Data-Driven Insights Into Vitex doniana Distribution: A Machine Learning Approach With Selected Abiotic Features

Peter E. Ugege [1] , Peace A. Ukoha [2] , Omokafe A. Ugbogu [3] , Jean C. Ganglo [4]

1. Informatics, Forestry Research Institute of Nigeria, Ibadan, NGA 2. Geographic Information System, Forestry Research Institute of Nigeria, Ibadan, NGA 3. Systematist, Forestry Research Institute of Nigeria, Ibadan, NGA 4. Biodiversity Informatics, University of Abomey-Calavi Calavi, Abomey, BEN

Corresponding author: Peter E. Ugege, pugege@frin.gov.ng

Abstract

Accurate species distribution modeling supports effective ecological conservation and management. This study applies machine learning to predict the distribution of *Vitex doniana* (black plum) in Benin Republic using selected topographical and environmental variables, including altitude, slope, aspect, hillshade, and landcover, alongside species occurrence data from the Global Biodiversity Information Facility. Four models, logistic regression (also known as generalized linear model), support vector machine, K-nearest neighbors, and gradient boosting machine (GBM), were developed and evaluated. Feature importance analysis led to the exclusion of aspect and landcover from some models due to limited predictive value, with altitude emerging as the most influential feature across all models. GBM achieved the highest performance (area under the receiver operating characteristic curve = 0.977, accuracy ≈ 94%) and outperformed others in precision, recall, and F1-score, confirming its robustness. Model outputs revealed that *V. doniana* is predominantly distributed in the southern and northern regions of Benin, where conditions, especially elevation, appear favorable. The central region showed lower suitability, possibly due to land-use patterns or climatic variability. These results underscore GBM's utility in ecological modeling and offer valuable insights for targeted conservation and habitat management strategies for *V. doniana*.

Introduction

Vitex doniana, commonly known as black plum, is a versatile and culturally significant tree species native to sub-Saharan Africa, valued for its nutritional, medicinal, and economic roles. The fruits are vitamin-rich and widely consumed, while its leaves and roots serve traditional medicinal purposes for treating digestive issues and infections. These attributes highlight the species' ecological and socioeconomic contributions, particularly in supporting food security and public health in rural communities [1,2].

Botanically, *V. doniana* is a deciduous tree that can reach heights of 4-15 m, with a thick cylindrical trunk and a wide root system that enhances its resilience in semi-arid environments. The species thrives in various soil types and plays a critical ecological role as a source of food, medicine, and fuelwood [3,4]. While its distribution spans diverse regions in Benin Republic (Figure *1*), its long-term sustainability is increasingly threatened by land-use changes, deforestation, and climate variability [5].
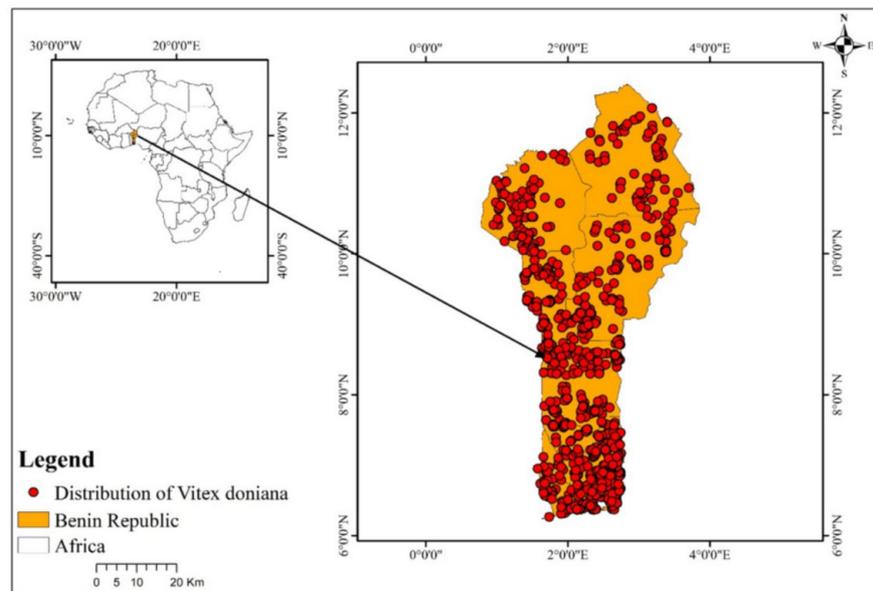
FIGURE 1: Occurrence distribution of Vitex doniana in Benin

To address these threats, effective conservation planning requires tools capable of accurately identifying suitable habitats. Species distribution models (SDMs) provide such capabilities, using environmental variables to map areas of potential occurrence. Although traditional SDMs like Maxent are widely used particularly with presence-only data, they often lack the flexibility to handle complex interactions or incorporate diverse datasets. Machine learning (ML) models offer a compelling alternative, excelling at modeling nonlinear ecological relationships and integrating multiple variable types with high accuracy [6,7].

Several ML algorithms such as Logistic Regression (LR), also known as Generalized Linear Model (GLM), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting Machines (GBM) have been successfully applied in ecological studies due to their adaptability and predictive strength. These models leverage both presence and absence data, improving habitat discrimination and reducing overprediction compared to presence-only approaches [8,9]. Compared to traditional tools like Maxent, ML methods offer enhanced customization, interpretability, and robustness in dynamic and data-rich ecological contexts.

In this study, we apply ML models to predict the distribution of *V. doniana* in Benin Republic using presence-absence data and selected abiotic variables: Altitude, Aspect, Slope, Hillshade, and Landcover. These features were chosen for their influence on habitat suitability through their effects on light availability, soil moisture, and microclimate [10]. By developing and evaluating ML-based SDMs, this research aims to generate spatially explicit insights that support conservation, habitat restoration, and sustainable management efforts for *V. doniana* in the region [11,12].

Materials And Methods

Study area

The study was on the distribution of *V. doniana* in Republic of Benin, a West African nation situated between latitudes 6°10′N and 12°25′N and longitudes 0°45′E and 3°55′E. Benin shares borders with Togo to the west, Nigeria to the east, Burkina Faso and Niger to the north, and the Atlantic Ocean to the south. The country's climate is predominantly tropical, characterized by alternating dry and rainy seasons, which significantly influence vegetation patterns and biodiversity distribution [13].

Benin's landscape exhibits considerable variation, encompassing dense forests in the south to savanna grasslands in the north. Approximately 65% of the country is covered by bushy vegetation, with grasslands accounting for about 93.1% of the highlands [14]. The nation's topography includes low-lying coastal plains, plateaus, and hills, with elevations ranging from sea level to approximately 800 m in the Atakora Mountains in the northwest [13]. These diverse ecological zones and topographical features play a crucial role in determining the distribution of plant species, including *V. doniana* (black plum).

Vitex doniana is a tree native to tropical Africa, commonly known as the black plum. It grows to heights of 4-8 m, occasionally reaching up to 15 m, and thrives in various habitats such as woodlands, savannas, and riverbanks. The species is widely distributed across tropical Africa, including Benin, and is considered common in many regions without major threats to its population [15].

In this study, key abiotic variables, Altitude, Aspect, Slope, Landcover, and Hillshade, were analyzed to assess their influence on the distribution of *V. doniana*. Understanding these factors is essential for effective biodiversity conservation and sustainable management strategies for this species in the region.

Data collection

This study utilized species occurrence data and abiotic variables to model the distribution of *V. doniana*. The data collection process involved two main components:

Species Occurrence Data

Species occurrence records for *V. doniana* were obtained from the Global Biodiversity Information Facility (GBIF) [16], a widely recognized open-access database that provides species distribution data from various sources, including research institutions like the National Herbarium of Benin Republic and citizen science projects. The dataset included georeferenced occurrence points (latitude and longitude) within Benin Republic, ensuring precise spatial mapping of the species. To improve data reliability, duplicate records, erroneous coordinates (e.g., points located in water bodies or outside the study area), and incomplete entries were removed through preprocessing following established SDM best practices [17]. A total of 2,596 occurrence records were downloaded. This reduced to 2,127 occurrence records after duplicate records, and those with incomplete location information were removed. Absence data were obtained based on ecological unsuitability, identified through expert knowledge and spatial analysis, by excluding areas beyond the known ecological range of *V. doniana*. This was implemented using the drawExtent() and crop() functions from the raster package in R to delimit environmentally improbable zones.

Abiotic Data

Five abiotic variables were used to model the distribution of *V. doniana*, comprising four topographical features (altitude, aspect, slope, and hillshade) and one environmental variable (landcover), all known to significantly influence species occurrence:

Altitude (m): Affects macro- and micro-climatic gradients such as temperature and precipitation, which are critical for plant physiology and growth [18].

Aspect (degrees): Indicates the direction a slope faces, influencing solar radiation, temperature regimes, and soil moisture retention [19].

Slope (degrees): Determines terrain steepness, which affects runoff, erosion rates, and soil formation [20].

Landcover: Represents surface vegetation and anthropogenic land use, key for identifying habitat suitability and landscape structure [21].

Hillshade: A terrain-derived variable measuring shading intensity based on sun angle, influencing local microclimatic conditions [22].

These layers were derived from public remote sensing datasets. Altitude, aspect, and slope were extracted from the Shuttle Radar Topography Mission 30 m Digital Elevation Model (DEM) [23], while hillshade was computed from the same DEM. Landcover data was sourced from the European Space Agency Climate Change Initiative Land Cover dataset [21]. All raster layers were resampled to a consistent spatial resolution using QGIS, and values were extracted for each occurrence and background point.

The final dataset was compiled by linking species presence and ecologically inferred absence records with the corresponding abiotic variables. This structured data matrix served as input for the ML models.

Data Splitting

The dataset was randomly partitioned into training (70%) and testing (30%) sets. The training data was further split into 80% for model training and 20% for validation. This hierarchical splitting allowed for robust model training, hyperparameter tuning, and unbiased performance evaluation.

Machine learning models

Four supervised ML algorithms, LR, SVM, KNN, and GBM, were selected to model the distribution of *V.*

---

doniana in Benin Republic. These algorithms were chosen to represent a diverse range of learning paradigms that align with the characteristics of the dataset. The dataset includes a mix of continuous features, such as altitude, and categorical variables, such as landcover. It also exhibits a mild class imbalance, with presence records outnumbering absence data, which was generated through ecologically informed background sampling. Furthermore, the data likely involve nonlinear interactions, a common feature in ecological datasets where complex relationships exist among predictors and species responses.

Each algorithm offers distinct advantages: LR provides interpretability and serves as a linear baseline; SVM is well-suited for high-dimensional, imbalanced data and can model nonlinear decision boundaries; KNN captures local patterns in the data without assuming any specific distribution; and GBM, an ensemble method, is highly effective at modeling complex interactions and iteratively correcting prediction biases [24,25].

Logistic Regression

LR estimates the probability of class membership through a linear combination of input features. As a baseline model, it provides interpretability of feature effects and is particularly useful for identifying variables strongly associated with species presence or absence [26,27].

Support Vector Machine

SVM constructs an optimal separating hyperplane in high-dimensional space. A radial basis function kernel was used in this study to account for nonlinearity in species-environment relationships. SVM is suitable for ecological datasets due to its robustness to feature scaling, ability to manage imbalanced data, and resistance to overfitting [28,29].

K-Nearest Neighbors

KNN is a non-parametric, instance-based learner that assigns class labels based on the majority class among the k closest data points. It does not assume any functional form and is effective in capturing local environmental patterns, making it well-suited for ecological modeling. The optimal k was determined using cross-validation [25,30].

Gradient Boosting Machine

GBM builds an ensemble of decision trees in a sequential manner, where each subsequent tree corrects the prediction errors of the previous one. GBM, implemented using the XGBoost package in R [31], is capable of handling mixed data types and modeling complex feature interactions. It is known to outperform many other algorithms in species distribution modeling, especially when the relationship between predictors and response is nonlinear and hierarchical [24,25].

Feature analysis

To evaluate the contribution of each variable, feature importance analysis was conducted across all four models. For LR, importance was derived from standardized model coefficients, while for SVM, KNN, and GBM, permutation importance was applied. This approach assesses the change in model performance measured by Area Under the Receiver Operating Characteristic Curve (AUC) when the values of a given feature are randomly shuffled.

Variables that consistently contributed little to model performance were removed to simplify the models and enhance their generalization. Specifically, Aspect was excluded from the LR model due to its low coefficient magnitude and lack of statistical significance. Similarly, Landcover was removed from both the SVM and KNN models based on its low permutation importance scores.

Base evaluation metric

Model performance was primarily assessed using AUC, as it effectively captures a model's ability to distinguish between presence and absence classes regardless of the classification threshold [32]. AUC also enables meaningful comparisons across models with different underlying assumptions [33] and supports robust model selection and hyperparameter tuning during cross-validation [28].

Despite its widespread use and strengths, AUC has certain limitations. It can be influenced by class prevalence and may not fully capture model performance in cases of substantial class imbalance. To address this, AUC results were interpreted alongside additional evaluation metrics derived from the confusion matrix, namely, precision, recall, and F1-score, which offer complementary perspectives on classification accuracy and balance.

Results

Model performance overview

To evaluate the predictive capabilities of different algorithms for modeling the distribution of *V. doniana*, four models, LR, SVM, KNN, and GBM, were compared based on their AUC scores (Figure 2). LR recorded an AUC of 0.85236, indicating moderate discriminative ability. SVM achieved an improved AUC of 0.90713, suggesting better classification performance. KNN further increased performance with an AUC of 0.92108, reflecting a strong capacity to capture spatial variability. GBM outperformed all the models, with a notably high AUC of 0.977, demonstrating superior predictive accuracy across the study area.
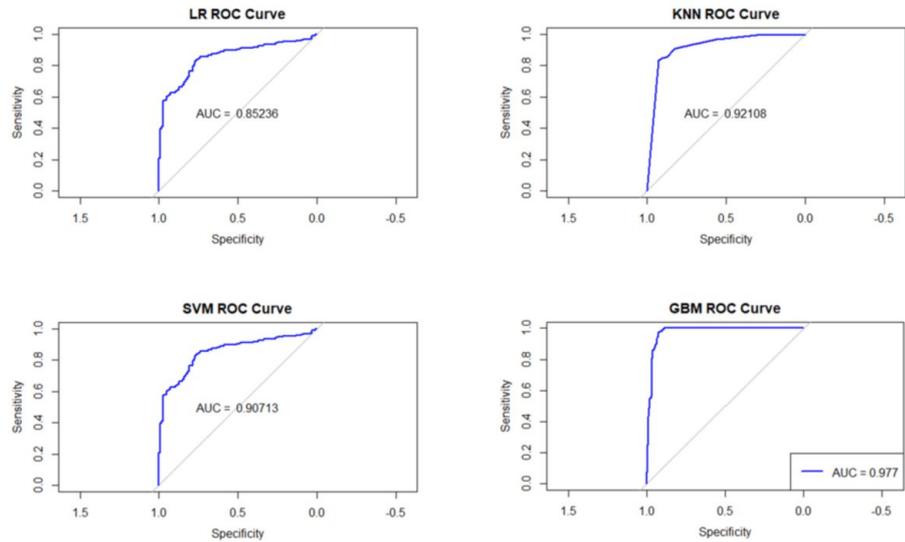

FIGURE 2: AUC scores for the various models

LR, Logistic Regression; KNN, K-Nearest Neighbors; SVM, Support Vector Machine; GBM, Gradient Boosting Machine; ROC, Receiver Operating Characteristic; AUC, Area Under the Curve

Comparative evaluation and model selection

The progressive improvement in AUC scores from LR to GBM highlights the advantage of more sophisticated algorithms. GBM emerged as the most effective model, not only by its highest AUC but also by consistently outperforming others in subsequent evaluations.

GBM evaluation with additional metrics

Beyond AUC, GBM achieved an accuracy of 93.96%, correctly predicting 731 out of 778 test instances (Figure 3). Its precision (0.96496), recall (0.96038), and F1-score (0.96266) affirm its robustness and balanced performance. These results underscore GBM's capacity to capture complex relationships between environmental variables and species presence.
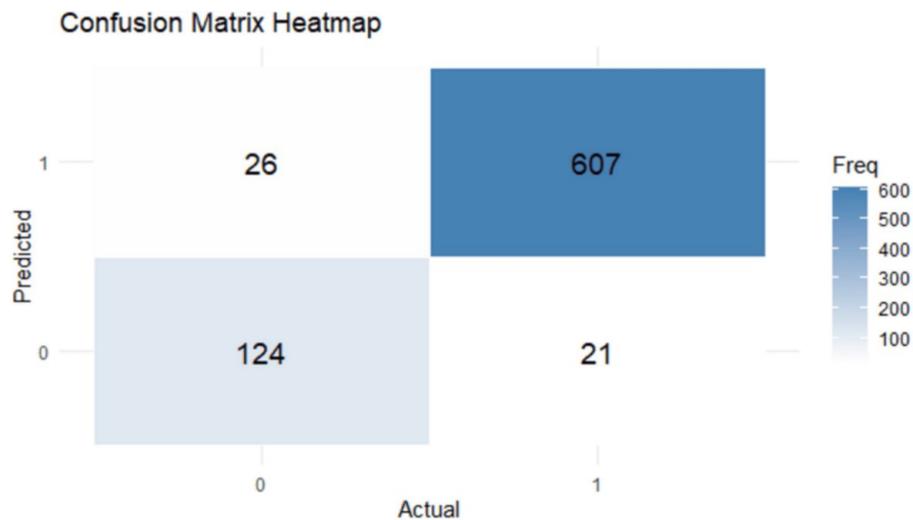
FIGURE 3: GBM confusion matrix

GBM, Gradient Boosting Machine

Predictive map comparison

Predictive maps were generated for all models to visualize habitat suitability for *V. doniana* across Benin Republic (Figure 4). Although all models utilized the same environmental inputs, their outputs differed in granularity and spatial coherence. LR (GLM) predicted widespread suitability, but its relatively lower accuracy suggests possible overgeneralization. The SVM model produced a more localized and refined distribution map, highlighting distinct high-probability zones while exhibiting some uncertainty along boundaries. KNN revealed well-defined clusters with abrupt transitions, reflecting sensitivity to local variations, though this may lead to overfitting in areas with sparse data. In contrast, GBM generated the most consistent and ecologically plausible distribution map, characterized by smooth transitions and strong alignment with known species habitats. Its predictions generalized effectively across diverse environmental gradients.
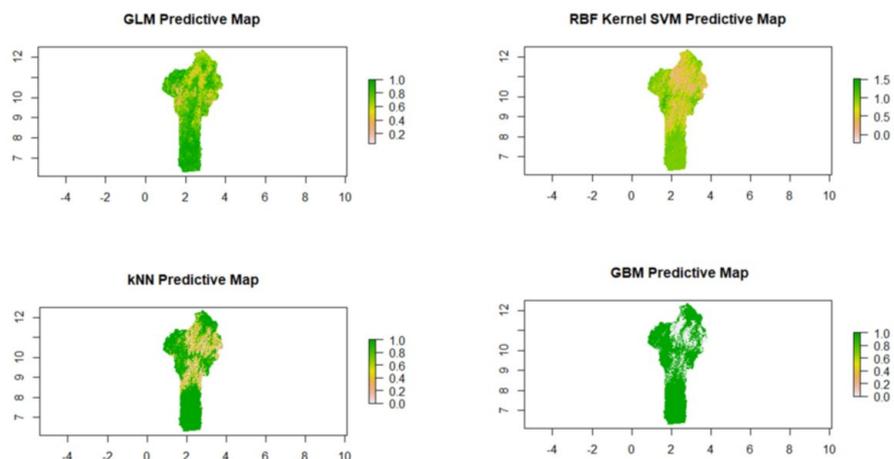

FIGURE 4: Predictive maps of the various models

GLM, Generalized Linear Model; RBF, Radial Basis Function; KNN, K-Nearest Neighbors; GBM, Gradient Boosting Machine

Summary of model rankings

A cross-model comparison ranks GBM highest in predictive performance, followed by KNN, SVM, and LR. The iterative boosting process in GBM likely enhances its ability to model complex, non-linear

relationships between abiotic factors and species distribution.

Feature importance

Feature contribution analysis across all models identified Altitude as the most influential variable (Figure 5). Aspect had minimal influence in LR and was excluded. Similarly, Landcover was removed in SVM and KNN due to negligible impact. These findings emphasize the role of topographical features particularly altitude in shaping species distribution.
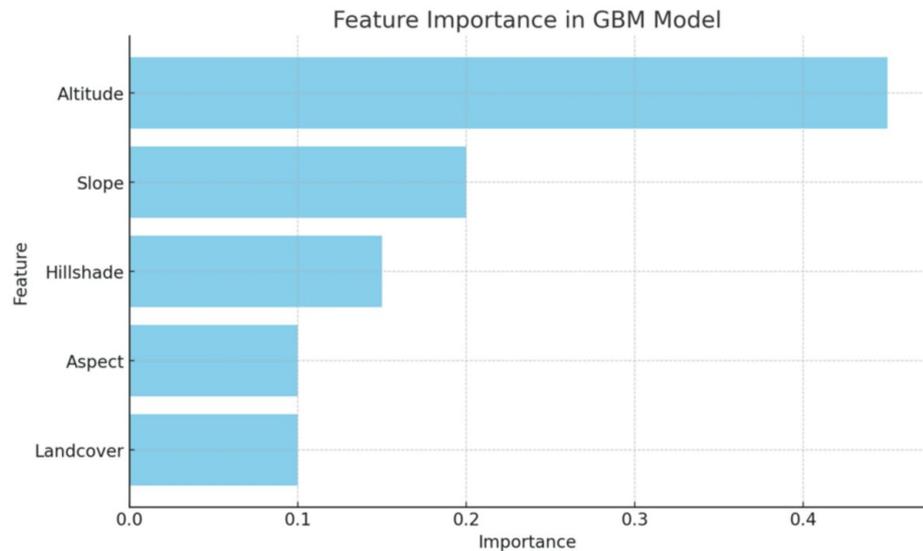

FIGURE 5: Feature importance of features in GBM model

GBM, Gradient Boosting Machine

Partial dependence analysis

Partial dependence plots (PDPs) generated from the GBM model (Figure 6) provide insights into how individual variables influence habitat suitability for *V. doniana*. Altitude shows the highest suitability at low elevations (0-200 m), followed by a sharp decline between 200 and 400 m, and then a gradual increase, suggesting environmental constraints or species adaptability linked to elevation. For slope, suitability increases at low to moderate gradients and then stabilizes, indicating a preference for gentle terrains. Hillshade reveals optimal suitability around a value of 0.64, likely associated with favorable microclimatic conditions such as sunlight exposure and moisture availability. Landcover demonstrates relatively stable suitability across most classes, but a pronounced peak occurs around classes 16-17, followed by a sharp decline, indicating selectivity for specific vegetative environments. Aspect does not exhibit any consistent pattern, reinforcing previous findings that it has minimal influence on the distribution of *V. doniana*.
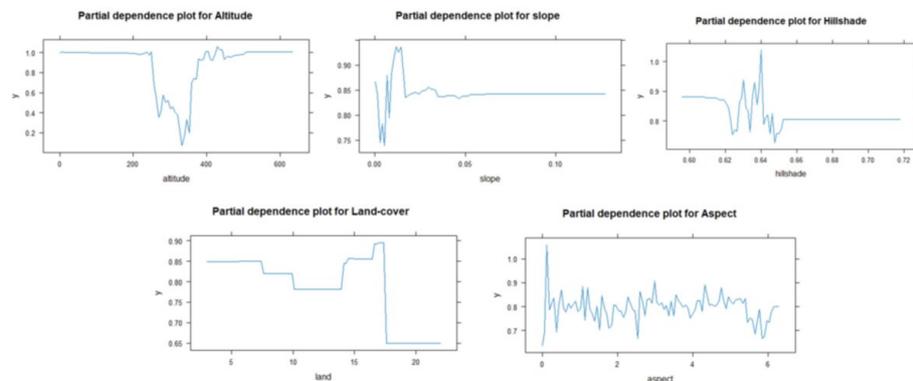

FIGURE 6: Partial dependence plots for the features

Discussion

Predictive performance of the models

This study compared four ML algorithms - LR (GLM), GBM, KNN, and SVM - for modeling the distribution of *V. doniana* across Benin Republic. Each model revealed distinct spatial patterns and predictive strengths.

LR, a baseline linear model, achieved an AUC of 0.85236 and provided broad suitability estimates but lacked the capacity to model complex environmental interactions. Despite this, its simplicity and interpretability make it a useful benchmark.

SVM (AUC: 0.90713) captured non-linear relationships more effectively, producing smoother transitions across habitat zones. However, its overall classification accuracy (~91%) was lower than that of KNN and GBM, and its suitability predictions showed some spatial uncertainty.

KNN performed better than both LR and SVM, with an AUC of 0.92108 and a relatively high accuracy (~92%). It produced sharper predictions but also displayed patchiness, likely due to sensitivity to local data density and neighborhood structure, which may compromise generalizability in sparsely sampled regions.

GBM outperformed all other models with an AUC of 0.977 and accuracy of 93.96%. Its ensemble-based iterative learning approach effectively captured non-linear relationships and complex interactions among variables, leading to precise and ecologically plausible predictions. Importantly, it consistently identified key habitat areas with minimal overprediction.

Comparative analysis

Comparing the four models highlights a performance gradient: GBM > KNN > SVM > LR. GBM provided the most refined and reliable distribution map, aided by its ability to handle high-dimensional interactions and reduce bias through boosting. KNN, while strong, showed limitations due to local overfitting and scattered predictions. SVM achieved good generalization but slightly underperformed in spatial precision. LR, while the least accurate, remains valuable in contexts prioritizing model simplicity and interpretability.

The comparative evaluation justifies the selection of GBM for focused distribution mapping, as it consistently outperformed others across all key performance metrics and produced ecologically coherent results.

Ecological interpretation and distribution mapping

Using the GBM model, we observed that *V. doniana* is predominantly distributed in the southern and northern regions of Benin Republic. These areas correspond to environmental conditions - particularly low to moderate altitudes (0-200 m) - that are favorable for the species' growth and survival.

Altitude emerged as the most influential feature across all models. This aligns with ecological understanding that *V. doniana*, a savanna and forest-edge species, favors lowland tropical zones with stable microclimates and fertile soils. Lower altitudes may also offer higher soil moisture retention and reduced temperature variability, which support seedling establishment and long-term viability.

Conversely, the central region exhibited lower suitability, which may reflect anthropogenic pressures such as land-use change, urban expansion, or shifts in microclimatic patterns that hinder natural regeneration. This geographic delineation highlights potential conservation hotspots and areas requiring reforestation or habitat protection efforts.

Limitations and implications

While GBM demonstrated high predictive performance, its complexity may limit interpretability and overfit smaller datasets if not properly tuned. Additionally, all models relied on available occurrence and environmental data, which may introduce spatial sampling bias or omit biotic interactions (e.g., pollinators, seed dispersers) that also influence species distribution.

Ecologically, the findings emphasize the importance of topographic and climatic variables in shaping species niches, supporting the need for spatially explicit conservation strategies. The high performance of GBM suggests that advanced ensemble methods are valuable tools in ecological modeling, especially when fine-scale habitat delineation is essential.

Conclusions

This study demonstrates the effectiveness of ML models for SDM, using *V. doniana* as a case study in

Benin Republic. Among the four models evaluated, LR (GLM), KNN, SVM, and GBM, GBM emerged as the most robust, achieving the highest accuracy (93.96%) and AUC (0.977), and generating ecologically consistent suitability maps. The model identified southern and northern Benin as high-suitability zones for *V. doniana*, largely influenced by altitude, which was the most significant predictor across all models. The strong correlation between altitude and species presence suggests that elevation-related factors, such as temperature gradients, soil moisture, and nutrient availability, critically shape the species' ecological niche. These insights are crucial for guiding conservation strategies, particularly in prioritizing regions for protection or restoration. In terms of algorithmic performance, GBM's ensemble learning approach proved highly effective in capturing complex, non-linear relationships among environmental variables. SVM also performed well in modeling non-linear interactions, while KNN provided localized pattern recognition but suffered from generalization issues. LR, though interpretable, lacked the capacity to model intricate ecological patterns and was the least effective in predictive accuracy. Given these findings, GBM is recommended for future SDM efforts, particularly where precision and spatial clarity are essential. However, future studies could benefit from exploring hybrid approaches, such as stacking or blending models (e.g., combining GBM with spatial autoregressive models or deep neural networks) to harness complementary strengths of different algorithms.

Despite the promising results, this study has several limitations. First, it relied solely on environmental variables, omitting biotic interactions (e.g., pollination, seed dispersal) and anthropogenic pressures, which also influence species distribution. Second, potential sampling biases in occurrence data and spatial resolution limitations in the environmental layers may have introduced predictive inaccuracies. Moreover, model interpretability particularly with GBM remains a challenge, which can hinder the ecological transparency of predictions. In conclusion, this research contributes to the growing body of knowledge advocating ML for ecological modeling. It underscores the utility of advanced algorithms in supporting evidence-based conservation planning. Future work should aim to integrate additional ecological and socio-environmental variables, employ bias correction techniques, and explore interpretable AI approaches to strengthen both the predictive and explanatory power of SDMs.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Peter E. Ugege, Omokafe A. Ugbogu, Jean C. Ganglo

Acquisition, analysis, or interpretation of data: Peter E. Ugege, Peace A. Ukoha, Omokafe A. Ugbogu, Jean C. Ganglo

Drafting of the manuscript: Peter E. Ugege

Critical review of the manuscript for important intellectual content: Peter E. Ugege, Peace A. Ukoha, Omokafe A. Ugbogu, Jean C. Ganglo

Supervision: Peter E. Ugege

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Imoisi C, Iyasele JU, Okhale SA: Phytochemical and antioxidant capability of Vitex doniana (black plum) fruit. Journal of Chemical Society of Nigeria. 2021, 46:0191-0196. 10.46602/jcsn.v46i1.589
2. Kamal N, Mio Asni NS, Rozlan INA, et al.: Traditional medicinal uses, phytochemistry, biological properties, and health applications of Vitex sp. Plants. 2022, 11:1944. 10.3390/plants11151944
3. Hounkpèvi A, Azihou AF, Kouassi ÉK, Porembski S, Kakaï RG: Climate-induced morphological variation of black plum (Vitex doniana Sw.) in Benin, West Africa. Genetic Resources and Crop Evolution. 2016, 63:1073-1084. 10.1007/s10722-016-0409-9
4. N'Danikou S, Achigan-Dako EG, Tchokponhoue DA, Agossou COA, Houdegbe CA, Vodouhe RS, Ahanchede A: Modelling socioeconomic determinants for cultivation and in-situ conservation of Vitex doniana Sweet (Black

plum), a wild harvested economic plant in Benin. Journal of Ethnobiology and Ethnomedicine. 2015, 11:28. 10.1186/s13002-015-0017-3

5. Soulémane N, N'Danikou S, Tchokponhoué DA, Ekué MR, Achigan-Dako EG: Production, domestication and genetic improvement of Vitex doniana Sweet: An overview. Genetic Resources and Crop Evolution. 2024, 72:1271-1287. 10.1007/s10722-024-02012-4

6. Pichler M, Hartig F: Machine learning and deep learning—A review for ecologists. Methods in Ecology and Evolution. 2023, 14:994-1016. 10.1111/2041-210x.14061

7. Bellin N, Tesi G, Marchesani N, Rossi V: Species distribution modeling and machine learning in assessing the potential distribution of freshwater zooplankton in Northern Italy. Ecological Informatics. 2022, 69:101682. 10.1016/j.ecoinf.2022.101682

8. Ahmad S, Yang L, Khan TU, Wanghe K, Li M, Luan X: Using an ensemble modelling approach to predict the potential distribution of Himalayan gray goral (Naemorhedus goral bedfordi) in Pakistan. Global Ecology and Conservation. 2020, 21:e00845. 10.1016/j.gecco.2019.e00845

9. Norberg A, Abrego N, Blanchet FG, et al.: A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecological Monographs. 2019, 89:e01370. 10.1002/ecm.1370

10. Sittaro F, Hutengs C, Vohland M: Which factors determine the invasion of plant species? Machine learning based habitat modelling integrating environmental factors and climate scenarios. International Journal of Applied Earth Observation and Geoinformation. 2023, 116:103158. 10.1016/j.jag.2022.103158

11. Tan C, Ferguson DK, Yang Y: Impacts of distribution data on accurate species modeling: A case study of Litsea auriculata (Lauraceae). Plants. 2024, 13:2581. 10.3390/plants13182581

12. Sofaer HR, Jarnevich CS, Pearse IS, et al.: Development and delivery of species distribution models to inform decision-making. BioScience. 2019, 69:544-557. 10.1093/biosci/biz045

13. Worldatlas. (2024). Accessed: January 12, 2025: https://www.worldatlas.com/articles/ecological-regions-of-benin.html?utm_source=chatgpt.com.

14. FAO. (2022). Accessed: January 10, 2025: https://openknowledge.fao.org/items/4c8bd12f-d6b8-4755-a82f-1284c41bf012.

15. Kew Science. (2023). Accessed: January 12, 2025: https://powo.science.kew.org/taxon/865694-1?utm_source=chatgpt.com.

16. GBIF. (2024). Accessed: January 12, 2025: https://www.gbif.org.

17. Elith J, Franklin J: Species distribution modeling. Encyclopedia of Biodiversity (Second Edition). Levin SA (ed): Academic Press, Cambridge, MA; 2013. 692-705. 10.1016/B978-0-12-384719-5.00318-X

18. Parker KC: Topography, substrate, and vegetation patterns in the northern Sonoran Desert. Journal of Biogeography. 1991, 18:151-163. 10.2307/2845289

19. Franklin J: Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography Earth and Environment. 1995, 19:474-499. 10.1177/030913339501900403

20. Guisan A, Zimmermann NE: Predictive habitat distribution models in ecology. Ecological Modelling. 2000, 135:147-186. 10.1016/s0304-3800(00)00354-9

21. ESA. (2023). Accessed: January 13, 2025: https://www.esa-landcover-cci.org/.

22. Wilson JP, Gallant JC: Terrain Analysis: Principles and Applications. John Wiley & Sons, New York; 2000.

23. Farr TG, Rosen PA, Caro E, et al.: The shuttle radar topography mission. Reviews of Geophysics. 2007, 45:RG2004. 10.1029/2005rg000183

24. Elith J, Leathwick JR, Hastie T: A working guide to boosted regression trees. Journal of Animal Ecology. 2008, 77:802-813. 10.1111/j.1365-2656.2008.01390.x

25. Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J: Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. Ecological Monographs. 2022, 92:e01486. 10.1002/ecm.1486

26. James G, Witten D, Hastie T, Tibshirani R: An Introduction to Statistical Learning: With Applications in R (2nd ed.). Sringer, New York; 2021. 10.1007/978-1-0716-1418-1

27. Guisan A, Thuiller W, Zimmermann NE: Habitat Suitability and Distribution Models: With Applications in R. Cambridge University Press, Cambridge, UK; 2017. 10.1017/9781139028271

28. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B: Support vector machines. IEEE Intelligent Systems and Their Applications. 1998, 13:18-28. 10.1109/5254.708428

29. Zhang P, Zhang H, Wang H, et al.: Warming alters juvenile carp effects on macrophytes resulting in a shift to turbid conditions in freshwater mesocosms. Journal of Applied Ecology. 2022, 59:165-175. 10.1111/1365-2664.14040

30. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB: Ecological Niches and Geographic Distributions. Princeton University Press, Princeton, NJ; 2011.

31. Chen T, Guestrin C: XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 785-794. 10.1145/2939672.2939785

32. Fawcett T: An introduction to ROC analysis. Pattern Recognition Letters. 2006, 27:861-874. 10.1016/j.patrec.2005.10.010

33. Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, et al.: Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography. 2015, 24:276-292. 10.1111/geb.12268