scientific reports

OPEN

Mapping global risk of bat and rodent borne disease outbreaks to anticipate emerging threats

Soushieta Jagadesh[1]✉, Claudia Cataldo[2], Wim Van Bortel[3], Esther Van Kleef[4,5,6], William Wint[7], Annapaola Rizzoli[8], Luca Busani[9] & Elena Arsevska[10]

Future epidemics and/or pandemics may likely arise from zoonotic viruses with bat- and rodent-borne pathogens being among the prime candidates. To improve preparedness and prevention strategies, we predicted the global distribution of bat- and rodent-borne viral infectious disease outbreaks using geospatial modeling. We developed species distribution models based on published outbreak occurrence data, applying machine learning and Bayesian statistical approaches to assess disease risk. Our models demonstrated high predictive accuracy (TSS = 0.87 for bat-borne, 0.90 for rodent-borne diseases), identifying precipitation and bushmeat activities as key drivers for bat-borne diseases, while deforestation, human population density, and minimum temperature influenced rodent-borne diseases. The predicted risk areas for bat-borne diseases were concentrated in Africa, whereas rodent-borne diseases were widespread across the Americas and Europe. Our findings provide geospatial tools for policymakers to prioritize surveillance and resource allocation, enhance early detection and rapid response efforts. By improving reporting and data quality, predictive models can be further refined and strengthen public health preparedness against potential future emerging infectious disease threats.

The World Health Organization (WHO) coined the term “Disease X” to describe any unknown infectious agent with the potential to cause a future epidemic or pandemic. Its hypothetical causative agent, Pathogen X, represents a yet-to-be-identified pathogen for which no therapeutic measures currently exist[1,2]. Various pathogens, including viruses, bacteria, fungi, parasites, and prions can potentially be causative agents of Disease X. Yet, viruses have become the predominant cause of recent outbreaks due to their transmission dynamics, host adaptability, and the limited availability of treatments and vaccines[3–5]. Notably RNA viruses, comprising ~ 94% of zoonotic viruses, have high mutation rates that facilitate spillover, immune evasion, and cross-species replication[6–8], though outcomes vary by ecological and evolutionary factors[9]. Hence, RNA viruses are the most common pathogens in emerging human infections, with a rate of two to three new viruses discovered annually[8].

The majority of emerging infectious diseases posing an international public health threat in the recent decades originated from wildlife reservoirs[10] as seen with COVID-19, the emergence of highly pathogenic H5N1 panzootic avian influenza[11], Severe Acute Respiratory Syndrome (SARS)[12], Middle Eastern Respiratory Syndrome (MERS)[13], Ebola virus disease (EVD)[14], and Mpox[15]. According to current evidence, bat- and rodent-borne viral diseases are among the most probable candidates for future emerging infectious disease (EID) events[16,17]. Furthermore, past outbreaks of bat- and rodent-borne viruses have demonstrated pandemic potential[17,18]. While a broader range of mammalian species can serve as intermediate hosts, such as dromedary camels and civets in the transmission of MERS[13] and SARS[12], both viruses are primarily bat-borne. Grouping taxa with similar transmission ecology helps identify broad-scale risk patterns[10,19], complementing pathogen-specific analyses by guiding targeted surveillance and research[20]. This study therefore focuses on these viral diseases of bat and rodent origin as high-priority threats for modeling and monitoring in the context of anticipating a potential Disease X.

[1]International Society of Infectious Diseases (ISID), Boston, MA 02116, USA. [2]Istituto Superiore Di Sanità, Viale Regina Elena, 299, 00161 Roma, RM, Italy. [3]Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium. [4]Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK. [5]Julius Centre for Health Sciences and Primary Care, UMCU, Utrecht University, Utrecht, The Netherlands. [6]Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium. [7]Environmental Research Group Oxford Limited (ERGO), Oxford, UK. [8]Fondazione Edmund Mach, Via Edmund Mach, 1, 38098 San Michele all’Adige, TN, Italy. [9]Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Roma, RM, Italy. [10]Centre de Coopération Internationale en Recherche Agronomique Pour le Développement (CIRAD), UMR ASTRE, Campus International Baillarguet, Montpellier, France. ✉email: soushie.jagadesh@gmail.com

| https://doi.org/10.1038/s41598-025-05588-8

The aim of this study is to identify the drivers and model the geographic distribution of the bat- and rodent-borne viral diseases outbreaks already reported in humans on a global scale and with a focus on Europe. By mapping these at-risk regions, we seek to enhance surveillance and preparedness for future outbreaks, epidemics, or pandemics, acknowledging the challenges posed by zoonotic spillover and the absence of effective therapeutic and immunization measures for many of these diseases. This work, studying the factors driving the emergence of bat- and rodent-borne viral diseases, can improve prevention and control strategies for future Disease X. We also developed an uncertainty map to simulate potential improvements in data collection in Europe. By reducing uncertainty in high-risk areas, we aim to demonstrate how enhanced reporting mechanisms and better data collaboration could lead to better predictive accuracy for future outbreaks by implementing the One Health principles in practice.

Results

Data sources and outbreak occurrences

We identified 183 articles and outbreak reports documenting occurrences of bat- (period: 1976–2023) and rodent-borne (1958–2022) viral disease outbreaks in humans. We extracted 119 and 150 distinct geographical locations for bat-borne and rodent-borne disease outbreaks in humans, derived from 94 (87 peer-reviewed publications, seven organizational websites) and 89 (78 peer-reviewed publications, 11 organizational websites) unique sources, respectively, to calibrate our models (Fig. 1).

High-risk areas for bat- and rodent-borne diseases

We developed ensemble models by incorporating predictions from three different models as meta covariates in binomial logistic regression models within a hierarchical Bayesian framework to account for spatial dependencies. The ensemble models of bat-borne and rodent-borne disease outbreaks had a high maxTSS of 0.87 and 0.90 respectively.

We identified 76,532 km^2 globally as having a high probability (0.8–1) of a bat-borne disease outbreak of zoonotic origin. The Democratic Republic of the Congo accounted for the largest high-risk area, with 19,864 km^2, representing 26% of total high-risk area. Across Africa, 59,902 km^2 was classified as high-risk, making up 78.3% of the continent's total high-risk area. Conversely, an area of 61,624 km^2 is linked to a high probability of rodent-borne diseases. Brazil in South America has the largest area at 24,010 km^2 (26.2% of its total area), followed by Europe with 15,844 km^2 (17.3%) at risk for rodent-borne diseases. Bahrain (0.82), Equatorial Guinea (0.76), and Gabon (0.56) had the highest mean probabilities for bat-borne diseases, while Montenegro (0.49), Sierra Leone (0.40), and Liberia (0.39) had the highest for rodent-borne diseases (Supplementary Table S9, 10). The high mean probability in Bahrain and Montenegro is likely driven by the model capturing the impact of MERS outbreaks in the Middle East and Hantavirus outbreaks in the Balkans.

Key environmental and socio-economical drivers of bat- and rodent-borne disease outbreaks

The covariates influencing the distribution of bat- and rodent-borne diseases are summarized in Table 1. The models identified that bat-borne disease outbreaks were mostly associated with precipitation, bushmeat activities and accessibility to the nearest city. We observed a positive correlation with bushmeat activities and minimum temperature, and a negative correlation with human population density. Mapping model-based distributions of rodent-borne disease outbreaks revealed predominant associations with deforestation, minimum temperature, and human population density. A positive correlation to rodent occurrence and negative correlation to altitude,

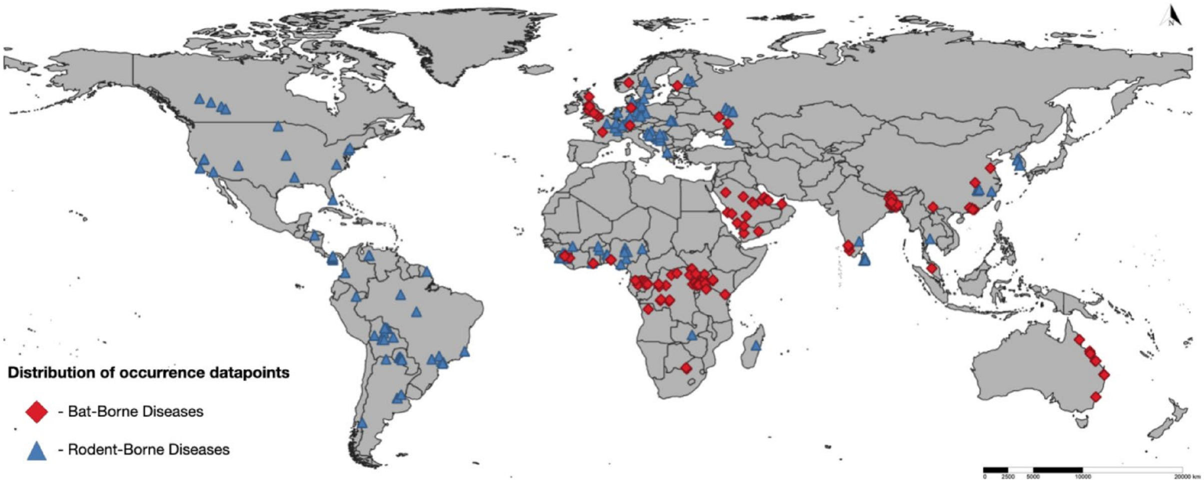

Fig. 1. Global distribution of bat-borne (blue triangles) and rodent-borne (red diamonds) viral disease outbreak emergence events resulting from zoonotic transmission, identified through a systematic literature search. Each symbol represents a distinct event in space and time, with only the first reported case per pathogen within a 10 km radius included to ensure spatial independence.

Corvariates	**Permutation importance/ correlation for bat-borne diseases**			**Permutation importance/ correlation for rodent-borne diseases**		
	RF	**BRT**	**BB**	**RF**	**BRT**	**BB**
Accessibility to the nearest city	**16.6**	**12.0**	Not significant	2.3	2.1	Not significant
Altitude	2.4	9.2	Not significant	0.0	7.3	**Negative**
Bat richness	1.0	5.2	Not significant	NA	NA	NA
Bushmeat activities	**21.4**	**13.0**	**Positive**	NA	NA	NA
Cropland distribution	NA	NA	NA	1.5	2.5	**Negative**
Forest cover loss	0.6	1.2	Not significant	**46.9**	**38.4**	Not significant
Global Human Modification of Terrestrial Systems	1.3	2.6	Not significant	2.3	2.5	Not significant
Human Population density	0.0	0.2	**Negative**	**19.9**	**25.5**	Not significant
Mammal richness index	**10.5**	9.9	Not significant	NA	NA	NA
Minimum temperature	0.7	6.1	Not significant	**27.0**	**11.4**	**Negative**
Rodent richness	NA	NA	NA	0.2	9.4	**Positive**
Total precipitation	**45.6**	**40.6**	Not significant	0.0	0.0	Not significant

Table 1. Summary of the covariates influencing the distribution of bat- and rodent-borne diseases, with permutation importance > 10 and significant correlations highlighted in bold.

cropland distribution, human population density, and minimum temperature were identified, thus factors driving the distribution of rodent-borne diseases.

The ensemble models for bat-borne (deviance = 185.6) and rodent-borne (deviance = 179.1) disease distributions outperformed the geographic null models (deviance = 765.9 and 888.8, respectively), which assume a uniform spatial distribution. Lower deviance indicates a better model fit, highlighting the importance of environmental and socio-economic predictors in explaining disease occurrence. We calculated the probability of global distribution of bat-borne (Fig. 2A) and rodent-borne diseases (Fig. 2B) from the mean value of the posterior distributions of probability per pixel of the ensemble model in each case.

Simulation methods for improved reporting

By reducing the uncertainty values in the uncertainty raster by 10%, 25%, and 50%, we simulated improved reporting scenarios. The results indicated that the ensemble model for bat-borne diseases exhibited increased maxTSS scores with reductions in uncertainty, suggesting that enhanced reporting could lead to more accurate predictions (Table 2). In contrast, the ensemble model for rodent-borne diseases showed similar maxTSS scores at 10% and 25% reductions compared to the original model. While the improvement in maxTSS from 0.87 to 0.91 might appear modest, it is important to consider that even slight increases in TSS can reflect a meaningful enhancement in predictive performance, particularly in regions where data are limited, and uncertainty is higher. This increase suggests a reduction in both false positives and false negatives, which can improve the practical reliability of the model in identifying true high-risk areas.

We visualized the before- and-after maps (Fig. 3A–F) focusing on Europe to show how reducing uncertainty improves the precision of the predicted areas at high-risk. The ensemble models with 10% and 25% reduction in uncertainty were near identical for both of bat- and -rodent borne diseases and are considered together in the visualizations (Fig. 3B,E). In Europe, the high-risk area for bat-borne diseases decreased from 198,125 km2 to 900 km2 when uncertainty was reduced by 50% in the ensemble model, demonstrating improved precision in regions with limited outbreak data (Supplementary Table S11). Similarly, for rodent-borne diseases, the high-risk area expanded from 14,600 km^2 in 7 Balkan countries to 106,321 km2 across all 11 endemic countries, suggesting that reducing uncertainty helps uncover underreported risk areas (Supplementary Table S12). Notably, potential surveillance gaps are indicated in regions where risk factors are present but reported occurrences are low, such as the Balkans (endemic for Hantavirus infections)[21], the Pyrenees, and Northern Spain. Addressing these gaps through targeted monitoring or improved reporting could enhance early detection and response efforts.

Discussion

Our comprehensive geospatial analysis of bat- and rodent-borne disease outbreaks already reported in humans advances understanding of their global distribution, key drivers, and implications for emerging infectious diseases, including potential Disease X. By integrating environmental, socio-economic, and surveillance uncertainty metrics, our study extends beyond previous efforts that primarily relied on ecological and population-based predictors. This approach refines risk estimates in underreported regions and provides a more nuanced understanding of outbreak dynamics. The observed differences in predicted hotspots compared to prior studies[22,23] underscore the influence of alternative bias corrections and additional risk factors considered in our model. By identifying high-probability areas and their driving factors, our work informs targeted public health interventions and resource allocation strategies for improved disease prevention and control.

The large proportion of at-risk territory for bat-borne diseases in Africa (78.3%) indicates a significant regional burden, which previous research showed to be driven by socio-economic and environmental factors like bushmeat practices[24], bat species richness[23], and climate conditions that favor spillover from bat reservoir hosts[25,26] Some of these factors have indeed been highlighted as significant covariates in the RF, BRT and Bayesian

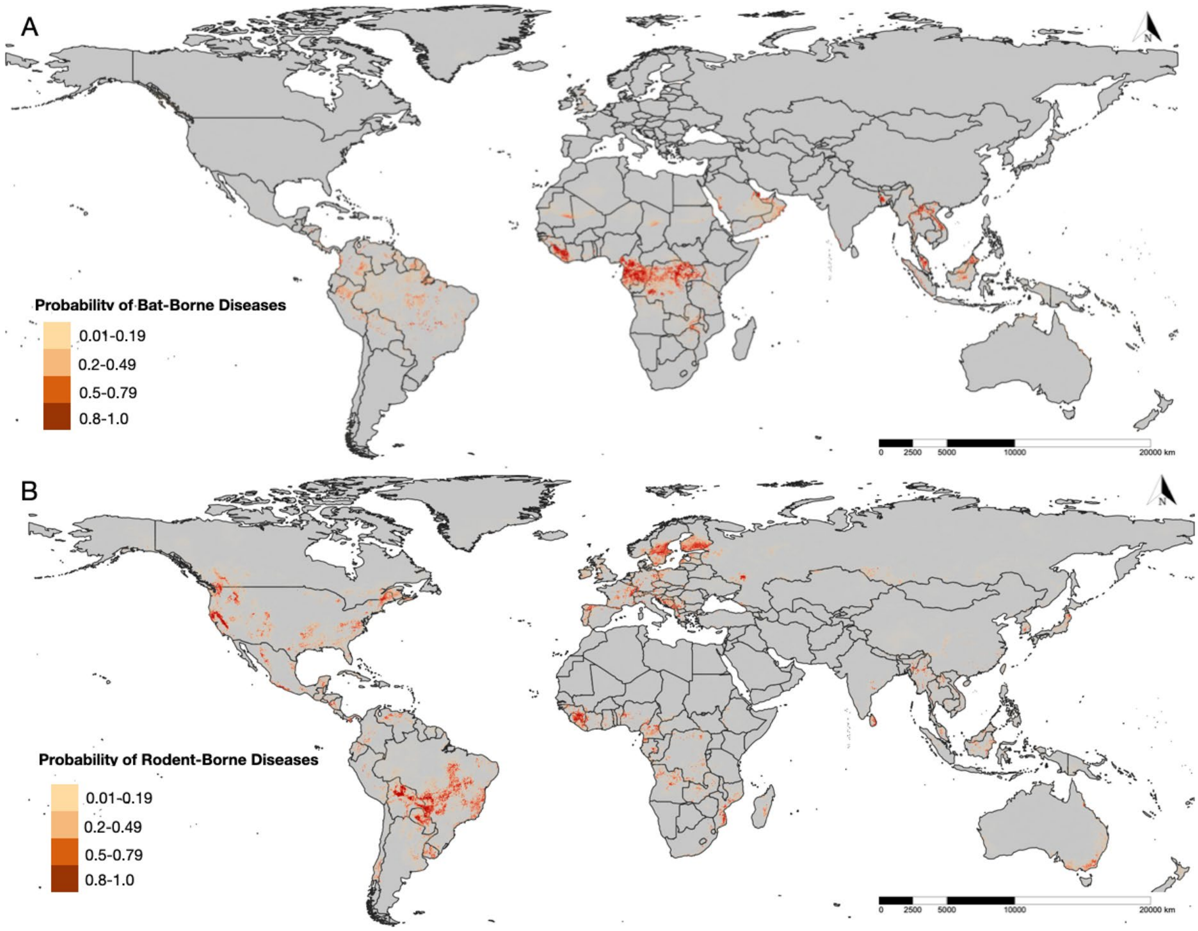

Fig. 2. (**A**) Global distribution of the probability of bat-borne diseases outbreaks from an ensemble of three model predictions using a hierarchical binomial model with spatial autocorrelation. (**B**) Global distribution of the probability of rodent-borne diseases outbreaks from an ensemble of three model predictions using a hierarchical binomial model with spatial autocorrelation. Gray areas indicate regions where probability is below 0.01. The maps were generated using QGIS (version 3.34.0-Prizren), an open-source geographic information system, *available at* https://qgis.org.

MaxTSS	Ensemble	Ensemble with 10% reduction in uncertainty	Ensemble with 25% reduction in uncertainty	Ensemble with 50% reduction in uncertainty
Bat-borne disease	0.87	0.91	0.91	0.92
Rodent-borne disease	0.90	0.90	0.90	0.91

Table 2. Comparison of the maxTSS of the ensemble model and the ensemble models with 10%, 25%, and 50% reduction in uncertainty of the probability of bat- and -rodent borne diseases.

Binomial models. Hunting, preparing, and selling of bushmeat is an established driver of bat-borne diseases as it increases direct human contact with bats and other wildlife[27,28]. For example, the 2007 Ebola outbreak in the Democratic Republic of the Congo was linked to bushmeat activities where villagers contracted the virus from infected fruit bats[29]. Furthermore, previous studies have shown that rainfall patterns, influenced by climate anomalies such as the El Niño Southern Oscillation (ENSO), correlate with the emergence of bat-borne diseases like Hendra virus in Australia and Nipah virus in Malaysia and Bangladesh[30–32].

In contrast, the high-risk areas of rodent-borne diseases detected in South America and Europe indicate a more diverse geographical spread. Deforestation and minimum temperature were key factors in the RF and BRT models. The Bayesian Binomial model's correlation with rodent occurrence and low altitude underscore ecological influences but does not contradict the dilution hypothesis. Our model includes multiple rodent-borne diseases, whereas outbreaks are typically driven by specific reservoir species. Future research focusing on pathogen-specific models could help refine these dynamics. Agricultural expansion and habitat fragmentation due to deforestation impact rodent populations, though species respond differently[33]. For instance, rats often

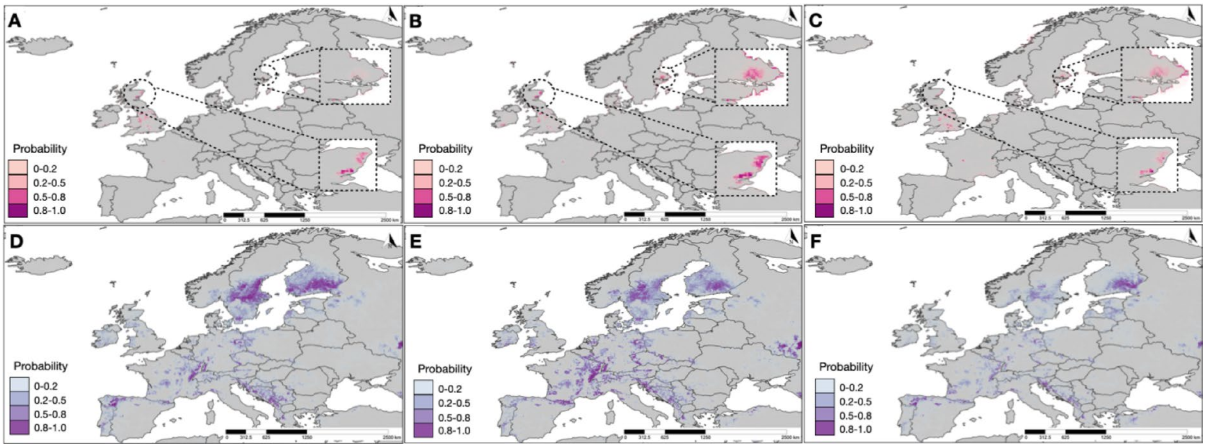

Fig. 3. (**A**) The ensemble model, (**B**) the ensemble models with 10% and 25%, (**C**) the ensemble model with 50% reduction in uncertainty of the probability of bat-borne diseases at the scale of Europe; (**D**) The ensemble model, (**E**) the ensemble models with 10% and 25%, (**F**) the ensemble model with 50% reduction in uncertainty of the probability of rodent-borne diseases at the scale of Europe. The maps were generated using QGIS (version 3.34.0-Prizren), an open-source geographic information system, *available at*https://qgis.org.

thrive in populated urban and peri-urban environments, whereas other small mammals, such as field mice, are more sensitive to rural habitat changes. Studies indicate that deforestation in West Africa and the Amazon brings rodents like *Mastomys natalensis* and *Oligoryzomys* closer to humans, heightening the risk of Lassa fever[34] and Hantavirus transmission[35–37]. Low-lying areas, another driver identified by our findings, provide favorable conditions for human settlements, agriculture, and rodent habitats, increasing the chances of rodent-to-rodent and rodent-to-human contact and subsequent viral transmission[38,39].

Our risk map for bat-borne diseases identified high-risk areas (Supplementary figure S21) in Southeast Asia (27,204 m^2) and Southern China (410 km^2), aligning with recent evidence on ecological niche of bat viruses closely related to SARS-CoV-2[40,41], highlighting areas requiring enhanced surveillance in animal hosts. These identified high-risk regions should prioritize strengthening wildlife surveillance systems. Surveillance is often underfunded or viewed as a costly system, particularly in low-density areas where wildlife health monitoring can be logistically challenging. To address these issues, we recommend developing cost-effective sustainable surveillance solutions, including educating local communities on the risks of increased exposure, wildlife trade, and encouraging practices that reduce contact with bat reservoir and intermediate host populations[42,43]. A community-based reporting approach, such as the Participatory One Health Disease Detection (PODD) system in Thailand, where local communities report unusual animal deaths or illnesses, could facilitate early outbreak detection and encourage awareness of the risks associated with wildlife exposure and trade[44]. Additionally, our risk maps provide valuable tools for human public health authorities in high-risk regions, helping to prioritize resource allocation for targeted surveillance, early detection, and disease prevention. Equitable surveillance is essential for ensuring accurate risk assessments and strengthening global preparedness for emerging infectious diseases. Furthermore, increased funding for research into high-risk pathogens and their transmission dynamics will improve understanding of risk factors and potential outbreaks, ultimately enhancing preparedness and response efforts.

Our simulation of improved reporting demonstrates how improvements in reporting can lead to more accurate disease risk predictions, which in turn would enhance public health preparedness. Even minor improvements in maxTSS could translate into more efficient public health interventions by narrowing down priority zones for monitoring and intervention. This approach not only reinforces the general importance of surveillance at high-risk areas but also provides a geographically explicit assessment of its impact. Our approach highlights areas where targeted surveillance efforts could significantly improve early detection and response strategies. The results demonstrate that increased reporting leads to a more precise prediction of the at-risk areas and identifying underreported risk areas, giving actionable insights for public health authorities in Europe. Although bat-borne zoonotic outbreaks are limited in Europe (Fig. 3A), our findings demonstrate multiple at-risk regions in the UK where a recent study[45] has demonstrated the presence of *Coronaviruses* circulating in bat populations with zoonotic potential underscoring the need for genomic surveillance of wildlife, particularly bats.

Several studies have attempted to map the geographic risk of zoonotic disease emergence, particularly for bat- and rodent-borne viruses[22,23,46]. Compared to these studies, our approach provides a geographically explicit assessment of bat- and rodent-borne disease risks based on reported outbreak occurrences rather than solely relying on host richness or viral sharing networks. Unlike the earlier studies, which primarily used ecological predictors[22,23], our models integrate both environmental and socio-economic drivers within a Bayesian ensemble framework, incorporating spatial dependencies through an intrinsic conditional autoregressive (iCAR) structure, allowing for a more comprehensive assessment of disease emergence risk. Additionally, our study identifies current high-risk areas, providing immediate relevance for surveillance and intervention planning. The differences in modeling approaches, data sources, and assumptions across these studies highlight the need for cross-validation and integration of multiple methodologies to refine zoonotic risk assessments. Unlike Allen et al. (2017)[22], which primarily used population metrics to correct for reporting bias, our study employs a novel uncertainty metric that accounts for spatial gaps, surveillance variability, and temporal inconsistencies in outbreak reporting. This approach provides a more dynamic correction for underreporting, particularly in regions with limited surveillance, leading to differences in predicted outbreak hotspots. Our findings complement these previous efforts by offering a practical framework for prioritizing high-risk areas for surveillance and public health interventions.

While this study provides valuable insights into the geographic risk of bat- and rodent-borne diseases, its limitations should be noted. We excluded avian influenza viruses as their transmission dynamics are primarily driven by farmed avian hosts, commerce and wild birds migratory patterns, differing fundamentally from rodent- and bat-borne diseases. Including AIVs would require a distinct eco- epidemiological framework beyond the scope of this study. Reliance on published reports may bias results due to underreporting in low-surveillance regions. To mitigate this, we weighted background points toward populous areas, included organizational reports, and suggest future use of bias layers to refine observation bias. While a 10 km resolution offers a global overview, using higher-resolution data (e.g., 500m–1km) in future studies may help incorporate finer-scale drivers, particularly for localized outbreak prediction and intervention planning. However, it is worth noting that even finer scales may fail to capture true spillover locations, as data often reflect where outbreaks are reported rather than where they originate and may also miss local habitat mosaics. Diagnostic capacity and health infrastructure influence disease emergence but lacked global data, so accessibility to the nearest city was used as a proxy. To improve the accuracy of surveillance proxies, future studies should consider incorporating data on specific local surveillance systems, such as community-based reporting or mobile health units, rather than relying solely on proximity to urban centers, as these systems may better reflect diagnostic capacity and outbreak detection in remote areas. Grouping all bat- and rodent-borne diseases may overlook pathogen-specific dynamics. We suggest that future work should develop pathogen-reservoir host specific models where data availability permits, to refine predictions and improve their applicability to targeted public health interventions. Increased reporting by reducing uncertainty may confirm known risk areas or introduce new uncertainties. While this method of simulation of improved reporting provides a structured way to assess potential reporting improvements, we acknowledge that real-world reporting enhancements may vary spatially and could be better captured using region-specific estimates in future research. Integrating field data, seroprevalence studies, and enhanced surveillance in high-risk areas will improve risk estimates and public health preparedness.

Nevertheless, the global maps of bat-borne and rodent-borne disease distributions geographically synthesize peer-reviewed outbreak occurrence data, providing valuable tools prioritizing surveillance and intervention. By identifying high-risk areas, resources can be strategically allocated to enhance disease prevention and control. These findings support a multi-sectoral approach, integrating animal health and ecological conservation, urban planning, and public health strategies. To strengthen surveillance, high-risk regions should implement targeted monitoring programs, drawing from successful models such as early warning systems and community-based reporting networks. Additionally, policy frameworks promoting habitat conservation and sustainable land use can mitigate human-wildlife interactions that drive spillover risks[47]. Understanding the environmental and socio-economic factors influencing disease emergence also aids in anticipating and managing emerging pathogens, facilitating proactive measures and rapid response strategies to reduce the risk of future pandemics[48]. Regular updates to risk maps, incorporating new surveillance data, are essential for maintaining the accuracy and relevance of risk assessments, ensuring that preparedness strategies remain effective and responsive to emerging threats.

Methods

We used a multistep procedure to model the geographical distribution of bat-borne and rodent-borne viral diseases (Fig. 4). We modeled the occurrence of zoonoses by using the following steps: collate outbreak datapoints from systematic literature search; prepare environmental and demographic covariates; fitting separate bat- and rodent-borne disease distribution models; conduct ensemble modeling for bat- and rodent-borne diseases respectively; and calculate the uncertainty of risk of these diseases and then simulate multiple scenarios for improved reporting.

Data collection

We searched for peer-reviewed reports for outbreaks of bat- and rodent-borne viral diseases (listed in Supplementary Table S1) by reviewing three electronic databases: Web of Science (https://www.webofscience.com), PubMed (https://pubmed.ncbi.nlm.nih.gov), and Google Scholar (https://scholar.google.com). We also searched organizational websites reporting disease outbreaks, WHO Disease Outbreak News (DONs; https://www.who.int/emergencies/disease-outbreak-news) and CDC Current Outbreak List (https://www.cdc.gov/outbreaks/index.html). The systematic literature search covered the period from January 1, 1965, to December 31, 2023 (PRISMA diagram found as Supplementary Figure S1). In this study, an outbreak event was defined as a reported occurrence of a bat- or rodent-borne viral disease affecting humans, as documented in peer-reviewed literature. Each outbreak represents a distinct event in space and time, rather than continuous transmission within an endemic setting. Our study selected the first reported emergence event (index human case location) for each pathogen at each unique geographical location (village, town, city within a 10 km radius) to ensure spatial independence and minimize bias from repeated reporting. A single confirmed human case was considered an occurrence event, particularly in cases of bat-borne diseases such as European Lyssavirus 1 and 2. For currently endemic diseases, only the earliest recorded outbreak in each location was included. Imported cases were excluded to focus on locally acquired outbreaks. This approach ensures that our dataset represents independent outbreak events while mitigating potential overrepresentation of frequently reported diseases.

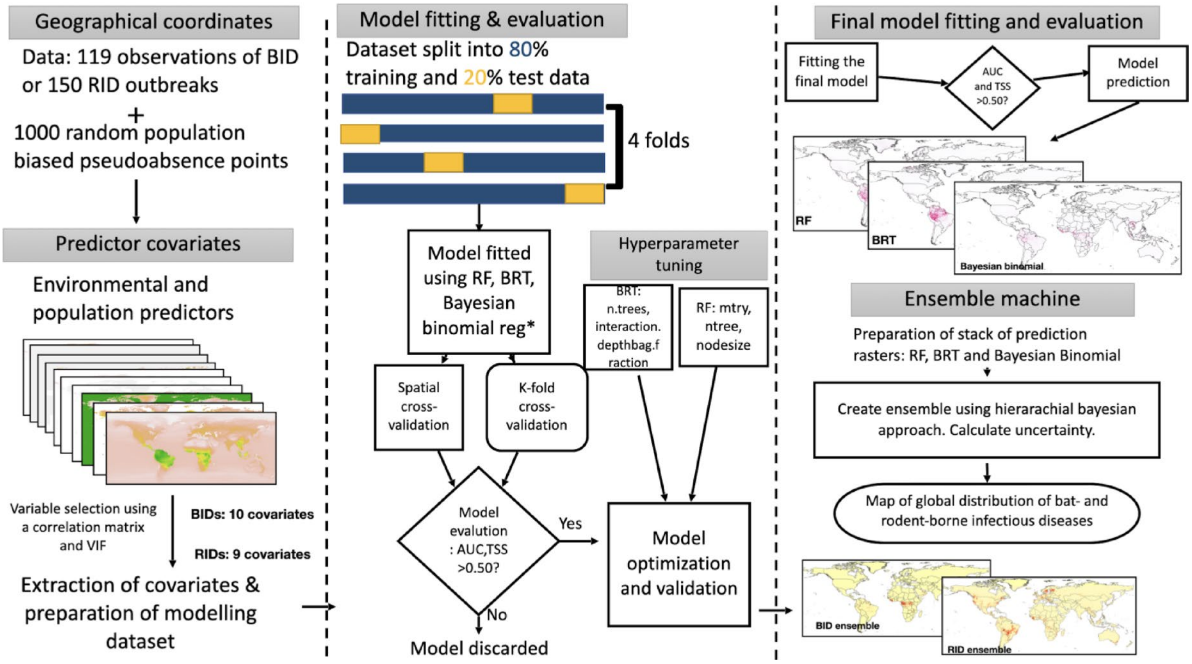

Fig. 4. Methodological workflow for mapping the global distribution of bat- (BID in the figure) and rodent-borne (RID in the figure) infectious diseases. RF- Random Forest model and BRT- Boosted Regression Tree.

Environmental and demographic covariates

We identified potential environmental and demographic covariates influencing the emergence and spread of bat- and rodent-borne diseases based on a previous publication[49](Table 3). The bat and rodent occurrence rasters represent the number of bat/rodent species per pixel based on aggregated species distribution data from two sources (IUCN and biodiversitymapping.org). We used data from World Geodetic System version 84 (GISGeography, https://gisgeography.com) to project all covariates and resampled by using a pixel resolution of 5 min of arc (0.083333 degrees), equating to ≈10 × 10 km resolution. The definitions of the covariates are detailed in the Supplementary Material S4.

Model fitting and evaluation

We selected 10 covariates for bat-borne diseases and 9 for rodent-borne diseases, ensuring each had a variance inflation factor (VIF) below 10[50], to reduce the risk of collinearity among the covariates. We used occurrence data obtained from our literature search, supplemented by 1,000 randomly generated pseudo-absence points that were intentionally weighed to reflect the density of more populated areas[51]. We use a presence-absence framework, where pseudo-absences were downweighted based on population density to partially account for observation bias, ensuring a more balanced representation of observed and background data. While population density has been used as a proxy for surveillance effort in previous studies, we acknowledge its limitations in fully capturing reporting bias, particularly in regions with well-documented outbreaks but relatively lower population densities. Future work could improve upon this approach by incorporating additional bias correction layers, such as accessibility metrics or disease-specific reporting biases, to refine pseudo-absence selection. Rather than using a presence-background approach, we adopted a binary classification framework, in which absence points were explicitly defined and incorporated into the model, allowing for clearer probabilistic inference and better control of biases[52].

We mapped the distribution by using three binomial models: random forest (RF), boosted regression tree (BRT), and Bayesian binomial model. Each of these models inherently assumes a binary response variable (presence = 1, absence = 0), distinguishing our approach from presence-background models like MaxEnt. Each model offers unique strengths: RF (R package “randomForest”) is known for its ability to handle complex interactions between variables[53], BRT (R package “gbm”) excels at capturing non-linear relationships[54], and the Bayesian binomial regression model (R package “hsdm”) accounts for uncertainty and incorporates prior information[55]. By integrating these diverse modeling approaches, we aimed to enhance the accuracy, reliability, and predictive capacity of our findings[24]. For each model, we used 80% of the datapoints (observed and background) for the training dataset; we used the remaining 20% of datapoints as the validation dataset. We fit and evaluated the base models by using area under the curve (AUC). By integrating these diverse modeling approaches, we aimed to enhance the accuracy, reliability, and predictive capacity of our findings.

We used 2 cross-validation (CV) methods and input covariates from R to prevent model overfitting: k-fold CV based on covariates from the SDMtune package[56] and environmental CV (EnvCV) with covariates from the blockCV package[57]. We split the training data into 4-folds (k = 4) for both approaches. We only chose models with an AUC and True Skill Statistic (maxTSS) > 0.5 after CV for hyperparameter tuning and development of the

Covariate	Model	Rationale	Source
Climatic covariates			
Monthly average minimum temperature	Bat and Rodent	Climate change such as increasing temperatures and precipitation changes causes shifts in the geographical range of pathogen hosts, vectors, and reservoirs	https://www.worldclim.org
Monthly average precipitation	Bat and Rodent		
Environmental covariates			
Altitude	Bat and Rodent	Altitude affects infectious disease emergence by influencing climate conditions, habitats, and pathogen viability	https://www.worldclim.org
Bat occurrence	Bat	Bats play a critical role as reservoirs	In-house raster developed from https://www.iucnredlist.org/resources/spatial-data-download and https://biodiversitymapping.org/index.php/download/
Cropland distribution	Rodent	The extension of agricultural cultivation into natural habitats increased exposure of humans to reservoirs	https://data.apps.fao.org/map/catalog/
Global Human Modification of Terrestrial Systems	Bat and Rodent	Land-use changes increase in the abundance of zoonotic host species and reduce diversity of non-hosts due to habitat loss	Global Human Modification of Terrestrial Systems. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/edbc-3z60
Mammal richness	Bat	Loss of biodiversity is found to play a major role with frequent emergence and transmission of zoonoses	Center For International Earth Science Information Network-CIESIN-Columbia University, & NatureServe. (2015). Gridded Species Distribution: Global Mammal Richness Grids, 2015 Release. SEDAC. https://doi.org/10.7927/H4N014G5
Rodent occurrence	Rodent	Rodents serve as host/reservoirs	In-house raster developed from https://www.iucnredlist.org/resources/spatial-data-download and https://biodiversitymapping.org/index.php/download/
Forest cover loss proxy for deforestation	Bat and Rodent	Deforestation leads to habitat fragmentation	Hansen, M. C., et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342 (2013): 850–53. https://glad.earthengine.app/view/global-forest-change
Demographic covariates			
Accessibility to the nearest city	Bat and Rodent	Accessibility to the closest city enables the transmission of the pathogen to a large seeming naïve urban population	Nelson, A., et al. A suite of global accessibility indicators. Sci Data 6, 266 (2019). https://doi.org/10.1038/s41597-019-0265-5
Bushmeat activities	Bat	Bushmeat has been linked with spillover due to frequent contact with infectious materials from wildlife	Jagadesh, S., et al. Mapping Global Bushmeat Activities to Improve Zoonotic Spillover Surveillance by Using Geospatial Modeling. Emerging infectious diseases vol. 29,4 (2023): 742–750. https://doi.org/10.3201/eid2904.221022
Human population density	Bat and Rodent	As population density increases, the opportunities increase for disease transmission due to closer and more frequent interactions between individuals	Center for International Earth Science Information Network—CIESIN—Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. SEDAC. https://doi.org/10.7927/H49C6VHW

Table 3. List of environmental and demographic covariates used in published models of bat- and rodent-borne diseases.

final RF and BRT models. We calculated the maxTSS for the validation of the Bayesian binomial model. We split data into training, validation, and testing sets for model optimization by tuning the appropriate hyperparameters for each RF and BRT model for bat – and rodent-borne diseases. We used the entire dataset in the optimized models to predict the global distribution of bat – and rodent-borne viral diseases, separately.

Ensemble modeling

We used the three model predictions from RF, BRT, and Bayesian binomial models as metacovariates for developing an ensemble model for of bat- and rodent-borne diseases, respectively. This approach leverages the complementary strengths of the above three models to improve predictive performance, accounting for both non-linear relationships and spatial dependencies, enhancing robustness compared to using any single model. The ensemble model was implemented using a hierarchical Bayesian framework with an intrinsic conditional autoregressive (iCAR) prior[55], which accounts for spatially structured random effects, ensuring that spatial autocorrelation is explicitly incorporated into the final predictions. Notably, while the individual binomial models for bat- and rodent-borne diseases did not use iCAR priors, the ensemble model integrates spatial dependencies through the iCAR structure applied at the ensemble level. We validated the output ensemble prediction by using maxTSS and comparing deviance with a geographic null model[58]. We generated the two 10 × 10 km resolution rasters mapping the risk of bat- and rodent-borne diseases, each from the mean probability from each pixel of the ensemble model.

To account for reporting biases more accurately, we developed an uncertainty metric that integrates multiple factors affecting disease surveillance rather than relying solely on population density as a proxy. While population-based corrections assume higher reporting accuracy in densely populated areas, this does not account for disparities in surveillance capacity across regions. Our approach incorporates spatial gaps in occurrence data, variability in surveillance efforts, and temporal inconsistencies in reporting to distinguish true disease absence from underreporting. This dynamic method reduces bias, prevents over-correction in high-population areas, and improves risk predictions, particularly in regions with limited surveillance. The uncertainty was calculated from the standard deviation (SD) of each pixel[24]. Uncertainty reflects the greater variation between model predictions suggests higher disagreement in risk estimates, which can arise from data sparsity or inconsistencies in environmental associations.

Calculation of predicted area

We reclassified the probability of bat – and rodent-borne diseases into four categories: very low probability (<0.2), low (0.2–0.5), intermediate (0.5–0.8), and high (>0.8). We then calculated the number of pixels per country in each category. We derived the area belonging to the high probability category by extracting the number of pixels >0.8 in each country and multiplying it by the area of the country. We defined high-risk areas as regions with the highest predicted probability of disease occurrence based on ensemble model outputs. These areas represent the top 20% of risk values, identified through spatial modeling of environmental and socio-economic covariates.

Simulation methods of improved reporting

Improved reporting refers to a hypothetical scenario where underreporting biases are reduced through enhanced surveillance, increased diagnostic capacity, and systematic data collection. In our simulations, this is modeled by adjusting reporting probabilities to better reflect potential disease presence in under-surveyed regions. We defined different levels of potential reporting improvements, where data collaboration, improved data quality or reporting frequency improves by 10%, 25%, and 50%. We decreased the values in the uncertainty raster by a fixed percentage (10%, 25%, and 50%) to simulate various scenarios of better reporting.

$$\text{New Probability Range} = \text{Probability} \pm (1.96 \times \text{Updated Uncertainty})$$

Using the updated uncertainty rasters, we calculated the new lower and upper confidence intervals (CI), to then calculate the new probability of bat- and rodent- borne diseases with increased reporting. We re-ran the ensemble modelling with the new probability raster as one of the meta-covariates (along with the RF, BRT and Bayesian Binomial raster). We compared the new maxTSS of the new ensemble with the original model.

The assumption that uncertainty is reduced with improved reporting is based on the rationale that better reporting leads to more precise and accurate data, which in turn reduces spatial uncertainty in disease occurrence. By using exact coordinates and not generalized or centroid locations, we hope to reduce the underlying limitations. Nevertheless, uncertainty can persist due to factors such as clustering, imprecise location data, and reporting biases.

Data availability

All data including geographical occurrence points, list of included articles/reports from which the data points were extracted, model evaluation (AUC, convergence plots) and codes to reproduce the model can be accessed through the GitHub repository: https://github.com/soushie13/Insights-for-Preventing-Disease-X/tree/main.

Received: 15 November 2024; Accepted: 3 June 2025

References

1. Prioritizing diseases for research and development in emergency contexts. https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts.
2. Simpson, S. K., Michael, C., Glozman, V. & Chakrabarti, A. Disease X: Accelerating the development of medical countermeasures for the next pandemic. *Lancet Infect. Dis.* **20**, e108–e115 (2020).
3. Morse, S. S. et al. Prediction and prevention of the next pandemic zoonosis. *Lancet Lond. Engl.* **380**, 1956–1965 (2012).
4. Kelly, T. R. et al. One health proof of concept: Bringing a transdisciplinary approach to surveillance for zoonotic viruses at the human-wild animal interface. *Prev. Vet. Med.* **137**, 112–118 (2016).
5. Institute of Medicine (US) Forum on MicrobialThreats. *Microbial Evolution and Co-Adaptation: A Tribute to the Life and Scientific Legacies of Joshua Lederberg: Workshop Summary*. (National Academies Press (US), Washington (DC), 2009).
6. Kreuder Johnson, C. et al. Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* **5**, 14830 (2015).
7. Carrasco-Hernandez, R., Jácome, R., López Vidal, Y. & Ponce de León, S. Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR J.* **58**, 343–358 (2017).
8. Rosenberg, R. Detecting the emergence of novel, zoonotic viruses pathogenic to humans. *Cell. Mol. Life Sci. CMLS* **72**, 1115–1125 (2015).
9. Ka-Wai Hui, E. Reasons for the increase in emerging and re-emerging viral infectious diseases. *Microbes Infect.* **8**, 905–916 (2006).
10. Jones, K. E. et al. Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
11. Joseph, U., Su, Y. C. F., Vijaykrishna, D. & Smith, G. J. D. The ecology and adaptive evolution of influenza A interspecies transmission. *Influenza Other Respir. Viruses* **11**, 74–84 (2017).
12. LeDuc, J. W. & Barry, M. A. SARS, the First Pandemic of the 21st Century1. *Emerg. Infect. Dis.* **10**, e26 (2004).
13. Han, H.-J., Yu, H. & Yu, X.-J. Evidence for zoonotic origins of Middle East respiratory syndrome coronavirus. *J. Gen. Virol.* **97**, 274–280 (2016).
14. Kalra, S. et al. The emergence of ebola as a global health security threat: From 'lessons learned' to coordinated multilateral containment efforts. *J. Glob. Infect. Dis.* **6**, 164–177 (2014).
15. Ren, S.-Y., Li, J. & Gao, R.-D. 2022 Monkeypox outbreak: Why is it a public health emergency of international concern? What can we do to control it?. *World J. Clin. Cases* **10**, 10873–10881 (2022).
16. Zhu, J. & Gack, M. U. Viral tolerance enabled by a bat-specific genomic tweak. *Nature* https://doi.org/10.1038/d41586-025-00081-8 (2025).
17. The Viral Most Wanted: The Arenavirus Family | CEPI. https://cepi.net//viral-most-wanted-arenavirus-family.
18. Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover and emergence. *Nat. Rev. Microbiol.* **18**, 461–471 (2020).
19. Russell, R. E., Katz, R. A., Richgels, K. L., Walsh, D. P. & Grant, E. H. A Framework for modeling emerging diseases to inform management. *Emerg. Infect. Dis.* **23**, 1 (2017).
20. Preparedness and Resilience for Emerging Threats (PRET). https://www.who.int/initiatives/preparedness-and-resilience-for-emerging-threats.
21. Avšič Županc, T., Korva, M. & Markotić, A. HFRS and hantaviruses in the Balkans/South-East Europe. *Virus Res.* **187**, 27–33 (2014).

22. Allen, T. et al. Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1124 (2017).
23. Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
24. Jagadesh, S., Zhao, C., Mulchandani, R. & Van Boeckel, T. P. Mapping global bushmeat activities to improve zoonotic spillover surveillance by using geospatial modeling. *Emerg. Infect. Dis.* **29**, 742–750 (2023).
25. Eby, P. et al. Pathogen spillover driven by rapid changes in bat ecology. *Nature* **613**, 340–344 (2023).
26. Becker, D. J., Eby, P., Madden, W., Peel, A. J. & Plowright, R. K. Ecological conditions predict the intensity of Hendra virus excretion over space and time from bat reservoir hosts. *Ecol. Lett.* **26**, 23–36 (2023).
27. Milbank, C. & Vira, B. Wildmeat consumption and zoonotic spillover: contextualising disease emergence and policy responses. *Lancet Planet. Health* **6**, e439–e448 (2022).
28. Plowright, R. K. et al. Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
29. Leroy, E. M. et al. Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Dis. Larchmt. N* **9**, 723–728 (2009).
30. Latinne, A. & Morand, S. Climate anomalies and spillover of bat-borne viral diseases in the Asia-Pacific Region and the Arabian Peninsula. *Viruses* **14**, 1100 (2022).
31. Giles, J. R. et al. Environmental drivers of spatiotemporal foraging intensity in fruit bats and implications for Hendra virus ecology. *Sci. Rep.* **8**, 9555 (2018).
32. Daszak, P. et al. Interdisciplinary approaches to understanding disease emergence: The past, present, and future drivers of Nipah virus emergence. *Proc. Natl. Acad. Sci.* **110**, 3681–3688 (2013).
33. Andreo, V. et al. Estimating hantavirus risk in southern Argentina: a GIS-based approach combining human cases and host distribution. *Viruses* **6**, 201–222 (2014).
34. Adetola, O. & Adebisi, M. Impacts of deforestation on the spread of Mastomys natalensis in Nigeria. *World Sci. News* **130**, 286–296 (2019).
35. Mills, J. N. Biodiversity loss and emerging infectious disease: An example from the rodent-borne hemorrhagic fevers. *Biodiversity* **7**, 9–17 (2006).
36. Mills, J. N. Regulation of rodent-borne viruses in the natural host: implications for human disease. In *Infectious Diseases from Nature: Mechanisms of Viral Emergence and Persistence* 45–57 (Springer-Verlag, Vienna, 2005). https://doi.org/10.1007/3-211-29981-5_5.
37. Prist, P. R. et al. Moving to healthier landscapes: Forest restoration decreases the abundance of Hantavirus reservoir rodents in tropical forests. *Sci. Total Environ.* **752**, 141967 (2021).
38. Engelthaler, D. M. et al. Climatic and environmental patterns associated with hantavirus pulmonary syndrome, Four Corners region, United States. *Emerg. Infect. Dis.* **5**, 87–94 (1999).
39. Ji, H., Li, K., Shang, M., Wang, Z. & Liu, Q. The 2016 severe floods and incidence of hemorrhagic fever with renal syndrome in the Yangtze River Basin. *JAMA Netw. Open* **7**, e2429682 (2024).
40. Latinne, A. et al. Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **15**, 10705 (2024).
41. Hassanin, A., Tu, V. T., Curaudeau, M. & Csorba, G. Inferring the ecological niche of bat viruses closely related to SARS-CoV-2 using phylogeographic analyses of Rhinolophus species. *Sci. Rep.* **11**, 14276 (2021).
42. Hayman, D. T. S. et al. Developing one health surveillance systems. *One Health* **17**, 100617 (2023).
43. McGowan, C. R. et al. Community-based surveillance of infectious diseases: a systematic review of drivers of success. *BMJ Glob. Health* **7**, e009934 (2022).
44. Yano, T. et al. A participatory system for preventing pandemics of animal origins: Pilot study of the participatory one health disease detection (PODD) system. *JMIR Public Health Surveill.* **4**, e25 (2018).
45. Tan, C. C. S. et al. Genomic screening of 16 UK native bat species through conservationist networks uncovers coronaviruses with zoonotic potential. *Nat. Commun.* **14**, 3322 (2023).
46. Carlson, C. J. et al. Climate change increases cross-species viral transmission risk. *Nature* **607**, 555–562 (2022).
47. Plowright, R. K. et al. Land use-induced spillover: a call to action to safeguard environmental, animal, and human health. *Lancet Planet. Health* **5**, e237–e245 (2021).
48. Gibb, R., Franklinos, L. H. V., Redding, D. W. & Jones, K. E. Ecosystem perspectives are needed to manage zoonotic risks in a changing climate. *BMJ* **371**, m3389 (2020).
49. Jagadesh, S. *et al.* Disease X. In *MOOD Project (MOnitoring Outbreak events for Disease surveillance in a data science context). Drivers and trends of zoonotic pathogens, antimicrobial resistance and Disease X in Europe.* (eds. Cataldo, C., Rizzoli, A. & Busani, L.) 80–94 (Rapporti ISTISAN 24/16, Roma: Istituto Superiore di Sanità, 2024).
50. Marcoulides, K. M. & Raykov, T. Evaluation of variance inflation factors in regression models using latent variable modeling methods. *Educ. Psychol. Meas.* **79**, 874–882 (2019).
51. Phillips, S. J. et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197 (2009).
52. Yackulic, C. B. et al. Presence-only modelling using MAXENT: when can we trust the inferences?. *Methods Ecol. Evol.* **4**, 236–243 (2013).
53. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
54. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
55. Vieilledent, G. et al. hSDM: Hierarchical Bayesian species distribution models. *R Package Version* **1**, (2014).
56. Vignali, S., Barras, A. G., Arlettaz, R. & Braunisch, V. SDMtune: An R package to tune and evaluate species distribution models. *Ecol. Evol.* **10**, 11488–11506 (2020).
57. Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**, 225–232 (2019).
58. Raes, N. & ter Steege, H. A null-model for significance testing of presence-only species distribution models. *Ecography* **30**, 727–736 (2007).

Acknowledgements

S.J. was supported by the International Society of Infectious Diseases (ISID) as part of a grant for the MOOD project. This publication is a product of the MOOD project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 874850 and is catalogued as MOOD 121. The contents of this publication are the sole responsibility of the authors and don't necessarily reflect the views of the European Commission. WVB, is a member of Outbreak Research Team of the Institute of Tropical Medicine. This Team is financially supported by the Department of Economy, Science, and Innovation of the Flemish government.

Author contributions

S.J., C.C., L.B. and A.R. were involved in the literature review. S.J. developed the geospatial model. S.J. and E.A. wrote the main manuscript text and S.J. and E.VK prepared the maps. All authors reviewed the manuscript.

Funding
S.J. was supported by the International Society of Infectious Diseases (ISID) as part of a grant for the MOOD project. This publication is a product of the MOOD project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 874850 and is catalogued as MOOD 121. The contents of this publication are the sole responsibility of the authors and don't necessarily reflect the views of the European Commission. WVB, is a member of Outbreak Research Team of the Institute of Tropical Medicine. This Team is financially supported by the Department of Economy, Science, and Innovation of the Flemish government.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-05588-8.

Correspondence and requests for materials should be addressed to S.J.

Reprints and permissions information is available at www.nature.com/reprints.