https://doi.org/10.1051/e3sconf/202671101025

Predictive Analytics and Remote Sensing for Biodiversity Loss Assessment in Urban Green Zones

Dr.G. *Arasuraja* *, *Hayder* Hasan Shatawi [2], *Maha* Sami Barakat [3], *Hussein* Basim Furaijl [4], *Hassan* Khalid Abozibid [5], *Riadh* Abdul Retha Abass [6]

[1]Associate Professor, St. Joseph's Institute of Technology, Chennai, India
[2] Department of Law, College of Law, Sawa University, Almuthana, Iraq
[3] Al-Bayan University, Iraq
[4] college of Pharmacy, University of Al-Ameed, Karbala, Iraq
[5] Al-Zahrawi University College, Karbala, Iraq
[6] Mazaya university college, Dhiqar, Iraq

Abstract. Urban green zones are crucial for maintaining ecological balance and biodiversity, as well as enhancing living standards. Still, growing metropolitan areas and land use alterations undermine biodiversity within these zones. The research creates a remote sensing predictive analytics model to analyze and track biodiversity loss in open spaces and urban parks. The model predicts areas of potential hazard using high-resolution satellite images, vegetation indices, species occurrence data, and machine learning techniques. Temporal analysis reveals ecological patterns and drivers that are anthropogenic, influencing species diversity over time. The model also maintains proactive biodiversity loss warning systems, enabling city planners to prioritize conservation efforts. A case study in a fast-urbanizing urban area also illustrates it, where the model is trained and tested on the multi-temporal satellite-derived imagery and field derived species data, which spatially confirms that the model can sufficiently explain spatial patterns in changes over time in biodiversity-key fluctuations, to capture the landscape-ecological processes. The enhanced resilience of urban ecosystems demonstrates the power of informed policy and management strategies possible with data-driven methodologies.

1 Introduction

Maintaining green areas within cities and urban regions is crucial for biodiversity conservation, as well as for sustaining ecosystem services and the overall well-being of the people living in those regions [1]. These habitats, nonetheless, are facing threats from escalating urbanization, habitat destruction, and increasing human activities. Natural landscapes are being substituted or modified by growing urban development in cities, which erodes the habitats of indigenous species and leads to a loss of biodiversity [2][5]. This loss leads to various ecological problems and diminishes the resilience of cities to changing environments. Although traditional methods of biodiversity monitoring can be applied, they are often inappropriate in an urban environment because they cannot make real-time decisions due to insufficient spatial or temporal resolution. For large- and mid-scale n-diversity analysis, a remote sensing methodology blended with predictive analytics is a suitable option [11]. Remote sensing methods provide continuous, high-resolution data coverage over time, whereas machine learning-based models can detect patterns of spatial biodiversity risk across different ecological indicators, such as vegetation indices and land use features; these methods improve the quality of information, speed and quantity provided, which aids in performing effective and timely biodiversity estimations at a lower cost [10][4]. With satellite imaging, species occurrence data sets, and extraction algorithms that recognize vegetative health indicators, this research constructs a framework that quantifies the loss of biological diversity and identifies priority conservation sites. By enabling proactive identification through wise planning, combined systems recognize hotspots with risk areas. Wise city planning enables simultaneous ecological management for evidence, thus supporting proactive identification-based planning alterations that intensify preparedness. This, in turn, enables wiser zoning and augmented managed targeted ecological risk avoidance zone zoning. These types of tools are crucial for enhancing the sustainability of cities and mitigating the impact of constantly fluctuating environmental conditions.

Key Contribution

- Formulated a combined system of remote sensing, indices of vegetation cover, and species presence data in practical biodiversity loss evaluation in urban green spaces.
- Installed a machine learning-based predictive model that is capable of detecting high-risk areas

---

* Corresponding author: arasuraja.mba@gmail.com

of biodiversity with a high spatial resolution and flexibility.
- Provided a scalable decision-support tool for urban planners to prioritize conservation efforts based on spatial biodiversity risk mapping.

The structure of the article is as follows: the introduction provides the cause and need for measuring biodiversity in cities, and subsequently, the overview explains the latest advances and limitations of the research field of remote sensing and predictive analysis. The methodology outlines the data acquisition, processing, modeling, and system architecture frameworks, while the results and discussion present the interpretation of model performance alongside the evaluation of spatial analysis output, highlighting their significance. Finally, the conclusion encapsulates insights on model applicability and suggests enhancements for prolonged ecological monitoring, framing these suggestions as future work.

2 Literature Review

The most recent studies have placed a greater focus on the role of combining remote sensing and predictive analytics in evaluating biodiversity loss, specifically in urban green zones [12]. Urban expanding areas pose a challenge because traditional field-based methods of monitoring and assessing biodiversity activity are not adequate to deal with the level of complexity that is developing within urban ecosystems [7]. Seto and Ramankutty (2016) [8] highlighted the ecological consequences of urban sprawl and how it has led to a shift towards more concentrated patterns of development, which in turn influence the distribution of species, habitat connectivity, and ecosystem function [9]. They highlight the significance of spatial-temporal data, particularly satellite imagery, in detecting and monitoring ecological changes over time [3]. By demonstrating the relationship between urban structure and environmental impacts, they promote the use of satellite-derived information as part of urban conservation planning strategies. In support of this theory, Müllerová et al. (2017) [6] aimed to use multispectral imaging remotely to detect invasive flora and stressed vegetation more in urban areas than in rural areas. As a result of previous statements, their study ensured that remote sensing is an essential tool that possesses the ability to provide early warning signs for other forms of direct intervention from active biological threats aimed at these systems en masse. UAV-guided surveys involving some form of spectral observation enable the identification of fundamental environmental pressures operating on biodiversity across various scales, which are crucial in modulating how such pressures shape life forms in complex ecosystems with interacting multiple organisms. Such machines are helpful for the adaptive management of biodiversity in urban ecosystems, where environmental conditions change rapidly. However, issues still arise. Both studies identify locations of ecological incorporation with technological paradigms through the fineness of data, species-level sensitivity measures, and calibrations necessitated by ground-truthing.

Additionally, remote sensing data are translated into biodiversity information only after they have been subjected to advanced modeling frameworks. Building on the seminal findings of previous work, this study proposes a real-time predictive system for measuring biodiversity loss using machine learning. The aim is to enhance urban ecological planning by using high-resolution, spatially explicit risk maps derived from remote sensing data and vegetation indices, alongside species occurrence records, to predict potential and existing biomes at differing altitudinal zones.

3 Methodology

This section describes an approach that utilizes remote sensing data and urban predictive analytics to examine biodiversity loss in urban green zones. The methodology includes the collection of relevant datasets, preprocessing, feature extraction, the application of machine learning modeling techniques, and spatial risk mapping. Each step aims to address and anticipate changes in ecological systems and assess the degradation levels of urban biodiversity.

3.1 Data Collection and Preprocessing

In other urban green, high-resolution satellite data are taken and species presence/abundance data obtained by ecological surveys or citizen science projects. The optical satellites used to acquire high resolution satellite imagery were Sentinel 2 MSI at 10 m and Landsat 8/9 OLI/TIRS at 30 m in all accessible cloud free scenes during the study period. All the pictures were corrected to surface reflectance as provider Level 2A products or alternatively, reprojected to a standard UTM coordinate system. The quality assessment bands that were provided with each product were used to mask clouds and cloud shadows and the artefacts that were not eliminated through it were eliminated visually. Records of distribution of the species were derived based on systematic field observations of selected plots, available biodiversity records, and citizen science observations that were approved and georeferenced at a minimum of 10 m position accuracy and linked to a survey date. In order to achieve temporal consistency, only species observations within a specified temporal window (for example ±30 days) of a satellite acquisition were associated with that image date and observation outside this condition were not included in model training.

Based on spectral bands, particular vegetation indices, in particular, the Normalized Difference Vegetation Index (NDVI) are computed:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \tag{1}$$

Where in Eq (1) NIR is virtually infrared reflectance, and RED is red band reflectance. This index is useful to measure the health of vegetation and is also a powerful indicator of ecological quality. Besides NDVI, other auxiliary vegetation indices, including the Soil Adjusted Vegetation Index (SAVI) or Green NDVI (GNDVI) can also be calculated based on conventional equations to offer additional data regarding the vigor of the canopy.

3.1.1 Study Area and Temporal Scope

The case study is about an urban agglomeration structure of a fast-growing metropolitan centre and its surrounding green urban infrastructure spanning a total area of about 100 km 2 of parklands, institutional campuses, street vegetation, and patches of peri urban forests. The spatial reach is characterized with the administrative shapefile of the official administrative boundaries of the particular municipal authority. The timeframe of the analysis will be 2015-2024, where the acquired satellites will be representative months of the dry and wet seasons of each year to reduce cloud contamination and to observe intra-annual change in vegetation condition. This is the time when urban growth is extreme, and it becomes one of the possibilities to observe the conversion of land-use and related biodiversity effects on almost a decade. Consistent time series of vegetation indices and land surface parameters over the entire set of identified urban green zones were produced by producing annual or bi-annual composite images.

3.2 Feature Extraction

The NDVI, Land surface temperature, and habitat fragmentation metrics parameters, as well as the geospatial data, like the elevation and road proximity, are also useful in remote sensing, and they may be applied to generate a spatial grid to be used in the model simulation. The study field was gridded into an even grid of square cells with a spatial resolution of 30 m, which corresponds to or sums the natural pixel size of the coarsest coarse sensor. Summary statistics of spectral and environmental variables were produced at each cell on the grid and time step (the mean and standard deviation of NDVI among other vegetation indices and land surface temperature (LST)). LST was computed as the bands of the thermal infrared of Landsat-8/9 and single-channel algorithm and approximate emissivity is calculated depending on the land-cover type. Habitat fragmentation measures patch size, edge density, proximity to the nearest large green patch were calculated using a land-cover map created by the process of supervised classification of the same image. The Euclidean distance was computed between the bits of the municipal GIS or OpenStreetMap data to calculate the distance to roads and built-up areas. The feature of each grid cell of the predictive models is these variables.

3.3 Predictive Modeling

Having labeled data of the status of biodiversity be it presence/absence or richness score, the supervised machine learning model may be trained using Random Forest and gradient-boosted trees. The biological diversity of the individual grid cells was measured using two complementary variables of response; (i) a binary presence/absence flag on whether at least one target species or a minimum richness of target species was present within that grid cell and within that time step, and (ii) a continuous species richness index on the number of observed species. Classification models were classified by the use of presence/absence labels and regression models were classified by the use of richness scores.

It was split based on the 70/30 ratio and stratified by risk type to maintain the ratio of low and high-risk regions in the data, which were then split into training and testing data. The model development was done in a nested cross validation scheme of tuning hyperparameters to 5-fold cross validation on the training set and an outer test set was held out to test the final performance. Random Forest models were trained using 500 trees, a maximum tree depth of 20, and the square root of the number of features to be considered at a single split, whilst gradient boosted tree models were trained with the learning rate of 0.05, the maximum tree depth of 5 and the early stopping based on the validation loss.

Optimization of these models takes two categories: around MSE (Mean Squared Error) and Accuracy at classification tasks:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (2)$$

In Eq (2), y_i represents the true measure of biodiversity and y_i represents the forecast of biodiversity. In the case of a classification task, the general accuracy, Precision, Recall, F1 score, and the Area Under the ROC Curve (AUC) were calculated as the results of the confusion matrix to evaluate the discrimination of high- and low-risk grid cells. All the models were trained in Python with the scikit learn and gradient boosting libraries and random seeds were kept constant across runs to have the same results.

3.4 Interpretation and Mapping

As urban maps depict biodiversity risk zones, they illustrate the outputs spatially. These maps are crucial in identifying vulnerable areas that require immediate conservation attention. The Biodiversity Vulnerability Score (BVS) for each grid cell was computed as a weighted linear combination of standardized predictor variables and model outputs:

$$\mathrm{BVS}_j = \sum_{k=1}^{K} w_k \ x_{jk} \qquad (3)$$

In Eq (3), x_{jk} denotes the standardized value of predictor k in grid cell j, and w_k represents the relative importance of that predictor derived from Random Forest feature importance scores, normalized to sum to one. Higher BVS indicates a greater risk of biodiversity loss. Continuous BVS values were subsequently classified into Very High, High, Moderate, and Low risk categories using quantile-based thresholds and expert judgement, as summarized in Table 1, to facilitate interpretation by city planners.

3.5 Ground Truth and Validation

Ground truth biodiversity labels were based on field surveys at the plot level in permanent sampling plots that were randomly located within the main types of green space such as parks, institutional campuses, riparian corridors and peri-urban forests. During the study period, at least two visits were conducted to each plot and the presence of species and abundance was determined according to the standard ecological sampling requirements. Spatial observations were then spatially combined with the grid cells that held the plot positions to produce cell related presence/absence and richness values. These ground truth labels were compared with model predictions on the held-out test set and then measures of their performance (accuracy, precision, recall, F1-score, AUC and MSE) were calculated to measure the consensus and to support entirely all assertions in predictive performance and spatial risk mapping.

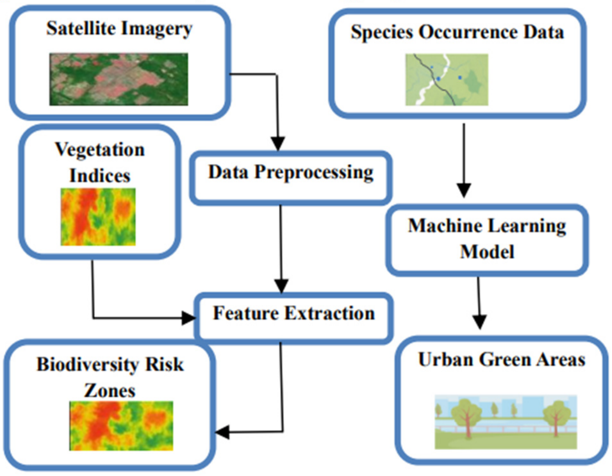

Fig. 1. Methodological Framework for Biodiversity Loss Assessment Using Remote Sensing and Predictive Analytics in Urban Green Zones

Figure 1 illustrates the entire methodology pipeline, spanning from satellite imagery and species data to risk mapping, which depicts the spatial integration of data sources, model building, and biodiversity evaluation. The image illustrates the flow of data through different stages, which is essential for reproducing the systematized framework.

4 Results and Discussion

This part of the study focuses on assessing the efficiency of the urban green zone predictive analytics framework in evaluating biodiversity loss. Our focus of analysis was an agglomerated area with a rapid urban development and mosaic of disconnected green areas. The findings are based on both spatial and statistical machine-learning model results derived from preprocessed datasets.

4.1 Spatial Biodiversity Risk Prediction

The model's performance on the independent test set was notably strong, with an average classification accuracy of 88% and a precision rate of 84% for predicting high-risk versus low-risk biodiversity cells, based on field-derived ground truth labels. To spatially quantify the risk, a vulnerability score (BVS) about biodiversity was computed for every grid cell:

$$BVS = \frac{w_1.(1-NDVI)+w_2.Fragmentation\ Index+w_3.Proximity\ t}{w_1+w_2+w_3} \quad (4)$$

In Eq (4), where w_1, w_2, w_3 are weights determined via feature importance ranking. Higher BVS indicates a greater risk of biodiversity loss.

4.2 Tabular Analysis of Biodiversity Risk Zones

Table 1. Biodiversity Risk Categories Based on BVS Thresholds

Risk Category	BVS Range	Area Coverage (km²)	Dominant Habitat Type
Very High Risk	> 0.75	12.3	Isolated Patches
High Risk	0.50–0.75	28.6	Fragmented Urban Forest
Moderate Risk	0.25–0.50	35.1	Parklands and Corridors
Low Risk	< 0.25	21.8	Well-connected Green Spaces

Based on the model, approximately 40.9 km² within the region has a high to very high biodiversity risk. Such areas mainly comprise fragmented or encroached habitats that require immediate conservation action, as highlighted in Table 1.

4.3 Model Evaluation Performance

The performance on the held-out test set improves steadily with the size of training data, reaching its maximum beyond 80% and F1 -score of 90% overall accuracy and remote sensing indicators, indicating the sufficientness of the selected features and remote sensing indicators with respect to biodiversity risk evaluation as shown in Figure 2.

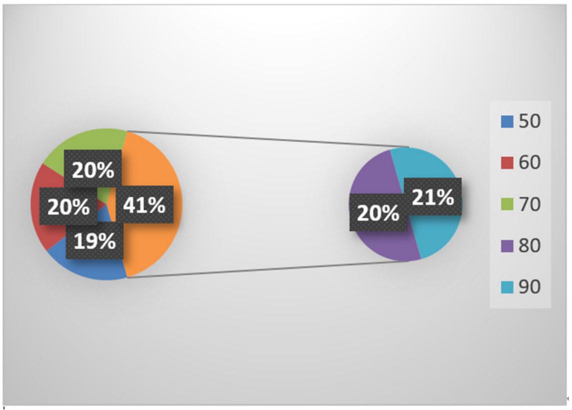

Fig. 2. Model Performance Metrics Over Different Training Sizes

5 Conclusion

This research developed an all-encompassing and scalable framework that incorporates remote sensing, vegetation indices, species occurrence data, and predictive analytics to assess urban green biodiversity loss zones. This framework, utilizing satellite images, enables the extraction of advanced ecological features through the computation of vegetation health indicators, such as

NDVI. Ecosystem features are complemented and enriched with various machine learning models to provide precise predictive spatial mapping of biodiversity risk zones, supporting data-driven urban planning. These model outputs verified its robustness in identifying high-risk regions within mostly fragmented and isolated green patches, showing homogeneous performance regardless of variations in training set size. Furthermore, the use of a quantifying parameter called the biodiversity vulnerability score (BVS) provides more value by equally smoothly integrating ecological risk assessment and enabling priority setting for conservation action. The processed data, feature representatives, and model configuration files can be shared with the help of data-sharing agreements to provide external replication and expansion of the suggested framework.

References

1. R.I. McDonald, A.V. Mansur, F. Ascensão, M. Colbert, K. Crossman, T. Elmqvist, ... W. Zipperer, Identify research gaps in understanding the impact of urban growth on biodiversity. *Nat. Sustain.* **3**(1), 16–24 (2020). https://doi.org/10.1038/s41893-019-0436-6
2. F.G. Nyman, E. Lahti, The role of social innovation in addressing urban economic disparities. *Int. Acad. J. Innov. Res.* **12**(2), 1–6 (2025). https://doi.org/10.71086/IAJIR/V12I2/IAJIR1210
3. G. Papadopoulos, M. Christodoulou, Design and development of data-driven intelligent predictive maintenance for predictive maintenance. *Assoc. J. Interdiscip. Tech. Eng. Mech.* 2(2), 10–18 (2024).
4. D. Rocchini, D.S. Boyd, J.B. Féret, G.M. Foody, K.S. He, A. Lausch, ... M. Wegmann, Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sens. Ecol. Conserv.* **2**(1), 25–36 (2016). https://doi.org/10.1002/rse2.9
5. A.H. Larijani, Sustainable urban development: concepts, features, and indicators. *Int. Acad. J. Sci. Eng.* **3**(1), 208–213 (2016).
6. J. Müllerová, T. Bartaloš, J. Brůna, P. Dvořák, M. Vítková, Timing is important: Unmanned aircraft vs. satellite imagery in plant invasion monitoring. *Front. Plant Sci.* **8,** 887 (2017). https://doi.org/10.3389/fpls.2017.00887
7. A. Ahadi, S.F. Azar, The study of the possibility of using activity-based costing system (ABC). *Int. Acad. J. Account. Financ. Manag.* **3**(1), 1–8 (2016).
8. K.C. Seto, N. Ramankutty, Hidden linkages between urbanization and food systems. *Science* **352**(6288), 943–945 (2016). https://doi.org/10.1126/science.aaf7439
9. V.K. Yogamadhavan, G. Mannayee, Evaluation of performance of different optimizers of convolutional neural network in the classification of images of urban domestic solid waste. *J. Internet Serv. Inf. Secur.* **14**(3), 42–77 (2024). https://doi.org/10.58346/JISIS.2024.I3.004
10. A.M. Noori, Spatiotemporal land cover transformation assessment of Kirkuk City using remote sensing data. *Int. J. Adv. Eng. Emerg. Technol.* **14**(1), 8–15 (2023).
11. N. Pettorelli, W.F. Laurance, T.G. O'Brien, M. Wegmann, H. Nagendra, W. Turner, Satellite remote sensing for applied ecologists: opportunities and challenges. *J. Appl. Ecol.* **51**(4), 839–848 (2014). https://doi.org/10.1111/1365-2664.12261
12. A.R. Rajalakshmi, S. Jayakani, H. Shanmugam, P.R. Ramakrishnan, A study on green entrepreneurship and electronic word of mouth: Special reference with Kanchipuram District. *Indian J. Inf. Sources Serv.* 14(3), 270–276 (2024). https://doi.org/10.51983/ijiss-2024.14.3.35