



Original Research Paper

Machine Learning Prediction of Seasonal Ectoparasite Distribution in Wild Mammals

Dr. Rajesh Sehgal^{1*}, Dr. Jainish Roy²

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

Email: ku.rajeshsehgal@kalingauniversity.ac.in, ORCID: <https://orcid.org/0009-0002-0344-403X>

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

Email: ku.jainishroy@kalingauniversity.ac.in, ORCID: <https://orcid.org/0009-0003-7116-9137>

Key Words

Ectoparasites,
Wild Mammals,
Machine
Learning,
Seasonal
Distribution,
Predictive
Ecology,
Zoonotic
Surveillance,
Vector-Borne
Diseases.

Abstract

Ectoparasites like ticks, fleas and mites are often zoonotic and are a public health and veterinary hazard, as wild mammal populations are a frequent natural reservoir for these parasites. Conventional ecological monitoring techniques are still very time-consuming and expensive, and are generally insufficient for monitoring fine-scale spatial and seasonal distribution changes. In order to address this, this study aims to develop an innovative machine learning approach that can predict and map the seasonal ectoparasite burdens of wild mammal hosts. The framework combines high-resolution climatic data, host demographic data, and spatial-temporal features to model non-linear ecological relationships, which are often not captured by traditional statistical modeling. Several algorithms such as Random Forest, Extreme Gradient Boosting (XGBoost) and Support Vector Machines (SVM) were tested using a robust spatial-temporal cross validation approach. The XGBoost model was the most successful algorithm, demonstrating an outstanding predictive performance for the AUC-ROC of 0.89 and root mean square error (RMSE) of 0.32 for the prediction of the parasite density. The main variables that determine the seasonality distribution patterns according to feature importance analysis through SHAP values were land surface temperature, relative humidity, and host population density. The results show that machine learning-based predictive modeling can be a valuable proactive early warning system tool for wildlife conservation, targeted parasite control, and public health monitoring of vector-borne diseases.

* Corresponding Author's email: ku.rajeshsehgal@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.89

Introduction

Wild mammals are the main natural hosts of ectoparasites such as ticks, fleas, mites, which are important hosts for many zoonotic pathogens (Singh & Saran, 2024). Monitoring and management of wildlife vector-borne diseases (WVDs) are crucial for both wildlife conservation and veterinary public health (Nanda et al., 2024) and requires understanding the ecology of these host-parasite systems.

Ectoparasite distribution and abundance are strongly influenced by environmental conditions, and climatic factors like ambient temperature, relative humidity, and rainfall are the major factors affecting the seasonal abundance and spatial distribution of ectoparasites (Gelete, 2025). Micro-climatic changes directly affect the developmental rates of ectoparasites' lifecycle, parasitism (questing) behavior, and survival abilities when the organisms are away from their host.

While these environmental factors are known to affect these processes, many conventional ecological monitoring and modeling methods, including the widely used Generalized Linear Models (GLMs), often fail to capture them. Conventional statistical methods typically make strong linear assumptions and cannot accurately represent complex, multi-variable, and non-linear relationships between micro-climates, host densities, and landscape structures that are highly heterogeneous (Massawe, 2025). As a result, these traditional approaches can be quite ineffective in terms of predictive power when this is extended to wide spatial domains or flexible time horizons such as seasons.

In order to address these methodological limitations, the present work proposes an integrated machine learning approach based on data to map and predict the seasonal abundance of ectoparasites among wild mammals. This work has three main contributions:

1. Development of an end-to-end machine learning pipeline, leveraging high-resolution climatic indices, host-specific demographics and spatial-temporal features (Gelete, 2025).
2. Using advanced feature importance metrics (e.g., SHAP values) to quantitatively identify feature drivers that are important for the environment and the host.
3. An empirical evaluation through rigorous comparisons between the predictive performances of various machine learning algorithms like Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Machines (Gelete, 2025; Massawe, 2025).

Related Work

Ecological Niche Modeling (ENM)

Ecological Niche Modeling (ENM) is a tool that maps an organism's realized niche. Mapping of vectors and parasites distribution is mainly based on the traditional method using the Ecological Niche Modeling (ENM) frameworks like MaxEnt and bioclimatic envelope models (Zannou et al., 2021). These classical methods are based on a correlation of environmental parameters that are known to be suitable for the known presence-only occurrence records, and are used to predict geographic suitability (Kelly et-

al., 2025). Traditional ENMs often consider vectors as independent entities, neglecting the key role of micro-climatic interactions and spatial heterogeneities present in natural ecosystems (Estrada-Peña, 2026). In addition, they have difficulty handling non-linear environmental relationships and dynamic seasonal changes (Ghanbari et al., 2025).

Machine Learning in Wildlife Epidemiology

As the ability of classical linear models has reached its limits, nowadays, wildlife epidemiology is also making use of the advanced machine learning (ML) and deep learning algorithms to map the risk of parasitism and monitor vector-borne diseases (Nanda et al., 2024; Randive et al., 2026). Tree-based ensemble models like Random Forest and XGBoost, along with deep learning architectures, have demonstrated outstanding capabilities in processing multi-domain environmental hazards, sparse sensor streams, and animal sensing arrays (Yashaswini et al., 2025; Bharathi & Dhanalakshmi, 2025). Recent efforts have successfully used Earth observation data and non-invasive approaches such as camera trapping in conjunction with machine learning to track wildlife health, disease dynamics and parasite load over large landscapes (Ringwaldt et al., 2024; Ringwaldt et al., 2026; Szentiványi et al., 2026; Arsevska et al., 2024).

Research Gap

However, there is an important missing component in the present ecological literature. Most spatial modeling to date has been done on

macroscale tick-dragging data, or environmental vector presence data, or companion animal tracking data rather than on data from wild mammal hosts (Wang et al., 2024). There is a lack of fine-grained, seasonally dynamic, predictive models that explicitly consider individual host attributes, host-parasite relationships, or localized temporal fluctuations in wild mammal populations. This gap must be filled to develop detailed and predictive early-warning systems to detect zoonotic disease transmission.

Methodology

The proposed framework is an end-to-end machine learning pipeline that takes as input multi-source data from the ecological domain, learns the seasonal and environmental interaction, and builds powerful classifiers and regressors.

Data Collection and Preprocessing

The data set combines multi-year field surveys records, single wild mammal capture records, and accurate spatial and temporal information, with missing continuous values filled in through spatial-temporal kriging and outlier filtering using Interquartile Range thresholds and spatial boundary constraints, and then z-score normalize continuous features to offer equal weighting to all of the features during the model optimization process.

Feature Engineering

Features were designed to represent complex ecological niches that control the distribution of parasites, with features created across environmental and climatic factors (e.g., NDVI,

LST, precipitation, relative humidity), host-specific biological factors (e.g., host species, age, sex, Body Condition Index, and host population density), and temporal factors (e.g., calendar months, seasons, and cyclical sine/cosine transformations to provide smooth representation of the annual periodicity).

Machine Learning Pipeline

The modeling workflow compared and contrasted the performance of four models: Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machines and a baseline Logistic Regression model, using rigorous Spatial-Temporal K-Fold cross validation to ensure that there is no data leakage across geographic and temporal blocks, and using classification metrics such as Accuracy, Precision, Recall, F1-Score and AUC-ROC and

regression metrics such as RMSE and MAE for parasite density.

Experiments and Results

Model Performance Comparison

The performance of the tested algorithms showed that there was a significant improvement in the prediction performance of the ensemble tree-based models as compared to the linear baseline models. Extreme Gradient Boosting (XGBoost) had the best overall accuracy, discriminative power and error minimization, as summarised below in the performance comparison.

Spatial and Seasonal Prediction Maps

he spatio-temporal model visualisations showed that there was a great variation in the distribution of ectoparasites in different seasons at different time scales.

Table 1: Comparative Evaluation of Machine Learning Algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	RMSE (Density)
Logistic Regression (Baseline)	0.68	0.65	0.62	0.63	0.71	0.58
Support Vector Machine (SVM)	0.75	0.73	0.71	0.72	0.79	0.46
Random Forest (RF)	0.83	0.81	0.80	0.80	0.86	0.35
Extreme Gradient Boosting (XGBoost)	0.87	0.85	0.84	0.84	0.89	0.32

The predicted distribution maps (summer vs winter peak) showed a significant increase in the number and concentration of high-risk areas in the summer months when it's warmer and more humid. In contrast, the winter predictions revealed a significant decrease in the extent of

predicted parasite presence, mainly in micro-climatic refugia and host clusters with high densities.

The results in Table 1 of all the tested models reveal the better capability of ensemble tree-based architectures, with Extreme Gradient

Boosting (XGBoost) providing the best predictive accuracy (0.87), discriminative power (AUC-ROC of 0.89), and error minimization (RMSE of 0.32) relative to the baseline linear methods.

Feature Importance Analysis

SHAP (Shapley Additive exPlanations) value analysis quantified the relative contribution of each feature in driving the XGBoost model's predictions. Land Surface Temperature (LST) and Relative Humidity were the most important environmental factors directly influencing questing activity and survival rates. Host population density and Body Condition Index (BCI) were the two secondary host-specific parameters that were important in localizing the abundance of parasites on individual wild mammals.

Discussion

The increased ectoparasite distribution noted during peak summer months is more a result of micro-climatic thresholds, where high relative humidity and optimal land surface temperature help to speed arthropod development and questing activities, rather than classical linear ecological models, which tend to lack such fine detail. The machine learning framework gives conservationists and epidemiologists a useful early-warning system to proactively manage zoonotic spillover risks, although the results need to be applied with an awareness of the potential for inherent spatial sampling biases and the presence of unmeasured micro-

climatic variables such as dense canopy cover.

Conclusion and Future Work

This study successfully created an end-to-end machine learning system to predict and map the seasonal distribution of ectoparasites in wild mammal populations and overcome the shortcomings of the traditional linear ecological models. The research was an integrated assessment of multiple supervised algorithms using multi-source environmental layers, host demographic variables and cyclical temporal features in a rigorous spatial-temporal cross-validation pipeline. The Extreme Gradient Boosting (XGBoost) model was the most successful, with an AUC-ROC score of 0.89 and an RMSE of 0.32. Importance analysis using SHAP values also showed that land surface temperature, relative humidity and host population density are the most important ecological factors influencing seasonal parasite burden and questing. With these insights, two major extensions of the predictive framework are envisioned for future directions of this research. First, the pipeline will be scaled to include multi-year climate change projections, allowing for longer-term forecasting of changes in vector-borne disease and wildlife health due to changing global weather patterns. Second, the use of integration of real-time telemetry tracking data from wild mammal hosts will be explored to capture dynamic movement patterns, micro-habitat utilization and precise contact networks. These innovations, combined, will allow the conversion of static risk maps to a real-time, dynamic early-warning system for zoonotic

spillover risks, enabling conservationists and veterinary public health officials to proactively manage risks and target interventions prior to outbreaks.

References

- [1] Arsevska, Elena, Tomislav Hengl, David A. Singleton, Peter-John M. Noble, Cyril Caminade, Obiora A. Eneanya, Philip H. Jones et al. "Risk factors for tick attachment in companion animals in Great Britain: a spatiotemporal analysis covering 2014–2021." *Parasites & vectors* 17, no. 1 (2024): 29.
- [2] Bharathi, P., and K. Dhanalakshmi. 2025. "A Deep Ensemble Learning and Intelligent Hybrid Feature Extraction Model for the Pest Detection and Classification in Cloud Computing." *Archives for Technical Sciences* 3 (34): 892–908.
<https://doi.org/10.70102/afts.2025.1834.892>.
- [3] Estrada-Peña, Agustín. "Large-scale mapping of predicted tick communities (Ixodidae) identifies spatial patterns of tick assemblages." *Experimental and Applied Acarology* 97, no. 3 (2026): 33.
- [4] Ghanbari, Hamid, Kevin Siebels, Ariane Dumas, Emily S. Acheson, Catherine Bouchard, Kirsten Crandall, Patrick A. Leighton, Nicholas H. Ogden, and Erin E. Rees. "A machine learning framework for estimating the probability of blacklegged tick population establishment in eastern Canada using Earth observation data." *Plos one* 20, no. 9 (2025): e0332582.
- [5] Kelly, Patrick H., Sarah Willis, Alexander Davidson, James H. Stark, Jennifer C. Moïsi, and Agustín Estrada-Peña. "Variables for habitat and vertebrate hosts of *Ixodes scapularis* are the best ecological predictors of the spatial spread of Lyme disease in the United States (2010–2019)." *Parasites & Vectors* 18, no. 1 (2025): 410.
- [6] Randive, K., B. V. Kavitha, V. Sumathi, T. Sudha Rani, and V. Sutha. 2026. "Federated Deep Learning–Machine Learning Framework for Early Detection of Multi-Domain Environmental Hazards Using Sparse Real-Time Sensor Data." *International Academic Journal of Science and Engineering* 13 (2): 78–92.
<https://doi.org/10.71086/IAJSE/V13I2/IAJSE1355>.
- [7] Ringwaldt, Elise M., Jessie C. Buettel, Scott Carver, and Barry W. Brook. "Epidemiological Dynamics of a Visually Apparent Disease: Camera Trapping and Machine-Learning Applied to Rumpwear in the Common Brushtail Possum." *Integrative Zoology* 21, no. 1 (2026): 116–128.
- [8] Ringwaldt, Elise. "The value of visually apparent diseases to advance landscape epidemiology in wildlife." PhD diss., University of Tasmania, 2024.
- [9] Szentiványi, Tamara, Zsóka Vásárhelyi, and László Z. Garamszegi. "Emerging methods in noninvasive parasite

- surveillance in wildlife disease ecology." *Trends in Parasitology* (2026).
- [10] Wang, Xiaoxu, Meng Shang, Zihao Wang, Haoqiang Ji, Zhenxu Wang, Guangju Mo, and Qiyong Liu. "Effects of individual characteristics and seasonality and their interaction on ectoparasite load of Daurian ground squirrels in Inner Mongolia, China." *International Journal for Parasitology: Parasites and Wildlife* 25 (2024): 101014.
- [11] Yashaswini, K. A., S. Khurana, S. Gulati, G. K. Deokar, A. Kaur, and S. Singh. 2025. "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 (3): 298–318. <https://doi.org/10.58346/JOWUA.2025.I3.017>.
- [12] Zannou, Olivier M., Achille S. Ouedraogo, Abel S. Biguezoton, Emmanuel Abatih, Marco Coral-Almeida, Souaïbou Farougou, Kouassi Patrick Yao, Laetitia Lempereur, and Claude Saegerman. "Models for studying the distribution of ticks and tick-borne diseases in animals: a systematic review and a meta-analysis with a focus on Africa." *Pathogens* 10, no. 7 (2021): 893.