Assessing the niche of *Rhododendron arboreum* using entropy and machine learning algorithms: role of atmospheric, ecological, and hydrological variables

Akash Anand[a], **Prashant K. Srivastava**[a,*], **Prem C. Pandey**[b], **Mohammed L. Khan**[c] and **Mukund D. Behera**[d]

[a]Banaras Hindu University, Institute of Environment and Sustainable Development, Remote Sensing Laboratory, Varanasi, Uttar Pradesh, India
[b]Shiv Nadar University, Greater Noida, Uttar Pradesh, India
[c]Dr. H. S. Gour University, Department of Botany, Sagar, Madhya Pradesh, India
[d]CORAL, IIT Kharagpur, IIT Kharagpur, West Bengal, West Bengal, India

Abstract. Species distribution models (SDMs) have been used extensively in the field of landscape ecology and conservation biology since its origin in the late 1980s. But there is still a void for a universal modeling approach for SDMs. With recent advancements in satellite data and machine learning algorithms, the prediction of species occurrence is more accurate and realistic. Presently, four machine learning and regression-based algorithms, namely, generalized linear model, maximum entropy, boosted regression tree, and random forest (RF) are used to model the geographical distribution of *Rhododendron arboreum*, which is economically and medicinally important species found in the fragile ecosystem of Himalayas. To establish complex relation between the occurrence data and regional climatic and land use parameters, several satellite products, namely, MODIS, Sentinel-5p, GPM, ECOSTRESS, and shuttle radar topography mission (SRTM), are acquired and used as predictor variables to the different SDM algorithms. The performance evaluation has been conducted using the area under curve (AUC), which showed the best result for Maxent (AUC = 0.871) and poor result was observed for RF (AUC = 0.755) among all. The overall prediction confirmed the distribution of *Rhododendron arboreum* in the mid to high altitudes of central and southern parts of the Garhwal Division. We provide crucial evidence that combining multisatellite products using machine learning algorithms can provide a much better understanding of species distribution that can eventually help the researchers and policymakers to take the necessary step toward its conservation. © *2022 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JRS.16.042402]

Keywords: species distribution model; Maxent; boosted regression trees; *Rhododendron arboreum*; Himalayan ecology; species occurrence.

Paper 210393SS received Jun. 30, 2021; accepted for publication Jan. 7, 2022; published online May 26, 2022.

1 Introduction

Since the start of the century, humans started to recognize the importance of regional ecosystem in explaining the distribution of flora and fauna. As a result, the understanding of species-specific geographical presence has become an important aspect especially considering the global concerns of climate change, altitudinal range shift, species invasions, and depletion of endangered species.[1,2] Modeling the potential distribution of a plant species is typically achieved by one (or more) of the several modeling methods. They use exploration of diversity patterns exploration to investigate the distribution depending upon species identity and based on different input parameters, such as climatic data, land use data, soil data, presence/absence data, climatic condition, and its projection data for the generation of suitability maps. These models are sensitive to abundance patterns, altitudinal variations,[3] latitudinal variation ranges, and climate change scenarios. The technique used to model species geographical distribution is termed as species

*Address all correspondence to Prashant K. Srivastava, prashant.iesd@bhu.ac.in; prashant.just@gmail.com

1931-3195/2022/$28.00 © 2022 SPIE

distribution models (SDM), also known as ecological niche modeling, bioclimatic envelop modeling, or bioclimatic modeling. It provided solutions to some of the core issues in ecology, evolution, and its conservation. With the advancement in SDM algorithms, there is still a need to better understand the nonlinear interactions of species with local parameters as prediction based on extrapolation was found to be nonrobust, especially with the conventional approaches.[4,5]

In recent years, several studies have been carried out to solve computational problems and implemented neural network[6] and machine learning models in the study, which are valuable tools for modeling many phenomena in ecology, mathematics, medical, economics, physics, and engineering.[6–8] Some of their significant applications were introduced in the research works of Jamali et al.,[6] Radmanesh and Ebadi,[7] Fouladi et al.,[8] Rafieipour et al.,[9] Heydarpour et al.,[10] and Altaher et al.[11] Also, the authors in Refs. 6 and 8 given several remarkable studies on the theory, analysis, and recent historical development of the neural network and computational studies.

According to the niche theory,[12] a species can only be found in a region where the combination of local bioclimatic gradients allows the species to have positive population growth. This theory conceptualizes the regional species environment and its occurrence considering the absence of immigration. While further extending the theory, it can also be realized that the variation in species traits allows them to inhabit different niches or cohabit in a particular spatial extent. These interactions are ecologically complex and nonlinear; therefore, the role of machine learning is crucial in understanding their distributions.[13,14] But before the introduction of machine learning in SDM, several theories and models were proposed by ecologists and researchers widely used to predict the distribution of plants and animals. BIOCLIM[15] and DOMAIN[16] are among the earliest SDMs that received global acceptance due to their less complex algorithm and easy to use interface. For establishing the nonlinear relation between input parameters, several machine learning iterative algorithms are proposed that give much better accuracy than the linear models. Boosted regression trees (BRT)[17] and random forest (RF)[18] are among the widely accepted iterative models, especially for modeling species distribution.[19] On the other hand, the maximum entropy (Maxent) model[20] is based on envelop model, which takes the presence-only data as its input parameter. Another one is based on the conventional regression-based learning technique called generalized linear model (GLM),[21] which can consider multiple measurement levels of response values using different link functions. Among the above mentioned SDM's, Maxent is widely expected algorithm due to its robust and nonlinear modeling techniques.[22] Maxent models are able to satisfy all known variables without any subjective assumptions, which is not present in earlier SDM models (such as Bioclim/DOMAIN). Therefore, it is more robust than earlier SDM because of the following inherent merits that involve improved mathematical modeling, machine learning, and statistical tools with better predictive accuracy. These SDM models have efficient deterministic algorithms that can be benefit to predict species' optimal probability distribution at the study sites. They are less sensitive to the various environmental variables and changes occurring in them. They consider interactions between environmental variables and minimize overfitting problems.

Major drawback of SDMs are the availability of non-uniform and relatively lesser field observations as compared to the area of interest, therefore models are generally extrapolated beyond their sampling sites. These spatial and geographical-extrapolations based on limited species sampling often lead to spurious results. A major limitation of macroecological SDMs is the inability to predict species identity and thus mainly involved species richness, i.e., emergent ecosystem properties implemented for exploring macroecological phenomena. Even the probability distribution is not uniform in earlier SDMs, thus the stability is lower than expected. Species distribution results depend on the spatial resolution chosen for the extent mapping, and also temporal aspects play a significant role in species. Therefore, models having functioning of species ecological distribution at the relevant scale are needed.[23] References 24 and 25 suggested that the unavailability of data or insufficient data-based predictions using extrapolation are limiting to the true species distribution in the region, which was supported by the study conducted by Ref. 26. Therefore, while using SDM, one has to understand data quality, sufficient data, predictor variables (hydrometeorology), and reliability of the models for distribution output. There is some development that happened in past, but there is still a lack of modeling techniques for understanding the complex relationship between different regional input parameters. As per

the fundamental assumptions of SDM, the target species is considered to be in equilibrium with the predictor variables, which is highly criticized in the past and still there is not a relevant alternative.[27] The recent developments in theoretical ecology, remote sensing techniques, and modeling algorithms have now enabled the ecologists to model near real-time distribution of the species, especially with the data coming at much finer spatial and temporal resolutions.[28–30] Also with the introduction of sensors such as Sentinel-5p and ECOSTRESS, it is now possible to assess the relation of greenhouse gases and evapotranspiration with species distribution, which was missing in the conventional studies. Earlier, ecologists preferred to use the data from Worldclim,[31] NCEP,[32] and ECMWF[33] as predictor variables that provide spatially interpolated atmospheric datasets at 0.1 deg to 2.5 deg of spatial resolution. The coarse resolution data were the major source of uncertainty and error, also they did not allow the model to predict the distribution of species at the regional scale. Particularly for the topographical conditions of the Himalayas, which varies drastically, it needed fine-scale satellite products.

The Himalayas being home to thousands of economically, medicinally, and rare flora and fauna is experiencing global climate change.[34,35] In their study, they reported that the overall warming in the Himalayas is consistently increasing for the past 100 years, and the rate is much higher than the global average of 0.74°C.[36,37] The temporal change in the distribution of species has been reported by several researchers and their impact on regional ecology.[38–40] One such species is *Rhododendron arboreum*, which comes from the Ericaceae family and dominantly found in the Himalayas, South India, Nepal, and Sri Lanka.[41,42] It is an economically and medicinally important species and sustains itself in the fragile ecotone of alpine and subalpine regions. The continuous change in regional climatic conditions is imperative to model the distribution of species so the biodiversity and conservation of the ecosystem can be maintained.

The main contribution of this work is to uncover the following:

- Impact of environmental variables on the distribution of *Rhododendron arboreum* at the study site.
- Linking variables, such as normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), evapotranspiration (ET), fraction of photosynthetically active radiation (*f*PAR), water vapor, leaf area index (LAI), land surface temperature (LST), precipitation, ozone, NOx, albedo, aerosol absorbing index (AAI), and digital elevation model (DEM) for understanding the distribution of species in Himalayan environment.
- Assessment of the variables on geographical distribution of *Rhododendron arboreum* through machine learning and entropy models.

Therefore, in purview of the above and considering uniformity in probability and stability of Maxent, this study is focused on establishing the relation between different bioclimatic and environmental parameters to model the distribution of *Rhododendron arboreum* within the study area. In Sec. 2, we provide an overview of the study site, specifications of target species, modeling algorithm employed in the study, and performance evaluation metrics. Section 3 explains the model result and discussion part, in which the species distribution maps along with the discussion are presented. Section 4 provides some conclusions and gives suggestions for future research.

2 Materials and Methods

2.1 *Study Area*

The Himalayas is one of the most complex and diverse ecoregions, and it offers rich biodiversity and has been home to thousands of floras and fauna. The complex topography allows some of the rare, medicinal, and economically important species to grow within this region. This study is conducted within the Garhwal Division of Uttarakhand state where the elevation ranges from 416 to 7801 m above mean sea level. As shown in Fig. 1, this region has several biomes, namely tropical evergreen and deciduous broadleaf forest, tropical and subtropical coniferous forest, temperate coniferous forest, temperate savanna, grassland, and shrubland as well as it has a significant number of glaciers as well. The region falls under subtemperate to temperate climate,

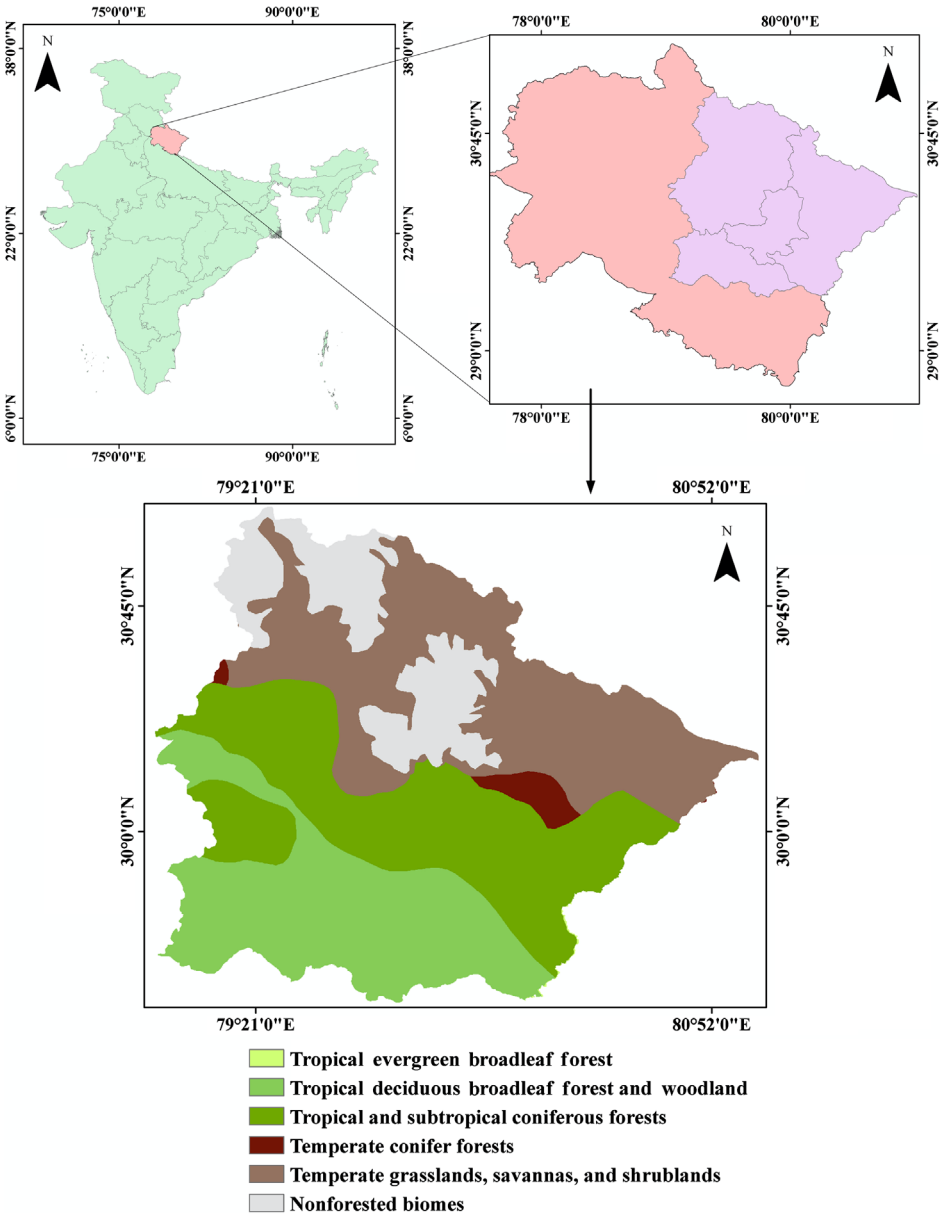

Fig. 1 Study area map showing the distribution of biomes.

and the hilly terrain with densely forested slopes receives significant rainfall from mid-June till September with several occasional rain events throughout the year, whereas ~20% of the area is covered with snow throughout the year. Due to its rich ecology and complex environmental conditions, it has always attracted the attention of researchers and scientists from around the world, especially with the recent trend in global warming, this region is showing early impacts of climate change that makes this region more important.

2.2 *Target Species and Occurrence Data*

The geographical distribution of *Rhododendron arboreum* is performed for the study area using *in situ* occurrence data and predictor variables. *Rhododendron arboreum* has a great biological significance and dominantly found in the Himalayas. The occurrence of this species was reported between 1200 and 4000 m above mean sea level.[42] *Rhododendron arboreum* is a high valued species both in terms of medicinal and economic importance, also it is reported by ecologists that it possesses some characteristics of invasive species.[43,44] Medicinally, *Rhododendron arboreum*

is anticancerous, antioxidant, antidiabetic hepatoprotective, antimicrobial, diarrhoea, and antinociceptive.[44] Whereas a study in Ref. 45 reported that the squash made of *Rhododendron arboreum* is used for the treatment of intellectual disabilities. Economically, the species used in making jams, local brews, squash, and in different juices, also its wood is used as fuel, and the leaves are used for treating the bedbug bites. With such importance, mapping its distribution is very crucial for ecosystem conservation and making necessary strategies for its protection.[46] The *in situ* data collection was conducted within the study area during September 2019 and Match 2021, covering the complex topography of the region. A total of 70 homogeneous patches of *Rhododendron arboreum* were identified at different elevations. Among the collected occurrence data, two-third used for model development and one-third is used for validation purpose. To mark the occurrence of the species, handheld Garmin GPS is used with a horizontal accuracy of $95\% \pm 9.3$ m.

2.3 *Predictor Variables*

For establishing complex relationship between the occurrence data and local ecosystem, several environmental predictor variables are acquired from different satellite sensors. The selection of predictor variables is performed on the basis of their direct impact on the vegetation. The environmental predictor variables include MODIS products, namely, NDVI, EVI, fPAR, and LAI. The sentinel-5p sensor provides information related to the greenhouse gases with a high spatial and temporal resolution that makes it very crucial in the regional ecological study considering climate change. The Sentinel-5p product used as predictors is AAI, water vapor, albedo, ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO), and sulfur dioxide (SO_2).[47] Apart from the climatic parameters provided by Sentinel-5p, LST and (ET) are acquired from ECOSTRESS sensor.[48] Precipitation data are obtained from GPM dataset. Topography plays a vital role in studying species distribution, especially in the topographically complex regions such as the Himalayas. Therefore, DEM provided by the SRTM[49] is used as an input parameter for the calculation of SDM. Also, the regional biome is taken into consideration as forest biomes directly influence the species distribution.[50]

2.4 *Species Distribution Modeling Algorithms*

The regional dynamics and distribution of species, especially their prediction under the influence of climate change, is a crucial issue for ecological biodiversity and conservation. The concept of SDM started in early 1980s, but before that, mapping of species distribution is only limited to *in situ* sampling. With the availability of satellite-based climatic data, several presence-only based model is designed and tested in the early 2000s, which includes BIOCLIM and DOMAIN. After that, several other models were introduced based on establishing the correlation between occurrence data and the climatic predictor variables. These models are easy to implement but did not provide significant accuracy and failed on the regional scale. As the interaction between species and respective ecological parameters is extremely complex, nonlinear models must establish their relationship. Therefore, some of the most popular and widely used machine learning algorithms are trained to model the geographical distribution of *Rhododendron arboreum* within the study area.

2.4.1 *Maxent*

Maxent is one of the widely used SDM that works on estimating the probability distribution through maximum entropy of input parameters. Reference 51 defined that entropy is the measure of the total number of choices involved in selecting a particular feature. The basic concept of Maxent was first introduced by Ref. 52, which stated that the best way to ensure the approximation is by testing the results with known positions, and the distribution must have maximum entropy, it is also known as the maximum-entropy principle. Maxent owes its success in species distribution because the entropy reaches a minimal value that is highest to that of a species among the probability distributions of all the species. This is achieved by predicting the occurrence of species through the distribution, which is mostly spread out or tends to have a uniform

distribution. This, in turn, necessitates having the information of all the environmental variables of the locations. The Maxent algorithm uses many points taken during ground data sampling that are referred to as background points, and these background points define the current environmental variables.[53] These variables are used as constraints that limit the rule for the predicted distribution. In general, Maxent considers linear/quadratic/product/threshold/hinge/categorical features as constraints that define the rule to confine the expected distribution. These features have different implications for the constraints. This work has used a categorical feature called "biome," which is defined as types of regional land use. This constraint specifies the proportion of predicted occurrences in each category to be as close to the proportion of observed occurrences in each category.

In this work, two probability densities are calculated. These densities provide the relative likelihood of all environmental variables over the range of background points. The algorithm then calculates the ratio between these probability densities to find the relative ecological suitability for the occurrence of the rhododendron for the given point in the study area. In this manner, the Maxent chooses the distribution that has maximum similarity between the environmental characteristics of the given climate and the locations where the required species are supposed to be abundant. This is the raw output of the algorithm, which is logically transformed by considering the prevalence value. In this work, the value is taken as 0.5, which implies that the species is present in half of all the possible locations. The limitation of this algorithm lies in the fact that it provides environmental suitability rather than the predicted probability of occurrence.[41]

The mathematical measure of the uniformity of a conditional distribution $\hat{\pi}(x)$, which is provided by the conditional entropy, is given as

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x). \tag{1}$$

The entropy is bounded from below by zero, the entropy of a model with no uncertainty at all, and from above by $\ln \hat{\pi}(x)$, the entropy of the uniform distribution over all possible $|\hat{\pi}(x)|$ of x. This acts as a base for presenting the principle of maximum entropy.

To select a model from a set S allowed probability distribution, choose the model $p^* \in S$ with maximum entropy $H(\hat{\pi})$:

$$p^* = \arg \max_{p \in S} H(\hat{\pi}). \tag{2}$$

It can be shown that p^* is always well-defined, that is, there is always a unique model p^* with maximum entropy in any constrained set S. The threshold chosen in the current method is the value for which the sum of sensitivity and specificity is highest, and the probability model prediction will be transformed to a binary score of the presence or absence of the species. The required solution is achieved by maximizing the gain function that is a penalized maximum likelihood function that is given as

$$\text{Gain} = \frac{1}{m} \sum_{i=1}^{M} z(x_j)\lambda - \log \sum_{i=1}^{N} Q(x_j) e^{z(x_j)\lambda} - \sum_{j=1}^{J} |\lambda_j| * \beta * \sqrt{s^2[z_j]/M}, \tag{3}$$

where the likelihood of the presence data are the sum of predicted values at presence locations is given by $\frac{1}{m}\sum_{i=1}^{M} z(x_j)\lambda$, the likelihood at all the background locations is the sum of the predicted values at background locations is given by $\sum_{i=1}^{N} Q(x_j) e^{z(x_j)\lambda}$ and overfitting penalty to be used in regularization is given as

$$\sum_{j=1}^{J} |\lambda_j| * \beta * \sqrt{\frac{s^2[z_j]}{M}}, \tag{4}$$

where β is the regularization coefficient, $s^2[z_j]$ represents the variance of feature j at presence locations, and $Q(x_j)$ is the prior distribution. A significant characteristic of Maxent is

regularization that helps in reducing the overfitting of the model. This is achieved by setting confidence intervals across the constraints and excluding the features that are not significant. The regularization used is the least absolute shrinkage and selection operator.

2.4.2 *Random forest*

Random forest algorithm[54] is developed using the classification and regression tree approach, and it has shown significant performances in different application of remote sensing, including forestry,[55] ecology,[56] classification,[57] and climate change.[58] Random forest is defined as a collection of weak learners having a tree structured interface with uniformly distributed random vectors where each tree provides a prediction for the resultant variable. While generating the group of weak learners based on the bootstrap of the data, the overall calculation converges on an optimal result avoiding the issues of general parametric, classification, and regression statistics. Bootstrap of training, independent variables (M) at each node, and retaining the variables that provided the most useful information sum up the significance of RF algorithm for both regression and classification purposes. To assess the information and purity of the node, Gini entropy index is used.

The algorithm for the prediction of a new variable can be explained as

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x), \tag{5}$$

where $\hat{f}_{rf}^{B}(x)$ is the bootstrapped RF function for predicting x, T_b is the RF tree, and B is the number of bootstraps.

2.4.3 *Boosted regression trees*

BRT is one of the most popular machine learning techniques that is widely used for SDM. It is based on improving the performance of a model by fitting multiple models and combine them for prediction. Broadly, BRT uses two algorithms, first one is the regression trees based on classification and regression of input parameters and the second builds the boosting and combines the collection of multiple models. Tree-based models divide the predictors into small clusters using a series of rules to identify the region having the most homogeneous response during prediction. The regression tree fits the mean response from predictors in any specific region. As per Ref. 59, the best way to fit a decision tree for growing a large tree and then trimming it by eliminating the weakest links identified through cross-validation. Decision trees are widely used because they are easy to implement, visualize, and one of the most flexible algorithms, especially for species distribution modeling.[60] At the same time, boosting technique is used for improving the accuracy of the model based on its background architecture of finding and averaging multiple rules of thumb rather than a single rule.[61] While other techniques include bagging, stacking, averaging, and merging the results from multiple models, boosting works as sequential models based on forward and stagewise procedures. The AdaBoost boosting algorithm is used to determine species distribution by fitting the predictors using sequential iteration technique.

2.4.4 *Generalized linear model*

GLM is also termed as an extension of the classical linear regression model, where the transformation is achieved to get a normal distribution for the dependent variable. In a GLM-based model, predictor variables are used to calibrate the model, and the link function is selected based on the statistical distribution of dependent variables. GLM being a parametric function is not optimized using the least-square method, rather it uses the maximum likelihood method for model optimization.[62–64] GLM model has a set of distribution function including binomial, Poisson, gamma, etc. in which gamma distribution having link function $f(x) = \mu$, where μ is the mean value of predictor variables, is used for establishing the relation between the predictors.

In GLM, the predictors $X_i (i = 1, 2, \ldots, n)$ are joined together to get a linear predictor (LP), which is related to $\mu = E(\gamma)$, where γ is the response variable, to the link function $f()$, which is represented as

$$f(E(\gamma)) = LP = \propto + X^T \beta, \tag{6}$$

where $\propto$ is the intercept, $X^T = X_1, X_2, \ldots, X_n$ is the vector of p variable, and $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ is the regression coefficient.

Therefore, for i'th observation,

$$g(\mu) = \propto + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_n X_{in}. \tag{7}$$

The error rate within the model is resolved using the least square algorithm[65] during the model fitting.

2.5 *Model Validation*

In any regression and classification-based machine learning model, the estimation of model performance and classification accuracy is an important task. Therefore, to evaluate the overall accuracy of the distribution modeling, AUC parameter evaluation matric is used for all four machine learning algorithms. Primarily, the data are divided into training and testing sets, in which one-third data are assigned for the model validation, and rest is allocated for model development. AUC is widely used for model validation, especially in binary classification models. AUC is a threshold-independent evaluation metric that validates the model performance at various discrimination thresholds. At each discrimination threshold, the true positive rate (TPR), also known as the probability of detection or sensitivity and the false positive rate (FPR), also known as the probability of false alarm, is estimated and plotted against each other. The TPR and FPR for each point (x, y) are plotted together to get the final AUC curve. It is also explained as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \tag{8}$$

whereas specificity is calculated as

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100, \tag{9}$$

$$\text{FPR} = 1 - \text{Specificity}, \tag{10}$$

where TP is the true positive, and FP is the false positive values. Specificity defines the true negative rate, whereas TPR calculates the percentage of correctly predicted values. The AUC value varies from 0 to 1, where the value closer to 1 shows the accurate classification, and the values close to 0 denote poor classification accuracy.

3 Results and Discussion

The geographical distribution of a particular species is dependent upon the regional climatic, topographical, and land cover conditions. To build a more relevant SDM, the predictor variable should directly influence the existence and growth of the species. The occurrence data for the target species, *Rhododendron arboreum*, are collected within the Garhwal Division of Uttarakhand state, where the topography and climatic conditions are complex. To establish a relationship between species occurrence and its regional climatic condition supporting its existence and growth, 16 predictor variables are considered input parameters to different machine learning-based SDMs. The yearly trend is analyzed for each input parameter so the generated relation can be widely accepted and irrespective of any short-term bias caused by local weather conditions. The machine learning algorithms used in developing SDMs are Maxent, GLM, RF, and BDT. The algorithms are intercompared and validated using statistical evaluation matrices.

3.1 *Input Variables*

Several satellite-based input variables are acquired for modeling the species distribution. With better spatial and temporal resolutions, products from the sensors such as MODIS, Sentinel-5p, GPM, and ECOSTRESS are widely used in analyzing the regional and global ecology. Presently, MODIS products, namely NDVI, EVI, LAI, and fPAR, are used in the SDM algorithms. These variables indicate the information related to the surface reflectance, productivity, energy transfer by the vegetation, water cycle processes, and other biophysical and biochemical properties of the vegetation, and the spatial resolution of these data varies from 250 to 500 m. The products from ECOSTRESS, namely ET and LST, is responsible for providing the information related to plant water consumption and regional temperature levels, the thing that makes the ECOSTRESS products more valuable is their spatial resolution, it provides data at 70 m spatial resolution and has comparatively better temporal revisit. Sentinel-5p being one of the most recent sensors that provide data of different atmospheric gases, including the greenhouse gases at global with a spatial resolution of 0.01 arc degree, which is better than other satellite sensors currently active. Sentinel-5p provides a wide range of atmospheric data, but the data that highly influence *Rhododendron arboreum* are taken as model inputs and are, namely, ozone, nitrogen dioxide, albedo, carbon monoxide, sulfur dioxide, and AAI. Precipitation is one of the most important parameters for SDM acquired from GPM at a spatial resolution of 0.1 arc degree, whereas DEM is used to consider the topography. As the Himalayas is made up of different biomes, this study area has five biomes as listed in Fig. 1, and the biome information is also used as a predictor variable to the SDM.

All the predictor variables are shown in Fig. 2, which is the yearly average to maintain the temporal consistency, and all are resampled so they can match each other on the pixel level. The estimated NDVI and EVI values are varying from −0.08 to 0.84 and −0.11 to 0.48, respectively, in which it was found that the southern part of the study area is having high vegetation content in the tropical evergreen and deciduous forest than the northern part where the temperate forests dominate the biome. NDVI is used to identify the green vegetation, and EVI can enhance the vegetation signal by reducing the canopy background noises. ET also supported NDVI and EVI results, as it varies from 3.34 to 37.2 kg/m^2 where the maximum values are observed around the boundary of tropical deciduous and tropical and subtropical conifer forest shoeing the latent heat flux coming from the earth surface. Also, the fPAR and LAI values are highly correlated with ET and varying from 0 to 0.87 and 0 to 5.48, respectively. The LST is varying from 259.16 to 300.89 K, the value of LST is higher in the southern part, and it is gradually decreasing in the northern direction due to increase in the elevation range, which is in between 416 and 7801 m, that shows the elevation drop in the region and its impact on LST. Precipitation is also very low in the northern part, as it almost covered with snow throughout the year, whereas the middle and southern regions have significant average rainfall in between 0.049 and 0.188 mm/h, higher values are seen in the Pithoragarh region, where the forest is dominated by tropical evergreen biome. The sentinel-5p-based parameters are also provided significant information regarding the regional climate condition throughout the year. Water vapor is one of the major greenhouse gases found to be higher in the highly forested regions and lower in the higher altitudes. The range of observed water vapor is between 151.86 and 1933.73 mol/m^2. The quantity of water vapor has a direct impact on plant growth and photosynthesis. The ozone layer is found to be higher in tropical forests and lesser in the higher altitude showing the thinner atmosphere in the northern part, it is varying from 0.1207 to 0.1258 mol/m^2. The higher nitrogen dioxide value affects the plant growth, and currently, it is ranging from 4.6e-005 to 6.2e − 005 mol/m^2 in which the higher values are found in the southwestern part where the forest density is low. The albedo and AAI have shown similar results with an observed value between 0.21 to 0.83 and −1.85 to 0.04, respectively. Albedo and AAI are higher in the higher altitudes and low values in forest area due to their high absorption by the dense vegetation. Carbon monoxide varies from 0.014 to 0.036 mol/m^2 with higher values over the forested region, whereas sulfur dioxide varies between −0.00055 and 0.00061 mol/m^2. Both carbon monoxide and sulfur dioxide are major greenhouse gases and have an impact on plant growth and distribution.

All the input parameters have major significance over the distribution of species. Although increasing the parameter may improve the accuracy, the model will become more

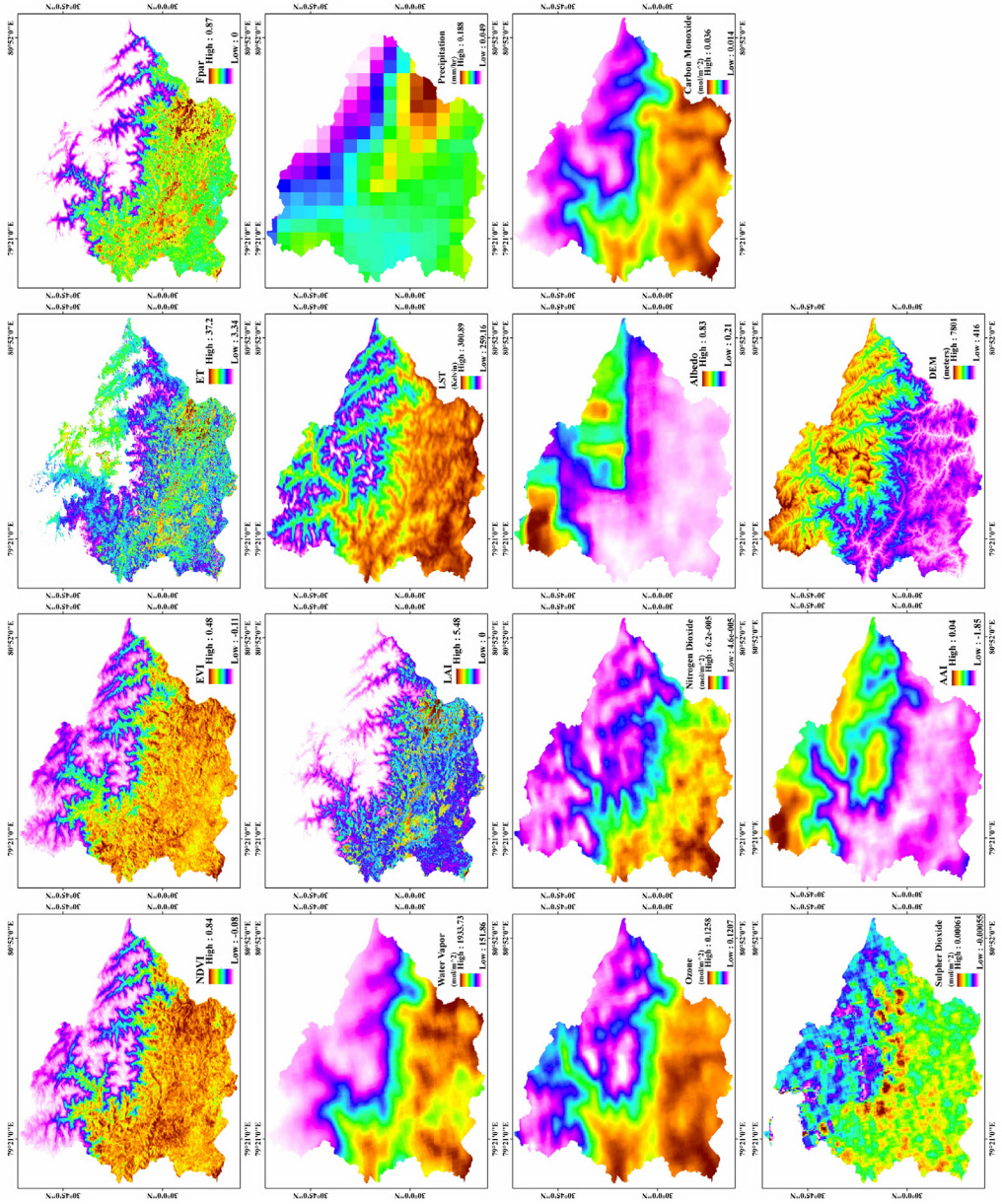

Fig. 2 Input variables for SDM.

computationally complex and parametric bias will also increase. Therefore, the predictor variable is limited to 16 in this study so a more robust model is designed. Rather than considering the satellite data during the sampling period, an overall trend is used to generate the mean value of each predictor variable. The trend of the predictors holds great importance as it demonstrates the change in climatic and land use condition throughout the year. The yearly trend of each variable is shown in Fig. 3. NDVI, EVI, and ET have demonstrated a similar trend, as the values are minimum in January, which linearly increased until the monsoon and gradually decreased in the winter. fPAR and LAI have similar trends as the maximum value is observed before and after the monsoon season when the insolation is on its peak. Precipitation and water vapor also followed a similar curve where the value is high in the monsoon period, and LST gradually increases until the monsoon and then starts decreasing. Atmospheric variables also have different responses to the local weather, and NO_2, albedo, and AAI values are high during monsoon, whereas O_3 is highest during January and gradually decreases the entire year. SO_2 value is lowest during May to September and progressively increases in winter.

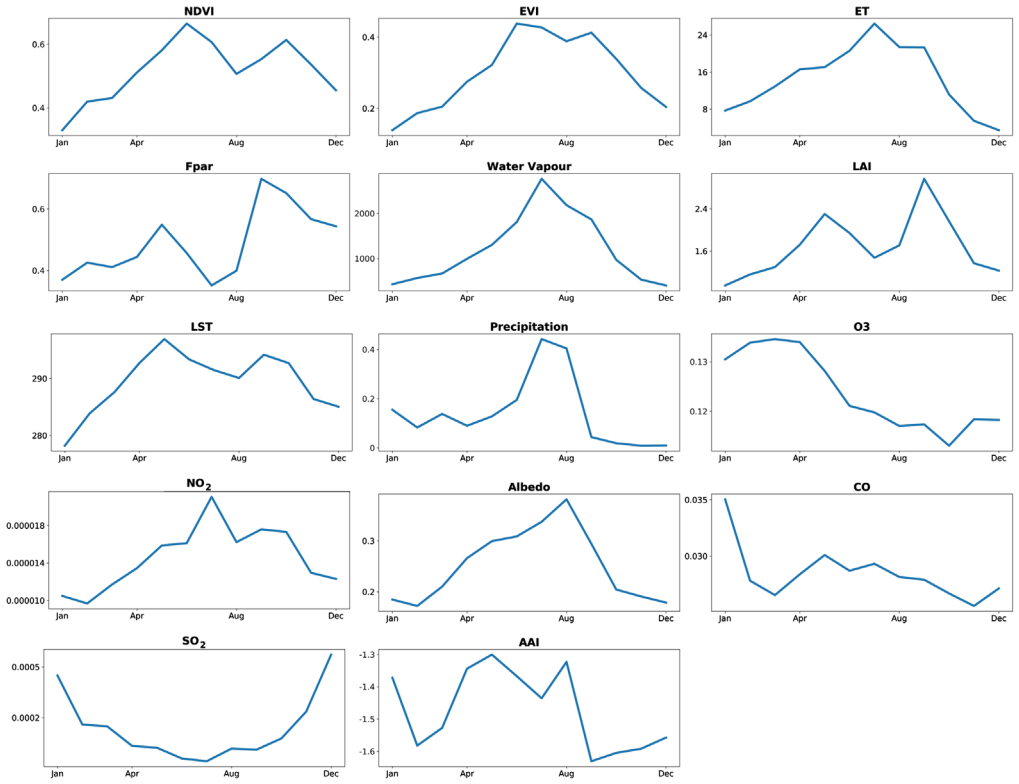

Fig. 3 Trend analysis of input variables for year 2020.

3.2 *Species Distribution*

The distribution of *Rhododendron arboreum* is predicted using four machine learning and regression-based algorithms considering 16 climatic and environmental datasets as input parameters. The prediction has been made for a spatial resolution of 100 m as shown in Fig. 4. As GLM is based on a regression-based linear modeling approach, the prediction given by GLM is showing overestimation with the higher probability of *Rhododendron arboreum* occurrence in the central region between tropical and subtropical coniferous and temperate conifer forest. The distribution predicted by Maxent is largely covering the mountains' tails and dominantly occurring in the central and southern parts of the study area. A similar result is shown by the BRT algorithm in which tree-based relation is generated within the input parameters. BRT is establishing major presence of *Rhododendron arboreum* in the southern part of the study area and some distributed patches on the northern side. On the other hand, RF vastly underestimated the prediction and has only shown higher probability of *Rhododendron arboreum* occurrences around the center of the study area and some occurrences on the south side. Overall, it is observed that GLM is overestimating the species distribution prediction, and RF is underestimating the same. But Maxent and BRT are showing promising results. The distribution of *Rhododendron arboreum* is largely found in the central and southern parts of the study area, and a higher probability can be seen near the tails of high topographic mountains.

3.3 *Model Validation*

The AUC curve is used as an evaluation metric to validate the prediction made to model the distribution of *Rhododendron arboreum* using four different machine learning and regression algorithms, as shown in Fig. 5. As AUC value varies from 0 to 1, with the value nearer to 1 is showing the high probability that the species is present. The maximum AUC value is recorded by Maxent and is 0.871, which shows that Maxent is the most promising machine learning model to assess species distribution on regional scale. After that, 0.835 AUC is recorded for GLM, which is also a considerable value in modeling species distribution, but the overestimation of the

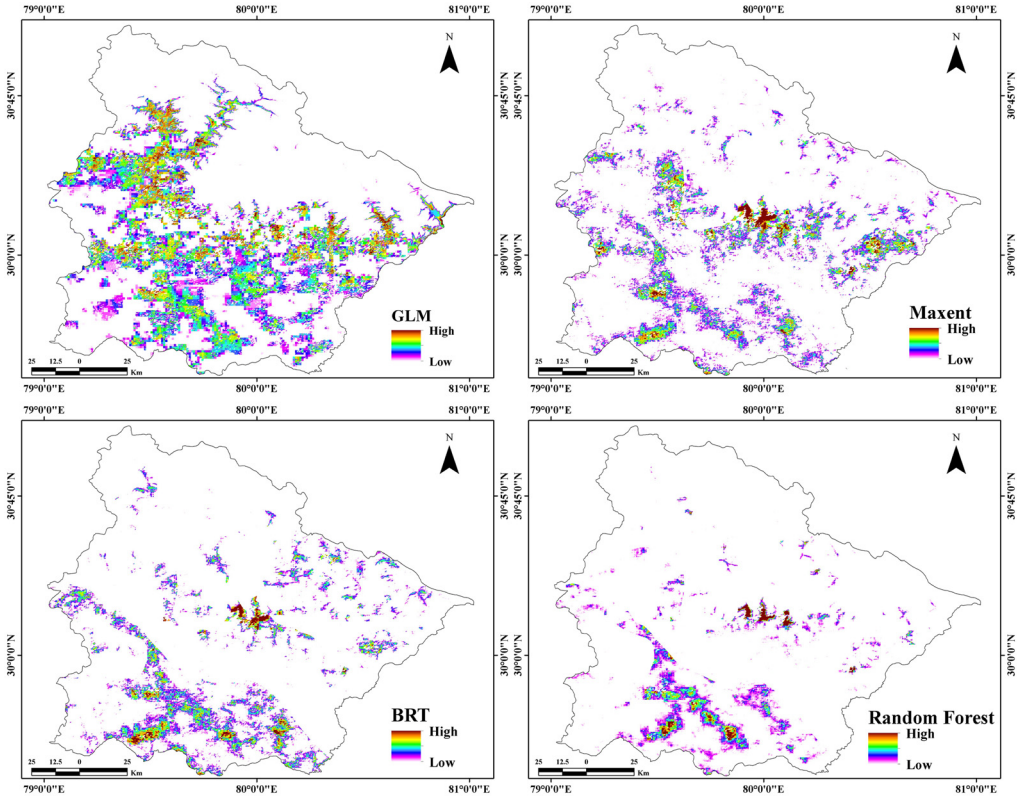

Fig. 4 Predicted species distribution using GLM, Maxent, BRT, and RF.

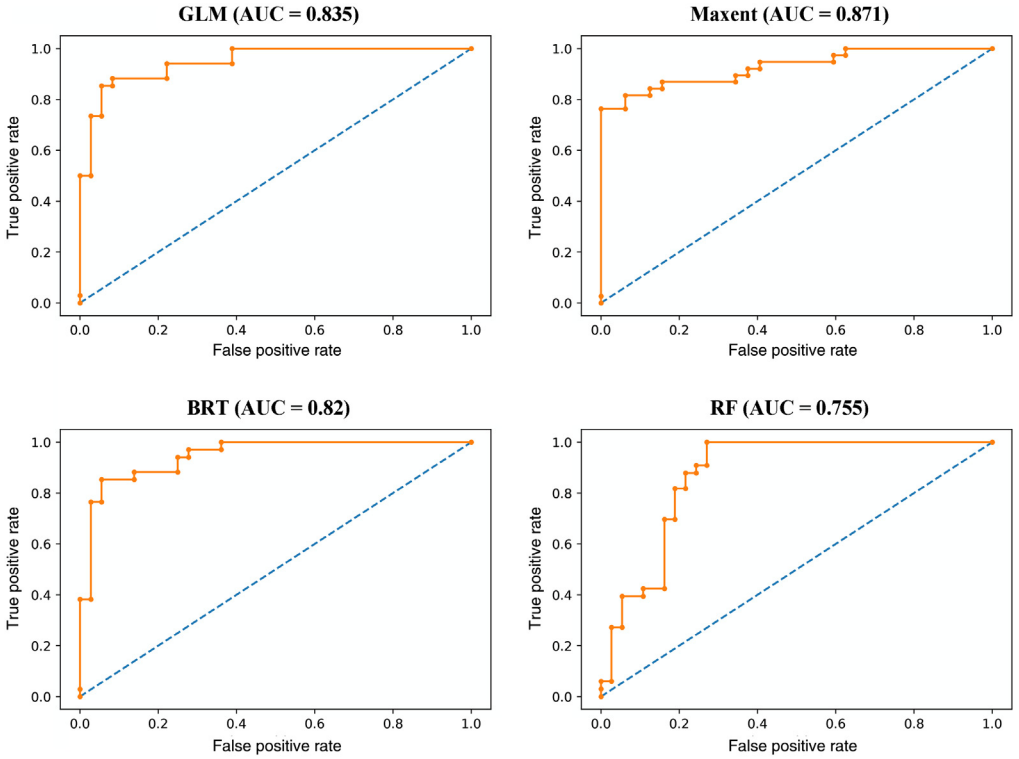

Fig. 5 AUC curve for each SDMs.

species in GLM is something that is more concerning. Apart from that, BRT has given an AUC value of 0.82 and according to its prediction, it has provided more precise values than even Maxent at some places. An underestimation is observed in RF with an AUC of 0.755, making it the worst-performing machine learning model among the four. Overall, it is observed that Maxent is performing well, and BRT has also shown promising results for modeling the distribution of *Rhododendron arboreum*.

The previous studies, such as Reiss et al.,[66] compared several models, such as MAXENT, RF, and SVM (support vector machine) for species distribution modeling and revealed that they have similar predictive performance. When compared with the other model such as BIOCLIM through their AUC, they found that values are significantly higher than BIOCLIM. In another study by Tsoar et al.,[67] they confirmed that Mahalanobis distance can even predict better than BIOCLIM and DOMAIN. In the study by Elith and Graham et al.,[68] they compered the performance of MAXENT with BIOCLIM and DOMAIN and pointed out that MAXENT gives a significantly higher predictive performance than the later. In Ref. 69, the authors divided the SDMs into two categories; at first they included the best performing one with higher stability such as Mahalanobis distance, RF, MAXENT, and SVM, whereas the second category composed of low stability one with lower performance such as BIOCLIM and DOMAIN. In other study by Giovanelli et al.,[70] they also confirmed the better performance of MAXENT as well as SVM for species distribution modeling and concluded that both SVM and MAXENT can be used. Overall, the above-mentioned studies indicate the superior predictive accuracy of MAXENT in SDM and recommended it for further use. The findings also revealed that the varying performance and stability of SDMs can be linked to changing environmental variables and climatic conditions. The results of this study are in agreement with the previous studies as mentioned above and hence can be used for prediction of *Rhododendron* sp. in the Himalayan region.

3.4 *Future Perspective and Challenges*

SDMs can predict the distribution of species on a regional scale and if well calibrated then for a global scale as well. Several models help in getting an insight into species distribution as well as establish linear and nonlinear interactions between the predictors, but they still lack establishing ecological theories and predefined assumptions. The confined understanding of species response to bioclimatic variables and limited statistical approaches bring error to the SDM. But with the advancement in spatial datasets and statistical algorithms, the prediction accuracy is continuously improving. The availability of satellite images providing bioclimatic, ecological, and eco-hydrological responses reduced the uncertainty related to the satellite-based biotic interactions. References 24 and 71 listed several actions to counter uncertainties, including the continuation of ecological and biological research that focuses on biotic interactions, regular and systematic collection of species occurrence data, temporal validation of retrieval models, selection of predictors, and algorithm improvements based on different climatic scenarios.

The integration of bioclimatic and atmospheric variables provides an exciting option for defining the consequences of global climate change on species distribution, especially for future scenarios. But, not all SDMs are optimally suited for predicting the species distribution based on various predictors. The earlier SDMs, namely BIOCLIM and DOMAIN, were based on the predefined hypothesis and did not have the option to integrate atmospheric gases and other satellite products. These early-stage models also lacked establishing complex nonlinear relation between the predictors. So, the data-driven models, namely GLM and BRT, are introduced to model the distribution based on observational data and substantially to integrate ecological hypothesis. These models are based on the observed realized niche and limited to the *in situ* observations and predictor variables. When combined with a certain degree of ecological knowledge, the data-driven models will act as the process-based approach in which the prediction will be more accurate and equally supported by the ecological hypothesis. Fot iterative models such as Maxent and RF, when supplied with process-based predictors, the generated prediction will have better accuracy, and the nonlinear relation will be more precise.

Despite ongoing improvements in SDM algorithms and the satellite-based predictors, the prediction is still influenced by the degree of uncertainty based on the biotic behavior of species and its interactions with changing climate and regional biota. Therefore, it is recommended to

use a process-based approach in modeling the distribution of species, which allows prediction beyond the observational data. Machine learning algorithms prove their importance in modeling process-based prediction, and with regular upgradation in the knowledge of species interaction, SDMs are continuously improving. However, additional research still needs to focus on the ecological understanding of species, ecological theories, and combining observational data rather than concentrating only on a data-driven approach. In the last few decades, with the high computing system and database technologies, now big datasets are available to users for deriving efficient outcomes.[72] With the advancement in deep learning, AI and data mining methods have entered a new age[8] that can help in analysis of high dimensional datasets with high accuracy, which now provides an enormous possibility in species distribution mapping also.[73]

4 Conclusion

In this study, we attempt to establish the relation between species occurrence data and their respective environmental predictor variables. The yearly tread of each parameter is analyzed to observe the variations throughout the year and a pixelwise mean value is calculated to be used in the SDM. Machine learning algorithms, namely Maxent, BRT, RF, and GLM, are implemented to establish the relation between predictor variables. The AUC-based performance evaluation matric is generated, and it is found that Maxent is performing better than others with an AUC of 0.871, also BRT has shown promising results with AUC = 0.82, but the GLM and RF are found to be overestimating and underestimating, respectively. The machine learning algorithms performed significantly well, and remote sensing data proved to be a vital source of information in ecological studies, which is continuously improving with regular upgradation in satellite data and algorithms. Although the SDMs are providing better results on regional studies but still lack in explaining the ecological background, the true meaning of their prediction and boundary conditions is a topic of research for future. To summarize, if SDMs are to be a standard tool, the background should be supported by good ecological understanding, and they give a new direction to the future research opportunities in SDM.

Acknowledgments

The authors are thankful to the National Mission for Himalayan Studies (NMHS), G.B. Pant National Institute of Himalayan Environment (NIHE) for the necessary financial assistance and support throughout. The authors declare no conflict of interest.

References

1. T. H. J. F. E. Booth, "Species distribution modelling tools and databases to assist managing forests under climate change," *For. Ecol. Manage.* **430**, 196–203 (2018).
2. C. D. Wilson, D. Roberts, and N. J. B. C. Reid, "Applying species distribution modelling to identify areas of high conservation value for endangered species: a case study using *Margaritifera* (L.)," *Biol. Conserv.* **144**(2), 821–829 (2011).
3. R. G. Mateo et al., "Do stacked species distribution models reflect altitudinal diversity patterns?" *PLoS One* **7**(3), e32586 (2012).
4. R. Kadmon, O. Farber, and A. J. E. A. Danin, "A systematic analysis of factors affecting the performance of climatic envelope models," *Ecol. Appl.* **13**(3), 853–867 (2003).
5. M. S. Wisz et al., "Effects of sample size on the performance of species distribution models," *Divers. Distrib.* **14**(5), 763–773 (2008).
6. N. Jamali et al., "Estimating the depth of anesthesia during the induction by a novel adaptive neuro-fuzzy inference system: a case study," *Neural Process. Lett.* **53**(1), 131–175 (2021).
7. M. Radmanesh and M. Ebadi, "A local mesh-less collocation method for solving a class of time-dependent fractional integral equations: 2D fractional evolution equation," *Eng. Anal. Boundary Elements* **113**, 372–381 (2020).

8. S. Fouladi et al., "Efficient deep neural networks for classification of COVID-19 based on CT images: virtualization via software defined radio," *Comput. Commun.* **176**, 234–248 (2021).
9. H. Rafieipour et al., "Study of genes associated with Parkinson disease using feature selection," *J. Bioeng. Res.* **2**(4), 1–12 (2020).
10. F. Heydarpour et al., "Solving an optimal control problem of cancer treatment by artificial neural networks," *Int. J. Interact. Multimedia Artif. Intell.* **6**(4), 18–25 (2020).
11. A. Altaher et al., "Using multi-inception CNN for face emotion recognition," *J. Bioeng. Res.* **3**(1), 1–12 (2020).
12. M. F. Udvardy, "Notes on the ecological concepts of habitat, biotope and niche," *Ecology* **40**, 725–728 (1959).
13. R. K. M. Malhi et al., "Synergistic evaluation of Sentinel 1 and 2 for biomass estimation in a tropical forest of India," *Adv. Space Res.* (2021).
14. R. K. M. Malhi et al., "An integrated spatiotemporal pattern analysis model to assess and predict the degradation of protected forest areas," *ISPRS Int. J. Geo-Inf.* **9**(9), 530 (2020).
15. J. R. Busby, "BIOCLIM-a bioclimate analysis and prediction system," *Plant Prot. Q.* **6**, 8–9 (1991).
16. G. Carpenter, A. Gillison, and J. Winter, "DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals," *Biodivers. Conserv.* **2**(6), 667–680 (1993).
17. J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *J. Anim. Ecol.* **77**(4), 802–813 (2008).
18. J. Peters et al., "Random forests as a tool for ecohydrological distribution modelling," *Ecol. Modell.* **207**(2-4), 304–318 (2007).
19. J. Franklin, *Mapping Species Distributions: Spatial Inference and Prediction*, Cambridge University Press (2010).
20. S. J. Phillips and M. Dudík, "Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation," *Ecography* **31**(2), 161–175 (2008).
21. J. Leathwick, J. Elith, and T. Hastie, "Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions," *Ecol. Modell.* **199**(2), 188–196 (2006).
22. M. Bobrowski et al., "Searching for ecology in species distribution models in the Himalayas," *Ecol. Modell.* **458**, 109693 (2021).
23. A. Guisan and C. Rahbek, "SESAM-A new framework for predicting spatio-temporal patterns of species assemblages: Integrating macroecological and species distribution models," *J. Biogeogr.* **38**, 1433–1444 (2011).
24. J. Elith and J. R. Leathwick, "Species distribution models: ecological explanation and prediction across space and time," *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).
25. E. Saupe et al., "Variation in niche and distribution model performance: the need for a priori assessment of key causal factors," *Ecol. Modell.* **237**, 11–22 (2012).
26. L. R. D. A. Carneiro et al., "Limitations to the use of species-distribution models for environmental-impact assessments in the Amazon," *PLoS One* **11**(1), e0146543 (2016).
27. R. G. Pearson and T. P. J. G. E. Dawson, "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" *Global Ecol. Biogeogr.* **12**(5), 361–371 (2003).
28. R. K. M. Malhi et al., "Synergetic use of in situ and hyperspectral data for mapping species diversity and above ground biomass in Shoolpaneshwar Wildlife Sanctuary, Gujarat," *Trop. Ecol.* **61**(1), 106–115 (2020).
29. P. Singh et al., "Delineation of ground water potential zone and site suitability of rainwater harvesting structures using remote sensing and *in-situ* geophysical measurements," in *Advances in Remote Sensing For Natural Resource Monitoring*, Vol. 1, P. C. Pandey and L. K. Sharma Eds., John Wiley & Sons Ltd. (2020).
30. P. C. Pandey, A. Anand, and P. K. Srivastava, "Spatial distribution of mangrove forest species and biomass assessment using field inventory and earth observation hyperspectral data," *Biodiver. Conserv.* **28**(8), 2143–2162 (2019).

31. R. J. Hijmans et al., "Very high resolution interpolated climate surfaces for global land areas," *Int. J. Climatol.: J. R. Meteorol. Soc.* **25**(15), 1965–1978 (2005).
32. E. Kalnay et al., "The NCEP/NCAR 40-year reanalysis project," *Bull. Am. Meteorol. Soc.* **77**(3), 437–472 (1996).
33. F. Molteni et al., "The ECMWF ensemble prediction system: methodology and validation," *Quart. J. R. Meteorol. Soc.* **122**(529), 73–119 (1996).
34. G. Negi et al., "Impact of climate change on the western Himalayan mountain ecosystems: an overview," *Trop. Ecol.* **53**(3), 345–356 (2012).
35. J. Salick, Z. Fang, and A. Byg, "Eastern Himalayan alpine plant ecology, Tibetan ethnobotany, and climate change," *Global Environ. Change* **19**(2), 147–155 (2009).
36. T. Yao et al., "δ 18 O record and temperature change over the past 100 years in ice cores on the Tibetan Plateau," *Sci. China Ser. D* **49**(1), 1–9 (2006).
37. IPCC, "Summary for policymakers," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker et al., Eds., Cambridge University Press, Cambridge, United Kingdom and New York (2014).
38. U. B. Shrestha and K. S. Bawa, "Impact of climate change on potential distribution of Chinese caterpillar fungus (Ophiocordyceps sinensis) in Nepal Himalaya," *PLoS One* **9**(9), e106405 (2014).
39. A. Fischer, M. Blaschke, and C. Bässler, "Altitudinal gradients in biodiversity research: the state of the art and future perspectives under climate change aspects," *Waldökol. Landschaft. Nat.* **11**, 35–47 (2011).
40. A. S. Jump, T. J. Huang, and C. H. Chou, "Rapid altitudinal migration of mountain plants in Taiwan and its implications for high altitude biodiversity," *Ecography* **35**(3), 204–210 (2012).
41. P. Kumar, "Assessment of impact of climate change on Rhododendrons in Sikkim Himalayas using Maxent modelling: limitations and challenges," *Biodivers. Conserv.* **21**(5), 1251–1266 (2012).
42. S. N. Veera et al., "Prediction of upslope movement of *Rhododendron arboreum* in Western Himalaya," *Trop. Ecol.* **60**, 518–524 (2020).
43. S. Ranjitkar et al., "Separation of the bioclimatic spaces of Himalayan tree rhododendron species predicted by ensemble suitability models," *Global Ecol. Conserv.* **1**, 2–12 (2014).
44. P. Rawat et al., "Review on *Rhododendron arboreum*-a magical tree," *Orient. Pharm. Exp. Med.* **17**(4), 297–308 (2017).
45. P. K. Sonar et al., "Isolation, characterization and activity of the flowers of *Rhododendron arboreum* (Ericaceae)," *E-J. Chem.* **9**(2), 631–636 (2012).
46. G. Secretariat, "GBIF backbone taxonomy," Checklist Dataset [cited 2017 Nov 14], Vol. 10 (2017).
47. J. Veefkind et al., "TROPOMI on the ESA Sentinel-5 precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," *Remote Sens. Environ.* **120**, 70–83 (2012).
48. J. B. Fisher et al., "ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station," *Water Resour. Res.* **56**(4), e2019WR026058 (2020).
49. J. J. Van Zyl, "The shuttle radar topography mission (SRTM): a breakthrough in remote sensing of topography," *Acta Astron.* **48**(5–12), 559–565 (2001).
50. T. Hengl et al., "Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential," *PeerJ.* **6**, e5457 (2018).
51. C. E. Shannon, "A mathematical theory of communication (concluded)," *Bell. Syst. Tech. J* **27**, 379–423 (1948).
52. E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.* **106**(4), 620 (1957).
53. E. Moreno-Amat et al., "Impact of model complexity on cross-temporal transferability in Maxent species distribution models: an assessment using paleobotanical data," *Ecol. Modell.* **312**, 308–317 (2015).
54. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).

55. J. Mascaro et al., "A tale of two "forests": random forest machine learning aids tropical forest carbon mapping," *PLoS One* **9**(1), e85993 (2014).
56. J. S. Evans et al., "Modeling species distribution and change using random forest," in *Predictive Species and Habitat Modeling in Landscape Ecology*, C. A. Drew, Y. F. Wiersma, and F. Huettmann, Eds., pp. 139–159, Springer, New York (2011).
57. A. Liaw and M. Wiener, "Classification and regression by random forest," *R. News* **2**(3), 18–22 (2002).
58. H. Hashimoto et al., "High-resolution mapping of daily climate variables by aggregating multiple spatial data sets with the random forest algorithm over the conterminous United States," *Int. J. Climatol.* **39**(6), 2964–2983 (2019).
59. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science & Business Media (2009).
60. G. De'ath and K. E. Fabricius, "Classification and regression trees: a powerful yet simple technique for ecological data analysis," *Ecology* **81**(11), 3178–3192 (2000).
61. R. E. Schapire, "The boosting approach to machine learning: an overview," in *Nonlinear Estimation and Classification*, D. D. Denison et al., Eds., pp. 149–171, Springer, New York (2003).
62. P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Routledge (2019).
63. F. E. Harrell, Jr, *Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, Springer (2015).
64. A. Anand, S. K. Singh, and S. Kanga, "Estimating the change in forest cover density and predicting NDVI for West Singhbhum using linear regression," *Int. J. Environ. Rehabil. Conserv.* **9**, 193–203 (2018).
65. A. Davison, "Biometrika centenary: theory and general methodology," *Biometrika* **88**, 13–52 (2001).
66. H. Reiss et al., "Species distribution modelling of marine benthos: a North Sea case study," *Mar. Ecol. Prog. Ser.* **442**, 71–86 (2011).
67. A. Tsoar et al., "A comparative evaluation of presence-only methods for modelling species distribution," *Divers. Distrib.* **13**(4), 397–405 (2007).
68. J. Elith et al., "Novel methods improve prediction of species' distributions from occurrence data," *Ecography* **29**(2), 129–151 (2006).
69. R.-Y. Duan et al., "The predictive performance and stability of six species distribution models," *PloS One* **9**(11), e112764 (2014).
70. J. G. Giovanelli et al., "Modeling a spatially restricted distribution in the Neotropics: how the size of calibration area affects the performance of five presence-only methods," *Ecol. Modell.* **221**(2), 215–224 (2010).
71. S. J. Sinclair, M. D. White, and G. R. Newell, "How useful are species distribution models for managing biodiversity under future climates?" *Ecol. Soc.* **15**(1), 8 (2010).
72. M. Rostami, K. Berahmand, and S. Forouzandeh, "A novel method of constrained feature selection by the measurement of pairwise constraints uncertainty," *J. Big Data* **7**(1), 83 (2020).
73. A. Anand et al., "Integrating multi-sensors data for species distribution mapping using deep learning and envelope models," *Remote Sens.* **13**(16), 3284 (2021).

Biographies of the authors are not available.