

Climatic Seasonality and Dry-Season Water Availability as Cross-Species Ecological Filters Shaping Medicinal Plant Suitability in Tropical Malaysia

JIYU YANG

jiyuyang0408@163.com

University of Malaya

Tutut Herawan

University of Malaya

Research Article

Keywords: Medicinal plants, Environmental filtering, Climatic seasonality, Dry-season water availability, Species distribution modelling

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9943094/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Climatic Seasonality and Dry-Season Water Availability as Cross-Species Ecological Filters Shaping Medicinal Plant Suitability in Tropical Malaysia

Authors:

Jiyu Yang^{1*}

Email: jiyuyang0408@163.com

Tutut Herawan²

Email: tutut@um.edu.my

Affiliations:

¹ Faculty of Computer Science and Information Technology,
Universiti Malaya,
Kuala Lumpur, Malaysia

² Faculty of Computer Science and Information Technology,
Universiti Malaya,
Kuala Lumpur, Malaysia

*Corresponding author: Jiyu Yang, Email: jiyuyang0408@163.com

Funding:

This research received no external funding.

Abstract

Background:

Medicinal plants represent important components of tropical biodiversity and provide valuable resources for traditional healthcare, conservation, and sustainable bioresource utilization. However, the environmental processes shaping medicinal plant distributions remain insufficiently understood, particularly in tropical regions where high annual temperature and precipitation may obscure the ecological importance of seasonal environmental variation. Understanding whether climatic variability and seasonal resource limitations act as shared environmental filters across species is essential for improving the ecological interpretation of medicinal plant distributions.

Results:

In this study, we investigated cross-species ecological filtering patterns among 30 medicinal plant species in Malaysia using a spatially validated and interpretable ensemble species distribution modelling framework. Occurrence records from 2014-2024 were combined with climatic and topographic predictors, and spatial block cross-validation was applied to reduce the influence of

spatial autocorrelation. Ensemble modelling, uncertainty assessment, and Shapley Additive Explanations (SHAP) were integrated to identify recurrent environmental constraints across species. The results showed that spatial sampling structure strongly influenced ecological inference, highlighting the importance of spatially rigorous evaluation for biodiversity-based modelling. Across species, temperature seasonality and dry-season precipitation variables consistently emerged as important predictors, suggesting that climatic variability and seasonal water availability represent shared environmental constraints influencing medicinal plant suitability patterns.

Conclusions:

This study demonstrates that biodiversity occurrence databases, despite inherent spatial sampling limitations, can provide insights into ecological processes when combined with appropriate validation and interpretation frameworks. Rather than being determined solely by average climatic conditions, medicinal plant distributions in tropical Malaysia appear to be associated with temporal variability in environmental resources. These findings highlight the importance of considering climatic seasonality and dry-season water availability when assessing species responses, conservation priorities, and sustainable management of tropical medicinal plant resources.

Keywords

Medicinal plants; Environmental filtering; Climatic seasonality; Dry-season water availability; Species distribution modelling

1. Introduction

Medicinal plants represent an important component of tropical biodiversity and provide valuable resources for traditional healthcare, cultural practices, and sustainable bioresource utilization. In biodiversity-rich regions such as Malaysia, medicinal plant diversity contributes not only to human well-being but also to conservation efforts and emerging plant-based bioeconomies (Tan et al., 2023). However, despite their ecological and socioeconomic importance, many medicinal plant resources are increasingly threatened by environmental and anthropogenic pressures, including climate change, habitat transformation, and unsustainable harvesting practices (Mykhailenko et al., 2025; Groner et al., 2022). These pressures may alter habitat availability, reshape species distributions, and increase uncertainty regarding the long-term persistence of valuable plant resources (Razak et al., 2024; Zulkeflee et al., 2022). Therefore, understanding the ecological processes underlying medicinal plant distributions is essential for identifying suitable habitats, improving conservation prioritization, and supporting sustainable utilization of plant resources. However, the factors determining why particular medicinal plant species occur in

specific environments remain insufficiently understood, especially across tropical landscapes where environmental conditions vary substantially across space and time.

Species distributions emerge from complex interactions among environmental constraints, physiological adaptation, dispersal processes, and historical factors. Among these processes, environmental filtering plays a fundamental role by restricting species occurrence to environments where ecological conditions match species requirements (Keddy, 1992; Philip, 2006; Weiher et al., 1998).

Environmental filtering provides an ecological framework for understanding why species are not randomly distributed across landscapes and why particular environmental conditions may favour some species while limiting others. However, identifying environmental filters in tropical ecosystems remains challenging because tropical regions are often characterized by high temperature and abundant precipitation, which may suggest relatively weak climatic constraints. In reality, tropical landscapes contain substantial spatial and temporal environmental heterogeneity, and species distributions may depend not only on average resource availability but also on the ability of organisms to tolerate environmental fluctuations. Thus, annual climatic averages alone may fail to capture the environmental pressures that shape species persistence and distribution patterns.

Climate represents one of the most important environmental factors influencing plant distributions, but its ecological influence extends beyond annual temperature and precipitation averages. In monsoonal tropical regions such as Malaysia, high annual rainfall does not necessarily indicate continuous water availability because precipitation is often unevenly distributed throughout the year. Seasonal reductions in precipitation may create temporary periods of water limitation, while temperature fluctuations may impose additional physiological constraints on plant growth, regeneration, and persistence. For perennial plants, including many medicinal species, successful persistence may depend not only on favourable average climatic conditions but also on the ability to withstand environmentally stressful periods. Previous ecological studies have demonstrated that climatic stress and resource variability can contribute to environmental filtering by restricting species persistence under unfavourable conditions (Csörgő et al., 2017; Le Bagousse-Pinguet et al., 2017). Therefore, climatic seasonality and dry-season water availability may represent important environmental constraints shaping medicinal plant distributions, particularly in tropical systems where annual climatic averages may conceal ecologically meaningful seasonal limitations.

Species distribution modelling (SDM) provides an effective framework for investigating relationships between species occurrences and environmental conditions and has been widely applied in biodiversity assessment, conservation planning, and ecological forecasting (Vasconcelos et al., 2024). By linking observed species occurrences with environmental predictors, SDMs allow researchers to estimate habitat suitability and explore potential environmental constraints underlying species

distributions. This approach is particularly valuable for medicinal plant research because comprehensive field surveys across large geographic regions are often limited by accessibility, sampling effort, and resource availability. Therefore, SDMs provide an opportunity to extract ecological information from available biodiversity datasets and support large-scale assessments of plant distribution patterns.

However, ecological interpretation from SDMs depends strongly on the quality, representativeness, and spatial structure of occurrence data. Public biodiversity databases, such as the Global Biodiversity Information Facility (GBIF), provide extensive occurrence records but are frequently affected by uneven sampling effort, with observations concentrated in accessible regions, protected areas, and locations with stronger research activity. Such spatial sampling biases may influence inferred species–environment relationships because models can capture patterns associated with observation effort rather than genuine ecological constraints. Furthermore, spatial autocorrelation represents an important challenge in ecological modelling because geographically close observations often share similar environmental conditions. When random data partitioning is applied, models may achieve artificially high performance by exploiting spatial similarity between training and testing data rather than learning transferable ecological relationships (Valavi et al., 2019; Roberts et al., 2017). Therefore, spatially explicit validation approaches are necessary to distinguish robust environmental signals from patterns generated by sampling structure and to improve confidence in ecological interpretation.

Recent advances in machine learning have expanded the ability of SDM approaches to capture nonlinear and complex species–environment relationships. Ecological systems often involve interactions among multiple environmental factors, threshold responses, and nonlinear relationships that may not be adequately represented by traditional statistical approaches. Tree-based algorithms, including Random Forest, XGBoost, and LightGBM, provide flexible modelling strategies capable of capturing complex interactions among environmental predictors (Cutler et al., 2007; Chen & Guestrin, 2016; Ke et al., 2017). These approaches have been increasingly applied in ecological modelling because they can accommodate heterogeneous environmental responses and improve predictive performance under complex ecological conditions (Liu et al., 2018). Ensemble modelling further provides a strategy to improve prediction stability by integrating multiple algorithms, which is particularly useful when modelling multiple species with contrasting occurrence structures and environmental representation (Huang et al., 2024; Jung et al., 2023; Pan et al., 2025). Nevertheless, predictive performance alone cannot reveal the ecological processes underlying species distributions. Interpretable machine learning approaches, such as Shapley Additive Explanations (SHAP), provide opportunities to identify the environmental variables contributing to model predictions and evaluate whether similar ecological constraints are repeatedly observed across multiple species (Luo et al., 2023; Roberts et al., 2017; Lundberg & Lee, 2017).

In this study, we investigate whether shared climatic filtering processes shape medicinal plant suitability patterns across multiple species in tropical Malaysia.

Rather than focusing solely on identifying the optimal predictive algorithm, we aim to determine whether consistent environmental constraints emerge across species when models are evaluated under spatially rigorous conditions. Specifically, we address three questions: (1) how does spatial sampling structure influence the reliability of species distribution inference from biodiversity occurrence data? (2) do medicinal plant species with contrasting occurrence structures exhibit shared environmental constraints despite differences in data availability? and (3) which climatic factors represent recurrent ecological filters shaping medicinal plant suitability across tropical Malaysia?

To answer these questions, we developed a spatially validated and interpretable ensemble species distribution modelling framework using 30 medicinal plant species based on occurrence records from 2014–2024, climatic and topographic predictors, spatial block cross-validation, explainable machine learning, and uncertainty assessment. By integrating prediction, interpretation, and uncertainty evaluation, this study aims to move beyond species suitability mapping and provide insights into the environmental processes associated with medicinal plant distributions under spatial sampling limitations. Specifically, we examine whether climatic variability, particularly temperature seasonality and dry-season water availability, represents a shared environmental constraint across multiple medicinal plant species.

2. Materials and Methods

2.1 Study design and ecological framework

This study developed a spatially validated multi-species ecological modelling framework to investigate whether shared climatic constraints act as ecological filters shaping medicinal plant suitability patterns in Malaysia.

Malaysia is located within the tropical region of Southeast Asia and represents one of the world's biodiversity-rich areas. The country contains diverse environmental gradients, including differences in temperature regimes, precipitation patterns, seasonal moisture availability, and topographic conditions. Although Malaysia generally experiences a humid tropical climate with high annual precipitation, seasonal variations in climate and regional environmental heterogeneity may impose different ecological constraints on plant persistence and habitat suitability.

The study was designed to move beyond species-level prediction and examine whether transferable environmental signals could be identified across multiple medicinal plant species under spatially constrained occurrence data. The overall workflow consisted of five major stages: (1) occurrence data compilation and quality assessment; (2) environmental variable integration; (3) spatially explicit species distribution modelling; (4) model interpretation and identification of ecological drivers; and (5) uncertainty assessment and robustness evaluation.

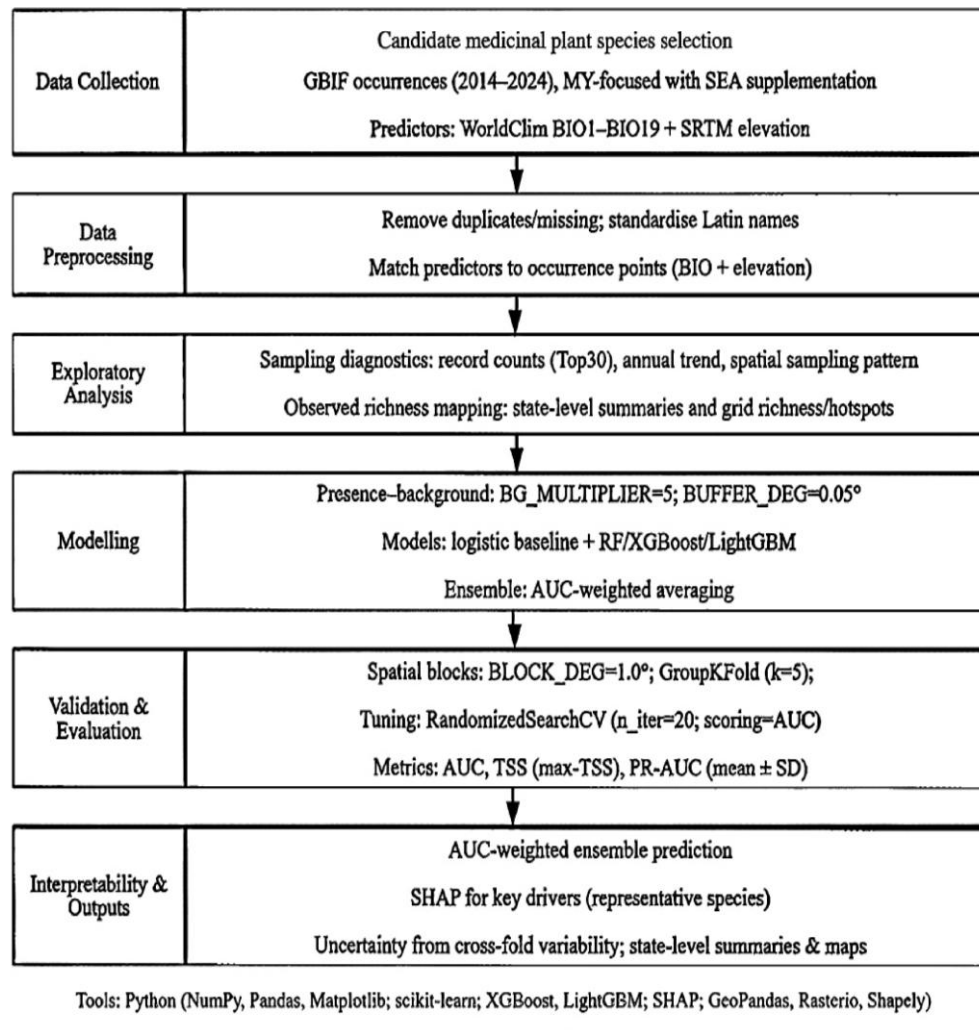


Figure 1. Conceptual framework of the spatially validated ecological inference workflow for identifying cross-species climatic filters of medicinal plants in Malaysia.

Figure 1 illustrates the complete analytical framework, from species occurrence data and environmental characterization to spatial validation, ecological interpretation, and uncertainty assessment. The workflow emphasizes that modelling was conducted not only for prediction, but also for identifying environmental constraints that may represent shared ecological filters across species.

2.2 Species selection and occurrence data processing

A multi-species dataset was established using medicinal plant species occurring in Malaysia. Candidate species were selected based on three criteria: (1) documented medicinal relevance; (2) ecological representation within Malaysian ecosystems; and (3) availability of sufficient occurrence records to support spatial modelling. Medicinal plant candidates were compiled from regional medicinal plant references and biodiversity resources. Species with insufficient geographic information or extremely limited occurrence records were excluded to ensure that the final dataset

supported comparative ecological analysis. The final dataset included 30 medicinal plant species. These species represented different occurrence structures, including species with abundant observations, broad geographic coverage, and species with relatively sparse or spatially clustered records. Rather than treating variation in record availability only as a limitation, differences in occurrence structure were considered part of the ecological modelling context because sampling intensity influences the amount of environmental information available for species distribution inference.

Occurrence records were obtained primarily from the Global Biodiversity Information Facility (GBIF). Only records collected between 2014 and 2024 were retained to represent recent distribution patterns. Occurrence data were subjected to a series of quality-control procedures, including the removal of duplicated observations, records lacking geographic coordinates, and records with insufficient spatial accuracy. Scientific names were standardized to ensure taxonomic consistency during data integration. After filtering, occurrence records were linked with environmental predictors according to geographic coordinates to generate species–environment datasets for subsequent modelling. Because public biodiversity databases often contain uneven spatial coverage, particularly for species with limited observations within specific regions, occurrence records from surrounding Southeast Asian countries were incorporated for selected species to improve environmental representation during model calibration. These regional records were used specifically to increase environmental coverage, reduce information limitations for species with sparse Malaysian observations, and better represent broader ecological gradients during model training. However, the ecological interpretation and evaluation framework remained Malaysia-focused, with model performance assessment and final suitability predictions restricted to Malaysian records to ensure that the results remained aligned with the national-scale objectives of this study.

Specifically, model performance was assessed using Malaysian records, and final suitability predictions, state-level summaries, and uncertainty analyses were interpreted within the Malaysian spatial context. This strategy allowed regional occurrence records to improve environmental learning while maintaining consistency with the national-scale objective of assessing medicinal plant suitability in Malaysia. The final cleaned occurrence–environment dataset used in this study has been deposited in Zenodo (Yang, 2026).

2.3 Environmental variables and ecological predictors

Environmental variables were selected to represent major ecological constraints potentially influencing medicinal plant suitability, including climatic variability and topographic heterogeneity. Because the objective of this study was to identify recurrent environmental filters rather than simply maximize predictive performance, predictor selection was designed to capture broad environmental dimensions associated with plant physiological responses and habitat requirements.

Nineteen bioclimatic variables from WorldClim version 2.1 were used to characterize temperature and precipitation conditions, including annual climatic averages, seasonal variability, and precipitation availability during biologically relevant periods (Fick & Hijmans, 2017). These variables represent different aspects of climatic conditions experienced by plants, including thermal variability, temperature extremes, annual moisture availability, and seasonal water constraints. In particular, variables describing climatic seasonality and precipitation during dry periods provide information on environmental fluctuations that may influence plant persistence beyond mean annual climatic conditions.

Elevation data were obtained from the Shuttle Radar Topography Mission (SRTM) dataset to represent topographic variation that may influence local habitat conditions, including environmental gradients associated with elevation and landscape heterogeneity.

The complete environmental variable set was retained during modelling rather than manually selecting predictors based on prior ecological assumptions. Important environmental drivers were subsequently identified through interpretable model analysis. This approach allowed ecological constraints to be inferred directly from species–environment relationships and reduced the possibility that potential filtering factors were predetermined before modelling.

2.4 Presence-background modelling and ensemble species distribution modelling

Because occurrence records obtained from biodiversity databases represent presence-only observations, a presence-background modelling framework was applied in this study. This approach is widely used in ecological modelling when reliable absence data are unavailable, as it allows comparison between environmental conditions associated with observed occurrences and the available environmental background (Merow et al., 2013). For each species, background points were generated using a multiplier of five times the number of presence records. To reduce potential spatial bias caused by background samples being located too close to known occurrences, a 0.05° exclusion buffer was applied around occurrence points before background generation.

Four modelling approaches were evaluated within the same ecological modelling framework. A regularized logistic regression (LR) model was used as a baseline to represent relatively simple species–environment relationships, while three tree-based machine learning algorithms, including Random Forest (RF), XGBoost, and LightGBM, were applied to capture nonlinear environmental responses and potential interactions among predictors. Random Forest was selected because of its established application in ecological classification problems (Cutler et al., 2007), while XGBoost

and LightGBM were included due to their ability to efficiently model complex nonlinear relationships through gradient boosting approaches (Tianqi & Carlos, 2016; Ke et al., 2017). These algorithms were considered complementary because different species may exhibit different levels of ecological complexity and data availability. Therefore, rather than assuming that a single algorithm would consistently provide the most reliable representation across all medicinal plant species, an AUC-weighted ensemble approach was applied to integrate model predictions and generate more stable suitability estimates under heterogeneous occurrence structures.

2.5 Spatial block validation and model evaluation

Spatial autocorrelation represents a major challenge in species distribution modelling because geographically clustered occurrence records may cause conventional random data splitting to produce overly optimistic estimates of model performance. In biodiversity datasets, nearby observations may share similar environmental conditions and spatial characteristics, allowing models to partially learn geographic patterns rather than transferable species-environment relationships. To reduce this potential bias and evaluate whether models captured ecological signals that could be transferred to unsampled regions, spatial block cross-validation was implemented.

The modelling region was divided into regular spatial blocks with a resolution of 1.0° , and GroupKFold cross-validation with five folds was applied to ensure geographic separation between training and testing datasets. This spatial validation framework was consistently applied across species and modelling algorithms. Figure 2 illustrates the spatial block assignment of five representative medicinal plant species with contrasting occurrence structures, including species with broad distributions, spatially clustered records, and relatively sparse observations. The purpose of this validation strategy was not only to evaluate predictive performance but also to determine whether identified environmental relationships remained robust when models were tested under spatially separated conditions.

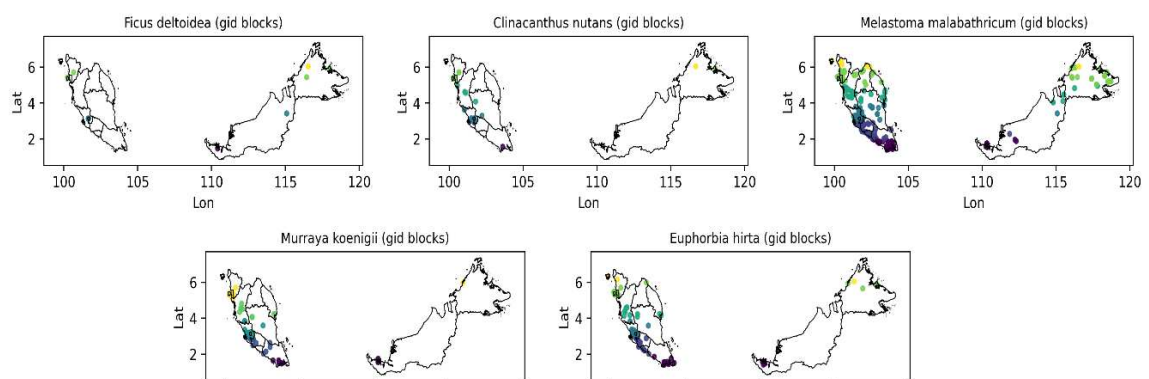


Figure 2. Spatial separation strategy used for blocked cross-validation of representative medicinal plant species.

Model performance was evaluated using multiple complementary metrics, including the area under the receiver operating characteristic curve (AUC), true skill statistic (TSS), and precision-recall area under the curve (PR-AUC). These metrics were selected because the dataset contained imbalanced presence-background samples and substantial variation in occurrence abundance among species. AUC was used as a general measure of discrimination ability, TSS was applied to evaluate threshold-dependent classification performance, and PR-AUC was considered particularly informative under imbalanced conditions. Hyperparameter optimization for tree-based models was conducted using RandomizedSearchCV within the spatial validation framework. Randomized search was selected because it provides an efficient strategy for exploring parameter space under limited computational resources (Bergstra & Bengio, 2012), while maintaining consistency with the spatial evaluation design.

2.6 Ecological interpretation, uncertainty assessment, and sensitivity analysis

To identify environmental variables associated with medicinal plant suitability patterns, model interpretation was conducted using Shapley Additive Explanations (SHAP). Unlike traditional model evaluation approaches that focus primarily on predictive performance, SHAP analysis provides insights into how individual environmental variables contribute to model predictions and enables identification of recurrent ecological signals among species. In this study, SHAP was used to examine whether climatic and topographic variables represented consistent environmental constraints across medicinal plant species.

Representative species were selected to capture contrasting occurrence conditions, including species with abundant records, broad geographic distributions, and sparse or spatially clustered observations. This design allowed evaluation of whether identified climatic signals remained consistent under different levels of data availability and sampling structure. Prediction uncertainty was assessed using variability among cross-validation folds, with higher variability interpreted as greater sensitivity to sampling structure, environmental extrapolation, or differences among modelling approaches. In addition, sensitivity analysis was conducted using alternative spatial block sizes (0.5° , 1.0° , and 2.0°) to evaluate whether the main ecological interpretations depended on the selected spatial validation scale.

By integrating spatial validation, interpretable modelling, uncertainty assessment, and sensitivity analysis, this framework aimed to distinguish robust environmental signals associated with medicinal plant suitability from patterns that may arise due to spatial sampling limitations. This approach allowed ecological interpretation of species distribution patterns while explicitly acknowledging uncertainty inherent in biodiversity occurrence datasets.

3. Results

3.1 Variation in occurrence structure determines the foundation of cross-species ecological inference

The final dataset included 30 medicinal plant species with substantial variation in occurrence availability, regional contribution, and spatial sampling structure. Although all species were evaluated under the same modelling framework, the amount and distribution of available occurrence information differed considerably among taxa. Some species contained relatively large numbers of observations distributed across broader geographic areas, whereas others were represented by sparse or spatially concentrated records.

This uneven occurrence structure reflects a fundamental characteristic of biodiversity datasets rather than simply a limitation of data availability. Species with more extensive records provide broader environmental coverage and stronger support for identifying species – environment relationships, while species with limited observations are associated with greater uncertainty because fewer environmental conditions are represented during model training and evaluation.

As shown in Figure 3, the Top30 medicinal plant dataset exhibited a clear long-tail distribution, with a small number of species contributing a large proportion of records while many species contained relatively limited observations. Species such as *Melastoma malabathricum* and *Euphorbia hirta* showed substantially greater record availability, whereas species such as *Ficus deltoidea* represented sparse-data cases. In addition, the relative contribution of Malaysia-only and Southeast Asian records varied among species, indicating differences in the availability of local observations.

Rather than treating these differences as a source of bias to be removed completely, the variation in occurrence structure was considered an important ecological and methodological context for interpreting model performance. The ability of models to identify environmental constraints depends not only on algorithm selection but also on how much ecological information is contained within available observations.

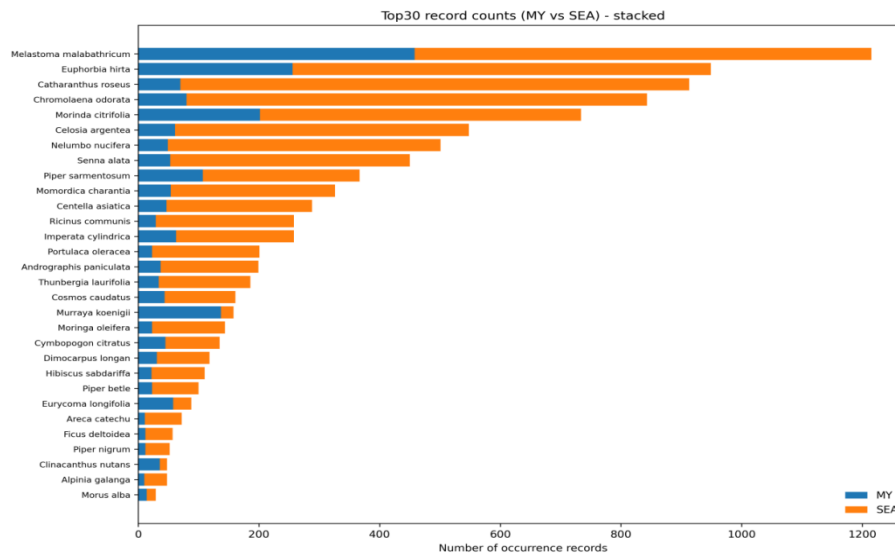


Figure 3. Variation in occurrence availability and regional record composition among the 30 medicinal plant species.

This figure summarizes differences in occurrence record availability among species and provides the ecological data context for interpreting subsequent modelling results.

3.2 Spatial validation reveals conservative but transferable suitability predictions

Under spatial block cross-validation, model performance varied substantially among medicinal plant species, demonstrating that the strength of inferred species–environment relationships differed across taxa with contrasting occurrence structures. These differences indicate that ecological predictability is not determined solely by algorithm selection, but also depends on how much environmental information is represented by available occurrence records and whether sampled conditions adequately capture the ecological niche of each species.

As summarized in Table 1, predictive performance differed among modelling approaches and evaluation stages rather than following a simple pattern of algorithm dominance. The MaxEnt-like logistic regression baseline provided competitive performance for some species, suggesting that relatively simple environmental relationships can explain part of the observed distribution patterns. Meanwhile, tree-based machine learning models improved the ability to represent nonlinear responses and interactions among environmental predictors, which may occur when species respond differently across climatic gradients. However, no individual algorithm consistently achieved superior performance across all species.

The performance-weighted ensemble achieved the most stable overall prediction performance, with a mean AUC of 0.629 under spatially strict evaluation. This value should be interpreted in the context of the validation framework. Compared with random cross-validation, spatial block validation provides a more conservative estimate because geographically separated testing areas reduce the influence of spatial autocorrelation and evaluate whether models capture transferable

species–environment relationships rather than local similarity among observations.

The results suggest that the primary value of the ensemble approach was not simply achieving the highest predictive score, but providing a stable framework for comparing ecological relationships among species with heterogeneous occurrence structures. By reducing dependence on individual algorithms, the ensemble allowed subsequent interpretation of environmental drivers to focus on recurrent ecological signals rather than model-specific patterns.

Table 1 Cross-species predictive performance of modelling approaches under spatial block validation.

model_stage	AUC_mean	AUC_std	TSS_mean	TSS_std	PRAUC_mean	PRAUC_std
B_LGBM	0.608	0.100	0.219	0.144	0.264	0.116
B_LR	0.612	0.101	0.259	0.148	0.249	0.086
B_RF	0.624	0.099	0.250	0.151	0.285	0.120
B_XGB	0.612	0.099	0.229	0.140	0.269	0.115
Ensemble	0.629	0.099	0.254	0.146	0.284	0.120
T_LGBM	0.610	0.105	0.221	0.146	0.267	0.119
T_RF	0.630	0.100	0.255	0.146	0.291	0.118
T_XGB	0.618	0.097	0.233	0.141	0.272	0.111

Figure 4 further demonstrates variation in predictive performance among models and stages. The cross-species comparison indicates a balance between model flexibility and prediction stability. Tree-based approaches generally provided greater capacity to represent nonlinear ecological responses; however, their advantages depended on the availability and quality of ecological information contained within occurrence records.

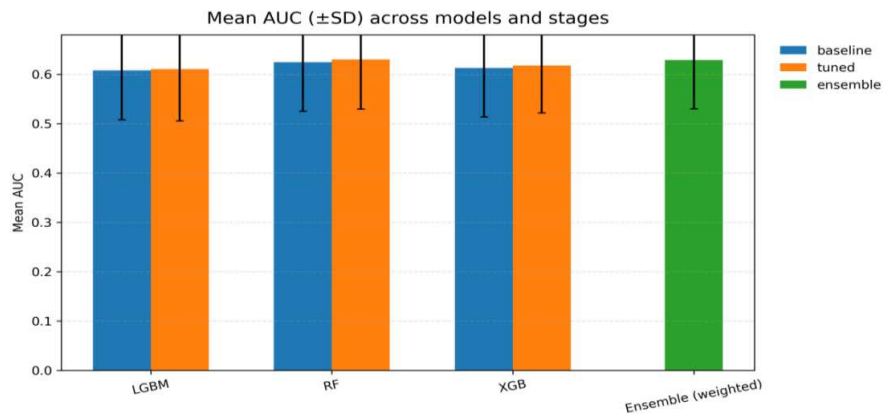


Figure 4. Cross-species model performance under spatially strict validation.

Figure 5 highlights the contribution of the ensemble approach. Although the ensemble did not outperform every tuned single model for every species, it reduced model-specific variability and maintained competitive performance across species with different levels of record availability and spatial structure.

In a heterogeneous multi-species dataset, the most reliable prediction strategy is not necessarily the model achieving the highest individual score, but rather the approach that minimizes instability across different ecological situations. Therefore, the ensemble model was retained as the final prediction layer.

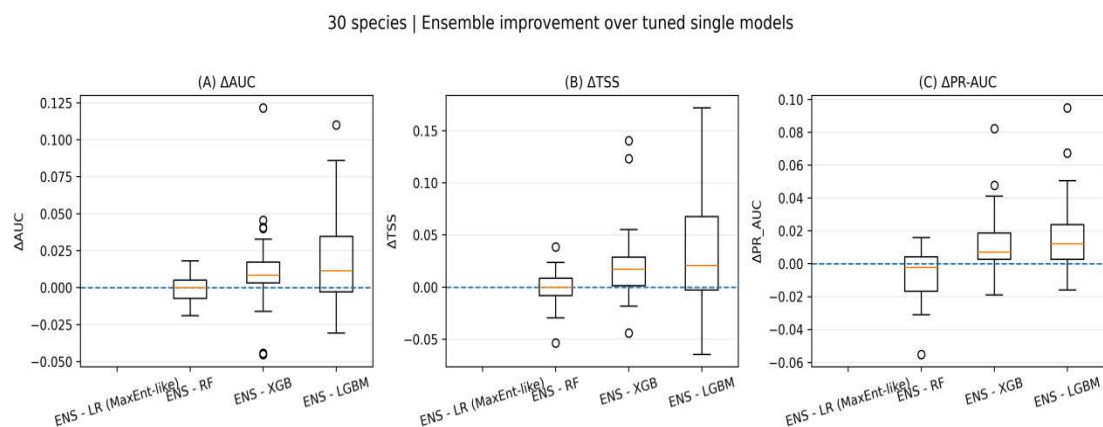


Figure 5. Stability of the performance-weighted ensemble compared with individual tuned models.

3.3 Representative species demonstrate how sampling structure influences model behaviour

To further understand how occurrence characteristics influence model outcomes, five representative species (Rep5) were selected to represent contrasting sampling conditions. These species included high-record species (*Melastoma malabathricum* and *Euphorbia hirta*), intermediate cases (*Cosmos caudatus*), Malaysia-dominant clustered records (*Murraya koenigii*), and sparse long-tail species (*Ficus deltoidea*).

Across these representative species, model behaviour reflected differences in environmental information availability. Species with broader occurrence coverage generally produced smoother suitability estimates and more consistent agreement among models, whereas sparse or spatially clustered species showed greater divergence among predictions.

This pattern indicates that uncertainty in species distribution modelling is not determined only by algorithm selection. Instead, it emerges from the interaction between sampling structure, geographic coverage, and the amount of transferable ecological information contained within occurrence records.

The ensemble approach provided a stable compromise across these contrasting situations. Although individual tuned models occasionally achieved higher scores for specific species, the ensemble reduced dependence on a single modelling assumption and provided a consistent basis for cross-species comparison.

Table 2. Rep5 sample structure and spatial characteristics

Rep5 species (Latin)	Total /MY /SEA	Spatial pattern in Malaysia	Rationale for case study selection
<i>Melastoma malabathricum</i>	1215 /458 /757	Broad coverage; signal also in East MY	High-sample reference species for stable prediction patterns
<i>Euphorbia hirta</i>	949 /256 /693	Widely distributed with dense observations	Common species supporting robustness comparison
<i>Cosmos caudatus</i>	161 /44 /117	More clustered in Peninsular MY	Medium–low sample size representing moderate sparsity and spatial clustering
<i>Murraya koenigii</i>	158 /137 /21	Strong clustering; MY-dominant records	MY-dominant sampling profile supporting state-level interpretation
<i>Ficus deltoidea</i>	57 12 /45	Very sparse and scattered	Long-tail hard case for uncertainty and extrapolation discussion

Better-sampled species produce smoother suitability surfaces and more stable blocked-fold performance, with modest gaps between the ensemble and the strongest tuned single model. Sparse or clustered species show greater divergence among models and more visible uncertainty. This contrast explains why the ensemble is retained as the final output layer: it provides a reliable summary strategy across contrasting data profiles rather than a narrow gain for one taxon.

Figure 6 compares the ensemble with the best single tuned model for Rep5. The ensemble does not always achieve the highest score, but it remains stable across species that differ strongly in sample abundance, spatial spread, and MY versus SEA support. Robustness across contrasting data conditions is more useful than isolated model wins.

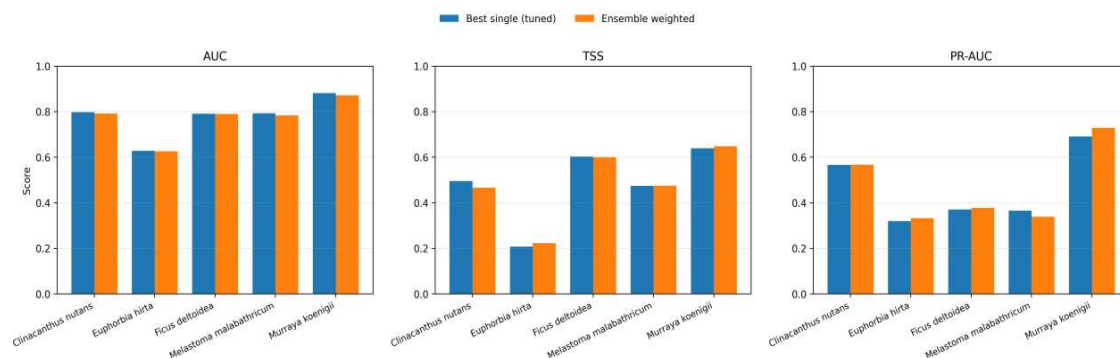


Figure 6. Ensemble stability across representative species with contrasting occurrence structures.

The Rep5 evidence also shows that performance is not a purely algorithmic property. A model that performs better on one representative species may simply benefit from richer, more spatially dispersed sampling; another model that appears weaker on a sparse species may still extract the limited transferable signal available. Model stability, uncertainty, and ecological interpretability need to be read together rather than as separate outputs.

3.4 Spatial suitability patterns, uncertainty, and recurrent ecological drivers

The ensemble modelling framework generated national-scale suitability predictions and uncertainty estimates for medicinal plant species across Malaysia. These outputs provide a broad-scale representation of how environmental conditions associated with observed occurrences are distributed across geographic regions. Rather than being interpreted as direct predictions of species presence, these suitability surfaces represent areas where environmental conditions are relatively consistent with the ecological requirements learned from occurrence data.

State-level aggregation further summarized the spatial variation of predicted suitability and provided a more interpretable perspective for regional ecological screening. Areas with higher mean suitability indicate regions where multiple

environmental predictors collectively supported higher model-estimated suitability, whereas areas with lower suitability represent environments that were less consistent with observed species–environment relationships.

However, suitability predictions should always be interpreted together with uncertainty estimates. Model uncertainty reflects the degree to which predictions vary among spatial folds and modelling procedures. Regions with lower uncertainty represent areas where models showed stronger agreement, whereas regions with higher uncertainty indicate locations where predictions were more sensitive to sampling limitations, environmental extrapolation, or differences among algorithms.

Therefore, uncertainty is not considered an additional limitation separated from prediction outputs, but rather an essential component for interpreting ecological suitability. Combining suitability estimates with uncertainty information provides a more cautious framework for identifying regions requiring further field validation or ecological investigation.

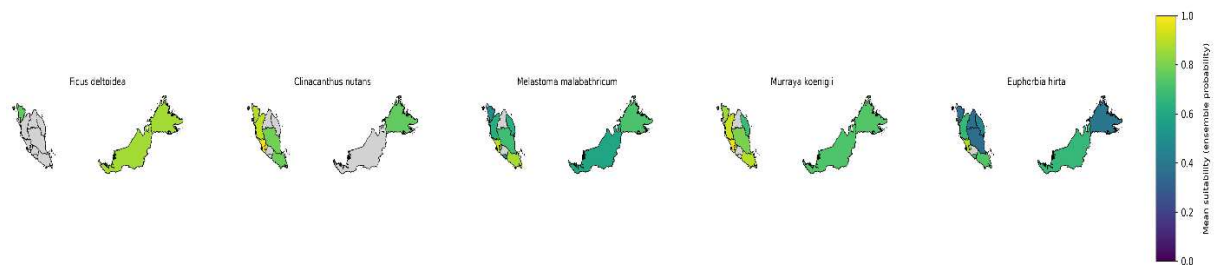


Figure 7. State-level patterns of ensemble-derived medicinal plant suitability across Malaysia.

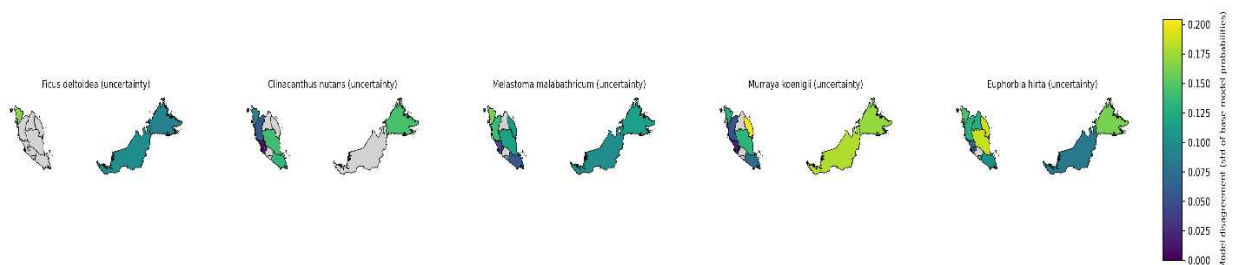


Figure 8. Spatial uncertainty associated with ensemble suitability predictions based on cross-fold variability.

SHAP interpretation was further used to identify whether environmental drivers were consistent across representative species. Figure 9 presents the pooled SHAP analysis for the representative species set and reveals recurrent environmental signals contributing to medicinal plant suitability.

Temperature seasonality showed the strongest and most consistent contribution among environmental predictors. This pattern suggests that medicinal plants may not simply respond to mean thermal conditions, but rather to the degree of climatic variability they experience across seasons. In tropical environments where average temperatures remain relatively high throughout the year, tolerance to thermal

fluctuations may represent an important ecological constraint influencing species persistence and habitat suitability.

Precipitation variables associated with dry periods, particularly BIO14 and BIO17, showed substantial contributions across species. Although Malaysia receives high annual precipitation, the distribution of rainfall throughout the year may create periods of seasonal water limitation. The recurrent contribution of dry-season precipitation variables suggests that plant suitability may depend on the ability of species to persist during temporary moisture deficits rather than simply their association with wet tropical conditions.

These results suggest that medicinal plant suitability in Malaysia is influenced not simply by total annual rainfall, but by seasonal patterns of thermal variability and water availability. In tropical ecosystems, high annual precipitation does not necessarily indicate continuous water availability because seasonal drought periods, rainfall distribution, and moisture stress may strongly influence plant persistence.

The relatively weaker and more variable contribution of BIO12 (annual precipitation) compared with dry-season precipitation variables further supports the interpretation that seasonal water limitation may be more ecologically informative than overall rainfall abundance. Differences; low variability indicates more stable predictions. A high-suitability region with low uncertainty is a stronger candidate for early-stage prioritization; a high-suitability region with high uncertainty should be verified more carefully.

SHAP analysis was used to examine which environmental variables most consistently influenced suitability across the representative species. Figure 9 presents the pooled global SHAP beeswarm for Rep5. Temperature seasonality (BIO4) shows the largest spread and strongest overall contribution, indicating that climatic variability repeatedly separates higher- from lower-suitability environments. Several precipitation-related variables also rank highly, especially BIO14 and BIO17, which show a clear directional pattern: low dry-season precipitation is associated with negative SHAP values, whereas higher values more often contribute positively to suitability. In contrast, BIO12 (annual precipitation) displays a more mixed contribution pattern, suggesting that total annual rainfall is less consistently informative than dry-season water availability.

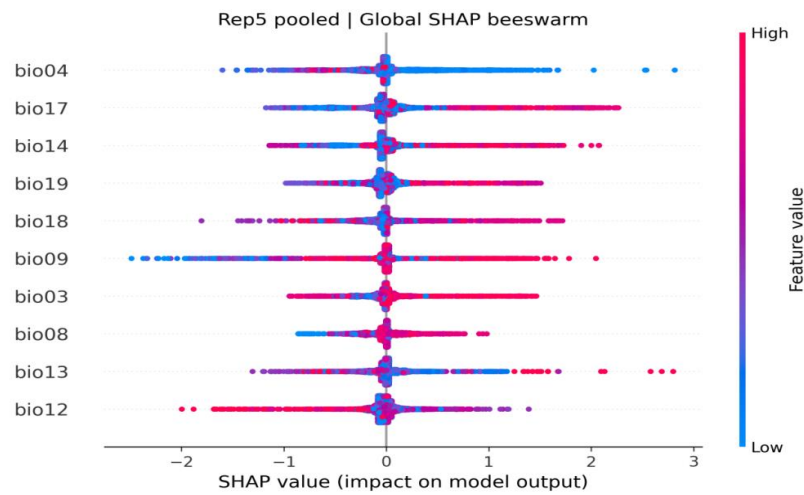


Figure 9. Global SHAP interpretation of recurrent environmental drivers across representative medicinal plant species.

Species-level SHAP analysis further demonstrated that shared ecological constraints do not imply identical responses among all medicinal plants. Figure 10 presents *Murraya koenigii* as an example of species-specific environmental responses.

For *Murraya koenigii*, BIO17 showed the strongest contribution, suggesting that precipitation availability during the driest quarter represents an important ecological constraint. Other variables, including BIO09, BIO04, and BIO14, also contributed to suitability prediction, indicating interactions between thermal conditions and seasonal moisture availability.

This species-level example supports the broader cross-species pattern while demonstrating that individual species retain distinct ecological responses. Therefore, the overall ecological signal identified in this study consists of two complementary components: shared climatic filtering across species and species-specific differences in the relative importance of individual environmental variables.

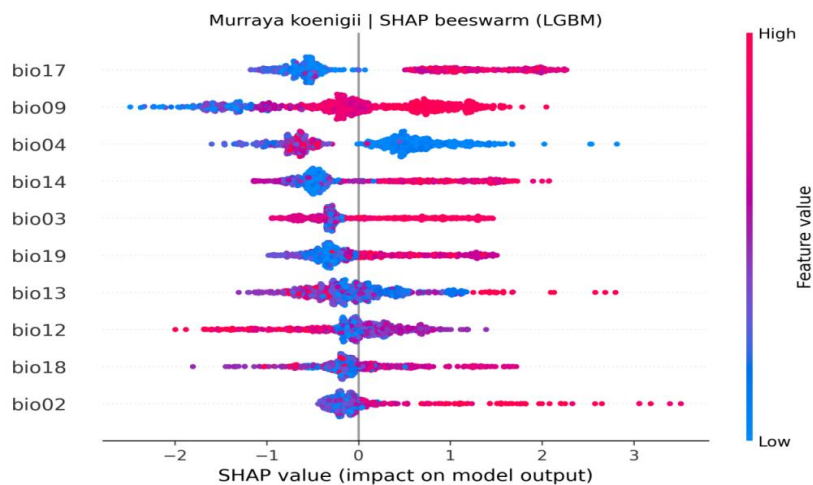


Figure 10. Species-specific environmental contributions illustrated by SHAP analysis of *Murraya koenigii*.

Overall, the SHAP results suggest that temperature seasonality and dry-season water availability are recurrent constraints on medicinal plant suitability in Malaysia. At the same time, the relative balance between thermal and moisture-related effects differs across species. This combination is important for interpretation: it supports a coherent regional ecological signal without reducing all taxa to a single uniform response pattern.

3.5 Robustness of ecological interpretation under alternative spatial validation scales

To evaluate whether the main ecological conclusions depended on the selected spatial validation configuration, sensitivity analysis was conducted using three spatial block sizes (0.5° , 1.0° , and 2.0°).

Mean AUC consistently decreased as spatial block size increased, reflecting increasingly conservative evaluation conditions caused by stronger geographic separation between training and testing data. Smaller spatial blocks allow greater similarity between training and testing environments, whereas larger blocks provide a more challenging assessment of model transferability.

Despite the reduction in absolute performance with increasing spatial separation, the relative ranking among models remained broadly stable across different block sizes. This indicates that the main conclusions regarding model stability and ecological interpretation were not driven by a single spatial validation scale.

The 1.0° block size was therefore retained as the primary validation configuration because it provided a balance between reducing spatial autocorrelation effects and maintaining sufficient environmental representation for regional-scale ecological inference.

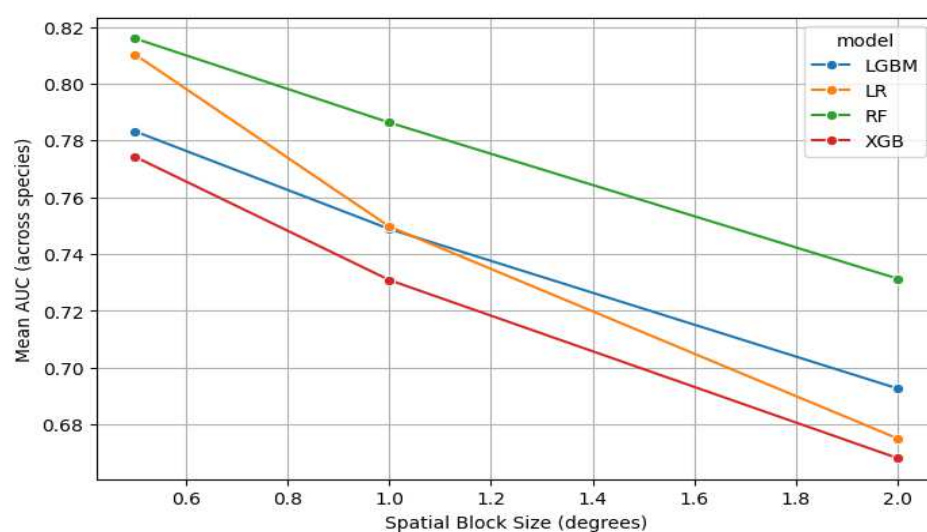


Figure 11. Sensitivity of cross-species model performance to alternative spatial block sizes.

4. Discussion

4.1 Spatial validation reveals transferable ecological signals beyond sampling structure

Species distribution modelling based on biodiversity databases faces an important challenge: occurrence records often reflect both ecological processes and human observation patterns. In publicly available datasets such as GBIF, observations are frequently concentrated in accessible regions, research hotspots, or areas with stronger survey effort. Consequently, conventional random train-test splitting may allow models to exploit spatial similarity rather than identify transferable relationships between species and environments.

In this study, spatial block cross-validation was applied to reduce the influence of spatial autocorrelation and provide a more realistic evaluation of ecological transferability. As expected, predictive performance under spatial validation was more conservative than approaches based on random partitioning. However, the objective of this framework was not to maximize predictive metrics, but to determine whether environmental relationships remained consistent when evaluated across geographically separated areas.

The stability of the ensemble predictions under spatially strict evaluation suggests that the identified suitability patterns were not solely generated by local sampling structure. Instead, the modelling framework captured ecological signals that were sufficiently transferable across space. This distinction is important because ecological interpretation requires separating environmental constraints from patterns created by uneven observation effort.

The sensitivity analysis further supported this interpretation. Although larger spatial blocks resulted in lower predictive performance because of increasingly challenging transfer conditions, the overall ranking among models remained stable. Therefore, the main ecological conclusions were not dependent on a single spatial validation configuration but reflected broader relationships between medicinal plant occurrences and environmental conditions.

4.2 Species distribution modelling across heterogeneous occurrence structures

The absence of a universally dominant algorithm across all medicinal plant species highlights an important characteristic of multi-species ecological modelling: species do not provide identical levels of ecological information. Differences in occurrence quantity, geographic coverage, and spatial clustering influence the complexity of the species–environment relationships that can be inferred from available data.

Species with abundant and geographically distributed observations provide broader environmental representation, allowing models to identify more stable relationships. In contrast, sparse or spatially clustered species contain fewer sampled environmental conditions and therefore

introduce greater uncertainty into ecological inference. These differences explain why individual algorithms performed differently among species.

The performance-weighted ensemble approach provided a practical solution under such heterogeneous conditions. Rather than assuming that one algorithm could represent all ecological responses equally well, the ensemble integrated complementary modelling perspectives and reduced dependence on the assumptions of individual learners.

Importantly, the value of the ensemble was not simply associated with achieving the highest prediction score for every species. Its ecological value lies in providing a stable cross-species framework that allows comparison among taxa with different sampling structures. In biodiversity applications, where multiple species must often be evaluated simultaneously, stability and interpretability may be more important than isolated improvements in predictive performance.

4.3 Climatic seasonality and dry-season water availability as cross-species ecological filters

The most important ecological finding of this study is that medicinal plant suitability in tropical Malaysia was associated not only with broad climatic conditions, but particularly with climatic seasonality and dry-season water availability. These results suggest that tropical plant distributions may be constrained by seasonal environmental stress rather than by average climatic conditions alone. Although tropical ecosystems are often considered energy-rich and water-abundant environments, seasonal fluctuations in temperature and precipitation can create periods of physiological limitation that influence plant establishment, persistence, and geographic distribution.

These findings suggest that tropical plant distributions cannot be fully explained by annual averages alone. Although Malaysia receives high annual precipitation, the timing and distribution of rainfall may create periods of seasonal water limitation that influence plant physiological processes, regeneration, and persistence. Therefore, dry-season precipitation variables may provide more ecologically meaningful information than total annual rainfall because they reflect potential constraints during stressful periods.

Temperature seasonality also emerged as an important predictor, suggesting that climatic variability itself may represent an ecological filter. Rather than simply representing mean thermal conditions, temperature variability reflects environmental fluctuations that species must tolerate. Species capable of maintaining physiological performance under variable thermal conditions may occupy broader climatic spaces, whereas species with narrower tolerance ranges may be restricted to more stable environments.

The consistency of these climatic signals across multiple medicinal plant species provides evidence for cross-species ecological filtering. However, the results do not imply that all species respond identically to climate. Species-level SHAP analysis demonstrated that individual taxa differed in the relative importance of specific predictors. Therefore, medicinal plant distributions appear to be shaped by a combination of shared environmental constraints and species-specific ecological strategies.

4.4 Implications for conservation-oriented ecological screening and sustainable medicinal plant assessment

The suitability predictions generated in this study provide a regional-scale ecological screening framework rather than a substitute for field-based assessment. By combining suitability estimates with uncertainty information, the framework identifies regions where environmental conditions appear consistently favourable while also highlighting areas where predictions require additional verification.

This approach is particularly relevant for medicinal plants because many species face increasing pressures from habitat transformation, climate change, and unsustainable harvesting. Identifying broad-scale environmental suitability patterns can support early-stage conservation prioritization, guide future field surveys, and provide preliminary information for sustainable resource management.

However, several limitations should be considered. First, occurrence records from biodiversity databases remain influenced by sampling bias despite quality control procedures and spatial validation. Second, the environmental predictors used in this study mainly represent broad climatic and topographic conditions and do not include fine-scale factors such as soil properties, land-use history, disturbance regimes, or biological interactions. Third, presence-background modelling estimates environmental suitability rather than confirmed population abundance or demographic persistence.

Therefore, the results should be interpreted as a macro-ecological assessment of potential suitability patterns rather than precise predictions of species distributions at local scales. Future research integrating field observations, soil variables, habitat structure, and temporal environmental changes will further improve understanding of the ecological processes controlling medicinal plant distributions.

5. Conclusion

This study provides an ecological modelling framework for understanding how environmental variability shapes medicinal plant suitability patterns under spatial sampling limitations. By integrating multi-species occurrence records, climatic and topographic predictors, spatial block validation, ensemble species distribution modelling, explainable machine learning, and uncertainty assessment, this study moves beyond evaluating predictive performance and investigates the environmental processes underlying species distributions.

Across 30 medicinal plant species in Malaysia, the results demonstrate that ecological inference from biodiversity databases is strongly influenced by occurrence structure and spatial sampling patterns. Spatially explicit validation provided a more conservative assessment of model transferability and showed that ensemble modelling could maintain stable performance across species with contrasting levels of environmental information. More importantly, the interpretation of model outputs revealed that medicinal plant distributions were not explained solely by broad climatic conditions or annual environmental averages.

Instead, climatic seasonality and dry-season water availability emerged as recurrent environmental constraints across multiple medicinal plant species. In tropical ecosystems where temperature and precipitation are generally abundant, temporal variability in environmental conditions may represent an important source of ecological stress. Temperature seasonality reflects the challenge of coping with thermal fluctuations, while dry-season precipitation variables indicate the importance of maintaining persistence during periods of reduced water availability. These findings suggest that tropical medicinal plant distributions are shaped not only by resource abundance, but also by the ability of species to tolerate seasonal environmental variation.

Although shared climatic constraints were identified across species, individual taxa showed distinct responses to specific environmental factors, indicating that common ecological filters operate together with species-specific ecological strategies. Therefore, cross-species patterns should not be interpreted as identical environmental requirements among all medicinal plants, but rather as evidence that different species respond to a common set of environmental pressures through different ecological pathways.

By combining spatially rigorous validation, interpretable modelling, and uncertainty assessment, this study demonstrates that biodiversity datasets affected by sampling limitations can still reveal meaningful ecological patterns when analysed within an appropriate framework. The findings provide new insights into the environmental processes shaping medicinal plant distributions in tropical Malaysia and highlight the importance of considering climatic variability, rather than only average climatic conditions, when assessing species responses under changing environmental conditions.

SDM: Species distribution modelling;

RF: Random Forest;

LR: Logistic regression;

XGBoost: Extreme Gradient Boosting;

LightGBM: Light Gradient Boosting Machine;

MaxEnt: Maximum entropy;

SHAP: Shapley additive explanations;

GroupKFold: Grouped K-fold cross-validation;

RandomizedSearchCV: Randomized search cross-validation.

AUC: Area under the receiver operating characteristic curve;

ROC-AUC: Area under the receiver operating characteristic curve;

PR-AUC: Area under the precision-recall curve;

TSS: True skill statistic;

SD: Standard deviation.

GBIF: Global Biodiversity Information Facility;

SRTM: Shuttle Radar Topography Mission;

BIO: Bioclimatic variable.

MY: Malaysia;

SEA: Southeast Asia;

Rep5: Five representative species.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The final cleaned occurrence-environment dataset for the 30 selected medicinal plant species in Malaysia, including occurrence records from 2014 to 2024 and associated bioclimatic and elevation predictors, is publicly available in the Zenodo repository: <https://doi.org/10.5281/zenodo.19224526>.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

J. Y. conceived the study, collected and processed the data, conducted the modelling and analysis, interpreted the results, prepared the figures and tables, and drafted the manuscript. T.H. supervised the study, provided methodological guidance, and reviewed and revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

Jiyu Yang and Tutut Herawan are affiliated with the Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia.

References

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of*

- Machine Learning Research*, 13(1), 281-305.
- Csergő, A. M., Salguero-Gómez, R., Broennimann, O., Coutts, S. R., Guisan, A., Angert, A. L., Welk, E., Stott, I., Enquist, B. J., & McGill, B. (2017). Less favourable climates constrain demographic strategies in plants. *Ecology letters*, 20(8), 969-980.
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11), 2783-2792. <http://www.jstor.org/stable/27651436>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12), 4302-4315.
- Groner, V. P., Nicholas, O., Mabhaudhi, T., Slotow, R., Akçakaya, H. R., Mace, G. M., & Pearson, R. G. (2022). Climate change, land cover change, and overharvesting threaten a widely used medicinal plant in South Africa. *Ecol Appl*, 32(4), e2545. <https://doi.org/10.1002/eap.2545>
- Huang, Y., Dong, Y., Huang, W., Guo, J., Hao, Z., Zhao, M., Hu, B., Cheng, X., & Wang, M. (2024). Predicting the Global Potential Suitable Distribution of Fall Armyworm and Its Host Plants Based on Machine Learning Models. *Remote Sensing*, 16(12), 2060.
- Jung, J. B., Park, G. E., Kim, H. J., Huh, J. H., & Um, Y. (2023). Predicting the Habitat Suitability for *Angelica gigas* Medicinal Herb Using an Ensemble Species Distribution Model. *Forests*, 14(3), 592.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree* https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
- Keddy, P. A. (1992). Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, 3(2), 157-164. <https://doi.org/https://doi.org/10.2307/3235676>
- Le Bagousse-Pinguet, Y., Gross, N., Maestre, F. T., Maire, V., de Bello, F., Fonseca, C. R., Kattge, J., Valencia, E., Leps, J., & Liancourt, P. (2017). Testing the environmental filtering concept in global drylands. *Journal of Ecology*, 105(4), 1058-1069. <https://doi.org/https://doi.org/10.1111/1365-2745.12735>
- Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., & Kneeshaw, D. (2018). Application of machine-learning methods in forest ecology: recent progress and future challenges. *Environmental Reviews*, 26(4), 339-350. <https://doi.org/10.1139/er-2018-0034>
- Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions* https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- Luo, W., Han, S., Yu, T., Wang, P., Ma, Y., Wan, M., Liu, J., Li, Z., & Tao, J. (2023). Assessing the suitability and dynamics of three medicinal *Sambucus* species in China under current and future climate scenarios [Original Research]. *Frontiers in Plant Science, Volume 14 - 2023*. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1194444>
- Merow, C., Smith, M., & Silander, J. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, 36. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Mykhailenko, O., Jalil, B., McGaw, L. J., Echeverría, J., Takubessi, M., & Heinrich, M. (2025). Climate change and the sustainable use of medicinal plants: a call for “new” research strategies [Policy and Practice Reviews]. *Frontiers in Pharmacology, Volume 15 - 2024*.

<https://doi.org/10.3389/fphar.2024.1496792>

- Pan, Y., Guan, Y., Lv, S., Huang, X., Lin, Y., Wei, C., & Xu, D. (2025). Assessing the Potential Distribution of *Lonicera japonica* in China Under Climate Change: A Biomod2 Ensemble Model-Based Study. *Agriculture*, 15(4), 393.
- Philip, G. J. (2006). Trait convergence and trait divergence in herbaceous plant communities: mechanisms and consequences. *Journal of Vegetation Science*, 17(2), 255-260.
- Razak, I., Wahab, A. Z., Nasir, D. M., & Ahmad, A. (2024). Predicting Habitat Suitability for Tarantula in Peninsular Malaysia by Using Species Distribution Modelling (SDM). *Tropical Natural History*, 24, 182-192.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929. <https://doi.org/https://doi.org/10.1111/ecog.02881>
- Tan, W.-N., Ali, R., Tong, W.-Y., Leong, C.-R., & Khaw, K.-Y. (2023). Essential Oils from Traditional Medicinal Plants in Malaysia: A Review on Their Traditional Uses, Bioactive Constituents and Therapeutic Potential. *Pharmaceutical Chemistry Journal*, 57(5), 703-711. <https://doi.org/10.1007/s11094-023-02941-x>
- Tianqi, C., & Carlos, G. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://api.semanticscholar.org/CorpusID:4650265>
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225-232. <https://doi.org/https://doi.org/10.1111/2041-210X.13107>
- Vasconcelos, R. N., Cantillo-Pérez, T., Franca Rocha, W. J. S., Aguiar, W. M., Mendes, D. T., de Jesus, T. B., de Santana, C. O., de Santana, M. M. M., & Oliveira, R. P. (2024). Advances and Challenges in Species Ecological Niche Modeling: A Mixed Review. *Earth*, 5(4), 963-989.
- Weiher, E., Clarke, G. P., & Keddy, P. A. (1998). Community assembly rules, morphological dispersion, and the coexistence of plant species. *Oikos*, 309-322.
- Yang, J. (2026). *Cleaned occurrence–environment dataset for 30 selected medicinal plant species in Malaysia*. <https://doi.org/10.5281/zenodo.19224526>
- Zulkeflee, N. H., Adek, N. B. N., Norzin, M. A. F., Awang, A., Ismail, N. A., Daliman, S., & Amaludin, N. A. (2022). Assessing the distribution of *Piper porphyrophyllum* (Sireh rimau) using species distribution model (SDM) in Kelantan, Peninsular Malaysia. *IOP Conference Series: Earth and Environmental Science*, 1102(1), 012066. <https://doi.org/10.1088/1755-1315/1102/1/012066>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryCode.zip](#)