

Article

Integrating Multi-Sensors Data for Species Distribution Mapping Using Deep Learning and Envelope Models

Akash Anand ¹, Manish K. Pandey ¹, Prashant K. Srivastava ^{1,*}, Ayushi Gupta ¹
and Mohammed Latif Khan ²

¹ Remote Sensing Laboratory, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi 221005, Uttar Pradesh, India; anand97aakash@gmail.com (A.A.); pandey.manish@live.com (M.K.P.); ayushi.gupta10@bhu.ac.in (A.G.)

² Department of Botany, Dr. Harisingh Gour Central University, Sagar 470003, Madhya Pradesh, India; mlkhan@dhgsu.edu.in

* Correspondence: prashant.iesd@bhu.ac.in

Abstract: The integration of ecological and atmospheric characteristics for biodiversity management is fundamental for long-term ecosystem conservation and drafting forest management strategies, especially in the current era of climate change. The explicit modelling of regional ecological responses and their impact on individual species is a significant prerequisite for any adaptation strategy. The present study focuses on predicting the regional distribution of *Rhododendron arboreum*, a medicinal plant species found in the Himalayan region. Advanced Species Distribution Models (SDM) based on the principle of predefined hypothesis, namely BIOCLIM, was used to model the potential distribution of *Rhododendron arboreum*. This hypothesis tends to vary with the change in locations, and thus, robust models are required to establish nonlinear complex relations between the input parameters. To address this nonlinear relation, a class of deep neural networks, Convolutional Neural Network (CNN) architecture is proposed, designed, and tested, which eventually gave much better accuracy than the BIOCLIM model. Both of the models were given 16 input parameters, including ecological and atmospheric variables, which were statistically resampled and were then utilized in establishing the linear and nonlinear relationship to better fit the occurrence scenarios of the species. The input parameters were mostly acquired from the recent satellite missions, including MODIS, Sentinel-2, Sentinel-5p, the Shuttle Radar Topography Mission (SRTM), and ECOSTRESS. The performance across all the thresholds was evaluated using the value of the Area Under Curve (AUC) evaluation metrics. The AUC value was found to be 0.917 with CNN, whereas it was 0.68 with BIOCLIM, respectively. The performance evaluation metrics indicate the superiority of CNN for species distribution over BIOCLIM.

Keywords: spatial distribution modelling; convolutional neural network; *Rhododendron arboreum*; biodiversity management; ecological responses



Citation: Anand, A.; Pandey, M.K.; Srivastava, P.K.; Gupta, A.; Khan, M.L. Integrating Multi-Sensors Data for Species Distribution Mapping Using Deep Learning and Envelope Models. *Remote Sens.* **2021**, *13*, 3284. <https://doi.org/10.3390/rs13163284>

Academic Editor: Parth Sarathi Roy

Received: 6 June 2021

Accepted: 11 August 2021

Published: 19 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Himalayan ecosystem is experiencing a continuous temperature rise, and the impact of climate change can be seen very clearly in the Himalayas, which demonstrates the need to monitor the Himalayan ecosystem even more [1,2]. The Himalayas are home to several medicinally and economically important plant species, *Rhododendron* species with botanical name *Rhododendron arboreum* Sm. from the family Ericaceae is among one of them [3–5]. It is widely spread in Himalayas, South India, and Sri Lanka [4]. With tremendous biological significance, it can sustain itself in the fragile ecotone between the alpine and subalpine biomes. Despite being identified as a medicinally important plant species, the geographical distribution and geospatial modelling of *Rhododendron arboreum* have not been explored to its fullest and needs to be deciphered, which will further benefit the formulation of conservation strategies [6]. The literature review of past

studies offered isolated information on the distribution pattern, frequency of the species, genetic diversity, and net productivity of *Rhododendron arboreum*, particularly in Himalayan regions. The studies primarily focus on the areas that are found over the Mussoorie hills of the Uttarakhand [7], Himachal Pradesh [8], and Garhwal division of Himalayas [9] and mainly showcased the threat of habitat fragmentation and frequency degradation of *Rhododendron arboreum* over the Himalayan region. Therefore, a holistic and cohesive research approach is needed to track the distribution of these species so as to generate baseline data for future research programmes, management practices, and conservation policies.

Throughout the centuries, researchers have observed and documented the linear relationship between species distribution and their local physical ecosystem and the role of climate and altitudinal variation in their occurrence, which can be found in the available scientific writings of the early 19th century [10–12]. Identifying the parameters for establishing the relationship between species and the environment is the core step for simulating the geographical distribution of any species [13,14]. Species Distribution Modelling (SDM) or Ecological Niche Modelling (ENM) is widely used in biogeography [15], macroecology [16], and biodiversity [17] research to model the geographical distribution of species. It is a statistical tool that performs habitat suitability analysis using high-dimensional digital data through regression or machine learning algorithms. Distribution modelling is achieved by establishing a linear or nonlinear relationship between regional climatic and ecological conditions. As each species is adapted to specific tolerance zones (also known as niches), SDM helps to identify the environmental constraints and simulates the n-dimensional input data to produce habitat suitability [18].

Generally, the SDM techniques take in geocoded input data of species distribution and establish a relationship with the regional environmental and climatic conditions to map its distribution throughout an area of interest [14,19,20]. One of the most commonly used SDM is BIOCLIM, which is also used in the present study. BIOCLIM is one of the earliest developed SDM algorithms, and it is mainly used by ecologists due to its easy-to-use algorithm, which is more accessible than other models. BIOCLIM was first introduced by reporting bioclimatic profiles and distribution maps of 73 species. After introducing BIOCLIM, several researchers applied the model in different bioclimatic conditions and received better results [21–24]. They found BIOCLIM to be quite consistent. The reason behind BIOCLIM's popularity is credited to its predefined assumptions and less complex training algorithm.

The theoretical aspect of SDM considers that a given species is likely to be found in a single privileged ecological niche under an ecosystem of unimodal distribution [25]. Still, an actual scenario is more complex and diverse than a hypothetical niche. To overcome this limitation, deep neural networks would be a good option, as their architecture favours high order multi-dimensional feature interactions without constraining their functional form [26]. Deep neural networks have shown significantly better results in image classification, and there are several cases where single-layered neural networks have been used for SDM [27,28]. Recently, a study by [29] has shown a better deep neural network prediction ability in SDM that has been found to be even better performing than the conventional ecological SDM models. A detailed discussion of convolution neural networks in handling and learning non-linear features are given in later sections.

In this paper, a study was conducted to map the distribution of *Rhododendron arboreum* using linear models, namely BIOCLIM, and Convolutional Neural Network (CNN) architecture has been proposed to establish nonlinear relationship between input parameters. A total of 16 environmental and climatic parameters acquired from different satellites are used as input parameters, which influence the distribution of a particular species to a greater extent in the given study area.

2. Materials and Methods

2.1. Study Area

The current study was conducted within four districts of Uttarakhand, India, namely Chamoli, Almora, Bagheshwar, and Pithoragarh, which is situated at the foothills of the Himalayas. Geographically, the study area lies between $28^{\circ}43'22.42''$ to $31^{\circ}27'22.06''$ N latitude and $77^{\circ}34'20.28''$ to $81^{\circ}2'34.35''$ E longitude with an area of around 20,736.99 km². The study area is shown in Figure 1. Being situated in the Himalayan Mountain range, the variation in ground elevation is very sharp, varying from 416 to 7801 m from mean sea level for the present study area. Due to the variation in elevation and unique climatic settings, this region is immensely rich with thousands of different plant species and has a remarkable diversity in flora and fauna [30,31]. This region experiences evenly distributed rainfall throughout the year and has an average temperature of 23.4 °C [32]. According to terrestrial ecoregion classifications that have been previously performed, the ecoregions found in the present study area are tropical and subtropical moist broadleaf forest, coniferous forest, temperate broadleaf and mixed forest, temperate conifer forest, and montane grassland and shrublands, and approximately 20% of the area is covered with snow throughout the year.

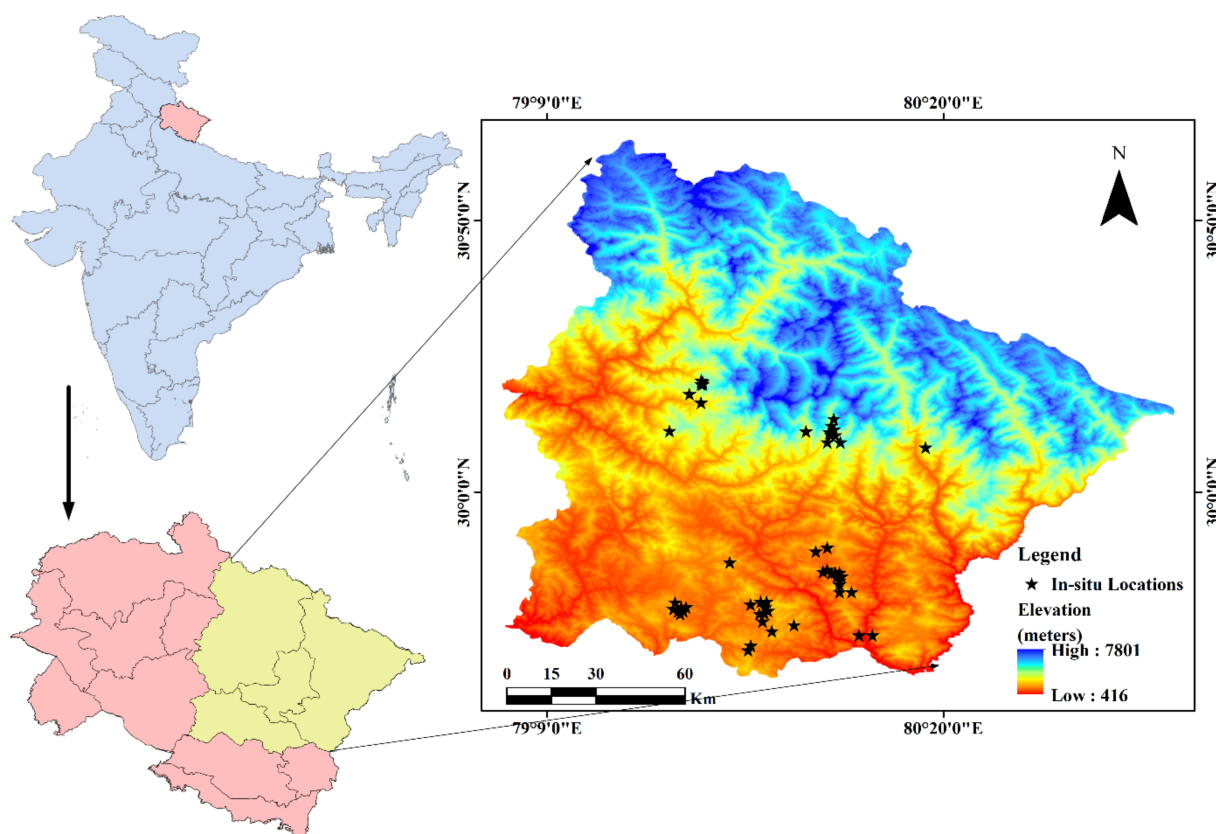


Figure 1. Location of the study area in the Western Himalayan region.

2.2. Target Species and Occurrence Data

Rhododendron species are found in the Himalayan range, which exhibits vast biological significance in the fragile ecotone of the alpine and subalpine zones [33]. Among the variety of Rhododendrons, *Rhododendron arboreum* is also a common species in the Western Himalayan region and can be found at an elevation range of 1200–4000 m above mean sea level. It shows some characteristics of invasive species and has a high medicinal value that makes the study of its distribution and the impact of climatic and ecological parameters on its growth very important [34]. Not only does this species carry medicinal value, but it is also very highly valued economically [35]. Medicinally, it is found to possess

anti-cancer, immunomodulatory, anti-inflammatory, hepatoprotective, antidiabetic, antioxidant, antidiarrheal, adaptogens, antimicrobial, and antinociceptive properties, among others [35]. Economically, it has found its usage in squash, local brew, jellies, jams, and *sherbet (rhodojuice)*. The juice from the leaves is used to encounter bed bugs bites. The wood from *Rhododendron* can be used to make tools used in agriculture such as ‘*Khukri*’ handles, etc. [36]. The leaves are used for decoration purposes in houses as well as in temples [37]. The wood is also used for the preparation of charcoal and can be used as a fuel. Some studies reported that consuming the squash made from the flowers can serve as a treatment for mental retardation [38,39], and flowers along with the roots and bark were found to be effective in treating digestive, heart, and respiratory complications [40]. The leaves of the plant burnt with juniper leaves are used to cleanse the air [41]. Menstrual cramps and heartaches are treated with the juice and squash made out of these flowers [42]. The extracts of the plants have also been utilized in curing nasal bleeding [43], headache, fever, rheumatism, wounds, dysentery [44], cough, skin diseases, liver malfunction, piles, worms, and jaundice as well as for preliminary cancer treatment [45].

Since phonological responses are better observed during the flowering season, the ground data sampling was done in September 2019, March 2020, and March 2021 at different elevation ranges [46]. The amount of ground data available for certain species, and particularly for *Rhododendron arboreum*, are limited, which makes it crucial to undertake the possibility of bias based on the sample size. As per the study by [47], they found that the SDM models with a smaller size consistently performed poorly and suggested that for reliable accuracy, the sample size must be greater than 30. In purview of the studies conducted by several researchers and considering the local geography, a total 65 homogenous patches of *Rhododendron arboreum* were identified and geotagged within the study area using the handheld Garmin GPS (Global Positioning System) with a horizontal accuracy of $95\% \pm 9.3$ m. According to Wisz et al. [47], a small sample size (>30) is generally useful in exploratory modelling, and considering that the present study is conducted at a regional scale for a single target species, 65 sample points are enough for regional distribution modelling. The observed multiple species occurrence within the pixels were removed by applying spatial rarefaction, which gave a single occurrence point per pixel. The photographs of the *Rhododendron arboreum* captured during field sampling are given in Figure 2.



Figure 2. Field photographs of *Rhododendron arboreum*.

2.3. Environmental Variables

The environmental variables used in this study include different bioclimatic variables acquired from various active satellites. Recent developments in satellite sensors have

enabled access to several ecological and climatic information derived from satellite observations at higher spatial and temporal resolutions. Current studies have used MODIS (<https://modis.gsfc.nasa.gov/data> accessed on 5 June 2021), Sentinel-2 (<https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products> accessed on 5 June 2021), Sentinel-5P (<https://sentinel.esa.int/web/sentinel/missions/sentinel-5/data-products> accessed on 5 June 2021), ECOSTRESS (<https://ecostress.jpl.nasa.gov/data> accessed on 5 June 2021), and SRTM (<https://www2.jpl.nasa.gov/srtm> accessed on 5 June 2021) satellite observations as an input parameter for the SDM algorithms. The satellite products are acquired during the sampling period (September 2019, March 2020, and March 2021) and then averaged, so that it can be used as a single product for each parameter. The Leaf Area Index (LAI) [48] and a fraction of Photosynthetically Active Radiation (fPAR) were retrieved from the MODIS data product. The motive behind using both LAI and fPAR establishes their direct relationship with the surface photosynthesis, evapotranspiration, and net primary productivity of the plants that is further utilized to estimate the water cycle processes, terrestrial energy, biophysical, and biochemical properties of the regional vegetation. Sentinel 2 optical data were used to estimate NDVI and EVI values at a fine scale of 10 m, which helped in understanding the vegetation status throughout the study area and were also used to mask the non-vegetative lands. Although there are a number of vegetation indices, which can play a crucial part in identifying the species distribution including Modified Soil Adjusted Vegetation Index (MSAVI) as well as other soil and ground surface adjusted indices, the target specie *Rhododendron Arboreum* is found in the dense forest cover of Himalayas and is independent of any soil and surface distortion; therefore, only NDVI and EVI were considered for SDM. Sentinel-5P is one of the most recent satellite missions from the European Space Agency (ESA), which is a combined mission of the European Union. It can take atmospheric measurements with the high spatial and temporal resolution and is utilized to retrieve several atmospheric parameters. Presently, eight different Sentinel-5P parameters have been used to establish a relationship between the existence of target species and to model species distribution which includes, the Aerosol Absorption Index (AAI), CO density, water vapor column, columnar NO₂, columnar O₃ level, SO₂ density, surface albedo, and tropospheric Formaldehyde (HCHO) density [49]. Evapotranspiration (ET) and Land Surface Temperature (LST) are also included to study the response of the target species with regional land processes. It was retrieved from the ECOSTRESS satellite. Elevation is an important parameter when studying the species distribution. It has a huge role in species growth and distribution due to the changing conditions at varying altitudes; therefore, SRTM Digital Elevation Model (DEM) data are used to retrieve the elevation factor at a spatial resolution of 30 m. A significant factor that is also included is terrestrial ecoregions to acquire the regional biome information to understand the diversity of the Himalayan ranges that coexist with each other. The ecoregion the vector data provided by [50] was used, which was attributed into 14 classes, 5 of which are used in the current study area that is found in the foothills of the Himalayas, namely tropical and subtropical moist broadleaf forests, tropical and subtropical coniferous forests, temperate broadleaf and mixed forests, temperate coniferous forests, and montane grassland/shrubland. The input parameters and their source mission are listed in Table 1.

All of the input parameters were used based on their linear and non-linear correlation with the occurrence data. The parameters are the yearly average for year 2020, considering the cloud free pixels only. Additionally, a correlation matrix was plotted to interpret the relationship between each parameter.

Table 1. Input parameters.

Satellite/Vector Data	Parameter	Unit	Spatial Resolution	Description
MODIS	LAI	Unitless	500 m	Defined as the projected area of leaves per unit of ground surface area.
	fPAR	Unitless	500 m	The fraction of photosynthetically active radiation (400–700 nm) absorbed by an integrated plant canopy.
Sentinel-5P	Aerosol Absorption Index	Unitless	0.01 arc degree	Indicates the elevated absorbed aerosols in the atmosphere.
	Vertically integrated CO column density	mol/m ²	0.01 arc degree	CO is an important atmospheric trace gas and a major atmospheric pollutant. A major source of CO is biomass burning and the oxidation of hydrocarbons.
	Water vapour column	mol/m ²	0.01 arc degree	A major greenhouse gas that directly impacts plant growth as well as photosynthesis.
	The total vertical column of NO ₂	mol/m ²	0.01 arc degree	A trace gas mostly found in the troposphere and stratosphere that can harm plant growth with an increase in its concentration
	The total atmospheric column of O ₃	mol/m ²	0.01 arc degree	Acts as a shield for the biosphere from solar ultraviolet radiation. It is an important greenhouse gas, and its high concentration can be harmful to the vegetation.
	SO ₂ vertical column density	mol/m ²	0.01 arc degree	Has a major impact on local and global climate change and is directly and indirectly related to plant growth and distribution.
	Surface Albedo	Unitless	0.01 arc degree	The flux per unit area received at the surface, and it shows low values in dense forest due to its high absorption.
	Tropospheric HCHO column number density	mol/m ²	0.01 arc degree	An intermediate gas in most of the oxidation chains of non-methane organic compounds. The inter-annual variations of HCHO distribution result from the oxidation in organic hydrocarbons from vegetation, fires, industrial sources, and temperature changes.
Sentinel-2	NDVI	Unitless	10 m	A simple indicator to assess whether or not the observed target contains green vegetation.
	EVI	Unitless	10 m	An optimized vegetation index to enhance the vegetation signal by decoupling the canopy background signal and reduction in atmospheric noises.
ECOSTRESS	Evapotranspiration	W/m ²	70 m	The latent heat flux coming from the earth's surface in the form of evaporation and plant transpiration.
	Land Surface Temperature	Kelvin	70 m	The radiative skin temperature of the earth's surface derived from solar radiation.
SRTM	DEM	Meters	30 m	An array of equally spaced elevation values referenced horizontally by a geographical coordinate system.
Terrestrial Ecoregions	Biome	Vector data		The classification of different types of forest present worldwide. The biome classification used for the present study has 14 different types of forest classes.

2.4. Ecological Niche Modelling

Broadly, there are two methods to model the ecological niche: a mechanistic approach and a correlative approach [47]. The mechanistic approach deals with the physiological limiting mechanisms of species intolerance in ecological conditions. In this approach, the growth parameters are taken into consideration, such as soil pH, nitrogen content in the soil, incoming solar radiation, carbon dioxide intake by plants, etc. [51]. The correlative

approach, also known as an empirical approach, uses the environmental variables that are reasonably expected to affect the growth of a particular species. The basis of the correlative approach is the interrelation between the observed parameters within the identified species location, which is used to establish a relationship between the parameters to model the species distribution for an entire area [52]. Having run the algorithm, a species distribution map can be generated using the established relationship. At this stage, the model's ability can also be tested using a set of species occurrence data that was not used in model development via suitable statistical parameters [53]. The SDM used in the present study only includes the presence of the BIOCLIM model and CNN based SDM. The representation of the conventional methodology is given in Figure 3.

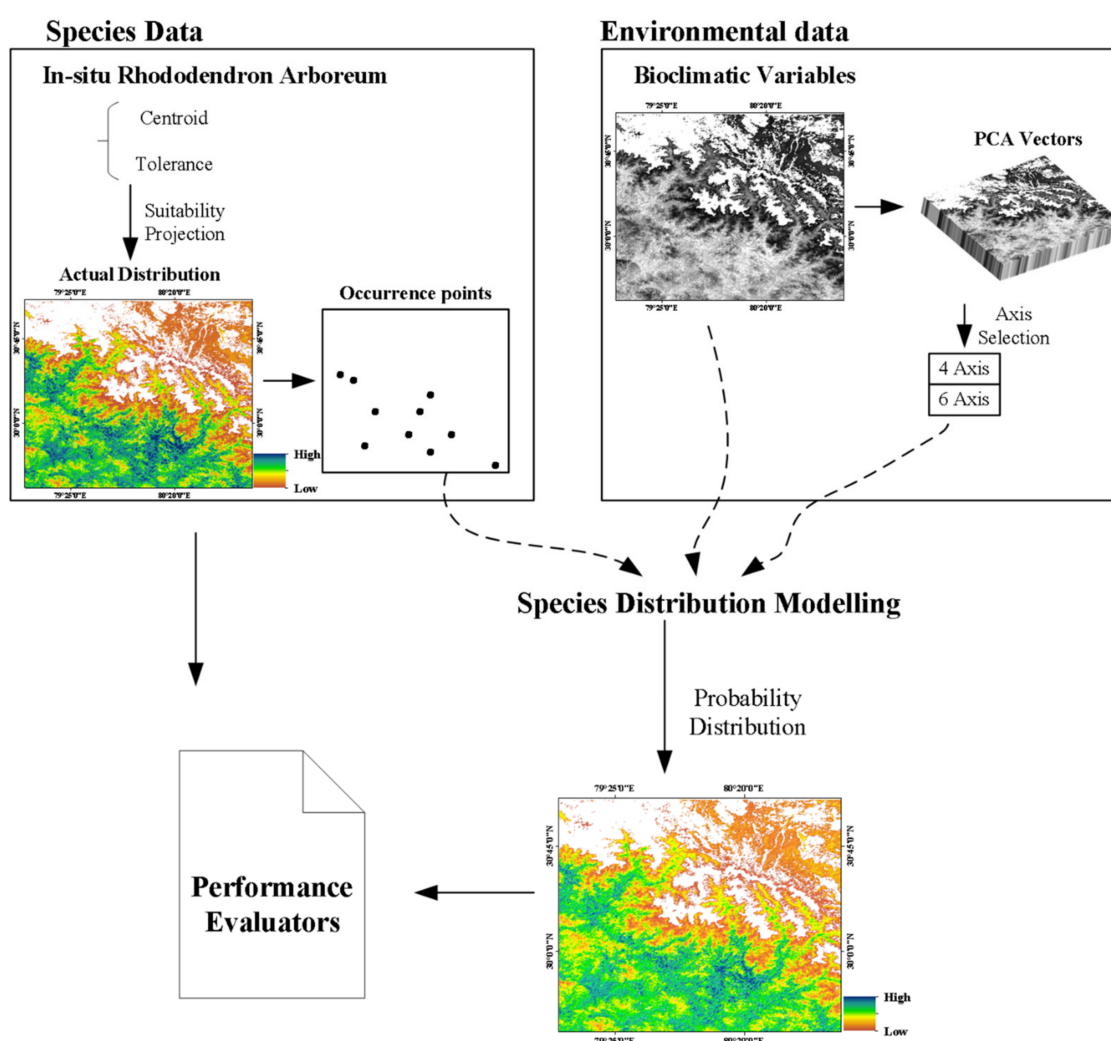


Figure 3. Flow diagram of the conventional approaches for species distribution modelling.

2.4.1. BIOCLIM

BIOCLIM is one of the first SDM algorithms to be introduced by [54]. The BIOCLIM is a widely used SDM due to its easy-to-use graphical user interface and wide application area. After the introduction of BIOCLIM, many researchers published their work using this algorithm, including the work of [55], in which they discussed the application of BIOCLIM in building ecographic regions and ways to improve the estimation of the ecological distance between patches in meta-population landscape dynamics. The study by [55] also pointed out some pros and cons, which include the error associated with the climatic parameters, defining the ranking of factors, and taxonomic uncertainty. Early BIOCLIM

applications occurred between 1984–1991 in terms of ecology and conservational biology and were addressed by [56].

BIOCLIM is based on a bioclimatic envelope model, which is widely used to predict the potential species distribution, which does not account for any possible interrelation between variables. Being an intuitively simple model, it assigns equal weight to each variable and produces binary predictions [57,58]. To predict the probability species distribution, BIOCLIM compares the values of input variables of known locations to the values of unknown pixels. The closer the value of an unknown pixel to the available pixel, the more suitable the location is for a particular species to be found. BIOCLIM is simple and intuitive. It is susceptible to over prediction and as specified, does not account for the interactions between input variables [59].

2.4.2. CNN

A deep neural network is a multi-layered model that can learn complex nonlinear relationships between the input parameters. The current study is an attempt that has been made to use the Convolutional Neural Network (CNN) architecture for SDM. During the last two decades, there has been a huge increase in deep learning and advanced machine learning algorithms in a variety of research fields [60,61]. Deep learning conducts high-level data abstraction using a hierarchical architecture consisting of multiple interconnected layers with multiple artificial neurons. The neurons receive the input values and multiply them with the specific weight obtained through optimization. Thereafter, the weighted sum is transformed through the nonlinear activation function to further pass it to the neurons of the next layer. The CNN architecture is represented in Figure 4 and the pseudocode is given in Table S1. Through this procedure, the network will learn through the optimal set of weights between the neurons in the adjoining layers and will maximize the network performance, which would help the neurons focus on specific patterns in the data. In the final layer, the parameters are passed through the SoftMax function, which transforms them into probabilities that sum to 1, as shown in Equation (1).

$$\hat{p}_k = \sigma(s(x))_k = \frac{\exp(s_k(x))}{\sum_{j=1}^K \exp(s_j(x))} \quad (1)$$

where K is the total number of classes, $s(x)$ is a vector with the weight of each class for instance x , and $\sigma(s(x))_k$ is the calculated probability of x belonging to class k as per the assigned weight. Although there have been several works that have been conducted with SDM using shallow networks containing a single hidden layer, their performance is not as good as that of the multi-layer networks [62]. The authors of [63], used a multi-layered network for distribution modelling and achieved a better performance compared to the single-layered networks.

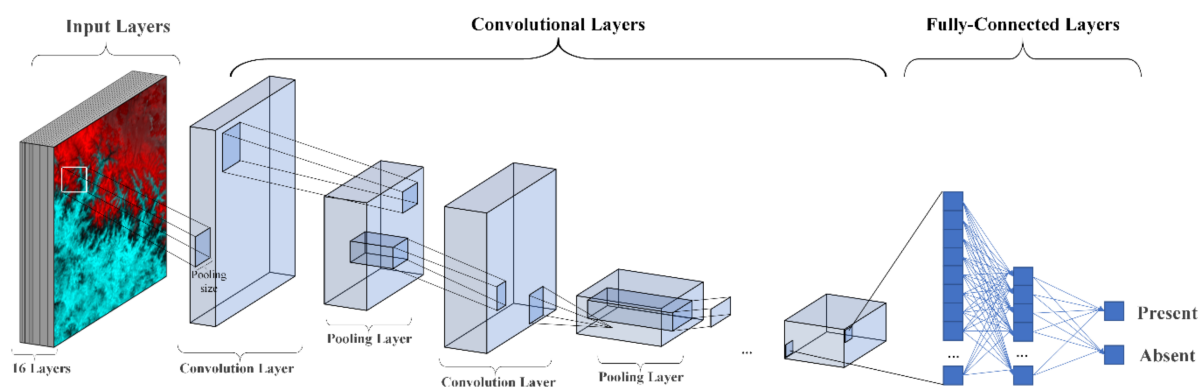


Figure 4. CNN architecture.

A Convolutional Neural Network (CNN) is a type of deep learning-based model for processing multidimensional data that follows a grid pattern [60]. The model is developed in such a way that the algorithm learns and adapts to the spatial hierarchies of features by itself from the lower to the higher levels of the pattern. Mathematically, it is composed of three layers or building blocks: convolution, pooling, and fully connected layers. Feature extraction is conducted using the first two layers and mapping the extracted features to the output is conducted by the third layer.

Convolution is used for feature extraction, in which a kernel is applied to an input tensor. A feature map is thus obtained through the product of kernel elements and tensor input. The procedure is then repeated on multiple kernels to obtain random feature maps that represent different feature extractors. The hyperparameters involved in convolution operations are the size and number of kernels. The size could be anything from 3×3 to 5×5 to 7×7 , and the kernel could be chosen randomly.

A pooling layer offers downsampling functionality that decreases the dimensionality of the feature maps to achieve translation invariance to the alterations and the biases incorporated and thus helps in reducing the number of learnable parameters. There are two types of pooling operations, namely Max Pooling and Global Average Pooling [64]. The first one extracts speckles from the input feature maps and offers maximum values in each of the speckles and leaves the remaining values unattended. The second one downsamples a feature map with a size equaling product of height and width into an array of a one cross one by averaging the elements of each feature map by retaining the depth of the feature map. The advantage of Global Average Pooling lies in reducing the number of learnable parameters along with offering the CNN with variable sized input.

The features extracted by the convolution layers followed by downsampling by the pooling layers are mapped using a subset of fully connected layers to the final output of the network. The fully connected layer is executed with the ReLU function [65,66]. Mathematically, the Rectifier can be described as:

$$f(x) = x^+ = \max(0, x) \quad (2)$$

where x is the input to the neuron. A unit employing the Rectifier is known as the Rectifier linear unit (ReLU).

The performance evaluation of the model is conducted by tuning the learnable parameters, kernels, and weight by a loss function through the forward propagation followed by updating these parameter values through an optimization algorithm either by backpropagation or gradient descent.

2.5. Model Validation

All of the input parameters are resampled in a single grid size of 100 m and are converted into the same file format. Out of the in-situ occurrences of *Rhododendron arboreum* at ground locations, only 70% of the data were used in calibrating the model, whereas the remaining 30% of the data were used to test the model. In any type of modelling, performance evaluation is an essential task. In terms of validation of species probability distribution, the AUC (Area Under ROC (Receiver Operating Characteristics) Curve) is one of the most used performance evaluation metrics [67]. The primary application of the ROC curve is in the threshold independent assessment that characterizes the model performance at various discrimination thresholds. This application was found in raster-based studies focusing on predicting land use and land cover, species distribution modelling, risk assessment, and other probability mappings.

The AUC is generated by plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) at varied thresholds. The TPR is also known as sensitivity, probability of detection, or recall, and the FPR is also known as the probability of false alarm. Therefore, an accurate model will generate a ROC curve away from the 1:1 line, and a less accurate model will have a ROC curve towards the 1:1 line. The range of the AUC varies from 0

to 1. The closer the value is to 1, the better the prediction is. The plots can be described mathematically as:

$$\text{TPR or Sensitivity or Recall or Probability of Detection} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \quad (3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \quad (4)$$

$$\text{FPR or Probability of false alarm} = 1 - \text{Specificity} \quad (5)$$

Here, TP stands for true positive, and FP is false positive, where specificity is also termed the true negative rate. The TPR provides the percentage of correctly predicted instances of species other than rhododendron, whereas specificity provides the percentage of correctly predicted instances of rhododendron distribution.

Thereafter, Cohen's kappa is also calculated to support the AUC value. Being one of the most popular performance evaluation indices, it is considered to be less complex and dependent on prevalence. The kappa value ranges from -1 to $+1$, where $+1$ indicates the perfect agreement. Other than kappa, the True Skill Statistic (TSS) is also incorporated, as it corrects the unimodel dependency of kappa. TSS is widely used in ecology, and it can be explained as

$$\text{TSS} = \text{Sensitivity} - \text{Specificity} - 1 \quad (6)$$

3. Results

The spatial species distribution is highly associated with regional environmental conditions, climatic variability, and land use [68,69]. The species distribution is simulated using the correlation models between the dependent as well as independent parameters. These models were generated through the presence-only data, presence/absence data, and pseudo-presence locations of the species. A total of 16 input parameters were taken from different satellite observations to model the potential distribution of *Rhododendron arboreum* confined to the current study area. The in situ species locations were recorded to be used as training and testing data and to retrieve the corresponding ecological and climatic satellite observations [70,71]. To understand the overall objective of the work, analysis was conducted on the distribution of the input parameters followed by the intercorrelation between them.

3.1. Assessing the Distribution of Input Parameters

As several input parameters were used from different satellite observations, a statistical downscaling was first performed to achieve a common spatial resolution of 100 m to be given as the model input. The statistically resampled images of different input parameters are shown in Figure 5. The yearly average was taken for each parameter to incorporate the overall variation throughout the year. AAI was found to range between -2.196 to 0.071 , in which the higher values were distributed where the higher altitudes have an upper limit of 7771 m and a lower limit of 379 m from mean sea level. This drastic variation in elevation permits rare species to grow in an extraordinary ecosystem, and it is the main reason for the higher species heterogeneity in this region. The EVI and NDVI derived from the Sentinel-2 optical data varies from -0.19 to 0.77 and -0.28 to 0.83 , respectively. The lower and lower-middle altitude locations tend to have higher NDVI/EVI values than the higher altitudes. As the LST has a linear relationship with altitude, a drastic variation in the upper and lower limits of LST can be found to be in the range of 242.2 to 306.1 Kelvin, respectively, which reflects the presence of glaciers on top of the Himalayan mountains. Due to the presence of dense vegetation at the lower altitudes, the values of water vapour, fPAR, LAI, and ET are also higher in the foothills and lower in the upper Himalayas. At the same time, atmospheric constituents like ozone, nitrogen dioxide, and carbon monoxide also show high values at the lower altitude, where the vegetation density and the presence of anthropogenic factors contributing to their concentration are relatively higher. However,

the concentration of SO_2 and HCHO are very low and are evenly distributed throughout, which directly relates to industrial and transportation activities, which is very low in these areas. The SO_2 varied from -0.0002 to 0.0004 mol/m^2 , and HCHO varied from -0.00005 to 0.00025 mol/m^2 .

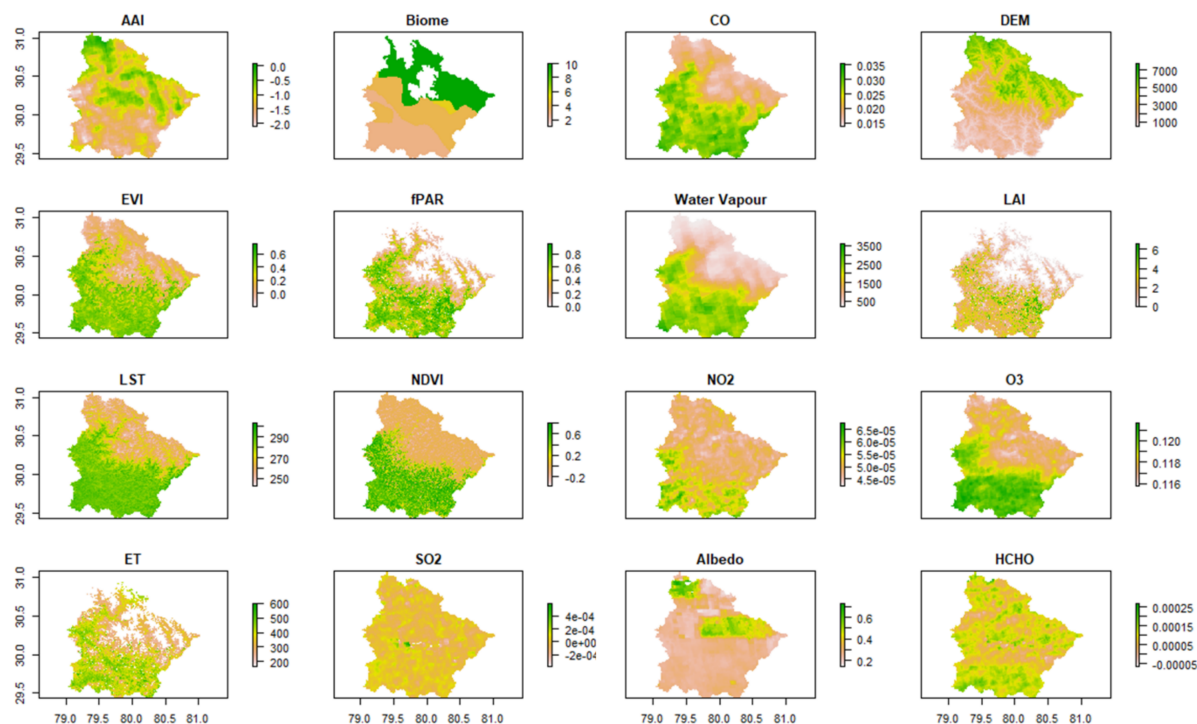


Figure 5. Spatial Distribution of various input parameters.

3.2. Understanding Parameter Intercorrelation

A correlation matrix plot was drawn as depicted in Figure 6 to understand the relationship between the input parameters. Total fifteen parameters except for the biome layer, which is in the vector form, were used in the correlation matrix. A highly linear or nonlinear relationship shows a relation/dependency or non-relation/non-dependency between the parameters. The representation can be explained in terms of the values varying from -1 to $+1$. The $-ve$ value represents the negative relationship, and the $+ve$ value describes the positive relationship. The depiction of the negative relationship is in orange, where the higher correlation value is visualized through the steeper circular shape, and vice versa for the positive relationship. No relationship is represented by the correlation value of zero that is represented by a perfect circular shape, and the colour becomes whitish. It can be observed that many parameters are related to each other. A highly linear relationship exists between Sentinel-5p based ozone and carbon monoxide as well as with ozone and water vapour with a correlation value of 0.99 . It can also be observed that the linear relationship of DEM with NDVI, LST, and water vapour is very high, which shows the variation in the local geometry and the influence of regional ecological and climatic parameters. A lower correlation value is observed between the atmospheric parameters and the vegetation indices, especially for EVI, LAI, and ET.

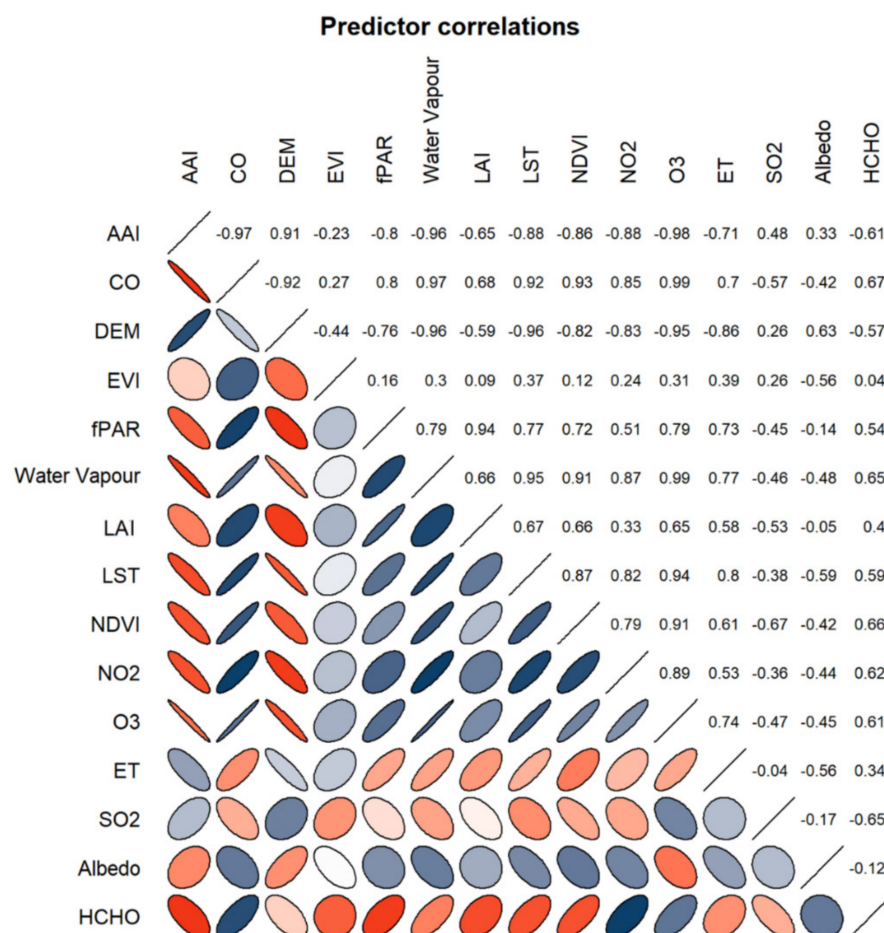


Figure 6. Square correlation matrix between input parameters.

3.3. Spatial Distribution of *Rhododendron arboreum*

To simulate the potential distribution of *Rhododendron arboreum*, linear and a nonlinear SDM were used in the current study using 16 a priori input parameters. The input parameters were taken from different satellite observations followed by data resampling to match their spatial resolution to achieve a standard resolution of 100 m. The probability distribution is classified into four classes, namely very low, low, high, and very high, according to their distribution. The presence, based only on the BIOCLIM model, predicts the probability distribution of species using a linear correlation, as shown in Figure 7a. Apart from a well-established presence only algorithm, a deep learning-based convolution neural network model was used to establish a nonlinear relationship between the input parameters to predict the probability of species distribution. A CNN based architecture was used to train the model according to the known locations and was fitted on different layers. The perfect combination of layers and activation functions was then used to predict the species distribution Figure 7b.

3.4. Model Validation and Comparison

The accuracy or performance of the probability distribution of the incorporated models was compared using the AUC, TSS, and kappa coefficient that characterize the performance of the models with an in situ validation dataset. Table 2 shows the statistical performance for BIOCLIM and CNN based the probability distribution of *Rhododendron arboreum*. A lower AUC value was obtained by the BIOCLIM models, which is 0.639, based on the in situ points reserved for the validation purpose. The AUC values for the CNN-based probability distribution was found to be 0.917, which is considered to be very good compared to the BIOCLIM that was given an AUC of 0.68. In addition to the AUC value, TSS and

kappa, with values 0.652 and 0.94, respectively, also gave support to the applicability of CNN in comparison to conventional SDM's such as BIOCLIM. These values showcase the superiority of deep learning models for species probability distribution using the given set of ecological and bioclimatic parameters.

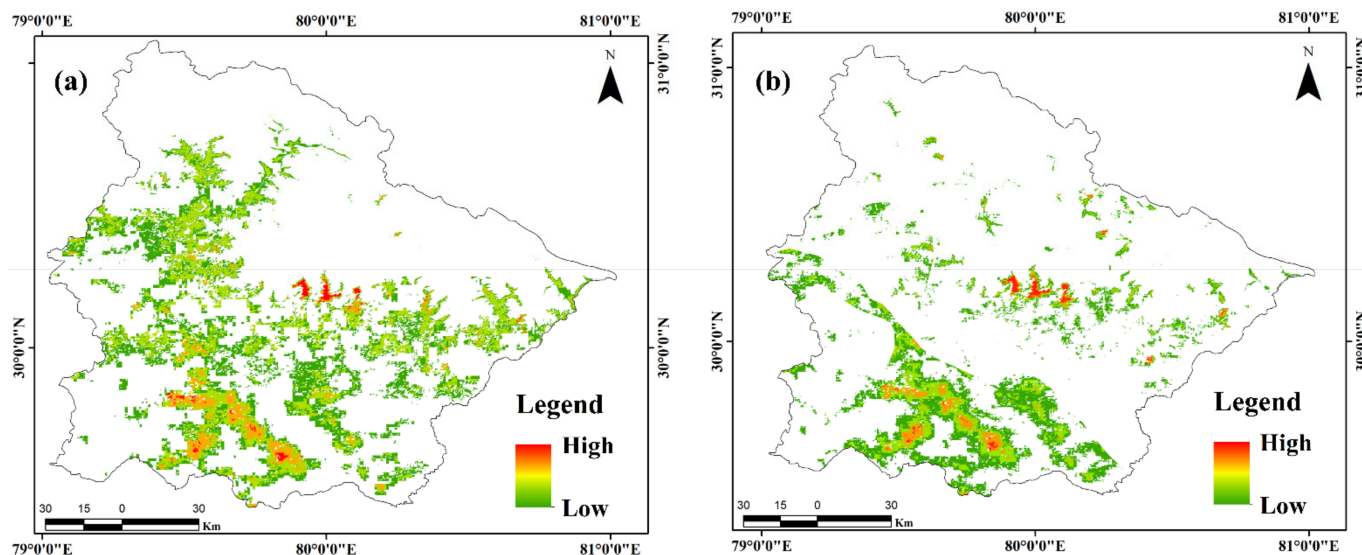


Figure 7. Probability distribution of *Rhododendron arboreum* using (a) BIOCLIM and (b) CNN Models.

Table 2. Statistical performance analysis of BIOCLIM and CNN.

	BIOCLIM	CNN
AUC	0.68	0.917
Kappa	0.76	0.94
TSS	0.44	0.652

4. Discussion

Understanding the dynamics and distribution of the forest ecosystem is a crucial step towards biodiversity conservation. The recent advancements in statistical machine learning models and the availability of reliable datasets help researchers build policies towards conservation and sustainable solutions to achieve this conservation. The SDMs came into existence in the mid-1980s with very limited ecological and climatic datasets, and from there, they have evolved with the regular integration of newer and more reliable datasets.

Recently, a great variety of SDMs have been used to model the distribution of species, in which the most popular are the ones that are based on the non-linear modelling approach followed by statistical and rule-based methods. Among machine learning models, Maxent is widely used due to its user-friendly interface and simple background algorithm. Maxent accuracy, as per [72], has provided robust predictions with an AUC of 0.75 and for BIOCLIM, an AUC of 0.65, whereas a similar result is achieved by [73] with an AUC of 0.73 and 0.66 for Maxent and BIOCLIM, respectively. Another well performing algorithm is the Boosted Regression Tree (BRT), which performs slightly better than Maxent. As per [74], they archived an overall AUC of 0.81 using BRT. Statistical and rule-based methods are among the conventional approaches that are not consistent with the changes in regional ecology.

The BIOCLIM model is one of the oldest yet most used SDMs due to its simple algorithm and easy parameterization [75]. It is based on a linear bioclimatic envelop model that assigns equal weight to each variable and offers a binary prediction. A previous study indicated the use of BIOCLIM to predict species distribution, and similar studies were conducted in the past. The introduction of machine learning and its integration in SDM revolutionized the probability distribution modelling approach. The variety in

modelling approaches and the increased number of datasets has made this model one of the most globally accepted SDMs. There is a growing concern for the establishment of the nonlinear relationship between the bioclimatic parameters through innovative approaches such as deep learning-based models. It has been observed that the BIOCLIM model is overestimates the species distribution and the higher probability of species occurrence at the higher altitude. Moreover, it has been observed that the distribution pattern of the predicted *Rhododendron arboreum* distribution using CNN architecture is quite different at some places than from conventional BIOCLIM models. The current work proposed deep learning-based CNN architecture for probability distribution modelling and proved to perform better than the traditional BIOCLIM model. There was an underestimation of species distribution observed in CNN than BIOCLIM. The distribution probability in CNN was precise, and it was found that the majority of *Rhododendron arboreum* is distributed in the southern part of the study area where the vegetation density is high. Some high probability patches at the higher altitudes are commonly predicted by both the models.

However, the scalability of the current outcome needs to be tested on a global scale. Apart from this, some limitations, namely uncertainties associated with the input data, the assigned weights, and some important biotic parameters, need to be handled for future work. An ensemble of all of these available methods needs to be explored in the future to establish the linear and nonlinear relationship between the dependent parameters to predict any one species out of the multiple species that are available in a location. Additionally, there is a need to perform sensitivity analysis to understand variable impact on the target variable, instead of forcing all of the variables into the model. This would reduce the algorithm complexity and computational demand.

In spite of achieving significant accuracy and popularity in the field of correlation modelling, there is still not a single algorithm that can be recommended. Deep neural networks are showing more promising results, but they are still to be tested in different ecological settings.

5. Conclusions

This study is a novel approach towards establishing a CNN architecture and testing the performance of CNN in SDM and its comparison with other well established SDMs namely, BIOCLIM. This study was conducted on the foothills of the Himalayas, where the altitudinal variation is very drastic and varies from 416 to 7801 m above mean sea level. This high-altitude Himalayan ranges constitutes a heterogeneous ecosystem and is home to many rare/endangered, medicinally, and economically important plant species. One of the major economically and medicinally important plant species, *Rhododendron arboreum*, was tracked and mapped in this study using different SDMs. Based on its occurrence and several ecological and bioclimatic satellite-based observations, the probability distribution of the *Rhododendron arboreum* was established. The CNN based probability distribution model outperformed the presence only based BIOCLIM model with an AUC score of 0.917. The CNN based prediction was also found to be more precise and accurate and with significantly less overestimation, whereas the AUC values of the BIOCLIM model were found to be 0.68 with a high overestimation. The superiority of CNN implies the role of nonlinear parameters in predicting the probability of species distribution. The scalability of the current solution on a global scale, the addition of some other important parameters, and an ensemble of all of the available SDMs need to be explored in future work. An increase in the presence of the *Rhododendron* species is an indication of strong soil retention, which, in turn, is fruitful for other vegetation to grow and flourish. Apart from this, an increase in the green vegetation fraction and a decrease in shade fraction was found to be associated with a higher likelihood of *Rhododendron*. This increased likelihood using the models would offer researchers an opportunity to understand the vegetation distribution and to contribute to the restoration of the ecology and biodiversity conservation in the protected areas so that a provision could be established for sustainable ecosystem services.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/rs13163284/s1>. Table S1: Methodology of the convolution and full connected layer.

Author Contributions: Conceptualization, data curation, formal analysis: A.A., M.K.P., and P.K.S.; funding acquisition: P.K.S.; supervision: P.K.S. and M.L.K.; validation, visualization, writing—original draft: A.A., M.K.P., and P.K.S.; writing—review and editing: A.G., P.K.S., and M.L.K. All authors have read and agreed to the published version of the manuscript.

Funding: National Mission on Himalayan Studies, G.B. Pant National Institute of Himalayan Environment (NIHE), Almora, Uttarakhand, India.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the satellite datasets used in this study are available free of charge. Links are given as follows: MODIS (<https://modis.gsfc.nasa.gov/data> accessed on 5 June 2021), Sentinel-2 (<https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products> accessed on 5 June 2021), Sentinel-5P (<https://sentinel.esa.int/web/sentinel/missions/sentinel-5/data-products> accessed on 5 June 2021), ECOSTRESS (<https://ecostress.jpl.nasa.gov/data>), and SRTM (<https://www2.jpl.nasa.gov/srtm> accessed on 5 June 2021).

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 this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sabin, T.; Krishnan, R.; Vellore, R.; Priya, P.; Borgaonkar, H.; Singh, B.B.; Sagar, A. Climate change over the Himalayas. In *Assessment of Climate Change over the Indian Region*; Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A., Chakraborty, S., Eds.; Springer: Singapore, 2020; pp. 207–222.
2. Kraaijenbrink, P.D.; Bierkens, M.; Lutz, A.; Immerzeel, W. Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers. *Nature* **2017**, *549*, 257–260. [CrossRef]
3. Kala, C.P. Status and conservation of rare and endangered medicinal plants in the Indian trans-Himalaya. *Biol. Conserv.* **2000**, *93*, 371–379. [CrossRef]
4. Veera, S.N.; Panda, R.M.; Behera, M.D.; Goel, S.; Roy, P.S.; Barik, S.K. Prediction of upslope movement of Rhododendron arboreum in Western Himalaya. *Trop. Ecol.* **2019**, *60*, 518–524. [CrossRef]
5. Srivastava, P. Rhododendron arboreum: An overview. *J. Appl. Pharm. Sci.* **2012**, *2*, 158–162.
6. Bhandari, M.S.; Meena, R.K.; Shankhar, R.; Shekhar, C.; Saxena, J.; Kant, R.; Pandey, V.V.; Barthwal, S.; Pandey, S.; Chandra, G. Prediction mapping through maxent modeling paves the way for the conservation of Rhododendron arboreum in Uttarakhand Himalayas. *J. Indian Soc. Remote Sens.* **2020**, *48*, 411–422. [CrossRef]
7. Jain, A.; Pandit, M.K.; Elahi, S.; Jain, A.; Bhaskar, A.; Kumar, V. Reproductive behaviour and genetic variability in geographically isolated populations of Rhododendron arboreum (Ericaceae). *Curr. Sci.* **2000**, *79*, 1377–1381.
8. Sharma, G. *Development and Characterization of UGMS Markers for Genetic Diversity Analysis in Rhododendron Arboreum*; Guru Kashi University: Punjab, India, 2013.
9. Chauhan, D.; Lal, P.; Singh, D. Composition, population structure and regeneration of Rhododendron arboreum Sm. temperate broad-leaved evergreen forest in Garhwal Himalaya, Uttarakhand, India. *J. Earth Sci. Clim. Chang.* **2017**, *8*, 430. [CrossRef]
10. Humboldt, A.V.; Bonpland, A. *Ideen Zu Einer Geographie Der Pflanzen Nebst Einem Naturgemälde Der Tropenländer*; Cotta: Tübingen, Germany, 1807.
11. De Candolle, A. *Géographie Botanique Raisonnée Ou Exposition Des Faits Principaux Et Des Lois Concernant La Distribution Géographique Des Plantes De L'époque Actuelle*; V. Masson: Paris, France, 1855; Volume 2.
12. Udvardy, M.F. Notes on the ecological concepts of habitat, biotope and niche. *Ecology* **1959**, *40*, 725–728. [CrossRef]
13. Priti, H.; Aravind, N.; Shaanker, R.U.; Ravikanth, G. Modeling impacts of future climate on the distribution of Myristicaceae species in the Western Ghats, India. *Ecol. Eng.* **2016**, *89*, 14–23. [CrossRef]
14. Adhikari, D.; Barik, S.; Upadhyaya, K. Habitat distribution modelling for reintroduction of Ilex khasiana Purk., a critically endangered tree species of northeastern India. *Ecol. Eng.* **2012**, *40*, 37–43. [CrossRef]
15. Franklin, J. Moving beyond static species distribution models in support of conservation biogeography. *Divers. Distrib.* **2010**, *16*, 321–330. [CrossRef]
16. Vasconcelos, T.S.; Rodríguez, M.Á.; Hawkins, B.A. Species distribution modelling as a macroecological tool: A case study using New World amphibians. *Ecography* **2012**, *35*, 539–548. [CrossRef]

17. Rodríguez, J.P.; Brotons, L.; Bustamante, J.; Seoane, J. The application of predictive modelling of species distribution to biodiversity conservation. *Divers. Distrib.* **2007**, *13*, 243–251. [\[CrossRef\]](#)
18. Lorena, A.C.; Jacintho, L.F.; Siqueira, M.F.; De Giovanni, R.; Lohmann, L.G.; De Carvalho, A.C.; Yamamoto, M. Comparing machine learning classifiers in potential distribution modelling. *Expert Syst. Appl.* **2011**, *38*, 5268–5275. [\[CrossRef\]](#)
19. Elith, J.; Leathwick, J.R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. of Ecol. Evol. Syst.* **2009**, *40*, 677–697. [\[CrossRef\]](#)
20. Rana, S.K.; Rana, H.K.; Luo, D.; Sun, H. Estimating climate-induced 'Nowhere to go' range shifts of the Himalayan *Incarvillea* Juss. using multi-model median ensemble species distribution models. *Ecol. Indic.* **2021**, *121*, 107127. [\[CrossRef\]](#)
21. Beaumont, L.J.; Hughes, L.; Poulsen, M. Predicting species distributions: Use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecol. Model.* **2005**, *186*, 251–270. [\[CrossRef\]](#)
22. Doran, B.; Olsen, P. Customizing BIOCLIM to investigate spatial and temporal variations in highly mobile species. In Proceedings of the 6th International Conference in GeoComputation, Brisbane, Australia, 24–26 September 2001.
23. Xu, Y.; Zhou, P.-Y.; Wang, Y.; Chen, Z.-X.; Ma, R.; Yu, S.-F. Assessment of risk of introduction of pine wood nematode, *bursaphelenchus xylophilus* in Yunnan Province using BIOCLIM ecological niche model. *J. Yunnan Agric. Univ.* **2008**, *23*, 746–753.
24. Bhatta, K.P.; Robson, B.A.; Suwal, M.K.; Vetaas, O.R. A pan-Himalayan test of predictions on plant species richness based on primary production and water-energy dynamics. *Front. Biogeogr.* **2021**, *13*, e49459.
25. Mangain, A.; Uniyal, P.L. Species Distribution Modelling of *Rhododendron arboreum* Sm.—A Keystone Species, in India and Adjoining Region. *Int. J. Ecol. Environ. Sci.* **2018**, *44*, 261–286.
26. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT press: Cambridge, MA, USA, 2016.
27. Lek, S.; Delacoste, M.; Baran, P.; Dimopoulos, I.; Lauga, J.; Aulagnier, S. Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.* **1996**, *90*, 39–52. [\[CrossRef\]](#)
28. Thuiller, W. BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change. *Glob. Chang. Biol.* **2003**, *9*, 1353–1362. [\[CrossRef\]](#)
29. Zhang, J.; Li, S. A Review of Machine Learning Based Species' Distribution Modelling. In Proceedings of the 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICIT), Wuhan, China, 2–3 December 2017; pp. 199–206.
30. Kumar, K. *Water Management in Himalayan Ecosystem: A Study of Natural Springs of Almora*; Indus Publishing: New Delhi, India, 1996.
31. Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. [\[CrossRef\]](#)
32. Tewari, A.P. Recent changes in the position of the snout of the Pindari glacier (Kumaon Himalaya), Almora District, Uttar Pradesh, India. In *Proceedings of the Role of Snow and Ice in Hydrology, Banff Symposia, September 1972*; WMO-IAHS-Unesco: Geneva, Switzerland, 1973; pp. 1144–1149.
33. Singh, K.; Rai, L.; Gurung, B. Conservation of rhododendrons in Sikkim Himalaya: An overview. *World J. Agric. Sci.* **2009**, *5*, 284–296.
34. Secretariat, G. GBIF backbone taxonomy. *Checklist Dataset* **2017**, *10*. Available online: <https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c> (accessed on 5 June 2021).
35. Rawat, P.; Rai, N.; Kumar, N.; Bachheti, R. Review on *Rhododendron arboreum*—A magical tree. *Orient. Pharm. Exp. Med.* **2017**, *17*, 297–308. [\[CrossRef\]](#)
36. Paul, A.; Khan, M.L.; Arunachalam, A.; Arunachalam, K. Biodiversity and conservation of rhododendrons in Arunachal Pradesh in the Indo-Burma biodiversity hotspot. *Curr. Sci.* **2005**, *89*, 623–634.
37. Chauhan, N.S. *Medicinal and Aromatic Plants of Himachal Pradesh*; Indus Publishing: New Delhi, India, 1999.
38. Watts, J.S. *When a Billion Chinese Jump: How China Will Save Mankind—Or Destroy It*; Simon and Schuster: New York, NY, USA, 2010.
39. Singh, V.K.; Ali, Z.A. *Herbal Drugs of Himalaya*; Today & Tomorrow's Printers and Publishers: Delhi, India, 1998.
40. Singh, N.; Ram, J.; Tewari, A.; Yadav, R. Phenological events along the elevation gradient and effect of climate change on *Rhododendron arboreum* Sm. in Kumaun Himalaya. *Curr. Sci.* **2015**, *108*, 106–110.
41. Paul, A.; Khan, M.L.; Das, A.K. *Utilization of Rhododendrons by Monpas in Western Arunachal Pradesh, India*; Assam University: Silchar, India, 2010.
42. Negi, V.S.; Maikhuri, R.; Rawat, L.; Chandra, A. Bioprospecting of *Rhododendron arboreum* for livelihood enhancement in central Himalaya, India. *Environ. We Int. J. J. Sci. Technol.* **2013**, *8*, 61–70.
43. Uniyal, S.K.; Singh, K.; Jamwal, P.; Lal, B. Traditional use of medicinal plants among the tribal communities of Chhota Bhangal, Western Himalaya. *J. Ethnobiol. Ethnomedi.* **2006**, *2*, 1–8. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Sharma, P.; Samant, S. Diversity, distribution and indigenous uses of medicinal plants in Parbati Valley of Kullu district in Himachal Pradesh, Northwestern Himalaya. *Asian J. Adv. Basic Sci.* **2014**, *2*, 77–98.
45. Zhasa, N.; Hazarika, P.; Tripathi, Y. Indigenous knowledge on utilization of plant biodiversity for treatment and cure of diseases of human beings in Nagaland, India: A case study. *Int. Res. J. Biol. Sci.* **2015**, *4*, 89–106.
46. Kumar, P. Assessment of impact of climate change on Rhododendrons in Sikkim Himalayas using Maxent modelling: Limitations and challenges. *Biodivers. Conserv.* **2012**, *21*, 1251–1266. [\[CrossRef\]](#)

47. Wisz, M.S.; Hijmans, R.; Li, J.; Peterson, A.T.; Graham, C.; Guisan, A.; NCEAS Predicting Species Distributions Working Group. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **2008**, *14*, 763–773. [\[CrossRef\]](#)
48. Yang, W.; Tan, B.; Huang, D.; Rautiainen, M.; Shabanov, N.V.; Wang, Y.; Privette, J.L.; Huemmrich, K.F.; Fensholt, R.; Sandholt, I. MODIS leaf area index products: From validation to algorithm improvement. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1885–1898. [\[CrossRef\]](#)
49. Martin, R.; Parrish, D.; Ryerson, T.; Nicks, D., Jr.; Chance, K.; Kurosu, T.; Jacob, D.J.; Sturges, E.; Fried, A.; Wert, B. Evaluation of GOME satellite measurements of tropospheric NO₂ and HCHO using regional data from aircraft campaigns in the southeastern United States. *J. Geophys. Res. Atmos.* **2004**, *109*, 1–11. [\[CrossRef\]](#)
50. Olson, D.; Dinerstein, E.; Wikramanayake, E.; Burgess, N.; Powell, G.; Underwood, E.; d'Amico, J.; Itoua, I.; Strand, H.; Morrison, J. Terrestrial ecoregions of the world: A new map of life on earth. *BioScience* **2001**, *51*, 933–938. [\[CrossRef\]](#)
51. Kamei, J.; Pandey, H.; Barik, S. Tree species distribution and its impact on soil properties, and nitrogen and phosphorus mineralization in a humid subtropical forest ecosystem of northeastern India. *Can. J. For. Res.* **2009**, *39*, 36–47. [\[CrossRef\]](#)
52. Kala, C.P.; Mathur, V.B. Patterns of plant species distribution in the Trans-Himalayan region of Ladakh, India. *J. Veg. Sci.* **2002**, *13*, 751–754. [\[CrossRef\]](#)
53. Pearson, R.G. Species' distribution modeling for conservation educators and practitioners. *Synth. Am. Mus. Nat. Hist.* **2007**, *50*, 54–89.
54. Nix, H.A. A biogeographic analysis of Australian elapid snakes. *Atlas Elapid Snakes Aust.* **1986**, *7*, 4–15.
55. Haydon, D.T.; Pianka, E.R. Metapopulation theory, landscape models, and species diversity. *Ecoscience* **1999**, *6*, 316–328. [\[CrossRef\]](#)
56. Guisan, A.; Thuiller, W. Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* **2005**, *8*, 993–1009. [\[CrossRef\]](#)
57. Parthasarathy, U.; Saji, K.; Jayarajan, K.; Parthasarathy, V. Biodiversity of Piper in South India—application of GIS and cluster analysis. *Curr. Sci.* **2006**, *91*, 652–658.
58. Rameshprabu, N.; Swamy, P. Prediction of environmental suitability for invasion of Mikania micrantha in India by species distribution modelling. *J. Environ. Biol.* **2015**, *36*, 565.
59. Booth, T.H.; Nix, H.A.; Busby, J.R.; Hutchinson, M.F. BIOCLIM: The first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Divers. Distrib.* **2014**, *20*, 1–9. [\[CrossRef\]](#)
60. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [\[CrossRef\]](#)
61. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [\[CrossRef\]](#)
62. Fukuda, S.; De Baets, B.; Waegeman, W.; Verwaeren, J.; Mouton, A.M. Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environ. Model. Softw.* **2013**, *47*, 1–6. [\[CrossRef\]](#)
63. Harris, D.J. Generating realistic assemblages with a joint species distribution model. *Methods Ecol. Evol.* **2015**, *6*, 465–473. [\[CrossRef\]](#)
64. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv Prepr.* **2013**, arXiv:1312.4400.
65. Agarap, A.F. Deep learning using rectified linear units (relu). *arXiv Prepr.* **2018**, arXiv:1803.08375.
66. Schmidt-Hieber, J. Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Stat.* **2020**, *48*, 1875–1897.
67. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [\[CrossRef\]](#)
68. Pandey, P.C.; Anand, A.; Srivastava, P.K. Spatial distribution of mangrove forest species and biomass assessment using field inventory and earth observation hyperspectral data. *Biodivers. Conserv.* **2019**, *28*, 2143–2162. [\[CrossRef\]](#)
69. Anand, A.; Malhi, R.K.M.; Pandey, P.C.; Petropoulos, G.P.; Pavlides, A.; Sharma, J.K.; Srivastava, P.K. Use of Hyperion for Mangrove Forest Carbon Stock Assessment in Bhitarkanika Forest Reserve: A Contribution Towards Blue Carbon Initiative. *Remote Sens.* **2020**, *12*, 597. [\[CrossRef\]](#)
70. Malhi, R.K.M.; Anand, A.; Srivastava, P.K.; Kiran, G.S.; Petropoulos, G.P.; Chalkias, C. An Integrated Spatiotemporal Pattern Analysis Model to Assess and Predict the Degradation of Protected Forest Areas. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 530. [\[CrossRef\]](#)
71. Malhi, R.K.M.; Anand, A.; Mudaliar, A.N.; Pandey, P.C.; Srivastava, P.K.; Sandhya Kiran, G. Synergetic use of in situ and hyperspectral data for mapping species diversity and above ground biomass in Shoolpaneshwar Wildlife Sanctuary, Gujarat. *Trop. Ecol.* **2020**, *61*, 106–115. [\[CrossRef\]](#)
72. Elith, J.; Graham, C.H.; Anderson, R.P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **2006**, *29*, 129–151. [\[CrossRef\]](#)
73. Graham, C.H.; Elith, J.; Hijmans, R.J.; Guisan, A.; Townsend Peterson, A.; Loiselle, B.A.; Group, N.P.S.D.W. The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* **2008**, *45*, 239–247. [\[CrossRef\]](#)
74. Hallman, T.A.; Robinson, W.D. Comparing multi-and single-scale species distribution and abundance models built with the boosted regression tree algorithm. *Landsc. Ecol.* **2020**, *35*, 1161–1174. [\[CrossRef\]](#)
75. Busby, J.R. BIOCLIM—a bioclimate analysis and prediction system. *Plant. Prot. Q.* **1991**, *61*, 8–9.