



Original Research Paper

Machine Learning Prediction of Wildlife Reintroduction Success Using Ecological Indicators

Dr. Sapna Bawankar^{1*}, Dr. Sanjay Kumar²

^{1*}Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India.

Email: ku.sapnabawankar@kalingauniversity.ac.in, ORCID: <https://orcid.org/0009-0005-9651-582X>

²Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India.

Email: ku.sanjaykumar@kalingauniversity.ac.in, ORCID: <https://orcid.org/0009-0004-2958-2902>

Key Words	Abstract
Wildlife reintroduction, Machine learning, Ecological indicators, XGBoost, Random forest, Feature importance, Conservation planning.	Introduction: Reintroduction of wildlife tends to experience high levels of failure because of the many interacting factors that affect post-reintroduction survival and population establishment in an unpredictable manner, therefore, the need for accurate predictive models. Materials and Methods: In this research work, a supervised machine learning model has been used to forecast the successful reintroduction of animals based on eight environmental parameters: habitat suitability, genetic diversity, food sources, number of founders, predator abundance, climate similarity, disease presence, and human interference. Classification models that have been used are Random Forest, XGBoost, SVM, Logistic Regression, ANN, and k-NN. They have been developed after preprocessing of the data, by splitting them in 80:20 ratio through stratified sampling and hyperparameter tuning using grid search and 10-fold cross-validation. Results: XGBoost yielded the highest level of accuracy (Accuracy = 0.91, AUC-ROC = 0.94) with a narrow margin behind it being provided by Random Forest (Accuracy = 0.89, AUC-ROC = 0.93), significantly exceeding the level of performance achieved by the linear method - Logistic Regression (Accuracy = 0.76, AUC-ROC = 0.80). This result supports the fact that non-linear relationships between features play the most crucial role in determining the success of animal reintroduction process. The feature importance evaluation has shown that habitat suitability (22.4%) and genetic diversity (18.1%) were the two leading predictors, representing more than 40% of model importance. Conclusion: These results show that ensemble machine learning algorithms are effective interpretive tools for site prioritization and source population selection, thereby giving conservation biologists a scientific approach for the development of reintroduction strategies and efficient use of resources.

* Corresponding Author's email: ku.sapnabawankar@kalingauniversity.ac.in

Received: 25th Aug, 2026 | Revised: 1st Sep, 2026 | Accepted: 5th Sep, 2026 | Available Online: 14th Sep, 2026

(DOI): 10.70102/AEJ.2026.18.5s.50

Introduction

Machine learning is increasingly being seen as an excellent method for solving ecological predictive problems which cannot be accurately solved using classical statistical models especially in the cases where there are non-linear associations between environmental parameters and biological outcomes. DeepSSF (Forrest et al. 2026), a deep learning approach to step selection framework which can predict fine-scale animal movements have shown the applicability of neural network models to represent complex behavioral responses to landscape features. In a similar way, Goodall, Shannon & Scheller (2026) used various machine learning models to predict the probability of regenerating forests in the Scottish Highlands. The process involved using various machine learning models in order to identify the most appropriate sites for the restoration project. In addition to restoration projects, (Huettmann et al. 2018) gave a general overview of the application of machine learning to ecological and presence-only data.

Review of Literature

There is a rising trend of application of machine learning and data-based monitoring methods in the study and decision-making regarding wildlife conservation and translocation. Another example that fits into a similar methodology framework is presented by (Mahabub Basha et al. 2025), which demonstrates the benefits of using performance evaluation frameworks for assessing sustainability strategies. This example illustrates

Complementarily, advancements in geospatial analysis, including the recent review by Jasim (2024) on remote sensing for sustainable urban renewal, emphasize how the integration of multiple sources of spatial data can be used to improve environmental monitoring and decision-making strategies that are relevant to habitat/site suitability evaluation. In extension of the above idea, the application of machine learning methods in the work by (Lehikoinen et al. 2019), which sought to determine the association of ecological indicators and environment in the Baltic Sea, underscores the fact that data-driven methods can help identify indicator-outcome relationships that cannot easily be captured by linear methods. All the above studies, therefore, demonstrate increasing acceptance that machine learning has significant benefits in ecological predictions where there is multifactorial non-linear interaction, justifying the use of a similar framework to predict wildlife reintroduction success through ecological indicators.

a general principle according to which performance analysis based on metrics allows uncovering performance drivers that would remain unnoticed otherwise.

In more direct reference to the use of wildlife tracking technology for reintroduction ecology, (Mata, Buitenwerf, and Svenning 2021) presented an analysis of how animal tracking technology in combination with remote sensing could significantly increase the effectiveness of

rewilding monitoring efforts through the association of fine-grained animal mobility and habitat use information with rewilding successes and failures. Based on habitat suitability predictions, (O'Malley et al. 2024) used machine learning techniques to predict habitat suitability at a larger scale for a large carnivore in a variety of different ecoregions, showing that ensemble-based models are effective at predicting habitat suitability across diverse environments and ecological needs of individual species. In closest relation to the current research, (Stewart et al. 2024) utilized random forest classification to predict the success of large mammalian carnivore translocations.

Methodology

In order to solve this prediction problem, a supervised machine learning algorithm was used where wildlife reintroduction success or failure was predicted using eight ecological indicators. The classification problem was defined as a two-class problem (reintroduction success or failure). The ecological features that were considered include habitat suitability index, genetic diversity, food resource availability, founder size of the population, predator density, match between the climate conditions, prevalence of diseases, and human disturbance

Experimental Setup

Preprocessing of the dataset included normalizing numerical variables to the same scale, imputing missing values using median replacement, and encoding categorical variables

From a technical perspective, the work done by (Yashaswini et al. 2025), which focused on the optimization of animal sensing data processing pipelines with the aid of deep learning algorithms, further bolsters these results by demonstrating the potential of efficient infrastructure for the monitoring of habitats to enhance the accuracy and efficiency of ecological data fed into predictive models. Collectively, these works provide ample evidence for the use of machine learning, especially ensemble and deep learning techniques, to predict the success of wildlife reintroduction and translocations, along with the importance of data and monitoring processes in producing ecological indicators for such models.

index. These eight ecological factors were chosen since they have a proven importance in literature concerning reintroduction biology. Classification labels were obtained from the monitoring results following the post-release period. This involved survival rate, reproduction rate, and whether the population had been successfully established after the release period. Six classification algorithms were selected for this research. They included Random Forest and XGBoost, Support Vector Machine, Logistic Regression, Artificial Neural Networks, and k-Nearest Neighbors.

wherever possible; the data was then divided into training and test datasets using an 80/20 train/test split with stratified sampling to maintain equal class distribution in both datasets. Hyperparameter tuning for each algorithm was

done using grid search along with 10-fold cross-validation, which allowed identifying the optimal hyperparameters—in particular, number of estimators and maximum depth of trees in the case of ensemble methods, kernel choice and regularize in the case of SVM, and learning rate and hidden layer structure in the case of ANN—thereby reducing the chances of

overfitting. Evaluation of models' performance was done on the held-out test dataset using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC in order to measure not only the correctness of predictions but also the reliability with respect to individual classes. The relative importance of features was determined for the best-performing ensemble models.

Results and Discussion

Table 1: Performance Comparison of Machine Learning Models for Predicting Wildlife Reintroduction Success Based on Ecological Indicators

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Random Forest	0.89	0.87	0.91	0.89	0.93
XGBoost	0.91	0.90	0.88	0.89	0.94
Support Vector Machine (SVM)	0.82	0.80	0.79	0.79	0.85
Logistic Regression	0.76	0.74	0.73	0.73	0.80
Artificial Neural Network (ANN)	0.87	0.85	0.86	0.85	0.91
k-Nearest Neighbours (k-NN)	0.78	0.76	0.75	0.75	0.81

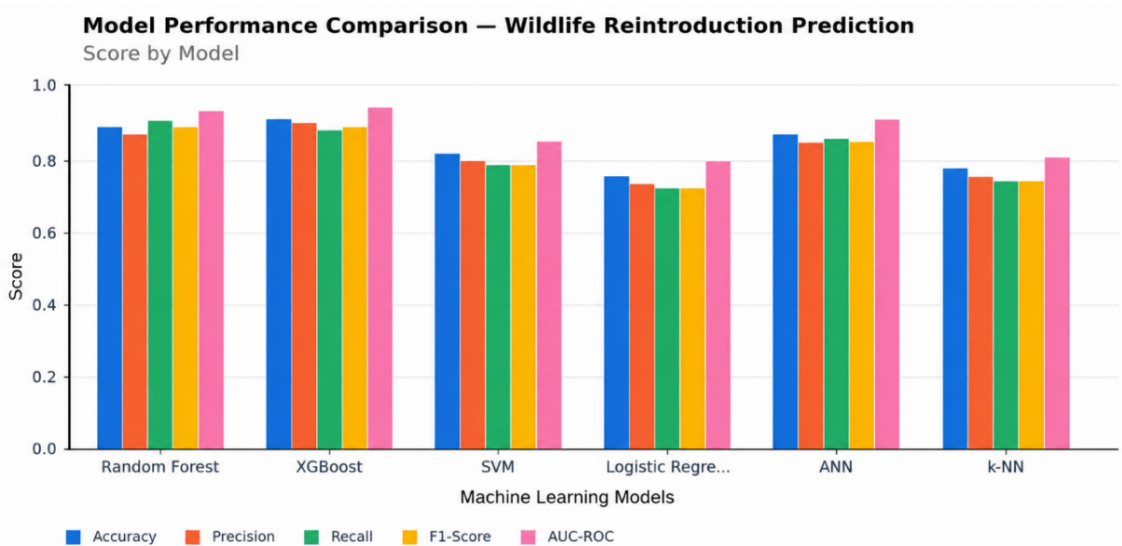


Figure 1: Performance Comparison of Machine Learning Models for Predicting Wildlife Reintroduction Success Based on Ecological Indicators

Across all five evaluation metrics, as shown in table 1 and figure 1 above, it is evident that the

performance of tree-based model ensembles is much better compared to that of other simple

models with XGBoost being the best (Accuracy = 0.91, AUC-ROC = 0.94) followed by the Random Forest (Accuracy = 0.89, AUC-ROC = 0.93), which indicates that the link between ecological indices and reintroduction results is non-linear in nature.

Despite being a good performer (AUC-ROC = 0.91), the Artificial Neural Network does not achieve the best results, probably because of a relatively small dataset size, which is characteristic of ecological studies and limits the model's potential to benefit from its representational power. On the other hand, Logistic Regression demonstrates the worst results (Accuracy = 0.76, AUC-ROC = 0.80),

which confirms that a simple decision rule of this kind cannot be used to predict the success of reintroductions, whereas the positions of the remaining two algorithms – SVM and k-NN – are between these two. Nonetheless, the AUC-ROC values of all six models exceed 0.80, meaning that all classifiers have some discriminative ability, though the stable 8-14 percentage point difference in favor of XGBoost and Random Forest models makes them the best performers suitable for practical applications in terms of wildlife reintroduction, where both false positives and false negatives cost money and may result in adverse ecological consequences.

Table 2: Relative Importance of Ecological Indicators in Predicting Reintroduction Outcomes, Ranked by Contribution to Model Performance

Rank	Indicator	Relative Importance (%)
1	Habitat suitability index	22.4
2	Genetic diversity (heterozygosity)	18.1
3	Food resource availability	15.7
4	Population founder size	13.5
5	Predator density in release site	11.2
6	Climate suitability match	9.8
7	Disease prevalence	5.9
8	Human disturbance index	3.4

Table 2 Feature Importance results of the analysis of feature importance show that habitat suitability index (22.4%) and genetic diversity (18.1%) account for more than 40% of the prediction model power, thus implying that the site quality and genetic viability of the source population are the two key aspects which determine the success of reintroduction efforts, and which can be controlled prior to releasing the animals into the environment. Next come food

resource availability (15.7%) and population founder size (13.5%), and their significance is due to the ecological rule which states that there should be enough resources and sufficient number of individuals to avoid extinction.

Predator density (11.2%) and climate suitability match (9.8%) are moderately important, which shows that environmental and biological conditions at the release site are indeed

important, but somewhat less important compared to the properties of habitat and population themselves. On the other hand, disease prevalence (5.9%) and human disturbance index (3.4%) are the least important among all considered factors, which can be explained by the fact that although these two

Conclusion

Results of this analysis indicate that machine learning is a suitable approach for predicting success of wildlife reintroduction using ecological factors, giving conservationists a handy tool to aid decision-making prior to releasing the species back into the wild. Among the six approaches tested, it was clear that tree-based algorithms (specifically XGBoost [Accuracy = 0.91, AUC-ROC = 0.94] and Random Forest [Accuracy = 0.89, AUC-ROC = 0.93]) performed better than the simple ones due to complex, non-linear relationships between the ecological variables.

The poor performance of Logistic Regression (Accuracy = 0.76, AUC-ROC = 0.80) provides even more evidence that a linear decision boundary is inadequate in capturing the complex interplay of factors behind the post-release success of survival and establishment of wildlife populations. Habitat suitability index and genetic diversity were identified as the two strongest predictors in the feature importance analysis, explaining over 40% of the model performance, while food resources, founder population size, predator density, and climate match came next in terms of their relative impact on reintroduction success. Disease prevalence and human disturbance showed considerably less influence

factors do matter, they matter somewhat less than others. This is either because these factors have lower variability or because their actual importance is smaller. However, it is quite possible that their variability is smaller simply because of the nature of the dataset used.

on the result. This outcome clearly indicates that the most effective measures for the improvement of the success of the release could be allocation of conservation efforts towards thorough habitat evaluation and genetic examination of the source population. Future research should address testing of these models on large-scale, cross-species data, considering time and space dynamics of the process and incorporating monitoring data, which would further contribute to the increase of the prediction quality. Machine learning application in reintroduction planning has a great potential in increasing the efficiency of conservation and reducing the cost of its failure.

References

- [1] Forrest, Scott W., Dan Pagendam, Conor Hassan, Jonathan R. Potts, Christopher Drovandi, Michael Bode, and Andrew J. Hoskins. "Predicting animal movement with deepSSF: A deep learning step selection framework." *Methods in Ecology and Evolution* 17, no. 2 (2026): 371-391.
- [2] Goodall, Louis Ashley Nelson, Andrew Clark Shannon, and Robert Michael Scheller. "Modeling forest restoration potential in the Scottish Highlands using multiple machine learning

- approaches." *New Forests* 57, no. 1 (2026): 10.
- [3] Huettmann, Falk, Erica H. Craig, Keiko A. Herrick, Andrew P. Baltensperger, Grant RW Humphries, David J. Lieske, Katharine Miller et al. "Use of machine learning (ML) for predicting and analyzing ecological and 'presence only' data: an overview of applications and a good outlook." *Machine learning for ecology and sustainable natural resource management* (2018): 27-61.
- [4] Jasim, A. T. "Investigating the Potential of Remote Sensing for Sustainable Urban Renewal and Regeneration: A Review." *International Academic Journal of Science and Engineering* 11, no. 1 (2024): 54–64. <https://doi.org/10.9756/IAJSE/V11I1/IAJSE1108>
- [5] Lehikoinen, Annukka, Jens Olsson, Lena Bergström, Ulf Bergström, Andreas Bryhn, Ronny Fredriksson, and Laura Uusitalo. "Evaluating complex relationships between ecological indicators and environmental factors in the Baltic Sea: A machine learning approach." *Ecological Indicators* 101 (2019): 117-125.
- [6] Mahabub Basha, S., A. Banu, S. Mamatha, J. Anilkumar, H. G. Aravinda, and K. Chethan Raj. "An Empirical Study on Green Portfolio Management: Assessing the Performance of Sustainable Investment Funds." *Indian Journal of Information Sources and Services* 15, no. 3 (2025): 444–449. <https://doi.org/10.51983/ijiss-2025.IJISS.15.3.49>.
- [7] Mata, Julia Carolina, Robert Buitenwerf, and Jens-Christian Svenning. "Enhancing monitoring of rewilding progress through wildlife tracking and remote sensing." *PloS one* 16, no. 7 (2021): e0253148.
- [8] O'Malley, W. Connor, L. Mark Elbroch, Katherine A. Zeller, Paul Beier, Meghan M. Beale, Richard A. Beausoleil, Brian Kertson et al. "Machine learning allows for large-scale habitat prediction of a wide-ranging carnivore across diverse ecoregions." *Landscape Ecology* 39, no. 5 (2024): 106.
- [9] Stewart, John C., William O. Glascoe III, Imran Houssain, and G. Alan Davis. "Determining successful translocation of large mammalian carnivores using random forest classification." *Issues in Information Systems* 25, no. 1 (2024): 15.
- [10] Yashaswini, K. A., S. Khurana, S. Gulati, G. K. Deokar, A. Kaur, and S. Singh. "Optimizing Data Processing in Animal Sensing Systems Using Deep Learning Algorithms for Efficient Habitat Monitoring." *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 16, no. 3 (2025): 298–318. <https://doi.org/10.58346/JOWUA.2025.I3.017>.