International Journal of Environment and Geoinformatics 10(1):146-152 (2023)

Research Article

Estimation of Urban Area Change in Eskişehir Province Using Remote Sensing Data and Machine Learning Algorithms

Dilek Küçük Matcı

Eskisehir Technical University, Institute of Earth and Space Sciences, Eskisehir, TURKIYE

E-mail: dkmatci@eskisehir.edu.tr

Received 15.08.2022
Accepted 13.03.2023

How to cite: Matçı, D.K. (2023). Estimation of Urban Area Change in Eskişehir Province Using Remote Sensing Data and Machine Learning Algorithms, *International Journal of Environment and Geoinformatics (IJEGEO)*, 10(1):146-152, doi. 10.30897/ijegeo.1162153

Abstract

Rapid population growth, natural events, and increasing industrialization are among the factors affecting land use. To keep this change under control and to make sound plans, it is necessary to control the changes. In this study, the spatial use change in the Eskişehir region between the years 1990-2018 was examined with CORINE data. Based on this determined change, an urban change model was created with the multivariate regression method. As a result of the evaluations, while an increase was observed in urban areas and pastures between 1990-2018, a decrease was determined in agricultural and forest areas. This change is defined as 43.74% in urban areas, 3.28% in agricultural areas, 7.78% in forest areas, and 60.10% in pasture areas. SMOReg, MLP Regressor, and M5P Model Tree methods were used for the estimation study to be carried out with the obtained spatial change data. Urban values for 2018 were estimated to find the best method. Finally, the areas of 2030 were estimated with the method that gave the best results. The results demonstrated the usability of modeling using CORINE data.

Keywords: Estimation, Land Use Change, Eskişehir, Remote sensing, Machine Learning

Introduction

Changes in land cover occur with human intervention as well as natural disasters (Jaiswal, ety al., 1999,Dubertret et al. 2022). Especially rapid and uncontrolled urbanization is a phenomenon that causes significant land use changes in both developing and developed countries (Yildiz et al. 2021). Detecting and estimating land use/ land cover (lulc) dynamics has become an important issue due to ensure proper use of limited resources (Wang et al. 2020). The transfer of nature to future generations without deterioration depends on monitoring, controlling, and planning the changes (Wang et al. 2020; Butt et al. 2015; Girma, et al., 2021).

Remote sensing and Geographic Information Systems are increasingly used in the assessment of land use/land cover changes. Remote sensing provides accurate and rapidly updatable information about natural resources and the environment (Fan, Weng, and Wang 2007; Seto and Kaufmann 2003). Using these data, air quality (Potts et al. 2021), agriculture (Dudu and Çakmak 2018; Ocer et al. 2020; Başaran, Matcı, and Avdan 2022), forest (Basaran, Matcı, and Avdan 2022), fire (Barton et al. 2023, Matcı and Avdan 2020), earthquake (Çömert, Matcı, and Avdan 2018). One of the data prepared with remote sensing systems is Environmental Information Coordination (CORINE) data. The CORINE data series was created by the European Community as a means of compiling geospatial environmental information (Feranec et al. 2007). Based on the interpretation of satellite images, Corine provides maps of lulc change over six-year periods.

Corine datasets are used in a wide variety of social and environmental studies. For example, Frenac et al. (2007) proposed a new change analysis method using Corine data. With this method, they determined the land use changes in the Netherlands and Slovakia (Feranec et al. 2007). In another study, Popovici et al. (2013) examined land use dynamics in Romania (Popovici, et al., 2013). In another study using Corine data, it was aimed to automatically label the classes obtained as a result of the uncontrolled classification of Sentinel images (Matcı and Avdan 2022). In another study, Corine data was used to examine urban heat islands in large cities in Greece (Stathopoulou and Cartalis 2007).

Modeling analysis is used to determine which parameters are affected by land use changes. Some of these models use past land use data to predict future land use scenarios along with environmental variables (Zadbagher, et al., 2018). The geospatial analysis utilizes a variety of statistical and rule-based modeling approaches to detect and predict land use changes (Overmars, et al., 2003). Among the models commonly used in lulcchange studies are; statistical models (Allbed, et al., 2014; Başaran, Matcı, and Avdan 2022), evolutionary models (Aitkenhead and Aalders 2009), cellular models (Kafy et al. 2021), Markov models (Mubea, et al., 2011), hybrid models (Weslati, et al, 2023, Küçük Matcı, et al., 2022).

The aim of this study is to determine the land cover changes in Eskişehir and to estimate the urban area changes in the region based on remote sensing data. In this direction, the CORINE data of the study area between 1990 and 2018 were used.

Materials and Methods

This study was carried out in Eskişehir, which is located in the northwest of the Central Anatolian region. 22% of Eskişehir' s surface area is covered by mountains and 26% by plains. The climate of the region shows the characteristics of Central Anatolia, the Western Black Sea, and the Mediterranean regions. The vegetation in the region can be counted as the Central Anatolian steppes, North Anatolian, and Western Anatolian forests (Url-3, 2022). The study area is shown in Figure 1.

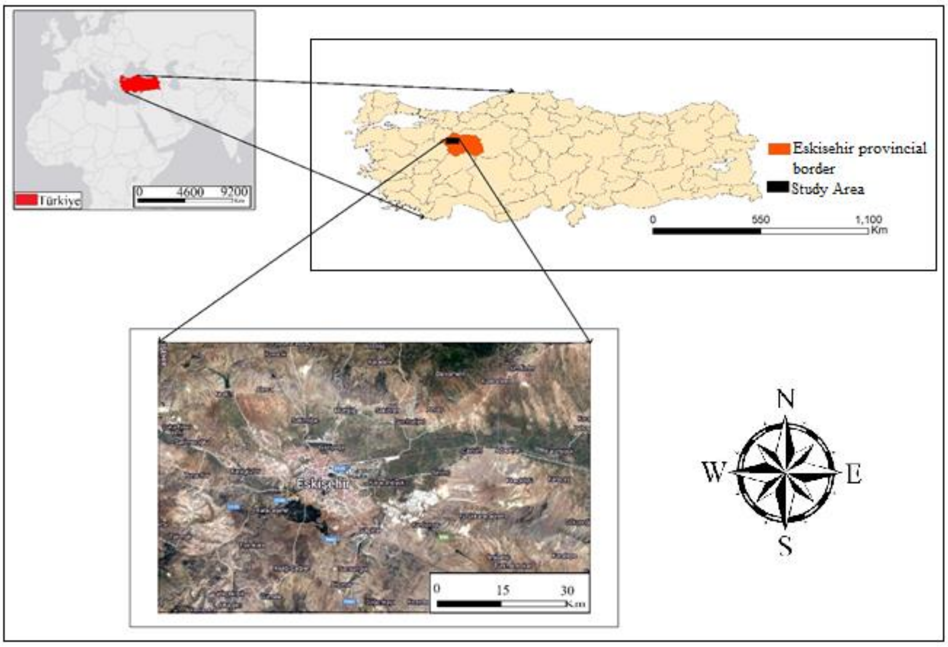

Fig. 1. Study Area

Within the scope of the study, it is aimed to determine the spatial change of the Eskişehir region between the years 1990-2018 with CORINE data and to model this change. In this direction, data sets published in 1990, 2000, 2006,

2012 and 2018 were used. These data are land cover maps with a spatial resolution of 100 m, based on remote sensing data, within the scope of the CORINE program (Url-1 2015). The method applied to determine and model the spatial change in the study area is given in Figure 2.

Accordingly, first of all, CORINE data were obtained using Google Earth Engine. The obtained data were cut according to the study area. To determine the spatial change, data between 1990 and 2018 were used. The land use classes in each data set obtained were determined and their areas were calculated. In the second stage of the study, urban area change was modeled with machine learning algorithms by using land use changes (Çelik and Gazioğlu, 2022; Gümüşçü, et al., 2023). One of the methods used for this purpose in the study is the Sequential Minimum Optimization Regressor (SMOReg) method. The SMOReg method uses support vector machines for regression. This method applies the sequential minimum optimization algorithm to train the support vector regression model. This application replaces all missing values globally and converts nominal attributes to binary values. It also normalizes all attributes by default (Li and Jiang 2006). While this method was used with the weka software in the study, the necessary parameters were determined as complexity parameter =0.1001, regressor optimizer = RegSMOImproved by trial and error method.

The second method used in the study is the multilayer perceptron (MLP) regressor model. This method consists of an input layer, one or more hidden layers, and an output layer. The artificial neurons in the layers are completely interconnected. The number of neurons in the input layer is the same as the size of the input vector, and the number of artificial neurons in the output layer is designed according to the structure of the output vector. The number of neurons in the hidden layer can be increased with several trials, at least one until optimal training is obtained. The model used in this method is given below:

$$\mathrm{Yn} = \mathrm{f0}\,\{\mathrm{b0} + \sum [\mathrm{wk} * \mathrm{fh}\,(\mathrm{bhk} + \sum_{i=1}^{m} \mathrm{w}_{ik} X_{ni})]\}$$

Here Yn is the normalized output, $f0$ is the output layer transfer, $b0$ is the output layer deviation, wk is k. where the link weight from the hidden layer to the output layer, $f0$ represents the hidden layer transfer function, and $b0k$ represents the deviation of the output layer. kth hidden layer terms, wik represents the connection from the i neuron in the input layer to the second neuron in the hidden layer, and Xni represents the normalized input vector (Koutras, Panagopoulos, and Nikas 2017). While this method was used with the Weka Software in the study, the necessary parameters were determined as ridge=0.101, seed=100 and numFunctions=3 by trial and error method. The last method used in the study is the M5P Model Tree method. This method is an extended version of the M5 algorithm. Model trees can efficiently process large numbers of datasets with large numbers of features and high dimensions. In addition, they can successfully work with missing data (Wang and Witten 1996).

Two different approaches were used to determine the accuracy of the estimation results obtained in the study. One is to estimate urban areas for 2018 using data from 1990-2012. To compare the obtained estimation result with the real urban area data of 2018. The second approach is to calculate Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Square Error (MSE).

From these values, the MAE is a measure of the errors between the predicted and actual data. Estimators with lower values are considered to perform better (Willmott and Matsuura 2005). Another value used in the study, RMSE, is the standard deviation of the estimation errors. It is a measure of how far the prediction errors are from the regression line data points. Estimators with a lower RMSE are considered to perform better (Willmott and Matsuura 2005). The MSE value measures the mean of the squares of the estimation errors. Estimators with an MSE value close to zero are considered to perform better (Gunst and Mason 1977).

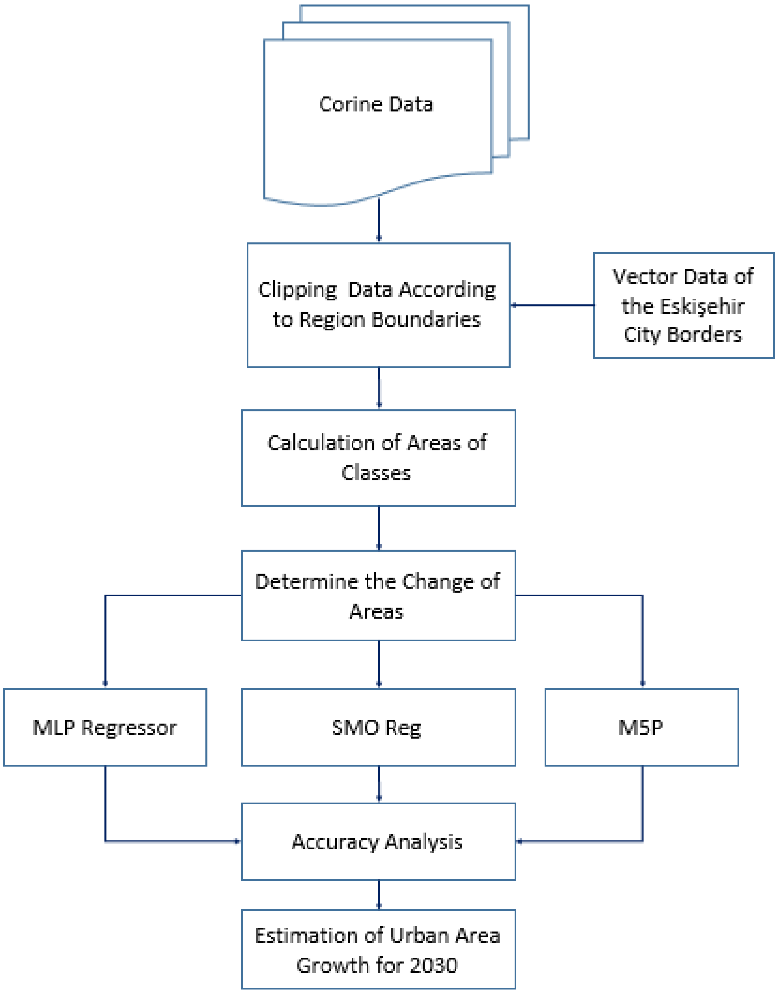

Fig. 2. Applied Method

Results and Discussion

In the research, all classes in the CORINE dataset were combined into main groups. Accordingly, changes in five classes (Urban/Artificial areas, agricultural areas, forests, water, and pasture areas) in Eskişehir were determined. The obtained land use maps are given in Figure 3. When the results are examined, the ones in the study area are urban, water, and pasture areas that increase in area, while the ones that decrease are agriculture and forest areas (Table 1).

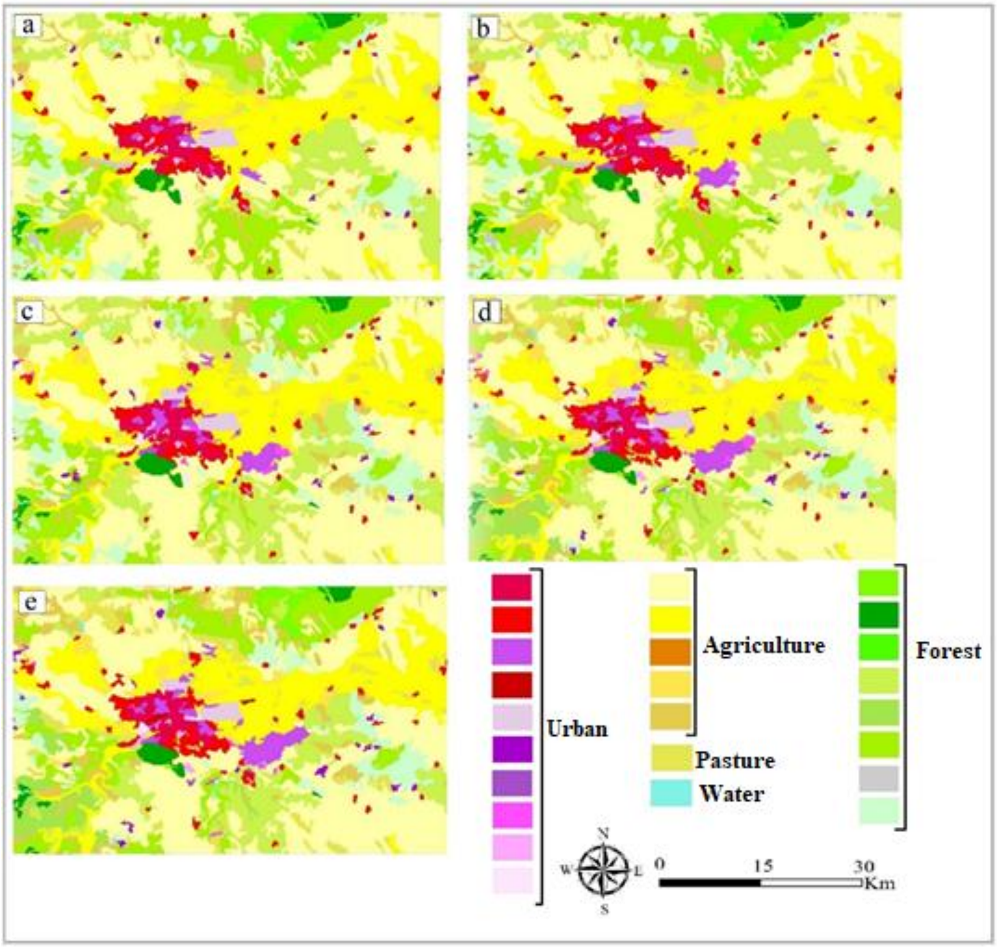

Fig. 3. LULC Maps a)1990 b)2000 c)2006 d)2012 e)2018

Table 1. Area Usage Amounts (Km 2)

Land Use Classes	1990	2000	2006	2012	2018	1990-2018 Change Rate
City	90.74	109.42	113.63	124.14	130.43	43.74
Agriculture	930.00	921.70	908.06	903.51	899.48	-3.28
Forest	488.70	480.18	476.56	451.73	450.66	-7.78
Water	0.28	0.61	1.27	1.43	1.43	410.71
Pastures	46.14	43.95	56.35	75.07	73.87	60.1

Urban areas in Eskişehir have increased by 43% in 28 years. When this change is analyzed based on periods, it increased by 20% between 1990 and 2000, 3.85% between 2000-2006, 9.24% between 2006-2012, and 5.07% between 2012-2018 (Figure 4).

Agricultural areas decreased by 3.28% and forest areas by 7.78%. Although the largest proportional change in the results obtained seems to be in the water masses, this change in area was 1.14 km2. This increase in water areas is due to the dams and ponds built in the region (Url-2, 2020). Besides the water areas, another class that shows a great change is the pastures. Although it showed a small decrease in the 1990-2006 period, it showed a large increase in the 2006-2018 period. This change can be explained by the increase in urban areas and the decrease in agricultural areas, and thus the transformation of these areas into pastures.

In the second stage of the study, urban area change in the examined period was analyzed with machine learning algorithms. In this direction, to determine the algorithm that gives the most accurate result, first of all, urban areas were estimated by SMOReg, MLPRegressor, and M5P methods. The results obtained are given in Table 2.

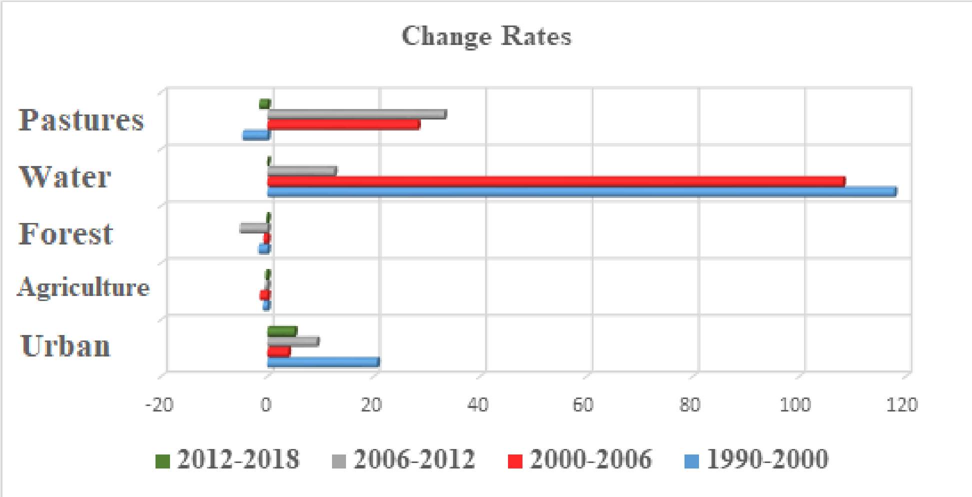

Fig. 4. Rates of Change of Land Use Classes

Table 2. Estimated Urban Area Usage Amounts for 2018 (Km ^{2})

	Urban Area in 2018	SMOReg	MLP Regressor	M5P
Area (Km ^{2})	130.43	136,73	125,22	133.62
MAE	-	0.51	2.02	0.92
RMSE	-	0.51	2.15	1.21
MSE	-	0.26	4.61	1.26

Table 3. Calculated Data for 2030

Method	2030 Area (km^2)	MAE	RMSE	MSE
M5P	148.77	2.57	2.63	6.9
SMOReg	138.82	2.02	2.19	4.77

These results show that when compared with real urban area data, the most accurate estimate is calculated with M5P. When the calculated MAE, RMSE, and MSE values were examined, it was observed that the SMOReg and M5P methods gave better results. For this reason, these two methods were used for the urban area estimation of 2030. The data obtained as a result of the process for the year 2030 are given in Table 3.

Monitoring land use change has become a critical issue for decision planners and conservationists due to inappropriate growth and its impact on natural ecosystems. For this reason, it is necessary to determine the changes that occur and to make the change predictions with the most accurate model. In this study, land use changes in Eskişehir city were examined. In addition, with the help of the data obtained, the urban area changes for 2030 was estimated.

When the results are examined, the results obtained show similarities with the studies in the literature. For example, in one study, Organic Carbon Content (SOC) in river floodplain soils was estimated using multiple linear regression (MLR), M5P, and random forest (RF) models, and the performance of these models was evaluated. As a result of the study, it was stated that the M5P and RF methods gave the most accurate results (Chen et al. 2021). Similarly, M5P and Gaussian Process methods were used in a study to estimate the soil permeability coefficient and it was determined that the M5P method gave more accurate results (Pham et al. 2021). In a study to model the spatial changes of Mangorove forests in the Philippines, SMOReg and MLPRegressor were used together with 16 other machine learning algorithms and the results were examined. Accordingly, SMOReg and MLPRegressor methods gave accurate results above the average compared to other methods (Castillo et al. 2017).

It is thought that the M5P method can clearly explain the relationship between data distributions, size, variable inputs and output parameters, a simple tree structure, and the ability to create linear equations applicable to multiple leaves, making this method more successful (Pham et al. 2021; Bui et al. 2018; Khosravi et al. 2018).

Conclusion

In this study, CORINE data were used to detect and model lulc changes in the Eskişehir region over an 18-year period. The results obtained in the study also show that the Eskişehir region has undergone significant lulc changes since 1990. When the changes in the region are examined, it is revealed that there is an increase in urban areas.

In addition, the spatial changes of the classrooms were examined with machine learning algorithms and the urban change was modeled. For this purpose, SMOReg, MLPRegressor, M5P methods were used. Under these circumstances, the method applied in the study and the assessment of lulc changes appear to provide highly accurate information that can help planners to plan development more sustainably.

References

Aitkenhead, M., Aalders, I. (2009). Predicting land cover using GIS, Bayesian and evolutionary algorithm methods. *Journal of Environmental Management,* 90(1), 236-250.

Allbed, A., Kumar, L., Sinha, P. (2014). Mapping and modelling spatial variation in soil salinity in the Al Hassa Oasis based on remote sensing indicators and regression techniques. *Remote Sensing,* 6(2), 1137-1157.

Barton, Andrew M, Helen M Poulos, George W Koch, Thomas E Kolb, Andrea E Thode. (2023). "Detecting patterns of post-fire pine regeneration in a Madrean Sky Island with field surveys and remote sensing." *Science of the Total Environment:*161517.

Başaran, N., Matcı, D. K., Avdan, U. (2022). Using multiple linear regression to analyze changes in forest area: the case study of Akdeniz Region. *International Journal of Engineering and Geosciences*, 7(3), 247-263.

Bui, D. T., Panahi, M., Shahabi, H., Singh, V. P., Shirzadi, A., Chapi, K.,……Li, S. (2018). Novel hybrid evolutionary algorithms for spatial prediction of floods. *Scientific reports,* 8(1), 1-14.

Butt, A., Shabbir, R., Ahmad, S. S., & Aziz, N. (2015). Land use change mapping and analysis using Remote Sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan. *The Egyptian Journal of Remote Sensing and Space Science,* 18(2), 251-259.

Castillo, J. A. A., Apan, A. A., Maraseni, T. N., & Salmo III, S. G. (2017). Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. Isprs *Journal of Photogrammetry and Remote Sensing*, 134, 70-85.

Chen, J., Zhang, H., Fan, M., Chen, F., Gao, C. (2021). Machine-learning-based prediction and key factor identification of the organic carbon in riverine floodplain soils with intensive agricultural practices. *Journal of Soils and Sediments,* 21(8), 2896-2907.

Celik, OI., Gazioglu, C. (2022). "Coast type based accuracy assessment for coastline extraction from satellite image with machine learning classifiers", *The Egyptian Journal of Remote Sensing and Space Science,* 25(1), 289-299, 2022.

Çömert, R., Matcı, D. K., Avdan, U. (2018). Detection of collapsed building from unmanned aerial vehicle data with object based image classification. *Eskişehir Teknik Üniversitesi Bilim ve Teknoloji Dergisi B-Teorik Bilimler,* 6, 109-116.

Dubertret, Fabrice, François-Michel Le Tourneau, Miguel L Villarreal, Laura M Norman. (2022). "Monitoring Annual Land Use/Land Cover Change in the Tucson Metropolitan Area with Google Earth Engine (1986–2020)." *Remote Sensing* 14 (9):2127.

Dudu, H., Çakmak, E. H. (2018). Climate change and agriculture: an integrated approach to evaluate economy-wide effects for Turkey. *Climate and Development*, *10*(3), 275-288.

Fan, F., Weng, Q., Wang, Y. (2007). Land use and land cover change in Guangzhou, China, from 1998 to 2003, based on Landsat TM/ETM+ imagery. *Sensors,* 7(7), 1323-1342.

Feranec, J., Hazeu, G., Christensen, S., Jaffrain, G. (2007). Corine land cover change detection in Europe (case studies of the Netherlands and Slovakia). Land use policy, 24(1), 234-247.

Girma, R., Fürst, C., Moges, A. (2021). Land Use Land Cover Change Modeling by Integrating Artificial-Neural-Network with Cellular Automata-Markov Chain Model in Gidabo River Basin, Main Ethiopian Rift. *Environmental Challenges,* 100419.

Gunst, R. F., Mason, R. L. (1977). Biased estimation in regression: an evaluation using mean squared error. *Journal of the American Statistical Association,* 72(359), 616-628.

Gümüşçü, İ., Altaş, F., Türkekul, B., Kaya, H.A., Erdem, F., Bakırman, T., Bayram, B. (2023). Water-body Segmentation in Heterogeneous Hydrodynamic and Morphodynamic Structured Coastal Areas by Machine Learning. *International Journal of Environment and Geoinformatics*, 10(1), 100-110, doi. 10.30897/ijegeo.1119096

Jaiswal, R. K., Saxena, R., Mukherjee, S. (1999). Application of remote sensing technology for land use/land cover change analysis. J*ournal of the Indian Society of Remote Sensing*, *27*(2), 123-128.

Kafy, A.-A., Naim, M. N. H., Subramanyam, G., Ahmed, N. U., Al Rakib, A., Kona, M. A., Sattar, G. S. (2021). Cellular Automata approach in dynamic modelling of land cover changes using RapidEye images in Dhaka, Bangladesh. *Environmental Challenges,* 4, 100084.

Khosravi, K., Mao, L., Kisi, O., Yaseen, Z. M., Shahid, S. (2018). Quantifying hourly suspended sediment load using data mining models: case study of a glacierized Andean catchment in Chile. *Journal of Hydrology,* 567, 165-179.

Koutras, A., Panagopoulos, A., Nikas, I. A. (2017). Forecasting tourism demand using linear and nonlinear prediction models. *Academica Turistica-Tourism and Innovation Journal,* 9(1).

Küçük Matcı, D., Çömert, R., Avdan, U. (2022). Analyzing and Predicting Spatiotemporal Urban Sprawl in Eskişehir Using Remote Sensing Data. *Journal of the Indian Society of Remote Sensing*, 1-14.

Li, C., Jiang, L. (2006). Using locally weighted learning to improve SMOreg for regression. Pacific Rim *International Conference on Artificial Intelligence,*

Matcı, D. K., Avdan, U. (2020). Comparative analysis of unsupervised classification methods for mapping burned forest areas. *Arabian Journal of Geosciences*, 13(15), 1-13.

Matcı, D. K., Avdan, U. (2022). Data-driven automatic labelling of land cover classes from remotely sensed images. *Earth Science Informatics,* 1-13.

Mubea, K., Ngigi, T., Mundia, C. (2011). Assessing application of Markov chain analysis in predicting land cover change: a case study of Nakuru municipality. *Journal of Agriculture, Science and Technology*, 12(2).

Ocer, N. E., Kaplan, G., Erdem, F., Kucuk Matci, D., Avdan, U. (2020). Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. *Remote sensing letters*, 11(9), 847-856.

Overmars, K. D., De Koning, G., Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological modelling*, 164(2-3), 257-270.

Pham, B. T., Ly, H.-B., Al-Ansari, N., Ho, L. S. (2021). A Comparison of Gaussian Process and M5P for Prediction of Soil Permeability Coefficient. *Scientific Programming*, 2021.

Popovici, E. A., Bălteanu, D., Kucsicsa, G. (2013). Assessment of changes in land-use and land-cover pattern in Romania using Corine Land Cover Database. *Carpathian Journal of Earth and Environmental Sciences*, 8(4), 195-208.

Potts, D. A., Marais, E. A., Boesch, H., Pope, R. J., Lee, J., Drysdale, W.,….., Moore, D. P. (2021). Diagnosing air quality changes in the UK during the COVID-19 lockdown using TROPOMI and GEOS-Chem. *Environmental Research Letters*, 16(5), 054031.

Seto, K. C., Kaufmann, R. K. (2003). Modeling the drivers of urban land use change in the Pearl River Delta, China: integrating remote sensing with socioeconomic data. *Land Economics*, 79(1), 106-121.

Stathopoulou, M., Cartalis, C. (2007). Daytime urban heat islands from Landsat ETM+ and Corine land cover data: An application to major cities in Greece. *Solar Energy*, 81(3), 358-368.

Url-1, 2015. TC. Tarım ve Orman Bakanlığı. "Corine Projesi." https://corine.tarimorman.gov.tr/ corineportal/ turkiyecalismalar .html.

Url-2, 2020. DSİ. "İşletmedeki Baraj ve Göletler." https:/ /bolge03. dsi.gov.tr/ Sayfa/Detay/898.

Url-3, 2022. Eskişehir Buyuksehir Belediyesi. 2020. "Coğrafya." http:// www. eskisehir. bel.tr/ sayfalar.php?

Wang, S. W., Gebru, B. M., Lamchin, M., Kayastha, R. B., Lee, W.-K. (2020). Land use and land cover change detection and prediction in the Kathmandu District of Nepal using remote sensing and GIS. *Sustainability*, 12(9), 3925.

Wang, Y., Witten, I. H. (1996). Induction of model trees for predicting continuous classes.

Weslati, Okba, Samir Bouaziz, and Mohamed Moncef Sarbeji. (2023). "Modelling and assessing the spatiotemporal changes to future land use change scenarios using remote sensing and CA-markov model in the mellegue catchment." *Journal of the Indian Society of Remote Sensing* 51 (1):9-29.

Willmott, C. J., Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82.

Yildiz, N. D., Avdan, U., Aytatlı, B., Kuzulugil, A., Enes, A. (2021). Determination of Field Use Changes by Using Landscape Metrics:"Erzurum City Example". *Journal of the Institute of Science and Technology*, 11(1), 661-671.

Zadbagher, E., Becek, K., Berberoglu, S. (2018). Modeling land use/land cover change using remote sensing and geographic information systems: case study of the Seyhan Basin, Turkey. *Environmental Monitoring and Assessment*, 190(8), 1-15.