Global NEST Journal, **28** (8), 08700
Available online: 19/10/2026

Machine Learning-Based Identification of Ecological Sensitivity Types and Associated Environmental Patterns in the Changsha–Zhuzhou–Xiangtan Metropolitan Area

Zijian Zheng[1], Xianchao Zhao[1,2]* and Ang Zhou[1]

[1]School of Civil and Environmental Engineering, Hunan University of Technology, Zhuzhou 412007, China
[2] School of Geography and Tourism, Hengyang Normal University, Hengyang 421010, China

*Correspondence e-mail: zhaoxianchao@hut.edu.cn

Abstract

Rapid urbanization has intensified ecological pressure and increased the complexity of ecological responses in metropolitan areas. This study identified ecological sensitivity types and their associated environmental patterns in the Changsha–Zhuzhou–Xiangtan metropolitan area. Twelve indicators representing natural geography, soil environment, ecological background, and anthropogenic disturbance were analysed using self-organizing maps, XGBoost, and SHAP. Four types were identified: anthropogenic activity-associated, hydrothermal constraint-associated, ecological background-associated, and natural–anthropogenic composite-associated areas, accounting for 8.78%, 38.85%, 21.16%, and 31.21% of the study area, respectively. $PM_{2.5}$ had the highest mean absolute SHAP value, followed by annual precipitation, nighttime light, GDP, animal diversity, and land-surface temperature. These variables should be interpreted as factors associated with model classification rather than as evidence of causal ecological drivers. The four types exhibited distinct association patterns, supporting differentiated ecological governance and spatial planning in rapidly urbanizing metropolitan areas.

Keywords: self-organizing map; XGBoost; SHAP; spatial clustering; environmental modelling; metropolitan area ecosystem.

OPEN ACCESS

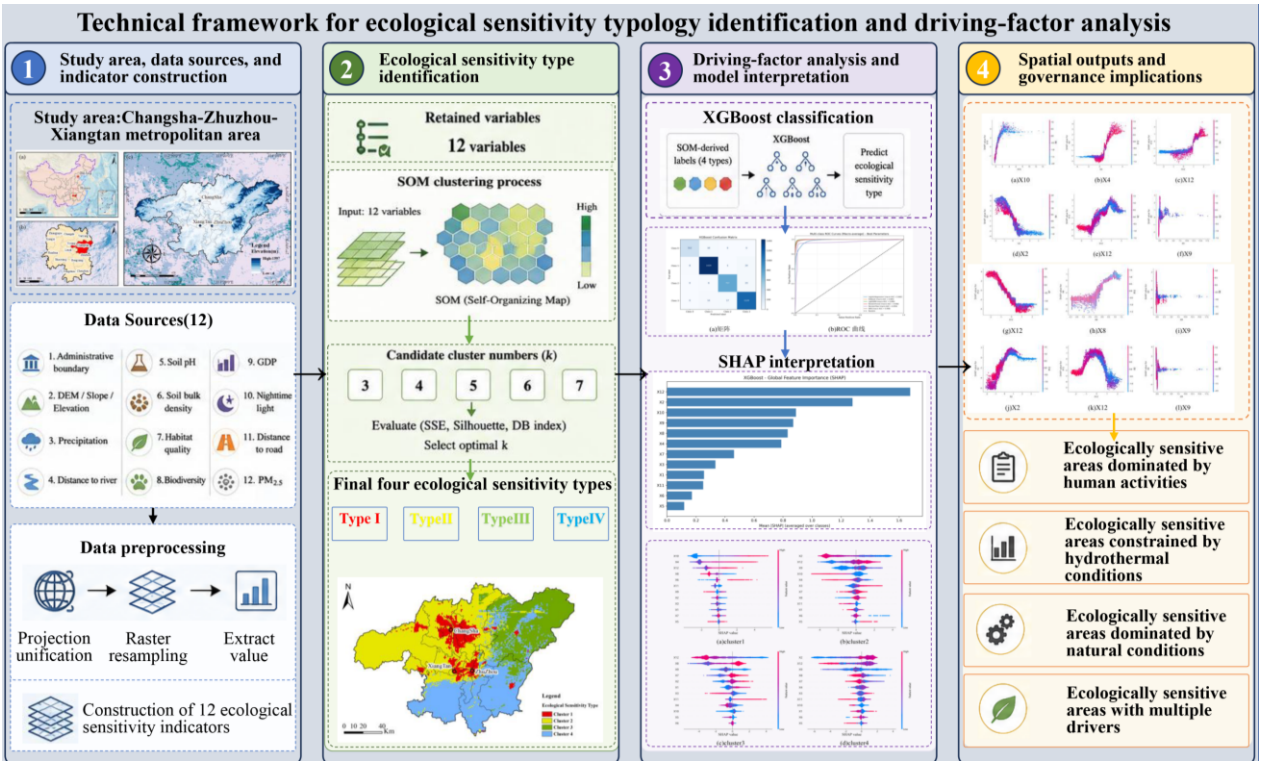
Received: 12/06/2026,

Accepted: 27/09/2026,

Available online: 19/10/2026

Copyright: © 2026 The Author(s). Published in Global NEST Journal. This article is licensed under the terms and conditions of the Creative Commons Attribution 4.0 International license (CC BY 4.0).

Graphical abstract



1. Introduction

Rapid urbanization and global environmental change have intensified pressures on urban ecosystems, including land development, built-up expansion, habitat fragmentation, pollutant emissions, and climate risks (Huang *et al.*, 2019; Chen *et al.*, 2020). These pressures reshape land-use patterns, ecological processes, and ecosystem service supply. Previous studies have shown that urban expansion affects animal diversity, carbon storage, ecosystem services, and urban ecological sustainability, and recent projections further suggest that future urban land expansion may generate substantial animal diversity conservation risks (McDonald *et al.*, 2020; Simkin *et al.*, 2022).Metropolitan areas are typical coupled human–natural systems, where population, industries, infrastructure, and ecological spaces concentrate within limited land resources (Liu *et al.*, 2007; Fang and Yu, 2017; Elmqvist *et al.*, 2019). Hence, identifying spatial patterns and associated environmental conditions of ecological sensitivity is essential for ecological conservation, territorial planning, and sustainable metropolitan development (Peng *et al.*, 2018a; Meerow, 2020; Zhang *et al.*, 2022; Wang *et al.*, 2023).

Ecological sensitivity describes the degree to which ecosystems respond to environmental change and anthropogenic disturbance. It is widely used in ecological security assessment, land-use suitability evaluation, ecological restoration zoning, and spatial governance. Existing studies indicate that ecological sensitivity is jointly influenced by topography, hydrothermal conditions, soil properties, vegetation cover, land use, human activity intensity, and pollutant emissions, and usually exhibits significant spatial heterogeneity, scale dependence, and nonlinear responses (Turner *et al.*, 2003; Adger, 2006; Xu *et al.*, 2021; Li *et al.*, 2022; Chen *et al.*, 2023).Research on ecological vulnerability, ecosystem services, and landscape change further suggests that regional ecological security is shaped by interactions among natural processes, human activities, and spatial structure (Eigenbrod *et al.*, 2011; Peng *et al.*, 2018b; Estoque *et al.*, 2021). Traditional ecological sensitivity assessments rely on methods such as the analytic hierarchy process, Delphi method, weighted overlay, and entropy weighting. These approaches are intuitive and operationally simple, but they usually depend on subjective or predefined weights and linear aggregation assumptions (Malczewski, 2006; Langemeyer *et al.*, 2020). Consequently, they may fail to capture nonlinear relationships, threshold effects, and factor interactions (Olden *et al.*, 2008; Ryo and Rillig, 2017; Meyer *et al.*, 2019). Moreover, most studies classify sensitivity into low, medium, or high levels, while fewer identify distinct sensitivity types or explain their formation patterns. In metropolitan areas with strong human–environment coupling, a single composite index can obscure internal heterogeneity, multiple stressor combinations, and type-specific association patterns (Wu, 2014; Zhou *et al.*, 2021; Wang *et al.*, 2024).

Machine learning offers useful tools because it can handle high-dimensional data, model nonlinear relationships, and identify complex interactions in environmental and ecological systems (Olden *et al.*, 2008; Zurell *et al.*, 2020; Hengl *et al.*, 2018). The self-organizing map (SOM) is an unsupervised neural network that identifies similarities among multidimensional samples while preserving topological relationships, making it suitable for ecological classification, environmental regionalization, and pattern recognition in complex spatial datasets (Vesanto and Alhoniemi, 2000; Kohonen, 2001; Kim *et al.*,2023). XGBoost, an ensemble learning algorithm based on gradient boosting decision trees, has strong predictive performance and is widely used in environmental modelling (Chen and Guestrin, 2016). However, the black-box nature of machine learning limits direct ecological interpretation. SHapley Additive exPlanations (SHAP), derived from cooperative game theory, quantifies variable contributions and explains model behaviour globally and locally (Lundberg and Lee, 2017; Lundberg *et al.*, 2020). Thus, integrating SOM, XGBoost, and SHAP provides a coherent approach for identifying ecological sensitivity types, validating their separability, and interpreting association patterns.

Recent work also shows that environmental governance increasingly depends on the interaction between analytical technologies, organizational capacity, and communication. Human capital can condition the success of environmental technologies, while corporate biodiversity attention and information networks can shape green transformation and innovation (Lei and Zhang, 2025; Yin *et al.*, 2026; Lei *et al.*, 2026b). AI-assisted studies have further used large language models and multimodal analytical frameworks to identify environmental leadership signals and interpret public-sector sustainability communication (Lei *et al.*, 2026a; Luo *et al.*, 2026). These developments reinforce the need for interpretable analytical frameworks that connect model outputs with ecologically and institutionally meaningful decisions.

Such an approach is particularly relevant where development and conservation pressures overlap. The Changsha–Zhuzhou–Xiangtan metropolitan area, in east-central Hunan Province, is a key part of the Yangtze River middle-reach urban agglomeration. Accelerated integration of the three cities has led to rapid urban expansion, industrial agglomeration, and transport development, increasing ecological pressure in the Xiangjiang River Basin. At the same time, the region's ecological green heart, the Xiangjiang River system, and the eastern mountainous barrier form major components of its ecological security pattern. These characteristics make the area a suitable case for studying ecological sensitivity under coupled pressures. Ecological sensitivity here is influenced by natural factors (topography, hydrothermal conditions, vegetation cover, habitat quality) and anthropogenic factors (GDP, nighttime light index, road density, $PM_{2.5}$). Nighttime light index indicates urban expansion and human activity intensity (Elvidge *et al.*, 1997; Levin and Duke, 2012), while $PM_{2.5}$ and land surface temperature reflect pollution exposure and urban thermal stress (van Donkelaar *et al.*, 2015).

To address this gap, this study selects 12 influencing factors from four dimensions: natural geography, soil environment, ecological background, and human activities. SOM clustering, XGBoost classification, and SHAP interpretation methods are applied to identify ecological sensitivity types and analyze their association patterns. Specifically, SOM was used to identify ecological sensitivity types without predefined sensitivity grades. XGBoost was then applied to test whether these types could be distinguished in a multidimensional indicator space. Finally, SHAP was used to quantify factor importance, contribution direction, and interaction effects for different sensitivity types. The objectives of this study are to: (1) identify the spatial types of ecological sensitivity in the Changsha–Zhuzhou–Xiangtan metropolitan area; (2) reveal the dominant factors and association patterns of different sensitivity types; and (3) provide an interpretable data-driven basis for ecological restoration zoning, differentiated ecological regulation, and sustainable territorial spatial planning in rapidly urbanizing metropolitan areas.

1.1. *Study area, data sources and preprocessing*

The Changsha–Zhuzhou–Xiangtan metropolitan area is located in the east-central part of Hunan Province and constitutes an important component of the middle reaches of the Yangtze River urban agglomeration. It is also a core region of Hunan Province with a high concentration of population, industries, and infrastructure. The study area includes the entirety of Changsha, the central urban area of Zhuzhou and Liling City, and the central urban area of Xiangtan and Shaoshan City, involving a total of 19 counties, county-level cities, and districts. The total area is approximately 18,900 km², accounting for 8.92% of the total land area of Hunan Province. By the end of 2021, the permanent resident population of the study area was about 14.84 million, accounting for 22.41% of the total provincial population. It is therefore an important spatial carrier of urbanization and regional integration in Hunan Province.

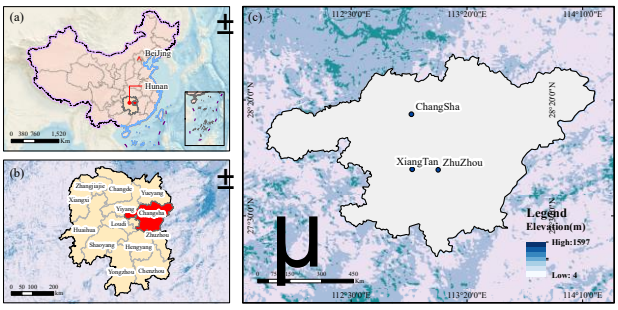


Figure 1. Study area.

The study area (**Figure 1**) is situated in the transitional zone from the central Hunan hills to the Dongting Lake Plain. The landform types include mountains, hills, tablelands, and plains, and the overall topography is characterized by relatively high elevations in the eastern and western parts and relatively flat terrain along the Xiangjiang River in the

central part. The Xiangjiang River and its tributaries run through the region and provide an important basis for water resource supply, urban spatial organization, and ecological corridor construction. The Changsha–Zhuzhou–Xiangtan ecological green heart, located at the junction of the three cities, covers an area of about 528.32 km² and represents an important ecological space for restraining continuous urban sprawl, maintaining ecological connectivity, and ensuring regional ecological security. With the advancement of regional integration, the expansion of construction land, industrial agglomeration, and the strengthening of transportation networks have continuously increased the pressure on ecological protection. Therefore, the region is characterized by both high-intensity development pressure and high ecological conservation value, making it a suitable case for ecological sensitivity studies at the metropolitan scale.

1.2. *Data sources*

This study used multi-source spatial datasets to characterize the natural, ecological, and socioeconomic conditions of the study area. The selected data included administrative boundary data, digital elevation model (DEM), annual precipitation, river network, land surface temperature, soil pH, soil bulk density, habitat quality, animal diversity, GDP, nighttime light index, road network, and $PM_{2.5}$ concentration. Administrative boundary data were used to define the study area. DEM data were used to calculate slope. River and road vector data were used to calculate distance to rivers and distance to roads through Euclidean distance analysis. Annual precipitation and land surface temperature were used to represent hydrothermal conditions. Soil pH and soil bulk density were used to describe soil physicochemical properties. Habitat quality and animal diversity were used to characterize ecological background conditions. GDP and nighttime light index data were used to represent economic development intensity and human activity intensity. $PM_{2.5}$ concentration was used to indicate air pollution pressure. All datasets were obtained from publicly available databases and geospatial data platforms. The main data sources and descriptions are shown in **Table 1**.

Table 1. Description of data sources.

Data type	Data name	Data name
Vector data	Administrative boundary	Hunan Standard Map Service Website (https://hunan.tianditu.gov.cn/)
Raster data	Slope	Geospatial Data Cloud (http://www.gscloud.cn/)
Raster data	Precipitation	National Earth System Science Data Center (http://www.geodata.cn/)
Raster data	Distance to rivers	National Basic Geographic Information Center (http://www.ngcc.cn/)
Raster data	Land surface temperature	Earth Resource Data Cloud (http://www.gis5g.com)
Raster data	Soil pH	Food and Agriculture Organization of the United Nations (FAO) official website (https://www.fao.org/)
Raster data	Soil bulk density	Food and Agriculture Organization of the United Nations (FAO) official website (https://www.fao.org/)
Raster data	Habitat quality	Zenodo platform (https://zenodo.org/)
Raster data	Animal diversity	Science Data Bank website (https://doi.org/10.57760/sciencedb.20221)
Raster data	GDP	Resource and Environmental Science Data Platform (https://www.resdc.cn/)
Raster data	Nighttime light index	Resource and Environmental Science Data Platform (https://www.resdc.cn/)
Raster data	Distance to roads	OpenStreetMap website (https://www.openstreetmap.org/)
Raster data	$PM_{2.5}$	National Earth System Science Data Center (http://www.geodata.cn/)

Table 2. Factors Influencing Ecological Sensitivity in the Changsha–Zhuzhou–Xiangtan Metropolitan Area.

Dimension	Factor	Code	Factor interpretation
Natural geography	Slope	X1	Extent of surface undulation of a spatial unit
	Annual precipitation	X2	Sum of the average monthly precipitation in a year
	Distance to rivers	X3	Reflects the influence of rivers on the adjacent ecological environment
	Land surface temperature	X4	Physical quantity reflecting the thermal conditions and spatial variation of the surface
Soil factors	Soil pH	X5	Indicates the strength of acidic or alkaline reactions in soil
	Soil bulk density	X6	Key parameter measuring soil structure, compactness, and changes in soil properties
Ecological background	Habitat quality	X7	Provides conditions and capabilities for the survival, reproduction, and maintenance of organisms and populations
	Animal diversity	X8	Reflects species richness, ecosystem complexity, and genetic diversity
Human activities	GDP	X9	Reflects the scale of regional economic development and overall economic strength
	Nighttime light index	X10	Reflects the intensity of human activities
	Distance to roads	X11	Reflects the impact of road construction and transportation on adjacent ecosystems
	$PM_{2.5}$ concentration	X12	Reflects air pollution level

1.3. *Indicator system of ecological sensitivity*

Ecological sensitivity is determined by the combined effects of natural environmental conditions, ecosystem status, and anthropogenic disturbances. Previous studies have generally developed ecological sensitivity or ecological vulnerability indicator systems using variables related to topography, hydrology, climate, soil, vegetation, land use, and socioeconomic activities (Turner *et al.*, 2003; Peng *et al.*, 2018b). Considering the strong coupling between natural processes and socioeconomic activities in metropolitan areas, 12 influencing factors were selected in this study based on the principles of scientific relevance, systematicity, representativeness, and data availability. These factors were grouped into four dimensions: natural geography, soil environment, ecological background, and human activities (**Table 2**).

The natural geography dimension comprises slope, annual precipitation, distance to rivers, and land surface temperature, which represent terrain conditions, water availability, proximity to water resources, and surface thermal environment, respectively. The soil environment dimension includes soil pH and soil bulk density, reflecting soil physicochemical properties and structural stability. The ecological background dimension consists of habitat quality and animal diversity, which characterize ecosystem quality, habitat integrity, species richness, and ecological conservation value. The human activities dimension includes GDP, nighttime light index, distance to roads, and $PM_{2.5}$ concentration, which indicate economic development intensity, human activity intensity, transportation disturbance, and air pollution pressure, respectively (**Figure 2**).

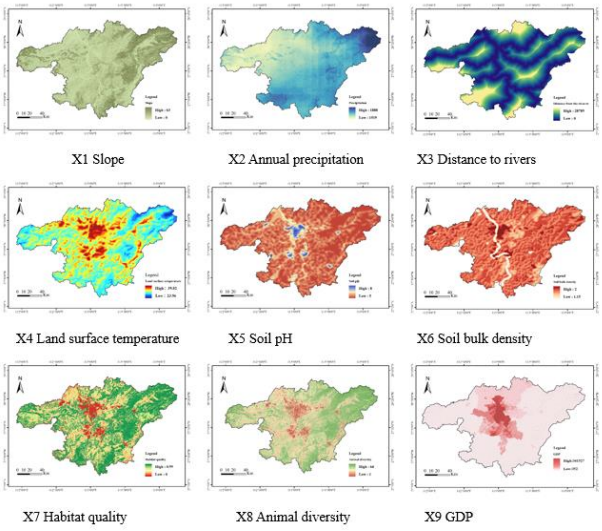


Figure 2. Spatial distribution of ecological sensitivity influencing factors in the Changsha–Zhuzhou–Xiangtan metropolitan area.

1.4. *Data preprocessing and multicollinearity test*

To ensure spatial consistency among multi-source datasets, all layers were projected to WGS 84 / UTM zone 48N (EPSG:32648), clipped to the study boundary, and aligned to a common 1 km × 1 km grid. Vector-derived distances were calculated in the projected coordinate system. The 12 ecological indicators were standardized using z scores, z = (x - mean)/standard deviation, before modelling. Projected x and y coordinates were standardized separately before their use in the spatially constrained SOM. Complete cases were retained, yielding 18,710 grid samples.

Before SOM, XGBoost, and SHAP analyses, multicollinearity among the 12 indicators was tested. Spearman rank correlation coefficients were used to examine pairwise correlations among variables. In addition, tolerance values and variance inflation factors (VIFs) were calculated to evaluate multicollinearity. The results show that all tolerance values were greater than 0.10 and the VIF values ranged from 1.106 to 3.283, which are below commonly used thresholds (O'Brien, 2007). This indicates that there was no serious multicollinearity among the selected indicators. Therefore, all 12 indicators were retained for subsequent analysis.

2. Methods

2.1. *Identification of ecological sensitivity types using SOM*

The Self-Organising Map (SOM) is an unsupervised neural network that maps high-dimensional data to a low-dimensional topological space while preserving neighbourhood relationships among samples (Kohonen, 1982). Through competitive learning and iterative weight updating, SOM can reveal latent clustering structures without predefined class labels and support the adaptive classification of complex multidimensional data (Vesanto and Alhoniemi, 2000). The calculation formula is as follows:

$$d_j = \|X - W_j\| = = \sqrt{\sum_{i=1}^{N}\left[x_i(t) - w_{ij}(t)\right]^2} \quad (1)$$

In **Equation (1),** X represents the input sample vector, W_j represents the weight vector, N represents the number of input neurons, X_i represents the i input neuron, W_{ij} represents the random initial value of the connection weight between the ith input neuron and the jth output neuron, and t represents the time step.

(4) Let the neighbourhood be defined as $S_k(t)$. The weights of the winning node and the nodes within its neighbourhood are updated, with the change in weights given by:

$$\Delta w_{ij} = \eta(t)\left[x_i(t) - W_{ij}(t)\right], j \in S_k(t) \quad (2)$$

In **Equation (2)**, η is a positive learning rate that decreases over time.

(5) The SOM input comprised the 12 z-standardized ecological indicators and separately standardized projected x and y coordinates, which were included to preserve spatial context. A 20 × 20 hexagonal grid (400 neurons) was trained in R using the kohonen package, random seed 123, 1,000 iterations, a learning-rate decay from 0.05 to 0.01, and a neighbourhood-radius decay from 3 to 1. The codebook vectors were clustered using Ward.D2

hierarchical clustering. Candidate solutions with k = 3-7 were evaluated using the full-sample average silhouette coefficient and cluster-size balance. To prevent either excessive aggregation or poorly represented groups, candidate solutions were required to have a minimum cluster proportion of at least 5% and a maximum cluster proportion of no more than 60%; the eligible solution with the highest silhouette coefficient was selected. Spatial continuity and ecological interpretability were then examined as supporting criteria.

2.2. *Classification analysis using XGBoost*

Extreme Gradient Boosting (XGBoost) is an ensemble learning algorithm based on gradient boosting decision trees that builds a strong predictive model by iteratively combining multiple weak learners (Chen and Guestrin, 2016). Owing to its regularisation strategy, parallel computing capacity, and ability to handle nonlinear relationships and complex interactions, XGBoost is widely used in ecological and environmental modelling (Ryo and Rillig, 2017). In this study, SOM clustering results were used as classification labels, and the 12 ecological sensitivity factors were used as explanatory variables to construct an XGBoost classification model for identifying the main drivers of different ecological sensitivity types. The objective of XGBoost is to minimise an objective function consisting of a loss function and a regularisation term, expressed as:

$$L(\theta) = \sum_i l(y_i, \overline{y_i}) + \sum \Omega\left[f(k)\right] \tag{3}$$

In **Equation 3**, $l(y_i, \overline{y_i})$ represents the loss function, i.e. the error between the predicted valueY_i and the actual value $\overline{y_i}$; $\Omega[f(k)]$ is used to control the model's complexity and prevent overfitting, and is defined as follows:

$$\Omega\left[f(k)\right] = \gamma T + \frac{1}{2}\lambda \sum_{\varphi} \Omega_{\varphi}^{\ 2} \tag{4}$$

In **Equation 4**, *T* represents the number of leaf nodes in the decision tree, Ω_{φ} represents the weight assigned to the leaves, and γ and λ are the correlation coefficients for the leaf nodes.

The retained XGBoost configuration used colsample_bytree = 0.5, gamma = 0.0, learning_rate = 0.1714, max_depth = 8, n_estimators = 300, and subsample = 1.0. The original evaluation used a class-preserving 80% training and 20% held-out test split. Model stability and overfitting risk were further examined using stratified random five-fold cross-validation and spatial-region five-fold validation. For the latter, standardized projected coordinates were clustered into five contiguous, Voronoi-like regions, and each complete region was held out once. A one-at-a-time sensitivity analysis varied maximum depth (6 and 10), learning rate (0.10 and 0.20), number of trees (200 and 400), colsample_bytree (0.70), and subsample (0.80) around the retained configuration. Accuracy, the training-validation accuracy gap, and one-versus-rest macro-average AUC were calculated. Because XGBoost used labels generated by SOM from the same ecological indicators, these metrics assess internal label separability and spatial transferability rather than independent ecological validity.

2.3. *Interpretation of Associated Factors Using SHAP*

Although XGBoost has strong predictive ability, its complex structure limits direct interpretation. SHAP was used to quantify how each feature contributed to the model output at global and type-specific levels (Lundberg and Lee, 2017). SHAP values describe associations learned by the fitted model; they do not establish causal effects or ecological mechanisms. Accordingly, the results are reported as model-explained association patterns, with ecological processes discussed as plausible interpretations that require independent validation.

$$\varphi_j = \sum_{S \subseteq \{x_1, \ldots x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} \left(f\left(S \cup \{x_i\}\right) - f(S) \right) \tag{5}$$

In **Equation 5**, S represents the feature subset in the model, *x* represents the feature vector of the instance to be interpreted,*P* represents the number of features, and $f(S)$ represents the prediction of the feature values in the set *S*.

Let *f* be the original model to be explained and *g* be the explanatory model. We interpret the output of the original model based on the idea of local explanations. The input to the original model *x*, is transformed via a mapping function $x = h_x(x')$ to yield the input to the explanatory model x' such that $g(z') \approx f\left(h_x(z')\right)$, where $z' \approx x'$. SHAP represents the interpretation of Shapley values as an additive feature attribution method; therefore, SHAP can define the explanation as:

$$g(z') = \varphi_0 + \sum_{j=1}^{M} \varphi_z z'_j \tag{6}$$

In **Equation (6)**, *g* represents the explanatory model, $z' \{0,1\}^M$ represents the feature combination vector, *M* represents the simplified number of input features, and φ_j represents the Shapley value of feature *j* in the feature attribution.

3. Results and Analysis

3.1. *Spatial distribution of ecological sensitivity types*

The quantitative comparison supported retention of four SOM clusters (**Table 3; Figure 3**). The three-cluster solution had the highest average silhouette coefficient (0.2000), but one cluster contained 13,110 samples (70.07%), exceeding the 60% balance threshold and obscuring substantial internal heterogeneity. The four-cluster solution had an average silhouette coefficient of 0.1516 and contained 1,635, 7,268, 3,965, and 5,842 samples (8.74%, 38.84%, 21.19%, and 31.22%), respectively. Increasing k to five produced only a small silhouette increase (0.1562) while creating a 130-sample cluster (0.69%); similarly small groups occurred for k = 6 and k = 7. Four clusters were therefore the only solution satisfying both balance constraints and provided the most defensible compromise among separation, representation, spatial continuity, and ecological interpretability. Their mapped areas were 1,633.29, 7,226.01, 3,934.55, and 5,803.95 km², accounting for 8.78%, 38.85%, 21.16%, and 31.21% of the study area.

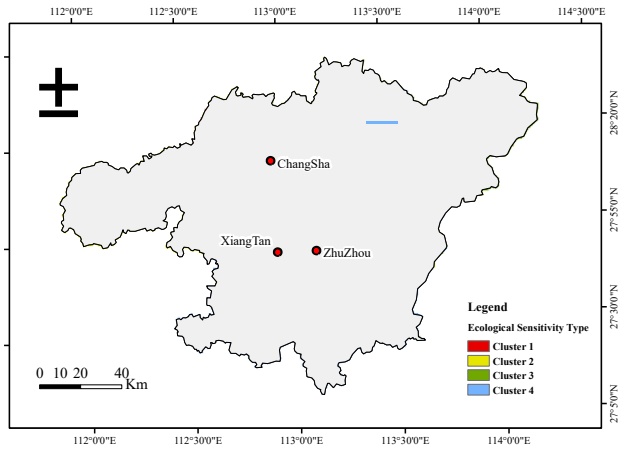


Figure 3. Cluster Classification.

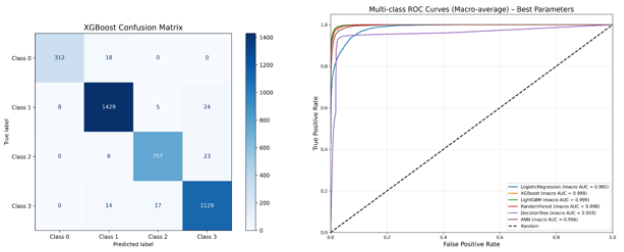
Spatially, Cluster 1 was concentrated in the central urban areas of Changsha, Zhuzhou, and Xiangtan and their surrounding built-up zones. Cluster 2 was mainly distributed in the western and northern parts of the study area and showed relatively continuous spatial coverage. Cluster 3 was concentrated in the eastern part, particularly in eastern Liuyang and northern Liling. Cluster 4 was mainly located in the southern part of the study area, especially in the transitional zone between ecological space and urban construction land.

Overall, the SOM results indicate marked regional differentiation in ecological sensitivity, with Cluster 1 corresponding to the metropolitan core, Cluster 2 to the western and northern periphery, Cluster 3 to the ecologically rich eastern area, and Cluster 4 to the southern transitional zone.

3.2. *Classification performance of the XGBoost model*

To assess the internal separability of the SOM-derived ecological sensitivity types, six classifiers were compared: XGBoost, LightGBM, random forest, artificial neural network, logistic regression, and decision tree. XGBoost achieved the highest performance on the original class-preserving 20% held-out test set, with a macro-average AUC of 0.999 and an overall accuracy of 96.93% (3,627 of 3,742 samples; **Figure 4(b)**). Recall for Classes 0–3 was 94.55%, 97.48%, 96.31%, and 97.33% (**Figure 4(a)**), respectively. The retained configuration was colsample_bytree = 0.5, gamma = 0.0, learning_rate = 0.1714, max_depth = 8, n_estimators = 300, and subsample = 1.0.

Stratified random five-fold cross-validation yielded a mean validation accuracy of 0.9680 ± 0.0012 and a macro-average AUC of 0.9988 ± 0.0001, with a mean training–validation accuracy gap of 0.0320. In the one-at-a-time parameter sensitivity analysis, validation accuracy ranged from 0.9701 to 0.9730 and macro-average AUC from 0.9987 to 0.9989, indicating limited sensitivity to parameter changes around the retained configuration. Under the stricter spatial-region five-fold validation, mean accuracy decreased to 0.7704 ± 0.1388 and macro-average AUC to 0.9220 ± 0.0937, with a mean training–validation accuracy gap of 0.2296. These results indicate strong and stable interpolation performance for spatially mixed samples but reduced and more variable transferability to entirely held-out regions. Therefore, the validation supports the internal separability of the SOM-derived labels rather than their independent ecological validity.



(a) Matrix (b) ROC Curve

Figure 4: Model Performance Evaluation Plot.

3.3. *Global importance of influencing factors*

The SHAP global interpretation results reveal obvious differences in the contribution of the 12 indicators to ecological sensitivity type identification. According to the mean absolute SHAP values, the six most important factors are $PM_{2.5}$, annual precipitation, nighttime light index, GDP, animal diversity, and land surface temperature (**Figure 5**).

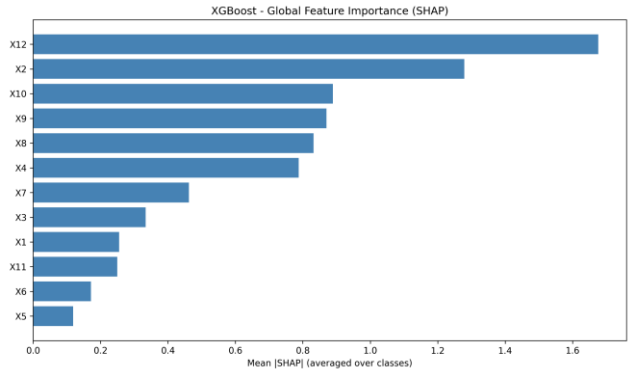


Figure 5. Importance of Ecological Sensitivity Factors in the Chang-Zhu-Tan Metropolitan Area.

$PM_{2.5}$ had the highest mean absolute SHAP value (1.68). This dominance may reflect direct pollution pressure, but it may also summarize the spatial footprint of industrial activity, traffic emissions, energy use, and dense construction. Its co-occurrence with nighttime light and land-surface temperature in central built-up areas indicates that $PM_{2.5}$ partly acts as a proxy for urbanization intensity in addition to representing atmospheric exposure. Annual precipitation ranked second (1.30), identifying water availability and local water–heat balance as major discriminators. Nighttime light (0.89) and GDP (0.87) characterized human activity and development intensity, whereas animal diversity (0.83) and land-surface temperature (0.80) represented ecological background and thermal stress. These rankings describe the fitted model and should not be interpreted as causal effect sizes.

By contrast, distance to rivers, distance to roads, slope, soil bulk density, and soil pH have relatively lower mean absolute SHAP values. This does not mean that these factors are ecologically unimportant, but rather that their independent contribution to distinguishing the four ecological sensitivity types is weaker in the current model.

Overall, the model associated ecological sensitivity differentiation with pollution exposure, hydrothermal conditions, human activity intensity, ecological background, and thermal environment. The co-occurrence of these factors highlights compound environmental association patterns in a rapidly urbanizing metropolitan area.

3.4. *Type-specific association patterns of ecological sensitivity*

SHAP analysis further revealed the basis on which the XGBoost model distinguished different ecological sensitivity clusters. The results showed that the four clusters were characterized by different combinations of dominant factors and effect directions (**Figure 6**).

For Cluster 1, high nighttime light index, high land surface temperature, and high $PM_{2.5}$ significantly increased the probability of sample membership. Combined with its concentration in the central urban areas and densely built-up zones of the Changsha–Zhuzhou–Xiangtan metropolitan area, this cluster was identified as an anthropogenic activity-associated ecological sensitive area.Cluster 2 was mainly influenced by annual precipitation, $PM_{2.5}$, GDP, nighttime light index, and land surface temperature. Among these factors, low precipitation showed a clear positive contribution to cluster identification, indicating that its sensitivity was mainly constrained by water availability and thermal environmental conditions. This cluster was therefore defined as a hydrothermal constraint-associated ecological sensitive area.

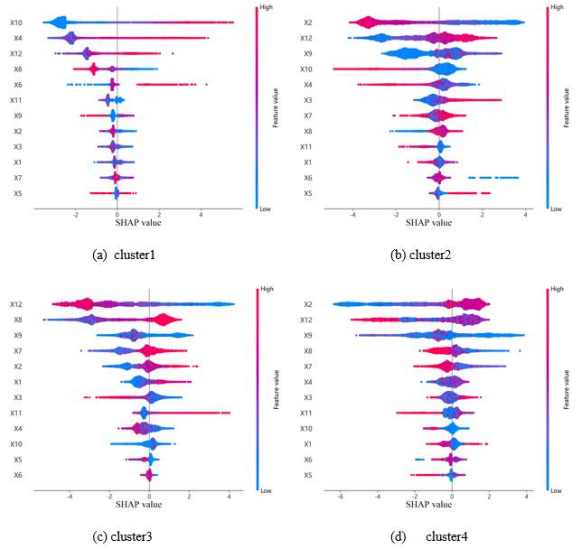


Figure 6. SHAP Analysis of Overall Characteristics of Ecological Sensitivity Factors in the Changsha-Zhuzhou-Xiangtan Metropolitan Area.

For Cluster 3, high animal diversity, high habitat quality, low GDP, and low $PM_{2.5}$ contributed positively to cluster identification, suggesting a relatively strong ecological background and weak human disturbance. Considering its distribution mainly in the ecologically resource-rich eastern part of the study area, this cluster was classified as an ecological background-associated ecological sensitive area. Cluster 4 was jointly affected by annual precipitation, $PM_{2.5}$, GDP, animal diversity, habitat quality, and land surface temperature, showing an interaction between natural ecological conditions and human activities. In combination with its location in the transitional zone between ecological space and built-up space in the southern part of the study area, it was identified as a natural-anthropogenic composite-associated ecological sensitive area.

Overall, three broad model-associated environmental patterns were identified in the Changsha–Zhuzhou–Xiangtan metropolitan area. The first was anthropogenic disturbance, represented by GDP, nighttime light index, and $PM_{2.5}$. The second was hydrothermal environmental constraint, represented by annual precipitation and land surface temperature. The third was ecological background response, represented by animal diversity and habitat quality.

3.5. *Nonlinear relationships and interaction effects among key factors*

The SHAP dependence plots indicated that ecological sensitivity in the Changsha–Zhuzhou–Xiangtan metropolitan area exhibited clear nonlinear responses and multi-factor coupling effects. $PM_{2.5}$, land surface temperature, annual precipitation, nighttime light index, GDP, animal diversity, and habitat quality were identified as the main interacting factors.

The SHAP dependence patterns differed among the ecological sensitivity types (**Figure 7**). In the anthropogenic activity-dominated sensitive area, high nighttime light index, high $PM_{2.5}$ concentration, and high land surface temperature jointly increased the probability of sample membership, indicating a model-associated co-occurrence of urban development, pollution exposure, and urban thermal conditions. In addition, the combination of high nighttime light index and low animal diversity further strengthened the identification of this type, consistent with an association between urbanization intensity, animal diversity, and habitat conditions

In the hydrothermal condition-constrained sensitive area, the interaction between annual precipitation and land surface temperature was dominant. The combination of low precipitation and high land surface temperature markedly increased the probability of membership, suggesting that insufficient water supply and thermal stress were the main limiting factors.

In the ecological background-dominated sensitive area, high animal diversity, high habitat quality, low GDP, and low $PM_{2.5}$ jointly increased the probability of sample membership, indicating that a favorable ecological background and weak anthropogenic disturbance shaped this type.

The natural–anthropogenic composite-driven sensitive area was characterized by complex nonlinear effects involving annual precipitation, $PM_{2.5}$, GDP, and animal diversity, reflecting the combined influence of natural ecological conditions and human activities.

Overall, three recurrent model-associated patterns were identified in the Changsha–Zhuzhou–Xiangtan metropolitan area. The first pattern was the co-occurrence of high

nighttime light index, high $PM_{2.5}$ concentration, and high land-surface temperature. The second pattern was the combination of low annual precipitation and high land-surface temperature. The third pattern was the co-occurrence of high animal diversity and habitat quality with low development intensity.

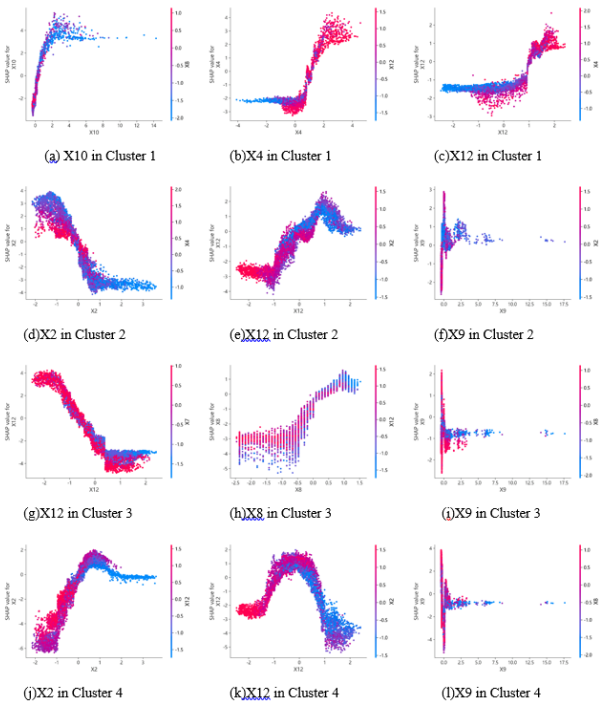


Figure 7. SHAP Feature Dependency Map of Ecological Sensitivity Factors in the Changsha-Zhuzhou-Xiangtan Metropolitan Area.

4. Discussion

4.1. *Spatial heterogeneity and associated environmental patterns*

The Changsha–Zhuzhou–Xiangtan metropolitan area exhibited clear spatial heterogeneity in ecological sensitivity, with four types associated with different combinations of anthropogenic disturbance, hydrothermal stress, ecological background, and transitional natural–human conditions. The types do not represent a simple low-to-high gradient. Rather, they distinguish contrasting multivariate contexts in which sensitivity may occur. This interpretation is deliberately associational because the cross-sectional SHAP analysis cannot establish causality.

The model associated overall spatial differentiation with natural conditions, ecological background, pollution exposure, and human activity intensity. This pattern is consistent with vulnerability frameworks that emphasize coupled exposure and sensitivity, and with regional studies reporting strong natural–human interactions in rapidly urbanizing Chinese metropolitan areas (Turner *et al.*, 2003; Adger, 2006; Xu *et al.*, 2021; Li *et al.*, 2022; Chen *et al.*, 2023; Wang *et al.*, 2023). Unlike weighted-overlay assessments that produce a single ordinal surface, the present type-based approach identifies multiple environmental configurations without prespecified indicator weights. Its advantage is the ability to represent nonlinear associations and type-specific combinations; its limitation is that data-driven classes remain sensitive to feature selection, spatial resolution, and algorithm choice.

In highly urbanized areas, the anthropogenic activity-associated type coincided with concentrated nighttime light, GDP, $PM_{2.5}$, and land-surface temperature. $PM_{2.5}$ plausibly reflects both atmospheric pollution pressure and the broader urban-industrial footprint because traffic, industrial energy use, dense construction, and limited dispersion tend to co-occur. Nighttime light independently represents the intensity and continuity of human activity, while land-surface temperature captures the thermal expression of impervious cover and reduced evapotranspiration. Their joint SHAP pattern is therefore consistent with compound urban pressure, but it cannot identify which component is causal (Levin *et al.*, 2020; Manoli *et al.*, 2019; Hammer *et al.*, 2020; Liu *et al.*, 2020).

The hydrothermal constraint-associated type was distinguished primarily by annual precipitation and land-surface temperature. Even in a humid subtropical region, spatially uneven precipitation, high evaporative demand, impervious surfaces, and fragmented vegetation can create local water–heat mismatches. Lower precipitation combined with higher surface temperature may therefore indicate reduced moisture buffering and vegetation recovery capacity. This interpretation is consistent with the importance of hydrothermal gradients reported in regional sensitivity studies, while remaining a plausible ecological explanation rather than a causal test.

The ecological background-associated type shows that ecological sensitivity should not be equated with poor environmental quality. Areas with high animal diversity, high habitat quality, low development intensity, and low pollution exposure may still be highly sensitive. This is because they have high conservation value and support important regional ecological functions. This finding is consistent with studies emphasizing the ecological importance of animal-diversity-rich habitats, ecological source areas, and low-disturbance landscapes. Urban expansion can generate substantial impacts on biodiversity and conservation-priority areas, particularly where development pressures approach ecologically valuable spaces (Seto *et al.*, 2012; Simkin *et al.*, 2022). Area-based conservation and ecological network protection therefore remain important for maintaining habitat connectivity, animal diversity, and ecosystem functions (Maxwell *et al.*, 2020).

The natural–anthropogenic composite-associated type represents a transitional association pattern. In this type, natural ecological conditions and anthropogenic disturbance coexist. This type showed lower construction intensity than the anthropogenic activity-associated type. However, its ecological processes were more susceptible to industrial activities, transport construction, and urban expansion than those of the ecological background-associated type. Transitional areas may not show the strongest single-factor pressure. Nevertheless, they can be sensitive because moderate human disturbance interacts with ecological background conditions. Such areas are important coordination zones between ecological conservation and moderate development.

The SHAP dependence analysis indicated nonlinear co-occurrence among nighttime light, $PM_{2.5}$, land-surface temperature, precipitation, animal diversity, and habitat quality. High nighttime light, $PM_{2.5}$, and land-surface temperature jointly characterized the urban type; low precipitation and high temperature characterized the hydrothermal type; and high animal diversity and habitat quality with low disturbance characterized the ecological-background type. SHAP interaction patterns describe the model response surface and do not prove synergistic ecological effects. Independent monitoring or quasi-experimental designs would be needed to test the proposed processes.

4.2. *Methodological implications*

Methodologically, SOM–XGBoost–SHAP complements conventional analytic hierarchy process, entropy weighting, and GIS weighted-overlay assessments. SOM identifies recurring multivariate types without prespecified sensitivity grades; XGBoost tests whether those types are separable; and SHAP reveals the variables associated with model decisions. This framework can capture nonlinear and type-specific patterns that linear aggregation may obscure. However, it does not remove subjectivity from indicator selection, and high classifier accuracy is not external validation. The method is therefore best used alongside field evidence, ecological redlines, conservation zoning, or independent ecosystem-condition datasets.

In this study, SOM identified ecological sensitivity types from multidimensional indicator similarities. XGBoost tested the separability of these types. SHAP quantified factor importance, contribution direction, and interaction effects. This integrated procedure links pattern identification with association interpretation. This link is particularly important for ecological sensitivity assessment in complex metropolitan areas, where natural conditions and human activities interact. Rather than replacing traditional assessment methods, this integrated approach complements them. It provides additional insights into nonlinear factor effects, multi-factor interactions, and the formation patterns of different ecological sensitivity types (Kohonen, 1982; Vesanto and Alhoniemi, 2000; Chen and Guestrin, 2016; Lundberg and Lee, 2017; Lundberg *et al.*, 2020; Meyer *et al.*, 2019; Molnar, 2022).

4.3. *Governance implications*

The differentiated association patterns have implications for ecological governance and territorial spatial planning. Management should shift from uniform sensitivity-level control toward type-specific measures while recognizing that the mapped boundaries are model-dependent. The recommendations below link each type to its principal location, associated pressures, and planning instruments.

Table 4. Type-specific and location-oriented ecological management strategies.

Type and principal location	Model-associated environmental pattern	Planning priority	Location-oriented measures
Anthropogenic activity-associated; central built-up areas of Changsha, Zhuzhou, and Xiangtan	High nighttime light, PM2.5, and land-surface temperature	Reduce compound pollution and heat exposure	Enforce urban growth boundaries; retrofit industrial and transport emission sources; protect ventilation corridors; expand connected blue-green infrastructure and cool-surface interventions.
Hydrothermal constraint-associated; western and northern peripheral areas	Low precipitation combined with high thermal stress	Strengthen water retention and drought-heat resilience	Restore riparian buffers and headwater vegetation; protect small wetlands and retention spaces; apply soil-water conservation and drought-resilient native planting.
Ecological background-associated; eastern Liuyang and northern Liling	High animal diversity and habitat quality with low disturbance	Prevent fragmentation of high-value habitat	Align strict protection with ecological redlines where applicable; maintain source-area and corridor connectivity; restrict new roads, mining, and high-intensity tourism in sensitive habitat.
Natural-anthropogenic composite-associated; southern ecological-urban transition zone	Moderate development pressure interacting with ecological background	Coordinate conservation and controlled development	Establish graded buffers around ecological sources and corridors; control construction-land expansion; require cumulative ecological-risk assessment for major infrastructure; restore corridor pinch points.

In anthropogenic activity-associated areas, management should focus on reducing development intensity, pollution exposure, and urban thermal stress. Compact land-use management, industrial emission control, urban ventilation corridor optimization, and blue–green infrastructure restoration may reduce compound pressures. These pressures include construction expansion, $PM_{2.5}$ exposure, and thermal stress.

Where hydrothermal constraints dominate, improving water retention capacity, vegetation resilience, and watershed conservation is important for reducing hydrothermal stress. Soil and water conservation, degraded vegetation restoration, and landscape moisture regulation may help enhance ecosystem stability under local water–heat imbalance.

Ecological background-associated areas require strict ecological protection, animal diversity conservation, habitat connectivity, and ecological corridor maintenance. Although these areas may have relatively good environmental quality, their high ecological value and sensitivity to disturbance require careful protection.

In natural–anthropogenic composite-associated areas, governance should balance ecological conservation and moderate development. This balance can be supported through ecological buffer zones, construction land boundary control, and ecological risk prevention. Overall, linking ecological sensitivity types with their dominant association patterns can support more targeted and adaptive planning policies. It can also help improve spatial differentiation in ecological governance (Frantzeskaki *et al.*, 2019; Liquete *et al.*, 2015; Meerow and Newell, 2017).

4.4. *Limitations and future research*

Several limitations should be acknowledged. First, the analysis used cross-sectional data and therefore cannot establish temporal ordering or causality; SHAP values explain fitted-model associations rather than ecological effects. Second, the 1 km grid may smooth fine-scale pollution, microclimate, and habitat variation. Third, including standardized coordinates improves spatial continuity but may strengthen geographic clustering. This limitation was evident in the spatial-region validation: accuracy and macro-average AUC were lower and more variable than under stratified random validation, indicating restricted transfer to entirely held-out regions. Future work should use multi-temporal monitoring, field observations, external ecological datasets, and sensitivity tests that omit coordinates or vary their weight.

Another limitation concerns the spatial resolution, statistical consistency, and retrieval accuracy of several indicators. $PM_{2.5}$ concentration, land surface temperature, animal diversity, and habitat quality may vary substantially at local scales. Available datasets may not fully capture these fine-scale differences. Gridded, remotely sensed, or modelled datasets may differ from ground-based observations. They may also fail to fully capture local pollution exposure, microclimatic conditions, or habitat characteristics. Animal diversity and habitat quality indicators may involve modelling assumptions or proxy variables. These factors can introduce uncertainty into the interpretation of ecological background conditions. Future research should integrate higher-resolution pollution emission data, ground-based monitoring data, refined human activity data, and field-based ecological observations to improve local-scale assessment accuracy.

Cross-algorithm comparison provided a partial robustness check: K-means with 20 independent initializations on the same standardized 14-dimensional feature space produced ARI = 0.3245 and NMI = 0.4270 relative to the SOM labels. This moderate agreement indicates that broad structure was partly reproducible but that detailed boundaries depended on the clustering algorithm. The result supports cautious use of the four-type framework and reinforces the need for external comparison with ecological redlines, protected areas, restoration zones, or field-based ecosystem-condition observations.

This study identified ecological sensitivity types and explained their association patterns. However, it did not construct a type-specific comprehensive evaluation index. Although the analytical procedure can reveal spatial types and associated factors, a comprehensive index system may still be needed for practical planning and policy implementation. Future studies could combine machine learning interpretation with ecological risk assessment, ecosystem service evaluation, and land-use scenario simulation. This would help develop differentiated assessment and governance frameworks. Scenario-based analysis considering climate change, urban expansion, and ecological restoration policies would also help evaluate how ecological sensitivity types may change under different planning and management pathways.

Finally, this study focused on a humid subtropical metropolitan area in central China. Therefore, the identified ecological sensitivity types and association patterns may not be directly transferable to regions with different climatic, geomorphological, ecological, or socio-economic contexts. The study boundary may also not fully capture cross-regional ecological processes, such as watershed connectivity, pollutant transport, and species movement. Comparative studies across metropolitan areas with different development stages and environmental backgrounds are needed to evaluate the generality and transferability of the proposed approach. These limitations qualify the interpretation of the results, but they do not diminish the value of the proposed approach for identifying and explaining ecological sensitivity types in the study area.

5. Conclusions

Ecological sensitivity in the Changsha–Zhuzhou–Xiangtan metropolitan area was spatially heterogeneous and associated with coupled natural and anthropogenic conditions. SOM, XGBoost, and SHAP identified four types: anthropogenic activity-associated, hydrothermal constraint-associated, ecological background-associated, and natural–anthropogenic composite-associated. These types represent contrasting multivariate contexts rather than causal mechanisms or a single low-to-high sensitivity gradient.

The dominant model-associated factors differed among the four types. $PM_{2.5}$, annual precipitation, nighttime light, GDP, animal diversity, and land-surface temperature were the strongest discriminators. Their nonlinear SHAP patterns suggest plausible links to pollution and urbanization intensity, local water-heat balance, and ecological background, but causal interpretation requires independent temporal or experimental evidence.

The integrated SOM–XGBoost–SHAP approach complements conventional level-based ecological sensitivity assessment by linking type identification, internal separability testing, and interpretable association analysis. Its planning value lies in supporting location-specific actions for central urban cores, western and northern hydrothermal-constrained areas, eastern high-

value ecological areas, and southern transition zones. Moderate K-means agreement also shows that results should be used with independent conservation and monitoring evidence rather than as stand-alone regulatory boundaries.

References

Adger W.N. (2006), Vulnerability, *Global Environmental Change*, **16**(3), 268–281.

Chen G., Li X., Liu X., Chen Y., Liang X., Leng J., Xu X., Liao W., Qiu Y., Wu Q. and Huang K. (2020), Global projections of future urban land expansion under shared socioeconomic pathways, *Nature Communications*, **11**, 537.

Chen H., Wang S., Wang R., Guo L., Deng X. and Yang Y. (2023), Spatial heterogeneity and driving mechanisms of ecological sensitivity in urban agglomerations: Insights from machine learning and SHAP interpretation, *Journal of Environmental Management*, **325**, 116558.

Chen T. and Guestrin C. (2016), XGBoost: a scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Eigenbrod F., Bell V.A., Davies H.N., Heinemeyer A., Armsworth P.R. and Gaston K.J. (2011), The impact of projected increases in urbanization on ecosystem services, *Proceedings of the Royal Society B: Biological Sciences*, **278**(1722), 3201–3208.

Elmqvist T., Andersson E., Frantzeskaki N., McPhearson T., Olsson P., Gaffney O., Takeuchi K. and Folke C. (2019), Sustainability and resilience for transformation in the urban century, *Nature Sustainability*, **2**, 267–273.

Elvidge C.D., Baugh K.E., Kihn E.A., Kroehl H.W. and Davis E.R. (1997), Mapping city lights with nighttime data from the DMSP Operational Linescan System, *Photogrammetric Engineering and Remote Sensing*, **63**(6), 727–734.

Estoque R.C., Ooba M., Seposo X.T., Togawa T., Hijioka Y., Takahashi K. and Nakamura S. (2021), Heat health risk assessment in Philippine cities using remotely sensed data and social-ecological indicators, *Nature Communications*, **12**, 1–12.

Fang C. and Yu D. (2017), Urban agglomeration: an evolving concept of an emerging phenomenon, *Landscape and Urban Planning*, **162**, 126–136.

Frantzeskaki N., McPhearson T., Collier M.J., Kendal D., Bulkeley H., Dumitru A., Walsh C., Noble K., van Wyk E., Ordóñez C., Oke C. and Pintér L. (2019), Nature-based solutions for urban climate change adaptation: Linking science, policy, and practice communities for evidence-based decision-making, *BioScience*, **69**(6), 455–466.

Hammer M.S., van Donkelaar A., Li C., Lyapustin A., Sayer A.M., Hsu N.C., Levy R.C., Garay M.J., Kalashnikova O.V., Kahn R.A., Brauer M., Apte J.S., Henze D.K., Zhang L., Zhang Q., Ford B., Pierce J.R. and Martin R.V. (2020), Global estimates and long-term trends of fine particulate matter concentrations, 1998–2018, *Environmental Science & Technology*, **54**(13), 7879–7890.

Hengl T., Nussbaum M., Wright M.N., Heuvelink G.B.M. and Gräler B. (2018), Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, *PeerJ*, **6**, e5518.

Huang K., Li X., Liu X. and Seto K.C. (2019), Projecting global urban land expansion and heat island intensification through 2050, *Environmental Research Letters*, **14**(11), 114037.

Kim S.H., Lee H.J. and Kim D.K. (2023), Classifying habitat characteristics of wetlands using a self-organizing map, *Ecological Informatics*, **74**, 102048.

Kohonen T. (1982), Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, **43**, 59–69.

Kohonen T. (2001), Self-organizing maps, *Springer Series in Information Sciences*, **30**, 1–501.

Langemeyer J., Wedgwood D., McPhearson T., Baró F., Madsen A.L. and Barton D.N. (2020), Creating urban green infrastructure where it is needed: A spatial ecosystem service-based decision analysis of green roofs in Barcelona, *Science of the Total Environment*, **707**, 135487.

Levin N. and Duke Y. (2012), High spatial resolution night-time light images for demographic and socio-economic studies, *Remote Sensing of Environment*, **119**, 1–10.

Levin N., Kyba C.C.M., Zhang Q., Sánchez de Miguel A., Román M.O., Li X., Portnov B.A., Molthan A.L., Jechow A., Miller S.D., Wang Z., Shrestha R.M. and Elvidge C.D. (2020), Remote sensing of night lights: A review and an outlook for the future, *Remote Sensing of Environment*, **237**, 111443.

Lei X. and Zhang Z. (2025), The algorithm cannot see everything: Human capital as the hidden variable in environmental technology success, *Journal of Environmental Management*, **397**, 128287.

Lei X., Jin X., Yue S. and Tufail M. (2026a), Can large language models detect green leaders? CEO environmental backgrounds and corporate ESG performance, *Technology in Society*, **88**, 103479.

Lei X., Khattak S.I., Qazi A. and Al-Mhdawi M.K.S. (2026b), When financial analysts become climate actors: How information networks shape corporate green innovation, *Sustainable Development*, advance online publication.

Li X., Liu Y., Zhang Z., Peng J. and Yang Q. (2022), Spatiotemporal dynamics of ecological sensitivity under rapid urbanization: Evidence from the Yangtze River Delta, China, *Ecological Indicators*, **134**, 108485.

Liquete C., Kleeschulte S., Dige G., Maes J., Grizzetti B., Olah B. and Zulian G. (2015), Mapping green infrastructure based on ecosystem services and ecological networks: A Pan-European case study, *Environmental Science & Policy*, **54**, 268–280.

Liu J., Dietz T., Carpenter S.R., Alberti M., Folke C., Moran E., Pell A.N., Deadman P., Kratz T., Lubchenco J., Ostrom E., Ouyang Z., Provencher W., Redman C.L., Schneider S.H. and Taylor W.W. (2007), Complexity of coupled human and natural systems, *Science*, **317**(5844), 1513–1516.

Liu X., Huang Y., Xu X., Li X., Li X., Ciais P., Lin P., Gong K., Ziegler A.D., Chen A., Gong P., Chen J., Hu G., Chen Y., Wang S., Wu Q., Huang K., Estes L. and Zeng Z. (2020), High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015, *Nature Sustainability*, **3**, 564–570.

Luo Y., Lei X., Zhang W., Zheng H., Xu X. and Xu J. (2026), Official or influencer? An AI-enhanced analytical framework for decoding multimodal persuasion in government marketing videos, *Information Processing & Management*, **64**(1), 105068.

Lundberg S.M. and Lee S.I. (2017), A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, **30**, 4765–4774.

Lundberg S.M., Erion G., Chen H., DeGrave A., Prutkin J.M., Nair B., Katz R., Himmelfarb J., Bansal N. and Lee S.I. (2020), From local explanations to global understanding with explainable AI for trees, *Nature Machine Intelligence*, **2**(1), 56–67.

Malczewski J. (2006), GIS-based multicriteria decision analysis: a survey of the literature, *International Journal of Geographical Information Science*, **20**(7), 703–726.

Manoli G., Fatichi S., Schläpfer M., Yu K., Crowther T.W., Meili N., Burlando P., Katul G.G. and Bou-Zeid E. (2019), Magnitude of urban heat islands largely explained by climate and population, *Nature*, **573**, 55–60.

Maxwell S.L., Cazalis V., Dudley N., Hoffmann M., Rodrigues A.S.L., Stolton S., Visconti P., Woodley S., Kingston N., Lewis E., Maron M., Strassburg B.B.N., Wenger A., Jonas H.D., Venter O. and Watson J.E.M. (2020), Area-based conservation in the twenty-first century, *Nature*, **586**, 217–227.

McDonald R.I., Mansur A.V., Ascensão F., Crossman K., Elmqvist T., Gonzalez A., Güneralp B., Haase D., Hamann M., Hillel O., Huang K., Kahnt B., Maddox D., Pacheco A., Pereira H.M., Seto K.C., Simkin R., Walsh B., Werner A.S. and Ziter C. (2020), Research gaps in knowledge of the impact of urban growth on biodiversity, *Nature Sustainability*, **3**, 16–24.

Meerow S. (2020), The politics of multifunctional green infrastructure planning in New York City, *Cities*, **100**, 102621.

Meerow S. and Newell J.P. (2017), Spatial planning for multifunctional green infrastructure: Growing resilience in Detroit, *Landscape and Urban Planning*, **159**, 62–75.

Meyer H., Reudenbach C., Wöllauer S. and Nauss T. (2019), Importance of spatial predictor variable selection in machine learning applications: moving from data reproduction to spatial prediction, *Ecological Modelling*, **411**, 108815.

Molnar C. (2022), *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed., independently published.

O'Brien R.M. (2007), A caution regarding rules of thumb for variance inflation factors, *Quality & Quantity*, **41**, 673–690.

Olden J.D., Lawler J.J. and Poff N.L. (2008), Machine learning methods without tears: a primer for ecologists, *The Quarterly Review of Biology*, **83**(2), 171–193.

Peng J., Pan Y., Liu Y., Zhao H. and Wang Y. (2018a), Linking ecological degradation risk to identify ecological security patterns in a rapidly urbanizing landscape, *Habitat International*, **71**, 110–124.

Peng J., Yang Y., Liu Y., Hu Y., Du Y., Meersmans J. and Qiu S. (2018b), Linking ecosystem services and circuit theory to identify ecological security patterns, *Science of the Total Environment*, **644**, 781–790.

Ryo M. and Rillig M.C. (2017), Statistically reinforced machine learning for nonlinear patterns and variable interactions, *Ecosphere*, **8**, e01976.

Seto K.C., Güneralp B. and Hutyra L.R. (2012), Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools, *Proceedings of the National Academy of Sciences of the United States of America*, **109**(40), 16083–16088.

Simkin R.D., Seto K.C., McDonald R.I. and Jetz W. (2022), Biodiversity impacts and conservation implications of urban land expansion projected to 2050, *Proceedings of the National Academy of Sciences of the United States of America*, **119**(12), e2117297119.

Turner B.L. II, Kasperson R.E., Matson P.A., McCarthy J.J., Corell R.W., Christensen L., Eckley N., Kasperson J.X., Luers A., Martello M.L., Polsky C., Pulsipher A. and Schiller A. (2003), A framework for vulnerability analysis in sustainability science, *Proceedings of the National Academy of Sciences of the United States of America*, **100**(14), 8074–8079.

van Donkelaar A., Martin R.V., Brauer M. and Boys B.L. (2015), Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter, *Environmental Health Perspectives*, **123**(2), 135–143.

Vesanto J. and Alhoniemi E. (2000), Clustering of the self-organizing map, *IEEE Transactions on Neural Networks*, **11**(3), 586–600.

Wang H., Zhang J., Li S., Yang Q. and Li X. (2023), Spatiotemporal heterogeneity and driving mechanisms of ecological risk in rapidly urbanizing metropolitan regions: a case study of the Changsha–Zhuzhou–Xiangtan metropolitan area, *Ecological Indicators*, **154**, 110712.

Wang X., Peng W. and Xiong H. (2024), Spatial-temporal evolution and driving factors of rural resilience in the urban agglomerations in the middle reaches of the Yangtze River, China, *Environmental Science and Pollution Research*, **31**(17), 25076–25095.

Wu J. (2014), Urban ecology and sustainability: the state-of-the-science and future directions, *Landscape and Urban Planning*, **125**, 209–221.

Xu Y., Wang Y., Liu Y., Wang J., Wang L. and Peng J. (2021), Integrating natural and anthropogenic driving forces for ecological sensitivity assessment in rapidly urbanizing regions: A case study of the Guangdong–Hong Kong–Macao Greater Bay Area, *Science of the Total Environment*, **789**, 147961.

Yin J., Lei X. and Xu J. (2026), The impact of corporate biodiversity attention on sustainable growth: Mediating roles of ESG performance, corporate reputation, and green transformation, *Sustainable Development*, advance online publication.

Zhang Y., Liu X., Chen D., Liu Z., Li H. and Yang Y. (2022), Identification and optimization of ecological security patterns in the Yangtze River Middle Reaches Urban Agglomeration: implications for territorial spatial planning, *Science of the Total Environment*, **815**, 152855.

Zhou K., Yin Y., Li H. and Shen Y. (2021), Driving factors and spatiotemporal effects of environmental stress in urban agglomeration: Evidence from the Beijing-Tianjin-Hebei region of China, *Journal of Geographical Sciences*, **31**(1), 91–110.

Zurell D., Franklin J., König C., Bouchet P.J., Dormann C.F., Elith J., Fandos G., Feng X., Guillera-Arroita G., Guisan A., Lahoz-Monfort J.J., Leitão P.J., Park D.S., Peterson A.T., Rapacciuolo G., Schmatz D.R., Schröder B., Serra-Diaz J.M., Thuiller W., Yates K.L., Zimmermann N.E. and Merow C. (2020), A standard protocol for reporting species distribution models, *Ecography*, **43**, 1261–1277.