DOI: http://dx.doi.org/10.26483/ijarcs.v16i1.7163

Volume 16, No. 1, January-February 2025

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Comparative Analysis of Land Use Classification Accuracy Using Maximum Likelihood Classification (MLC) and Spectral Angle Mapping (SAM) Methods

Ilada Aroonsri
Department of Business Digital, Faculty of Management Sciences, Valaya Alongkorn Rajabhat University Under the Royal Patronage
1, Moo 10, Klong 1, Klong Luang, Pathum Thani 13180,Thailand

Abstract: This study examines the effectiveness of Maximum Likelihood Classification (MLC) and Spectral Angle Mapping (SAM) methods in classifying land use types across a mixed-use landscape. Using an error matrix assessment based on 791 ground-reference points, the study evaluates the classification accuracy for different land use classes, including Paddy Fields, Water Bodies, Forests, Sugarcane, Cassava, Eucalyptus Plantations, and Built-up Areas. Results reveal that MLC achieved higher accuracy in classifying Paddy Fields and Sugarcane, demonstrating a strong overall accuracy of 83.944% with a Kappa coefficient of 0.808, indicating robust classification reliability. SAM, with an overall accuracy of 79.267% and a Kappa coefficient of 0.752, showed strengths in specific classes but struggled with spectral overlap in mixed or similar land cover types, particularly Built-up Areas and certain plantation classes. The study's findings suggest that MLC's probabilistic approach is more suitable for complex land cover patterns, whereas SAM performs best in distinguishing classes with distinct spectral properties. The comparative insights inform the selection of classification methods based on landscape characteristics, offering guidance for improved land use mapping accuracy in remote sensing applications.

Keywords: Land Use Classification; Remote Sensing; Maximum Likelihood Classification (MLC); Spectral Angle Mapping (SAM); Sentinel-2;

I. INTRODAUCTION

The study of land use and land cover (LULC) plays a crucial role in understanding the spatial dynamics of terrestrial ecosystems and human-environment interactions. By analyzing LULC, researchers and policymakers can assess environmental changes, monitor urban expansion, manage natural resources, and plan for sustainable development. Tracking these changes helps inform decision-making processes that affect biodiversity conservation, agriculture, water management, and climate adaptation efforts.

One key approach in LULC studies is the classification of land use types. Land use classification categorizes areas based on the observed human activities and natural characteristics, such as agricultural lands, forests, urban areas, and water bodies [1]. This structured approach enables a more systematic understanding of landscape patterns and their functions, assisting in resource allocation and ecological preservation [2]. Different classification schemes, including hierarchical and non-hierarchical systems, have been developed to accommodate various levels of detail and to address diverse research and policy needs.

In recent years, Borabue district in Maha Sarakham province has shown a high rate of expansion in land use and structural development. A significant factor in this growth is the presence of Highway No. 23, also known as Chaeng Sanit Road, which connects Ban Phai District in Khon Kaen Province through Borabue to Maha Sarakham City. This connectivity has led to increased density of residential areas, businesses, and retail establishments. Additionally, the area has notable agricultural land use, particularly in rice cultivation. These factors highlight the importance of studying and analyzing land use classification in this region.

With advancements in remote sensing technology, satellite imagery has become an indispensable tool for Land Use and Land Cover (LU/LC) classification, especially in rapidly developing areas like Borabue district. Remote sensing technology, when combined with satellite data, has revolutionized the study of natural resources, significantly reducing the need for extensive fieldwork and enabling efficient data collection over large areas [3]. This technology allows for the systematic analysis of different land use types by applying principles from various physics disciplines. Through the use of electromagnetic waves that interact with objects on Earth's surface and are captured by sensor systems, remote sensing provides detailed, temporal data [4]. By recording these interactions over multiple time intervals, remote sensing offers comprehensive insights into land use changes and trends [5],[6].

This study aims to utilize remote sensing in the classification of land use in Borabue district by analyzing Sentinel-2 satellite images with a resolution of 10 meters [7]. Through the use of the Semi-Automatic classification plugin, and suitable algorithms, the study seeks to understand current land use patterns in Borabue district and determine which algorithms are most effective for such classification [8],[9].

II. MATERIAL AND METHODOLOGY

A. Study Area

Borabue district is located in the southwestern part of Maha Sarakham province, approximately 26 kilometers from Maha Sarakham district and about 460 kilometers from Bangkok. The area consists of flat plains interspersed with low hills, sloping from west to east. The soil is primarily sandy loam, which has limited water retention capacity. The underlying geological structure is composed of Maha Sarakham rock formations and salt rock layers. The district has no major rivers, only small streams formed by rainwater runoff that flows down from the hilly areas into shallow creeks. Due to the topography, these streams have gradually become shallower over time, making it difficult to store water during the dry season. The primary land use in Borabue is agricultural, with most residents engaged in rice farming, crop cultivation, and livestock rearing [10].

B. *Data Collection for Training Land Use Classes*

The first step involves collecting representative training data for each land use class. Satellite imagery, such as Sentinel-2 or Landsat data, is selected and prepared to represent diverse land cover types present in the study area. Field data collection or visual interpretation from high-resolution images is then conducted to identify and label specific areas for training. This training dataset serves as a reference for distinguishing land use categories, such as built-up land (U), paddy field (A1), sugarcane (A203), casava (A204), eucalyptus (A3), forest (F) and water bodies (W), in the classification process.

C. *Maximum Likelihood Classification (MLC)*

The classifiers discussed rely on defining decision boundaries within a feature space by measuring the multispectral distances between training classes. The MLC method is probability-based: it assigns each pixel to the class for which it has the highest likelihood based on its features [11]. This probability is calculated across predefined classes, with the pixel being allocated to the class with the highest probability score. MLC remains a widely used supervised classification method [12-14], relying on Bayes' theorem (as shown in Equation 1) and assuming that training data statistics for each class are normally (Gaussian) distributed in each spectral band [15].

$$P(G_k/x) = \frac{P(x/G_k)P(G_k)}{P(x)} \qquad (1)$$

where

$P(G_k)$ is the prior probability of category k.

$P(x/G_k)$ is conditional probability of observing x from G_k.

$P(x)$ is the same for each category.

For data with multiple peaks (bi-modal or n-modal distributions), each peak may indicate distinct classes, which should ideally be trained separately to meet the Gaussian distribution assumption.

D. *Spectral Angle Mapping (SAM)*

SAM is an angle-based classification method that compares the spectral angle between a pixel's spectral signature and reference spectra. This technique is relatively unaffected by differences in illumination, making it particularly effective for distinguishing land cover types with similar reflectance properties. SAM works by measuring the angular difference between a test spectrum and a reference spectrum obtained from laboratory measurements, field data, or satellite observations. The algorithm assigns pixels to classes by evaluating the similarity between spectral vectors: a smaller angle indicates a closer match. In other words, the smaller the angle between two spectra, the higher their similarity; conversely, a larger angle suggests less similarity [16-20]. For example, the angle (α) between the test spectrum of category a and the reference spectrum of category b in a two-band image can be calculated as shown in Equation (2).

$$cos^{-1} = \left[\frac{\sum_{i-1}^{n} a_i b_i}{\left(\sum_{i-1}^{n} a_i^2\right)^{1/2}\left(\sum_{i-1}^{n} b_i^2\right)^{1/2}}\right] \qquad (2)$$

Where

n is number of bands

a_i is test spectrum

b_i is reference spectrum

E. *Accuracy Assessment with Error Matrix*

To assess classification accuracy, an error matrix (or confusion matrix) is constructed. This matrix compares the classified results with validation data or reference points, calculating overall accuracy, user's accuracy, producer's accuracy, and the Kappa coefficient. This step evaluates the reliability of each classification method in accurately representing the land use classes.

III. RESULTS

A. *Results of land use classification*

The land use classification results obtained from the Maximum Likelihood Classification (MLC) and Spectral Angle Mapping (SAM) methods revealed that the classified built-up areas (U), MLC classified 71.29 square kilometers, or 7,129 hectare, accounting for 10.19% of the total area, while SAM classified 39.77 square kilometers, or 3,977 hectare rai, representing 5.69% of the area. The classified water bodies (W), MLC classified 14.71 square kilometers, or 1,471 hectare, or 2.10% of the total area, whereas SAM classified 31.80 square kilometers, or 3,180 hectare, or 4.55% of the area. The classified of forests (F), MLC classified 35.58 square kilometers, or 3,558 hectare, which makes up 5.09% of the area, while SAM classified 20.95 square kilometers, or 2,095 hectare, representing 2.99% of the area. The classified of sugarcane plantations (A203), MLC classified 101.39 square kilometers, or 10,139 hectare, covering 14.49% of the area, compared to SAM's classification of 108.82 square kilometers, or 10,882 hectare, or 15.56% of the area. The classified of cassava fields (A204), MLC classified 133.26 square kilometers, or 13,326 hectare, amounting to 19.05% of the total area, whereas SAM classified 163.92 square kilometers, or 16,392 rai, making up 23.43% of the area. The classified of eucalyptus plantations (A3), MLC classified 68.80 square kilometers, or 6,880 hectare, which accounts for 9.83% of the area, while SAM classified 54.42 square kilometers, or 5,442 hectare, or 7.78% of the area. The classified of paddy fields (A1), MLC classified 274.52 square kilometers, or 27,452 hectare, covering 39.24% of the area, while SAM classified 279.85 square kilometers, or 27,985 hectare, representing 40.01% of the area. This comparison shows that the SAM method classifies water bodies, sugarcane, and cassava plantations with higher area percentages, whereas MLC identifies more extensive areas for built-up and eucalyptus plantations. Each classification technique provides a distinct perspective on land cover distribution in the study area. The results from both the MLC and SAM are presented in Tables 1 and 2, while the land use classification maps are shown in Figures 1 and 2, respectively.

Table 1: The areas and percentage values of land use classes from MLC classification

LU Class	Area		
	Sq.Km	Hectare	Percentage
Built-up (U)	71.29	7129	10.19
Water bodies (W)	14.71	1471	2.1
Forest (F)	35.58	3558	5.09
Sugarcane (A203)	101.39	10139	14.49
Casava (A204)	133.26	13326	19.05
Eucalyptus (A3)	68.8	6880	9.83
Paddy field (A1)	274.52	27452	39.24
Total	699.54	69954	100

Table 2: The areas and percentage values of land use classes from SAM classification

LU Class	Area		
	Sq.Km	Hectare	Percentage
Built-up (U)	39.77	3977	5.69
Water bodies (W)	31.8	318	4.55
Forest (F)	20.95	2095	2.99
Sugarcane (A203)	108.82	10882	15.56
Casava (A204)	163.92	16392	23.43
Eucalyptus (A3)	54.42	5442	7.78
Paddy field (A1)	279.85	27985	40.01
Total	699.54	69954	100

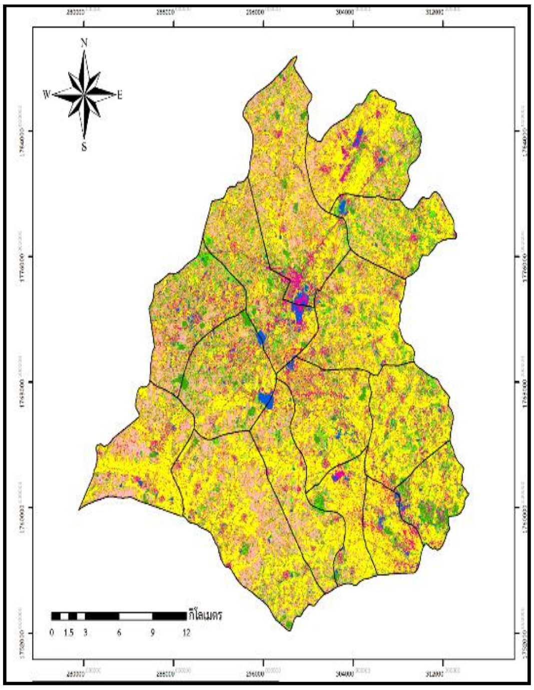

Figure 1: The land use classification maps from MLC

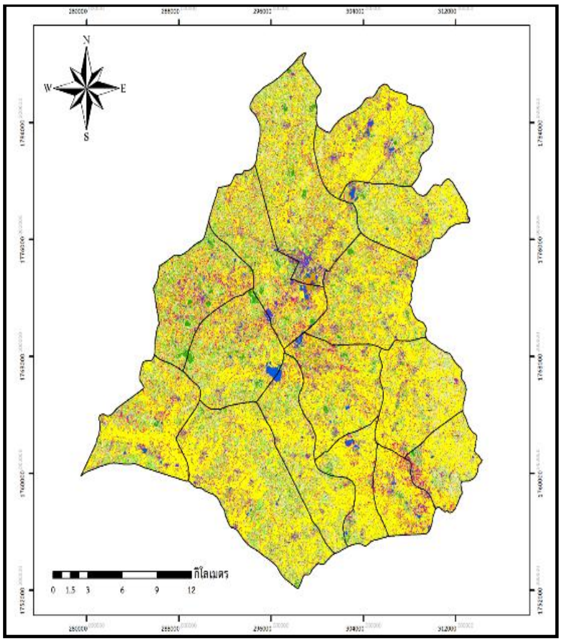

Figure 2: The land use classification maps from SAM

B. Accuracy Assessment with Error Matrix

The Error Matrix table is generated from an accuracy assessment, which overlays 791 ground-truth reference points collected from field surveys with the land classification results obtained from two classification algorithms: Maximum Likelihood Classification (MLC) and Spectral Angle Mapping (SAM). This table is used to evaluate both the accuracy and potential errors in classification results. Two key metrics are included in the assessment: Overall Accuracy, which represents the overall percentage of correctly classified points across all land use categories, and the Kappa Coefficient, a statistical measure of classification consistency or agreement. As shown in Tables 3 and 4, the results for MLC show an Overall Accuracy of 83.944% and a Kappa Coefficient of

0.808, indicating a high level of classification reliability. For SAM, the Overall Accuracy is 79.267%, with a Kappa Coefficient of 0.752, reflecting moderate agreement. These metrics allow for a quantitative comparison of the classification performance for each algorithm.

Table 3: Confusion matrices and kappa coefficient of land use classified from MLC

Class	Reference data							
	A1	A203	A204	A3	F2	U	W	Total
A1	162	0	3	0		20	0	185
A203	6	26	31	2	1	1	0	67
A204	1	17	115	0	0	0	0	133
A3	2	3	1	54	6	0	0	66
F2	0	0	5	5	82	0	0	92
U	13	0	1	1	1	118	5	139
W	0	0	0	0	0	2	107	109
Total	184	46	156	62	90	141	112	791

Table 4: Confusion matrices and kappa coefficient of land use classified from SAM

Class	Reference data							
	A1	A203	A204	A3	F2	U	W	Total
A1	159	1	4	0	0	39	1	204
A203	4	17	29	13	3	6	1	73
A204	4	13	74	21	28	2	3	145
A3	0	8	28	25	4	0	0	65
F2	0	5	18	3	55	0	0	81
U	9	1	1	0	0	46	8	65
W	8	1	2	0	0	48	99	158
Total	184	46	156	62	90	141	112	791

C. *Comparison of Classification Accuracy between MLC and SAM*

The comparative analysis between the MLC and SAM reveals notable differences in classification accuracy and error rates across land use types, as shown in the error matrix and associated kappa coefficients for each method.

High accuracy classes (paddy fields and sugarcane): In MLC, paddy fields (A1) and sugarcane (A203) are classified with a high degree of accuracy, similar to SAM. However, MLC provides a slightly higher accuracy for these classes, as reflected by the higher overall accuracy of 83.944% compared to SAM's 79.267%. Paddy fields particularly benefit from MLC's probabilistic approach, which utilizes both variance and covariance within each class, allowing it to effectively differentiate these large, uniform areas from other land use types. Sugarcane fields also exhibit high accuracy in MLC, with a lower misclassification rate compared to SAM, likely due to the method's ability to leverage the probabilistic distribution, thereby reducing spectral confusion.

Moderate accuracy classes (water bodies, deciduous forests, and cassava fields): water bodies (W), deciduous forests (F2), and cassava fields (A204) display moderate classification accuracy in both methods, but MLC generally shows higher classification consistency in these classes. MLC achieves this by better accounting for spectral variation within classes, minimizing misclassification with adjacent land cover types. SAM, however, occasionally misclassifies water bodies and deciduous forests due to the spectral similarity with vegetated and non-vegetated surfaces along water boundaries and forest edges. In SAM, cassava fields experience a moderate degree of confusion with other vegetation types, likely due to shared spectral properties. MLC, on the other hand, reduces this overlap by leveraging its probabilistic model, though some misclassification still occurs.

Low accuracy classes (built-up areas, sugarcane, and eucalyptus plantations): built-up areas (U) and eucalyptus plantations (A3) show lower classification accuracy in both methods, but MLC provides slightly better accuracy. For example, built-up areas are prone to misclassification in SAM due to spectral overlap with barren and vegetated surfaces, whereas MLC's probabilistic model allows for improved separation despite moderate overlap with other classes. Eucalyptus plantations face notable challenges in both methods due to spectral resemblance with other vegetation types. MLC slightly outperforms SAM here by offering marginally better separation, though SAM's angle-based classification approach struggles with mixed land covers where spectral angles overlap.

Overall accuracy and kappa coefficients: The overall accuracy of MLC (83.944%) surpasses that of SAM (79.267%), as MLC achieves a more refined classification in most land cover types, particularly those with spectral complexity or within mixed-use areas. This trend is supported by the kappa coefficient, which is higher for MLC (0.808) than SAM (0.752), indicating that MLC demonstrates stronger agreement between classified and reference data. The probabilistic approach of MLC tends to perform better in diverse land covers with mixed spectral signatures, while SAM's vector-based angle measurement has limited effectiveness when separating classes with subtle spectral differences.

IV. Discussion

The comparative analysis of land use classification results from MLC and SAM methods highlights the strengths and limitations of each approach across different land cover types. High accuracy in paddy fields and sugarcane: MLC demonstrated higher accuracy for paddy fields (A1) and sugarcane (A203) classifications due to its probabilistic model, which leverages class variance and covariance to handle similar spectral properties within these classes. SAM, although generally effective in distinguishing clear spectral patterns, yielded slightly lower accuracy in these areas, possibly due to its reliance on angular measurements, which can be limited when identifying classes with similar spectral characteristics. The distinction between these two methods indicates that MLC is better suited for classes with uniform land cover where spectral variance can be effectively managed.

The comparative analysis of land use classification results from MLC and SAM methods highlights the strengths and limitations of each approach across different land cover types. High accuracy in paddy fields and sugarcane: MLC demonstrated higher accuracy for paddy fields (A1) and sugarcane (A203) classifications due to its probabilistic model, which leverages class variance and covariance to handle similar spectral properties within these classes. SAM, although generally effective in distinguishing clear spectral patterns, yielded slightly lower accuracy in these areas, possibly due to its reliance on angular measurements, which can be limited when identifying classes with similar spectral

characteristics. The distinction between these two methods indicates that MLC is better suited for classes with uniform land cover where spectral variance can be effectively managed.

Low accuracy in built-up areas and eucalyptus plantations: built-up areas (U) and eucalyptus plantations (A3) experienced the lowest classification accuracy in both methods. SAM, in particular, struggled with Built-up Areas due to spectral overlap with barren and vegetated surfaces, leading to increased misclassification. The MLC approach, while slightly better, also faced challenges in these classes, as eucalyptus plantations often share spectral signatures with other vegetation types. This finding suggests that both methods have limitations in urban and plantation areas, where spectral similarities among mixed-use or single-crop plantations reduce classification reliability.

Overall accuracy and kappa coefficient: The overall accuracy of MLC (83.944%) exceeded that of SAM (79.267%), with MLC's higher kappa coefficient (0.808) indicating greater agreement between classified and reference data. This consistency implies that MLC is more robust for this dataset, particularly in areas where spectral variance is complex or mixed-use land covers create classification challenges. The SAM method, while effective in areas with unique spectral properties, had lower reliability for mixed land covers, as evidenced by its kappa coefficient (0.752).

The comparative results underscore the benefits of MLC in handling diverse land cover types due to its probabilistic approach, which enhances classification accuracy in spectrally complex regions. SAM, although a viable alternative in certain classes, is better suited to distinguish land cover types with clear and distinct spectral profiles. For future studies, combining MLC with ancillary data or other classifiers may enhance classification in challenging areas, such as urban zones and specific plantation types, to address spectral overlap and improve overall accuracy.

V. Acknowledgment

This research project supported by Valaya Alongkorn Rajabhat University Under the Royal Patronage.

VI. References

[1] Aroonsri, I., & Sangpradid, S. (2021). Orchard classification from satellite image data using fuzzy k-mean to estimate of carbon sequestration. Ecology Environment and Conservation, 27(3),1432-1438.

[2] Jensen, J.R. (2005). Introductory digital image processing: a remote sensing perspective. Pearson Prentice Hall.

[3] Sangpradid, S., Uttaruk, Y., Rotjanakusol, T., & Laosuwan, T. (2021). Forecasting time series change of the average enhanced vegetation index to monitoring drought condition by using TERRA/MODIS data. Agriculture & Forestry, 67(4),115-129.

[4] Lillesand, T., Kiefer, R.W., & Chipman, J. (2015). Remote sensing and image interpretation. John Wiley & Sons.

[5] Aroonsri, I., Sangpradid, S. (2021b). Artificial neural networks for the classification of shrimp farm from satellite imagery. Geographia Technica, 16(2), 149-159.

[6] Sangpradid, S. (2018). Change vector analysis using integrated vegetation indices for land cover change detection. International Journal of Geoinformatics, 14 (4), 71-77.

[7] ESA Vision . Sentinel-2's high-resolution multispectral instrument is based on well-established heritage from France's SPOT missions and the US Landsat satellites. https://www.esa.int/Applications/Observing_the_Earth /Copernicus/Sentinel-2/Instrument.

[8] Congedo, L. (2016). Semi-Automatic classification plugin documentation. GitHub.

[9] Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment, 120, 25-36.

[10] Department of Provincial Administration. Five-Year strategic development plan for Borabue district (2018-2022).https://dopacrr.dopa.go.th/plan_develop/app/main/upload/2564/25644406_1.pdf.

[11] Atkinson, P.M., & Lewis, P. (2000). Geostatistical classification for remote sensing: an introduction. Computers & Geosciences, 26(4), 361-371.

[12] Yonezawa, C. (2007). Maximum likelihood classification combined with spectral angle mapper algorithm for high resolution satellite imagery. International Journal of Remote Sensing, 28, 3729 - 3737.

[13] Safaralizade, E., Husseinzade, R., Pashazade, G., & Khosravi, B. (2014). Assessing the accuracy of the pixel-based algorithms in classifying the urban land use, using the multi spectral image of the IKONOS satellite (Case Study, Uromia City). International Letters of Natural Sciences, 11, 40-56.

[14] Dibs, H., Hasab, H., Al-Rifaie, J., & Al - Ansari, N. (2020). An optimal approach for land-use / land-cover mapping by integration and fusion of multispectral landsat OLI images: Case study in Baghdad, Iraq. Water Air and Soil Pollution, 231(9, 231.

[15] Sun, J., Yang, J., Zhang, C., Yun, W., & Qu, J. (2013). Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method. Mathematical and Computer Modelling, 58, 573-581.

[16] Rashmi, S., Addamani, S., Venkat & Ravikiran, S. (2014). Spectral angle mapper algorithm for remote sensing image classification. International Journal of Innovative Science, Engineering & Technology. 1(4), 201-205.

[17] Chakravarty, S.S., Mishra, R., Paikaray, B.k., & Dash, S. (2021). Hyperspectral image classification using spectral angle mapper. 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 87-90.

[18] De Carvalho, O. A. & Meneses, P. R. (2000). Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM). in Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18, (JPL publication Pasadena, CA, USA, 2000), 2.

[19] Yousefi, S., Mirzaee, S., Tazeh, M., Pourghasemi, H., & Karimi, H. (2015). Comparison of different algorithms

for land use mapping in dry climate using satellite images: a case study of the central regions of Iran. Desert, 20, 1-10.

[20] Mann, D., & Joshi, P. (2017). Evaluation of image classification algorithms on Hyperion and ASTER data for land cover classification. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 87, 855-865.