January 2026
Volume 2
Issue 1

DOI: 10.48308/IJCE.2026.244728.1024

Interdisciplinary Journal of Civil Engineering
Shahid Beheshti University

Machine Learning-Based Land Cover Mapping for Environmental and Urban Planning Applications: A Cloud Computing Framework Using Multi-Source Geospatial Data

Authors:

Alireza Habibi Razi[1], Zahra Azizi*[,2], Hossein Aghamohammadi[3], Ali Asghar Alesheikh[4], Alireza Gharagozlou[2]

Abstract

Accurate land cover information is essential for environmental management, urban planning, and sustainable development. This study presents a geospatial data fusion framework for land cover mapping using Google Earth Engine (GEE). The framework integrates Landsat 9 imagery, five spectral indices (NDVI, NDWI, NDBI, BSI, and SAVI), and ALOS Digital Surface Model (DSM) data. The methodology was applied in Hamedan Province, Iran, to classify four land cover classes: water bodies, vegetation, urban areas, and bare lands. Five supervised machine learning algorithms—Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Tree (CART), Gradient Tree Boosting (GTB), and Minimum Distance (MD)—were evaluated. Classification performance was assessed using overall accuracy, producer accuracy, user accuracy, and the Kappa coefficient. Results showed that integrating spectral and topographic features improved class separability and classification accuracy. RF achieved the best performance with an overall accuracy of 98% and a Kappa coefficient of 0.97, followed by GTB. In contrast, MD produced lower accuracy and was more affected by spectral confusion. The findings demonstrate the effectiveness of cloud-based machine learning and multi-source geospatial data fusion for accurate land cover mapping and support applications in environmental monitoring, resource management, and spatial planning

Keywords: Land Cover Mapping, Geospatial Data Fusion, Machine Learnining; SVM, Landsat 9

1. Department of Remote Sensing & GIS, SR. C., Islamic Azad University, Tehran, Iran

2. Department of Geoinformation and Geomatics Engineering, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran

3. Department of Civil Engineering, SR. C., Islamic Azad University, Tehran, Iran

4. Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

*Corresponding Author: zah_azizi@sbu.ac.ir

1. Introduction

Accurate and up-to-date land cover information is a fundamental requirement for environmental management, sustainable urban development, infrastructure planning, natural resource assessment, and spatial decision-making. Land surface monitoring has emerged as a critical scientific discipline that investigates the dynamics of terrestrial environments through the analysis of remotely sensed observations, including satellite imagery and aerial photographs (Borgogno-Mondino et al., 2018). Advances in Earth observation technologies have significantly expanded the availability of geospatial information, enabling continuous monitoring of environmental processes across multiple spatial and temporal scales (Nelson et al., 2022; Ouchra et al., 2022a).

Remote sensing applications have become increasingly important in a wide range of disciplines, including urban planning, agriculture, environmental monitoring, water resource management, disaster assessment, and mineral exploration (Nelson et al., 2022; Saraei et al., 2023; Ouchra et al., 2022b; Ouchra et al., 2022c). The ability of satellite sensors to provide repetitive, large-scale, and cost-effective observations has transformed the way land surface dynamics are analyzed. Consequently, land cover mapping has become one of the most active research areas in geospatial science and environmental engineering (Ouchra et al., 2022d; Ouchra et al., 2022e; Ouchra & Belangour, 2021; Hu et al., 2018; Xiao et al., 2021; Qiu et al., 2025). Despite substantial progress in image analysis techniques, accurate discrimination of land cover classes remains a challenging task because different surface features frequently exhibit similar spectral and textural characteristics. Furthermore, heterogeneous landscapes often contain complex spatial patterns where vegetation, urban areas, bare lands, and water bodies overlap spectrally, increasing classification uncertainty (Ouchra et al., 2022d; Ouchra et al., 2022e; Ouchra & Belangour, 2021; Hu et al., 2018).

In recent years, machine learning (ML) techniques have demonstrated remarkable success in managing large and high-dimensional datasets across various scientific and engineering domains (Torki, 2025). Compared with traditional statistical approaches, machine learning algorithms are capable of modeling nonlinear relationships and extracting complex patterns from remotely sensed data, thereby improving classification performance and robustness (Ouchra et al., 2022e; Xiao et al., 2021; Qiu et al., 2025). The increasing availability of satellite imagery, together with advances in computational capabilities, has accelerated the adoption of machine learning methods for automated land cover mapping and environmental monitoring applications (Dhingra & Kumar, 2019; Rabbi et al., 2020; Usama et al., 2019).

The production of accurate land cover maps generally involves two major stages: feature extraction and classification. Feature extraction aims to identify the most informative characteristics from remotely sensed imagery while reducing data redundancy and computational complexity. Once relevant features have been extracted, machine learning

classifiers can be employed to assign each pixel to a predefined land cover category. Among the most widely used supervised learning algorithms are Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Tree (CART), Gradient Tree Boosting (GTB), and other ensemble-based approaches that have shown promising performance in remote sensing applications (Xiao et al., 2021; Qiu et al., 2025).

Supervised classification techniques rely on representative training samples to learn the spectral and spatial characteristics of different land cover classes (Dhingra & Kumar, 2019; Rabbi et al., 2020; Usama et al., 2019). In contrast, unsupervised methods group pixels according to inherent similarities without requiring prior knowledge of class labels. Numerous studies have demonstrated that supervised machine learning algorithms generally outperform traditional classification methods, particularly when high-quality training data and complementary geospatial information are available (Basheer et al., 2022; Palanisamy et al., 2023; Yang et al., 2021; Wahbi et al., 2023).

Recent developments in Earth observation systems have further enhanced land cover mapping capabilities. Landsat 9, the latest mission in the Landsat program, provides improved radiometric performance and data continuity for environmental monitoring applications (NASA, 2023a; NASA, 2023b). However, relying solely on multispectral imagery may not be sufficient for distinguishing spectrally similar land cover classes. To overcome this limitation, researchers increasingly integrate spectral indices and auxiliary geospatial datasets into classification workflows (Feng et al., 2022; Palanisamy et al., 2023; Capolupo & Tarantino, 2023). Spectral indicators such as the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Difference Built-up Index (NDBI), Bare Soil Index (BSI), and Soil-Adjusted Vegetation Index (SAVI) enhance the representation of vegetation, water bodies, urban areas, and bare lands, thereby improving class separability and classification accuracy (Rouse et al., 1974; McFeeters, 1996; Zha et al., 2003; Diek et al., 2017; Huete, 1988).

In addition to spectral information, topographic variables derived from Digital Surface Models (DSMs) provide valuable complementary information for land cover characterization. Elevation-related features can reduce spectral confusion and improve the discrimination of heterogeneous land surface conditions. Consequently, the integration of multisource geospatial data has emerged as an effective strategy for enhancing classification performance and producing more reliable land cover products (Feng et al., 2022; Palanisamy et al., 2023; Ouchra et al., 2023).

Another major advancement in geospatial analysis is the emergence of cloud-based geospatial computing platforms. Google Earth Engine (GEE) has revolutionized remote sensing research by providing direct access to extensive satellite archives and large-scale computational resources through a cloud-based environment (Chen et al., 2024; Praticò et al., 2021; Yang et al., 2022). The platform enables efficient processing of massive geospatial datasets without the

need for local storage or high-performance computing infrastructure (Tamiminia et al., 2020). As a result, GEE has become one of the most widely used platforms for environmental monitoring, land cover mapping, climate studies, flood assessment, and natural resource management (Chen et al., 2024; Yang et al., 2021; Johary et al., 2023; Kazemi Garajeh et al., 2024).

Although numerous studies have investigated machine learning-based land cover classification using remotely sensed imagery (Basheer et al., 2022; Palanisamy et al., 2023; Ouchra et al., 2023), several research gaps remain. First, many studies focus on evaluating individual classifiers without comprehensively investigating the combined contribution of spectral indices and topographic information within a unified analytical framework. Second, comparative assessments of multiple machine learning algorithms using Landsat 9 imagery remain relatively limited, particularly in semi-arid environments characterized by diverse land cover patterns. Third, the potential of cloud-based geospatial data fusion approaches for supporting environmental management and sustainable urban planning requires further investigation.

To address these challenges, this study develops a cloud-based geospatial data fusion framework for land cover mapping in Hamedan Province, Iran. Landsat 9 multispectral imagery, spectral indices (NDVI, NDWI, NDBI, BSI, and SAVI), and ALOS Digital Surface Model (DSM) data were integrated within the Google Earth Engine environment. Five supervised classification algorithms, including Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Tree (CART), Minimum Distance (MD), and Gradient Tree Boosting (GTB), were evaluated for the classification of four major land cover classes, namely water bodies, vegetation, urban areas, and bare lands.

The main contributions of this study are threefold: (i) the development of a multi-source geospatial data fusion workflow for land cover mapping, (ii) a comprehensive comparison of five machine learning classifiers within a cloud-computing environment, and (iii) the identification of an effective framework for supporting environmental monitoring, spatial planning, and sustainable land management applications. The findings provide valuable insights into the role of machine learning and cloud-based geospatial analytics in improving land cover classification accuracy and supporting evidence-based decision-making.

2. Materials and Methods

2.1. Study Area

The proposed framework was implemented in Hamedan Province, western Iran, which is characterized by diverse topographic and land cover conditions. The province is located between 47°34′E and 49°36′E longitude and 32°59′N and 35°48′N latitude, covering an area of approximately 19,590 km². Hamedan includes mountainous regions associated with the Zagros Mountain system, agricultural lands, urban settlements, water bodies, and extensive

bare land areas, making it an appropriate case study for evaluating land cover classification algorithms under heterogeneous environmental conditions (Figure 1).

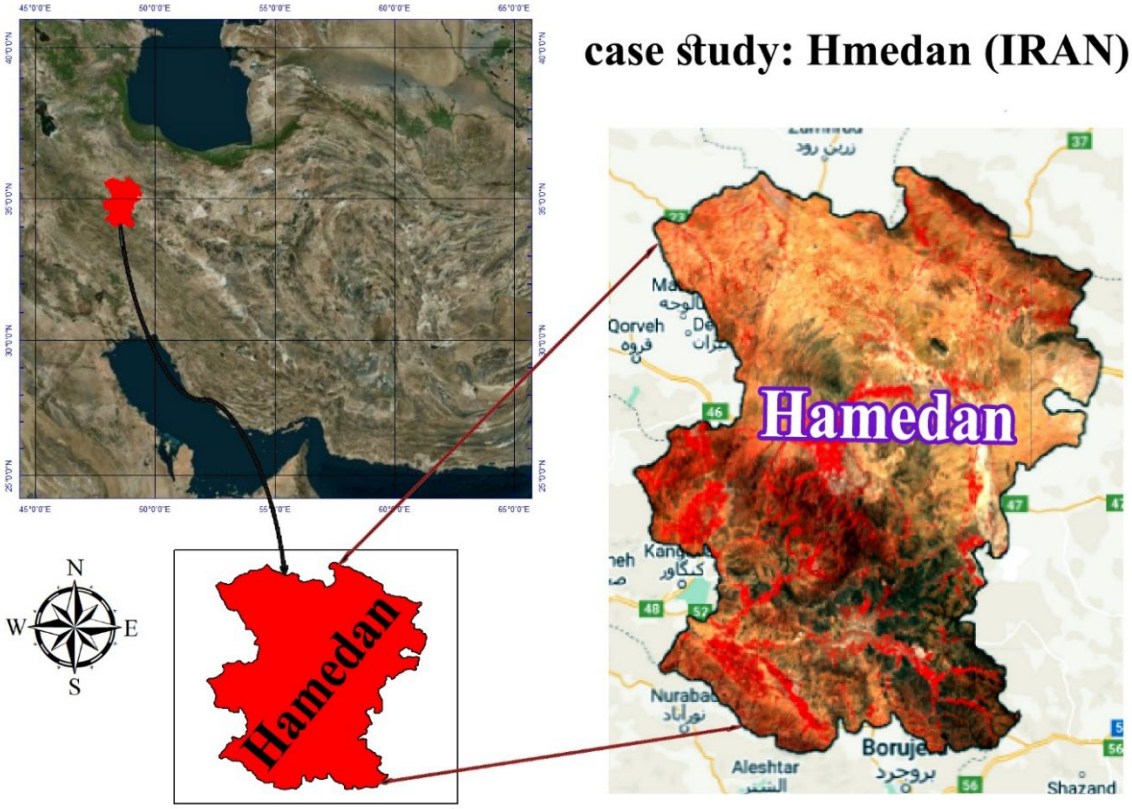

Figure 1. Geographic location of the study area (Hamedan Province, Iran).

2.2. Data Sources

Three categories of geospatial data were utilized in this study:

1. Landsat 9 Operational Land Imager (OLI) surface reflectance imagery;
2. Spectral indices derived from Landsat 9 bands;
3. ALOS Digital Surface Model (DSM) data.

Landsat 9 imagery was selected because of its improved radiometric quality, 30-m spatial resolution, and suitability for environmental monitoring applications (NASA, 2023a; NASA, 2023b; Masoumi et al., 2017; Capolupo & Tarantino, 2023). Surface reflectance products available through Google Earth Engine (GEE) were employed to minimize atmospheric effects and ensure consistency among observations.

To improve class separability, five spectral indices were generated:

- NDVI (Normalized Difference Vegetation Index) (Rouse et al., 1974)

- NDWI (Normalized Difference Water Index) (McFeeters, 1996)
- NDBI (Normalized Difference Built-up Index) (Zha et al., 2003)
- BSI (Bare Soil Index) (Diek et al., 2017)
- SAVI (Soil Adjusted Vegetation Index) (Huete, 1988)

Additionally, elevation information derived from ALOS-DSM was incorporated as an auxiliary variable to enhance discrimination among spectrally similar land cover classes.

2.3. Methodological Framework

The overall methodological workflow adopted in this study is illustrated in Figure 2.

The workflow consists of five sequential stages:

1. Acquisition and preprocessing of Landsat 9 imagery;
2. Generation of spectral indices and integration with DSM data;
3. Collection of training and validation samples;
4. Supervised classification using machine learning algorithms;
5. Accuracy assessment and comparative performance analysis.

The entire workflow was implemented within the Google Earth Engine cloud-computing environment, allowing large-scale geospatial processing without local computational constraints (Chen et al., 2024; Praticò et al., 2021; Yang et al., 2022; Ouchra & Belangour, 2021b; Tamiminia et al., 2020).

2.4. Image Pre-processing

Image preprocessing represents a critical stage in remote sensing applications because it directly affects the quality of subsequent classification results (Würsch et al., 2017).

In this study, Landsat 9 Surface Reflectance imagery acquired during 2025 was accessed through the Google Earth Engine platform. Cloud-contaminated observations were removed and image compositing procedures were applied to generate a representative cloud-free image for the study area.

The use of GEE enabled direct access to satellite archives and automated preprocessing operations, including image filtering, mosaicking, and clipping to the study area boundary (Chen et al., 2024; Yang et al., 2022).

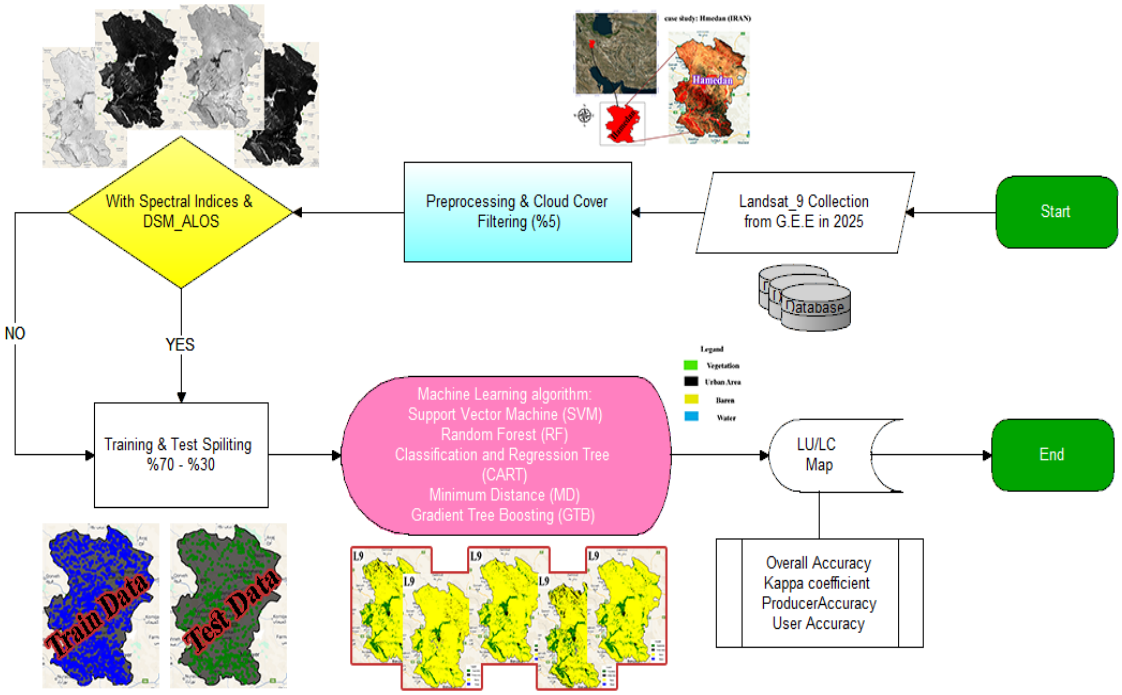

Figure 2. Overall methodological framework adopted for land cover mapping and classifier evaluation.

2.5. Feature Engineering and Geospatial Data Fusion

Feature engineering was performed to increase the discriminatory power of the classification process.

A multi-source geospatial data fusion approach was adopted in which spectral bands, spectral indices, and topographic information were integrated into a unified feature space.

The spectral indices were computed as follows:

Normalized Difference Vegetation Index (NDVI)

$$\frac{\text{Nir} - \text{Red}}{\text{Nir} + \text{Red}}$$

where NIR and Red denote the near-infrared and red spectral bands, respectively.

Normalized Difference Water Index (NDWI)

$$\frac{\text{Green} - \text{Nir}}{\text{Green} + \text{Nir}}$$

used to enhance the detection of water bodies.

Normalized Difference Built-up Index (NDBI)

$$\frac{\text{Nir} - \text{Swir}}{\text{Nir} + \text{Swir}}$$

which highlights urban and built-up features.

Bare Soil Index (BSI)

$$\frac{\text{Green} + \text{Nir}}{\text{Green} - \text{Nir}}$$

for identifying exposed soil surfaces.

Soil Adjusted Vegetation Index (SAVI)

$$\frac{(\text{Nir} - \text{Red})(1 + \text{L})}{(\text{Nir} + \text{Red} + \text{L})}$$

where L is the soil adjustment factor.

A total of 3,090 reference samples were collected and distributed among four land cover classes (water, vegetation, urban area, and bare land). The samples were randomly divided into 70% training data and 30% validation data (Table 1).

Table 1. Number of training samples used for each land cover class.

Class	Number of points
Water	303
Urban area	1110
Vegetation	737
Bareland	940
Total	3090

2.6. Machine Learning Classification

Five supervised classification algorithms were evaluated:

Support Vector Machine (SVM)

Support Vector Machine (Fig 3) identifies an optimal separating hyperplane that maximizes the margin between classes and has been widely applied in remote sensing classification tasks due to its strong generalization capability (Awad, 2021).

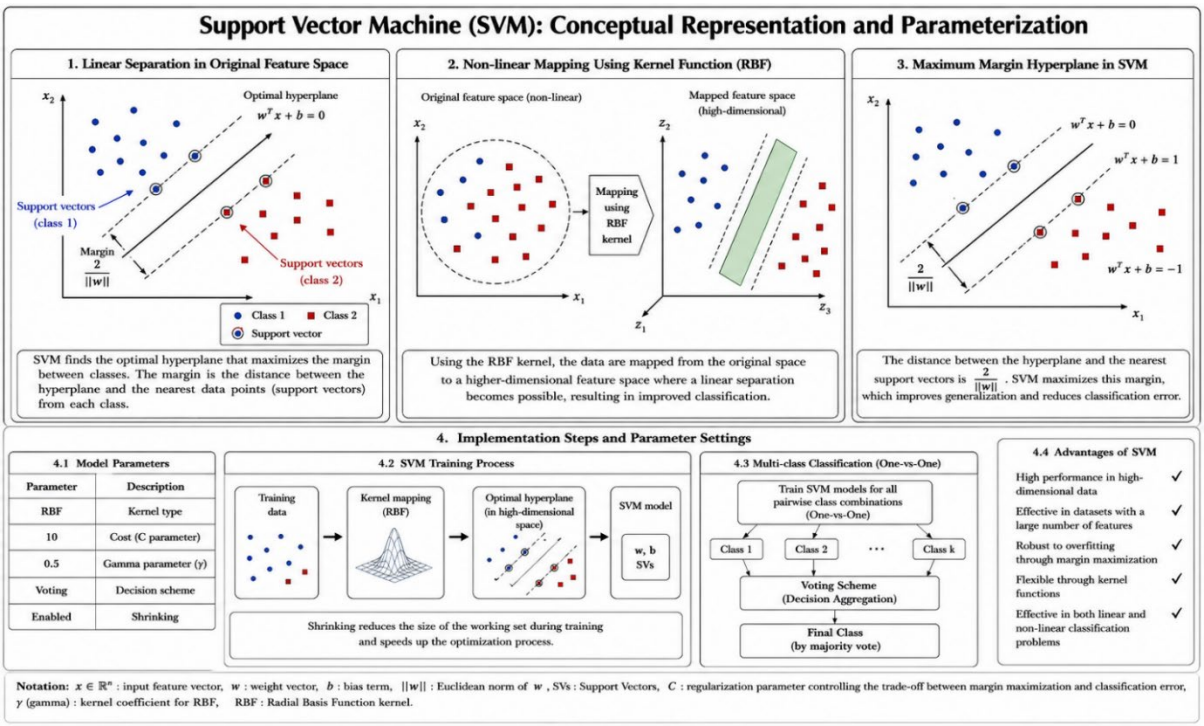

Figure 3. Conceptual representation and parameterization of the SVM classifier.

Minimum Distance (MD)

The Minimum Distance classifier (Figure 4) assigns each pixel to the class whose mean spectral signature is closest in Euclidean feature space. Although computationally efficient, its performance may decline in heterogeneous environments (Basheer et al., 2022).

Gradient Tree Boosting (GTB)

Gradient Tree Boosting (Figure 5) is an ensemble learning technique that iteratively improves classification performance by minimizing residual classification errors through sequential tree construction (Feng et al., 2022; Natekin & Knoll, 2013).

Random Forest (RF)

Random Forest (Figure 6) is an ensemble-based classifier composed of multiple decision trees generated using bootstrap sampling and random feature selection. It is recognized as one of the most robust algorithms for land cover mapping due to its resistance to overfitting and ability to model complex nonlinear relationships (Magidi et al., 2021; Phan et al., 2020).

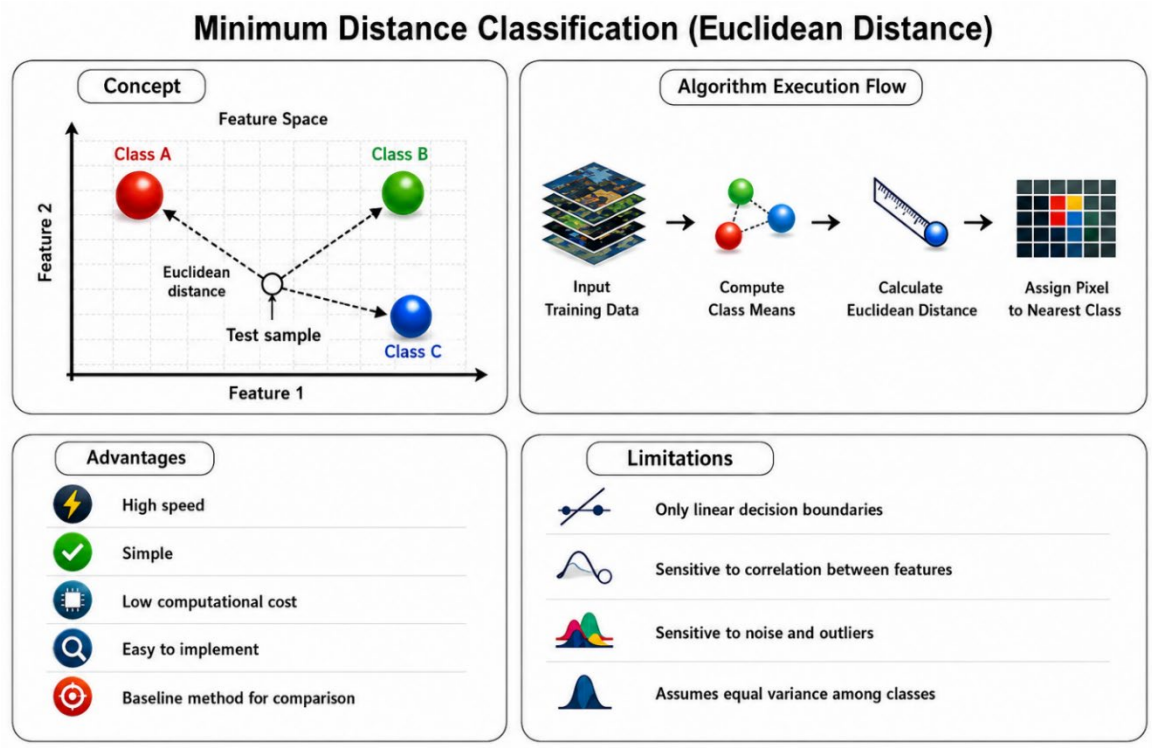

Figure 4. Conceptual workflow of the Minimum Distance classifier.

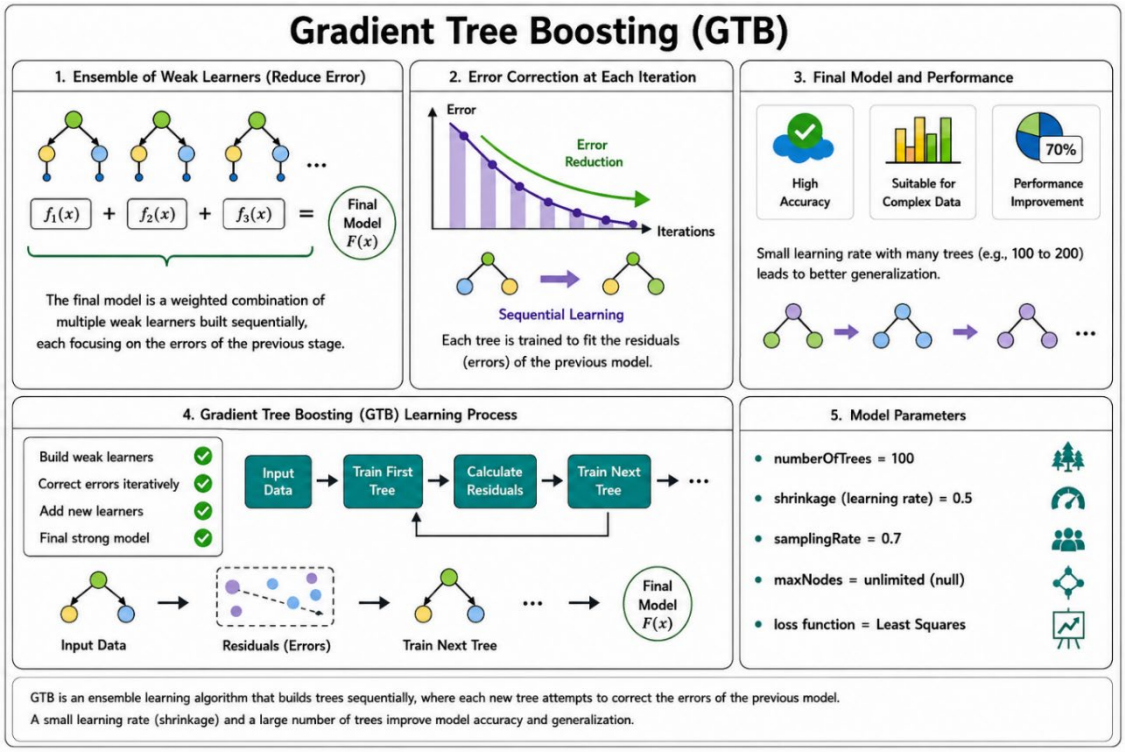

Figure 5. Structure of the Gradient Tree Boosting classifier.

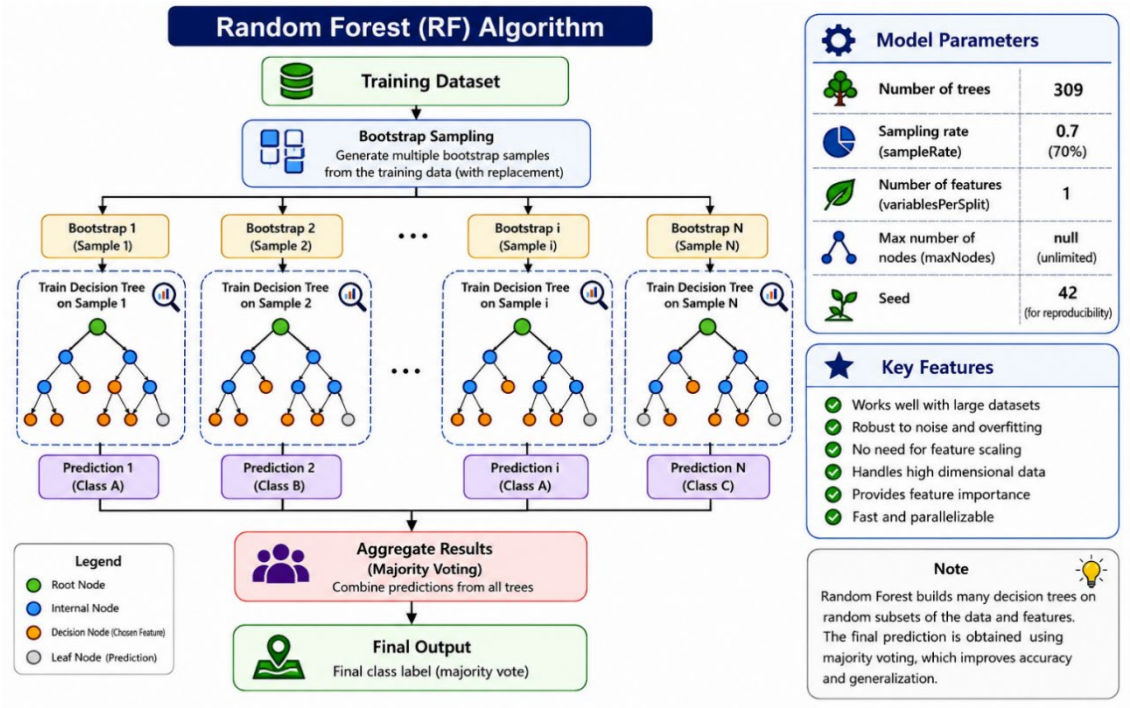

Figure 6. Random Forest architecture and parameter settings used in this study.

Classification and Regression Tree (CART)

CART constructs a hierarchical decision tree through recursive partitioning of the feature space (Figure 7). The algorithm provides interpretable classification rules and has demonstrated effectiveness in geospatial classification applications (Al Farikhi & Pramono, 2023).

2.7. Accuracy Assessment

Classification accuracy was evaluated using an independent validation dataset and confusion matrix analysis. Four standard performance metrics, namely Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA), and the Kappa coefficient, were calculated to assess the reliability of the classification results.

$$\text{Overall Accuracy (OA)} = (\Sigma\, xii) / N \quad (1)$$

where x_{ii} denotes the number of correctly classified samples in class i and N is the total number of validation samples.

$$\text{Producer's Accuracy (PA)} = xii / \Sigma\, xki \quad (2)$$

Producer's Accuracy measures the probability that a reference sample is correctly classified.

$$\text{User's Accuracy (UA)} = xii / \Sigma\, xik \quad (3)$$

User's Accuracy represents the reliability of a classified category from the user's perspective.

$$\text{Kappa Coefficient } (\kappa) = [N\Sigma xii - \Sigma(xi+ x+i)] / [N^2 - \Sigma(xi+ x+i)] \quad (4)$$

The Kappa coefficient evaluates the agreement between the classified map and reference data beyond chance agreement.

These metrics are widely used in land cover mapping studies and provide a comprehensive assessment of classification reliability and agreement beyond random chance (Zaraza Aguiler, 2020; Ouchra et al., 2023). The performance of all classifiers was compared using OA, PA, UA, and Kappa values to identify the most effective machine learning algorithm for land cover mapping in the study area.

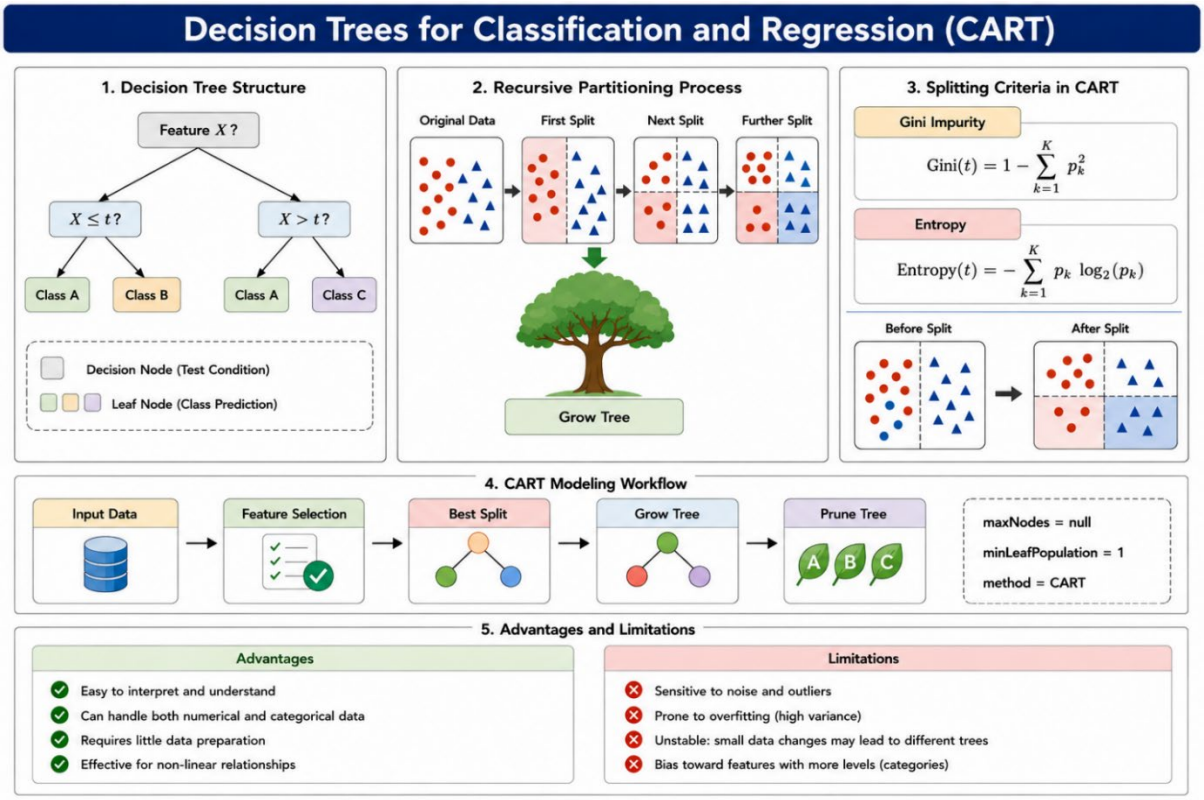

Figure 7. CART decision-tree structure utilized in the study.

3. Results

3.1. Land Cover Classification Results

Land cover maps of Hamedan Province were generated using five supervised machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), Gradient Tree Boosting (GTB), Classification and Regression Tree (CART), and Minimum Distance (MD). Each classifier was implemented under two experimental scenarios: (i) using only Landsat 9

spectral bands and (ii) using Landsat 9 imagery integrated with spectral indices (NDVI, NDWI, NDBI, BSI, and SAVI) and ALOS-DSM elevation data.

The classification results obtained from the five supervised machine learning algorithms are summarized in Figure 8. The figure presents a comprehensive comparison of the generated land cover maps under two experimental scenarios: (i) using Landsat 9 spectral bands only and (ii) integrating Landsat 9 imagery with spectral indices and ALOS-DSM data. For each classifier, including Support Vector Machine (SVM), Random Forest (RF), Gradient Tree Boosting (GTB), Classification and Regression Tree (CART), and Minimum Distance (MD), the corresponding classification maps produced under both scenarios are displayed side by side, enabling direct visual comparison.

A qualitative assessment of Figure 8 indicates that the integration of spectral indices and DSM information generally improved the spatial coherence of the classified maps and enhanced the delineation of major land cover classes. The additional features contributed to reducing spectral confusion among classes with similar reflectance characteristics, particularly between bare land and urban areas. Furthermore, ensemble-based classifiers, especially RF and GTB, produced more homogeneous class distributions and fewer isolated misclassified pixels compared with CART and MD, demonstrating their superior capability in capturing complex spatial patterns within the study area.

3.2. Accuracy Assessment of Classification Algorithms

The accuracy of each classification result was evaluated using an independent validation dataset and confusion matrix analysis. Overall Accuracy (OA), Kappa coefficient, Producer's Accuracy (PA), and User's Accuracy (UA) were computed according to Equations (1)– (4). The corresponding confusion matrices are reported in Tables 3.

Among the evaluated classifiers, Random Forest demonstrated the highest classification performance under both experimental scenarios. As shown in Table 3, RF achieved an Overall Accuracy of 98% and a Kappa coefficient of 0.97 regardless of whether spectral indices and DSM data were included. These results indicate the robustness of RF against variations in input feature space and confirm its suitability for heterogeneous land cover mapping applications.

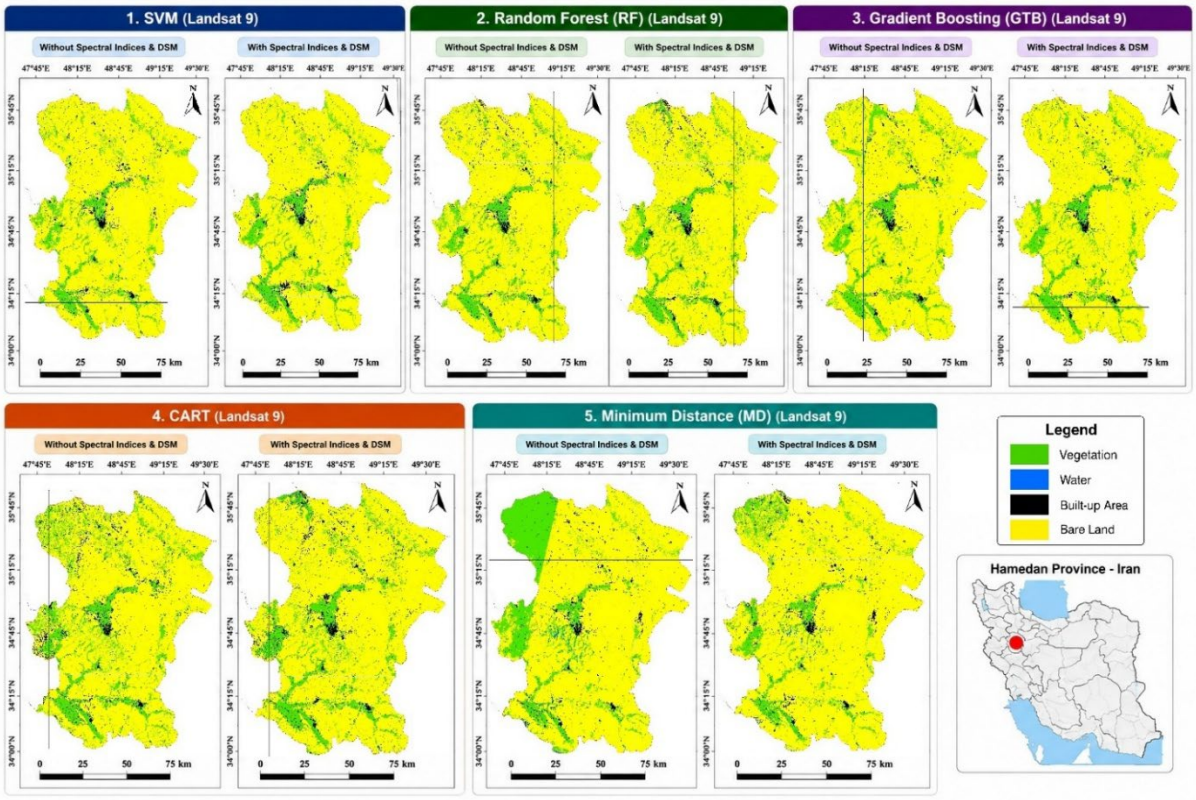

Figure 8. Comparative land cover classification results obtained from five machine learning classifiers under two input data scenarios: Landsat 9 only and Landsat 9 combined with spectral indices and DSM.

As summarized in Table 3, the Gradient Tree Boosting (GTB) classifier demonstrated highly competitive performance in both experimental scenarios. Using only Landsat 9 spectral bands, GTB achieved an Overall Accuracy (OA) of 98% and a Kappa coefficient of 0.97. Following the integration of spectral indices and DSM data, a marginal decrease in performance was observed, with OA and Kappa values slightly declining to 97% and 0.96, respectively. Despite this reduction, GTB remained among the most accurate classifiers evaluated in this study.

The Support Vector Machine (SVM) classifier also exhibited strong classification capability. The incorporation of spectral indices and elevation information improved the Overall Accuracy from 96% to 97%, while the Kappa coefficient increased from 0.95 to 0.96. These improvements indicate that the additional spectral and topographic features enhanced class separability and reduced classification uncertainty.

The CART classifier showed relatively stable performance under both classification scenarios. As presented in Table 3, the Overall Accuracy remained constant at approximately 96%, whereas the Kappa coefficient increased slightly from 0.94 to 0.95 after integrating spectral indices and DSM data. This finding suggests that the inclusion of auxiliary variables had a limited but positive effect on the classification performance of CART.

In contrast, the Minimum Distance (MD) classifier produced the lowest classification accuracies among all evaluated methods. When only Landsat 9 spectral bands were used, MD achieved an Overall Accuracy of 91% and a Kappa coefficient of 0.88. The addition of spectral indices and DSM information improved the OA to 94% and increased the Kappa coefficient to 0.92. Although this improvement demonstrates the positive contribution of multisource feature integration, the MD classifier remained less effective than the machine learning–based approaches.

Table 3. Comparative accuracy assessment of the evaluated classifiers using Landsat 9 imagery with and without spectral indices and DSM data.

Classifier	Scenario	Overall Accuracy	Kappa	PA Vegetation	PA Water	PA Urban	PA Bareland	UA Vegetation	UA Water	UA Urban	UA Bareland
SVM	Landsat 9	0.96	0.95	0.95	1.00	0.97	0.96	0.97	0.98	0.97	0.95
SVM	Landsat 9 + Spectral Indices + DSM	0.97	0.96	0.97	1.00	0.98	0.96	0.98	1.00	0.98	0.96
CART	Landsat 9	0.96	0.94	0.93	1.00	0.98	0.93	0.93	1.00	0.96	0.96
CART	Landsat 9 + Spectral Indices + DSM	0.96	0.95	0.95	1.00	0.97	0.96	0.93	1.00	0.97	0.97
RF	Landsat 9	0.98	0.97	0.97	1.00	0.98	0.98	0.98	1.00	0.97	0.98
RF	Landsat 9 + Spectral Indices + DSM	0.98	0.97	0.98	1.00	0.98	0.97	0.97	1.00	0.98	0.98
MD	Landsat 9	0.91	0.88	0.87	0.96	0.95	0.88	0.84	0.96	0.95	0.88
MD	Landsat 9 + Spectral Indices + DSM	0.94	0.92	0.91	0.98	0.95	0.95	0.97	0.97	0.96	0.90
GTB	Landsat 9	0.98	0.97	0.96	1.00	0.98	0.96	0.97	1.00	0.97	0.98
GTB	Landsat 9 + Spectral	0.97	0.96	0.98	1.00	0.98	0.96	0.95	1.00	0.97	0.98

	Indices + DSM								

Overall, the results reported in Table 3 indicate that ensemble learning methods, particularly Random Forest (RF) and GTB, consistently outperformed the conventional classifiers. Their superior performance can be attributed to their ability to model complex nonlinear relationships and effectively exploit the complementary information provided by spectral indices and topographic variables.

3.3. Class-Level Accuracy Analysis

Producer's Accuracy and User's Accuracy values reported in Table 3 reveal differences in classifier performance across land cover classes.

Water bodies consistently achieved the highest classification accuracies across all classifiers, frequently reaching Producer's Accuracy values close to 100%. This result can be attributed to the distinct spectral characteristics of water surfaces, particularly in the near-infrared and shortwave infrared wavelengths.

Vegetation classes also exhibited high classification accuracy, especially when spectral indices were incorporated into the feature space. The inclusion of vegetation-sensitive indices such as NDVI and SAVI improved class separability and reduced confusion with bare land surfaces.

The greatest classification challenges were observed for urban and bare land classes. These categories share similar spectral characteristics in semi-arid environments, resulting in increased misclassification rates. However, the addition of NDBI, BSI, and DSM information substantially reduced this spectral overlap and improved classification reliability.

3.4. Effect of Spectral Indices and DSM Data on Classification Performance

The influence of spectral indices and topographic information on land cover classification performance was evaluated by comparing the results obtained from two experimental scenarios: (i) using Landsat 9 spectral bands only and (ii) integrating Landsat 9 imagery with spectral indices (NDVI, NDWI, NDBI, BSI, and SAVI) and ALOS-DSM elevation data. The comparative classification maps are presented in Figure 8, while the corresponding quantitative accuracy measures are summarized in Table 3.

Visual inspection of Figure 8 indicates that the inclusion of spectral indices and DSM data improved the spatial coherence of the classified maps and reduced the occurrence of isolated misclassified pixels. The improvement is particularly evident in areas where urban surfaces and bare lands exhibit similar spectral characteristics. The additional spectral and topographic information enhanced class separability and contributed to a more realistic representation of land cover patterns across Hamedan Province.

The quantitative results presented in Table 3 further confirm the positive contribution of multisource data integration. Among the evaluated classifiers, the Support Vector Machine (SVM) showed a measurable improvement after incorporating spectral indices and DSM information, with Overall Accuracy increasing from 96% to 97% and the Kappa coefficient improving from 0.95 to 0.96. Similar trends were observed for the CART classifier, whose Kappa coefficient increased from 0.94 to 0.95, although the Overall Accuracy remained unchanged at 96%.

The most substantial improvement was observed for the Minimum Distance (MD) classifier. The integration of spectral indices and elevation data increased the Overall Accuracy from 91% to 94% and improved the Kappa coefficient from 0.88 to 0.92. This finding suggests that distance-based classifiers benefit considerably from additional features that enhance spectral discrimination among land cover classes. In particular, vegetation-sensitive indices and topographic variables reduced confusion between vegetation and bare land classes, leading to a notable increase in classification reliability.

In contrast, the ensemble-based classifiers Random Forest (RF) and Gradient Tree Boosting (GTB) demonstrated consistently high performance under both scenarios. RF achieved an Overall Accuracy of 98% and a Kappa coefficient of 0.97 regardless of the inclusion of additional variables, indicating that the classifier was already capable of extracting highly discriminative information from the original Landsat 9 spectral bands. GTB also maintained excellent classification performance; however, a slight decrease in Overall Accuracy (from 98% to 97%) and Kappa coefficient (from 0.97 to 0.96) was observed after integrating spectral indices and DSM data. Although the reduction was marginal, it suggests that the additional variables did not provide substantial new information beyond that already captured by the boosting ensemble.

A class-level examination of Producer's Accuracy and User's Accuracy values reveals that water bodies consistently achieved the highest accuracies across all classification scenarios, frequently approaching or reaching 100%. This outcome can be attributed to the distinctive spectral behavior of water surfaces in the near-infrared and shortwave infrared regions. Vegetation classes also benefited from the inclusion of NDVI and SAVI, whereas urban and bare land classes exhibited the greatest improvement due to the incorporation of NDBI, BSI, and elevation information.

Overall, the results demonstrate that the integration of spectral indices and DSM data generally enhanced classification performance and improved the spatial quality of the resulting land cover maps. Nevertheless, the magnitude of improvement varied among classifiers, with the greatest benefits observed for conventional distance-based approaches and relatively limited gains for advanced ensemble learning algorithms. These findings highlight the importance of considering both classifier characteristics and feature selection strategies when developing land cover classification frameworks based on multisource remote sensing data.

3.5. Comparative Performance of Machine Learning Algorithms

The overall performance comparison among classifiers is presented in Table 3. The results indicate that RF achieved the highest and most stable performance across all experimental configurations. GTB ranked second and demonstrated comparable accuracy levels, followed by SVM and CART. The Minimum Distance classifier consistently produced the lowest accuracies.

Based on Overall Accuracy and Kappa coefficient values, the classifiers can be ranked as follows:

RF > GTB > SVM > CART > MD

This ranking highlights the superiority of ensemble-based machine learning approaches for land cover mapping in heterogeneous environments. The ability of RF and GTB to model nonlinear relationships and exploit complex feature interactions appears to be a key factor contributing to their superior performance.

The results further demonstrate that the integration of multisource geospatial information, including spectral indices and DSM data, improves classification reliability and supports the production of more accurate land cover maps for environmental monitoring and spatial planning applications.

4. Discussion

The results demonstrate that integrating multisource geospatial data within a cloud-based machine learning framework improves land cover classification performance. The incorporation of spectral indices and ALOS-DSM data enhanced class separability and reduced spectral confusion, particularly between urban and bare land classes. Similar improvements resulting from the integration of spectral and topographic variables have been reported in previous studies (Feng et al., 2022; Palanisamy et al., 2023; Ouchra et al., 2023; Kazemi Garajeh et al., 2024).

Among the evaluated classifiers, Random Forest (RF) achieved the highest classification performance, with an Overall Accuracy of 98% and a Kappa coefficient of 0.97. The strong performance of RF can be attributed to its ensemble learning structure, which effectively models nonlinear relationships and reduces overfitting (Magidi et al., 2021). Maxwell et al. (Maxwell et al., 2018), and Pham et al. (Tamiminia et al., 2020), who identified RF as one of the most reliable classifiers for land cover mapping.

Gradient Tree Boosting (GTB) also produced highly competitive results, confirming the effectiveness of boosting-based ensemble approaches for remote sensing classification (Feng

et al., 2022; Al Farikhi & Pramono, 2023). Although a slight decrease in accuracy was observed after adding spectral indices and DSM data, GTB remained among the best-performing classifiers.

The incorporation of spectral indices and elevation information improved the performance of SVM and MD classifiers. For SVM, Overall Accuracy increased from 96% to 97%, while the Kappa coefficient improved from 0.95 to 0.96. The most notable improvement was observed for the MD classifier, whose Overall Accuracy increased from 91% to 94%. These findings indicate that additional spectral and topographic variables enhance class discrimination and are particularly beneficial for classifiers that rely heavily on spectral distance measures (Awad, 2021; Basheer et al., 2022; Zaraza Aguiler, 2020).

Class-level analysis revealed that water bodies consistently achieved the highest classification accuracies across all algorithms. This can be explained by the distinct spectral response of water surfaces and the effectiveness of NDWI for water extraction (McFeeters, 1996; Kazemi Garajeh et al., 2024). Vegetation classes also benefited from the inclusion of NDVI and SAVI, which improve vegetation discrimination and reduce soil background effects (Rouse et al., 1974; Zha et al., 2003). In contrast, urban and bare land classes represented the greatest classification challenge because of their spectral similarity in semi-arid environments. The incorporation of NDBI, BSI, and DSM data reduced this confusion and improved classification reliability (Zha et al., 2003; Maxwell et al., 2018).

The results further highlight the advantages of Google Earth Engine as a cloud-based geospatial platform. GEE enabled efficient image preprocessing, feature extraction, classification, and accuracy assessment within a unified environment, supporting previous findings regarding its effectiveness for large-scale remote sensing applications (Chen et al., 2024; Praticò et al., 2021).

Overall, the findings confirm that the integration of Landsat 9 imagery, spectral indices, and DSM data can significantly enhance land cover classification performance. The superior performance of RF and GTB demonstrates the potential of ensemble machine learning methods for generating reliable land cover information that can support environmental monitoring, urban planning, and sustainable land management.

5. Conclusion

This study developed a cloud-based geospatial data fusion framework for land cover mapping by integrating Landsat 9 imagery, spectral indices, and ALOS-DSM data within the Google Earth Engine platform. The performance of five supervised classification algorithms, including RF, GTB, SVM, CART, and MD, was evaluated for mapping four major land cover classes in Hamedan Province, Iran. The results demonstrated that the integration of spectral and topographic information generally improved classification accuracy and reduced spectral

confusion among land cover categories. Among the evaluated classifiers, Random Forest achieved the highest performance, with an Overall Accuracy of 98% and a Kappa coefficient of 0.97, followed closely by Gradient Tree Boosting. The findings confirm the effectiveness of combining multisource geospatial data and machine learning techniques for accurate land cover mapping. The proposed framework can support environmental monitoring, urban planning, natural resource management, and sustainable spatial decision-making at regional scales.

6. References

Al Farikhi, F., & Pramono, R.W.D. Perbandingan algoritma Classification and Regression Tree (CART) dan Random Forest (RF) untuk klasifikasi penggunaan lahan pada Google Earth Engine. *Jurnal Spatial Wahana Komunikasi dan Informasi Geografi, 2023, 23(2),* 170–179. https://doi.org/10.21009/spatial.232.09.

Awad, M. Google Earth Engine (GEE) cloud computing-based crop classification using radar, optical images and support vector machine algorithm (SVM). In Proceedings of the IEEE 3rd International Multidisciplinary Conference on Engineering Technology (IMCET), *Beirut, Lebanon, 2021*, 71–76. https://doi.org/10.1109/IMCET53404.2021.9665519.

Basheer, S., Wang, X., Farooque, A.A., Nawaz, R.A., Liu, K., Adekanmbi, T., & Liu, S. Comparison of land use/land cover classifiers using different satellite imagery and machine learning techniques. *Remote Sensing, 2022, 14(19)*, 4978. https://doi.org/10.3390/rs14194978.

Borgogno-Mondino, E., Lessio, A., Tarricone, L., Novello, V., & de Palma, L. A comparison between multispectral aerial and satellite imagery in precision viticulture. *Precision Agriculture, 2018, 19(2),* 195–217. https://doi.org/10.1007/s11119-017-9510-0.

Borra, S., Thanki, R., & Dey, N. Satellite Image Analysis: Clustering and Classification. Singapore: *Springer Nature*, 2019.

Capolupo, A., & Tarantino, E. Landsat 9 satellite images potentiality in extracting land cover classes in GEE environment using an index-based approach: The case study of Savona City. In: Innovations in GIS and Remote Sensing Applications. *Cham: Springer*, 2023. https://doi.org/10.1007/978-3-031-37114-1_17.

Chen, X., Wang, N., Peng, S., Meng, N., & Lv, H. Analysis of spatiotemporal dynamics of land desertification in Qilian Mountain National Park based on Google Earth Engine. *ISPRS International Journal of Geo-Information, 2024, 13(4),* 117. https://doi.org/10.3390/ijgi13040117.

Dhingra, S., & Kumar, D. A review of remotely sensed satellite image classification. *International Journal of Electrical and Computer Engineering, 2019, 9(3),* 1720–1731. https://doi.org/10.11591/ijece.v9i3.pp1720-1731.

Diek, S., Fornallaz, F., Schaepman, M.E., & De Jong, R. Barest Pixel Composite for agricultural areas using Landsat time series. *Remote Sensing, 2017, 9(12),* 1245. https://doi.org/10.3390/rs9121245.

Feng, K., Wang, T., Liu, S., Kang, W., Chen, X., Guo, Z., & Zhi, Y. Monitoring desertification using machine-learning techniques with multiple indicators derived from MODIS images in Mu Us Sandy Land, China. *Remote Sensing, 2022, 14(11),* 2663. https://doi.org/10.3390/rs14112663.

Hu, Y., Dong, Y., & Batunacun. An automatic approach for land-change detection and land updates based on integrated NDVI timing analysis and the CVAPS method with Google Earth Engine support. *ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 146,* 347–359. https://doi.org/10.1016/j.isprsjprs.2018.10.008.

Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 1988, 25(3),* 295–309. https://doi.org/10.1016/0034-4257(88)90106-X.

Johary, A.R.F., Révillion, C., Catry, T., Alexandre, C., Mouquet, P., Rakotoniaina, S., Pennober, G., & Rakotondraompiana, S. Detection of large-scale floods using Google Earth Engine and Google Colab. *Remote Sensing, 2023, 15(22*), 5368. https://doi.org/10.3390/rs15225368.

John, A., Cannistra, A.F., Yang, K., Tan, A., Shean, D., Hille Ris Lambers, J., & Cristea, N. High-resolution snow-covered area mapping in forested mountain ecosystems using PlanetScope imagery. *Remote Sensing, 2022, 14(14),* 3409. https://doi.org/10.3390/rs14143409.

Kazemi Garajeh, M., Haji, F., Tohidfar, M., et al. Spatiotemporal monitoring of climate change impacts on water resources using an integrated approach of remote sensing and Google Earth Engine. *Scientific Reports, 2024, 14*, 5469. https://doi.org/10.1038/s41598-024-56160-9.

Landsat 8 | Landsat Science. National Aeronautics and Space Administration (NASA). Available online: https://landsat.gsfc.nasa.gov/satellites/landsat-8/ (accessed 30 January 2023).

Landsat 9 | Landsat Science. National Aeronautics and Space Administration (NASA). Available online: https://landsat.gsfc.nasa.gov/satellites/landsat-9/ (accessed 30 January 2023).

Magidi, J., Nhamo, L., Mpandeli, S., & Mabhaudhi, T. Application of the Random Forest classifier to map irrigated areas using Google Earth Engine. *Remote Sensing, 2021, 13(5*), 876. https://doi.org/10.3390/rs13050876.

Masoumi, T., Eslamkish, T., Honarmand, M., & Abkar, A.A. A comparative study of Landsat-7 and Landsat-8 data using image processing methods for hydrothermal alteration mapping. *Resource Geology, 2017, 67(1),* 72–88. https://doi.org/10.1111/rge.12117.

Maxwell, A.E., Warner, T.A., & Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing, 2018, 39(9),* 2784–2817. https://doi.org/10.1080/01431161.2018.1433343.

McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 1996, 17(7*), 1425–1432. https://doi.org/10.1080/01431169608948714.

Natekin, A., & Knoll, A. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics, 2013*, 7, 21. https://doi.org/10.3389/fnbot.2013.00021.

Nelson, P.R., Maguire, A.J., Pierrat, Z., Orcutt, E.L., Yang, D., Serbin, S., Frost, G.V., Macander, M.J., Magney, T.S., Thompson, D.R., Wang, J.A., Oberbauer, S.F., Zesati, S.V., Davidson, S.J., Epstein, H.E., Unger, S., Campbell, P.K.E., Carmon, N., Velez-Reyes, M., & Huemmrich, K.F. Remote sensing of tundra ecosystems using high spectral resolution reflectance: Opportunities and challenges. *Journal of Geophysical Research: Biogeosciences, 2022, 127(2),* e2021JG006697. https://doi.org/10.1029/2021JG006697.

Ouchra, H., & Belangour, A. Object detection approaches in images: A survey. *Proceedings of SPIE, 2021*, *11878,* 132–141. https://doi.org/10.1117/12.2601452.

Ouchra, H., & Belangour, A. Satellite image classification methods and techniques: A survey. In Proceedings of the IEEE International Conference on Imaging Systems and Techniques (IST), *Kaohsiung, Taiwan, 2021*, 1–6. https://doi.org/10.1109/IST50367.2021.9651454.

Ouchra, H., Belangour, A., & Erraissi, A. A comparative study on pixel-based classification and object-oriented classification of satellite image. *International Journal of Engineering Trends and Technology, 2022, 70(8*), 206–215. https://doi.org/10.14445/22315381/IJETT-V70I8P221.

Ouchra, H., Belangour, A., & Erraissi, A. A comprehensive study of using remote sensing and geographical information systems for urban planning. *InterNetworking Indonesia Journal, 2022, 14(1*), 15–20.

Ouchra, H., Belangour, A., & Erraissi, A. Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. *IEEE Access, 2023, 11,* 71127–71142. https://doi.org/10.1109/ACCESS.2023.3293828.

Ouchra, H., Belangour, A., & Erraissi, A. Machine learning for satellite image classification: A comprehensive review. *In Proceedings of the International Conference on Data Analytics for Business and Industry (ICDABI), 2022*, 1–5. https://doi.org/10.1109/ICDABI56818.2022.10041606.

Ouchra, H., Belangour, A., & Erraissi, A. Satellite data analysis and geographic information system for urban planning: A systematic review. *In Proceedings of the International Conference on Data Analytics for Business and Industry (ICDABI), 2022.* https://doi.org/10.1109/ICDABI56818.2022.10041487.

Ouchra, H., Belangour, A., & Erraissi, A. Spatial data mining technology for GIS: A review. *In Proceedings of the International Conference on Data Analytics for Business and Industry (ICDABI), 2022*, 655–659. https://doi.org/10.1109/ICDABI56818.2022.10041574.

Palanisamy, P.A., Jain, K., & Bonafoni, S. Machine learning classifier evaluation for different input combinations: A case study with Landsat 9 and Sentinel-2 data. *Remote Sensing, 2023, 15(13*), 3241. https://doi.org/10.3390/rs15133241.

Phan, T.N., Kuch, V., & Lehnert, L.W. Land cover classification using Google Earth Engine and Random Forest classifier: The role of image composition. *Remote Sensing, 2020, 12(15),* 2411. https://doi.org/10.3390/rs12152411.

Praticò, S., Solano, F., Di Fazio, S., & Modica, G. Machine learning classification of Mediterranean forest habitats in Google Earth Engine based on seasonal Sentinel-2 time-series and input image composition optimisation. *Remote Sensing, 2021, 13(4),* 586. https://doi.org/10.3390/rs13040586.

Qiu, Z., Liu, D., Yan, N., Yan, Y., Yang, C., Zhang, C., & Duan, H. Landsat and dual random forest modelling reveal sediment fining in the Yellow River shaped by ecological restoration on China's Loess Plateau. *Remote Sensing of Environment, 2025,* 330, 114994. https://doi.org/10.1016/j.rse.2025.114994.

Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., & Chao, D. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sensing, 2020, 12(9),* 1432. https://doi.org/10.3390/rs12091432.

Rouse, J.W., Haas, R.H., Schell, J.A., & Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden, S.C., Mercanti, E.P., & Becker, M.A. (Eds.), *Third Earth Resources Technology Satellite-1 Symposium, Vol*. 1. NASA Special Publication SP-351, Greenbelt, MD, USA, 1974, pp. 309–317.

Salas, E.A.L., Kumaran, S.S., Bennett, R., Willis, L.P., & Mitchell, K. Machine learning-based classification of small-sized wetlands using Sentinel-2 images. *AIMS Geosciences, 2024, 10(1),* 62–79. https://doi.org/10.3934/geosci.2024005.

Saraei, M., Almodaresi, S.A., & Raghebian Hanzaie, F. Using satellite imagery to assess urban growth in Yazd City from 1996 to 2016. *Journal of Radar and Optical Remote Sensing and GIS, 2023, 6(3),* 45–60. https://doi.org/10.71593/jrors.2023.1103307.

Shafaey, M.A., Salem, M.A.M., Ebied, H.M., Al-Berry, M.N., & Tolba, M.F. Deep learning for satellite image classification. *In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics (AISI),* 2018.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L.J., Adeli, S., & Brisco, B. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 164*, 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001.

Torki, M. Application of machine learning in spatial data analysis. *Journal of Radar and Optical Remote Sensing and GIS, 2025,* 8(4).

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. Unsupervised machine learning for networking: *Techniques, applications and research challenges. IEEE Access, 2019,* 7, 65579–65615. https://doi.org/10.1109/ACCESS.2019.2916648.

Wahbi, M., El Bakali, I., Ez-Zahouani, B., Azmi, R., Moujahid, A., Zouiten, M., Alaoui, O.Y., Boulaassal, H., Maatouk, M., & El Kharki, O. A deep learning classification approach using high spatial satellite images for detection of built-up areas in rural zones: Case study of Souss-Massa Region, Morocco. *Remote Sensing Applications: Society and Environment, 2023,* 29, 100898. https://doi.org/10.1016/j.rsase.2022.100898.

Würsch, L., Hurni, K., & Heinimann, A. Google Earth Engine image pre-processing tool: User guide. Centre for Development and Environment (CDE), *University of Bern, Bern, Switzerland, 2017.* Available online: https://www.cde.unibe.ch/e65013/e542846/e707304/e707386/e707390/CDE_PreprocessingToolUserGuide_eng.pdf (accessed 13 December 2022).

Xiao, W., Xu, S., & He, T. Mapping paddy rice with Sentinel-1/2 and a phenology-based object-oriented algorithm: An implementation in the Hangjiahu Plain, China, using the Google Earth Engine platform. *Remote Sensing, 2021, 13(5),* 990. https://doi.org/10.3390/rs13050990.

Yang, L., Driscol, J., Sarigai, S., Wu, Q., Chen, H., & Lippitt, C.D. Google Earth Engine and artificial intelligence (AI): A comprehensive review. *Remote Sensing, 2022, 14(14),* 3253. https://doi.org/10.3390/rs14143253.

Yang, Y., Yang, D., Wang, X., Zhang, Z., & Nawaz, Z. Testing accuracy of land cover classification algorithms in the Qilian Mountains based on the Google Earth Engine cloud platform. *Remote Sensing, 2021, 13(24),* 5064. https://doi.org/10.3390/rs13245064.

Zaraza Aguiler, M. Classification of land cover through machine learning algorithms for fusion of Sentinel-2A and PlanetScope imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020, XLII-3/W12, 361–368. https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-361-2020.

Zha, Y., Gao, J., & Ni, S. Use of normalized difference built-up index (NDBI) in automatically mapping urban areas from TM imagery. International *Journal of Remote Sensing, 2003, 24(3),* 583–594. https://doi.org/10.1080/01431160304987.