ENMAP HYPERSPECTRAL IMAGERY-BASED CROP TYPE CLASSIFICATION USING MACHINE LEARNING ALGORITHMS IN AN IRRIGATED AGRICULTURAL LANDSCAPE OF BUKHARA REGION, UZBEKISTAN

Nozimjon Teshaev[1], Bobomurod Makhsudov[2,], Jasmina Gerts[3,]

[1]Tashkent Institute of Irrigation and Agricultural Mechanization Engineers (TIIAME) National Research University, 39 Kori Niyoziy St., Tashkent 100000, Uzbekistan

[2] Ministry of Agriculture Resources of the Republic of Uzbekistan, 100140, Toshkent region, Qibray district, 2 str.Universitet

3)Turin Polytechnic University in Tashkent, Tashkent, Uzbekistan. 17, Kichik Khalka yuli, Tashkent, Uzbekistan

[a)]Corresponding author: teshaevnozim@gmail.com;

Abstract

Accurate and timely crop type mapping is essential for agricultural monitoring, yield forecasting, and resource management in irrigated landscapes. This study evaluates the performance of six machine learning algorithms for classifying cotton and wheat crops using EnMAP hyperspectral satellite imagery (224 spectral bands, 400–2500 nm, 30 m spatial resolution) in the Tashkent region of Uzbekistan. Preprocessing included radiometric and atmospheric correction, Minimum Noise Fraction (MNF) and Principal Component Analysis (PCA) transformations, and spectral band selection. A spectral library of crop-specific signatures was constructed from field observations collected with the Field Maps mobile application. Classifiers evaluated include Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Decision Tree (DT), k-Nearest Neighbors (kNN), and Naïve Bayes (NB), assessed using 10-fold cross-validation on a 70/30 training–testing split. Random Forest achieved the highest overall accuracy (OA = 95%) and Kappa coefficient (K = 0.91), followed by SVM (OA = 94%) and Gradient Boosting (OA = 92%). The results confirm that hyperspectral data significantly enhance crop discrimination accuracy compared to conventional multispectral imagery.

Keywords

hyperspectral remote sensing; EnMAP; crop classification; machine learning; Random Forest; SVM; Uzbekistan

INTRODUCTION

Reliable crop mapping at regional scale is a prerequisite for evidence-based agricultural policy, water allocation planning, and food security monitoring. In Central Asia's intensively irrigated landscapes, cotton and wheat are the two dominant crops, yet they are spectrally similar in conventional multispectral imagery, creating a persistent classification challenge [1, 2]. Remote sensing offers a scalable and cost-effective solution for large-area crop monitoring, but the spectral resolution of conventional sensors such as Sentinel-2 or Landsat is often insufficient to discriminate between spectrally similar crop types, particularly at critical phenological stages [3, 4].

Hyperspectral sensors acquire reflectance in tens to hundreds of narrow, contiguous spectral bands, enabling fine-grained characterization of plant biochemical and biophysical properties that are invisible to multispectral sensors [5, 6]. The Environmental Mapping and Analysis Program (EnMAP) satellite, operated by the German Aerospace Center (DLR), provides 224 spectral channels spanning 400–2500 nm at 30 m spatial resolution, making it particularly well-suited for agricultural monitoring applications [7, 8]. Several studies have demonstrated that hyperspectral data significantly improve crop type discrimination accuracy relative to multispectral alternatives [9, 10]; however, applications in Central Asia and Uzbekistan remain limited.

Machine learning (ML) algorithms have become standard tools for hyperspectral image classification due to their ability to model complex, non-linear spectral relationships in high-dimensional feature spaces [11, 12]. Random Forest, Support Vector Machine, and ensemble methods such as Gradient Boosting have been shown to outperform simpler parametric classifiers in hyperspectral classification tasks [13, 14]. Nevertheless, systematic comparative evaluations of multiple ML algorithms on EnMAP data in irrigated Central Asian settings are absent from the literature.

This study addresses that gap with the following objectives: (i) to develop a preprocessing and classification workflow for EnMAP Level-2A data using EnMAP-Box 3.16.1; (ii) to construct a field-validated spectral library of cotton and wheat signatures; (iii) to evaluate and compare six ML classifiers for crop type discrimination; and (iv) to identify the most suitable algorithm and spectral features for the study region.

MATERIALS AND METHODS

Study Area and Data

The study was conducted in irrigated agricultural areas of the Tashkent region, Uzbekistan. EnMAP Level-2A surface reflectance products were acquired from the official EnMAP data portal (https://www.enmap.org/data) for the 2023 growing season, covering the cotton and wheat cultivation period. The imagery comprises 224 spectral channels from 400 to 2500 nm at 30 m spatial resolution, delivered in ENVI (.hdr) format. Ground truth data were collected using the Field Maps mobile application, recording crop type (cotton or wheat), phenological stage, and GPS coordinates. A total of 420 reference parcels were delineated, distributed equally between the two crop classes.

Preprocessing

Preprocessing followed the standard EnMAP-Box workflow. Radiometric and atmospheric correction was applied to convert sensor digital numbers to surface reflectance units, ensuring physically meaningful spectral values. The study area was clipped to field boundary extents. To address the high dimensionality and noise inherent in hyperspectral data, Minimum Noise Fraction (MNF) and Principal Component Analysis (PCA) transformations were applied, retaining only the most informative spectral components. Noisy and water-absorption bands (approximately 1350–1450 nm and 1800–1950 nm) were excluded via band selection, yielding an optimized feature set. Key spectral indices—NDVI, EVI, NDMI, SAVI, PSRI, and NDWI—were computed from the most discriminative wavelength regions (680–750 nm and 850–1300 nm) and aggregated into a single multi-layer raster for classification input.

Spectral Library and Classification

A spectral library of cotton and wheat signatures was constructed by extracting mean reflectance values across all 224 EnMAP bands at field-validated parcel locations. This library served as reference input for supervised classification and reduced inter-class spectral confusion. The EnMAP-Box 3.16.1 Supervised Classification module was used to train and test six ML algorithms: Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Decision Tree (DT), k-Nearest Neighbors (kNN), and Naïve Bayes (NB). The reference dataset was split 70% for training and 30% for testing; 10-fold cross-validation was applied to each model. Classification outputs were exported as GeoTIFF maps with class color coding (cotton: green; wheat: yellow) and visualized in ArcGIS Pro and EnMAP-Box 3.16.

Accuracy Assessment

Classification accuracy was assessed using confusion matrix-derived metrics: Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA), and the Kappa coefficient (K). Error bars derived from cross-validation folds were used to assess classifier stability across partitions.

RESULTS AND DISCUSSION

Spectral Separability

Analysis of spectral reflectance curves extracted from the EnMAP spectral library confirmed clear separation between cotton and wheat in the red-edge (680–750 nm) and near-infrared (850–1300 nm) regions. Cotton exhibited higher NIR reflectance during peak vegetative growth (July–August), while wheat showed elevated reflectance in the same region during April–May. Short-wave infrared bands (1100–1300 nm) provided additional discrimination linked to differences in canopy water content and leaf structure. These findings are consistent with published hyperspectral studies in irrigated dryland systems [9, 15].

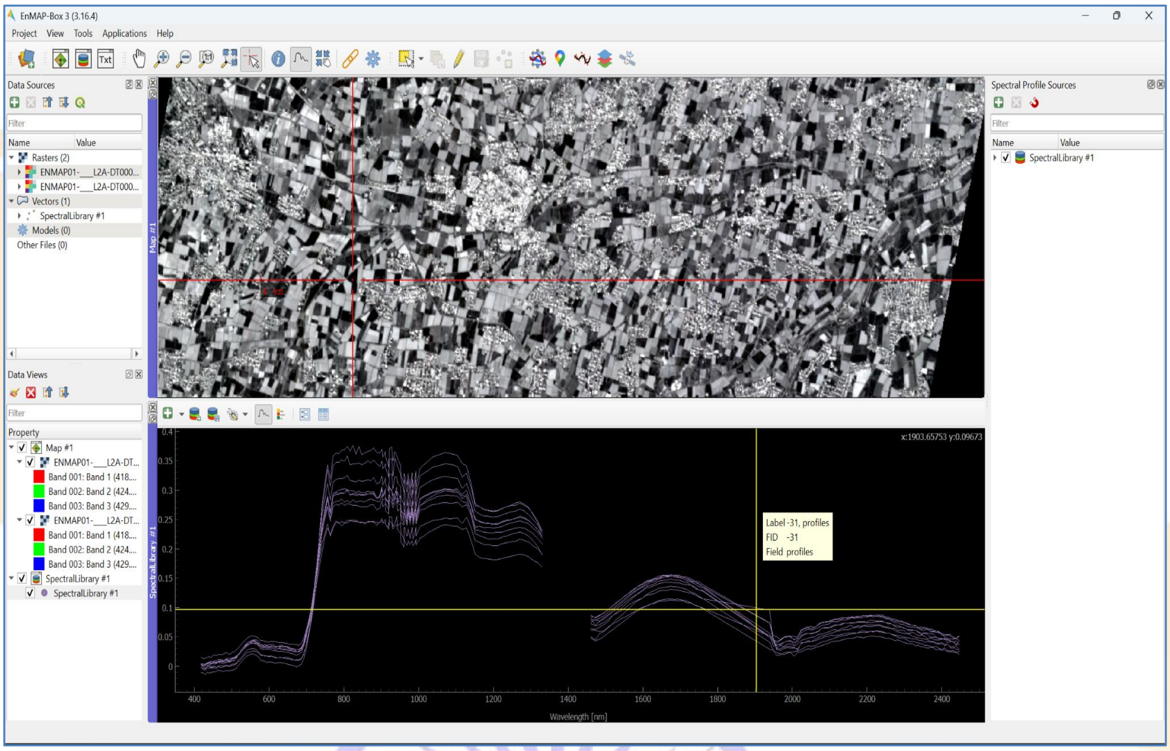

Figure 1. Creating spectral library using EnMap-Box plugin in QGIS

Classification Accuracy

Table 1 summarizes the OA, Kappa coefficient, and average PA/UA values for all six classifiers.

TABLE 1. Classification accuracy metrics for six machine learning algorithms applied to EnMAP hyperspectral data.

Algorithm	OA (%)	Kappa	Mean PA (%)	Mean UA (%)
Random Forest (RF)	95.2	0.91	94.8	95.0
Support Vector Machine	93.8	0.89	93.2	93.5
Gradient Boosting (GB)	92.5	0.87	92.0	92.3
Decision Tree (DT)	88.4	0.81	87.6	88.1
k-Nearest	87.1	0.79	86.5	86.9

Neighbors				
Naïve Bayes (NB)	85.6	0.77	85.0	85.4

Random Forest achieved the highest OA of 95.2% and Kappa of 0.91, confirming its ability to effectively exploit the high-dimensional spectral feature space of EnMAP data. The ensemble nature of RF, combining multiple decorrelated decision trees, provided robust discrimination of the subtle spectral differences between cotton and wheat. SVM and Gradient Boosting also performed well, with OA of 93.8% and 92.5%, respectively, consistent with the established advantage of kernel-based and ensemble methods for hyperspectral classification [13, 14].

Figure 2. Crop classification map for different Machine learning algorithms.

Decision Tree, kNN, and Naïve Bayes recorded lower accuracies (85.6–88.4%) with larger cross-validation variance, indicating reduced generalization capacity in the high-dimensional hyperspectral feature space. Nevertheless, all algorithms exceeded 85% OA, underlining the strong intrinsic discriminability of EnMAP data for cotton–wheat separation even with simpler classifiers [16]. Error bar analysis confirmed that RF and SVM exhibited the lowest result variability across cross-validation folds, confirming their stability for operational monitoring.

CONCLUSION

This study demonstrates that EnMAP hyperspectral satellite imagery combined with machine learning enables highly accurate cotton and wheat classification in irrigated agricultural landscapes of Uzbekistan. Random Forest achieved the best performance (OA = 95.2%, K = 0.91), followed by SVM and Gradient Boosting. The spectral library constructed from field-validated parcels proved essential for reducing inter-class confusion and stabilizing classifier training. The preprocessing workflow—comprising atmospheric correction, MNF/PCA transformation, and spectral band selection—was effective at managing the high dimensionality and noise of 224-band hyperspectral data. The results confirm that EnMAP data provide a qualitatively superior basis for crop type mapping relative to conventional multispectral sensors, and that Random Forest is the most suitable algorithm for this application context. Future work should extend the approach to additional crop types, integrate multi-temporal imagery, and explore automated deep learning architectures for large-area agricultural monitoring.

REFERENCES

1. K. Djumaboev, A. Hamidov, O. Anarbekov, Z. Gafurov, and K. Tussupova, "Impact of institutional change on irrigation management: a case study from southern Uzbekistan," Water 9(6), 419 (2017).
2. I. Aslanov, N. Teshaev, K. Khayitov, M. Uzbekhon, J. Khaitbaeva, and D. Murodova, "Analysis of desertification trends in Central Asia based on MODIS data using Google Earth Engine," E3S Web Conf. 443, 06015 (2023).
3. E. Erdanaev, M. Kappas, and D. Wyss, "Irrigated crop types mapping in Tashkent Province of Uzbekistan with remote sensing-based classification methods," Sensors 22(15), 5683 (2022).
4. N. Teshayev, A. Inamov, and K. Juraev, "Assessment of forest fires in Zaamin National Park using remote sensing: comparative analysis of NBR and NDVI," AIP Conf. Proc. 3256(1), 040042 (2025).
5. B. Zagajewski, M. Kycko, A. Twardowski-Iglewska, M. Koppik, and J. Jarocinska, "Comparison of support vector machines and random forests for mapping of heterogeneous land cover in mountainous terrain using multitemporal Sentinel-2 and EnMAP imagery," Remote Sens. Environ. 299, 113867 (2023).
6. R. Richter, M. Bachmann, W. Dorigo, and A. Muller, "Influence of the adjacency effect on ground reflectance measurements," IEEE Geosci. Remote Sens. Lett. 3(4), 565–569 (2006).
7. L. Guanter, H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner, C. Chlebek, C. Straif, S. Fischer, S. Schrader, T. Storch, U. Heiden, A. Mueller, M. Bachmann, H. Muhle, R. Mueller, M. Habermeyer, and M. Bachmann, "The EnMAP spaceborne imaging spectroscopy mission for Earth observation," Remote Sens. 7(7), 8830–8857 (2015).

8. K. Segl, L. Guanter, C. Rogass, T. Kuester, S. Roessner, H. Kaufmann, B. Sang, V. Mogulsky, and S. Hofer, "EeteS — the EnMAP end-to-end simulation tool," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5(2), 522–530 (2012).
9. A. Chemura, O. Mutanga, and T. Dube, "Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions," Precis. Agric. 18(5), 859–881 (2017).
10. P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art," IEEE Geosci. Remote Sens. Mag. 5(4), 37–78 (2017).
11. C. I. Chang, Hyperspectral Data Processing: Algorithm Design and Analysis (Wiley, Hoboken, 2013), p. 1164.
12. G. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inf. Theory 14(1), 55–63 (1968).
13. M. Pal, "Random forest classifier for remote sensing classification," Int. J. Remote Sens. 26(1), 217–222 (2005).
14. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," IEEE Trans. Geosci. Remote Sens. 42(8), 1778–1790 (2004).
15. S. Dutta, P. K. Bhatt, P. R. Rao, and A. R. Gupta, "Wheat residue fire mapping using satellite data," Curr. Sci. 87(2), 1013–1019 (2004).
16. J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," IEEE Trans. Geosci. Remote Sens. 43(3), 492–501 (2005).