Journal of Environmental Health and Sustainable Development

Investigating the Impact of Environmental Factors on Electricity Consumption Using Spatial Data Mining and Artificial Neural Network: A Case Study in Yazd City

Alireza Sarsangi [1], Ara Toomanian [1], Najmeh Neysani Samany [1], Majid Kiavarz [1], Mohammad Hossein Saraei [2]*

[1] Department of Remote Sensing and GIS, Faculty of Geography, University of Tehran, Tehran, Iran.
[2] Department of Geography, Yazd University, Yazd, Iran.

ARTICLE INFO

ORIGINAL ARTICLE

Article History:
Received: 25 May 2024
Accepted: 10 July 2024

***Corresponding Author:**
Ara Toomanian

Email:
a.toomanian@ut.ac.ir

Tel:
+98 912 2782641

Keywords:
Artificial neural network,
Space syntax,
Electricity consumption,
Remote sensing,
Spatial data mining,
Yazd City.

ABSTRACT

Introduction: Modeling energy demand in different energy consuming sectors is a crucial measure for effective management of the energy sector and appropriate policies to increase productivity. The rising importance of energy resources in economic development is evident. Sustainable energy use is crucial for environmental protection and social progress. Understanding the factors affecting energy consumption is essential for effective energy management. Therefore, the purpose of the current study is to investigate the impact of environmental factors on household electricity consumption in Yazd city.

Materials and Methods: In the present research, various environmental factors affecting electricity consumption, including air pollution, air temperature in homes, ground surface temperature, and green space were investigated. The effects of these factors on electricity consumption of subscribers were investigated with ANN and apriori methods.

Results: Among the environmental factors, the distance to the regional park, the area of the park, and the amount of vegetation at a distance of 300m have the greatest impact, respectively, and the average summer air temperature, the amount of vegetation at a radius of 500 m, the distance from the local park, and the average summer NDVI have had the smallest effect. Unlike neural network methods, apriori presents relationships between parameters affecting electricity consumption transparently in the form of rules.

Conclusion: It's used to identify the most frequently occurring elements and meaningful associations in a dataset. Greenspace can be a mitigation strateegy for reduction of energy consumption.

Citation: Sarsangi A, Toomanian A, Neysani Samany N, et al. ***Investigating the Impact of Environmental Factors on Electricity Consumption Using Spatial Data Mining and Artificial Neural Network: A Case Study inYazd City.*** J Environ Health Sustain Dev. 2024; 9(3): 2354-68.

Introduction

Today, electricity is considered as one of the most important energy carriers of the country, an effective factor in production and a vital commodity in consumption [1]. Environmental issues caused by the use of fossil fuels, which is one of the problems facing the world today, have led to an increase in the global tendency to use less polluting and healthier fuels such as electricity [2]. Modeling energy demand in different energy consuming sectors is one of the crucial measures for effective management of the energy sector and appropriate policy making in order to increase productivity in this sector. Energy efficiency has always been one of the main goals of energy policy makers [3]. In Iran, most of the studies conducted in the field of electricity consumption have focused on electricity consumption forecasting [4]. Omidi et al. studied modeling and forecasting of electricity production and consumption in Iran between 1967-

[DOI: 10.18502/jehsd.v9i3.16583]

2013 and evaluated the artificial neural network method as the most accurate method for forecasting electricity consumption [5]. Fatahi et al. evaluated the impact of the structure of the high electricity consumption population (a case study comparing the electricity consumption of western and eastern provinces of the country) [6]. Their results showed the positive effect of urbanization in eastern provinces due to the high income of the easterners and the presence of urban facilities, as well as the residents' use of low-consumption electrical appliances; as a result, electricity consumption has been much lower.

Undoubtedly, climate changes, especially temperature, are effective in increasing or decreasing the use of heating and cooling devices [7]. Numerous studies have been conducted on the influence of climatic factors on electricity consumption [8]. Salmani and Mojarad investigated the relationship between weather variables and electricity consumption and electricity demand forecasting using circulation models in west of Iran [9, 10]. The relationship between climatic variables and electricity consumption in thirteen stations of the region in a 28-year period was modeled using multiple regression. The results showed that the temperature of cold and hot days and relative humidity have the most significant effect on increasing electricity consumption. The results of a similar study by Wu et al. in Australia showed that the temperature of hot days, humidity, evaporation, and wind speed had the greatest impact on electricity consumption in Australia [11]. The results by Basak and Foucault's study showed that there was a non-linear relationship between electricity consumption and temperature, especially in countries with hot climates [12]. Vine also considered climate change as an important and effective challenge on household electricity demand in the state of California [13]. Zheng et al. examined spatial granularity in electricity consumption forecasting and found the use of LSTM recurrent neural network to be effective [14]. Through hourly electricity consumption information, Ramos et al. clustered various consumption patterns [15]. Brunen et al. also investigated the effect of family behavior, awareness, and literacy on energy consumption with regard to their expenses [16].

In human societies, development becomes possible by using more energy, and in this way, in order to achieve development, human beings change the physical, chemical, biological, social and traditional characteristics of their environment. The production, transmission, and consumption of energy has important environmental effects on the earth's ecosystem. Energy production and use policies play a central role in local and regional environmental issues. Therefore, the need to determine the complex relationship between environmental issues and energy has become more tangible. The increasing importance of energy resources in formation and growth of economic processes as well as the necessity of exploiting these resources based on environmental considerations and sustainable economic and social development highlights the issue of identifying and examining the factors affecting energy consumption. It is clearly defined that this energy provides the possibility of economic development and progress. In the meantime, the abnormal cycle developed in the form of economic growth, energy use, and environmental problems should be eliminated. The most necessary action in the initial stages is to examine the factors affecting energy consumption, mostly consumed in urban areas, in this caseelectricity. One of the most important factors is the environmental factors investigated in this research. Therefore, the aim of the current study is to investigate the impact of environmental factors on household electricity consumption in Yazd city.

Material and Method

Introducing the study area

Yazd city, the center of Yazd Province with an area of over 100 square kilometers, is located in the center of the province between 47° 22° 54° to 33° 54° 24° east longitude, and 39° 47° 31° to 51° 56° 31° north latitude. Its altitude is 1215 m. According to the 2015 census, it has a population of more than 535,000 and is one of the first 15 cities in the country in terms of population (Figure 1). In six

[DOI: 10.18502/jehsd.v9i3.16583]

months of the year, the temperature of this city is very hot, so that the temperature reaches 50 centigrade in the summer season, therefore, in this research, the focus is on the data estimated in the summer season, when the electricity consumption increases due to the use of cooling systems.

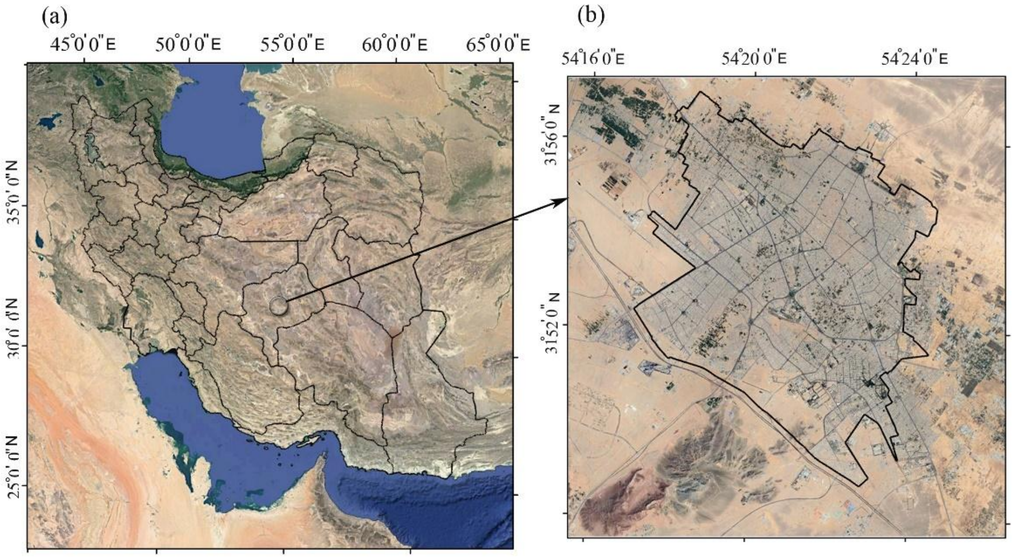

Figure 1: The location of study area; a) Iran, b) Yazd City

Data and descriptive statistics

A: Electricity consumption data

Data about electricity consumption of Yazd city subscribers was obtained from the province's electricity distribution company for the years 2016 to 2019. The information provided includes the data of more than 350,000 electricity subscribers in 2-month withdrawals. The information collected was at the zip code and parcel level.

B: Satellite images

In the present study, the images of Landsat 8 satellite were used. Landsat 8 was launched as part of the LDCM data continuity mission on February 11, 2013. This satellite carries OLI and TIRS sensors. With the help of two bands in the atmospheric windows of 10.6 to 11.2 µm for band 10 and 11.5 to 12.5 µm for band 11, the TIRS sensor is able to record thermal infrared radiations with a spatial resolution of 100 meters [17-19]. The fact that Landsat 8 is equipped with two thermal bands has distinguished it from other satellites in the Landsat series.

C: Air temperature in houses

The current research was conducted using an automatic thermometer that has the ability to measure and store the temperature in the desired sequence, at 60 points with proper distribution in the entire study area. Using these thermometers, the air temperature in home was measured hourly.

D: Air pollution

To evaluate the impact of pollution, the amount of PM_{10} was measured in 85 points in the courtyards of the houses.

Data processing

A parcel is described as the main cell of urban design structures, which determines the shape of the surrounding road network and the structure of internal buildings. Urban parcel data is one of the cornerstones of contemporary urban planning. In the present study, first, urban parcel data related to Yazd city was obtained from the municipality, which included 229,571 parcels. Since the aim of the current study is spatial data mining of electricity consumption, it is necessary to consider data as the basis of spatial distribution and other descriptive information and quantitative parameters to be placed on the location. Therefore, urban parcel data was used as the basis of spatial data. All urban parcels are not made up of residential houses and some include unbuilt land,

[DOI: 10.18502/jehsd.v9i3.16583]

government centers, schools, etc. The data on electricity consumption of urban parcels was obtained from Yazd Electricity Distribution Company. This data is recorded in a period of two months from the beginning of the solar (Persian) year.

Therefore, in this research, there were limitations in preparing electricity data because there was no access to information in time series with less sequence. In studies of electricity consumption monitoring in developed countries, hourly information on peak electricity consumption has been used. The number of electricity subscribers, in addition to directly affecting electricity consumption, indicates the number of people who use the public electricity grid to meet their needs. The presence of a large number of electricity subscribers can increase the electricity load in power plant network, and this may be due to the simultaneous use of electrical equipment at a certain time.

Algorithms

A: Spatial data mining

Data mining is considered as an intelligent data processing tool in order to understand the structures, patterns, and connections between large and complex data sets and to take advantage of the knowledge in data [20]. The problem of exploring alternating item sets was first presented by Agrawal in 1993 in the form of exploring association rules among item sets [21].

Association rules are one of the main techniques in the field of data mining that can be used in different applications. It is also a method of data mining and is used to extract useful patterns from huge databases ; they are descriptive and unsupervised data mining methods, which search the data set to find the relationship between features. In fact, this method studies features associated with each other, while reducing the relationship between these features [22]. These rules show the interdependencies between a large set of data items [23].

In this study, the problem of association rules is expressed as follows: $I = \{i_1, i_2, ..., i_m\}$ is a set of data items, and $T = \{i_1, i_2, ..., t_n\}$ is a set of transactions, each of which includes data items from the set of data items *I*. Therefore, each transaction t_i consists of a set of data items; so that $t_j \subseteq I$. If X and Y are assumed to be data items, the semantic association rule is of the form $X \to Y$ where , $X \subset I$ ‘$Y \subset I$, , and $X \cap Y = \phi$ [24]. In association, the rule that is in the form $X \to Y$ is called "precedence", and Y is called "result". It is clear that the value of the antecedent includes the value of the result. The range of support for the rule and the level of confidence in the rule are the most important qualitative criteria for evaluating the interestingness of the rule.

B: Apriori

So far, various algorithms have been presented to explore association rules, which differ in how to discover frequent items. The most famous algorithm in exploring association rules is the apriori algorithm [25]. Apriori is a data mining method used to identify and extract relationships, structures, and patterns between items that occurr simultaneously in a database but are not clear [26]. In apriori algorithm, association rules are extracted based on three indicators including confidence, support, and lift. In fact, each rule is evaluated through these three indicators, and the effective rules are selected from the set of possible rules. Support is estimated as follows Equation 1:

$$\sup(A \to B) = \frac{A \cap B}{X} \qquad (1)$$

Where, *A* and *B* are two different data types in the database, and *X* is the total number of items in the database. This rule extracts all the transactions in which there are two data items *A* and *B* and compares them with the minimum value of the rule's support range specified by the user, and then, only selects transactions. Make sure that the support range of their rule is greater than or equal to the minimum support range of the rule, and the rest of the transactions are non-applicable rules that are deleted. The confidence component of the

[DOI: 10.18502/jehsd.v9i3.16583]

rule is calculated as follows Equation 2:

$$cnf\ (A \rightarrow B) = P(B / A) = \frac{\sup(A \rightarrow B)}{\sup(A)};$$
$$\sup(A) = \frac{A}{X} \tag{2}$$

First, transactions containing data type A are calculated, and transactions containing data type B are extracted from them. Then, a comparison of the output of this relationship with the minimum level of confidence in the rule will be carried out. The lift component greater than 1 indicates a positive correlation that infers the occurrence of data B in the occurrence of data A ,and is estimated as follows Equation 3[26]:

$$Lift\,(A \rightarrow B) = \frac{Conf\ (A \rightarrow B)}{\sup(B)} = \frac{\sup(A \rightarrow B)}{\sup(A).\sup(B)} \tag{3}$$

C: Artificial neural network

Artificial Neural Networks (ANNs) created by modeling the human body are composed of cells connected to each other, just like the human body. One of the most widely-used artificial neural networks in modeling and forecasting is the multilayer perceptron or MLP neural network. The MLP network consists of several layers of input, output, and hidden, where the output of the first layer is the input vector of the second layer. Similarly, the output of the second layer forms the input vector of the third layer. The outputs of the second layer show the real response of the network. In the multi-layer perceptron neural network, data processing is carried out by the activation function or the transfer function. The basic question is with what strength and quality should a neuron transmit the signal to the adjacent neuron. Adjusting neural network parameters is the main issue in model design.

Statistical analysis

In this section, the results of using multi-layered perceptron neural network for modeling and predicting electricity consumption are presented. First, the architecture of multilayer perceptron neural network is explained: this architecture includes three layers: input, hidden and output. The activation function used in the hidden layer is the sigmoid tangent function. In addition, the number of neurons in the hidden layer was 5. For network training, the Levenberg-Marquardt algorithm was chosen as the training function, which is usually the fastest and best algorithm for network training problems. For data preprocessing, two functions named "removeconstantrows" and "mapminmax" were applied to inputs and outputs. The "removeconstantrows" function was used in the data preprocessing stage. The purpose of this function is to remove columns whose value is the same in all rows and have a fixed value. These columns usually do not provide useful data for training the network, and removing them can lead to improved network performance. The "mapminmax" function is also used for input and output data and is used to normalize the data. Using this function, data values are converted to a specified range. For example, by applying this function to the data, input and output values are converted to the range between 0 and 1; this makes the influence of variables on the performance of the network more balanced and improves the performance of the trained network. The use of these two pre-processing functions is aimed at improving the quality and efficiency of the neural network. By applying these changes, data is presented to the network more readily, and the network can better learn patterns and meaningful relationships between inputs and outputs. This is more important when there are different scales and ranges for variables, inputs, and outputs, and they should be transfered to a common range so that the network is properly trained. Moreover, the number of epochs or training courses of the network was considered 200 iterations. Also, the goal of stopping the network was estimated to be 10^{-9}. The correct setting of these parameters is very important in performance and quality of network training. According to the type of data and the desired problem, different requirements can be considered for these parameters. For example, the number of epochs should be large enough for the network to be sufficiently trained, but overfitting should be avoided. Furthermore, training target

[DOI: 10.18502/jehsd.v9i3.16583]

should be small enough for the network to train well, but not too small for the network to suffer during training

Results

The results of the survey regarding the number of electricity subscribers and electricity meters in the building showed that 61.1% of the buildings had an electricity meter, and buildings with 2 and 3 meters, respectively, included 33.7% and 4.1% of the total surveyed parcels. Also, a small number of buildings had more than 3 electricity subscriptions, which was smaller than 1% in total.

In the studied area, there were 354,903 subscriptions for electricity consumption, and due to the multi-story nature of some buildings, there were several subscriptions in some parcels. Electricity consumption data was prepared in the period of two months between 2016 and 2019. Due to the heat of the air in spring and summer and the use of water coolers, the information related to this season was used for the review. Figure 2 shows the position of the basic information of electricity meters and selected parcels.

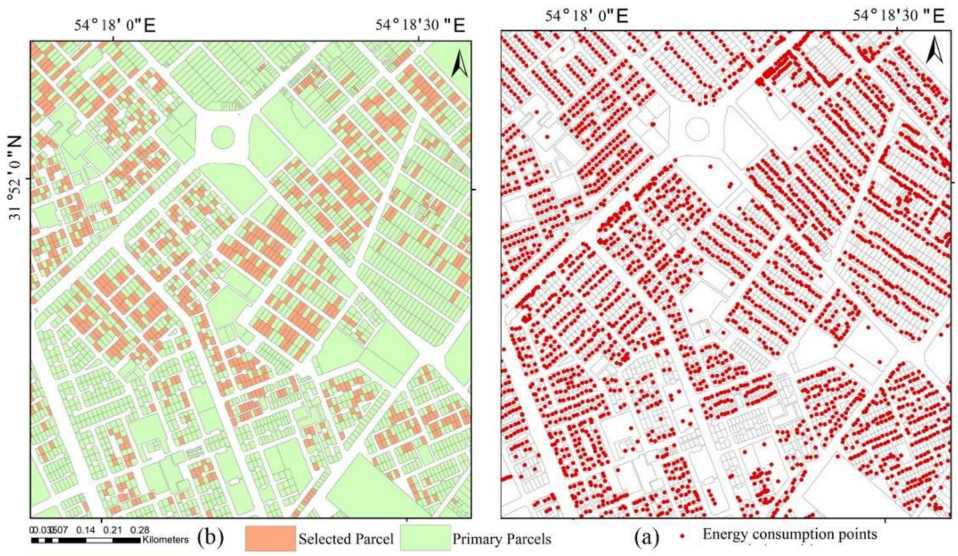

Figure 2: a) Displaying the location of energy consumption points on urban parcels b) Displaying selected parcels based on electricity consumption data in the period from 2016 to 2019 in one part of the study area

The environmental factors investigated in the present study included: air temperature in houses, air pollution, Summer NDVI, park area within a distance of 300 and 500 meters, NDVI at 300 and 500 meter radius, distance to local and regional park, and land surface temperature in summer and area of the yard.

The results of monitoring air temperature in houses showed that at the beginning of hot season, air temperature in the houses had a peak and is associated with an increase in temperature, but after turning on the cooling devices in the houses, the air temperature had a constant pattern throughout the hot season of the year. Therefore, throughout the year, with the increase in air temperature, the temperature of houses did not increase, and the temperature of houses was almost constant from the beginning of summer until autumn. As an example, air temperature changes in a house with hourly sequence and in the summer season are shown in Figure 3.

[DOI: 10.18502/jehsd.v9i3.16583]

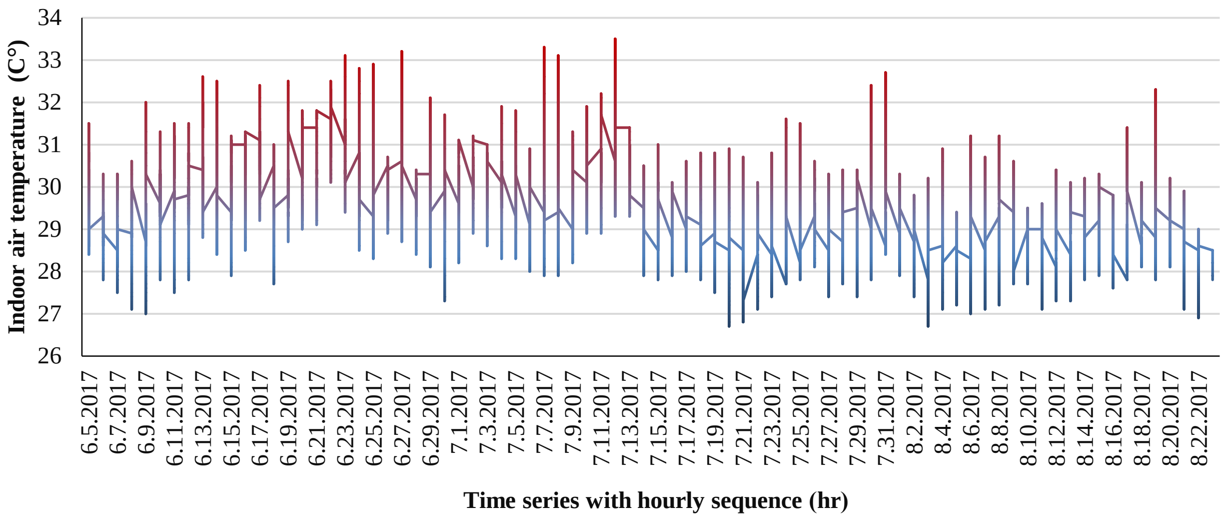

Figure 3: Analyzing Hourly Air Temperature Changes in Homes

The highest recorded temperature in this house was 34.5°C, and the highest temperature difference in one day and night around the house was 4.5°C. According to the location of the studied area, located in the dry belt of central Iran, in the summer, the temperature difference of the ground surface in one day and night is around 30°C. Examining the average summer air temperature in houses showed that the highest temperature was 34.5°C, and the lowest was 25.3°C. Also, 33% of the houses had an average summer temperature of less than 30°C, 46% had an air temperature of 30 to 32°C, and the rest had an air temperature of more than 32°C (Figure4).

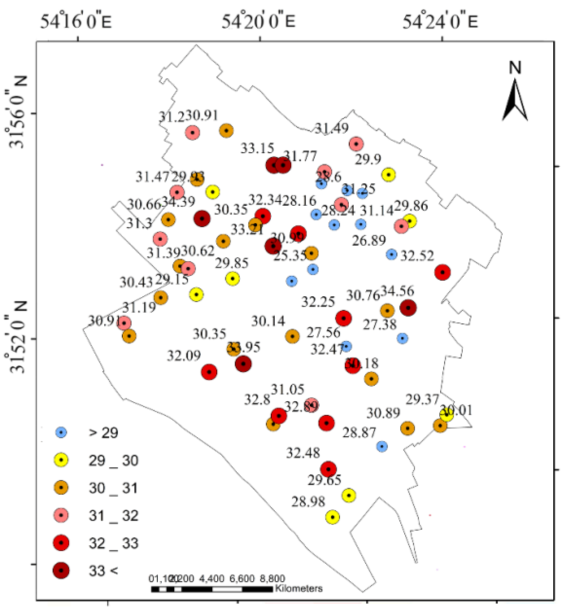

Figure 4: Air temperature measurement points inside residential houses

[DOI: 10.18502/jehsd.v9i3.16583]

To check the impact of pollution, the amount of PM_{10} was measured in 85 points in the courtyards of the houses. These values were estimated using interpolation method for all the houses (Figure 5).

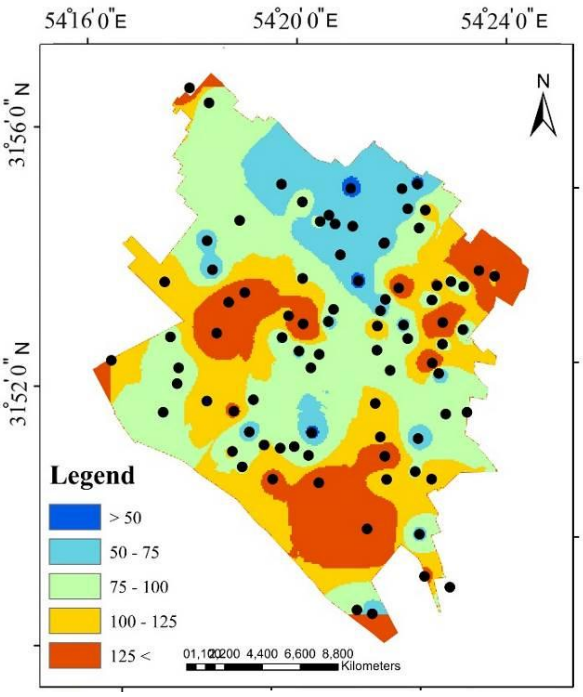

Figure 5: Location of air pollution measurement points and interpolated image of PM_{10}

The results of identification and investigation of local parks showed that there were 135 parks or local green spaces in the studied area. Average local green space was 6334 square meters. The local green space was divided into different categories according to the area, and the results of the number survey showed that 24% of them had an area of less than 2000 square meters, and 27% of the green spaces had an area of 2000 to 4000 square meters; in addition, 18% of the total number of green spaces had an area of more than 10,000 square meters. The prevalence of green space with an area of 6000 to 10000 was less than other categories, and it accounted for about 10% of the total number of local green spaces. The examination of the percentage of local green space also showed that green spaces with an area of more than 10,000 square meters constituted 64% of the local green space of the entire study area. Also, there were 9 green spaces or regional parks in the study area, whose area varied from 1 to 24 hectares, and the average area in these parks was 12.6 hectares, and the total area of local parks in the study area was 113 hectares (Figure 6). Using Landsat 8 satellite images, land surface temperature, and the NDVI index in summer in the study area were investigated, and their effects on electricity consumption were evaluated.

[DOI: 10.18502/jehsd.v9i3.16583]

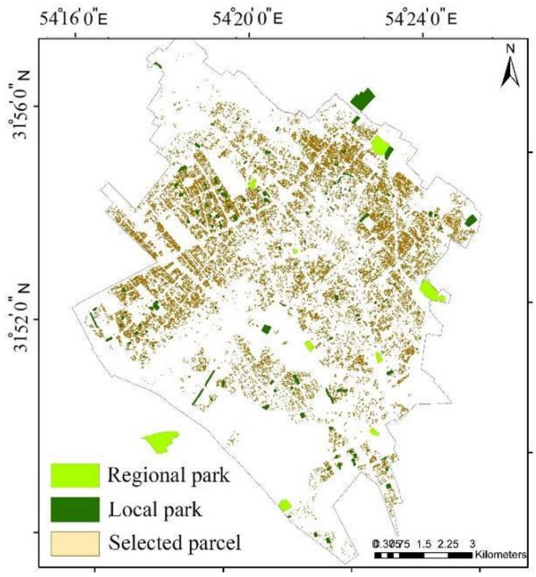

Figure 6: Location of local and regional parks in the studied area

Considering that the area of the parcel alone cannot be valid data because the area of the parcel or piece of land under investigation may be large, a small part of it was covered by buildings, or in a parcel with a large area, the built-up land was equal to the parcel with smaller area, it was necessary to specify the area of the yard in addition to the area of the parcel. Building area represents the built part of the parcel, and yard represents the unbuilt part of the parcel, which usually includes the courtyard. The average, maximum and minimum area values in parcel, yard ,and building were investigated (Figure 7). The results showed that in the studied area and the selected parcels, the maximum area of the parcel is about 500 square meters, and the average area of the parcel is 203 and the minimum is 49 square meters. The maximum and minimum area of the mansion or building was estimated to be 301 and 30 square meters, respectively. The average unbuilt land or yard in the study area was 135 square meters ; the maximum area was 251, and the minimum was 68 square meters. In other words, in the studied area, the average area of the courtyard was 68 square meters. The presence of the yard and its green space had an effect on the amount of electricity consumption, which was further investigated.

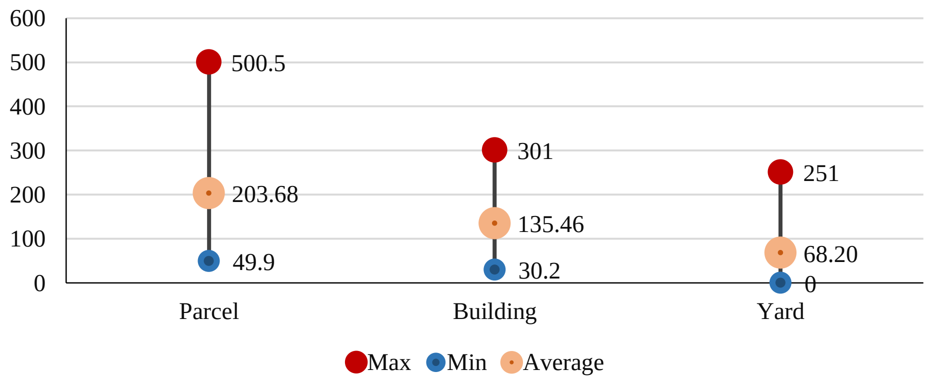

Figure 7: The average, maximum, minimum area of parcel, building ,and yard in the studied area

[DOI: 10.18502/jehsd.v9i3.16583]

Examining factors affecting electricity consumption using artificial neural network algorithm.

In the next step, the data was divided into training, validation, and test ratios. The training set included 80%, the validation set included 10%, and the test set included 10% of the data. The accuracy criteria of mean square error (MSE) and correlation (R) were used to evaluate the performance of the network. Also, graphs and tables related to the results of network training and evaluation were reviewed.

Table 1: MSE value estimated using neural network

	(Train)	(Validation)	(Test)	Total
MSE	0.0542	0.0701	0.0605	0.0564
R	0.9305	0.9308	0.94	0.9315

According to the results, the MSE value for train data was equal to 0.0542 and for evaluation data and test data, MSE values were 0.0701 and 0.0605, respectively. This showed that the neural network was able to predict the amount of electricity consumption based on the parameters used.

Regarding the correlation values or *R*, for all cases, the value of *R* was equal to 0.93. These results showed the acceptable performance of network in predicting the amount of electricity consumed by buildings. The value of *R* indicated a direct and linear relationship between the expected outputs of network and the actual values of the amount of electricity consumed. This positive and strong relationship suggested that the network can accurately predict the electricity consumption of buildings.

According to these results and the proper performance of network, it was concluded that the proposed method using neural network and the 12 selected parameters could be effectively used to predict the amount of electricity consumed by buildings. Due to the high value of *R* and the performance results close to 0, the network was able to learn various patterns and relationships from the data and reliably predict the amount of electricity consumed by buildings. Moreover, the importance of parameters for predicting the electricity consumption of buildings was determined using neural network. In the following, the table and graph of the parameters with greatest impact on the forecast are given.

Examining the results of neural network showed that among the environmental factors, distance to the regional park, the area of the park, and the amount of vegetation at a distance of 300 meters had the greatest effect, and the area of the park within a radius of 500 meters, the value of NDVI within a radius of 500 meters, the distance to the local park, and the average summer NDVI had a smaller effect. Using satellite images, two indices of land surface temperature and NDVI were estimated, and the results of the neural network showed that temperature had a greater impact on electricity consumption. (Figure 8).

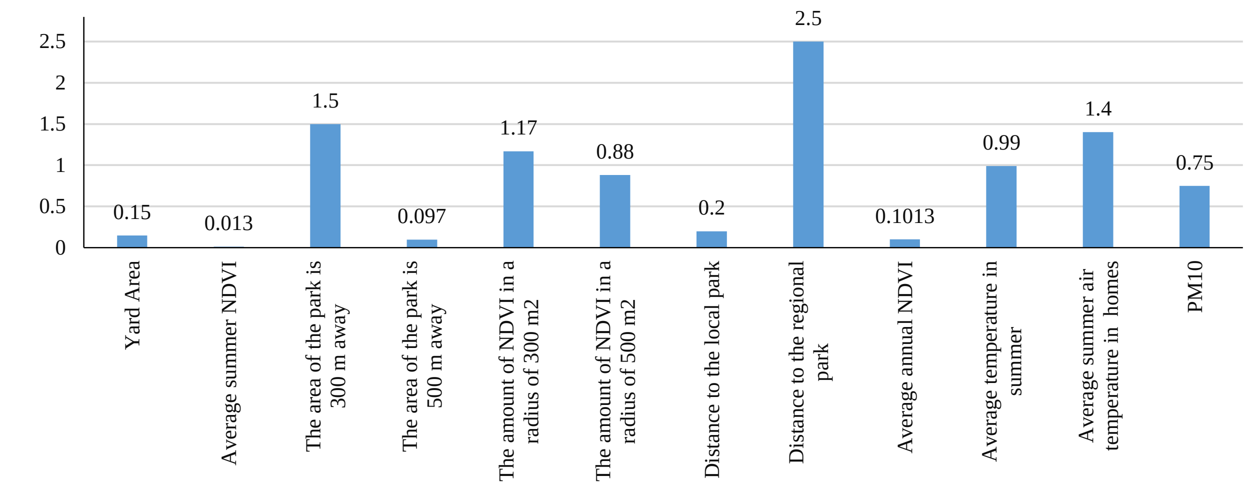

Figure 8: Investigating the effects of different variables on electricity usage using a neural network model

[DOI: 10.18502/jehsd.v9i3.16583]

Examining factors affecting electricity consumption using spatial data mining

The Apriori algorithm had previously been used in non-spatial applications. In order to produce maps of factors affecting electricity consumption using spatial data mining (apriori method), it is necessary to extract frequent and reliable rules. Therefore, by defining support and confidence thresholds as 1.0, 5 and lift 1 respectively, frequent rules were identified for subscribers with high and very high consumption (Figure 9). Since each rule is composed of different combinations of items, their spatial pattern is different. Among the examined components in determining the rules related to subscribers with high and very high consumption, the following items had the greatest impact. The areas where vegetation covered index within the radius of 300 and 500 meters indicated the absence of cover as well as the areas where the area was large, medium, and greater, were included in the category of subscribers with high consumption. Unlike neural network methods, this algorithm presents relationships between parameters affecting electricity consumption transparently in the form of rules. These rules entail support and confidence. Greater support indicates that those rules occur more widely. Therefore, the reproducibility and reliability of the rules allow the rules to be clearly understood and interpreted. On the other hand, the rules of the apriori algorithm can be easily extended to areas with similar environmental conditions, while other statistical methods based on neural networks do not have such capabilities.

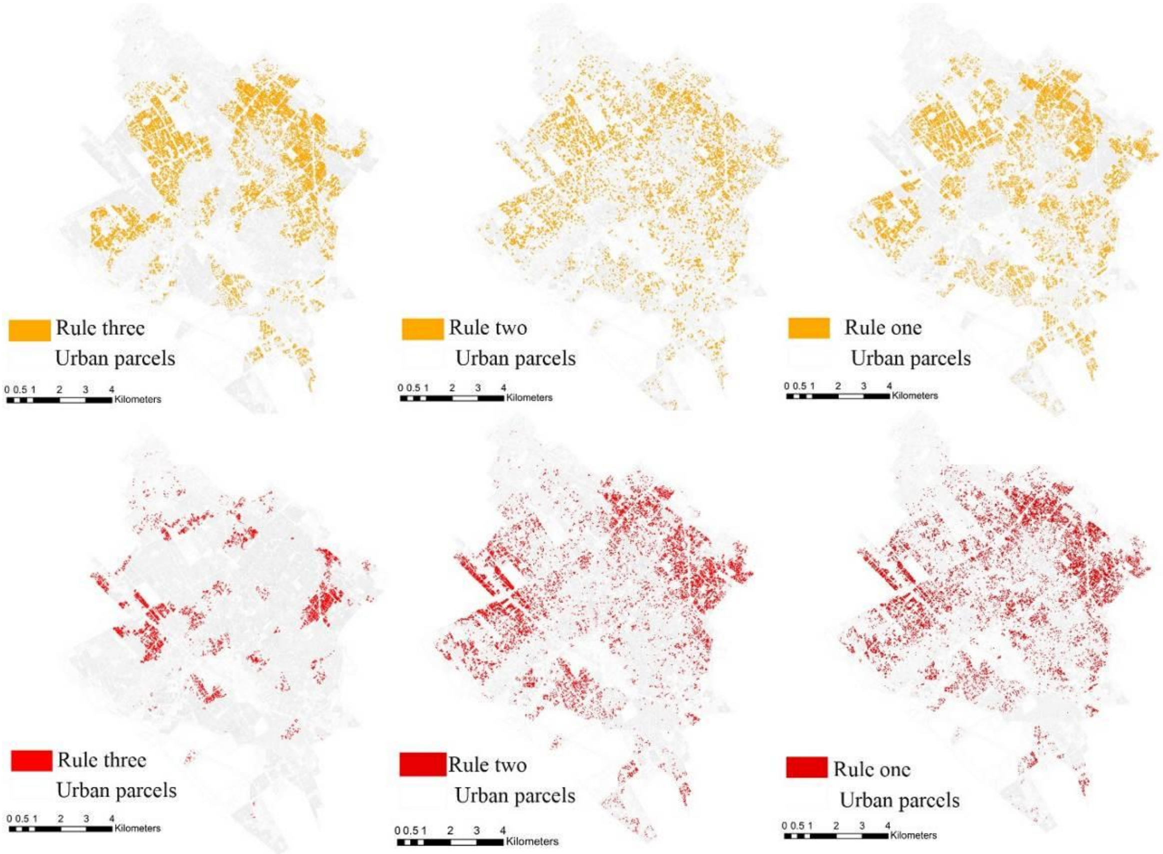

Figure 9: A) Rules associated with subscribers with high consumption; B) Rules associated with subscribers with very high consumption

[DOI: 10.18502/jehsd.v9i3.16583]

Discussion

Electricity consumption in residential houses is one of the most important types of energy consumption. For this reason, monitoring and checking electricity consumption in residential houses is of particular importance. Several factors affect electricity consumption in residential houses, and the investigation of these factors can be effective in the optimal management of electricity consumption, which includes urban physical, social and economic factors, environment, etc. In this study, the impact of environmental factors on electricity consumption of residential houses was evaluated. Considering that the studied area had a hot and dry climate and in six months of the year, it is necessary to use cooling equipment that consumes a lot of electricity. Only data of electricity consumption in summer was used. It should be noted that one of the primary limitations in the current research was access to residential electricity consumption data because these data were stored in a two-month sequence. If there was a daily or monthly consumption amount, it would be much more appropriate to investigate their relationship with environmental components. The investigation of air temperature in residential houses showed that in 80% of the houses, the average air temperature during the day and night was less than 32. The difference in air temperature in different houses indicated different amounts of electricity consumption, which was affected by various factors, including environmental factors. Another environmental component investigated in this research was air pollution. In a study conducted by Yi et al. (2020) [27] the results showed that people prefered to stay at home when air was polluted, and this factor increased household electricity consumption. Also, You et al. (2017) [28] reported a significant relationship between monthly household electricity consumption and outdoor PM2.5 concentrations in Singapore and Shanghai. Among other environmental parameters investigated in the current research was the impact of urban green space on the electricity consumption of residential houses, which was also investigated in the studies by Gargiulo et al. (2018) [29]. To examine the effect of green space, the distance of parks from urban parcels, the area of house yards, and NDVI vegetation cover index were investigated. The average yard area of houses was estimated to be 68 square meters. Considering that usually only a part of it is used for the green space inside the houses, the area of the green space of the houses in the study area is small. This showed that the neural network was able to predict the amount of electricity consumption based on the parameters used. The value of R indicated a direct and linear relationship between the expected outputs of the network and the actual values of the amount of electricity consumed. This positive and strong relationship demonstrated that the network could accurately predict the electricity consumption of buildings. According to these results and the proper performance of network, it can be concluded that the proposed method using neural network and 12 selected parameters could be effectively used to predict the amount of electricity consumed by buildings. Examining the results of neural network showed that among the environmental factors, the distance to the regional park, the area of the park, and the amount of vegetation at a distance of 300 m had the greatest effect, and the area of the park within a radius of 500 meters, the value of NDVI within a radius of 500m, the distance to local park, and average summer NDVI had less influence. Using satellite images, two indexes of land surface temperature and NDVI were estimated, and the neural network results showed that temperature had a greater effect on electricity consumption. Therefore, the investigation of environmental parameters affecting electricity consumption in residential houses using artificial neural network showed that the distance from the park, the value of NDVI index, and the distance from the park within a radius of 300 meters had the greatest impact on electricity consumption. Spatial data mining was also done using Apriori algorithm, which does not require training data compared to neural network; however, it does not work locally. The pattern between different parameters was identified in this method. 500 meters indicated the lack of coverage,

[DOI: 10.18502/jehsd.v9i3.16583]

which was among the high-consumption subscribers.

Conclusion

Nowadays, electricity plays a significant role in the process of economic decision-making and advancing the development goals of countries due to its connection with other sectors and economic institutions. Saving energy and reducing greenhouse gas production is the main goal of governments all over the world, and Iran is no exception. In recent years, in particular, various attempts have been taken, and many incentives have been considered to increase energy efficiency. In the current study, various factors affecting electricity consumption were examined, and after examining each of these parameters in urban parcels and giving spatial entity to them, their impact on electricity consumption of subscribers was investigated. Two methods of artificial neural network and Apriori were examined. Using the artificial neural network method, the effect of parameters on electricity consumption was estimated, and using the Apriori method, the rules governing electricity consumption were extracted. One of the problems of Apriori algorithm and other algorithms in the field of associative rule extraction was that the user had to specify a minimum threshold for the range of rule support. Apriori inputs should be discrete classes while most environmental parameters are continuous in nature. Therefore, the process of cutting input parameters causes the data within the class to be lost. Certainly, many environmental variables can affect electricity consumption. However, there are limitations in obtaining information. In the present study, these data were selected based on two criteria. 1) Information that could be obtained from satellite images, such as the temperature of the earth's surface and vegetation indices. 2) Information that could be measured on the ground, especially on an urban scale. The more environmental information is used in the model, the better the results will be. Data such as dust and the presence of aerosols and wind data such as speed, air humidity, etc. can also be examined, but in this research, access to this information was not available.

Acknowledgements

The authors would like to thank all the staff and students of the Faculty of Geography of Tehran University for their cooperation

Conflict of interest

The authors declared no conflict of interests.

Funding

The work was unfunded.

Ethical Considerations

Authors complied with the best practice regarding publication ethics specifically concerning the authorship (avoidance of guest authorship), dual submission, manipulation of figures, and competing interests. Authors adhered to publication requirements that the submitted work was original and had not been published elsewhere in any language.

Author’s contributions

Alireza Sarsangi was involved with methodology; Alireza Sarsangi and Ara Toomanian wrote the original draft, and; Najmeh Neysani Samany, Majid Kiavarz , Mohammad Hossein Saraei was involved with reviewing and editing, all the authors read and agreed to the final manuscript.

References

1. Yoo S-H. The causal relationship between electricity consumption and economic growth in the ASEAN countries. Energy policy. 2006; 34(18):3573-82.
2. Lean HH, Smyth R. CO_2 emissions, electricity consumption and output in ASEAN. Applied energy. 2010;87(6):1858-64.
3. Rahimi Ga, khadem f, Shahiki Tash MN. Estimation of energy demand function on Iran's

most energy intensive industries. Quarterely Energy Economics Review. 2014;10(40):149-72.
4. Soleimani P, Yaghobi Z. Forecasting of Iran power demand network by hybrid of support vector regression model and Fruit fly optimization algorithm. Advances in Industrial Engineering. 2018;52(3):405-20.
5. Omidi M, Omidi N, Asgari H, et al. Modeling and forecasting electricity production and consumption in Iran. Future study Management. 2016; 106(27):71-83.
6. Fatahi S, Baharipour S, Rezaee E. Evaluation of the effects of population structure of electricity consumption (case study comparing electricity consumption provinces west and east on the country). Iranian Electric Industry Journal of Quality and Productivity. 2017;5(3):71-90.
7. Tejero-González A, Andrés-Chicote M, García-Ibáñez P, et al. Assessing the applicability of passive cooling and heating techniques through climate factors: an overview. Renew Sustain Energy Rev. 2016;65:727-42.
8. Erfani EE, A. Baaghideh, M. Babaeiyan, M. Future perspective of electricity consumption in Iran during the period 2011-2100 under climate change scenarios using downscaling of general circulation models. Geography and Environmental Planning. 2017;27(4):131- 44.
9. Salmani A, Mojarrad F. Analysis of relationship between climatic variables and electricity consumption and estimated demand by general circulation models in western Iran. Physical Geography Research. 2019;51(2):301-15.
10. Salmanzadeh-Meydani N, Ghomi SF. The causal relationship among electricity consumption, economic growth and capital stock in Iran. J Policy Model. 2019;41(6):1230-56.
11. Wu X, Benjamin Zhan F, Zhang K, et al. Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China. Environmental Earth Sciences. 2016;75:1-16.
12. Bessec M, Fouquau J. The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. Energy Economics. 2008;30(5):2705-21.
13. Vine E. Adaptation of California's electricity sector to climate change. Clim Change. 2012;111:75-99.
14. Zheng Z, Chen H, Luo X. Spatial granularity analysis on electricity consumption prediction using LSTM recurrent neural network. Energy Procedia. 2019;158:2713-8.
15. Ramos S, Soares J, Cembranel SS, et al. Data mining techniques for electricity customer characterization. Procedia Comput Sci. 2021;186:475-88.
16. Brounen D, Kok N, Quigley JM. Energy literacy, awareness, and conservation behavior of residential households. Energy Economics. 2013;38:42-50.
17. Kovalskyy V, Roy DP. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat data product generation. Remote Sens Environ. 2013;130:280-93.
18. Wulder MA, White JC, Loveland TR, et al. The global Landsat archive: status, consolidation, and direction. Remote Sens Environ. 2016;185:271-83.
19. Roy DP, Wulder MA, Loveland TR, et al. Landsat-8: science and product vision for terrestrial global change research. Remote Sens Environ. 2014;145:154-72.
20. Agrawal R, Imieliński T, Swami A, editors. Mining association rules between sets of items in large databases. InProceedings of the 1993 ACM SIGMOD international conference on Management of data; 1993. p. 207-216.
21. Agarwal S, editor. Data mining: data mining concepts and techniques. In2013 international conference on machine intelligence and research advancement; 2013: IEEE.(p. 203-207).
22. Akhondzadeh-Noughabi L, Albadvi A, Aghdasi M. Mining customer dynamics in designing customer segmentation using data mining techniques. Journal of information technology management. 2014;6(1):1-30.
23. Kim YS, Yum B-J. Recommender system based on click stream data using association rule mining. Expert Systems with Applications.

[DOI: 10.18502/jehsd.v9i3.16583]

2011;38(10):13320-7.
24. Liu B. Web data mining: exploring hyperlinks, contents, and usage data. Berlin: Springer; 2011;1.
25. Tan PS, M. Kumar, V. Introduction to data mining. New Jersey: Pearson Education .Springer; 2006. p. 2-12.
26. Hunter H, Cervone G. Analysing the influence of African dust storms on the prevalence of coral disease in the Caribbean Sea using remote sensing and association rule data mining. Int J Remote Sens. 2017;38(6):1494-521.
27. Yi F, Ye H, Wu X, et al. Self-aggravation effect of air pollution: evidence from residential electricity consumption in China. Energy Economics. 2020;86:104684.
28. You S, Neoh KG, Tong YW, et al. Variation of household electricity consumption and potential impact of outdoor PM2. 5 concentration: a comparison between Singapore and Shanghai. Applied energy. 2017;188:475-84.
29. Gargiulo C, Ayad A, Tulisi A, et al. Effect of urban greenspaces on residential buildings' energy consumption: case study in a Mediterranean climate. Smart Planning: Sustainability and Mobility in the Age of Change. 2018:109-25.

[DOI: 10.18502/jehsd.v9i3.16583]