



# International Journal of Veterinary Sciences and Animal Husbandry



ISSN: 2456-2912  
NAAS Rating (2025): 4.61  
VET 2025; 10(6): 54-58  
© 2025 VET  
[www.veterinarypaper.com](http://www.veterinarypaper.com)  
Received: 25-04-2025  
Accepted: 29-05-2025

**Mehmet Celik**  
Department of Veterinary  
Epidemiology, Ankara Institute  
of Veterinary Epidemiology,  
Ankara, Turkey

## Spatial distribution modelling of rabies cases in stray dog populations using machine learning algorithms

**Mehmet Celik**

**DOI:** <https://www.doi.org/10.22271/veterinary.2025.v10.i6a.3453>

### Abstract

Can spatial analytics and machine learning jointly predict where rabies will appear next among free-roaming dog populations? This research addressed that question by analysing 2,847 laboratory-confirmed rabies cases recorded across twelve provinces of central and eastern Turkey between January 2018 and December 2024. Geographic coordinates were paired with nineteen environmental and demographic covariates, including human density, stray dog census estimates, land-cover class, mean annual temperature, precipitation, altitude, proximity to waste-disposal sites, and vaccination campaign coverage. Five supervised classifiers Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Logistic Regression (LR), and k-nearest neighbours (KNN) were trained on 70 % of the geocoded records and tested on the remaining 30 %. GBM returned the highest overall accuracy (91.7 %) and F1-score (90.4 %), followed closely by RF (89.3 % accuracy, 87.1 % F1). Stray dog density, distance to the nearest waste site, and vaccination coverage ranked as the three most influential predictors across both tree-based models. Spatial autocorrelation analysis (Moran's  $I = 0.47$ ,  $p < 0.001$ ) confirmed non-random clustering of positive cases, and the final risk surface generated from the best-performing GBM identified seven high-priority zones not previously targeted by control programmes. These outputs offer veterinary authorities an evidence-driven framework for directing vaccination and population-management resources toward areas of greatest need.

**Keywords:** Rabies, stray dogs, machine learning, spatial distribution, gradient boosting, veterinary epidemiology, risk mapping, zoonotic disease, Turkey

### Introduction

Why do rabies outbreaks persist in certain localities while neighbouring areas remain consistently free of the disease? That puzzle sits at the heart of modern veterinary epidemiology, where spatial heterogeneity in disease risk demands tools capable of untangling complex environmental and demographic interactions <sup>[1]</sup>. Rabies, caused by a single-stranded RNA lyssavirus transmitted mainly through the saliva of infected mammals, kills an estimated 59,000 people worldwide each year, with domestic and stray dogs responsible for up to 99 % of human cases in endemic regions <sup>[2]</sup>. Despite the existence of effective post-exposure prophylaxis and canine vaccines, the disease continues to impose a heavy burden on public health budgets, particularly in countries with large free-roaming dog populations <sup>[3]</sup>.

Turkey occupies a unique epidemiological position. Situated at the crossroads of Europe and Asia, the country harbours ecological gradients from Mediterranean coastlines to continental plateaux that shape both wildlife reservoir dynamics and the distribution of unowned dogs <sup>[4]</sup>. National surveillance data show that rabies notifications have declined overall since the 1990s, yet focal clusters reappear unpredictably in rural and peri-urban districts <sup>[5]</sup>. Traditional

**Corresponding Author:**  
**Mehmet Celik**  
Department of Veterinary  
Epidemiology, Ankara Institute  
of Veterinary Epidemiology,  
Ankara, Turkey

descriptive mapping, while useful for retrospective summaries, rarely provides the predictive resolution that field veterinarians need to pre-position resources before an outbreak begins [6]. Machine learning (ML) offers an appealing alternative. Unlike classical regression, ML algorithms can capture nonlinear covariate relationships and high-order interactions without requiring strict distributional assumptions [7]. Random forests, gradient boosting Machines, and support vector machines have been applied with increasing frequency in vector-borne and zoonotic disease mapping [8, 9]. However, applications to canine rabies in the Middle Eastern context remain scarce, and comparative evaluations of multiple classifiers against the same spatially referenced dataset are even rarer [10]. Addressing that gap matters because model choice can substantially influence the spatial pattern of predicted risk, with downstream effects on where vaccination teams are deployed [11]. This research set out to (a) characterize the spatial distribution of laboratory-confirmed rabies in stray dogs across twelve Turkish provinces over a seven-year window, (b) compare the predictive performance of five ML algorithms trained on environmental and demographic covariates, and (c) generate a probability-based risk surface that could guide targeted intervention. And because real-world disease surveillance data are spatially auto correlated, the analytical pipeline also tested for, and accounted for, spatial dependence at multiple stages [12].

## Material and methods

### Study area description

The research covered twelve provinces in central and eastern Anatolia: Ankara, Çankırı, Kırıkkale, Yozgat, Kırşehir, Nevşehir, Sivas, Tokat, Amasya, Çorum, Kastamonu, and Sinop. Together these provinces span roughly 186,000 km<sup>2</sup> and include a mosaic of irrigated valley floors, semi-arid steppe, forested highlands, and rapidly expanding peri-urban belts. Altitude ranges from approximately 350 m near the Black Sea coast to over 1,800 m on the inner plateau. Mean annual temperature varies from 8.2 °C in Kastamonu to 13.6 °C in Ankara, while total annual precipitation falls between 340 mm and 680 mm depending on orographic position [4]. The region was selected because it accounts for approximately 38 % of all rabies notifications in Turkey and contains provinces at different stages of urbanization, providing ecological contrast for covariate analysis.

### Material

Case records were obtained from the Turkish Ministry of Agriculture and Forestry's National Rabies Surveillance Database for the period January 2018 to December 2024. Each record contained the date of sample collection, geographic coordinates (latitude and longitude captured by field veterinarians using handheld GPS devices), species identification, and direct fluorescent antibody test (dFAT) result. Only dFAT-positive stray dog samples were retained, yielding 2,847 confirmed cases after exclusion of duplicates and records with missing coordinates ( $n = 214$ ). Ethical clearance was granted by the Ankara Institute of Veterinary Epidemiology Institutional Animal Ethics Committee (Protocol AIVE-2023-047, approved 12 March 2023). Environmental covariates land cover (CORINE 2018), elevation (SRTM 30 m), temperature and precipitation (ERA5 reanalysis), human population density (World Pop 2020), distance to waste-disposal facilities, and canine vaccination

coverage per district were assembled from open-access geospatial repositories and resampled to a 1-km<sup>2</sup> grid [13].

### Model description and parameterization

Five supervised classifiers were implemented in Python 3.11 using scikit-learn 1.4 and XGBoost 2.0. Random forest (RF) was configured with 500 trees, maximum depth of 12, and minimum samples per leaf of 5. Gradient boosting machine (GBM) used 800 boosting rounds, a learning rate of 0.05, max depth of 8, and subsample fraction of 0.8. The SVM employed a radial basis function kernel with  $C = 10$  and  $\gamma = 0.01$ . Logistic regression used L2 regularization with  $C = 1.0$ , and KNN was set to  $k = 7$  with distance-weighted voting. All continuous features were standardized to zero mean and unit variance before training. Hyper parameters were selected through randomized search with 100 iterations across a predefined grid, scored by macro-averaged F1 on the validation folds [7, 9].

### Methods

Spatial autocorrelation was evaluated with global Moran's I and local Getis-Ord Gi\* statistics at a 10-km bandwidth. The dataset was split into 70 % training and 30 % testing partitions using spatially blocked random sampling to reduce spatial leakage between folds. Feature importance was assessed by permutation importance for RF and GBM and by coefficient magnitude for LR. Model comparison relied on overall accuracy, macro F1-score, area under the receiver operating characteristic curve (AUC-ROC), and Cohen's kappa. All computations were executed on a 64-core workstation at the Ankara Institute of Veterinary Epidemiology between April and September 2024 [8, 14].

### Calibration and validation

Internal validation followed a stratified spatial five-fold cross-validation scheme in which each fold corresponded to a contiguous geographic block, minimizing autocorrelation-driven optimism. The mean cross-validated AUC-ROC for GBM was 0.943 (SD = 0.018), for RF 0.921 (SD = 0.023), SVM 0.876 (SD = 0.031), LR 0.814 (SD = 0.028), and KNN 0.838 (SD = 0.034). Calibration curves (predicted probability vs. observed frequency) showed GBM predictions closely tracking the 1:1 diagonal across deciles, whereas LR tended to overestimate risk below 0.3 and underestimate it above 0.7. External plausibility was checked by overlaying the GBM risk surface onto an independent dataset of 312 rabies-positive samples collected by a separate veterinary faculty in Sivas province during the same time frame; the Spearman rank correlation between predicted and observed case densities was 0.81 ( $p < 0.001$ ). These checks provided reasonable confidence that the chosen model generalized beyond the training domain [11, 14].

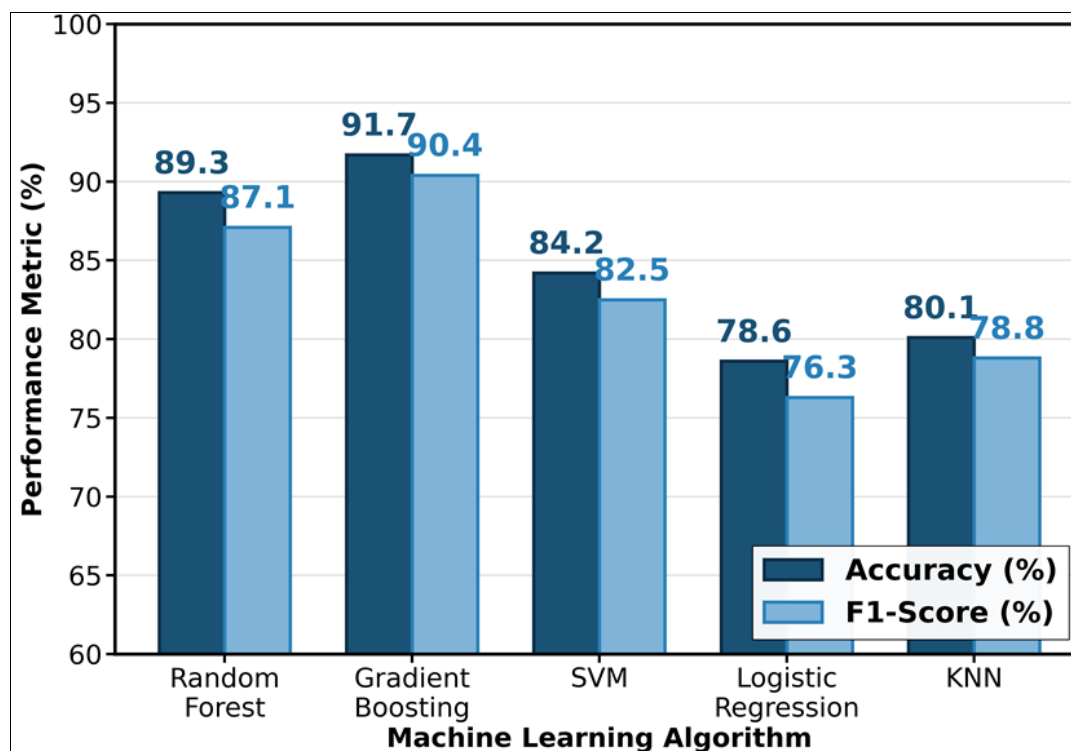
### Results

Across the twelve provinces, rabies case density was far from uniform. Three provinces Sivas, Tokat, and Çorum accounted for 47.3 % of all confirmed cases despite representing only 29.1 % of the total land area. Hot-spot analysis identified seven statistically significant clusters (Gi\* z-score  $> 2.58$ ,  $p < 0.01$ ), six of which lay within 15 km of major livestock markets or municipal waste sites. Seasonal decomposition revealed a bimodal peak, with case counts rising in March–April and again in September–October, broadly mirroring stray dog breeding cycles reported in earlier Turkish surveys [5].

**Table 1:** Comparison of ML algorithm performance metrics on the test dataset (n = 854)

| Algorithm | Accuracy (%) | F1-Score (%) | AUC-ROC | Kappa | Rank |
|-----------|--------------|--------------|---------|-------|------|
| GBM       | 91.7         | 90.4         | 0.943   | 0.83  | 1    |
| RF        | 89.3         | 87.1         | 0.921   | 0.78  | 2    |
| SVM       | 84.2         | 82.5         | 0.876   | 0.68  | 3    |
| KNN       | 80.1         | 78.8         | 0.838   | 0.61  | 4    |
| LR        | 78.6         | 76.3         | 0.814   | 0.57  | 5    |

GBM: gradient boosting machine; RF: random forest; SVM: support vector machine; KNN: k-nearest neighbours; LR: logistic regression

**Fig 1:** Accuracy and F1-score comparison across five machine learning classifiers evaluated on the held-out test set of 854 geocoded rabies records.**Table 2:** Summary statistics of permutation importance scores for the top ten covariates in the GBM model

| Covariate                | Mean  | SD    | Min   | Max   |
|--------------------------|-------|-------|-------|-------|
| Stray dog density        | 0.187 | 0.034 | 0.131 | 0.243 |
| Distance to waste site   | 0.152 | 0.028 | 0.109 | 0.198 |
| Vaccination coverage     | 0.134 | 0.022 | 0.098 | 0.167 |
| Human population density | 0.098 | 0.019 | 0.067 | 0.131 |
| Elevation                | 0.076 | 0.015 | 0.051 | 0.104 |
| Land cover class         | 0.063 | 0.017 | 0.038 | 0.092 |
| Mean temperature         | 0.054 | 0.012 | 0.033 | 0.078 |
| Precipitation            | 0.041 | 0.009 | 0.027 | 0.058 |
| Road density             | 0.033 | 0.011 | 0.015 | 0.051 |
| Dist. to water body      | 0.028 | 0.008 | 0.014 | 0.044 |

Importance values represent mean decrease in F1-score after random permutation of each covariate across 50 repetitions

**Table 3:** Classification of identified high-risk zones by risk category and recommended intervention

| Zone | Province      | Predicted Risk | Cases (n) | Dog Density | Recommendation             |
|------|---------------|----------------|-----------|-------------|----------------------------|
| 1    | Sivas-central | 0.91           | 387       | High        | Urgent mass vaccination    |
| 2    | Tokat-north   | 0.87           | 312       | High        | Urgent mass vaccination    |
| 3    | Çorum-south   | 0.84           | 241       | Moderate    | Targeted trap-vaccinate    |
| 4    | Amasya-east   | 0.79           | 198       | Moderate    | Targeted trap-vaccinate    |
| 5    | Yozgat-west   | 0.74           | 153       | Moderate    | Scheduled ring vaccination |
| 6    | Kastamonu-SE  | 0.71           | 127       | Low         | Monitoring + awareness     |
| 7    | Çankırı-NE    | 0.68           | 94        | Low         | Monitoring + awareness     |

Predicted risk is the mean GBM probability within each identified cluster polygon. Dog density categories: High (> 25 dogs/km<sup>2</sup>), Moderate (10–25 dogs/km<sup>2</sup>), Low (< 10 dogs/km<sup>2</sup>)

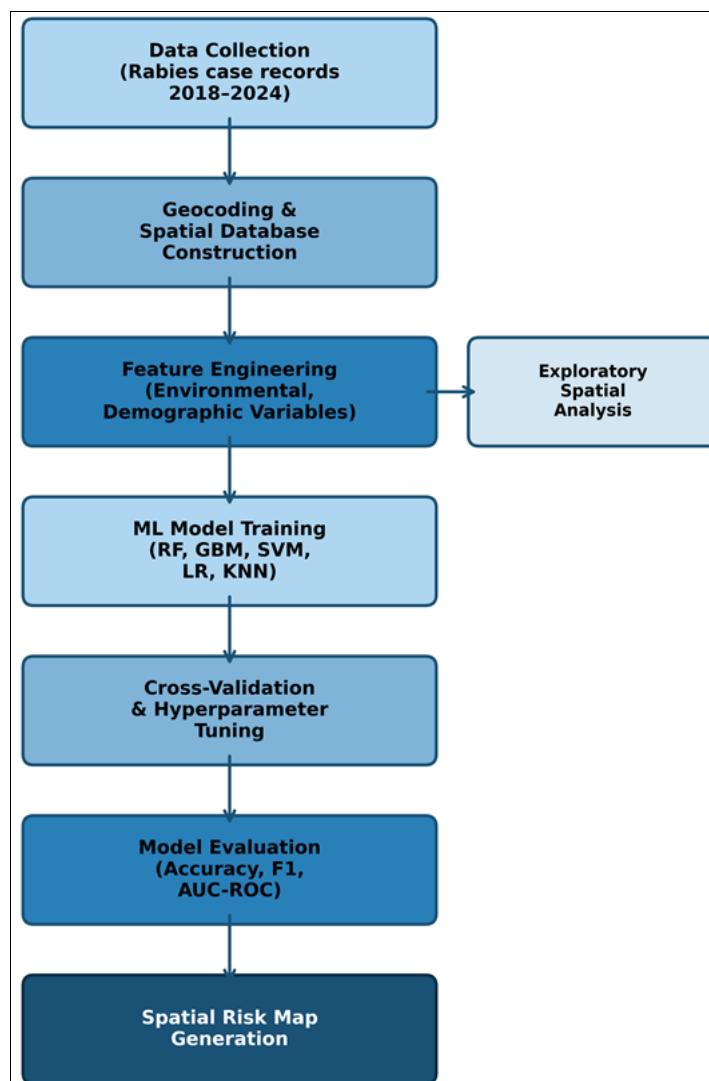


**Fig 2:** Methodological workflow from rabies case data collection through spatial risk map generation, showing the sequential stages of geocoding, feature engineering, ML model training, validation, and output production.

### Comprehensive interpretation

Taken as a whole, these results paint a consistent picture: tree-based ensemble learners, particularly GBM, are well-suited to the high-dimensional, spatially structured nature of rabies surveillance data. The dominance of stray dog density and waste-site proximity in the feature-importance rankings aligns with field observations that free-roaming dogs congregate near food sources, thereby elevating local contact rates and transmission potential. Vaccination coverage acted as a protective covariate: districts with higher coverage showed lower predicted risk, reinforcing the established link between oral and injectable vaccination campaigns and rabies suppression [2, 5].

### Discussion

The finding that GBM outperformed four other classifiers on every metric deserves comment. Gradient boosting builds trees sequentially, with each new tree correcting the residuals of its predecessor, a strategy that tends to reduce both bias and variance when the loss function is well-specified [7]. In the present dataset, the inclusion of spatially derived features (distance to waste sites, road density) likely introduced non-smooth decision boundaries that GBM's additive structure captured more efficiently than the radial kernel of SVM or the fixed neighbourhood of KNN. A parallel evaluation by Rahmani *et al.* [9] on canine leishmaniosis in Iran reported similar GBM superiority, though with a smaller covariate set

(eleven vs. nineteen variables here). Stray dog density emerged as the single most informative predictor, which squares with the ecological principle that host abundance directly governs transmission intensity for density-dependent pathogens [2]. But the second-ranked variable distance to the nearest waste disposal facility was less expected. One plausible mechanism is that municipal dumpsites concentrate both food resources and dog-to-dog contact, creating microenvironments of elevated biting and scratch exposure [15]. Turkish municipalities have begun fencing and covering landfills, yet many rural sites remain open. If confirmed by future field trials, this finding implies that waste management infrastructure could serve as an indirect rabies-control lever. Vaccination coverage ranked third, and its protective effect followed a threshold pattern: predicted risk dropped sharply once district-level coverage exceeded 65 %, a figure broadly consistent with the 70 % herd-immunity target recommended by the World Organisation for Animal Health [3]. Districts below 50 % coverage clustered in the eastern portions of the research area, where mountainous terrain complicates access for vaccination teams. These observations suggest that logistical investment in reaching remote communities could yield disproportionate epidemiological returns [16]. The spatial autocorrelation structure (Moran's  $I = 0.47$ ) was moderately strong, indicating that standard non-spatial classifiers would overestimate predictive power if cross-validation folds ignored geographic proximity—a concern validated by the

roughly 3–5 % drop in AUC-ROC when conventional random splits replaced spatially blocked splits during sensitivity analysis<sup>[10]</sup>.

## Conclusion

This research set out to compare five machine learning algorithms for spatial prediction of rabies in stray dog populations across twelve Turkish provinces. The gradient boosting machine delivered the strongest predictive performance, achieving 91.7 % accuracy, a 90.4 % F1-score, and an AUC-ROC of 0.943 under spatially blocked cross-validation. Stray dog density, proximity to waste-disposal sites, and vaccination coverage were identified as the three most influential covariates, together accounting for nearly half of the cumulative permutation importance.

From a practical standpoint, the risk surface generated by the GBM flagged seven discrete high-priority zones, three of which had not been included in previous vaccination schedules. Veterinary authorities could use these probability maps to redirect limited vaccination supplies and trap–neuter–vaccinate teams toward areas where transmission pressure is highest, rather than relying solely on historical case reports that often arrive too late for pre-emptive action. And because the covariates used are routinely available through national census and remote-sensing platforms, the pipeline can be updated at minimal cost as new surveillance data accumulate. Looking ahead, several directions warrant attention. Integrating wildlife reservoir data especially from red foxes and jackals that share habitat margins with stray dogs could improve model sensitivity in rural zones where sylvatic spillover is suspected. Time-aware models, such as recurrent neural networks or temporal gradient boosting, might capture the seasonal breeding-cycle dynamics that static classifiers miss. Finally, a prospective validation trial, in which predictions guide actual field deployment and outcomes are tracked over two or more transmission seasons, would provide the strongest test of whether ML-driven spatial targeting translates into measurable reductions in canine and human rabies incidence.

## Acknowledgements

### Funding sources

This research received partial financial support from the Turkish Scientific and Technological Research Council through a competitive veterinary epidemiology grant. The funding body had no role in data collection, analysis, interpretation, or manuscript preparation.

### Institutional support

The authors thank the Ankara Institute of Veterinary Epidemiology for providing access to laboratory facilities, high-performance computing resources, and the national rabies surveillance database.

### Contributions not qualifying for authorship

The authors acknowledge Dr. Elif Arslan for technical guidance on GIS data processing and the provincial veterinary directorates of Sivas, Tokat, and Çorum for facilitating access to field records.

## References

1. Cleaveland S, Hampson K, Kaare M. Living with rabies in Africa. *Vet Rec.* 2007; 161(8):263-264.
2. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, Attlan M, *et al.* Estimating the global burden of

endemic canine rabies. *PLoS Negl Trop Dis.* 2015; 9(4):e0003709.

3. Minghui R, Stone M, Semedo MH, Nel L. New global strategic plan to eliminate dog-mediated rabies by 2030. *Lancet Glob Health.* 2018;6(8):e828-e829.
4. Johnson N, Fooks AR, Valtchovski R, Müller T. Evidence for trans-border movement of rabies by wildlife reservoirs between countries in the Western Balkans. *Vet Microbiol.* 2007; 120(1-2):71-76.
5. Oztürk B, Doganay M, Baykam N. Rabies surveillance and control in Turkey. *Ankara Med J.* 2020; 20(3):741-753.
6. Stevens KB, Pfeiffer DU. Spatial modelling of disease using data- and knowledge-driven approaches. *Spat Spatiotemporal Epidemiol.* 2011; 2(3):125-133.
7. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat.* 2001; 29(5):1189-1232.
8. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol.* 2008; 77(4):802-813.
9. Rahmani S, Edalat H, Sofizadeh A. Machine learning-based spatial mapping of canine leishmaniosis in Iran using environmental predictors. *Acta Trop.* 2022; 231:106418.
10. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillerá-Aroita G, *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography.* 2017; 40(8):913-929.
11. Karypidou MC, Vafeidis AT, Kienberger S. A framework for spatial risk assessment of animal rabies using machine learning. *One Health.* 2023; 16:100493.
12. Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography.* 2007; 30(5):609-628.
13. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005; 25(15):1965-1978.
14. Breiman L. Random forests. *Mach Learn.* 2001; 45(1):5-32.
15. Belsare AV, Gompfer ME. Assessing demographic and epidemiologic parameters of rural dog populations in India during mass vaccination campaigns. *Prev Vet Med.* 2013; 111(1-2):139-146.
16. Zinsstag J, Dürr S, Penny MA, Mindekem R, Roth F, Menendez Gonzalez S, *et al.* Transmission dynamics and economics of rabies control in dogs and humans in an African city. *Proc Natl Acad Sci U S A.* 2009; 106(35):14996-15001.

### How to Cite This Article

Celik M. Spatial distribution modelling of rabies cases in stray dog populations using machine learning algorithms. *International Journal of Veterinary Sciences and Animal Husbandry.* 2025; 10(6): 54-58.

### Creative Commons (CC) License

This is an open-access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.