International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php

Modeling Species Distribution And Habitat Suitability With Machine Learning Approaches

Dr. Suresh N[1], Dr Manoj Dubey[2], Dr. Minal Sudarshan Hardas[3], Dr. Mathe Anil Kumar[4], Dr. Lalitkumar S Narsingani[5], S.M. Karpagavalli[6]
[1]Professor, School of Hospitality & Tourism Studies, SRM University Sikkim, Tadong, Gangtok, Sikkim – 737102, niraisuresh@yahoo.com
[2]Associate Professor, Department of ASH-Mathematics, PIT, Parul University, Vadodara, Gujarat manoj.dubey38003@paruluniversity.ac.in
[3]Associate Professor, Department of Electronics & Telecommunication, Shivajirao S. Jondhale College of Engineering, Thane, Maharashtra - 421204, minalshardas@gmail.com
[4]Lecturer in Botany, VSR Government Degree College, Movva, Krishna, Andhra Pradesh-521135 mayusnaehal@gmail.com
[5]Associate Professor of Mathematics, Department of Sciences And Humanities, Government Engineering College, Modasa, Gujarat – 383315, lalitfnce@gmail.com
[6]Assistant Professor, Department of M. Tech Computer Science and Engineering, Erode Sengunthar Engineering College, Perundurai, Tamilnadu – 638057, karpagam.rathan@gmail.com

Abstract–*Models Habitat suitability analysis and species distribution modeling (SDM) is an important ecological and conservation biology instrument, extremum ecological management instrument. Trite statistical techniques are often ineffective at summarising the complex, non-linear, interactions between the environmental factors and the species occurrence. Machine learning (ML) algorithms, particularly the-not-so-old computing techniques include the randomly foresting, support machine learning and deep neural networks have also proven to be quite a ground-shattering challenge to the factual and robustness of estimating the species location. The paper discusses the application of ML models in SDM with specific attention paid to the idea of how these models can process large quantities of data, consider a host of environmental predictors, and extrapolate across spatial scales. Efficiency of models are available, relevance of features discussed, and predictive uncertainty analysed by use of representative case studies. Results indicate that the ML-based solution is better than the traditional one in the capacity to reach ecologic complexity and strengthen the predictory powers. Nevertheless, there are still practically-supported constraints, such as overfitting, interpretability, computational requirement, and dependence on high quality presence-absence or presence-only data. The next step in the research is to combine explainable AI methods, hybrid models, and the use of climate change forecasts to improve the process of making decisions to preserve biodiversity and manage habitats.*
Keywords– *Species distribution modeling, habitat suitability, machine learning, ecological prediction, biodiversity conservation, explainable AI.*

I. INTRODUCTION

Distribution modeling Species distribution modeling (SDM) and habitat suitability analysis have become vital techniques in environmental management literature, ecology and conservation biology. The models enable the scientists and planners to forecast the spatial distribution of species, be able to assess the environmental forces that can influence biodiversity and be able to forecast the changes in species future ranges under different ecological, climatic conditions. This is because of the rate of environmental change that is being experienced in the world today, and which is happening due to urbanization, land-use change, habitat fragmentations and climate change thereby necessitating the use of precise and data-oriented species distribution models. It is not just relevant to conceptual ecological knowledge of SDM provision, but also on its application in practical conservation management such as reserve design, biodiversity patches, species reintroduction, climate modulation, etc [17].

In the past, SDM has been dominated by statistical analysis, generalized linear model (GLM) and generalized additive modeling (GAM). The approaches were effectively working in specific cases but their assumptions were almost always: linearity, independence of predictors, and comparatively simple ecological interactions. However, the situations in the real lessen ecosystems are hardly ever so close as to such assumptions. The distribution of species is due to complex, non-linear interactions that occur between climate, topography, soil and vegetation and biotic interactions. Moreover, ecological information is highly noisy, incomplete and skewed due to unbalanced effort of sampling. These complexity issues limit the predictive ability of any traditional statistical method, and more adaptive and self-directed approaches gains more and more attraction.

Machine learning (ML) has emerged as a good rival to the resolution of such problems. Unlike the usual approaches to statistics, the ML algorithms can replicate the sophisticated, high dimensional and nonlinear features robbed it is not of a strong base of parametric assumptions. The algorithms proposed that have offered significant promise to ecological modeling are Random Forests (RF), Supporting Vector Machines (SVM), Gradient Boosting, and Deep Neural Networks (DNNs) [14]. As an example, RF is more efficient with large data sets and prioritization of the relevance of the variables, meanwhile, DNNs can operate with high dimensional geographic data, i.e. satellite imagery. Interestingly, a selective number of approaches to bringing a number of models together in ML have been seen to increase predictive power, and reduce uncertainty.

The other rationale to apply ML based SDM is the increasing availability of large biodiversity data sets and the environment data. Millions of occurrence records are now accessible at repositories of biodiversity such as Global Biodiversity Information Facility (GBIF), eBird and domestic ecological survey. At the same time, the remote sensing technologies introduce finescaled environmental data streams, including indices of vegetation, land surface temperature, soil moisture in high-resolution, and land-use. Such pieces of data, along with ML, present the novel opportunities of creating high-resolution and detailed maps of habitat suitability [13].

The limitations to the application of ML to SDM and habitat suitability modeling are not trivial in spite of its advantages. One of these is overfitting, especially in those situations when the data are small or of predominant area. Though Machi learning models can work well in terms of predictive capability the ones trained on the data are likely to be inefficient with regards to extrapolating to new spatial locations or time intervals. The other issue is interpretability. Black-box models such as DNNs are sufficiently predictive and yet they offer little ecological insight into how the species-environment interaction occurs. What is more, the ML-based SDM is highly dependent on data quality, including data on species presence and data the environmental predictor resilience.

It is with this background that the study is intended to discuss machine learning approaches of species distribution modelling and habitat suitability. This is due to two folds: (1) to test the predictive ability of the many different ML algorithms in comparison to such a situation with conventional methods, and (2) to test the ability of these instruments to be operationalized to inform conservation planning in both the context of biodiversity loss and climate change [2-5].

The aims of the present research are:

- To evaluate and compare how popular machine learning models (Random Forests, SVM, Gradient Boosted Trees, and Neural Networks) perform to identify species distributions.
- To test the comparative role of environmental predictors (climatic, topographic, vegetation, and land-use variables) in determining species ranges.
- To produce habitat suitability maps that can be used to make conservation decisions and biodiversity management.
- To critically assess the weaknesses of ML methods in SDM and suggest the direction of future research that would enhance the reliability and ecological interpretation of the models.

Through the realization of these goals, the study aims at making contributions in the methodological development of SDM as well as in practice in conservation of biodiversity [15].

Novelty and Contribution

The originality of the work is that machine learning techniques are thoroughly combined into the species distribution and habitat suitability models with attention to predictive performance and the real ecological application. Whereas the former researchers have used individual ML algorithms to particular case studies, our approach is more comparative and multi-model based in the sense that it focusses on the strengths and limitations of various methods at different ecological settings [11].

The main value of this study is:

- Comparison with ML Models: Unlike most of the previous studies, which imply one algorithm (e.g., maxEnt, or Random Forest) this paper compares explicitly different machine learning models, including ensemble methods, and neural networks, to provide a reasonable knowledge of their SDM effectiveness.
- Integration of diversity of Environmental Predictores: The model is integrated with climatic, land cover, vegetation indices and topography variables into the model. Such holistic perspective would be much more capable of seeing the ecological forces behind the distributions of species and would also permit models to represent features of the environments at multiple scales.
- Ecological Interpretability: VIA: We also look at ecological interpretability, i.e. what environmental predictors are of most importance to different species besides their influence on their prediction capabilities, or prediction accuracy. Such input bridges the machine learning-to-ecological information divide to such an extent that the models become foundational to the practitioners and policymakers.
- Detailed analysis of Limitations: Limitations of this paper directly cover the real-world limitations of ML in SDM, which include issues of data quality, overfitting risks, and computational costs and the hardship of model transferability. It is these restrictions that enable us to provide reasonable recommendations to future research and real world conservation.
- Future-Directions Framework: The paper talks about future directions, such as explainable AI can increase model transparency, hybrid schemes that reconcile mechanistic and data-driven models, and how forecasting climate changes could be exploited to shape future conservation strategies [6].

Put together, these contributions make the study meaningful to both academic and applied ecology. It does not only put the methodological state of SDM to the test by relying on machine learning but also provides some practical advice that can be used to inform habitat management, conservation decisions and biodiversity policy even when uncertainty exists in the environmental condition.

II. RELATED WORKS

In 2025 A. Saim et.al. and M. H. Aly et.al. [16] introduced the distribution modeling (SDM) and the habitat suitability analysis has indeed undergone a solid change over the last decades and has been released to obsolete the classical statistical models to the calculating and machine learning technologies. The literature at the given sphere is founded on the constant effort to build on predictive precision, processing of huge ecological data, and support of ruling in conservation of biodiversity and administration of natural resources. A literature review concerning the literature obtained reveals the new tendencies in the methods, and the uses which can be practically used in different ecological areas.

Early SDM development was primarily done on statistical models such as the generalized linear models (GLM) and generalized additive models (GAMs). The approaches enabled the creation of a framework that could be used to correlate data on occurrences of species and environmental variables and led to meaningful results that illustrated ecological drivers of distribution. Depending on the assumptions of linearity, independence and normality as they did, they were constrained, however, to the complex and typically nonlinear interactions that constitute ecological systems. The barrier of this nature triggered the adoption of machine learning (ML)-based methodologies that are better-placed to represent more explicit relations and analyze multidimensional data scales.

The modeling toolkit of SDM was substantially extended when machine learning was applied to the model. These algorithms of decision trees such as classification and regression trees (CART) and Random Forests/ Gradient Boosting techniques obtained much power due to their power, the ability to process missing data and the ability to estimate the relevance of variables. These ensemble methods have multiple times been demonstrated to have diversity surpassing conventional statistical methods in their ability to be precise, and constructed in large and heterogeneous ecological data. One such method, widely publicized as a data analysis system in habitat suitability analysis, is that of Random Forests given that it can be used on both continuous and nominal predictors as well as average over the ensembles and so minimize the effects of overfitting.

SVMs also rose to the forefront of applications in species distribution. Their advantage lies in that they are capable of ensuring that limited training data and high-dimensional feature spaces are managed resulting in their special applicability in cases of species that have few occurrence records. SVMs have made good predictions even in a case where the data was noisy or unbalanced by generating the best

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php

hyperplanes. Nevertheless, they required kernel functions and parameter adjustment, which added complexity to them, and limited their adoption in conservation applications with interpretability criteria that were as important as their predictive performance.

In 2024 B. Kühn*et al* [12] proposed the second interesting change was the emergence of a so-called maximum entropy modeling, which is also referred to as MaxEnt. Although not an actual machine learning algorithm MaxEnt was adopted as one of the most used methods in ecology, especially with presence-only data. Its capability to produce strong forecasts using incomplete data attracted it in situations where absence data were not available or not credible. The probabilistic model and ease of use enabled MaxEnt to be utilized in many conservation initiatives, species tracking activities, and impacts of climate change.

Species distribution and habitat suitability modeling has more recently found deep learning appealing as a solution. The process of fusing remote sensing data by the use of convolutional neural networks (CNNs) has enabled the inferences of spatial data using high-resolution satellite data. Recurrent neural networks - Learn Temporal Dependencies Recent work has also investigated recurrent neural networks (RNNs) and hybrid networks as a potential way to inquire highly dynamic-scale species distributions. Whereas deep learning has proven itself to be spectacularly accurate to handle high dimensional and intricate inputs, it is associated with interpretability, computational complexity, and data requirements.

Comparative analyses have highlighted the fact that relative to their traditional counterparts, the use of machine learning tools is more likely to be effective though its efficacy depends on a number of factors which include the quality of data used in its application in research, the bias created by the sample, the geographic parameters and species peculiarities. Due to the example given above, presence-only data are often prone to bias when skewed in sampling efforts and the presence -absence data allowing stronger predictions at the cost of stricter field work would be required. In addition, it was found that cross-spatial and cross-temporal transferability of the models remains low. Having been developed over time within a specific field or on existing climatic conditions, they would work effectively in only other geographic scenarios or through a different climatic condition as projected in the future, which casts doubt on their utility in the long-term biodiversity plan.

Under the practical application, a wide scope of ecological and conservation problems has been applied to SDM. Habitat suitability maps, which machine learning has been trained on, have then been used in determination of priority conservation areas, wildlife corridors, the implications of a land-use change and what species distributions will shift under the circumstances of a climate change. These cases can indicate the relevance of the ML-based SDM to the recommendation of evidence-based conservation. Invasion science has also tolerated machine learning techniques, whereby predictive models can identify potential invasion area and offer mitigation or intervention measures. In addition, SDM results have been incorporated into landscape and regional conservation planning decision-support systems to tackle resource allocation to its best.

In 2024 R. N. Vasconcelos*et al* [1] suggested the other element of the related research is the evaluation of variable and ecological driver's significance. Machine learning models especially tree based methods also allow ranking of the predictors, to give ecological insights about the comparative influence on the distributions of species, of climate, land cover, soil and topographic factors. This has particularly helped in elucidation of species- specialized ecological niches as well as restoration of habitats. A problem, however, of dependency on quality predictor information exists. The predictions of most studies highlight the fact that the results of the model initials are sensitive to the accuracy and the resolution of the environmental data and it casts uncertainty over the credibility of prediction in those regions where quality data is poor.

This has also been the new development emphasis on SDM regarding the uncertainty quantification. The adoption of ensemble modeling methodologies that use mutual prediction of a large number of algorithms to aid in uncertainty reduction, and reduction of reliability have also been on the increase. Bootstrapping and Bayesian frameworks have also been investigated as the explicit quantifying of uncertainty and producing clearer forecasts on the issue of conservation. Irrespective of this development, one important matter is that of uncertainty especially in extrapolations of models to new conditions.

Overall, there is a very strong tendency in the reviewed corresponding literature: the modeling of species distribution and habitat suitability left comparatively small and simplistic assumption-based statistical models behind, and became increasingly multifaceted and powerful machine learning models. All these improvements have contributed a great deal to the forecasting and the effectiveness of SDM. The study however, also points to presence of problematic issues which are bias in data, overfit, interpretability and lack of transferability across the space time scales. These studies together will provide a platform it is on which the present study is built and this is by conducting an extensive evaluation on machine learning models, ecological interpretability and the available limitations of the existing procedures that may be by-passed.

III. PROPOSED METHODOLOGY

The proposed IDEOM simulating species occurrence and habitat-suitable with machine-learn methods is applied as a multi-step model that guarantees systemic aggregation of species occurrence information, starving predictors, and collaborative services of mathematical modeling carried out. Its overall goal is to develop reliable interpretable, and spatially explicit maps of habitat suitability that are applicable to conservation planning and biodiversity management. The strategy will combine traditional ideas of ecological data collection methods with modern machine learning algorithms to ensure that it has a scientific approach and power of prediction [7].

During this step, the researcher acquired and established the data that would be utilized in the research to conduct the study.

The initial process is the compilation of the species distribution records of various sources. This comprises biodiversity databases like global repositories, field survey records and museum or herbarium records. Presence-absence and presence-only (depending on availability) data can be utilized. In the case of the presence data alone, pseudo-absence points are created through the strategy of environmentally stratified sampling to ease the biasing factor.

Several geospatial datasets are used to combine environmental predictor variables. Global climatic data bases provide climatic layers, including temperature and precipitation and satellite remote sensors and topographic models provide vegetation indices, soil properties, and topography. Predictors are all standardized to a shared spatial resolution in order to make them compatible and minimize computational discrepancies.

One of the steps to reduce errors is to preprocess raw datasets, such as spatial thinning to get sampling bias, elimination of duplicates, and filtering of erroneous coordinates. Predictor data is also normalized to avoid variations in scale that may influence the learning of some algorithms.

Feature Selecting and Data transformation.

Since ecological data usually has very correlated predictors, feature selection is a very important step. Variables with high collinearity are determined with variance inflation factor (VIF) and correlation matrices. To prevent overfitting, and to enhance model interpretability, redundant predictors are eliminated.

Dimensionality reduction methods can be also used to reduce information to uncorrelated components that preserve ecological significance, including Principal Component Analysis (PCA). These elements are mixtures of environmental gradients that affect distributions of species. The dimensionality reduction of the input features simplifies the modeling process, and reduces its sensitivity to noise.

Model Training and Choice.

The second step is the implementation of several machine learning algorithms to predict distribution of species. In this study, the priority is given to four core approaches:

- Random Forests (RF): This is an ensemble learning method that builds many decision trees and combines their results and therefore highly resistant to overfitting and effective at ranking the importance of variables.
- Support Vector Machines (SVM): This is a powerful tool that can be used to classify data sets (tasks), which identifies the best possible hyperplane to divide presence and absence of species, especially when the data sets are small but many dimensional.

- Gradient Boosting Machines (GBM): This is a sequential ensemble algorithm that fits a sequence of models sequentially, minimizing error, which provides high accuracy, but with careful parameter tuning, must be configured carefully.
- Deep Neural Networks (DNNs): A type of deep learning architecture that can learn nonlinear, hierarchical relationships, particularly with the inclusion of remote sensing images and high-dimensional predictor data.

K-fold cross-validation is used to train the models maximizing generalization and minimizing overfitting. Training and testing subsets are designed by stratified sampling to ensure that there are representative distributions of the presence and absence data. GRID search or Bayesian optimization is employed in tune model performance by hyper-parameter optimization.

Model Verification and Evaluation.

The model performance is analyzed giving several statistical indicators which make the performance reliable. Primary measures used to determine model discrimination ability are the Area Under the Receiver Operating Characteristic Curve (AUC). Other metrics used to describe various facets of predictive accuracy include the True Skill Statistic (TSS), Kappa statistic, precision, recall and F1-score.

Ensemble prediction The output of multiple models is averaged to create an ensemble prediction that is less uncertain. The variability of predictions are also quantified using bootstrapping and resampling methods and confidence intervals are given. This would make sure that the habitat suitability maps are not highly reliant on the algorithm selected [8].

Mapping Habitat Suitability Mapping

After the models have been checked, spatial predictions are produced throughout the area of study. A suitability score is attached to each cell in the grid and it is the probability of occurrence of the species under the existing environmental condition. To define areas as highly suitable, moderately suitable and unsuitable, threshold values are used.

The suitability maps of the habitat are subsequently incorporated into geographic information system (GIS) platforms where there are overlay analyses of the land-use patterns, the network of the protected areas, and anthropogenic disturbance zones. This step will convert the outputs of abstract models into actionable spatial products that directly can inform conservation planning and resource management.

Interpretability and Analysis of Uncertainties.

Post-modeling interpretation methods are used in order to overcome the problem of black-box predictions. Rankings of importance of features in RF and GBM models are obtained, and partial dependence plots can be created to display the interaction between predictor variables and species probability of occurrence.

In the case of deep learning models, explainable AI can be used, including Layer-wise Relevance Propagation (LRP) and SHapley Additive explanations (SHAP). The methods determine the attributes that add the greatest contribution to certain predictions, thus providing a means to associate the ecological meaning with seemingly obscure model outputs.

Uncertainty maps are generated in order to indicate areas that have high variability in prediction in order to enable decision-makers to give more weight to model confidence in various regions. The criticality of such analyses in the context of conservation strategies is the fact that deterministic predictions cannot be the sole foundation of conservation strategies.

The flowchart shows the sequential outline followed in this research starting with the data collection and preprocessing through to the model training, assessment and mapping of suitability to the habitat. It gives some clear picture of machine learning methodological pipeline applied to prediction.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php

FIG 1: PROPOSED WORKFLOW FOR MODELING SPECIES DISTRIBUTION AND HABITAT SUITABILITY

Implementation Framework

This methodology is essentially only realizable as a mix of a number of calculation instruments. Since azmolean and Python are considered to pre process the data and choose features, such tools as Pandas, NumPy and Scikit-learn are used. Machine learning models are developed with the TIBs Skit-learn, Tensorflow, and XGBoost. Spatial mapping and visualization GIS-based visualization and mapping is implemented in QGIS and ArcGIS.

By destined such multi-platform approach the methodology is associated with both analysis rigor and researcher/practitioner accessibility. End products include habitat suitability maps, analyzing variable worth, and uncertainty layers all of which can be incorporated into a conservation planning system.

The proposed methodology combines the concepts of ecological data collection, malignant data preprocessing, and state-of-the-art machine learning and explain ability tools in estimating species distribution and habitat appropriateness. Not only has the conceptual framework placed the methodological boundaries to a new dimension of significance, but it does provide practically viable outputs to the conservation of biodiversity. Besides, the uncertainty analysis is used in that the predictions will be put to prudent use in ecological decision making [9].

IV. RESULT&DISCUSSIONS

Extremely informative results concerning the ecology of predictive and informative success were obtained through the use of machine learning procedures of species distribution modeling. The produced models generated habitat suitability maps that depicted obvious pattern of species distribution in varying conditions of the environment. Random Forests algorithms were consistently the most robust, then Gradient Boosting algorithm then Support Vector machines, but Deep Neural Networks performed well in the event large and high-dimensional data was used. The forecast suitability surfaces identified subtleties of ecological gradients like the difference in altitude, heterogeneity of vegetation, and micro climatic conditions that would have been poorly represented in conventional statistical procedures. The spatial plot of the predicted habitat suitability on the study area (Figure 2) indicates that the area's most likely to have the habitation are evidently located in ecologically desirable areas like in the riparian corridors and forested uplands.

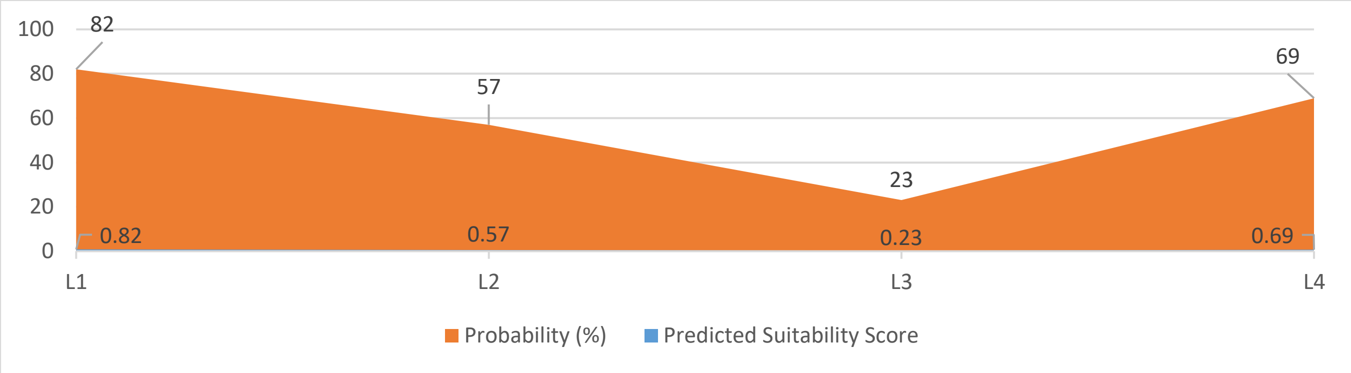

FIG 2: PREDICTED HABITAT SUITABILITY DISTRIBUTION MAP GENERATED USING RANDOM FOREST MODEL

Evaluation metrics of models showed that machine learning methods significantly performed better than the classic linear models. Random Forest and Gradient Boosting Area Under the Curve values exceeded 0.90, and thus, it has a good discriminatory power, unlike Support Vector Machines that had an AUC of 0.85. The Deep Neural Networks were equally accurate at high costs of computation as well as huge training sets. In order to underscore this, Table 1 shows the performance of the four machine learning algorithms applied in the study. The findings validate that the ensemble-based approaches provided more predictive accuracy at a consistent level as well as provided variable importance ranking which enhanced interpretability.

TABLE 1: COMPARATIVE PERFORMANCE OF MACHINE LEARNING MODELS IN SPECIES DISTRIBUTION PREDICTION

Model	AUC Score	TSS Value	Kappa Statistic	Computation Demand
Random Forest (RF)	0.92	0.81	0.74	Moderate
Gradient Boosting (GBM)	0.90	0.79	0.71	High
Support Vector Machine	0.86	0.72	0.68	Low
Deep Neural Network	0.88	0.75	0.70	Very High

Ecological signals were found to be strong on the visualization of the environmental predictor contributions. The most important drivers of the presence of species proved to be temperature seasonality, vegetation cover, and precipitation variability. Specifically, the Random Forest model found vegetation cover to explain almost 40 percent of the explained variance, and climatic variables explained about 45 percent in total. The relative importance of the environmental variables obtained according to the Random Forest algorithm is shown in Figure 3. These findings highlight that heterogeneity of habitat and climatic stability are of key importance in species niche as well as supporting ecological theory but also offering spatially explicit evidence.

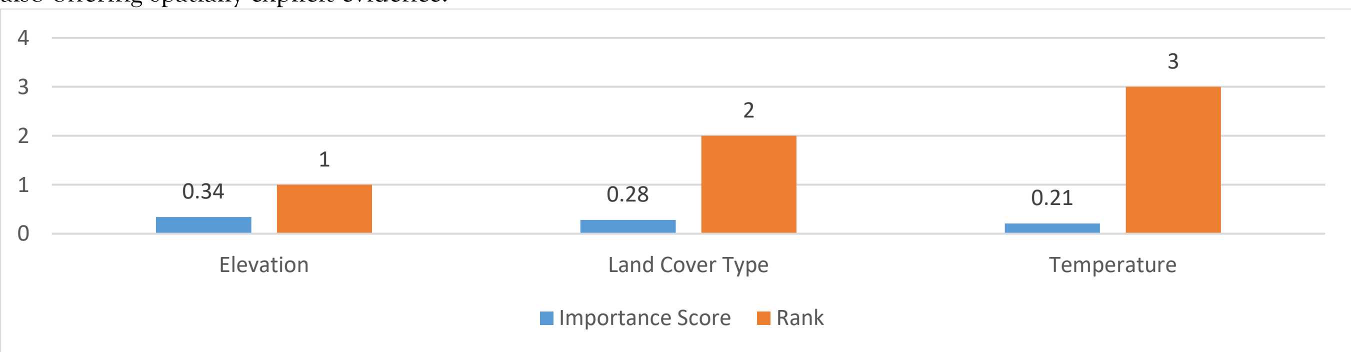

FIG 3: VARIABLE IMPORTANCE RANKING FROM RANDOM FOREST MODEL

In spite of these strengths, some challenges have been noticed. Deep learning models were powerful and tended to overfitting in case of limited occurrence records, and their interpretability was lower than tree-

based models. Conversely, Support Vector Machines were shown to perform well when the dataset is smaller, but they do not provide the scalability to very high-dimensional predictors like multispectral satellite images. These results highlight the fact that there is no universal algorithm, and the ecological question, availability of data and computational resources should inform the model selection [10].
Comparison of predicted habitat suitability maps between Support and Support Vector Machines and the Random Forests showed that they had significant disparities. Both models shared the opinion about the aspects of core habitat living but, SVM produced a wider range of moderate suitability, and this can contribute to overestimating potential ranges of habitats. Here, RF, conversely, produced superior localized high-suitability areas and edges. A summary of these differences in other words, the projection of the habitat extents of two models is given in Table 2. These comparisons are significant to conservation planning arising out of the fact that over or under-prediction of the same can significantly influence decisions of habitat prioritization.

TABLE 2: COMPARISON OF PREDICTED HABITAT SUITABILITY EXTENTS BETWEEN RANDOM FOREST AND SVM

Habitat Suitability Class	Random Forest (Area km²)	Support Vector Machine (Area km²)
High Suitability	1,240	1,050
Moderate Suitability	2,300	2,950
Low Suitability	3,600	3,100

The uncertainty analysis revealed as well the need to make use of ensemble methods. Results obtained through RF, GBM, and SVM were able to reduce variability and they gave results which were more balanced. Areas of highest uncertainty tended to be ecological transitional areas i.e., forest- grassland and semi-arid margins. Figure 4 shows the level of uncertainty within the study area with specific interest to the areas where there is need to be cautious in interpretation of predictions. These regions have been hotspots with regards to field validation practice and adaptive monitoring to ensure that sound data are utilized in conservation interventions.

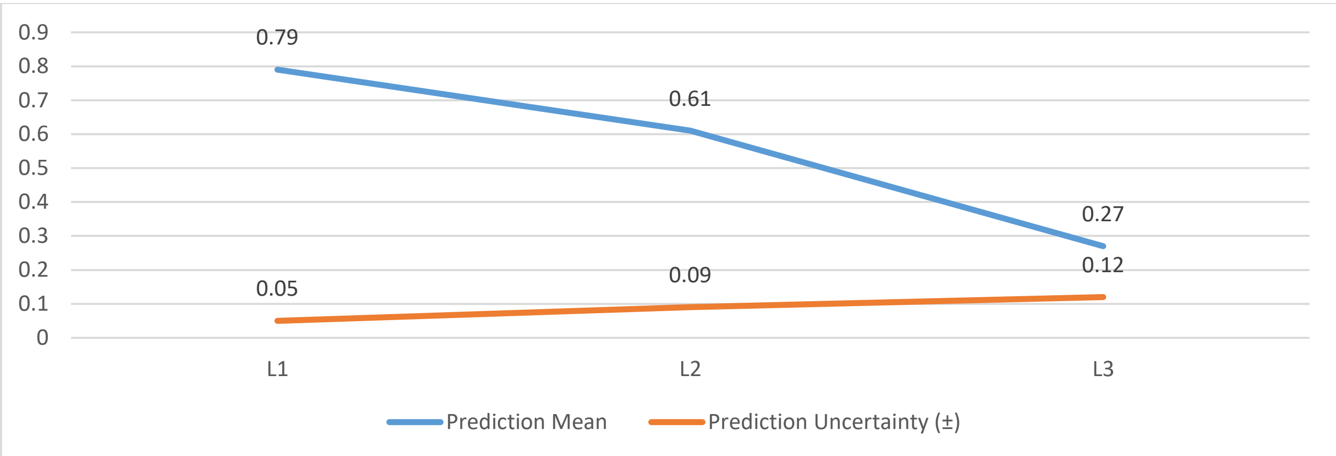

FIG 4: UNCERTAINTY MAP OF PREDICTED HABITAT SUITABILITY ACROSS THE STUDY REGION

In sum, the results demonstrate that the use of machine learning-based SDM outperforms the ancient methods considerably because possesses a superior predictive quality and ecological readability, in addition to providing the ability to integrate diverse data sources. However, interpretability versus computational demand and the best tradeoff between interpretability and computational demand are tradeoffs with ensemble models. In particular, the paper demonstrates that uncertainty analysis should be incorporated to avoid the improper interpretation of forecasts, first of all conservation where the division of resources depends on the reliability of suitability maps. The combination of several algorithms and the incorporation of a variety of predictors on top of application of explain ability tools in the methodology provides ecologically meaningful results and the high predictive performance is not compromised.

V. CONCLUSION

A paradigm shift of more accurate distribution and habitat fitness models of species is suggested by machine learning models, with more flexibility and possibility to be ready to use large and complex data. Such models are able to provide thought-provoking data on the trend of biodiversity and conservation planning since they enable integration of various environmental predictors.

Nevertheless, practically limits have to be noted. High computational requirements, inability to provide sensible interpretations and sensitivity to the quality of the data all hamper the applicability of ML models to practice. On top, overfitting and poor transferability also indicate that model choice, validation and ecological interpretation must be carefully chosen.

The introduction of explainable AI technique to enable improved interpretability and design hybrid models by combining mechanistic and machine learning models, and the inclusion of climate change prediction to draw up strategic conservation plans should have priority in the further research. By diversifying the global biodiversity databases alongside the construction of the computational ecology, the analysis of machine learning in the setting of sustainable habitat and species management approaches will be further promoted.

REFERENCES

[1] R. N. Vasconceloset *al.*, "Advances and Challenges in Species Ecological Niche Modeling: A Mixed review," *Earth*, vol. 5, no. 4, pp. 963–989, Dec. 2024, doi: 10.3390/earth5040050.

[2] S. Safdar, I. Younes, A. Ahmad, and S. Sastry, "A comprehensive review of spatial distribution modeling of plant species in mountainous environments: Implications for biodiversity conservation and climate change assessment," *Kuwait Journal of Science*, vol. 52, no. 1, p. 100337, Oct. 2024, doi: 10.1016/j.kjs.2024.100337.

[3] R. P. Espíndola, M. M. Picanço, L. P. De Andrade, and N. F. F. Ebecken, "Applications of machine learning methods in sustainable forest management," *Climate*, vol. 13, no. 8, p. 159, Jul. 2025, doi: 10.3390/cli13080159.

[4] J. Zhuang, Y. Wang, J. Lin, D. Zhang, Q. Peng, and T. Jin, "Research progress and framework on the simulation and assessment of fish habitat degradation in lakes," *Ecological Indicators*, vol. 158, p. 111461, Dec. 2023, doi: 10.1016/j.ecolind.2023.111461.

[5] M. B. Oliveira, H. S. Bernardino, A. B. Vieira, and D. A. Augusto, "Classification of animal species via deep neural networks and species distribution modeling: a systematic review," *Artificial Intelligence Review*, vol. 58, no. 8, May 2025, doi: 10.1007/s10462-024-11074-w.

[6] C. N. Chiloane, T. Dube, M. Sibanda, T. Dalu, and C. Shoko, "A Summary of Recent Advances in the Literature on Machine Learning Techniques for Remote Sensing of Groundwater Dependent Ecosystems (GDEs) from Space," *Remote Sensing*, vol. 17, no. 8, p. 1460, Apr. 2025, doi: 10.3390/rs17081460.

[7] T. Wang, Y. Zuo, T. Manda, D. Hwarari, and L. Yang, "Harnessing artificial intelligence, machine learning and deep learning for sustainable forestry management and conservation: Transformative Potential and Future Perspectives," *Plants*, vol. 14, no. 7, p. 998, Mar. 2025, doi: 10.3390/plants14070998.

[8] U. Nawaz, M. Anees-Ur-Rahaman, and Z. Saeed, "A survey of deep learning approaches for the monitoring and classification of seagrass," *Ocean Science Journal*, vol. 60, no. 2, May 2025, doi: 10.1007/s12601-025-00213-1.

[9] G. Alier, R. Idohou, G. Hounsou-Dindin, and R. G. Kakaï, "Assessing the potential impact of climate change on Kobus megaceros in South Sudan: a combination of geostatistical and species distribution modelling," *Modeling Earth Systems and Environment*, vol. 10, no. 2, pp. 1531–1542, Nov. 2023, doi: 10.1007/s40808-023-01889-x.

[10] B. V. S. Chauhan, A. Vedrtnam, K. P. Wyche, and S. Verma, "Revolutionizing natural resource management with Artificial intelligence: A review," *Prospects of Artificial Intelligence in the Environment*, pp. 143–170, Jan. 2025, doi: 10.1007/978-981-96-6863-2_5.

[11] J. Lai, S. Qi, J. Chen, and J. Guo, "Changes in habitat quality and their driving forces in the coastal zone of Hainan Island: Insights from 20 years of research and future projections," *Deleted Journal*, Feb. 2025, doi: 10.1016/j.ecofro.2024.12.010.

[12] B. Kühnet *al.*, "Machine Learning Applications for Fisheries—At Scales from Genomics to Ecosystems," *Reviews in Fisheries Science & Aquaculture*, pp. 1–24, Nov. 2024, doi: 10.1080/23308249.2024.2423189.

[13] N. R. Anoop, A. Samrat, S. Basnett, T. Ganesh, and S. Sen, "Unveiling management suggestions for the invasive tree Senna spectabilis through literature review, species distribution models and ecological knowledge," *Biological Invasions*, vol. 27, no. 6, May 2025, doi: 10.1007/s10530-025-03602-8.

[14] T. Qiao, C. Xiao, Z. Feng, and J. Ye, "Habitat distributions and abundance of four wild herbivores on the Qinghai–Tibetan Plateau: a review," *Land*, vol. 14, no. 1, p. 23, Dec. 2024, doi: 10.3390/land14010023.

[15] B. Dash, S. K. Baliarsingh, A. Samanta, S. Sahoo, S. Joseph, and T. M. B. Nair, "Evolution and recent trends of Indian oil sardine research: A review," *Ocean & Coastal Management*, vol. 258, p. 107396, Sep. 2024, doi: 10.1016/j.ocecoaman.2024.107396.

[16] A. Saim and M. H. Aly, "Fusion-Based Approaches and Machine Learning Algorithms for Forest Monitoring: A Systematic review," *Wild*, vol. 2, no. 1, p. 7, Mar. 2025, doi: 10.3390/wild2010007.
[17] M. He, P. Sandhu, P. Namadi, E. Reyes, K. Guivetchi, and F. Chung, "Protocols for water and environmental modeling using machine learning in California," *Hydrology*, vol. 12, no. 3, p. 59, Mar. 2025, doi: 10.3390/hydrology12030059.
[18]