

Assessing urbanisation and ecological integrity coupling in Malaysia using interpretable machine learning

Received: 26 February 2026

Accepted: 20 April 2026

Published online: 12 May 2026

Cite this article as: Shi Q., Yusoff M.M., Noor N.M. *et al.* Assessing urbanisation and ecological integrity coupling in Malaysia using interpretable machine learning. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-50326-3>

Qinyu Shi, Mariney Mohd Yusoff, Nisfariza Mohd Noor, Jinyu Zhang, Xiaoya Li, Zhichao Wang, Ting Guo & Peiyuan Bai

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Assessing urbanisation and ecological integrity coupling in Malaysia using interpretable machine learning

Authors

Qinyu Shi^a,

e-mail: s2184461@siswa.um.edu.my

ORCID: <https://orcid.org/0009-0008-7629-112X>

Mariney Mohd Yusoff^{a,*},

Nisfariza Mohd Noor^a,

e-mail: nish@um.edu.my

ORCID: <https://orcid.org/0000-0001-6303-9509>

Jinyu Zhang^a,

e-mail: s2153494@siswa.um.edu.my

ORCID: <https://orcid.org/0009-0005-9258-4289>

Xiaoya Li^a,

e-mail: s2128614@siswa.um.edu.my

ORCID: <https://orcid.org/0000-0002-8712-0392>

Zhichao Wang^a,

e-mail: cchaoswang@gmail.com

ORCID: <https://orcid.org/0000-0002-0865-6544>

Ting Guo^b,

e-mail: 30420220156680@stu.xmu.edu.cn

Peiyuan Bai^c

e-mail: baipeiyuan@pku.edu.cn

Author Affiliation:

^a Department of Geography, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

^b a Key Laboratory of Ministry of Education for Coastal Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Fujian, 361102, China

^c College of Urban and Environmental Sciences, Peking University, Beijing, 100080, China

***Corresponding Author:**

Mariney Mohd Yusoff

email: mariney@um.edu.my

ORCID: <https://orcid.org/0000-0003-2758-4715>

Declarations

Competing interests

The authors of this article declare no potential competing interests with other people or organizations.

Data available statement

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

No funding was obtained for this study.

Authors' contributions

Q.S. led the study and prepared the full manuscript, including data collection, data analysis, figure and table preparation, and drafting and revising the paper. J.Z. and X.L. supported the data analysis and interpretation of results. Z.W. assisted with data preparation and visualization. T.G. and B.P. contributed to the study design and provided revisions to improve the manuscript. M.M.Y. and N.M.N. supervised the research and provided overall guidance and editorial input.

Abstract

Urbanisation-environment interactions are increasingly non-linear, yet traditional Coupling Coordination Degree (CCD) frameworks often rely on static, linear assumptions that fail to capture complex feedback loops. This study proposes an integrated framework combining CRITIC-weighted CCD assessment with interpretable machine learning (Random Forest and XGBoost) to decode the co-evolution of urbanisation and ecological integrity. Taking Malaysia—a country characterized by rapid expansion and high spatial disparity—as a critical case study, we utilized multi-source remote sensing and longitudinal statistics across 16 states. Our XGBoost model (Test $R^2 \approx 0.87$) outperformed traditional regressions, confirming significant non-linear and dynamic coupling effects. By applying SHAP (SHapley Additive exPlanations), we move beyond mere prediction to quantify the "tipping points" of sustainability. Key findings reveal: (1) a distinct spatial decoupling in northern and east-coast regions; (2) a "hump-shaped" threshold effect for built-up expansion where coordination peaks before declining; and (3) the critical role of forest/water assets as non-linear ecological buffers. These results demonstrate that sustainable transitions depend more on spatial structure and asset management than on income growth alone. This interpretable AI-CCD framework provides a scalable, evidence-based toolkit for low- and middle-income countries to navigate the trade-offs between development and ecological preservation.

Keywords: Coupling Coordination Degree (CCD); Interpretable Machine Learning; SHAP; Urbanisation; Ecological Environment; Malaysia

1. Introduction

Rapid urbanisation has become a defining feature of contemporary development, reshaping economic structures, settlement patterns and environmental conditions [1]. In many low- and middle-income countries, cities function as engines of growth, concentrating investment, infrastructure and employment. United Nations projections suggest that almost all future urban growth will occur in these countries, with their urban population expected to increase by about 75% between 2015 and 2030, and nearly 90% of the additional 2.5 billion urban residents anticipated by 2050 concentrated in Asia and Africa [2]. At the same time, the international policy agenda has placed cities at the centre of sustainable development. The 2030 Agenda highlights the need to reconcile economic growth, social inclusion and environmental protection within rapidly urbanising regions [3]. However, urban expansion has outpaced the provision of basic services, environmental infrastructure and effective land use control. This has resulted in rising greenhouse gas emissions, deteriorating air and water quality, and the loss of green and blue spaces, which together erode urban liveability and resilience [4][5][6]. Understanding how urbanisation can be steered towards more sustainable and climate responsive pathways is therefore an urgent analytical and policy task.

The relationship between urbanisation and the ecological environment is neither linear nor uniform across space. In some contexts, early stages of urban growth are associated with rapid environmental degradation, followed by partial improvement as income rises, technology upgrades and environmental regulations strengthen, resembling an Environmental Kuznets Curve pattern [7]-[8]. In others, density, infrastructure form and land use configuration interact in ways that amplify or dampen environmental pressures. The same increase in population or built-up area can have very different ecological consequences depending on the quality of planning, the energy mix, the protection of green infrastructure and the underlying biophysical setting. These nonlinear responses and feedbacks mean that simple linear models or single indicators are often inadequate to capture how urban development and environmental quality co-evolve [9]-[12].

Coupling coordination degree (CCD) models have become one of the most widely used tools for evaluating the interaction between urbanisation and the ecological environment and other multi-system couplings [13]-[16]. By aggregating multiple indicators into

composite indices and feeding them into a standard coupling-coordination formula, these studies classify regions into stages ranging from severe disorder to high coordination and trace spatio-temporal evolution in coordination levels. This framework has generated useful insights in many settings. However, most CCD applications still construct subsystem indices through linear weighting and apply fixed, pre-specified functional forms for the coupling and coordination calculation, which can obscure nonlinear interactions and feedbacks among indicators [17]-[18]. In addition, CCD outputs are often interpreted descriptively, while the underlying driving mechanisms are examined only with relatively simple techniques such as correlation analysis, regression or Geodetector, which provide limited information on threshold effects, interactions and spatially varying influences [19]-[21].

Recent advances in machine learning offer a way to address this limitation. Tree-based ensemble methods, together with interpretability techniques such as feature importance and Shapley Additive Explanations (SHAP), are adept at modelling the nonlinearities and high-dimensional interaction effects that govern complex system dynamics, while still providing policy-relevant insights into the contribution of individual drivers. Empirical studies have demonstrated these advantages. For instance, Hong et al [22] used a random forest and SHAP to reveal threshold and interaction effects between urban form, meteorological variables and urban heat island intensity. Similarly, Chen et al [23] applied a geographically weighted random forest combined with SHAP to show how built-environment and environmental factors influence the green view index in complex and spatially varying ways. These successful applications underscore the potential of integrating machine learning to capture the co-evolution of urban development and environmental quality more comprehensively than traditional linear models. Even so, applications that systematically integrate CCD analysis with interpretable machine learning to analyse nonlinear and spatially heterogeneous drivers remain scarce, particularly in the context of urbanisation-ecological environment coordination in low- and middle-income countries.

As a rapidly urbanising upper-middle-income country, Malaysia now has roughly 77 % of its population living in urban areas, and urbanisation rose from around 66 % in 2005 to over 77 % by 2025 — illustrating the speed and scale of urban transformation [24]-[25]. Urban expansion in major economic corridors such as Kuala Lumpur, Penang and Johor has coincided with widespread land-use change:

between 2000 and 2020 the country lost about 1.1 million hectares of tree cover ($\approx 3.8\%$ of forest area). At the same time, in Kuala Lumpur green cover declined by nearly 3% between 2001 and 2013. Similar green-space losses have been documented in other major cities including Penang and Johor Bahru. In several major Malaysian cities, substantial tree-cover losses were recorded between 2000 and 2012 [26]-[27]. Rapid land conversion, rising emissions and declining green cover have raised concerns about the sustainability of Malaysia's urban future. At the same time, there is marked heterogeneity among states in terms of development trajectories, environmental conditions and institutional capacity. Some states have diversified economies and relatively advanced environmental governance, while others remain more dependent on primary sectors or face constraints in planning and enforcement [28]. This diversity creates a useful natural laboratory for investigating how the coordination between urbanisation and the ecological environment varies within a single national context.

To address the methodological and regional gaps identified above, this study integrates coupling coordination assessment with interpretable machine-learning techniques to enhance the understanding of urban sustainability dynamics in Malaysia. It contributes to the literature in four ways. First, it provides the first systematic state-level evaluation of the coordination between urbanisation and the ecological environment in Malaysia. Second, it combines the CCD framework with machine-learning models (Random Forest and XGBoost) and SHAP interpretation to uncover nonlinear relationships, threshold effects and state-specific drivers that are not easily detectable using traditional linear methods. Third, it examines the interactions between multiple dimensions of urbanisation (demographic, economic and land-based) and ecological indicators within an integrated analytical framework. Fourth, by identifying the relative importance of drivers across contrasting development contexts, the study offers insights that can inform locally responsive sustainability strategies in Malaysia, while also providing a transparent and transferable analytical framework that may be applied in other rapidly urbanising regions. Collectively, these contributions strengthen the empirical foundation for promoting coordinated and sustainable urban development.

Therefore, this study aims to address the following research questions:

- RQ1: How did the coupling coordination degree (CCD) between urbanisation and the eco-environment change across Malaysia?

- RQ2: To what extent can machine learning models explain and predict variation in CCD?
- RQ3: Do key urbanisation and eco-environment indicators exhibit nonlinear or threshold-like relationships with CCD?

2. Methodology

2.1 Study area

Malaysia is located in Southeast Asia and consists of two geographically separated regions: Peninsular Malaysia and East Malaysia (Figure 1). Peninsular Malaysia lies on the Malay Peninsula and contains eleven states and two federal territories (Kuala Lumpur and Putrajaya). East Malaysia, situated on the island of Borneo, comprises the states of Sabah and Sarawak along with the federal territory of Labuan. The country spans a land area of approximately 330,000 square kilometres, with coastal plains and urban corridors in the west and more mountainous, forested landscapes in the east [29].

Peninsular Malaysia is more urbanized and economically concentrated, with major metropolitan centres such as Kuala Lumpur, Selangor, and Pulau Pinang forming the core of national development. East Malaysia features lower population density, larger proportions of natural forest, and greater ecological variability. These regional contrasts create distinct patterns of land use, urban growth, and environmental conditions across the country. For the purpose of this study, all thirteen states and three federal territories are included to capture the full range of spatial heterogeneity in Malaysia's urban and ecological systems.

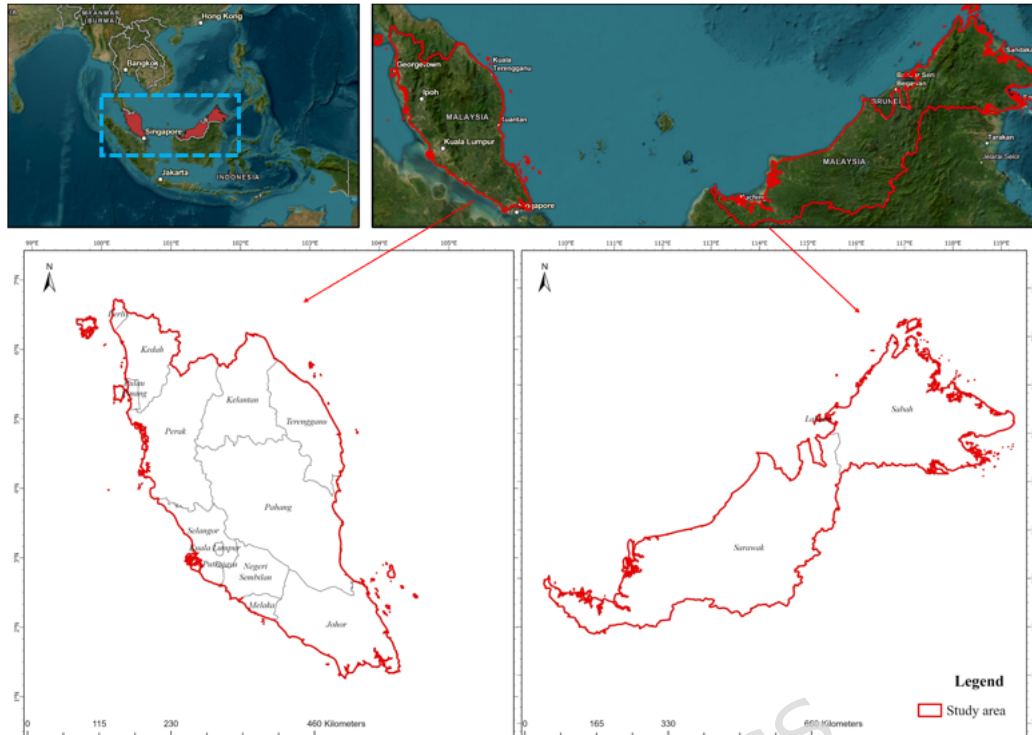


Figure 1. Location of Malaysia. The map was created by the authors using ArcGIS Pro version 3.0.2 (Esri, Redlands, CA, USA; <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>).

2.2 Indicator system and data sources

The indicator system for this study was developed to capture the core mechanisms through which urbanisation both drives and interacts with environmental sustainability across Malaysian states. Two subsystems were constructed: the urbanisation subsystem (U), which reflects demographic, economic, and land-use dynamics, and the ecological environment subsystem (E), which represents environmental pressures, ecosystem structure, and biophysical status. The complete set comprises nine indicators in total (Table 1).

The urbanisation subsystem captures three structural dimensions essential to sustainable urban development. Demographic urbanisation is represented by population density, a key indicator of settlement intensity and the associated demands on infrastructure, public services and resource use. Economic urbanisation is measured through GDP per capita, which reflects the level of economic productivity and the financial capacity to support environmental management, while also signalling consumption levels and associated emissions. Land urbanisation is quantified using the built-up area ratio, a direct measure of physical land conversion that influences hydrological regulation, ecosystem

fragmentation and urban thermal conditions. Together, these indicators characterize the complementary roles of population, economic activity, and physical expansion in shaping Malaysia's urban transformation.

The ecological environment subsystem incorporates six indicators that represent environmental pressures, ecosystem status and climatic stressors particularly relevant in rapidly urbanising tropical regions. CO₂ emission intensity indicates atmospheric pressure arising from industrial activity and energy use, whereas PM_{2.5} concentration reflects exposure to air pollution, with direct consequences for human health and environmental quality. Mean Normalised Difference Vegetation Index (NDVI) serves as a proxy for vegetation condition and ecosystem functioning, while forest cover ratio represents the structural integrity of natural ecosystems and their carbon sequestration potential, which is especially significant given Malaysia's ecological importance. Forest cover ratio was derived using a 30% tree cover threshold, adopted as a conservative operational definition to distinguish relatively continuous forest cover from more sparsely wooded land. A lower threshold would classify more pixels as forest and increase the estimated forest cover ratio, whereas a higher threshold would yield a more restrictive estimate of forest extent. This cutoff is stricter than the FAO minimum forest definition of 10% canopy cover [30]. However, 30% thresholds have been used in remote-sensing based forest mapping employing global tree-cover products under study-specific forest definitions [31]. Previous research further suggests that the most appropriate threshold may vary across ecological zones and is therefore context-dependent [32]. Accordingly, the 30% cutoff is interpreted here as a pragmatic and internally consistent choice for state-level comparison, not a universally fixed ecological boundary. Surface water area ratio denotes the availability of blue infrastructure that supports climate resilience and hydrological regulation. Lastly, Land Surface Temperature (LST) reflects thermal stress associated with urban morphology and the intensity of heat island effects.

The final set of nine indicators was selected based on three criteria: theoretical relevance to the urbanisation-eco-environment coupling framework, availability of consistent annual data for all 13 states and 3 federal territories over 2005-2019, and comparability across administrative units and data sources. Priority was given to indicators that could capture key demographic, economic, land-use, ecological, and environmental dimensions while remaining

methodologically consistent at the state level. Other potentially relevant variables, such as waste management and direct energy consumption, were considered but not included because complete and comparable annual panel data were not consistently available across all states for the full study period, and because some of their effects were partly reflected by selected indicators such as CO₂ emission intensity and PM_{2.5} concentration. This approach ensured a parsimonious but balanced indicator system suited to the Malaysian research context.

To ensure consistency across all variables, the study period was standardized to 2005–2019, based on the overlap of all datasets. Original datasets span longer periods (e.g., WorldPop since 2000, GHSL since 1975, JRC water data since 1984, EDGAR emissions until 2023), allowing harmonization and temporal alignment before conducting coupling coordination analysis.

Table 1. Indicator Definitions, Data Sources, and Study Periods

System	Dimension	Indicator	Data source	State-level definition	Unit	Dir.	Original years	Study years
U1	Demographic urbanisation	Population density	WorldPop Annual Population Raster	Total population in each state ÷ state land area	persons/km ²	+	2000 – 2020	2005–2019
U2	Economic urbanisation	GDP per capita	State GDP (official stats) + WorldPop population	State GDP ÷ total population	currency/person	+	2005 – 2023	2005–2019
U3	Land urbanisation	Built-up area ratio	GHSL Built-up / Impervious	Built-up pixel area ÷ total state	%	+	1975 – 2030	2005–2019*

			Surface Raster	land area				
E1	Atmospheric pressure	CO ₂ emission intensity	EDGAR	Annual GHG emissions ÷ state land area	tonnes/km ²	-	2000 - 2023	2005-2019
E2	Air pollution pressure	Annual mean PM2.5	Global Annual PM2.5 Grids (MODIS)	Area-weighted mean PM2.5 concentration	µg/m ³	-	1998 - 2019	2005-2019
E3	Vegetation status	Mean NDVI	MODIS NDVI (MOD13A2)	Annual mean NDVI (built-up pixels optional)	dimensionless	+	2000 - now	2005-2019
E4	Forest ecosystem structure	Forest cover ratio	Hansen Global Forest Change	Forest area (>30% tree cover) ÷ state land area	%	+	2000 - now	2005-2019
E5	Surface water status	Surface water area ratio	JRC Global Surface Water	Permanent water area ÷ state land area	%	+	1984 - 2021	2005-2019

E6	Thermal environment	Mean land surface temperature (LST)	MODIS LST (MOD11A2)	Annual mean daytime LST average in each state	°C	–	2000–now	2005–2019
-----------	---------------------	-------------------------------------	---------------------	-----------------------------------------------	----	---	----------	-----------

2.3 Indicator weighting method (CRITIC)

To ensure that each indicator objectively reflects its contribution to the urbanisation and ecological environment system, this study applies the CRITIC (Criteria Importance Through Inter-criteria Correlation) method to determine indicator weights. Unlike subjective approaches such as AHP or expert scoring, CRITIC derives weights directly from the statistical properties of the data, which enhances reliability and reduces personal bias [33].

Because the indicators differ in units and scales, the raw values were first normalized. Positive indicators were standardized using min-max transformation, and negative indicators were transformed accordingly to maintain consistent direction. This step allows all indicators to be evaluated on a comparable scale. As formulas shown:

(1) Positive indicator

When a higher value represents a better performance:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

(2) Negative indicator

When a higher value represents a worse performance:

$$x'_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)}$$

Where:

- x_{ij} is the original value of indicator j in year/state i
- x'_{ij} is the normalized value
- $\max(x_j)$ and $\min(x_j)$ are the maximum and minimum of indicator j

CRITIC evaluates the importance of each indicator from two dimensions. The first is variation, represented by the standard deviation:

$$\sigma_j = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_{ij}' - \bar{x}_j')^2}$$

The second is conflict, which captures how much unique information an indicator contributes relative to others. It is calculated using the correlation matrix:

$$R_j = \sum_{k=1}^n (1 - r_{jk})$$

In the correlation matrix, n denotes the total number of indicators included in the weighting system, and r_{jk} is the Pearson correlation coefficient computed from the normalized data for indicators j and k (the remaining indicators in the system).

Then the comprehensive information content of indicator j is computed as:

$$C_j = \sigma_j \times R_j$$

Finally, the objective CRITIC weight is obtained through normalization:

$$w_j = \frac{C_j}{\sum_{j=1}^n C_j}$$

These weights reflect both the distinguishing ability of the indicator and the extent to which it provides non-redundant information.

2.4 Coupling coordination degree

Following the construction of the weighted indicator system, this study evaluates the interactive relationship between urbanisation and the ecological environment through the coupling coordination degree (CCD) model. The CCD model captures both the strength of interaction between the two subsystems and the extent to which they develop harmoniously over time [34].

The normalized indicators within each subsystem were aggregated using their respective CRITIC weights to obtain the comprehensive development index for the urbanisation subsystem U and the ecological environment subsystem E :

$$U = \sum_{j=1}^{n_U} w_j x_j'; \quad E = \sum_{k=1}^{n_E} w_k x_k'$$

The coupling degree measures the level of interaction between the two subsystems. It describes whether the development of one

subsystem promotes or suppresses the development of the other. The coupling degree C is calculated using:

$$C = \frac{2\sqrt{U \times E}}{U + E}$$

The value of C ranges from 0 to 1, with higher values indicating stronger interaction and closer linkage between urbanisation and the ecological environment.

Although the coupling degree reflects the strength of interaction, it does not capture the overall development status of the two subsystems. Therefore, a coordination degree model is applied by constructing a composite evaluation index T :

$$T = \alpha U + \beta E$$

Where α and β represent the contribution coefficients of each subsystem. In most CCD studies, including this research, urbanisation and ecological environment are considered equally important, thus the baseline setting of $\alpha = \beta = 0.5$, two alternative schemes were tested: $\alpha = 0.4, \beta = 0.6$ and $\alpha = 0.6, \beta = 0.4$.

The coupling coordination degree combines the interaction strength and the overall development level of the two subsystems:

$$D = \sqrt{C \times T}$$

The value of D indicates how harmoniously the two subsystems are evolving together. A higher value reflects stronger coordination and healthier system interaction. Based on standard classification thresholds, the CCD is further categorized into different grades such as disorder, basic coordination, primary coordination, and high coordination (Table 2).

Table 2. Coupling Coordination Grades (CCG)

Grade code	Label	CCD range	Interpretation (plain language)
G1	Severe disorder	0.00–0.20	Systems poorly matched; coordination is very low
G2	Moderate disorder	0.20–0.30	Significant mismatch; limited mutual support

G3	Mild disorder	0.30-0.40	Partial mismatch; some positive interactions
G4	Basic coordination	0.40-0.50	Entry to coordination; fragile and easily reversed
G5	Primary coordination	0.50-0.60	Stable but still improving; coordination taking shape
G6	Intermediate coordination	0.60-0.80	Good alignment; most interactions are synergistic
G7	Good-high coordination	0.80-1.00	High synergy; structure and environment are well matched

2.5 Machine learning model development and validation

This study employed two machine learning algorithms: Random Forest (RF) and Extreme Gradient Boosting (XGBoost), to identify the key determinants of the coupling coordination degree (CCD) and capture nonlinear relationships among indicators. These models were selected because they perform well in high-dimensional settings, handle multicollinearity effectively, and are capable of modelling complex interactions that traditional regression methods may overlook [35].

A series of time-aware predictor variables was constructed to incorporate the dynamic characteristics of the urbanisation and ecological environment systems. For each indicator, three types of features were generated: current-year values to represent immediate subsystem conditions, one-year lagged values to reflect delayed effects and temporal inertia, and trailing three-year means to capture cumulative exposure and smoothed historical patterns relevant to longer-term changes in coordination performance. To avoid temporal leakage, lagged and averaged features were constructed using only information available prior to the prediction target. The models were trained on data from 2005 to 2016 and evaluated on the unseen 2017-2019 hold-out set.

Model development followed standard procedures for tree-based ensemble methods [36]-[37]. The RF model constructed multiple independent decision trees through bootstrap sampling, with final predictions generated by averaging outputs across trees to reduce

variance. XGBoost, by contrast, built trees sequentially, with each new tree fitted to the residual errors of the preceding ensemble. In addition to this boosting structure, XGBoost incorporates regularization terms that help constrain model complexity and reduce overfitting.

For both models, hyperparameters were tuned exclusively within the training set using grid search with five-fold random cross-validation (RCV). For RF, core structural parameters such as the number of trees and maximum tree depth were tuned. For XGBoost, the search additionally considered regularization-related parameters, including `gamma`, `reg_alpha`, and `reg_lambda`, together with `n_estimators`, `max_depth`, `learning_rate`, `subsample`, and `colsample_bytree`. After tuning, the RF and XGBoost models were first compared based on hold-out test performance on the 2017-2019 test set, while the RCV results within the training data were reported to summarize conventional validation performance.

Model accuracy on the hold-out test set was evaluated using the coefficient of determination (R^2) and mean absolute error (MAE). To provide a stricter assessment of generalizability under potential spatial dependence, block cross-validation (BCV) was additionally implemented using state-level folds, such that all observations from the same state were assigned to the same validation block. BCV was not used for hyperparameter tuning; instead, it was applied after tuning as an additional and more conservative validation framework to assess spatial generalizability. The mean and standard deviation of validation R^2 under both RCV and BCV were reported to reflect model stability under conventional and spatially constrained validation settings. The better-performing model was then retained for subsequent interpretation, which is presented in the next section through SHAP (SHapley Additive exPlanations)-based analysis.

2.6 SHAP

To interpret the influence of individual indicators on the model's predictions, SHAP (SHapley Additive exPlanations) was applied. SHAP is a game-theoretic interpretability method that decomposes a model's output into additive contributions from each predictor, allowing both the direction and magnitude of influence to be quantified [38]. This makes it suitable for the indicator system used in this study, which includes current-year values, lagged indicators, and multi-year averages that may interact in nonlinear ways.

In this study, SHAP was used to evaluate how the predictive model

incorporates these temporal and subsystem features when estimating the CCD. Two forms of insight are obtained from the SHAP values. First, by aggregating contributions across all observations, global importance patterns and potential nonlinear effects can be identified, forming the basis for examining global drivers in the Results section. Second, SHAP can be analysed at the level of individual states to explore state heterogeneity, revealing how the influence of a given indicator may vary across different spatial contexts.

To avoid bias and ensure that the interpretability reflects the model's generalization ability, all SHAP analyses were conducted on the hold-out test period covering 2017 to 2019. These years were not used during model training, so the resulting SHAP values represent the model's behaviour on unseen data, which is the recommended practice when working with temporally ordered datasets. Computing SHAP on the test set ensures that the interpretations are not influenced by overfitting on the training period.

For each prediction, SHAP assigns a value to every indicator, indicating whether it increases or decreases the predicted coordination level relative to a model baseline. The baseline is defined as the expected value of the model output, corresponding to the mean predicted coordination level across the entire dataset. A positive SHAP value implies that an indicator contributes to a higher predicted coordination level than the baseline, while a negative value indicates a lowering effect [39]. Summary and dependence plots derived from these values allow visualization of both the relative importance of indicators and how their effects vary across their observed ranges.

By providing a transparent explanation of the predictive model's behaviour, the SHAP analysis complements the CRITIC weighting and CCD modelling. It offers a data-driven understanding of how different indicators, including lagged and cumulative features, contribute to coordination outcomes and supports the interpretation presented in the Results section. It should be noted that SHAP values indicate the relative contribution of features within the model's predictive structure and are used here to support interpretation of feature influence and nonlinear patterns. They do not, however, constitute direct evidence of causal relationships between the predictors and CCD.

3. Results

3.1 CCG distribution

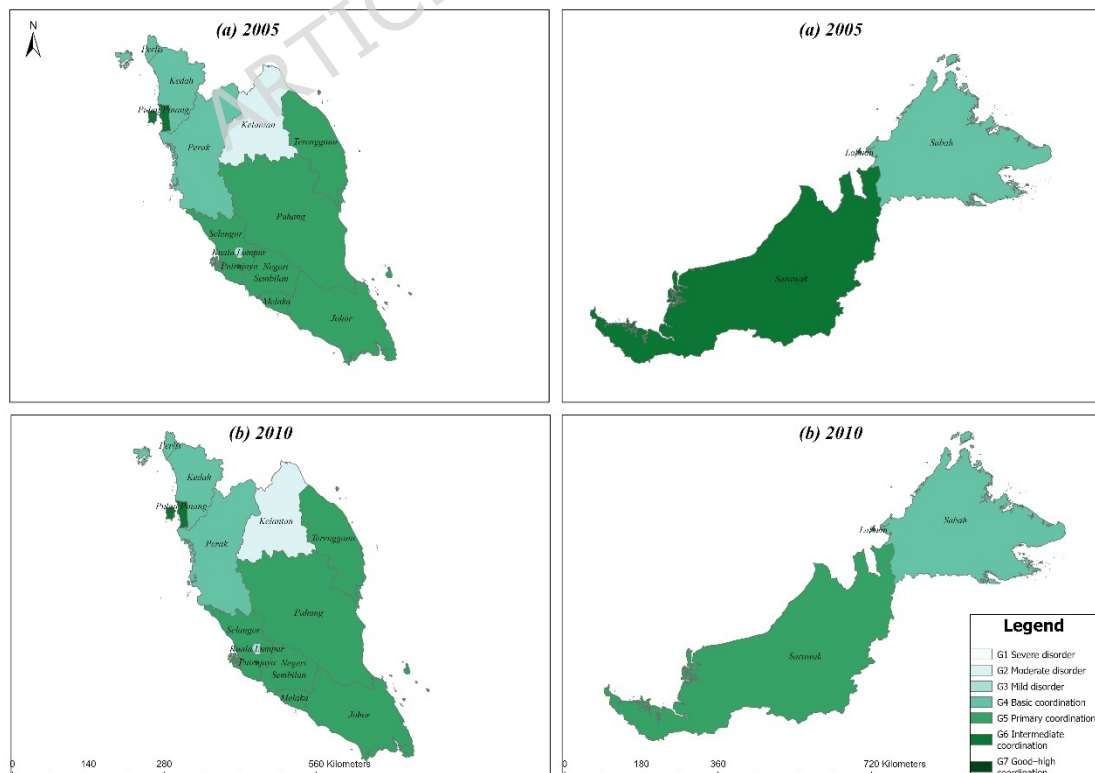
Figure 2 illustrates clear spatiotemporal differences in the Coupling Coordination Grades (CCG) across Malaysia from 2005 to 2019. In 2005, Malaysia was already in a generally coordinated but uneven state. Six states (Johor, Melaka, Negeri Sembilan, Pahang, Selangor and Terengganu) were classified as Basic coordination (G4), while four states (Kedah, Perak, Perlis and Sabah) showed Mild disorder (G3). Two states, Kelantan and Kuala Lumpur, still fell into Moderate disorder (G2), indicating relatively weak coupling between urbanisation and the ecological environment. The highest grade was Intermediate coordination (G6), observed in Labuan, Pulau Pinang, Putrajaya and Sarawak, which formed small high-coordination pockets in both Peninsular and East Malaysia. At this early stage, the pattern suggests that relatively strong economic concentration or ecological buffering helped some states maintain higher coordination, while structurally weaker states remained constrained.

Between 2005 and 2010, the overall spatial pattern remained broadly stable, with only minor shifts in grades. The number of states in Basic coordination increased slightly from six to seven, mainly due to Sarawak moving down from Intermediate to Basic coordination while other central and southern states (Johor, Melaka, Negeri Sembilan, Pahang and Selangor) maintained their Basic status. Four states (Kedah, Perak, Perlis and Sabah) continued to exhibit Mild disorder, and Kelantan and Kuala Lumpur remained in Moderate disorder, suggesting that long-standing structural constraints in these areas had not yet been resolved. In Kelantan, this persistence is broadly consistent with weak income-related momentum and slower improvement in development capacity, whereas Sarawak's downgrade may reflect the gradual erosion of the ecological buffer associated with its relatively large forest endowment. High-coordination areas persisted in Labuan, Pulau Pinang and Putrajaya, which stayed in Intermediate coordination, likely supported by concentrated economic or administrative functions and relatively efficient spatial organization.

In 2015, the distribution became slightly more polarized. Six states (Johor, Melaka, Negeri Sembilan, Sarawak, Selangor and Terengganu) were in Basic coordination, while the number of Mild disorder states increased to six (Kedah, Kuala Lumpur, Pahang, Perak, Perlis and Sabah). Only three states (Labuan, Pulau Pinang and Putrajaya) retained Intermediate coordination, and Kelantan remained the only state in Moderate disorder. The maps therefore

suggest that, although most states stayed within the coordinated category, several regions, particularly on the northern and east-coast side of Peninsular Malaysia and in Sabah, experienced relatively weaker coordination compared with the early period. This weakening is broadly consistent with a combination of limited economic momentum in lower-performing states and rising ecological pressure in states where forest and other natural buffers play an important stabilizing role.

By 2019, Malaysia's CCG pattern showed further consolidation around the middle grades rather than a uniform upward shift. Seven states (Kedah, Kuala Lumpur, Pahang, Perak, Perlis, Sabah and Terengganu) were now classified as Mild disorder, while six states (Johor, Melaka, Negeri Sembilan, Pulau Pinang, Sarawak and Selangor) maintained Basic coordination. Only Labuan and Putrajaya remained in Intermediate coordination, and Kelantan continued to be the sole state in Moderate disorder. This evolution indicates a gradual narrowing of differences between states, but at the cost of a slight decline in the highest coordination grades, with some initially better-performing states drifting from Intermediate to Basic or Mild coordination over time. Taken together, these shifts suggest that sustained improvements in CCG depend not only on economic growth, but also on whether ecological buffers are preserved and whether built-up expansion remains within a range compatible with coordinated development.



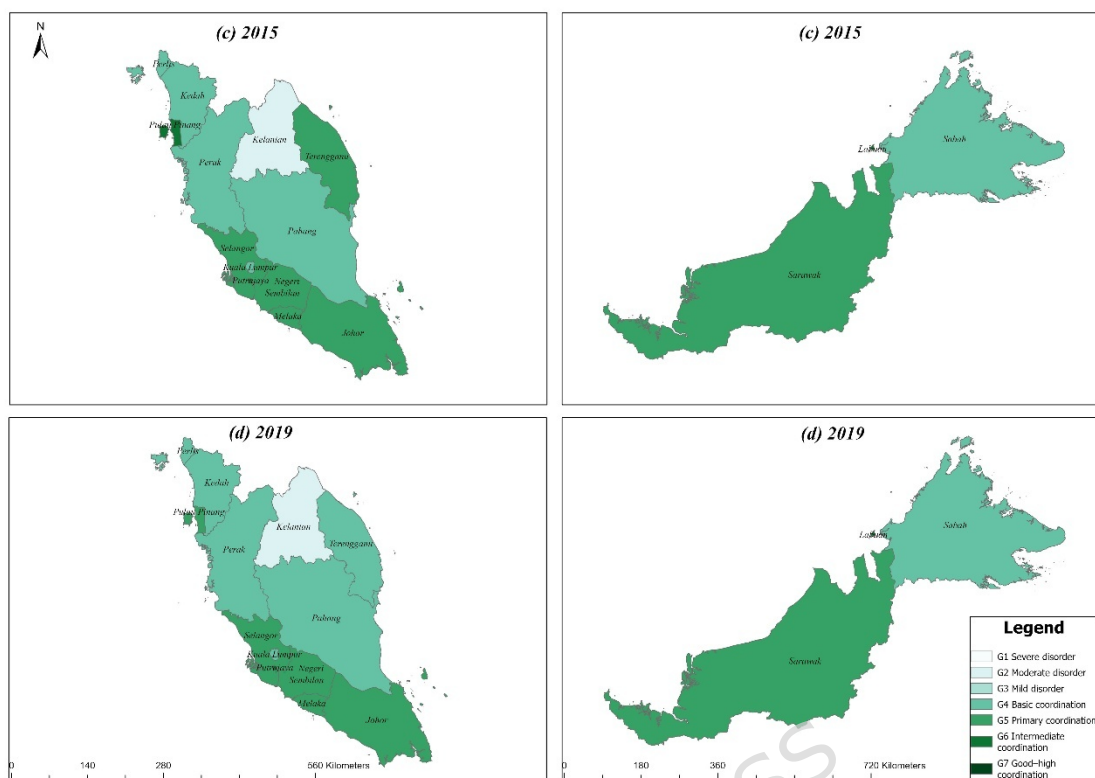


Figure 2. Spatiotemporal evolution of Coupling Coordination Grades (CCG) in Malaysia for the years (a) 2005, (b) 2010, (c) 2015, and (d) 2019. The map was created by the authors using ArcGIS Pro version 3.0.2 (Esri, Redlands, CA, USA; <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>).

3.2 Sensitivity analysis

To test the robustness of the state-level CCG classifications, a sensitivity analysis was performed by varying the contribution coefficients in the composite development index from the baseline setting of $\alpha = 0.5$ and $\beta = 0.5$ to two alternative schemes: $\alpha = 0.4$ and $\beta = 0.6$; $\alpha = 0.6$ and $\beta = 0.4$. The resulting CCD values and grade classifications are summarized in Table 3.

As shown in Table 3, the overall classification pattern remained broadly stable under the alternative coefficient settings. The mean absolute CCD differences from the baseline were 0.027 and 0.026, respectively, while the maximum absolute differences were 0.042 and 0.043. Between 72.08% and 75.83% of state-year observations retained the same grade as in the baseline classification. Although 24.17% to 27.92% of observations changed grade, all changes were confined to adjacent categories, and no shifts of two or more grades were observed.

In substantive terms, the direction of these shifts also helps clarify

the relative roles of the two subsystems. Under $\alpha = 0.4$ and $\beta = 0.6$, most transitions were downward, suggesting that stronger weighting of the eco-environment subsystem lowers the coordination grades of states where environmental performance is relatively weaker than urbanisation. Under $\alpha = 0.6$ and $\beta = 0.4$, several states shifted upward, indicating that their urbanisation subsystem contributes more positively to the overall coordination result when given greater emphasis. Overall, these findings suggest that moderate changes in subsystem contribution coefficients do not alter the broad state-level pattern of coupling coordination, although observations close to grade thresholds remain somewhat sensitive to weighting adjustments.

Table 3. Sensitivity analysis of state-level CCD classifications under alternative contribution coefficients

Scenario	Mean absolute CCD difference	Maximum absolute CCD difference	Unchanged grades (%)	Changed grades (%)	Adjacent grade shifts (n)	Shifts ≥ 2 grades (n)
$\alpha=0.5$, $\beta=0.5$	0	0	100	0	0	0
$\alpha=0.4$, $\beta=0.6$	0.027	0.042	72.08	27.92	67	0
$\alpha=0.6$, $\beta=0.4$	0.026	0.043	75.83	24.17	58	0

3.3 Model comparison

To quantify the factors driving the observed CCD trends and validate their robustness, an ensemble learning approach was employed. Two predictive models, Random Forest and Extreme Gradient Boosting (XGBoost), were implemented and compared based on their performance metrics on the out-of-sample test set (2017–2019). As reported in Table 4, XGBoost slightly outperformed Random Forest in terms of Test R^2 , achieving a value of 0.870 compared with 0.860 for Random Forest. Random Forest yielded a slightly lower Test MAE (0.020) than XGBoost (0.023), but the overall difference in predictive accuracy was small. Under conventional random cross-validation, XGBoost also achieved a higher mean validation R^2 (0.967 ± 0.010) than Random Forest (0.940 ± 0.025), indicating stronger predictive consistency in the standard validation setting.

Concerns regarding overfitting were addressed by expanding the XGBoost tuning procedure to include both structural and regularization-related parameters. In addition to `n_estimators`, `learning_rate`, `max_depth`, `min_child_weight`, `subsample`, and `colsample_bytree`, the search space explicitly included the regularization terms `gamma`, `reg_alpha`, and `reg_lambda`. The final selected configuration was `n_estimators = 800`, `learning_rate = 0.05`, `max_depth = 2`, `min_child_weight = 1`, `subsample = 0.7`, `colsample_bytree = 0.85`, `gamma = 0`, `reg_alpha = 0`, and `reg_lambda = 1`. This selected configuration suggests that model complexity was explicitly constrained during tuning. In particular, the choice of a shallow tree depth (`max_depth = 2`) suggests that the tuning procedure favoured a simpler and more generalizable model structure.

Block cross-validation (BCV) was implemented using state-level folds to provide a stricter assessment of spatial generalizability and to reduce overly optimistic evaluation arising from spatial autocorrelation. Under this stricter validation framework, XGBoost achieved a mean BCV R^2 of 0.503 (SD = 0.363), whereas Random Forest achieved a mean BCV R^2 of 0.265 (SD = 0.384) (Table 4). Although predictive performance declined substantially for both models relative to the random-validation setting, XGBoost retained clearly stronger predictive ability than Random Forest under cross-state validation. The relatively large variation across BCV folds likely reflects the substantial spatial heterogeneity among Malaysia's states, including the differing development trajectories of Peninsular Malaysia and East Malaysia. Under cross-geographical validation, lower predictive consistency is expected because models calibrated in one regional context may not fully capture the distinct development patterns of another. Taken together, the hold-out test results, the RCV results, and the stricter BCV results support the retention of XGBoost as the final model for subsequent SHAP-based interpretation of the key drivers of CCD.

Table 4. Performance metrics of the two machine learning models

Model	Train R^2	Test R^2	Train MAE	Test MAE	RCV R^2 (Mean \pm SD)	BCV R^2 (Mean \pm SD)
Random Forest	0.9954 07	0.8597 46	0.0049 50	0.0204 83	0.940 4 \pm 0.025	0.264 6 \pm 0.384

					4	0
XGBoost	0.9997 08	0.8700 08	0.0013 16	0.0233 13	0.967 3 ± 0.009 9	0.503 3 ± 0.363 2

3.4 Global drivers and nonlinear effects

The global importance of predictors derived from the trained XGBoost model (evaluated on the 2017–2019 test set) is presented in Figure 3. The ranking is determined by the mean absolute SHAP value for each feature, representing its average magnitude of contribution to the predicted CCD score across all test observations.

This ranking shows a clear dominance of income variables: GDP per capita (lag 1) is the top driver, followed by the 3-year mean of GDP per capita and current GDP per capita. Together these reflect both level and momentum effects of economic development on coordination. Among urbanisation and ecological variables, built-up ratio and forest ratio contribute next, then population density. Surface water ratio and PM2.5 have smaller but non-trivial influence, while their rolling means and short-run thermal and NDVI signals are comparatively weaker.

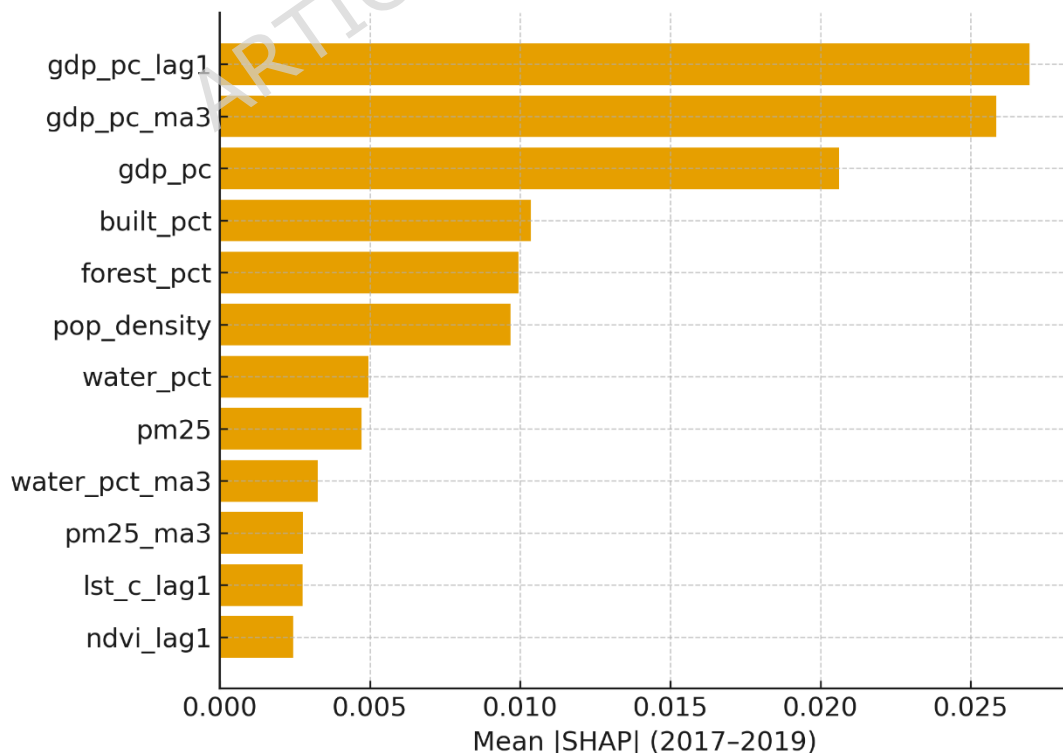


Figure 3. Mean absolute SHAP values for predictors in the XGBoost model for Malaysia, evaluated on the 2017-2019 hold-out test set. Bars represent the average absolute contribution of each feature to the predicted coupling coordination degree (CCD) across state-year observations; larger values indicate greater overall importance in explaining variation in coordination.

Figure 4 shows SHAP dependence plots for the four strongest influential features from the XGBoost model evaluated on the hold-out years 2017-2019. Each blue point represents a state-year observation. The x-axis is the raw value of the predictor. The y-axis is the SHAP value, which is the model-estimated contribution of that predictor to the predicted CCD for that observation. Positive SHAP values increase CCD, negative values decrease CCD. The light grey bars along the x-axis are a small histogram showing where the predictor's values are most common in the data; they are not affects themselves, but help the reader see where data points are concentrated.

For GDP per capita (lag 1 year), the relationship is clearly positive with diminishing returns. At low-income levels, SHAP values are negative, indicating that low past income pulls CCD down. As income rises into the mid-range (roughly 20-40 thousand), SHAP values become strongly positive and reach their largest magnitudes. At very high-income levels the points flatten near zero, showing weaker marginal gains. The grey histogram peaks in the mid-range, so most observations lie where the positive effect is strongest.

For the 3-year mean of GDP per capita, the pattern is similar to the lag-1 variable. Smoothing income over time still yields positive contributions to CCD that are largest at low to mid income levels and then taper as income rises. The current-year GDP per capita also shows a positive, concave pattern: rising income improves coordination, but the incremental benefit declines once a state reaches higher development levels, consistent with basic services and environmental management already being established.

For the built-up ratio, the effect is hump-shaped. Very low built-up shares tend to have small or slightly negative SHAP values, suggesting under-urbanized structure is not conducive to coordination. At moderate built-up levels (about 80-120 on the scale), SHAP values become positive, indicating that compact development supports CCD. At very high built-up shares (extreme outliers above ~200-250), SHAP values turn negative, implying that extensive land

conversion reduces ecological performance and lowers CCD. The grey histogram shows most observations are in the low to moderate range, where effects are small to moderately positive.

Overall, the four panels indicate that income level and its recent dynamics are the dominant, consistently positive drivers of CCD, with the strongest gains at low to middle income levels, and that urban form matters nonlinearly: moderate compactness helps, while excessive expansion harms coordination.



Figure 4. SHAP dependence plots for the four most influential predictors in the XGBoost model, evaluated on the 2017-2019 hold-out test set. Each point represents a state-year observation; the x-axis shows the original predictor value and the y-axis shows its SHAP contribution to the predicted coupling coordination degree (CCD). Grey histograms indicate the distribution of observations and help identify nonlinear patterns, thresholds, and diminishing returns.

3.5 State heterogeneity

While economic drivers dominate the overall global ranking, a detailed state-by-state analysis using SHAP reveals significant cross-state heterogeneity in the relative influence of specific predictors.

The SHAP heatmap (Figure 5) visually illustrates the distribution of influence across all features and states (Test Years 2017–2019). The heatmap highlights that while the first few columns (income terms) are uniformly bright, ecological and urban-form features exhibit varying, localized intensity, indicating that policy levers must be state-specific. Kelantan shows the largest |SHAP| magnitudes for income terms, suggesting an extreme sensitivity to economic momentum. Kuala Lumpur and Perak show distinct signatures in their ecological and urban-form features.

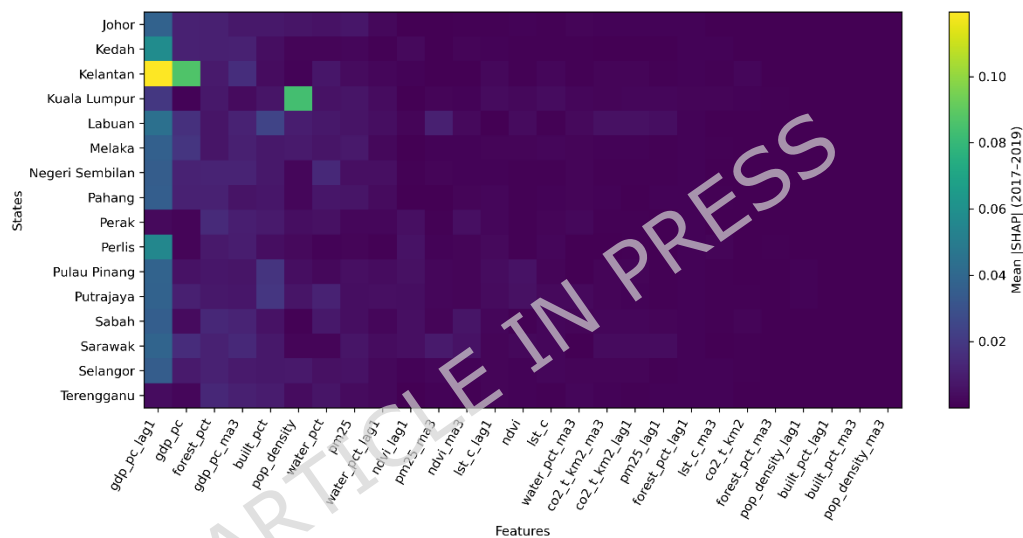


Figure 5. State-level SHAP heatmap of feature influence in the XGBoost model across Malaysia for the 2017-2019 test years. Cell intensity represents the mean absolute SHAP value of each feature within each state, highlighting cross-state heterogeneity in the relative importance of economic, urban-form, and ecological drivers of coordination. The figure was generated by the authors using Python version 3.10.0 with matplotlib version 3.10.0 and pandas version 2.2.3 (Python Software Foundation, Wilmington, DE, USA; <https://www.python.org>; matplotlib: <https://matplotlib.org>; pandas: <https://pandas.pydata.org>).

The Table 5 below quantifies these findings by presenting the Top-3 predictors by mean absolute contribution (|SHAP|) for each state, providing the necessary precision to confirm the observed patterns.

The ranking confirms the primary role of economic dynamics (GDP

per capita, lag-1) as the top driver in most states, including Johor, Melaka, and the two East Malaysian states. However, several critical exceptions and complementary drivers emerge:

- (1) Urban Form Signature: Kuala Lumpur is a clear exception, where population density ranks first, reflecting its unique, highly urbanized context. The built-up share is also highly ranked (Top-3) in compact, high-density areas like Labuan, Pulau Pinang, and Putrajaya, reinforcing the importance of managing urban form where land conversion pressures are intense.
- (2) Ecological Buffers: Ecological structure plays a leading role in several states: forest share leads in Perak and Terengganu, and ranks second in Sabah and Selangor. Similarly, surface water ratio is highly influential in Negeri Sembilan and Putrajaya. This indicates that while income establishes the foundation for coordination, maintaining green and blue infrastructure is the critical marginal step for lifting or buffering CCD in ecologically sensitive or resource-rich states.
- (3) Sensitivity to Income: Kelantan shows an exceptionally high sensitivity to income, with the |SHAP| values for its Top-3 drivers being significantly larger than any other state, suggesting that any positive economic momentum has a magnified effect on its low CCD.

Table 5. Top three predictors ranked by mean absolute SHAP value for each state

State	Top 1	Top 2	Top 3
Johor	gdp_pc_lag1 (0.0377)	gdp_pc (0.0111)	forest_pct (0.0103)
Kedah	gdp_pc_lag1 (0.0582)	gdp_pc (0.0111)	gdp_pc_ma3 (0.0108)
Kelantan	gdp_pc_lag1 (0.1195)	gdp_pc (0.0867)	gdp_pc_ma3 (0.0154)
Kuala Lumpur	pop_density (0.0836)	gdp_pc_lag1 (0.0193)	forest_pct (0.0077)
Labuan	gdp_pc_lag1 (0.0439)	built_pct (0.0243)	gdp_pc (0.0165)
Melaka	gdp_pc_lag1 (0.0365)	gdp_pc (0.0186)	gdp_pc_ma3 (0.0107)
Negeri Sembilan	gdp_pc_lag1 (0.0352)	water_pct (0.0138)	forest_pct (0.0121)
Pahang	gdp_pc_lag1 (0.0357)	forest_pct (0.0110)	gdp_pc (0.0109)
Perak	forest_pct	gdp_pc_ma3	built_pct

	(0.0143)	(0.0100)	(0.0089)
Perlis	gdp_pc_lag1 (0.0552)	gdp_pc_ma3 (0.0095)	forest_pct (0.0080)
Pulau Pinang	gdp_pc_lag1 (0.0383)	built_pct (0.0179)	forest_pct (0.0075)
Putrajaya	gdp_pc_lag1 (0.0376)	built_pct (0.0188)	water_pct (0.0114)
Sabah	gdp_pc_lag1 (0.0358)	forest_pct (0.0132)	gdp_pc_ma3 (0.0119)
Sarawak	gdp_pc_lag1 (0.0387)	gdp_pc (0.0154)	gdp_pc_ma3 (0.0138)
Selangor	gdp_pc_lag1 (0.0354)	forest_pct (0.0109)	gdp_pc_ma3 (0.0086)
Terengganu	forest_pct (0.0132)	gdp_pc_ma3 (0.0104)	built_pct (0.0089)

4. Discussion

4.1 Interpretation of spatiotemporal coupling dynamics

The spatial analysis of the Coupling Coordination Grades (CCG) in Malaysia from 2005 to 2019 points to a development trajectory that is neither purely degrading nor strongly transformative, but one of partial and uneven adjustment. This unevenness is geographically patterned: High-coordination pockets (Intermediate G6) were established in geographically small administrative territories—Labuan, Pulau Pinang, and Putrajaya—which benefit from highly concentrated economic or administrative functions. This combination of centralized resource deployment and limited land area enables the successful, efficient usage of resources for optimal urban-ecological coupling [40]. Conversely, persistent disorder (Mild G3 and Moderate G2) remained concentrated in the northern and eastern Peninsular states, notably Kelantan, which consistently fell into Moderate disorder, points to long-standing structural constraints such as limited economic diversification, weaker infrastructure provision and slower institutional capacity improvements. These conditions hinder the ability of urban development to translate into ecological gains, resulting in coordination levels that remain consistently lower than in more developed regions.

The Federal Territory of Kuala Lumpur (KL), despite being the national capital and largest economic hub, remains an outlier in terms of coordination performance. Its persistent Mild to Moderate disorder can be interpreted through urban systems theory,

particularly the concepts of ecological footprint and urban metabolism, which emphasise how dense economic centres consume resources and generate waste at rates that exceed local environmental carrying capacity [41]. KL's concentrated economic activity, intense land development and limited ecological buffers contribute to a continuous environmental deficit, where pollution, congestion and heat accumulate faster than natural systems or infrastructure upgrades can absorb [42]-[43]. This pattern is consistent with broader research showing that as cities become more densely populated and economically concentrated, environmental pressures and infrastructure congestion tend to intensify [44]-[45]. Furthermore, the effort required to maintain environmental quality often rises non-linearly, or more rapidly, than in less intensive urban systems, reflecting significant environmental threshold effects and the high cost of managing urban externalities [46]-[47]. For Kuala Lumpur, this theoretical relationship implies that part of its economic growth is effectively internalised as continuous environmental and infrastructural stress, thereby slowing its progression toward higher coordination levels compared with smaller or less densely developed cities.

Temporally, the overall CCG pattern showed a consolidation around the middle grades (Basic G4 and Mild disorder G3) rather than a uniform upward shift. This evolution came at the expense of a slight decline in the highest coordination grades. Viewed together with the later SHAP results, this suggests that improvements in coordination were not determined by economic growth alone, but also by whether states could retain ecological buffers and avoid excessive land development pressure. For instance, Sarawak (East Malaysia) was classified as Intermediate coordination (G6) in 2005, likely reflecting the ecological buffer provided by its high forest coverage offsetting nascent urbanisation pressures. Its subsequent move down to Basic coordination (G4) by 2010 indicates that even moderate development may have begun to consume this buffer faster than the state's coordination mechanisms could adjust. More broadly, the temporal pattern implies that states with weaker income momentum remained trapped in lower coordination categories, while states facing stronger built-up or ecological pressure found it difficult to sustain higher coordination levels over time. This trend underscores that while economic resources help prevent severe disorder, sustaining higher levels of coordination requires both continued development capacity and the protection of forest, water, and other ecological assets as urban pressures intensify. The observed dynamic of intensified pressure outpacing adaptive capacity finds a

formal explanation in the recently proposed “Control-Entropy Paradox”, which formalises how efforts to maintain highly ordered socio-ecological states through intensified control and resource use eventually encounter thermodynamic limits, with rising energetic costs and growing externalised disorder that make further gains in coordination increasingly difficult [48].

4.2 Dominance and nonlinearity of key drivers

The machine learning and SHAP results confirmed that economic development is the dominant universal driver of CCG dynamics. This strong economic influence suggests that access to financial resources is a primary determinant of a state's capacity to invest in infrastructure and environmental governance, especially when moving out of lower coordination stages.

The relationship between income and CCG in Malaysia is nonlinear, with clear signs of diminishing returns. The SHAP dependence plots show that the strongest positive contributions to CCG occur as states move from low to mid-range income levels, whereas the effect becomes much weaker at very high-income levels. This indicates that improvements in coordination are most pronounced during the shift from low to middle income, when basic infrastructure, public services and environmental management systems can be established or substantially upgraded. Once these foundations are in place, further increases in income have a more limited impact on coordination unless they are accompanied by deeper changes in production structure, land use and ecological protection. This finding is consistent with stage-based views of urbanisation-environment coupling, in which systems move from disorder to basic coordination before facing more complex constraints at higher development levels [49]-[50], and echoes the logic of Environmental Kuznets Curve hypotheses that emphasize nonlinear relationships between economic growth and environmental outcomes [7]-[8].

Furthermore, urban form exhibits a critical nonlinear effect, showing a hump-shaped relationship with CCG. The SHAP dependence for built-up ratio indicates that moderate built-up levels, corresponding to relatively compact development, are generally conducive to coordination, whereas extremely low or extremely high built-up shares are associated with weaker coordination. This pattern is broadly consistent with compact city and optimal density perspectives, which suggest that moderate concentration can improve the efficiency of infrastructure, transport, and land use. However, these benefits are not unlimited. From the perspective of

the Control-Entropy Paradox, in highly intensified urban systems, greater demands for cooling, waste treatment, pollution control, drainage regulation, and ecological restoration can cause environmental maintenance costs to increase faster than the gains from agglomeration. As a result, further land conversion may weaken rather than strengthen coordination. In other words, the relationship between density and sustainability is subject to context-specific thresholds, beyond which additional physical intensification can undermine environmental quality rather than improve it [48], [51]-[53]. This reinforces the need for compact but carefully managed urban planning if high levels of CCG are to be sustained.

4.3 Policy implications of state-level heterogeneity

The findings challenge the assumption that additional economic growth is the primary route to improving coordination in all Malaysian states. More broadly, the state-level differences identified by the SHAP analysis should be understood within Malaysia's wider development context, where national planning has long sought to balance economic growth, infrastructure provision, urban expansion, and environmental protection through successive Five-Year Malaysia Plans, alongside sectoral instruments such as the National Urbanisation Policy. In this setting, the factors shaping CCG are not only technical or environmental, but are also closely tied to uneven patterns of development, land use governance, and the distribution of ecological assets across states.

In densely developed and geographically constrained territories such as Kuala Lumpur, Labuan, Pulau Pinang and Putrajaya, the SHAP analysis shows that population density and built-up share, rather than income, are among the most influential drivers of CCG. This suggests that in high-density, land-limited contexts the main constraint is not a lack of economic resources but the way space is organised and used. Recent planning instruments already recognise this. For example, the Kuala Lumpur Structure Plan 2040 emphasizes a "green, healthy and vibrant" city with a focus on green networks and efficient mobility, while the National Urbanisation Policy introduced standards for compact, well-serviced urban growth. The state-level results here reinforce these policy directions by indicating that the most effective levers for lifting CCG are those that manage urban form: controlling outward sprawl, promoting vertical and transit-oriented development, and safeguarding urban green corridors and open spaces so that environmental pressures do not outpace economic capacity [54]-[55].

Ecological structure emerges as an equally important, non-income driver in several states. Forest cover is the leading or second-ranked influence in states such as Perak, Terengganu, Sabah and Selangor, while surface water ratio is highly influential in Negeri Sembilan and Putrajaya. This pattern suggests that in ecologically sensitive or resource-rich states, green and blue infrastructure acts as an active buffer that conditions how urbanisation and economic activity translate into coordination outcomes [56]. In the Malaysian context, this is especially important because forests, rivers, wetlands, and reservoirs are not only ecological assets, but are also embedded in state-level development priorities and land use decisions. For these states, policies that protect and restore forests, rivers, wetlands and reservoirs are therefore not secondary environmental add-ons. They are central instruments for sustaining or improving CCG once a basic level of economic stability has been achieved.

Recent controversies over converting forest reserves to development land in Selangor further highlight the tension between economic development and ecological protection. This case illustrates that land and forest governance in Malaysia is not simply a matter of environmental management, but also part of a broader political economy in which natural land may be viewed simultaneously as a conservation asset and a development resource. The model results reinforce the view that, for states where forest and water variables are key drivers, decisions over ecological protection can materially shape the long-term coordination between urbanisation and the eco-environment system. In this sense, forest and water protection are not marginal environmental concerns, but central policy instruments for sustaining or improving CCG.

Kelantan represents a different policy regime. The model identifies an exceptionally high sensitivity of CCG to income-related indicators, with the magnitude of its leading SHAP contributions exceeding those of other states. Combined with its persistent Moderate disorder classification, this suggests that Kelantan's low coordination is driven less by the mismanagement of an already dense, complex system and more by a chronic deficit in economic opportunities and core infrastructure. This interpretation is consistent with assessments that Kelantan continues to lag behind in infrastructure, healthcare, education, and economic diversification [57]-[58]. In such a context, well-targeted development policies that raise incomes, expand essential services and upgrade core infrastructure are likely to produce disproportionately large improvements in CCG, provided that new

growth is accompanied by basic safeguards against environmental degradation [59]. This contrasts with the policy priorities in already affluent and dense territories, underscoring that Malaysia requires differentiated strategies that reflect state-specific combinations of economic capacity, urban form and ecological assets rather than a uniform national recipe.

5. Conclusion

This study advances research on urbanisation-environment interactions by integrating the Coupling Coordination Degree (CCD) framework with interpretable machine learning, specifically XGBoost and SHAP, to examine how urbanisation and ecological integrity co-evolve across Malaysia's 16 states and federal territories during 2005-2019. Rather than treating coordination as a static outcome, the proposed framework captures nonlinear relationships, temporal effects, and state-level heterogeneity, thereby extending conventional CCD analysis beyond descriptive classification toward mechanism-oriented interpretation. In this sense, the study contributes not only a Malaysian case analysis, but also a more flexible analytical strategy for understanding sustainability transitions in rapidly urbanising contexts.

Several findings highlight the novelty of this approach. In response to RQ1, the results show that urbanisation-ecological coordination in Malaysia did not follow a simple upward trajectory, but instead remained uneven across states and over time, with most regions concentrated in middle coordination grades rather than progressing uniformly. In response to RQ2, the machine learning analysis shows that models such as XGBoost can explain and predict variation in CCD with strong performance, while also offering interpretable insight into the relative importance of different drivers. In response to RQ3, the results further demonstrate that key drivers do not operate in a purely linear manner: economic development remains the dominant overall driver of CCD, but with diminishing returns as states move from low to middle income levels, while the built-up ratio exhibits a hump-shaped relationship in which moderate compact development can support coordination but excessive land conversion undermines ecological performance. Finally, the SHAP analysis reveals substantial state heterogeneity, showing that the drivers of coordination differ markedly across dense metropolitan territories, ecologically sensitive states, and structurally lagging regions. Together, these findings show that sustainable urban transitions cannot be explained by income growth alone, but depend on the interaction between economic capacity, spatial form, and ecological

assets.

The broader significance of this study lies in the generalizability of the ML-CCD framework. Many low- and middle-income countries face a similar combination of rapid urban expansion, uneven regional development, ecological vulnerability, and limited planning capacity. In such settings, traditional linear or purely descriptive approaches are often insufficient to identify the thresholds, trade-offs, and place-specific drivers that shape coordination outcomes. By combining CRITIC-weighted CCD measurement with interpretable machine learning, the framework developed here offers a scalable, evidence-based toolkit that can be transferred to other rapidly urbanising regions where multi-source remote sensing and official statistics are available. Its value lies not only in prediction, but in its ability to identify which factors matter most, where they matter, and whether their effects are linear, nonlinear, or context dependent. This makes the framework especially useful for supporting differentiated policy design in settings where development pressures and ecological constraints are closely intertwined.

From a policy perspective, the findings suggest that there is no single pathway to improving urbanisation-environment coordination. In some regions, especially lower-income and structurally constrained states, improving economic opportunities and infrastructure may still generate substantial coordination gains. In others, particularly high-density or land-constrained territories, the key challenge is to manage urban form more carefully and prevent further ecological stress from excessive built-up expansion. In ecologically sensitive states, forest and water assets function not as secondary amenities but as core stabilising mechanisms within the coupled system. The main policy implication, therefore, is that coordinated development depends on matching interventions to state-specific combinations of economic conditions, urban structure, and ecological endowment, rather than applying a uniform national model of sustainability governance.

Several limitations should be noted. The analysis is limited to 2005-2019 because of dataset overlap and therefore does not capture post-2020 changes, including possible shifts associated with the COVID-19 period. In addition, the indicator system relies mainly on remote sensing and official statistics, which cannot fully reflect social and institutional dimensions such as governance capacity or public perception. The study is also subject to the Modifiable Areal Unit Problem (MAUP) because it is conducted at the state level. While

this scale supports national comparison, it may mask important intra-state variation, especially for the built-up ratio, whose SHAP effect should therefore be interpreted as a state-level average rather than a uniform local process. Future research could extend the time period, incorporate broader social and institutional indicators, and test these relationships at finer spatial scales.

Acknowledgements

We are grateful to the anonymous reviewers and editors for their professional feedback and recommendations.

References

- [1] Mukherjee, S. & Bairwa, H. Environmental issues arising from urbanization: A study on the ecological consequences of rapid urban growth. *J. Humanit. Educ. Dev.* **7**, 618268; 10.22161/jhed.7.3.3 (2025).
- [2] United Nations, Department of Economic and Social Affairs, Population Division. *World urbanization prospects: The 2018 revision (ST/ESA/SER.A/420)*. (United Nations, 2018).
- [3] Walsh, P. P., Banerjee, A. & Murphy, E. The UN 2030 agenda for sustainable development. In *Partnerships and the sustainable development goals 1-12* (Springer International Publishing, 2022).
- [4] Mahendra, A. et al. Seven transformations for more equitable and sustainable cities. *World Resour. Rep.* (2021).
- [5] Grimm, N., Faeth, S. H., Golubiewski, N. E., Redman, C., Wu, J., Bai, X. et al. Global change and the ecology of cities. *Science* **319**, 756-760 (2008).
- [6] Wang, D., Xu, P. Y., An, B. W. & Guo, Q. P. Urban green infrastructure: Bridging biodiversity conservation and sustainable urban development through adaptive management approach. *Front. Ecol. Evol.* **12**, 1440477 (2024).
- [7] Grossman, G. M. & Krueger, A. B. Economic growth and the environment. *Q. J. Econ.* **110**, 353-377 (1995).
- [8] Stern, D. I. The rise and fall of the environmental Kuznets curve. *World Dev.* **32**, 1419-1439 (2004).
- [9] Iungman, T. et al. The impact of urban configuration types on urban heat islands, air pollution, CO₂ emissions, and mortality in Europe: A data science approach. *Lancet Planet. Health* **8**, e489-e505 (2024).
- [10] Liu, Y., Fan, C. & Xue, D. A review of the effects of urban and green space forms on the carbon budget using a landscape

- sustainability framework. *Sustainability* **16**, 1870 (2024).
- [11] Alberti, M. The effects of urban patterns on ecosystem function. *Int. Reg. Sci. Rev.* **28**, 168-192 (2005).
- [12] Wang, Q. & Wang, L. The nonlinear effects of population aging, industrial structure, and urbanization on carbon emissions: A panel threshold regression analysis of 137 countries. *J. Clean. Prod.* **287**, 125381 (2021).
- [13] Zhang, Q., Shen, J. & Sun, F. Spatiotemporal differentiation of coupling coordination degree between economic development and water environment and its influencing factors using GWR in China's province. *Ecol. Model.* **462**, 109794 (2021).
- [14] Li, L., Fan, Z., Feng, W., Yuxin, C. & Keyu, Q. Coupling coordination degree spatial analysis and driving factor between socio-economic and eco-environment in northern China. *Ecol. Indic.* **135**, 108555 (2022).
- [15] Dong, G., Ge, Y., Liu, J., Kong, X. & Zhai, R. Evaluation of coupling relationship between urbanization and air quality based on improved coupling coordination degree model in Shandong Province, China. *Ecol. Indic.* **154**, 110578 (2023).
- [16] Wan, H. et al. Coupling coordination relationship between urbanization and water environment in China. *J. Clean. Prod.* **472**, 143423 (2024).
- [17] Aghajani, H., Sarkari, F. & Borhani, M. Coupling coordination analysis between urbanization and ecology in Iran. *Front. Urban Rural Plan.* **2**, 5 (2024).
- [18] Ye, C., Pi, J. & Chen, H. Coupling coordination development of the logistics industry, new urbanization and the ecological environment in the Yangtze River Economic Belt. *Sustainability* **14**, 5298 (2022).
- [19] Xiao, S. et al. Research on the coupling coordination and driving role of urbanization and ecological resilience in the middle and lower reaches of the Yangtze River. *PeerJ* **11**, e15869 (2023).
- [20] Guo, S. et al. The coupling coordination relationship between urbanization and ecosystem health in the Yellow River Basin: A spatial heterogeneity perspective. *Land* **14**, 801 (2025).
- [21] Luo, Q., Guan, Y. & Lei, X. Spatiotemporal coupling coordination between ecological environmental quality and urbanization level and influencing factors of eco-environment: A case study of Jiangsu, China. *Front. Environ. Sci.* **13**, 1537903 (2025).
- [22] Hong, T., Yim, S. H. & Heo, Y. Interpreting complex relationships between urban and meteorological factors and

- street-level urban heat islands: Application of random forest and SHAP method. *Sustain. Cities Soc.* **126**, 106353 (2025).
- [23] Chen, C. et al. Unraveling nonlinear effects of environment features on green view index using multiple data sources and explainable machine learning. *Sci. Rep.* **14**, 30189 (2024).
- [24] World Bank. Urban population (% of total population), Malaysia (SP.URB.TOTL.IN.ZS). *World Development Indicators* (2024).
- [25] Worldometer. Malaysia population. Available at: <https://www.worldometers.info/world-population/malaysia-population/> (2025).
- [26] Kanniah, K. D. Quantifying green cover change for sustainable urban planning: A case of Kuala Lumpur, Malaysia. *Urban For. Urban Green.* **27**, 287-304 (2017).
- [27] Kanniah, K. D. & Ho, C. S. Urban forest cover change and sustainability of Malaysian cities. *Chem. Eng. Trans.* **56**, 673-678 (2017).
- [28] Hutchinson, F. Evolving paradigms in regional development in Malaysia. Preprint at 10.13140/RG.2.2.19324.95364 (2016).
- [29] Azari, M., Billa, L. & Chan, A. Multi-temporal analysis of past and future land cover change in the highly urbanized state of Selangor, Malaysia. *Ecol. Process.* **11**, 2 (2022).
- [30] Food and Agriculture Organization of the United Nations. *Global forest resources assessment 2020: Main report.* (Food and Agriculture Organization of the United Nations, 2020).
- [31] Sannier, C., McRoberts, R. E. & Fichet, L. V. Suitability of Global Forest Change data to report forest cover estimates at national level in Gabon. *Remote Sens. Environ.* **173**, 326-338 (2016).
- [32] Lwin, K. K., Ota, T., Shimizu, K. & Mizoue, N. Assessing the importance of tree cover threshold for forest cover mapping derived from global forest cover in Myanmar. *Forests* **10**, 1062 (2019).
- [33] Alinezhad, A. & Khalili, J. CRITIC method. In *New methods and applications in multiple attribute decision making (MADM)* 199-203 (Springer International Publishing, 2019).
- [34] Li, Y., Li, Y., Zhou, Y., Shi, Y. & Zhu, X. Investigation of a coupling model of coordination between urbanization and the environment. *J. Environ. Manage.* **98**, 127-133 (2012).
- [35] Fatima, S., Hussain, A., Amir, S. B., Ahmed, S. H. & Aslam, S. M. H. XGBoost and random forest algorithms: An in depth analysis. *Pakistan Journal of Scientific Research (PJSR)*. **3**, 26-31 (2023).
- [36] Breiman, L. Random forests. *Mach. Learn.* **45**, 5-32 (2001).

- [37] Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785-794 (ACM, 2016).
- [38] Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. Preprint at 10.48550/arXiv.1705.07874 (2017).
- [39] Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56-67 (2020).
- [40] Su, X., Fan, Y. & Wen, C. Systematic coupling and multistage interactive response of the urban land use efficiency and ecological environment quality. *J. Environ. Manage.* **365**, 121584 (2024).
- [41] Rees, W. E. Ecological footprints and appropriated carrying capacity: What urban economics leaves out. In *The Earthscan Reader in Rural-Urban Linkages* 285-297 (Routledge, 2018).
- [42] Elsayed, I. S. Mitigation of the urban heat island of the city of Kuala Lumpur, Malaysia. *Middle-East J. Sci. Res.* **11**, 1602-1613 (2012).
- [43] Yasin, M. Y., Zain, M. A. B. M. & Hassan, M. H. B. Urbanization and growth of Greater Kuala Lumpur: Issues and recommendations for urban growth management. *Southeast Asia Multidiscip. J.* **22**, 4-19 (2022).
- [44] Sarkar, C. & Webster, C. Healthy cities of tomorrow: The case for large scale built environment-health studies. *J. Urban Health* **94**, 4-19 (2017).
- [45] McGranahan, G. An overview of urban environmental burdens at three scales: intra-urban, urban-regional and global. In *The Earthscan Reader in Rural-Urban Linkages* 298-319 (2018).
- [46] Kahn, M. E. *Green cities: Urban growth and the environment*. (Rowman & Littlefield, 2007).
- [47] Duranton, G. & Puga, D. The economics of urban density. *J. Econ. Perspect.* **34**, 3-26 (2020).
- [48] Hlabisa, S. The Control-Entropy Paradox: Modeling the thermodynamic limits of environmental governance. *Syst. Res. Behav. Sci.* 1-8; 10.1002/sres.3178 (2025).
- [49] Huang, J. C. & Fang, C. L. Analysis of coupling mechanism and rules between urbanization and eco-environment. *Geogr. Res.* **22**, 211-220 (2003).
- [50] Zhang, X. et al. Coupling coordination between the ecological environment and urbanization in the middle reaches of the Yangtze River urban agglomeration. *Urban Clim.* **52**, 101698 (2023).

- [51] Van Der Waals, J. The compact city and the environment: A review. *Tijdschr. Econ. Soc. Geogr.* **91**, 111-121 (2000).
- [52] Lehmann, S. Sustainable urbanism: Towards a framework for quality and optimal density? *Future Cities Environ.* **2**, 8 (2016).
- [53] Bibri, S. E., Krogstie, J. & Kärrholm, M. Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability. *Dev. Built Environ.* **4**, 100021 (2020).
- [54] Organisation for Economic Co-operation and Development. *Compact city policies: A comparative assessment*. (OECD, 2012).
- [55] Ali, L. et al. Dynamics of transit oriented development, role of greenhouse gases and urban environment: A study for management and policy. *Sustainability* **13**, 2536 (2021).
- [56] McNabb, T., Charters, F. J., Challies, E. & Dionisio, R. Unlocking urban blue-green infrastructure: An interdisciplinary literature review analysing co-benefits and synergies between bio-physical and socio-cultural outcomes. *Blue-Green Syst.* **6**, 217-231 (2024).
- [57] Hassan, C. H. B. Poverty in a Malay state in Malaysia: A socio-demographic study. In *Society of Interdisciplinary Business Research (SIBR) 2011 Conference on Interdisciplinary Business Research*. Preprint at 10.2139/ssrn.1867885 (2011).
- [58] Mustapa, W. N. B. W., Al Mamun, A. & Ibrahim, M. D. Economic impact of development initiatives on low-income households in Kelantan, Malaysia. *Soc. Sci.* **7**, 118 (2018).
- [59] Straub, S., Serebrisky, T., Bagnoli, L. S. & Rojas, C. *Infrastructure and poverty reduction: Innovative policies for effective access*. 10.18235/0013445 (2025).