International Journal of Geography and Geography Education

Research Article/Araştırma Makalesi

**To Cite This Article:** Canpolat, F. A. (2022). The use of machine learning to identify suitable areas for urban growth in mountainous areas: Tunceli city example. *International Journal of Geography and Geography Education (IGGE)*, 47, 210-232. http://dx.doi.org/10.32003/igge.1119297

# THE USE OF MACHINE LEARNING TO IDENTIFY SUITABLE AREAS FOR URBAN GROWTH IN MOUNTAINOUS AREAS: TUNCELİ CITY EXAMPLE

## Dağlık Alanlarda Makine Öğrenmesi ile Kentsel Büyümeye Uygun Alanların Belirlenmesi, Tunceli Kenti Örneği

**Fethi Ahmet CANPOLAT***

**Abstract**

One of the most important trigger factors contributing to increased human intervention in space in many regions of the world is urbanization. To manage and plan urbanization in harmony with other human activities, it is necessary to manage and plan it accordingly. Even though urbanization studies tend to focus on large cities, small-scale cities are quite common throughout the world, both in terms of their numbers and regarding their population density. Moreover, small cities can contribute to a more homogeneous distribution of development at the national and regional levels. It may, however, be hindered by a variety of limitations, including the hinterlands and the unused potential of these settlements. The city of Tunceli is also a small settlement with natural and human factors limiting its growth. In this study, based on machine learning algorithms, "support vector machines", "artificial neural networks" and "random forest" models were used to determine urban growth zones. In the city, the most suitable sites for primary growth are those which are suited for peripheral growth and inward-stacked growth (12 km$^2$). While more than 90% of predictions were accurate, regarding the spatial equivalents of the findings, the best results respectively, came from "random forests", "artificial neural networks", and finally "support vector machines".

**Keywords:** Tunceli; urban growth; support vector machines; artificial neural networks; random forest

**Öz**

Dünyanın birçok bölgesinde kentleşme, insanın mekâna olan müdahalesini arttıran en önemli tetikleyici unsurlardan birine dönüşmüş durumdadır. Dolayısıyla kentleşme sürecinin, diğer beşeri faaliyetlere göre yönetilmesi ve planlanması öncelik arz etmektedir. Kentleşme konusundaki çalışmalar ağırlıklı olarak büyük şehirler üzerinde yoğunlaşmasına rağmen, küçük ölçekli kentler hem nüfus miktarı hem de sayı açısından, dünya genelinde oldukça fazladır. Ayrıca küçük kentler, kalkınmanın ulusal ve bölgesel düzeyde daha homojen dağılmasında etkili olabilecek alanlardır. Ancak bu yerleşmelerin büyümesinde, situasyonu, hinterlandı ve potansiyelin kullanılamaması gibi çeşitli sınırlılıklar engel oluşturabilmektedir. Tunceli kenti de küçük ölçekli ve hem doğal hem de beşeri faktörler tarafından büyümesinde kısıtlılıkları olan bir yerleşmedir. Bu çalışmada makine öğrenmesi algoritmalarından "destek vektör makineleri", "yapay sinir ağları" ve "rastgele orman" modelleri kullanılarak kentsel büyümeye uygun alanlar tespit edilmiştir. Kentte içe doğru yığılmalı büyüme ile periferik büyümeye elverişli alanların öncelikli büyümeye uygun olduğu (12 km$^2$) tespit edilmiştir. Kullanılan modellerde, %90'ın üzerinde tahmin doğruluğuna ulaşılmasına rağmen, sonuçların mekânsal karşılığı açısından en iyi sonuçların sırasıyla "rastgele orman", "yapay sinir ağları" ve son olarak "destek vektör makineleri" modelinde elde edilmiştir.

**Anahtar Kelimeler:** Tunceli; kentsel büyüme; destek vektör makineleri; yapay sinir ağları; rastgele orman

* Assist. Prof., Fırat University, facanpolat@firat.edu.tr

**Geliş Tarihi/Recieved:** 20.05.2022 **Kabul Tarihi /Accepted:** 28.07.2022 **Ç. Yayınlanma Tarihi/Online Published:** 20.09.2022

## INTRODUCTION

As complex ecosystems dominated by humanity, urbanization is accelerated by current global economic conditions which have literally changed the face of the planet in many parts of the world. Urban Growth, which continues on an increasing scale throughout the world, is influenced by competitive conditions related to global economic integration. In addition, it is a multifaceted dynamic process which is contingent upon the driving forces related to politics, the environment, population movements, technology and climate change. Therefore, managing urban growth became one of the most important problems of the 21st century as it increased in terms of both scope and complexity (Vitousek et al. 1997, Cohen, 2004, Goncalves et al., 2019).

Although small and medium-sized cities are important in terms of both their numbers and populations, research on the economic, social, environmental and technological changes of cities focuses mainly on metropolitan areas (Fridemann, 1986; Sassen, 1991; Scott, 2001; Dadashpoor et al. , 2019). In recent years, however, in the context of their importance and impact, studies on small and medium-sized cities have become more popular (Henderson, 1997; Üzmez, 2012; Çalışkan & Tezer, 2018).

In a study conducted in the USA, a small city was defined to have a population of less than 50 thousand (Brennan & Hoene, 2003). In a study conducted for Turkish cities, cities with a central population of at least 50,000 and with surrounding areas between 100,000 and 750,000 people were defined as being moderate in size, while those below these figures were referred to as small sized cities (Üzmez, 2012). Based on this, Tunceli city can be considered a small-scale city as the city center's population is 39.000 and the provincial total is 83.000 as of 2021 (TUIK, 2022).

Tunceli is a relatively young and small city located in the Upper Euphrates section of the Eastern Anatolia Region, within the Tunceli Mountainous Region, on the slopes and terraces of the valleys where Munzur Stream and Pülümür Stream merge to form Munzur River (Uzunçayır Dam) (Figure 1, Photo 1). In the selection of the settlement's location, the need for a strategic location was decisive, as required for the security concerns of the period. There are areas which are not suitable for settlement around the establishment area corresponding to the steep and rocky slopes on both sides of the Munzur River flowing in the north-south direction, in connection to both the increased slop which brings the risk of a natural disaster (Tuncel, 2012; Tunceli Municipality, 2017).

While the city is located in a very rugged landscape, the hydrographic elements that accompany this landscape prevent the city from having a compact structure. Therefore, the city's physiographic structure is a restrictive factor for the development areas and growth potential of the city. For this reason, the city has shown a rather slow growth rate since the time of its foundation. Today, this slow growth trend continues due to both the demographic growth capacity of the city and the restrictions related to the location of foundation of the city.

Due to the isolated structure of the city, the economy of scale could not be exploited and production was directed only towards domestic consumption, thus it has not been possible to compete with companies producing nationally within areas where the economy of scale was possible (İzmen, 2014). In terms of population, it has a very low growth trend compared to other provincial and district centers throughout the country. As of 2021, it ranks 399th out of 973 districts from large to small, and 80th among provincial centers.

The area of the neighborhoods connected to the settlement is about 95 $km^2$ while, the area covered by the city with different densities within this area is approximately 27 $km^2$. The maximum elevation in the settlement area of the city is 1567 meters, and the minimum elevation is 857 meters. The average elevation above sea level is 1076 meters.

The land mass of the city located in the eastern part of the Munzur River is very limited. Only public housings were built in the Esentepe neighborhood, close to the village of Batman. The remaining land consists of land that is not suitable for settlement due to fragmentation, loose ground and high slope values, which are often torn apart by streams due to Eocene-era calcareous marn (Aslan, 2016).

Limiting the location of the city of Tunceli only to its immediate surroundings is not enough to understand the current state of the city. The city, which is located in the "Tunceli Mountainous Region", has been shaped depending on the conditions of the region. Therefore, from the upper scale, the city is located in a very rugged terrain surrounded by the Karasu River in the west, Peri Stream in the east, Keban Dam in the south (Murat River before that) and Mercan and Karagöl mountains in the north. Having such clear borders and the rugged nature of the terrain has caused the city to be away from external factors just like an island (Figure 1).


**Figure 1:** Location map of the City of Tunceli

**Photo 1:** Situation of the City from the Northern Slopes of the Yeni Mahalle Neighborhood

The largely mountainous structure of the city and difficult transportation opportunities made Tunceli's reach to the surrounding cities very limited until the 1950s, and for many years, a closed and stagnant economic structure prevailed within the city (Tunceli Municipality, 2017). Therefore, throughout the city, agriculture and livestock have been the most important economic activities for a long time and even these activities have been sparse and scattered due to land conditions and could not be done efficiently. Apart from the shortcomings of the basic dynamics that ensure the start and development of industrial and commercial activities, past security problems have had an impact on the sluggish economic structure of the city.

Tunceli is a city with no historical past and whose foundation was laid during the Republican era of Turkey. With the law no. 2885 adopted on 25 December 1935, it was decided to establish a new city called Tunceli and the area, which remained almost empty until then with a village called Mameki nearby, was chosen for the city area of establishment. The main factor in choosing this location was its proximity to the road connecting Erzurum, Elazığ and Malatya provinces to the mediterranean coast to İskenderun port (Tuncel, 2012).

In the city, which was called Mameki-Kalan at the time of its foundation and had a single neighborhood settlement, the number of neighborhoods increased to three in 1967 and the population increased to about 6 thousand. These are Alibaba, Atatürk and Cumhuriyet neighborhoods. In 1975, the population increased to 10 thousand and the number of neighborhoods increased to six with the addition of Moğultay, Esentepe and Yeni neighborhoods. After 1980, military areas and other official institution areas increased within the city. Some of the new settlements and housing areas were built in today's Cumhuriyet neighborhood. Fırat University Vocational School, which was opened in 1983 in the village of Aktuluk, was transformed into Munzur University in 2008 with the addition of various other faculties (Aslan, 2016). Aktuluk village, located at the southern end of the city, and İsmet İnönü, located at the eastern end, are the most recently added neighborhoods.

Looking at the functional characteristics of the neighborhoods, we can say that there are generally low-rise residential areas in Alibaba, Aktuluk, İnönü, Yenimahalle and Cumhuriyet neighborhoods. These residential areas are accompanied by military areas and lodgings in the Cumhuriyet neighborhood, the university campus area in the Aktuluk neighborhood and the organized industrial zone in the İnönü neighborhood. In the Moğultay neighborhood, where the central commercial complexes are located, commercial areas, a bus station and some public organizations accompany a small number of residential areas. Atatürk Neighborhood is located on the main development axis of the city, which has a new and developing central business district, small industrial site and numerous public institutions. New and high-rise buildings are more common in this neighborhood. Apart from the Mameki bridge, which provides the road connection between the old and new residential areas of the city, the Batman bridge, completed in 2018, provides the connection to the public housing areas in the Esentepe neighborhood. There are also bridges to the west of the Mameki bridge that pedestrians can use.

Among the elements that restrict the areas suitable for urban settlement are those that hinder the development of the city to the north. The risk of rock fall due to the downpours at the foot of Döndül Tepe and Zeytin Tepe, the risk of flooding along the south coast of Munzur River, and the risk of landslides in Esentepe, Cumhuriyet and Alibaba Neighborhoods have led to a ban on construction in these areas (Canpolat, 2019). In addition, large military areas in the neighborhoods of Atatürk and Cumhuriyet, Uzunçayır Dam in the east side of the city and the high sloping and fragmented land structure around the city are other restrictive factors. In this context, the city is a structure whose growth is limited both from a physical and social point of view.

Based on one study, it is predicted that the most important economic development area in the city, in particular, will be the construction industry. The demands of Tunceli residents living in Turkey and abroad to have a second residence in the city, the increase in demand for housing with the establishment of Munzur University and the need to renovate older and low-rise buildings due to urban transformation are expressed as the main determinants of the growth in the construction industry (İzmen, 2014).

## DATA AND METHOD

The urbanization process basically refers to the capture of non-urban lands, that is, natural and rural areas. Therefore, the development of the city depends on the opportunities and resources offered by these lands. In other words, urbanization gains strength and continuity if natural environment factors (distance, geological structure, landforms, water and soil resources, vegetation, etc.) provide sufficient opportunities (Karadağ & Koçman, 2007; Li vd. 2018).

In terms of appearance, cities are densely inhabited areas with man-made structures, characterized by constant growth. The concepts of urban growth, urban development, urban growth and, in part, urban sprawl are often used as synonyms. Urban growth refers to the increase and concentration of urban lands (Mendiratta and Gedam, 2018; Pratyush et al. 2018; Viana et al. 2019).

The study aims to predict the growth of Tunceli as a small-scale and young city, thereby helping to identify suitable places for planning. In this context, by using the positional characteristics of the housing areas built since the establishment of Tunceli, the study estimates in which direction or towards which areas the city can expand to and which areas are suitable for urbanization. Among the important algorithms used in machine learning for prediction processes, "support vector machines", "artificial neural networks" and "random forest" algorithms were used comparatively.

The dependent variable of the study consists of the housing areas available as of 2008. The reason why residential areas are preferred instead of the whole city is that there is a potential settlement area problem depending on the restrictive characteristics of the city. Again, the use of all housing areas instead of the residential areas that have developed in the last 10 or 20 years is due to the fact that the housing areas built in the recent period have a homogeneous structure. In this context, the study aims to project where the areas suitable for housing construction can be in all settlement areas preferred since the day it was established within the scope of the challenging physical characteristics of the city, at the points where the conditions are forced to the maximum. Of course, there are functional areas in the city other than residential areas, but there is still a need for residential areas in Tunceli as a priority for the development of the city. For this reason, the framework of the study consists of the identification of areas suitable for urban development, rather than making estimations about it. Thus, depending on the current development pattern of the city, which has a history of about 60 years, it is aimed to identify possible settlement areas using machine learning algorithms.

The boundary of the research area was determined by including the villages outside the administrative boundaries of the city which cover the possible development areas in the vicinity. In this limited area, for the determination of land suitable for urban growth, the curves in the sheets with 25 thousand vectors belonging to the General Directorate of Maps (HGM) were turned into points and were interpolated, 10 meters apart to determine the elevation. This digital elevation model (DEM) was used for the purpose of this study. Again, by using DEM, slope, slope exposure and topographic humidity index variables were created. The distance to the main roads, other roads and the dam was obtained using Open Street Map data (GEOFABRIK, 2022). The distance to educational and healthcare institutions was determined from the master plan and the distance to the central business districts was determined according to the implementary development plan explanation report and the observations in the field (Figure 2).

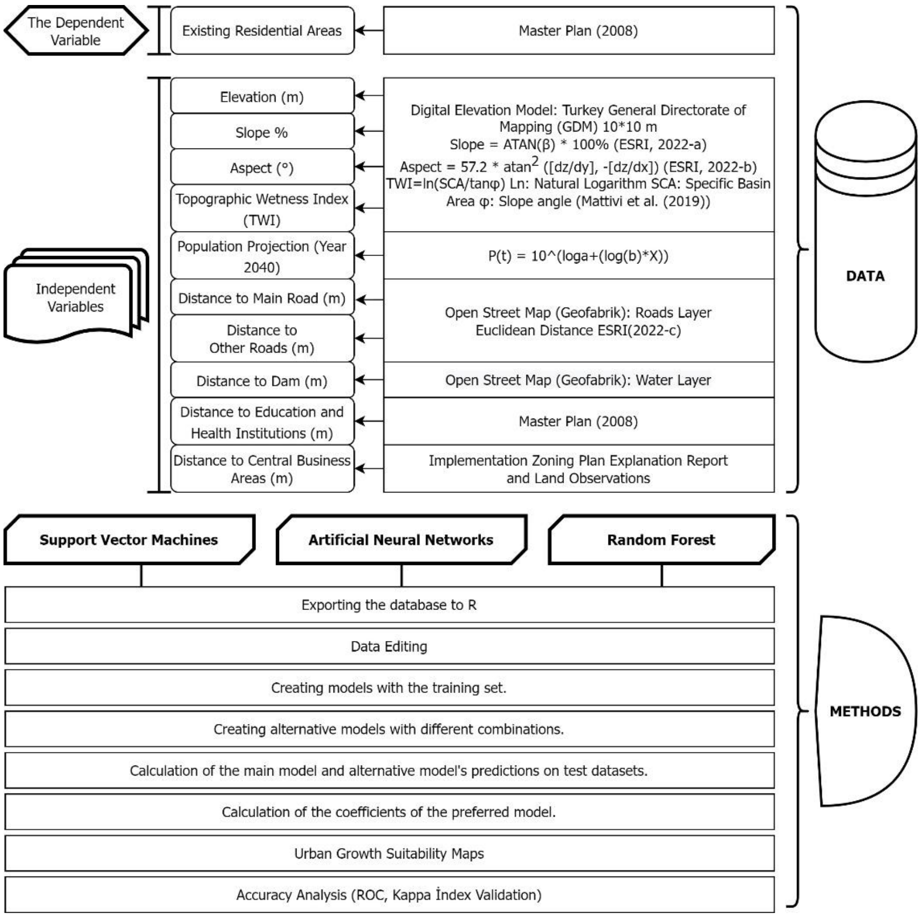



**Figure 2:** Flow chart of urban growth suitability analysis for models

Three different methods were used comparatively to create maps of land suitable for urban growth. All raster images belonging to the variables used in the methods are in the WGS 1984 UTM Zone 37N projection and have a size of 10*10 m. All three models are common methods used in machine learning. For the implementation of the models, the point data (pixel count in total 1666*1824) the dependent and independent variables were separated as training and test data and this data was transferred to the R Studio program. By trying different combinations of each model, the model that better predicted the test data was applied to the entire data set, the predictions were realized and the result maps were created (Figure 2).

## Parameters Affecting the Areas Suitable for Urban Settlement in the City of Tunceli

Settlements have a way of life like any living thing. Settlements are born, grow and finally die in terms of the same principles in living beings (Tunçdilek, 1986). In this respect, as Thompson (1966:26) said, explaining the organic form, "the changes in form of any part of matter, whether alive or dead, seen in its movements and growth, can be described in every case by the act of a force. That is, the form of an object is the diagram of forces acting on that object". We can derive from this judgment that the formal properties of settlements are a living result of the relationship of spatial and other forces to each other.

Apart from the natural environment features such as the landform, climate, water and soil properties that have an impact on the distribution and establishment places of settlements, human factors such as proximity to the dense transportation network and having a large hinterland are important as well for their development (Özdemir, 1996). However, the city of Tunceli has an unsuitable location neither in terms of the natural environment characteristics that affect the place of its establishment nor in terms of human factors necessary for its development. The location of the city was chosen and established based on strategic objectives, especially related to security. Therefore, this "difficult city", which was established due to compulsory conditions and by forcing the conditions, has been around for approximately 75 years, far from sight and has not shown a significant development in terms of urbanization and urban functions.

In terms of the development of the settlement, the city, which is within the Tunceli mountainous region, was shaped based on the conditions of the region. Therefore, when viewed from the upper scale, it is located in a very rugged terrain surrounded by Karasu in the west, Peri Suyu in the east, Keban Dam in south, Mercan and Karagöl mountains in the north. The fact that its borders are so sharp and its rugged structure due to the terrain conditions has caused it to remain away from external influences just like an island.

One of the important parameters affecting urban growth the interaction with other settlements. Although it is necessary to add the distance to the surrounding provincial and district centers to the models in order to demonstrate this interaction, this variable has not been formed in this case, due to the isolated structure of our city. As a matter of fact, Tunceli's interaction is weak not only with other cities, but even with its own districts. It is noteworthy that Ovacık and Nazımiye districts have interactions only with Tunceli city, while Çemişgezek, Pertek and Hozat districts have interactions with Elazığ city and Pülümür district has most interactions with Erzincan city.

In order to determine the areas suitable for urban settlement in the studied area, a large number of variables that were thought to be effective were tested while creating the models and in the end, 11 variables with dominant characters were selected among them (Figure 3). The variables in question and the considerations that required their inclusion in the models are as follows:

***Elevation:*** In order to ensure integration, which is one of the most important elements of the urban economy, the physical conditions in the settlement must be suitable to the development of various functional areas. Especially in small and medium-sized cities, the proximity to the central business districts is an important attraction element, so the increase in elevation and slope and hilliness has been marked as an inverse factor to urban growth.

***Slope*:** One of the important elements affecting the cost of making residential buildings and creating urban function areas is the slope of the land. Since slope also affects the erosion and fragmentation of the land, it is an element that is primarily taken into account in the distribution of land suitable for settlement in cities with rugged terrain. In a study examining the geomorphological factors (slope, elevation and exposure) that have an impact on the historical development of the city of Tunceli, it is emphasized that the most effective variable for each of the three periods evaluated is the "slope" (Esen, 2021).

***Exposure*:** Exposure, which is an element of climatic attraction in the choice of the location of residential buildings, is a categorical variable. In cases where categorical/nominal variables are to be included in the model, either each category in the form of a "dummy variable" should be used as a binary form (yes/1, no/0) or it should be ensured that they are entered into the model by a method such as the density approach (Patriche et al., 2015). "Exposure", which is the only nominal variable among the existing variables, was added to the models both as a dummy variable and as a continuous data with the density approach, but since it did not make a significant difference as a dummy variable, only the continuous data aspect was used.

***The Topographic Wetness Index (TWI)*:** describes a cell's tendency to accumulate water using numerical elevation models. The main variables used in the index are the basin area and slope (Mattivi et al. 2019). In the physiographic structure of Tunceli city, hydrographic elements have an important place as well as morphological elements. As a matter of fact, the city is located on a land torn apart by numerous streams that pour into Munzur River and Peri River. For this reason, the areas suitable for the development of the city are concentrated on fragmentary plains and terraces, which are allowed by hydrographic units. Due to the fact that v-shaped valleys with high slopes are generally not suitable for settlement, areas far from these elements are effective in choosing a location. In order to include the effect of the hydrographic elements that cause the emergence of this fragmented structure, over to urban growth in the model, the topographic humidity index instead of distance to the streams was added to the models as a more sensitive variable.

***Population Projection*:** The critical element for determining how much the city will expand in which parts. Urban areas will increase in relation to the direction and magnitude of this change. In this context, in order to add the estimate for the urban development axis to the model, population projections were made on a neighborhood basis. For this projection, the population projection of the villages that are likely to develop in and around the existing neighborhoods of the city until 2040 was calculated by different methods (Exponential increase, Least squares method, linear regression, etc.), but among them, the formula of logarithmic increase which has been the highest in terms of prediction accuracy rate was used. For this, firstly, population data on the basis of neighborhoods and villages obtained during the address-based population registration system period (2007-2021) were compiled. Then, the population for 2021 was estimated using the logarithmic population projection formula from these data between 2011 and 2016. The accuracy rate of these estimates was deemed appropriate and the population projection for year 2041 was created based on the data between 2011-2021. The results obtained were transferred to the neighborhood-village centers in Arcmap software and converted into raster data using the Inverse Distance Weight (IDW) tool.

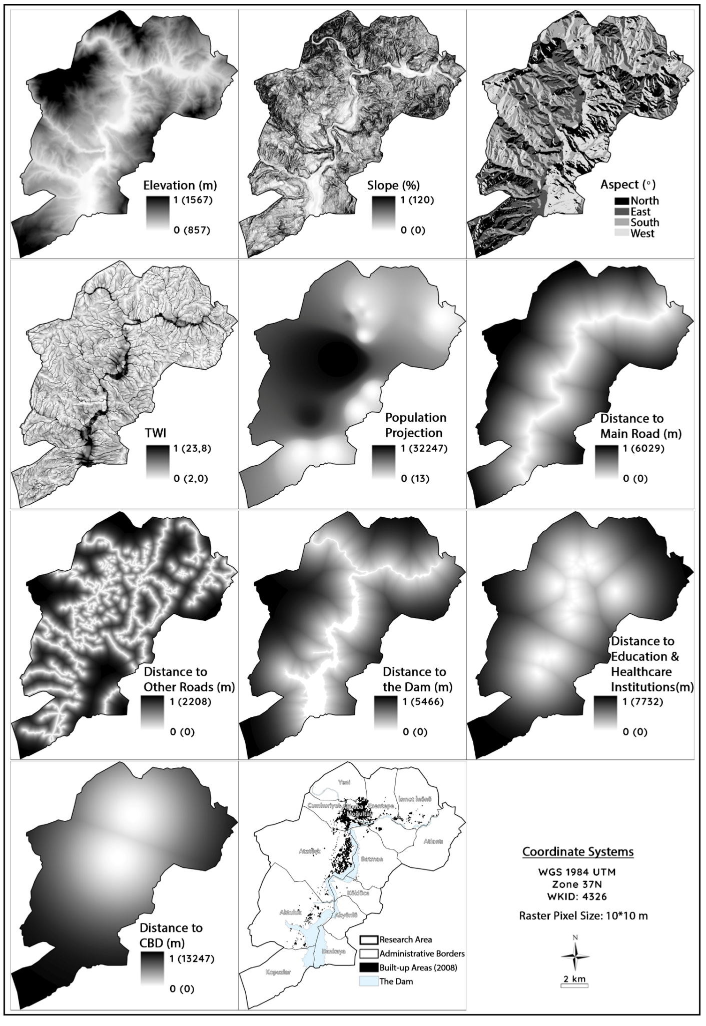



**Figure 3:** Independent Variables Used in the Determination of Lands Suitable for Urban Growth and the Distribution of Housing Areas in 2008

***Distance to the Main Road*****:** Although the city does not have an airport and rail connection, the road connection with the surrounding cities is also essentially not very good. The most important connection with the city of Elazığ, which is located in the area of influence, is the ferry services carried out through the Keban dam. The highway connected to Erzurum and Erzincan via the Pülümür Stream can become problematic with snowfalls, especially in winter. However, the impact of these roads on the current urban growth is significant. As a matter of fact, the development of the Elazığ and Erzincan highways stands out for the isolated city, breaking its shell. Along this highway route, which also forms the main development axes of the settlement, it is possible to see new residential areas. Therefore, the fact that the lands suitable for urban settlement are especially close to this highway is effective in choosing a place. Thus, the main highway obtained from the GEOFABRIK data was converted into a continuous numerical variable with the tool "Euclidian Distance" in the Arcmap environment and added to the models.

***Distance to Other Roads*****:** Although it is not as effective as the distance to the main road, secondary and tertiary roads in urban transportation axes are another attractive element in the growth of the settlement. These transportation lines, which are described as avenues, streets or alleys, are taken from the same source as the main road data and included in the models as another predictive variable.

***Distance to The Dam*****:** Uzunçayır Dam, which extends to the south and east of the city, can be shown as one of the characteristic elements unique to the city of Tunceli. As a restrictive element, the dam made it even more difficult to settle on high-sloping and fragmented lands to the east of the Munzur River. As a geographical element limiting the settlement, the distance to the dam is an important criterion determining the suitable areas of the city for settlement. The dam area in the master zoning plan was removed and added to the model as one of the distance elements.

***Distance to Educational and Healthcare Institutions*****:** One of the important factors that are considered in the construction of housing in cities and taken into account in the choice of location is the distance to education and healthcare institutions. Due to the fact that both these functional areas are located close to each other in Tunceli and it is not a very large city in terms of its area, these two elements were combined and included in the analysis as a single factor.

***Distance to Central Business Districts*****:** Two central business districts have been identified, one in the first settlement area of Tunceli (Ali Baba Neighborhood) and one in the new settlement area (Atatürk Neighborhood) that emerged after the 90s. In medium – and small-sized cities, the distance to the city center (bazaar) is an effective site selection parameter in the choice of settlement. For this reason, the distance to the central business districts was selected and used as another human factor.

In order to check whether there is a positive or negative correlation between the variables used, a correlation analysis was performed. However, since the available data do not correspond to normal distribution, Spearman correlation method was used. While the correlation coefficients vary between – 1 and 1, 1 indicates the strongest correlation in the positive direction, – 1 indicates the strongest correlation in the negative direction, and 0 indicates a non-existent correlation. Accordingly, while the highest positive correlation was found between the distance to central business districts and the distance to educational and healthcare institutions (0.82), the highest negative correlation (-0.65) was seen between slope and topographic humidity index. The lowest correlation values were observed in population projection and TWI, while the variables that correlate the most with other variables were the distance to the main road and the distance to the dam (Figure 4, Table 1).

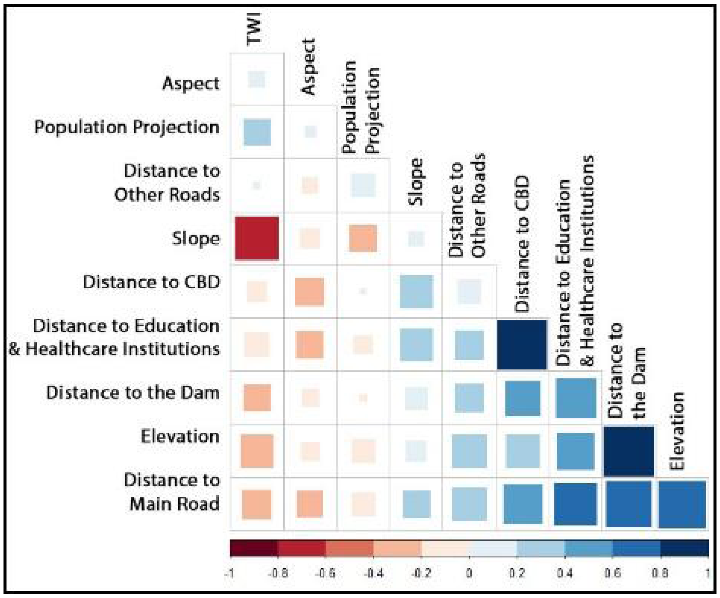



**Figure 4:** The Correlation (Spearman) Coefficients of the Variables Used in the Models

**Table 1:** The Correlation (Spearman) Coefficients of the Variables Used in the Models and their VIF (Variance Inflation Factor)

| Variables | Distance to main road | Exposure | Distance to the dam | Elevation | Distance to other roads | Distance to education and healthcare institutions | Distance to CBD | Population Projection | Slope | TWI | *VIF* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Distance to main road** | 1.00 | -0.22 | 0.73 | 0.76 | 0.25 | 0.62 | 0.52 | -0.18 | 0.39 | -0.28 | *4.6* |
| **Exposure** | -0.22 | 1.00 | -0.10 | -0.11 | -0.09 | -0.25 | -0.27 | 0.05 | -0.13 | 0.09 | *1.1* |
| **Distance to the dam** | 0.73 | -0.10 | 1.00 | 0.87 | 0.17 | 0.55 | 0.42 | -0.02 | 0.28 | -0.25 | *5.5* |
| **Elevation** | 0.76 | -0.11 | 0.87 | 1.00 | 0.14 | 0.48 | 0.37 | -0.18 | 0.39 | -0.37 | *3.6* |
| **Distance to other roads** | 0.25 | -0.09 | 0.17 | 0.14 | 1.00 | 0.35 | 0.37 | 0.20 | 0.07 | 0.01 | *1.4* |
| **Distance to education and healthcare institutions** | 0.62 | -0.25 | 0.55 | 0.48 | 0.35 | 1.00 | 0.82 | -0.11 | 0.28 | -0.20 | *5.9* |
| **Distance to MİA** | 0.52 | -0.27 | 0.42 | 0.37 | 0.37 | 0.82 | 1.00 | 0.01 | 0.19 | -0.13 | *4.1* |
| **Population Projection** | -0.18 | 0.05 | -0.02 | -0.18 | 0.20 | -0.11 | 0.01 | 1.00 | -0.25 | 0.23 | *1.4* |
| **Slope** | 0.39 | -0.13 | 0.28 | 0.39 | 0.07 | 0.28 | 0.19 | -0.25 | 1.00 | -0.65 | *1.7* |
| **TWI** | -0.28 | 0.09 | -0.25 | -0.37 | 0.01 | -0.20 | -0.13 | 0.23 | -0.65 | 1.00 | *1.6* |

The number of variables used to reduce confusion and prevent over-learning should be optimal when preparing models. The number of factors that are overused can make the accuracy values high when the model is created, but it can also fail when it encounters the actual values. Therefore, it is essential that there are no factors that perform the same function between the variables. The index that controls this is expressed as VIF (variance inflation factor) and is expected to be a maximum of 10 (Hair et al. 2014). Accordingly, it can be said that there is no colinearity problem among the variables used (Figure 4).

In the database used to create models, the total number of dependent and independent variables is 13428. Of this, 20% is reserved as test data (2685) and 80% (10743) as training data. The models created with training data were subjected to accuracy analysis with test data.

## Support Vector Machines

In this classification method, the data is first moved to a multidimensional attribute space where it can be separated linearly, using kernel functions, and the classes are separated from each other by an optimal hyperplane. Support vector machines can be used to distinguish between both linearly separable and linearly non-separable classes (Vapnik, 2000). If the classes can be separated linearly from each other, the one whose distance is the largest from the planes separating the classes determines the planes and creates a linear distinguishing function using these planes. If these classes cannot be separated linearly, they are moved to another space of higher dimensions where the classes can be linearly separated using a positive C (cost) parameter and kernel functions that will make the classification error minimal and the distance between the planes maximum. The classification process takes place in this space. Because it performs better, the most commonly used kernel function is the radial-based function (Akar and Görmüş, 2019). Therefore, the model basically serves to classify the elements that cannot be separated in a two-dimensional plane by adding a third dimension. In doing so, the parameters of the hyperplane and the formulation arrangements required for the use of kernel functions must be determined by the user. Meyer (2001:7) states that there are four different kernel functions: linear ($u^\top v$,polynomic($d$), radial-based ($exp\{-\gamma \vee u - v \quad ^2\}$and sigmoid function ($tanh\{\gamma u^\top v + c_0\}$. As a result, in solving a two-class problem that cannot be separated linearly, the parameters of the support vector machines are as follows (Karimi et al. 2019).

$$L(a) = \sum_{i=1}^{n} \quad \alpha_i - \frac{1}{2}\sum_{i=1}^{n} \quad \sum_{j=1}^{n} \quad \alpha_i \alpha_j y_i y_j k(x_i . x_j)$$

"Support vector machines", one of the basic algorithms in machine learning (Osuna et al., 1997), has been used for the estimation and modeling of events in the field of geography such as urban growth (Huang et al., 2010; Karimi et al., 2019; Şahin, 2021), disaster risk analyses (Liong & Sivapragasam, 2002; Aydın and Çelik, 2013), remote sensing (Chapelle et al. 1999; Pal & Mather, 2005; Kavzoğlu & Çölkesen, 2010).

## Artificial Neural Networks

ANN works with the working principle of a human nerve cell. An event encountered or a situation that needs to be perceived comes to the nerve cell with signals. The nerve cell creates an action with these signals. ANN learns by imitating this phenomenon that takes place in nerve cells (Çakır, 2020). This learning process is based mainly on the fact that the available collected data is processed using a predetermined function with link weights, bringing it closer to the desired values. At the end of a training process in which sufficient conditions are met, the performance is determined by the margin of error between the target value and the value obtained (Özcan, 2008). Artificial neural networks, as a nonlinear machine learning method, are convenient algorithms for solving the complex structure of spatial phenomena, as they can be used for both regression and classification problems.

In an ANN cell, there are 5 layers: inputs, ($X_1$, $X_2$, $X_3$ ...), weights ($W_1$, $W_2$, $W_3$ ...), coupling function, activation function ($\varphi$ ) and output (y) (Haykin, 1999). In the input layer, there are as many nerve cells as there are inputs from outside. Weights indicate the importance of the information coming into the cell and its effect in a positive or negative way. With the merge function, the inputs and weights are multiplied first, then the NET (Network) value is calculated. The activation function determines the output that the network will produce by processing the extreme values obtained at the previous stage. This is done using one of the different functions such as sigmoid, hyperbolic tangent, gaussian. Finally, the information obtained in the output layer will be input to the cells or will come out of the cell as direct information (Çakır, 2020). The best artificial neural network model is considered to be the network with the least amount of total output error, and the total output error adopted is expressed as 'E'. In the model, 'E' is calculated as follows.

$$E = \frac{1}{S}\sum_{k=1}^{S} \quad E_k = \frac{1}{S}\sum_{k=1}^{S} \quad \sum_{j=1}^{m} \quad \left(y_{kj} - d_j\right)^2$$

Where S is the number of samples in the formula, Ek is the output error for each training iteration, $m$ is the number of output layer neurons, $y_{kj}$ is the output according to the input example. $d_j$ (j = 1, 2, ..., m ) is the scale (Tong et al. 2010: 1489).

There are different use examples of artificial neural networks in various branches of science. This model was used widely for various factors in geography such as land use and change (Liu & Lathrop, 2002; Cheng et al. 2003), urban growth, expansion and change (Alkheder, 1999; Pijanowski et al. 2001; Burbridge & Zhang, 2003), appropriate location selection (Sung et al. 2001; Park et al. 2011).

## Random Forest

Batch classification methods are learning algorithms that produce multiple classifiers instead of one classifier and then classify the new data with votes from their estimates. The most commonly used batch classifiers are Bootstrap, Acceleration, and Random Forest. In the Bootstrap algorithm, multiple preloaded training datasets are created to train a classifier, interchangeably, using the original training data set, and a tree is generated for each preloaded training dataset. Successive trees are independent of the previous one and the largest vote is taken as the basis for the prediction (Akar and Güngör, 2012).

The random forest (RF) method is an improved version of decision trees. In RF, the decision tree is enlarged by adding the randomness property, without placement in the bootstrap. In the model, each node is separated using the best division among all variables. In a random forest, each node is allocated using the best among a subset of randomly selected predictors (arguments) on that node. This planned strategy performs better compared to many other classes, including discriminant analysis, support vector machines, and artificial neural networks. It is also a durable method against the problem of overfitting (Breiman, 1996; Breiman, 2001; Liaw & Wiener, 2002).

In the Random Forest algorithm, the user is asked for two parameters. These are the number of trees (N) and the number of variables (m) to be used in each node to create the tree structure (Breiman, 2001). After the parameters are selected, bootstrap is used to create a sample for each tree. Then tree development begins for each sample. In each node, the best branching is determined by using the variables randomly selected as 'm' out of all variables (Erdem et al.2018). By voting between the designated trees and their branching, the tree with the highest number of votes is assigned to a class.

The most commonly used attribute selection criteria in decision tree branching are the Knowledge Gain Rate index and the GINI index. The random forest classifier uses the GINI index as an attribute selection measure that measures the purity of an attribute by classes. By randomly selecting a case for a given training set 'T' and specifying that it belongs to the class Ci, the Gini index is written as:

$$\sum\sum j \neq i\left(f(C_i,T)/|T|\right)\left(f\left(C_j,T\right)/|T|\right)$$

In the equation, T is the training data set, Ci is the class to which the pixel belongs, indicates the probability that the selected pixel belongs to the Ci class (Pal, 2005).

The use of the random forest model in the field of geography, like the other two algorithms, is focused on land use (Gislason et al. 2006), urban growth and change (Jun, 2021; Frimpong & Molkenthin, 2021), disaster risk forecasting (Stumpf & Kerle, 2011; Chen et al. 2012) and remote sensing (Akar & Güngör, 2012; Erdem et al. 2018).

## RESULTS AND DISCUSSION

Urban growth forecasting is actually a matter of selecting location. In this context, there are many algorithms that model this process, which includes determining the optimal location within the current land conditions. The aim of site selection analysis is to identify the best site for some activity given the set of potential (feasible) sites. In this type of analysis all the characteristics (such as location, size, relevant attributes, etc.) of the candidate sites are know. The problem is to rank or rate the alternative sites based on their characteristics so that the best site can be identified (Malczewski, 2004: 4). In terms of models, the starting point is to interpolate the probability of similar situation in other places based on the common properties of known points.

### Modeling of Land Suitable for Urban Growth with Support Vector Machines Model

The implementation of the model was carried out by downloading the 'e1071' package in R Studio and operating the normal support vector machines procedure. In the model, the radial function was determined as the classifier core function, the cost value (C) was 2.5, and the gamma value was 0.75. 25 different combinations were tried to determine the cost and gamma values. The lowest error rate in these combinations was obtained at these values. The cost value usually ranges from 1 to 10. This value is kept small for the model to maximize the margin and be more tolerant for misclassifications.

$$\text{SVM\_Model <- } svm(grid\_code \sim ., data = train, kernel = "radial", cost = 1, gamma = 0.25$$

The results obtained by the model from the training data and the results obtained from the test data are consistent with each other. Both training and test data have shown successful performance with 96% accuracy. The kappa value indicating the consistency of the model indicates that the errors in the classification were reduced by 91% (Table 2).

**Table 2:** Cross-Validation Matrix of Training and Test Data of the Support Vector Machines Model

| Data Type | Type Of Class | 0 | 1 | Class. Error | Accuracy value | Kappa value |
|---|---|---|---|---|---|---|
| **Training Data** | 0 (Non-urban) | 5012 | 33 | 0.7 | 96% | 0.92 |
| | 1 (Urban) | 378 | 5320 | 6.6 | | |
| **Test Data** | 0 (Non-urban) | 1260 | 27 | 2.1 | 96% | 0.91 |
| | 1 (Urban) | 87 | 1311 | 6.2 | | |

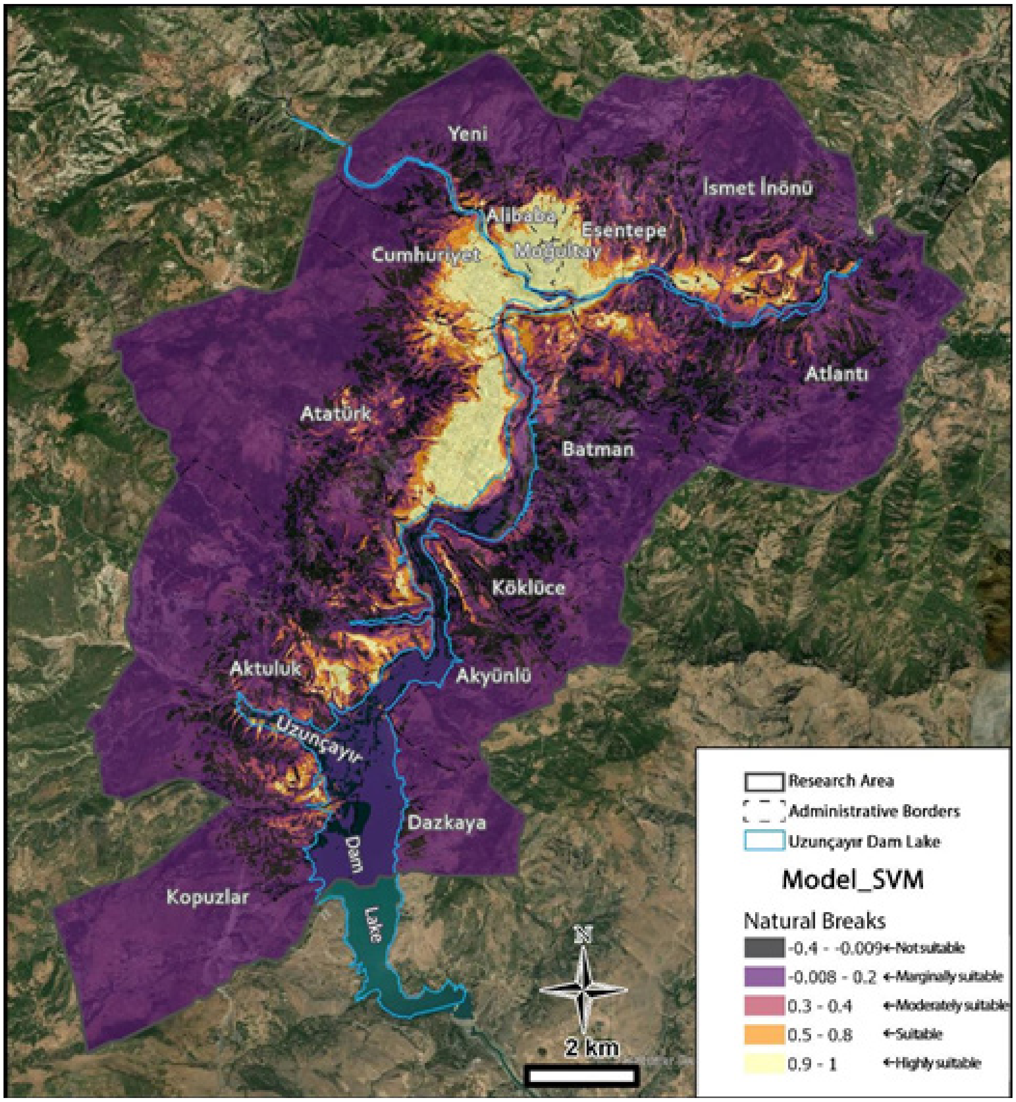



**Figure 5:** The Resulting Map of Support Vector Machines Model

In the results map, the outer line on the close periphery of the existing housing areas was identified as the most suitable area for urban growth. In this context, it is foreseen that the growth of the city will first take place in empty parcels within the settlement area, then this region will be limited to the periphery. Areas that are slightly suitable and not suitable for settlement appear to be the areas far away from the city, where the slope and topographic humidity increase. The most important problem that draws attention in the result map of the model is seen in the reservoir area. Most of this area has been included as slightly suitable lands instead of the not suitable class, and some parts of it have been shown as suitable and very suitable lands for growth. (Figure 5).

## Using the Artificial Neural Network Model for the Urban Growth Prediction

Having the ability to learn with the examples given, the ANN consists of a large number of interconnected nodes known as neurons. Learning takes place by following a non-linear route through the nodes. Because there are a lot of parameters used in the creation of the model (error, activation, number of repetitions, threshold value, etc.), creating and trying different combinations of them slows down the process quite a lot. In addition, the creation time of each different trial takes much more time compared to the other two models.

The implementation of the model was carried out by downloading the ‘neuralnet’ package in R Studio and setting up an equation with 10 hidden layers. A forward-fed algorithm (rprop) was used to create the model because it gives faster results. The learning threshold was .05 and the maximum number of attempts was 1e+06.

In the network diagram, the positive weights between the layers are shown as black lines and the negative weights are shown as gray lines. The line thickness is proportional to the absolute size of each weight (Beck, 2018). In the diagram, I(input) refers to the independent variable inputs, B (Bias) refers to the bias, H (hidden) refers to the hidden layers, and O (output) refers to the dependent output variable. In this context, while the distance to the main road, population projection, distance to educational and healthcare institutions and distance to other roads contributed to the model in terms of positive weight, distance to the dam, elevation and TWI variables contributed to the model in terms of negative weight (Figure 6).

$$\text{ANN\_Model} <\!\!-neuralnet(grid\_code \sim ., data = train, algorithm = "rprop+", hidden = c(10), threshold = 0.05, stepmax = 1e+06)$$

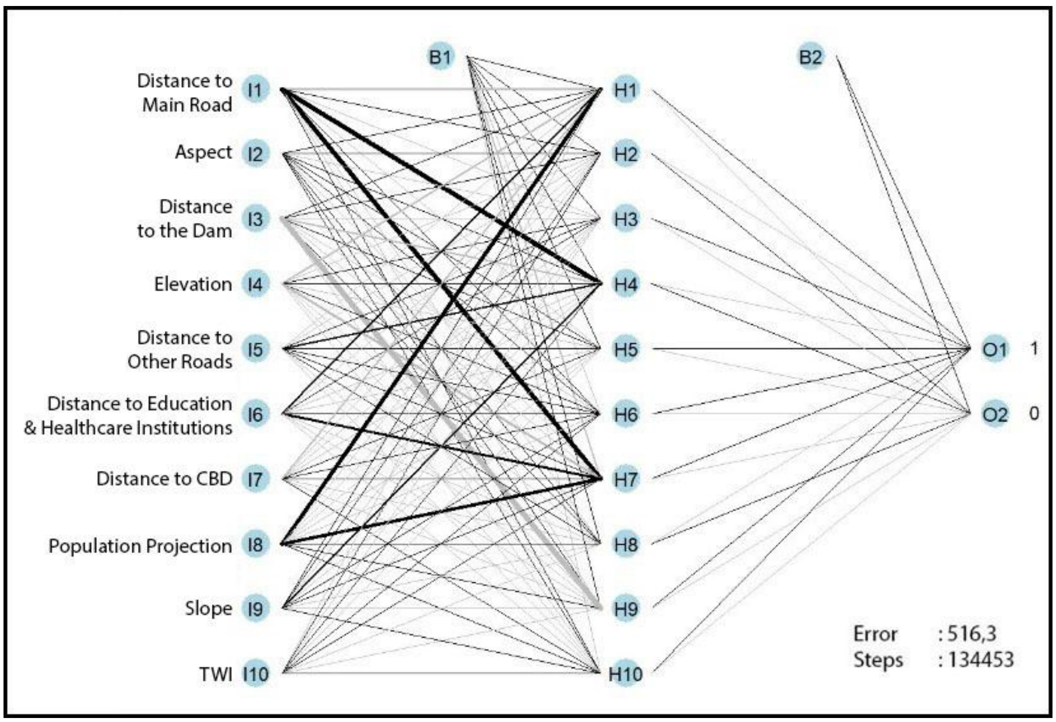



**Figure 6:** ANN Model Network Structure

The results of the accuracy and consistency analysis of the model show similar results to SVM. Also, the estimates obtained from the training data and the estimates obtained from the test data show consistency with each other. The overall accuracy rate turned out to be lower (95%) than the previous model. The values obtained in terms of kappa value, which shows the consistency of the model, indicate that the classification is free from errors around 90% according to random estimates (Table 3).

**Table 3:** Cross-Validation Matrix of Training and Test Data of the Artificial Neural Networks Model

| Data Type | Type Of Class | 0 | 1 | Class. Err. | Accuracy value | Kappa value |
|---|---|---|---|---|---|---|
| Training Data | 0 (Non-urban) | 4946 | 444 | 8.2 | 94% | 0.89 |
| | 1 (Urban) | 159 | 5194 | 3.0 | | |
| Test Data | 0 (Non-urban) | 1259 | 88 | 6.5 | 95% | 0.90 |
| | 1 (Urban) | 42 | 1296 | 3.1 | | |

The ANN result map seems to have simpler and clearer distinctions compared to the SVM. In this context, the areas belonging to the existing urban land were primarily shown as the most suitable class for settlement / growth, and the lands with low fragmentation (topographic humidity) in the immediate vicinity were determined as lands suitable for development. It is worth noting that there is a potential for development in the form of a belt, especially in the west of the Atatürk neighborhood, which can be considered partially suitable. In addition, the lands suitable for settlement around the existing public housing areas around Esentepe Neighborhood and Batman Village are designated as suitable and partially suitable areas. A similar situation was observed in the direction of Pülümür road to the east of Esentepe neighborhood. One of the most important advantages of the model over the SVM is that the dam lake and the river valleys are included in the class of unsuitable and slightly suitable lands, with some exceptions (Figure 7).

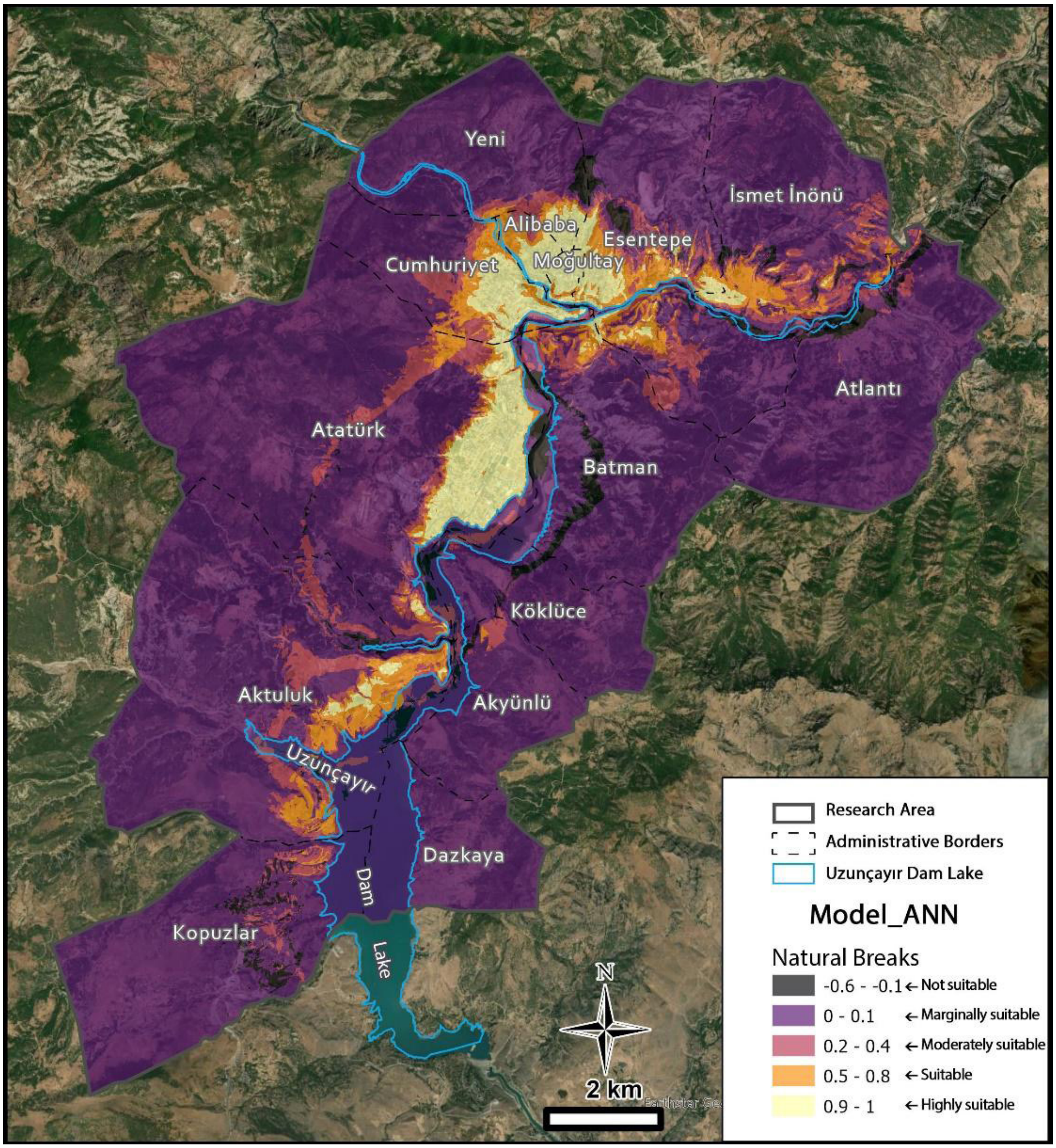



**Figure 7:** The Resulting Map of Artificial Neural Networks

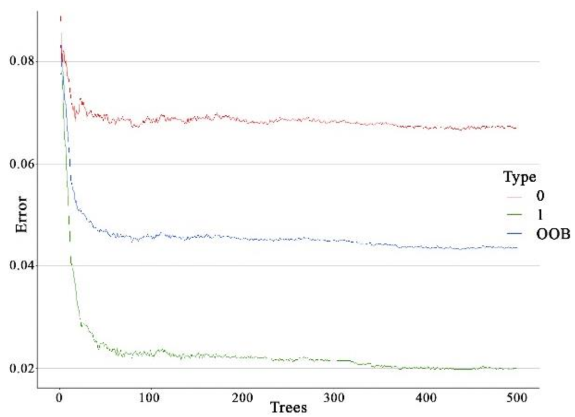
## Urban Growth Forecasting Model with Random Forest Model

The advantage of the Random Forest model, which performs best compared to other models, is the choices it makes randomly. In this way, it remained resistant to the problem of over-fitting and produced better predictions. For the implementation of the model, the "randomForest" package was downloaded in R studio and classification was selected as the model type. The number of trees in the model was determined to be 500 and the number of variables to be divided was determined to be 3.

The most important criterion determining the performance of a random forest model is the number of trees. As can be seen from the OOB, 250 trees gave optimal results. However, in order to make the model predict better, it was preferred to have 500 trees (Figure 8).

RF_Model <- $randomForest(formula = grid\_code \sim ., data = train,\ method = \text{"}rf\text{"}, metric = \text{"}Accuracy\text{"}, tuneGrid = tuning, importance = TRUE, nodesize = 3, ntree = 500,\ trControl = trainControl,\ method = cv, number = 5,\ proximity = TRUE))$

The OOB Overall Prediction Error Rate of the Model is 4.64%. So the correct prediction rate is 95%. This was also confirmed in the cross-validation matrix. As a matter of fact, in the results of the model created with training data, the overall accuracy rate was determined as 95%, the kappa value was determined as 0.90, while the accuracy rate increased to 96% and the kappa value to 0.93 in the test data. With these values, the highest accuracy and classification values were obtained according to support vector machines and artificial neural networks (Figure 8).




| Data Type | Type Of Class | 0 | 1 | Classification Error | Overall Accuracy Rate | Kappa value |
|---|---|---|---|---|---|---|
| Training Data | 0 (Non-urban) | 5028 | 362 | 6.7% | 95% | 0.90 |
| | 1 (Urban) | 137 | 5216 | 2.5% | | |
| Test Data | 0 (Non-urban) | 1290 | 29 | 2.2% | 96% | 0.93 |
| | 1 (Urban) | 57 | 1309 | 4.1% | | |

**Figure 8:** Change of Overall Error Rate in a Random Forest Model in accordance with the Number of Trees and Cross-Validation Matrix of Training and Test Data

In the results of the independent variables used in the random forest model in order of importance, the variables that contributed the most to the model were determined as the distance to education and healthcare institutions and the distance to central business districts. The variables with the least impact were determined as distance to other roads outside the main road in the city and exposure values (Table 4).

**Table 4:** Hierarchical Importance Ranking of Variables in the Random Forest Model

| Variables | 0 | 1 | Average Reduction Accuracy | Gini index of Average Reduction |
|---|---|---|---|---|
| Distance to Roads | 13.1 | *104,0* | 102,8 | 1473.6 |
| Distance to MİA | 14.7 | *100,3* | 87.8 | 1390.8 |
| Slope | 8.4 | *49.2* | 51.4 | 110,1 |
| Distance to the dam | 2.5 | *48.8* | 51.3 | 467,6 |
| Elevation | -8.9 | *48.5* | 49.9 | 285,5 |
| Population Projection | 12.4 | *46.6* | 49.7 | 220,1 |
| Distance to main road | 12.9 | *44.3* | 48.2 | 697,5 |
| TWI | -0.06 | *43.3* | 43.3 | 75.9 |
| Distance to other Roads | -7.4 | *34.6* | 35.4 | 181,9 |
| Exposure | -5.3 | *19.1* | 19.0 | 35.1 |

The first thing that draws attention in the result map of the model is that the places shown as the most suitable lands pass through the outer boundaries of the existing built-up areas. Therefore, the model considers bare lands within these areas as areas suitable for priority development. In other words, it is estimated that the inward stacked growth form in the city will be the first alternative. The class of lands shown as appropriate (0.4-0.7) consists of the immediate vicinity of the most suitable areas and the vacant lands between them. Partially suitable lands (0.2-0.4) consist of areas on the periphery of the lands in the previous two classes and areas close to roads. Non-suitable lands in the area studied (0.06-0.2) indicates areas where the city could expand into by making extra arrangements. In this context, the places between the outer periphery of the city and the fragmented residential areas are included in this class. Among the lands that are not suitable are the whole of the reservoir and river valleys with flows, as well as places away from the urban characteristics and where the elevation and slope increase (Figure 9).

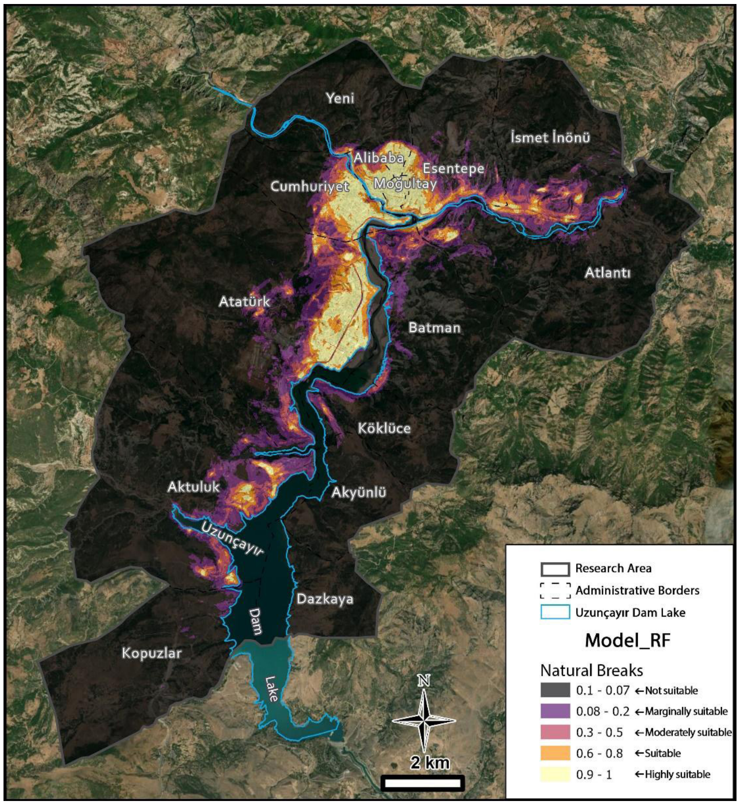



**Figure 9:** The Resulting Map of Random Forest Model

In the selection of the lands in Batman and Köklüce villages by the model, their proximity to the central business districts, the village roads passing through them and the relatively low slope values were the effective factors (Figure 10).

## Comparative Analysis of Models

The RMSE (Root Mean Square Error) value, which is one of the determining criteria for the estimation errors of the models, indicates the standard deviation of the errors made in the estimation. The approximation of this value to 0 means that the error rate of the model is low (Willmott & Matsuura, 2005). In this context, the RMSE value was determined as 0.2 in the SVM and ANN model and 0.17 in the RF model. Thus, although it was understood that the prediction errors of the models are generally close to each other and they give accurate results, it was seen that the lowest error rate was in the random forest model.

According to the results of spatial accuracy analysis of the result maps, the method with the highest prediction ratio was 'Random Forest' (0.997), while 'Support Vector Machines' and 'Artificial Neural Networks' had similar accuracy values (Figure 10). Although there does not appear to be a huge difference between the accuracy results of the models, it was concluded that the RF model with the highest AUC value was the one that overlapped the most with the field observations.

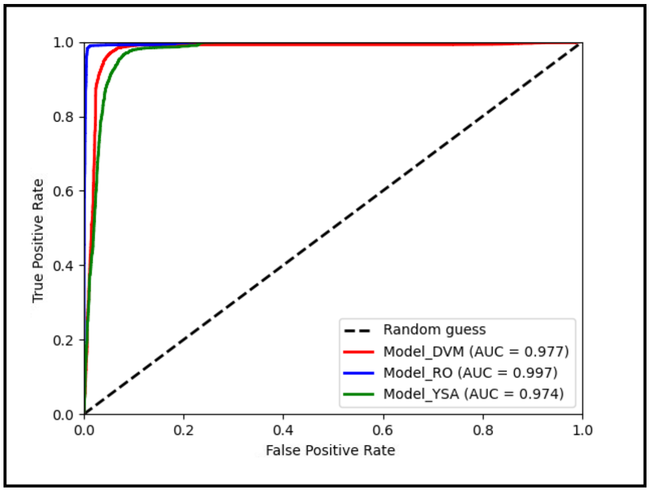



**Figure 10:** AUC Values of Models and Accuracy Analysis

**Table 5:** According to the Result Maps, the Amount of Land Suitable for Urban Growth in Tunceli (km2)

| Availability | SVM | ANN | RF | Average |
|---|---|---|---|---|
| **Highly suitable** | 7.9 | 6.9 | 5.0 | 6.6 |
| **Suitable** | 5.4 | 6.6 | 3.5 | 5.2 |
| **Moderately suitable** | 8.7 | 11.4 | 5.1 | 8.4 |
| **Marginally suitable** | 104,9 | 122,6 | 12.2 | 79.9 |
| **Not suitable** | 24.7 | 4.2 | 125,8 | 51.5 |

The amount of land in the studied area that is most suitable for urban settlement was determined as approximately 6.6 $km^2$ by taking the average of all three models and the sum of suitable and partially suitable land is approximately 13.6 $km^2$. The total amount of land not suitable for urban growth is about 130 $km^2$. The results of the models of support vector machines and artificial neural networks show values close to each other, while in the random forest model, especially the suitable areas in the first 3 steps, have lower values. While there is a parity in SVM and ANN in the surface area of land that is slightly suitable and not suitable for urban growth in the models, there is a significant difference in the RF model in the class that is not suitable (Table 5).

## CONCLUSION

In order for the controlled and planned growth of cities, it is a necessity to conduct suitability analyzes for the correct selection of lands for growth. With these studies, it may be possible to develop infrastructure, take precautions against possible risks and prepare for potential problems by foreseeing the possible development axes of urban settlements. Thus, the paradigm of building resilient cities, one of the important paradigms of the 21st century, can become easier.

Forecasting is not planning. As part of the planning process, this is just one step (Cuhls, 2003). It is therefore intended to assist urban planning studies with the results of this research. As a small-scale city, Tunceli is still at an early stage of development in terms of urbanization. In this context, it would be useful to prevent distorted and irregular urbanization with the right planning and to determine the possible settlement areas correctly.

In addition to the many restrictive physical elements present in the city of Tunceli, the limited opportunities in human factors also make it difficult for the city to expand. In order to identify growth areas, 'support vector machines', 'artificial neural networks', and 'random forests' machine learning models were used to create prediction maps. Based on the critical factors that are/can be effective in the emergence of urban areas built in the city and its immediate surroundings since the 1950s, these models have been developed.

Support vector machines, one of the machine learning methods, transfer the data to a multidimensional plane to predict the dependent variable, where the values are grouped in a linear or nonlinear way and the prediction is made. It is both easier and faster to create a model on support vector machines and perform calibration operations with different combinations.

The model of artificial neural networks, on the other hand, is based on a rather complex process. By simulating human learning with inputs and weighting, the model creates a network of artificial neurons and is able to make nonlinear predictions. An artificial neural network model is an algorithm that requires more time and experimentation to create and calibrate compared to the other two models. However, with numerous arrangements such as the number of hidden layers, the learning threshold value and the number of repetitions required to train neural networks, it provides more accurate results.

Finally, the random forest model is an improved version of the "decision trees" algorithm. By adding randomness to decision trees, the margin of error is reduced. In this model, nodes (trees) are formed to make decisions, voting is made between the designated trees and their branching, and the tree with the highest number of votes is assigned to a class. Of the three algorithms used, the random forest model was the most resistant to the colinearity problem, it was faster and less calibration dependent, and had the highest accuracy and lowest error rate. In addition, the spatial equivalent of the result map created with the weighing coefficients obtained was fairly satisfactory. The fact that there are areas in the result map that are not included in the most suitable areas class despite being within the city, indicates that the distinctive aspect of the model is high. In this model, distance to educational and healthcare institutions and distance to central business districts were determined to be the most important variables. Other variables, on the other hand, have a degree of weighting close to each other.

The hydrographic elements that accompany the landforms of the city and the public spaces in addition to the forests, create significant obstacles in the expanson of the settlement and the creation of a compact structure. Depending on the current conditions of the region, the areas of the city suitable for growth were envisaged as stock lands within the city, and later on, as favorable areas in its immediate vicinity.

## REFERENCES


Akar, Ö. & Güngör, O. (2012). Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi* , (106), 139-146 . https://doi.org/10.9733/jgg.241212.1t

Akar, Ö., & Görmüş, E. T., (2019). Göktürk-2 ve Hyperion EO-1 uydu görüntülerinden rastgele orman sınıflandırıcısı ve destek vektör makineleri ile arazi kullanım haritalarının üretilmesi. *Geomatik*, 4(1), 68-81. https://doi.org/10.29128/geomatik.476668

Alkheder, S., (1999). Urban growth simulation using remote sensing imaginary and neural networks, retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.8120&rep=rep1&type=pdf

Aslan, S. (2016). Şehir içi arazi kullanım yönünden Tunceli. (*Yüksek Lisans Tezi*, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara). s.

Aydın, M., & Çelik, E. (2013). Destek vektör makineleri ve yapay sinir ağları kullanarak türkiye'deki tehlikeli hava durumlarının uydu görüntüleri ile erken tespiti. *21st Signal Processing and Communications Applications Conference (SIU 2013) : North Cyprus Turkish Republic, 24 – 26 April 2013*, Haspolat.

Beck M. W. (2018). Neuralnettools: Visualization and analysis tools for neural networks. *Journal of statistical software*, 85(11), 1–20. https://doi.org/10.18637/jss.v085.i11

Breiman , L. (1996). Bagging predictors. *Machine Learning*, 24 (2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning,* 45 (1), 5–32.

Brennan, C., & Hoene, C. (2003). Research brief on Americas cities: Demographic change in small cities, 1990-2000. *National League of Cities*. Washington, DC.

Burbridge, S., & Zhang, Y. Z. Y. (2003). A neural network based approach to detecting urban land cover changes using Landsat TM and IKONOS imagery. *2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas* (pp. 157-161). IEEE.

Çakır, F. S. (2020). *Yapay sinir ağları matlab kodları ve matlab toolbox çözümleri*. Nobel Yayıncılık, Ankara.

Çalışkan, O.Ç. ve Tezer, A. (2018). Türkiye kentleşmesinin çok boyutlu sürdürülemezliğinde yeni bir yol arayışı: orta ölçekli kentler üzerinden kır-kent dayanışma ağları, *Planlama (Ek 1)*: 73-90. https://doi.org/10.14744/planlama.2018.66376

Canpolat, F. A. (2019). Tunceli kentinin nüfus özellikleri. *Uluslararası Bilimsel Araştırmalar Dergisi (IBAD),* 4(2), 183-200. https://doi.org/10.21733/ibad.537457

Chapelle, O., Haffner, P., & Vapnik, V. N. (1999). Support vector machines for histogram-based image classification. *IEEE transactions on Neural Networks*, 10(5), 1055-1064. https://doi.org/10.1109/72.788646

Chen, J., Li, M., Wang, W. (2012). Statistical uncertainty estimation using random forests and its application to drought forecast. *Mathematical Problems in Engineering*, https://doi.org/10.1155/2012/915053

Cheng J., & Masser, I. (2003). Understanding urban growth system: Theories and methods. *In 8th International Conference on Computers in Urban Planning and Urban Management*, Sendai City, Japan, pages229–237.

Church, R. L. (2002). Geographical information systems and location science. *Computers & Operations Research*, 29(6), 541-562. https://doi.org/10.1016/S0305-0548(99)00104-5

Cohen, B. (2004). Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts, *World development*, 32(1), 23-51. https://doi.org/10.1016/j.worlddev.2003.04.008

Cuhls, K. (2003). From forecasting to foresight processes—new participative foresight activities in Germany. *Journal of forecasting*, 22(2-3), 93-111. https://doi.org/10.1002/for.848

Dadashpoor, H., Azizi, P., Maghadasi, M. (2019) Analyzing spatial patterns, driving forces and predicting future growth scenarios for supporting sustainable urban growth: evidence from Tabriz metropolitan area, Iran, *Sustainable Cities and Society*, 47, 1-32. https://doi.org/10.1016/j.scs.2019.101502

Erdem, F. , Derinpınar, M. A. , Nasırzadehdızajı, R. , Oy, S. , Şeker, D. Z. & Bayram, B. (2018). Rastgele orman yöntemi kullanılarak kıyı çizgisi çıkarımı İstanbul örneği, *Geomatik*, 3(2), 100-107. https://doi.org/10.29128/geomatik.362179

Esen, F. (2021). Jeomorfolojik özelliklerin Tunceli şehrinin gelişimine etkileri. *Jeomorfolojik Araştırmalar Dergisi* , (7) , 109-131 . https://10.46453/jader.948540

ESRI (2022-a). How slope works. Retrieved from, on 10 February 2022, https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-aspect-works.htm

ESRI (2022-b). How aspect works. Retrieved from, on 10 February 2022, https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-analyst/how-aspect-works.htm

ESRI (2022-c). Understanding euclidean distance analysis, Retrieved from, on 10 February 2022, https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/understanding-euclidean-distance-analysis.htm

Fridemann, J. R. (1986). The world city hypothesis: development and change. *Urban Studies*, 23(2), 59-137.

Frimpong, B. F., & Molkenthin, F. (2021). Tracking urban expansion using random forests for the classification of landsat imagery (1986–2015) and predicting urban/built-up areas for 2025: A Study of the Kumasi Metropolis, Ghana. *Land,* 10(1), 44. https://doi.org/10.3390/land10010044

GEOFABRIK. (2022). Open street map, Retrieved from, on 10 February 2022, https://download.geofabrik.de/europe/turkey.html,

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300. https://doi.org/10.1016/j.patrec.2005.08.011

Goncalves, T. M., Zhong, X., Ziggah, Y. Y., & Dwamena, B. Y. (2019). Simulating urban growth using cellular automata approach (SLEUTH)-A case study of Praia City, Cabo Verde. *IEEE Access*, 7, 156430–156442. https://doi.org/10.1109/ACCESS.2019.294.9689

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. (2014). Multivariate data analysis (Vol. 7), *Pearson Education Limited*, USA. ISBN: 1-292-02190-X

Haykin, S. (1999). *Neural networks, A comprehensive foundation*, Pearson Education, Singapore.

Henderson, V. (1997). Medium size cities, *Regional Science and Urban Economics*, 27(6), 583-612.

Huang, B., Xie, C., & Tay, R. (2010). Support vector machines for urban growth modeling, *Geoinformatica*, 14(1), 83-99. https://doi.org/10.1007/s10707.009.0077-4

İzmen, Ü. (2014). Bölgesel kalkınmada yerel dinamikler: Tunceli modeli ve 2023 senaryoları, *Fam Yayınları*, İstanbul, ISBN: 978.605.6487-90-3

Jun, M. J. (2021). A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the Seoul metropolitan area. *International Journal of Geographical Information Science*, 35(11), 2149-2167. https://doi.org/10.1080/13658.816.2021.1887490

Karadağ, A., & Koçman, A. (2007). Coğrafi çevre bileşenlerinin kentsel gelişim süreci üzerine etkileri: Ödemiş (İzmir) örneği, *Ege Coğrafya Dergisi*, 16(1-2), 3-16.

Karimi, F., Sultana, S., Babakan, A. S., & Suthaharan, S. (2019). An enhanced support vector machine model for urban expansion prediction. *Computers, Environment and Urban Systems*, 75, 61-75. https://doi.org/10.1016/j.compenvurbsys.2019.01.001

Kavzoğlu, T., & Çölkesen, İ. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi, *Harita Dergisi*, 144(7), 73-82.

Li, C., Li, J., & Wu, J. (2018). What drives urban growth in China? A multi-scale comparative analysis. *Applied Geography*, 98, 43–51. https://doi.org/10.1016/j.apgeog.2018.07.002

Liaw, A. ve Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

Liong, S. Y., & Sivapragasam, C. (2002). Flood stage forecasting with support vector machines, *Journal of the American Water Resources Association*, 38(1), 173-186. https://doi.org/10.1111/j.1752-1688.2002.tb01544.x

Liu, X. ve Lathrop R. G. (2002) Urban change detection based on an artificial neural network, *International Journal of Remote Sensing*, 23:12, 2513-2518, https://doi.org/10.1080/014.311.60110097240

Malczewski, J. (2004). GIS-based land-use suitability analysis: a critical overview. *Progress in planning*, 62(1), 3-65. https://doi.org/10.1016/j.progress.2003.09.002

Mattivi, P., Franci, F., Lambertini, A., & Bitelli, G. (2019). TWI computation: a comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, 4(1), 1-12. https://doi.org/10.1186/s40965.019.0066-y

Mendiratta, P., & Gedam, S. (2018). Assessment of urban growth dynamics in Mumbai Metropolitan Region, India using object-based image analysis for medium resolution data. *Applied Geography*, 98, 110–120. https://doi.org/10.1016/j.apgeog.2018.05.017

Meyer, D. (2001). Support vector machines. *R News*, 1(3), 23-26. https://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf

Osuna, E., Freund, R., and Girosi, F. (1997). Support vector machines: Training and applications. A.I. Memo 1602, MIT Artificial Intelligence Laboratory.

Özcan, H. (2008). İstanbul'da kentsel yayılmanın yapay sinir ağları ile öngörüleri, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi. Fen Bilimleri Enstitüsü, İstanbul. 67 s.

Özdemir, M. A. (1996). Türkiye'de büyük yerleşme alanlarının seçiminde jeomorfolojik esaslar. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 8(2), 209-222.

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), 1007-1011. https://doi.org/10.1080/014.311.6051233.131.4083

Pal, M., (2005) Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26:1, 217-222, https://doi.org/10.1080/014.311.6041233.126.9698

Park, S., Jeon, S., Kim, S., & Choi, C. (2011). Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea. *Landscape and Urban Planning*, 99(2), 104-114. https://doi.org/10.1016/j.landurbplan.2010.09.001

Patriche, C. V., Vasiliniuc, I., & Biali, G. (2015). Quantitative evaluation of landslide susceptibility in the Bârlad basin. *Environmental Engineering and Management Journal*, 14(9), 2229-2236.

Pijanowski, B., Shellito, B., Bauer, M., & Sawaya, K. (2001). Using GIS, artificial neural networks and remote sensing to model urban change in the Minneapolis–St. Paul and Detroit Metropolitan areas, *In Proceedings, American Society of Photogrammetry and Remote Sensing annual conference*, April 23–27, 2001, St. Louis, Missouri, USA, 1-13.

Pratyush, R. R., Bandopadhyay, A., & Singh, S. K. (2018). Urban growth modeling using logistic regression and geo-informatics: a case of Jaipur, India. *International Journal of Science & Technology*, 13(2018), 47–62.

Şahin, E. (2021). Kentsel büyüme simülasyon modelleri. *International Geoinformatics Student Symposium (IGSS)*, 1(1), 13–18,

Sassen, S., (1991). The global city: London, New York, Tokyo. *Princeton: Princeton University Press*, ISBN: 978.069.1078-66-3

Scott, A.J,Agnew, J.,Soja, E.W and Storper, M. (2001*). 'Global city-regions'*, In Scott, A. (ed): Global City-Regions: Trends, Theory, Policy. Oxford: Oxford University Press, 11–30.

Stumpf, A., & Kerle, N. (2011). Combining Random Forests and object-oriented analysis for landslide mapping from very high resolution imagery. *Procedia Environmental Sciences*, 3, 123-129. https://doi.org/10.1016/j.proenv.2011.02.022

Sung, D.G., Lim, S.H., Ko, J.W. and Cho, G.S., (2001). Scenic evaluation of landscape for urban design purposes using GIS and ANN, *Landscape and Urban Planning*, 56, 75-85. https://doi.org/10.1016/S0169-2046(01)00174-8

Thompson, D. W. (1966). On growth and form: An abridged edition. Edited by John Tyler Bonner. *Cambridge University Press*, Cambridge MA.

Tong, X., Zhang, X., & Liu, M. (2010) Detection of urban sprawl using a genetic algorithm-evolved artificial neural network classification in remote sensing: a case study in Jiading and Putuo districts of Shanghai, China, *International Journal of Remote Sensing*, 31:6, 1485-1504, https://10.1080/014.311.60903475290

TUIK (2022). ADNKS (Adrese Dayalı Nüfus Kayıt Sistemi) Sonuçları, Retrieved from, on 10 February 2022, https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr

Tunçdilek, N. (1986). Türkiye'de yerleşmenin evrimi, *İstanbul Üniversitesi Deniz Bilimleri ve Coğrafya Enstitüsü Yayınları*, No:3367, İstanbul.

Tuncel, M. (2012). "Tunceli". TDV İslâm Ansiklopedisi. Retrieved from, on 10 February 2022, https://islamansiklopedisi.org.tr/tunceli#1

Tunceli Municipality, (2017). Tunceli ilave-revizyon nazım ve uygulama imar planı plan açıklama raporu, Tunceli.

Üzmez, U. (2012). Türkiye'de orta ölçekli kentsel alanlar sorununa çözüm arayışları: Zonguldak örneği. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(2), 127-158.

Vapnik, V. (2000). The nature of statistical learning theory. *Springer science & business media*, New York, ISBN: 978.144.1931-60-3

Viana, C. M., Oliveira, S., Oliveira, S. C., & Rocha, J. (2019). Land use/land cover change detection and urban sprawl analysis. In Spatial Modeling in GIS and R for Earth and Environmental Sciences (pp. 621–651). Elsevier. https://doi.org/10.1016/b978-0-12-815226-3.00029-6

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, 277(5325), 494-499.

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. https://doi.org/10.3354/cr030079