



Harnessing machine learning to predict habitat suitability of medicinal plants: insights from *Oliveria decumbens* under environmental change

Emran Dastres¹ · Hamidreza Rabiei-Dastjerdi² · Hassan Esmaeili¹ · Mahdis Amiri³ · Akbar Karami⁴ · Mehdi Gheisari^{5,6,7,8}

Received: 26 February 2025 / Revised: 29 May 2025 / Accepted: 30 May 2025 / Published online: 29 June 2025
© The Author(s) 2025

Abstract

Accurate habitat suitability modeling is essential for the conservation and sustainable cultivation of medicinal plants, particularly under changing environmental conditions. This study evaluates the habitat suitability of *Oliveria decumbens* in Fars Province, Iran, using five machine learning algorithms—Naive Bayes, Quadratic Discriminant Analysis, Linear Discriminant Analysis, Generalized Additive Models, and eXtreme Gradient Boosting. A total of 168 occurrence points and 18 environmental predictors, including topographic, climatic, and edaphic factors, were analyzed. Feature selection using Boruta and LASSO methods identified annual rainfall, elevation, and soil silt content as the most influential variables. QDA and LDA models outperformed others in spatial cross-validation, with AUC scores of 0.90 and 0.89, respectively, demonstrating high predictive accuracy and low uncertainty. Suitability maps indicated that optimal habitats are concentrated in the northern and northwestern regions of the province. Despite the robustness of the models, limitations related to environmental extrapolation and landscape dynamics remain. The findings provide a data-driven foundation for conservation planning and sustainable cultivation of *O. decumbens*. Future research should incorporate climate and land-use change projections to improve long-term resilience. This study demonstrates how integrating machine learning with spatial data supports evidence-based decision-making in ecological modeling.

Keywords Habitat suitability modeling · *O. decumbens* · Machine learning algorithm · GIS · Environmental

1 Introduction

Oliveria decumbens Vent., a member of the Apiaceae family and commonly known as Moshkoorak, Den, or Denak in Iran, is an annual medicinal and aromatic plant (MAP) with a natural distribution across Iran, Syria, Iraq, and southeastern Anatolia [1]. Traditionally, *O. decumbens* has

been utilized for treating ailments such as fever, diarrhea, dyspepsia, and stomach pain, highlighting its value in traditional medicine [2]. The plant's essential oils are rich in phenolic compounds—particularly carvacrol and thymol—which impart significant biological activities, including antioxidant, anticancer, insecticidal, antibacterial, antidiabetic, antiviral, and acetylcholinesterase-inhibitory properties

✉ Hamidreza Rabiei-Dastjerdi
Hamid.rabiei@dcu.ie; hamid.rabiei555@gmail.com

✉ Hassan Esmaeili
h_esmaeili@sbu.ac.ir; h.esmaili_6007@yahoo.com

¹ Department of Agriculture, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, 1983969411, Tehran, Iran

² School of History and Geography, Faculty of Humanities and Social Sciences, Dublin City University, Dublin, Ireland

³ Department of Watershed and Arid Zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

⁴ Department of Horticultural Science, School of Agriculture, Shiraz University, Shiraz, Iran

⁵ Institute of Artificial Intelligence, Shaoxing University, Zhejiang, China

⁶ Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Chennai, Tamil Nadu, India

⁷ Department of Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran

⁸ Department of R&D, Shenzhen BKD Co LTD, Shenzhen, China

[3]. These bioactive characteristics render *O. decumbens* a highly sought-after species in the food, feed, cosmetic, and pharmaceutical industries. However, the limited geographic distribution of *O. decumbens* and growing industrial demand threaten its availability, creating a strong need to explore sustainable production methods to meet these demands without further compromising its natural habitats.

Wild populations currently provide approximately 75% of MAPs, which contributes to these valuable resources' depletion along with habitat destruction and ecosystem degradation [4]. The unregulated harvesting of MAPs from their native environments poses a serious risk of extinction for many species. Therefore, sustainable cultivation of *O. decumbens* is essential to meet industrial requirements and protect the plant's natural habitats from overharvesting pressures. To support such initiatives, it is necessary to develop accurate and reliable methods for habitat suitability modeling (HSM), which can help identify optimal sites for cultivating *O. decumbens* and promote sustainable agricultural practices [5].

Advances in artificial intelligence (AI) and machine learning (ML) have opened new possibilities for habitat suitability modeling, providing powerful tools for analyzing the environmental variables that influence the distribution of MAPs [6]. Recent applications of machine learning and remote sensing have significantly enhanced habitat predictions and risk assessments for various species [7]. Recent applications of ML and AI have successfully supported habitat predictions and risk assessments for various MAPs and other species [8]; for instance, extreme gradient boosting has been applied to assess habitat suitability for cassava, sorghum, and maize, while Bayesian network models have proven effective in complex environmental data analysis and problem-solving [9]. In particular, the Naïve Bayes (NB) model has performed reliably across diverse scenarios despite its structural simplicity [9]. Other models, such as quadratic discriminant analysis (QDA), generalized additive models (GAM), and linear discriminant analysis (LDA), have also been applied effectively in the field of habitat suitability modeling [10]. Moreover, random forest (RF) algorithms have been used to predict habitat vulnerability for *Posidonia oceanica* meadows, reinforcing the utility of ML for assessing the environmental parameters that define suitable habitats for plant species [11].

While these studies highlight the growing potential of ML in habitat suitability modeling, there is still a need for further research to apply these techniques specifically to identify optimal growth conditions for *O. decumbens*, a species that has not yet been extensively studied in this context. Additionally, there remains a gap in understanding how multiple environmental variables interact to influence the habitat suitability of MAPs in diverse geographical regions. This study aims to address these gaps by using a range of ML

algorithms to predict habitat suitability for *O. decumbens*, offering insights into sustainable cultivation practices and conservation strategies. Sustainable practices in managing medicinal and aromatic plants are essential for balancing industrial demands with conservation goals [12].

In this study, we apply ML algorithms, including NB, QDA, GAM, LDA, and Extreme Gradient Boosting (XGB), to predict the habitat suitability of *O. decumbens* based on an array of environmental factors such as elevation, mean annual temperature, mean annual rainfall, slope aspect, slope degree, profile curvature, plan curvature, distance from rivers, distance from roads, clay percent, sand percent, silt percent, pH, electrical conductivity, nitrogen, phosphorus, potassium, and organic matter. The primary research questions of this study are: (1) Which environmental factors have the greatest influence on the habitat suitability of *O. decumbens*? (2) How accurately can ML algorithms predict habitat suitability for *O. decumbens* in different regions? (3) What is the potential for identifying optimal sites for sustainable cultivation of *O. decumbens*? This research aims to support sustainable cultivation initiatives for *O. decumbens* by identifying optimal sites for its growth, contributing to its conservation, and ensuring a steady supply of raw materials for various industries.

2 Materials and methods

2.1 Study area

The research region is in southwestern Iran between latitudes 28° 00' and 30° 00' N and longitudes 51° 00' and 54° 00' E within the Fars province (Fig. 1). This vast province is home to various animal and plant species due to its climatic and ecosystem diversity [13]. A dynamic environmental gradient is created between the arid and semi-arid areas of the plains and the Zagros Mountains because of their temperate climate and rich biodiversity [14]. This climatic diversity has led to the growth of various medicinal species in this province, such as *O. decumbens*, which grows in limited areas.

2.2 Methodology

In this study, the distribution areas of the *O. decumbens* plant were first identified throughout the province. Then, the geographical characteristics of 168 plant growth areas, including their elevation, longitude, and latitude, were recorded using a GPS device. These 168 sites were carefully selected to encompass the full range of environmental and geographical conditions in which the plant is known to grow. By covering diverse habitats, including variations in topography, climate, and soil conditions, the selected sites represent a comprehensive sampling of the species' distribution across

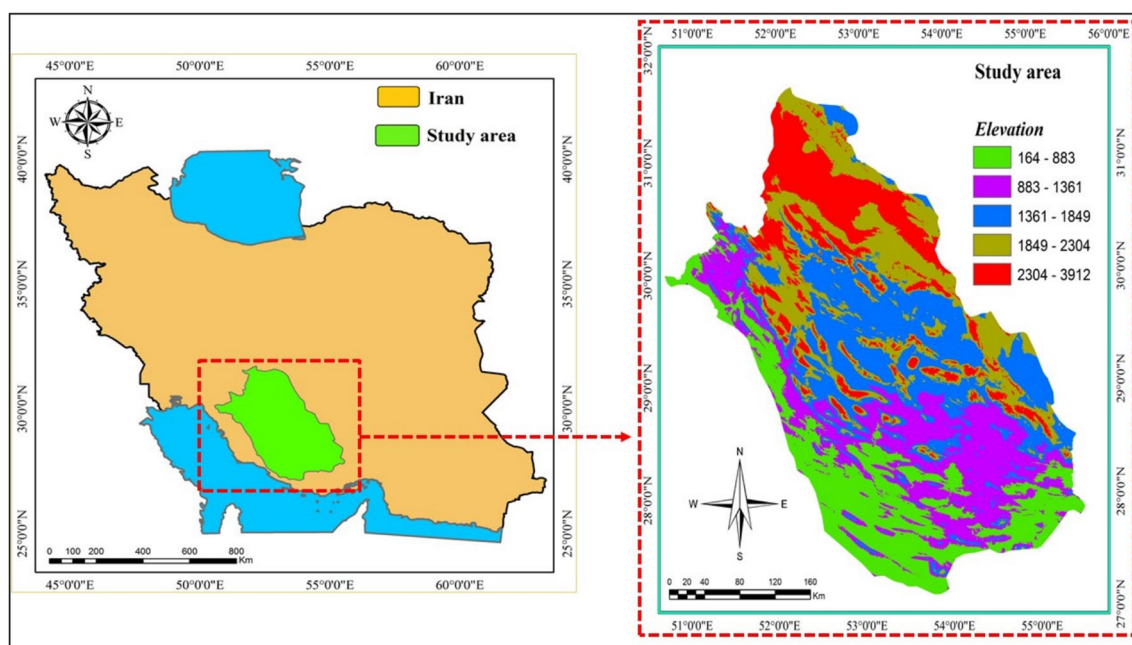


Fig. 1 The study area for habitat suitability assessment of *Oliveria decumbens* Vent

its natural range. This approach ensures that the dataset captures the variability of the plant's ecological preferences and provides a robust basis for modeling its habitat suitability with machine learning techniques.

Meteorological information such as average annual precipitation and temperature was obtained from the 25 nearest synoptic stations to the plant growth areas. These factors were selected because temperature and precipitation are among the most critical climatic variables influencing the growth, distribution, and habitat suitability of many plant species, including medicinal and aromatic plants (MAPs). Precipitation and temperature not only affect the physiological processes of plants but also determine the presence of water resources, which is essential for the plant's survival and reproduction [15]. Using data from the nearest synoptic stations, we ensured that the climatic conditions surrounding the selected plant growth areas were accurately represented, providing a reliable basis for understanding the environmental conditions that influence *O. decumbens* (Fig. 2a, b).

Additionally, soil samples were collected from each wild-growing habitat at depths of 0–30 cm from all 168 sites. This depth was chosen because it represents the primary rooting zone for many plant species, including *O. decumbens*, where nutrient uptake and water absorption are critical. The physical properties of the soil, including percentages of sand, silt, and clay, were analyzed using the hydrometer method. These properties were selected because they are crucial in determining soil texture, water retention, drainage, and fertility—key factors influencing plant growth. Soil texture, in particular, affects the ability of the soil to retain moisture

and nutrients, which can significantly impact the distribution and productivity of the species [16]. By incorporating these soil properties, we can account for the variations in habitat suitability from different soil conditions, ensuring a comprehensive assessment of the environmental factors influencing *O. decumbens* growth.

The soil pH, electrical conductivity (EC), and organic matter (OM) of the samples were recorded according to Regasa method [17]. Nitrogen (N) content was determined using the Kjeldahl method, in which the soil samples are digested with sulfuric acid to convert organic nitrogen into ammonium sulfate [18]. Phosphorus (P) was measured with the Olsen method [19]. Potassium (K) was extracted using ammonium acetate and measured with a flame photometer [20]. Soil data were analyzed using the Inverse Distance Weighting (IDW) interpolation algorithm and converted into 30-m resolution raster maps within a GIS environment, to be used in subsequent modeling steps (Fig. 3a–i) [21].

The slope degree, plan and profile curvature, slope aspect, and elevation layer maps were obtained from a 30-m spatial-resolution digital elevation model (DEM). The inverse distance weighting (IDW) algorithm was applied to convert the mean annual temperature and rainfall data into 30-m-resolution raster maps. Using geographic information system (GIS) software, the Euclidean distance (ED) algorithm was employed to generate raster maps indicating the distances from rivers and roads (Fig. 4a–g).

GIS and spatial analysis tools have been critical in modeling ecological distributions, as demonstrated in recent studies [22]. To assess the multi-collinearity among independent

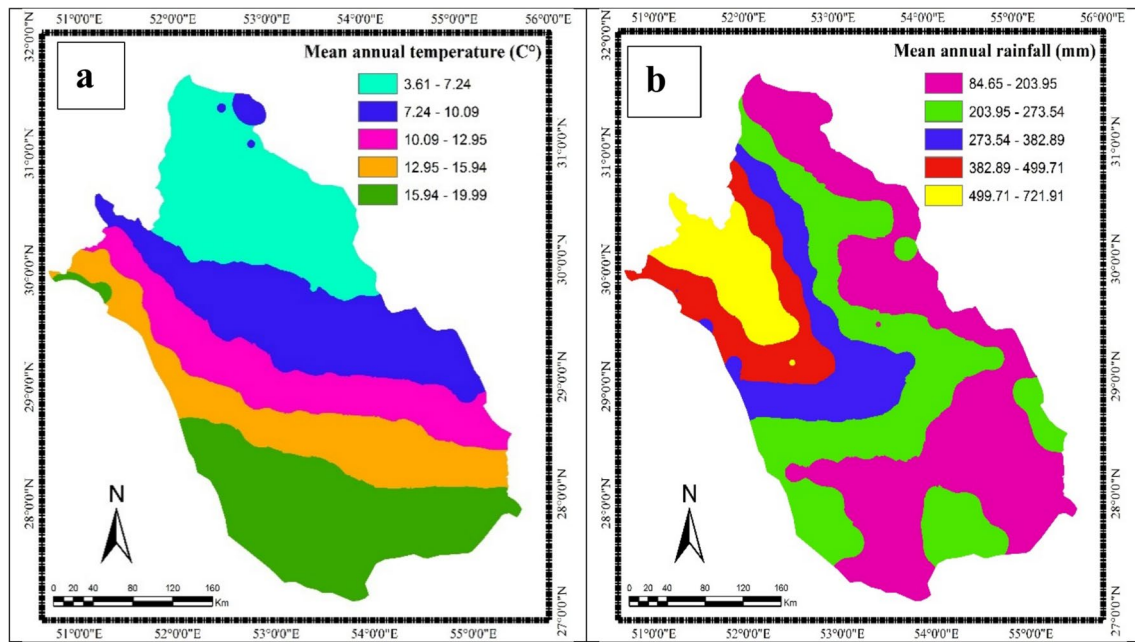


Fig. 2 Raster maps of climatic factors used in the modeling process: **a** Mean annual temperature, **b** Mean annual rainfall

variables, the tolerance (TOL) and variance inflation factor (VIF) collinearity tests were employed. A comprehensive dataset of environmental factors and locations of *O. decumbens* was then provided.

The LASSO (Least Absolute Shrinkage and Selection Operator) and Boruta algorithms were utilized to identify the factor with the greatest impact value on the modeling process. The LASSO algorithm is a regression analysis technique used in machine learning and statistics for variable selection and regularization [23]. It enhances the prediction accuracy and interpretability of the statistical model. The Boruta algorithm is a feature selection method specifically designed to identify all features that carry useful information for prediction [24]. The Boruta algorithm is a powerful feature selection tool that identifies all relevant features in a dataset, leveraging the strengths of random forests to provide robust and comprehensive insights into feature importance.

The five algorithms, including GAM, LDA, NB, QDA, and XGB, were employed to assess the habitat suitability of *O. decumbens*. One of the non-parametric statistical models is the generalized additive model (GAM), which evolved from the development and expansion of generalized linear models (GLMs) [25]. Machine learning algorithms have been increasingly adopted in environmental and habitat modeling, offering robust tools for analyzing complex ecological systems. Unlike GLMs, where the relationship between independent and dependent variables is expressed through regression equations and coefficients, GAMs allow the data to dictate the shape of the response curve [26]. Linear discriminant analysis (LDA), a statistical method for

dimension reduction and classification, maximizes the ratio of between-group to within-group dispersion. This technique is employed in machine learning to find the linear combination of features that best separates multiple classes. Another significant advancement is the eXGB algorithm, introduced by, also known as the regular reinforcement method. Renowned for its high predictive power, eXGB is considered the best option for accuracy across various applications. This algorithm leverages both linear models and tree learning algorithms, offering a performance approximately ten times faster than traditional gradient boosting methods [27]. Naive Bayes (NB) is a family of simple yet effective probabilistic classifiers based on Bayes' theorem. Despite its simplicity, it often performs surprisingly well for various complex tasks, particularly in text classification and natural language processing. Naive Bayes remains a cornerstone algorithm in machine learning, particularly due to its interpretability and efficiency in high-dimensional datasets [28]. Quadratic discriminant analysis (QDA) is a statistical technique used for classification. It is a variant of linear discriminant analysis (LDA). Still, unlike LDA, which assumes that the covariance matrices of different classes are identical, QDA allows each class to have its own covariance matrix. This flexibility enables QDA to model more complex class boundaries [29]. Finally, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) were used to assess the effectiveness of the models.

To predict the suitability of *O. decumbens*, a systematic workflow was developed as outlined in Figure 5. The methodological steps are described below:

Fig. 3 Raster maps of soil-related factors used in the modeling process: **a** Clay percent, **b** Sand percent, **c** Silt percent, **d** pH, **e** Electrical conductivity (EC), **f** Nitrogen (N), **g** Phosphorus (P), **h** Potassium (K), **i** Organic matter (OM)

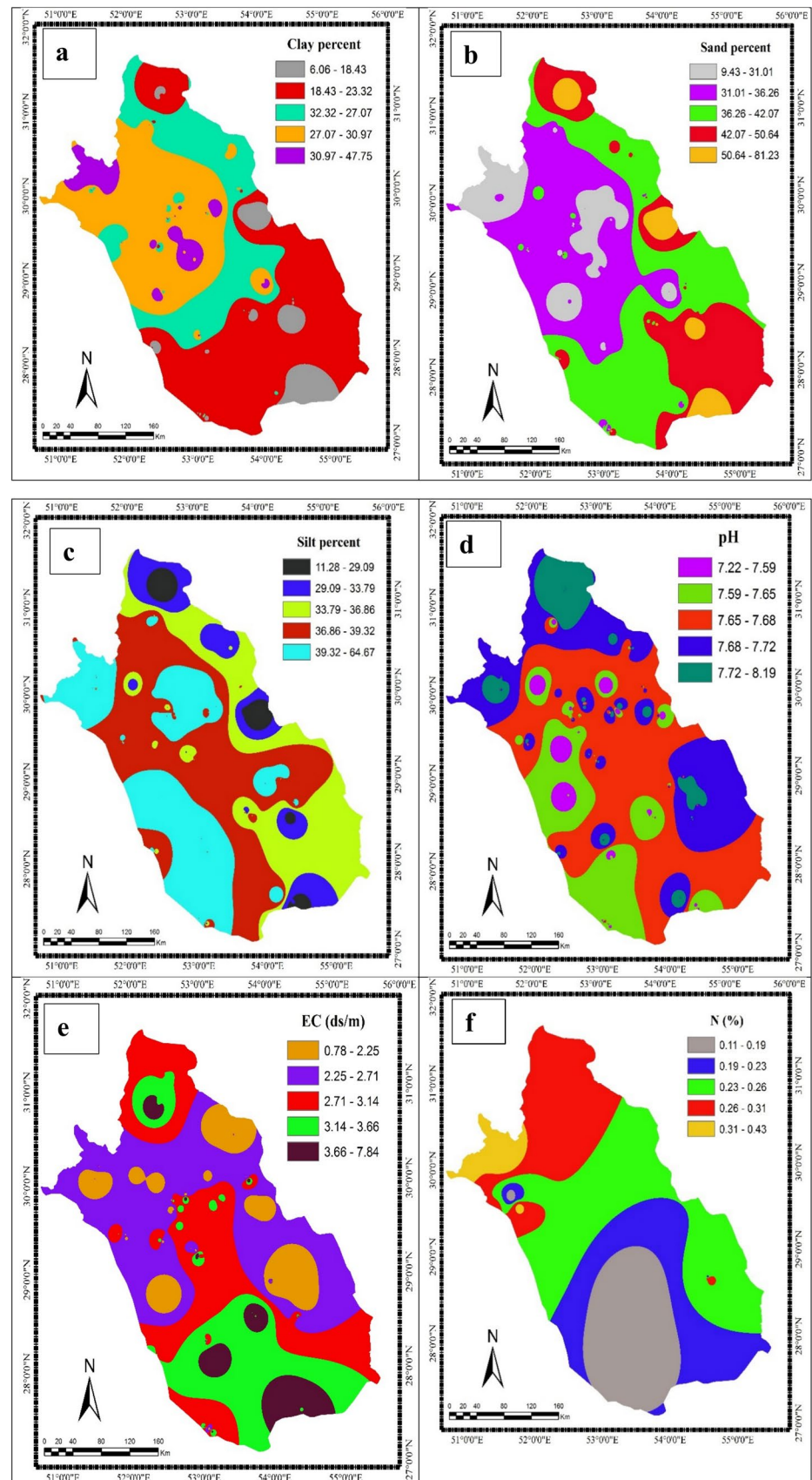
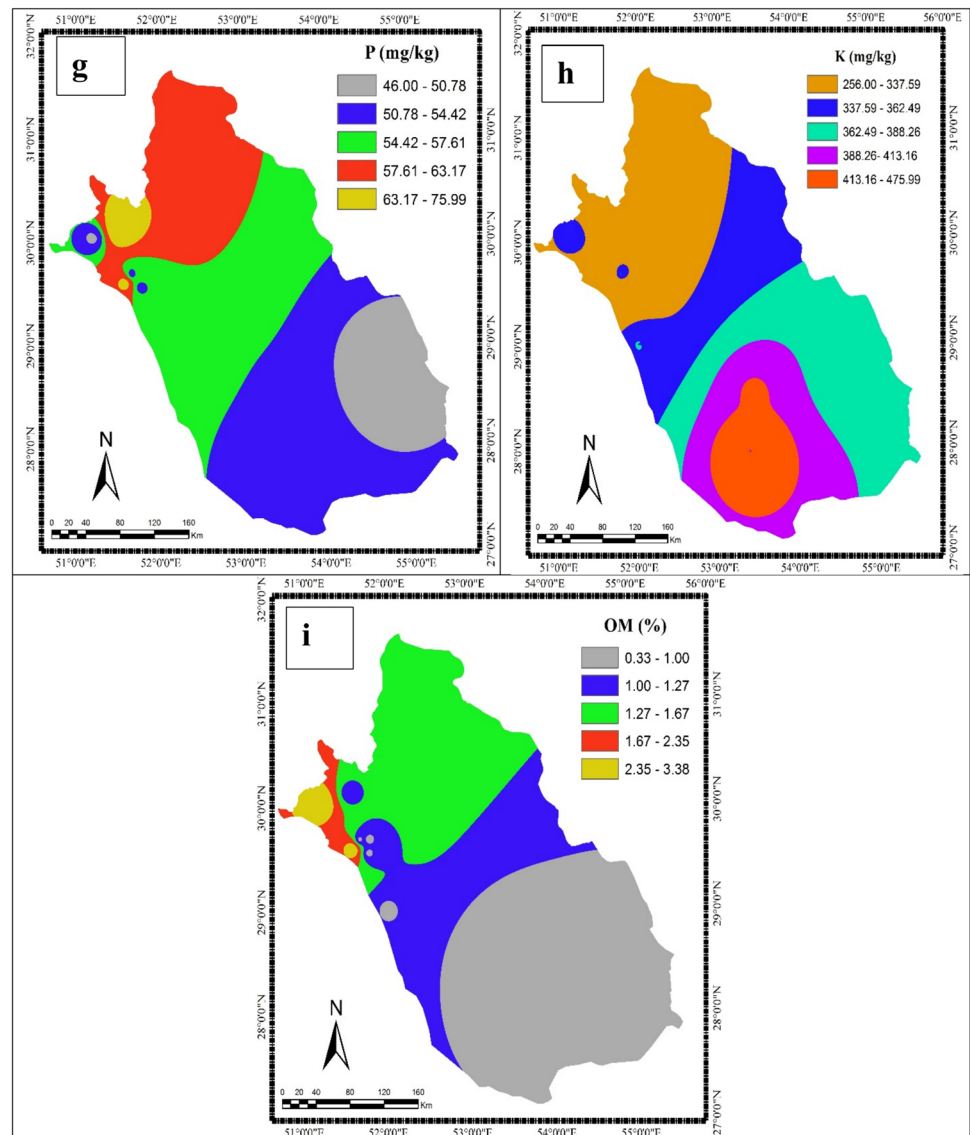


Fig. 3 (continued)



(a) Plant sample data collection

The recorded geographical characteristics of 168 plant growth areas—including elevation, longitude, and latitude—served as the basis for this study. These samples were randomly divided into two subsets: 70% (118 points) for model training and 30% (50 points) for validation, ensuring generalizability. In addition to the presence data, 118 pseudo-absence points were randomly generated from areas outside the known distribution range of *O. decumbens*, ensuring spatial separation from presence records. This approach created a balanced and spatially independent presence/absence dataset.

To reduce potential spatial bias and avoid overfitting to known hotspots, pseudo-absence points were spatially stratified across environmentally diverse areas of Fars Province. Furthermore, to statistically assess the presence of spatial

autocorrelation, we performed a Moran's I analysis on the predicted occurrence probabilities across all sample locations. The observed Moran's I was 0.041, with a Z-score of 1.29 and a p value of 0.197, indicating no significant spatial autocorrelation (Fig. 6, Table 1). These findings confirm the spatial reliability of our sampling design and modeling dataset, supporting robust performance across the five algorithms employed (NB, QDA, LDA, eXGB, and GAM).

(b) Predictive environmental factors

Eighteen predictive factors were considered (Table 2), representing key environmental and soil properties that influence habitat suitability [30]. These factors included:

1. Climatic variables: Average annual temperature and precipitation.

2. Soil physical and chemical properties: Sand, silt, and clay percentages; pH; electrical conductivity (EC); nitrogen (N); phosphorus (P); potassium (K); and organic matter (OM).
3. Topographical features: Elevation, slope degree, slope aspect, plan curvature, and profile curvature.
4. Proximity indices: distance from rivers and roads.
5. Variable screening and multi-collinearity check

To ensure the predictive factors were not collinear, a variance inflation factor (VIF) analysis was performed. This step identified and excluded variables with high multi-collinearity, enhancing model stability and interpretability.

(d) Model selection and habitat suitability mapping

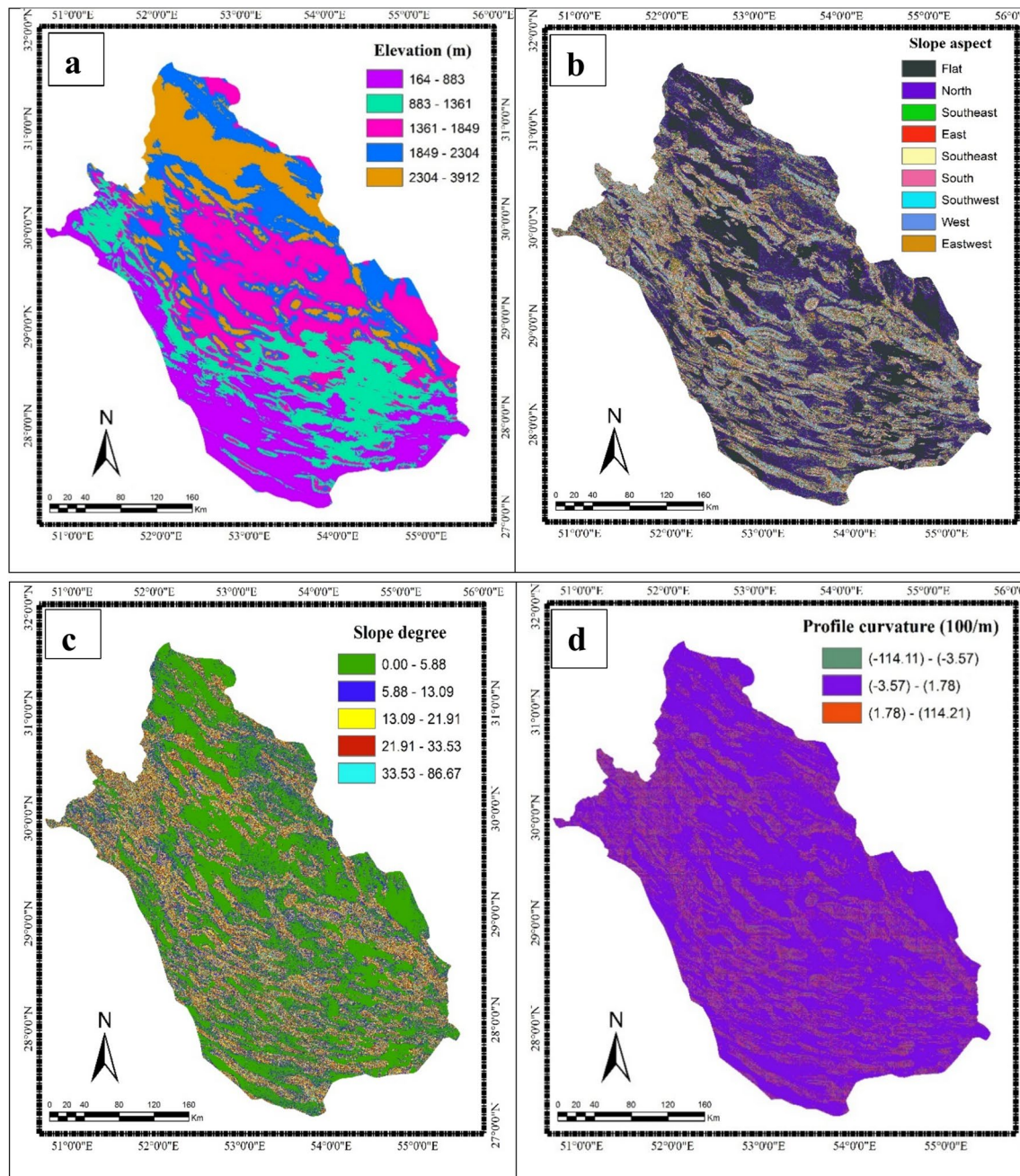


Fig. 4 Raster maps of topographic and proximity factors used in the modeling process: **a** Elevation, **b** Slope degree, **c** Slope aspect, **d** Profile curvature, **e** Plan curvature, **f** Distance from rivers, **g** Distance from roads

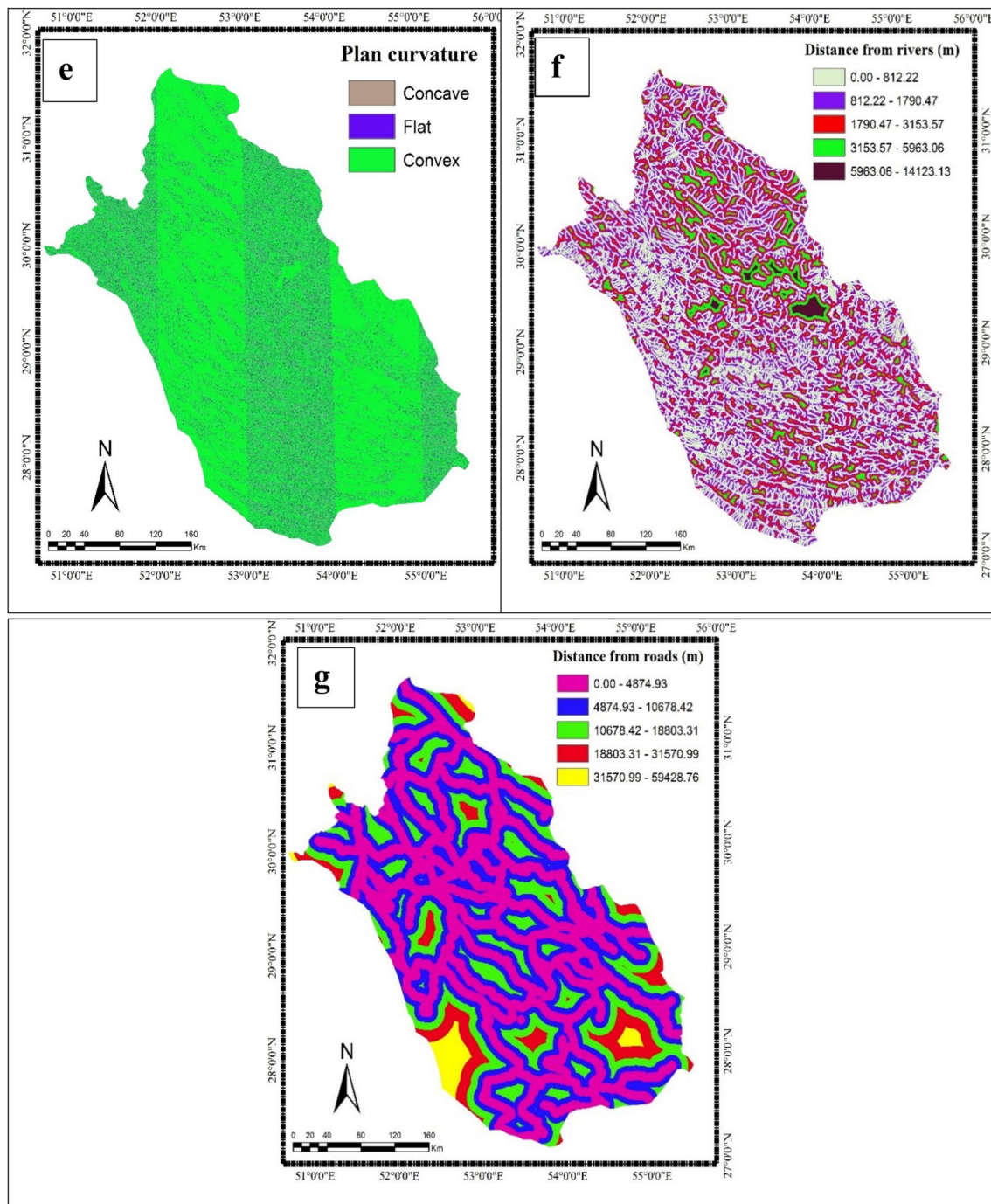


Fig. 4 (continued)

Five machine learning algorithms—Naïve Bayes (NB), quadratic discriminant analysis (QDA), generalized additive models (GAM), linear discriminant analysis (LDA), and extreme gradient boosting (XGB)—were employed to predict habitat suitability. These models were selected for their established performance in ecological modeling and ability to capture complex relationships between predictors and species distribution.

(e) Accuracy assessment

A total of 50 presence points (based on confirmed occurrences of *O. decumbens*) and 50 pseudo-absence points (randomly sampled from outside known presence areas) were spatially stratified across the blocks to represent the full environmental and geographic variability of the region. The random allocation of pseudo-absence points helped prevent

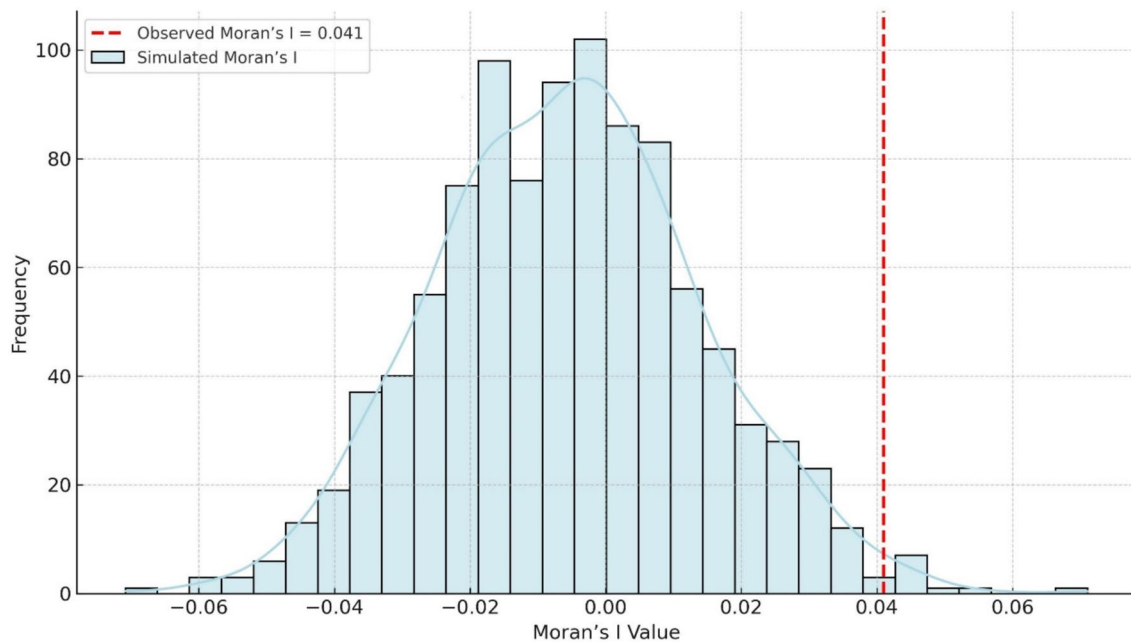
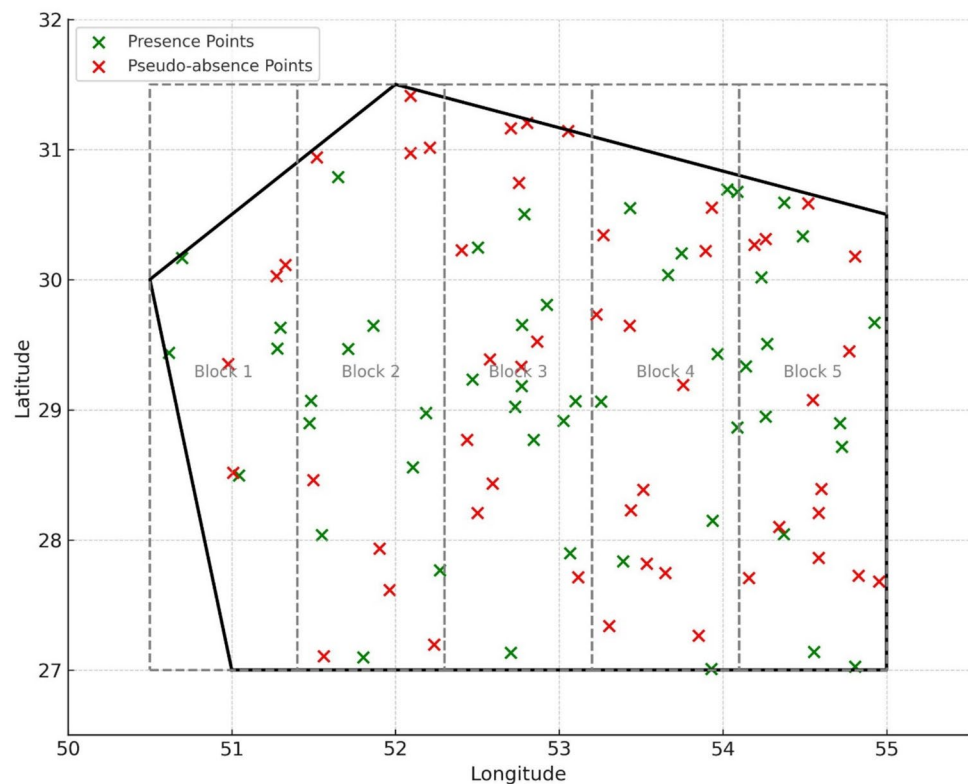


Fig. 5 Diagram displaying the stages involved in the modeling procedure

Fig. 6 Monte Carlo simulation of Moran's I test for model-predicted occurrence values across sample points



sampling bias and reduced the chance of the model overfitting to known species hotspots.

A visual representation of this spatial partitioning is shown in Fig. 7, where red points indicate pseudo-absences and green points represent known species presences. The

dashed lines show the borders of the five cross-validation blocks. This setup minimizes the risk of spatial leakage, allowing for a more reliable evaluation of model generalization in unsampled areas.

Table 1 Summary of Moran's I test results for predicted occurrence values

Metric	Value
Observed Moran's I	0.041
Expected Moran's I	− 0.006
Z-score	1.23
P value	0.218
Interpretation	No significant spatial autocorrelation

For each model and fold, performance metrics including the area under the curve (AUC) and the receiver operating characteristic (ROC) were calculated. The mean AUC and its standard deviation across folds were then used to evaluate the spatial extrapolation capability of the models. The model with the highest mean AUC and lowest variability was considered the most robust and suitable for habitat suitability.

(f) Feature importance analysis

Feature selection techniques, such as LASSO regression and the Boruta algorithm, were applied to identify the most influential factors in determining habitat suitability. These methods quantified the contribution of each variable to the modeling process, providing insights into the ecological drivers of *O. decumbens* distribution.

(g) Habitat suitability maps

Based on the selected model, habitat suitability maps were developed, highlighting the most suitable areas for *O. decumbens*. These maps provide a practical tool for guiding conservation efforts and optimizing sustainable cultivation practices.

3 Result

3.1 Evaluating multi-collinearity in habitat suitability models

In this study, we assessed the dataset for the possibility of multi-collinearity among the variables using the variance inflation factor (VIF) and tolerance (T) index. Multi-collinearity occurs when independent variables in a regression model are highly correlated, resulting in unreliable estimates of regression coefficients, increased standard errors, and difficulty in determining the individual effects of each predictor. It is crucial to detect and address multi-collinearity to ensure the stability and interpretability of the model, as high correlation among predictors can distort the results and affect model performance. The VIF and Tolerance indices are widely used to quantify the degree of multi-collinearity, with high VIF values indicating potential multi-collinearity among variables. Addressing this issue is essential to ensure that the selected predictive factors contribute independently to the model, improving the overall accuracy and reliability of the habitat suitability predictions. As shown in Table 3, all studied variables in this study had VIF and T values within the acceptable ranges, confirming the absence

Table 2 Factors influencing *O. decumbens* and their justification

Factors	Description
Climatic variables	1. Average annual temperature 2. Average annual precipitation Average annual temperature influences species' growth, survival, and distribution. Precipitation determines water availability, affecting vegetation, soil quality, and ecosystem productivity [31]
Soil variables	1. Silt percentage 2. Sand percentage 3. Clay percentages 4. pH 5. Electrical conductivity (EC) 6. Nitrogen (N) 7. Potassium (K) 8. Phosphorus (P) 9. Organic matter (OM) Soil texture impacts fertility: silt retains water and nutrients, enhancing fertility, while sandy soils drain quickly and retain fewer nutrients. Clay soils have high nutrient retention but poor drainage. Soil pH affects nutrient availability; neutral pH (6–7) is optimal. Electrical conductivity (EC) indicates salinity levels that can reduce fertility. Nitrogen is vital for plant growth and soil fertility, while potassium aids water uptake and plant health. Phosphorus is essential for root, and organic matter improves soil structure and microbial activity [32, 33]
Topographical features	1. Elevation 2. Slope degree 3. Slope aspect 4. Plan curvature 5. Profile curvature Elevation affects temperature, precipitation, and soil properties, with higher elevations showing cooler climates and reduced nutrients. Slope steepness influences erosion and fertility, while slope aspect affects microclimates and evaporation rates. Curvature (both plan and profile) influences water flow and soil moisture [34]
Proximity indices	1. Distance from rivers 2. Distance from roads Proximity to rivers affects erosion, sediment, and vegetation distribution, while roads influence local ecosystems and vegetation diversity [35, 36]

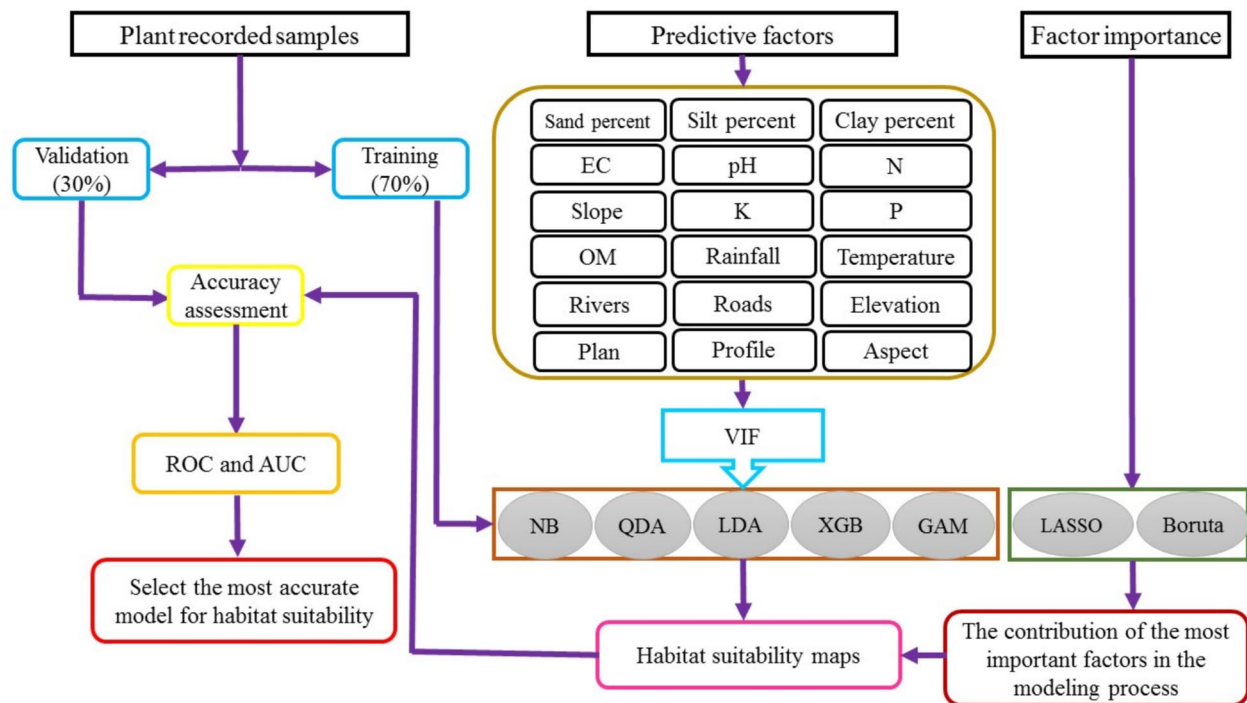


Fig. 7 Spatial distribution of presence (green) and pseudo-absence (red) sample points across five spatial blocks in Fars Province, Iran

Table 3 Multi-collinearity analysis of factors influencing habitat suitability using variance inflation factor (VIF) and tolerance index (TI)

Factors affecting habitat suitability	Collinearity statistics*	
	Tolerance	VIF
Slope aspect	0.24	4.05
Slope degree	0.54	1.95
Elevation (m)	0.32	4.01
Profile curvature (100/m)	0.65	1.34
Plan curvature (100/m)	0.19	4.89
Silt (%)	0.22	4.15
Sand (%)	0.26	3.89
Clay (%)	0.13	4.96
pH	0.43	3.26
EC (ds/m)	0.18	4.85
N (%)	0.76	1.11
K (mg/kg)	0.25	3.81
P (mg/kg)	0.18	4.85
OM (%)	0.16	4.92
Mean annual temperature (°C)	0.35	4.76
Mean annual rainfall (mm)	0.49	3.15
Distance from roads (m)	0.28	3.49
Distance from rivers (m)	0.51	2.18

*The VIF > 5 and T < 0.1 indicate high multi-collinearity

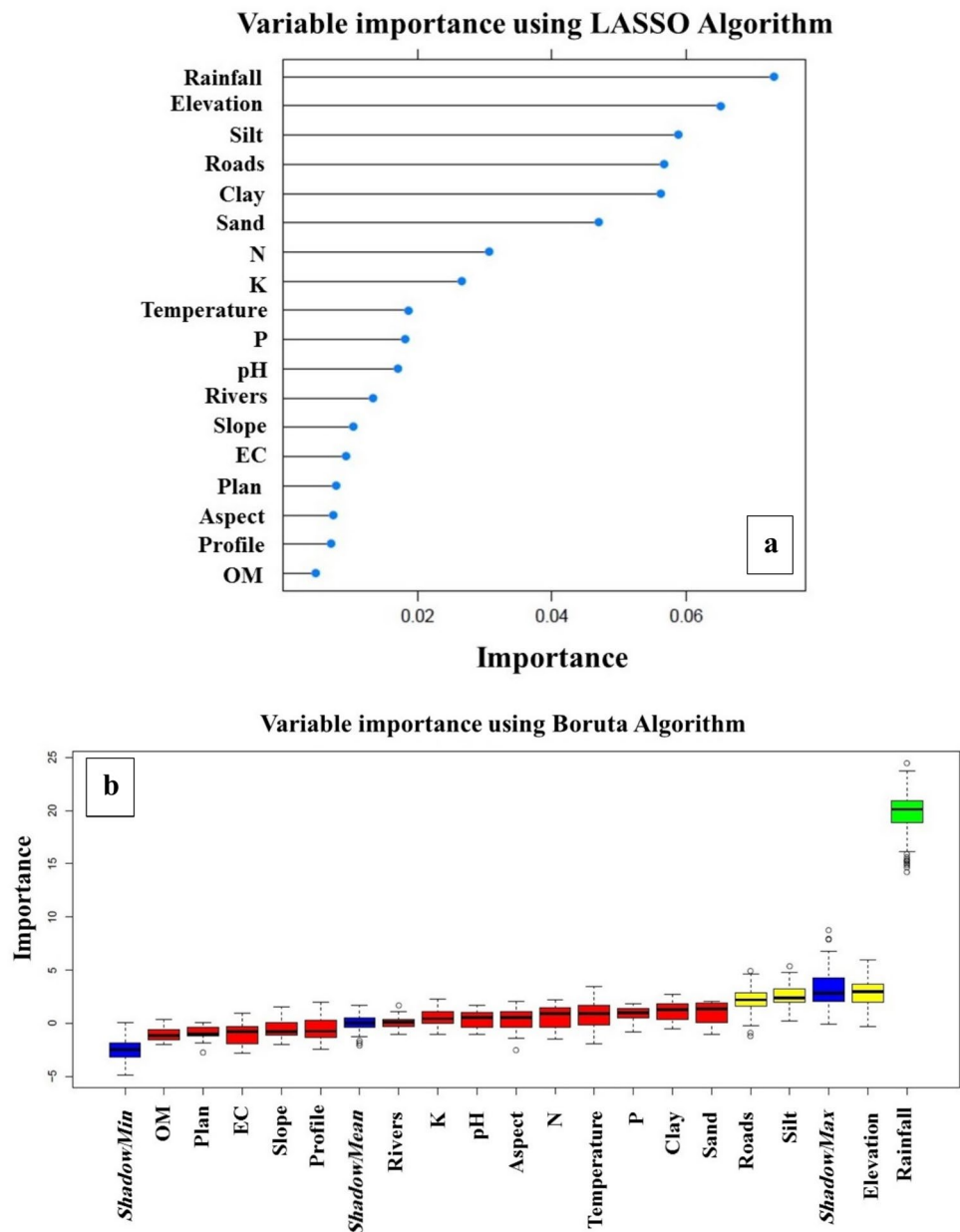
of multi-collinearity. For example, although the Clay (%) variable had the highest VIF (4.96), it was still within an acceptable range (VIF < 5).

3.2 Evaluating the key factors and analyzing partial response curves

This study revealed that mean annual rainfall, elevation, and silt percent were the major factors influencing the habitat suitability of *O. decumbens* in Fars Province, as determined consistently by both the Boruta and LASSO algorithms (Fig. 8a, b). These variables contributed most significantly to the predictive models, underscoring their ecological relevance. In contrast, variables such as organic matter (OM), profile curvature, and slope aspect (according to LASSO), as well as OM, plan curvature, and electrical conductivity (EC) (according to Boruta), were identified as the least influential.

The top 10 predictive variables obtained from each method are listed in Table 4. According to Boruta, the most important variables are rainfall, elevation, silt, roads, sand, clay, phosphorus (P), temperature, nitrogen (N), and aspect. LASSO identified rainfall, elevation, silt, roads, clay, sand, nitrogen (N), potassium (K), temperature, and phosphorus (P) as top contributors. These results indicate a strong overlap between the two approaches in identifying core environmental drivers, particularly for rainfall, elevation, and silt content.

Fig. 8 The importance of factors affecting habitat suitability based on LASSO (a) and Boruta (b) algorithms. OM: Organic matter; EC: Electrical conductivity; K: Potassium; N: Nitrogen; P: Phosphorus



The habitat suitability and environmental factor maps further support these findings, showing that optimal habitats for *O. decumbens* occur in areas with specific environmental thresholds: annual rainfall between 382.89–721.91 mm (Fig. 2b), elevation from 164 to 1361 m (Fig. 4a), and silt percentage between 36.86–64.67% (Fig. 3c). These data suggest that the species thrives in wetter regions with moderate elevation and finer-textured soils, which are conducive to water retention and root penetration.

However, ongoing drought conditions in Fars Province pose a substantial threat to these optimal zones, underscoring the need for targeted conservation planning. Additionally, spatial analysis revealed that the most suitable areas for the species' development are located on relatively flat lands

with higher silt concentrations, predominantly in the north-west regions of the province. This analysis offers valuable insights for prioritizing conservation and informing cultivation strategies for *O. decumbens* under current and future climatic scenarios.

3.3 Habitat suitability maps and model accuracy

The habitat suitability maps generated by the five machine learning models (NB, QDA, LDA, XGB, and GAM) revealed both consistent regional trends and model-specific spatial nuances in predicting suitable areas for *O. decumbens* across Fars Province (Fig. 9). A prominent pattern shared among all models was the concentration of highly

Table 4 Top 10 variables from each method

Rank	Boruta (Based on Z-score Importance)	LASSO (Coefficient Importance)
1	Rainfall	Rainfall
2	Elevation	Elevation
3	Silt	Silt
4	Roads	Roads
5	Sand	Clay
6	Clay	Sand
7	P	N (Nitrogen)
8	Temperature	K (Potassium)
9	N (Nitrogen)	Temperature
10	Aspect	P (Phosphorus)

suitable habitats in the northern and northwestern parts of the province—areas likely characterized by favorable climatic, edaphic, and topographic conditions. In contrast, the central, eastern, and southeastern regions were generally classified as low to very low suitability zones, though the extent and intensity of this classification varied by model. Notably, QDA and LDA provided the most spatially discrete delineation of habitat classes, suggesting a stronger sensitivity to underlying environmental gradients. These two models distinctly separated high-suitability zones from marginal areas, while the XGB model produced a more graduated and smoother suitability transition across the landscape. The NB model showed a slightly more optimistic distribution of high-suitability areas, potentially indicating a broader tolerance for environmental variation in its predictions. Meanwhile, the GAM model emphasized both highly suitable and unsuitable extremes, reinforcing its capacity to detect both optimal and restrictive zones based on continuous relationships. Collectively, the ensemble of model outputs illustrates the complex spatial heterogeneity of habitat conditions in the region and demonstrates the value of comparative modeling in identifying consensus areas of ecological importance and uncertainty.

The habitat suitability values were categorized into five classes: very low, low, moderate, high, and very high using the natural breaks method (Fig. 10). The comparison of habitat suitability classes across five machine learning models reveals considerable variation in the classification patterns of environmental favorability. Notably, GAM, QDA, and LDA allocated a large proportion of the area to the “Very low” suitability class (over 40%), indicating a more conservative or restrictive prediction pattern. In contrast, the XGB model showed a more balanced distribution, with higher representation in the “Low” and “Moderate” classes, suggesting a nuanced differentiation between marginally suitable and moderately favorable areas. The NB model demonstrated a

relatively higher percentage in the “High” suitability class, highlighting its potential sensitivity in identifying regions with greater ecological value. Interestingly, the “Very high” class was predominantly emphasized by the GAM model, while XGB classified the fewest areas under this category. These discrepancies highlight the inherent differences in model behavior and underscore the importance of selecting an algorithm based on the specific objectives of habitat suitability assessments.

The initial model performance was evaluated using 30% of the presence points ($n = 50$) and 50 randomly generated absence points across the province. These points were mixed and used in a value-to-point accuracy assessment. To address concerns regarding overestimation due to spatial autocorrelation, we implemented spatial block k-fold cross-validation ($k = 5$), partitioning the study area into geographically distinct blocks. Mean AUC scores and their standard deviations were computed across folds for each algorithm to assess spatial transferability. The cross-validation results confirmed that QDA and LDA remained the top-performing models (mean AUC = 0.90 ± 0.05 and 0.89 ± 0.06 , respectively), followed by GAM (0.86 ± 0.08), NB (0.84 ± 0.07), and XGB (0.83 ± 0.06) (Table 5, Fig. 11).

To further examine model reliability and uncertainty, we visualized the variability in model performance using both random split and spatial cross-validation AUC scores along with their associated errors (Fig. 12). This analysis highlights the robustness and generalization capacity of the evaluated models. Notably, QDA and LDA exhibited the lowest uncertainty, indicating consistent performance across spatial partitions. In contrast, GAM and NB showed slightly higher variability, reflecting sensitivity to spatial heterogeneity. Such uncertainty assessment is crucial in habitat modeling to ensure spatially reliable predictions and to guide decision-making in conservation and land-use planning.

3.4 Model limitations, projection constraints, and future research directions

While the modeling framework presented in this study offers valuable insights into the spatial distribution and key environmental drivers of *O. decumbens* in Fars Province, several limitations and considerations must be acknowledged to ensure appropriate interpretation and application of the results.

First, although all predictors passed multi-collinearity checks using VIF and Tolerance thresholds, and the models exhibited high accuracy and low uncertainty (especially QDA and LDA), it is essential to recognize that the environmental range and spatial distribution of the training data constrain the model projection. Extrapolating predictions beyond these ranges, particularly under future climate or land-use scenarios, may yield unreliable outputs. Second,

our models do not explicitly account for dynamic processes such as ongoing climate change, increasing anthropogenic disturbances, or shifts in land use, all of which could significantly alter habitat suitability in the coming decades. Although current results suggest that northwestern regions with higher silt content and rainfall are most suitable, these zones may contract or shift under drought intensification or

habitat fragmentation. Moreover, incorporating socio-economic data or land-use change forecasts (e.g., agriculture intensification) would also improve real-world relevance.

Future research should consider ensemble modeling across multiple global climate models (GCMs), the integration of temporal layers, and multi-criteria decision analysis to identify resilient conservation zones. Furthermore,

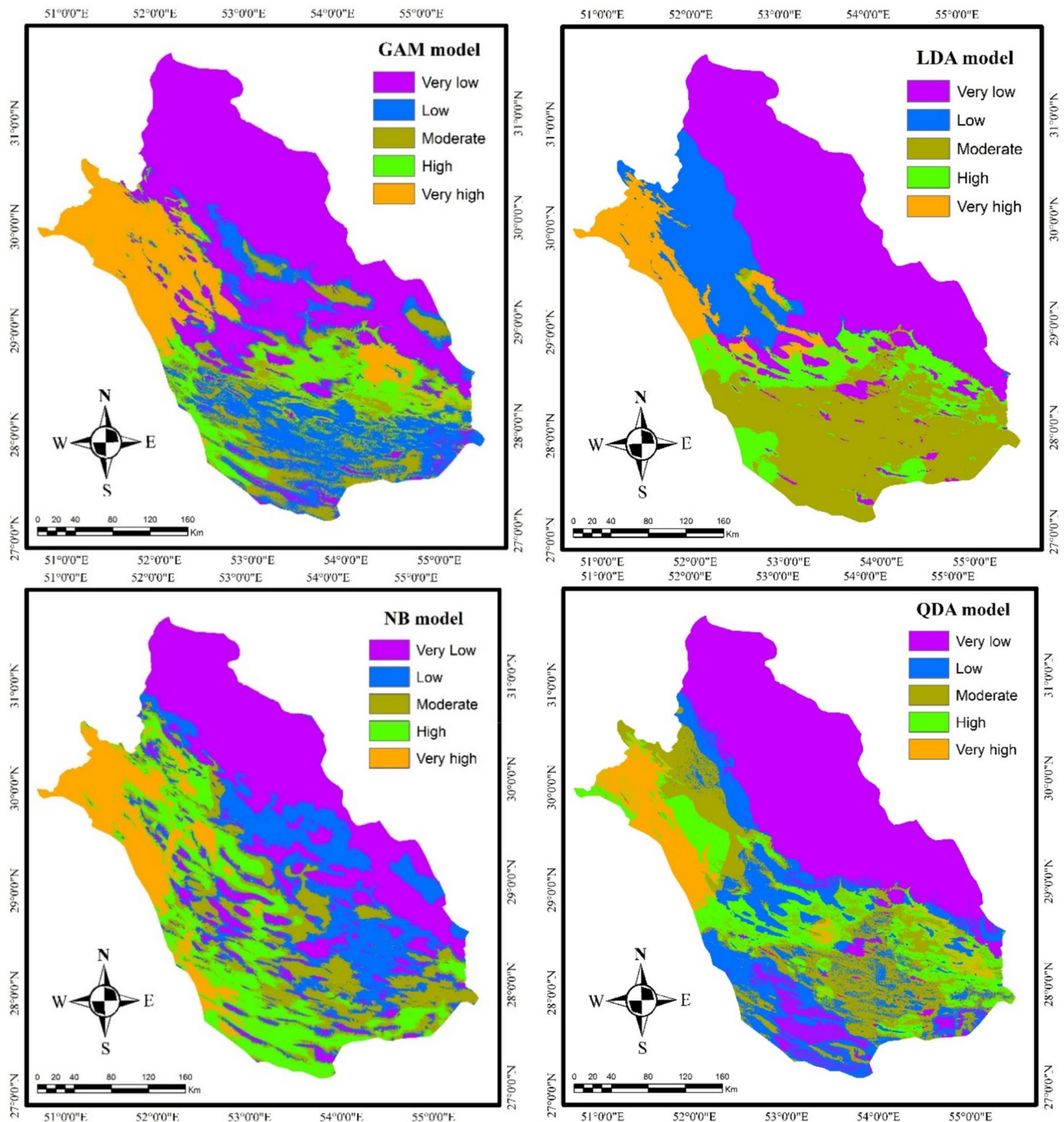


Fig. 9 The habitat suitability maps of *O. decumbens* generated by the QDA, LDA, GAM, XGB, and NB models

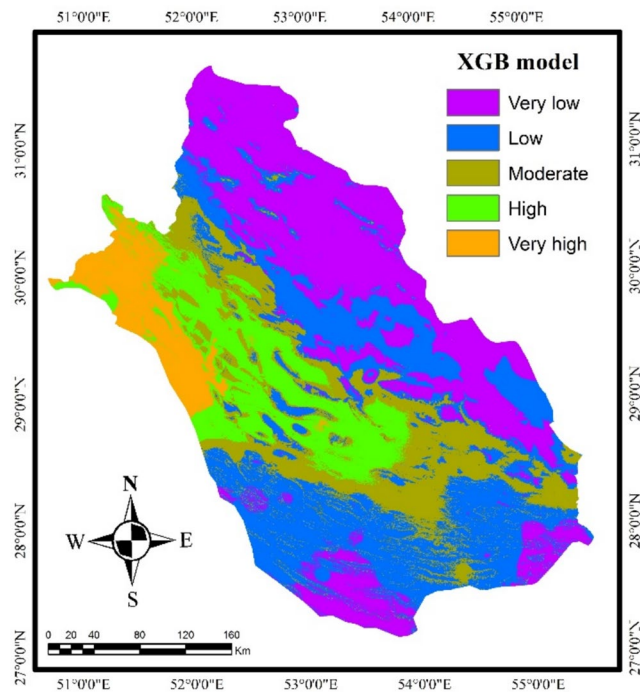


Fig. 9 (continued)

Fig. 10 Similarity percentage of five models across various habitat suitability classes

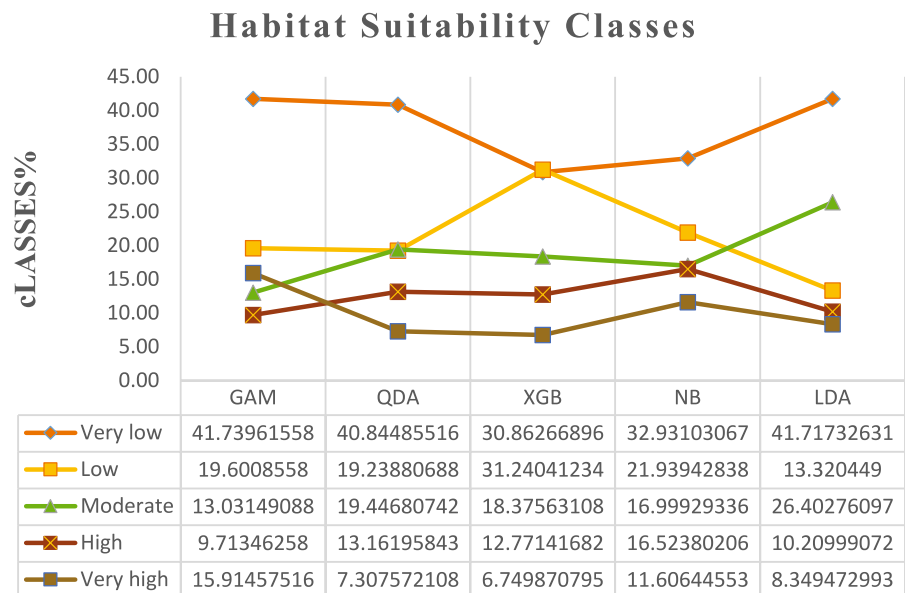


Table 5 Model evaluation results using the AUC metric

Model	AUC (random split)	Std. error	95% CI (lower–upper)	Mean AUC (spatial CV)	SD (spatial CV)
GAM	0.94	0.04	0.87–1.00	0.86	0.08
QDA	0.98	0.02	0.95–1.00	0.90	0.05
XGB	0.90	0.06	0.79–1.00	0.83	0.06
NB	0.90	0.05	0.81–1.00	0.84	0.07
LDA	0.97	0.03	0.92–1.00	0.89	0.06

monitoring programs should be established in predicted high-suitability areas to validate model outputs over time and refine habitat suitability thresholds under real environmental dynamics. While our current models provide a robust baseline for understanding species distribution, informed conservation and land-use planning will require integrating uncertainty, spatial dynamics, and future scenarios to ensure ecological resilience and sustainability.

4 Discussion

In ecological modeling, accurate habitat suitability predictions and environmental factors' influence are essential [37]. A key challenge in achieving these goals is multi-collinearity, where predictor variables show high correlations, leading to unreliable estimates and inflated standard errors [31]. Collinearity diagnostics, using methods like the variance inflation factor (VIF) and Tolerance indices, are crucial for ensuring the robustness of ecological models [38]. The collinearity tests in this study confirmed the absence of significant collinearity among the selected factors, thereby enhancing the model's reliability and interpretability.

Fig. 11 ROC curve analysis is used to validate the habitat suitability models

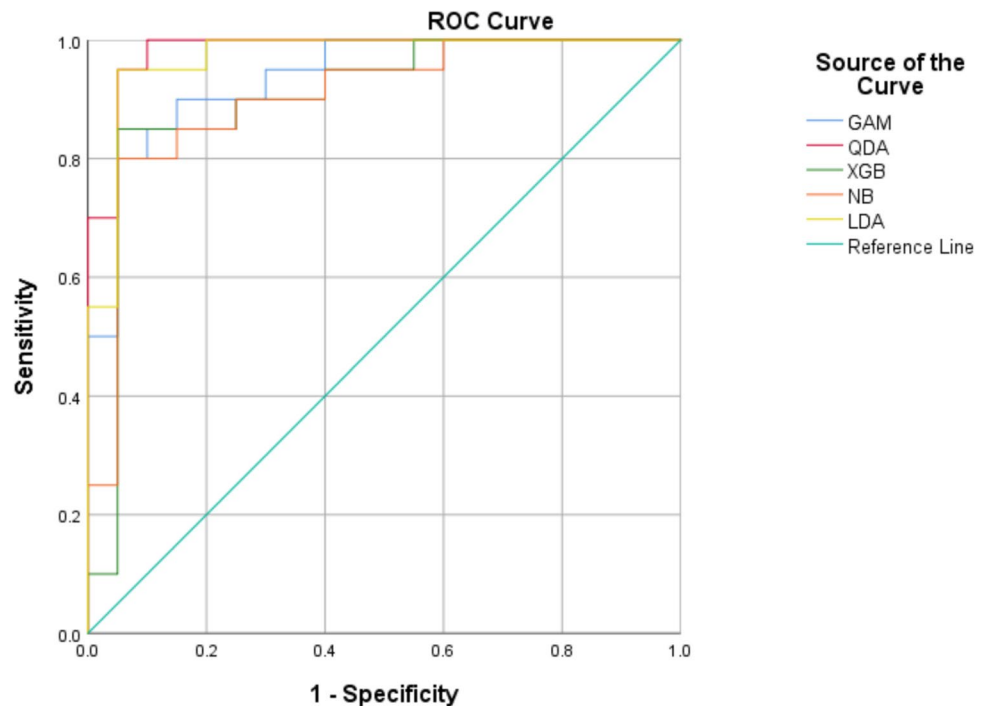
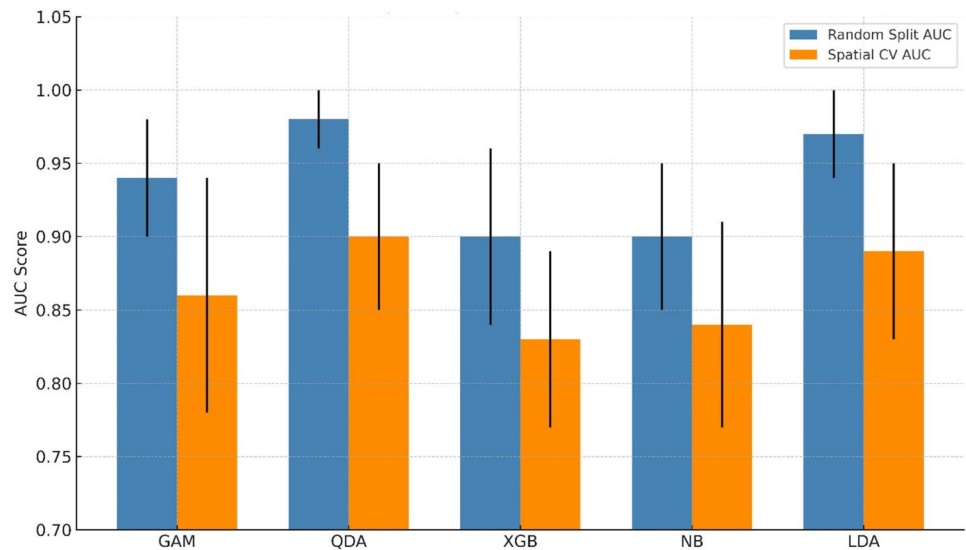


Fig. 12 Model uncertainty analysis based on AUC



Advanced algorithms, specifically Boruta and LASSO, significantly enhanced model performance by selecting the most relevant environmental factors [39]. Boruta is an effective feature selection tool, preserving all influential variables by contrasting their importance with random comparisons, ensuring no significant factor is omitted. Conversely, LASSO fine-tunes the model by shrinking less relevant coefficients toward zero, thus removing non-essential variables [40]. This approach is particularly useful in ecological data with numerous predictors, as it maintains the model's explanatory power while simplifying it. Combined, Boruta and LASSO allowed precise identification of key predictors,

such as rainfall, elevation, and silt percentage, critical for determining the habitat suitability of *O. decumbens*.

Among the environmental factors, rainfall emerged as a major determinant for the growth and distribution of medicinal and aromatic plants (MAPs), influencing processes such as nutrient absorption and photosynthesis. Previous research has highlighted that rainfall changes, particularly in arid regions, can alter plant phenology and impact pollinator resources [41]. Additionally, rainfall variations directly affect the habitat distribution and resilience of MAPs to climate-induced stress [42]. For *O. decumbens*, shifts in rainfall patterns due to climate change pose substantial challenges,

underscoring the need for habitat models that inform adaptive management and conservation efforts.

Elevation significantly influences habitat suitability, affecting variables such as temperature and humidity, which in turn impact soil properties and plant nutrient uptake [43]. In regions such as the Himalayas, MAP diversity decreases with elevation, which affects species distribution [44]. Studies have shown that MAPs often produce varying levels of secondary metabolites, such as flavonoids, at different elevations, with potential implications for the medicinal potency of these plants [45]. This study found that *O. decumbens* showed a stronger suitability at lower elevations, suggesting that conservation efforts should prioritize these regions.

Soil properties, specifically silt percentage, also impacted the distribution of *O. decumbens*. Silt-rich soils often enhance moisture retention, nutrient availability, and organic carbon mass, providing favorable microhabitats, especially in arid regions [46]. The study found that higher silt content contributed to increased habitat suitability, as seen in other MAPs, where soil texture influences water infiltration and plant growth.

This research implemented multiple machine learning models, including Naïve Bayes (NB), quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), gradient boosting (XGB), and generalized additive models (GAM), to enhance habitat suitability predictions. LDA and QDA exhibited superior predictive power, effectively capturing non-linear relationships and distinguishing habitat suitability classes [47]. NB, known for its computational efficiency, was well-suited for the large datasets characteristic of ecological studies, as evidenced by previous research on invasive species mapping [48]. QDA's strength in modeling complex data relationships was particularly beneficial for capturing ecological nuances in habitat distributions [49].

The ROC-AUC metric provided an unbiased measure of model accuracy, despite potential class imbalances in the data, an important consideration in ecological studies where data availability is often limited [50, 51]. Including models like XGB and GAM facilitated accurate predictions by handling diverse ecological contexts and non-linear relationships among predictors, ensuring robust and interpretable results.

This study utilizes machine learning algorithms to enhance the ecological modeling of MAPs, particularly *O. decumbens*, by facilitating precise habitat suitability assessments. These models are invaluable for identifying conservation priority areas and guiding efforts toward regions with optimal conditions for species survival and sustainable management. Comparative studies have demonstrated that advanced machine learning techniques, similar to those employed in this study, are effective in habitat suitability modeling. This research highlights the potential of machine learning to address evolving environmental challenges and inform proactive conservation strategies for MAPs, fostering resilience in ecologically and economically valuable species.

5 Conclusion

This study highlights the importance of incorporating machine learning algorithms into habitat suitability modeling for the medicinal plant *O. decumbens*, particularly in response to environmental variability. By applying five robust models—Naive Bayes (NB), quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), generalized additive models (GAM), and extreme gradient boosting (XGB)—the research identified rainfall, elevation, and silt content as the most influential environmental predictors, consistently selected by both Boruta and LASSO feature selection methods. Among the models, QDA and LDA exhibited the highest spatial predictive performance and lowest uncertainty, suggesting their strong potential for ecological applications in heterogeneous landscapes. The generated habitat suitability maps revealed spatial heterogeneity, with optimal zones primarily located in the north and northwest of Fars Province. These outputs offer practical insights for targeted conservation and sustainable cultivation strategies. While the models showed high accuracy, limitations related to environmental extrapolation and dynamic land-use changes should be considered when interpreting the predictions. Overall, this study highlights the potential of data-driven approaches in spatial information science to support informed decision-making for medicinal plant management under current and future environmental conditions.

Author contributions Emran Dastres and Hamidreza Rabiei-Dastjerdi conceived and designed the study, and led the writing with input from Hassan Esmaili, Mahdis Amiri, Akbar Karami, and Mehdi Gheisari. Emran Dastres performed simulations and data analyses, and Hamidreza Rabiei-Dastjerdi developed analytical results. All authors contributed critically to drafts and gave final approval for publication.

Funding Open Access funding provided by the IReL Consortium.

Data availability Upon acceptance, all data associated with this study will be archived in a publicly accessible repository, such as Dryad or Figshare, to ensure transparency and reproducibility. The dataset will include raw environmental variables, processed data, and model outputs. The specific DOI and link to the archived data will be provided in the Data Accessibility Statement within the manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Mirahmad, A., Ghoran, S. H., Alipour, P., Taktaz, F., Hassan, S., Naderian, M., et al. (2024). *Oliveria decumbens* Vent. (Apiaceae): Biological screening and chemical compositions. *Journal of Ethnopharmacology*, 318, 117053. <https://doi.org/10.1016/j.jep.2023.117053>
- Esmaili, H., Karami, A., & Maggi, F. (2018). Essential oil composition, total phenolic and flavonoids contents, and antioxidant activity of *Oliveria decumbens* Vent. (Apiaceae) at different phenological stages. *Journal of Cleaner Production*, 198, 91–95. <https://doi.org/10.1016/j.jclepro.2018.07.029>
- Esmaili, H., Mirjalili, M. H., Karami, A., & Nejad Ebrahimi, S. (2024). Introducing the glycyrrhizic acid and glabridin rich genotypes from the cultivated Iranian licorice (*Glycyrrhiza glabra* L.) populations to exploit in production systems. *Scientific Reports*, 14(1), 11034. <https://doi.org/10.1038/s41598-024-61711-1>
- Rabiei-Dastjerdi, H., Brereton, F., & O'Neill, E. (2025). Towards designing a comprehensive composite index for social vulnerability to natural hazards in the big data era: Potential challenges and partial solutions. *Natural Hazards*, 121, 3885–3913. <https://doi.org/10.1007/s11069-024-06874-w>
- Rahmati, B., Rabiei-Dastjerdi, H., Bibri, S. E., et al. (2024). Reducing urban energy consumption and carbon emissions: A novel GIS-based model for sustainable spatial accessibility to local services and resources. *Computational Urban Science*, 4, 37. <https://doi.org/10.1007/s43762-024-00139-9>
- Zareie, S., Mohammadi, S., & Rabiei-Dastjerdi, H. (2025). Determining the effects of large dams and urbanization on soil salinity and surface temperature using satellite images in a Middle Eastern country. *Discover Sustainability*, 6, 393. <https://doi.org/10.1007/s43621-025-01168-y>
- Sadeghi, M., Karimi, M., Rabiei-Dastjerdi, H., & Sarkar, D. (2025). Multi-scale dynamic population estimation: An Adaptive Inverse Distance Weighting (AIDW) model incorporating spatial characteristics. *Cartography and Geographic Information Science*. <https://doi.org/10.1080/15230406.2025.2492670>
- Amini, S., Saber, M., Rabiei-Dastjerdi, H., & Homayouni, S. (2022). Urban land use and land cover change analysis using random forest classification of Landsat time series. *Remote Sensing*, 14(11), 2654. <https://doi.org/10.3390/rs14112654>
- Gil-Marin, J. K., Shirazi, M., & Ivan, J. N. (2024). Assessing the Negative Binomial-Lindley model for crash hotspot identification: Insights from Monte Carlo simulation analysis. *Accident Analysis & Prevention*, 199, 107478. <https://doi.org/10.1016/j.aap.2024.107478>
- Alves, M. V., Maciel, L. I., Passos, J. O., Morais, C. L., Dos Santos, M. C., Lima, L. A., et al. (2023). Spectrochemical approach combined with symptoms data to diagnose fibromyalgia through paper spray ionization mass spectrometry (PSI-MS) and multi-variate classification. *Scientific Reports*, 13(1), 4658. <https://doi.org/10.1038/s41598-023-31565-0>
- La Manna, G., Guala, I., Pansini, A., Stipcich, P., Arrostituto, N., & Ceccherelli, G. (2024). Soundscape analysis can be an effective tool in assessing seagrass restoration early success. *Scientific Reports*, 14(1), 20910. <https://doi.org/10.1038/s41598-024-71975-2>
- Sudhakaran, G. (2024). Impact of climate change on the yield of medicinal plants in recent years. *Natural Product Research*. <https://doi.org/10.1080/14786419.2024.2372845>
- Hassanpouraghdam, M. B., Ghorbani, H., Esmaeilpour, M., Alford, M. H., Strzemeski, M., & Dresler, S. (2022). Diversity and distribution patterns of endemic medicinal and aromatic plants of Iran: Implications for conservation and habitat management. *International Journal of Environmental Research and Public Health*, 19(3), 1552. <https://doi.org/10.3390/ijerph19031552>
- Azizi, K., Soltani, Z., Aliakbarpour, M., Rezanezhad, H., & Kalantari, M. (2022). Bionomics of phlebotomine sand flies in different Climates of Leishmaniasis in Fars Province, Southern Iran. *Journal of Arthropod-Borne Diseases*, 16(2), 148. <https://doi.org/10.18502/jad.v16i2.11805>
- Saixiyala, A., Chen, L., Yi, F., Qiu, X., Sun, H., Cao, H., et al. (2023). Warming in combination with increased precipitation mediate the sexual and clonal reproduction in the desert steppe dominant species *Stipa breviflora*. *BMC Plant Biology*, 23(1), 474. <https://doi.org/10.1186/s12870-023-04439-w>
- Chen, X., Wang, Y., Wang, Y., Zhang, Y., Shen, Y., He, X., & Xiao, C. (2023). A natural moisture gradient affects soil fungal communities on the South Shore of Hulun Lake, Inner Mongolia, China. *Journal of Fungi*, 9(5), 549. <https://doi.org/10.3390/jof9050549>
- Regasa, A., Haile, W., & Abera, G. (2024). Assessment of soil acidity and fertility status under different land uses types in Sayo District of Western Ethiopia. *Russian Agricultural Sciences*, 50(2), 172–184. <https://doi.org/10.3103/S1068367424700137>
- Jiafeng, W., & Qiuliang, C. (2024). Nitrogen migration and transformation characteristics of the soil in karst areas under the combined application of oxalic acid and urea inhibitors. *Frontiers in Plant Science*, 15, 1386912. <https://doi.org/10.3389/fpls.2024.1386912>
- Hewitt, K. G., Hofmann, R. W., Ball, O. J., Luo, D., Finch, S. C., Bryant, R. H., & Popay, A. J. (2024). Phosphorus fertiliser is associated with reduced grass grub (*Costelytra giveni*) fitness in *Epichloë* endophyte-infected meadow fescue and perennial ryegrass. *Pest Management Science*, 80(12), 6409–6423. <https://doi.org/10.1002/ps.8369>
- Selassie, Y. G., Molla, E., Muhabie, D., Manaye, F., & Dessie, D. (2020). Response of crops to fertilizer application in volcanic soils. *Heliyon*, 6(12), e05629. <https://doi.org/10.1016/j.heliyon.2020.e05629>
- Dastres, E., Rabiei-Dastjerdi, H., Esmaeili, H., Amiri, M., Sonboli, A., & Mirjalili, M. H. (2025). Modeling habitat suitability for endangered herb (*Salvia lerifolia* Benth) using innovative hybrid machine learning algorithms. *Environmental and Sustainability Indicators*. <https://doi.org/10.1016/j.indic.2025.100694>
- Rabiei-Dastjerdi, H., Amini, S., McArdle, G., & Homayouni, S. (2023). City-region or city? That is the question: Modeling sprawl in Isfahan using geospatial data and technology. *GeoJournal*, 88(Suppl 1), 135–155. <https://doi.org/10.1007/s10708-021-10554-8>
- Garcia-Carretero, R., Barquero-Perez, O., Mora-Jimenez, I., Soguero-Ruiz, C., Goya-Esteban, R., & Ramos-Lopez, J. (2019). Identification of clinically relevant features in hypertensive patients using penalized regression: A case study of cardiovascular events. *Medical & Biological Engineering & Computing*, 57, 2011–2026. <https://doi.org/10.1007/s11517-019-02007-9>
- Shaheen, A., & Iqbal, J. (2018). Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and random forest. *Boruta algorithm*. *Sustainability*, 10(3), 799. <https://doi.org/10.3390/su10030799>
- Coupé, C. (2018). Modeling linguistic variables with regression models: Addressing non-Gaussian distributions, non-independent observations, and non-linear predictors with random effects and

- generalized additive models for location, scale, and shape. *Frontiers in Psychology*, 9, 513. <https://doi.org/10.3389/fpsyg.2018.00513>
26. Sadeghi, M., Karimi, M., Rabiei-Dastjerdi, H., & Sarkar, D. (2023). Adaptive weighted least squares (AWLS): A new vector-based model to improve urban population estimation at small-area scale using morphology and attractiveness criteria. *Applied Geography*, 158, 103050. <https://doi.org/10.1016/j.apgeog.2023.103050>
27. Zhou, C. M., Li, H., Xue, Q., Yang, J. J., & Zhu, Y. (2024). Artificial intelligence algorithms for predicting post-operative ileus after laparoscopic surgery. *Heliyon*, 10(5), e26580. <https://doi.org/10.1016/j.heliyon.2024.e26580>
28. Weng, W. H., Waghholkar, K. B., McCray, A. T., Szolovits, P., & Chueh, H. C. (2017). Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC Medical Informatics and Decision Making*, 17, 1–13. <https://doi.org/10.1186/s12911-017-0556-8>
29. Sosulski, J., & Tangermann, M. (2022). Introducing block-Toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain-computer interfaces. *Journal of Neural Engineering*, 19(6), 066001. <https://doi.org/10.1088/1741-2552/ac9c98>
30. Dastres, E., Edalat, M., Moayedi, G., Jahangiri, E., Pourghasemi, H. R., & Tiefenbacher, J. P. (2022). An evaluation of alternative statistical models for predicting habitat suitability for weeds. *Weed Research*, 62(6), 404–421. <https://doi.org/10.1111/wre.12553>
31. Dastres, E., Jahangiri, E., Edalat, M., Zamani, A., Amiri, M., & Pourghasemi, H. R. (2023). Habitat suitability modeling of *Descurainia sophia* medicinal plant using three bivariate models. *Environmental Monitoring and Assessment*, 195(3), 392. <https://doi.org/10.1007/s10661-023-10996-2>
32. Kahsay, A., Haile, M., Gebresamuel, G., Mohammed, M., & Okolo, C. C. (2023). Dynamics of soil properties as impacted by contrasting lithology, slope class, and land use types: A case study in semi-arid highlands of northern Ethiopia. *Environmental Monitoring and Assessment*, 195(9), 1091. <https://doi.org/10.1007/s10661-023-11706-8>
33. Maqbool, Z., Farooq, M. S., Rafiq, A., Uzair, M., Yousuf, M., Khan, M. R., & Huo, S. (2024). Unlocking the potential of biochar in the remediation of soils contaminated with heavy metals for sustainable agriculture. *Functional Plant Biology*. <https://doi.org/10.1071/FP23257>
34. Skudienė, R., Matyžiūtė, V., & Šiaudinis, G. (2023). Soil seed banks and their relation to soil properties in hilly landscapes. *Plants*, 13(1), 104. <https://doi.org/10.3390/plants13010104>
35. Rincón-Aranguri, M., Toro-Cardona, F. A., Galeano, S. P., Roa-Fuentes, L., & Urbina-Cardona, N. (2023). Functional diversity of snakes is explained by the landscape composition at multiple areas of influence. *Ecology and Evolution*, 13(7), e10352. <https://doi.org/10.1002/ece3.10352>
36. Shabani, M., Darvishi, S., Rabiei-Dastjerdi, H., Alavi, A. S., Choudhury, T., & Solaimani, K. (2022). An integrated approach for simulation and prediction of land use and land cover changes and urban growth (case study: Sanandaj city in Iran). *Journal of the Geographical Institute Jovan Cvijic, SASA*, 72(3), 273–289
37. Dastres, E., Esmaili, H., & Edalat, M. (2025). Species distribution modeling of *Malva neglecta* Wallr. weed using ten different machine learning algorithms: An approach to site-specific weed management (SSWM). *European Journal of Agronomy*, 167, 127579. <https://doi.org/10.1016/j.eja.2025.127579>
38. Marcoulides, K. M., & Raykov, T. (2019). Evaluation of variance inflation factors in regression models using latent variable modeling methods. *Educational and Psychological Measurement*, 79(5), 874–882. <https://doi.org/10.1177/0013164418817803>
39. Zhu, J., Luo, J., & Ma, Y. (2024). Screening of serum exosome markers for colorectal cancer based on Boruta and multi-cluster feature selection algorithms. *Molecular & Cellular Toxicology*, 20(2), 343–351. <https://doi.org/10.1007/s13273-023-00348>
40. Hayah, I., Ababou, M., Botti, S., & Badaoui, B. (2021). Comparison of three statistical approaches for feature selection for fine-scale genetic population assignment in four pig breeds. *Tropical Animal Health and Production*, 53(3), 395. <https://doi.org/10.1007/s11250-021-02824-x>
41. Kunwar, R. M., Thapa-Magar, K. B., Subedi, S. C., Kutal, D. H., Baral, B., Joshi, N. R., et al. (2023). Distribution of important medicinal plant species in Nepal under past, present, and future climatic conditions. *Ecological Indicators*, 146, 109879. <https://doi.org/10.1016/j.ecolind.2022.109879>
42. Thuma, J. A., Duff, C., Pitera, M., Januario, N., Orians, C. M., & Starks, P. T. (2023). Nutrient enrichment and rainfall affect plant phenology and floral resource availability for pollinators. *Frontiers in Ecology and Evolution*, 11, 1150736. <https://doi.org/10.3389/fevo.2023.1150736>
43. Jiang, D. Q., Wang, H. Y., Kang, C. Z., Jiang, J. Y., Zhang, Y., et al. (2020). Influence and mechanism of stress combination on medicinal plants secondary metabolism. *Zhongguo Zhong Yao Za Zhi*, 45(9), 2009–2016. <https://doi.org/10.19540/j.cnki.cjcmm.20200302.106>
44. Samant, S. S., & Mohinder Pal, M. P. (2003). Diversity and conservation status of medicinal plants in Uttaranchal State. *Indian Journal of Forestry*, 26(4), 397–406. <https://doi.org/10.5555/20043012153>
45. Bisht, R., Bhattacharyya, A., Shrivastava, A., & Saxena, P. (2021). An overview of the medicinally important plant type III PKS derived polyketides. *Frontiers in Plant Science*, 12, 746908. <https://doi.org/10.3389/fpls.2021.746908>
46. Dimou, V., Demertzis, K., & Kantartzis, A. (2024). Harvesting wind damaged trees: A study of prediction of windthrow damage in mixed-broadleaf stands via a machine learning model. *International Journal of Forest Engineering*, 35(1), 43–57. <https://doi.org/10.1080/14942119.2023.2171599>
47. Oh, H. J., Syifa, M., Lee, C. W., & Lee, S. (2019). *Ruditapes philippinarum* habitat mapping potential using SVM and Naïve Bayes. *Journal of Coastal Research*, 90(1), 41–48. <https://doi.org/10.2112/SI90-006.1>
48. Marvasti-Zadeh, S. M., Goodsman, D., Ray, N., & Erbilgin, N. (2023). Early detection of bark beetle attack using remote sensing and machine learning: A review. *ACM Computing Surveys*, 56(4), 1–40. <https://doi.org/10.1145/3625387>
49. Chaibutr, N., Pongpanitanont, P., Laymanivong, S., Thanchomnang, T., & Janwan, P. (2024). Development of a machine learning model for the classification of *Enterobius vermicularis* egg. *Journal of Imaging*, 10(9), 212. <https://doi.org/10.3390/jimaging10090212>
50. Dastres, E., Sarvestani, G. S., Edalat, M., & Pourghasemi, H. R. (2025). Habitat suitability modeling of dominant weed in rapeseed (*Brassica napus*) fields using machine learning techniques. *Weed Science*. <https://doi.org/10.1017/wsc.2025.5>
51. Akter, S. B., Akter, S., Tuli, M. D., Eisenberg, D., Lotvola, A., Islam, H., Fernandez, J. F., Hüttemann, M., & Pias, T. S. (2025). Fair and explainable Myocardial Infarction (MI) prediction: Novel strategies for feature selection and class imbalance correction. *Computers in biology and medicine*, 184, 109413. <https://doi.org/10.1016/j.compbiomed.2024.109413>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.