, sixty training samples (locations) were used to train MLC classifiers to categorize the LU/LC classes of the study area. The overall accuracy of the classified image compares how each of the pixels is classified versus the definite land cover conditions obtained from their corresponding ground truth data. The overall accuracy is calculated as the percentage of all validation pixels correctly classified to the total number of pixels in all the ground truth classes (Rwanga and Ndambuki, 2017).

3. Results and Discussion

The correlation matrix was developed to investigate the correlation between the different S2 bands used in this study. The results are shown in Table 2, 3, 4 and 5 for dates 18/10/2020, 27/12/2020, 17/03/2021 and 6/04/2021 respectively. Within this matrix, the correlation coefficient between the reflectance of each pair of satellite bands is computed for each pixel in the images. In the four studied images, the pair of bands 11 and 12, which are both in the shortwave infrared region of the spectrum, had the highest correlation.

Table 2: The correlation matrix between the S2 bands acquired on 18th of October

Band	B2	B3	B4	B5	B6	B7	B8	B11	B12
B2	1.00								
B3	0.97	1.00							
B4	0.95	0.97	1.00						
B5	0.89	0.93	0.94	1.00					
B6	0.51	0.62	0.52	0.66	1.00				
B7	0.36	0.47	0.37	0.51	0.97	1.00			
B8	0.35	0.47	0.36	0.47	0.90	0.93	1.00		
B11	0.78	0.83	0.88	0.93	0.60	0.47	0.45	1.00	
B12	0.82	0.86	0.91	0.93	0.52	0.38	0.36	0.98	1.00

Table 3: The correlation matrix between the S2 bands acquired on 27th of December

Band	B2	B3	B4	B5	B6	B7	B8	B11	B12
B2	1.00								
B3	0.96	1.00							
B4	0.95	0.96	1.00						
B5	0.88	0.94	0.94	1.00					
B6	-0.04	0.14	-0.06	0.17	1.00				
B7	-0.19	-0.02	-0.22	-0.01	0.98	1.00			
B8	-0.19	-0.02	-0.21	-0.01	0.93	0.95	1.00		
B11	0.85	0.87	0.93	0.94	-0.04	-0.19	-0.19	1.00	
B12	0.81	0.86	0.89	0.94	0.13	-0.03	-0.03	0.98	1.00

Table 4: The correlation matrix between the S2 bands acquired on 17th of March

Band	B2	B3	B4	B5	B6	B7	B8	B11	B12
B2	1.00								
B3	0.98	1.00							
B4	0.96	0.98	1.00						
B5	0.89	0.94	0.95	1.00					
B6	0.28	0.39	0.30	0.44	1.00				
B7	0.00	0.10	0.02	0.14	0.98	1.00			
B8	-0.01	0.09	-0.01	0.10	0.87	0.94	1.00		
B11	0.83	0.88	0.91	0.95	0.38	0.09	0.07	1.00	
B12	0.86	0.90	0.93	0.96	0.30	0.00	-0.02	0.99	1.00

Table 5: The correlation matrix between the S2 bands acquired on 6th of April

Band	B2	B3	B4	B5	B6	B7	B8	B11	B12
B2	1.00								
B3	0.98	1.00							
B4	0.95	0.98	1.00						
B5	0.89	0.94	0.95	1.00					
B6	0.22	0.33	0.22	0.37	1.00				
B7	0.02	0.12	0.01	0.15	0.96	1.00			
B8	-0.01	0.10	-0.02	0.11	0.90	0.94	1.00		
B11	0.82	0.86	0.89	0.93	0.37	0.17	0.14	1.00	
B12	0.86	0.89	0.92	0.94	0.26	0.04	0.02	0.98	1.00

There was also a high correlation between other pairs of bands from 2 & 3 & 4 (blue and green and red), i.e. all the bands of the visible range. High correlation was also observed between bands 6 and 7 (both bands area in the red edge range). The lowest correlation was between pairs of bands 2, 3, 4 and band 5, i.e. the visible bands and first red edge band, and bands 6, 7 and 8, i.e. the second and third red edge bands and the near infrared (NIR) band. Furthermore, the same trend was observed between the later three bands and the two shortwave infrared bands (band 11 & 12). Nevertheless, the above-mentioned low correlations were not prominent in the first image of 18th of October. This could be rendered to the influence of substantial presence of vegetation in the study area in the other three dates, as that portion of the spectrum (second and third red edge bands and the near infrared) the reflection is highest and correlated to the presence vegetation and affected by its different kind, type, health and growth stages. Whereas the image acquired on 18th of October represented the beginning of the winter season and was characterized by the absence of aboveground vegetation cover in the case of the field crops and vegetables. The high correlations between the pair of bands suggest that possibility of removing some of these bands from further analyses.

Based on fieldwork nine LU/LC classes were found. These classes included wheat, Egyptian clover (barssem), green beans, potato, citrus, strawberry, guava, fishponds and settlements. Using the ENVI software the training sets (locations) were developed into region of interests. The separability of the region of interests (ROI), was evaluated to examine spectral separability between each two classes as more separable training data will result in better classification. The separability was determined by using the Jeffries-Matusita measures (Bhaskaran *et al.*, 2010) for the images. The separability between each two classes should be more than 1.8 to indicate that each class is properly separated (Zhu *et al.*, 2007). In case of citrus, strawberry, guava, fish ponds and settlements, the separability between classes exceeded 1.8 thus indicating that each class is properly separated from each other. On the other hand, the separability between the remaining LU/LC classes was inadequate especially between wheat and barssem. Thereafter, the MLC was applied and results revealed that the overall classification accuracy ranged between 70.9 and 72.6% for each of the studied images, which reflected the effect of the relatively low separability established between classes.

As the results were not satisfying, a different approach was developed for processing the data. This approach included the use of the principal component analysis (PCA) to reduce the redundancy in multispectral data and allow better discrimination between classes. The PCA of the four images revealed that the first principal component in all images contained about 66.5-76.5 % of the data variability, while the second encompassed about 18.5-29 % (Table 6). Therefore, both the first and second PCA of each of the four images was used in the classification and the rest of the PCAs, which covered about 3-5% of the variance, were discarded and omitted from the classification process. These two first PCAs of each of the four images were stacked into an eight bands image (hereafter referred to as PCA4*2 image) and classified using the MLC. The overall classification accuracy increased to reach 82.6% (Fig. 2).

Furthermore, another PCA was performed to further enhance the classification accuracy and utilized the effect of the different characteristics of the vegetation cover throughout the growing season. This PCA was applied on the nine images, which were developed from stacking the matching band of each of the four dates into one image. The first PCA account for about 62.4-79.0 % of the data variability, while the second covered only 12.4 to 21.3 % of the data variability (Table 7). Therefore, the first two PCAs of each image was used to develop a nine bands image and used in classification and the resulting image had an overall classification accuracy of 80.0% when using the MLC.

Moreover, this PCA image was stacked with the previously developed PCA image (PCA4*2) into one image and classified using the MLC. This image had an overall classification accuracy that reached 86.8% (Fig. 3). The classification results revealed that wheat and berssem cover most of the studied area (about 67%), followed by green beans,

which covered about 12% of the studied area. Guava and strawberry covered less than 1 % while potato and citrus covered 3.6 and 5.7 % of the studied area, respectively (Table 8).

Table 6: The eigenvectors of the covariance matrix of the four S2 images

Date	PCA1	PCA2	PCA 3-9
18/10/2020	76.4	18.5	5.1
27/12/2020	72.9	23.4	3.7
17/03/2021	70.0	25.9	4.1
06/04/2021	66.5	29.1	4.4

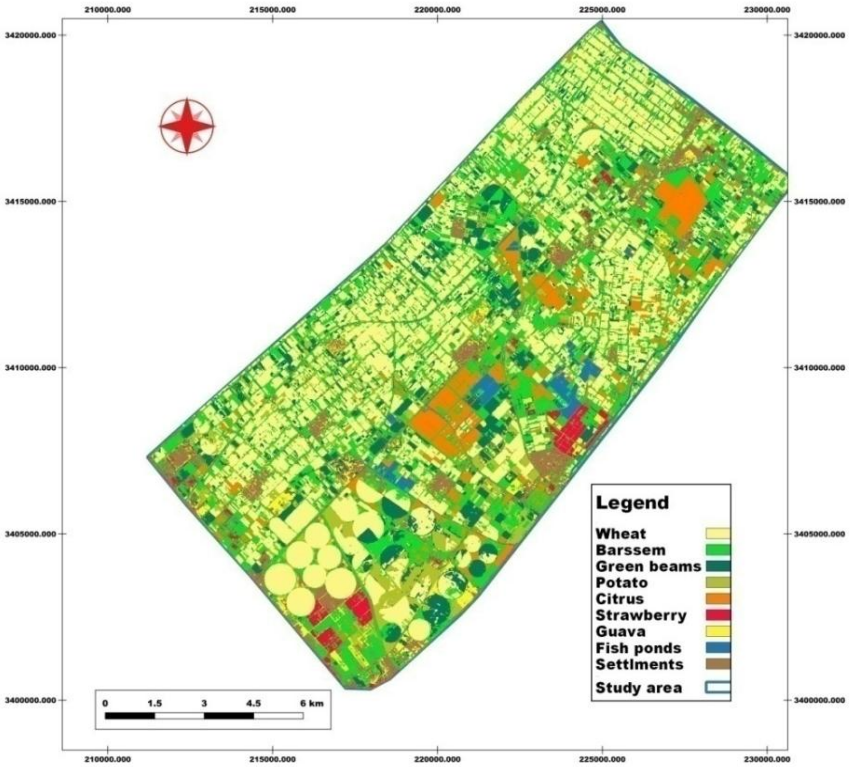

Fig. 2: Image classification of PCA4*2 image

Table 7: The eigenvectors of the covariance matrix of the nine images

Band\PCA	PCA1	PCA2	PCA 3-4
B2	79.0	12.4	8.6
B3	77.89.	13.7	8.5
B4	75.0	14.2	10.8
B5	76.6	14.3	9.1
B6	66.2	19.5	14.3
B7	62.4	21.3	16.3
B8	63.0	20.4	16.7
B11	73.4	17.1	9.5
B12	72.9	16.1	11.0

Table 8: Acreage of the LU/LC units

Class	Km²	%
Wheat	60.1	34.2
Barssem	57.3	32.6
Green beans	21.0	12.0
Potato	6.3	3.6
Citrus	10.0	5.7
Strawberry and Guava	1.7	0.9
Fish ponds	3.5	2.0
Settlements	15.9	9.0
Total	175.8	100.0

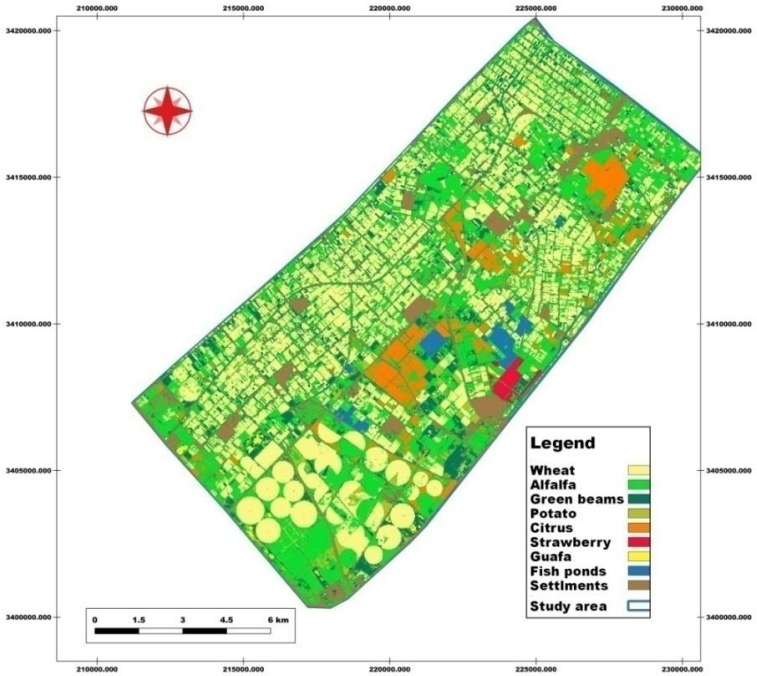

Fig. 3: Land use/ Land cover classification map of the study area

4. Conclusion

PCA has been widely used in remote sensing especially in LU/LC classification to increase the accuracy of classification. A PCA based classification technique has been developed in this research to classify the different LU/LC classes using sequential set of Sentinel-2 images covering the winter season of the studied area. The integration of various PCA types enhanced the accuracy of classification compared to the use of single images. The overall accuracy of classification increased from 70.9-72.6 % for single images to reach 86.8% utilizing the proposed technique. The developed PCA approach took advantage of the difference between the selected bands or wavelengths as well as the difference within each band throughout the growing season of each class.

5. Acknowledgment

The authors would like to express their deepest gratitude to the National Research Center of Egypt for funding this research, which was the extremely important for our research.

References

Abdu, H.A., 2019. Classification accuracy and trend assessments of land cover- land use changes from principal components of land satellite images. International Journal of Remote Sensing, 40(4):1275-1300. https://doi.org/10.1080/01431161.2018.1524587.

Abedini, M., M.A.M. Said and F. Ahmad, 2012. Clustering approach on land use land cover classification of Landsat TM over Ulu Kinta Catchment. World Applied Sciences Journal, 17 (7): 809-817.

Ahmad, A. and S. Quegan, 2012. Analysis of maximum likelihood classification on multispectral data. Applied Mathematical Sciences, 6 (129): 6425 – 6436.

Bhaskaran, S., S. Paramananda and Maria Ramnarayan, 2010. Per-Pixel and object-oriented classification methods for mapping urban features using Ikonos satellite data. Applied Geography, 30(4):650–665. https://doi.org/10.1016/j.apgeog.2010.01.009.

Bolstad, P.V. and T.M. Lillesand, 1991. Rapid maximum likelihood classification. Photogrammetric Engineering & Remote Sensing, 57(1): 67-74.

Dadon, A., M. Mandelmilch, E. Ben-Dor, and E. Sheffer, 2019. Sequential PCA-based classification of Mediterranean Forest Plants using airborne hyperspectral remote sensing. Remote Sens., 11, 2800. https://doi.org/10.3390/rs11232800 .

Deb, S.K. and R.K. Nathr, 2012. Land use/cover classification- An introduction review and comparison. Global Journal of Researches in Engineering, 12(1): 5-16.

Desai, V., 2014. Land use/Land cover classification of remote sensing data and their derived products in a heterogeneous landscape of a Khan-Kali Watershed, Gujarat. Asian Journal of Geo-informatics, 14:1-12.

Fonji, S.F. and G. Taff, 2014. Using satellite data to monitor land-use land-cover change in North-eastern Latvia. Springer Plus, 3:61-76. https://doi.org/10.1186/2193-1801-3-61.

Gonzales, R.C. and R.E. Woods, 2002. Digital image processing. Prentice Hall, second edition, 795 pages, ISBN 0-201-18075-8.

Jolliffe, I.T. and J. Cadima, 2016. Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A374:20150202. https://doi.org/10.1098/rsta.2015.0202.

Li, C., L. Zhou and W. Xu, 2021. Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China. Remote Sens., 13(8): 1595-1612. https://doi.org/10.3390/rs13081595.

Naik, G.R., 2018. Advances in Principal Component Analysis: Research and Development; Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-10-6704-4 .

Rwanga, Sophia S. and J.M. Ndambuki, 2017. Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. International Journal of Geosciences, 8: 611-622.

Santiesteban, I.A., 2003. Multi-sensor image fusion applied to morphodynamics of coastal environments in Belize. M.Sc. Thesis, International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede, The Netherlands.

Satir, O. and S. Berberoglu, 2012. Land use/cover classification techniques using optical remotely sensed data in landscape planning, landscape planning, Dr. Murat Ozyavuz (Ed.), Ch. 2, pp.21-54, InTech. https://www.intechopen.com/chapters/37553.

Tripathy, B.K., S. Sahu, and M.B.N.V. Prasad, 2013. PCA classification technique of remote sensing analysis of colour composite image of Chillika Lagoon, Odisha. International Journal of Advanced Research in Computer Science and Software Engineering, 3:1-6.

Vajsová, B. and P. Aastrand, 2015 New sensors benchmark report on Sentinel-2A; EUR27674EN. https://doi.org/10.2788/544302.

Zhu, X.F., Y. Pan, J.S. Zang, S. Wang, X.H. Gu, and C. Xu, 2007. The effects of training samples on the wheat planting area measure accuracy in TM scale (I): The accuracy response of different classifiers to training samples. Journal of Remote Sensing, 6:826-836.