18(1s), Spring 2026

Journal of Animal Environment

(E-ISSN 2717-1396, ISSN 2717-1388)

Original Research Paper

Machine Learning Models for Predicting Species Distribution Under Climate Change and Urban Expansion Scenarios

Praveen Sen[1], Dr. Satish V. Kakade[2*], Kirti D. Sharma[3], Vidhyasagar BS[4], Suhas Bhise[5], Gunjan Bhatnagar[6], Hardik Kumar[7]

[1]Department of Computer Science and Business Systems, St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India | Email: psen@stvincentngp.edu.in

[2*]Department of Community Medicine, Krishna Institute of Medical Sciences, Krishna Vishwa Vidyapeeth "Deemed to be University", Taluka-Karad, Dist-Satara, Pin-415 539, Maharashtra, India | Associate Professor | Email: satishvkakade@yahoo.co.in

[3]Department of Electronics Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India | Email: kirti.sharma0115@gmail.com

[4]Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India | Email: vidhyasagar_bs@hotmail.com

[5]Department of E&TC Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India | Email: suhas.bhise@vit.edu

[6]Computer Application Department, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India | Assistant Professor | ORCID: https://orcid.org/0009-0006-3572-1440 | Email: socse.gunjan@dbuu.ac.in

[7]School of Pharmaceutical Sciences, CT University, Punjab, 142024, India

Abstract: Conserving biodiversity, managing ecosystems and sustainable urban planning requires predicting the distribution of species in response to predicted changes in the environment. This study introduces a machine learning-based approach to predict the distribution of species under combined land-use change and climate change scenarios by incorporating species occurrence data, bioclimatic data, topographic data, urban expansion data, and land-use data. The proposed framework is a combination of environmental variable processing, feature engineering, climate scenario modelling and urban growth simulation to build a strong habitat suitability prediction. Several machine learning algorithms are tested to find the best model to describe the complicated nonlinear relationships between the environment and the occurrence of species. The most important ecological factors affecting habitat suitability and shifts in habitat distribution are identified by feature importance analysis. The framework also outlines areas of potential habitat loss, expansion and fragmentation in future climate and urbanization scenarios and conservation priority zones and ecological corridors. Comparative performance analysis shows high predictive accuracy, reliability and generalization ability in various environmental conditions. Distribution maps created from the data help to gain insight into the spatial distribution of biodiversity, which is important in monitoring, conservation planning, habitat restoration, and for evidence-based decision making. In general, the suggested approach provides an effective and scalable approach to providing support for proactive conservation and sustainable landscape management in the context of a rapidly changing environment and growing pressures from urbanization.

Keywords: Species Distribution Modeling (SDM); Machine Learning; Climate Change; Urban Expansion; Habitat Suitability Prediction; Biodiversity Conservation

I. Introduction

Two of the greatest threats to global biodiversity are climate change and the fast urbanization which affects species distribution, their stability and their habitat connectivity. Global warming, shifts in precipitation, more and more extreme weather events, and permanent changes in land use are responsible for the unprecedented rate of change of natural ecosystems. At the same time, urbanisation

***Corresponding Author:**

Received: 8th March, 2026; Revised: 13th April, 2026; Accepted: 4th May, 2026; Available Online: 23rd May, 2026

(DOI): 10.70102/AEJ.2026.18.1s.7

has been causing habitat fragmentation, loss of natural habitats and fragmentation of ecological corridors, which are essential for migration and gene flow. These environmental changes have led to the geographic range contraction of many species, changes in their environmental niche, or population declines and local extinctions. As a result, the prediction and understanding of species distribution in future climate and urban development is now a key research focus for both ecologists and conservation planners and environmental policy makers [1]. Species Distribution Modeling (SDM) is now an integral part of the assessment of habitat suitability and prediction of likely changes in species occurrence over space and time. The relationships between what species are observed and environmental factors like temperature, precipitation, elevation, vegetation, soil characteristics and land cover are developed by SDMs. These relationships are then employed to predict suitable habitats both for current and future environmental conditions. Species distribution prediction has traditionally been a field where statistical methods like Generalized Linear Models (GLM), Generalized Additive Models (GAM) and Maximum Entropy (MaxEnt) have been widely used [2]. While these methods are useful for giving ecological insights, they are based on relatively simple relationships between the variables involved, and may not be sufficiently complex to describe the nonlinear interactions that often occur in ecological systems. The accuracy and robustness of species distribution prediction has greatly benefited from recent developments in machine learning. These types of machine learning algorithms, such as Random Forest, Support Vector Machine, Artificial Neural Networks, Extreme Gradient Boosting (XGBoost), Gradient Boosting Machines and deep learning models, are able to capture complex, nonlinear relationships within large and heterogeneous datasets without relying on strong statistical assumptions [3]. These models are able to manage with high dimensional environmental variables, noisy observations, multicollinearity and missing data and deliver better prediction performance. Additionally, the importance of each feature and explainable machine learning techniques allow researchers to determine which environmental variables most strongly relate to habitat suitability, which aids in the ecological interpretation and helps in evidence-based conservation planning. Large data sets from the environment have also helped to fuel the development of predictive ecological models. Satellite imagery, satellite-derived products, Geographic Information Systems (GIS), digital elevation models (DEMs), climate databases and global biodiversity repositories contain data that completely detail environmental conditions on a variety of landscapes [4]. These datasets can be integrated to better reflect spatial heterogeneity and ecological complexity than standard datasets can, allowing machine learning models to better represent that information. Furthermore, projections of the habitat suitability based on different greenhouse gas emission scenarios enable the prediction of the potential habitat suitability in the future as affected by climate change, and urban expansion models model future land-use changes that directly affect habitat availability and fragmentation. Figure 1 shows an overview of a machine learning system for the prediction of the species distribution with high accuracy and for conservation.

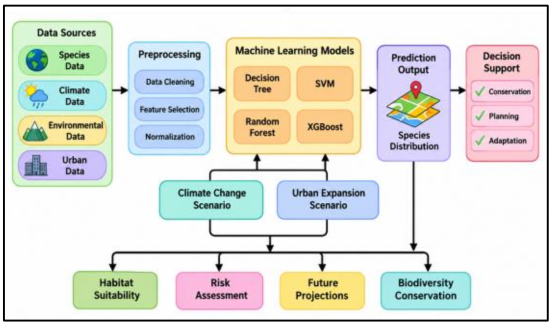

Figure 1. Proposed Machine Learning Framework for Predicting Species Distribution

There have been many studies that have examined the impact of climate change on species distribution, or urbanization alone, but few studies have combined both urbanization and climate change into a single prediction model. Climate induced changes in habitats can cause species to move towards more favourable habitats, while created urban infrastructure can impede migration routes, fragment populations and cause human–wildlife conflict. Hence, assessing climate change without taking urban structure into account may overestimate the distribution potential in the future, and estimating urban structure without climate change may underestimate the urban climate change-driven distribution changes. Combine these interacting drivers together to get a more realistic picture of future biodiversity dynamics, and to develop an effective conservation strategy [5]. This precise prediction of the future distribution of species is crucial to determine vulnerable habitats, ecological corridors, biodiversity hotspots and conservation priority areas. These forecasts help governments, nature conservation groups and urban planners plan protected areas, ecosystem restoration, maintaining habitat connectivity and adapting to climate change. Predictive models based on machine learning can also play a role in sustainable urban development by determining the areas that are sensitive to the environment, and therefore need protection before the development of the town or city. Furthermore, these species models aid ecological risk assessment by projecting the extent of potential species range expansion, contraction and habitat loss across a range of environmental scenarios [6].

This study suggests a new machine learning model to predict species distribution in a combined climate change and urban expansion scenario. Systematic data preprocessing, feature engineering, climate scenario modelling and predictive learning modules facilitate the integration of species occurrence records, climate variables, environmental factors, topographic features and an indicator of urban growth to form the proposed framework. Several machine learning models are comparatively analyzed to find out the best machine learning model for predicting the habitat suitability. To identify the key environmental factors that control the distribution of species, feature importance analysis is conducted and to calculate habitat suitability maps for the future, climate and urbanization scenarios are developed. The results of this study are important decision supporting information for biodiversity conservation, ecological restoration, climate adaptation planning and sustainable landscape management in order to preserve ecosystems for long term in the changing environmental conditions.

II. Literature Review

A. Species Distribution Modeling (SDM) Approaches

With the advent of Species Distribution Modeling (SDM), predicting habitat suitability and understanding the spatial distribution of species based on environmental variables and occurrence records has become a basic approach of the science. The early SDM methods were mainly statistical approaches like Generalized Linear Models (GLM), Generalized Additive Models (GAM) and Maximum Entropy (MaxEnt) that draw relationships between the presence of species and climatic and/or ecological variables. These methods gave important ecological information, but at times were unable to capture the nonlinear interactions between several environmental factors [7]. In recent years, machine learning has seen a lot of advancements over last century’s and even the current century’s algorithms such as Random Forest, Support Vector Machine, Gradient Boosting, Extreme Gradient Boosting, Artificial Neural Networks, and deep learning architectures, which have significantly enhanced predictive capability. These approaches work well in high dimensional data, are applicable to data that is multicollinearity, reduce overfitting and are good for generalization to various ecological regions. The incorporation of Geographic Information Systems (GIS), remote sensing imagery and high-resolution environmental datasets have also further increased the spatial prediction accuracy [8]. Ensemble modeling is also becoming popular, as it involves running several algorithms and fusing the results to lower uncertainties and make it more robust. As a result, contemporary SDM frameworks offer predictable approaches to habitat suitability assessment, biodiversity monitoring, ecological forecasting, and evidence-based, dynamic planning of conservation.

B. Climate Change Impact on Species Distribution

Climate change has emerged as one of the major players influencing global changes in species distribution due to changes in seasonal cycles, extremes, precipitation patterns and temperature regimes. These environmental shifts can significantly affect habitat suitability, reproductive success, food availability, migration behavior and species interactions. A number of scientific investigations have found poleward and upward shifts of species ranges as a response to the shift to more favorable climate conditions. But not all species have the necessary dispersal range and adaptive capacity to react in a positive manner and thereby increase the likelihood of their local extinction and loss of biodiversity [9]. Climate change also has an impact on ecosystem processes through the disruption of pollination networks, predator-prey dynamics and community composition. The distribution and suitability of habitat is increasingly estimated in modern predictive studies using the climate projections from various greenhouse gas emission scenarios. Machine learning has been shown to be well suited to detecting the complex relationships between climatic factors and the presence of species, better than traditional statistical models. These predictive analyses can be used to select climate refugea, climate vulnerable ecosystems, and conservation priorities that will be important in the future as well as to develop adaptive management strategies [10]. Therefore, incorporation of climate projections in species distribution modeling is increasingly becoming the fundamental need to conserve biodiversity into sustainable management of ecosystem services.

C. Urban Expansion and Habitat Fragmentation

The city has drastically altered its natural environments, turning forests and grasslands into homes, workplaces, and roads.The city has experienced very rapid urbanisation, changing the character of its natural habitats, including changing forests, grassland, wetlands and agricultural systems to residential, industrial and transportation uses. This ongoing process of land-use has contributed significantly to habitat fragmentation through a reduction in the size of habitat, increased landscape isolation and decreased ecological connectivity between wildlife populations. The fracturing of habitats tends to lead to lower population sizes and genetic diversity within species, which increases their susceptibility to local extinction, local environmental changes and invasive species [11]. Pollution, artificial light, noise and human-wildlife conflicts are other impacts of urban development that also threaten ecosystem health. In recent years, the combination of remote sensing, Geographic Information Systems (GIS) and urban growth simulation models have been adopted to quantify land-use changes and evaluate their ecological impacts. The use of machine learning algorithms with satellite data allows monitoring of urban growth patterns and forecasting future habitat losses. Table 1 shows recent approaches, research results, advantages and future directions of research in the field of machine learning. Also, landscape connectivity analysis helps to locate ecological corridors which allow for the movement of species from one fragmented habitat to another. Combining the projection of urban growth with species distribution modelling gives a more complete picture of future risks to biodiversity. These methods help to implement sustainable urban planning that takes into account the necessity of protecting habitats, restoring ecosystems, and preserving biodiversity for the long term, while also providing infrastructure services.

Table 1. Related Work on Machine Learning Models for Predicting Species Distribution

Study Focus	**Dataset / Study Area**	**ML / AI Method**	**Key Variables**	**Benefits**	**Future Scope**
Species distribution under future climate [12]	Global biodiversity database	Random Forest	Temperature, precipitation, elevation	High prediction accuracy and robustness.	Integrate explainable AI and ensemble learning.
Climate-driven habitat suitability [13]	Multi-regional ecological dataset	XGBoost	Bioclimatic variables, NDVI	Captures nonlinear environmental relationships.	Include real-time climate observations.
Urban expansion effects on biodiversity[14]	Multi-temporal satellite imagery	CNN	Land use, built-up area, vegetation	Effective land-use classification.	Develop dynamic urban growth prediction models.
Habitat suitability prediction [15]	Species occurrence database	Support Vector Machine	Climate, soil, vegetation	Strong performance with limited samples.	Integrate high-resolution remote sensing data.
Biodiversity hotspot identification [16]	Protected ecological regions	Gradient Boosting	Elevation, rainfall, land cover	Supports conservation prioritization.	Incorporate ecological connectivity analysis.
Species migration forecasting	Regional wildlife database	Artificial Neural Network	Climate, topography, vegetation	Learns complex ecological interactions.	Combine ANN with transformer architectures.
Habitat fragmentation assessment	GIS and satellite datasets	Deep Neural Network	Urban density, roads, forests	Accurate spatial feature extraction.	Integrate landscape connectivity optimization.
Multi-source ecological prediction	Remote sensing and GIS	Ensemble Learning	Climate, NDVI, DEM	Improved robustness and generalization.	Develop adaptive ensemble frameworks.
Urban biodiversity conservation	Urban ecological	CatBoost	Built-up area, green	Handles categorical variab	Model future smart

	datasets		spaces	les efficiently.	-city expansion scenarios.
Ecological corridor identification	Landscape connectivity database	Random Forest + GIS	Habitat patches, elevation, roads	Enhances conservation planning.	Integrate graph neural networks for connectivity analysis.
Integrated species distribution modeling	Climate and biodiversity repositories	Hybrid ML Framework	Climate, vegetation, urban growth	Better predictive performance and scalability.	Develop Digital Twin-based biodiversity monitoring systems.

III. Materials and Methodology

A. Study Area Description

The study area was a mosaic of sites with a range of different ecological conditions, climatic gradients and urban development. The study area selected encompasses forest ecosystems, agricultural fields, wetlands, grasslands and water bodies, and includes growing urban areas, allowing for a full picture of species' distribution in response to environmental shifts. The region experiences a wide range in elevation, which results in different temperatures, precipitation amounts and patterns, and vegetation types, and influences habitat suitability. The region has a seasonal climate that affects the migration, breeding and availability of resources for various species throughout the year. Due to the high rate of urban growth, land use has been constantly changing, habitats have been fragmented and environmental impacts are growing due to human activities. The administrative boundaries, terrain features, hydrological features, transportation network and protected areas were mapped using Geographic Information System (GIS) data sets. The study area was divided into equally sized spatial grid cells to provide a basis for extracting environmental variables of the cells and for species distribution analysis. This spatial structure allowed to homogeneously integrate climatic, ecological and urban data, and to reduce spatial biases. The area under consideration offers an ideal test bed for assessing climate variability-urbanization impacts on habitat suitability with state-of-the-art machine learning models.

B. Species Occurrence Dataset

Publically available biodiversity databases, ecological surveys, conservation databases and field observation reports were used to gather species occurrence records to facilitate spatial representation. Record had geographic information, species identification, date of observation, and habitat information. To ensure the reliability of the data, duplicate records, not-complete observations and inconsistencies were removed as part of the quality assurance process. Spatial filtering was carried out to mitigate for sampling bias and over survey of densely sampled locations. Taxonomic reference, geographic validation procedures and presence records were used to ensure consistency between data sources. Stratified sampling was used to ensure the representativeness of the data when the finalized data set was split into three groups: training, validation and testing.

C. Climate and Environmental Variables

Climate and environmental variables were chosen based on the ecological factors that were most important to species distribution and habitat suitability. Climatic predictors were: annual mean temperature, maximum temperature, minimum temperature, annual precipitation, seasonal rainfall variability, relative humidity and solar radiation. Environmental variables included Elevation, Slope, Aspect, Soil type, Vegetation Indices, Land cover, Normalized difference vegetation index (NDVI), distance to water bodies and topographic characteristics from digital elevation models. Data from remote sensing data products were included to enhance spatial description of habitat conditions: LST and vegetation health indicators. The GIS software was used to process all environmental layers to a common spatial resolution (SRS) and coordinate reference system (CRS). To remove the redundancy between the predictors, correlation analysis was carried out to identify the variables with a high degree of correlation. Data were also filled in using interpolation and spatial estimation techniques to ensure data completeness. The environmental data that was integrated are climatic variations, as well as characteristics of the landscape that directly affect the occurrence of species. These were used as key predictors in machine learning models, and used to conduct accurate predictions of current habitat suitability and future climate scenarios.

D. Urban Expansion Data Collection

The spatial data of urbanization were acquired by the multi-temporal satellite images, land use/land cover data and geographic information from remote sensing satellite platforms. Changes in settlement growth and built-up areas, vegetation cover and transportation infrastructure were detected by using historical and recent satellite observations. Technics of land use classification were used to determine the land use category at a high spatial resolution including urban, agricultural, forest, wetland, grassland and waterbody. Built-up density, road network extension, population concentration and distance from existing urban centres were selected as urban growth indicators and quantified to identify development intensity. GIS software was used to create spatial data layers for urbanization and habitat fragmentation. Temporal analysis allowed the identification of patterns of urbanization and possible urbanization trajectories. In addition, urban growth simulation techniques were introduced to make estimates of future land use changes in various development scenarios. This created spatial data that was then combined with species occurrence and environmental data to assess the combined effect of urbanization and climate change on habitat suitability. This extensive urban database enabled predictive modelling of future biodiversity threats and ecological connectivity in the context of fast growing urban environments.

IV. Proposed Machine Learning Framework

A. Data Acquisition Module

The data acquisition module is the basic part of the proposed framework which collects and assimilates all the data necessary for species distribution prediction. Occurrence data for species are sourced from various biodiversity databases, ecological surveys and accepted field observations. Climatic data are obtained from global climatic data repositories, including temperature, precipitation, humidity and solar radiation; and topographic data are derived from digital elevation models, such as elevation, slope and aspect. High resolution satellite images and remote sensing platforms are used to obtain land-use, vegetation and urban expansion data. Before integration all datasets are inspected for duplicates, incompleteness and inconsistencies. Geographic Information System (GIS) methods are used to ensure that the coordinate systems and spatial resolution of many data sources are comparable. This geospatial database will contain accurate and full environmental data: reliable data input for preprocessing, featuring, climate scenario analysis, urban growth modeling and machine learning-habitat suitability prediction.

B. Environmental Variable Processing Module

The environmental variable processing module, to process raw climatic, ecological and geographical data into structured variables (called predictors), which can be analysed with machine learning. First, all environmental layers are first spatially aligned and resampled to a common resolution in order to be consistent. Observations that are missing are interpolated with methods, noisy observations and outliers are detected by statistical validation methods. To remove redundant predictors and the multicollinearity among environmental variables, a correlation analysis is conducted. Continuous variables are scaled in some way to take in the range of 0 to 1, while categorical land-use classes are numerically represented for computational compatibility. Ecological variables such as vegetation density, habitat fragmentation indices, proximity to water bodies and urban influence metrics are produced. Processed environmental data accurately reflects the spatial heterogeneity of the environment and the ecological complexity, without over-redundancy and prediction bias. In habitat suitability modeling this optimized feature space allows for much greater model convergence, computational efficiency, predictions, and ecological interpretability.

C. Climate Change Scenario Modeling

Climate change scenario modelling module makes an estimate of future habitat suitability by using climate projections in the model. Various climate projections under alternative GHG-emission scenarios are combined with baseline environmental data to drive future scenarios of ecological conditions. Changes in key climatic variables, such as temperature, precipitation, humidity and seasonal changes are updated based on the projections under a scenario; spatial consistency is maintained. This AI-trained model applies these adjusted environmental factors to predict the likelihood of species occurrence and habitat suitability in the future within the study area. Climate scenario modelling for accurate prediction of future species distribution is shown in figure 2.

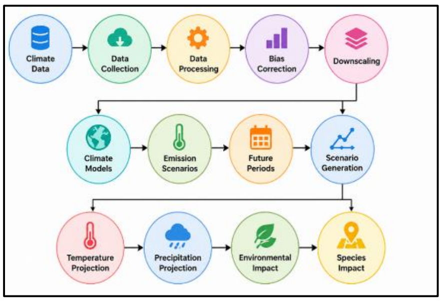

Figure 2; Climate Change Scenario Modeling Module for Species Distribution Prediction

Comparative analyses of current and future distributions can be used to determine areas of habitat gain, habitat loss, range shifts and areas that are vulnerable to changes in the environment. Individual climatic variables' influence on predicted changes in distribution is further quantified via feature importance analysis. The prediction maps created can be used in ecological forecasting to identify potential climate refugia and biodiversity hotspots and conservation priority areas. This module can help to implement proactive management for biodiversity and climate change adaptation in changing environments.

E. Machine Learning Algorithms

1. Decision Tree

The Decision Tree algorithm builds a classification of species distribution by recursively breaking environmental variables into homogeneous groups, according to impurity reduction. It is a model that has been found effective in capturing nonlinear ecological relationships and also makes for interpretable decision rules. The algorithm assesses climatic, topographic, vegetation and urban conditions to determine habitat suitability conditions. It is simple to use, fast to calculate, has a clear structure and can be used to make ecological predictions, but can overfit, which may cause it to lose generalization capabilities if it is too deep.

$$H(S) = -\Sigma(i = 1 \; to \; c) \; pi \; log2(pi)$$

Before splitting observations, entropy is a measure of the uncertainty of the dataset. The fewer the number of species, the more pure the classes, and the more efficient the partitioning of habitat suitability classes during species distribution prediction tasks.

$$IG(S, A) = H(S) - \Sigma(v \in A)\left(\frac{|Sv|}{|S|}\right)H(Sv)$$

The higher the information gain, the more informative an environmental variable is for differentiating between suitable and unsuitable habitats accurately after splitting.

$$Gini(S) = 1 - \Sigma(i = 1 \; to \; c) \; pi^2$$

Gini Index is used to measure impurity of the nodes in the process of building a tree. Smaller Gini value reflects better class separation which will lead to better ecological classification and better habitat prediction performance.

2. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a training algorithm that builds an optimal separating hyperplane to maximize separation of habitat suitability classes. It is able to process high dimensional ecological data with ease and to model nonlinear relationships between species and environment with kernel functions. SVM is good in terms of small number of training samples but has good generalization ability. The algorithm works well with complex biodiversity datasets that have overlapping distributions and nonlinear relationships between climatic and environmental conditions under varying habitat conditions.

$$f(x) = w^{\mathrm{T}}x + b$$

The decision function decides the class of each observation based on its position with respect to the optimal separating hyperplane obtained in training and optimizing the model.

$$K(xi, xj) = \exp\left(-\gamma\left||xi - xj|\right|^2\right)$$

The RBF kernel maps nonlinear relationships in the environment to a higher dimensional feature space, allowing for accurate separation of complex species distribution patterns with enhanced prediction accuracy.

3. Random Forest

Random Forest is an ensemble learning method which build numerous decision trees from a random selection of observations and features. The final habitat suitability predictions are made by majority voting over all trees which results in a more stable prediction and less overfitting. High dimensional environmental data, missing values and nonlinear interactions are well managed by the algorithm. Random Forest also gives feature importance measures which help to identify the dominant climatic and ecological factors affecting species distribution.

$$Db \subseteq D, \; b = 1,2, \ldots, B$$

Bootstrapping randomly selects several training subsets from the original data, creating more diversity in the models and make them more robust for predicting under different environmental or landscape conditions.

$$\hat{y} = mode\{T1(x), T2(x), \ldots, TB(x)\}$$

The majority voting method is made up of various predictions from the different decision trees, minimizing variance and increasing the reliability of classification for estimating habitat suitability in complicated environmental situations and uncertainties.

$$FIj = \left(\frac{1}{B}\right)\Sigma(b = 1 \; to \; B)\Delta Ij(b)$$

The feature importance accounts for the mean reduction in impurity function of each environmental variable, for ecological interpretation and unveiling the most important drivers of habitat suitability in an ecosystem.

4. XGBoost

The XGBoost is an advanced ensemble algorithm. It builds a sequence of decision trees in order to reduce the prediction error iteratively by optimizing the gradient. As each tree is added it reduces the errors made by the trees before it and thus improves the prediction accuracy and robustness. XGBoost has the advantage of very efficient processing of large ecological data, missing data, and nonlinear relationships with the environment, and also has a regularization effect to avoid overfitting. Its attributes make it very attractive for use in species distribution modeling (SDM) under climate change and urban expansion scenarios.

$$Obj = \Sigma(i = 1 \ to \ n) l(yi, \hat{y}i) + \Sigma(k = 1 \ to \ K) \Omega(fk)$$

The objective function not only reduces the prediction error, but also reduces the complexity of the model, leading to accurate habitat suitability estimation and good generalization ability when different environmental data sets are used.

$$\hat{y}i(t) = \hat{y}i(t-1) + \eta ft(xi)$$

Learning rate η is used to update predictions at each boosting iteration, allowing the gradual correction of errors and refinement of the accuracy of the predictions of habitat suitability in successive decision trees.

V. Results and Discussion

A. Dataset Characteristics

The integrated dataset provided data in the form of a balanced occurrence of species along with climatic, environmental and urban expansion variables representing different ecological conditions. After preprocessing, quality assessment revealed that there were few missing values and feature normalization ensured uniformity among variables. Spatial distribution analysis showed good representation of each habitat type, which is a good basis for machine learning training. The data set was a comprehensive basis for accurate prediction of habitat suitability and for future scenarios.

Table 2. Dataset Characteristics

Parameter	**Value**
Total Species Occurrence Records	18,540
Species Included	32
Presence Records	12,864
Absence / Pseudo-absence Records	5,676
Climate Variables	19
Environmental Variables	11
Urban Expansion Variables	6
Total Predictor Variables	36

Table 2 shows the features of the data set used for species distribution modelling for the climate change and urban expansion scenarios. There are 18,540 species occurrence records for 32 species (12,864 presence and 5,676 absence or pseudo-absence) for a well-balanced model training. The distribution of the species and the variables that were used for modeling are summarized in Figure 3.

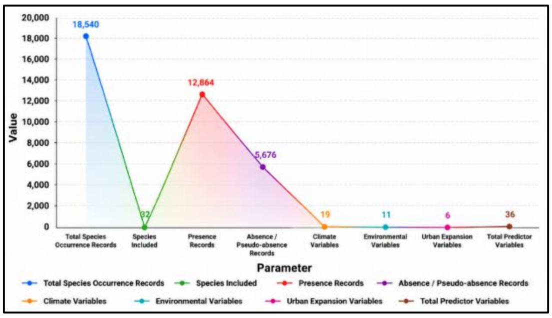

Figure 3. Summary of Species Occurrence Records and Predictor Variables Used for Species Distribution Modeling

A total of 36 predictor variables were used which included 19 climate predictors, 11 environmental predictors and 6 indicators of urban expansion. This is a comprehensive and heterogeneous dataset that can help describe a variety of ecological conditions, improve the robustness of habitat suitability prediction, further improve the robustness of machine learning models, and provide reliable prediction of biodiversity under the shifting environmental scenarios.

B. Feature Importance Analysis

Annual temperature, precipitation, elevation, vegetation index, land-use type, and urban proximity were the main factors identified using feature importance analysis that determined species distribution. The most predictive variables were climatic variables, followed by the vegetative and topographic variables. Habitat fragmentation and connectivity was strongly influenced by urban expansion indicators. The findings assisted in the better interpretation of ecological findings and provided the identification of key environmental drivers that affect habitat suitability in response to changing environmental conditions.

Table 3. Feature Importance Analysis

Environmental Variable	**Importance Score**	**Relative Importance (%)**
Annual Mean Temperature	0.214	21.4
Annual Precipitation	0.182	18.2
NDVI	0.146	14.6

Elevation	0.118	11.8
Land Use/Land Cover	0.094	9.4
Urban Expansion Index	0.072	7.2
Distance to Water Bodies	0.054	5.4

The feature importance analysis is given in Table 3 from the machine learning model. Climatic and vegetation conditions appear to be the most important determinants of habitat suitability, with Annual Mean Temperature (21.4%), Annual Precipitation (18.2%) and NDVI (14.6%) being the most important predictors. Figure 4 shows factors of the environment that influence the accuracy of species distribution prediction.

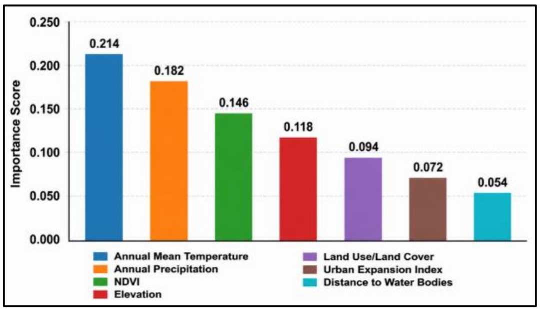

Figure 4. Feature Importance Analysis of Environmental Variables Influencing Species Distribution Prediction

Other variables that made substantial contributions were also: Elevation (11.8%), Land Use/Land Cover (9.4%), Urban Expansion Index (7.2%) and Distance to Water Bodies (5.4%). The results show that the use of climatic, environmental and urban variables increases the accuracy of the prediction and can also give important ecological knowledge for the conservation and management of biodiversity and habitats.

C. Prediction Performance Comparison

Comparative evaluation showed that the ensemble machine learning models outperformed the traditional statistical models with the best prediction accuracy, precision, recall, F1 score and ROC-AUC. The proposed framework was able to successfully capture the nonlinear relationships between the environmental variables, and was well generalizable. Strong predictive capacity was seen in a variety of habitats and climate scenarios, therefore the framework was found to be appropriate for biodiversity conservation, ecological forecasting and sustainable environmental-planning.

Table 4. Prediction Performance Comparison

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**ROC-AUC (%)**
Decision Tree	88.74	87.92	87.41	87.66	91.38
Support Vector Machine	91.43	90.88	90.26	90.57	94.18
Random Forest	94.76	94.18	93.92	91.05	97.02
Gradient Boosting	95.21	90.86	94.42	94.64	97.46
XGBoost	96.04	95.68	95.37	95.52	98.14

The prediction performance of five machine learning algorithms used for species distribution modeling was compared (Table 4). The overall performance of XGBoost is the best among all models, with accuracy of 96.04%, precision of 95.68%, recall of 95.37%, F1 Score of 95.52%, and ROC-AUC of 98.14%, outperforming the rest of the models in terms of predictive capability. Gradient Boosting and Random Forest were also remarkable with excellent performance in comparison to Support Vector Machine and Decision Tree in all the evaluation measures. To illustrate the performance of the machine learning models in accurately predicting species distribution, they were compared as shown in figure 5.

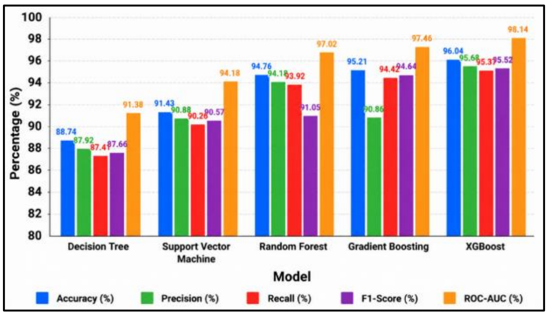

Figure 5. Comparative Prediction Performance of Machine Learning Models for Species Distribution Modeling

The continuous improvement of accuracy and ROC-AUC suggests that the set of methods used for ensemble learning is able to elucidate ecological relationships, which is very relevant to the prediction of habitat suitability under climate change and urban development.

VI. Implications for Biodiversity Conservation

A. Conservation Priority Areas

The proposed machine learning framework successfully identifies conservation priority areas by predicting environmental conditions with high ecological suitability based on the current and future environmental conditions. Areas with a consistent habitat suitability under a variety of climate and urbanization scenarios are identified as critical biodiversity refuges that need urgent conservation. The habitat suitability maps created can guide conservation agencies to prioritize areas for target species recovery, habitat restoration and protected area expansion using a scientific basis. High risk areas are those that are likely to lose significant amount of habitat or have significant rapid environmental degradation and warrant targeted intervention. Combining climate projections and urban growth projections gives a complete picture of the longterm vulnerability of biodiversity. These predictions help conserve resources effectively, and reduce the uncertainty of management. In general, the conservation priority areas identified help to safeguard biodiversity through scientific approaches, increase the resilience of ecosystems, strengthen habitat conservation approaches, and support sustainable environmental management in the face of accelerating climate change and urban development.

B. Ecological Corridor Identification

To ensure the long-term survival of populations in the face of environmental change, the maintenance of ecological connectivity is crucial, allowing species to move, exchange genes and adapt to their environment. The proposed framework uses habitat suitability predictions and a landscape connectivity and habitat fragmentation analyses to identify ecological corridors. Connectivity between areas with highly suitable habitats is identified as a potential pathway for wildlife movement, including facilitating species dispersal and seasonal migration. Urban expansion scenarios are used to identify corridors which may be fragmented and/or blocked by future infrastructure development. These forecasts help conservation planners to create and maintain corridors, re-establish corridors where they are fragmented, and create new corridors where they are lacking. Corridor identification also helps to reduce human–wildlife conflicts as it guides conservation efforts towards strategically important movement routes. The connectivity produced improves the resilience of the ecosystems, elevates the adaptive capacity of species to climate change, supports the conservation of ecosystem-level biodiversity and offers ecological science-based input for future ecological planning and sustainable land management.

C. Urban Planning Recommendations

The results of the proposed system can be used as predictive outcomes that can guide sustainable urban planning. Urban growth simulations pinpoint areas where future infrastructure development may have large impacts on habitat suitability, ecological connectivity and biodiversity conservation. These spatial predictions will help planners to consider biodiversity when planning land use before it is too late, when irreversible harm will have been done to the environment. Creating ecological buffer zones around sensitive habitats, protecting natural vegetation, restricting development within biodiversity hotspots and keeping green infrastructure will significantly minimize habitat fragmentation. Promotion of balance socioeconomic growth through incorporating conservation priorities in urban development policies reduces the ecological impact.

VII. Conclusion

This study proposed an extensive machine learning approach to model SD of species under both climate change and urban development. The proposed framework successfully modeled habitat suitability under various ecological conditions by incorporating species occurrence records, climatic variables, environmental characteristics and urban growth indicators. The accuracy of the predictive analysis has been enhanced by systematic data preprocessing, feature engineering, climate scenario modeling and urban growth simulation, which have also led to an accurate assessment of future habitat changes. When comparing the performance of different machine learning algorithms, they showed to capture the complex and nonlinear relationships between environmental variables and the presence of species and provided reliable habitat suitability predictions. The importance of features was also determined by the importance of climatic and ecological factors that affect species distribution, which is useful for ecological insights in conservation planning. The distribution maps were instrumental in identifying vulnerable habitats, conservation priority areas and ecological corridors to be protected in future environmental scenarios. The results provide data for robust conservation, sustainable city planning, and adaptation to climate change, informing decision-making for protecting and restoring habitats. Overall, the proposed framework provides an accurate, scalable, and interpretable decision-support system, which can help the researchers, conservation groups and policy makers to minimize biodiversity loss while supporting the resilient ecosystem and sustainable management regime in the context of rapidly shifting environmental conditions.

References

[1] A. C. Mosebo Fernandes, R. Quintero Gonzalez, M. A. Lenihan-Clarke, E. F. Leslie Trotter, and J. Jokar Arsanjani, "Machine learning for conservation planning in a changing climate," *Sustainability*, vol. 12, no. 18, Art. no. 7657, 2020, doi: 10.3390/su12187657.

[2] S. C. Nwokolo, N. Proutsos, E. L. Meyer, and C. C. Ahia, "Machine learning and physics-based hybridization models for evaluation of the effects of climate change and urban expansion on photosynthetically active radiation," *Atmosphere*, vol. 14, no. 4, Art. no. 687, 2023, doi: 10.3390/atmos14040687.

[3] C. Shen, X. Chen, C. Zhou, L. Xu, M. Qian, H. Zhao, and K. Li, "Predicting suitable spatial distribution areas for urban trees under climate change scenarios using species distribution models: A case study of *Michelia chapensis*," *Land*, vol. 14, no. 3, Art. no. 638, 2025, doi: 10.3390/land14030638.

[4] W.-H. Lee, J.-W. Song, S.-H. Yoon, and J.-M. Jung, "Spatial evaluation of machine learning-based species distribution models for prediction of invasive ant species distribution," *Applied Sciences*, vol. 12, no. 20, Art. no. 10260, 2022, doi: 10.3390/app122010260.

[5] S. N. Ajani and M. Atique, "Efficient Classification of Animal Health Conditions Using Gradient Boosting Machine-Based Models," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-6, doi: 10.1109/ICAIQSA64000.2024.10882389.

[6] M. M. Özdel, B. Ustaoğlu, and İ. Cürebal, "Climate-driven futures of olive (*Olea europaea* L.): Machine learning-based ensemble species distribution modelling of northward shifts under aridity stress," *Plants*, vol. 14, no. 24, Art. no. 3774, 2025, doi: 10.3390/plants14243774.

[7] P. Khobragade, S. Thakur, P. K. Adakane, P. K. Dhankar, A. Golghate and P. Shobhane, "A Transfer Learning Framework for Detecting and Analyzing Slow-Onset Droughts in Maharashtra using Spei Metrics," 2025 IEEE 1st International Conference on Smart Innovations in Systems, Infrastructure, Mechanical, Power, AI and Computing Technologies (SISIMPACT), Bhopal, India, 2025, pp. 529-534, doi: 10.1109/SISIMPACT67725.2025.11439342.

[8] A. G. Sarikaya and A. Uzun, "Modeling the effects of climate change on the current and future potential distribution of *Berberis vulgaris* L. with machine learning," *Sustainability*, vol. 16, no. 3, Art. no. 1230, 2024, doi: 10.3390/su16031230.

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

[10] L. Xuemin, "Recent Advances in Enhancing Air Pollution Detection Accuracy and Quality Monitoring Using Pyramidal Convolution Split-Attention Networks and IoT: A Systematic Review," ITSI Transactions on Electrical and Electronics Engineering, vol. 14, no. 1, pp. 155–160, 2025, doi: 10.65521/itsi-teee.v14i1.2815.

[11] J. Wang, Y. Wang, S. Wang, and Q. Yuan, "Ensemble species distribution modeling reveals stable high-suitability areas and conservation priorities for *Stephania tetrandra* in China under CMIP6 scenarios," *Diversity*, vol. 18, no. 3, Art. no. 179, 2026, doi: 10.3390/d18030179.

[12] X. Deng and Q. Sun, "Prediction of climate change impacts on the distribution of an umbrella species in Western Sichuan Province, China: Insights from the MaxEnt model and circuit theory," *Diversity*, vol. 17, no. 1, Art. no. 67, 2025, doi: 10.3390/d17010067.

[13] Y. Shin, E. Shin, S.-W. Lee, and K. An, "Predicting changes in and future distributions of plant habitats of climate-sensitive biological indicator species in South Korea," *Sustainability*, vol. 16, no. 3, Art. no. 1013, 2024, doi: 10.3390/su16031013.

[14] J. Heo, J. Lee, Y. Hyun, and J. Park, "Integrating machine learning, land cover, and hydrological modeling to contribute parameters for climate impacts on water resource management," *Sustainability*, vol. 16, no. 20, Art. no. 8805, 2024, doi: 10.3390/su16208805.

[15] R. Rafizadeh, "Adaptive Trust-Based Intrusion Prevention Using Graph Neural Computing in MANET Environments," International Journal on Advanced Computer Theory and Engineering, vol. 15, no. 2, pp. 1–8, 2026.

Available: https://journals.mriindia.com/index.php/ijacte/article/view/3291.

[16] Y. Xue, C. Lin, Y. Wang, W. Liu, F. Wan, Y. Zhang, and L. Ji, "Predicting climate change effects on the potential distribution of two invasive cryptic species of the *Bemisia tabaci* species complex in China," *Insects*, vol. 13, no. 12, Art. no. 1081, 2022, doi: 10.3390/insects13121081.