Modelling habitat suitability for the leopard in southern India through Ensemble Species Distribution Modelling

DAKSH SINGH

June 2023

SUPERVISORS:

Assoc. Prof. Dr. Tiejun Wang

Drs. Raymond Nijmeijer

Modelling habitat suitability for the leopard in southern India through Ensemble Species Distribution Modelling

DAKSH SINGH
Enschede, The Netherlands, June 2023

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Natural Resources Management

SUPERVISORS:
Assoc. Prof. Dr. Tiejun Wang
Drs. Raymond Nijmeijer

THESIS ASSESSMENT BOARD:
Assoc. Prof. Dr. Thomas A. Groen (Chair, University of Twente)
Dr. Shadrack Ngene (External Examiner, Wildlife Research and Training Institute, Kenya)
Assoc. Prof. Dr. Tiejun Wang (First supervisor, University of Twente)
Drs. Raymond Nijmeijer (Second supervisor, University of Twente)

DISCLAIMER

This document describes work undertaken as part of a program of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author and do not necessarily represent those of the faculty.

ABSTRACT

In the southern Indian provinces of Tamil Nadu and Kerala, anthropogenic influences and habitat fragmentation have exerted significant impacts on the Indian leopard's environment within the biodiverse Western Ghats. As a globally recognized UNESCO World Heritage site, the Western Ghats' diverse ecosystems and rich biodiversity are integral to the survival of these leopards. This study analyses the implications of integrating anthropogenic variables with environmental variables, assessing their cumulative impact on the precision of Species Distribution Models (SDMs). Four distinct SDMs utilizing techniques Maximum Entropy (Maxent), Random Forests, Generalized Linear Models (GLM), and Boosted Regression Trees (BRT) were employed. Initial model execution solely incorporated environmental variables, and the outputs were incorporated into a comprehensive ensemble model, with recorded accuracy measures. In subsequent runs, anthropogenic variables were introduced, enhancing the integrated model's overall accuracy, thereby underscoring their pivotal role in forecasting habitat suitability. Moreover, this investigation estimates the viable habitat area for leopards outside the prescribed protected zones within Tamil Nadu and Kerala. An exploration into the influence of dataset selection on leopard habitat suitability modelling was also undertaken, utilizing two divergent datasets: the Global Biodiversity Information Facility (GBIF) and the official Indian government report on leopard status. A significant discrepancy was noted when each dataset was applied individually, resulting in the deployment of a combined dataset for the final analysis. The ensemble model predicts a total suitable habitat area of 21,797 square kilometres within these southern Indian states. From this, 7,426.015 square kilometres are within protected areas, leaving 14,370.985 square kilometres of appropriate habitat situated outside these zones. It also provides a district-wise breakdown of the predicted leopard habitat and protected areas in both states. Representing the pioneering application of machine learning techniques in Tamil Nadu and Kerala for predicting suitable habitats for leopards, this research significantly contributes to the region's conservation initiatives.

Keywords: Leopards, Species Distribution Modelling, Ensemble Model, Habitat Suitability, Human-Wildlife Conflict, Western Ghats, Tamil Nadu, Kerala

ACKNOWLEDGEMENTS

From a young age, I found myself enchanted by the world of wildlife conservation, a fascination largely ignited by my idol, Steve Irwin. His dynamic presence on television not only stirred a deep-seated interest in the magnificent creatures that inhabit our Earth but also instilled a lifelong commitment to their protection and preservation.

As a native of India, I have always been enthralled by the extraordinary wildlife the country hosts, ranging from the royal Bengal Tiger and stealthy leopards to mighty Gaurs and majestic Indian elephants. In essence, India encapsulates the entire cast of 'The Jungle Book,' and this diverse and rich wildlife of my homeland has always held me captive. Drawing inspiration from Steve Irwin, my childhood dream was to contribute to the conservation of these unique and wonderful creatures in their natural habitats. The opportunity to realize this dream was granted to me by ITC, where I have been able to work toward wildlife conservation in earnest. My expertise in species distribution modelling has been greatly enhanced by the Natural Resources Management (NRM) program. I am profoundly grateful to Dr. Tiejun Wang and Dr. Raymond Nijmeijer for their unwavering belief in me and their invaluable guidance throughout this two-year journey. I extend my deepest respect and gratitude to the entire faculty of the Department of Natural Resources. Their dedication to shaping my thought processes and analytical skills has been instrumental in my growth. Special mention goes to Dr. Thomas A. Groen, who made the intricate concepts of species distribution models accessible to me, thus opening the door to the intersections of machine learning and wildlife conservation. This knowledge, I believe, will prove to be an invaluable asset in my conservation efforts. A heartfelt thanks to my NRM colleagues who challenged me and contributed to my personal and professional development, helping me become the best version of myself.

I am sincerely thankful for my father, whose unwavering belief in my abilities has been a driving force in my journey. I'm proud to have fulfilled his expectation of completing my MSc from ITC, Europe's most prestigious institution in GIS. Lastly, but by no means least, I dedicate this thesis to my mother. Her relentless motivation and support have shaped me just as a leopard in the wild nurtures her cubs. Her influence has been instrumental in my journey, and for that, I am eternally grateful.

Table of contents –

LIST OF FIGURES

LIST OF TABELS

LIST OF ABBREVIATIONS

BRT	Boosted Regression Trees
GIS	Geographic Information System
IUCN	International Union for Conservation of Nature
Maxent	Maximum Entropy Model
NASA	National Aeronautics and Space Administration
RF	Random Forest
SDM	Species Distribution Model
GLM	Generalized Linear Model
USDM	Uncertainty Analysis for Species Distribution Models
SECR	Spatially Explicit Capture-Recapture
ENM	Ecological Niche Modelling
TSS	True Skill Statistics
AUC	Area Under the Curve
NDVI	Normalised Vegetation Index
UNESCO	United Nations Educational, Scientific and Cultural Organization
GBIF	Global Biodiversity Information Facility
VIF	Variance Inflation Factor

1. Introduction

1.1 Background

One of the most diverse families of mammals, carnivores (order Carnivora), has managed to inhabit all continents and a wide range of habitats, including deserts, tropical forests, savannas, rivers, and oceans (Eizirik et al., 2010). Globally, large carnivores are acknowledged as flagship species for conservation initiatives (Macdonald et al., 2015). As top predators, they control primary consumers (herbivores) both directly and indirectly, which has a cascading effect on the ecosystem (Carter & Linnell, 2016).

Large carnivores provide various ecological services like biodiversity improvement, native plant diversity restoration, disease control, and carbon storage to mitigate climate change, also reducing the amount of their herbivore prey; large predators may improve carbon storage in particular environments, allowing plants to thrive (Ripple et al., 2014). Therefore, conserving carnivores globally can help restore declining forests by aiding in carbon storage, especially in dense tropical forests, where plant biomass declines after removing the large predator from those forests (Terborgh et al., 2001).

However, because of their predatory habits, many of these species have had severe losses in their populations worldwide. They may be more vulnerable to anthropogenic threats than other animal species (Fernández-Sepúlveda & Martín, 2022). Over the past 200 years, the majority have seen significant population decreases and range reductions over the globe (Ripple et al., 2014). These carnivores frequently require giant prey and broad habitats due to endothermy's high metabolic demands and large body size (Ripple et al., 2014). They often clash with people and animals due to their varied temperaments and feeding needs. This makes them vulnerable to extinction, in addition to human intolerance.

Even though protected zones have been constructed worldwide to save endangered species, large carnivores are often seen occupying human-dominated areas outside protected reserve limits to fulfil their food needs (Naha et al., 2021). A significant fraction of the surviving geographic range of carnivores globally is represented by such shared landscapes (Carter & Linnell, 2016). Most threatened, and declining carnivore species are distributed in East and South Asia's tropical forest, shrubland, and grassland habitat. The primary dangers to the Carnivora order are hunting and trapping of terrestrial species and habitat degradation due to deforestation and agricultural expansion (Fernández-Sepúlveda & Martín, 2022).

As part of the Asia Pacific, India has a diverse climate, varied topography, at least ten unique bio-geographical areas, sustains a wide range of forest types, and is home to three hotspots for terrestrial biodiversity worldwide (Kumar & Verma, 2020). As other terrestrial habitats have lost their natural status, the forest currently holds most of the terrestrial biodiversity (Kumar & Verma, 2020). An outstanding protected area network includes 99 national parks (18 biosphere reserves), 514 wildlife sanctuaries, and many sacred groves preserved by indigenous tribes (Kumar & Verma, 2020). Nevertheless, despite a supportive forest policy and a comprehensive regulatory framework, the ever-increasing human population's growing needs, land use changes, and the spread of foreign invasive species all contribute to the ongoing degradation of forests and loss of biodiversity.

Despite having many national parks, wildlife sanctuaries, and protected areas, India still struggles with issues like habitat destruction, human-wildlife conflicts, illegal poaching, etc. The Indian leopard (*Panthera pardus fusca*) is one large carnivore species impacted by these activities. The Indian subspecies of leopard is distributed in all the forested habitats of the country; it is absent only in deserts and the upper Himalayas (Jhala et al., 2018). According to information provided by the Wildlife Protection Society of India, at least one leopard dies each day in India. Leopards are widely distributed across India, with at least 13000 in number, stated in the leopard report, 2018. Compared to other carnivores, leopards are better able to survive in increasingly human-dominated areas, mainly because they have highly adaptable behaviour and protection provided by the Indian government (Jhala et al., 2018). In most of India's forested areas, the Indian leopard serves as the top predator in addition to the tiger and lion (Jhala et al., 2018). Although leopards are widely distributed throughout the country, their habitat is becoming increasingly fragmented. This loss of suitable habitat and wild prey causes leopards to venture into areas with a high human presence in search of food, which leads to human-leopard conflict (Jhala et al., 2018). The surveys conducted across four major tiger conservation landscapes in India estimated the abundance of the Indian leopard in 1) Shivalik Hills and Gangetic plains, 2) Central India and the Eastern Ghats, 3) the Western Ghats, and 4) North-eastern Hills and Brahmaputra Flood Plains stated by Status of leopard report, 2018 as shown in Figure 1.

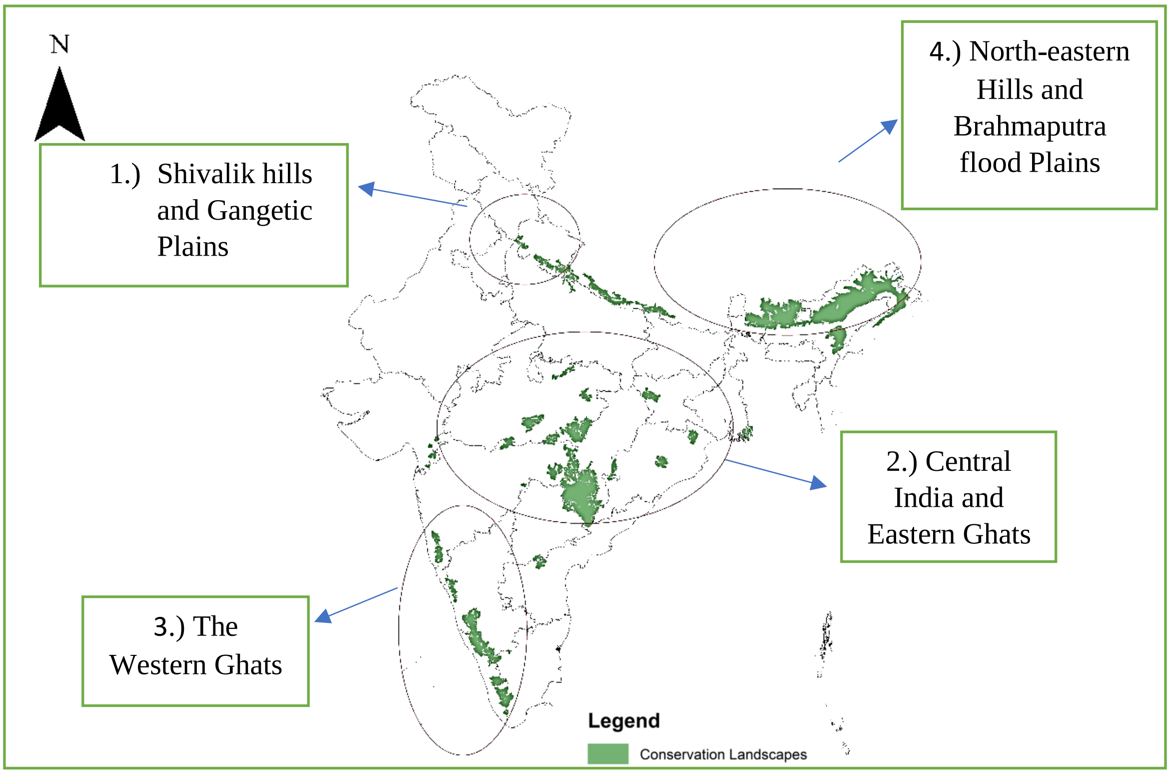

Figure 1: The figure depicts the 4-four tiger conservation landscapes across India where the survey for the Indian leopard abundance was conducted.

The southern states of India have been enriched with the Western Ghat's exceptionally high biological diversity and endemism which made this biodiversity paradise a part of the world heritage site by the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2021). This mountain range is one of the eight "hottest hotspots" for biological diversity in the world (UNESCO, 2021). Some of the world's best non-equatorial tropical evergreen forest specimens can be found in the Western Ghats. The Ghats are home to at least 323 species that are included on the IUCN Red Data List as threatened. They are home to 229 plant species, 31 animal species, 15 bird species, 43 amphibian species, and 5 reptile species, all of which are internationally threatened. Of the 325 species that are threatened by extinction globally and found in the Western Ghats, 129 are considered vulnerable, 145 are endangered, and 51 are critically endangered. Therefore, leopards can be an umbrella species for biodiversity conservation in areas devoid of other large carnivores. Studies on leopards have been quite scarce despite their ecological importance and the problems they confront; the literature review suggests that the study of human-leopard conflict and leopard habitat suitability is very limited in the southern states of Tamil Nadu and Kerala. Most studies conducted in India are from the northern part, including human-leopard conflict, species, and prey distribution. Hence, the conservation of leopards is essential to maintain an ecological balance in the forested landscape of south India.

To understand the distribution of Indian leopards in the forested landscape of southern India, Ecological niche models can be used for modelling the habitat suitability of leopards as these models are popular in ecology and used globally to address fundamental questions like where a species is likely to be found, what factors are involved in the distribution of a species, and what challenges climate change imposes on different species. Ecological niche models are also known as species distribution models. They are extensively used to model habitat suitability and understand the distribution of other species in broad environments concerning space and time. Ecological niche models are also used to build prediction maps by relating the occurrence of a species to the corresponding environmental variables (Franklin, 2012). It was first used in the 1920s to assess the role of climatic determinants in species distribution with the help of predictive output maps (Guisan & Thuiller, 2005). The advancement in data science has resulted in the development of complex machine learning algorithms, which are being integrated and used in these niche models to develop more accurate maps and provide advanced decision-making for the conservation of endangered species globally (Elith et al., 2006; Kindt, 2018; Mi et al., 2017; Woodman et al., 2019).

1.2 Problem Statement

Leveraging their remarkable adaptability towards varying habitats and diets, leopards have demonstrated a resilient presence in environments heavily populated by humans and subjected to intensive farming (Nowell et al., 1996). These creatures exhibit robust growth, exceeding a rate of 10% annually, and are identified as prolific breeders (Kumar et al., 2019). However, anthropogenic factors such as habitat degradation, conflicts with humans, natural prey depletion, and illegal poaching have severely impacted their populations, causing a significant contraction in their global range over the last century (Jhala et al., 2018).

Existing studies of leopard distribution and status have reported alarming reductions in leopard territory, with losses amounting to 48-67% in Africa and an even more devastating 83-87% in Asia (Jacobson et al., 2016). Correspondingly, recent genetic studies conducted in India have attributed human activities to a dramatic decline of 75-90% in leopard populations over the past 120-200 years. This alarming trend has prompted the International Union for Conservation of Nature (IUCN) to revise the status of leopards from "Near Threatened" to "Vulnerable" (IUCN, 2022). Despite India's efforts to safeguard these creatures by granting them maximum protection under the Wildlife Protection Act of 1972 and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), the persistent misperception of their abundance owing to their widespread visibility continues to threaten their survival.

In the southern regions of India, escalating human-wildlife conflicts are largely a consequence of the increasing human population and the subsequent fragmentation of habitats. Tea estates, due to their proximity to forests, are frequently regarded as hotspots for human-leopard conflicts (Bali et al., 2007). Additionally, the Western Ghats are increasingly grappling with the pressures of population expansion and development (Nath et al., 2022). The management of these sites faces immense challenges due to their extensive size, complex territories, and cluster formations (Nath et al., 2022). Furthermore, an increase in human-wildlife conflicts, predominantly involving Asian elephants and leopards, has been noted across these southern states, underlining the stress exerted by human encroachments (Hills et al., 2017).

The integrity of the forest corridors, critical to the protection of these sites, is at risk of being eroded due to demographic pressures (Nath et al., 2022). Given these pressing circumstances, it becomes imperative to investigate the correlation between leopard habitat preferences and instances of human-leopard conflicts. This examination will provide valuable insights to inform conflict mitigation strategies, promoting a sustainable model of human-leopard coexistence.

1.3 Research objective

- The overall aim of this study is to model the habitat suitability for the Indian leopard in two south Indian states (i.e., Tamil Nadu and Kerala) using the ensemble model.

1.4 Research questions

1. Does integrating anthropogenic variables into the ecological niche model improve the prediction accuracy of the suitable habitat for leopards?
2. What is the extent of potentially suitable habitats available for leopards outside protected areas as predicted by the ensemble model?
3. What is the extent of the difference in prediction accuracy between the GBIF dataset and the Status of Leopard Report dataset when used in species distribution modelling?

1.5 Research hypotheses

H_0: Incorporating anthropogenic variables into ecological niche modelling cannot significantly improve the prediction accuracy of the suitable habitat for leopards.

H_1: Incorporating anthropogenic variables into ecological niche modelling will significantly improve the prediction accuracy of the suitable habitat for leopards.

2. Materials and Methods

2.1 Study area

The geographic focus of this thesis encompasses the southern Indian states of Kerala and Tamil Nadu, spanning areas of 38,863 km² and 130,058 km² respectively. The specific coordinates and intricate topographical details of the region are illustrated in Figure 2. The selection of this study area was guided by its unique climatic conditions, rich ecological profile, and significant leopard populations. According to the 'Status of Leopards 2018' report, along with information sourced from the Ministry of Environment, the Western Ghats - which extend across these two states - provide a habitat for an estimated population of 3,387 leopards. Of these, Tamil Nadu, and Kerala host 868 and 650 leopards, respectively, largely within the confines of designated tiger reserves.

Internationally recognized for its ecological value, the Western Ghats region surpasses even the Himalayas in antiquity and stands as a notable geomorphic feature (Nath et al., 2022). The Ghats encapsulate the tropical monsoon system, supporting distinctive montane forest ecosystems that significantly modulate the Indian monsoon weather patterns while tempering the region's tropical temperatures (UNESCO, 2021). Furthermore, the Ghats function as a vital meteorological barrier, intercepting the monsoon winds originating from the southwest during late summer and early autumn, thereby preserving the region's unique climate (UNESCO, 2021). Leopards, as key predators within this habitat, play an necessary role in shaping the population dynamics of other species, fostering ecological balance and health. In addition, their status as an umbrella species implies that conservational initiatives targeting leopards would concurrently shield a countless of other cohabiting species.

Considering the leopards' ecological importance, their vulnerable conservation status, and the unique climatic and ecological features of Kerala and Tamil Nadu within the Western Ghats, this study area becomes a vital setting for examining leopard population dynamics and conservation needs. Gaining insights into these aspects could augment broader conservation strategies, thereby enriching the biodiversity of this globally significant region.

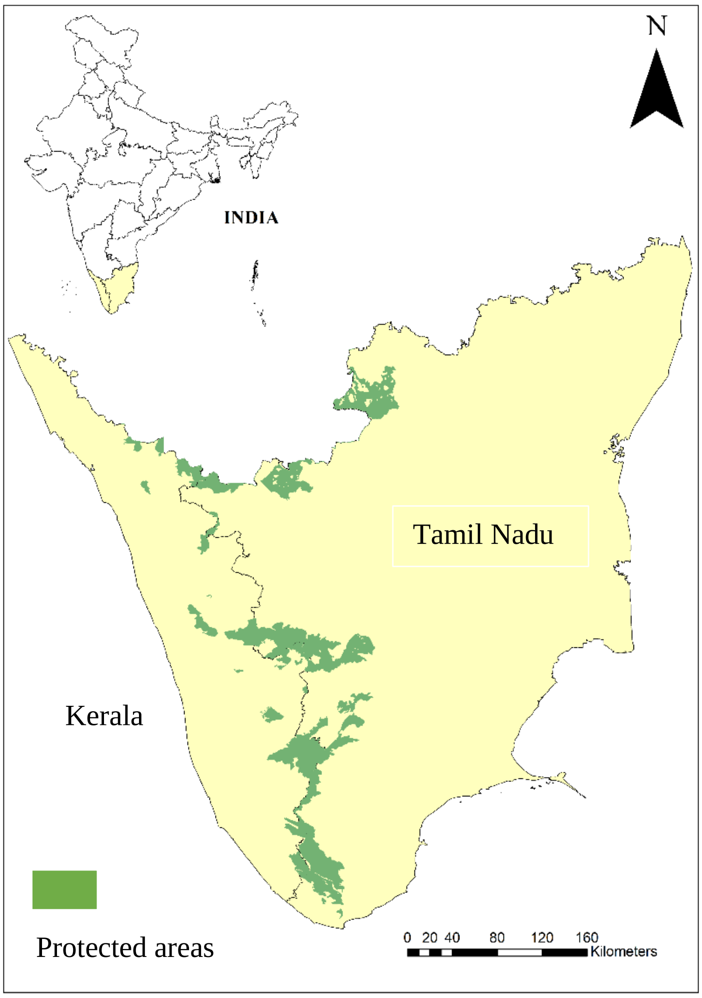

Figure 2: The location of the study area and the distribution of the protected areas in the southern Indian states of Kerala and Tamil Nadu

2.2 Datasets

A Combined dataset (GBIF and Leopard Report) of 300 leopard presence points was used in this study.

2.2.1 GBIF (Global biodiversity information facility) dataset

The Global Biodiversity Information Facility (GBIF) is an international open data infrastructure that allows anyone, anywhere, to access data about all types of life on Earth, shared across national boundaries via the internet. Its rich repository of biodiversity data, contributed by a surplus of participating institutions and organizations from across the world, forms a pivotal resource for scientific research, conservation, and sustainable development. The GBIF dataset encompasses a wide array of biodiversity data, including taxonomic information, species distribution records, and observational data, among others. This dataset, being dynamic and regularly updated, represents a high asset for global biodiversity studies and related research endeavours. A total of 80 leopard presence points were clipped and collected from the GBIF dataset DOI https://doi.org/10.15468/dl.d4srbz between 2017 and 2022 (figure 3) for the region of southern India.

The procedure for Extracting Location Data of Leopards from GBIF is:

1. Navigate to the GBIF homepage at https://www.gbif.org.
2. Use the 'Search' function located at the top of the webpage and enter 'Leopard' into the search field.
3. A drop-down menu will appear; select 'Species'. This will lead you to a new page containing information about various leopard species.
4. Select the specific species of leopard for which you wish to extract location data. For instance, if you are interested in the Panthera pardus, select it.
5. Once you have navigated to the chosen species page, click on the 'Occurrences' tab.
6. Here, you will find a map illustrating the global distribution of the selected species based on the occurrence data recorded in the GBIF dataset.
7. Above the map, there is an option to 'Download data'. Click on this button.
8. You will then be prompted to log in or create a GBIF account if you do not already have one. After logging in, you can proceed with the download request.
9. After submitting the download request, you will have to wait for the data to be prepared. The duration of this process can vary depending on the volume of data requested.
10. Once the data is ready, you will receive an email notification with a link to download the data. This dataset can be downloaded in various formats, such as CSV or Darwin Core Archive, and will include detailed location data for the selected leopard species.
11. After downloading the data, it can be imported into suitable data analysis software for further study and analysis.

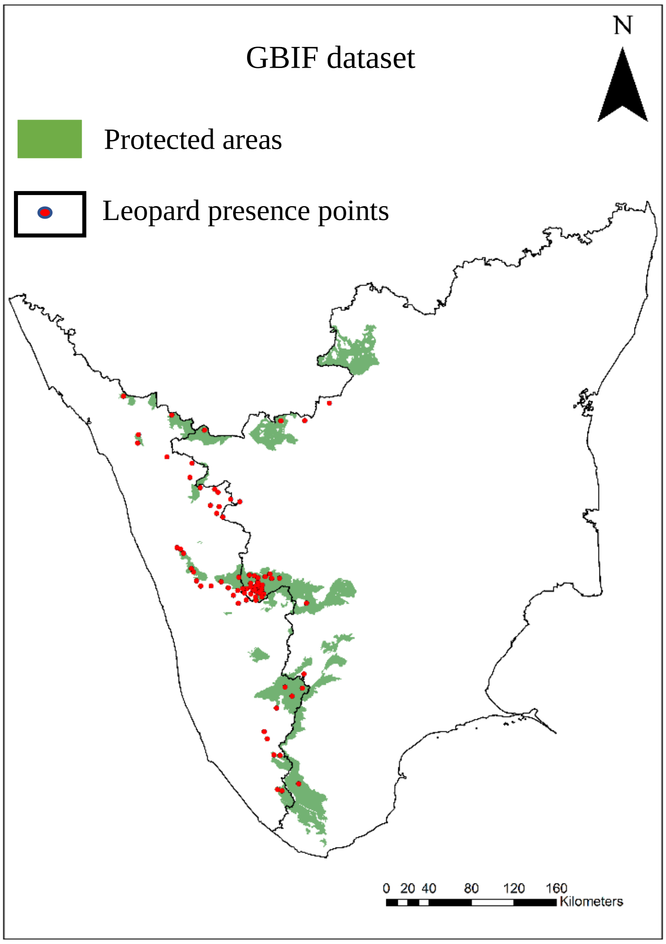

Figure 3: Spatial representation of GBIF dataset

2.2.2 Status of leopard report dataset

The data compilation for the report drew upon an extensive nationwide leopard census utilizing a broad-spectrum camera-trap survey, conducted across 141 diverse sites spanning 21 Indian states (Jhala et al., 2018). The strategic implementation of camera traps facilitated the capture and individual identification of leopards, leveraging their unique spot patterns. This enabled a robust estimation of leopard populations in the surveyed regions. The survey encompassed a sweeping area of approximately 26,838 square kilometres, inclusive of several protected zones and reserves, culminating in a robust dataset. Alongside camera-trapping, ancillary data sources were employed, encompassing field observations, track signs, local interviews, and incidental leopard sightings, to gain a holistic perspective on the distribution and status of leopards (Jhala et al., 2018). The amassed dataset was subjected to sophisticated statistical scrutiny, including the Spatially Explicit Capture-Recapture (SECR) models, to extrapolate leopard densities and population sizes.

The "Status of Leopards in India 2018" report serves as a potent foundation for species distribution modelling (Jhala et al., 2018). First, the curated data presents crucial insights into the spatial distribution of leopards across India, empowering researchers to identify potential habitats and comprehend the factors driving leopard distribution. Next, the report delineates leopard population densities across different regions, offering a pivotal understanding of habitat-carrying capacities. This is integral in forecasting potential shifts in leopard distribution instigated by factors such as habitat loss, prey availability, or human encroachment (Jhala et al., 2018).

Finally, the standardized approach to data collection across different states ensures data compatibility and reliability, fostering the development of robust, credible, and accurate species distribution models. These, in turn, yield critical perspectives, essential for effective conservation planning and management (Jhala et al., 2018).

A total number 230 leopard presence points were collected using the report (see figure 4). For spatial representation of camera trap data in the southern Indian states refer (appendix I).

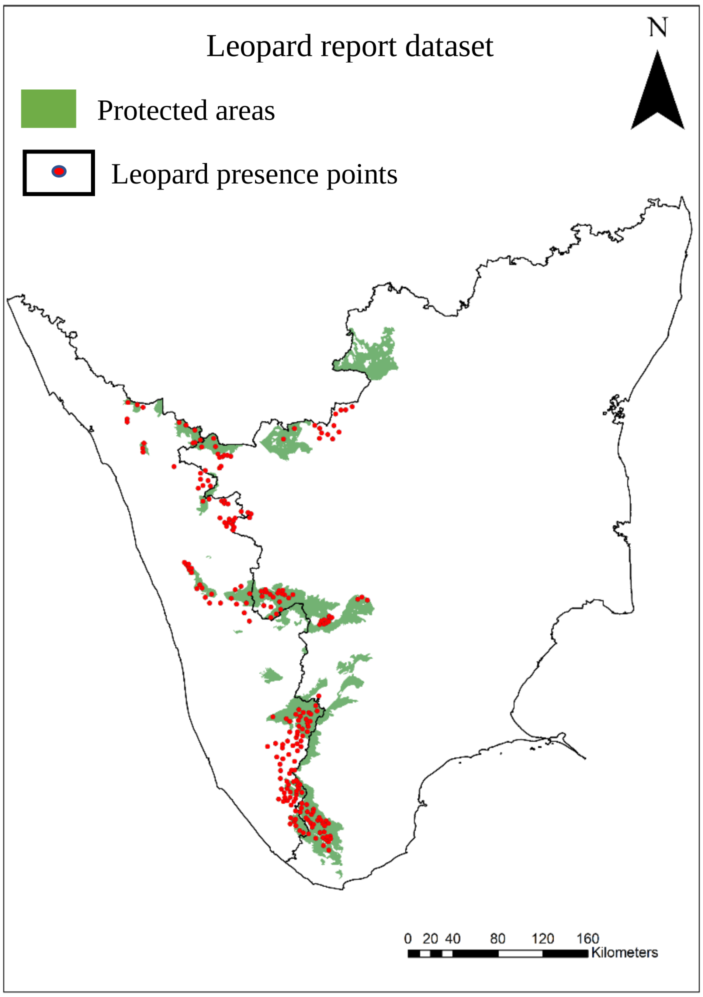

Figure 4: Spatial representation of Status of leopard report dataset

2.3 Predictor variables

In the current study, predictor variables for the Southern Indian leopards were meticulously selected based on a comprehensive review of existing literature, research studies, local news reports, national publications, as well as expert knowledge in the field. Predictor variables, or covariates, play a critical role in species distribution modelling . They serve as the independent factors that influence the presence, absence, or abundance of a species in specific locations (Guisan & Zimmermann, 2000). The correct selection and incorporation of predictor variables are vital in SDMs as they facilitate an understanding of how different environmental and anthropogenic factors may affect species distribution (Austin, 2007).

For example, variables such as temperature, precipitation, and land use might influence a species' ability to survive and reproduce, thereby determining its distribution (Elith & Leathwick, 2009). Additionally, understanding these relationships allows for the projection of potential shifts in species distributions in response to environmental change, a fundamental component in conservation planning and management (Guisan et al., 2013).

2.3.1 Environmental variables

The environmental variables used in this study were obtained from publicly accessible sources (refer to Table 1) and underwent pre-processing in ArcGIS software. Pre-processing involved converting the variables into the appropriate format, namely ASCII, and ensuring that they had a consistent spatial resolution of 1 km. Additionally, variables that originally had vector features, such as points and lines, were transformed into raster format while maintaining the same 1 km resolution.

Bio-climatic layers

Climate sensitivity is a fundamental aspect of all living organisms. Bio-climatic variables serve as biologically relevant indicators for characterizing species distribution across both continental scales (Blach-Overgaard et al., 2015) and regional scales (Kandel et al., 2015). To acquire the necessary bio-climatic data, the WorldClim database (http://worldclim.org/) was accessed. The WorldClim database, specifically version 2, consists of a comprehensive collection of global climate layers. The database includes annual time series with metrics such as annual means, seasonality, as well as extreme or limiting temperature and precipitation data (Hijmans et al., 2005). In this study, 19 bio-climatic layers and elevation (obtained from Worldclim database), each possessing a spatial resolution of 1 km, were employed (refer to Table 1).

Creation of additional environmental layers

To provide a more comprehensive evaluation of the habitat, additional environmental layers were generated encompassing water bodies, herbaceous vegetation, shrub, forest cover, and bare or sparsely vegetated areas. These layers were extracted from the Copernicus Global Land Service's Land Cover 100m data repository, specifically the version 3 Globe spanning the years 2015-2019. This dataset represents a high-resolution depiction of land cover on a global scale, generated through the analysis of satellite data and providing a vital resource for ecological modelling (Buchhorn et al., 2020). The Copernicus data has been recognized for its consistent, reliable, and comprehensive global coverage which is critical in studying the dynamics of land cover changes over time. By incorporating the environmental features from this dataset, our model can accurately consider factors such as the availability of water bodies, the extent of vegetated areas, and the presence of forest cover, which are integral elements of the leopards' habitat (Tuanmu & Jetz, 2014). Once obtained, these layers were then resampled to 1km spatial resolution using ArcGIS 10.8. Next, The forest height data employed in this study was sourced from the Global Forest Canopy Height dataset of 2019, distributed by the Global Land Analysis & Discovery (GLAD) group (Townshend et al., 2019). This dataset is characterized by its high spatial resolution of 30 meters, allowing for comprehensive and intricate insights into forest structure across global landscapes. For the purpose of this investigation, the original resolution of the data was modified; the layer was resampled to a coarser resolution of 1 km using ArcGIS 10.8.

Preparation of Normalized Difference Vegetation Index (NDVI) layers

A key aspect of this study involves the creation of multiple Normalized Difference Vegetation Index (NDVI) layers – namely, NDVI_maximum, NDVI_standard deviation, NDVI_minimum, and NDVI_mean – utilizing the Google Earth Engine platform. The Earth Engine allows for the development of a customizable program capable of computing NDVI for any geographic region, contingent upon user-specified geometry parameters. The output layers were rendered as Geo TIFF files with integrated coordinates and a spatial resolution of 1 km. These contain the corresponding NDVI value for each category (max, min, mean, and standard deviation) and can be directly incorporated into Geographic Information System (GIS) software such as ArcGIS and Q GIS.

The NDVI was calculated using the MOD13A2.061 Terra Vegetation Indices product, a MODIS (Moderate Resolution Imaging Spectroradiometer) dataset with a 16-day interval and global 1 km resolution. This dataset delivers various vegetation index layers that offer a precise representation of Earth's photosynthetically active vegetation and facilitate the observation and monitoring of vegetation conditions worldwide. Employing this dataset for NDVI computation helps in generating highly accurate environmental layers, crucial in species distribution modelling.

2.3.2 Anthropogenic variables

Anthropogenic variables represent those elements within an ecosystem substantially influenced or modified by human activities. These variables, in turn, exert profound implications on the flora and fauna, including species like the leopard, by altering habitats, reshaping landscapes, and changing ecological dynamics (Laurance et al., 2014). Particularly in the context of southern Indian states such as Kerala and Tamil Nadu, several anthropogenic variables pose considerable threats to leopard populations and their habitats. In this study, seven key anthropogenic variables considered include population density, distance to roads, distance to paths, distance to settlements, distance to protected areas, distance to farmlands, and croplands. Each of these variables reflects a different aspect of human-induced environmental changes impacting leopards and their habitats. Population density can exert pressure on habitats through activities like deforestation, pollution, and urbanization, directly affecting leopards' natural habitats (Ripple et al., 2014). Similarly, the distance to roads, paths, and settlements represents the encroachment of human-made structures and activities into leopard habitats, leading to habitat fragmentation and increased human-wildlife conflicts (Carter et al., 2012).

The proximity to protected areas and farmlands portrays the tension between conservation efforts and agricultural activities. Protected areas can serve as refuges for leopards, but nearby farmlands can become potential conflict zones due to livestock predation and crop damage (Athreya et al., 2013). Finally, croplands not only replace natural habitats but can also disrupt wildlife corridors, leading to further isolation and fragmentation of leopard populations (Karanth et al., 2010). These layers of anthropogenic variables, created at a resolution of 1 km using ArcGIS, serve as critical tools in understanding the spatial dynamics of these human-induced factors, enabling more precise and context-specific conservation strategies for leopards in southern India.

Table 1: All Predictor variables

Data Sources	Categories	Variables	Abbreviation	Units
Worldclim	Environmental Variables	Annual mean temperature	bio1	◦ C
		Mean diurnal range (mean of monthly (Max temp – min temp))	bio2	◦ C
		Isothermality (BIO2/BIO7)	bio3	Dimensionless
		Temperature seasonality (standard deviation)	bio4	◦ C
		Max temperature of the warmest month	bio5	◦ C
		Min temperature of the coldest month	bio6	◦ C
		Temperature annual range (BIO5-BIO6)	bio7	◦ C
		Mean temperature of wettest quarter	bio8	◦ C
		Mean temperature of driest quarter	bio9	◦ C

	Mean temperature of warmest quarter	bio10	◦ C
	Mean temperature of coldest quarter	bio11	◦ C
	Annual precipitation	bio12	mm
	Precipitation of wettest month	bio13	mm
	Precipitation of driest month	bio14	mm
	Precipitation seasonality coefficient of variation)	bio15	Dimensionless
	Precipitation of wettest quarter	bio16	mm
	Precipitation of driest quarter	bio17	mm
	Precipitation of warmest quarter	bio18	mm
	Precipitation of coldest quarter	bio19	mm
	Elevation	elevation	m
GEOFABRIK	Distance to water bodies		km
GLAD	Forest Height		m
Copernicus Global Land Service's Land Cover	Bare and Sparse Vegetation		Dimensionless
	Shrubs		Dimensionless
	Herbaceous vegetation		Dimensionless
	Distance to water bodies		km
	Bare and Sparse Vegetation		Dimensionless
Google Earth Engine	Annual minimum NDVI	ndvi_min	Dimensionless
	Annual mean NDVI	ndvi_mean	Dimensionless
	Annual maximum NDVI	ndvi_max	Dimensionless

		Standard deviation NDVI	NDVI_standard deviation	Dimensionless
NASA (SEDAC)	Anthropogenic Variables	Population density		Population per square km
GEOFABRIK		Distance to roads		km
		Distance to Railways		km
		Distance to built-up areas		km
		Distance to farmlands		km
GEONODE		Distance to Protected areas		km
Copernicus Global Land Service's Land Cover		Cropland		km

2.4 Collinearity analysis of predictor variables

Multicollinearity analysis is a critical procedure in statistical modelling, including ecological niche modelling (ENM), which addresses the issue of collinearity, or the high correlation among predictor variables (Dormann et al., 2013). It aims to identify and mitigate the influence of interrelated variables that may otherwise confound the results and interpretations of a model, thereby ensuring robust and credible model outcomes.

Within the context of ENMs multicollinearity can lead to an overestimation or underestimation of the effects of different environmental variables on species distributions (Elith et al., 2011). This can mislead the interpretation of species-environment relationships and can undermine the predictive performance of the model. Hence, a careful multicollinearity analysis is fundamental to the successful application of ENM.

In this study, multicollinearity analysis was performed using the Variance Inflation Factor (VIF) rule, specifically identifying variables with a VIF greater than 10. This rule stipulates that variables with a VIF exceeding 10 are considered highly collinear and thus should be omitted from the model (O'Brien, 2007). This step ensures that the final model is free from the undue influence of correlated variables, thereby increasing its interpretability and accuracy

The stepwise multicollinearity analysis conducted in this study leveraged the VIF step function (table 3), a tool developed by Babak Naimi (Naimi, 2018). This function operates on an iterative mechanism, initiating the process by determining the VIF of each variable incorporated in the model. Should any of the variables manifest a VIF value surpassing the predefined threshold, which is typically set at 10, the function identifies the variable with the most significant VIF value and eliminates it from the model. This procedure continues iteratively, systematically eliminating variables that demonstrate a VIF value beyond the threshold. The process ceases when all remaining variables in the model have VIF values within the acceptable range (VIF>10). The outcome of this process is a model with multicollinearity significantly reduced, thereby enhancing the interpretability and credibility of the model results.

Table 2 – VIF test for all variables

S.N.	Variables	VIF
1	Bare and Sparse Vegetation	1.385
2	Human population density	1.413
3	Distance to water bodies	1.419
4	Bio 15	1.441
5	Shrubs	1.442
6	Herbaceous vegetation	1.462
7	Distance to roads	1.473
8	Forest height	1.504
9	NDVI_minimum	1.544
10	Distance to farmlands	1.575
11	Bio 11	1.589
12	Elevation	1.622
13	Distance to build up area	1.708
14	Distance to railway	1.842
15	Bio 3	1.846
16	Bio 1	2.171
17	Bio 18	2.525
18	Bio 14	2.707
19	Bio 9	3.065
20	Bio 13	3.105
21	Bio 2	3.385
22	Cropland	3.701
23	Bio 17	3.713
24	Bio 7	3.831
25	Bio 8	3.856
26	Bio 4	3.981
27	Bio 10	4.067
28	Bio 12	4.295
29	Bio 6	4.581
30	NDVI_maximum	5.625
31	Forest cover	5.718
32	Bio 5	5.953
33	NDVI_standard deviation	6.027
34	Bio 16	8.228
35	Bio 19	8.238
36	Distance to Protected areas	8.899
37	NDVI_mean	8.988

Table 3 - VIF values after applying the VIF step function

S.N.	Variables	VIF
1	Human population density	1.256
2	Distance to water bodies	1.305
3	Bare and Sparse Vegetation	1.363
4	Herbaceous vegetation	1.375
5	Shrubs	1.418
6	Distance to roads	1.422
7	Distance to farmlands	1.468
8	NDVI_minimum	1.503
9	Forest height	1.530
10	Distance to build up area	1.606
11	Distance to railway	1.666
12	Bio 8	2.661
13	Cropland	3.496
14	Bio 19	3.551
15	Bio 14	4.167
16	Bio 15	4.266
17	Distance to Protected areas	4.595
18	NDVI_standard deviation	4.879
19	Bio 2	4.966
20	NDVI_maximum	5.084
21	Bio 3	5.108
22	Forest cover	5.927
23	Bio 18	5.984
24	NDVI_mean	7.756

Table 4 – Environmental variables after VIF Test

S.N.	Environmental Variables	VIF
1	Distance to water bodies	1.305
2	Bare and Sparse Vegetation	1.363
3	Herbaceous vegetation	1.375
4	Shrubs	1.418
5	NDVI_minimum	1.503
6	Forest height	1.530
7	Bio 8	2.661
8	Bio 19	3.551
9	Bio 14	4.167
10	Bio 15	4.266
11	NDVI_standard deviation	4.879
12	Bio 2	4.966
13	NDVI_maximum	5.084
14	Bio 3	5.108
15	Forest cover	5.927
16	Bio 18	5.984

17	NDVI_mean	7.756

Table 5 – Anthropogenic variables after VIF Test

S.N.	Anthropogenic Variables	VIF
1	Human population density	1.256
2	Distance to roads	1.422
3	Distance to farmlands	1.468
4	Distance to build up area	1.606
5	Distance to railway	1.666
6	Cropland	3.496
7	Distance to Protected areas	4.595

2.5 Ecological niche modelling

Ecological Niche Modelling (ENM), also known as species distribution modelling or habitat suitability modelling, is a predictive tool that utilizes statistical algorithms to identify relationships between species occurrences and their environmental and spatial characteristics (Elith & Leathwick, 2009). It operates on the fundamental ecological principle that species are not randomly distributed across landscapes but are confined to suitable environments dictated by their evolutionary adaptations and ecological needs (Soberón, 2007).

In the context of wildlife and habitat conservation, ENM has emerged as a significant tool owing to its ability to predict the potential distribution of species under different environmental scenarios, providing critical insights for biodiversity management and conservation planning (Franklin, 2013). It assists in identifying areas of high conservation value, informing reserve design, assessing the impact of climate change on species distributions, and guiding efforts to reintroduce species or mitigate human-wildlife conflicts (Guillera-Arroita et al., 2015).

The application of ENM to leopards in the southern Indian states of Kerala and Tamil Nadu holds considerable promise. By combining occurrence records with a range of environmental and anthropogenic variables, ENM can identify areas that offer suitable conditions for leopards. This can enhance our understanding of the species ecological requirements and potential distribution within these regions, which can be used to prioritize areas for conservation (Phillips et al., 2017). Furthermore, ENM can predict the impacts of future land-use changes on leopard habitats, thereby providing valuable guidance for land-use planning and conflict mitigation strategies to ensure the coexistence of humans and leopards in these areas (Elith et al., 2010).

2.5.1 Model scenario

Ensemble modelling is a methodological approach that combines the predictions from multiple individual models to generate a single, more robust prediction. This methodology capitalizes on the strength of consensus, reducing the likelihood of decision-making based on anomalous results from a single model (Araújo & New, 2007). By integrating a variety of predictions, ensemble modelling can mitigate the uncertainties inherent in single-model predictions and improve the precision and reliability of the final outputs (Marmion et al., 2009).

Role in Species Distribution Modelling:

Ensemble modelling, by combining outputs from multiple SDMs, offers a more comprehensive and reliable view of species distributions. It allows for a more better understanding of the relationships between species and their environment, making it a particularly valuable tool in ecology and conservation biology. The ability to integrate and compare different model outputs has made ensemble modelling a preferred choice for predicting the potential impacts of climate change, land use change, or other environmental changes on species distribution patterns (Buisson et al., 2010). Consequently, ensemble modelling has become instrumental in formulating effective conservation strategies and managing biodiversity under global change scenarios (Araújo & New, 2007).

In this study, 4 SDM techniques are used –

1. Maxent –

- The Maximum Entropy Model, more commonly known as Maxent, is a widely utilized tool in the field of species distribution modelling. Developed by Phillips, Anderson, and Schapire (2006). Within the realm of species distribution modelling, Maxent is used to infer potential distributions of species based on the environmental constraints of the locations where they have been observed. It operates under the assumption that the species under study is at distributional equilibrium with its environment, i.e., it is found wherever suitable conditions exist and does not occur where conditions are unsuitable (Elith et al., 2011).
- Maxent is particularly notable for its applicability to presence-only data, which makes it useful when absence data are not available or reliable. This ability to operate efficiently with limited data inputs makes it highly suitable for applications such as the study of rare or elusive species and in regions where comprehensive survey data is lacking. Maxent also provides an array of output formats that are interpretable and useful for conservation planning, such as logistic output for habitat suitability and response curves to elucidate relationships between variables and probability of presence (Phillips & Dudík, 2008).
- Overall, Maxent is a robust, flexible, and user-friendly tool that has been shown to perform well compared to other species distribution modelling methods, especially when using small sample sizes or incomplete data, making it a significant asset in modern ecological research.

2. Generalized Liner Model –

- The Generalized Linear Model (GLM) is a fundamental and widely employed statistical method in the field of species distribution modeling (Guisan et al., 2002). As a flexible extension of ordinary linear regression, GLMs allow for response variables with error distributions other than the normal distribution, making them a powerful tool for modelling a range of biological phenomena (McCullagh & Nelder, 1989).

- GLMs are particularly beneficial for examining the relationships between a response variable (e.g., species presence or abundance) and multiple explanatory variables, such as environmental or anthropogenic factors. The relationships are formulated in a systematic component that linearly combines the effects of these factors on a suitable scale for the response (Hastie & Tibshirani, 1986).
- The application of GLMs in species distribution modelling allows ecologists to determine the influence of each environmental factor on species distribution while controlling for others, thereby providing insights into the species' ecological requirements and tolerances (Austin, 2002). This understanding can be harnessed to predict a species' potential distribution under different environmental scenarios, which is fundamental for conservation planning and management. Notably, GLMs are also capable of handling both presence-absence and presence-only data, further increasing their utility in ecological studies where complete datasets might not always be available (Yee & Mitchell, 1991).

3. Boosted Regression Tree –

- Boosted Regression Trees (BRTs) amalgamate the strengths of two machine learning techniques: regression trees and boosting, providing a powerful tool for understanding species-environment relationships and predicting species distribution (Elith et al., 2008). Regression trees generate simple, interpretable decision rules, but their predictive performance is relatively modest. In contrast, boosting improves the accuracy by combining many simple models to create a single, highly accurate prediction model. BRT models can represent complex nonlinear relationships and interactions between predictors, thus enabling a nuanced understanding of species-environment relationships (De'ath, 2007).
- In species distribution modelling, BRTs allow for flexible response shapes, provide robust predictive performance, and handle different types of predictor variables, making them highly versatile (Elith et al., 2008). Furthermore, BRTs can handle missing data and are resistant to outliers, both common challenges in ecological datasets (Leathwick et al., 2006). BRTs also can quantify variable importance, providing insight into which environmental factors are most influential in determining the species' distribution (Hastie et al., 2009). This can provide crucial information for guiding conservation actions and understanding the potential impacts of environmental changes.

4. Random Forest –

- Random Forest (RF) is an ensemble machine-learning technique that builds multiple decision trees and merges their predictions to produce a final, more accurate prediction (Breiman, 2001). This method has gained widespread recognition in the field of species distribution modelling due to its ability to handle complex ecological data effectively (Cutler et al., 2007). A key advantage of RF lies in its robustness to overfitting and its ability to handle large datasets with many predictor variables, even when they are highly correlated or contain missing values (Breiman, 2001). It also efficiently models non-linear relationships and interactions between predictors without requiring explicit specification (Hastie et al., 2009).

- In species distribution modelling, RF offers insights into the relative importance of predictor variables in determining species' distribution, providing crucial knowledge for habitat management and conservation planning (Prasad et al., 2006). For example, variable importance measures can reveal which environmental factors are most influential for a species' presence or absence, helping to prioritize conservation actions. Furthermore, RF can be used to predict species' distribution under current conditions as well as under future scenarios of environmental change, offering valuable foresight for biodiversity conservation strategies in the face of climate change (Evans et al., 2011).

Subsequently, an ensemble modelling approach was adopted to integrate the four Species Distribution Models (SDMs). This approach was executed in two distinctive manners:

- Incorporation of environmental variables exclusively.
- Incorporation of both environmental and anthropogenic variables.

The former examines the potential natural distributions based on environmental conditions, whereas the latter considers human-induced changes, thereby providing a more realistic projection of species distribution in the current human-dominated landscapes.

All the 4 models (Maxent, GLM, BRT and RF) were incorporated via the Species Distribution Modelling (SDM) package, a tool devised by Babak Naimi (Naimi, 2015). This package was selected for its capacity to generate comprehensive ecological niche and species distribution predictions. Next, each of the four SDMs were executed independently ten times. Subsequently, the results from these iterations were integrated into an ensemble model, which served as the basis for further analysis. This approach was designed to enhance the reliability and robustness of our predictive models.

2.5.2 Threshold selection for predicting suitable leopard habitat

For ensemble modelling, the Uncertainty Analysis for Species Distribution Models (USDM) R-package was utilized; the USDM package in R is an instrumental tool in ecological studies, particularly for species distribution modelling. The package is designed to assess uncertainties associated with these models, integrate several modelling algorithms to create ensemble models, estimate variable importance, and evaluate correlations among predictors (Naimi & Araújo, 2016). Understanding these uncertainties is vital in biodiversity conservation as it influences planning and decision-making processes. Ensemble forecasting, facilitated by USDM, is a sophisticated strategy employed in species distribution modelling. It amalgamates predictions from multiple individual models to provide a comprehensive, robust, and more accurate prediction (Araújo & New, 2007). This approach counteracts the weaknesses of individual models, thus enhancing overall model performance and reliability.

In the current study, the 'weighted,' 'stat,' and 'opt' parameters were utilized to generate a habitat suitability map for the leopards in Southern India. Specifically, 'Weighted Averaging of True Skill Statistics (TSS) was employed. This method leverages the TSS from each model, averaging them based on their respective weights to produce the final ensemble result ('weighted'). The 'stat' parameter, set to 'TSS,' indicates that the True Skill Statistics is used for model evaluation. TSS is a highly regarded performance measure in species distribution modelling as it accounts for both omission and commission errors and is not affected by prevalence (Allouche et al., 2006). The 'opt' parameter was set to '2', which signifies the usage of the TSS value in the ensemble model. The selection of '2' corresponds to maximizing the sum of sensitivity and specificity, aligning with recommended practices in species distribution modelling (Liu et al., 2013).

In summary, the USDM package, and specifically its capability for ensemble forecasting, offers a valuable approach to creating robust and reliable species distribution models. The weighted averaging of TSS to generate the final ensemble model ensures an accurate representation of the habitat suitability for leopards in Southern India.

2.6 Model performance assessment

The evaluation of model performance is a cornerstone in the application of machine learning algorithms and modelling procedures, serving as a critical indicator of a model's predictive capabilities (Guisan et al., 2017). In the present study, the predictive effectiveness of the machine learning algorithms employed for habitat suitability modelling was evaluated using two widely adopted accuracy assessment metrics in species distribution modelling: the 'Area Under the Curve' (AUC) of the Receiver Operating Characteristics (ROC) function (Elith et al., 2006), and the True Skill Statistics (TSS) (Allouche et al., 2006).

The AUC, a threshold-independent metric, gauges the model's aptitude to distinguish species presence from absence (Elith et al., 2006; Guisan et al., 2017). The value of AUC ranges between 0 and 1, where 1 denotes perfect discrimination, 0.5 implies predictive accuracy is no better than a random estimate, and less than 0.5 indicates subpar performance, worse than a random estimate (Elith et al., 2006). While widely utilized, AUC has faced criticism for its accuracy measurements (Lobo et al., 2008). In conjunction with AUC, the TSS was employed as an additional model performance assessment metric. Unaffected by prevalence and factoring in both omission and commission errors, TSS ranges from +1 to -1. A score of +1 signifies flawless performance, while a score of zero or less represents a classifier performance that is no better than a random prediction (Allouche et al., 2006). As TSS is a threshold-dependent metric, the threshold value was determined by maximizing the sum of sensitivity and specificity, as advocated by previous research (Liu et al., 2013).

3. Results

3.1 Leopard habitat suitability modelling

3.1.1 Predicted suitable habitat for leopards using only environmental variables

The ensemble model only with environmental variables (figure 6) was modelled by integrating four distinct SDM techniques: (a) Maxent, (b) GLM, (c) BRT, and (d) RF (figure 5). The ensemble model (figure 6) represents the habitat suitability for leopards with an AUC - 0.95 and TSS – 0.82

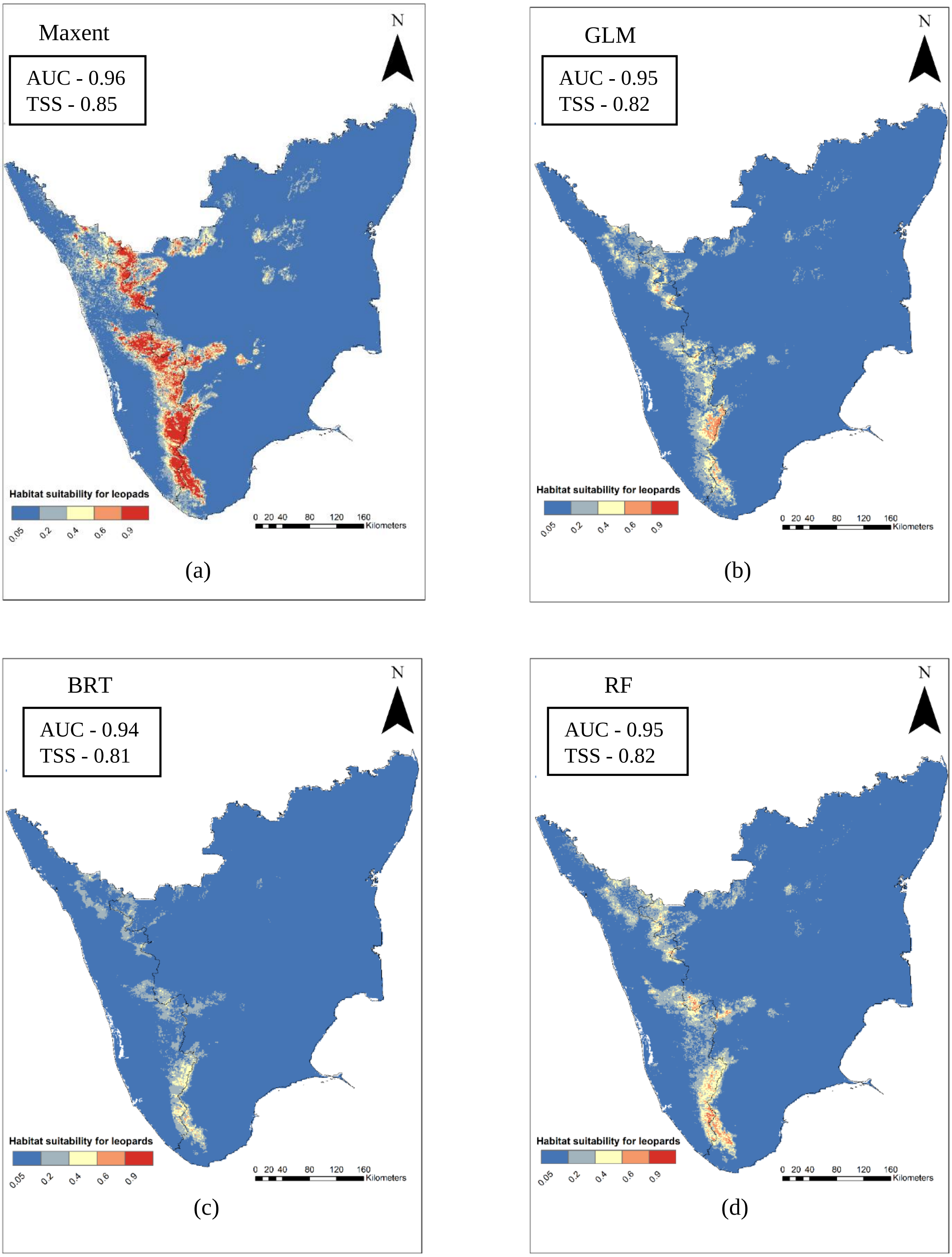

Figure 5: Depicted here are the outcomes of four distinct SDM techniques, namely, (a) Maxent, (b) GLM, (c) BRT, and (d) RF. These models were generated solely using environmental variables. For a detailed list of the employed environmental variables (table 4).

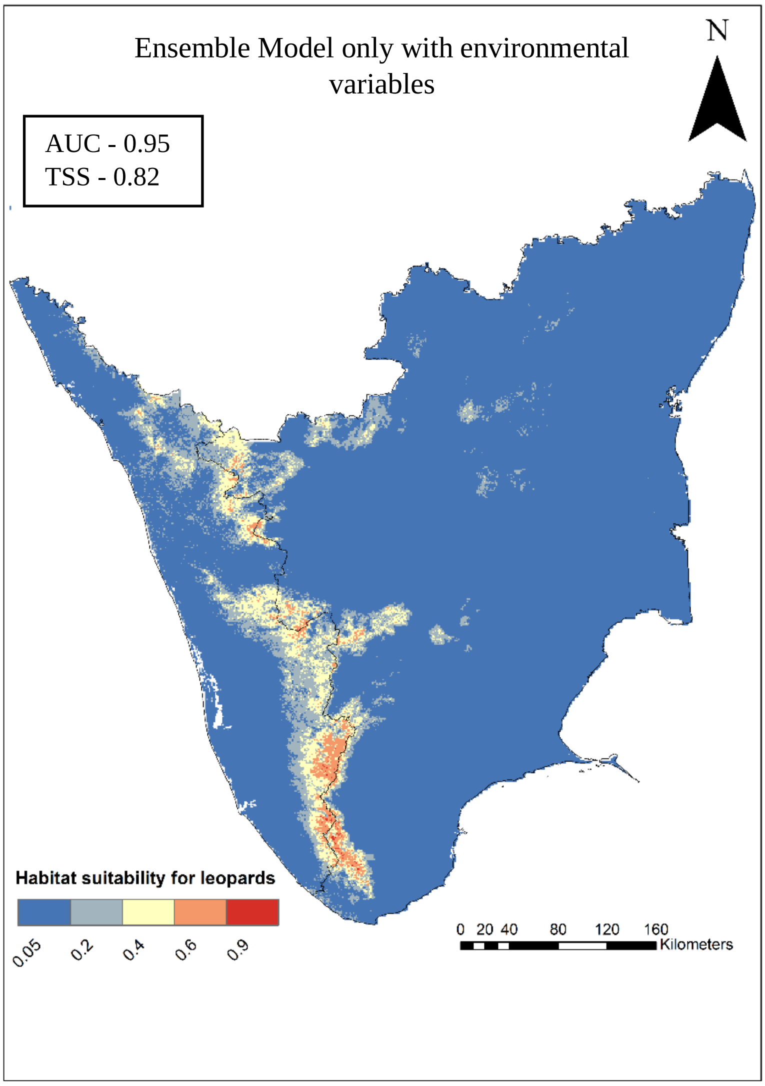

Figure 6: Depicted here is a habitat suitability map, a product of the ensemble model. This model unifies the four distinct SDM techniques, as shown in Figure 5, with an exclusive focus on environmental variables. The resulting predictive accuracy and performance of this environmental variable-based ensemble model are indicated by an AUC score of 0.95 and a TSS score of 0.82, respectively.

3.1.2 Predicted suitable habitat for leopards using environmental variables and anthropogenic variables

The ensemble model with environmental variables and anthropogenic variables (figure 8) was modelled by integrating four distinct SDM techniques: (a) Maxent, (b) GLM, (c) BRT, and (d)

RF (see figure 7). The ensemble model (figure 8) represents the habitat suitability for leopards with an AUC - 0.96 and TSS – 0.85.

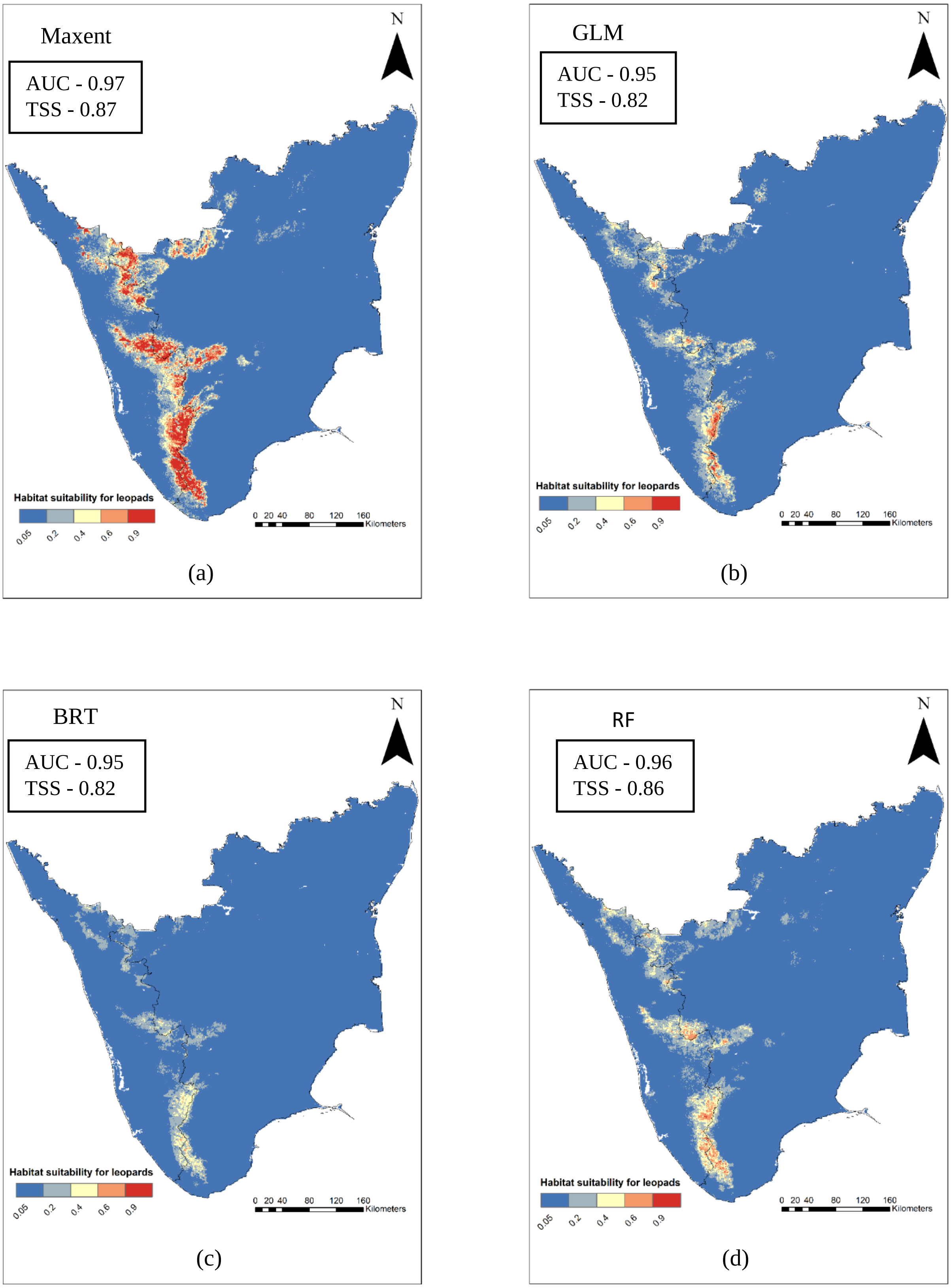

Figure 7: Depicted here are the outcomes of four distinct SDM techniques, namely, (a) Maxent, (b) GLM, (c) BRT, and (d) RF. The development of these models was conducted by integrating both environmental and anthropogenic variables. For a detailed list of the employed environmental variables and anthropogenic variables (Table 4 and 5)

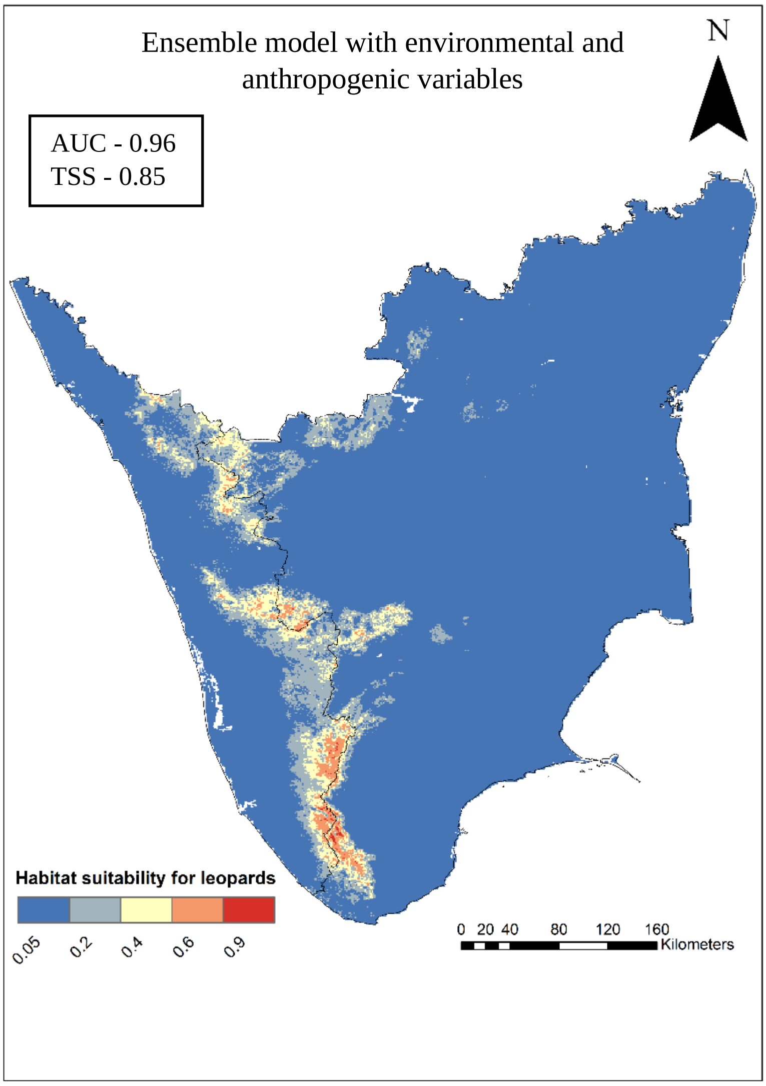

Figure 8: Depicted here is a habitat suitability map, a product of the ensemble model. This model unifies the four distinct SDM techniques, as shown in Figure 7, employing both environmental and anthropogenic variables. The resulting predictive accuracy and performance of this environmental variable-based ensemble model are indicated by an AUC score of 0.96 and a TSS score of 0.85, respectively

3.2 The distribution of suitable leopard habitat in Southern India

The evaluation of suitable leopard habitat in Southern India (figure 9) was conducted using a multifaceted approach, integrating both Environmental and Anthropogenic variables and employing four SDM techniques - Maxent, GLM, BRT, and RF. This comprehensive analysis was achieved through the utilization of 'weighted,' 'stat,' and 'opt' parameters, which collectively facilitated the development of a habitat suitability map for Southern Indian leopards. A specific technique, 'Weighted Averaging of TSS, was deployed, leveraging the TSS derived from each model. These were then averaged in accordance with their relative weights to yield the final ensemble outcome ('weighted') (Araújo & New, 2007).

The 'stat' parameter was configured to 'TSS, for the evaluation of the model. As a performance metric, TSS is highly esteemed in species distribution modelling due to its capacity to account for both omission and commission errors, and its immunity to prevalence (Allouche et al., 2006). The 'opt' parameter was set to '2,' signifying the application of the TSS value in the ensemble model. This choice aligns with best practices in species distribution modelling, as it maximizes the sum of sensitivity and specificity (Liu et al., 2013).

Criteria	Threshold	Sensitivity	Specificity	TSS
Max (se+sp)	0.04768	0.984	0.891	0.874

The ensemble model (figure 10) predicts a total suitable habitat area of 21,797 square kilometres within the southern states of India. Of this, a total of 7,426.015 square kilometres falls within protected areas, leaving 14,370.985 square kilometres of suitable habitat situated outside protected zones. In terms of individual states, Tamil Nadu has a predicted suitable habitat area of 9,587 square kilometres, while Kerala has a larger area of 12,132 square kilometres (Appendix II).

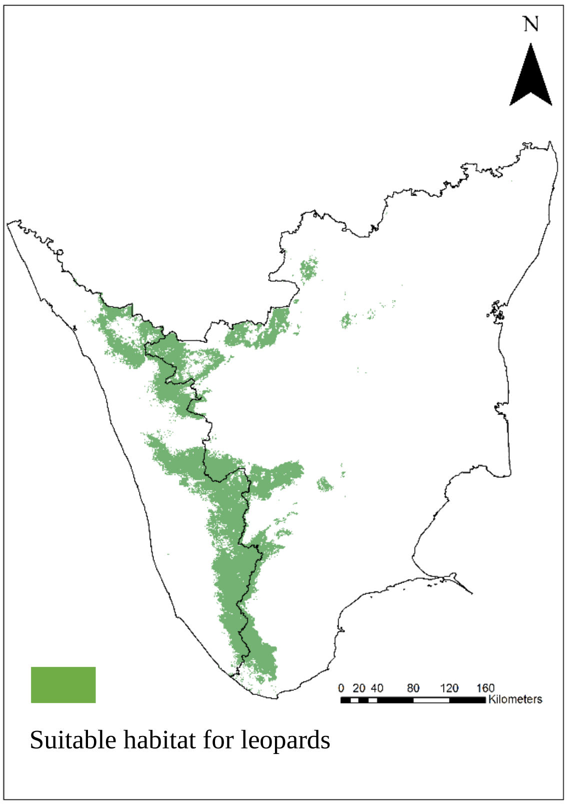

Figure 9: Predicted suitable habitat for leopards

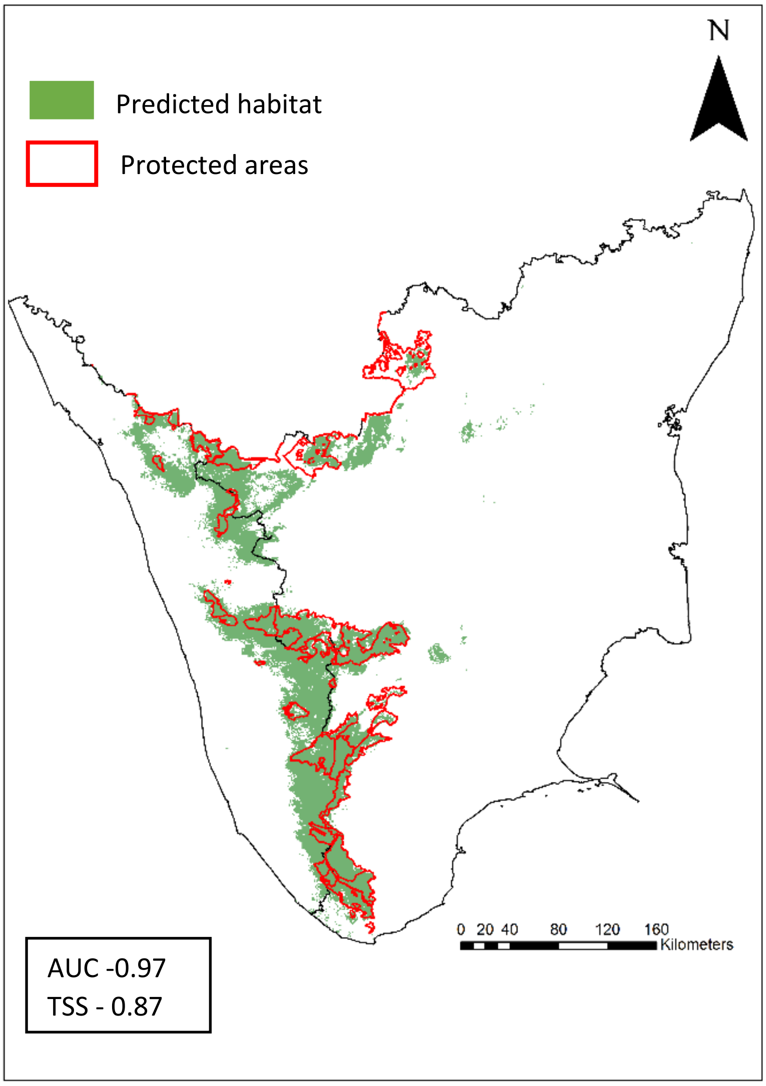

Figure 10 - Spatial representation of predicted Leopard Habitat within and beyond protected areas in the southern states of Tamil Nadu and Kerala

3.3 Distribution of suitable leopard habitat using different datasets

3.3.1 Predicted suitable habitat using the GBIF dataset

Ensemble model (figure 11) with both the environmental and anthropogenic variables was used to generate habitat suitability map relying on data obtained from the GBIF pertaining to leopard sightings. The dataset constituted approximately 80 presence points gathered between 2017 and 2022, providing a valuable basis for mapping leopard habitats in the study area. This method adheres to current best practices in species distribution modelling, which emphasize the importance of considering a broad spectrum of influencing factors to attain an accurate depiction of species habitats (Guisan & Thuiller, 2005).

The performance of this habitat suitability model was assessed using two widely accepted evaluation metrics: AUC and TSS. The AUC value achieved was 0.94, indicating an excellent model performance, as values closer to 1 suggest a near-perfect ability of the model to distinguish between presence and absence areas (Swets, 1988). The TSS value, at 0.78, also demonstrates a high level of accuracy, as values closer to +1 denote perfect agreement between observed and predicted species presence (Allouche et al., 2006).

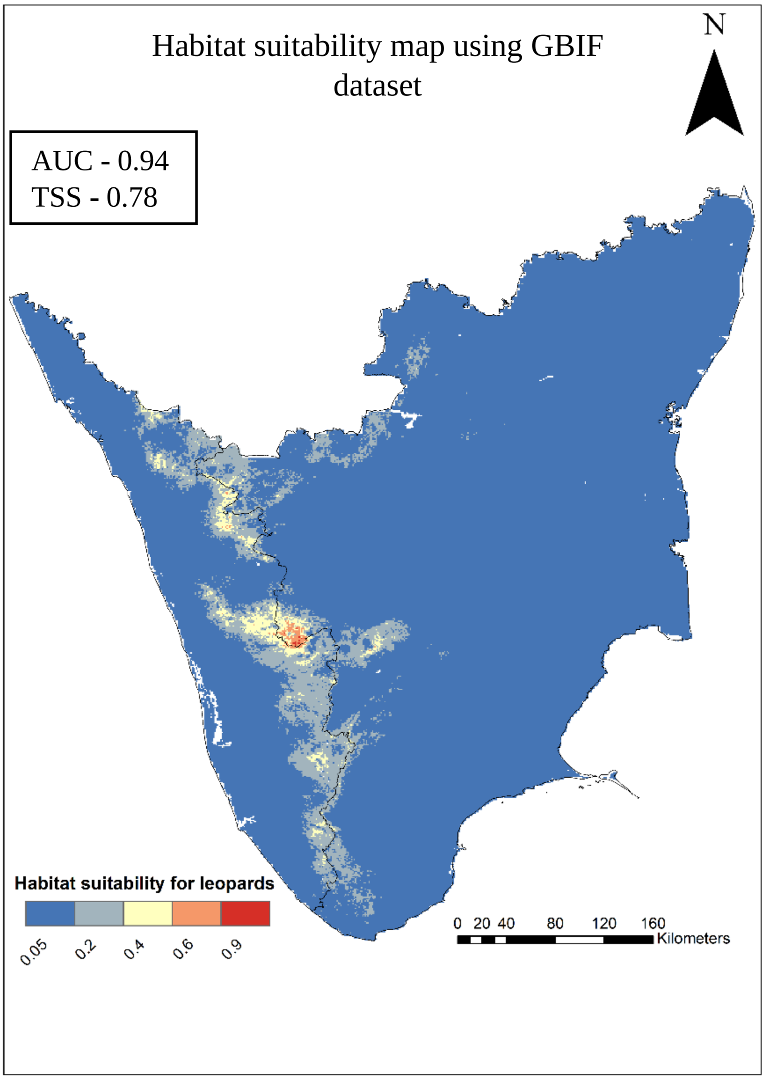

Figure 11: Habitat suitability map using the GBIF dataset

3.3.2 Predicted suitable habitat using the status of leopard report dataset

Ensemble model (figure 12)with both the environmental and anthropogenic variables was used to generate habitat suitability map relying on data obtained from the Status of the Leopard report, India 2018. This invaluable report provided approximately 230 presence points of leopards, allowed us to build a robust and holistic understanding of leopard habitat suitability in the southern states of India.

The efficacy of the generated model was rigorously evaluated using two commonly employed statistical measures: the AUC and TSS. The model performed commendably, achieving an AUC score of 0.95 and a TSS score of 0.83. An AUC score of 0.95 signifies an exceptional discriminatory power of the model, as values closer to one represent excellent discrimination between presence and absence locations (Swets, 1988). Similarly, a TSS score of 0.83 suggests a high level of predictive accuracy, with values closer to +1 indicating flawless agreement between observed and predicted species presence (Allouche et al., 2006).

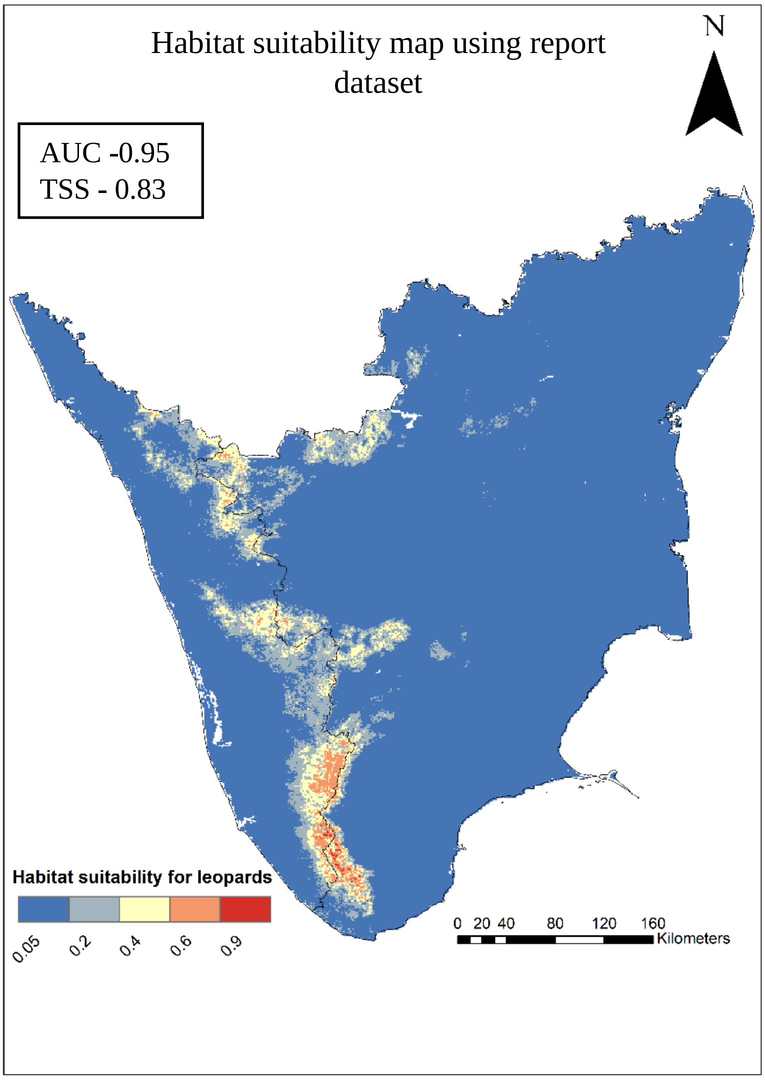

Figure 12: Habitat suitability map using the status of the leopard report dataset

4. Discussion

Leopards (Panthera pardus) is integral to the intricate ecosystems of the Western Ghats in India, playing a vital role in maintaining the stability and diversity of this complex ecological landscape (Ripple et al., 2014). As apex predators, leopards regulate prey populations, thus controlling the population dynamics and contributing to the balance of trophic interactions within the system (Terborgh et al., 2001). This leads to 'trophic cascades,' which significantly influence community structure and biodiversity (Estes et al., 2011).

Conserving leopards in the Western Ghats is crucial due to their role as an umbrella species. Essentially, their conservation can indirectly protect the multitude of species that share their habitat, leading to wider biodiversity conservation (Noss et al., 1996). Leopards also have a vast home range, and preserving these expansive habitats can lead to larger, more robust, and interconnected conservation areas that benefit overall ecological integrity (Riggio et al., 2013). Moreover, leopards may have cultural and economic importance for local communities through tourism, enhancing local support for wildlife conservation initiatives (Goodrich, 2010). However, leopards in the Western Ghats face increasing threats due to anthropogenic factors such as habitat fragmentation, prey depletion, and retaliatory killings due to human-wildlife conflicts (Jacobson et al., 2016). Therefore, it is of utmost importance to implement effective conservation strategies and promote harmonious human-leopard coexistence to safeguard the ecological health and resilience of the Western Ghats.

4.1 Assessing the Influence of anthropogenic factors on leopard habitat suitability

This study's findings illustrate the integral role that anthropogenic variables can play in enhancing the accuracy of ecological niche models. Particularly in the context of leopards, widely recognized as one of the most adaptable large cat species, the integration of anthropogenic variables with environmental variables in the above models presented a detailed understanding of their potential habitats (Athreya et al., 2013). Increasing reports from the southern Indian states of Tamil Nadu and Kerala reveals that leopards have demonstrated a remarkable ability to subsist in human-influenced environments, sometimes even exploiting these conditions for their benefit (Athreya et al., 2016). A salient example is the ease with which leopards’ prey on domesticated animals due to their ready accessibility, a phenomenon likely exacerbated by forest fragmentation, and unauthorized cattle grazing in protected areas (Bhatnagar et al., 2013). Such trends highlight the influence of anthropogenic factors on leopard ecology, and by extension, their habitat suitability. In fact, the increased accuracy observed in the above models (figure 7 and 8) upon the inclusion of anthropogenic variables provides empirical support for this relationship. These findings underscore the necessity of considering anthropogenic influences, alongside environmental parameters, in studies aiming to predict leopard habitat suitability accurately (Guisan et al., 2013).

This increased understanding can significantly contribute to wildlife management strategies and conservation plans, especially in regions like Tamil Nadu and Kerala where human-leopard interactions are commonplace (Athreya et al., 2016). As human-dominated landscapes continue to expand, integrating knowledge of anthropogenic influences into ecological niche modelling will become increasingly important in ensuring the survival of adaptable species like leopards (Guisan et al., 2013).

4.2 An ensemble model's perspective on leopard habitat suitability in Southern India

The ensemble model (figure 9) has provided a comprehensive estimate of suitable leopard habitats in the southern states of India, notably Tamil Nadu and Kerala. Covering a total area of 21,797 square kilometers, the model enables us to understand the spatial distribution of these habitats and can be instrumental in future wildlife conservation planning and management efforts.

Interestingly, a significant portion of these suitable habitats - around 7,426.015 square kilometers - is encompassed by protected areas. This is a promising indicator, suggesting that current conservation efforts might be partly aligned with leopard habitat requirements. Nonetheless, the model (figure 10) also predicts 14,370.985 square kilometers of suitable habitat outside these protected zones, hinting at the potential for spatial conflicts between leopards and human activities. This also highlights the need for enhanced conservation strategies beyond existing protected areas. In a state-wise comparison, Kerala has been found to harbor a larger expanse of suitable leopard habitat, 12,132 square kilometers, as opposed to Tamil Nadu 9,587 square kilometers (Appendix III). Such findings could guide state-specific conservation policies, considering the differing ecological and anthropogenic contexts of these regions. Nonetheless, it is essential to note that our model's accuracy and precision could be significantly augmented with the inclusion of live presence data of leopards, especially if procured from the wildlife departments and national parks of Tamil Nadu and Kerala. The ensemble model is data-dependent; hence, its predictive power is invariably tied to the quality and quantity of data utilized. Live presence data can offer a real-time snapshot of leopard distributions, enabling the model to fine-tune its predictions and present a more accurate picture of their habitats. Accurate models are integral to effective conservation strategies. They offer reliable insights into species' habitat requirements and their spatial distribution, facilitating the formulation of data-driven, targeted, and efficient conservation plans. Therefore, these findings recommend future research collaborations with wildlife departments and national parks to enhance data collection efforts and consequentially, the accuracy of habitat suitability models.

In conclusion, this study underscores the importance of habitat suitability assessments in wildlife conservation, the potential of ensemble models in such assessments, and the necessity for incorporating high-quality, real-time data to improve these models. Further research in these directions would undoubtedly contribute to our collective goal of effective leopard conservation.

4.3 The Impact of dataset selection on habitat suitability modelling for leopards

This investigation shows the importance of the quality and comprehensiveness of presence-only data used in habitat suitability modelling, particularly about leopards in the southern Indian states of Tamil Nadu and Kerala. Comparative analysis of habitat suitability maps derived from the GBIF dataset (figure 11) and the Status of the Leopard 2018 Report (figure

12) revealed noteworthy differences. While both sets of data contributed valuable insights into leopard habitat preferences, the GBIF-sourced map was marked by lower AUC and TSS values. This discrepancy may be linked to the disparate methods of presence point collection. The GBIF dataset, heavily weighted towards the central regions of Tamil Nadu and Kerala, indicated a concentrated 'red zone' of leopard presence. However, data drawn from the 2018 Leopard Report, collected via extensive camera trap deployments throughout the tiger reserves, revealed a more geographically dispersed leopard presence (Phillips et al., 2009).

Interestingly, the GBIF dataset exhibited some spatial inaccuracies. Specifically, presence points marked within large urban areas of Tamil Nadu and Kerala, which are unlikely leopard habitats, had to be discarded. This issue underlines the limitations of exclusive reliance on the GBIF dataset for species distribution modelling, reinforcing the importance of supplemental data sources for a more accurate and comprehensive depiction of habitat suitability.

5. Conclusion

This study has furthered our understanding of the factors influencing leopard habitat suitability, most notably the role of anthropogenic elements. This research revealed the remarkable resilience of leopards, exhibiting their adaptability to thrive even in human-influenced environments, exploiting these conditions to their advantage. Hence, the inclusion of anthropogenic variables, along with environmental factors, in the ensemble models has undeniably enriched our understanding of potential leopard habitats. Figure 9 shows clearly that a significant portion of suitable leopard habitats fall within protected areas, affirming the value of these conservation efforts. However, a large expanse of habitat lies outside these zones, which calls for a broader outlook on conservation strategies. State-specific findings, especially regarding Tamil Nadu and Kerala, could be valuable in formulating localized conservation policies, aligning them with the ecological and anthropogenic specificities of the respective regions. The key to more precise habitat suitability models lies in integrating higher quality, real-time data. Live presence data, sourced from wildlife departments and national parks, can present an accurate snapshot of leopard distributions, allowing the model to fine-tune its predictions. Emphasizing the importance of reliable data, Steve Irwin once said, "I believe our biggest issue is the same biggest issue that the whole world is facing, and that's habitat destruction." Hence, utilizing authentic data will not only refine the models but also adopt more effective conservation strategies. In the realm of habitat suitability modelling, dataset selection holds utmost importance. The differences noted in habitat suitability maps drawn from the GBIF dataset (figure 11) and the Status of the Leopard 2018 report dataset (figure 12) highlight this. Spatial inaccuracies detected in the GBIF dataset reiterate the need for careful data selection and supplementing data sources for more comprehensive habitat suitability assessments.

To quote an old African proverb, "A leopard cannot change its spots." We must remember that while leopards may adapt to changing environments, they remain bound by their inherent ecological requirements. Their survival, like all wildlife, hinges on the preservation of suitable habitats, and it's our responsibility to provide them with this. In conclusion, this study accentuates the significance of habitat suitability assessments, the potential of ensemble models, and the value of incorporating authentic, real-time data in improving these models. It prompts us to envisage innovative conservation strategies that extend beyond protected areas, bearing in mind the unique adaptability of leopards. The path towards effective leopard conservation is arduous, but our journey is inspired by the hope of hearing the leopard's roar echoing in the wilderness for generations to come.

6. Recommendation

This study highlights the value of integrating machine learning and statistical techniques, in the form of Species Distribution Models (SDMs), into the domain of wildlife conservation. Emerging as a potent tool amidst the escalating global biodiversity crisis, SDMs transcend purely academic relevance, offering actual insights beneficial to policymaking, wildlife management strategies, and real-world conservation initiatives.

These models adeptly map the potential geographical distribution of species, using environmental and, as our study highlights, anthropogenic variables. Such information is paramount for the conservation of rare and endangered species, whose habitats might be fragmented, insufficiently understood, or under threat from human activities. Furthermore, SDMs serve as predictive tools, anticipating future distributions in the wake of changing climatic and land-use conditions. This predictive prowess allows conservation managers to proactively counter potential threats and adopt timely mitigation strategies. A nuanced understanding of biodiversity's spatial distribution, as facilitated by SDMs, steers conservation efforts towards areas boasting high species richness, or those sheltering endangered or endemic species. In this context, SDMs inform critical conservation decisions, such as the designation of protected areas, prioritization of zones for restoration, or the identification of potential wildlife corridors facilitating species movement. Notably, SDMs promote evidence-based decision-making in conservation, providing governments, conservation organizations, and society at large with the means to comprehend and visualize the potential conflicts between biodiversity conservation and other land-use interests. This understanding can stimulate the development of balanced policies and management strategies that harmonize biodiversity conservation with sustainable development goals. As we navigate the challenges of accelerated anthropogenic changes and declining biodiversity, the importance of robust, data-driven tools like SDMs cannot be overstated. By bridging the divide between scientific knowledge and conservation practice, they represent indispensable allies in our collective quest for a sustainable co-existence with the planet's diverse range of species.

Moreover, our study recommends the integration of live presence data into these models, which would enhance their accuracy. The use of real-time data, particularly from wildlife departments and national parks, provides a more accurate picture of species distribution, allowing for more precise predictive modelling. Finally, the study emphasizes the value of state-specific analyses in forming localized conservation policies, aligning them with the ecological and anthropogenic contexts of the respective regions. This targeted approach, enriched by reliable data, and powered by the precision of SDMs, promises to heighten the effectiveness of conservation efforts.

In conclusion, the current study has reaffirmed the potential of SDMs as crucial tools in contemporary wildlife conservation, driving policymaking and management strategies with data-driven insights. By addressing the evolving challenges to biodiversity, especially in the context of anthropogenic influences, and advocating for the integration of higher-quality, real-time data, this research sets the stage for more effective, sustainable, and inclusive conservation strategies in the future.

LIST OF REFERENCES

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa, and the true skill statistic (TSS). Journal of Applied Ecology, 43(6), 1223-1232.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa, and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223– 1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Almond, R. E. A., Grooten, M., Juffe Bignoli, D., & Petersen, T. (2022). Living Planet Report 2022. https://livingplanet.panda.org/en-us/

Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33(10), 1677–1688. https://doi.org/10.1111/J.1365-2699.2006.01584.X

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology & Evolution, 22(1), 42-47.

Araújo, M.B., New, M., 2006. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Athreya, V., Odden, M., Linnell, J. D., & Karanth, K. U. (2016). A cat among the dogs: leopard Panthera pardus diet in a human-dominated landscape in western Maharashtra, India. Oryx, 50(1), 156-162.

Athreya, V., Odden, M., Linnell, J.D., Krishnaswamy, J., & Karanth, U. (2013). Big cats in our backyards: persistence of large carnivores in a human-dominated landscape in India. PloS one, 8(3), e57872.

Athreya, V., Odden, M., Linnell, J.D., Krishnaswamy, J., & Karanth, U. (2013). Big cats in our backyards: persistence of large carnivores in a human dominated landscape in India. PLoS ONE, 8(3), e57872.

Austin, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling, 157(2-3), 101-118.

Bai, D. F., Chen, P. J., Atzeni, L., Cering, L., Li, Q., & Shi, K. (2018). Assessment of habitat suitability of the snow leopard (Panthera uncia) in Qomolangma National Nature Reserve based on MaxEnt modeling. Zoological Research, 39(6), 373. https://doi.org/10.24272/J.ISSN.2095-8137.2018.057

Bai, D.F., Chen, P.J., Atzeni, L., Cering, L., Li, Q., Shi, K., 2018. Assessment of habitat suitability of the snow leopard (Panthera uncia) in Qomolangma National Nature Reserve based on MaxEnt modeling. Zool. Res. 39, 373–386. https://doi.org/10.24272/j.issn.2095-8137.2018.057

Baldwin, R.A., 2009. Use of maximum entropy modeling in wildlife research. Entropy 11, 854– 866. https://doi.org/10.3390/e11040854

Bali, A., Kumar, A., & Krishnaswamy, J. (2007). The mammalian communities in coffee plantations around a protected area in the Western Ghats, India. Biological Conservation, 139(1–2), 93–102. https://doi.org/10.1016/J.BIOCON.2007.06.017

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011

Belgiu, M., Drăgu, L., 2016. "Random forest in remote sensing: A review of applications and future directions." ("Random forest in remote sensing: A review of applications and future ...") ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. Ecology Letters, 15(4), 365-377.

Bhatnagar, Y. V., Wangchuk, R., & Jackson, R. (2013). Trans-Himalayan large carnivores: their status and conservation. Biodiversity Conservation in the Himalaya. G. S. Rawat and Y. V. Bhatnagar (Eds.). Nature Conservation Foundation, Mysore, 171-191.

Bhatt, S., Biswas, S., Karanth, K., Pandav, B., & Mondol, S. (2020). Genetic analyses reveal population structure and recent decline in leopards (Panthera pardus fusca) across the Indian subcontinent. PeerJ, 2020(2), e8482. https://doi.org/10.7717/PEERJ.8482/SUPP-5

Bradie, J., & Leung, B. (2017). A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. Journal of Biogeography, 44(6), 1344–1361. https://doi.org/10.1111/JBI.12894

Bradie, J., Leung, B., 2017. "A quantitative synthesis of the importance of variables used in MaxEnt species distribution models." ("Sci-Hub | A quantitative synthesis of the importance of variables used ...") J. Biogeography. 44, 1344–1361. https://doi.org/10.1111/jbi.12894

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Breiman, L., 2001. Random Forests, in: Machine Learning. pp. 5–32.

Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. Global Change Biology, 16(4), 1145-1157.

Carter, N. H., & Linnell, J. D. C. (2016). Co-Adaptation Is Key to Coexisting with Large Carnivores. Trends in Ecology & Evolution, 31(8), 575–578. https://doi.org/10.1016/J.TREE.2016.05.006

Carter, N.H., Shrestha, B.K., Karki, J.B., Pradhan, N.M., & Liu, J. (2012). Coexistence between wildlife and humans at fine spatial scales. Proceedings of the National Academy of Sciences, 109(38), 15360-15365.

Constant, N. L., Bell, • S, & Hill, • R A. (2015). The impacts, characterisation, and management of human-leopard conflict in a multi-use land system in South Africa. Biodiversity Conservation, 24, 2967–2989. https://doi.org/10.1007/s10531-015-0989-2

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007). Random forests for classification in ecology. Ecology, 88(11), 2783–2792.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., Mcclean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop.). 36, 27–46. https://doi.org/10.1111/j.1600- 0587.2012.07348.x

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Leitão, P.J. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1), 27-46.

Eizirik, E., Murphy, W. J., Koepfli, K. P., Johnson, W. E., Dragoo, J. W., Wayne, R. K., & O'Brien, S. J. (2010). Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. Molecular Phylogenetics and Evolution, 56(1), 49–63. https://doi.org/10.1016/J.YMPEV.2010.01.033

Elith, J., & Leathwick, J.R. (2009). Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40, 677-697.

Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models. Ecography (Cop.). 32, 66–77. https://doi.org/10.1111/j.1600-0587.2008.05505.x

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve the prediction of species' distributions from occurrence data. Ecography (Cop.). 29, 129– 151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. Methods in Ecology and Evolution, 1(4), 330-342.

Elith, J., Kearney, M., & Phillips, S. (2011). The art of modeling range-shifting species. Methods in Ecology and Evolution, 2(4), 330-342.

Elith, J., Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40.

Elith, J., Leathwick, J.R., & Hastie, T. (2008). A working guide to boosted regression trees. Journal of Animal Ecology, 77(4), 802-813.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for

Environ. Model. Softw. 100, 136–145. https://doi.org/10.1016/j.envsoft.2017.11.009

Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., Carpenter, S. R., Essington, T. E., Holt, R. D., Jackson, J. B., Marquis, R. J., Oksanen, L., Oksanen, T., Paine, R. T., Pikitch, E. K., Ripple, W. J., Sandin, S. A., Scheffer, M., Schoener, T. W., ... Wardle, D. A. (2011). Trophic downgrading of planet Earth. Science, 333(6040), 301-306.

Feng, X., Park, D.S., Liang, Y., Pandey, R., Papeş, M., 2019. Collinearity in ecological niche modeling: Confusions and challenges. Ecol. Evol. 9, 10365–10376. https://doi.org/10.1002/ece3.5555

Fernández-Sepúlveda, J., & Martín, C. A. (2022). Conservation status of the world's carnivorous mammals (order Carnivora). Mammalian Biology, 1, 1–15. https://doi.org/10.1007/S42991-022-00305-8/FIGURES/5

Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods in Ecology and Evolution, 6(4), 424-438.

Franklin, J., 2012. Species distribution modeling, in: Mapping Species Distributions. Cambridge University Press, Cambridge, pp. 3–20. https://doi.org/10.1017/cbo9780511810602.003

Genuer, R., Poggi, J.-M., Tuleau-malot, C., 2010. Variable selection using Random Forests.

Gobeyn, S., Mouton, A.M., Cord, A.F., Kaim, A., Volk, M., Goethals, P.L.M., 2019. Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. Ecol. Modell. 392, 179–195. https://doi.org/10.1016/j.ecolmodel.2018.11.013

Gubbi, S., Sharma, K., & Kumara, V. (2020). Every hill has its leopard: Patterns of space use by leopards (Panthera pardus) in a mixed-use landscape in India. PeerJ, 8. https://doi.org/10.7717/peerj.10072

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8(9), 993–1009. https://doi.org/10.1111/J.1461-0248.2005.00792.X

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8(9),

Guisan, A., Thuiller, W., 2005. Predicting species distribution: Offering more than simple habitat models. Ecol. Lett. 8, 993–1009. https://doi.org/10.1111/j.1461- 0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., ... & Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. Ecology letters, 16(12), 1424-1435.

Hamilton, P. H. (1976). The movements of leopards in Tsavo National Park, Kenya, as determined by radio-tracking (Doctoral dissertation, University of Nairobi).

Hills, A., Author, I., Sidhu, S., Raghunathan, G., Mudappa, D., & Shankar Raman, T. R. (2017). Conflict to Coexistence: Human-Leopard Interactions in a Plantation Landscape in Anamalai Hills, India. 15(4), 474–482. https://doi.org/10.4103/cs.cs_16_35

Jacobson, A. P., Jr, J. R., Gerngross, P., Lemeris Jr, J. R., Schoonover, R. F., Anco, C., Breitenmoser-Würsten, C., Durant, S. M., Farhadinia, M. S., Henschel, P., Kamler, J. F., Laguardia, A., Rostro-García, S., Stein, A. B., & Dollar, L. (2016). Leopard (Panthera pardus) status, distribution, and the research efforts across its range. https://doi.org/10.7717/peerj.1974

Jha, K. K., & Jha, R. (2021). Study of Vulture Habitat Suitability and Impact of Climate Change in Central India Using MaxEnt. Https://Doi.Org/10.5814/j.Issn.1674-764x.2021.01.004, 12(1), 30–42. https://doi.org/10.5814/J.ISSN.1674-764X.2021.01.004

Jhala, Y. V., Qamar, Q., & Yadav, S. P. (2018). Status Of Leopards in India. https://wii.gov.in/status_of_leopard_2018

Kamyo, T., & Asanok, L. (2019). Modeling habitat suitability of Dipterocarpus alatus (Dipterocarpaceae) using MaxEnt along the Chao Phraya River in Central Thailand. Https://Doi.Org/10.1080/21580103.2019.1687108, 16(1), 1–7. https://doi.org/10.1080/21580103.2019.1687108

Kindt, R., 2018. Ensemble species distribution modelling with transformed suitability values.

Kumar, A., & Verma, A. K. (2020). BIODIVERSITY LOSS AND ITS ECOLOGICAL IMPACT IN INDIA. www.nesa-india.org

Macdonald, E. A., Burnham, D., Hinks, A. E., Dickman, A. J., Malhi, Y., & Macdonald, D. W. (2015). Conservation inequality and the charismatic cat: Felis felicis. Global Ecology and Conservation, 3, 851–866. https://doi.org/10.1016/J.GECCO.2015.04.006

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

Mpakairi, K., Ndaimani, H., Vingi, K., Madiri, T.H., Nekatambe, T., 2018. Ensemble modelling predicts Human Carnivore Conflict for a community adjacent to a protected area in Zimbabwe. Afr. J. Ecol. 56, 957–963. https://doi.org/10.1111/aje.12526

Na, X., Zhou, H., Zang, S., Wu, C., Li, W., & Li, M. (2018). Maximum Entropy modeling for habitat suitability assessment of Red-crowned crane. Ecological Indicators, 91, 439–446. https://doi.org/10.1016/J.ECOLIND.2018.04.013

Naha, D., Dash, S. K., Kupferman, C., Beasley, J. C., & Sathyakumar, S. (2021). Movement behavior of a solitary large carnivore within a hotspot of human-wildlife conflicts in India. Scientific Reports 2021 11:1, 11(1), 1–14. https://doi.org/10.1038/s41598-021-83262-5

Nath, A., Chakraborty, P., Sarkar, V., Joshi, C., Bhadouria, B. S., Kakati, N., Ghosh, S., & Talukdar, G. (2022). Monitoring Outstanding Universal Value: An Analysis of the Status of Natural World Heritage Sites in India. Https://Doi.Org/10.1177/24559296221091470, 7(1), 37–62. https://doi.org/10.1177/24559296221091470

Nowell, K., Jackson, P., & IUCN/SSC Cat Specialist Group. (1996). Wild cats: status survey and conservation action plan. 382. https://portals.iucn.org/library/node/6998

Pattern Recognit. Lett. Elsevier 31, 2225–2236.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S.J.., Dudík, M., E. Schapire, R., 2004. A Maximum Entropy Approach to Species Distribution Modeling, in: Proceedings of the Twenty-First International Conference on Machine Learning. pp. 655–662.

Probst, P., Wright, M.N., Boulesteix, A.L., 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 9, 1–15. https://doi.org/10.1002/widm.1301

Rather, T. A., Kumar, S., & Khan, J. A. (2020). Multi-scale habitat modelling and predicting change in the distribution of tiger and leoProtrpard using random forest algorithm. Scientific Reports 2020 10:1, 10(1), 1–19. https://doi.org/10.1038/s41598-020-68167-z

Ripple, W. J., Estes, J. A., Beschta, R. L., Wilmers, C. C., Ritchie, E. G., Hebblewhite, M., Berger, J., Elmhagen, B., Letnic, M., Nelson, M. P., Schmitz, O. J., Smith, D. W., Wallach, A. D., & Wirsing, A. J. (2014). Status and ecological effects of the world’s largest carnivores. Science, 343(6167). https://doi.org/10.1126/SCIENCE.1241484/SUPPL_FILE/RIPPLE.SM.PDF

Terborgh, J., Lopez, L., Nuñez, P. v., Rao, M., Shahabuddin, G., Orihuela, G., Riveros, M., Ascanio, R., Adler, G. H., Lambert, T. D., & Balbas, L. (2001). Ecological Meltdown in Predator-Free Forest Fragments. Science, 294(5548), 1923–1926. https://doi.org/10.1126/SCIENCE.1064397

Townshend, J., Hansen, M., Carroll, M., Dimiceli, C., Sohlberg, R., White, P., & Wulder, M. (2019). Global Forest Canopy Height, 2019. Global Land Analysis & Discovery (GLAD). Retrieved from http://www.glad.umd.edu/dataset/global-forest-canopy-height-2019

UNESCO. (2021). https://whc.unesco.org/en/list/1342/

Wolpert, D.H., 1992. Stacked generalization. Neural Networks 5, 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Woodman, S.M., Forney, K.A., Becker, E.A., DeAngelis, M.L., Hazen, E.L., Palacios, D.M., Redfern, J. V., 2019. ESDM: A tool for creating and exploring ensembles of predictions from species distribution and abundance models. Methods Ecol. Evol. 10, 1923–1933. https://doi.org/10.1111/2041-210X.13283

Yadav, N., Areendran, G., Sarma, K., Raj, K., & Sahana, M. (2021). Susceptibility assessment of human–leopard conflict in Aravalli landscape of Haryana using geospatial techniques. Modeling Earth Systems and Environment, 7(3), 1459–1473. https://doi.org/10.1007/S40808-020-00858-Y/FIGURES/12

Appendix I – Spatial representation of camera trap data in southern Indian states

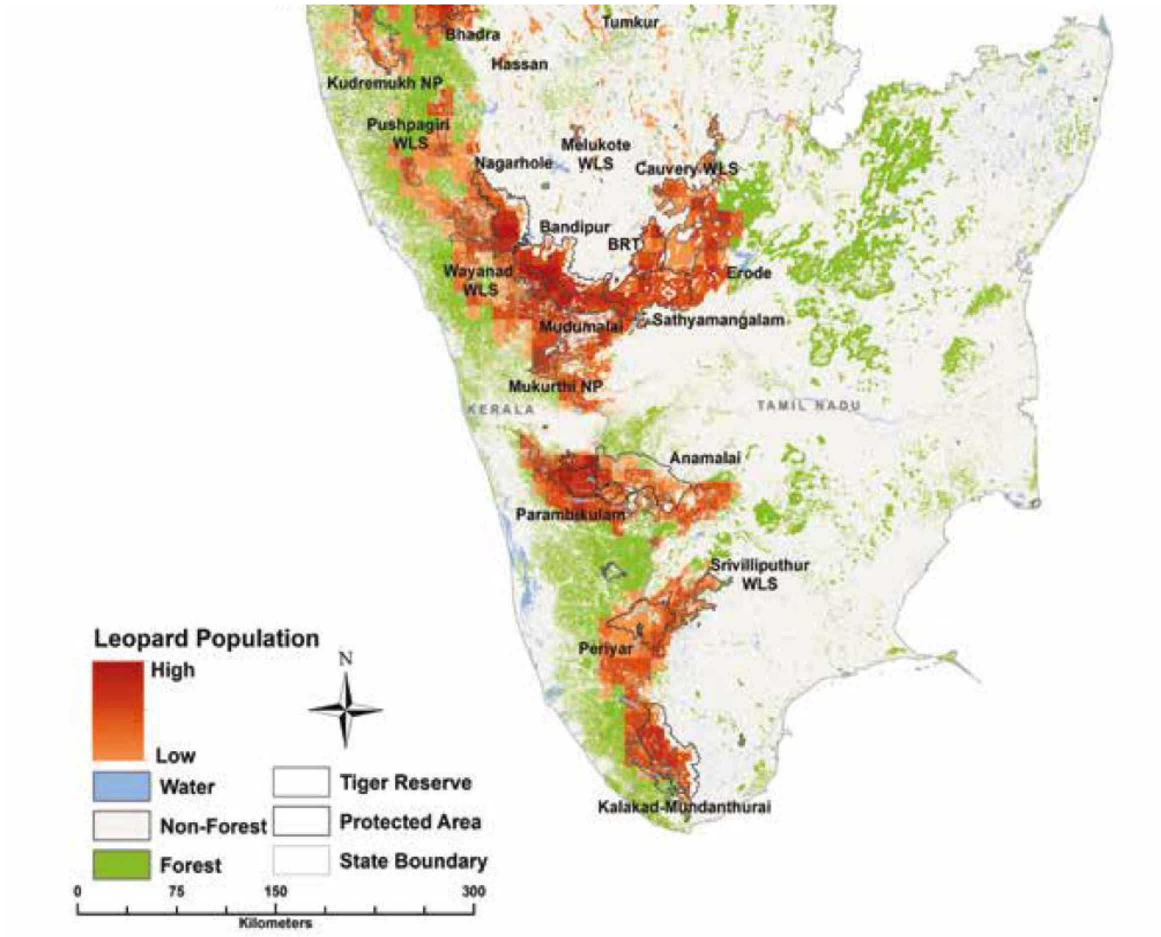

Figure 13: This illustration represents a density map of the leopard population across southern Indian states, generated utilizing camera trap data collated from various tiger reserves.

Appendix II – Protected areas of southern India

Table 6 – Protected areas of Kerala and Tamil Nadu

Name of the Protected Area	**Type**	**State**	**Area in Km^2**
Mathikettan Shola National Park	National Park	Kerala	12.82
Pambadum Shola National Park	National Park	Kerala	1.318
Aralam WLS	Wildlife Sanctuary	Kerala	55
Kottiyoor WLS	Wildlife Sanctuary	Kerala	30.38
Peechi-Vazhani WLS	Wildlife Sanctuary	Kerala	125
Chimmony WLS	Wildlife Sanctuary	Kerala	85
Chinnar WLS	Wildlife Sanctuary	Kerala	90.44
Peppara WLS	Wildlife Sanctuary	Kerala	53
Parambikulam WLS/Tiger Reserve	Wildlife Sanctuary	Kerala	285
Thattekad Bird WLS	Wildlife Sanctuary	Kerala	25
Chulannur Peafowl WLS	Wildlife Sanctuary	Kerala	3.42
Shendurney WLS	Wildlife Sanctuary	Kerala	100.32
Malabar WLS	Wildlife Sanctuary	Kerala	74.215
Neyyar WLS	Wildlife Sanctuary	Kerala	128
Idukki WLS	Wildlife Sanctuary	Kerala	70
Kurinjimala WLS	Wildlife Sanctuary	Kerala	32
Anamudi Shola National Park	National Park	Kerala	7.5
Periyar WLS/Tiger Reserve	Wildlife Sanctuary	Kerala	427
Wayanad WLS	Wildlife Sanctuary	Kerala	344.44
Periyar National Park/Tiger Reserve	National Park	Kerala	350
Silent Valley	National Park	Kerala	89.52
Eravikulam National Park	National Park	Kerala	97
Sathyamangalam WLS/Tiger Reserve	Wildlife Sanctuary	Tamil Nadu	1411.61
Srivilliputhur Grizzled Squirrel WLS	Wildlife Sanctuary	Tamil Nadu	485.2
Kannyakumari WLS	Wildlife Sanctuary	Tamil Nadu	457.78
Nellai WLS	Wildlife Sanctuary	Tamil Nadu	356.73
Cauvery North WLS	Wildlife Sanctuary	Tamil Nadu	504.334
Kodaikanal WLS	Wildlife Sanctuary	Tamil Nadu	608.95
Kalakad WLS	Wildlife Sanctuary	Tamil Nadu	223.58
Mundanthurai WLS	Wildlife Sanctuary	Tamil Nadu	567.38
Indira Gandhi WLS (Anamalai)	Wildlife Sanctuary	Tamil Nadu	841.49
Megamalai WLS	Wildlife Sanctuary	Tamil Nadu	269.11
Mukurthi	National Park	Tamil Nadu	78.46
Mudumalai	National Park	Tamil Nadu	103.23
		Total Area	**7426.015 km^2**

Appendix III - District-Wise Distribution and Quantitative Analysis of Predicted Habitat Suitability and Protected Areas in Kerala and Tamil Nadu

Kerala has a predicted suitable habitat area of 12,132 square kilometres

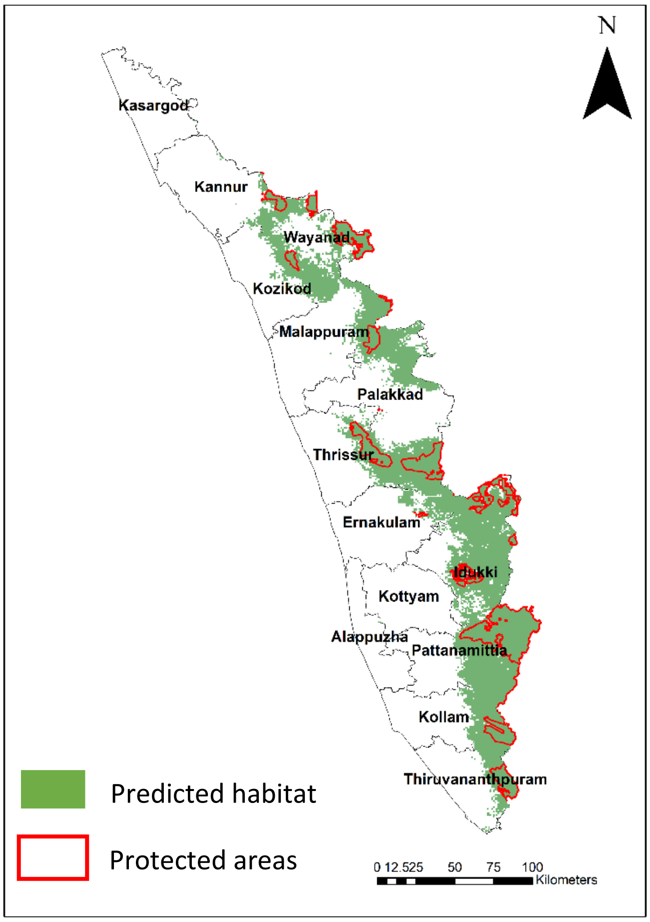

Figure 14: Geospatial distribution of predicted habitats and protected areas across Kerala's districts

Table 7: District wise distribution of predicted habitat and protected areas across Kerala

District Name	Protected Area		Predicted Area	
	Sq. KM	% In Dist. Area	Sq. KM	% In Dist. Area
Idukki	1182.355	26	3392.547	75
Wayanad	427.802	19	1261.685	56
Pattanamittia	208.034	7.7	1304.734	47
Palakkad	403.622	8.8	1731.038	37
Thrissur	241.074	7.8	990.702	32

Kollam	185.226	7.3	794.775	31.5
Thiruvananthpuram	199.745	9	635.634	28
Malappuram	14.097	0	762.138	20
Ernakulam	16.845	0.5	568.297	18
Kozikod	0.0437	0	410.855	16
Kannur	86.690	3	234.179	7
Kottyam	4.408	0	41.346	1.9
Kasargod	0	0	2.5679	0
Alappuzha	0	0	1.6293	0

Tamil Nadu has a predicted suitable habitat area of 9,587 square kilometres

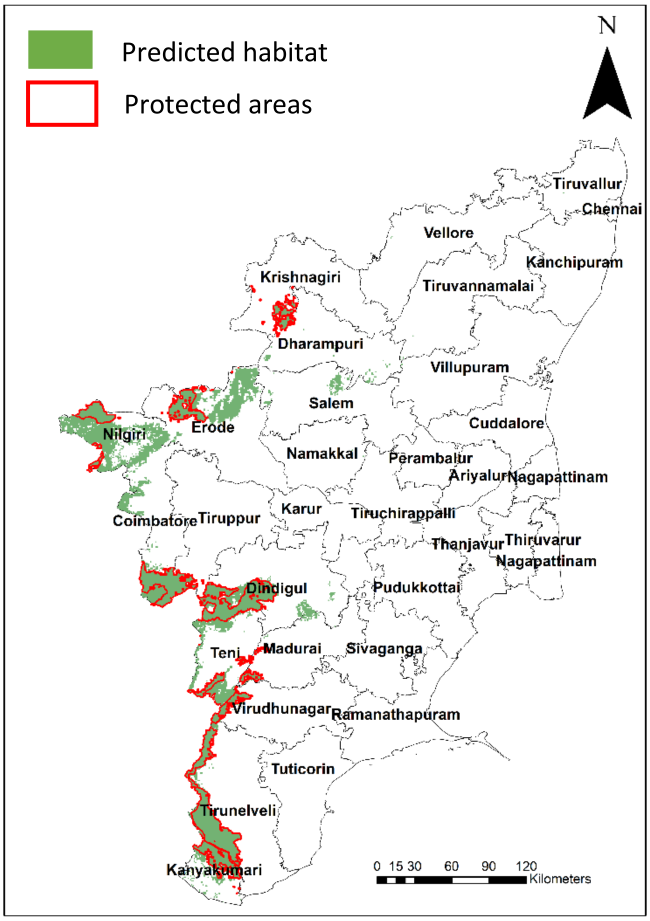

Figure 15: Geospatial distribution of predicted habitats and protected areas across Tamil Nadu's districts

Table 8: District wise distribution of predicted habitat and protected areas across Tamil Nadu

District Name	Protected Area		Predicted Area	
	Sq. KM	% In Dist. Area	Sq. KM	% In Dist. Area
Kanyakumari	513.832	30	724.638	63
Krishnagiri	919.406	17.5	170.264	42
Tirunelveli	1171.176	17	1293.074	28
Nilgiri	440.814	16.8	1689.196	27
Dindigul	988.532	16	1294.342	23
Coimbatore	519.883	13	1134.597	20
Erode	796.016	13	1434.967	18.5
Teni	328.171	11	826.432	5.3
Tiruppur	470.684	8	290.389	4.7
Dharampuri	325.163	7	151.806	3.2
Madurai	257.107	6	196.498	3.1
Virudhunagar	240.838	5.5	232.631	2.5
Salem	0	0	136.573	0
Namakkal	0	0	2.983	0
Tiruvannamalai	0	0	4.919	0
Vellore	0	0	1.989	0
Tiruvallur	0	0	0.994	0
Villupuram	0	0	1.473	0