Kirkuk University Journal for Agricultural Sciences
ISSN:2958-6585
https://kujas.uokirkuk.edu.iq

RESEARCH ARTICLE https://doi.org. 10.58928/ku25.16404

Investigating The Green Space Of Erbil City And Determining Its Effective Parameters Using Machine Learning

Tarza Jabar Baqi [1] **Srwsht Mohammed jaza** [1] **Hiwa Farajpanah** [2]

[1]*Surveying Department,Darbandikhan technical institute,sulaimani polytechnic university,sulaimani,KGR,IRAQ.*
[2]*Civil Engineering and Architecture Faculty, Shahid Chamran University of Ahvaz, Ahvaz- IRAN.*
*Corresponding Author: tarza.jabar@spu.edu.iq.

Received:01/09/2025 **Revised: 20/11/2025** **Accepted: 30/11/2025** **Published: 02/12/2025**

ABSTRACT

Accurate prediction of environmental events is a result of the combination of climatic and environmental variables. The Random Forest model was employed in this study to examine the contribution of various climatic and environmental variables to the phenomenon in question under three scenarios: (1) all variables combined, (2) climatic variables only, and (3) environmental variables only. Training was conducted using 70% of the data and testing on 30%. Measures of performance like R-squared, MAE, MSE, RMSE, and MAPE were used to assess predictive accuracy, whereas the relative contribution of each variable was calculated by employing Gini importance index.

Results indicate that optimal prediction performance was achieved when all variables were included (R-squared = 0.98, MAE = 0.0271, RMSE = 0.0414), but the performance was lower when climatic variables alone were used (R-squared = 0.89). Environmental variables alone gave performance which was very close to that of the full-parameter (R-squared = 0.97) case, reflecting their better role in prediction of the model. Variable importance analysis revealed the top contributing drivers of the variables as urbanization (NDBI) and vegetation condition (VCI), then Albedo, PDSI, AOD, and LST. Soil moisture and actual evapotranspiration were climatic drivers with average contribution, while atmospheric pressure and wind speed contributed the least.

These findings demonstrate that Random Forest models can effectively capture complex interactions in environmental systems and emphasize the core role of environmental variables, particularly urbanization and vegetation condition, in predictive modeling. This research provides insights for future environmental monitoring and management strategies in arid and semi-arid regions.

***Keywords*:** Random Forest, Environmental prediction, Climatic variables, Urbanization (NDBI), Vegetation Condition Index (VCI).

INTRODUCTION

The rapid urbanization trend of the past decades has profoundly altered the physical and environmental landscape of cities around the world [1,2,3]. Horizontal and vertical expansion of urban fabrics, rising housing and public infrastructure needs, and massive land use conversions have placed great pressure on urban ecosystems and natural resources [4,5,6,7]. Out of all elements of urban architecture, urban green spaces are an essential element in realizing environmental sustainability as well as enhancing the well-being of citizens [8,9,10].

Urban greenspaces have impacts not only on the visual appearance and landscape of cities but also provide a number of ecosystem services, including air pollution mitigation, microclimate management, biodiversity enhancement, and social and recreational space provision [11,12,13,14,]. Nevertheless, population increase, land use pressures, and unregulated urban sprawl have degraded per capita open green spaces in most cities, disrupting the connection between humans and nature [15,16,17]. Lack of green space is likely to degrade air quality, cause ecological imbalance, and affect social behavior and public health.

One of the most critical environmental issues associated with reduced urban green cover is enhanced urban heat island (UHI) effect [18,19,20,21]. Substitution of natural surfaces by impermeable urban surfaces increases heat storage and alters the urban energy balance, leading to higher air and surface temperatures, reduced wind flow, and more arid urban atmospheres [22,23,24]. Climate change intensifies these impacts further and poses risks to human thermal comfort, public health, and energy consumption. For dry and hot climate conditions, these effects are typically worse due to intense solar irradiation and high temperatures during the day [25,26].

Urban green spaces are governed by the interplay of climatic, environmental, and anthropogenic factors. Climatic factors such as land surface temperature, rainfall, humidity, and wind systems have a direct influence on vegetation development and stress, particularly in arid and semi-arid ecosystems where moisture limitations limit plant potential. Environmental conditions like soil characteristics, topography, and surface albedo govern the spatial pattern and vegetation cover in cities [27,28,29,30,31,32,33]. Interactions between such factors can either increase or decrease stress on urban greenspaces; for example, high temperatures and low humidity can increase evapotranspiration and decrease the healthfulness of plants [34,35]. It is necessary to understand the relative contributions of these climate-forced and environmental factors in order to plan, manage, and restore urban green infrastructure effectively and maintain ecosystem services and improve urban microclimates [36].

Urban vegetation cover mapping and monitoring have traditionally relied on field surveys, accurate though they are, taking time and often not practical for extensive cities. Remote sensing has been a cost-effective alternative, allowing the extraction of vegetation cover from vegetation indices such as the normalized difference vegetation index (NDVI) [37,38]. With the incorporation of sophisticated image classification methods, remote sensing enables high-precision urban green space mapping across large spatial scales [39,40]. In addition, green cover and composition monitoring for urban areas through landscape metrics—patch size, number of patches, and aggregation index—yield data on green patch spatial arrangement and connectivity [41,42]. These analyses, in association with environmental driver information, can guide site-specific urban planning and environmentally sustainable green infrastructure management.

Various research work all over the world has been conducted with an emphasis on the impact of various parameters on urban greens.[21] conducted massive research work in Puebla, Mexico, to examine the greenness of vegetation and land surface temperature (LST) in urban greens for long durations. Using Landsat-based vegetation and LST indicators for 80 urban green spaces during 1986-2019 and 2000-2019, the study indicated that Indian laurel-dominant green spaces had larger and more stable vegetation index values than mixed-dominant green spaces. Larger-sized green spaces were considerably cooler, with size explaining nearly 30% of temperature variability. Secondly, greener regions with greater vegetation index values always possessed lower temperatures, and the association between greenness and cooling effect became even stronger over time. The findings suggest the critical significance of vegetation type, green space area, and vegetation density in alleviating urban heat island effects and hold significant implications for urban planners who desire to construct more robust urban green infrastructure in fast-growing cities. Several research studies have examined the impact of land surface temperature on urban vegetation, including: [43,44,45].

Other studies have tried targeting the specific subsets of the parameters, such as climatic parameters.[32] carried out a study on significant Moroccan cities to investigate diurnal surface temperature behavior during the vegetative growth period and interactions among urban and vegetation surfaces. The study introduced the Urban Thermal Impact Ratio (UTIR), a new index that is set to quantify the urban thermal comfort from the intensity of the surface urban heat island (SUHI). Using a land surface model (LSM) from 2010, that integrated high-resolution Landsat and MODIS data for twelve land cover classes, active winter vegetation was the cause of moderate urban–vegetation temperature differences in coastal cities like Tangier and Rabat. Urban form dominated in the case of inland locations, where Fes exhibited a less intense SUHI compared to Meknes but with a higher surface area of impervious surfaces. Urban desert Dakhla exhibited the typical trend, with nighttime SUHI of 2.1 °C and daytime cooling of −0.7 °C, under the influence of irrigated parks and lawns boosting evapotranspiration and shading. The study highlighted vegetation as the governor of urban surface temperature, evened-out urban–rural thermal gradients, and provided a useful, reproducible indicator of urban heat stress. These findings are useful for climate-resilient city planning, effective use of energy, and the design of public health early warning systems. [46] conducted studies to test the impact of physical indicators of urban green spaces on the thermal environment of the areas. Temperature data and physical indicators of 36 green spaces in Xi'an were collected, and correlation analysis was used to explore the relationship between green space characteristics and internal temperature. The results indicated that to achieve the lowest internal temperature, the area of green space should be between 0.6–0.7 km^2, and the perimeter should range from 4000 to 4500 m. Similarly, for green urban water bodies, the size should be between 0.3–0.4 km^2 and the perimeter about 5000 m so that internal temperature is reduced. Urban planners and policymakers can directly apply these metrics to improve urban thermal comfort.

Given these results, and due to the limited studies which simultaneously emphasize climatic and environmental conditions, the principal objective of this study is to detect and measure the most efficient factors leading to Arbil city's urban green areas. Drawing on remote sensing data as well as advanced machine learning techniques, the research explores both environmental and climatic variables to determine what has the greatest impact on urban green spaces' size, distribution, and health. Outputs are expected to provide evidence-based basis for sustainable urban design and green infrastructure management.

Study area

Erbil, the capital of the Kurdistan Region of Iraq, is one of the world's oldest continuously occupied cities, situated around 36.19°N latitude and 44.01°E longitude, with an approximate area of 2,200 km² (Figure. 1). The city has a semi-arid climate with hot dry summers and cool wet winters, and most of its mean annual rainfall of 400–500 mm is received from November to April, while the summer temperatures exceed 40°C, resulting in widespread urban heat stress. The city has a flat topography with some low hills surrounding it, which influence the wind regime and the urban microclimate. High urbanization in the recent decades, triggered by an increasing population, increased development activities, and land use alteration, has put strong pressure on natural and green areas. Arbil is richly provided with various parks, gardens, and urban green areas and thus is an appropriate case study in investigations of interlinks between urban green spaces, environmental characteristics, and climatic conditions. Selecting Arbil as the site for study makes it possible to examine the dynamics of urban parks in an intensively urbanizing semi-arid city, which is the typical condition for most Middle Eastern cities sharing the same climatic and environmental concerns.

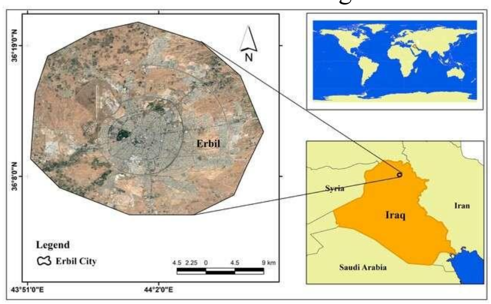

Figure. 1: Study area: Arbil city, Iraq

Material And Methods

Research methodology involves the following major steps:

1- Downloading climatic and environmental data from remote sensing images.
2- Preparing the data for input into machine learning, including tests for multicollinearity, variance inflation, and normalization.
3- Running machine learning models.
4- Identifying the most influential factors affecting urban green spaces.

The overall methodology is illustrated in Figure 2, and the detailed explanation of each step is provided in the following subsections.

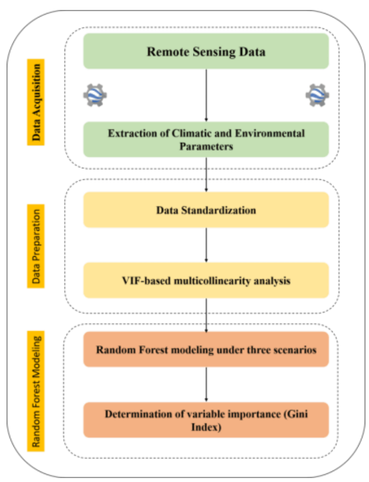

Figure 2. Workflow of the present study

In the present study, several climatic and environmental parameters were utilized to examine their trends and determine their impact on urban green spaces. For getting comprehensive and high-resolution data, three datasets were utilized: MODIS, providing high-resolution information on land surface characteristics and vegetation processes; TerraClimate, providing high-resolution monthly climate and climatic water balance data; and CHIRPS, providing reliable estimates of precipitation over spatially variable domains. The merging of these datasets provided a sound determination of temporal and spatial changes in environmental conditions and thus allowed a comprehensive examination of their influence on the distribution, status, and dynamics of urban greenspaces. Data processing was all performed in the Google Earth Engine (GEE) environment, utilizing its cloud-based geospatial computational power [47]. The data were preprocessed to achieve monthly time series data from 2000 to 2025 in support of continuous temporal monitoring and assessment of environmental and climatic parameters. GEE's high-performance computing platform facilitates efficient manipulation of remotely sensed large-scale data, thus enabling rapid analysis, spatial summation, and retrieval of relevant parameters for large urban regions.

Here, the urban green cover was assessed using the Normalized Difference Vegetation Index (NDVI) obtained from MODIS imagery. Climatic indicators employed are land surface temperature, precipitation, wind speed, potential evapotranspiration, actual evapotranspiration, soil moisture, atmospheric vapor pressure, and total incoming surface energy. Environmental indicators are Normalized Difference Built-up Index (NDBI), Vegetation Condition Index (VCI), Palmer Drought Severity Index (PDSI), land surface albedo, and aerosol optical depth. The specifications of the individual products and datasets used in detail are presented in Table 1.

Table 1. Environmental and Climate variables used in the current study.

Variable	Sensor	Units
NDVI	MOD13Q1	----
LST	MOD11A2	Kelvin
Albedo	MCD43A3	----
NDBI	MOD09A1	----
Precipitation (PreC)	CHIRPS	mm
VCI	MOD13A2	----
Aerosol Optical Depth (AOD)	MCD19A2	----
Actual evapotranspiration (aet)	TERRACLIMATE	mm
Reference evapotranspiration (pet)	TERRACLIMATE	mm
Wind Speed (WS)	TERRACLIMATE	m/s
Vapor Pressure (VP)	TERRACLIMATE	kPa
Downward surface shortwave radiation (srad)	TERRACLIMATE	W/m^2
Soil moisture (SM)	TERRACLIMATE	mm
PDSI	TERRACLIMATE	----

Methodology

Data standardization

Standardization is a significant preprocessing phase, particularly in the case of high-dimensional data, and a requirement prior to collinearity analysis and model construction. By eliminating the impact of different scales of various variables, this method enhances the performance and accuracy of machine learning models [48,49]. All the variables in the present study were standardized using Equation (1) as follows:

$$Z_i = \frac{X_{i-}\mu}{\sigma} \qquad (Eq. 1)$$

In the formula above, Z_i is the standard score for data X_i, μ is the mean and σ is the standard deviation of the data. By doing this, the Z_i's will have a mean of 0 and a variance of 1.

Multicollinearity Analysis

Collinearity arises when a multiple regression model independent variable has a linear relationship with one or more other independent variables, thereby making it, in effect, a linear combination of them. Similarly, multicollinearity arises if many explanatory variables are linearly related and can be expressed as linear combinations of one another. The presence of multicollinearity or collinearity in a regression equation undermines the validity of the estimated coefficients because the effect of every variable on the dependent variable becomes enmeshed with the effect of other variables. This results in inflated variances for coefficient estimates and can even generate biased forecasts. Thus, even minimal changes to the input data can initiate extreme oscillations in the regression coefficients [50,51]. The Variance Inflation Factor (VIF) is often used to quantify the level of such inflation in coefficients' variance and is calculated as demonstrated in Equation (2), providing a diagnostic statistic for detecting collinearity in regression models.

$$VIF = \frac{1}{TC} = \frac{1}{(1-R_i^2)} \qquad \text{(Eq.2)}$$

In this context, Ri^2 represents the unadjusted coefficient of determination obtained when the independent variable is regressed on all other independent variables. The tolerance coefficient (TC) is defined as the reciprocal of the VIF. A low TC value (TC < 0.2) signals a strong correlation among the independent variables, whereas TC values above 0.2 suggest minimal multicollinearity effects. To mitigate multicollinearity, one common strategy is to remove variables that are highly correlated with others—a procedure adopted in this study.

Random Forest (RF)

The Random Forest algorithm is one of the most used embedded machine learning methods for variable importance and feature selection. The accuracy of this method greatly depends on the number of trees constructed [52,53]. Feature importance is derived using the Gini Index (or Shannon Entropy), which provides an effective measure for the selection of predictors contributing the most. At every node z, in this process, a subset of the variables is randomly selected as splitting candidate variables. For each splitting candidate variable, heterogeneity reduction is calculated. When a variable splits a node, the heterogeneity index indicates by how much the split alters the Gini Index. Upon each split of a variable, the corresponding heterogeneity reduction is saved. These values are then summed over all nodes and scaled by the number of trees in the forest to determine the global Gini index, as illustrated in Equation (3) [54,55]. This method facilitates stable detection and ordering of the most influential features at the topmost level in the model.

$$Gini\ index = \frac{1}{n}\sum_{z}[d(x,z) \cdot I(x,z)] \qquad (Eq.3)$$

Here, I(h,z) is a function that equals 1 if variable x-th variable is used for splitting at the node z, and equals 0 if it is not used.

Results

In Figure 3, overall temporal trends of 14 parameters are presented, with Normalized Difference Vegetation Index (NDVI) as the explained variable and 13 environmental and climatic variables as explaining variables. The figure presents an overall view of the dynamics and relationships between the green cover of urban areas and its potential driving forces over time. Before passing the data to Random Forest machine learning algorithm, the normalization has to be carried out. Then, Variance Inflation Factor (VIF) is acquired, and independent variables with lowest multicollinearity are retained in the model only. The calculation of VIF is done iteratively: the variable with the highest VIF is eliminated in each iteration, and this continues until all variables remaining have a VIF of less than 10. Table 2 shows the Variance Inflation Factor (VIF) analysis, which was conducted to check for the degree of multicollinearity between predictor variables. A three-step systematic process was employed to identify and remove highly collinear variables. For step one, PET was identified to have the highest VIF value and therefore exhibited high multicollinearity and therefore it was removed from the data. For step two, the SRAD and LST were compared, and it was identified that SRAD had a higher VIF value and therefore removed and the model re-run. Following these iterative processes, all the remaining variables had VIF values less than 10, indicating that multicollinearity was successfully addressed. These findings affirm that the chosen variables are suitable for application in the Random Forest model, yielding consistent estimation of variable importance and stable model performance.

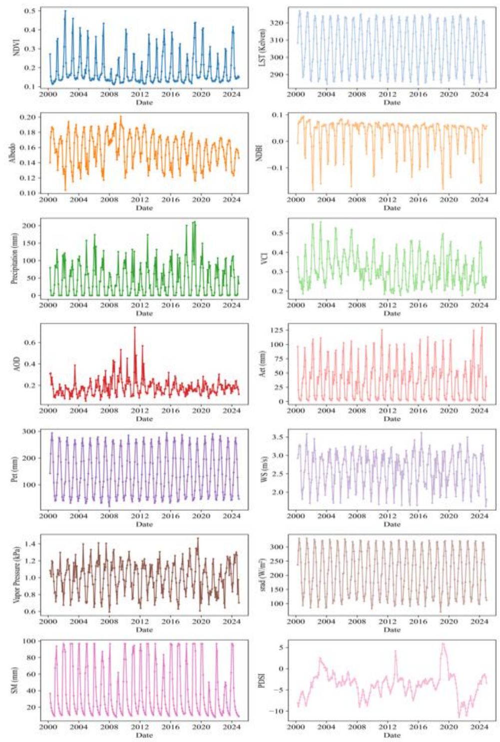

Figure 3: Temporal variations of the dependent and independent environmental and climatic variables.

Table 2 – Stepwise Calculation and Presentation of Variance Inflation Factor (VIF)

Step-1		Step-2		Step-2	
VIF		VIF		VIF	
LST	59.59	LST	45.59	LST	6.5
Albedo	5.27	Albedo	4.75	Albedo	4.06
NDBI	4.9	NDBI	4.84	NDBI	4.77
PreC	5.09	PreC	4.87	PreC	4.43
VCI	2.99	VCI	2.85	VCI	2.78
AOD	1.8	AOD	1.66	AOD	1.57
AET	5.79	aet	4.56	aet	4.35
PET	75.09	WS	5.32	WS	4.56
WS	6.4	VP	4.59	VP	4.42
VP	4.77	srad	30.8	SM	4.23
srad	44.32	SM	4.39	PDSI	1.27
SM	4.4	PDSI	1.27		
PDSI	1.28				

Table 3 illustrates the prediction accuracy of the machine learning model for a 70%-30% training and testing dataset split in three scenarios. In Scenario 1, when climatic as well as environmental factors were considered, the model achieved the highest degree of predictive accuracy with an R-squared of 0.98, indicating that nearly all the variance within the provided data was explained. The corresponding error measurements (MAE = 0.0271, MSE = 0.0017, RMSE = 0.0414, MAPE = 0.2962) were very low, which justified the outstanding predictive capability of the model. This suggests that the interaction between climatic and environmental variables provides a comprehensive description of the system, and their interaction tends to enhance model performance.

In Scenario 2, where only climatic parameters were used, the model's predictability decreased as evident in lower R-squared value of 0.89 and larger error measures (MAE = 0.0453, MSE = 0.0053, RMSE = 0.0728, MAPE = 0.4833).

The results indicate that climatic variables are not sufficient to capture all the variability in the data set and that environmental variables must be added to make more accurate predictions.
In Scenario 3 with environmental variables alone, the model had an R-squared of 0.97 with error metrics (MAE = 0.0278, MSE = 0.0017, RMSE = 0.0416, MAPE = 0.3137) similar to those of the full parameter scenario. This again indicates how the environmental variables alone are highly informative and can nearly equal the predictive accuracy when using the full variables.
In general, it appears that environmental parameters play a greater role compared to climatic parameters in determining the model's predictive ability. While climatic variables also make their contributions, their contributions are secondary, as evidenced by the decisive role of environmental parameters in this instance.

Table 3: Performance metrics obtained from the Random Forest model for three different scenarios

	All parameter	Climatic parameter	Environmental parameter
R2	0.98	0.89	0.97
MAE	0.0271	0.0453	0.0278
MSE	0.0017	0.0053	0.0017
RMSE	0.0414	0.0728	0.0416
MAPE	0.2962	0.4833	0.3137

The results of the Gini importance index for the independent variables across the three scenarios are presented in Figure 4. In Scenario 1, where both climatic and environmental variables were considered, the NDBI (Normalized Difference Built-up Index), which reflects urban growth, exhibited the highest influence and importance. Following NDBI, the drought-related index VCI (Vegetation Condition Index) demonstrated substantial importance. The relative importance of the remaining variables, in descending order, was as follows: NDBI, VCI, Albedo, PDSI, AOD, SM, AET, LST, PreC, WS, and VP. Among all variables, atmospheric pressure (VP) exhibited the lowest importance. These results highlight that urbanization and vegetation condition are the primary drivers in the model, while other environmental and climatic factors contribute to varying degrees.

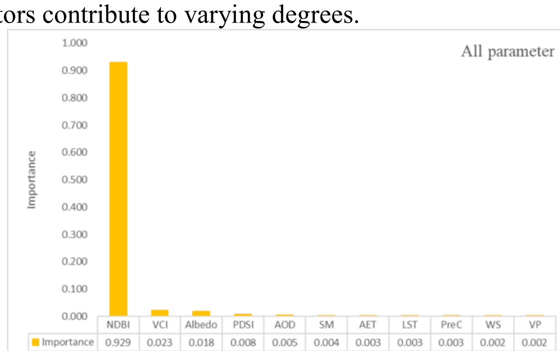

Figure 4: Importance coefficients (Gini index) for all environmental and climatic variables
In Scenario 2, considering only climatic variables, the highest influence was observed for by actual evapotranspiration (AET) in the region. Given the arid and semi-arid climatic conditions of the study area, this result is consistent with expectations. The relative importance of the climatic parameters, in descending order, was as follows: AET, SM (soil moisture), PreC (precipitation), VP (vapor pressure), and WS (wind speed), as illustrated in Figure 5.

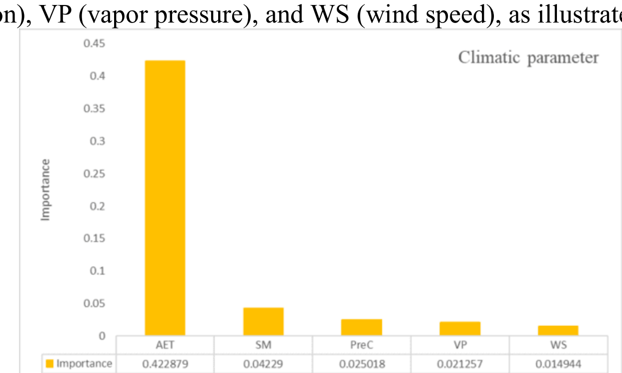

Figure 5: Importance coefficients (Gini index) for climatic variables

For the next scenario, considering only environmental parameters, the Normalized Difference Built-up Index (NDBI) exhibited the highest importance coefficient, as shown in Figure 6. Based on the Gini importance values, the relative importance of the environmental parameters, in descending order, was as follows: NDBI, VCI (Vegetation Condition Index), Albedo, PDSI (Palmer Drought Severity Index), AOD (Aerosol Optical Depth), and LST (Land Surface Temperature).

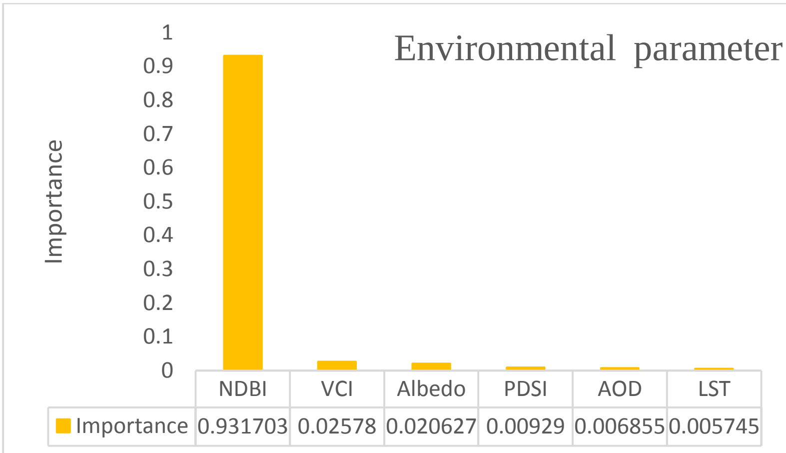

Figure 6: Importance coefficients (Gini index) for environmental variables

Conclusions

The present study investigated the predictive power of a Random Forest model under three sets of circumstances, with climatic and environmental factors separately and in combination. The model showed considerable predictive potential, particularly with the use of the two sets of variables combined. In this high-parameter scenario, the model generated the greatest R-squared value (0.98) and smallest error metrics (MAE = 0.0271, MSE = 0.0017, RMSE = 0.0414, MAPE = 0.2962), indicating that the combined information from climatic and environmental indices provides an extensive representation of the system behind it. These findings accentuate the importance of integrating different types of variables to increase the strength and reliability of environmental predictive models.

Variable importance analysis using the Gini index revealed urbanization, as represented by the Normalized Difference Built-up Index (NDBI), to be consistently in the top position for environmental variables [56,57,58,59,60]. The most significant climatic variable was Vegetation Condition Index (VCI), followed by actual evapotranspiration (AET), soil moisture (SM), and precipitation (PreC). The dominance of NDBI and VCI stresses the main role of land cover change and vegetation status in controlling the simulated phenomena, particularly over arid and semi-arid regions where these factors are very sensitive to climate fluctuation.

Based on climatic parameters only, model performance was decreased (R-squared = 0.89), mirroring the poor predictive capability of climatic factors alone. Using environmental factors only gave results nearly identical to the full-parameter case (R-squared = 0.97), suggesting that in regions with strong anthropogenic and environmental gradients, environmental indicators such as urbanization and surface characteristics will be more controlling than climate-related variables for individual prediction tasks.

Overall, the study shows how Random Forest models can integrate heterogeneous environmental and climatic data to achieve superior predictive performance. From the results, it is clear that proper choice of influential variables, particularly the environmental variables, is key to generating robust model performance. The Gini importance analysis also provides insightful knowledge on the relative relevance of individual variables, which is valuable in giving recommendations for possible future monitoring and management strategies for urbanized and ecologically sensitive regions.

Reference

[1]. Gu C. Urbanization: Processes and driving forces. Science China Earth Sciences. 2019 Sep;62(9):1351-60.
[2]. Cheshmehzangi A, Tang T. Chongqing: Navigating rapid urbanization and regional development challenges. InChina Under Construction: Shaping Cities Through Recent Urban Transformation 2024 Dec 12 (pp. 81-106). Singapore: Springer Nature Singapore.
[3]. Teklemariam N. Historic preservation as sustainable urban development in African cities: A technical and technological framework. Sustainability. 2024 Jul 12;16(14):5949.

[4]. Abed AD, Khalaf HK, Al-Mussawy HA, Azamathulla HM. Land-use planning between master plans and reality: analysing land-use transformation and urban planning challenges in Samarra District, Iraq. Town Planning Review. 2025 Jun 30:1-22.
[5]. Hama A, Djeffal D. Analyzing the impact of urban growth on urban quality of life in the city of El Eluma: A multidimensional study. Glasnik Srpskog geografskog drustva. 2025;105(1):457-74.
[6]. Mansour S, Ghoneim E, El-Kersh A, Said S, Abdelnaby S. Spatiotemporal monitoring of urban sprawl in a coastal city using GIS-based Markov chain and artificial neural network (ANN). Remote Sensing. 2023 Jan 19;15(3):601.
[7]. Al Tarawneh WM. Urban sprawl on agricultural land (literature survey of causes, effects, relationship with land use planning and environment): a case study from Jordan (Shihan Municipality Areas). Journal of Environment and Earth Science. 2014;4(20):97-124.
[8]. Ojobo H, Oluwagbemiga PA, Shamang KJ. Unveiling the impact of urban green landscape on quality of life in kaduna, Nigeria: residents' perceptions and sustainable strategies. Journal of Contemporary Urban Affairs. 2024 Jan 21;8(1):16-36.
[9]. Priya UK, Senthil R. Framework for enhancing urban living through sustainable plant selection in residential green spaces. Urban Science. 2024 Dec 1;8(4):235.
[10]. Addas A. Influence of urban green spaces on quality of life and health with smart city design. Land. 2023 Apr 26;12(5):960.
[11]. Vargas-Hernández JG, Pallagst K, Zdunek-Wielgołaska J. Urban green spaces as a component of an ecosystem. InHandbook of engaged sustainability 2018 (pp. 1-32). Springer, Cham.
[12]. Elderbrock E, Enright C, Lynch KA, Rempel AR. A guide to public green space planning for urban ecosystem services. Land. 2020 Oct 14;9(10):391.
[13]. Ćwik A, Wójcik T, Ziaja M, Wójcik M, Kluska K, Kasprzyk I. Ecosystem services and disservices of vegetation in recreational urban blue-green spaces—Some recommendations for greenery shaping. Forests. 2021 Aug 12;12(8):1077.
[14]. Edeigba BA, Ashinze UK, Umoh AA, Biu PW, Daraojimba AI, Edeigba BA, Ashinze UK, Umoh AA, Biu PW, Daraojimba AI. Urban green spaces and their impact on environmental health: A Global Review. World J. Adv. Res. Rev. 2024 Feb 28;21(2):917-27.
[15]. Artmann M, Inostroza L, Fan P. Urban sprawl, compact urban development and green cities. How much do we know, how much do we agree?. Ecological indicators. 2019 Jan 1;96:3-9.
[16]. Ramaiah M, Avtar R. Urban green spaces and their need in cities of rapidly urbanizing India: A review. Urban science. 2019 Aug 25;3(3):94.
[17]. Yasin MY, Yusoff MM, Abdullah J, Noor NM, Noor NM. Urban sprawl literature review: Definition and driving force. Geografia. 2021;17(2).
[18]. Sun R, Chen L. Effects of green space dynamics on urban heat islands: Mitigation and diversification. Ecosystem Services. 2017 Feb 1;23:38-46.
[19]. Arshad S, Ahmad SR, Abbas S, Asharf A, Siddiqui NA, ul Islam Z. Quantifying the contribution of diminishing green spaces and urban sprawl to urban heat island effect in a rapidly urbanizing metropolitan city of Pakistan. Land use policy. 2022 Feb 1;113:105874.
[20]. Zhou W, Yu W, Zhang Z, Cao W, Wu T. How can urban green spaces be planned to mitigate urban heat island effect under different climatic backgrounds? A threshold-based perspective. Science of the total environment. 2023 Sep 10;890:164422.
[21]. Elmarakby E, Elkadi H. Prioritising urban heat island mitigation interventions: Mapping a heat risk index. Science of the total environment. 2024 Oct 20;948:174927.
[22]. Ibsen PC, Crawford BR, Corro LM, Bagstad KJ, McNellis BE, Jenerette GD, Diffendorfer JE. Urban tree cover provides consistent mitigation of extreme heat in arid but not humid cities. Sustainable Cities and Society. 2024 Oct 15;113:105677.
[23]. Liu C, Lu S, Tian J, Yin L, Wang L, Zheng W. Research overview on urban heat islands driven by computational intelligence. Land. 2024 Dec 13;13(12):2176.
[24]. Luo Y, Wang Q, Yang K, Zhang Y, Shang C, Pei X, Chen R, Li D, Peng C. The formation process, mechanism, and attribution of urban impervious surface thermal runoff. Water Resources Research. 2025 Feb;61(2):e2024WR037696.
[25]. Patt A, Pfenninger S, Lilliestam J. Vulnerability of solar energy infrastructure and output to climate change. Climatic change. 2013 Nov;121(1):93-102.

[26]. Marzo A, Trigo-Gonzalez M, Alonso-Montesinos J, Martínez-Durbán M, López G, Ferrada P, Fuentealba E, Cortés M, Batlles FJ. Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. Renewable Energy. 2017 Dec 1;113:303-11.
[27]. Maimaitiyiming M, Ghulam A, Tiyip T, Pla F, Latorre-Carmona P, Halik Ü, Sawut M, Caetano M. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. ISPRS Journal of Photogrammetry and Remote Sensing. 2014 Mar 1;89:59-66.
[28]. Abulibdeh A. Analysis of urban heat island characteristics and mitigation strategies for eight arid and semi-arid gulf region cities. Environmental Earth Sciences. 2021 Apr;80(7):259.
[29]. Roy PS, Ramachandran RM, Paul O, Thakur PK, Ravan S, Behera MD, Sarangi C, Kanawade VP. Anthropogenic land use and land cover changes—A review on its environmental consequences and climate change. Journal of the Indian Society of Remote Sensing. 2022 Aug;50(8):1615-40.
[30]. Norouzi M, Chau HW, Jamei E. Design and Site-Related Factors Impacting the Cooling Performance of Urban Parks in Different Climate Zones: A Systematic Review. Land. 2024 Dec 13;13(12):2175.
[31]. Priya UK, Senthil R. Framework for enhancing urban living through sustainable plant selection in residential green spaces. Urban Science. 2024 Dec 1;8(4):235.
[32]. Ed-dahmany N, Bounoua L, Lachkham MA, Yacoubi Khebiza M, Bahi H, Messouli M. Interplay Between Vegetation and Urban Climate in Morocco—Impact on Human Thermal Comfort. Urban Science. 2025 Jul 25;9(8):289.
[33]. Siddiqui A, Maske AB, Khan A, Kar A, Bhatt M, Bharadwaj V, Kant Y, Hamdi R. An Urban Climate Paradox of Anthropogenic Heat Flux and Urban Cool Island in a Semi-Arid Urban Environment. Atmosphere. 2025 Feb 1;16(2).
[34]. Sadok W, Lopez JR, Smith KP. Transpiration increases under high□temperature stress: Potential mechanisms, trade□offs and prospects for crop resilience in a warming world. Plant, Cell & Environment. 2021 Jul;44(7):2102-16.
[35]. Otchere-Darko E, Atuah L, Opoku R, Koranteng C. Redefining green roof systems with climbers: simulation of a conceptual model for thermal-radiative performance and plant vitality. International Journal of Building Pathology and Adaptation. 2025 May 13;43(3):528-47.
[36]. Pandey B, Ghosh A. Urban ecosystem services and climate change: a dynamic interplay. Frontiers in Sustainable Cities. 2023 Oct 30;5:1281430.
[37]. Huang S, Tang L, Hupy JP, Wang Y, Shao G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. Journal of forestry research. 2021 Feb;32(1):1-6.
[38]. Zhao Q, Qu Y. The retrieval of ground ndvi (normalized difference vegetation index) data consistent with remote-sensing observations. Remote Sensing. 2024 Mar 29;16(7):1212.
[39]. Chen K, Wang Y, Huang C, Wang J, Li SL, Guan H, Ma L. GreenNet: A dual-encoder network for urban green space classification using high-resolution remotely sensed images. International Journal of Applied Earth Observation and Geoinformation. 2025 Aug 1;142:104709.
[40]. Neyns R, Canters F. Mapping of urban vegetation with high-resolution remote sensing: A review. Remote sensing. 2022 Feb 21;14(4):1031.
[41]. Mostafazadeh R, Alaei N, Mirchooli F, Hussain S. Changes in urban green space configuration and connectivity using spatial graph-based metrics in Ardabil developing city, Iran. Environmental Monitoring and Assessment. 2024 Sep;196(9):778.
[42]. Aleixo C, Branquinho C, Laanisto L, Tryjanowski P, Niinemets Ü, Moretti M, Samson R, Pinho P. Urban Green Connectivity Assessment: A Comparative Study of Datasets in European Cities. Remote Sensing. 2024 Feb 22;16(5):771.
[43]. Asgarian A, Amiri BJ, Sakieh Y. Assessing the effect of green cover spatial patterns on urban land surface temperature using landscape metrics approach. Urban Ecosystems. 2015 Mar;18(1):209-22.
[44]. Yao L, Li T, Xu M, Xu Y. How the landscape features of urban green space impact seasonal land surface temperatures at a city-block-scale: An urban heat island study in Beijing, China. Urban Forestry & Urban Greening. 2020 Jun 1;52:126704.
[45]. Hasan M, Hassan L, Al MA, Abualreesh MH, Idris MH, Kamal AH. Urban green space mediates spatiotemporal variation in land surface temperature: a case study of an urbanized city, Bangladesh. Environmental Science and Pollution Research. 2022 May;29(24):36376-91.
[46]. Zhang Q, Zhou D, Xu D, Cheng J, Rogora A. Influencing factors of the thermal environment of urban green space. Heliyon. 2022 Nov 1;8(11).

[47]. Velastegui-Montoya A, Montalván-Burbano N, Carrión-Mero P, Rivera-Torres H, Sadeck L, Adami M. Google Earth Engine: a global analysis and future trends. Remote Sensing. 2023 Jul 23;15(14):3675.
[48]. Kwon A, Lee HY, Shin SY, Yang K, Sung Y, Lee K, Byeon NS, Lim TH, Lee JH. Current health data standardization project and future directions to ensure interoperability in Korea. Healthcare Informatics Research. 2024 Apr 30;30(2):93-102.
[49]. Chew EY, Burns SA, Abraham AG, Bakhoum MF, Beckman JA, Chui TY, Finger RP, Frangi AF, Gottesman RF, Grant MB, Hanssen H. Standardization and clinical applications of retinal imaging biomarkers for cardiovascular disease: a Roadmap from an NHLBI workshop. Nature Reviews Cardiology. 2025 Jan;22(1):47-63.
[50]. Kalnins A, Praitis Hill K. The VIF score. What is it good for? Absolutely nothing. Organizational research methods. 2025 Jan;28(1):58-75.
[51]. Kyriazos T, Poga M. Dealing with multicollinearity in factor analysis: the problem, detections, and solutions. Open Journal of Statistics. 2023 Jun 15;13(3):404-24.
[52]. Salman HA, Kalakech A, Steiti A. Random forest algorithm overview. Babylonian Journal of Machine Learning. 2024 Jun 8;2024:69-79.
[53]. Parmar A, Katariya R, Patel V. A review on random forest: An ensemble classifier. InInternational conference on intelligent data communication technologies and internet of things 2018 Aug 7 (pp. 758-763). Cham: Springer International Publishing.
[54]. Algehyne EA, Jibril ML, Algehainy NA, Alamri OA, Alzahrani AK. Fuzzy neural network expert system with an improved Gini index random forest-based feature importance measure algorithm for early diagnosis of breast cancer in Saudi Arabia. Big Data and Cognitive Computing. 2022 Jan 27;6(1):13.
[55]. Daniya T, Geetha M, Kumar KS. Classification and regression trees with gini index. Advances in Mathematics: Scientific Journal. 2020 Sep 23;9(10):8237-47.
[56]. Lane KJ, Stokes EC, Seto KC, Thanikachalam S, Thanikachalam M, Bell ML. Associations between greenness, impervious surface area, and nighttime lights on biomarkers of vascular aging in Chennai, India. Environmental health perspectives. 2017 Aug 2;125(8):087003.
[57]. Kuang W. Mapping global impervious surface area and green space within urban environments. Science China Earth Sciences. 2019 Oct;62(10):1591-606.
[58]. Yan Z, Teng M, He W, Liu A, Li Y, Wang P. Impervious surface area is a key predictor for urban plant diversity in a city undergone rapid urbanization. Science of the Total Environment. 2019 Feb 10;650:335-42.
[59]. Dou Y, Kuang W. A comparative analysis of urban impervious surface and green space and their dynamics among 318 different size cities in China in the past 25 years. Science of the Total Environment. 2020 Mar 1;706:135828.
[60]. Zhong Q, Li Z. Long-term trends of vegetation greenness under different urban development intensities in 889 global cities. Sustainable Cities and Society. 2024 Jul 1;106:105406.

دراسة المساحة الخضراء لمدينة اربيل وتحديد معاييرها الفعالة باستخدام التعلم الآلي

تەرزه جبار باقي[1]   سروشت محمد جزا [1]   هيوا فرج پناه[2]
[1] جامعة السليمانية التقنية، المعهد دربنديخان التقني، قسم المساحة، السليمانية، إقليم كردستان، العراق.
[2] جامعة الشهيد شمران، كلية الهندسة المدنية والعمارة، الأهواز- ايران.

الخلاصة
يُعزى دقة التنبؤ بالأحداث البيئية إلى الجمع بين المتغيرات المناخية والبيئية. استُخدم نموذج الغابة العشوائية في هذه الدراسة لدراسة مساهمة مختلف المتغيرات المناخية والبيئية في الظاهرة المعنية ضمن ثلاثة سيناريوهات: (1) جميع المتغيرات مجتمعة، (2) المتغيرات المناخية فقط، و(3) المتغيرات البيئية فقط. أُجري التدريب باستخدام 70% من البيانات، واختبار 30% منها. استُخدمت مقاييس الأداء مثل ***R-squared*** ، و***MAE***، و***MSE***، و***RMSE***، و ***MAPE***لتقييم دقة التنبؤ، بينما حُسبت المساهمة النسبية لكل متغير باستخدام مؤشر جيني المهم.
تشير النتائج إلى تحقيق أفضل أداء للتنبؤ عند تضمين جميع المتغيرات (***R-squared = 0.98***، ***MAE = 0.0271***، ***RMSE = 0.0414***)، ولكن كان الأداء أقل عند استخدام المتغيرات المناخية وحدها (***R-squared = 0.89***). أعطت المتغيرات البيئية وحدها أداءً قريبًا جدًا من أداء حالة المعلمة الكاملة (***R-squared = 0.97***)، مما يعكس دورها الأفضل في التنبؤ بالنموذج. كشف تحليل أهمية المتغيرات عن أهم العوامل المساهمة في المتغيرات مثل التحضر (***NDBI***) وحالة الغطاء النباتي (***VCI***)، ثم البياض، و***PDSI***، و***AOD***، و***LST***. كانت رطوبة التربة والتبخر النتح الفعلي من العوامل المناخية ذات المساهمة المتوسطة، بينما كان الضغط الجوي وسرعة الرياح أقل مساهمة. تُظهر هذه النتائج أن نماذج الغابات العشوائية يمكنها التقاط التفاعلات المعقدة في الأنظمة البيئية بشكل فعال وتؤكد على الدور الأساسي للمتغيرات البيئية، وخاصة التحضر وحالة الغطاء النباتي، في النمذجة التنبؤية. يوفر هذا البحث رؤى لاستراتيجيات الرصد والإدارة البيئية المستقبلية في المناطق القاحلة وشبه القاحلة.

الكلمات المفتاحية: الغابات العشوائية، التنبؤ البيئي، المتغيرات المناخية، التحضر (***NDBI***)، مؤشر حالة الغطاء النباتي (***VCI***).