DOI: 10.35595/2414-9179-2025-1-31-519-532

Nuriddin Mirjalalov[1], Nozimjon Teshaev[2], Eshkabul Safarov[3], Jasmina Gerts[4], Adbujalil Mominov[5], Anvar Pardaboyev[6]

COMPARATIVE ANALYSIS OF RANDOM FOREST AND SUPPORT VECTOR MACHINE FOR LULC CLASSIFICATION IN TASHKENT REGION USING LANDSAT-8 IMAGERY

ABSTRACT

Accurate Land Use/Land Cover (LULC) classification is essential for effective environmental monitoring, sustainable agricultural management, and informed urban planning. With increasing land transformation driven by urbanization, deforestation, and climate variability, reliable classification methods are needed to support data-driven decision-making. This study presents a comparative analysis of two widely used machine learning algorithms — Support Vector Machine (SVM) and Random Forest (RF) — for LULC classification in the Tashkent Region, Uzbekistan, using Landsat-8 Operational Land Imager (OLI) data for the June–July of 2024. The workflow involved critical preprocessing steps, cloud filtering, and band selection, followed by classification using six dominant LULC classes: water, bare land, built-up areas, cropland, snow, and forest. Ground truth samples were used to train and validate the models. Accuracy assessment was conducted using a confusion matrix, and performance was evaluated based on Overall Accuracy (OA) and the Kappa Coefficient (KC). The results revealed that the SVM classifier slightly outperformed RF, achieving an OA of 96.78 % and a KC of 0.96, compared to RF's OA of 94.95 % and KC of 0.94. The superior performance of SVM is likely due to its effectiveness in handling high-dimensional feature spaces and separating non-linear class boundaries, especially in heterogeneous landscapes like Tashkent Region. While both algorithms showed strong potential, SVM demonstrated better precision in classifying cropland and forest areas. These findings highlight the importance of algorithm selection in remote sensing-based LULC studies. The study contributes to ongoing efforts to enhance land cover mapping accuracy using machine learning and offers valuable insights for land managers, urban developers, and environmental policymakers. Future research may consider the integration of multi-seasonal imagery, ancillary environmental data, and deep learning frameworks to further improve classification performance.

KEYWORDS: land use/land cover (LULC), random forest (RF), support vector machine (SVM), Landsat-8, remote sensing, Tashkent Region

---

[1] Mirzo Ulugbek National University of Uzbekistan, 4, University str., Tashkent, 100174, Republic of Uzbekistan, *e-mail:* **mirnur748@gmail.com**

[2] National Research University, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, 39, Koriy Niyoziy str., Tashkent, 100000, Republic of Uzbekistan, *e-mail:* **n.teshayev@tiiame.uz**

[3] Mirzo Ulugbek National University of Uzbekistan, 4, University str., Tashkent, 100174, Republic of Uzbekistan, *e-mail:* **e.safarov@nuu.uz**

[4] Turin Polytechnic University, 17, Kichik Halka Yuli str., Tashkent, 100095, Republic of Uzbekistan, *e-mail:* **jasminagerts@tiiame.uz**

[5] Alfraganus University, 2a, Yuqori Karakamish str., Tashkent, 100190, Republic of Uzbekistan, *e-mail:* **mirnur748@gmail.com**

[6] Tashkent State Technical University, 2b, University str., Tashkent, 100000, Republic of Uzbekistan, *e-mail:* **mirnur748@gmail.com**

INTRODUCTION

Land Use/Land Cover (LULC) classification is a fundamental component of remote sensing and geographic information system (GIS) applications, playing a crucial role in environmental monitoring, natural resource management, and sustainable urban planning [*Dash*, *Maity*, 2023; *Bhungeni* et al., 2024]. Accurate LULC mapping is essential for understanding land dynamics, evaluating the impacts of climate change, and supporting decision-making in agriculture, forestry, and water resource management [*Belgiu*, *Drăguţ*, 2016]. The increasing availability of satellite imagery, particularly from sensors like Landsat-8 Operational Land Imager (OLI), has facilitated large-scale and high-resolution LULC classification, allowing for more precise monitoring of landscape transformations. Various machine learning approaches have been employed to enhance classification accuracy, with Random Forest (RF) and Support Vector Machine (SVM) emerging as two widely used techniques due to their robustness and adaptability in remote sensing applications [*Huang* et al., 2002; *Pal*, 2005; *Gislason* et al., 2006; *Mountrakis* et al., 2011]. However, the comparative effectiveness of these methods for LULC classification in the Tashkent Region remains partly unexplored.

Uzbekistan, particularly the Tashkent Region, has been undergoing significant land cover changes due to rapid urbanization, agricultural expansion, and climate-induced environmental shifts [*Teshaev* et al., 2020; *Aslanov* et al., 2023; 2024]. As the capital and largest economic center of Uzbekistan, Tashkent has experienced considerable urban sprawl, leading to the conversion of agricultural and forested areas into built-up environments [*Alikhanov* et al., 2020; *Aslanov* et al., 2021; *Erdanaev* et al., 2022; *Sharipov*, *Khayitmurodov*, 2024]. Additionally, climate variability has influenced seasonal snow cover, water availability, and cropland productivity, necessitating accurate and up-to-date LULC classification for sustainable land management policies [*Erdanaev* et al., 2022; *Farmonov* et al., 2023]. Given these dynamics, implementing advanced machine learning-based classification techniques is essential for improving land cover monitoring in the region.

Traditional LULC classification methods, such as Maximum Likelihood Classification (MLC) and K-Nearest Neighbors (KNN), have been widely used in remote sensing applications [*Foody*, 2002; *Mirzaei* et al., 2023]. However, these approaches often suffer from limited classification accuracy, especially in complex landscapes with mixed land cover types [*Foody*, 2002]. Several studies have assessed the performance of machine learning algorithms for land cover mapping using satellite imagery [*Farmonov* et al., 2024]. For instance, [*Belgiu*, *Drăguţ*, 2016] conducted a comparative analysis of RF and SVM, demonstrating that RF generally provides better classification accuracy due to its ability to handle high-dimensional datasets and reduce overfitting. Similarly, [*Pal*, 2005] explored the advantages of SVM in remote sensing and found that its performance varies depending on the study area's complexity and the chosen spectral features. More recent studies, such as those by [*Zhang* et al., 2022] have integrated spectral indices and multi-temporal imagery, highlighting the significance of proper feature selection in improving classification outcomes.

Indeed, in recent years, machine learning algorithms, particularly Random Forest (RF) and Support Vector Machine (SVM), have emerged as robust and efficient alternatives due to their ability to handle high-dimensional data and non-linear relationships [*Pal*, 2005; *Rodriguez-Galiano* et al., 2012]. Random Forest (RF) is an ensemble learning algorithm that constructs multiple decision trees to improve classification accuracy and reduce overfitting [*Rodriguez-Galiano* et al., 2012; *Tian* et al., 2016; *Peng* et al., 2023; *Alikhanov* et al., 2024]. It has been widely adopted in LULC classification studies due to its robustness in handling spectral variability and missing data [*Breiman*, 2001]. Several studies have demonstrated RF's superior performance in remote sensing applications, particularly in heterogeneous landscapes [*Gislason* et al., 2006; *Belgiu*, *Drăguţ*, 2016]. Support Vector Machine (SVM), on the other hand, is a supervised learning algorithm that finds an optimal hyperplane to separate different land cover classes [*Vapnik*, 2000]. It is particularly effective for small sample sizes and high-dimensional datasets and has been successfully applied in various LULC studies [*Huang* et al., 2002; *Mountrakis* et al., 2011]. However, SVM's performance is highly dependent on parameter tuning, and it may struggle with complex class distributions and overlapping spectral signatures [*Maxwell*, *Fang*, 2018].

Several comparative studies have evaluated RF and SVM for LULC classification using satellite imagery, highlighting their advantages and limitations. While RF has been found to be computationally efficient and accurate, SVM often performs well in cases where clear class separability exists [*Adam* et al., 2010]. However, limited research has been conducted on the application of these algorithms in Uzbekistan, specifically in the Tashkent Region, where land cover patterns are rapidly evolving. Given the increasing demand for high-accuracy LULC classification in the Tashkent Region, this study aims to compare the performance of RF and SVM for LULC classification using Landsat-8 OLI imagery. Evaluate the classification accuracy of each model by assessing Overall Accuracy (OA) and Kappa Coefficient (KC). Identify the most suitable algorithm for large-scale land cover mapping in Uzbekistan, focusing on urban, snow, bare land, forest, water, and croplands. Contribute to existing knowledge on machine learning-based LULC classification in Uzbekistan by providing a case study in the Tashkent Region.

Landsat-8, launched by NASA and the USGS in 2013, provides high-quality multispectral and thermal data, making it an essential resource for LULC analysis. Compared to previous Landsat missions, Landsat-8 offers improved radiometric resolution (16-bit), enhanced signal-to-noise ratio, and two additional spectral bands, the Coastal/Aerosol band (B1) and the Cirrus band (B9) — which improve atmospheric correction and land cover discrimination. The Operational Land Imager (OLI) sensor, capturing data across nine spectral bands, ensures high accuracy in distinguishing vegetation, water bodies, urban areas, and bare land. Additionally, the satellite's 30-meter spatial resolution and 16-day revisit cycle provide an optimal balance between spatial coverage and temporal frequency, making it suitable for both seasonal and long-term LULC monitoring. Compared to Sentinel-2, which has higher spatial resolution but a shorter historical archive, Landsat-8's long-term data availability supports multi-year trend analysis in rapidly changing environments like Tashkent [*Schmidt* et al., 2016; *Korhonen* et al., 2017; *Sajib*, *Wang*, 2020].

This study aims to enhance the understanding of machine learning applications in LULC classification and provide valuable insights for urban planners, environmental managers, and policymakers in Uzbekistan. By comparing two classification algorithms, this research seeks to improve land cover monitoring efforts and support sustainable land management strategies in the region.

RESEARCH MATERIALS AND METHODS

The study was conducted in the Tashkent Region, which serves as the capital and the most populous administrative division of Uzbekistan. Covering an area of approximately 15 300 km², the region is strategically significant due to its economic, agricultural, and environmental importance. Tashkent is in the northeastern part of the country, bordered by Kazakhstan to the north and the Syr Darya River to the west, with varying topography that includes plains, foothills, and mountainous areas in the east, where it meets the Chatkal Mountain Range [*Alikhanov* et al., 2020].

Tashkent experiences a continental climate, characterized by hot, dry summers and cold winters, leading to significant seasonal variations in land cover. The average temperature in summer can exceed 35 °C, while winter temperatures often drop below freezing, with occasional snowfall. These climatic variations contribute to seasonal shifts in vegetation, changes in water availability, and fluctuations in snow cover, which must be accounted for in LULC classification efforts [*Erdanaev* et al., 2022; *Sharipov*, *Khayitmurodov*, 2024].

The land cover of the Tashkent Region has been rapidly changing due to multiple factors, including urbanization, agricultural expansion, deforestation, and climate-induced environmental changes [*Aslanov* et al., 2021]. The growth of urban areas is primarily driven by population increase, infrastructure development, and industrialization, leading to the conversion of agricultural lands and natural vegetation into built-up environments. Meanwhile, agriculture remains a dominant land use in the region, with large areas dedicated to wheat, cotton, orchards, and vegetable cultivation, often supported by extensive irrigation systems fed by the Chirchiq River and other smaller tributaries [*Gafforov* et al., 2020]. The diversity in land cover types, coupled with rapid urban and environmental transformations, makes the Tashkent region an ideal study site for LULC classification using remote sensing techniques.

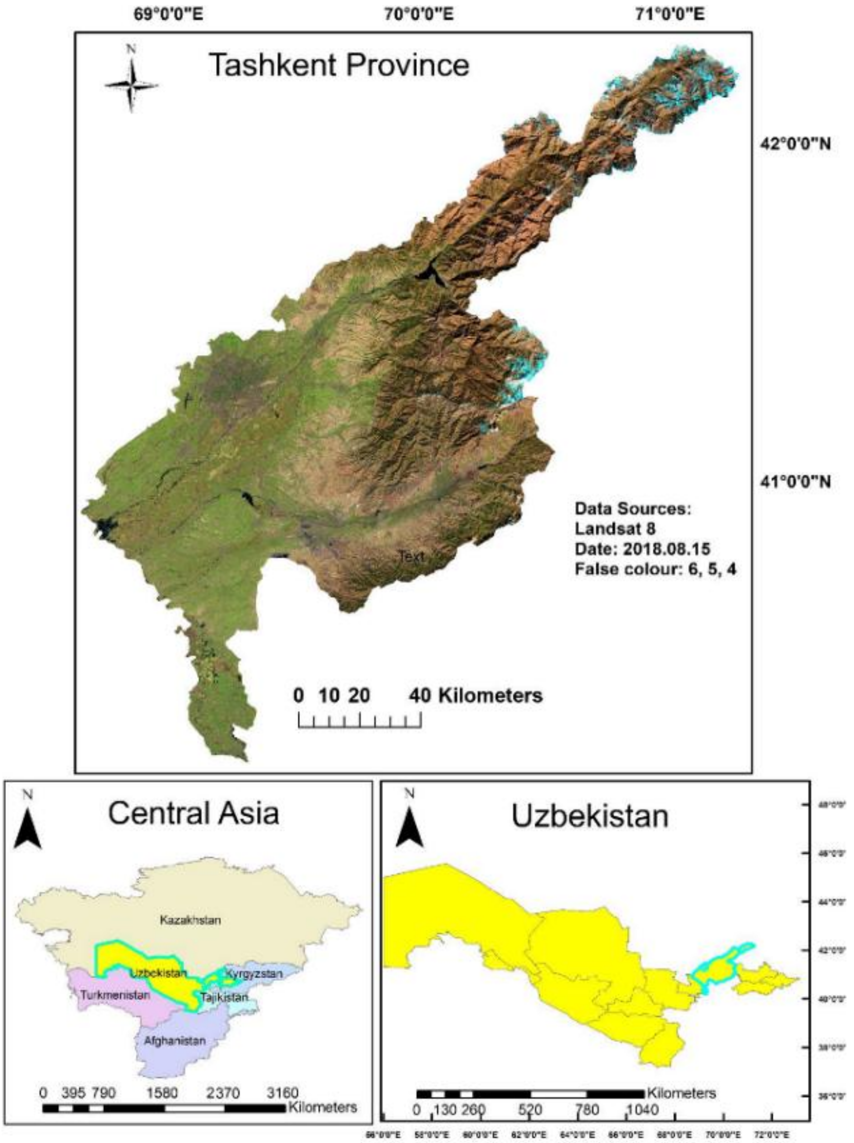

Fig. 1. The study area location map [Erdanaev et al., 2022]

The study categorizes the landscape of the Tashkent Region into six major Land Use/Land Cover (LULC) classes on their distinct spectral, spatial, and environmental characteristics. These classes include urban areas, croplands, bare land, forests, water bodies, and snow cover. Each category represents a significant component of the region's land dynamics, influenced by urbanization, agricultural practices, natural ecosystems, and climatic variations. The classification provides insights into land cover distribution, resource management, and environmental sustainability. The table below summarizes the key LULC classes, their descriptions, and their environmental significance.

Table 1. Land Use/Land Cover (LULC) Classes in the Tashkent Region

LULC Class	Description	Key Characteristics	Environmental Importance
Urban Areas	Built-up environments include residential, commercial, and industrial zones	High-density buildings, paved surfaces, minimal vegetation	It affects local climate (heat island effect), reduces permeable surfaces, and impacts air quality
Croplands	Agricultural fields cultivated with various seasonal and perennial crops	Dominated by wheat, cotton, and horticultural crops; irrigated and rain-fed agriculture	Essential for food security, affected by climate and land-use policies
Bare land	Unvegetated surfaces including exposed soil, rocky areas, and degraded lands	Found in construction zones, deforested areas, and eroded landscapes	Indicator of land degradation and soil erosion; potential for land reclamation
Forests	Natural and plantation forests provide ecological and economic benefits	Found mainly in mountainous regions; composed of juniper, pine, and mixed deciduous trees	Helps in carbon sequestration, biodiversity conservation, and soil erosion prevention
Water Bodies	Lakes, rivers, reservoirs, and irrigation channels critical for the region's hydrology	Includes Chirchiq River, Charvak Reservoir, and artificial irrigation networks	Supports agriculture, drinking water supply, and hydropower generation
Snow Cover	Seasonal or permanent snow-covered surfaces, mainly in high-altitude areas	Observed during winter months; fluctuates with climate variability	Important for water resource availability, glacier monitoring, and climate change studies

Satellite data

This study utilized Landsat-8 Operational Land Imager (OLI) imagery, a widely used satellite dataset for remote sensing applications. Landsat-8 OLI provides high-quality multispectral data suitable for Land Use/Land Cover (LULC) classification. The key specifications of the Landsat-8 OLI sensor are as follows:

- spatial resolution: 30 meters for multispectral bands;
- temporal resolution is 16-day revisit cycle, ensuring regular data updates;
- spectral coverage has contained 11 bands, including visible, near-infrared (NIR), short-wave infrared (SWIR), and thermal infrared (TIR) bands, allowing for detailed land cover differentiation.

Landsat-8 data and scene selection

Landsat-8 Collection 2 Level 2 surface reflectance scenes covering the Tashkent study area were retrieved from Google Earth Engine (GEE). Images were filtered to the period 2024-06-01 to 2024-07-30 and constrained to scenes with cloud cover <30 %, resulting in 12 scenes (Table 2). Because the scenes partially overlap (adjacent WRS path/row tiles), a per-pixel median composite of the filtered images was computed and used as the input raster for classification. The June–July window corresponds to the regional peak growing season, maximizing spectral separability of major classes (cropland, forest, grassland). Median compositing reduces residual cloud/shadow effects and radiometric noise; however, compositing across a multi-week window can smooth short-term phenological differences (e. g., early wheat harvest vs. later cotton growth).

Table 2. Landsat-8 scenes used to build the June–July 2024 median composite (Tashkent Region)

No.	Landsat_Scene_ID	Acquisition_Date	Cloud_Cover (%)
1	LC08_153031_20240613	2024-06-13	5.02
2	LC08_153031_20240629	2024-06-29	12.19
3	LC08_153032_20240613	2024-06-13	7.35
4	LC08_153032_20240629	2024-06-29	6.48
5	LC08_154031_20240604	2024-06-04	1.76
6	LC08_154031_20240620	2024-06-20	2.21
7	LC08_154031_20240706	2024-07-06	3.35
8	LC08_154031_20240722	2024-07-22	0.41
9	LC08_154032_20240604	2024-06-04	20.18
10	LC08_154032_20240620	2024-06-20	7.97
11	LC08_154032_20240706	2024-07-06	2.17
12	LC08_154032_20240722	2024-07-22	1.01

Classification algorithms

In this study, two widely used machine learning algorithms — Random Forest (RF) and Support Vector Machine (SVM) — were applied for Land Use/Land Cover (LULC) classification. Both algorithms are non-parametric, capable of handling high-dimensional datasets, and have been extensively validated in remote sensing applications [*Huang* et al., 2002; *Mountrakis* et al., 2011; *Arpitha* et al., 2023].

RF is an ensemble-based algorithm that constructs a large number of decisions trees and aggregates their results through majority voting. Each tree is trained using a bootstrap sample of the training dataset, while at each node only a random subset of predictor variables is considered. This mechanism reduces overfitting and enhances generalization ability [*Gislason* et al., 2006; *Rodriguez-Galiano* et al., 2012; *Belgiu*, *Drăguţ*, 2016]. The classification in RF is determined using the majority voting approach (1):

$$f(x) = majority\ vote\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (1),$$

where y — the predicted class label,
$h_i(x)$ —the output of the i^{th} decision tree,
n — is the total number of trees in the forest [*Breiman*, 2001].

RF is particularly effective in handling noisy data and complex class distributions, making it suitable for heterogeneous landscapes such as the Tashkent Region. RF has become one of the most commonly used classifiers due to its ensemble learning approach, where multiple decision trees are constructed and combined to produce more stable results. Its main advantages include robustness to overfitting, tolerance to irrelevant variables, and computational efficiency. RF also provides variable importance measures, allowing researchers to assess which bands or indices contribute most to classification. However, a known limitation is that RF may be biased toward classes with larger sample sizes, and in some cases, it may misclassify minority land cover types if the training dataset is imbalanced.

Support Vector Machine (SVM) is a supervised learning algorithm that classifies data by finding the optimal hyperplane that maximizes the margin between different land cover classes. SVM is particularly effective for small sample sizes and high-dimensional data [*Cortes*, *Vapnik*, 1995; *López* et al., 2022]. For a given dataset (x_i, y_i) where x_i is the feature vector and y_i is the class label, SVM (Kernelized SVM) aims to solve the following optimization problem (2):

$$f(x) = sign(\sum_{i=1}^{n} a_i y_i K(x_i, x) + b) \qquad (2),$$

where $K(x_i, x)$ — the kernel function (e. g., Radial Basis Function, Polynomial).

In this study, we used RBF kernel, which is (3):

$$K(x_i, x) = \exp\,(-\gamma \left|\left|x_i - x_j\right|\right|^2) \qquad (3).$$

SVM, on the other hand, excels in cases with smaller training datasets and complex class distributions. By using kernel functions, such as the Radial Basis Function (RBF), SVM can project data into higher-dimensional spaces where classes become linearly separable. This makes SVM particularly effective for distinguishing spectrally similar land cover classes (e. g., croplands vs. forests). Nevertheless, SVM requires careful parameter tuning, specifically the regularization parameter (C) and kernel coefficient (γ), and may become computationally intensive for very large datasets. Despite these challenges, numerous studies have demonstrated its superior accuracy in remote sensing classification tasks.

By employing both RF and SVM, this study not only compares their classification performance but also highlights their respective strengths and limitations. The dual application allows for a more comprehensive evaluation of LULC mapping accuracy in the study area and contributes to methodological transparency for future research.

Accuracy assessment

To evaluate the performance of the classification results, an accuracy assessment was conducted using a confusion matrix derived from the validation dataset. The assessment included the calculation of Overall Accuracy (OA), Kappa Coefficient (KC) for each land cover class. These metrics provide a comprehensive understanding of classification reliability by comparing predicted classes with ground truth labels. To ensure accurate classification and validation, ground truth data were collected from high-resolution Google Earth imagery and field observations. More than 700 sample points were selected and manually classified into six LULC categories. The dataset was divided into 70 % for training and 30 % for validation, ensuring a balanced and statistically significant assessment of classification accuracy. This evaluation ensures the robustness and generalizability of the classification models.

Software and tools

All classification experiments, including Random Forest (RF) and Support Vector Machine (SVM), were carried out in the Google Earth Engine (GEE) cloud platform, which provides access to the full Landsat archive and offers built-in machine learning algorithms (ee.Classifier.smile RandomForest() and ee.Classifier.libsvm()). This enabled efficient preprocessing, model training, accuracy assessment and sampling processes directly within a cloud-based environment, ensuring reproducibility and eliminating the need for local data storage or high-performance computing resources. However, GEE's implementations of RF and SVM have limited parameter tuning compared to Python libraries (e. g., scikit-learn), which might constrain optimization for specific datasets [*Sultan* et al., 2025]. Built-in preprocessing does not automatically normalize reflectance in mountainous areas, which may affect classification accuracy in regions like northeastern Tashkent [*Zhou* et al., 2016]. While GEE is powerful for analysis, final visualization and figure preparation required licensed ArcGIS Pro, adding a dependency on commercial software.

For visualization purposes, the classified raster outputs generated in GEE were exported and further processed in ArcGIS Pro (licensed version). ArcGIS was only used for cartographic representation, map layouts, and preparation of figures for publication. All accuracy statistics, including confusion matrices, Overall Accuracy (OA), and Kappa Coefficient (KC), were computed inside GEE. The combination of GEE and ArcGIS ensured both computational efficiency and high-quality visualization. While GEE offers robust scalability for processing large datasets, one limitation is the restricted flexibility for hyperparameter tuning compared to local Python-based workflows.

RESEARCH RESULTS AND DISCUSSION

The Land Use/Land Cover (LULC) classification for the Tashkent Region was performed using Landsat-8 imagery acquired between June and July 2024. Two machine learning algorithms — Support Vector Machine (SVM) and Random Forest (RF) — were implemented and evaluated for classification accuracy. Six major land cover classes were identified: Water, Bare land, Built-up Area, Cropland, Snow, Forest. Ground truth samples were used for supervised classification, and the models were trained using 70 % of the samples, while the remaining 30 % were reserved for validation. The performance of each model was assessed using confusion matrices, overall accuracy (OA), and the kappa coefficient (KC).

The SVM classifier achieved the highest performance with an Overall Accuracy of 96.78 % and a Kappa Coefficient of 0.9604. The Random Forest classifier also performed well, with an Overall Accuracy of 94.95 % and a Kappa Coefficient of 0.9377. These results indicate a strong agreement between the classified outputs and the reference data, with SVM showing slightly better reliability.

Table 3. Overall Accuracy (OA), and Kappa Coefficient (KC) for the classifiers

Metric	RF	SVM
Overall Accuracy	94.95 %	96.78 %
Kappa Coefficient	93.77 %	96.04 %

Confusion matrix analysis revealed that SVM produced fewer misclassifications, particularly in complex classes such as Cropland, and Forest. In contrast, the RF classifier showed slightly higher confusion between Cropland and Forest, as reflected in its error matrix. The higher performance of the SVM classifier can be attributed to its ability to handle high-dimensional data and perform well in cases where class boundaries are not linearly separable. RF, though slightly less accurate, remains a robust classifier known for its efficiency and capability to handle large datasets.

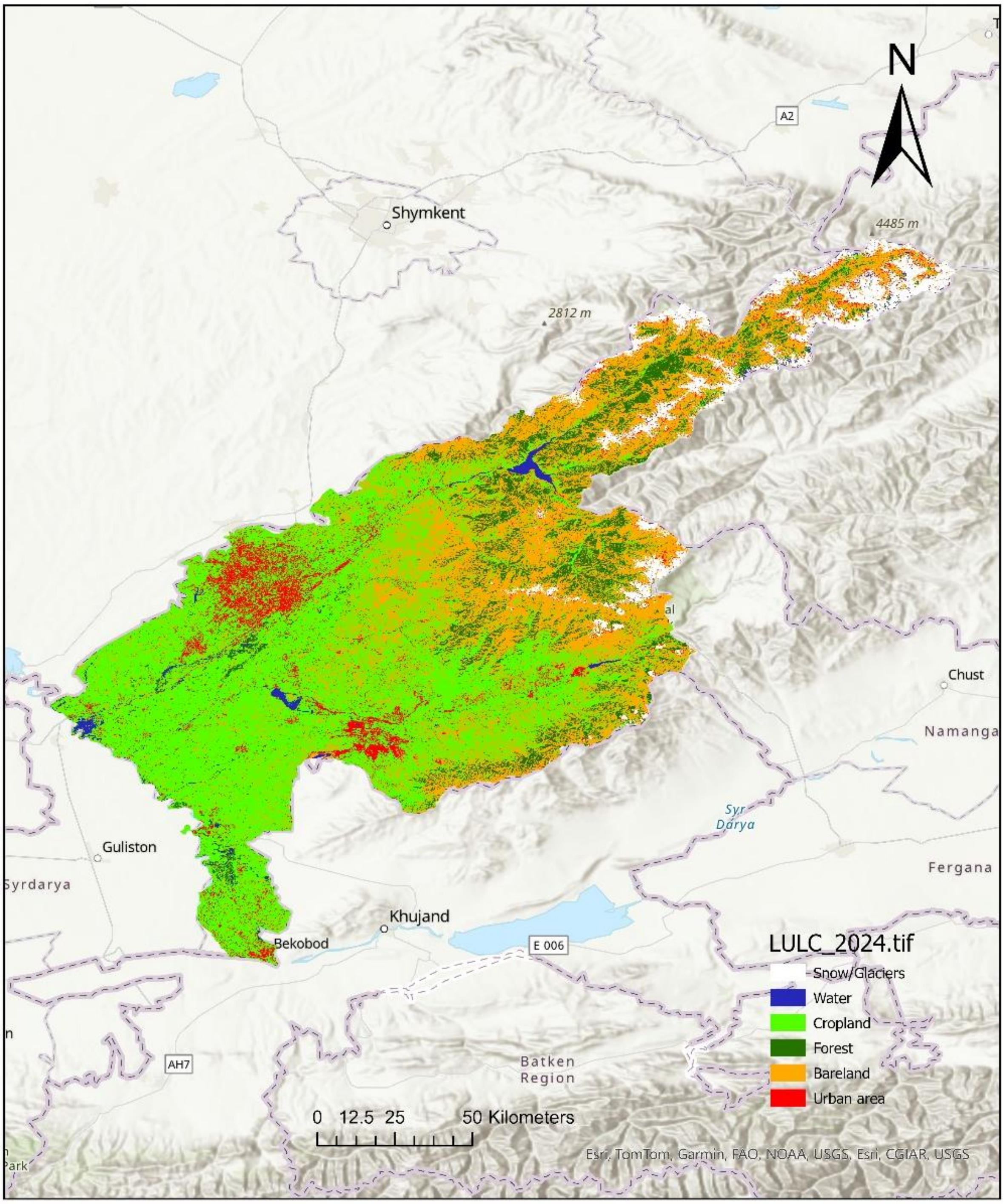

Fig. 2. Land Use/Land Cover (LULC) classification maps of the Tashkent Region generated using Landsat-8 imagery (June 2024). The classifications were performed using Random Forest (RF) algorithm. The map shows seven major classes: Water, Bare land, Built-up Area, Cropland, Snow, Forest

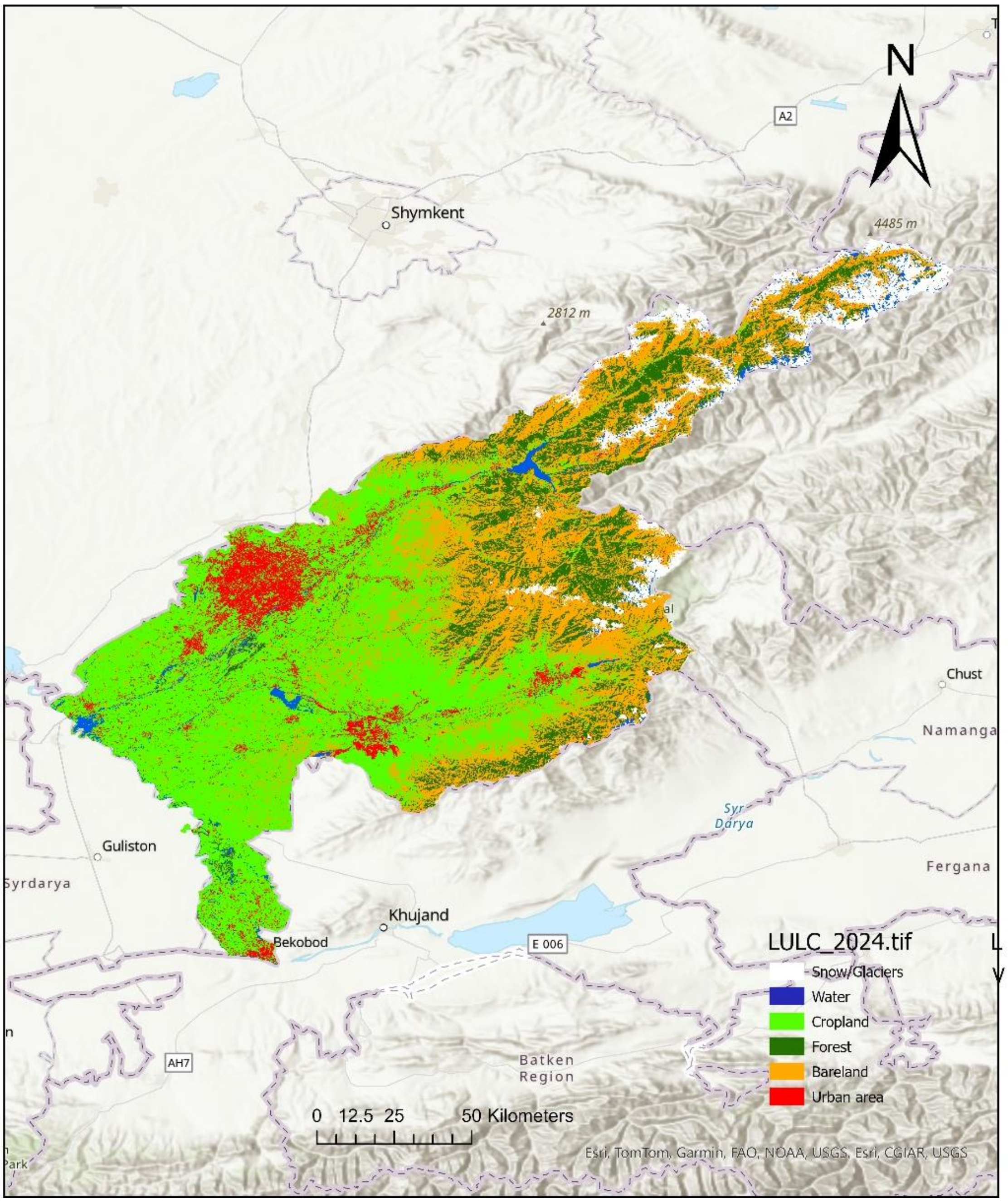

Fig. 3. Land Use/Land Cover (LULC) classification maps of the Tashkent Region generated using Landsat-8 imagery (June 2024). The classifications were performed using Support Vector Machine (SVM)

Table 4. Confusion Matrix for the Support vector machine (SVM) classifier

Actual/Predicted	**Water**	**Bare land**	**Built-up**	**Cropland**	**Snow**	**Forest**
Water	14	0	0	0	0	0
Bare land	0	22	0	0	0	0
Built-up	0	0	43	1	2	2
Cropland	0	0	0	47	0	0
Snow	0	0	0	0	39	0
Forest	0	0	0	0	2	46

Table 5. Confusion Matrix for the Random Forest (RF) classifier

Actual/Predicted	Water	Bare land	Built-up	Cropland	Snow	Forest
Water	14	0	0	0	0	0
Bare land	0	22	0	0	0	0
Built-up	0	0	45	1	2	0
Cropland	0	0	0	47	0	0
Snow	0	0	5	0	34	0
Forest	0	0	2	0	1	45

CONCLUSIONS

The accurate classification of Land Use/Land Cover (LULC) is a foundational component of environmental management, agricultural monitoring, and urban planning. In this study, we assessed and compared the effectiveness of two well-established machine learning algorithms — Support Vector Machine (SVM) and Random Forest (RF) — for classifying LULC types using Landsat-8 OLI satellite imagery for the summer season of 2024 in the Tashkent Region of Uzbekistan. The methodology integrated a comprehensive workflow that included preprocessing steps such as filtering for cloud-free images, selecting relevant spectral bands, clipping to the study region, and using a well-prepared set of ground truth samples representing six dominant LULC classes: water, bare land, built-up areas, croplands, forest, and snow. Classification performance was assessed using confusion matrices, and metrics such as Overall Accuracy (OA) and Kappa Coefficient (KC) were used to evaluate model robustness and reliability.

The results demonstrated that both classifiers provided high classification performance, reinforcing the applicability of machine learning techniques in satellite-based land cover mapping. However, the Support Vector Machine (SVM) classifier outperformed the Random Forest (RF) algorithm, achieving an OA of 96.78 % and a Kappa Coefficient of 0.96, compared to RF's OA of 94.95 % and KC of 0.93. The SVM algorithm showed particularly strong accuracy in distinguishing between spectrally similar classes, such as cropland and forest, which are often difficult to separate using traditional classification techniques. These findings suggest that SVM is more suitable for LULC classification in regions with complex and heterogeneous landscapes, such as the Tashkent Region. The SVM's capacity to work well with high-dimensional input data and its ability to define optimal class boundaries using kernel functions played a key role in delivering superior results. While Random Forest remains a powerful and interpretable classifier with good performance, especially in handling large datasets and noisy variables, it was slightly less precise in this particular case study.

Another important takeaway from this study is the importance of high-quality training data. The performance of both classifiers was heavily dependent on the representativeness and spatial distribution of the training samples. The integration of accurate ground truth data allowed for the development of reliable models, reducing the risk of misclassification and enhancing the generalizability of the results. In addition, the study demonstrated the utility of Google Earth Engine (GEE) as a powerful cloud-based platform for scalable and efficient satellite image analysis. GEE enabled the automation of the workflow, seamless integration of various data sources, and rapid processing of large-scale remote sensing datasets, which is particularly beneficial for research and operational land monitoring in developing regions.

The implications of this study are significant for land managers, policymakers, urban planners, and environmental researchers. High-accuracy LULC maps can support a wide range of applications, including land resource management, monitoring agricultural expansion, deforestation tracking, urban growth assessment, and the evaluation of climate impacts on land systems.

In conclusion, this study provides empirical evidence of the effectiveness of machine learning algorithms in satellite-based LULC classification. The SVM classifier demonstrated strong potential for delivering reliable and precise land cover maps, which can play a pivotal role in supporting sustainable land management strategies and evidence-based decision-making in Uzbekistan and similar regions. As remote sensing technology and machine learning methods continue to advance, their integration will become increasingly vital for addressing pressing environmental and socio-economic challenges in the face of rapid global change.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the United States Geological Survey (USGS) for providing open access to the Landsat-8 satellite imagery and the Google Earth Engine (GEE) platform for enabling large-scale image processing, classification, and analysis. We also thank Tashkent Irrigation and Agricultural Mechanization Engineers National Research University for academic and technical support during this research. The use of ArcGIS Pro software for visualization is acknowledged as part of the institutional license. The constructive comments and suggestions from the reviewers are sincerely appreciated, as they have significantly improved the clarity and quality of this manuscript.

REFERENCES

Alikhanov B. S., Alikhanova R., Oymatov Z. F., Pulatov A. Land Cover Change in Tashkent Province during 1992–2018. IOP Conference Series: Materials Science and Engineering, 2020. V. 883. Art. 012088. DOI: 10.1088/1757-899X/883/1/012088.

Alikhanov B., Pulatov B., Samiev L. The Detection of Past and Future Land Use and Land Cover Change in Ugam Chatkal National Park, Uzbekistan, Using CA-Markov and Random Forest Machine Learning Algorithms. Forum Geografi, 2024. V. 38. No. 2. P. 121–137. DOI: 10.23917/forgeo.v38i2.4221.

Arpitha M., Sohrab A., Harishnaika N. Land Use and Land Cover Classification Using Machine Learning Algorithms in Google Earth Engine. Earth Science Informatics, 2023. V. 16. P. 3057–3073. DOI: 10.1007/s12145-023-01073-w.

Aslanov I., Mukhtorov U., Mahsudov R., Makhmudova U., Alimova S., Djurayeva L., Ibragimov O. Applying Remote Sensing Techniques to Monitor Green Areas in Tashkent Uzbekistan. E3S Web of Conferences, 2021. V. 258. Art. 04012. DOI: 10.1051/e3sconf/202125804012.

Aslanov I., Teshaev N., Jabbarov Z., Opp C., Oymatov R., Karimov Yu., Henebry G. Characterizing Land Surface Dynamics in Aral Sea Basin of Uzbekistan Using Climatic and Remote Sensing Data to Project Future Conditions. E3S Web of Conferences, 2024. V. 575. DOI: 10.1051/e3sconf/202457504009.

Aslanov I., Teshaev N., Khayitov Kh., Mukhtorov U., Khaitbaeva J., Murodova D. Analysis of Desertification Trends in Central Asia Based on MODIS Data Using Google Earth Engine. E3S Web of Conferences, 2023. V. 443. Art. 06015. DOI: 10.1051/e3sconf/202344306015.

Belgiu M., Drăguţ L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. ISPRS Journal of Photogrammetry and Remote Sensing, 2016a. V. 114. P. 24–31. DOI: 10.1016/j.isprsjprs.2016.01.011.

Bhungeni O., Ashadevi R., Gebreslasie M. Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal. Remote Sensing, 2024. V. 16. Art. 2219. DOI: 10.3390/rs16122219.

Breiman L. Random Forests. Machine Learning, 2001. V. 45. Iss. 1. P. 5–32. DOI: 10.1023/A:1010933404324.

Cortes C., Vapnik V. Support-Vector Networks. Machine Learning, 1995. V. 20. Iss. 3. P. 273–297. DOI: 10.1007/BF00994018.

Dash S. S., Maity R. Effect of Climate Change on Soil Erosion Indicates a Dominance of Rainfall over LULC Changes. Journal of Hydrology: Regional Studies, 2023. V. 47. Art. 101373. DOI: 10 .1016/j.ejrh.2023.101373.

Elhadi A., Mutanga O., Rugege D. Multispectral and Hyperspectral Remote Sensing for Identification and Mapping of Wetland Vegetation: A Review. Wetlands Ecology and Management, 2010. V. 18. Iss. 3. P. 281–296. DOI: 10.1007/s11273-009-9169-z.

Erdanaev E., Kappas M., Wyss D. Irrigated Crop Types Mapping in Tashkent Province of Uzbekistan with Remote Sensing-Based Classification Methods. Sensors, 2022. V. 22. Iss. 15. Art. 5863. DOI: 10.3390/s22155683.

Farmonov N., Amankulova Kh., Khan Sh. N., Abdurakhimova M., Szatmári J., Khabiba T., Makhliyo R., Khodicha M., Mucsi L. Effectiveness of Machine Learning and Deep Learning Models at County-Level Soybean Yield Forecasting. Hungarian Geographical Bulletin, 2023. V. 72. No. 4. P. 383–398. DOI: 10.15201/hungeobull.72.4.4.

Farmonov N., Esmaeili M., Abbasi-Moghadam D., Sharifi A., Amankulova Kh., Mucsi L. Hyps-LiDNet: 3D-2D CNN Model and Spatial-Spectral Morphological Attention for Crop Classification with DESIS and LiDAR Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024. P. 1–34. DOI: 10.1109/JSTARS.2024.3418854.

Gafforov Kh. Sh., Bao A., Rakhimov Sh., Liu T., Abdullaev F., Jiang L., Durdiev Kh., Duulatov E., Rakhimova M., Mukanov Ye. The Assessment of Climate Change on Rainfall-Runoff Erosivity in the Chirchik-Akhangaran Basin, Uzbekistan. Sustainability, 2020. V. 12. Iss. 8. Art. 3369. DOI: 10 .3390/su12083369.

Giles M. F. Status of Land Cover Classification Accuracy Assessment. Remote Sensing of Environment, 2002. V. 80. Iss. 1. P. 185–201. DOI: 10.1016/S0034-4257(01)00295-4.

Gislason P. O., Benediktsson J. A., Sveinsson J. R. Random Forests for Land Cover Classification. Pattern Recognition Letters, Pattern Recognition in Remote Sensing (PRRS 2004), 2006. V. 27. Iss. 4. P. 294–300. DOI: 10.1016/j.patrec.2005.08.011.

Huang C., Davis L. S., Townshend J. R. G. An Assessment of Support Vector Machines for Land Cover Classification. International Journal of Remote Sensing, 2002. V. 23. Iss. 4. P. 725–49. DOI: 10.1080/01431160110040323.

Korhonen L., Hadi, Packalen P., Rautiainen M. Comparison of Sentinel-2 and Landsat-8 in the Estimation of Boreal Forest Canopy Cover and Leaf Area Index. Remote Sensing of Environment, 2017. V. 195. P. 259–74. DOI: 10.1016/j.rse.2017.03.021.

López O. A. M., López A. M., Crossa J. Support Vector Machines and Support Vector Regression. Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer Nature, 2022. P. 337–378. DOI: 10.1007/978-3-030-89010-0_9.

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. V. 39. Iss. 9. P. 2784–2817. DOI: 10.1080/01431161.2018.1433343.

Mirzaei A., Hossein B., Iman Kh. Enhancing Crop Classification Accuracy through Synthetic SAR-Optical Data Generation Using Deep Learning. ISPRS International Journal of Geo-Information, 2023. V. 12. Iss. 11. Art. 450. DOI: 10.3390/ijgi12110450.

Mountrakis G., Im J., Ogole C. Support Vector Machines in Remote Sensing: A Review. ISPRS Journal of Photogrammetry and Remote Sensing, 2011. V. 66. Iss. 3. P. 247–259. DOI: 10.1016/j.isprsjprs.2010.11.001.

Pal M. Random Forest Classifier for Remote Sensing Classification. International Journal of Remote Sensing, 2005. V. 26. Iss. 1. P. 217–222. DOI: 10.1080/01431160412331269698.

Rodriguez-Galiano V. F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J. P. An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification. ISPRS Journal of Photogrammetry and Remote Sensing, 2012a. V. 67. P. 93–104. DOI: 10.1016/j.isprsjprs.2011.11.002.

Sajib Md. Q. U., Wang T. Estimation of Land Surface Temperature in an Agricultural Region of Bangladesh from Landsat-8: Intercomparison of Four Algorithms. Sensors, 2020. V. 20. Iss. 6. Art. 1778. DOI: 10.3390/s20061778.

Schmidt M., Pringle M., Devadas R., Denham R., Tindall D. A Framework for Large-Area Mapping of Past and Present Cropping Activity Using Seasonal Landsat Images and Time Series Metrics. Remote Sensing, 2016. V. 8. Iss. 4. Art. 312. DOI: 10.3390/rs8040312.

Sharipov Sh., Khayitmurodov A. The Impacts of Green Spaces on Mitigating the Urban Hot Island Effect in the City of Tashkent. BIO Web of Conferences, 2024. V. 105. Art. 06013. DOI: 10.1051/bioconf/202410506013.

Sultan M., Saleous N., Issa S., Dahy B., Sami M. Optimizing Land Use Classification Using Google Earth Engine: A Comparative Analysis of Machine Learning Algorithms. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2025. V. X-G-2025. P. 863–869. DOI: 10.5194/isprs-annals-X-G-2025-863-2025.

Teshaev N., Mamadaliyev B., Ibragimov A., Khasanov S. The Soil-Adjusted Vegetation Index for Soil Salinity Assessment in Uzbekistan. InterCarto. InterGIS. Proceedings of the International Conference. Moscow: Lomonosov Moscow State University, Faculty of Geography, 2020. V. 26. Part 3. P. 324–333. DOI: 10.35595/2414-9179-2020-3-26-324-333.

Tian Sh., Zhang X., Tian J., Sun Q. Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. Remote Sensing, 2016. V. 8. Iss. 11. Art. 954. DOI: 10.3390/rs8110954.

Vapnik V. N. Introduction: Four Periods in the Research of the Learning Problem. The Nature of Statistical Learning Theory. Springer Nature, 2000. P. 1–15. DOI: 10.1007/978-1-4757-3264-1_1.

Wei P., Ye H., Qiao Sh., Liu R., Nie Ch., Zhang B., Song L., Huang Sh. Early Crop Mapping Based on Sentinel-2 Time-Series Data and the Random Forest Algorithm. Remote Sensing, 2023. V. 15. Iss. 13. Art. 3212. DOI: 10.3390/rs15133212.

Zhang Ch., Dong J., Xie Ya., Zhang X., Ge Q. Mapping Irrigated Croplands in China Using a Synergetic Training Sample Generating Method, Machine Learning Classifier, and Google Earth Engine. International Journal of Applied Earth Observation and Geoinformation, 2022. V. 112. Art. 102888. DOI: 10.1016/j.jag.2022.102888.

Zhou Ya., Jiang H., Wang Zh., Yang X., Geng E. Assessment of Four Typical Topographic Corrections in Landsat TM Data for Snow Cover Areas. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016. V. XLI-B1. P. 157–162. DOI: 10.5194/isprs-archives-XLI-B1-157-2016.