Quantitative Analysis of Jogging Behavior Spatial Differentiation Mechanism in Washington, D.C. Streets Based on the XGBoost-SHAP Method and Urban Construction Characteristics

Hanlin Xiong
Academy of Geography and Information Engineering, China University of Geosciences, Wuhan, China

Abstract: With the acceleration of urbanization, outdoor sports, especially jogging, have gradually gained popularity among urban residents. Urban greening, particularly street trees, as an important means to improve the quality of life, has a significant impact on residents' sports behavior. However, there is a lack of research on the relationship between street tree characteristics and the willingness to jog. Based on the characteristics of street trees in Washington, D.C., this study quantitatively analyzes 11 street - related characteristic variables such as tree pit length, crown width, crown length, and the number of street trees. The XGBoost regression model is used to predict the willingness to jog, and the SHAP method is combined for result interpretability analysis. The study aims to reveal the potential relationship between these street characteristics and residents' willingness to jog. The goal of this study is to provide scientific basis for urban planners to optimize the design of street trees and urban greening, thereby promoting residents' healthy behaviors and participation in jogging activities. By filling the research gap in the relationship between street trees and the willingness to jog, this paper provides theoretical support and practical guidance for the construction of healthy cities and the improvement of residents' quality of life.

Keywords: Component; Urban Construction Environment; Street Trees; Xgboost-SHAP Interpretation; Jogging Willingness

1. Introduction

Built envirUrban greening is an important means to improve the quality of urban life and the ecological environment. As a core component of urban greening, street trees not only have functions such as beautifying the environment, purifying the air, and providing shade, but may also have a profound impact on residents' lifestyles and behaviors. In recent years, with the acceleration of urbanization and the improvement of residents' health awareness, more and more urban residents choose to engage in outdoor sports such as jogging on the roads. Therefore, the characteristics of street trees may have an important impact on residents' willingness to jog, but this relationship has not been systematically explored.

This study aims to explore the potential impact of various biological characteristics of street trees on the willingness to jog in urban areas through quantitative analysis. Based on the eleven collected street - related characteristic variables, including tree pit length, crown width, crown length, number of street trees, intersecting area with parks, width of street tree pits, number of buildings, number of households, degree of construction, aesthetic degree of construction, and entertainment index, we constructed multiple variables to describe the relationship between these characteristics and the willingness to jog. Through these variables, we hope to comprehensively reveal the potential connection between tree characteristics and the willingness to jog.

In terms of data analysis methods, we first divided the continuous jogging willingness index into three categories - low, medium, and high - through hierarchical classification, representing different intensities of jogging willingness. Then, we used the random forest model to screen out the most influential independent variables and employed the XGBoost model to conduct classification prediction of jogging willingness. Finally, we combined the SHAP method for interpretable analysis of the results. Through this series of methods, this study aims to provide scientific evidence for urban planners to assist in optimizing the planting and management of street trees, thereby promoting the healthy behaviors of urban residents, especially their participation in jogging activities.

This study not only fills the research gap in the relationship between urban greening and residents' behavior but also provides a new perspective for promoting the construction of healthy cities and improving residents' quality of life.

2. Research Background

With the continuous advancement of urbanization, jogging, as a low-cost and high-benefit aerobic exercise, has gradually become the main fitness method for urban residents. In recent years, scholars have conducted extensive research on the influence of the urban environment andonment on jogging behavior, especially the exploration of green environmental factors has increased. Gao et al. proposed a framework for evaluating the "runnability" of urban streets based on big data, emphasizing the use of geospatial big data and regression analysis methods to assess the running friendliness of urban streets. The framework covers the influencing factors of three aspects: the built environment, street environment, and natural environment, especially road accessibility, public transport convenience, greening, and sense of security [1]. Zhang Ying et al. studied the influence of the urban environment on jogging behavior through pedometers and questionnaires, and found that factors such as fitness facilities, traffic safety, and road design have a significant impact on jogging activities, especially the accessibility of fitness facilities and traffic safety [2].

In the study of the impact of green spaces on jogging, Feng Ziyi et al. proposed that green space accessibility and spatial greening level are the key factors affecting the selection of jogging routes, and suggested optimizing the running environment by improving the layout of green spaces and enhancing lighting [3]. Zhang et al. used the AHP expert scoring method and combined it with fitness application recommendation data to explore the impact of 22 environmental factors on running routes. They found that environmental factors such as green spaces and waterfront spaces are particularly concerned by the public, especially in high - frequency running routes [4].

In specific environmental impact studies, Liu et al. investigated the impact of the urban park environment in Chongqing on running behavior and found that park trails and water surface landscapes had a positive effect on the running flow, while the landscape shape index and the distance from the city center had a negative impact [5]. Yang et al. adopted interpretable machine learning methods (GBDT and SHAP) in their multiple studies and combined big data and geospatial analysis to explore the non - linear impact and interaction of the urban environment on jogging behavior. For example, Yang et al. proposed an analysis framework based on GBDT and SHAP through analyzing jogging data in Beijing, which could well explain the multi - level and non - linear impact of environmental factors on jogging behavior, especially factors such as green space coverage and public facility density [6 - 8].

In terms of micro - environmental characteristics, Jiang et al. found that street safety, the size of open spaces, and road width significantly promote running behavior. In contrast, factors such as high work density, point - of - interest density, canopy density, and air pollution impede running activities [11]. Huang et al. investigated the relationship between street green view index (GVI) and running intensity and found that street greening can significantly enhance running intensity and has a positive correlation with running frequency and duration [13]. Chen et al. used the geographically weighted random forest and SHAP models to explore the non - linear relationship between urban green view index (GVI) and jogging behavior, further revealing the impact of green spaces on running behavior [14].

Yang Wei et al.'s research focused on the interaction between the urban built environment and jogging behavior, especially the spatial differences at the block scale. The study found that factors such as population density, land - use mix, and bus stop density are the main factors influencing outdoor jogging behavior, and emphasized the spatial non - stationarity and interaction effects of these factors [8]. Liu et al. studied facilities such as the number and distance of parks and running tracks and found that they have a significant impact on jogging behavior. Particularly within a specific quantity range, an increase in the number of parks and running tracks can significantly promote running activities [15].

Although numerous studies have explored the impact of the urban environment on jogging behavior, most of them focus on the urban scale or relatively macro - environmental factors. There is limited in - depth analysis of the influence of specific green space features at the block scale (such as the number of road buildings, the total number of households, the intensity of road construction, the aesthetic degree of construction, the area of parks within the field of vision, the species of street trees, tree crowns, tree diameters, etc.). The innovation of this project lies in starting from the block scale, with a focus on studying the impact of urban construction and green space environmental factors on jogging behavior, especially conducting a systematic analysis of the differentiation mechanism of green space. By using the XGBoost regression model and SHAP interpretation method, this study aims to explore and predict the influence of different street environmental characteristics on street jogging behavior and further quantify the intensity of these influencing factors under different environmental conditions.

3. Predict and Analysis

The Xgboost-Shap method was used to perform the classification fitting and analysis between feature variables and the jogging willingness index, and a dependence graph was generated. First, the ROC table of the classification was output and evaluated. Then, the random forest was utilized to select the optimal number of the best feature indicators, which were then input into the xgboost model to achieve classification fitting, and the prediction set was predicted.

3.1 Indicators of ROC Classification Results

The classification models of these three categories all have a certain ability to distinguish positive and negative samples, and their performances are relatively close. The model of class 1 has the highest AUC value and performs relatively the best; the AUC value of class 0 is the lowest, but the difference is not significant.

3.2 Feature Selection Using Random Forest

Random forest was used to select variables most strongly correlated with the jogging willingness index, and 10 feature variables were obtained from 11 feature indicators as strongly correlated feature indicators. The feature indices are shown in the figure 1 below.

Ranked from high to low, they are the width of road tree pits, the number of road trees, the length of tree canopies, the width of tree canopies, the number of buildings, the length of tree pits, the number of households, the intersecting area of parks, the degree of construction, and the degree of construction beautification. Relatively speaking, the influence of the number of entertainment facilities on the jogging willingness index is not obvious.

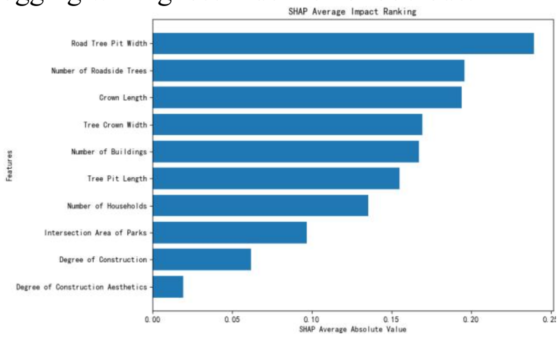

Figure 1. Characteristic Index

3.3 Grid Search for Hyperparameters and XGBoost Fitting

Using the top ten best feature indicators obtained, the optimal hyperparameters were obtained through grid search. The results are shown in the following table 1.

Table 1. Stylesresult of Hyperparameters

<table>
<tr><td rowspan="3">Parameter</td><td colspan="3">The best parameter results</td></tr>
<tr><td>learning_rate</td><td>max_depth</td><td>n_estimators</td></tr>
<tr><td>0.3</td><td>7</td><td>500</td></tr>
</table>

Through grid search, the optimal hyperparameters were determined to be learning_rate=0.3, max_depth=7, and n_estimators=500. A learning_rate of 0.3 strikes a balance between convergence speed and model stability, avoiding excessive fitting while ensuring efficient parameter updating. A max_depth of 7 enables the model to capture complex nonlinear relationships among features without introducing excessive complexity, preventing overfitting and maintaining good generalization ability for classifying jogging willingness indices.

Among them, learning_rate represents the optimal learning rate of the XGBoost model, max_depth represents the optimal depth of the model, and n_estimators represents the optimal number of iterations of the model.

After completing the search for the optimal hyperparameters, 1143 road data points from the training set were input into the XGBoost model for classification fitting. The calculation results of the SHAP values were output, and the model was used to predict 12665 data points in the prediction set. The fitting results are shown in the following table 2.

Table 2. Model Metrics

Indicator name	Classification result		
	0	1	2
Precision	0.76	0.78	0.77
Recall	0.78	0.79	0.79
f1-score	0.77	0.75	0.75

The model results indicate that the Xgboost has achieved the fitting of eleven feature indicators and the willingness to jog to a certain extent. Among them, the F1 score performs well in the three classifications. Moreover, there are no obvious differences or imbalance problems among the three classes (0, 1, and 2).

After completing the model fitting, the prediction set is predicted, and the following map of jogging willingness at the street scale in Washington, D.C. predicted by the xgboost model is obtained (see Figure 2).

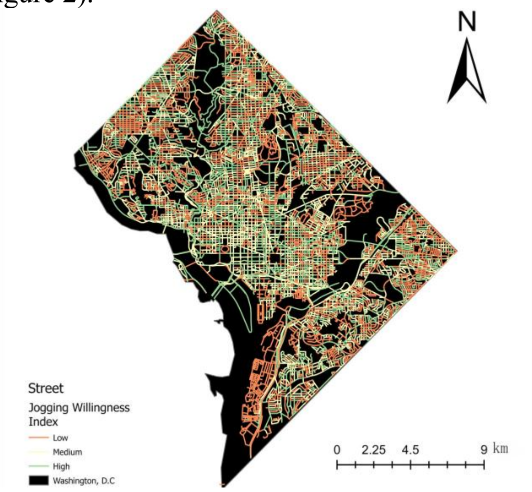

Figure 2. Predict Result

3.4 SHAP Explanation and Prediction

The SHAP interpretation was used for this model to explore the correlations and relationships among its various feature variables. The SHAP feature ranking plot is shown as follows Figure 3.

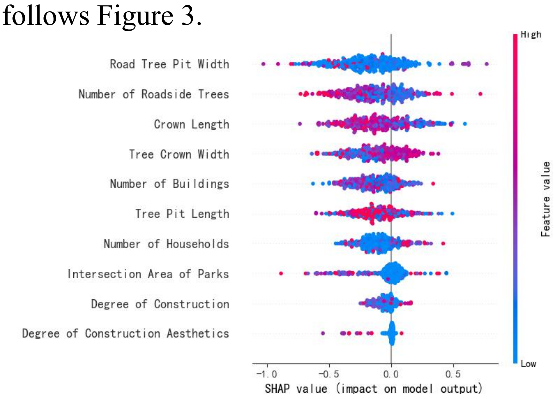

Figure 3. SHAP Value

As can be seen from the figure, the degree of feature influence is consistent with the ranking results of feature selection by the random forest. The width of road tree pits has both positive and negative impacts on the willingness to jog, showing a scattered distribution at different levels. However, the low values of this feature tend to reduce the index of jogging willingness. That is, according to the model, roads with a low width of road tree pits have a lower willingness to jog, areas with a large number of road trees have a higher willingness to jog, roads with a longer and wider tree crown have a higher willingness to jog, and areas with a large number of buildings have a lower willingness to jog. In contrast, the intersection area of parks, the degree of construction, and the degree of construction beautification have little impact on the willingness to jog.

The dependency graph of each of its indicators is shown in the following figure 4.

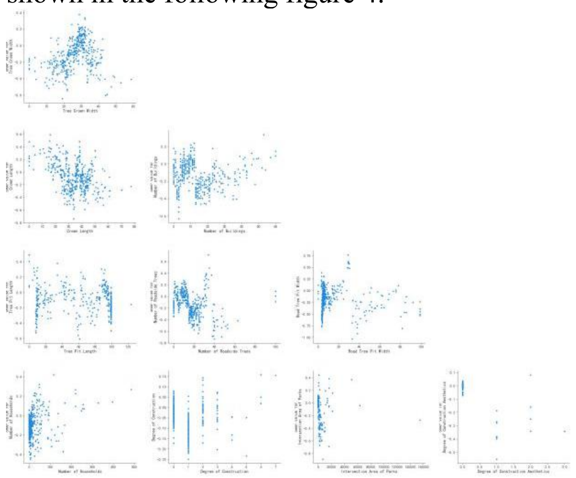

Figure 4. Dependence Diagram

From the dependency graph, it can be found that when the width of the road tree pool is between 0-30 m, it has a positive impact on the willingness to jog. After that, as the width of the tree pool increases, it has a negative impact on the willingness to jog. The number of road trees and the crown length in the concentrated area have a negative impact on the willingness to jog, while the tree crowns with a width between 20-40 m have a significant positive impact on the willingness to jog. The increase in the number of road buildings and the number of families has a significant impact on the willingness to jog. However, the intersecting area of the park, the degree of construction, and the degree of construction beautification have little impact on the willingness to jog.

4. Donclusion

This study provides valuable insights for urban planners and policymakers by quantitatively analyzing the relationship between the urban construction characteristics of streets and the spatial differentiation mechanism of jogging behavior. Through dissecting SHAP value distributions and feature dependencies, clear actionable directions emerge. For street tree characteristics, planners should prioritize cultivating species with crown lengths in the 20–50 m range and crown widths of 20–40 m (aligning with positive impact zones in the data), and strategically increase planting density in jogging-frequent areas like park perimeters and residential paths while balancing shade provision and pedestrian flow. When designing tree pits, setting widths within 0–30 m can maximize positive effects on jogging willingness; for larger pits (e.g., to accommodate mature trees), supplementing with adjacent small-scale greenery patches can offset potential negative influences. Regarding built elements, controlling building density in jogging-focused zones and adopting a "low-rise + open-space corridor" layout in residential blocks helps preserve continuous green pathways. Even for factors with relatively minor direct impacts, like park intersection areas and construction aesthetics, integrating signed "green connectors" between parks and streets and adding subtle design elements (e.g., art-themed jogging stations) can incrementally enhance the jogging experience. These targeted measures, rooted in the research findings, offer a scientific basis for optimizing urban greening design-directly addressing how physical space shapes residents' health-promoting behaviors like jogging and empowering planners to craft more activity-friendly streetscapes.

In summary, this study provides a scientific basis for urban planners to optimize the planning of block construction and green space environment, especially the management of the quantity and morphology of street trees, by quantitatively analyzing the influence of street tree characteristics on the spatial differentiation mechanism of jogging behavior. Future research can be expanded in terms of data sources, model optimization, comprehensive multi-factor analysis, and field surveys to further deepen the understanding of the relationship between urban greening and residents' healthy behaviors. For instance, long-term monitoring could track how street tree growth stages dynamically affect jogging patterns, and cross-regional comparative studies might reveal climate or cultural context-specific optimization strategies, ensuring urban greening efforts evolve to better support diverse communities' health needs.

References

[1] Huagui Guo, Shuyu Zhang, Yufei Liu, et al. Building running-friendly cities: effects of streetscapes on running using 9.73 million fitness tracker data in Shanghai, China[J]. BMC Public Health, 2024, 24(1):2251-2251.

[2] Zhang Ying, Chen Liang, Liu Xin. Grey Relational Analysis of Urban Human Settlement Environment Factors Affecting Pedestrian Suitability in Coastal Cities. Journal of Environment and Health, 2010. 27(12): p. 1106-1108.

[3] Feng Ziji, Ma Danya. Research on the Strategy of Creating Urban Jogging Environments Based on Mobile Phone Fitness Data: A Case Study of Tianjin. In 2022/2023 China Urban Planning Conference. 2023. Wuhan, Hubei, China.

[4] Yujia Zhong, Meng Guo, Menghan Zhang, et al. Identifying Street Environmental Factors That Attract Public Attention from the Jogger's Perspective: A Multiscale Spatial Exploration[J]. Buildings, 2024, 14(7):1935-1935.

[5] Liu Yong, Hu Jie, Yang Wei, et al. Effects of urban park environment on recreational jogging activity based on trajectory data: A case of Chongqing, China[J]. Urban Forestry & Urban Greening, 2022, 67

[6] Yang Wei, Li Yingpeng, Liu Yong, et al. Environmental factors for outdoor jogging in Beijing: Insights from using explainable spatial machine learning and massive trajectory data[J]. Landscape and Urban Planning, 2024, 243:104969.

[7] Wei Yang, Jun Fei, Yingpeng Li, et al. Unraveling nonlinear and interaction effects of multilevel built environment features on outdoor jogging with explainable machine learning[J]. Cities, 2024, 147:104813.

[8] Yang Wei, Hu Jie, Liu Yong. The Correlation and Interaction between Built Environment and Outdoor Jogging Supported by Crowdsourced Geographic Information. Landscape Architecture, 2024. 31(04): p. 44-52.

[9] Nadine Schuurman, Leah Rosenkrantz, Scott

A Lear. Environmental Preferences and Concerns of Recreational Road Runners[J]. International Journal of Environmental Research and Public Health, 2021, 18(12): p.6268.

[10] Mingke Xie, Zhangxian Feng, Wang Long, et al. What are the environmental preferences of runners? Evidence from Guangzhou[J]. Applied Geography, 2025, 174: p.103469.

[11] Jiang, H., D. Lin, Q. Bing, How Are Macro-Scale & Micro-Scale Built Environments Associated with Running Activity? The Application of Strava Data & Deep Learning in Inner London. ISPRS Int. J. Geo Inf., 2022. 11: p. 504.

[12] Yang Linchuan, Yu Bingjie, Liang Pengpeng, et al. Crowdsourced Data for Physical Activity-Built Environment Research: Applying Strava Data in Chengdu, China
[J]. Frontiers in Public Health, 2022, 10:883177-883177.

[13] Huang Dengkai, Tian Meng, Yuan Lei. Sustainable design of running friendly streets: Environmental exposures predict runnability by Volunteered Geographic Information and multilevel model approaches[J]. Sustainable Cities and Society, 2023, 89.

[14] Cai, C., et al., Unraveling nonlinear effects of environment features on green view index using multiple data sources & explainable machine learning. Scientific Reports, 2024. 14.

[15] Liu Yong, Li Yingpeng, Yang Wei, et al. Exploring nonlinear effects of built environment on jogging behavior using random forest[J]. Applied Geography, 2023, 156.