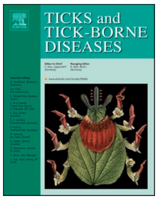
Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

journal homepage: www.elsevier.com/locate/ttbdis



Original article

Insights into the contribution of multiple factors on *Ixodes ricinus* abundance across Europe spanning 20 years using different machine learning algorithms

Samantha Lansdell[a], Abin Zorto[b], Misaki Seto[a], Edessa Negera[a], Saeed Sharif[b], Sally Cutler[a,*]

[a] Department of Health, Sport and Bioscience. University of East London, Water Lane, Stratford E15 4LZ, United Kingdom
[b] Department of Architecture, Computing and Engineering. University of East London, University Way, London E16 2RD, United Kingdom

ARTICLE INFO

Keywords:
Tick
Ixodes ricinus
Machine learning
Model
Europe
Temperature
Rainfall
Habitat

ABSTRACT

The interplay of biotic and abiotic factors driving *Ixodes ricinus* abundance trends are not fully understood. Machine learning (ML) approaches are being increasingly used to explore this and predict future abundance patterns of this species, however, the studies focusing on this to date have had limitations (including short study duration, limited sample size, narrow geographical range and use of a single ML model). This study was undertaken to address these limitations by applying 11 predictive ML models (across three data clustering techniques) to a large *I. ricinus* occurrence dataset (27,150 records) containing geographical and temporal data from a 20-year period across 30 European countries, coupled with data covering a range of climatic and habitat features (temperature, rainfall, Normalised Difference Vegetation Index (NDVI), percentage of discontinuous urban fabric and land use category). To assess which ML model was most suited to prediction of *I. ricinus* abundance, four performance metric values were calculated per model: Normalised Root Mean Square Error (NRMSE), Scatter Index (SI), Mean Absolute Percentage Error (MAPE) and R^2, all of which describe the statistical relationship between predicted and actual *I. ricinus* abundance values. Furthermore, using a Random Forest (RF) model across three clustering methods, we determined which features most significantly impacted upon *I. ricinus* abundance. The study demonstrated that Agglomerative Hierarchical Clustering (AC) methods and Linear Regression (LR) modelling performed best with this dataset. Our findings revealed that land use and rainfall were the primary contributors to *I. ricinus* abundance, with temperature playing a lesser role. This was measured according to the extent of prediction error increase following exclusion of that factor from the analysis. We provide a summary of the factors most strongly linked to *I. ricinus* abundance, which can be used to guide interventions to aid the control of ticks and tick-borne disease across Europe.

1. Introduction

When examining *Ixodes ricinus* abundance patterns, the complexities of an ecosystem comprised of both biotic and abiotic factors must be fully considered (Stachurski et al., 2021). Prevention of *I. ricinus* desiccation requires a minimum of 80 % humidity (Gray et al., 2021) making abiotic factors such as rainfall and temperature essential to consider (Estrada-Peña et al., 2013). However, we cannot attribute abundance entirely to these abiotic factors since biotic influences can also affect tick survival and proliferation (Stachurski et al., 2021). For example, vegetation-based ground cover features help to retain moisture in the tick's surroundings, thus aiding survival of *I. ricinus* (Medlock et al., 2013). Other important environmental factors impacting upon tick survival include host availability, biodiversity and land fragmentation (Estrada-Peña, 2003).

Due to these ecological complexities, a robust and multifaceted approach is required to fully understand which factors most strongly influence tick abundance. Machine learning (ML) approaches are rapidly evolving and possess the capacity to process large multifactorial and complex data sets to not only identify most significant contributing factors towards a trend, but also to predict future scenarios (Alanazi, 2022) making ML approaches highly suitable for deciphering the driving

[*] Corresponding author.
E-mail address: s.cutler@uel.ac.uk (S. Cutler).

https://doi.org/10.1016/j.ttbdis.2025.102437
Received 28 August 2024; Received in revised form 28 November 2024; Accepted 4 January 2025

forces for tick abundance.

Others have used similar approaches to understand the factors impacting upon *I. ricinus* abundance patterns in Europe. For example, Boulanger et al. (2024) assessed *I. ricinus* ticks collected from 40 sites in France. They examined soil type, land use, forest type and host presence in relation to *I. ricinus* tick (and pathogen) abundance. The sampling process was repeated on four different dates between the years of 2020 and 2021. The data were analysed using the XGB Regressor model to predict tick abundance, which was shown to perform best out of the seven predictive ML models tested according to Root Mean Square Error (RMSE) and R^2 values. Results indicated that the strongest predictors of tick abundance were related to silt content, soil moisture and presence of Sciuridae, deer and birds. However, it could be argued that this study is not fully representative of wider tick populations because it was focused on just two regions within 20 km of each other and only sampled ticks over a single year.

Scandinavian researchers used Boosted Regression Tree (BRT) modelling to explore and predict abundance of *I. ricinus* ticks, using field derived data coupled with a variety of environmental variables, including land surface temperature, Normalised Difference Vegetation Index (NDVI), altitude and precipitation (Kjær et al., 2019). Results showed that *I. ricinus* abundance increases were related to land surface temperature and vegetation index, whereas land cover and fragmentation of land were shown to be poorly linked to tick abundance. Furthermore, they evaluated the predictive power of the model by using the data from the first year to predict tick abundance in the subsequent year, with findings verified by repeating their sampling over the same time frame in the following year, comparing actual abundance to predicted abundance values. However, a limitation of this study was its limited duration (which took place over August-September in a single year).

Other studies have used ML approaches, but in other contexts, such as prediction of binary *I. ricinus* presence/absence in particular locations, or to measure overall habitat suitability rather than abundance. For example, in a study carried out in Northern Italy by Signorini et al. (2019) *I. ricinus* ticks were collected from a total of 52 sites over the course of two years (2009–2010). Environmental variable data were added for each site location (including altitude, land cover, rainfall, NDVI and land surface temperature). The study focused on habitat suitability mapping for *I. ricinus* ticks using a MaxEnt ML model, providing a measure of likelihood of tick presence in a certain location (on a scale of 0–1) based on these environmental variables – with the results indicating that the factor most important for prediction of tick presence was NDVI. However, this was a small-scale study, with only 341 tick occurrence records used, all of which were localised to a single region in a short timeframe, introducing various sources of potential bias.

Estrada-Peña and de la Fuente (2024) carried out a similar habitat suitability study, but this time spanning Europe between the years of 1990–2022. A total of five ML approaches (Random Forest (RF), Neural Networks, Naïve Bayes, Gradient Boosting and Adaboost) were used to calculate probability of *I. ricinus* presence based upon habitat features, including vertebrate host distribution, climatic features (temperature and water vapour deficit) and landscape features (related to vegetation, land fragmentation and land use). Model performance was mostly assessed using area under the receiver-operating characteristic curve (AUC). Results indicated that Gradient boosting, RF and Neural Networks performed best for habitat suitability mapping, and that model performance was impacted by climatic variables more significantly than landscape variables. However, the climate data used were calculated as a mean value per month as opposed to daily values, which may have resulted in skewed data which is less representative of true climatic conditions.

As a further example, a study carried out by Noll et al. (2023) used *I. ricinus* and *Dermacentor reticulatus* tick occurrence data from European locations, obtained from a combination of sources (GBIF and relevant publications between the years of 1970–2021). Each of these occurrence entries were linked to several explanatory variables describing the environment (including temperature, precipitation, NDVI, land surface temperature and soil moisture). Although the importance of each environmental variable was not determined, habitat suitability mapping (ecological niche modelling) was carried out using three machine learning approaches: Generalised Additive Modelling (GAM), RF and MaxENT and performance was measured using: AUC, true skill statistic (TSS), omission rate (OR), Miller's calibration slope (MCS) and Continuous Boyce Index (CBI). The authors, however, reported that the maps produced for all models did not accurately reflect tick distribution. A further limitation was their limited use of habitat-related variables, which may have impacted negatively upon the accuracy of their predictions.

A similar approach was also used by Lihou and Wall (2022) who predicted *I. ricinus* presence/absence utilising questionnaire-generated data from farmers across Britain, indicating whether ticks were present on their land in the previous nine months. Data were added for other variables related to habitat, climate and animal density at each site. An RF model was then used to process the data to predict which areas are most likely to be at risk of high tick presence – with the results showing that certain areas are at higher risk (such as central Scotland, Northern England and Wales). The results also showed that the most crucial predictors of tick presence were centred around precipitation. However, in addition to the limited duration of data collection, these findings were only representative of British farmland, excluding other land types and locations. Furthermore, the retrospective nature of this study and requirement for farmers to recognise a tick may introduce recall and reporting bias.

It is evident from these previous studies that ML-based algorithms have potential to model and predict *I. ricinus* abundance (as well as presence/absence or specific habitat/ecological suitability in a particular location). However, the full potential of abundance prediction studies specifically might have been somewhat compromised by limited sample size, geographical range, time span and the biotic/abiotic factors evaluated. In addition, many previous studies have employed a single ML model, providing little opportunity for comparison of model performance. To overcome these limitations, our objective was to apply 11 predictive machine learning approaches to data obtained from 30 European countries over a 20-year period. This data will not only dramatically increase our understanding of which ML methods are best suited to *I. ricinus* abundance prediction, but it will also provide a comprehensive insight into the factors increasing *I. ricinus* abundance (and associated risks) across Europe, thus leading to enhanced knowledge and better-informed intervention strategies.

2. Materials and methods

2.1. Tick occurrence dataset preparation

A dataset of 27,150 *I. ricinus* tick occurrence records was used in this study, with records originating from various sources. Firstly, 5365 *I. ricinus* tick occurrence records (from Europe between the years of 2000 and 2019) were obtained from three online data repositories: Global Biodiversity Information Facility (GBIF), National Biodiversity Network (NBN) Atlas and Vectormap (explained in further detail below). The countries included in the boundaries of Europe were selected according to a document published by the United Nations (Countries [WWW Document] n.d available at: https://www.ohchr.org/en/countries). For each individual record exported, full date of occurrence, data collection method and WSG84 decimal coordinates (latitude and longitude) were noted. Occurrence records were only included if a tick was present (no absence records were included) with a focus on quantitative data (tick occurrence count) at each site as opposed to qualitative presence/absence data.

For GBIF, an occurrence search filtered to *I. ricinus* was undertaken, and a total of 7945 occurrences were exported in .csv format (GBIF.org (16 September 2023) GBIF Occurrence Download. https://doi.org/10.15468/dl.kmp6em). The DOI for each individual contributor of data can be found in https://doi.org/10.15468/dl.kmp6em. To initially clean the data, records prior to 2000 and after 2019 were removed. Occurrences in countries outside of Europe were removed. Entries missing latitude or longitude coordinates or day, month or year of occurrence were also removed. Following this initial cleaning process, 3766 records remained. A summary of the raw data provided can be found in the Supplementary Material (Table S1).

For NBN Atlas (which only contains records from the UK) *I. ricinus* records were searched, and filtered to exclude unconfirmed entries (Search: SPECIES: Ixodes (Ixodes) ricinus | Occurrence records | NBN Atlas [WWW Document] n.d https://records.nbnatlas.org/occurrences/search?*q*=lsid%3ANHMSYS0000730335&fq=occurrence_status%3Apresent&fq=(identification_verification_status%3A%22Accepted%22%20OR%20identification_verification_status%3A%22Accepted%20-%20considered%20correct%22%20OR%20identification_verification_status%3A%22Accepted%20-%20correct%22)&nbn_loading=true). A total of 358 records were exported in .csv format. DOI information for each contributor can be found in the Supplementary Material (Table S2). As above, entries before 2000 or after 2019 were removed. Entries missing latitude or longitude coordinates or day, month or year of occurrence were removed. A total of 165 records remained following this process (summarised in Table S1).

For Vectormap, a search was conducted for *I. ricinus* ticks. A total of 5189 records were exported in .csv format. Data were obtained from records held in the VectorMap data portal (https://experience.arcgis.com/experience/5f95c3edfbea4634b8347fec0bd1dcd6/) on 16 September 2023 (VectorMap [WWW Document] n.d). The institutions contributing towards the dataset used were: Argyll Biological Records Centre, BIS, BRERC, Buglife, Cofnod – North Wales, Environmental Information Service, Highland Biological Recording Group, Lancashire, Environment Record Network, Leicestershire and Rutland Environmental Records Centre, National Trust, Natural England, Natural Resources Wales, Naturespot, Biological Records Centre, North East Scotland Biological Records Centre, OHBR, Rotherham Biological Records Centre, Steve Woodward and Yorkshire Wildlife Trust. Data were cleaned by removal of entries outside of Europe, before 2000 or after 2019. Additionally, entries missing latitude or longitude coordinates or day, month or year of occurrence were removed. Following this process, a total of 1434 records remained (summarised in Table S1).

These 5365 *I. ricinus* occurrences were plotted on a map using Tableau (Tableau [WWW Document] n.d https://www.tableau.com/products) to demonstrate geographical range and tick density per site (as shown in Fig. 1) with larger dot size indicating increased tick density (demonstrating the quantitative aspect of the work). A further map was created using Tableau (Tableau [WWW Document] n.d https://www.tableau.com/products) with colour-coded dots representing month of year (to demonstrate seasonal patterns) as shown in Fig. 2.

A further 225 records were obtained from Institute of Public Health in Albania, who provided data from sites in Albania between the start of 2000 and end of 2019. We were not given permission to share any further details regarding the data provided (consequently we are unable to include raw data or details regarding occurrence mapping, geographical range or collection method).

Finally, 21,560 records were provided by Professor Roy Brown, Senior Visiting Research Associate at University of Bangor. Ticks were collected using blanket dragging and noted as individual occurrences, between the years of 2000 and 2019 only. Although we do not have permission to share the raw occurrence data from this provider, these occurrences were plotted on a map using Tableau (Tableau [WWW Document] n.d https://www.tableau.com/products/tableau) shown in Fig. 3, documenting the geographical range covered, with larger dot size indicating higher tick density at a particular site.

2.2. Selection and addition of climatic and environmental variables to the dataset

Data relating to several environmental and climatic variables were added to the occurrence record dataset. The following variables were selected (in addition to 'Latitude', 'Longitude' and 'DayofYear' variables already included in the dataset): 'Temperature', 'Rainfall', 'NDVI', 'Land use' and 'Percentage of discontinuous urban fabric'. These variables were selected based on an extensive literature search, which indicated that they are key contributors to tick abundance (Kjær et al., 2019; Signorini et al., 2019; Lihou and Wall, 2022; Estrada-Peña, 2003)

For each *I. ricinus* occurrence record, maximum temperature and precipitation amount were added. This was achieved by using an online data repository available at: https://www.visualcrossing.com/weathe

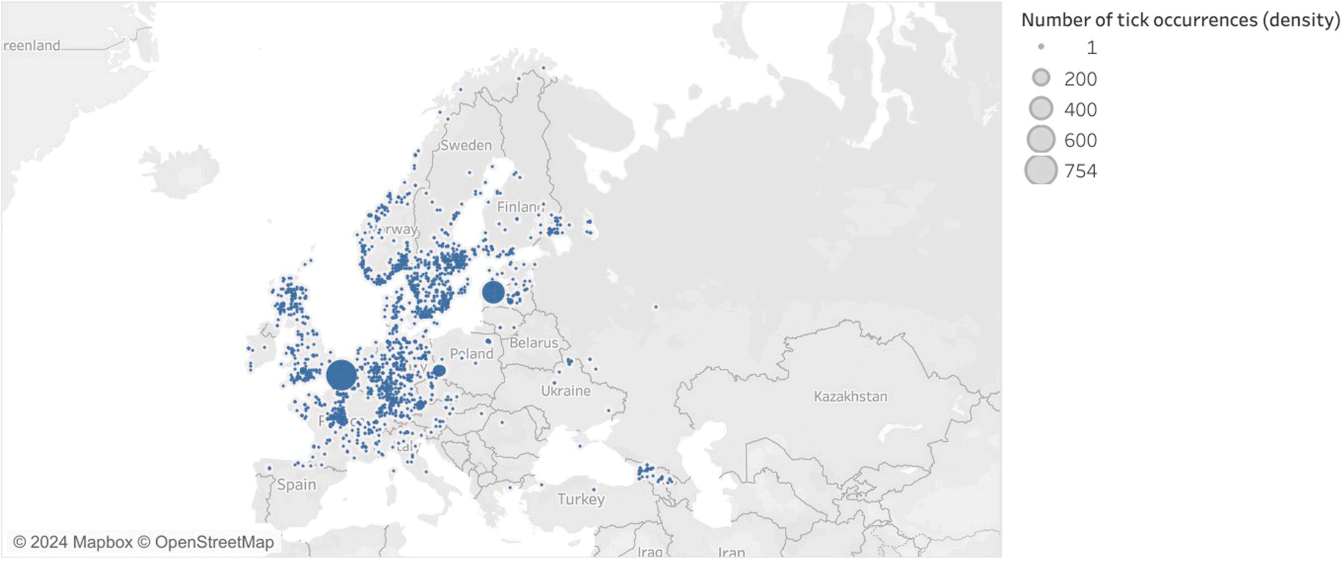


Fig. 1. Map showing the distribution and density of *I. ricinus* occurrences across European countries between the years of 2000–2019. The map shows *I. ricinus* occurrence locations across Europe. Larger tick count values in a single location are represented by a larger dot size (as shown in the legend to the right of the map). Data were obtained from NBN Atlas, GBIF and Vectormap repositories (previously cited) and plotted using Tableau.
Created using Tableau (2024) https://www.tableau.com/products (Tableau [WWW Document] n.d).

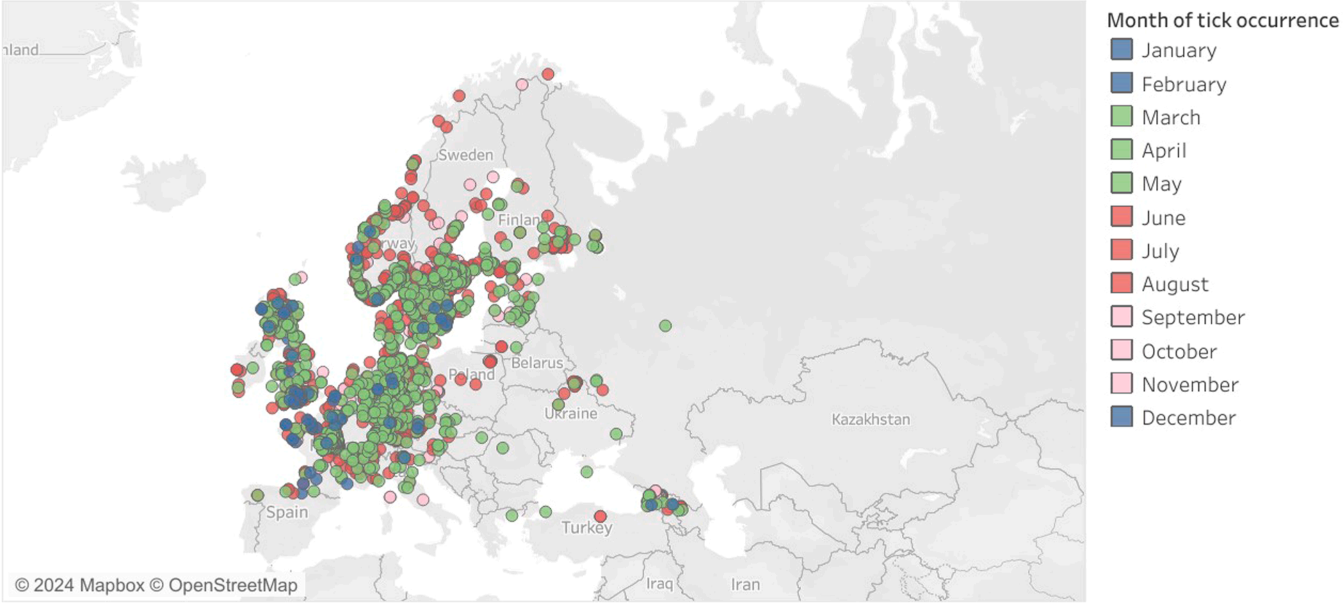


Fig. 2. Map showing the distribution and seasonality of *I. ricinus* occurrences across European countries between the years of 2000–2019.
The map shows *I. ricinus* occurrence locations across Europe, with occurrence points colour-coded to represent different months of the year (as shown in the legend to the right of the map). Data were obtained from NBN Atlas, GBIF and Vectormap repositories (previously cited) and plotted using Tableau.
Created using Tableau (2024) https://www.tableau.com/products (Tableau [WWW Document] n.d)

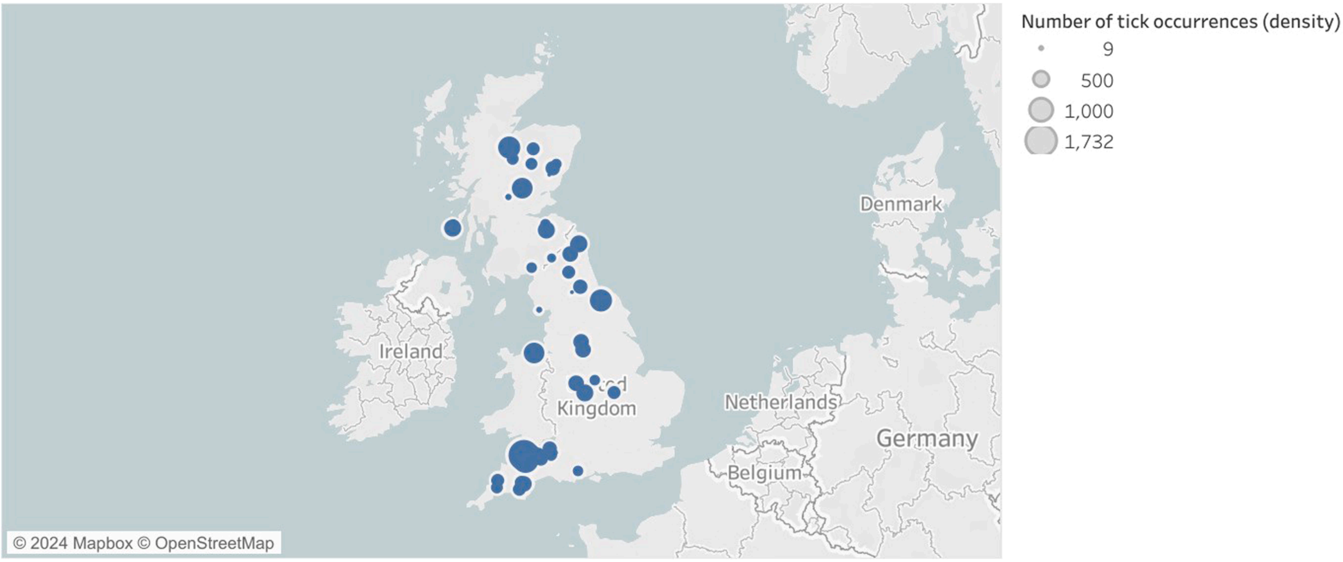

Denmark
Ireland
Kingdom
Netherlands
Germany
Belgium
© 2024 Mapbox © OpenStreetMap
Number of tick occurrences (density)
9
500
1,000
1,732

Fig. 3. Map showing the distribution and density of *I. ricinus* occurrences across the United Kingdom between the years of 2000–2019.
The map shows *I. ricinus* occurrence locations across the United Kingdom. Larger tick count values in a single location are represented by a larger dot size (as shown in the legend to the right of the map). Data were obtained from Professor Roy Brown (previously cited) and plotted using Tableau (Tableau [WWW Document] n.d).

r/weather-data-services ("Weather Data Services | Visual Crossing," n. d.) to search for temperature and rainfall measurements according to the specific location coordinates and date of the tick occurrence. The maximum values for the specified day were used.

Data corresponding to three habitat variables (land cover category, NDVI value and percentage of discontinuous urban fabric) were then added to each record from an online data repository called EcoDataCube (Open Environmental Data Cube viewer [WWW Document] n.d. https://ecodatacube.eu/) a resource provided by Open Data Science Europe ("Open Data Science Europe – EU-wide automated mapping system for harmonization of Open Data based on FOSS4G and ML," n. d.). For land cover, the layer search function of EcoDataCube was used to locate the 'Land Cover' data page. For each occurrence record, the location coordinates were searched within this data page and the land cover category was noted specific to the year of the occurrence. For the NDVI part, the occurrence records were firstly split according to month of occurrence and assigned to an occurrence season (spring, summer, autumn or winter month). The layer search function was then used on EcoDataCube to locate the 'NDVILandsat (quarterly)' data pages (with one data page available for each season). For each occurrence, the location coordinates were searched in the relevant 'NDVILandsat (quarterly)' data page (according to season) and the NDVI value was noted (specific to the year of occurrence). Finally, the 'Discontinuous urban fabric' data page was located on EcoDataCube using the layer search function. For each occurrence, the location coordinates were searched in this data page and the percentage value was noted specific to the year of the occurrence.

2.3. Dataset cleaning and preprocessing

The 27,150 *I. ricinus* data records (with climatic and habitat data variables added) were further cleaned by removing invalid spatial data entries. Missing data relating to any of the climate or habitat variables (where 'No Data Available' was written in any of these columns because of missing data in the data repositories) were handled by replacing these with 'np.NaN'. Data rows containing these entries were then removed to avoid impact upon subsequent modelling processes. In addition, inconsistent data entries (not following expected format or with values contradictory to other data points) were removed. A total of 2804 records were removed during these cleaning stages, leaving a total of 24,346 records in the analysis.

Finally, the dataset was simplified by merging the day, month and year columns, changing this to a single entry containing day of year. Numerical data entries (for 'Temperature', 'Rainfall', 'NDVI' and 'Percentage of discontinuous urban fabric') were converted to more precise numerical formats. In cases where the variable was in the form of categorical data (such as land use and data collection method) this was transformed into a format suitable for numerical analysis, which was carried out for the purpose of creating binary data, thus enabling the model to process this categorical data without assuming any inherent order.

2.4. Feature set engineering

The data relating to 'Latitude', 'Longitude', 'Temperature', 'Rainfall', 'NDVI' and 'Land use' variables were arranged in two feature sets:

- Geo-climatic variables (temperature, rainfall, NDVI, latitude and longitude data)
- Land use category variables

Data from both datasets were used for subsequent processes described below.

2.5. Clustering and data aggregation methodology

Using Python version 3.6 (Python Software Foundation. Python Language Reference, version 3.6. Available at http://www.python.org) clustering techniques were used to group and summarise the cleaned and prepared dataset of *I. ricinus* observations according to location. The dataset was clustered in three different ways (resulting in three separate datasets, with each one representing a single clustering method).

The three clustering methods employed (and resulting datasets) were:

- K-Nearest Neighbour (KNN)
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- Agglomerative Hierarchical Clustering (AC)

KNN clustering groups data points based on their proximity to each other. This method is highly suited to large data sets (Dhanabal and Chandramathi, 2011). DBSCAN clustering operates in a different manner, focusing on density of data points. It is advantageous due to its ability to ignore noise (Deng, 2020). AC clustering involves the merging of multiple neighbouring clusters (Wang et al., 2017). This method can handle datasets presented in a variety of shapes, dimensions and sizes (Chidananda Gowda and Krishna, 1978).

Further features were then added to each of the clusters:

- Number of *I. ricinus* observations (tick count)
- Average values for each of the geographic variables (latitude and longitude) and environmental variables (temperature, rainfall, NDVI and land use category)
- Data collection information

2.6. Outlier removal and data filtering

To enhance data quality and increase precision by eliminating extreme data, a robust method of outlier detection and removal was applied. Firstly, the 5th and 95th percentiles of the "Tick Count" variable were calculated (representing lower and upper data extremes, beyond which the data are considered outliers). Data points beyond these data extremes were excluded from further analysis to most accurately represent tick abundance without skew. As a result, only the central 90 % of the data was considered (maximising the level of representation). Outlier removal varied according to each clustering method. For KNN clustering, 5588 (22.95 %) of the records were excluded leaving 18,758 records remaining. For the DBSCAN clustering, 19,418 (79.76 %) of records were excluded, leaving 4928 records remaining. Finally, for AC clustering, 5587 (22.95 %) of the records were excluded, leaving 18,759 records remaining.

Models and clustering methods were also evaluated in the presence of the outlier data, to demonstrate the impact of outlier removal upon model performance as a point of comparison (although this was not the main focus of our work).

2.7. Model training and evaluation (part one of analysis)

For each clustering dataset of *I. ricinus* records, the data were split into ten equal parts for model training and testing. K-fold cross validation (Leave-one-out Cross Validation) processes were used, during which nine parts of the data were used for model training and the remaining one part was used for model testing. This was repeated a total of ten times so that each individual part formed part of the testing data once, acting as a highly advantageous method of cross-validation covering all parts of the data (Yadav and Shukla, 2016).

This process was applied to each of the following models:

- Random Forest (RF)
- XGBoost (XGB)
- LightGBM (LGBM)
- CatBoost (CB)
- Voting Regressor (VR)
- Bagging Regressor (BR)
- Stacking Regressor (SR)
- AdaBoost Regressor (AB)
- Linear Regression (LR)
- Decision Tree (DT)
- Support Vector Machine (SVM)

To determine which of the models showed the best performance when applied to *I. ricinus* occurrence records, various model performance metric values were calculated for each of the models, across all three clustering methods. These performance metrics were: Normalised Root Mean Squared Error (NRMSE), Scatter Index (SI), Mean Absolute Percentage Error (MAPE) and R^2. Guidance on the definition and interpretation of these values can be found in Table 1, all relating to the statistical relationship between predicted and actual values.

The model evaluation process had two key aspects based on these performance metrics:

1. **Determination of the best-performing clustering method**
   For each clustering method, the average value for each performance metric was calculated across all models. The clustering method showing the lowest average NRMSE, SI and MAPE values

Table 1
Performance metric definitions and interpretation.

Performance metric	Definition	Interpretation	References
Normalised Root Mean Squared Error (NRMSE)	RMSE measures absolute error between predicted and actual values. To calculate the NRMSE value, the RMSE is divided by the range of the actual values observed.	Values closer to zero represent a lower error rate.	Kambezidis, 2012; Shcherbakov et al., 2013.
Scatter Index (SI)	SI examines the frequency of observations within a dataset. It measures how scattered these are, thus showing levels of dispersion within the dataset, acting as a measure of consistency.	Lower values indicate less dispersion in the data (thus representing higher consistency).	Bhattacharya and Sinha, 2022.
Mean Absolute Percentage Error (MAPE)	MAPE measures the predictive accuracy of a model by calculating the average percentage difference between predicted and actual values.	Values closer to zero represent lower percentage of error (higher accuracy).	Chicco et al., 2021.
R^2	The coefficient of determination (R^2) measures how closely actual values fit with predicted values. It acts as a measure of data variance.	Values closer to one represent a better fit between predicted and actual values.	Håkanson, 1995; Chicco et al., 2021.

This table outlines the definition of each performance metric (NRMSE, SI, MAPE and R^2) and specifies how the values should be interpreted.

(and the highest R^2 values) was deemed to be the best performing approach.

2. **Determination of the best-performing model for each clustering method**

   For each clustering method, the model showing the lowest NRMSE, SI and MAPE values (and the highest R^2 values) was deemed to be the best performing model.

2.8. *Feature importance analysis (part two of analysis)*

Each clustering dataset (containing the following variables: 'Latitude', 'Longitude', 'Temperature', 'Rainfall', 'DayOfYear', 'NDVI', 'Percentage of discontinuous urban fabric' and each 'Land Use' category) was subjected to RF modelling (for predicting *I. ricinus* abundance, as described in part one). RF modelling was selected due to the ability of this model to combine multiple decision trees, as well as its high suitability for variable importance assessments (Boulesteix et al., 2012). Success of the prediction model for this part of the work was measured according to the MAPE value only (which represents predictive accuracy, as described in Table 1). To determine importance of each feature in the prediction of *I. ricinus* abundance, the modelling process was repeated multiple times, but with a single variable excluded each time, enabling evaluation of the numerical impact of this exclusion on the MAPE value. The extent of impact of each factor exclusion on the MAPE value was converted to a coefficient value (between 0 and 1) with values closer to 1 indicating a more significant impact on MAPE value (thus signifying higher feature importance).

As an extension of this, correlation and potential collinearity between environmental variables (in relation to their impact upon tick count) were explored by systematic evaluation via the calculation of a correlation matrix (focusing on the strength and direction of the relationship between environmental factors and tick count). Furthermore, scatter plots were generated to visually explore these relationships. However, these have not been included in this paper.

2.9. *ODMAP protocol*

A full in-depth summary of the processes followed in this work can be found in Supplementary Material (Fig. S3) completed according to the ODMAP protocol (defined by Zurell et al., 2020).

3. Results

3.1. *Determination of the most suitable data clustering method*

We initially determined which of the three clustering methods (with outliers removed) resulted in the best average model performance metric values across the 11 models (lowest MAPE, NRMSE and SI values, and highest R^2 value, as described in Table 1). As shown in Fig. 4, a graph was firstly plotted using the average MAPE, NRMSE, SI and R^2 value for each clustering method. Overall, AC clustering performed best for prediction of *I. ricinus* abundance in the absence of outliers. Results are fully described in the corresponding figure legend. This same process was also repeated for the three datasets, but with outliers included, with results included in the Supplementary Material (see Fig. S4) with a full description provided in the figure legend.

3.2. *Determination of the most suitable ML model (using the best-performing clustering method)*

Using the AC clustering method (with outlier removal) values for each algorithm were plotted in a graph to demonstrate which of the 11 models showed the best performance (lowest MAPE, NRMSE and SI, and highest R^2, as per Table 1). The performance metric values for each ML model have been plotted in Fig. 5. Using AC clustering, DT performed best in terms of MAPE, whereas, LR performed best in terms of NRMSE and SI. For R^2, the best performance was shown for LR. Overall, LR was considered to be the best-performing model. Full description of these

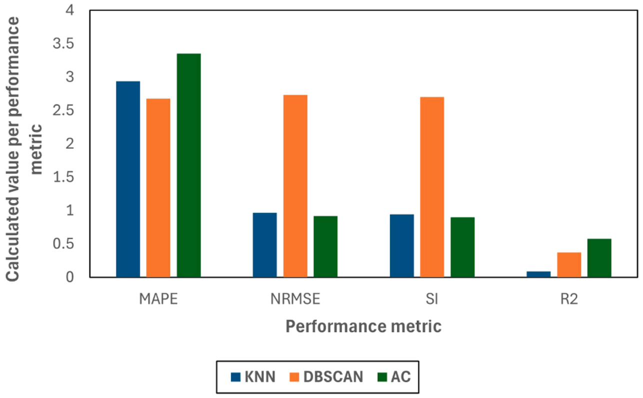


Fig. 4. Bar chart showing average MAPE, NRMSE, SI and R^2 value per clustering method (with outliers removed).
Comparative performance analysis showed that for average MAPE value, DBSCAN clustering generated the value closest to zero, indicating greatest accuracy. However, for average NRMSE and SI, the values closest to zero were generated using the AC clustering method, indicating the best error rate and consistency. For R^2, the highest average value was shown for AC clustering, indicating that this clustering method shows the lowest variance (indicating a good fit between predicted and actual values).

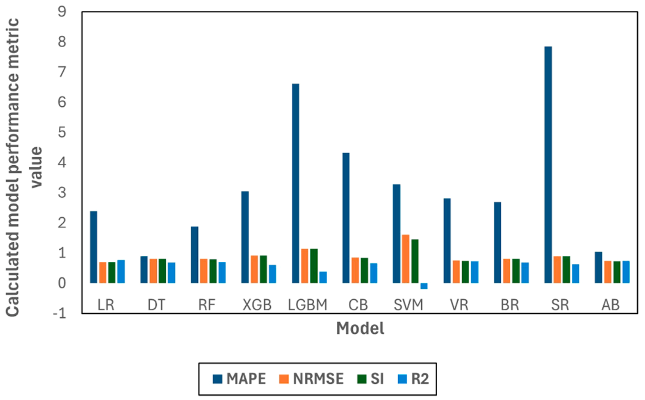


Fig. 5. Bar chart showing MAPE, NRMSE, SI and R^2 values for each model following AC clustering (with outliers removed).
Using MAPE, the best performing model was DT, which generated the value closest to zero, indicating best accuracy. However, for NRMSE and SI, superior results were found for LR in terms of error rate and consistency when compared to other models. For R^2, the highest value was shown for LR, indicating the lowest variance (the best fit between predicted and actual values).

results can be found in the corresponding figure legend. The same process was repeated for the AC clustering dataset, but with outliers included. The results for this have been included in the Supplementary Material (see Fig. S5) with a full description provided in the figure legend.

Finally, the same process was applied for the KNN and DBSCAN clustering datasets (both with and without outliers). The results from this part of the work are shown in the Supplementary Material (Fig. S6 for KNN data without outliers, Fig. S7 for KNN data with outliers, Fig. S8 for DBSCAN data without outliers, Fig. S9 for DBSCAN data with outliers) with a full description provided in the figure legend.

3.3. *Feature importance analysis*

Using AC clustering, the coefficient values for feature importance are shown in Fig. 6, revealing that land use (particularly moors and heathland), rainfall and temperature showed greatest importance when predicting tick abundance. Full description of these results can be found in the corresponding figure legend. This analysis was also conducted for KNN and DBSCAN clustered data (shown in Figs. S10 and S11 of the Supplementary Material respectively) with a full description provided in the corresponding figure legends.

4. Discussion

Our main objective was to determine which factors played the most important role in prediction of *I. ricinus* abundance. Furthermore, given the somewhat contradictory conclusions of prior work, we sought to undertake a side-by-side comparison of different ML analytical approaches using a comprehensive dataset to assess their ability to handle *I. ricinus* occurrence and environmental data. As our study focused only on *I. ricinus* abundance, this section discusses our findings in relation to previous abundance studies only (excluding presence/absence studies).

4.1. *Feature importance*

Our analysis revealed that land use, precipitation and temperature had greatest impact for driving tick abundance. The top-ranking factor was land use (moors and heathland) which could relate to the water retention of moors providing a controlled level of humidity that is critical for tick survival (Lihou et al., 2020). Similarly, heathland is characteristically covered with bracken which provides a deep leaf litter to help promote tick survival during hostile climatic conditions and a safe haven for a diverse range of potential hosts for ticks when conditions are conducive for questing (Heylen et al., 2013).

In contrast to our findings, the previous study examining land use as a driving factor for tick abundance, carried out by Kjær et al. (2019) found that land use had no impact on *I. ricinus* abundance. However, in

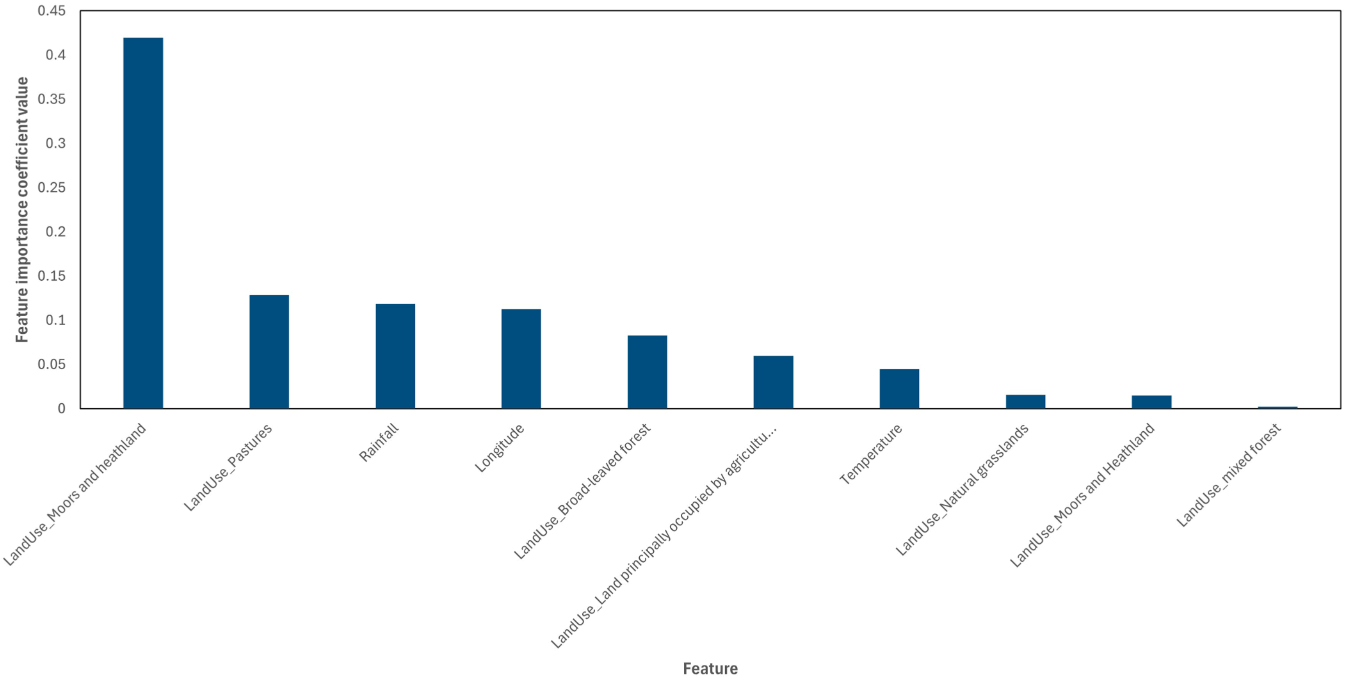


Fig. 6. Results of RF model feature importance analysis for AC clustering.
Feature importance coefficients closer to 1 represent a higher level of importance in terms of tick abundance prediction. For AC clustering, land use (especially moors and heathland), rainfall and temperature were shown to have the most significant impact on *I. ricinus* abundance.

addition to their limitations regarding study duration, their land use category range was limited (only categorising forests and meadows) which could have compromised their findings, whereas, our study explored a broad range of land use categories over a 20 year time period, providing a more detailed analysis.

Precipitation was shown to be the second most important factor following land use. The study by Kjær et al. (2019) examined the role of precipitation specifically in relation to *I. ricinus* abundance – with results indicating that this was not a contributing factor. However, they used precipitation values from annual or quarterly data averages, which could result in data skew from calculation of average values. In contrast, our study examined precipitation specific to date of occurrence, giving better precision, which is likely to be more representative of true precipitation values.

Temperature was a significant contributing factor to *I. ricinus* abundance but not to the same extent as either land use or rainfall. Supporting our study findings, Kjær et al. (2019) concluded that temperature played a key role in prediction of tick abundance. However, they used average temperature values, which could potentially misrepresent true climatic conditions on a given day, whereas we used daily temperature readings specific to the tick report, again adding precision. Several studies unrelated to ML-based predictions have also implied that climatic factors have dramatically influenced both abundance and range of tick species (Estrada- Peña et al., 2012; Gray et al., 2009).

Climatic factors are clearly important and may govern aspects of tick phenology including host-seeking behaviour (such as timing of seasonal questing) and diapause, however, we must consider that overestimation of the importance of temperature may have arisen from failure to appreciate that much of the life of a tick is spent deep in the leaf litter where the assemblage of bracken and other vegetation provide a microclimate that protects ticks from the more extreme variability of temperature and humidity. In response to unfavourable climatic conditions, ticks are capable to deploy their 'sit and wait' strategy until conditions are favourable to quest for potential hosts, thus limiting their exposure to climate factors. Furthermore, they may avoid desiccation in hotter climates by seeking alternative hosts (such as lizards) in the leaf litter rather than questing at higher levels. In consequence, the impact of climatic change might be obscured, resulting in subtle behavioural shifts or decreased over-winter mortality rates, rather than binary presence or absence of ticks (Stachurski et al., 2021). As such, we may see a shift from regions where ticks can survive to areas where ticks can thrive, thus increasing tick abundance.

Our study has limitations, with a key limitation being that questing tick numbers have been used to gauge overall abundance, which being a behavioural trait rather than absolute count, could be a source of error. Interpretation of *I. ricinus* abundance data is further complicated by the longevity of this tick species, with the tick life cycle potentially spanning up to 4–6 years (Kahl and Gray, 2023). Furthermore, with the year-by-year variability of tick numbers, measurement of tick density would ideally need to extend over several years to mitigate aberrant fluctuations. In addition, it should be noted that our data is open to potential bias based upon its requirement for reporting of ticks, either through tick biting or active collections of ticks from their habitat. Often this latter source of data is dictated by proximity to researchers with an active interest in tick abundance rather than being collected through systematic surveillance (Kugeler and Eisen, 2020b).

Our study is further limited by the fact that there are many other factors which may influence tick survival and perpetuation which have not been considered. For example, tick survival will undoubtedly be subject to local ecological drivers such as host availability and biodiversity (Estrada-Peña, 2003). As witnessed in our study, data on host presence and biodiversity is challenging to obtain. Consequently, the impact of biodiversity upon tick numbers has not been adequately addressed to date. This should include the impact of not only blood meal local hosts, but also dispersal hosts such as birds and amplification hosts such as deer.

In future work, other features could also be introduced in modelling to address some of the limitations of this current study; for example, features which more accurately represent the microclimate (such as soil temperature and soil moisture) could be included. Additionally, incorporation of host and biodiversity data would strengthen our ML model for the future.

Despite these limitations, our comprehensive analysis spanning Europe across a 20-year duration determined which features were most important for *I. ricinus* abundance prediction, providing novel insights which are more widely representative across Europe and can be used to better explain tick density patterns. However, there is scope to improve this in future work by application of more extensive model evaluation procedures.

4.2. ML model evaluation

In addition to our findings regarding feature importance, evaluation of models revealed that LR modelling in combination with AC clustering showed the best performance values, suggesting that this is the most suitable approach for handling *I. ricinus* data in the context of abundance prediction.

The principal strategy for this current work was to focus on results obtained from the data following outlier removal, for the purpose of ensuring that extreme data which may result in *I. ricinus* distribution skew were excluded from the analysis. By including the central 90 % of the data (with minimal exclusion of outliers at the upper and lower data extremes), we have ensured maximum inclusion of tick occurrence sites. However, the model performance results from data with outliers included have been added into the Supplementary Material to demonstrate model performance using the entire tick dataset, allowing scope for further ecological interpretation. Reassuringly, when using the combination of approaches found to be most successful in this study (AC and LR) outlier inclusion or exclusion had very little impact on model performance.

Previous studies comparing the performance of multiple ML models in relation to tick abundance prediction have been limited. A previous study carried out by Boulanger et al. (2024) compared seven different models according to RMSE and R^2 performance values, with the finding that XGBoost showed the best performance. This conflicts with our findings, which showed that LR was the best-performing model. This discrepancy might have arisen from differences in clustering techniques as they did not disclose the clustering method used. Furthermore, in addition to the wider range of models evaluated in our study, our data was considerably more comprehensive regarding numbers, timespan and geographical range evaluated.

During our comprehensive cross-validation process, each model was trained on nine parts of the data to make predictions about tick count in the locations present in the remaining one part of the data. In this way, some of the models showed excellent predictive abilities in terms of accurately predicting tick count in a specifically defined area. However, the predictive abilities of the model were not tested on any areas not included in the original dataset, making this a further limitation of our work. Resource limitations precluded systematic assessment of new locations to further optimise and evaluate the machine learning methods in new areas. This would be a natural future progression of this study.

Furthermore, the use of a broad geographical range with outputs which are generalised to the whole of Europe, could be considered both a strength given its inclusive applicability, but also a limitation through the potential risk of failure of the model to account for individual tick conducive eco-areas that might exist adjacent or in close proximity to refractory areas for *I. ricinus* ticks. Due to the constraints of our study, we have not been able to address this to date, but exploration of spatial resolution of ML models is a priority for future work.

In summary, we undertook extensive comparative analysis to assess performance of modelling approaches, to determine those best suited to tick abundance data, providing key insights which can inform future modelling work. The successful application of modelling has excellent predictive potential, allowing us to identify or even predict which locations are likely to face the highest risk of high tick abundance, whilst also creating an opportunity for potential expansion in the future to other aspects (such as prediction of tick-borne pathogen abundance). Overall, this will be invaluable for guiding public health policies and strategies for tick-borne disease prevention and control across Europe.

5. Conclusion

This study provides a novel insight into the factors driving *I. ricinus* abundance patterns in Europe over a 20-year period. The results showed that land use and rainfall measurements play a significant role in tick abundance. In contrast to many previous studies, temperature appeared to play less of a significant role. The results of the study also showed that the most suitable approach to handling tick occurrence datasets involves a combination of AC clustering and LR modelling, which show an excellent ability to predict tick occurrence in previously unsampled areas. The findings from this study provide several key insights which will help to guide future tick control approaches.

Author contributions

SC & SS conceptualised the study and secured funding. SL & AZ conducted data collation, cleaning and analysis under supervision of SC & SS respectively. SL drafted the original manuscript. All authors (SC, SS, SL, AZ, EG, MS) participated in review and revision of the manuscript. All authors approved the submitted manuscript.

Funding

This project has been funded by University of East London (Impact Funding Scheme 2023).

Ethics

Not required.

CRediT authorship contribution statement

Samantha Lansdell: Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Abin Zorto:** Writing – review & editing, Investigation, Formal analysis. **Misaki Seto:** Writing – review & editing. **Edessa Negera:** Writing – review & editing. **Saeed Sharif:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Sally Cutler:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Conflict of interest

All authors declared no conflict of interest.

Acknowledgements

We thank the Impact Funding Scheme 2023 at UEL for funding our research. We are also grateful to the online repositories providing the tick occurrence record data (National Biodiversity Network Atlas, Global Biodiversity Information Facility and Vectormap). We would also like to acknowledge Enkelejda Dikolli and Perparim Kadriaj from Institute of Public Health in Albania, who kindly shared tick occurrence data for use in this study. We are hugely indebted to Professor Roy Brown (Senior Visiting Research Associate at University of Bangor) for provision of tick data and helpful discussions. This work has been done within the framework of the COST Action CA21170 'Prevention, anticipation and mitigation of tick-borne disease risk applying the DAMA protocol (PRAGMATICK)'.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ttbdis.2025.102437.

Data availability

The authors do not have permission to share data.

References

Alanazi, A., 2022. Using machine learning for healthcare challenges and opportunities. Inform. Med. Unlocked. 30, 100924. https://doi.org/10.1016/j.imu.2022.100924.

Bhattacharya, D., Sinha, N., 2022. Scatter index: an alternative measure of dispersion based on relative frequency of occurrence of observations. pp. 65–72. https://doi.org/10.1007/978-981-19-1559-8_7.

Boulanger, N., Aran, D., Maul, A., Camara, B.I., Barthel, C., Zaffino, M., Lett, M.-C., Schnitzler, A., Bauda, P., 2024. Multiple factors affecting *Ixodes ricinus* ticks and associated pathogens in European temperate ecosystems (northeastern France). Sci. Rep. 14, 9391. https://doi.org/10.1038/s41598-024-59867-x.

Boulesteix, A.-L., Janitza, S., Kruppa, J., König, I.R., 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. WIREs Data Min. Knowl. Discov. 2, 493–507. https://doi.org/10.1002/widm.1072.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623. https://doi.org/10.7717/peerj-cs.623.

Chidananda Gowda, K., Krishna, G., 1978. Agglomerative clustering using the concept of mutual nearest neighbourhood. Pattern. Recognit. 10, 105–112. https://doi.org/10.1016/0031-3203(78)90018-3.

Countries [WWW Document], n.d. OHCHR. URL https://www.ohchr.org/en/countries (accessed 11.24.24).

Deng, D., 2020. DBSCAN Clustering Algorithm Based on Density, in: 2020 7th International Forum on Electrical Engineering and Automation (IFEEA). In: Presented at the 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), pp. 949–953. https://doi.org/10.1109/IFEEA51475.2020.00199.

Dhanabal, S., Chandramathi, S., 2011. A review of various k-nearest neighbor query processing techniques. Int. J. Comput. Appl. 3.

Download [WWW Document], n.d. URL https://www.gbif.org/occurrence/download/0018649-230828120925497 (accessed 11.24.24).

Estrada-Peña, A., 2003. The relationships between habitat topology, critical scales of connectivity and tick abundance *Ixodes ricinus* in a heterogeneous landscape in northern Spain. Ecography. 26, 661–671. https://doi.org/10.1034/j.1600-0587.2003.03530.x.

Estrada-Peña, A., Ayllón, N., de la Fuente, J., 2012. Impact of climate trends on tick-borne pathogen transmission. Front. Physiol. 3, 64. https://doi.org/10.3389/fphys.2012.00064.

Estrada-Peña, A., de la Fuente, J., 2024. Machine learning algorithms for the evaluation of risk by tick-borne pathogens in Europe. Ann. Med. 56, 2405074. https://doi.org/10.1080/07853890.2024.2405074.

Estrada-Peña, A., Farkas, R., Jaenson, T.G.T., Koenen, F., Madder, M., Pascucci, I., Salman, M., Tarrés-Call, J., Jongejan, F., 2013. Association of environmental traits with the geographic ranges of ticks (Acari: *ixodidae*) of medical and veterinary importance in the western Palearctic. A digital data set. Exp. Appl. Acarol. 59, 351–366. https://doi.org/10.1007/s10493-012-9600-7.

Gray, J., Kahl, O., Zintl, A., 2021. What do we still need to know about *Ixodes ricinus*? Ticks. Tick. Borne Dis. 12, 101682. https://doi.org/10.1016/j.ttbdis.2021.101682.

Gray, J.S., Dautel, H., Estrada-Peña, A., Kahl, O., Lindgren, E., 2009. Effects of climate change on ticks and tick-borne diseases in Europe. Interdiscip. Perspect. Infect. Dis. 2009, 593232. https://doi.org/10.1155/2009/593232.

Håkanson, L., 1995. Optimal size of predictive models. Ecol. Modell. 78, 195–204. https://doi.org/10.1016/0304-3800(93)E0103-A.

Heylen, D., Adriaensen, F., Van Dongen, S., Sprong, H., Matthysen, E., 2013. Ecological factors that determine *Ixodes ricinus* tick burdens in the great tit (*Parus major*), an avian reservoir of *Borrelia burgdorferi* s.l. Int. J. Parasitol. 43, 603–611. https://doi.org/10.1016/j.ijpara.2013.02.007.

Kahl, O., Gray, J.S., 2023. The biology of Ixodes ricinus with emphasis on its ecology. Ticks Tick Borne Dis. 14, 102114. https://doi.org/10.1016/j.ttbdis.2022.102114.

Kjær, L., Soleng, A., Edgar, K.S., Lindstedt, H.E.H., Paulsen, K.M., Andreassen, Å.K., Korslund, L., Kjelland, V., Slettan, A., Stuen, S., Kjellander, P., Christensson, M., Teräväinen, M., Baum, A., Klitgaard, K., Bødker, R., 2019. Predicting the spatial abundance of *Ixodes ricinus* ticks in southern Scandinavia using environmental and climatic data. Sci. Rep. 9, 18144. https://doi.org/10.1038/s41598-019-54496-1.

Kambezidis, H.D., 2012. 3.02 - The Solar Resource. In: Sayigh, A. (Ed.), Comprehensive Renewable Energy. Elsevier, Oxford, pp. 27–84. https://doi.org/10.1016/B978-0-08-087872-0.00302-4.

Kugeler, K.J., Eisen, R.J., 2020b. Challenges in predicting Lyme disease risk. JAMA Netw. Open 3 (3), e200328, 3.

Lihou, K., Rose Vineer, H., Wall, R., 2020. Distribution and prevalence of ticks and tick-borne disease on sheep and cattle farms in Great Britain. Parasites Vec. 13, 406. https://doi.org/10.1186/s13071-020-04287-9.

Lihou, K., Wall, R., 2022. Predicting the current and future risk of ticks on livestock farms in Britain using random forest models. Vet. Parasitol. 311, 109806. https://doi.org/10.1016/j.vetpar.2022.109806.

Medlock, J.M., Hansford, K.M., Bormane, A., Derdakova, M., Estrada-Peña, A., George, J.-C., Golovljova, I., Jaenson, T.G.T., Jensen, J.-K., Jensen, P.M., Kazimirova, M., Oteo, J.A., Papa, A., Pfister, K., Plantard, O., Randolph, S.E., Rizzoli, A., Santos-Silva, M.M., Sprong, H., Vial, L., Hendrickx, G., Zeller, H., Van Bortel, W., 2013. Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe. Parasites Vec. 6, 1. https://doi.org/10.1186/1756-3305-6-1.

Noll, M., Wall, R., Makepeace, B.L., Newbury, H., Adaszek, L., Bødker, R., Estrada-Peña, A., Guillot, J., da Fonseca, I.P., Probst, J., Overgaauw, P., Strube, C., Zakham, F., Zanet, S., Rose Vineer, H., 2023. Predicting the distribution of Ixodes ricinus and Dermacentor reticulatus in Europe: a comparison of climate niche modelling approaches. Parasites Vect. 16, 384. https://doi.org/10.1186/s13071-023-05959-y.

Open Data Science Europe – EU-wide automated mapping system for harmonization of Open Data based on FOSS4G and ML, n.d. URL https://opendatascience.eu/ (accessed 11.10.24).

Open Environmental Data Cube viewer [WWW Document], n.d. URL https://ecodatacube.eu/?base=OpenStreetMap%20(grayscale)&layer=NDVI%20Landsat%20(quarterly)%20-%20autumn&zoom=4&eye=5000000&center=57.2806,-4.2519&opacity=99&time=201909 (accessed 11.24.24).

Search: SPECIES: Ixodes (Ixodes) ricinus | Occurrence records | NBN Atlas [WWW Document], n.d. URL https://records.nbnatlas.org/occurrences/search?q=lsid%3ANHMSYS0000730335&fq=occurrence_status%3Apresent&fq=(identification_verification_status%3A%22Accepted%22%20OR%20identification_verification_status%3A%22Accepted%20-%20considered%20correct%22%20OR%20identification_verification_status%3A%22Accepted%20-%20correct%22)&nbn_loading=true (accessed 11.24.24).

Shcherbakov, M., Brebels, A., Shcherbakova, N.L., Tyukov, A., Janovsky, T.A., Kamaev, V.A., 2013. A survey of forecast error measures. World Appl. Sci. J. 24, 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032.

Signorini, M., Stensgaard, A.-S., Drigo, M., Simonato, G., Marcer, F., Montarsi, F., Martini, M., Cassini, R., 2019. Towards improved, cost-effective surveillance of *Ixodes ricinus* ticks and associated pathogens using species distribution modelling. Geospat. Health 14. https://doi.org/10.4081/gh.2019.745.

Stachurski, F., Boulanger, N., Blisnick, A., Vial, L., Bonnet, S., 2021. Climate change alone cannot explain altered tick distribution across Europe: a spotlight on endemic and invasive tick species. pp. 125–131. https://doi.org/10.1079/9781789249637.0018.

Tableau [WWW Document], n.d. URL https://www.tableau.com/products/tableau (accessed 11.24.24).

VectorMap [WWW Document], n.d. URL https://experience.arcgis.com/experience/5f95c3edfbea4634b8347fec0bd1dcd6/(accessed 11.24.24).

Wang, J., Zhu, C., Zhou, Y., Zhu, X., Wang, Y., Zhang, W., 2017. From partition-based clustering to density-based clustering: fast find clusters with diverse shapes and densities in spatial databases. IEEE Access 6, 1718–1729. https://doi.org/10.1109/ACCESS.2017.2780109.

Weather Data Services | Visual Crossing [WWW Document], n.d. URL https://www.visualcrossing.com/weather/weather-data-services (accessed 11.24.24).

Welcome to Python.org [WWW Document], 2024. Python.org. URL https://www.python.org/(accessed 11.24.24).

Yadav, S., Shukla, S., 2016. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). Presented at the 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 78–83. https://doi.org/10.1109/IACC.2016.25.

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A standard protocol for reporting species distribution models. Ecography. 43, 1261–1277. https://doi.org/10.1111/ecog.04960.