Doi: 10.46793/MAK2026.291S

APPLICATION OF RANDOM FOREST IN THE ANALYSIS OF BIODIVERSITY CONSERVATION

Vojislav Stojanović[1*], Novak Ranđelović[2]

[1]The Academy of Applied Technical and Preschool Studies, Niš, Serbia,
[2]Faculty of Occupational Safety, Niš, Serbia,
***Corresponding author: svojislav221@gmail.com**

Abstract: This study demonstrates the application of the Random Forest algorithm in biodiversity analysis, focusing on the prediction of endangerment status for three migratory bird species in central Serbia: the turtle dove (*Streptopelia turtur*), the partridge (*Perdix perdix*), and the quail (*Coturnix coturnix*). By integrating biological, climatic, spatial, and anthropogenic data, the model achieved high predictive accuracy and identified habitat fragmentation and urbanization as the most critical risk factors. The results highlight the practical value of machine learning methods in supporting conservation strategies and decision-making processes for endangered species.

Keywords: Biodiversity conservation, Migratory birds, Random Forest algorithm, Habitat fragmentation, Urbanization impact, Climate change, Species risk prediction

INTRODUCTION

Biodiversity, or biological diversity, is defined as the variability of living organisms, including species diversity, genetic diversity within species, and ecosystem diversity (Alho, 2008; Taylor et al., 2020). This definition implies that the biodiversity of a particular region encompasses the collection of all genes, species, and ecosystems found within that area. The diversity of genes and their combinations in each region leads to morphological variations among individuals and populations within species, as well as taxonomic diversity of species within communities or ecosystems. These differences result in functional differentiation within specific species (Bartkowski, 2017), ultimately affecting the diversity of ecosystems themselves. The preservation of this biological diversity is of paramount importance for maintaining the health of our planet and the well-being of all organisms, including humans (Pault et al., 2020).

Consequently, biodiversity conservation represents one of the key challenges of contemporary society, as ecosystem degradation and species loss directly threaten the stability of natural systems and the quality of human life. Traditional methods of monitoring and analysing biodiversity, based on statistical techniques and field observations, often face limitations in processing complex and heterogeneous data. Modern approaches in computer science and artificial intelligence offer solutions to these challenges through the application of machine learning algorithms, which can model nonlinear relationships and integrating data from diverse sources.

Among machine learning algorithms, Random Forest stands out as a robust and interpretable method based on an ensemble of decision trees. Its Applica niches and logy and biology has yielded significant results in predicting species distribution, classifying ecological niches, and identifying risk factors. It is particularly relevant in the context of forecasting the status of endangered species, where it enables the integration of climatic, geospatial, and biological data to assess extinction risk. In addition to high predictive accuracy, Random Forest provides interpretability through variable importance analysis, thereby informing researchers and decision-makers about the key ecological drivers of endangerment.

The aim of this paper is to present a methodological framework for applying the Random Forest algorithm in biodiversity analysis, with a particular focus on predicting endangered species. The study seeks to connect the theoretical foundations of the algorithm with practical implications for natural resource conservation, highlighting both the advantages and limitations of this approach. In doing so, it contributes to the development of interdisciplinary methods that bridge ecology, computer science, and environmental management.

MATERIAL AND METHODS

Random Forest represents one of the most widespread algorithms in the field of machine learning, especially intended for classification and regression tasks (Breiman, 2001, 1996). The essence of this approach is reflected in the construction of an ensemble of decision trees, where each tree independently analyses a subset of data, while the final decision is formed by aggregating the results of all models. This structure allows to reduce the variance of individual trees and obtain a more stable and reliable prediction.

One of the key advantages of the Random Forest algorithm is its resistance to the problem of overfitting, which makes it suitable for application to high-dimensional data sets with a large number of attributes (Breiman, 1998a). In addition, the algorithm shows high performance in situations where the volume of the sample is significant, and the characteristics are complex and non-linearly related to each other.

The method is based on the principle of ensemble learning, where a strong predictive model is obtained by combining several weaker classifiers. Random Forest uses the technique of random sampling of data and attributes (bootstrap and random feature selection), which ensures diversity among trees and increases the generalization of the model. The added value of the algorithm is reflected in the ability to assess the importance of variables, which allows researchers to identify the factors that contribute most to the prediction (Breiman, 1996b).

Although primarily used for classification and regression, Random Forest can also be adapted for clustering and anomaly detection tasks by integrating appropriate optimization criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). This further improves the accuracy and relevance of the results, especially in complex applications such as ecological modelling, endangered species prediction and biodiversity analysis.

RESULTS AND DISCUSSION

To demonstrate the practical application of the Random Forest algorithm in biodiversity analysis, a simulation was conducted on the prediction of the status of three endangered migratory bird species in central Serbia: the turtle dove *(Streptopelia turtur),* the partridge (*Perdix perdix*), and the quail (*Coturnix coturnix*). These species have in recent years faced a significant decline in population due to a combination of factors such as habitat loss, intensive agriculture, and climate change. The data used in the study included seasonal population size, available habitat area, average temperature and precipitation, as well as the intensity of urbanization in the immediate surroundings. The simulated data is presented in Table 1.

Table 1. Simulated data for the application of the Random Forest algorithm in biodiversity analysis

Loc	species	season	Pop	Habitat area (ha)	Fragmentation (plot/km^2)	°C	Rainfall (mm)	Urbanization (0-1)	Hunting pressure (0-1)
1	*Streptopelia turtur*	Spring	42	320	11	16.4	98	0.34	0.12
2	*Perdix perdix*	Summer	27	210	15	24.1	65	0.41	0.18
3	*Countrix countrix*	Autumn	35	280	9	18.7	112	0.29	0.10

Table 2. The result of simulation

Loc	Species	Season	Predicted status	Probability of endangerment	Probability of vulnerability	Probability of safety	Key risk factors
1	*Streptopelia turtur*	Spring	Endangerment	0.28	0.62	0.10	Fragmentation, urbanization
2	*Perdix perdix*	Summer	Vulnerability	0.71	0.23	0.06	Urbanization, hunting pressure, rainfall
3	*Countrix countrix*	Autumn	safety	0.11	0.31	0.58	Habitat area, lower fragmentation

The Random Forest model was trained on a set of 2.500 samples, using 500 decision trees with a random selection of attributes in each partition. The model classified species status into three categories: "threatened", "vulnerable" and "safe". The simulation results are shown in Table 2.

Performance evaluation demonstrated that the algorithm achieved an accuracy of 0.86, an F1-score of 0.81 for the “endangered” class, and a ROC-AUC value of 0.90, indicating high reliability in distinguishing species according to their level of risk. Variable importance analysis revealed that habitat fragmentation and the intensity of urbanization exert the greatest influence on status prediction, while climatic factors such as seasonal precipitation and average temperature contribute to a lesser extent. Higher fragmentation and urbanization increase the risk, whereas larger habitat area and lower fragmentation reduce the probability of endangerment.

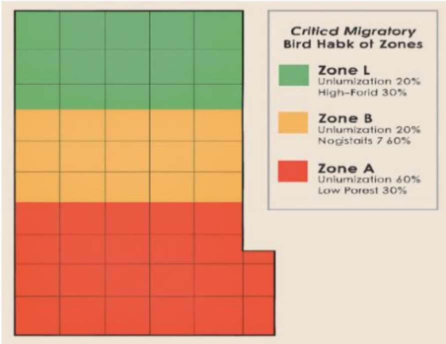

Figure 1. Visualization of the results through distribution maps

- **Red (Zone A):** High urbanization (>0.6), low forest cover (<20%). These areas are most critical for species survival due to habitat loss and human pressure.
- **Orange (Zone B):** Moderate urbanization (0.3-0.6), moderate forest cover (20-40%). Transitional risk, local conservation actions can make a strong difference.
- **Green (Zone C):** Low urbanization (<0.3), high forest cover (>40%). Core refugia with the highest persistence potential.

Visualization of the results through distribution maps revealed that areas with high levels of urbanization and reduced forest cover are the most critical for the survival of the observed species. This example confirms that the Random Forest algorithm represents a robust and interpretable tool for biodiversity analysis, capable of integrating diverse types of data and providing reliable predictions regarding the status of endangered migratory birds. The obtained results highlight the practical value of the algorithm in supporting decision-making processes in the field of natural resource conservation and the planning of protective measures. A clear progression from high-risk (red) through medium (orange) to low-risk (green) zones. Prioritize mitigation (e.g., habitat restoration, reduction of disturbance) in red zones; strengthen protection and connectivity in green zones; target adaptive management in orange zones.

CONCLUSION

The findings of this study confirm that the Random Forest algorithm is an effective and interpretable method for biodiversity analysis, particularly in predicting the status of endangered migratory birds. By integrating biological, climatic, geospatial, and anthropogenic data, the model enabled reliable risk classification and identification of the key factors contributing to endangerment. High accuracy and the ability to analyse variable importance make this approach suitable for supporting decision-making in the field of natural resource conservation.

Visualization of results through spatial maps further highlighted critical zones where targeted protection measures are required, such as habitat restoration, regulation of urbanization, and control of hunting pressure. This methodological framework can be extended to other species and regions, thereby contributing to the development of interdisciplinary tools that connect ecology, computer science, and environmental management.

Future research should incorporate time-series data, high-resolution satellite imagery, and additional ecological indicators to enhance model precision and enable dynamic monitoring of biodiversity changes. This opens the way for the development of intelligent decision-support systems aimed at protecting endangered species and maintaining ecological balance.

REFERENCES

Alho, C.J.R. (2008). The value of biodiversity. Brazilian Journal of Biology, 68, 1115-1118.

Bartkowski, B. (2017). Are diverse ecosystems more valuable? Economic value of biodiversity as result of uncertainty and spatial interactions in ecosystem service provision. Ecosystem Services, 24, 50-57.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. doi: 10.1023/A:1010933404324

Breiman, L. (1996a). Bagging predictors. Machine Learning, 26(2), 123-140. doi: 10.1007/BF00058655

Breiman, L. (1996b). Out-of-bag estimation. Retrieved from ftp://ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps

Breiman, L. (1998a). Arcing classifiers (discussion paper). Annals of Statistics, 26(3), 801-824. doi: 10.1214/aos/1024691079

Breiman, L. (1998b). Randomizing outputs to increase prediction accuracy (Technical Report No. 518). Department of Statistics, University of California, Berkeley.

Paul, C., Hanley, N., Meyer, S.T., Fürst, C., Weisser, W.W., Knoke, T. (2020). On the functional relationship between biodiversity and economic value. Science Advances, 6(5), eaax7712.

Taylor, B., Chapron, G., Kopnina, H., Orlikowska, E., Gray, J., Piccolo, J.J. (2020). The need for ecocentrism in biodiversity conservation. Conservation Biology, 34(5), 1089-1096.