Journal of Geomatics Science and Technology, vol. 13, no. 4, pp. 57-66, June 2024 - Research article

A hybrid approach to urban growth assessment using K-Nearest Neighbour, Support Vector Machine, Random Forest, and Maximum Likelihood (Case study: West Tehran)

Hossein Joulaei[1*], Alireza Vafaeinajad[2], Mehrnoush Sharifzadeh[3]

[1] MSc Student in Land Administration Systems, Faculty of Civil, Water, and Environmental Engineering, Shahid Beheshti University (SBU), Tehran, Iran, hossein.joulaei98@gmail.com

[2] Assistant Professor Geographic Information System, Faculty of Civil, Water, and Environmental Engineering, Shahid Beheshti University (SBU), Tehran, Iran, a_vafaei@sbu.ac.ir

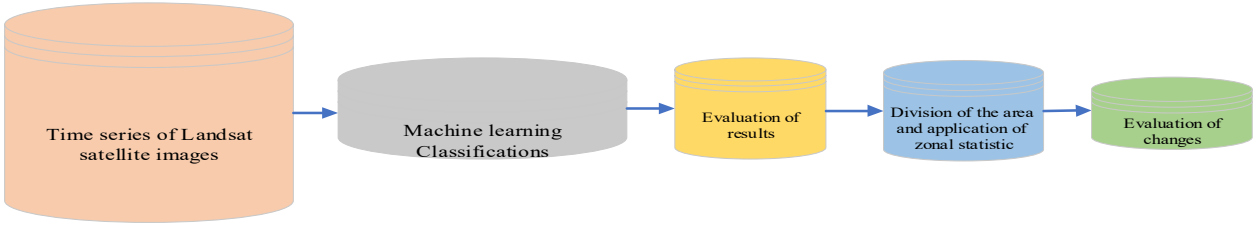
[3] MSc Student in Photogrammetry, Faculty of Civil, Water, and Environmental Engineering, Shahid Beheshti University (SBU), Tehran, Iran, m.sharifzadeh@sbu.ac.ir

(Received: December 2023, Accepted: February 2024)

Abstract

Urbanization is a growing concern, and satellite images play a crucial role in assessing urban growth. The availability and time series characteristics of satellite images make them a powerful tool for evaluating changes in phenomena. To begin working with satellite images, it is necessary to take samples and classify the images according to the region's complications. Using a specific classification method for time series of images may not produce accurate results to evaluate the changes in a phenomenon, and much depends on the dispersion of the samples taken from the images. In this study, 4 machine learning algorithms (K-Nearest Neighbour, Support Vector Machine, Random Forest (Random Trees), and Maximum Likelihood) were used to classify images from three periods of Landsat satellite imagery (Landsat 7, 8, 9) at two 10-year intervals (2003, 2013, and 2023). In four areas of Tehran (2, 5, 21, 22), this has been applied to urban growth. For the classification results, accuracy and Kappa coefficient were used. Therefore, Using the KNN method with a Kappa coefficient of 91%, Landsat image 7 performed best due to the uniformity of the samples. Additionally, Landsat images 8 and 9 were successfully analysed with the SVM method with an accuracy of 97% and 94%, respectively, as well as a Kappa coefficient of 95% and 89%. Urban growth is also evaluated using selected methods for each image. For this purpose, to evaluate the changes in a specific area, the study area is divided into equal parts, and using zonal statistics, the area of each element of the changes is applied to the divided areas. Therefore, between 2003 and 2013, urban growth was 10%, between 2013 and 2023, it was 24%, and as a result, between 2023 and 2003, it was 34%. Additionally, we examine the change in barren and green lands in this study. Our study offers the most accurate hybrid approach to image classification for urban growth, and it can provide valuable information to urban planners and policymakers for managing urban growth and promoting sustainable development in cities.

Keywords: Machine learning algorithms, Classification, Urben growth, KNN, SVM, RF, MLC



Graphical Abstract

* Corresponding author

1- Introduction

Urbanization has become one of the major concerns of early societies in recent years. A number of recent studies have demonstrated that urban environments can also create significant inequalities, despite the fact that urbanization has long been associated with human progress and development [1]. In 1950, less than one-third of the world's population lived in urban areas, which is expected to increase to nearly two-thirds by 2050[2]. Rapid population growth, the expansion of road networks, urbanization, and social and economic inequalities all contribute to urbanization [3]. As a result of urbanization, in addition to a lack of public use in allocating resources and providing services, there is a loss of habitat and a change in land use [3], [4]. Thus, it is essential to pay particular attention to these issues and use all the tools available for managing and developing urbanization. In today's society, remote sensing and GIS provide a means to identify and monitor changes over time [5]. In these images, urban growth can be identified among other phenomena [6]. In recent years, machine learning algorithms have emerged that are capable of classifying and predicting future events based on these images, which has made this a topic of interest to researchers [7]– [9]. For this purpose, machine learning algorithms differ depending on the quality of the data obtained and the type of algorithm employed [7]. Machine learning algorithms commonly used include artificial neural networks, support vector machines, and random forests [10]. In recent years, a great deal of research has been done on remote sensing and machine learning methods of classifying images or assessing the growth and development of a phenomenon as it occurs over time.

Andreas Rienow et al calibrated Cellular Automata (CA) models using RF and SVM classifiers. Both RF and SVM classifiers improved urban sprawl simulations. RF and SVM-based CA models perform better than other classification methods [6]. For remote sensing image classification, Sheikhmuso et al developed RF and SVM. Precision and accuracy were assessed in their research. RF has superior accuracy and low variance when quantifying land and cover changes with low-perspective images. They concluded that SVM performed better on datasets with large numbers of features. Further, RF performs better at classifying hyperspectral and multispectral images [7]. Yuho et al utilized Landsat 8 images and Landsat ETM+ satellite data for the analysis of land cover change and land use change, utilizing KNN, SVM, ANN, and RF algorithms. According to their results, the RF model was 90% more accurate than other classification models [8].

According to Gilbert and Shi, a study measured and projected changes in land cover and land use. Using Landsat 7 satellite images, populations were classified and correlations were found. In their study, the population increased by 86.24 percent between 2002 and 2012, 53.87 percent between 2012 and 2022, and 2033 by 2033[11]. Mustaquim and Woheeul Islam examined land cover and land use from 1991 to 2021. A neural network, automated Markov cells, and a classification algorithm (MLC) were used to predict the landscape in 2031. The imagery was taken from Landsat 5TM, Landsat 8 OLI/TIRS, and Landsat 9[12]. Aplan et al used Random Forest (RF), KNN (K-Nearest Neighbours), and SVM (Support Vector Machine). Each classification was ranked based on its overall accuracy with SVM, RF, and KN classifiers [13].

Sadegh Khalifa Hanoun has developed a method combining dynamic land use and land cover change (LULCCs) with Sustainable Development Goals (SDGs). This method uses multiple machine learning (ML) techniques, such as random forests (RF) and artificial neural networks (ANN) with automated cells (CA), as well as ensemble models based on (KNN), RF, and logistic regression (LR) algorithms. In this report, they presented a forecast for urban growth through 2092[14]. Abd al-Qadir Rash et al performed a land use/land cover change (LULC) analysis using multi-temporal Landsat images. Several machine learning algorithms were used including SVM, RF, ANN, KNN, and XGBoost. Their results showed that the RF algorithm produces a high Kappa coefficient (0.93-0.97), which outperforms other algorithms (0.91-0.96) [15]. A study conducted by Vikas Kumar Rana et al used Sentinel 2 satellite images, maximum likelihood, random forest, and support vector machines to classify the images. Using a random sampling method, they evaluated the accuracy of the obtained maps. For calculating performance, they used criteria such as the Kappa coefficient, overall accuracy, producer's accuracy, and user's accuracy. Based on the results of their study, they found that

support vector machine algorithms resulted in accurate maps [16]. Hemant Singh Pokhariya evaluated four different classification algorithms using Sentinel-2 imagery: random forest (RF), support vector machine (SVM), decision tree (DT), and CART. According to this study, algorithm accuracy varies with sample conditions for land cover classification [17]. Ghairohamkaran et al. Compare five surveillance classification algorithms for producing land cover and land use maps based on Sentinel 2 and Landsat 8 satellite imagery. As a result of optimizing the algorithms, they chose SVM as the most effective method for classification [18].

Most studies evaluate several algorithms to determine the most effective algorithm choose one of them and utilize the selected algorithm for classification in all applications, such as forecasting and modelling the effects of urbanization and land use changes. We point out that these algorithms can have different results depending on the type of data, the dispersion of samples, and the changes in a phenomenon like urban growth over time. In this research, we used four machine learning algorithms on Landsat images to evaluate urban growth. Based on each image, we checked four algorithms in terms of accuracy and selected the most accurate method for each image.

2- Case Study

To evaluate urban growth, districts 2, 5, 21, and 22 of Tehran (Iran) were chosen as a case study. These areas include Tehran's northwest and west and its area is about 215 KM2.

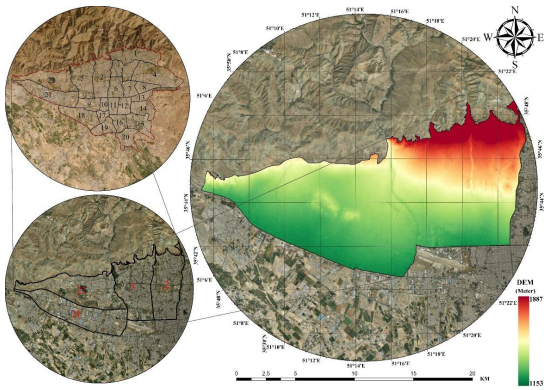


Figure 1. The studied area along with 22 areas and the digital elevation model (DEM)

The growth of the city of Tehran has shifted to the west due to an increase in population and other economic and social factors. In the last census, the four districts had a combined population of 1,922,817 people, and the DEM indicates the highest altitude in the northern regions, while the altitude decreases gradually towards the south and west. There is no difference in the slope of the regions except in the northern region and Region 2. In addition, longitude varies from 51º06' to 51º23'N and latitude from 35º41' to 35º49'N (Fig. 1).

3- Methods

As part of this study, four general steps will be followed using the ArcGIS Pro and Google Earth Pro software, as well as Python programming language. As a first step, satellite images are collected from the USGS site and the necessary corrections are made (Data preparation). As a second step, machine learning algorithms are used to classify images (Classification using machine learning algorithms). As a third step, the accuracy of all methods used on each image is evaluated, and the best image and the most accurate method are selected (Accuracy evaluation and selection of the most accurate method). As a fourth step (The result section), to better understand the urban growth of the study area, it is divided into equal areas, and the results are joined to all these parts in step 3 (Divided of the study area and zonal Statistic) (Fig. 2).

3-1- Data preparation

The chosen satellite images for the study area are Landsat series images. To evaluate urban growth, we have chosen three time periods of 10 years. From the (earthexplorer.usgs.gov) website, we obtained the Landsat 7(ETM+) image in 2003, the Landsat 8 (OLI) image in 2013, and the Landsat 9(OLI2) image in 2023[19]. These three images are all from the Collection2 Level 1 image package. Unlike Collection Level 2 packages, this package includes a panchromatic band (band 8) in the collection of images, but atmospheric corrections and other corrections have not been applied to the images [20], [21]. To achieve this, after combining RGB and NIR bands, we apply atmospheric corrections from the downloaded metadata file to each image using ArcGIS Pro processing tools. It is important to note that the RGB and Near-infrared bands for Landsat 7 images are bands 1, 2, 3, and 4, and for Landsat 8 and 9 images, bands 2, 3, 4, and 5.

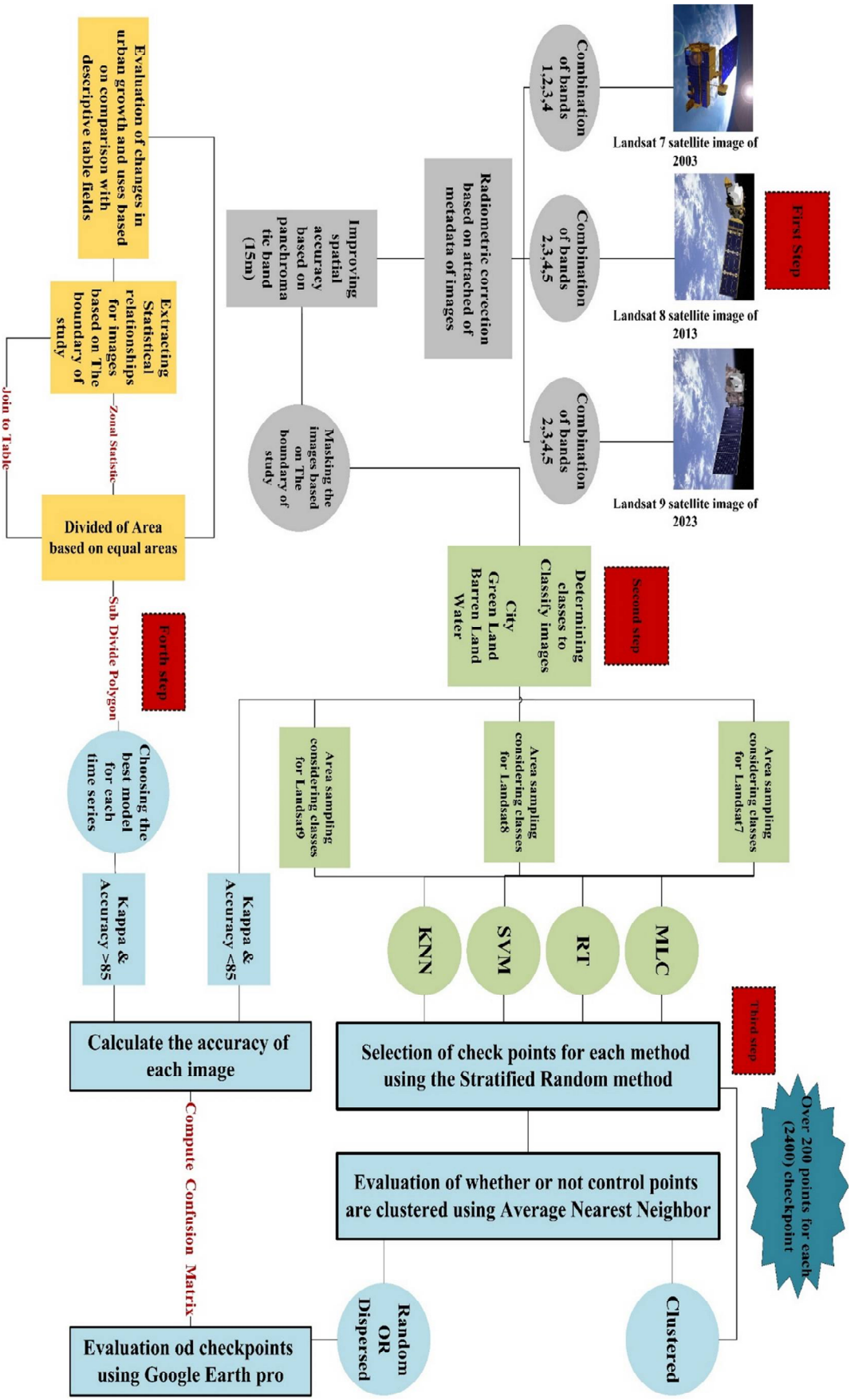


Figure 2. Flowchart of the methodology

Also, the reason for using the NIR band with the RGB bands is to identify vegetation in the area [22] (Table 1).

Table 1. An overview of the information contained in the captured images

	COLLECTION & LEVEL	YEAR	PIXEL SIZE	BAND
LANDSAT 7	2 &1	2003	30M & 15M	1, 2, 3, 4, 8
LANDSAT 8	2 &1	2013	30M & 15M	2, 3, 4, 5, 8
LANDSAT9	2 &1	2023	30M & 15M	2, 3, 4, 5, 8

After preparing the images, we improved their spatial resolution. Band 8 images from the Landsat satellite collection downloaded have a resolution of 15 meters [22]. Using the Pan sharp processing tool in ArcGIS Pro software, the resolution is increased from 30 meters to 15 meters. Each image is masked based on the study boundary while maintaining spatial resolution (Fig. 3).

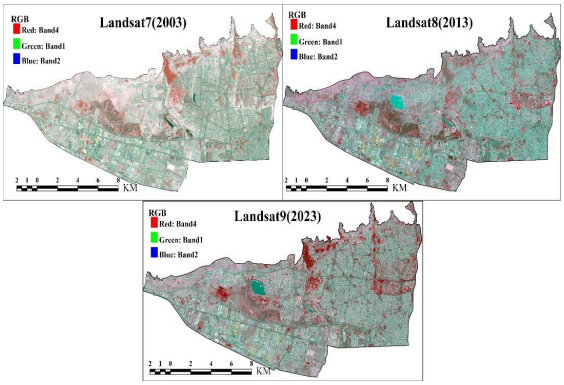


Figure 3. Preparation of Landsat time series images

3-2- Classification using machine learning algorithms

In general, due to the variety of existing classification methods, which use different algorithms to classify images. Also, the types of output that classification methods create cannot provide a comprehensive and complete definition of classification. However, it can be said that all classification methods seek to discover a complication (class) that matches the desired pixel with spectral characteristics [23]. Regardless of the classification algorithm used, the following steps should be followed: 1- Definition of class 2- Selection of the data 3- Preprocessing of the data 4- Selection of the algorithm and its application 5- Postprocessing of the data 6- Assessment of accuracy [24], [25]. According to the situation of the region, we have formed four classes for images: City, Green Land, Barren Land, and Water. There is an artificial lake in the area, and to separate its corresponding pixel from other pixels, we have considered creating a separate class for water. As this lake was constructed in 2012, it is not visible on image 7 of Landsat. Regardless, all algorithms are implemented in ArcGIS Pro version 3.2.

3-3- K-Nearest Neighbour

To apply classification, we first use the KNN algorithm. A KNN algorithm is based on calculating the distance between a sample and a training sample of a known category, then finding the K neighbors that are most similar to that sample; then, based on the category to which these neighbors belong, the type of sample data to be classified is determined [23], [26].

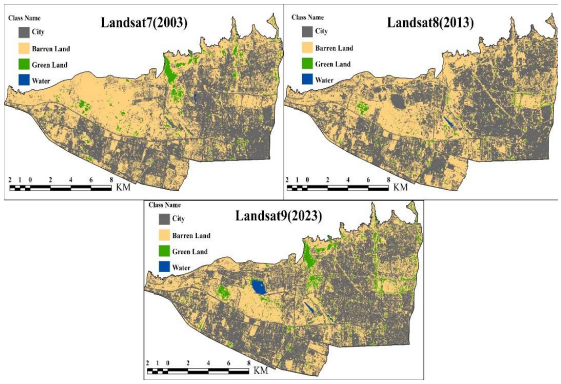


Figure 4. Applying the KNN method on the Landsat time series

In this process, the training set S is searched and the k-nearest neighbor samples closest to the object D are found. As a distance metric, "Nearest Neighbour" uses the Euclidean distance, which is a mathematical formula. Measurement of distance depends on the choice of distance method, and other distances can also be used [27]. Between two objects $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$, the Euclidean distance is:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \qquad (1)$$

This algorithm was applied to the relevant satellite images (Fig.4):

3-4- Support Vector Machine

The SVM method has been widely used in remote sensing applications as a non-parametric supervised learning approach [28]. An SVM model is used to determine which decision boundary is most appropriate between two classes in order to maximize the margin between them [29], [30]. The hyperplane is the term used to describe this decision boundary. Hyperplanes are used to create two distinct classes by determining the distance between the data points for each class [17]. This algorithm was applied to the relevant satellite images (Fig.5):

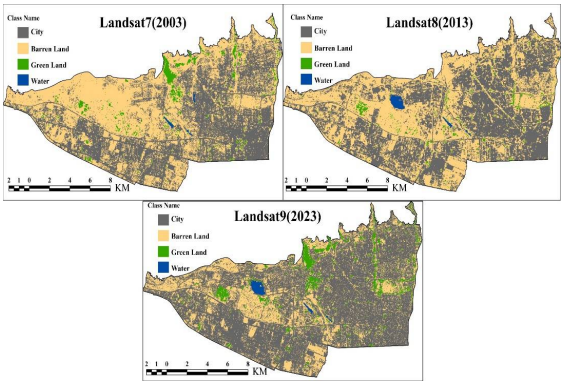


Figure 5. Applying the SVM method on the Landsat time series

3-5- Random Trees (Forest)

A random tree or random forest classification method generates individual decision trees from different samples and subsets of training data [31], [32].

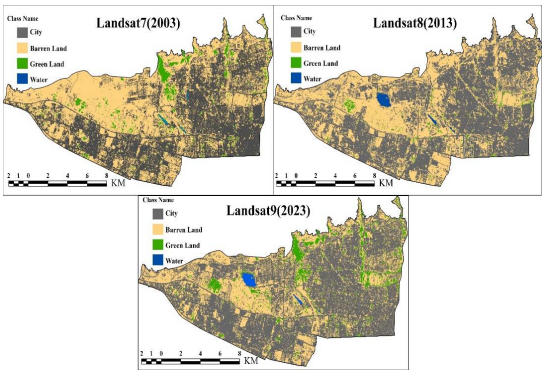


Figure 6. Applying the RT method on the Landsat time series

A decision tree is constructed by ranking the decisions according to the importance of each pixel. We see a branch when we graph these for a pixel [31]. This algorithm was applied to the relevant satellite images (Fig.6):

3-6- Maximum Likelihood

A maximum likelihood classifier (MLC) is a classification method based on probability theory [23]. When training data, MLC assumes the training data statistics for each class in each band are Gaussian in nature [33]. This algorithm was applied to the relevant satellite images (Fig.7):

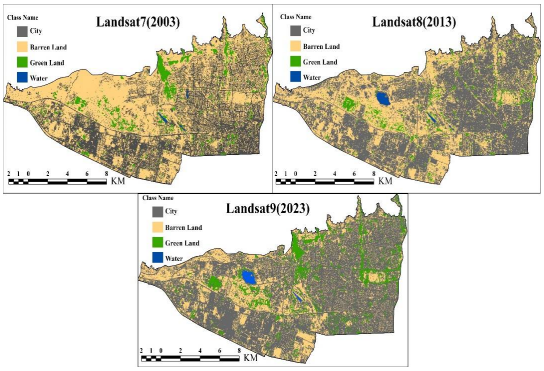


Figure 7. Applying the MLC method on the Landsat time series

3-7- Accuracy evaluation and selection of the most accurate method

To evaluate the best algorithm for each image, we create approximately 200 points for each method based on the sampling method Stratified Random Sampling for each image. Using this sampling method, points are created based on the area of the classes. For example, a class with a larger area creates more points in that area [23]. We also examine the distribution of points using Average Nearest Neighbour (AVG) analysis, and if there are clusters of points, we have repeated the procedure so many times that the distribution is random and dispersed. Then, all the points (Ground True) are evaluated using Google Earth Pro software, and the confusion and error matrix is calculated using the scikit-learn library in Python.

According to the matrix, the numbers on the main diameter indicate the number of pixels that match the two data series. In other words, the pixels that are correctly classified are placed on the main diameter, while the elements other than the main diameter represent the errors [23]. There is also an

overall accuracy in this matrix, which is an average of classification accuracy based on the ratio of correctly classified pixels to total pixels [34].

$$O.A = \frac{\sum_{i=1}^{c} E_{ii}}{N} \quad (2)$$

Where C is the number of classes, N is the total number of known pixels, E_{ii} is the principal diameter members and O.A or U.A is the overall classification accuracy.

The overall accuracy does not require a complex operation for calculation and is a general estimate of classification accuracy and treats all classes the same and does not consider the difference between classes, so we consider the Kappa coefficient as the accuracy parameter. The Kappa coefficient calculates the overall classification accuracy compared to random classification [23], [34].

$$K = \frac{N \sum_{i=1}^{r} X_{ii} - \sum^{r} X_{i+} X_{+i}}{N^2 - \sum^{r} X_{i+} X_{+i}} \quad (3)$$

where N is the total number of ground reality pixels, X_{i+} is the sum of the elements of the i-th row, and X_{+i} is the sum of the elements of the j-th column.

Based on the evaluations made by the confusion matrix, about the Landsat 7 image in 2003, the KNN method with an overall accuracy of 0.94 and a kappa of 0.91, the Landsat 8 image in 2013, the SVM method with an overall accuracy of 0.97 and a kappa of 0.95, and the Landsat 9 image in 2023, the SVM method had the best results with overall accuracy of 0.94 and a kappa of 0.89, and it has been chosen as a criterion for calculating urban growth (Fig.8):

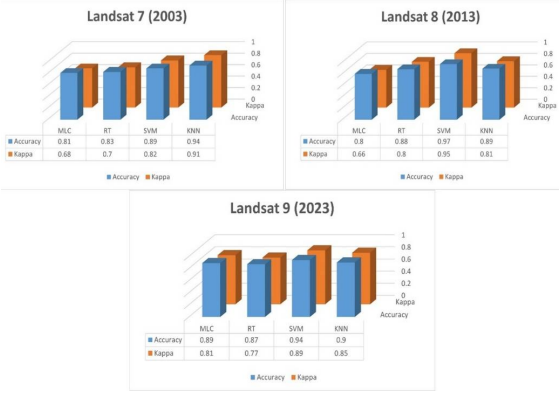


Figure 8. Check accuracy and kappa coefficient

4- Results and Discussion

In this study, we applied 4 classification methods using machine learning algorithms to Landsat images. The most suitable algorithm for each image was selected in the previous section. Urban growth is evaluated by examining the difference in pixels related to 2003 and 2013, 2013 and 2023, and finally 2003 and 2023. To calculate the areas of the samples, we multiply the area of the corresponding pixels by their number (the dimensions of each pixel after spatial improvement are 15x15) (Fig. 9).

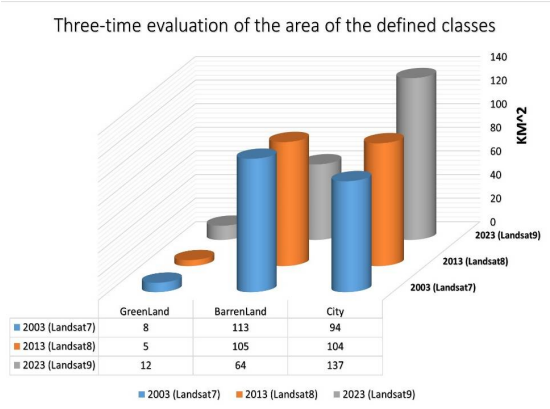


Figure 9: Evaluation of the area of the defined classes

Also to better understand the urban growth in a certain region, we divided our region into equal areas and used raster statistical analysis (Zonal Statistic) to check the area of pixels in those divisions. Additionally, we will examine changes to Greenland pixels and Barren land pixels. Following the previous steps, urban green spaces, agricultural lands, and hilly areas were regarded as green lands (Since water-related pixels have a small area, they are not evaluated).

During the period 2003 to 2013, urban growth occurred primarily in the northwest and west, while barren land was reduced in the same vicinity. It is also possible that the increase in barren land in the central and southern parts of the district is due to the destruction of worn-out fabric, the commencement of construction work for the upcoming years, or the destruction of agricultural land. Regarding green lands, it can be said that where urbanization has expanded, agricultural lands and green spaces have completely disappeared, but in the centre and south of the region, we saw an increase, due to the expansion of urban green spaces (Fig. 10).

Regarding urban expansion between 2013-2023, there has been an increase in the entire district, and a decrease in barren land has also occurred in the entire district.

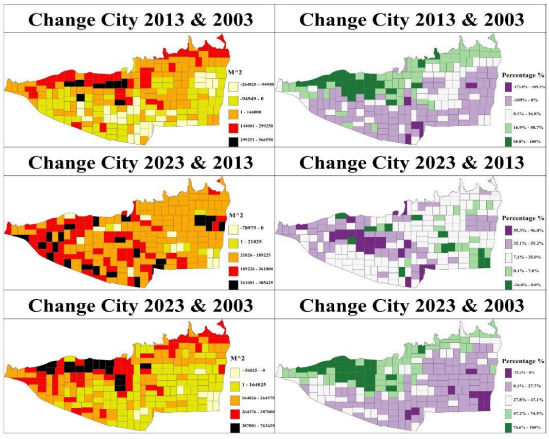


Figure 10. Evaluation of urban growth (City)

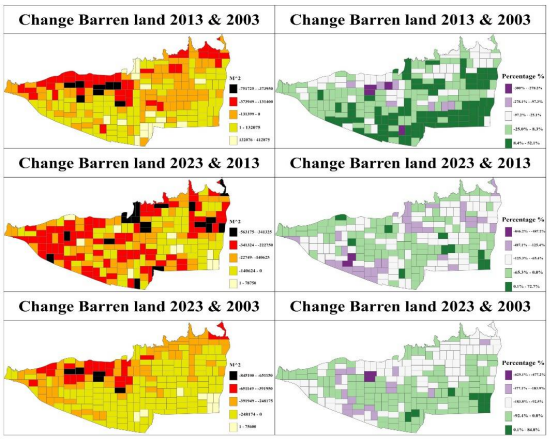


Figure 11. Evaluation of urban growth (Barren land)

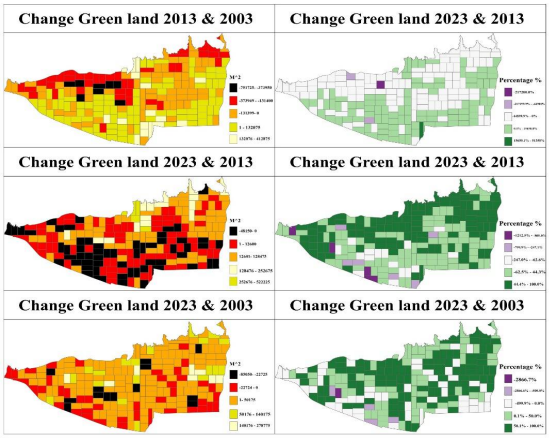


Figure 12. Evaluation of urban growth (Green land)

About green lands, it can be said that there has been a decrease in the southern part of the district and a slight decrease in the western part of the district, but in general, we have seen an increase in these lands due to the expansion of urban parks in the region (Fig. 11).

In general, over the past 20 years (2003-2023), urban expansion has taken place throughout the region, and barren land has been reduced as well. There has, however, been a very sharp decline in green land in western regions. Due to the development of urban parks, there is an obvious increase in green space in the district and the centre (Fig. 12).

5- Conclusion

Previous studies examined the growth of a phenomenon by evaluating several classification methods on one image and applying the chosen method to other images. Other methods of classification on other images should be considered if possible. We used four machine learning algorithms to classify images of urban growth in Tehran's western region over two ten-year periods. The use of one method for each time interval may result in incorrect evaluation of the results when it comes to image classification. Accordingly, we tested all four methods for each image, and two methods, SVM and KNN, were accepted for three images. For the Landsat 7 satellite image in 2003, method KNN with 94% accuracy and 91% Kappa coefficient, and for Landsat 8 and 9 images in 2013 and 2023, method SVM with 97% and 94% accuracy and 95% and 89% better Kappa coefficient, respectively and had the best performance and entered the urban growth calculations. It can be said that tree-based methods (RF, SVM) perform better with images with dispersion between samples (Also, the RF method produces an almost identical Kappa coefficient to the SVM method as well as accuracy), while distance-based methods (KNN) perform better with images that are uniform between samples. Additionally, the MLC method resulted in low accuracy and a low Kappa coefficient. As a result of combining these methods, between 2013 and 2003, we saw a 10% increase in the expansion of the city, a 13% decrease in barren land, and a 60% decrease in green land. Also, between 2013 and 2023, we saw a 24% growth in urban development, a 64% decrease in barren land, and a 50% decrease in green land. As a result, between 2003 and 2023, there will be a 34% growth in the city's expansion, a 76% decrease in barren land, and a 33% increase in green land.

As suggestions for future studies: 1- the XGBoost algorithm can be used for classification. XGBoost is a boosting algorithm that uses bagging, which trains multiple decision trees and then combines the results. It allows XGBoost to learn more quickly than other algorithms but also gives it an advantage in situations with many features to consider. 2- A better understanding of these changes can also be achieved by analysing the driving factors of change and modelling them using various regression methods.

Data availability statement

The data used in this study was obtained from the United States Geological Survey (USGS): https://earthexplorer.usgs.gov. It is also possible to obtain processed data on demand: (hossein.joulaei98@gmail.com)

References

[1] M. A. Kuddus, E. Tynan, and E. McBryde, "Urbanization: A problem for the rich and the poor?," *Journal of Public Health Rev*, vol. 41, no. 1, Jan. 2020, doi: 10.1186/s40985-019-0116-0.

[2] "Urbanization." Accessed: Dec. 09, 2023. [Online]. Available: https://un.org/development/desa/pd/content/urbanization-0

[3] A. Talib, "International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Bienn... Potentiality of ecotourism development of Kirala Kele Partial-Nature-Based wetland the Southern Sri Lanka View project The Use of Probiotics to Optimize Mud Crab Scylla paramamosain Larval Culture. View project Noresah Mohd Shariff Sanjay Gairola Environment and Protected Areas Authority Sharjah," 2010. [Online]. Available: https://www.researchgate.net/publication/235931337

[4] X. Sun, C. Zhang, and Q. Tan, "Factors Influencing the Coordinated Development of Urbanization and Its Spatial Effects: A Case Study of Beijing-Tianjin-Hebei Region," *Sustainability (Switzerland)*, vol. 15, no. 5, Mar. 2023, doi: 10.3390/su15054137.

[5] Muhammad Nasar-u-minAllah Bhalli and Abdul Ghaffar, "Use of Geospatial Techniques in Monitoring Urban Expansion and Land Use Change Analysis: A Case of Lahore, Pakistan," *Journal of Basic & Applied Sciences*, vol. 11, pp. 265–273, Jan. 2015, doi: 10.6000/1927-5129.2015.11.38.

[6] A. Rienow, A. Mustafa, L. Krelaus, and C. Lindner, "Modeling urban regions: Comparing random forest and support vector machines for cellular automata," *Transactions in GIS*, vol. 25, no. 3, pp. 1625–1645, Jun. 2021, doi: 10.1111/tgis.12756.

[7] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13. Institute of Electrical and Electronics Engineers Inc., pp. 6308–6325, 2020. doi: 10.1109/JSTARS.2020.3026724.

[8] Y. G. Yuh, W. Tracz, H. D. Matthews, and S. E. Turner, "Application of machine learning approaches for land cover monitoring in northern Cameroon," *Ecol Inform*, vol. 74, May 2023, doi: 10.1016/j.ecoinf.2022.101955.

[9] Y. O. Ouma, A. Keitsile, B. Nkwae, P. Odirile, D. Moalafhi, and J. Qi, "Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach," *Eur J Remote Sens*, vol. 56, no. 1, 2023, doi: 10.1080/22797254.2023.2173659.

[10] Y. Qian, W. Xing, X. Guan, T. Yang, and H. Wu, "Coupling cellular automata with area partitioning and spatiotemporal convolution for dynamic land use change simulation," *Science of the Total Environment*, vol. 722, Jun. 2020, doi: 10.1016/j.scitotenv.2020.137738.

[11] K. M. Gilbert and Y. Shi, "Land use/land cover change detection and prediction for sustainable urban land management in Kigali City, Rwanda," vol. 2023, no. 2, pp. 62–75, [Online]. Available: https://publish.mersin.edu.tr/index.php/alm

[12] W. I. Md.Mustaquim, "Assessment of Land Use/Land Cover Change and its Future Prediction Using CA-Markove with ANN Simulation for Berhampore, West Bengal, India," *Res Sq*, 2023, doi: https://doi.org/10.21203/rs.3.rs-3407386/v1.

[13] P. Thanh Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors (Basel)*, vol. 18, no. 1, Dec. 2017, doi: 10.3390/s18010018.

[14] S. K. Hanoon, A. F. Abdullah, H. Z. M. Shafri, and A. Wayayok, "Urban Growth Forecast Using Machine Learning Algorithms and GIS-Based Novel Techniques: A Case Study Focusing on Nasiriyah City, Southern Iraq," *ISPRS Int J Geoinf*, vol. 12, no. 2, Feb. 2023, doi: 10.3390/ijgi12020076.

[15] A. Rash, Y. Mustafa, and R. Hamad, "Quantitative assessment of Land use/land cover changes in a developing region using machine learning algorithms: A case study in the Kurdistan Region, Iraq," *Heliyon*, vol. 9, no. 11, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21253.

[16] V. K. Rana and T. M. Venkata Suryanarayana, "Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands," *Remote Sens Appl*, vol. 19, Aug. 2020, doi: 10.1016/j.rsase.2020.100351.

[17] H. S. Pokhariya, D. P. Singh, and R. Prakash, "Evaluation of different machine learning algorithms for LULC classification in heterogeneous landscape by using remote sensing and GIS techniques," *Engineering Research Express*, vol. 5, no. 4, p. 045052, Dec. 2023, doi: 10.1088/2631-8695/acfa64.

[18] L. Ghayour *et al.*, "Performance evaluation of sentinel-2 and landsat 8 OLI data for land cover/use classification using a comparison between machine learning algorithms," *Remote Sens (Basel)*, vol. 13, no. 7, Apr. 2021, doi: 10.3390/rs13071349.

[19] "USGS- Scince for changing world." Available: https://www.usgs.gov/news/news-releases

[20] "Landsat Collection 2 Level-2 Science Products." Available: https://www.usgs.gov/landsat-missions/landsat-collection-2-level-2-science-products

[21] "Landsat Collection 2 Level-1 Data." Accessed: Dec. 07, 2023. [Online]. Available: https://www.usgs.gov/landsat-missions/landsat-collection-2-level-1-data

[22] R. Goldblatt, A. Rivera Ballesteros, and J. Burney, "High Spatial Resolution Visual Band Imagery Outperforms Medium Resolution Spectral Imagery for Ecosystem Assessment in the Semi-Arid Brazilian Sertão," *Remote Sens (Basel)*, vol. 9, no. 12, p. 1336, Dec. 2017, doi: 10.3390/rs9121336.

[23] F. Seyyed Bagher and R. Yosof, *Principles of Remote Sensing*. Iran-Isfahan: Azadeh, 2015.

[24] S. Abburu and S. B. Golla, "Satellite Image Classification Methods and Techniques: A Review," 2015.

[25] I. Nurwauziyah, U. D. Sulistyah, I. Gede, B. Putra, M. I. Firdaus, and D. S. Umroh, "Satellite Image Classification using Decision Tree, SVM and k-Nearest Neighbor," 2018. [Online]. Available: https://www.researchgate.net/publication/326316293

[26] R. Li and S. Li, "Multimedia Image Data Analysis Based on KNN Algorithm," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/7963603.

[27] "Lecture 2: k-nearest neighbors." Accessed: Dec. 10, 2023. [Online]. Available: https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote02_kNN.html#:~:text=The%20k%2DNN%20algorithm&text=Denote%20the%20set%20of%20the,furthest%20point%20in%20Sx).

[28] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int J Remote Sens*, vol. 23, no. 4, pp. 725–749, Feb. 2002, doi: 10.1080/01431160110040323.

[29] "Support Vector Machine - an overview | ScienceDirect Topics." Accessed: Dec. 10, 2023. [Online]. Available: https://www.sciencedirect.com/topics/immunology-and-microbiology/support-vector-machine

[30] N. Rezaei and P. Jabbari, "Support vector machines in R," *Immunoinformatics of Cancers*, pp. 143–156, 2022, doi: 10.1016/B978-0-12-822400-7.00013-0.

[31] Esri, "Train Random Trees Classifier (Spatial Analyst)," 2023, Accessed: Dec. 10, 2023. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/train-random-trees-classifier.htm

[32] A. Ghosh, R. Sharma, and P. K. Joshi, "Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion," *Applied Geography*, vol. 48, pp. 31–41, Mar. 2014, doi: 10.1016/j.apgeog.2014.01.003.

[33] B. R. Shivakumar and S. V. Rajashekararadhya, "Investigation on land cover mapping capability of maximum likelihood classifier: A case study on North Canara, India," in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 579–586. doi: 10.1016/j.procs.2018.10.434.

[34] A. Rash, Y. Mustafa, and R. Hamad, "Quantitative assessment of Land use/land cover changes in a developing region using machine learning algorithms: A case study in the Kurdistan Region, Iraq," *Heliyon*, vol. 9, no. 11, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21253.