Landcover class extraction from Prisma hyperspectral data using Fuzzy Machine Learning Techniques

Harshaditya Gaur [1,2], Saloni Avaghade [1], Vinay Kumar [1], Devendra Bhavsar [2]

[1] Photogrammetry and Remote Sensing Department, Indian Institute of Remote Sensing - ISRO, Dehradun, Uttarakhand - 248001, India – gaurharsh24@gmail.com, saloniavaghade1999@gmail.com, vinayiirs@gmail.com
[2] Computer Science Department, J K Lakshmipat University, Jaipur, Rajasthan – 302026, India – devendrabhavsar@jklu.edu.in

Keywords: Landcover, Prisma, Hyperspectral, Fuzzy, Machine Learning, Panchromatic, Gram-Schmidt, Remote Sensing

Abstract

Hyperspectral remote sensing enables highly detailed spectral discrimination between landcover types. However, conventional classification methods often struggle with spectral variability, mixed pixels, heterogeneous landscapes or when limited training data is provided. This study presents a fuzzy machine learning framework for extracting landcover classes from Prisma hyperspectral data, using the strength of soft classification to address classification related challenges. The spaceborne hyperspectral dataset, with its rich spectral resolution spanning across VNIR to SWIR region, serves as an ideal dataset for detecting subtle spectral differences between surface materials/features. By implementing fuzzy machine learning approach, the study moves beyond the conventional binary/hard classification, enabling the identification of partial class memberships and improving the mapping accuracy in heterogeneous and spectrally overlapping landcover types. After pre-processing and MNF dimensionality reduction, different fuzzy techniques (Fuzzy C-Means, Possibilistic C-Means and Modified Possibilistic C-Means) were applied to different vegetation types, built-up area and water body. The model was trained with limited ground truth, and results show that fuzzy techniques achieve higher class precision and spatial consistency particularly in urban and vegetation-based landscape. The study highlights the potential of fuzzy based machine learning as a robust soft classification technique for hyperspectral data achieving a classification accuracy of 87.01% (MPCM-HSI) and 92.38% (MPCM-HSI+PAN). This demonstrates, how the spectral resolution of Prisma hyperspectral sensor can be efficiently used for enhanced landcover mapping in real-world situations where only partial class knowledge is known.

1. Introduction

In recent years, advancements in remote sensing technology have significantly increased the use of hyperspectral imaging, which offers highly detailed spectral discrimination among different land cover classes. Unlike multispectral sensors, which capture non-continuous and discrete spectral bands, hyper-spectral sensors operate within continuous spectral intervals (Li et al., 2021). This capability allows for the identification of subtle, often overlapping, spectral features that are critical for precise land cover classification. However, conventional classification models used in hyperspectral data analysis often encounter challenges such as spectral variability, mixed pixels, and heterogeneous landscapes can complicate the classification process. Additionally, the effectiveness of these models decreases when limited training data is available, emphasising the necessity of improved methodologies for interpreting high-dimensional hyperspectral datasets with minimal training data (Delogu et al., 2023). As remote sensing technology continues to evolve, innovative computational frameworks are essential to address these challenges, improving the reliability and accuracy of hyperspectral data classification and providing scientific insights to support decision-making (Akar & Tunc Gormus, 2021). To overcome these limitations, this study integrates Prisma hyperspectral data with fuzzy-based soft classification techniques. Fuzzy logic offers a flexible and uncertainty-tolerant framework particularly well-suited for hyperspectral datasets (Huo et al., 2018). It allows for partial class membership which accommodates the inherent ambiguities associated with spectrally mixed pixels. This an approach is advantageous in complex environments, where class boundaries are not well-defined, and material signatures overlap. The study utilises the Prisma hyperspectral dataset, which is characterized by its high spectral resolution of 12nm, with the wavelength region from 400 to 2505nm across 293 spectral bands, in-combination with fuzzy based machine learning technique (soft classification) which offers efficiently handles the complexities of hyperspectral data, leading to more reliable environmental monitoring and assessment outcomes. The fuzzy machine learning approach effectively addresses the integral challenges related with landcover classification by utilizing the advantages of soft classification techniques (Vishnoi et al., 2022). Different fuzzy-based techniques such as Fuzzy C-Means, Possibilistic C-Means, and Modified Possibilistic C-Means were applied to classify different landcover classes such as cropland, urban areas, water, riverbeds, forests, and scrublands. To further test the capabilities of soft classification techniques, the hyperspectral dataset was pansharpened using the 5-meter panchromatic band. This increases the spatial resolution resulting in improved classification accuracy, especially in complex and fragmented environments where class separability can be become challenging due to coarse resolution (Hadi Mahdipour et al., 2024). Improving the spatial features of the data enhances class separability and contributes to higher classification accuracy and more precise mapping outputs. This study demonstrates the capabilities of the Prisma spaceborne hyperspectral sensor for landcover classification through comparative analysis of different fuzzy-based soft classifiers.

2. Study Area

The study focuses on the city of Rishikesh and the surrounding reserved forest regions in the state of Uttarakhand, India. The area extends geographically from 30.0° to 30.2° N latitude and 78.2° to 78.4° E longitude. The area lies at the base of the Himalayan foothills and the upper Ganga basin. The terrain includes flat river plains, gently rising slopes, and moderate elevation changes. The landcover in the area comprises a mix of urban infrastructure, agricultural fields, riverbeds, and dense forest zones. The location experiences seasonal climate variation, affecting vegetation patterns and surface conditions throughout the year. Overall, this area is ideal for evaluating land cover classification techniques using hyperspectral data.

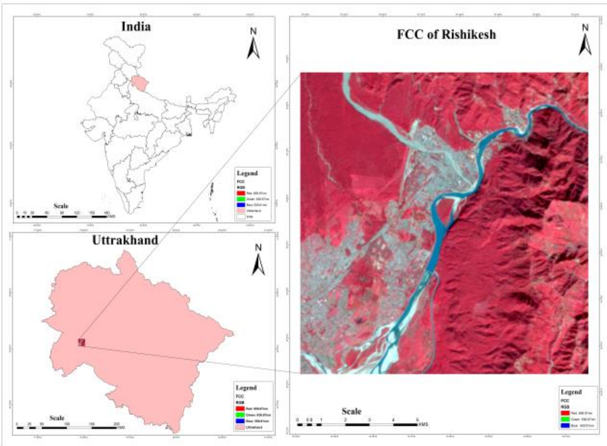


Figure 1. Study area map of Rishikesh City, Uttarakhand.

3. Dataset and Sensor Specification

The study utilizes spaceborne hyperspectral and panchromatic datasets acquired by the Prisma sensor operated by the Italian Space Agency (ASI). The dataset covers the latitudinal extent of 30.0° to 30.2°N and longitudinal extent of 78.2° to 78.4°E. It was acquired on 20th November 2022 and with a spatial resolution of 30m for hyperspectral and 5m for panchromatic. The hyperspectral dataset spans across the visible to shortwave infrared range (400–2500nm) with a spectral interval of 12 nm, across 239 contiguous bands (66 visible in near-infrared bands from 400 to 1010nm and 173 shortwave infrared bands from 920 to 2505nm) and 1 panchromatic band (400-700nm). Sensor operates in a sun-synchronous orbit at an altitude of 615km with a 29-day revisit and swath width of 30km (E. Vangi et al. 2021)

PRISMA Sensor Specification	
Spatial Resolution	30 meters (Hyperspectral), 5 meters (Panchromatic)
Wavelength Range	400 to 2505 nm
Spectral Resolution	12 nm
Spectral Bands	240 bands (66 VNIR, 173 SWIR, 1 PAN)
Orbital Altitude	615 km
Revisit Cycle	29 days
Calibration	Radiometric and Geometric
Data Level	L2D, Surface Reflectance
Sensor Swath	30km (nadir)

Table 1. PRISMA Spaceborne Sensor Specification

4. Methodology

The study used atmospherically corrected PRISMA L2D surface reflectance data of the city of Rishikesh, Uttarakhand. For data pre-processing, the noisy bands removal (from 1-9, 99-108, 143-161, 163-167, 211-230 based on standard deviation metrics) which were affected by atmospheric absorption and sensor error were removed, while the remaining bands were used for further analysis. MNF transformation was performed to reduce the data dimensionality and improve the signal-to-noise (SNR). Following the MNF transformation, the first 16 components were used for data analysis using different variation of Fuzzy Machine Learning techniques such as Fuzzy C-Means (FCM), Possibilistic C-Means (PCM) and Modified Possibilistic C-Means (MPCM). Fuzzy-based algorithms are particularly effective in processing hyperspectral data, where pixel-level spectra often represent mixed land cover types due to coarse spatial resolution and spectral similarity. Unlike traditional hard classifiers, fuzzy models assign degrees of membership to each class, allowing for more refine interpretation of sub-pixel heterogeneity and spectral ambiguity. However, by using the possibility theory-based models (PCM and MPCM) which are robust to outliers and effectively manage uncertainty and imprecision. A pansharpened hyperspectral imagery was also processed using the same fuzzy machine learning techniques to assess the impact of enhanced spatial resolution on classification performance. Classification results obtained from the pan-sharpened data were quantitatively compared with the original MNF-reduced dataset to evaluate spatial coherence, and sub-pixel mapping accuracy at optimized fuzzifier value. Accuracy assessment and comparative output assessment of all the classification techniques in terms of fuzzy error matrix (User Accuracy and Producer Accuracy and Overall Accuracy) was conducted to determine the best approach for landcover class extraction using spaceborne PRISMA hyperspectral data.

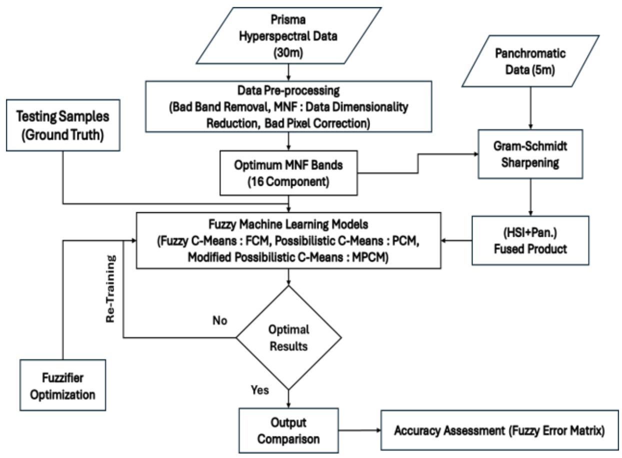


Figure 2. Methodology flowchart of Fuzzy Machine Learning Techniques.

4.1 Gram-Schmidt Pan-Sharpening

Gram-Schmidt (GS) pan-sharpening technique is an image fusion approach that enhances the spatial resolution of multispectral and hyperspectral imagery by integrating it with a high-resolution panchromatic (PAN) band. The algorithm generates a synthetic low-resolution panchromatic band from the input multi-band dataset using a weighted linear combination. This synthetic data serves as the first component in the Gram-Schmidt orthogonalization process. Then, the original multi-band dataset is the transformed into an uncorrelated set of components through the GS transformation. The original high-resolution panchromatic image then replaces the first component, effectively fusing the spatial detail. An inverse transformation is applied to reconstruct sharpened multi-band data now have the enhanced spatial detail while preserving the original spectral integrity (Laben et al., 2000).

4.2 MNF Transformation

MNF (Minimum Noise Fraction) transformation is a two-step orthogonal linear transformation technique that maximizes the signal-to-noise ratio (SNR) of remote sensing data, especially of hyperspectral imagery (HSI). Conceptually, it is very similar to principal component analysis (PCA), but unlike PCA which emphasizes on variance. MNF focuses on separating noise from signal in high-dimensional data by transforming it into low-dimensional data while preserving information by eliminating random noise and reliably estimating the data from the signal using co-variance matrix – eigen statistics (Green et al., 1988).

4.3 Fuzzy C-Means (FCM)

Fuzzy C-Means is an unsupervised clustering algorithm that assigns membership values to data points relative to each cluster center. Unlike hard clustering, the FCM allows each point to belong to multiple clusters with varying degrees of membership.

Given a dataset $X = \{x_1, x_2, \ldots, x_N\} \in R^{NxD}$, the objective function minimized in FCM (Bezdek, 1988) is:

$$J_{FCM} = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad (1)$$

where u_{ij} = degree of membership of point x_i in class j
$m > 1$ = fuzzifier parameter (smoothness)
c_j = centroid of cluster j
N = number of data points
C = number of clusters

4.4 Possibilistic C-Means (PCM)

Possibilistic C-Means (PCM) algorithm is an extension of the traditional FCM technique. It overcomes the limitations of FCM by introducing a possibility-based model that relaxes the probabilistic membership constraint when classifying high dimensional and noisy datasets. In PCM the sum of membership values for a pixel across all classes is constrained to 1, it introduces possibilistic memberships that are independent across the classes, enabling greater flexibility in modelling uncertainty and outliers (Krishnapuram and Keller, 1996).

Given a dataset $X = \{x_1, x_2, \ldots, x_N\} \in R^{NxD}$, the objective function minimized in PCM is:

$$J_{PCM} = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 + \eta_j \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left(1 - u_{ij}^{m}\right) \quad (2)$$

where $u_{ij} \in [0,1]$ = degree of membership of pixel x_i to class j
c_j = centroid of cluster j
$m > 1$ = degree of membership smoothness
η_j = scale parameter that adjusts cluster influence
C = number of clusters
n = number of data points

4.5 Modified Possibilistic C-Means (MPCM)

Modified Possibilistic C-Means (MPCM) algorithm is a modified version of the PCM, developed to address the issues in PCM such as coincident clusters and membership degeneracy. MPCM introduces a hybrid objective function that integrates both fuzzy and possibilistic components, balancing data partitions and noise tolerance efficiently (Saad & Alimi, 2009).

Given a dataset $X = \{x_1, x_2, \ldots, x_N\} \in R^{NxD}$, the objective function minimized in MPCM is:

$$J_{MPCM} = \sum_{j=1}^{C}\sum_{i=1}^{N}\left[\alpha u_{ij}^{m} + (1-\alpha) t_{ij}^{m}\right] \left\| x_i - c_j \right\|^2 + \sum_{j=1}^{C} \eta_j \sum_{i=1}^{N} \left(1 - t_{ij}\right)^{m} \quad (3)$$

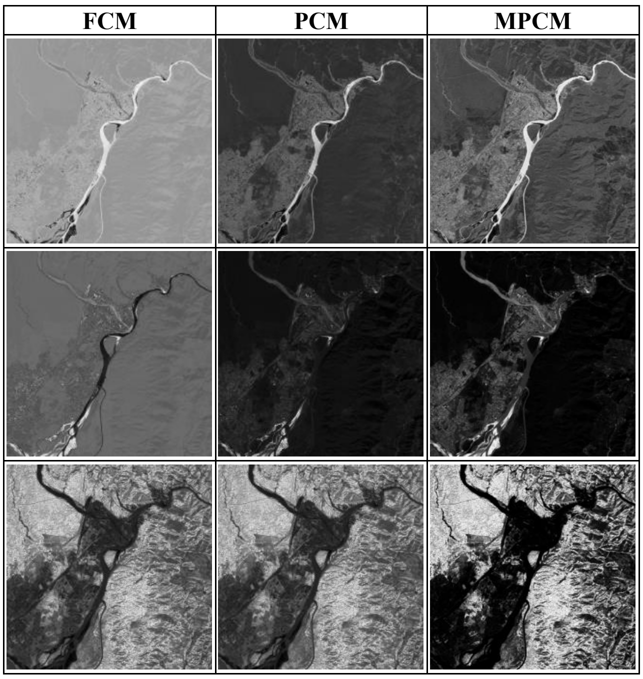
where $u_{ij} \in [0,1]$ = degree of membership of pixel x_i to class j
c_j = centroid of cluster j
$m > 1$ = degree of membership smoothness
η_j = scale parameter that adjusts cluster influence
C = number of clusters
n = number of data points

4.6 Fuzzy Error Matrix (FERM)

Fuzzy Error Matrix is a generalized extension of conventional confusion matrix, to evaluate the performance of soft classification algorithms, where each pixel can belong to multiple classes with varying degrees of membership. Unlike hard classification, where each pixel is assigned exclusively to a single class, fuzzy classification assigns continuous membership values to all classes, leading to a more nuanced representation of class uncertainty, especially in high dimensional datasets like multispectral and hyperspectral (Binaghi et al., 1999).

5. Results and Discussion

The classified outputs of all the fuzzy machine learning models (FCM, PCM, and MPCM), demonstrate clear differences across the landcover classes: Water, Riverbed, Forest, Agriculture, Urban, and Scrubland (Figure 3.). The output from FCM shows that while the majority of river was correctly classified as Water class, some smaller tributaries were partial detected. A similar pattern was observed for the riverbed class, where partial misclassification was evident. Vegetation-related classes such as forest, agriculture, and scrubland exhibited poor class separability. This is primarily due to their similar spectral characteristics, which the FCM failed to resolve effectively, despite the continuous band information from hyperspectral data. The urban class showed better classification results than the vegetation classes. However, some misclassification with the riverbed class was still present, due to their similar reflectance profiles in certain regions. In comparison, the PCM



This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-189-2025 |

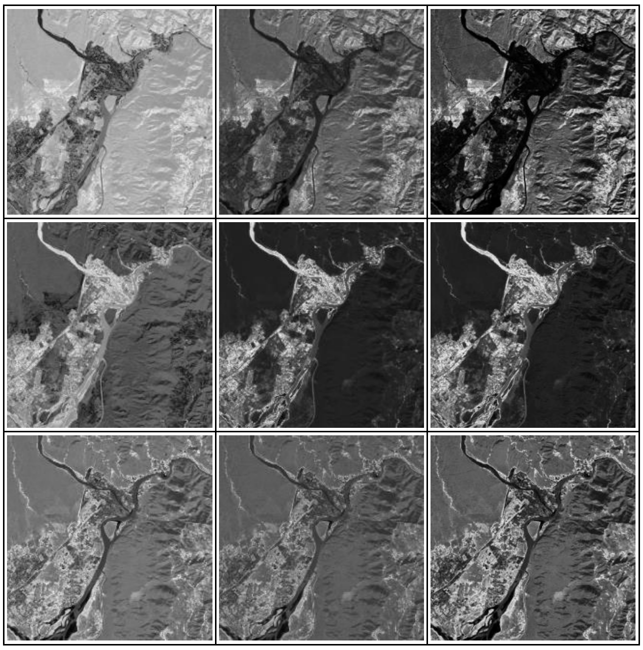


Figure 3. Classified output (1) Water, (2) Riverbed, (3) Forest, (4) Agriculture, (5) Urban and (6) Scrubland

performed better than FCM across most of the classes. Probabilistic constraints allowed the PCM to better handle uncertainty, improving classification in mixed and edge regions. Improvements were particularly noticeable for riverbed, water, agriculture, and urban class. However, the outputs for forest and scrubland classes were very similar to FCM outputs, with only minor improvements in class delineation. This indicates that while PCM improves robustness to ambiguity, but it still struggles with spectrally overlapping vegetation classes. MPCM results showed a further improvement in classification, particularly for the spectrally similar vegetation classes. The algorithm demonstrated better separation between the forest and agriculture classes, indicating its ability to utilize both fuzzy and probabilistic constraints effectively. The urban and riverbed classes also exhibited improved spatial coherence and boundary definition compared to previous methods. The classification improvement observed in MPCM can be attributed to its combined objective function, which addresses the overlapping and coincident clusters issues in PCM. However, some misclassifications were still observed in the scrubland class, where the results closely resembled those from FCM and PCM. This may be due to intra-class variability and limited spectral distinctiveness of scrub vegetation. Overall, MPCM technique provided the most balanced and accurate classification, particularly in complex regions with spectral similarity and mixed landcover. The comparison of results across all the models highlights the advantage of integrating both fuzzy and probabilistic constraint for enhance data classification accuracy.

5.1 Comparison of Hyperspectral v/s Panchromatic Fused Hyperspectral classified using MPCM Technique

The classification results obtained from the original Prisma hyperspectral imagery using the Modified Possibilistic C-Means (MPCM) algorithm were compared with those from the pansharpened hyperspectral dataset, which was also classified using MPCM. The comparative analysis underlines the importance of spatial information in enhancing classification performance. In the pansharpened dataset, the spatial correlation among neighbouring pixels was effectively utilized, resulting in

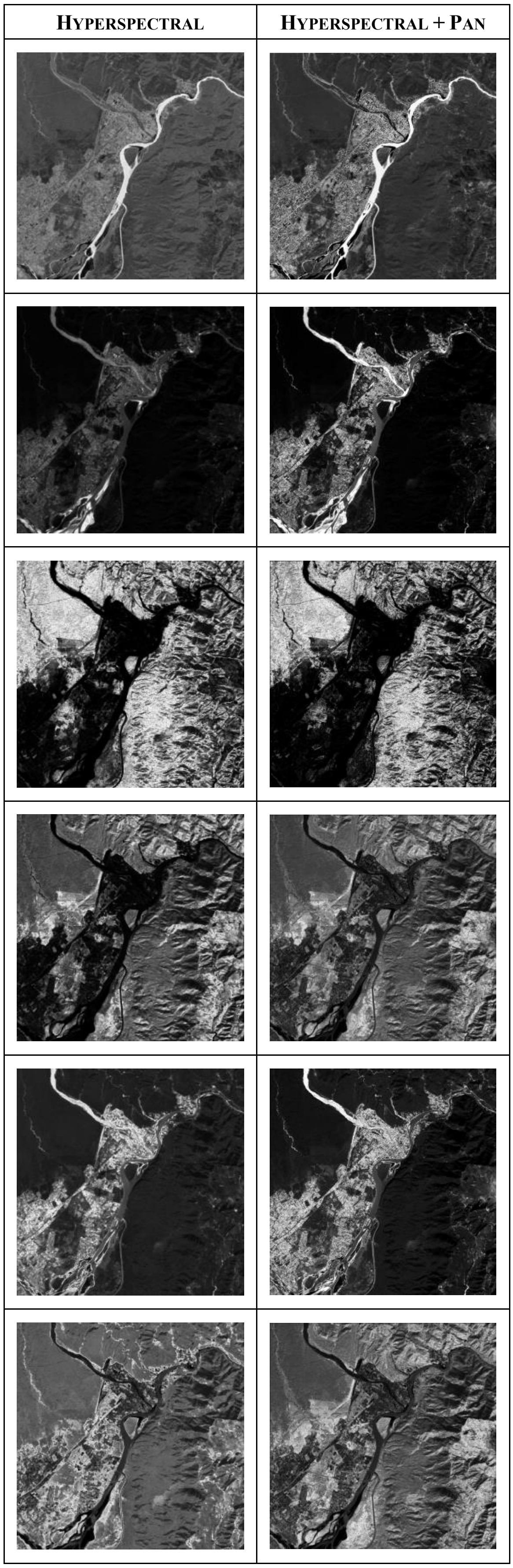


Figure 4. Classified output (1) Water, (2) Riverbed, (3) Forest, (4) Agriculture, (5) Urban and (6) Scrubland

the improved classification. This approach not only helped reduced the noise but also enhanced the delineation between spectrally similar classes. The noticeable improvement was observed in the riverbed class, which previously showed partial classification but was more accurately mapped post-pansharpening. Similarly, class separation for water and urban also had improvements. While the forest and agriculture class results were similar to those of the original HSI dataset, the scrubland class showed significant improvement with clearer spatial boundaries. Overall accuracy improved from 87.01% (HSI) to 92.38% (pansharpened HSI), which validates the critical role of spatial regularization/constraint. These results highlight the importance of incorporating spatial information for reliable and accurate hyperspectral classification and analysis.

	Fuzzy C-Means	**Possibilistic C-Means**	**Modified Possibilistic C-Means**
Hyperspectral (User Accuracy)	78.53	84.27	88.42
Hyperspectral (Producer Accuracy)	73.96	82.58	85.61
Overall Accuracy	76.24	81.42	87.01

Table 2. Accuracy Assessment via Fuzzy Error Matrix (FERM)

	Hyperspectral	**Pansharpened Hyperspectral**
Modified Possibilistic C-Means	87.01	92.38

Table 3. Overall Accuracy of Pansharpened (HSI) v/s Hyperspectral

6. Conclusion

This study demonstrated the applicability of soft classification for landcover mapping using Prisma hyperspectral imagery. The effectiveness of these models was evaluated using the Fuzzy Error Matrix (FERM), which highlighted their ability to manage the inherent uncertainty and mixed pixel effects of hyperspectral dataset. The classification results indicated that both FCM and PCM encountered limitations, particularly due to the coarse spatial resolution of the hyperspectral sensor. These limitations resulted as misclassification in spectrally overlapping classes and regions with fine-scale spatial variation. However, MPCM demonstrated improved performance by addressing the shortcomings of FCM and PCM, resulting in better separation of spectrally similar landcover classes. While, spectral information was effectively utilized, the absence of spatial context within the models remains a constraint. However, when the MPCM technique was applied to pansharpened dataset, the classification accuracy was further improved. The enhanced spatial resolution of the fused dataset provided sharper class boundaries and reduced confusion in spectrally similar region. This highlights the importance of spatial information in hyperspectral data classification. While spectral richness is critical, incorporating spatial constraints or context-aware regularization into fuzzy classification frameworks will further improve the classification accuracy and reliability, particularly in heterogeneous and spatially complex mapping environments.

7. Future Scope

To extend the usability and accessibility of fuzzy classification techniques for hyperspectral data analysis, a Python-based graphical user interface is being developed. This will streamline the classification, and accuracy assessment steps, enabling users to interact with the models without requiring programming expertise. The modular nature of tool will allow the integration of additional fuzzy-based machine learning algorithms such as kernel-based clustering, neural-based clustering and hybrid fuzzy techniques. Support for adjusting clustering parameters, incorporating spatial-constraints, and real-time visualization of classification outputs would further increase the tool's efficacy.

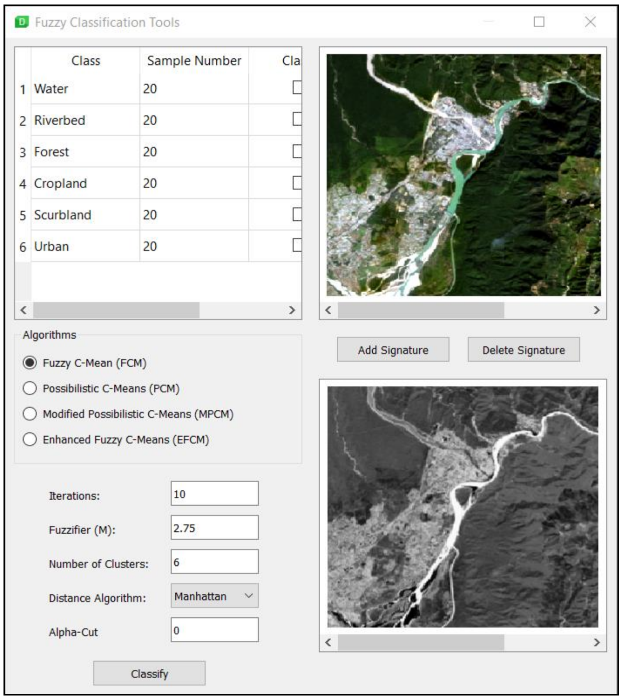


Figure 5. Python QT5 based GUI of Fuzzy Classification Tool.

Acknowledgements

The authors acknowledge the Italian Space Agency (ASI) for providing the access to Prisma spaceborne imagery utilized in this research. Access to this dataset was instrumental in achieving the analysis and outcomes presented in this study. The authors express their gratitude to ASI for supporting scientific research by providing access to earth observation data.

References

Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing, 26(1), 65–74. https://doi.org/10.1109/36.3001

Elisabetta Binaghi, Brivio, P. A., Ghezzi, P. P., & Rampini, A. (1999). Assessing the accuracy of soft thematic maps using fuzzy set-based error matrices. Proceedings of SPIE, the

International Society for Optical Engineering/Proceedings of SPIE, 3871, 188–194. https://doi.org/10.1117/12.373256

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3), 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Krishnapuram, R., & Keller, J. M. (1996). The possibilistic C-means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems, 4(3), 385–393. https://doi.org/10.1109/91.531779

Saad, M. F. & Alimi, A. M. (2009). Modified fuzzy possibilistic c-means. In *Proc. of the Int. Multiconference of Engineers and Computer Scientists*, 1: 18-20.

Laben, Craig A., and Bernard V. Brower. Process for Enhancing the Spatial Resolution of Multispectral Imagery using Pan-Sharpening. US Patent 6,011,875, filed April 29, 1998, and issued January 4, 2000.

Vangi, E., D'Amico, G., Francini, S., Giannetti, F., Lasserre, B., Marchetti, M., & Gherardo Chirici. (2021). The New Hyperspectral Satellite PRISMA: Imagery for Forest Types Discrimination. Sensors, 21(4), 1182–1182. https://doi.org/10.3390/s21041182

Delogu, G., Caputi, E., Perretta, M., Maria Nicolina Ripa, & Boccia, L. (2023). Using PRISMA Hyperspectral Data for Land Cover Classification with Artificial Intelligence Support. Sustainability, 15(18), 13786–13786. https://doi.org/10.3390/su151813786

Huo, H., Guo, J., & Li, Z.-L. (2018). Hyperspectral Image Classification for Land Cover Based on an Improved Interval Type-II Fuzzy C-Means Approach. Sensors, 18(2), 363. https://doi.org/10.3390/s18020363

Hadi Mahdipour, Sharifi, A., Mehdi Sookhak, & Carlos Rubio Medrano. (2024). Ultra-fusion: optimal fuzzy fusion in land-cover segmentation using multiple panchromatic satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1–12. https://doi.org/10.1109/jstars.2024.3360648

Vishnoi, S., Sisodia, A., Mehrotra, T., Khenwar, M., Upadhyay, S., & Agrawal, S. (2022). Land Cover Classification of Satellite Imagery using Fuzzy Hybrid Approaches by incorporating Contextual Information. Computer Integrated Manufacturing Systems, 28(12), 1572-1587.

Akar, O., & Tunc Gormus, E. (2021). Land use/land cover mapping from airborne hyperspectral images with machine learning algorithms and contextual information. Geocarto International, 1–28. https://doi.org/10.1080/10106049.2021.1945149

Li, K., Xu, J., Zhao, T., & Liu, Z. (2021). A fuzzy spectral clustering algorithm for hyperspectral image classification. IET Image Processing, 15(12), 2810–2817. https://doi.org/10.1049/ipr2.12266

S. Manju, & K. Helenprabha. (2022). Sub-band Selection-Based Dimensionality Reduction Approach for Remote Sensing Hyperspectral Images. Lecture Notes in Networks and Systems, 167–178. https://doi.org/10.1007/978-981-16-7018-3_14