frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fevo.2023.1116083

Check for updates

OPEN ACCESS

EDITED BY
Wei Zhao,
Institute of Geographic Sciences and Natural Resources Research (CAS),
China

REVIEWED BY
Babar Zahoor,
Smithsonian Conservation Biology Institute (SI),
United States
Yingxin Huang,
Northeast Institute of Geography and Agroecology (CAS),
China

*CORRESPONDENCE
Wei Mao
✉ maowei@hainanu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Population, Community, and Ecosystem Dynamics, a section of the journal Frontiers in Ecology and Evolution

RECEIVED 05 December 2022
ACCEPTED 16 January 2023
PUBLISHED 02 February 2023

CITATION
He B, Zhao Y, Liu S, Ahmad S and Mao W (2023) Mapping seagrass habitats of potential suitability using a hybrid machine learning model.
Front. Ecol. Evol. 11:1116083.
doi: 10.3389/fevo.2023.1116083

Mapping seagrass habitats of potential suitability using a hybrid machine learning model

Bohao He[1,2†], Yanghe Zhao[1,2†], Siyu Liu[1,2], Shahid Ahmad[1,2] and Wei Mao[1,2*]

[1]School of Ecology and Environment, Hainan University, Haikou, China, [2]Key Laboratory of Agro-Forestry Environmental Processes and Ecological Regulation of Hainan Province, Hainan University, Haikou, China

Seagrass meadows provide essential ecosystem services globally in the context of climate change. However, seagrass is being degraded at an accelerated rate globally due to ocean warming, ocean acidification, aquaculture, and human activities. The need for more information on seagrasses' spatial distribution and health status is a serious impediment to their conservation and management. Therefore, we propose a new hybrid machine learning model (RF-SWOA) that integrates the sinusoidal chaos map whale optimization algorithm (SWOA) with a random forest (RF) model to accurately model the suitable habitat of potential seagrasses. This study combines *in situ* sampling data with multivariate remote sensing data to train and validate hybrid machine learning models. It shows that RF-SWOA can predict potential seagrass habitat suitability more accurately and efficiently than RF. It also shows that the two most important factors affecting the potential seagrass habitat suitability on Hainan Island in China are distance to land (38.2%) and depth to sea (25.9%). This paper not only demonstrates the effectiveness of a hybrid machine learning model but also provides a more accurate machine learning model approach for predicting the potential suitability distribution of seagrasses. This research can help identify seagrass suitability distribution areas and thus develop conservation strategies to restore healthy seagrass ecosystems.

KEYWORDS

seagrass, machine learning, species distribution model, hybrid model, habitat suitability, niches

1. Introduction

Seagrasses are large submerged angiosperms with the general characteristics of vascular plants, a fully adapted aquatic environment, and the only angiosperm that can flower, fruit, and germinate in seawater (Hemminga and Duarte, 2000). Seagrasses are one of the extremely important marine resources that provide significant ecological value (Fourqurean et al., 2012; Cullen-Unsworth and Unsworth, 2018). Globally, seagrass habitats are rapidly degrading, and the loss of seagrass habitats will lead to multiple risks, such as the increased impacts of global climate change, shoreline destruction, and declining biodiversity (Orth et al., 2006; Waycott et al., 2009; Kendrick et al., 2019; Moksnes et al., 2021). Seagrass suitability habitat distribution patterns in the world are changing as the effects of global change severely threaten seagrass suitability habitats. The accurate knowledge of seagrass habitat and understanding of what factors limit or even threaten seagrass distribution has become an urgent issue (Short and Wyllie-Echeverria, 1996; Waycott et al., 2009). Unfortunately, many seagrass habitats around the world do not have clear spatial information (McKenzie et al., 2001; Short et al., 2007; He et al., 2022b), which seriously hinders marine environmental management and seagrass conservation. The traditional experimental method of mapping seagrass distribution requires large-scale field investigations, which are time consuming and cost-effective (McKenzie et al., 2001;

Krause-Jensen et al., 2004). In recent years, due to the development of remote sensing technology, a mushrooming number of data and methods have been applied to marine predictive modeling, such as satellite data, unmanned aerial vehicles (UAV), acoustic surveys, and Geographic Information Systems (GIS; Picart et al., 2014; Ouellette and Getinet, 2016; Fingas, 2019; Belkin, 2021).

Species distribution models (SDMs) are used to predict regional distribution maps (Gonzalez-Irusta et al., 2015; Bittner et al., 2020) and to assess the degree of habitat suitability (Miller, 2010; Zimmermann et al., 2010; Pollock et al., 2014). As SDM has been intensively studied, more and more studies have chosen to use machine learning for SDM modeling and have produced excellent results (Evans et al., 2011; Li and Wang, 2013). Downie et al. (2013) used GAM and MaxEnt models to predict seagrass distribution, and their results showed that machine learning could accurately predict seagrass distribution. However, Bittner et al. (2020) found the differences in the relative importance of environmental factors in predicting the distribution of seagrasses between machine learning models when predicting the distribution of species. Therefore, a more accurate and robust machine learning model should be selected for prediction, such as random forest model that is widely used in SDM modeling methods. As a very representative tree modeling algorithm, random forest model can provide high prediction accuracy and stability (Kosicki, 2017; Mi et al., 2017; Kosicki, 2020; Luan et al., 2020; Saranya et al., 2021).

In recent years, with the development of machine learning, hybrid machine learning models have been widely used (Bies et al., 2006; Ardabili et al., 2019). Meta-heuristic algorithms have been found to improve the classification accuracy of models significantly (Beheshti and Shamsuddin, 2013; Singh and Kottath, 2021). Further, population-based hybrid optimization algorithms can dramatically increase the speed and power of search algorithms by moving from many individuals to collaborative groups (Abdel-Basset et al., 2018; Dokeroglu et al., 2019). The excellent performance and optimal solutions of metaheuristic algorithms solve the puzzles of multidisciplinary research, ranging from engineering and social sciences to ecology (Yang, 2009, 2013). This led to the widespread use of metaheuristics in many studies (de Melo and Carosio, 2013; Talbi, 2016; Dokeroglu et al., 2019; Hassan and Pillay, 2019; Cruz-Duarte et al., 2021; Moya et al., 2021), e.g., the whale optimization algorithm (WOA; Mirjalili and Lewis, 2016; Kaur and Arora, 2018; He et al., 2022a).

Some applications have demonstrated the usability of hybrid machine learning models in providing insights into various knowledge's domains (Tsai and Chen, 2010; Pinter et al., 2020). Still, few have explored the use of hybrid machine learning models to predict species suitability distributions. This study combines WOA into a Sinusoidal (S) chaotic graph and couples it with Random Forest (RF) to form a new hybrid machine learning model (RF-SWOA). The model is able to more accurately model seagrass habitat suitability. Thus, the objectives of this study are: (1) to develop a hybrid machine learning model for more accurately predicting potential seagrass habitat; (2) to explore the effects of environmental variables on seagrass habitat; and (3) to evaluate the predictive advantages and limitations of the hybrid machine learning model.

2. Materials and methods

2.1. Seagrass occurrence data

Hainan Province, located in the southernmost island of China, is the largest province in China in terms of land area (land plus sea). Hainan Province has a latitude and longitude range of 3.30°N to 20.07°N and 108.15°E to 120.05°E, respectively. The climate of Hainan Island belongs to the monsoon tropical climate, which is between the two temperature zones of the tropics and subtropics. Its annual average temperature is 24°C. Hainan Island is rich in plant and animal resources, of which seagrass is one of the main aquatic seed plant resources.

Hainan Island accounts for 64% of China's total seagrass area (Zheng et al., 2013). Therefore, this study conducted a field survey to determine the distribution of seagrass on Hainan Island from March to August 2021 (Figure 1). The presence of seagrass was marked with latitude and longitude, and samples were collected to identify seagrass species according to the method advocated by international seagrass researchers (Kuo and Den Hartog, 2001). The literature and field survey data were also combined to form the known distribution of seagrass beds on Hainan Island. We used GPS (ICEGPS 610) to record seagrass bed boundaries, as well as the latitude and longitude coordinates at low tide, to determine the spatial extent of seagrass distribution on Hainan Island, when the large areas of seagrass beds are more easily exposed at low tide (Yang and Yang, 2009; Jiang et al., 2017). A total of 42 actual seagrass distribution sites were used in this study, including seven species of seagrass (i.e., *Halophila ovalis*, *Halophila minor*, *Thalassia hemprichii*, *Halodule uninervis*, *Halodule pinifolia*, *Enhalus acoroides*, and *Halophila beccarii*), while a random sampling of pseudo-absence in the Hainan seagrass distribution area produced a total of 31,700 records (Barbet-Massin et al., 2012).

2.2. Environmental data

The distribution of seagrasses is regulated by external environmental factors and key physiological processes. A total of 15 environmental prediction layers were used in this study (Table 1), each of which was important in determining the prediction of potentially suitable habitat for seagrasses (Dennison, 1987; Duarte, 1990, 1991; Nguyen et al., 2021).

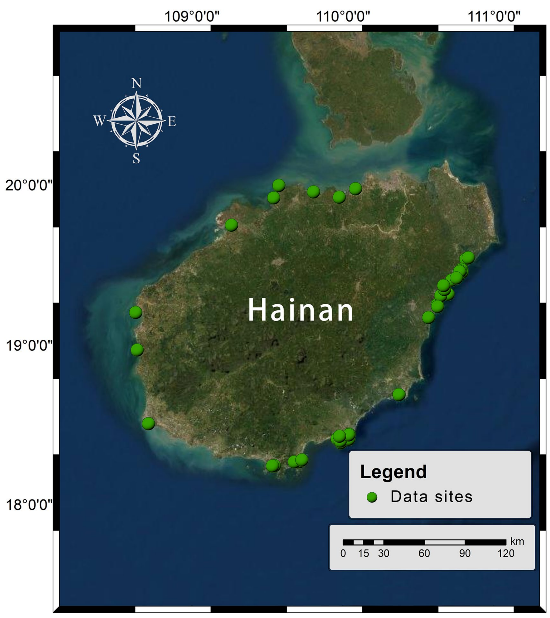

FIGURE 1
Study area and seagrass field distribution location sites.

TABLE 1 Environmental variables used in this study.

Notation	Description	Units
Silicate	Ocean silicate concentration	$mol.m^{-3}$
Attenuation	Diffuse attenuation	m^{-1}
Calcite	Constituent minerals in the ocean	$mol.m^{-3}$
Chlorophyll	Ocean chlorophyll-a concentration	$mol.m^{-3}$
Depth	Ocean depth	F
Land	Distance from land	F
Nitrate	Ocean nitrate concentration	$mol.m^{-3}$
Par	Photosynthetically active radiation	$E.m^{-2}.day^{-1}$
pH	Hydrogen ion concentration	1
Phosphate	Ocean phosphate concentration	$mol.m^{-3}$
Phytoplankton	Phytoplankton in the ocean	$\mu mol.m^{-3}$
Salinity	Ocean salinity	PSS
Slope	Ocean slope	F
Temperature	Ocean surface temperature	°C
Current	Currents velocity	m^{-1}

Temperature, salinity, velocity, nitrate, phosphate, silicate, phytoplankton, calcite, pH, and attenuation were obtained by Bio-ORACLE 2.2 version.[1] Ocean slope data are from GMED 2.0 version.[2] Ocean chlorophyll-*a* concentration data are from Google Earth Engine.[3] Photosynthetically active radiation data are from the Moderate-resolution Imaging Spectroradiometer (MODIS) aqua sensor.[4] Distance to nearest-shore data is from NASA's Ocean Biology Processing Group.[5] Bathymetric dataset is from The General Bathymetric Chart of the Oceans (GEBCO) global network.[6] Table 2 shows the minimum (MIN), maximum (MAX), mean (MEAN), and standard deviation (STD) of the 15 different environmental variables. All environmental variables were interpolated to 1 km spatial resolution using kriging interpolation in the ArcGIS 10.8 version of geostatistical analysis. To reduce spatial autocorrelation between variables (Legendre, 1993; Koenig, 1999; Dormann et al., 2007), correlation coefficients ($r > 0.7$) were excluded using *spdep* R package (Bivand et al., 2015).

2.3. Machine learning models and evaluation

2.3.1. Random forest model

Random Forest (RF) algorithm is an extension of Bagging (Breiman, 1996, 2001), in which base learners are fixed as decision trees and a forest is made up of multiple trees (Figure 2). Compared with bagging integration of decision trees, RF has poor starting performance. However, as the number of base learners increases, RF tends to converge to a lower generalization error. Also, unlike bagging, in which the decision tree selects the optimal division attributes from all attribute sets, RF selects the division attributes from only a subset of the attribute set and thus is more efficient to train. In this study, the gini importance built-in algorithm of random forests was used to calculate the importance of the environmental features of the potentially suitable habitats for seagrasses.

2.3.2. Hybrid model

Whale Optimization Algorithm (WOA) was introduced by Mirjalili and Lewis (2016). Inspired by the way whales hunt, the predation behavior is organized into three mathematical models: prey encirclement, bubble net attack, and prey search (Mirjalili and Lewis, 2016; Aljarah et al., 2018). A whale encircles its prey while locating the best search position with an increasing number of iterations, while updating in real time. The mathematical expression of this behavior is

$$D = \left|C * X_L(t) - X(t)\right|, \quad X(t+1) = X_L(t) - A * D, \tag{1}$$

where D represents the distance between whale and prey, A and C are the coefficient vectors, t indicates the current iteration, X_L is the position vector of the best solution obtained so far, X is the position vector, $|\ |$ is the absolute value, and * is an element-wise multiplication. A and C are calculated *via*

$$A = 2a * r - a, \quad C = 2 * r \tag{2}$$

where r is a random vector, and a is linearly decreased from 2 to 0 during iterating. A new position must be defined between the initial search position and the optimal search position so as to adjust the parameters. In this case, it is described as follows:

$$X(t+1) = \mathrm{D} * e^{bl} \cdot \cos(2\pi l) + X_L(\mathrm{t}), \tag{3}$$

where b is a constant coefficient, and l is a random vector whose items are all within [0, 1]. The whale contraction or spiral model approach is selected based on a 50% probability. Based on the mathematical model, the whale's prey is simulated in a spiral circle as follows:

$$X(t+1) = \begin{cases} X_L(t) - A * D & \text{if } p < 0.5 \\ D * e^{bl} * \cos(2\pi l) + X_L(t) & \text{if } p \geqslant 0.5 \end{cases} \tag{4}$$

Contraction envelope and spiral position updates are performed simultaneously, with contraction according to p and spiral wandering according to $1 - p$, where $p \in [0,1]$.

As the whale searches for prey, it moves toward the local optimal location while expanding the global optimal location search, and this phase can be described as follows:

$$D = \left| C * X_{\text{rand}} - X(t) \right|, \quad X(t+1) = X_{\text{rand}} - A * D, \tag{5}$$

1 https://www.bio-oracle.org/index.php, accessed on February 5, 2022.
2 https://gmed.auckland.ac.nz/index.html, accessed on February 6, 2022.
3 https://earthengine.google.com/, accessed on February 5, 2022.
4 https://oceancolor.gsfc.nasa.gov/data/aqua/, accessed on February 6, 2022.
5 https://oceancolor.gsfc.nasa.gov/docs/distfromcoast/, accessed on February 6, 2022.
6 https://www.gebco.net/, accessed on February 6, 2022.

where X_{rand} is a vector of random locations. A more detailed explanation of the WOA algorithm can be found in Mirjalili and Lewis (2016).

WOA becomes SWOA after adding a chaotic map to optimize global search capabilities. SWOA is mathematically described as follows:

$$p_{k+1} = ap_k^2 \sin(\pi p_k), p_0 \in [0,1], 0 < a \leqslant 4, \quad (6)$$

where k is the number of iterations, and a is the description parameter within $0 < a \leqslant 4$. For more information on SWOA algorithm, please refer to (He et al., 2022a).

The model in this study randomly selects 80% of the seagrass occurrence data for training and the remaining 20% for testing. The RF and RF-SWOA models were developed in Python 3.8 (Python, 2021).

TABLE 2 Statistical analysis results for different environmental variables.

Notation	Min	Max	Mean	Std
Silicate	5.87	13.12	8.13	1.96
Attenuation	0.04	0.27	0.15	0.06
Calcite	0.00	0.04	0.01	0.01
Chlorophyll	0.12	0.56	0.25	0.10
Depth	−100.09	−1.11	−30.82	18.70
Land	0.02	0.28	0.14	0.05
Nitrate	0.02	1.93	0.52	0.51
Par	36.04	42.37	39.21	1.54
pH	8.18	8.19	8.19	0.00
Phosphate	0.00	0.10	0.05	0.03
Phytoplankton	0.92	2.97	1.53	0.45
Salinity	32.94	33.31	33.18	0.08
Slope	0.02	0.19	0.09	0.04
Temperature	24.97	27.13	26.20	0.67
Current	0.12	0.55	0.24	0.09

2.3.3. Model evaluation

A comprehensive evaluation of the model was conducted using six evaluation metrics. They are AUC, Omission rate, Correct classification rate, Sensitivity, Specificity, Kappa.

$$\text{AUC} = \frac{1 + \frac{\text{True positive}}{\text{True positive} + \text{False negative}} - \frac{\text{False positive}}{\text{False positive} + \text{True negative}}}{\text{True number} * \text{False number}} \quad (7)$$

$$\text{Omission rate} = \frac{\text{False negative}}{\text{False negative} + \text{True negative}} \quad (8)$$

$$\text{Correct classification rate} = \frac{\text{True number}}{\text{Total sample}} \quad (9)$$

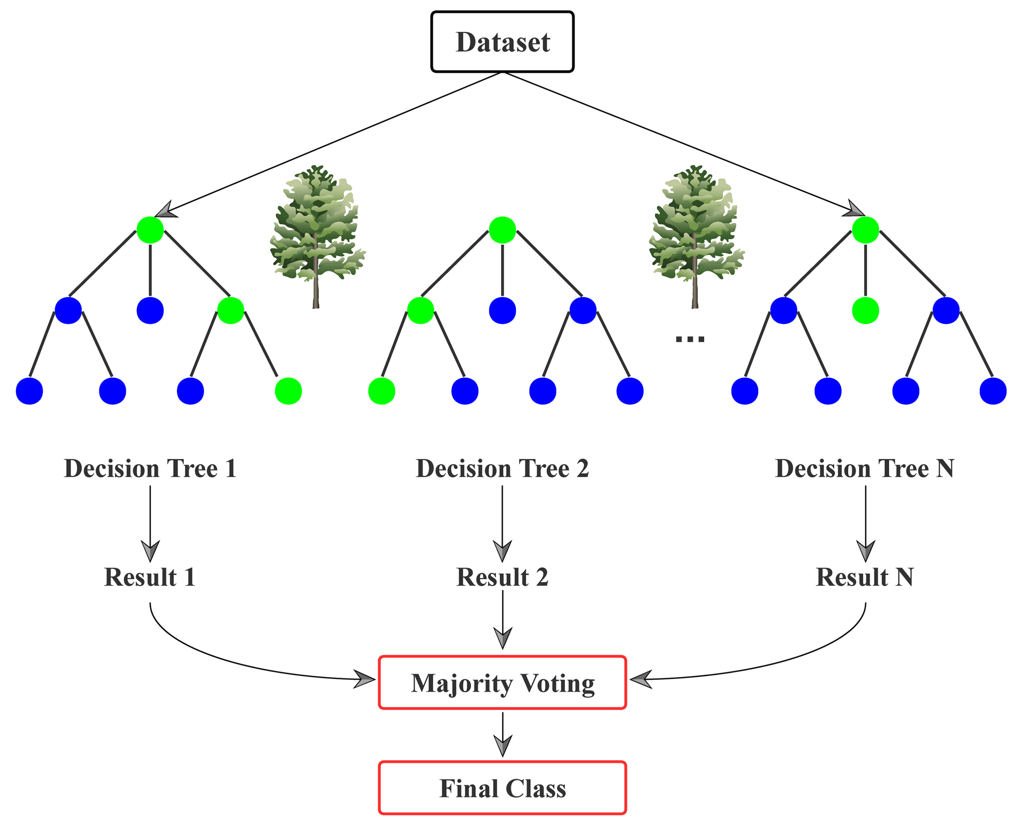

FIGURE 2
Random forest model structure.

10.3389/fevo.2023.1116083

$$\text{Sensitivity} = \frac{\text{True positive}}{(\text{True positive} + \text{False negative})} \quad (10)$$

$$\text{Specificity} = \frac{\text{True positive}}{(\text{False positive} + \text{True negative})} \quad (11)$$

$$\text{Kappa} = \frac{P_{\text{observed}} - P_{\text{bychance}}}{1 - P_{\text{bychance}}} \quad (12)$$

3. Results

3.1. Correlation analysis between environments

The results of the study clearly show the spatial autocorrelation among all environmental variables (Figure 3). Correlation analysis of environmental variables was used to identify and remove variables with high multicollinearity ($r > 0.7$) in order to prevent model over-fitting. After removing phosphate, phytoplankton, par, and attenuation environmental variables, and the remaining variables were introduced into the model training.

3.2. Importance of environment features

The results of the importance of environmental characteristics showed that the most important ones to predict the potential habitat of seagrass were the distance to land (38.2%) and the depth of the ocean (25.9%). The rest of the environmental variables showed small contribution values (<6%) to the prediction of the potential habitat of seagrass (Figure 4).

3.3. Potential seagrass habitat

Both models (RF and RF-SWOA) mapped potential seagrass habitat areas (Figure 5). RF model overestimates the potential habitat of

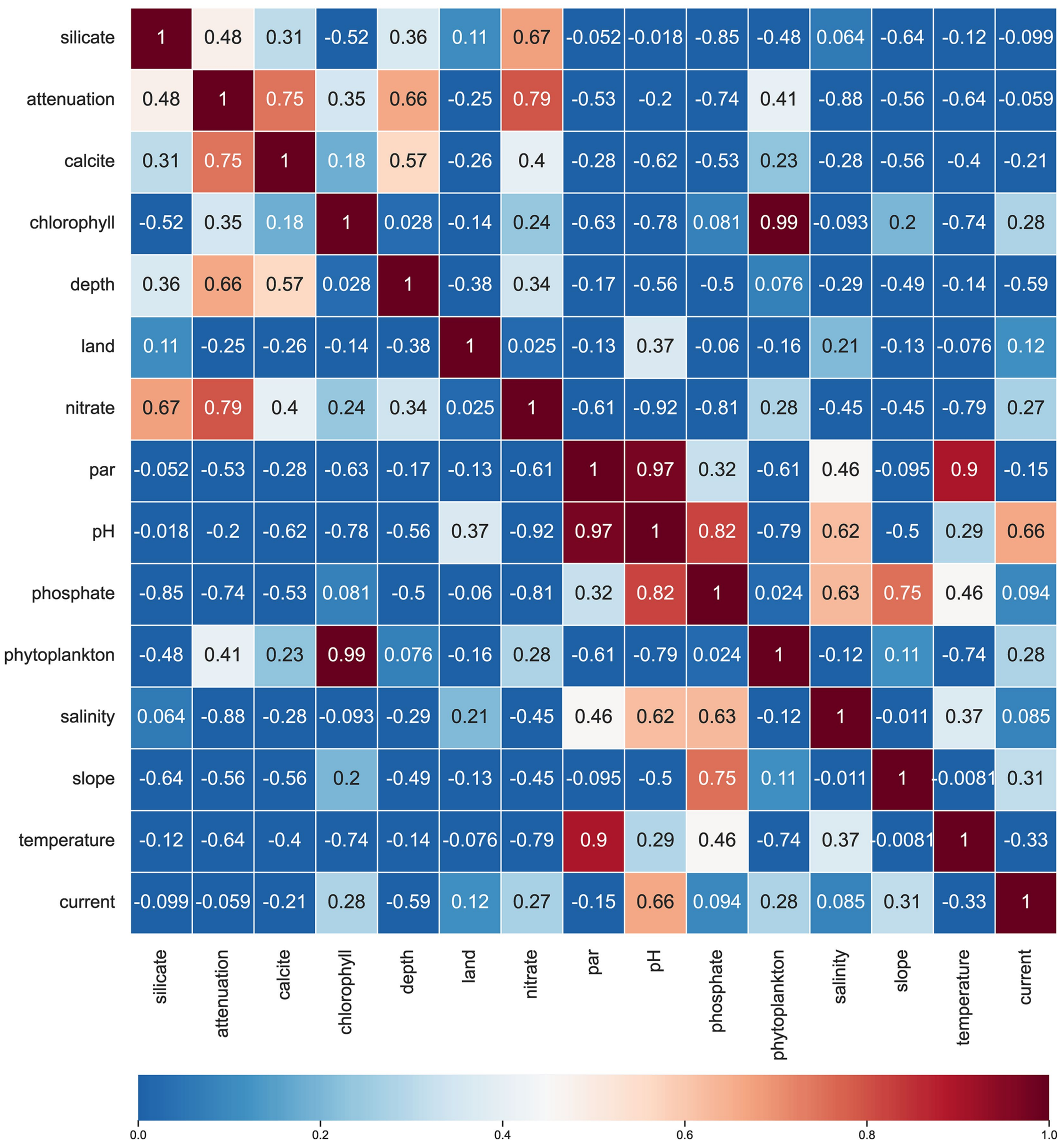

FIGURE 3
Correlation analysis matrix for different environmental variables.

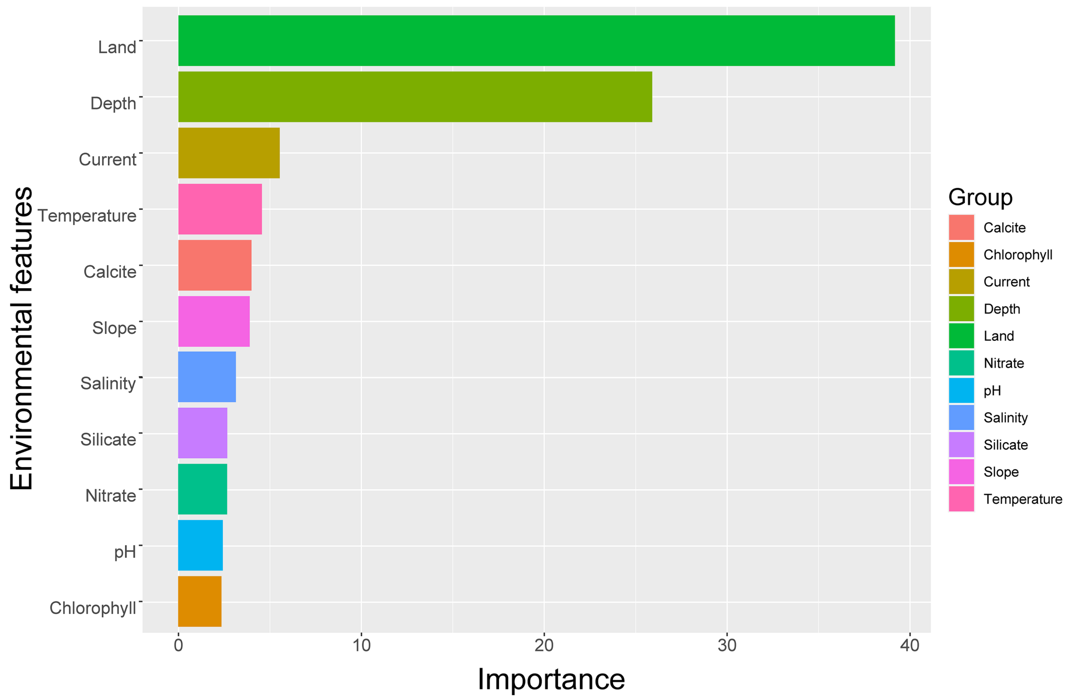

FIGURE 4
Importance analysis of 11 environmental features.

seagrass and makes a more optimistic prediction, but this is not consistent with actual observations (Figure 1). In contrast, the potential seagrass habitat areas estimated by RF-SWOA model are closer to actual observations. From Figure 5, it can be found that the further is the potential seagrass habitat from land, the less likely it exists. This is reflected in both models.

3.4. Model performance evaluation

RF-SWOA and RF models are compared in Figure 6. Their results show that RF-SWOA has a higher AUC, correct classification rate, Kappa, and lower omission rate than RF. RF-SWOA produced a more accurate and stable prediction of seagrass habitat than RF. In Figure 7, the sensitivity and specificity of the proposed model (RF-SWOA) are better than those of RF model. Hybrid machine learning algorithms with higher sensitivity and specificity in prediction can reduce errors in the potential distribution of seagrass, achieving more reliable results.

4. Discussion

4.1. SWOA hybrid model evaluation

Intelligent optimization algorithms are widely used in various engineering practices (Su et al., 2014; Wang et al., 2020; Li et al., 2021), and simple operation is one of the advantages of WOA algorithm. It has excellent optimization capabilities and few parameters, which can dramatically increase the accuracy of the solution and convergence speed in the process of optimizing machine learning functions (Sun et al., 2018; Chakraborty et al., 2021). Although WOA has obvious advantages compared with other intelligent algorithms, it has similar problems like other intelligent algorithms, such as being easily trapped into a local optimum. The SWOA algorithm proposed in this paper can update its position according to its adaptive parameter strategy while updating the optimal individual to achieve the ability of optimizing global search. This study further verified the performance of the SWOA algorithm through simulation experiments. Four standard test functions (Table 3) were used to assess the performance of the algorithm. F1 and F2 test functions were used to determine how quickly and efficiently the SWOA algorithm finds an optimal value (Figures 8A,B). F3 and F4 test functions were used to see if the algorithm can jump out of its local optimum (Figures 8C,D). Each simulation test is to solve the performance of a 1,000-dimensional test function. By testing the performance of the SWOA and WOA algorithms through simulations, the SWOA algorithm has better global convergence and robustness (Figure 8). In this study, a random forest model was also used, which is increasingly being used in ecology due to its predictive accuracy and stability (Cutler et al., 2007; Evans et al., 2011). In particular, random forest models are still very robust at predicting species distributions with limited sample sizes (Luan et al., 2020). After coupling SWOA with RF, we found that the SWOA algorithm greatly influenced the performance of RF on classification. Based on the above findings, hybrid machine learning models can improve predictions of marine species distributions (e.g., seagrasses).

4.2. Environmental drivers of seagrass habitat

The potential adaptability of seagrass habitat is influenced by a combination of environmental elements. In this study, those environmental variables were combined to model potential seagrass habitats. Our results show that the most critical environmental factors affecting seagrass habitat are the distance from land, ocean depth, and current velocity. It reflects the particular importance of physical environmental variables for seagrass habitats. However, this does not mean that chemical and biological types of environmental

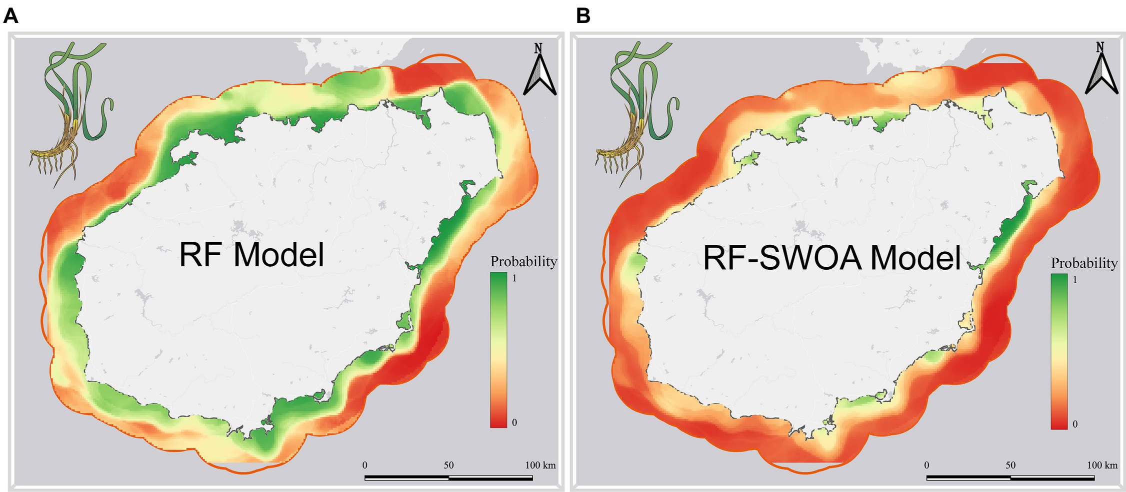

FIGURE 5
Potential habitat areas (Predicted by **(A)** RF model and **(B)** RF-SWOA model).

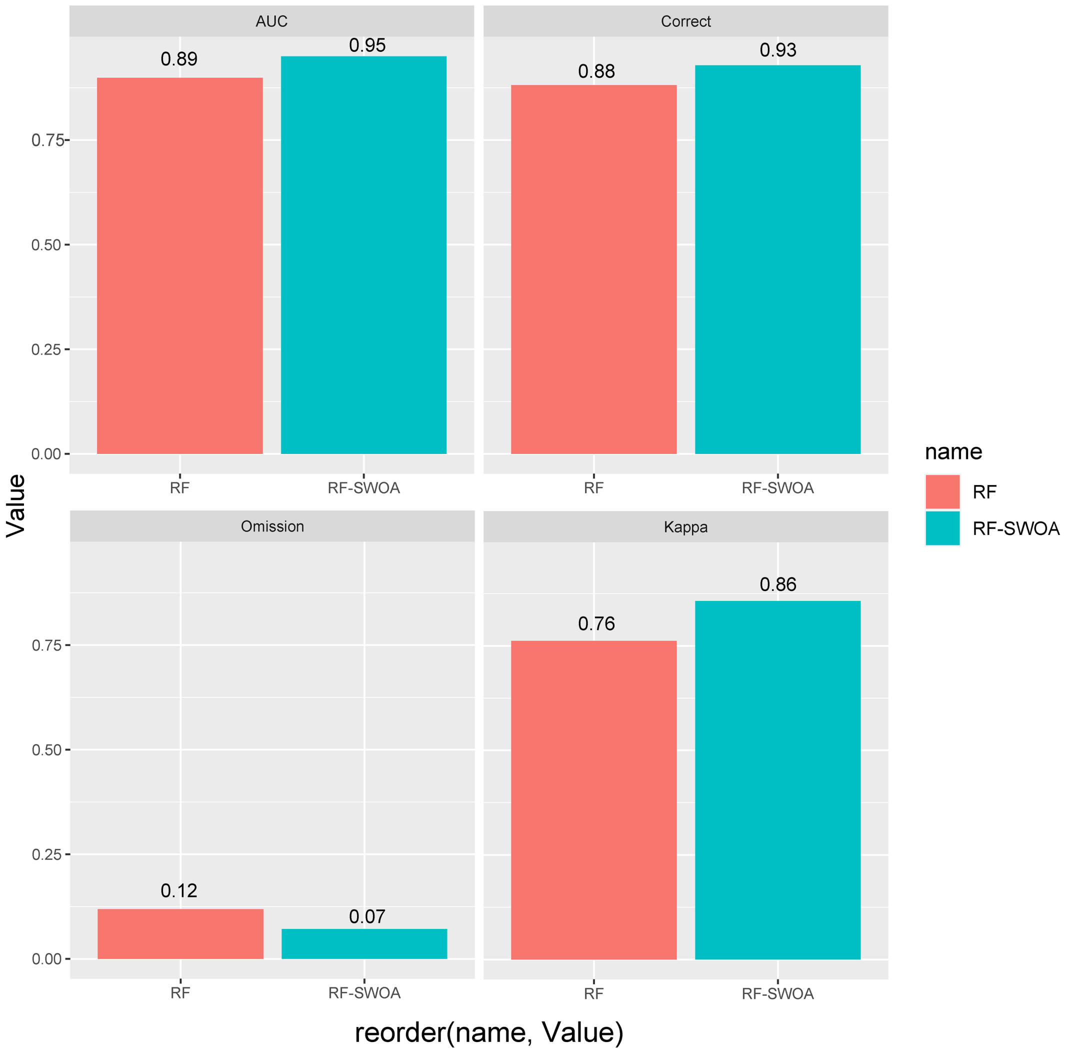

FIGURE 6
RF and RF-SWOA model performance evaluation.

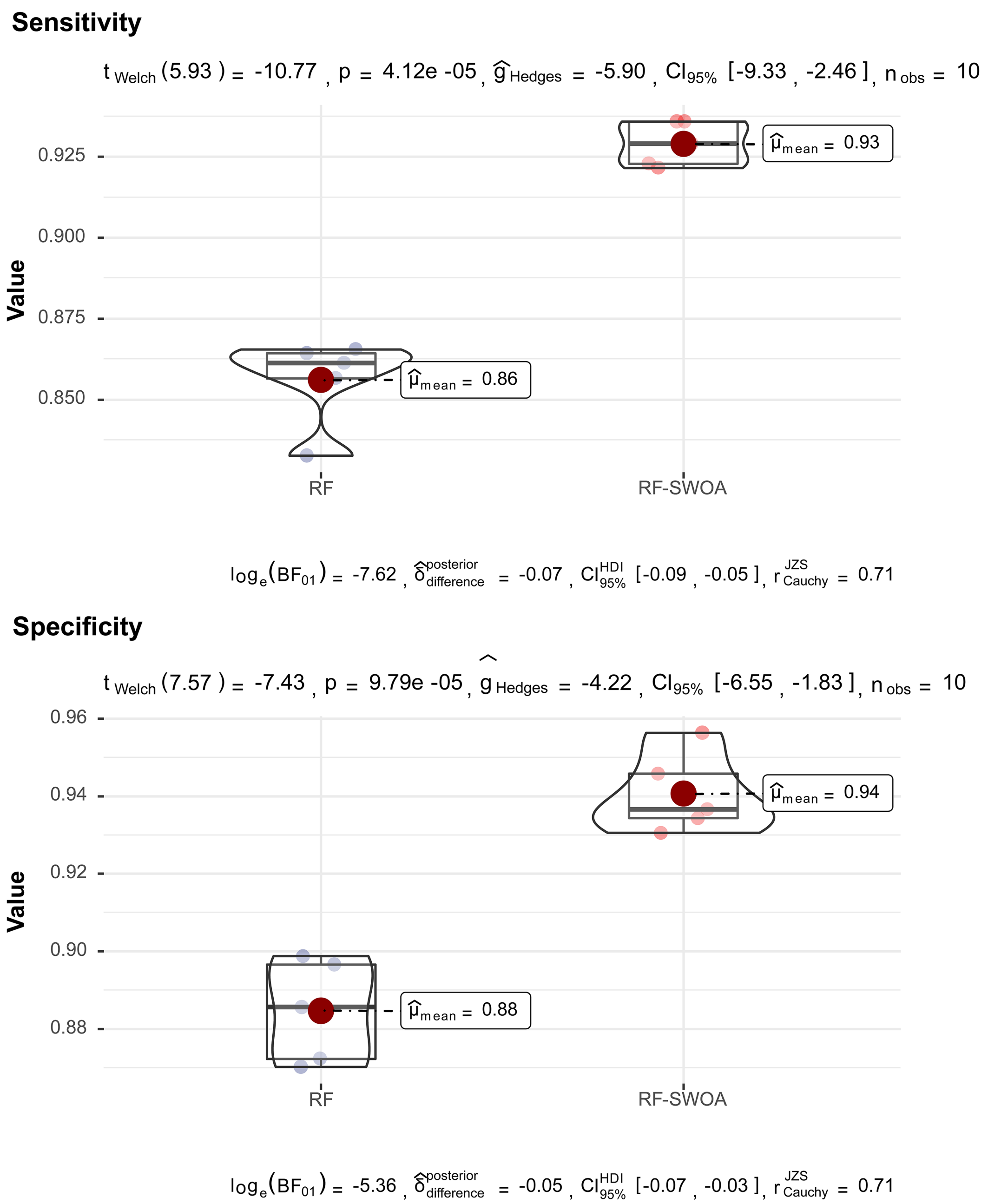

FIGURE 7
Sensitivity and specificity tests of RF and RF-SWOA models. The upper part of the panel shows the statistical test results of frequentist analysis, and the lower part of the panel shows the statistical test results of Bayesian analysis. The results follow the gold standard of statistical reporting (Patil, 2021).

variables do not affect seagrass survival. We found that modeling the distribution of seagrasses in different study areas and scales was influenced by different environmental drivers. A global model showed that the temperature of the sea surface and the distance to the land were the most important environmental variables to predict the distribution of seagrass (Jayathilake and Costello, 2018). At a regional scale, surface nitrate concentration and the availability of benthic light became the most important environmental variables for predicting seagrass distribution in a model of seagrass species distribution in the Gulf of Mexico, while in another sea area, the distance to the sandy shore and depth were the most important environmental drivers (Downie et al., 2013; Bittner et al., 2020). Therefore, we proposed to establish a seagrass habitat simulation in the local study area to identify which environmental factors will lead to seagrass distribution limitation in order to better target local seagrass conservation and restoration (Mao et al., 2022).

5. Conclusion

This study proposed a new hybrid machine learning model (RF-SWOA) to accurately predict suitable habitats for potential seagrasses. The results of this study indicated that the RF-SWOA model could effectively be applied to model seagrass distribution. The results of the RF-SWOA model compared with RF model showed that RF-SWOA was able to identify potential seagrass habitats more accurately and stably. This hybrid machine learning model was demonstrated to be effective in improving the prediction of SDM. The

TABLE 3 Four simulation test functions.

Simulation function expression	Function name	Search space	Global optimum	Characteristic
$F_1 = 1 + \sum_{i=1}^{n} \frac{x_i^2}{4000} - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) \sqrt{b^2 - 4ac}$	Griewank	[−600, 600]	0	Unimodal function
$F_2 = \sum_{i=1}^{n} \lvert x_i \rvert$	Schwefel 2.20	[−100, 100]	0	Unimodal function
$F_3 = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + \exp(1)$	Ackley	[−32, 32]	0	Bimodal function
$F_4 = 10n + \sum_{i=1}^{n}\left(x_i^2 - 10\cos(2\pi x_i)\right)$	Rastrigin	[−5.12, 5.12]	0	Bimodal function

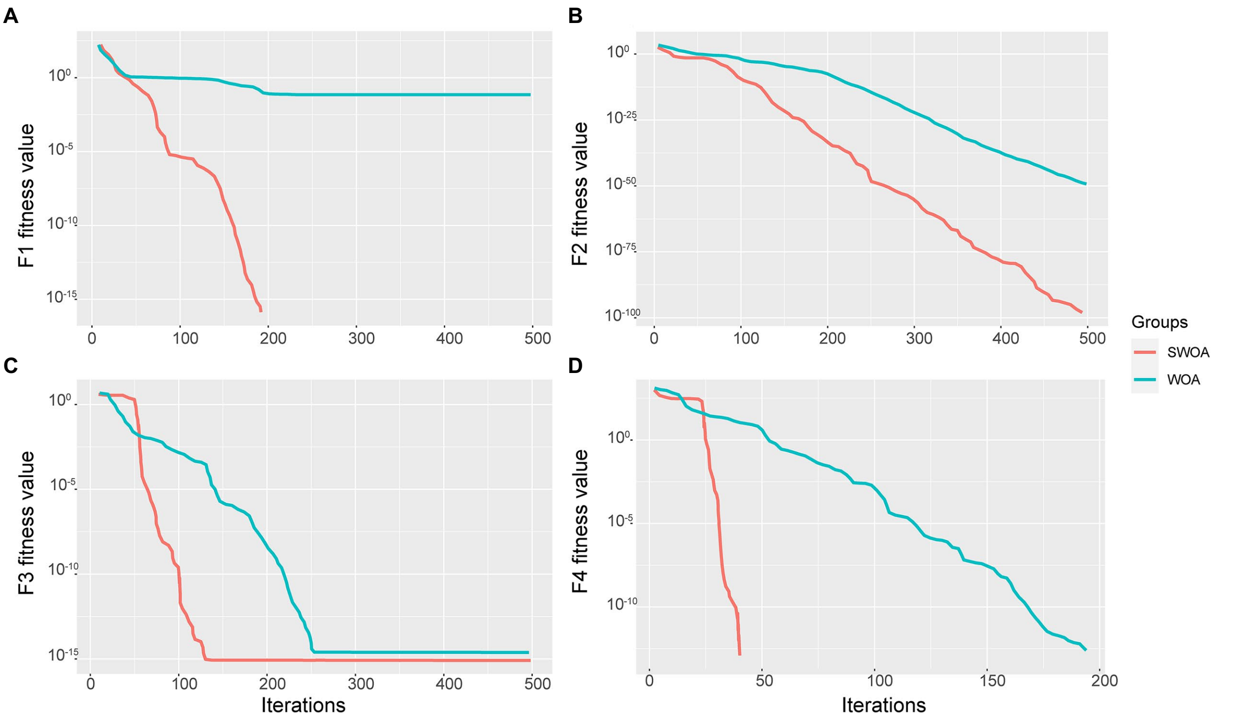

FIGURE 8
RF and RF-SWOA model performance evaluation. **(A)** Griewank simulation function; **(B)** Schwefel 2.20 simulation function; **(C)** Ackley simulation function; **(D)** Rastrigin simulation function.

most important environmental factors affecting seagrass distribution were the distance from land, ocean depth, and current velocity. Therefore, seagrass potential adaptability habitat maps based on the RF-SWOA model can assist in the adequate conservation and restoration of seagrass and provide scientific guidance for seagrass area planning.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BH: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, and funding acquisition. YZ: investigation, resources, data curation, and writing—original draft. SL: writing—review and editing, and supervision. SA: writing—review and editing, supervision, and project administration. WM: conceptualization, resources, data curation, writing—review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Major Science and Technology Project of Hainan Province (ZDKJ202008-1-2), the National Natural Science Foundation of China (42276235), the Start-up funding from Hainan University (kyqd20035), the Innovative Research Projects for Graduate Students in Hainan Province (Qhys2021-16).

Acknowledgments

We would like to thank Biying Jia and Peng Wang for their help in data acquisition. BH sincerely thanks Yingshuai Liu for his help and advice on GIS and remote sensing. The authors would thank university of Maryland's integration and application network.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdel-Basset, M., Abdel-Fatah, L., and Sangaiah, A. K. (2018). "Chapter 10 - metaheuristic algorithms: a comprehensive review," in *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*. eds. A. K. Sangaiah, M. Sheng and Z. Zhang (Academic Press) 185–231.

Aljarah, I., Faris, H., and Mirjalili, S. (2018). Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft. Comput.* 22, 1–15. doi: 10.1007/s00500-016-2442-1

Ardabili, S., Mosavi, A., and Várkonyi-Kóczy, A. R. (2019). "Advances in machine learning modeling reviewing hybrid and ensemble methods", in *International Conference on Global Research and Education* (Springer), 215–227.

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* 3, 327–338. doi: 10.1111/j.2041-210X.2011.00172.x

Beheshti, Z., and Shamsuddin, S. M. H. (2013). A review of population-based meta-heuristic algorithms. *Int. J. Adv. Soft Comput. Appl.* 5, 1–35.

Belkin, I. M. (2021). Remote sensing of ocean fronts in marine ecology and fisheries. *Remote Sens.* 13:883. doi: 10.3390/rs13050883

Bies, R. R., Muldoon, M. F., Pollock, B. G., Manuck, S., Smith, G., and Sale, M. E. (2006). A genetic algorithm-based, hybrid machine learning approach to model selection. *J. Pharmacokinet. Pharmacodyn.* 33, 195–221. doi: 10.1007/s10928-006-9004-6

Bittner, R. E., Roesler, E. L., and Barnes, M. A. (2020). Using species distribution models to guide seagrass management. *Estuar. Coast. Shelf Sci.* 240:106790. doi: 10.1016/j.ecss.2020.106790

Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., et al. (2015). Package 'spdep'. The Comprehensive R Archive Network.

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Chakraborty, S., Saha, A. K., Sharma, S., Mirjalili, S., and Chakraborty, R. (2021). A novel enhanced whale optimization algorithm for global optimization. *Comput. Ind. Eng.* 153:107086. doi: 10.1016/j.cie.2020.107086

Cruz-Duarte, J. M., Amaya, I., Ortiz-Bayliss, J. C., Conant-Pablos, S. E., Terashima-Marin, H., and Shi, Y. (2021). Hyper-heuristics to customise metaheuristics for continuous optimisation. *Swarm Evol. Comput.* 66:100935. doi: 10.1016/j.swevo.2021.100935

Cullen-Unsworth, L. C., and Unsworth, R. (2018). A call for seagrass protection. *Science* 361, 446–448. doi: 10.1126/science.aat7318

Cutler, D. R., Edwards, T. C. Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology* 88, 2783–2792. doi: 10.1890/07-0539.1

De Melo, V. V., and Carosio, G. L. C. (2013). Investigating multi-view differential evolution for solving constrained engineering design problems. *Expert Syst. Appl.* 40, 3370–3377. doi: 10.1016/j.eswa.2012.12.045

Dennison, W. C. (1987). Effects of light on seagrass photosynthesis, growth and depth distribution. *Aquat. Bot.* 27, 15–26. doi: 10.1016/0304-3770(87)90083-0

Dokeroglu, T., Sevinc, E., Kucukyilmaz, T., and Cosar, A. (2019). A survey on new generation metaheuristic algorithms. *Comput. Ind. Eng.* 137:106040. doi: 10.1016/j.cie.2019.106040

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30, 609–628. doi: 10.1111/j.2007.0906-7590.05171.x

Downie, A.-L., Von Numers, M., and Boström, C. (2013). Influence of model selection on the predicted distribution of the seagrass *Zostera marina*. *Estuar. Coast. Shelf Sci.* 121–122, 8–19. doi: 10.1016/j.ecss.2012.12.020

Duarte, C. M. (1990). Seagrass nutrient content. Marine ecology progress series. *Oldendorf* 67, 201–207. doi: 10.3354/meps067201

Duarte, C. M. (1991). Seagrass depth limits. *Aquat. Bot.* 40, 363–377. doi: 10.1016/0304-3770(91)90081-F

Evans, J. S., Murphy, M. A., Holden, Z. A., and Cushman, S. A. (2011). "Modeling species distribution and change using random forest," in *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*, eds. C. A. Drew, Y. F. Wiersma and F. Huettmann (New York, NY: Springer), 139–159.

Fingas, M. (2019). "Chapter 5 - remote sensing for marine management," in *World Seas: An Environmental Evaluation (Second Edition)*. ed. C. Sheppard (Academic Press), 103–119.

Fourqurean, J. W., Duarte, C. M., Kennedy, H., Marbà, N., Holmer, M., Mateo, M. A., et al. (2012). Seagrass ecosystems as a globally significant carbon stock. *Nat. Geosci.* 5, 505–509. doi: 10.1038/ngeo1477

Gonzalez-Irusta, J. M., Gonzalez-Porto, M., Sarralde, R., Arrese, B., Almon, B., and Martin-Sosa, P. (2015). Comparing species distribution models: a case study of four deep sea urchin species. *Hydrobiologia* 745, 43–57. doi: 10.1007/s10750-014-2090-3

Hassan, A., and Pillay, N. (2019). Hybrid metaheuristics: an automated approach. *Expert Syst. Appl.* 130, 132–144. doi: 10.1016/j.eswa.2019.04.027

He, B., Jia, B., Zhao, Y., Wang, X., Wei, M., and Dietzel, R. (2022a). Estimate soil moisture of maize by combining support vector machine and chaotic whale optimization algorithm. *Agric. Water Manag.* 267:107618. doi: 10.1016/j.agwat.2022.107618

He, B., Zhao, Y., and Mao, W. (2022b). Explainable artificial intelligence reveals environmental constraints in seagrass distribution. *Ecol. Indic.* 144:109523. doi: 10.1016/j.ecolind.2022.109523

Hemminga, M. A., and Duarte, C. M. (2000). *Seagrass Ecology*. Cambridge: Cambridge University Press.

Jayathilake, D. R., and Costello, M. J. (2018). A modelled global distribution of the seagrass biome. *Biol. Conserv.* 226, 120–126. doi: 10.1016/j.biocon.2018.07.009

Jiang, Z., Liu, S., Zhang, J., Zhao, C., Wu, Y., Yu, S., et al. (2017). Newly discovered seagrass beds and their potential for blue carbon in the coastal seas of Hainan Island, South China Sea. *Mar. Pollut. Bull.* 125, 513–521. doi: 10.1016/j.marpolbul.2017.07.066

Kaur, G., and Arora, S. (2018). Chaotic whale optimization algorithm. *J. Comput. Des. Eng.* 5, 275–284. doi: 10.1016/j.jcde.2017.12.006

Kendrick, G. A., Nowicki, R. J., Olsen, Y. S., Strydom, S., Fraser, M. W., Sinclair, E. A., et al. (2019). A systematic review of how multiple stressors from an extreme event drove ecosystem-wide loss of resilience in an iconic seagrass community. *Front. Mar. Sci.* 6:455. doi: 10.3389/fmars.2019.00455

Koenig, W. D. (1999). Spatial autocorrelation of ecological phenomena. *Trends Ecol. Evol.* 14, 22–26. doi: 10.1016/S0169-5347(98)01533-X

Kosicki, J. Z. (2017). Should topographic metrics be considered when predicting species density of birds on a large geographical scale? A case of random Forest approach. *Ecol. Model.* 349, 76–85. doi: 10.1016/j.ecolmodel.2017.01.024

Kosicki, J. Z. (2020). Generalised additive models and random Forest approach as effective methods for predictive species density and functional species richness. *Environ. Ecol. Stat.* 27, 273–292. doi: 10.1007/s10651-020-00445-5

Krause-Jensen, D., Quaresma, A. L., and Cunha, A. H. (2004). How are seagrass distribution and abundance monitored. *European seagrasses: an introduction to monitoring and management*, 45–53.

Kuo, J., and Den Hartog, C. (2001). Seagrass taxonomy and identification key. *Glob. Seagrass Res. Methods* 33, 31–58. doi: 10.1016/B978-044450891-1/50003-7

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659–1673. doi: 10.2307/1939924

Li, X., and Wang, Y. (2013). Applying various algorithms for species distribution modelling. *Integr. Zool.* 8, 124–135. doi: 10.1111/1749-4877.12000

Li, W., Wang, G.-G., and Gandomi, A. H. (2021). A survey of learning-based intelligent optimization algorithms. *Arch. Comput. Methods Eng.* 28, 3781–3799. doi: 10.1007/s11831-021-09562-1

Luan, J., Zhang, C., Xu, B., Xue, Y., and Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fish. Res.* 227:105534. doi: 10.1016/j.fishres.2020.105534

Mao, W., Zhao, Y., He, B., Jia, B., and Li, W. (2022). Review on degradation mechanism and restoration strategies of seagrass ecosystem. *J. Desert Res.* 42:87.

Mckenzie, L. J., Finkbeiner, M. A., and Kirkman, H. (2001). "Chapter 5 - Methods for mapping seagrass distribution," in *Global Seagrass Research Methods*. eds. F. T. Short and R. G. Coles (Amsterdam: Elsevier Science), 101–121.

Mi, C., Huettmann, F., Guo, Y., Han, X., and Wen, L. (2017). Why choose random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ.* 5:e2849. doi: 10.7717/peerj.2849

Miller, J. (2010). Species distribution modeling. *Geogr. Compass* 4, 490–509. doi: 10.1111/j.1749-8198.2010.00351.x

Mirjalili, S., and Lewis, A. (2016). The whale optimization algorithm. *Adv. Eng. Softw.* 95, 51–67. doi: 10.1016/j.advengsoft.2016.01.008

Moksnes, P. O., Röhr, M. E., Holmer, M., Eklöf, J. S., Eriander, L., Infantes, E., et al. (2021). Major impacts and societal costs of seagrass loss on sediment carbon and nitrogen stocks. *Ecosphere* 12:e03658. doi: 10.1002/ecs2.3658

Moya, I., Bermejo, E., Chica, M., and Cordon, O. (2021). Coral reefs optimization algorithms for agent-based model calibration. *Eng. Appl. Artif. Intell.* 100:104170. doi: 10.1016/j.engappai.2021.104170

Nguyen, H. M., Ralph, P. J., Marín-Guirao, L., Pernice, M., and Procaccini, G. (2021). Seagrasses in an era of ocean warming: a review. *Biol. Rev.* 96, 2009–2030. doi: 10.1111/brv.12736

Orth, R. J., Carruthers, T. J., Dennison, W. C., Duarte, C. M., Fourqurean, J. W., Heck, K. L., et al. (2006). A global crisis for seagrass ecosystems. *Bioscience* 56, 987–996. doi: 10.1641/0006-3568(2006)56[987:AGCFSE]2.0.CO;2

Ouellette, W., and Getinet, W. (2016). Remote sensing for marine spatial planning and integrated coastal areas management: achievements, challenges, opportunities and future prospects. *Remote Sens. Appl.* 4, 138–157. doi: 10.1016/j.rsase.2016.07.003

Patil, I. (2021). Visualizations with statistical details: The'ggstatsplot'approach. *J. Open Source Softw.* 6:3167. doi: 10.21105/joss.03167

Picart, S. S., Sathyendranath, S., Dowell, M., Moore, T., and Platt, T. (2014). Remote sensing of assimilation number for marine phytoplankton. *Remote Sens. Environ.* 146, 87–96. doi: 10.1016/j.rse.2013.10.032

Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., and Gloaguen, R. (2020). COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics* 8:890. doi: 10.3390/math8060890

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., et al. (2014). Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods Ecol. Evol.* 5, 397–406. doi: 10.1111/2041-210X.12180

Python, W. (2021). Python. *Python Releases for Windows* 24.

Saranya, K. R. L., Lakshmi, T. V., and Reddy, C. S. (2021). Predicting the potential sites of Chromolaena odorata and Lantana camara in forest landscape of eastern Ghats using habitat suitability models. *Eco. Inform.* 66:101455. doi: 10.1016/j.ecoinf.2021.101455

Short, F., Carruthers, T., Dennison, W., and Waycott, M. (2007). Global seagrass distribution and diversity: a bioregional model. *J. Exp. Mar. Biol. Ecol.* 350, 3–20. doi: 10.1016/j.jembe.2007.06.012

Short, F. T., and Wyllie-Echeverria, S. (1996). Natural and human-induced disturbance of seagrasses. *Environ. Conserv.* 23, 17–27. doi: 10.1017/S0376892900038212

Singh, P., and Kottath, R. (2021). An ensemble approach to meta-heuristic algorithms: comparative analysis and its applications. *Comput. Ind. Eng.* 162:107739. doi: 10.1016/j.cie.2021.107739

Su, Z., Wang, J., Lu, H., and Zhao, G. (2014). A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. *Energy Convers. Manag.* 85, 443–452. doi: 10.1016/j.enconman.2014.05.058

Sun, Y., Wang, X., Chen, Y., and Liu, Z. (2018). A modified whale optimization algorithm for large-scale global optimization problems. *Expert Syst. Appl.* 114, 563–577. doi: 10.1016/j.eswa.2018.08.027

Talbi, E.-G. (2016). Combining metaheuristics with mathematical programming, constraint programming and machine learning. *Ann. Oper. Res.* 240, 171–215. doi: 10.1007/s10479-015-2034-y

Tsai, C.-F., and Chen, M.-L. (2010). Credit rating by hybrid machine learning techniques. *Appl. Soft Comput.* 10, 374–380. doi: 10.1016/j.asoc.2009.08.003

Wang, J., Du, P., Hao, Y., Ma, X., Niu, T., and Yang, W. (2020). An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting. *J. Environ. Manag.* 255:109855. doi: 10.1016/j.jenvman.2019.109855

Waycott, M., Duarte, C. M., Carruthers, T. J., Orth, R. J., Dennison, W. C., Olyarnik, S., et al. (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12377–12381. doi: 10.1073/pnas.0905620106

Yang, X.-S. (2009). "Harmony search as a metaheuristic algorithm," in *Music-Inspired Harmony Search Algorithm: Theory and Applications*. ed. Z. W. Geem (Berlin, Heidelberg: Springer), 1–14.

Yang, X.-S. (2013). Optimization and metaheuristic algorithms in engineering. *Metaheuristics in Water, Geotechnical and Transport Engineering* 1:23.

Yang, D., and Yang, C. (2009). Detection of seagrass distribution changes from 1991 to 2006 in Xincun Bay, Hainan, with satellite remote sensing. *Sensors* 9, 830–844. doi: 10.3390/s90200830

Zheng, F., Qiu, G., Fan, H., and Zhang, W. (2013). Diversity, distribution and conservation of Chinese seagrass species. *Biodivers. Sci.* 21, 517–526. doi: 10.3724/SP.J.1003.2013.10038

Zimmermann, N. E., Edwards, T. C. Jr., Graham, C. H., Pearman, P. B., and Svenning, J. C. (2010). New trends in species distribution modelling. *Ecography* 33, 985–989. doi: 10.1111/j.1600-0587.2010.06953.x