Kelly *et al.*
International Journal of Health Geographics (2025) 24:3
https://doi.org/10.1186/s12942-025-00388-9

International Journal of
Health Geographics

RESEARCH **Open Access**

Different environmental factors predict the occurrence of tick-borne encephalitis virus (TBEV) and reveal new potential risk areas across Europe via geospatial models

Patrick H. Kelly[1*], Rob Kwark[2], Harrison M. Marick[2], Julie Davis[3], James H. Stark[4], Harish Madhava[5], Gerhard Dobler[6] and Jennifer C. Moïsi[7]

Abstract

Background Tick-borne encephalitis (TBE) is the most serious tick-borne viral disease in Europe. Identifying TBE risk areas can be difficult due to hyper focal circulation of the TBE virus (TBEV) between mammals and ticks. To better define TBE hazard risks and elucidate regional-specific environmental factors that drive TBEV circulation, we developed two machine-learning (ML) algorithms to predict the habitat suitability (maximum entropy), and occurrence of TBEV (extreme gradient boosting) within distinct European regions (Central Europe, Nordics, and Baltics) using local variables of climate, habitat, topography, and animal hosts and reservoirs.

Methods Geocoordinates that reported the detection of TBEV in ticks or rodents and anti-TBEV antibodies in rodent reservoirs in 2000 or later were extracted from published and grey literature. Region-specific ML models were defined via K-means clustering and trained according to the distribution of extracted geocoordinates relative to explanatory variables in each region. Final models excluded colinear variables and were evaluated for performance.

Results 521 coordinates (455 ticks; 66 rodent reservoirs) of TBEV occurrence (2000–2022) from 100 records were extracted for model development. The models had high performance across regions (AUC: 0.72–0.92). The strongest predictors of habitat suitability and TBEV occurrence in each region were associated with different variable categories: climate variables were the strongest predictors of habitat suitability in Central Europe; rodent reservoirs and elevation were strongest in the Nordics; and animal hosts and land cover contributed most to the Baltics. The models predicted several areas with few or zero reported TBE incidence as highly suitable (≥ 60%) TBEV habitats or increased probability (≥ 25%) of TBEV occurrence including western Norway coastlines, northern Denmark, northeastern Croatia, eastern France, and northern Italy, suggesting potential capacity for locally-acquired autochthonous TBEV infections or possible underreporting of TBE cases based on reported human surveillance data.

Conclusions This study shows how varying environmental factors drive the occurrence of TBEV within different European regions and identifies potential new risk areas for TBE. Importantly, we demonstrate the utility of ML models to generate reliable insights into TBE hazard risks when trained with sufficient explanatory variables and to provide high resolution and harmonized risk maps for public use.

Keywords Tick-borne encephalitis, TBE, Ixodes, Machine-learning, Infectious disease modeling, TBE virus, TBEV

---

*Correspondence:
Patrick H. Kelly
patrick.kelly2@pfizer.com
Full list of author information is available at the end of the article

Background

Tick-borne encephalitis (TBE) is a major public health problem and the most serious tick-borne viral disease in Europe and Asia. Over the past decade, notification rates of TBE cases have increased throughout Europe including countries in the Nordics (Norway, Sweden, and Finland), Central Europe (Germany, Czech Republic, Austria) and the Baltics (Lithuania and Latvia) [1]. TBE is caused by infection with the TBE virus (TBEV), a single-stranded RNA virus belonging to the genus *Orthoflavivirus,* and most often spread via the bite of an infected tick but occasionally through the consumption of contaminated dairy products [2]. Two of the five confirmed TBEV subtypes are endemic in Europe and cause human disease: (i) the European subtype (TBEV-Eu) which is broadly distributed across the continent and transmitted mostly by *Ixodes ricinus* and *Dermacentor reticulatus*; and (ii) the Siberian subtype (TBEV-Si) which is endemic in more northern latitudes in eastern Europe and Asia transmitted predominantly by *Ixodes persulcatus* [3].

Circulation of TBEV in nature occurs between vector ticks and rodent reservoirs in hyper-localized, heterogeneous areas known as ‘foci’ which are determined by the interplay of environmental factors of climate, ecology, density of infected host-seeking ticks, and abundance of mammalian tick hosts and reservoirs [4]. Due to their strict focality in the environment, sometimes less than 100 m^2, defining TBE risk areas is challenging [5]. Usually, microfoci are identified based on TBE patients’ memory indicating potential locations in which the infected tick bites occurred [5–7]. This approach, however, is subject to substantial recall bias. Active tick surveillance via tick dragging is labor intensive and typically requires the collection of thousands of tick specimens to detect new TBEV foci due to the low prevalence of TBEV infection in host-seeking ticks [8, 9]. Sentinel surveillance of anti-TBEV antibodies or viral detection of RNA in rodent TBEV reservoirs can be more useful to define TBE risk areas by indicating current or recent TBEV transmission from infected ticks and identifying potential foci [10–12].

Robust statistical tools like machine-learning (ML) models are increasingly utilized to predict distributions of vector ticks and estimate spatiotemporal risk of TBE by linking their magnitude of occurrence to local environmental features of climate, habitat, animal hosts and reservoirs, and ecologies [13–16]. Therefore, ML model studies measuring associations between the environment and TBEV occurrence can help define TBE hazard risks by identifying suitable habitat areas and estimating the probability of local TBEV circulation [6, 7, 17–20]. Previous ML models, however, often lack regional specificity or do not include all relevant explanatory variables which limits data reliability and generalizability. For example, most predictive models are geographically broad in scope (e.g., European-wide) or fail to incorporate the distribution or abundance of animal hosts or rodent TBEV reservoirs as explanatory variables, though these are required for foci establishment and local circulation of the TBEV between ticks and animals [21, 22].

The primary objective of this study was two-fold. First, we sought to identify the environmental factors associated with habitat suitability and occurrence of TBEV across Europe via the development of predictive models within three geographically distinct regions: Central Europe, Nordics, and Baltics. We achieved this by building region-specific, novel ML models that utilize different approaches incorporating comprehensive and epidemiologically relevant explanatory variables of TBE zoonosis including features of climate, habitat, vector ticks, and animal hosts and TBEV reservoirs. We then leveraged these ML models to predict the distribution of suitable habitat areas and local probability of TBEV occurrence to illustrate high risk areas across varying population densities across the three regions.

Methods

Identification of source data and target territory

The target territory for source data and ML model development included the following countries: Austria, Belgium, Croatia, Czech Republic, Denmark, England, Estonia, Finland, France, Germany, Italy, Latvia, Lithuania, Netherlands, Norway, Poland, Scotland, Slovakia, Slovenia, Sweden, and Switzerland. Source records with coordinate locations of reported TBEV occurrence in the defined countries were retrieved from PubMed, grey literature (congress abstracts and unpublished data) in August 2023 and April 2024. Other records were identified during full-text reviews of retrieved studies that cited additional studies in the bibliography sections which we termed as “snowballing”. Literature searches of published records between 01 January 2000 and 28 May 2024 were guided by the Preferred Reporting Items for Systematic Review and Meta-Analyses Explanatory variables [23] using the following search terms and their analogues in various combinations in any language: tick-borne encephalitis, TBE, TBEV, foci, seroprevalence, antibody, Ixodes, ricinus, persulcatus, Dermacentor, reticulatus, Europe, Scandinavia, Baltics, Balkans, and target territory countries. Titles and abstracts from retrieved records were screened by two researchers to assess for eligibility and model inclusion. Original studies, case reports, narrative and comprehensive reviews, and systematic literature review articles were included for screening (Fig. 1). Additionally, five publications were used to assess the comprehensiveness and confirm the spatial specificity of extracted geocoordinates [6, 7, 20, 24, 25].

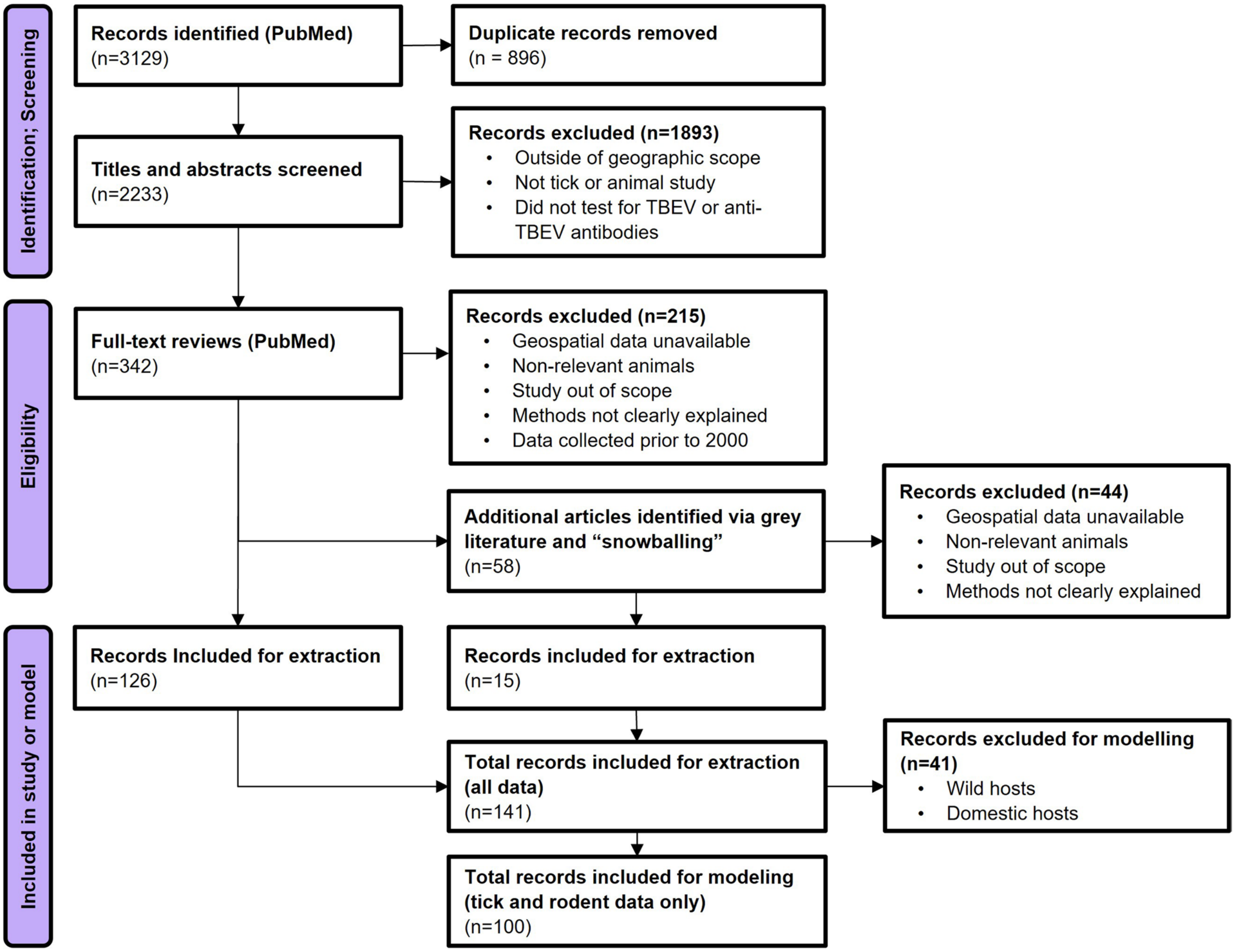

Fig. 1 Flow diagram of source records with coordinate locations with tick-borne encephalitis virus (TBEV) occurrence included for extraction and modeling

Inclusion criteria and extraction of TBEV occurrence geocoordinates

Geocoordinates of TBEV occurrence were grouped according to the biological categories involved in TBE zoonosis: "vector ticks", "rodent reservoirs", and "animal hosts." Coordinates were considered a location of "TBEV occurrence" and extracted for analysis if the original records detected TBEV RNA or anti-TBEV IgM/IgM antibodies in at least one individual vector tick, rodent reservoir, or animal host within the target territory countries via (i) detection of TBEV RNA in nymph or adult competent TBEV vector tick or competent TBEV rodent reservoir by molecular sequencing methods of PCR, RT-PCR, nested RT-PCR, quantitative PCR, pyrosequencing, or in vivo cellular culture; or (ii) serological detection of anti-TBEV IgM or IgG antibodies in a rodent reservoir or large mammalian tick-feeding host via ELISA, neutralization test, hemagglutinin inhibition, indirect fluorescent antibodies, or cellular culture. To account for potential short-term stability and possible disappearance for some coordinate locations of TBEV occurrence, sites identified prior to the year 2000 were excluded from analysis. Ticks and animals collected by either active or passive surveillance methods including tick dragging of host-seeking ticks, field trapping of rodent reservoirs and feeding ticks, or citizen science and community-associated data (hunting, veterinary clinics, etc.) were included for extraction.

Vector tick species included were *Ixodes ricinus, Ixodes persulcatus,* and *Dermacentor reticulatus.* Competent rodent reservoirs of TBEV included the yellow-necked mouse (*Apodemus flavicollis*), wood mouse (*Apodemus sylvaticus*), bank vole (*Myodes glareolus*), and European pine vole (*Microtus subterraneus*). Records that reported non-specific mammalian "rodents" or "*Ixodes* spp." were also included. Animal hosts were mostly larger mammalian hosts that are utilized for tick feeding and reproduction and categorized as either "domestic hosts" or "wild hosts." Domestic hosts included dogs (*Canis familiaris*) and agricultural animals such as cow and cattle species (*Bos* spp.), sheep (*Ovis aries*), goat (*Capra aegagrus hircus*), horse (*Equus ferus caballus*), pig (*Sus domesticus*), alpaca (*Vicugna pacos*); localities with documented TBEV-contaminated dairy products (e.g., cow milk) were also considered. Wild hosts were defined as species within the family *Cervidae* including roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), fallow deer (*Dama dama*), reindeer (*Rangifer tarandus*), moose (*Alces alces*), or any non-specified "deer" species, species within the family *Bovidae* including mouflon (*Ovus gmelini*) and European bison (*Bison bonasus*), and other wild animals including ibex (*Capra ibex*), red fox (*Vulpes vulpes*), wild boar (*Sus scrofa*), and any wild monkey or non-migratory bird species.

Geocoordinates were extracted and recorded in a Microsoft Excel database via the following methods:

1. Locations with associated decimal latitude and longitude coordinates were extracted directly as reported in the original records. Coordinates reported as degrees, minutes, seconds (DMS), Universal Transverse Mercator (UTM), or other non-typical coordinate systems were standardized to decimal latitude and longitude coordinates.
2. TBEV occurrence site(s) that did not provide geocoordinates and were described solely by location (e.g., address) or subdistrict administrative region had the centroid geocoordinates of the highest resolution administrative region extracted.
3. Among records that illustrated the spatial distribution of coordinate locations with TBEV occurrence but provided no additional geocoordinate specificity, geocoordinates were subjectively determined by scaling Google maps to the corresponding area and selecting the closest locality on the map.

The following data associated with the geocoordinates were also extracted: record source, publication year, collection site (country, state, NUTS-3 district), collection years (range), collection method, detection method, sample size of collected and tested organisms, number of organisms with TBEV detected, and mean infection rate. The entire master datafile of TBEV occurrence coordinates used in the study is provided (File S1).

Explanatory variables

We used an epidemiological-relevant comprehensive set of 28 environmental features important for the enzootic circulation of TBEV as explanatory covariates for ML model development based on previously published ecological observations [3, 4, 22, 26–28] and modeling approaches [1, 6, 7, 15–18, 20, 29–32] for TBE. The 28 individual explanatory covariates were grouped according to the environmental categories of "climate" (n = 19), "topography" (elevation and land cover type), competent rodent "reservoir" species of TBEV (n = 4), and large mammalian "host" species ideal for tick feeding and reproduction (n = 3) (Table 1). The geospatial data associated for all 28 explanatory variables were extracted across the entire target territory with the addition of Luxembourg to build a comprehensive explanatory variable dataset for ML model development. Daily records of variables for climate and weather were obtained from the Copernicus E-Obs dataset (https://surfobs.climate.copernicus.eu/surfobs.php) and summarized at monthly, seasonal, or yearly intervals over the temporal period (2000–2022) at 0.1-degree grids (accessed 12 October 2023) [33]. Gridded land cover data at a spatial resolution of 100 m were obtained from the European Environmental Agency (EEA) (https://www.eea.europa.eu/) dataset classified under the EUNIS Level 2 system last updated in 2018 (v3.1) (accessed 31 March 2024) [34]. Data for "Elevation" were obtained from the EuroDEM dataset (https://www.mapsforeurope.org/datasets/euro-dem) updated in 2023 at 2 arcsecond resolution (accessed 12 October 2023) [35]. Explanatory variables for the distribution for red deer (*C. elaphus*) and roe deer (*C. capreolus*) animal hosts were obtained from high resolution maps of predicted abundance (0–1) at a spatial scale of < 0.01 degrees from published literature [36, 37]. Explanatory variables of the distributions of the animal host fallow deer (*D. dama*) and TBEV rodent reservoirs of yellow-necked mouse (*A. flavicollis*), European pine vole (*A. sylvaticus*), bank vole (*M. glareolus*), and wood mouse (*M. subterraneus*) were based on each species' reported historical spatial occurrence data beginning in the year 1600 through September 2023, publicly available from the Global Biodiversity Information Facility (GBIF) (https://www.gbif.org) [38] to calculate their respective relative habitat suitability (0–1) at a 3 km resolution across the target territory and provided in March 2024 (courtesy of Dr. Agustin Estrada-Pena). Further details and definitions on all explanatory variables used for modeling are provided (File S2).

Data structure and processing

Extracted geocoordinates were prepared and mapped to their respective NUTS-3 administrative region. Geocoordinates located in offshore locations (sea, ocean, or otherwise not on land) were verified for accuracy according to the reported coordinates in the original records. If these coordinates were still considered to be reported in error, the coordinates were mapped to the nearest land coordinate in the collection country. Rarely, nonsensical or otherwise unreliable geocoordinates were discarded.

Table 1 Variables and data sources included for modeling

Model component (n)	Variable (n)	Spatial resolution	Data range (years)	Interval	Source(s)
Training dataset (1)	Geocoordinates of TBEV occurrence (1)	GPS coordinates	2000–2022	Annual	Literature search; personal communication; congress abstracts
Explanatory variables (28)	Climate; weather (19)	0.1°	2000–2022	Daily	Copernicus E-Obs Dataset [33]
	Relative habitat suitability of rodent TBEV reservoir species: *Apodemus flavicollis; Myodes glareolus* (2)[a]	~ 55 m	1736-September 2023	Annual reported occurrences	GBIF; https://www.gbif.org/species/2437756; https://www.gbif.org/species/5706764 [38]
	Relative habitat suitability: *Apodemus sylvaticus* (1)[a]	~ 55 m	1600-September 2023	Annual reported occurrences	GBIF; https://www.gbif.org/species/2437760 [38]
	Relative habitat suitability: *Microtus subterraneus* (1)[a]	~ 55 m	1838-September 2023	Annual reported occurrences	GBIF; https://www.gbif.org/species/2438660 [38]
	Relative habitat suitability of fallow deer: *Dama dama* (1)[a]	~ 55 m	1575-September 2023	Annual reported occurrences	GBIF; https://www.gbif.org/species/5220136 [38]
	Predicted distribution and abundance: *Capreolus capreolus* (1)	< 0.01°	Uploaded April 2014	Static	Alexander et al., 2014 [37]
	Predicted distribution and abundance: *Cervus elaphus* (1)	< 0.01°	Uploaded April 2014	Static	Wint et al., 2014 [36]
	Elevation (1)	2 arcseconds (60 m)	Updated 2023	Static	EuroDEM [35]
	Habitat and land cover types (1)	100 m	1 Jan 2012 – 31 Dec 2012	Static	European Environmental Agency Land cover (EUNIS Level 2 Classification—v3.1 2018 [34]

GBIF Global Biodiversity Information Facility

[a] Data provided March 2024

Geocoordinates were classified as either present or absent (e.g., geocoordinates with one TBEV-infected tick out of 100 sampled specimens and geocoordinates with 99 TBEV-infected ticks out of 100 sampled specimens were both counted and contributed equally to the model outputs). Spatial raster files corresponding to the explanatory variables were harmonized using QGIS software (version 3.32.1-Lima) at the 0.005-degree resolution to create a composite raster image. Variables at lower resolutions were joined to the nearest raster geocoordinate of the composite raster image to create a final explanatory variable predictor spatial dataset at 0.005-degrees resolution. Missing values were only present for lower resolution explanatory variable data, which was resolved by matching data from the nearest coordinate. TBEV occurrence associated with vector ticks and rodent reservoirs exclusively were used for model training due to increased spatial precision of potential microfoci. TBEV occurrence associated with animal hosts were excluded due to surveillance bias of domestic host species and poor spatial specificity among wild host collection sites. Geocoordinates were de-duplicated to eliminate multiple collection records at each coordinate location and minimize model bias.

A modified K-means algorithm was implemented across the target territory using the distribution of the entire dataset of extracted geocoordinates of TBEV occurrence to develop separate ML models for three defined regions. The rationale for this procedure is due to the hypothesis that the eco-epidemiology of TBE is region-specific and has different associations with ecological, climatological, and habitat-related features in the environment across regions. The K-means clustered algorithm considered the latitude and longitude coordinates of the TBEV occurrence and whether the coordinates were located north or south of the Baltic Sea to ensure resulting clusters did not cross major bodies of water (remote islands and the United Kingdom (UK) were excluded from this rule). TBEV occurrence and explanatory variables were split according to the defined geographic clusters, aggregated, and matched to the nearest explanatory variable datum within each defined region to construct separate regional datasets for model training ("training dataset").

Machine-learning model algorithm approaches

Three ML algorithm approaches were selected to create models of TBEV "risk" based on their reliability and utility to measure environmental associations with TBEV occurrence: Maximum entropy (Maxent), regularized logistic regression (RLR), and extreme gradient boosting (XGBoost). Maxent was chosen since it does not require absence (negative) geospatial data and due to its demonstrated utility for ecological niche modeling when predicting spatial distribution and occurrence [39]. In this study, we refer to the Maxent model outputs as "habitat suitability" which compares the predicted spatial suitability of TBEV occurrence relative to other areas within the model region. While linear models need non-linear relationships and variable-outcome interactions implicitly defined, we developed extreme gradient boosted decision tree (XGBoost) models that can implicitly handle non-linear and multifactorial relationships between explanatory variables and target outcomes [40]. XGBoost models are extremely powerful and can provide highly accurate predictive outputs when tuned and validated appropriately. Predictive outputs from XGBoost models were defined as "probability of TBEV occurrence" in each region.

Model-specific input data processing

Due to the lack of available absence data of TBEV foci localities, we incorporated 10,000 randomly distributed "pseudoabsence" points for the XGBoost and RLR models across the entire study area proportional to each cluster regions' landmass (Central Europe = 4972; Nordics = 3868; Baltics = 1160). Pseudoabsence points were incorporated to create a meaningful contrast with the extracted geospatial TBEV occurrence points, enabling the model to better distinguish areas where the species is likely to occur from areas where it is absent [41]. This approach enhances the models' ability to better capture the environmental conditions associated with the species' distribution ultimately increasing the models' performance. Furthermore, this method aligns with established practices in species distribution modeling, as several studies have successfully utilized pseudoabsence points to achieve robust results [18, 20]. Absent land cover types were systematically sampled and incorporated as pseudoabsence points into the training datasets. One hot encoding was used to extract each categorical land cover type (n = 45) into separate binary explanatory variables. Both the RLR and XGBoost algorithms used the 'saga' solver with a maximum iteration of 10,000 with L1 regularization to reduce overfitting using the scikit-learn package in Python (v. 1.4.2). With the Maxent models, the number of background points used for each region was the same as the number of pseudoabsence points for the XGBoost and regularized logistic regression models. The Maxent model was developed using a regularization value = 2 with 1000 maximum iterations using the Maxent software package (version 3.4.4).

Machine-learning model training and feature selection

Ultimately, with three modeling approaches across three European regions, nine models were developed in total. To improve interpretability of the models and reduce multicollinearity bias among explanatory variables, feature selection was implemented using Pearson's correlation coefficient (R-value). We implemented a correlation coefficient threshold of $R = 0.95$ and a statistical significance threshold of $p = 0.01$ to identify correlated pairwise explanatory variables. The ML models within each region were then trained and tuned to account for potential overfitting using a grid search approach to find optimal hyperparameters. The list of explanatory variables excluded from the final algorithms in each region is provided (File S3).

Feature importance scores for each respective model were obtained. Due to the differences between the XGBoost, RLR, and Maxent models, we used different approaches to measure variable importance for each model type. The decision to use different importance metrics for each model stemmed from the structural difference between them: our XGBoost is a tree-based model, RLR is a generalized linear model, and Maxent is an entropy-based model. For the Maxent models, we used Permutation Importance, which measures variable importance by assessing the percent drop in AUC when randomly permuting the values of a given predictor. For the XGBoost models, we used the traditional feature importance metric, calculated as the average gain of splits in a tree using a given feature. For the RLR models, since training data was normalized, we used the absolute value of the traditional linear coefficient values to determine feature importance.

Among the pairwise variable associations that met the defined R-value threshold for collinearity, the explanatory variable that possessed the lower permutation importance or feature score was dropped from each model. Afterwards, final models were developed using the remaining explanatory variables within each region to obtain output predictions and feature importance (%). To compare the importance of habitat land cover types between the XGBoost or RLR models and the Maxent models within each region, the variable scores of the one hot encoded binary variables for all land cover types in the XGBoost and regularized logistic regression models were aggregated into a single "land cover" variable score by computing the weighted mean of their scores, with weights proportional to the frequency of each land cover type in the region (e.g., training dataset). Feature importance of the explanatory variables within the categories for climate, rodent reservoirs, and animal hosts were aggregated to provide insights into the proportional (%) model contributions of each category across regions.

ML model output predictions and performance evaluation

The spatial distribution of Point predictions (0–100) for habitat suitability (Maxent) and probability of TBEV occurrence (XGBoost and RLR) were output at 0.01-degree resolution scale and 0.005-degree resolution scale, respectively. Final validation of each model was performed using fivefold cross-validation by NUTS-3 districts in each region to measure model Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) results and assess model evaluation and performance. The 5 folds were constructed based on the clustered geographic regions (Central Europe, Nordics, and Baltics) and split according to their respective NUTS-3 districts. For each fold of the fivefold cross-validation process, 80% of the NUTS-3 districts were randomly selected within each region/model for training and the remaining 20% of the NUTS-3 districts were used as a test set to measure the regions' model accuracy. This procedure was repeated another four instances (five total) for each fold to ensure the entire dataset across all NUTS-3 districts had been used for both model training and testing without artificially shrinking the training dataset. The final AUC scores were computed using aggregated predictions across the five test sets and evaluated for model selection. Point predictions were further aggregated up to their respective NUTS-3 levels to produce mean prediction and probability of TBEV occurrence in each NUTS-3 district for further visualization and analytical purposes. All programming language and resulting code are available for download. A schematic describing the procedure for data mining and extraction, aggregation and cleaning of explanatory variables, development of the ML algorithms, and evaluation of model performances is provided (Fig. 2).

Association between human population density and predicted risk

Model results were used to further investigate the relationship between population density and predicted probability of TBEV occurrence (XGBoost) among NUTS-3 districts in the target territory. Data for the total population and landmass were gathered from Eurostat database (2024) (https://ec.europa.eu/eurostat/en/) (accessed 02 July 2024) [42] to calculate population density (km^2) at the NUTS-3 level for all countries in the target territory (2024) and the UK (2019). Missing NUTS-3 region data for the UK were supplemented with population density data from the Office of National Statistics (ONS) in 2024 (https://www.ons.gov.uk/) (accessed 02 July 2024) [43]. Descriptive summaries of the absolute and proportion of NUTS-3 districts with population densities of $<300/\text{km}^2$, $300–1500/\text{km}^2$, and $>1500/\text{km}^2$ relative to thresholds for predicted probabilities of TBEV occurrence of $<1\%$, 1–5%, 5–15%, and $>15\%$ were calculated and visualized for each region.

Results

Summary of TBEV foci data and model regions

Among 2233 titles and abstracts screened, 400 (17.9%) records (342 from online database search and 58 from "snowballing") were full-text reviewed (Fig. 1). In total, 141 records were included (File S4) for the extraction of 899 coordinate locations with reported TBEV occurrence (2000–2022) (Table 2; Fig. 1). Of the entire extracted dataset, 343 (38%) coordinates were associated with domestic host species (214) and wild animal hosts (129) from 41 records and excluded from modeling resulting in a total of 100 records included for analyses (Table 2; Fig. 1). After de-duplication of co-localized

Table 2 Summary of coordinate locations with reported occurrence of tick-borne encephalitis virus (TBEV) included in final models, by region (2000–2022)

Cluster region model	Total coordinates extracted[a]	No. coordinates for modeling (source records)[b]	*Ixodes ricinus*	*Ixodes persulcatus*	*Ixodes* spp.	*Dermacentor reticulatus*	Rodent reservoirs	Range years collection
Central Europe	565	294 (55)	248	0	1	1	44	2000–2022
Nordics	166	100 (24)	68	23	7	0	2	2003–2021
Baltics	168	127 (21)	49	0	48	10	20	2000–2021
Total	899	521 (100)	365	23	56	11	66	2000–2022

[a] Total number of coordinates for TBEV occurrence extracted for all ticks and animals (domestic hosts, wild hosts, and rodent reservoirs)

[b] Coordinates of TBEV occurrence for model training included ticks and rodent reservoirs, exclusively. Locations with TBEV occurrence were represented by one set of coordinates only

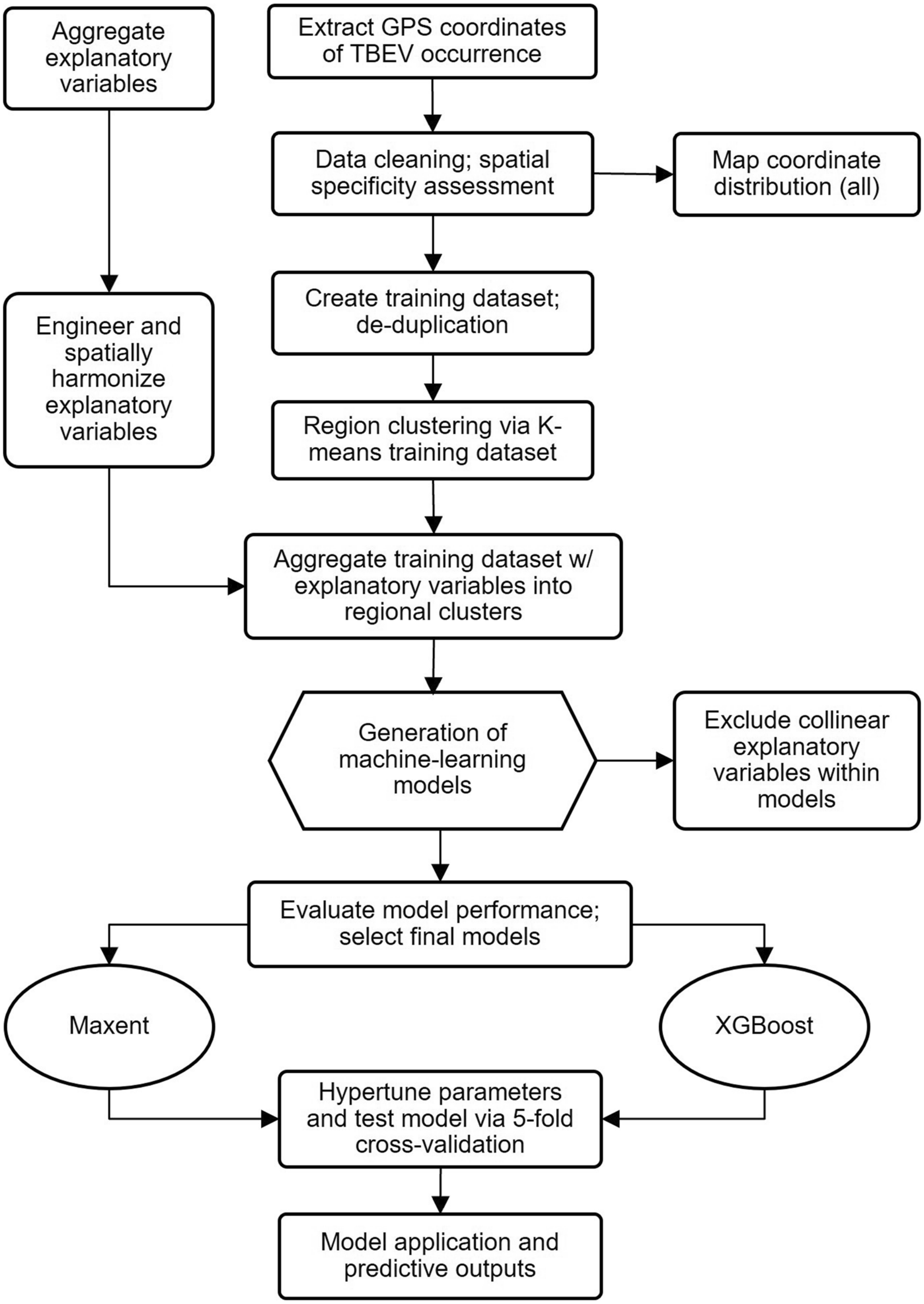

Fig. 2 Diagram of study procedures for modeling predicted habitat suitability and occurrence of tick-borne encephalitis virus (TBEV) via maximum entropy (Maxent) and extreme gradient boosting (XGBoost) machine-learning algorithms

geocoordinates, the final models were trained on a total of 521 coordinate locations (training dataset) of TBEV occurrence from vector tick species (455; 87.3%) and rodent reservoirs (66; 12.7%) (Table 2; Fig. 1). The distribution of all 899 extracted geocoordinates TBEV occurrence is illustrated (Figure S1) and summarized by country (File S5).

A spatial clustering analysis to define the geographic scope for model training revealed three geographic clusters encompassing countries in the regions of "Central Europe", the "Nordics", and "Baltics" (Fig. 3A). The "Central Europe" regional model covered the UK, France, Belgium, the Netherlands, Denmark, Germany, Luxembourg, Switzerland, most of Czech Republic and Slovakia, and western Poland; the "Nordics" region included Norway, Sweden, and Finland; and the "Baltics" model ranged from central and eastern Poland, the easternmost part of Czech Republic, northeastern Slovakia, and the three Baltic states of Lithuania, Latvia, and Estonia. More than half (n = 294; 56.4%) of all coordinates of TBEV occurrence were distributed within Central Europe (2000–2022), 127 (24.4%) were within the Baltics (2000–2021), and 100 (19.2%) were in the Nordics (2003–2021) (Fig. 3A; Table 2). Overall, 455 (87.3%) coordinates were identified by the collection and detection of TBEV-infected vector ticks, mostly in *I. ricinus* (n = 365) followed by *Ixodes* spp. (n = 56), *I. persulcatus* (n = 23) and *D. reticulatus* (n = 11) and 66 (12.7%) coordinates were identified by the detection of TBEV RNA or anti-TBEV antibodies in rodent reservoirs (Table 2). Based on land cover type, more coordinates occurred in coniferous woodland (n = 123; 23.6%) and broadleaved deciduous woodland habitats (n = 96; 18.4%) than any other land cover types, followed closely by arable land and gardens (n = 84; 16.1%), buildings of cities, towns, and villages (n = 66; 12.7%), and mesic grasslands (n = 61; 11.7%) (Fig. 3C; File S6). By region, coordinates of TBEV occurrence from the Baltics were more equally distributed across vector tick species and rodent reservoirs compared to Central Europe and Nordics regions (Fig. 3B). Conversely, coordinates of TBEV occurrence were distributed across more land cover types in Central Europe relative to the Nordics and Baltics (Fig. 3C).

Evaluation of algorithm performance and feature selection

The Maxent algorithms were superior in Central Europe (AUC: 0.885) and Baltics (AUC: 0.774) while the XGBoost algorithm was the best performer in the Nordics (AUC: 0.918) (File S7). The RLR models had the lowest performance characteristics across all regions and models and were excluded for further analyses (File S7). Ultimately, both the Maxent and XGBoost models were selected for predictive outputs in each region. Four to nine explanatory variables were excluded from the regions' final models based on multicollinearity assessments (File S3). To illustrate regional disparities among individual predictor features across the modeled regions, the spatial distributions and magnitudes of a subset of explanatory variables corresponding to categories of topography (elevation; Fig. 4A), climate (mean daily temperature in driest annual quarter, mean ratio days per month > 5 °C, mean annual "wet" days; Fig. 4B–D), hosts (*D. dama*; Fig. 4E), and reservoirs (*A. flavicollis*; Fig. 4F) are provided.

Environmental variable associations with habitat suitability and occurrence of TBEV

The predicted relative habitat suitability (Maxent) for TBEV was broadly driven by different category environmental variables, although some similarities were observed. Climate variables contributed (permutation importance) the most (72.2%) for habitat suitability in Central Europe compared to contributing less than half (43.6%) and less than one-third (33.2%) in the Nordics and Baltics regions, respectively (Fig. 5A). Variables of animal hosts contributed the most (27.9%) to habitat suitability in the Baltics and reservoirs (20%) and topography (26.2%) were the strongest in the Nordics. Overall, model contributions from reservoirs (14.1%), topography (10.9%), andhosts (2.9%) were the lowest in Central Europe. Similarly, with the Maxent models, variables of climate had the largest contributions (feature scores) to the XGBoost models to predict the absolute probability of TBEV occurrence across the regions but the proportional contributions across categories were mostly similar (Fig. 5B). Overall, climate variables scored the highest in the XGBoost model for Central Europe (66.3%) compared to the Nordics (53.5%) and Baltics (58.7%) models. Conversely, contributions from variables of topography (13.6%), hosts (14.8%), and reservoirs (18.1%) were the highest in the Nordics.

Model contributions respective to each individual predictor were often similar between the Maxent and XGBoost models within each region but some discrepancies were observed (Fig. 6). In Central Europe, "mean monthly ratio days > 5 °C" was the strongest predictor for habitat suitability (30.3%) but excluded from the XGBoost model, while "mean June temperature" was the top-ranked predictor in the XGBoost model (7%) and excluded from the Maxent model. For the Baltics, the fallow deer abundance was the top predictor for relative habitat suitability (21.3%) while roe deer abundance was the top driver for the predicted absolute probability of TBEV occurrence (5.9%). Elevation was ranked in the top-three for all regional habitat suitability models and the top-ranked predictor in the Nordics for both habitat suitability (23.5%) and absolute probability of TBEV occurrence (10.2%). Land cover was the second-ranked variable for both Maxent (11.3%) and XGBoost models (5.6%) in the Baltics. Variables for reservoirs contributed the most in the Nordics for both models but were the weakest category of predictors overall across all regions. The three variables of relative humidity were among the lowest ranked predictors across all regions and models.

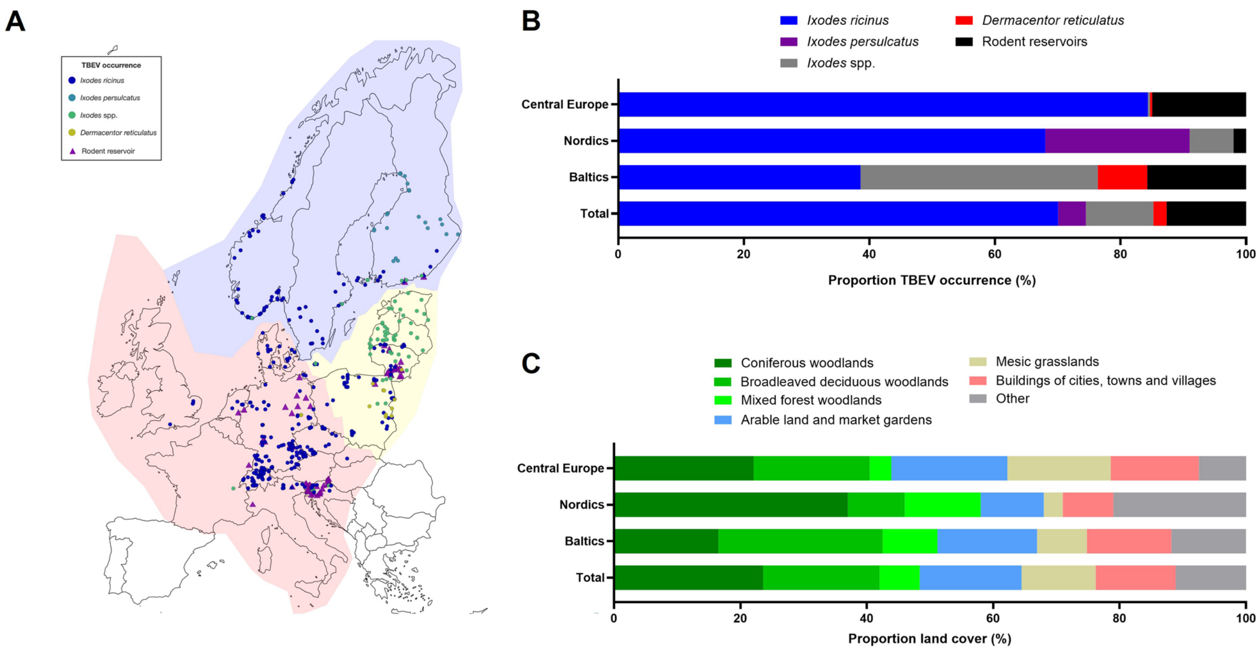

Fig. 3 Locations with reported occurrence of tick-borne encephalitis virus (TBEV) in vector ticks and rodent reservoirs within European regions (2000–2022). Clustered regions for "Central Europe" (red), "Nordics" (blue), and "Baltics" (yellow) and their associated models are illustrated with coordinate locations of TBEV occurrence for the main TBEV vector tick species and rodent reservoirs (**A**). The proportion of TBEV occurrence locations associated with each of the vector tick species and rodent reservoirs (**B**) or land cover habitat types (**C**) by region are provided

Although variable contributions from the explanatory variable categories were broadly similar across the XGBoost modeled regions, different individual predictors within each category were responsible for their respective contributions in each region (Fig. 6A, C). Among climate variables, "mean June temperature" was the strongest predictor in the overall XGBoost model in Central Europe (7%) and the fourth-ranked predictor (5%%) in the Baltics models but was excluded from the Nordics. Further, "mean spring 'warm' days" was the second strongest individual predictor in the Central Europe but was excluded from the XGBoost models for the other regions (Fig. 6A). Conversely, "mean monthly minimum temperature" was the second strongest predictor in the Nordics (7%) but was the 13th-ranked variable for both other regions and "mean monthly ratio days > 5 °C" was the top predictor of climate in the Baltics (5.1%) but excluded from the other regions' XGBoost models (Fig. 6A, C). Additionally, the top ranked animal host predictor for each region was a different species: roe deer abundance was top-ranked animal host variable in the Baltics while fallow deer abundance (5.8%) and red deer abundance (4.6%) were the highest-ranked animal host variables in Nordics (4th overall) and Central Europe (8th overall), respectively (Fig. 6C). The contributions from the category and individual explanatory variables in the Maxent and XGBoost models for each region are listed in Table S2.

Predicted habitat suitability and probability of TBEV occurrence

Spatial predictions for the relative habitat suitability (Fig. 7A-B) and absolute probability of TBEV occurrence (Fig. 7C-D) were generated at high-resolution (0.01 degree) and administrative NUTS-3 regions. For the Maxent models, 90% of the predictive outputs (0.01 degree) were between 0 and 52.2% relative habitat suitability, 95% of predictions between 0 and 68.2% relative habitat suitability, and 99% of predictions between 0 and 92.0% relative habitat suitability (Fig. 7A). Areas with the highest predictions of habitat suitability (≥ 60%) were southern Germany, northern regions of Switzerland, Italy, and Croatia, eastern Latvia, and several coastal regions of the Nordics including western and southern Norway, western Finland, and southern Sweden (Fig. 7A-B). Other notable regions with increased (10–60%) habitat suitability for TBEV were eastern France, most of southern Sweden and Finland, some eastern regions of Scotland and England, and nearly all of Poland, Lithuania, and Estonia (Fig. 7A-B). For the XGBoost models, 90% of the predictive outputs

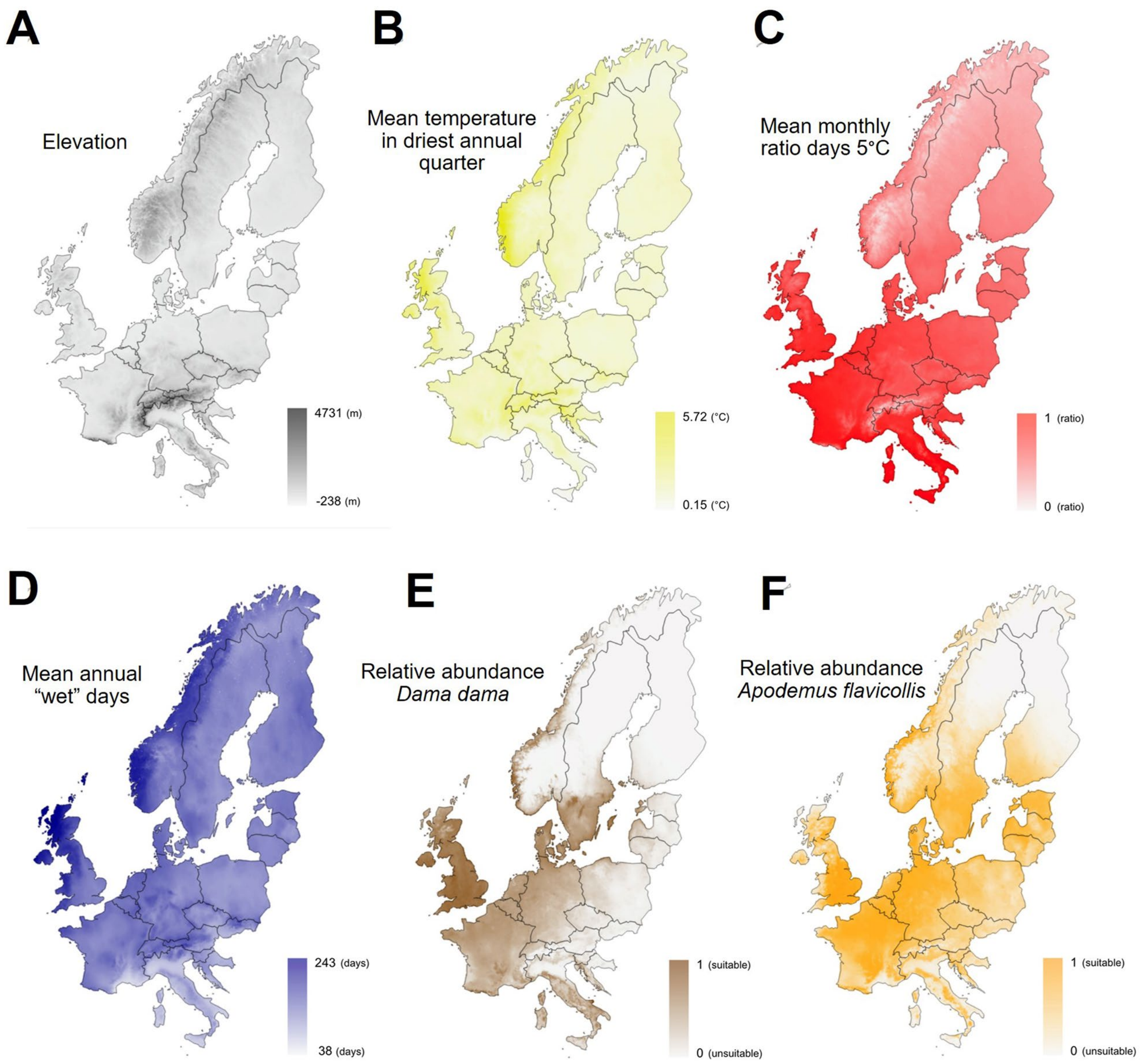

Fig. 4 Spatial distributions and magnitudes of a subset of individual predictors within environmental variable categories of topography, climate, animal hosts, and rodent reservoirs for machine-learning models. Regional disparities across the modeled regions can be seen in the maps for the individual predictors of elevation (**A**), "mean daily temperature in driest annual quarter" (**B**), "mean ratio days per month > 5 °C" (**C**), "total annual 'wet' days" (**D**), and relative suitability of fallow deer (*Dama dama*) (**E**) and yellow-necked mouse (*Apodemus flavicollis*) (**F**). Further details on the definitions of individual predictor variables are provided in File S2

were between 0 and 10.1% absolute probability of TBEV occurrence, 95% of predictions falling between 0 and 15.5%, and 99% of predictions were between 0 and 29.5% (Fig. 7C). The areas with the highest predicted absolute probabilities of TBEV occurrence (≥ 25%) across the modeled territory included southern Germany, most of Switzerland but especially concentrated in the northern regions, eastern Slovenia and bordered areas with Croatia, northeastern Poland, and western and central Baltics, as well as the coastal regions of western Norway, southern Finland, and Danish islands (Fig. 7C-D).

We assessed the total number and proportion (%) of NUTS-3 districts (n = 1155) with population densities < 300/km^2 (n-761), 300–1500/km^2 (n = 266), and ≥ 1500/km^2 (n = 128) that had predicted absolute probabilities of TBEV occurrence of < 1%, 1- < 5%, 5- < 15%, and ≥ 15% in each region (Fig. 7D; Fig. 8A-B).

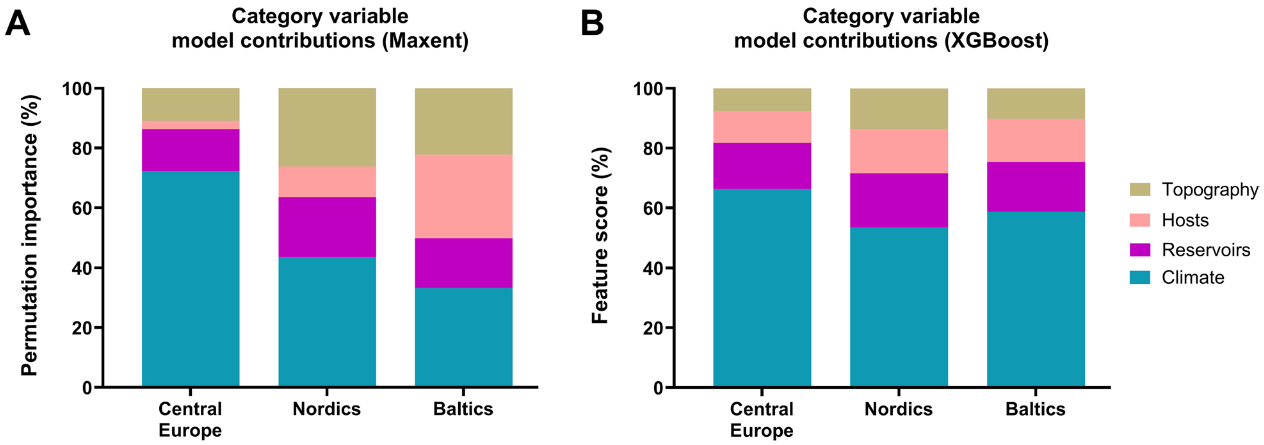

Fig. 5 Categorical explanatory variable contributions and ranking in maximum entropy (Maxent) and extreme gradient boosting (XGBoost) models in Central Europe, Nordics, and Baltics regions. Total contributions from the variable categories to predict habitat suitability (Maxent) (**A**) and probability of tick-borne encephalitis virus (TBEV) occurrence (XGBoost) (**B**) are shown in aggregate for each region. Further details on the variable contributions for each region and model are provided in Supplemental Materials

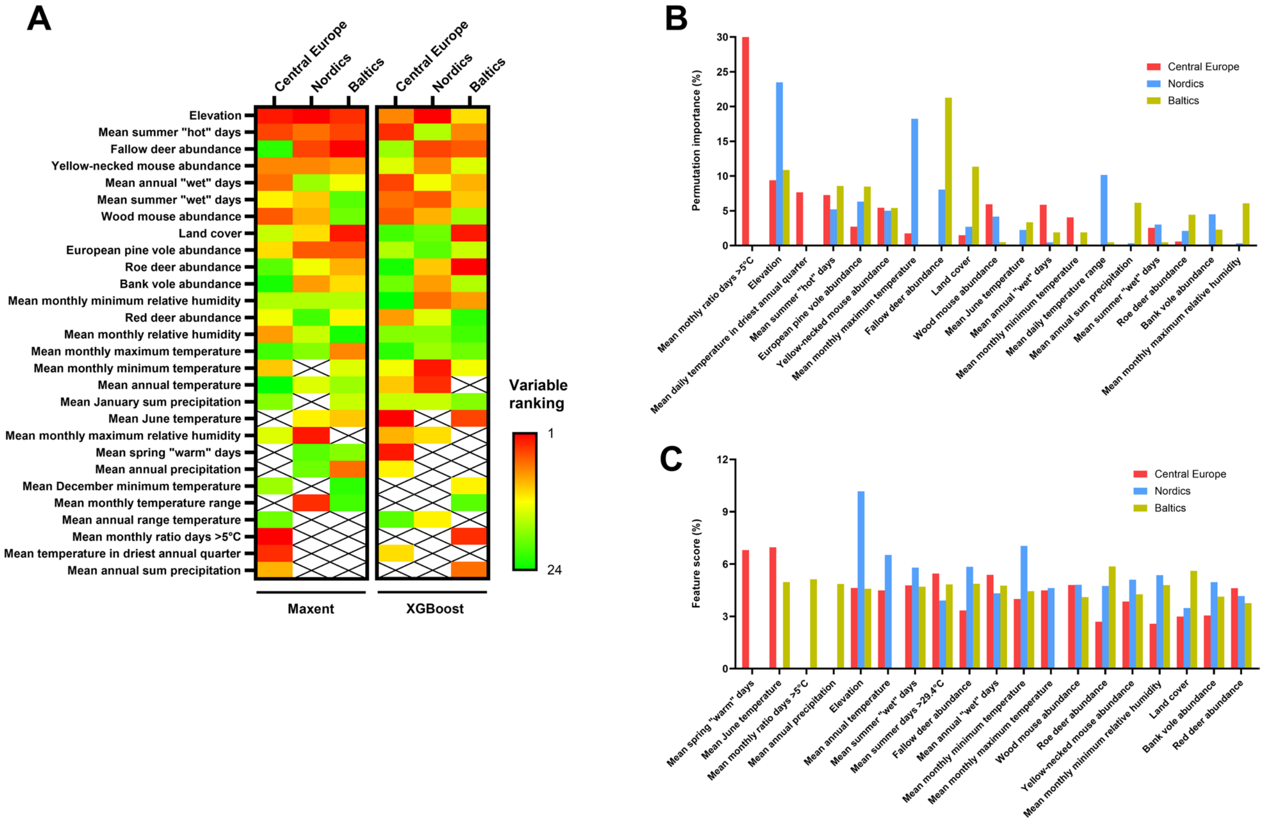

Fig. 6 Individual explanatory variable contributions and ranking in maximum entropy (Maxent) and extreme gradient boosting (XGBoost) models in Central Europe, Nordics, and Baltics regions. Variables included in each model are individually ranked according to their respective contributions to predict habitat suitability (permutation importance) or probability of tick-borne encephalitis virus (TBEV) occurrence (feature score) (**A**). The individual contributions of the top ten ranked variables for each region are illustrated in Maxent (**B**) and XGBoost (**C**) models. Further details on the variable contributions for each region and model are provided in Supplemental Materials

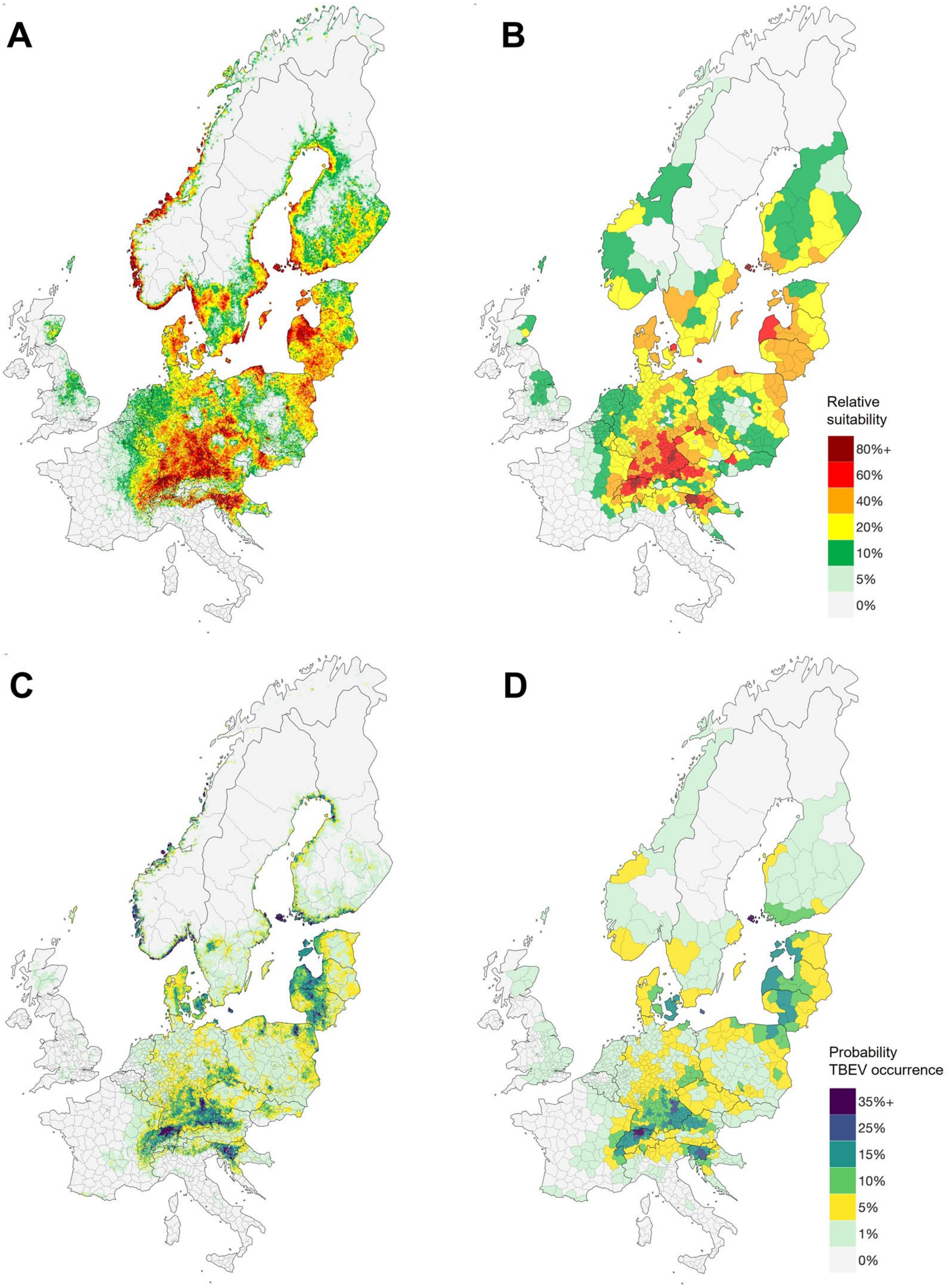

Fig. 7 Predicted habitat suitability (**A**, **B**) and probability of tick-borne encephalitis virus (TBEV) occurrence (**C**, **D**) in Europe based on machine-learning algorithms. Predicted outputs for the relative habitat suitability (0–100) and absolute probability of TBEV occurrence were generated via maximum entropy (Maxent) and extreme gradient boosting (XGBoost), respectively. Maps are shown at spatial resolutions of 100 m (**A** and **C**) and aggregate NUTS-3 (**B** and **D**) scales. Note: maps in each figure are the combined predictive outputs generated from each regional model associated with "Central Europe", "Nordics", and "Baltics"

Across all regions and districts, the XGBoost model predicted the majority (72.3%) to have a 1-<5% (n = 467) or 5-<15% (n = 368) absolute probability of TBEV occurrence (Fig. 8A). Twenty-one NUTS-3 districts in Central Europe with population densities ≥ 1500/km^2 had predicted absolute probabilities of TBEV occurrence

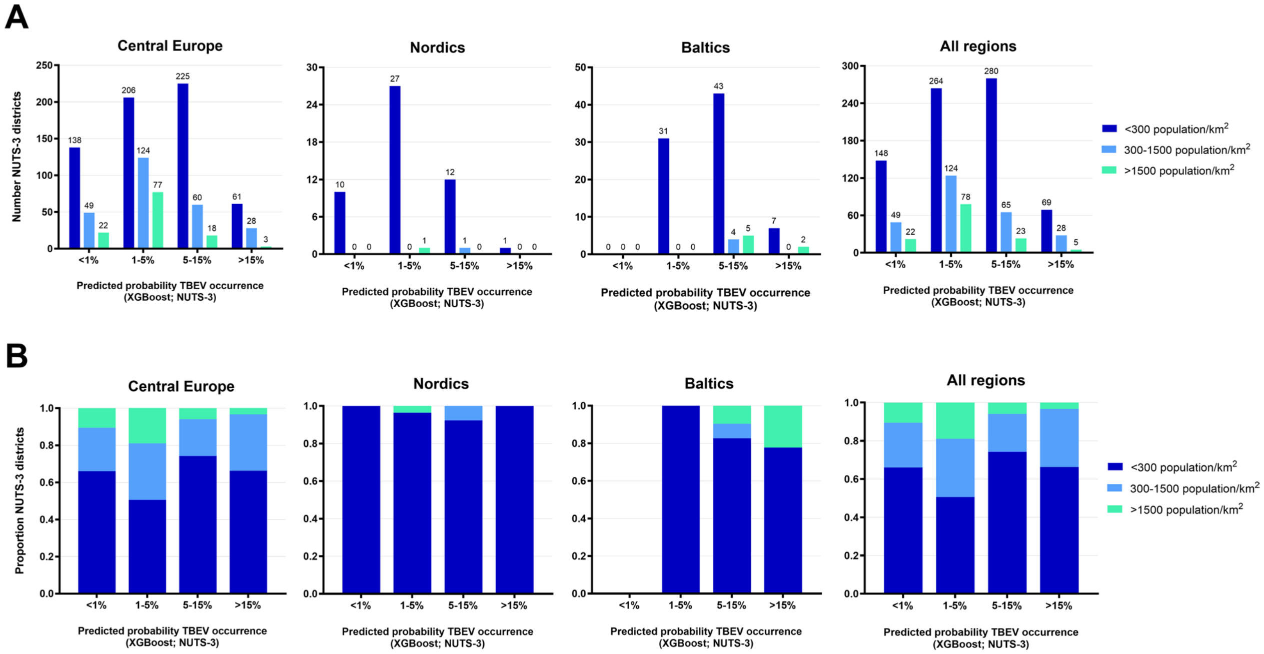

Fig. 8 Qualitative assessment of potential risk and local exposure to tick-borne encephalitis virus (TBEV) based on the predicted absolute probability of TBEV occurrence relative to population density among NUTS-3 districts in Europe. Predictive outputs for the probability of TBEV occurrence (x-axis) were generated at the NUTS-3 administrative level via extreme gradient boosting (XGBoost) machine-learning models and plotted relative to their respective population densities (/km^2) (y-axis) as reported by the European Union (2019 and 2023). NUTS-3 districts were categorized as population densities of < 300/km^2, 300–1500/km^2, and ≥ 1500/km^2 and grouped accordingly to the corresponding predicted probability of TBEV occurrence of < 1%, 1– < 5%, 5– < 15%, and ≥ 15%

of 5- < 15% (n = 18) or ≥ 15% (n = 3) (Fig. 8A) and not a single district in the Baltics regions (n = 92) had a predicted absolute probability of TBEV occurrence < 3.2% (Fig. 8A-B).

Discussion

In this study, we used multiple ML model approaches to generate high-resolution risk maps for TBE illustrating the predicted habitat suitability and probability of TBEV occurrence within three distinct regions of Europe. By incorporating sufficient explanatory variables related to features of climate, habitat, and abundance of animal hosts and rodent reservoirs, we demonstrated high ML model performance and accuracy to identify the region-specific environmental factors associated with the occurrence of TBEV-infected ticks and rodent reservoirs. We also showed how different environmental variables were differentially associated with habitat suitability and TBEV occurrence across regions, emphasizing that multiple habitat ecotypes throughout Europe can support the establishment and stability of TBE foci. This novel and comprehensive approach demonstrates the utility of ML models to improve the precision and accuracy of TBE risk maps compared those available from public health agencies and national governments [3].

Comparing the explanatory variables associated with habitat suitability and predicted TBEV occurrence among regions

We observed differences in the environmental category variables associated with habitat suitability and probability of TBEV occurrence among model regions. Among abiotic variables, features of climate contributed the most predictive power in Central Europe while elevation and land cover types more strongly contributed to the models in the Nordics and Baltics, respectively. Our results corroborate previous ML models in the central European countries of Germany and Czech Republic that report weather variables, like evapotranspiration, hot summer days, and increased temperature, are the strongest predictors of *I. ricinus* densities or TBEV foci relative to other environmental variables [6, 16, 44, 45] as well as ML models in the Baltic countries of Latvia and Lithuania that report greater influence of landscape on TBE incidence [46–48]. Elevation was ranked in the top-three variable predictors for habitat suitability in all regions and was the strongest predictor of TBEV occurrence in the Nordics. Although XGBoost and Maxent models do not provide directional (positive vs. negative) coefficients for feature scoring and model importance, we can reliably conclude from our results that local TBEV transmission is unlikely to occur at altitudes over 600 m asl in the Nordics due to decreased abundances of TBEV vectors in these environments [49]. We note, however, that the altitudinal threshold for the presence of *Ixodes* vectors varies according to latitude, explaining why elevation was a lower-ranked predictor of TBEV occurrence in Central Europe (7th) and the Baltics (11th) [50]. For example, increasing temperatures over the past few decades have allowed the altitudinal expansion of *I. ricinus* to elevations between 1000–1650 m asl and multiple coordinates of TBEV-infected *I. ricinus* were detected in Czech Republic at altitudes of > 900 m asl and included in this study [51, 52].

Logically, areas with increased abundance of highly competent rodent reservoirs of TBEV are highly suitable habitat for local occurrence of TBEV. Aside from their moderate associations with habitat suitability and TBEV occurrence in the Nordics, the variables for rodent reservoir abundance were the weakest categorical variable predictors across regions and models. Interestingly, the models identified different species as the top ranked animal host predictors of TBEV occurrence in each region. It is likely these animal host species are variably abundant (or absent) across the continent which influences host selection for blood feeding by vector ticks. Nevertheless, large mammals like red deer, roe deer, and fallow deer are ideal blood sources for adult ticks and amplify local TBEV circulation and enhance TBEV infection prevalence. In the Boreal and Alpine biogeographies in the Nordics, land use patterns, habitat types (forests), and variables for the abundance of deer and hare populations strongly predict TBE incidence and potential risk areas more so than variables of climate [18, 53, 54]. The unique environmental landscape in the Baltics supports the continued expansion and merger of the three main TBEV vector species: *I. ricinus*, *I. persulcatus*, and *D. reticulatus* [55]. The incidence of TBE in Baltic countries is the highest in the world, which cannot be accounted for by environmental and climate variables alone [46, 56]. Overall, these findings emphasize the varying interplay between local ecological factors to drive TBE foci and TBE risk and strengthen the argument of regional specificity of suitable habitat ecotypes capable of supporting foci.

Previous studies have utilized ML approaches to measure spatial and temporal associations between the environment and epidemiological factors of TBE, with varying results. Our results contrast with a continental-wide ML model study that found only minor variation among environmental factor associations with self-reported point-of-infection of TBE across Europe [7]. The differences between studies largely stems from different approaches for model development such as which explanatory variables are included for interpretation (climate, habitat, anthropogenic, animal reservoirs, etc.), quality and congruence across data inputs (temporal interval, spatial unit, etc.), the geographic scope of the target territory (country, region, or continent), and the selected outcome or target variable for model prediction (habitat suitability; TBEV occurrence; TBE incidence, etc.). To generate the most reliable insights into which variables predict habitat suitability and TBEV occurrence, our model included sufficient explanatory variable data, comprehensive of the most relevant environmental factors for TBE foci. These variables included features of climate such as temperature, precipitation, and humidity, habitat and land cover types, elevation, abundance of large mammalian hosts (deer species) important for tick reproduction, and abundance of rodent reservoirs of TBEV required for foci establishment. Importantly, models lacking the multitude of environmental variables that account for the multifactorial influence on TBE foci and incidence cannot properly assess for variable collinearity nor reliably decouple individual variable associations with the epidemiological predictors of TBE. Ultimately, these models can lead to over- or under-estimations of redundant outputs or data misinterpretation. For example, models in Germany, the Baltics, and continental Europe that have attempted to predict spatial or temporal associations between the climate and the occurrence of TBEV vectors and TBE incidence report conflicting results, which complicates our understanding of the climatic factors that influence TBE incidence [14, 19, 46, 57, 58]. Similarly, the development of a broad, continental-wide model would likely have reduced our ability to accurately predict habitat suitability and TBEV occurrence at high resolutions due to the greater variance in environmental variables across vast regions for model training. Models that predict the spatial distribution of *I. ricinus* in Germany, for example, would not be optimal to predict the distribution of *I. persulcatus* in Finland due to the latter species' preference for and resistance to colder, drier temperatures [59]. For these reasons and to better account for the regional variability among environmental factors in our study, we developed separate ML models for each geographic region and used tick and animal-related endpoints of TBEV occurrence instead of human-related endpoints (incidence, cases, etc.).

Ecological niche models utilizing Maxent and XGBoost algorithms are increasingly being used to model hazard risk metrics like the occurrence of TBEV-infected ticks and anti-TBEV seropositive animals offers several advantages compared to endpoints related to human surveillance data (e.g., reported cases or incidence). Models that rely on human surveillance for TBE likely introduce bias due to the challenges in distinguishing imported versus autochthonous cases both during clinical diagnosis and reporting to surveillance databases [1]. Human surveillance data are also a lagging indicator of risk and prone to under-ascertainment due to low clinical awareness of TBE, and in certain countries, lack of mandatory notification [1, 2, 60, 61]. Measuring relationships between environmental variables and TBE incidence, therefore, may result in false-positive or false-negative associations. Modeling the occurrence of TBEV via TBEV-infected ticks and animal surveillance improves cross-country comparisons through a harmonized and consistent approach across borders and regions which increases data congruence (spatially and temporally) and overall model precision.

Providing detailed risk maps is important due to the variable perception of tick-borne disease risk, especially in travelers who may have low awareness, knowledge, and experience with ticks and thus unknowingly increase their exposure to tick populations and susceptibility to disease [62]. Leveraging ML algorithms to identify new TBE risk areas can also help support evidence-based decisions to immunize residents of endemic areas, where vaccination rates remain low despite the availability and recommendations for use of two licensed TBE vaccines (including FSME-Immun® and Encepur®) [63, 64]. Although we did not incorporate human surveillance data in our models, our results provide insight into areas where incidence for TBE may be under-reported or more likely to rise in the future. Overall, the models' predictions of increased probabilities of TBEV occurrence are consistent with publicly-available human surveillance data reporting human incidence of TBE [65] in southern Germany, Latvia, and Lithuania, but are notably inconsistent for some regions with minimal or no reported cases including Bergen and the western coastlines of Norway, Zealand and Northern Jutland in Denmark, northeastern Croatia, and the Bolzano and Trentino provinces in Italy. Our models also corroborate prior ML algorithm model studies that predicted spatiotemporal TBEV foci risks broadly across Europe [15] and locally in Poland [31, 66], Sweden [17, 67], Finland [18], northern Italy [29], and align with a recent study that leveraged ML algorithms to predict the occurrence of human cases of TBE throughout Europe [68] and observed similar spatial distribution outputs. We note that modeled outputs of hazard risk for vector-borne pathogens, however, including those in this study, are unable to capture human behaviors and movement patterns that influence individual exposure to TBEV-infected ticks and may not necessarily be predictive or indicative of tick-borne disease incidence which underscores the importance of model validation and incorporation of comprehensive explanatory variables to increase model reliability [27, 44].

Over the past thirty years, climate change has impacted the relationship between local environmental and epidemiological factors and tick-borne diseases. *Ixodes persulcatus* and other tick species have expanded across northern latitudes of Sweden and Finland [59, 69, 70], the duration and interannual variation of TBE cases is increasing [71, 72], and shifting (earlier) seasonal trends in host-seeking ticks and peak disease incidence have been observed among Lyme borreliosis cases in Norway [73] as millions of migratory birds continually disseminate new ticks across the world [74]. The emphasis on climate-based models to predict future trends of the potential burden of tick-borne diseases according to climate change scenarios is problematic since these projections are unable to comprehensively consider the influence of other biotic variables (e.g., animal hosts, reservoirs, and vectors) required for the zoonotic disease transmission. Therefore, a comprehensive understanding of the environmental drivers that result in increased human exposure and infection of TBEV is critical for effective public health communication and public preparedness. Predictive models should incorporate other explanatory covariates and passive indicators which cannot be evaluated through real-world data such as internet-based search trends [75], social media [76], risk perception studies [77], and human movement patterns [78–80]. Importantly, we emphasize increased data transparency and sharing of model inputs to support broader collaboration and development of optimized models to make greater impact(s) on public health policy [21, 81, 82].

Limitations

This study has some limitations. Surveillance bias of collected ticks and animals likely contributed to the extraction of TBEV occurrence geocoordinates and subsequent cluster selection of regional models [60]. We did not validate the continued persistence of TBEV occurrence over time among the geocoordinates included for model training as TBE foci are often fragile and can disappear (and reappear) over time unnoticed [3, 26]. We accounted for the potential short-term stability and likely disappearance for some of the locations with reported TBEV occurrence sites by limiting the inclusion of extracted coordinates to those identified in 2000 or later. The period from 2000 onward allowed for the extraction of required observations to sufficiently power the models and generate more up-to-date risk maps. Studies that collect and report the historical locations of TBE foci or with previous detection of TBEV over long-term periods (e.g., pre-2000 through 2024) are useful to describe temporal associations with TBE incidence but are less precise when modeling the data as static variables in aggregate.

However, we note the importance of studies to collect data over long temporal durations to confirm the stability of identified foci [83].

We excluded coordinates of TBEV occurrence associated with large animal hosts (Cervids) and domesticated animals (dogs, cows, sheep, etc.) from the model as these species are non-competent reservoirs for TBEV and would have decreased model precision due to their ability to cover vast ranges or their non-relevance in the enzootic cycle. Expanding the target territory to countries such as Romania, Hungary, and Turkey would also have provided additional data for model training. Although we included sufficiently broad, comprehensive epidemiologically-relevant environmental factors as explanatory variables in the models, other factors that likely contribute to TBE risk such as human behavior, socio-economics, and public health policies were not included for model training. For example, it is well documented that the incidence of TBE in the Baltics and Eastern Europe is related to socio-economics hardship, changing land use patterns and evolving public health policies due to political changes since the fall of the Soviet Union [47, 48, 84–87]. The observed disparities in model results across European regions could also be explained by additional ecological or seasonal variables (either directly or indirectly) that were not included for modeling such as beech tree masting, tick phenology, or tick co-feeding dynamics [88–92]. Our geographic clustered regional models may not be ideal for comparisons across specific borders (e.g. between Germany and Poland) or within specific geographies (e.g., between Denmark and Sweden). Alternative approaches could model these countries together given their geographic proximity and greater spatial continuity [93]. Predictive spatial distribution outputs and identified associations between environmental factors and TBEV occurrence may also have been influenced by the "changing the modifiable area unit" problem according to the spatial area and unit selected for modeling [94]. Although powerful ML approaches like XGBoost can measure non-linear relationships and handle high levels of data variance, we provided illustrative risk maps at high resolutions (Fig. 7A and C) and broader spatial scales of local administrative units or NUTS-3 districts (Fig. 7B and D), which can be more useful for dissemination of public health information. The manual extraction of geocoordinates of TBEV occurrence using the centroid of reported administrative districts are also approximations of the potential site location of the original study which could bias model results. Precision of output targets is similarly diluted when aggregating point-specific (0.01 degree) mapping predictions at the NUTS-3 level which is most relevant for NUTS-3 districts across vast landmasses like those in northern Scandinavian countries.

Conclusions

This study demonstrates the utility of robust ML model approaches to reliably predict where TBEV may be circulating and define the habitat areas that can support viral occurrence due to local features in the environment. Our work also emphasizes the importance of region-specific ML model approaches that incorporate sufficient explanatory variables to accurately inform which ecological factors are most critical for TBEV enzootic risks. Importantly, our models identified different environmental variables as the most important predictors for the occurrence of TBEV in three separate European regions, indicating how region-specific ecosystems support the establishment and stability of TBEV foci. Ultimately, the results from our ML models can drive increases in the public's awareness of potential TBE risk and help national governments and public health agencies prioritize surveillance efforts and resources to areas with potential under-reporting of TBE cases.

Abbreviations

TBE	Tick-borne encephalitis
TBEV	Tick-borne encephalitis virus
Maxent	Maximum entropy
RLR	Regularized logistic regression
XGBoost	Extreme gradient boosting
NUTS-3	Nomenclature of Territorial Units for Statistics-level 3
ML	Machine-learning
asl	Above sea level

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12942-025-00388-9.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements

We would like to thank Dr. Agustín Estrada-Peña for his help with data interpretation and providing the habitat suitability models for the predicted relative abundance of host reservoir species as predictor variables in the models.

Author contributions

PHK was responsible for the study concept, design, literature review, data extraction, interpretation, data visualization, and writing the original draft manuscript. RK and HMM identified, collated, and harmonized the explanatory variable datasets, developed and built the models, conducted the formal analyses, and contributed to data visualization and writing the original draft manuscript. JD contributed to the literature review, data extraction, data visualization, and manuscript revision. JHS provided project supervision, funding and manuscript revisions. GD oversaw project development and study design, analytical procedure, provided data, and manuscript revisions. HM provided the initial funding, methodological procedure, and manuscript revisions. JCM supervised the entire project, contributed funding and data visualization, and writing the original draft manuscript.

Funding

This study was partly supported and jointly funded by Valneva and Pfizer as part of their co-development of a Lyme disease vaccine.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and available for download in the supplementary information or in the GitHub repository, https://github.com/hmarickTiber/TBEV_Europe_Modeling.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

PHK, JHS, HM, and JCM are employees of Pfizer and may hold shares and/or stock options in the company. RK and HMM are employees of Tiber Solutions, which received funding from Pfizer in connection with the development of this manuscript. JD is an employee of Clarivate Analytics, which received funding from Pfizer in connection with the development of this manuscript. GD was a paid contractor to Pfizer in connection with the development of this manuscript. Pfizer owns a tick-borne encephalitis vaccine (FSME-IMMUN© / Ticovac©). Part of this work was supported and jointly funded by Valneva and Pfizer as part of their co-development of a Lyme Disease vaccine.

Author details

[1]Vaccines and Antivirals Medical Affairs, Pfizer US Commercial Division, Pfizer, Inc., 66 Hudson Yards Blvd E, New York City, NY, USA. [2]Tiber Solutions, Fairfax, VA, USA. [3]Clarivate Analytics, Phoenix, AZ, USA. [4]Vaccines and Antivirals Medical Affairs, Pfizer US Commercial Division, Cambridge, MA, USA. [5]Vaccines and Antivirals Medical Affairs, Pfizer US Commercial Division, London, UK. [6]Bundeswehr Institute for Microbiology, National TBEV Consultant Laboratory, 80937 Munchen, Germany. [7]Vaccines and Antivirals Medical Affairs, Pfizer US Commercial Division, Paris, France.

Received: 15 September 2024 Accepted: 20 February 2025

References

1. Van Heuverswyn J, Hallmaier-Wacker LK, Beaute J, Gomes Dias J, Haussig JM, Busch K, Kerlik J, Markowicz M, Makela H, Nygren TM, Orlikova H, Socan M, Zbrzezniak J, Zygutiene M, Gossner CM. Spatiotemporal spread of tick-borne encephalitis in the EU/EEA, 2012 to 2020. Euro Surveill. 2023. https://doi.org/10.2807/1560-7917.ES.2023.28.11.2200543.
2. Dobler G, Gniel D, Petermann R, Pfeffer M. Epidemiology and distribution of tick-borne encephalitis. Wien Med Wochenschr. 2012;162(11–12):230–8. https://doi.org/10.1007/s10354-012-0100-5.
3. Krol N, Chitimia-Dobler L, Dobler G, Kiewra D, Czulowska A, Obiegala A, Zajkowska J, Juretzek T, Pfeffer M. Identification of new microfoci and genetic characterization of tick-borne encephalitis virus isolates from Eastern Germany and Western Poland. Viruses. 2024. https://doi.org/10.3390/v16040637.
4. Borde JP, Kaier K, Hehn P, Matzarakis A, Frey S, Bestehorn M, Dobler G, Chitimia-Dobler L. The complex interplay of climate, TBEV vector dynamics and TBEV infection rates in ticks-Monitoring a natural TBEV focus in Germany, 2009–2018. PLoS ONE. 2021;16(1): e0244668. https://doi.org/10.1371/journal.pone.0244668.
5. Dobler G, Hufert F, Pfeffer M, Essbauer S. Tick-borne encephalitis: from microfocus to human disease. Progress Parasitol. 2011;2:323–31. https://doi.org/10.1007/978-3-642-21396-0_17.
6. Borde JP, Glaser R, Braun K, Riach N, Hologa R, Kaier K, Chitimia-Dobler L, Dobler G. Decoding the geography of natural TBEV microfoci in Germany: a geostatistical approach based on land-use patterns and climatological conditions. Int J Environ Res Public Health. 2022. https://doi.org/10.3390/ijerph191811830.
7. Friedsam AM, Brady OJ, Pilic A, Dobler G, Hellenbrand W, Nygren TM. Geo-spatial characteristics of 567 places of tick-borne encephalitis infection in Southern Germany, 2018–2020. Microorganisms. 2022. https://doi.org/10.3390/microorganisms10030643.
8. Honig V, Svec P, Halas P, Vavruskova Z, Tykalova H, Kilian P, Vetiskova V, Dornakova V, Sterbova J, Simonova Z, Erhart J, Sterba J, Golovchenko M, Rudenko N, Grubhoffer L. Ticks and tick-borne pathogens in South Bohemia (Czech Republic)–spatial variability in Ixodes ricinus abundance, Borrelia burgdorferi and tick-borne encephalitis virus prevalence. Ticks Tick Borne Dis. 2015;6(5):559–67. https://doi.org/10.1016/j.ttbdis.2015.04.010.
9. Stefanoff P, Pfeffer M, Hellenbrand W, Rogalska J, Rühe F, Makówka A, Michalik J, Wodecka B, Rymaszewska A, Kiewra D, Baumann-Popczyk A, Dobler G. Virus detection in questing ticks is not a sensitive indicator for risk assessment of tick-borne encephalitis in humans. Zoonoses Public Health. 2013;60(3):215–26. https://doi.org/10.1111/j.1863-2378.2012.01517.x.
10. Simkute E, Pautienius A, Grigas J, Sidorenko M, Radzijevskaja J, Paulauskas A, Stankevicius A. The prevalence of tick-borne encephalitis virus in wild rodents captured in tick-borne encephalitis foci in highly endemic Lithuania. Viruses. 2024. https://doi.org/10.3390/v16030444.
11. Pascoe EL, de Vries A, Esser HJ, Koenraadt CJM, Sprong H. Detection of tick-borne encephalitis virus in ear tissue and dried blood spots from naturally infected wild rodents. Parasit Vectors. 2023;16(1):103. https://doi.org/10.1186/s13071-023-05717-0.
12. Imhoff M, Hagedorn P, Schulze Y, Hellenbrand W, Pfeffer M, Niedrig M. Review: sentinels of tick-borne encephalitis risk. Ticks Tick Borne Dis. 2015;6(5):592–600. https://doi.org/10.1016/j.ttbdis.2015.05.001.
13. Brugger K, Boehnke D, Petney T, Dobler G, Pfeffer M, Silaghi C, Schaub GA, Pinior B, Dautel H, Kahl O, Pfister K, Suss J, Rubel F. A density map of the tick-borne encephalitis and Lyme Borreliosis Vector Ixodes ricinus (Acari: Ixodidae) for Germany. J Med Entomol. 2016;53(6):1292–302. https://doi.org/10.1093/jme/tjw116.
14. Cunze S, Glock G, Klimpel S. Spatial and temporal distribution patterns of tick-borne diseases (Tick-borne Encephalitis and Lyme Borreliosis) in Germany. PeerJ. 2021;9: e12422. https://doi.org/10.7717/peerj.12422.
15. Dagostin F, Tagliapietra V, Marini G, Cataldo C, Bellenghi M, Pizzarelli S, Cammarano RR, Wint W, Alexander NS, Neteler M, Haas J, Dub T, Busani L, Rizzoli A. Ecological and environmental factors affecting the risk of tick-borne encephalitis in Europe, 2017 to 2021. Euro Surveill. 2023. https://doi.org/10.2807/1560-7917.ES.2023.28.42.2300121.
16. Daniel M, Maly M, Danielova V, Kriz B, Nuttall P. Abiotic predictors and annual seasonal dynamics of Ixodes ricinus, the major disease vector of Central Europe. Parasit Vectors. 2015;8:478. https://doi.org/10.1186/s13071-015-1092-y.
17. Kjaer LJ, Johansson M, Lindgren PE, Asghar N, Wilhelmsson P, Fredlund H, Christensson M, Wallenhammar A, Bodker R, Rasmussen G, Kjellander P. Potential drivers of human tick-borne encephalitis in the Orebro region of Sweden, 2010–2021. Sci Rep. 2023;13(1):7685. https://doi.org/10.1038/s41598-023-34675-x.
18. Uusitalo R, Siljander M, Dub T, Sane J, Sormunen JJ, Pellikka P, Vapalahti O. Modelling habitat suitability for occurrence of human tick-borne encephalitis (TBE) cases in Finland. Ticks Tick Borne Dis. 2020;11(5):101457. https://doi.org/10.1016/j.ttbdis.2020.101457.
19. Voyiatzaki C, Papailia SI, Venetikou MS, Pouris J, Tsoumani ME, Papageorgiou EG. Climate changes exacerbate the spread of Ixodes ricinus and the occurrence of Lyme Borreliosis and tick-borne encephalitis in Europe-how climate models are used as a risk assessment approach for tick-borne diseases. Int J Environ Res Public Health. 2022. https://doi.org/10.3390/ijerph19116516.
20. Walter M, Vogelgesang JR, Rubel F, Brugger K. Tick-borne encephalitis virus and its european distribution in ticks and endothermic mammals. Microorganisms. 2020. https://doi.org/10.3390/microorganisms8071065.
21. Maxwell SP, Brooks C, Kim D, McNeely CL, Cho S, Thomas KC. Improving surveillance of human tick-borne disease risks: spatial analysis using multimodal databases. JMIR Public Health Surveill. 2023;9: e43790. https://doi.org/10.2196/43790.
22. Poisson A, Boulinier T, Bournez L, Gonzalez G, Migne CV, Moutailler S, Faivre B, Metras R. Tick-borne zoonotic flaviviruses and Borrelia infections in wildlife hosts: What have field studies contributed? One Health. 2024;18: 100747. https://doi.org/10.1016/j.onehlt.2024.100747.
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hrobjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P,

Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71. https://doi.org/10.1136/bmj.n71.

24. Bakker JW, Esser HJ, Sprong H, Godeke GJ, Hoornweg TE, de Boer WF, Pijlman GP, Koenraadt CJM. Differential susceptibility of geographically distinct Ixodes ricinus populations to tick-borne encephalitis virus and louping ill virus. Emerg Microbes Infect. 2024;13(1):2321992. https://doi.org/10.1080/22221751.2024.2321992.
25. Moraga-Fernandez A, Munoz-Hernandez C, Sanchez-Sanchez M, Fernandez de Mera IG, de la Fuente J. Exploring the diversity of tick-borne pathogens: the case of bacteria (Anaplasma, Rickettsia, Coxiella and Borrelia) protozoa (Babesia and Theileria) and viruses (Orthonairovirus, tick-borne encephalitis virus and louping ill virus) in the European continent. Vet Microbiol. 2023;286: 109892. https://doi.org/10.1016/j.vetmic.2023.109892.
26. Bournez L, Umhang G, Moinet M, Boucher JM, Demerson JM, Caillot C, Legras L, Devillers E, Hansmann Y, Velay A, Richomme C, Moutailler S, Boue F. Disappearance of TBEV circulation among rodents in a natural focus in Alsace Eastern France. Pathogens. 2020. https://doi.org/10.3390/pathogens9110930.
27. Lang D, Chitimia-Dobler L, Bestehorn-Willmann M, Lindau A, Drehmann M, Stroppel G, Hengge H, Mackenstedt U, Kaier K, Dobler G, Borde J. The emergence and dynamics of tick-borne encephalitis virus in a new endemic region in Southern Germany. Microorganisms. 2022. https://doi.org/10.3390/microorganisms10112125.
28. Michelitsch A, Wernike K, Klaus C, Dobler G, Beer M. Exploring the reservoir hosts of tick-borne encephalitis virus. Viruses. 2019. https://doi.org/10.3390/v11070669.
29. Dagostin F, Tagliapietra V, Marini G, Ferrari G, Cervellini M, Wint W, Alexander NS, Zuccali MG, Molinaro S, Fiorito N, Dub T, Rocchini D, Rizzoli A. High habitat richness reduces the risk of tick-borne encephalitis in Europe: a multi-scale study. One Health. 2024;18: 100669. https://doi.org/10.1016/j.onehlt.2023.100669.
30. Kjaer LJ, Klitgaard K, Soleng A, Edgar KS, Lindstedt HEH, Paulsen KM, Andreassen AK, Korslund L, Kjelland V, Slettan A, Stuen S, Kjellander P, Christensson M, Teravainen M, Baum A, Jensen LM, Bodker R. Spatial patterns of pathogen prevalence in questing Ixodes ricinus nymphs in southern Scandinavia, 2016. Sci Rep. 2020;10(1):19376. https://doi.org/10.1038/s41598-020-76334-5.
31. Stefanoff P, Rubikowska B, Bratkowski J, Ustrnul Z, Vanwambeke SO, Rosinska M. A predictive model has identified tick-borne encephalitis high-risk areas in regions where no cases were reported previously, Poland, 1999–2012. Int J Environ Res Public Health. 2018. https://doi.org/10.3390/ijerph15040677.
32. Uusitalo R, Siljander M, Linden A, Sormunen JJ, Aalto J, Hendrickx G, Kallio E, Vajda A, Gregow H, Henttonen H, Marsboom C, Korhonen EM, Sironen T, Pellikka P, Vapalahti O. Predicting habitat suitability for Ixodes ricinus and Ixodes persulcatus ticks in Finland. Parasit Vectors. 2022;15(1):310. https://doi.org/10.1186/s13071-022-05410-8.
33. Copernicus E-OBS Climate Dataset. https://surfobs.climate.copernicus.eu/surfobs.php. Accessed 12 Oct 2023.
34. European Environmental Agency (EEA). EUropean Nature Information System (EUNIS). Ecosystem types of Europe 2012 - Level 2 Classification (v3.1). https://www.eea.europa.eu/. Accessed 31 Mar 2024.
35. EuroDEM dataset (digital elevation model). https://www.mapsforeurope.org/datasets/euro-dem. Accessed 12 Oct 2023.
36. Wint W, Morley D, Medlock J, Alexander N. A first attempt at modelling red deer (Cervus elaphus) distributions over Europe. 2014. Open Health Data. https://doi.org/10.6084/m9.figshare.1008334.v1.
37. Alexander N, Morley D, Medlock J, Searle K, Wint W. A first attempt at modelling roe deer (Capreolus capreolus) distributions over Europe. 2014. Open Health Data. https://doi.org/10.6084/m9.figshare.1008335.v1.
38. The Global Biodiversity Information Facility (GBIF). https://www.gbif.org/. Accessed 5 Dec 2023.
39. Warren DL, Seifert SN. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol Appl. 2011;21(2):335–42. https://doi.org/10.1890/10-1171.1.
40. Arji G, Ahmadi H, Nilashi M, Rashid TA, Hassan Ahmed O, Aljojo N, Zainol A. Fuzzy logic approach for infectious disease diagnosis: a methodical evaluation, literature and classification. Biocybern Biomed Eng. 2019;39(4):937–55. https://doi.org/10.1016/j.bbe.2019.09.004.
41. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol. 2012;3(2):327–38. https://doi.org/10.1111/j.2041-210X.2011.00172.x.
42. Eurostat. https://ec.europa.eu/eurostat/en/. Accessed 02 Jul 2024.
43. Office for National Statistics (ONS). Population estimates for the UK, England, Wales, Scotland, and Northern Ireland: mid-2022. https://www.ons.gov.uk/. Accessed 02 Jul 2024.
44. Daniel M, Danielova V, Fialova A, Maly M, Kriz B, Nuttall PA. Increased relative risk of tick-borne encephalitis in warmer weather. Front Cell Infect Microbiol. 2018;8:90. https://doi.org/10.3389/fcimb.2018.00090.
45. Daniel M, Danielova V, Kriz B, Ruzek D, Fialova A, Maly M, Materna J, Pejcoch M, Erhart J. The occurrence of Ixodes ricinus ticks and important tick-borne pathogens in areas with high tick-borne encephalitis prevalence in different altitudinal levels of the Czech Republic. Part I Ixodes ricinus ticks and tick-borne encephalitis virus. Epidemiol Mikrobiol Imunol. 2016;65(2):118–28.
46. Sumilo D, Asokliene L, Bormane A, Vasilenko V, Golovljova I, Randolph SE. Climate change cannot explain the upsurge of tick-borne encephalitis in the Baltics. PLoS ONE. 2007;2(6): e500. https://doi.org/10.1371/journal.pone.0000500.
47. Sumilo D, Bormane A, Asokliene L, Vasilenko V, Golovljova I, Avsic-Zupanc T, Hubalek Z, Randolph SE. Socio-economic factors in the differential upsurge of tick-borne encephalitis in Central and Eastern Europe. Rev Med Virol. 2008;18(2):81–95. https://doi.org/10.1002/rmv.566.
48. Vanwambeke SO, Sumilo D, Bormane A, Lambin EF, Randolph SE. Landscape predictors of tick-borne encephalitis in Latvia: land cover, land use, and land ownership. Vector Borne Zoonotic Dis. 2010;10(5):497–506. https://doi.org/10.1089/vbz.2009.0116.
49. Jore S, Viljugrein H, Hofshagen M, Brun-Hansen H, Kristoffersen AB, Nygard K, Brun E, Ottesen P, Saevik BK, Ytrehus B. Multi-source analysis reveals latitudinal and altitudinal shifts in range of Ixodes ricinus at its northern distribution limit. Parasit Vectors. 2011;4:84. https://doi.org/10.1186/1756-3305-4-84.
50. Kahl O, Gray JS. The biology of Ixodes ricinus with emphasis on its ecology. Ticks Tick Borne Dis. 2023;14(2): 102114. https://doi.org/10.1016/j.ttbdis.2022.102114.
51. Martello E, Mannelli A, Ragagli C, Ambrogi C, Selmi M, Ceballos LA, Tomassone L. Range expansion of Ixodes ricinus to higher altitude, and co-infestation of small rodents with Dermacentor marginatus in the Northern Apennines. Italy Ticks Tick Borne Dis. 2014;5(6):970–4. https://doi.org/10.1016/j.ttbdis.2014.07.021.
52. Danielova V, Daniel M, Schwarzova L, Materna J, Rudenko N, Golovchenko M, Holubova J, Grubhoffer L, Kilian P. Integration of a tick-borne encephalitis virus and Borrelia burgdorferi sensu lato into mountain ecosystems, following a shift in the altitudinal limit of distribution of their vector, Ixodes ricinus (Krkonose mountains, Czech Republic). Vector Borne Zoonotic Dis. 2010;10(3):223–30. https://doi.org/10.1089/vbz.2009.0020.
53. Dub T, Ollgren J, Huusko S, Uusitalo R, Siljander M, Vapalahti O, Sane J. Game Animal Density, Climate, and Tick-Borne Encephalitis in Finland, 2007–2017. Emerg Infect Dis. 2020;26(12):2899–906. https://doi.org/10.3201/eid2612.191282.
54. Jaenson TGT, Petersson EH, Jaenson DGE, Kindberg J, Pettersson JH, Hjertqvist M, Medlock JM, Bengtsson H. The importance of wildlife in the ecology and epidemiology of the TBE virus in Sweden: incidence of human TBE correlates with abundance of deer and hares. Parasit Vectors. 2018;11(1):477. https://doi.org/10.1186/s13071-018-3057-4.
55. Capligina V, Seleznova M, Akopjana S, Freimane L, Lazovska M, Krumins R, Kivrane A, Namina A, Aleinikova D, Kimsis J, Kazarina A, Igumnova V, Bormane A, Ranka R. Large-scale countrywide screening for tick-borne pathogens in field-collected ticks in Latvia during 2017–2019. Parasit Vectors. 2020;13(1):351. https://doi.org/10.1186/s13071-020-04219-7.
56. Sidorenko M, Radzijevskaja J, Mickevicius S, Bratcikoviene N, Paulauskas A. Prevalence of tick-borne encephalitis virus in questing Dermacentor reticulatus and Ixodes ricinus ticks in Lithuania. Ticks Tick Borne Dis. 2021;12(1): 101594. https://doi.org/10.1016/j.ttbdis.2020.101594.
57. Beermann S, Dobler G, Faber M, Frank C, Habedank B, Hagedorn P, Kampen H, Kuhn C, Nygren T, Schmidt-Chanasit J, Schmolz E, Stark K, Ulrich RG, Weiss S, Wilking H. Impact of climate change on vector- and rodent-borne infectious diseases. J Health Monit. 2023;8(Suppl 3):33–61. https://doi.org/10.25646/11401.

58. Noll M, Wall R, Makepeace BL, Newbury H, Adaszek L, Bodker R, Estrada-Pena A, Guillot J, da Fonseca IP, Probst J, Overgaauw P, Strube C, Zakham F, Zanet S, Rose VH. Predicting the distribution of Ixodes ricinus and Dermacentor reticulatus in Europe: a comparison of climate niche modelling approaches. Parasit Vectors. 2023;16(1):384. https://doi.org/10.1186/s13071-023-05959-y.
59. Sormunen JJ, Saaksjarvi IE, Vesterinen EJ, Klemola T. Crowdsourced tick observation data from across 60 years reveals major increases and northwards shifts in tick contact areas in Finland. Sci Rep. 2023;13(1):21274. https://doi.org/10.1038/s41598-023-48744-8.
60. Stefanoff P, Orlikova H, Prikazsky V, Bene C, Rosinska M. Cross-border surveillance differences: tick-borne encephalitis and lyme borreliosis in the Czech Republic and Poland, 1999–2008. Cent Eur J Public Health. 2014;22(1):54–9. https://doi.org/10.21101/cejph.a3937.
61. Stefanoff P, Polkowska A, Giambi C, Levy-Bruhl D, O'Flanagan D, Dematte L, Lopalco PL, Mereckiene J, Johansen K, D'Ancona F, Venice project gatekeepers cpg. Reliable surveillance of tick-borne encephalitis in European countries is necessary to improve the quality of vaccine recommendations. Vaccine. 2011;29(6):1283–8. https://doi.org/10.1016/j.vaccine.2010.11.077.
62. Marano C, Moodley M, Melander E, De Moerlooze L, Nothdurft HD. Perceptions of tick-borne encephalitis risk: a survey of travellers and travel clinics from Canada, Germany, Sweden and the UK. J Travel Med. 2019;26(Suppl 1):S10–6. https://doi.org/10.1093/jtm/tay063.
63. Angulo FJ, Zhang P, Halsby K, Kelly P, Pilz A, Madhava H, Moisi JC, Jodar L. A systematic literature review of the effectiveness of tick-borne encephalitis vaccines in Europe. Vaccine. 2023;41(47):6914–21. https://doi.org/10.1016/j.vaccine.2023.10.014.
64. Ghiani M, Hagemann C, Friedrich J, Maywald U, Wilke T, von Eiff C, Malerczyk C. Can risk area designation help increase vaccination coverage for Tick-Borne Encephalitis? Evidence from German claims data. Vaccine. 2022;40(50):7335–42. https://doi.org/10.1016/j.vaccine.2022.10.075.
65. Angulo FJ, Halsby K, Davidson A, Ravikumar S, Pilz A, Stark JH, Moisi JC. Publicly available surveillance data on tick-borne encephalitis in Europe, 2023. Ticks Tick Borne Dis. 2024;15(6): 102388. https://doi.org/10.1016/j.ttbdis.2024.102388.
66. Stefanoff P, Zielicka-Hardy A, Hlebowicz M, Konior R, Lipowski D, Szenborn L, Siennicka J, Orlikova H. New endemic foci of tick-borne encephalitis (TBE) identified in districts where testing for TBE was not available before 2009 in Poland. Parasit Vectors. 2013;6:180. https://doi.org/10.1186/1756-3305-6-180.
67. Jaenson TG, Hjertqvist M, Bergstrom T, Lundkvist A. Why is tick-borne encephalitis increasing? A review of the key factors causing the increasing incidence of human TBE in Sweden. Parasit Vectors. 2012;5:184. https://doi.org/10.1186/1756-3305-5-184.
68. Dagostin F, Erazo D, Marini G, DaRe D, Tagliapietra V, Avdicova M, Avšič-Županc T, Dub T, Fiorito N, Knap N, Gossner CM, Kerlik J, Mäkelä H, Markowicz M, Olyazadeh R, Richter L, Wint W, Zuccali MG, Žygutienė M, Dellicour S, Rizzoli A. Modeling spatio-temporal annual changes in the probability of human tick-borne encephalitis (TBE) occurrence across Europe. BioRxiv (Pre-print). 2024. https://doi.org/10.1101/2024.10.18.619031.
69. Kjaer LJ, Soleng A, Edgar KS, Lindstedt HEH, Paulsen KM, Andreassen AK, Korslund L, Kjelland V, Slettan A, Stuen S, Kjellander P, Christensson M, Teravainen M, Baum A, Isbrand A, Jensen LM, Klitgaard K, Bodker R. A large-scale screening for the taiga tick, Ixodes persulcatus, and the meadow tick, Dermacentor reticulatus, in southern Scandinavia, 2016. Parasit Vectors. 2019;12(1):338. https://doi.org/10.1186/s13071-019-3596-3.
70. Sormunen JJ, Andersson T, Aspi J, Back J, Cederberg T, Haavisto N, Halonen H, Hanninen J, Inkinen J, Kulha N, Laaksonen M, Loehr J, Makela S, Makinen K, Norkko J, Paavola R, Pajala P, Petaja T, Puisto A, Sippola E, Snickars M, Sundell J, Tanski N, Uotila A, Vesilahti EM, Vesterinen EJ, Vuorenmaa S, Ylonen H, Ylonen J, Klemola T. Monitoring of ticks and tick-borne pathogens through a nationwide research station network in Finland. Ticks Tick Borne Dis. 2020;11(5): 101449. https://doi.org/10.1016/j.ttbdis.2020.101449.
71. Randolph SE. Tick-borne encephalitis virus, ticks and humans: short-term and long-term dynamics. Curr Opin Infect Dis. 2008;21(5):462–7. https://doi.org/10.1097/QCO.0b013e32830ce74b.
72. Rubel F, Walter M, Vogelgesang JR, Brugger K. Tick-borne encephalitis (TBE) cases are not random: explaining trend, low- and high-frequency oscillations based on the Austrian TBE time series. BMC Infect Dis. 2020;20(1):448. https://doi.org/10.1186/s12879-020-05156-7.
73. Goren A, Viljugrein H, Rivrud IM, Jore S, Bakka H, Vindenes Y, Mysterud A. The emergence and shift in seasonality of Lyme borreliosis in Northern Europe. Proc Biol Sci. 1993;2023(290):20222420. https://doi.org/10.1098/rspb.2022.2420.
74. Wilhelmsson P, Jaenson TGT, Olsen B, Waldenstrom J, Lindgren PE. Migratory birds as disseminators of ticks and the tick-borne pathogens Borrelia bacteria and tick-borne encephalitis (TBE) virus: a seasonal study at Ottenby Bird Observatory in South-eastern Sweden. Parasit Vectors. 2020;13(1):607. https://doi.org/10.1186/s13071-020-04493-5.
75. Sulyok M, Richter H, Sulyok Z, Kapitany-Foveny M, Walker MD. Predicting tick-borne encephalitis using Google Trends. Ticks Tick Borne Dis. 2020;11(1): 101306. https://doi.org/10.1016/j.ttbdis.2019.101306.
76. Boligarla S, Laison EKE, Li J, Mahadevan R, Ng A, Lin Y, Thioub MY, Huang B, Ibrahim MH, Nasri B. Leveraging machine learning approaches for predicting potential Lyme disease cases and incidence rates in the United States using Twitter. BMC Med Inform Decis Mak. 2023;23(1):217. https://doi.org/10.1186/s12911-023-02315-z.
77. Vasic A, Bjekic J, Veinovic G, Mihaljica D, Sukara R, Poluga J, Filipovic SR, Tomanovic S. Knowledge, attitudes, and practices on tick-borne encephalitis virus and tick-borne diseases within professionally tick-exposed persons, health care workers, and general population in Serbia: a questionnaire-based study. Int J Environ Res Public Health. 2022. https://doi.org/10.3390/ijerph19020867.
78. Fellin E, Varin M, Millien V. Risky business: human-related data is lacking from Lyme disease risk models. Front Public Health. 2023;11:1113024. https://doi.org/10.3389/fpubh.2023.1113024.
79. Lippi CA, Ryan SJ, White AL, Gaff HD, Carlson CJ. Trends and opportunities in tick-borne disease geography. J Med Entomol. 2021;58(6):2021–9. https://doi.org/10.1093/jme/tjab086.
80. Martin LJ, Hjertqvist M, Straten EV, Bjelkmar P. Investigating novel approaches to tick-borne encephalitis surveillance in Sweden, 2010–2017. Ticks Tick Borne Dis. 2020;11(5): 101486. https://doi.org/10.1016/j.ttbdis.2020.101486.
81. Ocias LF, Waldeck M, Hallen I, Norgaard M, Krogfelt KA. Transnational exchange of surveillance data reveals previously unrecognized TBEV microfocus. Eur J Public Health. 2019;29(4):631–3. https://doi.org/10.1093/eurpub/ckz059.
82. Urcuqui-Bustamante AM, Leahy JE, Sponarski C, Gardner AM. Collaborative modeling of the tick-borne disease social-ecological system: a conceptual framework. EcoHealth. 2023;20(4):453–67. https://doi.org/10.1007/s10393-023-01669-0.
83. Sormunen JJ, Klemola T, Hanninen J, Makela S, Vuorinen I, Penttinen R, Saaksjarvi IE, Vesterinen EJ. The importance of study duration and spatial scale in pathogen detection-evidence from a tick-infested island. Emerg Microbes Infect. 2018;7(1):189. https://doi.org/10.1038/s41426-018-0188-9.
84. Godfrey ER, Randolph SE. Economic downturn results in tick-borne disease upsurge. Parasit Vectors. 2011;4:35. https://doi.org/10.1186/1756-3305-4-35.
85. Randolph SE. Tick-borne encephalitis incidence in Central and Eastern Europe: consequences of political transition. Microbes Infect. 2008;10(3):209–16. https://doi.org/10.1016/j.micinf.2007.12.005.
86. Randolph SE, team E-Ts-p. Human activities predominate in determining changing incidence of tick-borne encephalitis in Europe. Euro Surveill. 2010;15(27):24–31. https://doi.org/10.2807/ese.15.27.19606-en.
87. Sumilo D, Bormane A, Vasilenko V, Golovljova I, Asokliene L, Zygutiene M, Randolph S. Upsurge of tick-borne encephalitis in the Baltic States at the time of political transition, independent of changes in public health practices. Clin Microbiol Infect. 2009;15(1):75–80. https://doi.org/10.1111/j.1469-0691.2008.02121.x.
88. Bregnard C, Rais O, Herrmann C, Kahl O, Brugger K, Voordouw MJ. Beech tree masting explains the inter-annual variation in the fall and spring peaks of Ixodes ricinus ticks with different time lags. Parasit Vectors. 2021;14(1):570. https://doi.org/10.1186/s13071-021-05076-8.
89. Marini G, Tagliapietra V, Cristofolini F, Cristofori A, Dagostin F, Zuccali MG, Molinaro S, Gottardini E, Rizzoli A. Correlation between airborne pollen

data and the risk of tick-borne encephalitis in northern Italy. Sci Rep. 2023;13(1):8262. https://doi.org/10.1038/s41598-023-35478-w.

90. Palo RT. Tick-borne encephalitis transmission risk: its dependence on host population dynamics and climate effects. Vector Borne Zoonotic Dis. 2014;14(5):346–52. https://doi.org/10.1089/vbz.2013.1386.
91. Randolph SE, Green RM, Peacey MF, Rogers DJ. Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. Parasitology. 2000;121(Pt 1):15–23. https://doi.org/10.1017/s0031182099006083.
92. Rosa R, Tagliapietra V, Manica M, Arnoldi D, Hauffe HC, Rossi C, Rosso F, Henttonen H, Rizzoli A. Changes in host densities and co-feeding pattern efficiently predict tick-borne encephalitis hazard in an endemic focus in northern Italy. Int J Parasitol. 2019;49(10):779–87. https://doi.org/10.1016/j.ijpara.2019.05.006.
93. Ribeiro R, Eze JI, Gilbert L, Wint GRW, Gunn G, Macrae A, Medlock JM, Auty H. Using imperfect data in predictive mapping of vectors: a regional example of Ixodes ricinus distribution. Parasit Vectors. 2019;12(1):536. https://doi.org/10.1186/s13071-019-3784-1.
94. O'Connor C, Prusinski MA, Aldstadt J, Falco RC, Oliver J, Haight J, Tober K, Sporn LA, White J, Brisson D, Backenson PB. Assessing the impact of areal unit selection and the modifiable areal unit problem on associative statistics between cases of tick-borne disease and entomological indices. J Med Entomol. 2024;61(2):331–44. https://doi.org/10.1093/jme/tjad157.

Publisher's Note