

Spatiotemporal Dynamics and Predictive Modeling of Meningitis in Cameroon: Integrating Bayesian Hierarchical and Machine Learning Approaches for Public Health Surveillance

Ishmael Bakpianefene Ayeng

ishmael.ayeng1@gmail.com

African Institute for Mathematical Sciences

Solange Whegang Youdom

Université de Dschang

Boris Kompguep Mipo

Ministry of Public Health

Ancel Emmanuel Songo

Catholic University of Central Africa

Henri E. Z Tonnang

University of KwaZulu-Natal

Berges Tsanou

Université de Dschang

Jules Brice Tchatchueng Mbougua

Université de Yaoundé I

Research Article

Keywords: meningitis, spatiotemporal modeling, Bayesian hierarchical models, LISA, machine learning, district, Cameroon, disease surveillance

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11029443/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Spatiotemporal Dynamics and Predictive Modeling of Meningitis in Cameroon: Integrating Bayesian Hierarchical and Machine Learning Approaches for Public Health Surveillance

Ishmael Bakpianefene Ayeng^{1,2*}, Solange Whegang Youdom³, Boris Kompguep Mipo⁵, Ancel Emmanuel Songo^{5,7}, Henri E. Z. Tonnang^{9,10}, Berges Tsanou⁸, Jules Brice Tchatchueng Mbougua^{4,6}

***Corresponding author**

Abstract:

Background: Meningitis remains a severe public health threat across the African meningitis belt, where recurrent seasonal epidemics strain already resource-limited health systems. In Cameroon, the burden is most acute in the North and Far North regions, where climatic conditions, population density, and proximity to porous international borders create conditions favorable to transmission. Although the introduction of MenAfriVac in 2011 achieved remarkable reductions in serogroup~A disease, district-level spatially heterogeneous risk persists, and the emergence of non-A serogroups introduces new surveillance challenges. To this end, national surveillance systems remain predominantly reactive, with resources mobilized only after epidemic thresholds are formally exceeded.

Methods: We analyzed national meningitis surveillance secondary data from the Ministry of Public Health (MINSANTE), covering 206 health districts across all ten administrative regions of

Cameroon from 2017 to 2025 ($n = 88,168$ district-week observations). A Bayesian spatiotemporal framework incorporating the reparameterized Besag--York--Mollié (BYM2) model and Local Indicators of Spatial Association (LISA) was employed to characterize spatial clustering and temporal trends in district-level incidence. Ensemble machine learning models (Random Forest, XGBoost, and LightGBM) were implemented for outbreak prediction, using 21 systematically engineered features. Performance was evaluated using R^2 , RMSE, MAE, and AUC-ROC on a 70:15:15 split.

Results: All ensemble models achieved demonstrably robust predictive accuracy: $RF(R^2 = 0.94, MAE = 0.002)$, $XGBoost(R^2 = 0.95, MAE = 0.003)$, and $LightGBM(R^2 = 0.97, MAE = 0.002)$. For next-outbreak classification ($C \geq 5$ cases within a 2–4-week horizon), models achieved AUC-ROC of 0.95–0.97 and AUC-PR of 0.28–0.41, reflecting strong discrimination but modest precision given the rarity of true outbreak events (1.3% of district-weeks). The BYM2 model confirmed a significant and sustained declining temporal trend $\hat{\beta} = -0.089$; 95%CrI: -0.113 to -0.065 ; $IRR = 0.915$), corresponding to an 8.5% annual incidence reduction consistent with the impact of MenAfriVac and associated interventions. Despite this national improvement, substantial spatial heterogeneity persisted ($\rho = 0.682$; 95%CrI: $0.456 - 0.892$), with 15 high-risk districts concentrated in the Far North (8), North (4), and Adamaoua (3). Garoua exhibited a 2.8-fold elevated risk (95% CrI: $2.1 - 3.8$) relative to the national average. LISA analyses identified 4-6 statistically significant hotspot clusters per year, with no consistent spatial or temporal downward trend.

Conclusion: Combining Bayesian spatiotemporal modeling with machine learning provides a complementary analytical framework for meningitis surveillance in Cameroon. Our findings revealed that substantial national progress can mask persistent, localized hotspots that aggregate statistics systematically obscure. Geographically targeted interventions, routine immunization integration, and embedding predictive analytics within national surveillance are essential for achieving the WHO goal of defeating meningitis by 2030.

keywords: meningitis; spatiotemporal modeling; Bayesian hierarchical models; LISA; machine learning; district, Cameroon; disease surveillance

1. BACKGROUND

1.1 The global burden and geography of bacterial meningitis

Bacterial meningitis is an acute inflammation of the protective membranes surrounding the brain and spinal cord, caused predominantly by *Neisseria meningitidis*, *Streptococcus pneumoniae*, and *Haemophilus influenzae* (1–3). The disease is notable for the speed with which it can progress from early symptoms---fever, headache, and neck stiffness---to coma, organ failure, and death, often within 24 to 48 hours of onset (4). Even with appropriate treatment, (5) reiterated that approximately one in six patients dies, and one in five survivors is left with permanent disabilities, including hearing loss, cognitive impairment, and motor deficits.

Bacterial meningitis imposes a combined burden of high mortality and lifelong morbidity, making it one of the most consequential infectious diseases worldwide. Its epidemiology, however, is markedly unequal: in high-income countries, annual incidence rarely exceeds 2 cases per 100,000

population, whereas in sub-Saharan Africa, epidemic periods can produce 100–800 cases per 100,000 population. This represents a disparity of approximately 50- to 400-fold (1,6–9). According to (10), the African meningitis belt remains a hyperendemic zone, despite the growing population. The geographical concentration of this burden within this zone is distinctly exacerbated by the sporadic disease transmission and recurring epidemic phenomenon, shaped jointly by the combined effects of ecological, climatic, and socioeconomic determinants that together create conditions uniquely favorable to large-scale transmission. (11) reiterated that the dry Harmattan winds of the Saharan season, from October to May across the belt, are the single most consistent environmental driver of epidemic cyclicality. Low relative humidity, typically falling below 40% during this period, desiccates the mucosal surfaces of the upper respiratory tract, compromising the physical barrier that would ordinarily limit bacterial colonization and invasion (12–14). Simultaneously, dust-laden aerosols carried by Harmattan winds directly irritate respiratory epithelium, triggering an inflammatory response that both facilitates bacterial adherence and further suppresses mucosal immunity (6,15,16). (6) emphasized that these mechanisms do not operate in isolation since they are amplified by seasonal crowding in confined spaces during cold nights, concurrent respiratory infections that further weaken host defenses, and the movement of highly susceptible populations across porous borders. The result is a transmission dynamic that is both predictable in its seasonal timing and devastating in its human consequences.

1.2 Meningitis in Cameroon: a country within the belt

The near-elimination of serogroup A has altered the competitive conceptual frame among circulating strains/serogroups, enabling the relative emergence of non-A serogroups---particularly C, W, and X---for which the existing MenAfriVac vaccine offers no protection and for which population immunity across Cameroon is largely absent (4,17,18). At the same time, a critical

structural gap has emerged in the failure to integrate MenAfriVac into the routine Expanded Programme on Immunization (EPI) in Cameroon: unlike rotavirus, pneumococcal, and human papillomavirus vaccines introduced around the same period, MenAfriVac has not been systematically incorporated into the childhood immunization schedule, leaving all birth cohorts since 2011 unprotected against serogroup A (19–21). Mathematical models project that, without routine immunization, 40–50% of the population in northern high-risk districts could be vulnerable to serogroup A resurgence by 2030 (22). This highlights how post-campaign vulnerability in new birth cohorts creates an ongoing need for strong surveillance.

1.3 Why are integrated spatiotemporal analytics needed now

Despite the evolving epidemiological dynamics, the operational and organizational structure and framework of public health response to meningitis in Cameroon have remained largely unchanged. Like many countries in the belt, district-level surveillance still operates around the WHO alert threshold of five cases per 100,000 population per week and the epidemic threshold of ten cases per 100,000 per week, which together define the conditions for escalating public health action (23). While these thresholds are important for resource-limited settings, their fundamental limitation is that they only ensure an inherently reactive response, which does not allow adequate preparatory measures to predict and prepare for emerging outbreaks. To address this, Besag and colleagues (24) established the foundational spatial statistical methods that could be applied to this problem. Moreover, for common challenges about data sparsity, overdispersed counts with strong spatial autocorrelation, and a high proportion of zeros, Bayesian hierarchical spatial models, in particular the reparameterized Besag--York--Mollié (BYM2) formulation, are well-suited (24–26). By simultaneously estimating district-specific relative risks and borrowing statistical strength from

neighboring areas, BYM2 models distinguish persistent structural hotspots from stochastic fluctuations (26), while Local Indicators of Spatial Association (LISA) extend this spatial characterization to identify statistically significant cluster configurations on an annual basis, which provides the granularity necessary for resource targeting (27) . To complement these efforts, machine learning approaches offer complementary strengths in the predictive domain by learning the complex nonlinear relationships across large feature spaces, including lagged epidemiological signals, seasonal harmonics, and historical burden indicators, without the parametric assumptions that constrain traditional statistical models, and they can be updated in near-real-time as new surveillance data arrive (26,28).

We acknowledge the contributions of earlier studies that have applied distinct methodological traditions to meningitis surveillance in sub-Saharan Africa. (28) developed a multi-state Markov framework for categorized meningitis incidence, (26) applied spatiotemporal clustering methods to Niger's surveillance data, and (29) modeled meningococcal transmission dynamics through compartmental frameworks. To the best of our knowledge, no national study has integrated Bayesian spatiotemporal risk mapping and machine-learning outbreak prediction using district-level meningitis surveillance data from Cameroon. Hence, this study focuses on addressing that gap.

Drawing on 88,168 district-level weekly observations across all ten administrative regions of Cameroon from 2017 to 2025, a period that simultaneously captures the post-MenAfriVac incidence decline, as well as the emergence of non-A serogroup outbreaks, and the persistence of spatial heterogeneity, we develop, implement, and evaluate an integrated analytical framework combining Bayesian spatiotemporal modeling with ensemble machine learning for meningitis surveillance and prediction. Specifically, we pursue three objectives: (1) to characterize the spatial

clustering and temporal trajectory of district-level meningitis risk using BYM2 and LISA analyses; (2) to develop and evaluate predictive machine learning models for outbreak detection; (3) to identify persistent high-risk districts and derive actionable evidence for geographically targeted public health interventions in Cameroon; (4) build an interactive dashboard to support the ministry of public health, Cameroon for their decision on intervention by providing real-time alert prediction.

2. Methods

Cameroon is administratively organized into ten regions: Adamaoua, Centre, East, Far North, Littoral, North, Northwest, South, Southwest, and West, subdivided into approximately 206 health districts. The districts serve as the primary operational units for epidemiological surveillance, outbreak threshold evaluation, and the planning and delivery of vaccination campaigns (4,23). It is at this level that weekly case reports are compiled and where reactive public health actions are formally initiated and operationalized. This justifies health districts as both the natural spatial unit of our analysis for this surveillance-oriented study and the scale at which our findings can most directly inform operational decisions.

The dataset for this research was obtained as secondary data from the national meningitis surveillance system coordinated by the Direction de la Lutte contre la Maladie, les Epidemies et les Pandemies (DLMEP) under the Ministry of Public Health (MINSANTE), Cameroon. The dataset comprises cumulative weekly reports of the suspected number of meningitis cases, deaths, and population size from all health districts across all regions in Cameroon. The dataset spans epidemiological week 1 of 2011 through week 12 of 2025; however, our analytical focus was restricted to the period 2017-2025 (epidemiological weeks 1-428). This is to ensure a minimal confounding effect of major changes in surveillance system structure that occurred around 2016-

2017, ensure internal consistency across the analytical period, and align the dataset with the epidemiological era defined by post-MenAfriVac dynamics and the emerging non-A serogroup burden (17,18)

Moreover, spatial boundary files for all 206 health districts and the 10 administrative regions were also obtained from MINSANTE's Data Management Unit for the construction of a spatial neighborhood metric for Bayesian spatial models, the calculation of geographic centroids for distance-based analyses, and the creation of choropleth risk maps. These files were georeferenced to the WGS84 coordinate system (EPSG:4326).

2.1 Epidemiological indicators

From the three core raw surveillance variables, weekly case counts ($C_{\{ijt\}}$) death counts ($D_{\{ijt\}}$) and population denominators ($P_{\{ijt\}}$) for district i , region j , at epidemiological week t , we derived three standardized indicators that enable comparisons across districts of varying population size and across time periods with changing denominators.

The **Attack Rate (AR)** shows disease incidence per 100,000 population:

$$AR_{\{ijt\}} = \left(\frac{C_{\{ijt\}}}{P_{\{ijt\}}} \right) \times 100000$$

Case Fatality Ratio (CFR) captures disease severity and serves as a proxy for the quality and timeliness of clinical management:

$$CFR_{\{ijt\}} = \left(\frac{D_{\{ijt\}}}{C_{\{ijt\}}} \right) \times 100, C_{\{ijt\}} > 0$$

Finally, Cumulative Incidence over any seasonal window $[t_1, t_2]$ provides an integrated measure of seasonal burden independent of week-to-week stochastic variation:

$$AR_{\{ij\}}^{\{cum\}} = \left(\frac{\sum_{t=t_1}^{t_2} C_{\{ijt\}}}{\{P\}_{\{ij\}}} \right) \times 100000$$

These three indicators collectively constitute the basis for both the descriptive epidemiology and the outcome variables in subsequent modeling tasks.

2.2 Data Preprocessing and Feature Engineering

Our data preprocessing pipeline was implemented entirely in R and proceeded through distinct sequential steps designed to progressively improve data quality while preserving analytical transparency.

The steps addressed missing values, which arose from three distinct mechanisms requiring different treatments. Missing case counts in districts that reported population denominators in the same week were imputed as zero, reflecting the true absence of detected disease rather than a data gap. Weeks with missing population denominators were flagged and excluded from rate-based calculations while being retained for count-based analyses where population size is not required. Consecutive sequences of missing values exceeding four weeks in any district were classified as probable surveillance reporting lapses, arising from administrative disruptions, instability, or logistical failures common in remote northern districts, and were excluded from temporal trend analyses to prevent their confounding effect on time-series models.

Districts that underwent genuine boundary changes---typically splits or mergers---were consolidated into temporally consistent spatial units by merging data from the constituent districts, accepting a small reduction in spatial resolution in exchange for analytical consistency across the full study period.

195 The final dataset comprised 206 health districts, 428 epidemiological weeks, and 88,168 district-
 196 week records, with a completeness rate of 99.2% of expected observations after excluding
 197 structural missingness attributable to the above-described reporting lapses. We applied standard
 198 quality checks to all retained observations, validating that $D_{\{ijt\}} \leq C_{\{ijt\}}$ for all records confirmed
 199 in all cases; flagging statistical outliers exceeding $\mu \pm 5\delta$ for manual clinical review, and cross-
 200 validating population denominators against INS projections (discrepancies below 5% across all
 201 districts and years).

202 The feature engineering focused on enabling our machine learning models to capture the
 203 underlying temporal, seasonal, and spatial patterns in the surveillance data. Drawing on established
 204 transmission theory, the documented seasonality of the African meningitis belt, and empirical
 205 patterns identified in the surveillance data, we systematically engineered 21 features across six
 206 functional categories (see Table~1).

207 **Temporal features** (4) captured calendar time at multiple resolutions—calendar year,
 208 epidemiological week number, and quarter—allowing the model to track both long-term secular
 209 trends and within-year phase.

210 **Seasonal features** (2) were constructed to encode the Harmattan-driven epidemic seasonality of
 211 the meningitis belt (13,16) through a sine–cosine cyclical encoding of the epidemiological week,
 212 $Sin(w) = \sin\left(\frac{(2\pi w)}{52}\right), Cos(w) = \cos\left(\frac{(2\pi w)}{52}\right)$ so that week 52 and week 1 are represented as
 213 adjacent rather than distant, without a discontinuity at the year boundary.

214 **Lag features** (6) captured the strong temporal autocorrelation that is an inherent property of
 215 infectious disease dynamics: current case counts are inherently linked to recent infection history
 216 through ongoing transmission chains, progressive susceptible depletion within affected

communities, and the persistence of the environmental conditions that initiated transmission. Case and death counts were lagged at $\tau \in \{1,2,4\}$ weeks, spanning the 2- to 10-day incubation period of *N. meningitidis* (15) through multi-week outbreak trajectories:

$$CASES_{LAG_{\tau(t)}} = C_{\{i,j,t-\tau\}}$$

All lag features use strictly weeks $t-1$ and earlier relative to the prediction week, so that no same-week information enters the model.

Moving average features (4) provided temporally smoothed representations of epidemiological trends by computing rolling means and standard deviations over 4- and 8-week windows. The rolling means captured the underlying trend in case counts while filtering out week-to-week stochastic noise inherent in passive surveillance data. Each rolling-window feature at week t was computed using only weeks $t-1$ and earlier, so that a district's own current-week case count could never leak into its own predictors.

District-level features (3) captured the heterogeneous epidemiological profiles of individual health districts that cannot be inferred from regional or national patterns alone. These included a district identifier (encoded), population size, and each district's own expanding-window historical outbreak frequency, using only weeks strictly before t . This distinguishes persistently high-burden districts in the north from sporadically or chronically low-burden districts in the south, a distinction invisible to any model that treats all districts as exchangeable.

Regional features (2) complemented district-level context with a region identifier (encoded) and a spatial-neighbor signal: the mean lagged case count (week $t-1$) among spatially contiguous neighboring districts, using the same queen-contiguity adjacency structure used for the BYM2 and LISA analyses. This allows the model to learn that a district's near-term risk is elevated when its

239 immediate neighbors are already experiencing rising case counts, consistent with the spatial
 240 clustering documented in Section 3.2.

241 **Table 1: Categories and descriptions of the engineered features used in machine learning**
 242 **models for meningitis outbreak prediction**

Category	n	Description
Temporal	4	Year, epidemiological week number, quarter
Seasonal	2	Cyclical encoding of epidemiological week (sine and cosine transforms)
Lag features	6	Lagged cases and deaths at $\tau = 1, 2, 4$ weeks (strictly week $t-1$ and earlier)
Moving averages	4	Rolling means and standard deviations over 4- and 8-week windows, computed using only weeks up to $t-1$
District-level	3	District identifier (encoded), population, and each district's own expanding-window historical outbreak frequency (weeks strictly before t)
Regional	2	Region identifier (encoded); mean lagged case count among spatially contiguous neighboring districts (queen contiguity)
Total	21	

243 *Note: See appendix for feature importance plot (Figure 4)*

2.3 Bayesian spatiotemporal modeling

2.3.1 BYM2 hierarchical Poisson model

The Besag–York–Mollié (BYM) model (24) is a Bayesian Poisson hierarchical model widely used for fitting spatial count data, particularly in the field of disease mapping. The BYM model extends the generalized linear Poisson model by including both spatially structured and unstructured random effects in the regression model. We modeled the district-level meningitis counts within a Bayesian hierarchical Poisson framework (24,25,32). This enables the model to accommodate overdispersion and excess zeros, borrow statistical strength across spatially adjacent districts to improve estimation in sparsely observed areas, and decompose observed spatial variation into components attributable to neighborhood context versus district-specific factors. The model is specified as follows:

$$Y_{\{i,t\}} \sim \text{Poisson}(N_{\{i,t\}} \cdot \lambda_{\{i,t\}})$$

$$\log(\lambda_{\{i,t\}}) = x_{\{i,t\}}^T \beta + \varphi_i + \delta_t + \eta_{\{i,t\}}$$

where $Y_{\{i,t\}}$ is the observed case count, $N_{\{i,t\}}$ is the population at risk, $\lambda_{\{i,t\}}$ is the incidence rate to be estimated. $X_{it}^T \beta$ is the linear predictor of fixed effects (including a temporal trend variable for years since vaccination). φ_i is a district-specific spatial random effect, δ_t is a temporal random effect following an AR(1) structure, and $\eta_{\{i,t\}}$ captures residual space-time interaction.

The reparametrized BYM2 formulation (24) improves upon earlier Besag--York--Mollié specifications by replacing two separate variance parameters with a single mixing parameter ρ that directly quantifies the proportion of spatial variation attributable to structured neighborhood effects versus unstructured district heterogeneity:

$$\eta_i = \alpha + x_i^T \beta + \left(\frac{1}{\sqrt{\tau}}\right) (\sqrt{\rho} \varphi_i^* + \sqrt{1-\rho} \theta_i^*)$$

where φ_i^* and θ_i^* are scaled spatial (ICAR) and unstructured random effects, respectively; τ controls overall precision; and $\rho \in [0,1]$ is interpreted as, when ρ approaches 1, a district's risk is predominantly determined by its geographical context, and when ρ approaches 0, idiosyncratic district-level factors dominate. The joint distribution of the intrinsic CAR (ICAR) spatial effects is:

$$\varphi \sim N(0, \sigma^2 (D - W)^{-1})$$

Where W is the binary spatial adjacency matrix encoding which district pairs share a boundary, and $D = \text{diag}(n_{\{\delta_1\}}, \dots, n_{\{\delta_n\}})$ is the diagonal matrix of neighborhood sizes. Bayesian inference was performed via Integrated Nested Laplace Approximation (INLA) using the R-INLA package (Finley et al., 2014), which provides accurate marginal posterior distributions without the convergence challenges associated with Markov Chain Monte Carlo methods for large spatial datasets. We used the Deviance Information Criterion (DIC) and the Widely Applicable Information Criterion (WAIC) to evaluate the model's fit and out-of-sample outbreak detection performance via cross-validation.

2.3.2 Spatial autocorrelation and LISA cluster detection

We also assessed global spatial autocorrelation using Moran's I, which quantifies the degree to which districts with similar observed values are spatially proximate. To provide district-level cluster granularity, we use the Local Indicators of Spatial Association (LISA) annually from 2017 to 2025 to identify statistically significant High-High hotspot clusters (high-burden districts surrounded by high-burden neighbors), Low-Low coldspot clusters, and spatial outliers. Crucially,

286 computing LISA annually rather than for the aggregate study period allows us to distinguish
 287 between hotspots that are geographically stable across years---and therefore likely driven by
 288 persistent structural factors amenable to standing surveillance investment---and clusters that shift
 289 in location from year to year, as well as those requiring adaptive surveillance allocation.

290 Moran's I global statistic was computed using:

$$291 \quad I = \frac{N}{W} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

292 Whereas, for each local indicator across the respective years, the test statistic was:

293 LISA (Local Moran's I)

$$294 \quad I_i = \frac{(x_i - \bar{x})}{m_2} \cdot \sum_{j=1}^N w_{ij} (x_j - \bar{x})$$

295 Variance (m_2)

$$296 \quad m_2 = \frac{\sum_{k=1}^N (x_k - \bar{x})^2}{N}$$

297 Where N represents the number of spatial units, x_i , observed cases at district i, $\bar{x}$ represents the
 298 mean cases, w_{ij} represents the spatial weight between location i and j, which defines the
 299 neighborhood structure.

300 **2.4 Machine learning models**

301 For the outbreak prediction task, we trained Random Forest models (33) with B = 500 decision
 302 trees, each constructed on an independently bootstrapped subsample of the training data with a

random subset of candidate features considered at each split node. The ensemble prediction averages across all trees:

$$\hat{f}_{RF(x)} = \left(\frac{1}{B}\right) \sum_{b=1}^B T_{b(x)}$$

Our bagging strategy ensures a reduction in variance without inflating bias, yielding predictions more stable than any single tree. Variable importance was quantified through the mean decrease in Gini impurity across all trees and feature-involving splits, providing a data-driven ranking of which engineered features most strongly discriminate between outbreak and non-outbreak district-weeks (34,35).

Finally, our XGBoost algorithm (36) takes a fundamentally different approach, building an ensemble sequentially by training each new tree on the residual errors of all preceding trees. This additive procedure optimizes a regularized objective that penalizes model complexity:

$$L(\varphi) = \sum_i \ell(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \Omega(f) = \gamma T + \left(\frac{1}{2}\right) \lambda ||w||^2$$

where ℓ is a differentiable convex loss function, T is the number of tree leaves, and w are leaf weights. At each boosting round, a computationally efficient second-order Taylor approximation reduces the objective to a tractable form:

$$L^t \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \left(\frac{1}{2}\right) h_i f_t^2(x_i) \right] + \Omega(f_t)$$

where g_i and h_i are the first and second derivatives of the loss with respect to the prediction from the previous round. The regularization terms γ and λ prevent overfitting, which is an essential consideration in a surveillance dataset where the high proportion of zero observations creates the risk for models to predict zeros everywhere and achieve superficially impressive metrics.

2.5 Model performance and evaluation

We formulated two complementary machine learning prediction tasks to capture both the magnitude and the binary occurrence of outbreaks. For the regression task, our outcome variable was the raw weekly case count

$$Y_{\{ijt\}} = C_{\{ijt\}},$$

enabling us to quantify predicted outbreak intensity. For the classification task, the prediction outcome was a binary outbreak indicator for the next week:

$$Y_{ijt} = I(C_{ijt} \geq 5 \text{ cases per district} - \text{week})$$

Given how few next-week outbreak events remain, the classification target was defined over a 2–4-week forward horizon, $Y_{i,j,[t+1,t+4]} = I(\max(C_{i,j,t+1}, \dots, C_{i,j,t+4}) \geq 5)$, which reflects an operationally actionable early-warning window while providing enough positive events for a stable evaluation.

This threshold aligns with the WHO alert criterion for populations exceeding 30,000 (23) and with MINSANTE's operational protocols for initiating enhanced surveillance and outbreak investigation. Meanwhile, the regression performance was evaluated using R^2 , RMSE, MAE, and MAPE; classification performance was assessed using the F1-score and AUC-ROC. The data were split chronologically by year into training (70%), validation (15%), and test (15%) sets to prevent temporal data leakage; the validation set was used for hyperparameter tuning, and all evaluations reported below were conducted on the held-out 15% test set.

344

345 **2.6 Temporal trend analysis of vaccination impact**

346 To complement the spatially adjusted BYM2 trend estimate, we fitted a log-linear regression model
347 to annual national incidence rates as a function of time elapsed since MenAfriVac introduction in
348 2011:

$$349 \quad \log(IncidenceRate_t) = \beta_0 + \beta_1 t_{vaccine} + \varepsilon_t$$

350 where $t_{\{vaccine\}}$ is the number of years since campaign initiation. This aggregate-level analysis
351 provides an accessible summary of the vaccination program's temporal effectiveness and serves as
352 a complement to the spatially stratified BYM2 estimates. All analyses were implemented in R
353 (version 4.6) and Python (version 3.10), using *the INLA, spdep, sf, randomForest, xgboost, and*
354 *lightgbm* packages.

355 **2.7 Ethical considerations**

356 This study used secondary meningitis surveillance data obtained from the Ministry of Public
357 Health (MINSANTE), Cameroon. The analysis was conducted on aggregated surveillance data
358 and did not involve direct recruitment or interaction with patients, collection of biological samples,
359 or access to identifiable individual information.

360 Because the study relied solely on secondary analysis of routinely collected surveillance data,
361 ethical approval and individual informed consent were not required under the institutional and
362 data-governance arrangements governing the use of MINSANTE data. The study was conducted
363 in accordance with applicable ethical principles and relevant regulations for the responsible use of
364 health surveillance data.

3. Results

The final dataset comprised 88,168 district-week records from the entire population across all age groups spanning 206 health districts over 428 epidemiological weeks from 2017 through the first quarter of 2025. The distributional profile of the data shows approximately 94.2% of all district-week observations recorded zero suspected cases.

Among the 5.8% of non-zero observations, the case count distribution was markedly right-skewed, with a mean of approximately two cases and a median of one case per district-week, indexed by occasional extreme values from localized outbreaks. The mean case fatality ratio across the study period was 5.41%, which is substantially lower than the historical pre-MenAfriVac CFRs exceeding 10%; nonetheless, it signals that a substantial fraction of detected cases is progressing to fatal outcomes.

Against this baseline profile, the aggregate temporal trajectory shows sustained progress with national case counts declining from 1,588 cases and an incidence of 2.08 per 100,000 population in 2017 to 816 cases, an incidence rate of 0.98 in 2024, and 495 cases and 0.59 per 100,000 by the first quarter of 2025---an overall reduction of 73.4% across eight years (Table~2). We quantified this reduction rate using the log-linear trend model, and the results showed a statistically significant annual decline of 12.14% ($\beta_1 = -0.129$; *standarderror* = 0.035; $t = -3.67$; $p = 0.008$; $R^2 = 0.658$), consistent with the sustained impact of MenAfriVac vaccination campaigns and associated community sensitization.

387 **Table 2: Annual Meningitis Surveillance Summary (Cameroon, 2017–2025)**

Year	Cases	Incidence / 100,000	CFR (%)	% Change
2017	1,588	2.08	4.98	Baseline
2018	1,025	1.31	8.20	-35.4%
2019	2,025	2.56	5.23	+27.5%
2020	1,128	1.41	3.10	-29.0%
2021	1,053	1.30	3.32	-33.7%
2022	1,227	1.50	2.61	-22.7%
2023	1,128	1.36	4.08	-29.0%
2024	816	0.98	1.84	-48.6%
2025 (Q1)	495	0.59	0.81	-68.8%

388

389 *Note: Overall decline in incidence, mortality, and CFR across years 6--14 post-MenAfriVac*

390 *introduction, with year-to-year fluctuation. The 2019 spike is attributed to a localized outbreak*

391 *concentrated in the Far North region. % change is calculated relative to the 2017 baseline*

392

3.1 Bayesian Spatiotemporal Model Results

3.1.1 Bayesian Spatial Heterogeneity Analysis

We further decomposed this national temporal trend using the BYM2 model while simultaneously accounting for the spatial autocorrelation structure of district-level data.

The posterior mean coefficient for years elapsed since the introduction of MenAfriVac was $\beta = -0.089$ (95%CrI: -0.113 to -0.065), corresponding to an Incidence Rate Ratio (IRR) of 0.915 (95% CrI: 0.893- 0.937) and an estimated 8.5% annual reduction in district-level incidence after spatial adjustment. Our test of the hypothesis to examine whether the coefficient was negative showed that the posterior probability was effectively 1.0 ($Pr(\beta < 0|y) > 0.998$).

This provides extremely strong Bayesian evidence that the observed national improvement is not a statistical artifact of changing spatial dynamics or geographic redistribution of cases. However, it reflects a genuine, sustained, and spatially pervasive reduction in transmission risk (Table~3).

The mixing parameter $\rho = 0.682$ (95%CrI: $0.456 - 0.892$) indicates that 68.2% of the total spatial variation in district-level meningitis risk was spatially structured---that is, attributable to neighborhood context and shared regional exposures rather than to idiosyncratic district-specific factors.

Across the 206 districts, the spatial random effects ranged from $\varphi_1 = -0.89$ in Kribi (Littoral region) to $\varphi_i = +1.23$ in Garoua (North region), translating into relative risks from 0.41 to 3.42 after adjustment for the temporal trend. In total, 15 districts (7.6% of all districts) were found to exhibit significantly elevated/higher-than-normal risk, defined as a posterior mean relative risk exceeding 1.5 with a 95% credible interval that excluded 1.0. Also, 8 of these districts were found

in the Far-North, 4 in the North, and 3 in the Adamaoua regions. The highest-risk district, Garoua, depicts a 2.8-fold increase in risk (95% CrI: 2.1--3.8) relative to the national average.

This is particularly concerning because Garoua, among other areas, has been a focal point for consistent vaccination exercises. Therefore, an elevated risk 2.8 times higher than the national average raises important questions. However, this scenario could only be a hypothesis, demanding additional data to verify whether this signals the emergence of serogroup shifts toward non-A strains not targeted by MenAfriVac, whether coverage statistics overestimate the proportion of the population actually protected due to population mobility and displacement, or whether local transmission factors operate with sufficient intensity to sustain elevated risk even in partially immunized communities.

Conversely, 12 districts (6.1%) showed significantly reduced risk ($RR < 0.67$), primarily in the South and Southwest regions. We found heterogeneous but geographically patterned vaccine protection across the country.

The model demonstrated a strong statistical fit ($DIC = 3,456.8$; $WAIC = 3,459.2$), with out-of-sample validation performance and correctly identified 85% of observed outbreak weeks in cross-validation.

Table 3: Results of the Bayesian Spatiotemporal BYM2 Model for Meningitis Risk in Cameroon (2017–2025)

Component	Parameter	Estimate	95% CrI
Fixed Effects	Intercept	-1.235	(-1.477, -0.988)
	Years since vaccine	-0.089	(-0.113, -0.065)
Random Effects	Spatial SD	0.346	(0.278, 0.432)
	Mixing parameter (ρ)	0.682	(0.456, 0.892)
	Temporal SD	0.123	(0.089, 0.178)
Model Fit	DIC	3,456.8	—
	WAIC	3,459.2	—

434

435 *Notes: IRR for years since vaccine = 0.915 (95% CrI: 0.893–0.937), corresponding to an 8.5%*
436 *annual decline in meningitis incidence after adjusting for spatial autocorrelation.*
437 *CrI = Credible Interval; IRR = Incidence Rate Ratio.*

438

439 **3.2 Spatial autocorrelation and the annual LISA cluster dynamics**

440 We also provided an initial characterization analysis of spatial dependence across the district
441 network using the global Moran's I. Total meningitis case counts showed statistically significant
442 positive spatial autocorrelation ($I = 0.080$; $Z = 2.27$; $p = 0.022$), with a weak but significant
443 clustering of high-burden neighbors rather than random distribution across the country. However,

when we normalized the case counts by population size to derive the incidence rates, the autocorrelation was statistically insignificant and weakened considerably ($I = 0.054$; $p = 0.098$).

Our analysis suggests that a portion of the observed geographic clustering of cases is attributable to the spatial concentration of the at-risk population in regions where districts are larger, more densely populated, and more ecologically exposed, rather than to uniformly elevated per-capita transmission risk across the entire north (Table~4). Population-adjusted incidence, therefore, reveals a more nuanced spatial risk surface than raw case counts alone would imply.

Table 4: Global Moran’s I Test for Spatial Autocorrelation of Meningitis Outcomes Across Health Districts, Cameroon (2017–2025)

Variable	Moran’s I	Z-score	p-value	Interpretation
Total Cases	0.0804	2.27	0.022	Significant positive spatial autocorrelation
Incidence Rate	0.0541	1.32	0.098	Weak, non-significant spatial clustering
Total Deaths	0.0549	1.44	0.084	Marginal spatial dependence

The LISA analyses revealed the district-level spatial granularity and their annual implementation (2017--2025) to track how the cluster configuration evolved through time.

Our findings revealed that there were between 3 and 6 statistically significant High-High (hotspot) clusters identified each year throughout the study period. Moreover, the hotspot cluster numbers show no consistent downward trend despite the apparent decline in national incidence.

The distribution of Low-Low (cold spot) clusters, by contrast, expanded substantially over the study period, from 4 in 2017 to 16 in 2025, and this was consistent with the progressive incidence reduction in the historically lower-risk southern and western regions where vaccine uptake and surveillance infrastructure may be more complete.

The year-to-year variation analysis in the number, location, and spatial extent of hotspot clusters further reveals that meningitis risk in Cameroon is not static even at the local level.

Specific districts that appear as hotspots in one year may not be depicted in the succeeding year, and new hotspot configurations emerge at different locations depending on the particular combination of seasonal, demographic, and immunological features prevailing in that year. This temporal instability in the spatial risk surface is consistent with the known sensitivity of meningitis transmission to inter-annual variations in Harmattan timing and intensity, as documented in neighboring countries in the belt (16,30) and has direct implications for the design of surveillance networks as well as the allocation of resources between predictable standing infrastructure in persistently high-risk areas and flexible reactive capacity for the unpredictable components of the risk surface.

Figure 1: Local Indicators of Spatial Association (LISA): Meningitis incidence cluster by year (2017 to 2025) (See Appendix)

3.3 Machine learning for outbreak prediction: regression and classification

3.3.1 Regression task

Our machine learning model for predicting meningitis outbreaks for the next week in Cameroon showed strong robust predictive performance.

Before the feature engineering, all models substantially outperformed the baseline (Table~5) with Linear Regression achieving the best R^2 ($R^2 = 0.20$, RMSE = 0.333, MAPE = 80.4%), narrowly ahead of LightGBM ($R^2 = 0.19$, RMSE = 0.336, MAPE = 78.3%), XGBoost ($R^2 = 0.15$, RMSE = 0.342, MAPE = 78.1%), and Random Forest ($R^2 = 0.12$, RMSE = 0.350, MAPE = 76.2%). The high MAPE values across all models reflect the difficulty of proportional error metrics on a strongly zero-inflated series rather than a change in absolute prediction error.

Table 5: Performance of Machine Learning Models on the Regression Task before Feature Engineering (Case Count Prediction) (Held-Out Test Set)

Model	RMSE	MAE	R^2	MAPE (%)
Baseline	0.3773	0.1796	-0.034	89.54
Linear Regression	0.3334	0.1117	0.197	80.36
Random Forest	0.3498	0.1304	0.116	76.18
LightGBM	0.3357	0.1202	0.186	78.35
XGBoost	0.3424	0.1177	0.153	78.14

492

493 After incorporating the futures, all three ensemble models substantially outperformed the baseline
 494 linear regression model across all the evaluation metrics (Table~6). LightGBM achieved the
 495 strongest overall predictive performance ($R^2 = 0.97$; $RMSE = 0.021$; $MAE =$
 496 0.002 ; $MAPE = 2.52\%$) followed closely by XGBoost ($R^2 = 0.95$; $RMSE =$
 497 0.026 ; $MAPE = 3.54\%$) and Random Forest ($R^2 = 0.94$; $RMSE =$
 498 0.030 ; $MAPE = 2.15\%$). The models achieved prediction errors (RMSE) below 0.03 for
 499 case counts and are suitable to be operationally meaningful for a real-time surveillance alert
 500 system.

501 **Table 6: Performance of Machine Learning Models on the Regression Task after Feature**
 502 **Engineering (Case Count Prediction) (Held-Out Test Set)**

Model	RMSE	MAE	R ²	MAPE (%)
Baseline	0.3773	0.1796	-0.034	89.54
Linear Regression	0.1834	0.0438	0.756	33.44
Random Forest	0.0295	0.0019	0.94	2.15
LightGBM	0.0211	0.0020	0.97	2.52
XGBoost	0.0256	0.0029	0.95	3.54

503

3.3.2 Classification task

For the binary outbreak classification task, the goal was to predict whether a district would experience an outbreak ($C \geq 5$ cases) within a forward-looking 2–4 week horizon, using only predictor information available through the current week, thereby supporting proactive decision-making. The ensemble models showed strong discrimination but modest precision at the operating thresholds typically used for classification: Logistic Regression achieved the best AUC-PR (0.41) and AUC-ROC (0.97), followed closely by Random Forest (AUC-PR = 0.41, AUC-ROC = 0.95), LightGBM (AUC-PR = 0.31, AUC-ROC = 0.96), and XGBoost (AUC-PR = 0.28, AUC-ROC = 0.96). True next-outbreak windows are rare in this data (49 of 21,419 test-set district-week windows, 0.23%), so F1 at a standard 0.5 threshold remained low across all models (0.04–0.25); recall could be raised substantially at the cost of precision by lowering the decision threshold, and we report the precision–recall tradeoff explicitly (Figure~2), since the appropriate threshold depends on the relative operational cost of a missed outbreak versus a false alarm. Taken together, these results indicate that the engineered feature set (short-lag case counts, strictly-lagged rolling epidemiological averages, seasonal Fourier terms, each district's historical outbreak frequency, and spatially lagged case counts from contiguous neighboring districts) carries genuine predictive signal for outbreak risk.

Figure 2: Precision-Recall and ROC curves for next-outbreak classification (2–4 week horizon, $C \geq 5$) (See Appendix)

4. Discussion

4.1 Vaccination Impact and Temporal Trends

Our central findings are threefold: (1) The observed decline is consistent with the post-MenAfriVac period and associated interventions, but causal attribution is limited by the absence of district-level vaccination coverage and serogroup-specific data; (2) persistent spatial heterogeneity with localized high-risk hotspots in the northern regions despite national vaccination progress; and (3) robust predictive accuracy of ensemble machine learning models for outbreak risk detection.

From our combined temporal and spatial analyses, there is evidence of substantial and statistically robust national progress in the effort to combat meningitis. This progress is accompanied by an uneven spatial distribution. The 8.5% annual decline in district-level incidence estimated by the BYM2 model, the 73.4% overall reduction in national case counts between 2017 and 2025, and the near-certainty of the posterior probability of a declining trend ($Pr(\beta < 0) > 0.998$) together constitute compelling evidence that MenAfriVac has contributed to a remarkable and sustained public health benefit to Cameroon.

Our findings are consistent with findings from the wider nine-country program documented by (10), and with the district-level analyses of vaccine impact published by (21,37) specifically for Cameroon. We also found a parallel reduction in the case fatality ratio---from 4.98% in 2017 to

0.81% in 2025- and this is likely attributable to the improvements in health system responsiveness alongside reduced case incidence (17).

4.2 Spatial heterogeneity and Persistent Hotspots

The BYM2 mixing parameter ($\rho = 0.682$) indicates that the spatial distribution of meningitis risk in Cameroon is predominantly structured rather than driven by district-specific idiosyncrasies. Specifically, two-thirds of the spatial variation in district-level risk is explained by a district's geographical position relative to high-burden neighbors. This suggests that risk in the northern districts is not merely the sum of independent local factors; however, they are likely spatially connected risk environments shaped by exposure to the influence of seasonal fluctuations like the Harmattan, cross-border transmission dynamics, and geographically concentrated infrastructure deficits (13,38). A practical implication of this finding is that if interventions are designed at the level of individual districts without consideration of spatial context, they may be less effective than those deployed with awareness of the connected risk field. For example, a district-level vaccination campaign achieving high coverage in a focal district may see imported reintroductions rapidly erode local gains if neighboring high-burden districts are not simultaneously addressed. Conversely, the identification of districts with relative risk significantly below the national average, such as the 12 low-risk districts in the South and Southwest, offers an opportunity to reallocate surveillance and response resources away from areas where the structural risk environment does not warrant intensive investment.

The persistence of Garoua as the highest-risk district ($RR = 2.8$; $95\%CrI: 2.1 - 3.8$), despite documented high vaccination coverage, is in this context one of the most notable epidemiological findings of our study. The pattern confirms observations from Burkina Faso (15) and northern Nigeria (27), where geographically localized risk factors, including prison populations, refugee

camps, occupationally crowded settings, and disrupted healthcare access, sustain elevated transmission independently of district-level vaccination coverage statistics. While these observations are possible, an alternative explanation is that reported coverage in Garoua may not accurately reflect the immunological protection of the actual at-risk population, particularly given the substantial numbers of internally displaced persons associated with the ongoing Boko Haram insurgency in the Far North (38) . Displaced individuals may have been separated both from their registered health district and from vaccination campaigns conducted there. Investigating this potential discrepancy between reported vaccine coverage and observed risk would require data linkages between surveillance, vaccination records, and displacement tracking that are not currently available in Cameroon at the time of this study, an important priority for future research.

Rather than revealing a fixed configuration of persistent hotspot districts, the annual LISA maps reveal a spatially unstable risk surface, with the specific districts identified as hotspots shifting from year to year. Moreover, hotspot cluster numbers ranged from 3 to 6 annually without any consistent downward trend. This temporal instability, superimposed on the geographically structured background risk captured by the BYM2 model, aligns with interannual variability in Harmattan onset, intensity, and duration— factors identified as major drivers of year-to-year fluctuation in meningitis attack rates across the belt (16,30), . A practical implication is that any surveillance strategy based on a fixed map of high-risk districts— regardless of the method used to produce it- will be partially mis-specified in any given year and should be complemented by flexible, data-driven reallocation mechanisms capable of tracking the shifting risk surface in near-real time.

4.3 Predictive Performance and Complementary Modeling Approaches

The predictive performance of our ensemble machine learning models—LightGBM, achieving $R^2 = 0.97$ on regression, and Logistic Regression and Random Forest achieving an AUC-PR of 0.41 for 2–4-week outbreak-window classification—demonstrate genuine considerably modest predictive performance, though alert thresholds could be tuned to the precision–recall tradeoff shown in Figure~2. This supports continued development of the predictive dashboard to support policy decisions in Cameroon. The dashboard is to leverage machine learning approaches for healthcare intervention decisions using defined alert thresholds that notify the surveillance unit of potential outbreaks and recommend next steps.

Figure 3: Weekly Meningitis outbreak prediction Dashboard (*See Appendix*)

It is worth noting, however, that while machine learning models excel at pattern recognition and short-term projection of known dynamics, they do not explain the causal mechanisms generating those patterns, nor do they provide credible intervals for district-level relative risks or decompose variation into spatially structured and unstructured components.

This is why our two-way modeling approaches are significantly relevant in their complementary approaches. The BYM2 model characterizes where and why risk concentrates spatially and produces credible-interval risk maps suitable for strategic resource allocation; the machine learning models generate near-real-time predictive alerts that can trigger operational preparedness responses from days to weeks before outbreak conditions formally materialize. This will offer an upgrade from relying absolutely on the current threshold-based reactive system toward a proactive, timely, prepared approach.

The lag features and rolling averages constitute the most important predictors, consistent with epidemiological dynamics; current case counts strongly depend on recent infection history through susceptible depletion and serial transmission chains (29) . Consistent with(16,30) , the seasonal Fourier terms that captured Harmattan periodicity also provide a strong predictive signal.

As recommended by(25,28,32,39), our integrated approach, combining Bayesian spatial models for risk characterization with machine learning for predictive modeling , mirrors recommendations from the broader disease modeling literature.

4.4 Policy Implications

Our findings carry several direct actionable policy implications for Cameroon and other countries within the meningitis belt:

Proactive surveillance: Embedding our predictive machine learning framework within MINSANTE's routine surveillance platform could transition the national response from threshold-triggered reaction to prospective risk stratification, enabling earlier mobilization of outbreak response resources.

Geographically targeted campaigns: The 15 persistently high-risk districts (primarily in the Far North, North, and Adamaoua regions) should receive prioritized vaccination booster campaigns, enhanced laboratory diagnostic capacity, and pre-positioned medical stockpiles during weeks 1—16 (peak transmission season).

Routine immunization integration: The growing immunity gap in post-2011 birth cohorts—estimated to potentially affect 40--50% of the population in high-transmission areas by 2030

without routine vaccination (22)— highlights the urgency of integrating MenAfriVac into Cameroon's routine Expanded Program on Immunization (EPI) (19–21,37).

Multivalent vaccine introduction: Given the epidemiological transition from serogroup A dominance toward non-A serogroups (C, W, X), the introduction of multivalent conjugate vaccines covering serogroups A, C, W, Y, and X should be prioritized (18)

4.5 Limitations and implications for future research

We recognize the limitations of the research, which may influence both the interpretation and implications of the results.

First, the surveillance dataset likely underestimates true disease burden due to passive case detection, incomplete laboratory confirmation, and reporting lags.

Second, the modeling lacked direct vaccination coverage data at the district-week level, precluding formal causal mediation analysis of vaccine impact. Third, key environmental covariates (humidity, dust concentration, and rainfall) were not integrated into the predictive models; their inclusion is expected to further improve model performance, as demonstrated by (13,16). This will enable the surveillance system to generate reliable early warnings based on climatological forecasts as well as epidemiological observations.

5. Conclusion

This study demonstrates that integrating Bayesian spatiotemporal modeling with machine learning approaches provides an evidence-based framework for meningitis surveillance in Cameroon.

The predictive value of ensemble machine learning models, LightGBM ($R^2 = 0.97$) for case-count regression and Logistic Regression and Random Forest (AUC-PR = 0.41) for 2–4 week outbreak-window classification, is genuine and considerably modest, which supports integration into surveillance systems, provided alert thresholds are calibrated to an operationally acceptable precision–recall tradeoff beyond the assumption from the models' discriminative ability alone. Future work should incorporate environmental covariates, serogroup-stratified analyses, and validation against prospective outbreak events to further strengthen the operational utility of this framework for national and regional meningitis control in alignment with the WHO's goal of defeating meningitis by 2030.

List of Abbreviations

AR Attack Rate

AUC-ROC Area Under the Receiver Operating Characteristic Curve

BYM2 Besag York Mollié (reparameterised model)

CAR Conditional Autoregressive

CFR Case Fatality Ratio/Rate

CrI Credible Interval

DIC Deviance Information Criterion

DLMEP	Direction de la Lutte contre la Maladie, les Épidémies et les Pandémies
EPI	Expanded Programme on Immunization
ICAR	Intrinsic Conditional Autoregressive
INLA	Integrated Nested Laplace Approximation
INS	Institut National de la Statistique
IRR	Incidence Rate Ratio
LISA	Local Indicator of Spatial Association
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MINSANTE	Ministère de la Santé Publique (Ministry of Public Health, Cameroon)
RF	Random Forest
RMSE	Root Mean Square Error
WAIC	Widely Applicable Information Criteria
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine

668

669 **Declarations**

670 **Ethics approval and consent to participate:** This study used secondary meningitis surveillance
671 data obtained from the Ministry of Public Health (MINSANTE), Cameroon. The analysis was
672 conducted on aggregated surveillance data and did not involve direct recruitment or interaction
673 with patients, collection of biological samples, or access to identifiable individual information.

674 Because the study relied solely on secondary analysis of routinely collected surveillance data,
675 ethical approval and individual informed consent were not required under the institutional and
676 data-governance arrangements governing the use of MINSANTE data. The study was conducted
677 in accordance with applicable ethical principles and relevant regulations for the responsible use of
678 health surveillance data

679 **Consent for publication:** Not applicable

680 **Availability of data and materials:** Aggregated district-level data and analysis code are available
681 upon request to the corresponding author, subject to MINSANTE data-sharing approval.

682 **Competing interests:** The authors declare no competing interests.

683 **Funding:** This study was supported by the Mastercard Foundation Scholars Program through the
684 African Institute for Mathematical Sciences (AIMS), Cameroon. The funders had no role in study
685 design, data collection, analysis, interpretation, or the decision to submit.

686

687 **Authors' contributions:** IBA conceptualized the study, led data preprocessing, feature
688 engineering, Bayesian modelling, machine learning implementation, and drafted the manuscript.

SW provided supervisory guidance on statistical methodology and spatial modelling. JBT reviewed the draft and guided modeling strategies. AES and BKM facilitated data access through MINSANTE/DLMEP and provided epidemiological context. HT and BT provided expert guidance on modeling frameworks. All authors reviewed, revised, and approved the final manuscript;

Acknowledgements

The authors gratefully acknowledge MINSANTE and the DLMEP data management team for access to national meningitis surveillance data, and the Mastercard Foundation Scholars Program through AIMS-Cameroon for financial support.

Authors' Affiliation

1. African Institute for Mathematical Sciences (AIMS), Limbe, Cameroon.
2. Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
3. Research Unit in Biostatistics, Data Science, Health Information Systems and Technologies (URBioDataSIT), Center for Research in Human Health and Drug Development (CRESHDEM), Faculty of Medicine and Pharmaceutical Sciences, University of Dschang, Dschang, Cameroon
4. University of Yaoundé I, Yaoundé, Cameroon, Faculty of Medicine and Biomedical Sciences
5. Direction de la Lutte contre la Maladie, les Épidémies et les Pandémies (DLMEP), Ministry of Public Health (MINSANTE), Yaoundé, Cameroon
6. Centre Pasteur du Cameroun, Service d'Épidémiologie et Santé Publique, Yaoundé, Cameroon

- 710 7. Catholic University of Central Africa, School of Health Sciences, Yaoundé, Cameroon
711 (UCAC/École des Sciences de la Santé)
- 712 8. Department of mathematics, faculty of science, University of Dschang, Dschang,
713 Cameroon
- 714 9. School of Agricultural, Earth, and Environmental Sciences, University of KwaZulu-Natal,
715 Pietermaritzburg 3209, South Africa
- 716 10. International Institute of Tropical Agriculture (IITA), PMB, 5320, Oyo Road, Idi-Oshe,
717 Ibadan, Nigeria

718

719 ***Corresponding author**

720 **Email addresses of Authors**

721 IBA*: Ishmael.ayengl@gmail.com

722 SW: whegang@gmail.com

723 BKM: kkompguepboris@gmail.com

724 AES: ancelsongo@gmail.com

725 HT: htonnang@gmail.com

726 BT: bergetsanou@gmail.com

727 JBT: tchatchueng@pasteur-yaounde.org

728

729 **References**

- 730 1. Runde TJ, Anjum F, Hafner JW. Bacterial Meningitis. In: StatPearls [Internet]. Treasure
731 Island (FL): StatPearls Publishing; 2026 [cited 2026 Aug 5]. Available from:
732 <http://www.ncbi.nlm.nih.gov/books/NBK470351/> PubMed PMID: 29261975.
- 733 2. Tunkel AR, Hartman BJ, Kaplan SL, Kaufman BA, Roos KL, Scheld WM, et al. Practice
734 guidelines for the management of bacterial meningitis. *Clin Infect Dis Off Publ Infect Dis*
735 *Soc Am.* 2004 Nov 1;39(9):1267–84. doi:10.1086/425368 PubMed PMID: 15494903.
- 736 3. van de Beek D, Brouwer MC, Koedel U, Wall EC. Community-acquired bacterial
737 meningitis. *Lancet.* 2021 Sep 25;398(10306):1171–83. doi:10.1016/S0140-6736(21)00883-7
738 PubMed PMID: 34303412.
- 739 4. World Health Organization. Meningitis [Internet]. 2025 [cited 2026 Jul 30]. Available from:
740 <https://www.who.int/news-room/fact-sheets/detail/meningitis>
- 741 5. Mohanty S, Johansson Kostenniemi U, Silfverdal SA, Salomonsson S, Iovino F, Sarpong
742 EM, et al. Increased Risk of Long-Term Disabilities Following Childhood Bacterial
743 Meningitis in Sweden. *JAMA Netw Open.* 2024 Jan 19;7(1):e2352402.
744 doi:10.1001/jamanetworkopen.2023.52402 PubMed PMID: 38241045; PubMed Central
745 PMCID: PMC10799263.
- 746 6. Edmond K, Clark A, Korczak VS, Sanderson C, Griffiths UK, Rudan I. Global and regional
747 risk of disabling sequelae from bacterial meningitis: a systematic review and meta-analysis.
748 *Lancet Infect Dis.* 2010 May;10(5):317–28. doi:10.1016/S1473-3099(10)70048-7 PubMed
749 PMID: 20417414.
- 750 7. Franklin K, Kwambana-Adams B, Lessa FC, Soeters HM, Cooper L, Coldiron ME, et al.
751 Corrigendum to: Pneumococcal Meningitis Outbreaks in Africa, 2000–2018: Systematic
752 Literature Review and Meningitis Surveillance Database Analyses. *J Infect Dis.* 2023 May
753 15;227(10):1220. doi:10.1093/infdis/jiac002
- 754 8. Greenwood B. Manson Lecture. Meningococcal meningitis in Africa. *Trans R Soc Trop Med*
755 *Hyg.* 1999;93(4):341–53. doi:10.1016/s0035-9203(99)90106-2 PubMed PMID: 10674069.
- 756 9. Villalpando-Carrión S, Henao-Martínez AF, Franco-Paredes C. Epidemiology and Clinical
757 Outcomes of Bacterial Meningitis in Children and Adults in Low- and Middle-Income
758 Countries. *Curr Trop Med Rep.* 2024 Jun;11(2):60–7. doi:10.1007/s40475-024-00316-0
759 PubMed PMID: 39006487; PubMed Central PMCID: PMC11244613.
- 760 10. Trotter CL, Lingani C, Fernandez K, Cooper LV, Bitá A, Tevi-Benissan C, et al. Impact of
761 MenAfriVac in nine countries of the African meningitis belt, 2010–15: an analysis of
762 surveillance data. *Lancet Infect Dis.* 2017 Aug;17(8):867–72. doi:10.1016/S1473-
763 3099(17)30301-8 PubMed PMID: 28545721.

- 764 11. Codjoe SNA, Nabie VA. Climate Change and Cerebrospinal Meningitis in the Ghanaian
765 Meningitis Belt. *Int J Environ Res Public Health*. 2014 Jul;11(7):6923–39.
766 doi:10.3390/ijerph110706923 PubMed PMID: 25003550; PubMed Central PMCID:
767 PMC4113853.
- 768 12. Guarnieri G, Olivieri B, Senna G, Vianello A. Relative Humidity and Its Impact on the
769 Immune System and Infections. *Int J Mol Sci*. 2023 May 29;24(11):9456.
770 doi:10.3390/ijms24119456 PubMed PMID: 37298409; PubMed Central PMCID:
771 PMC10253274.
- 772 13. Molesworth AM, Thomson MC, Connor SJ, Cresswell MP, Morse AP, Shears P, et al.
773 Where is the meningitis belt?: Defining an area at risk of epidemic meningitis in Africa.
774 *Trans R Soc Trop Med Hyg*. 2002 Aug 15;96(3):242–9. PubMed PMID: 12174770.
- 775 14. Qiu Y, Zhou Y, Chang Y, Liang X, Zhang H, Lin X, et al. The Effects of Ventilation,
776 Humidity, and Temperature on Bacterial Growth and Bacterial Genera Distribution. *Int J*
777 *Environ Res Public Health*. 2022 Jan;19(22):15345. doi:10.3390/ijerph192215345
- 778 15. Mueller JE, Gessner BD. A hypothetical explanatory model for meningococcal meningitis in
779 the African meningitis belt. *Int J Infect Dis*. 2010 Jul 1;14(7):e553–9.
780 doi:10.1016/j.ijid.2009.08.013
- 781 16. Sultan B, Labadi K, Guégan JF, Janicot S. Climate drives the meningitis epidemics onset in
782 west Africa. *PLoS Med*. 2005 Jan;2(1):e6. doi:10.1371/journal.pmed.0020006 PubMed
783 PMID: 15696216; PubMed Central PMCID: PMC545199.
- 784 17. Fernandez K, Lingani C, Aderinola OM, Goumbi K, Bicaba B, Edea ZA, et al.
785 Meningococcal Meningitis Outbreaks in the African Meningitis Belt After Meningococcal
786 Serogroup A Conjugate Vaccine Introduction, 2011–2017. *J Infect Dis*. 2019 Oct 31;220(220
787 Suppl 4):S225–32. doi:10.1093/infdis/jiz355 PubMed PMID: 31671449; PubMed Central
788 PMCID: PMC6822966.
- 789 18. Mustapha MM, Marsh JW, Harrison LH. Global epidemiology of capsular group W
790 meningococcal disease (1970–2015): Multifocal emergence and persistence of hypervirulent
791 sequence type (ST)-11 clonal complex. *Vaccine*. 2016 Mar 18;34(13):1515–23.
792 doi:10.1016/j.vaccine.2016.02.014
- 793 19. Amani A, Saidu Y, Asaah CT, Cheuyem FZL, Njoh AA, Kabir HA, et al. Immunization in
794 Cameroon: Uncovering Progress, Confronting Challenges, and Paving the Path towards
795 Achieving Sustainable Development Goal 2030 Targets. *J Vaccines Vaccin*. 2023 Apr
796 19;14(3):1–10. doi:10.35248/2157-7560.23.14.520
- 797 20. Fouda AAB, Durupt A, Bwaka AM, Benissan CT, Latt A, Lingani C, et al. Rollout of the
798 meningococcal A conjugate vaccine in the African meningitis belt countries from 2010 to
799 2023: progress, challenges, and future directions. *PAMJ Clin Med*. 2025 Aug 19;51(14).
800 doi:10.11604/pamj.suppl.2025.51.1.47497

21. Gake B, Babinne Graobe B, Abdouraman B, Satou Ngah C, Aissatou A, Gake NM, et al. Impact of MenAfriVac on Meningococcal A Meningitis in Cameroon: A Retrospective Study Using Case-by-Case-Based Surveillance Data from 2009 to 2015. *J Trop Med*. 2021 Sep 27;2021:4314892. doi:10.1155/2021/4314892 PubMed PMID: 34616456; PubMed Central PMCID: PMC8490062.
22. Karachaliou A, Conlan AJK, Preziosi MP, Trotter CL. Modeling Long-term Vaccination Strategies With MenAfriVac in the African Meningitis Belt. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2015 Nov 15;61 Suppl 5(Suppl 5):S594-600. doi:10.1093/cid/civ508 PubMed PMID: 26553693; PubMed Central PMCID: PMC4639487.
23. World Health Organization. Meningitis Outbreak Response in Sub-Saharan Africa: WHO Guideline [Internet]. Geneva: World Health Organization; 2014 [cited 2026 Aug 3]. (WHO Guidelines Approved by the Guidelines Review Committee). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK274193/> PubMed PMID: 25674656.
24. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;43(1):1–20. doi:10.1007/BF00116466
25. Moraga P. Chapter 3 Bayesian inference and INLA | Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny [Internet]. 2019 [cited 2026 Aug 6]. Available from: <https://www.paulamoraga.com/book-geospatial/sec-bayesianinference.html>
26. Paireau J, Girond F, Collard JM, Maïnassara HB, Jusot JF. Analysing spatio-temporal clustering of meningococcal meningitis outbreaks in Niger reveals opportunities for improved disease control. *PLoS Negl Trop Dis*. 2012;6(3):e1577. doi:10.1371/journal.pntd.0001577 PubMed PMID: 22448297; PubMed Central PMCID: PMC3308932.
27. Hassan A, Mustapha GU, Lawal BB, Na'uzo AM, Ismail R, Womi-Eteng Oboma E, et al. Time delays in the response to the *Neisseria meningitidis* serogroup C outbreak in Nigeria - 2017. *PloS One*. 2018;13(6):e0199257. doi:10.1371/journal.pone.0199257 PubMed PMID: 29920549; PubMed Central PMCID: PMC6007901.
28. AGIER L, STANTON M, SOGA G, DIGGLE PJ. A multi-state spatio-temporal Markov model for categorized incidence of meningitis in sub-Saharan Africa. *Epidemiol Infect*. 2013 Aug;141(8):1764–71. doi:10.1017/S0950268812001926 PubMed PMID: 22995184; PubMed Central PMCID: PMC9155280.
29. Irving TJ, Blyuss KB, Colijn C, Trotter CL. Modelling meningococcal meningitis in the African meningitis belt. *Epidemiol Infect*. 2012 May;140(5):897–905. doi:10.1017/S0950268811001385 PubMed PMID: 21781369.
30. Yaka P, Sultan B, Broutin H, Janicot S, Philippon S, Fourquet N. Relationships between climate and year-to-year variability in meningitis outbreaks: A case study in Burkina Faso and Niger. *Int J Health Geogr*. 2008 Jul 2;7:34. doi:10.1186/1476-072X-7-34 PubMed PMID: 18597686; PubMed Central PMCID: PMC2504476.

31. Stolwijk AM, Straatman H, Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis. *J Epidemiol Community Health*. 1999 Apr;53(4):235–8. doi:10.1136/jech.53.4.235 PubMed PMID: 10396550; PubMed Central PMCID: PMC1756865.
32. Morris M, Wheeler-Martin K, Simpson D, Mooney SJ, Gelman A, DiMaggio C. Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in Stan. *Spat Spatio-Temporal Epidemiol*. 2019 Nov;31:100301. doi:10.1016/j.sste.2019.100301 PubMed PMID: 31677766; PubMed Central PMCID: PMC6830524.
33. Breiman L. Random Forests. *Mach Learn*. 2001 Oct 1;45(1):5–32. doi:10.1023/A:1010933404324
34. Gouda HF, Abdallah FDM. Comparative performance of bagging and boosting ensemble models for predicting lumpy skin disease with multiclass-imbalanced data. *Sci Rep*. 2025 Nov 10;15:39373. doi:10.1038/s41598-025-23846-7 PubMed PMID: 41214187; PubMed Central PMCID: PMC12603216.
35. Kübler R. Understanding the Effect of Bagging on Variance and Bias visually. *Towards Data Science* [Internet]. 2019 Dec 21 [cited 2026 Aug 6]. Available from: <https://towardsdatascience.com/understanding-the-effect-of-bagging-on-variance-and-bias-visually-6131e6ff1385/>
36. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cited 2026 Jul 30]. p. 785–94. (KDD '16). Available from: <https://dl.acm.org/doi/10.1145/2939672.2939785> doi:10.1145/2939672.2939785
37. Viviani S. Efficacy and Effectiveness of the Meningococcal Conjugate Group A Vaccine MenAfriVac® in Preventing Recurrent Meningitis Epidemics in Sub-Saharan Africa. *Vaccines*. 2022 Apr;10(4):617. doi:10.3390/vaccines10040617
38. Lingani C, Bergeron-Caron C, Stuart JM, Fernandez K, Djingarey MH, Ronveaux O, et al. Meningococcal Meningitis Surveillance in the African Meningitis Belt, 2004–2013. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2015 Nov 9;61(Suppl 5):S410. doi:10.1093/cid/civ597 PubMed PMID: 26553668.
39. Agarwala S, Sufyan MA, Zedni MostAA, Agarwala S, Kar R, Sagor OJ, et al. Improving Monkeypox Outbreak Prediction Through Time-Series Forecasting with Machine Learning Models. In: *2024 27th International Conference on Computer and Information Technology (ICCIT)* [Internet]. 2024 [cited 2026 Aug 4]. p. 2068–73. Available from: <https://ieeexplore.ieee.org/abstract/document/11022396> doi:10.1109/ICCIT64611.2024.11022396

LIST OF FIGURES

Title: Spatiotemporal Dynamics and Predictive Modeling of Meningitis in Cameroon: Integrating Bayesian Hierarchical and Machine Learning Approaches for Public Health Surveillance

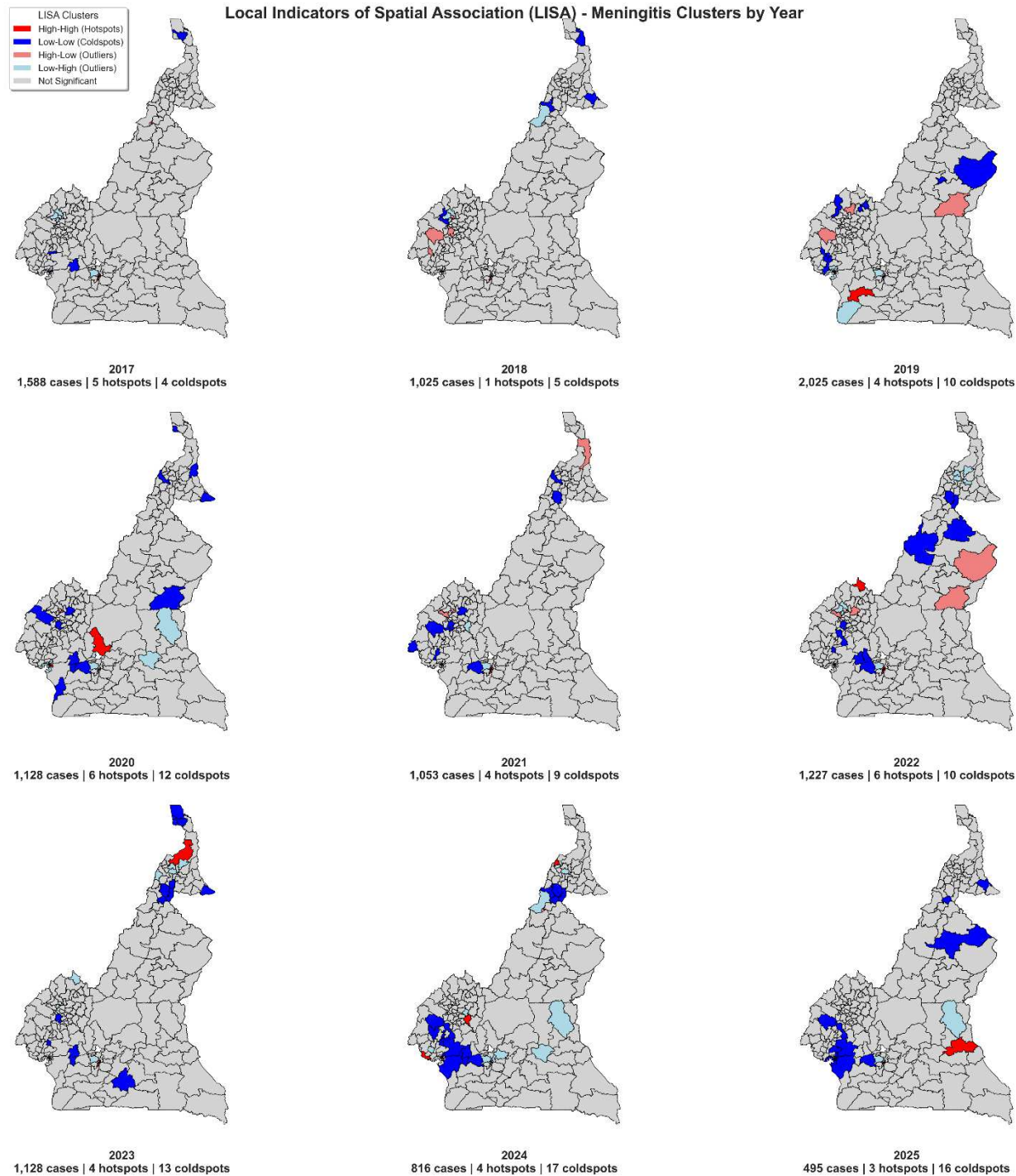


Figure 1: Local Indicators of Spatial Association (LISA): Meningitis incidence cluster by year (2017 to 2025)

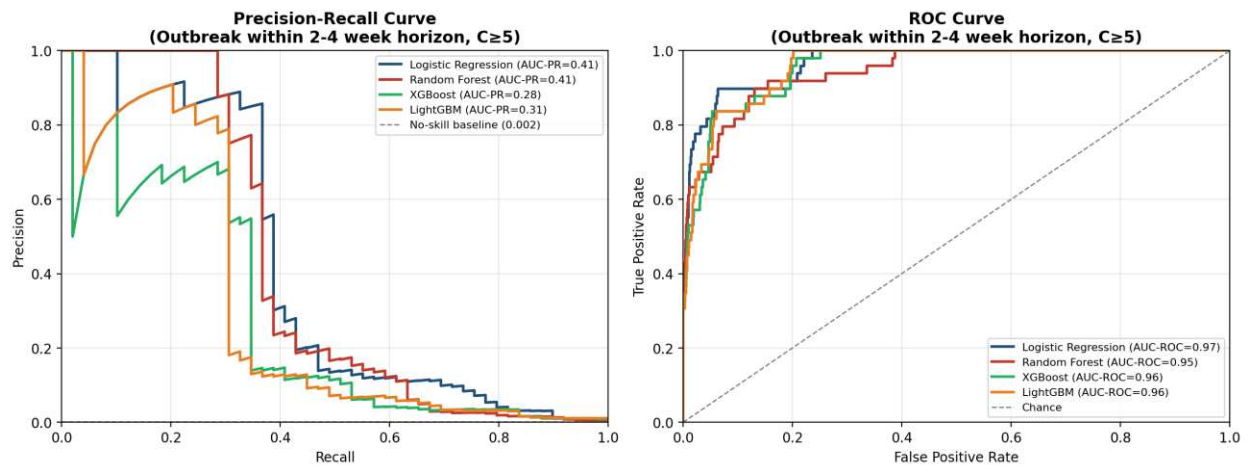


Figure 2: Precision-Recall and ROC curves for next-outbreak classification (2–4 week horizon, $C \geq 5$).

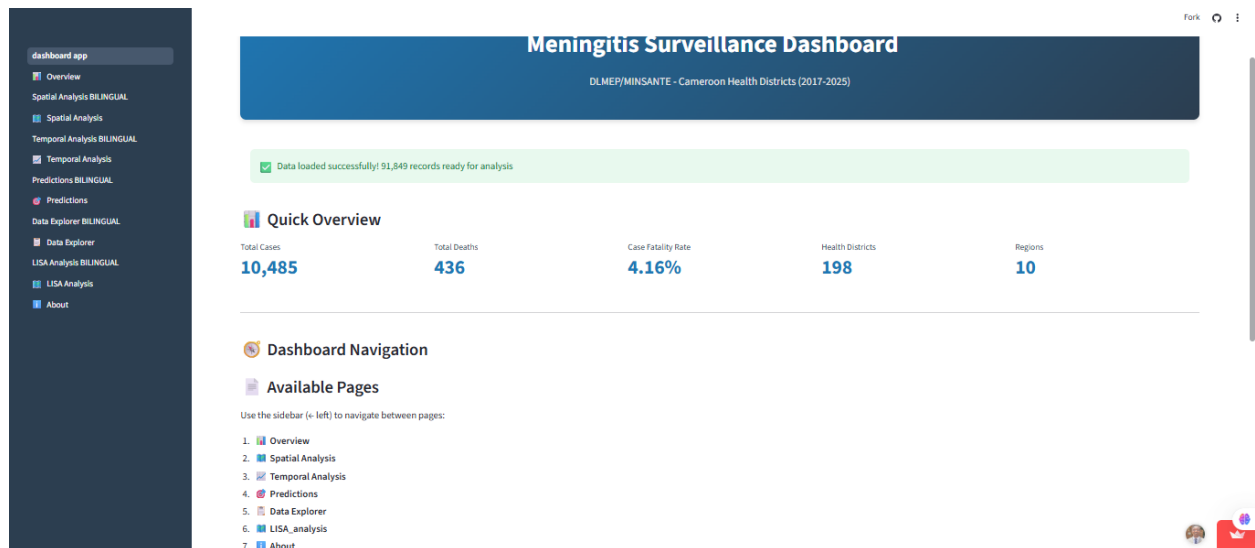


Figure 3: Weekly Meningitis outbreak prediction Dashboard

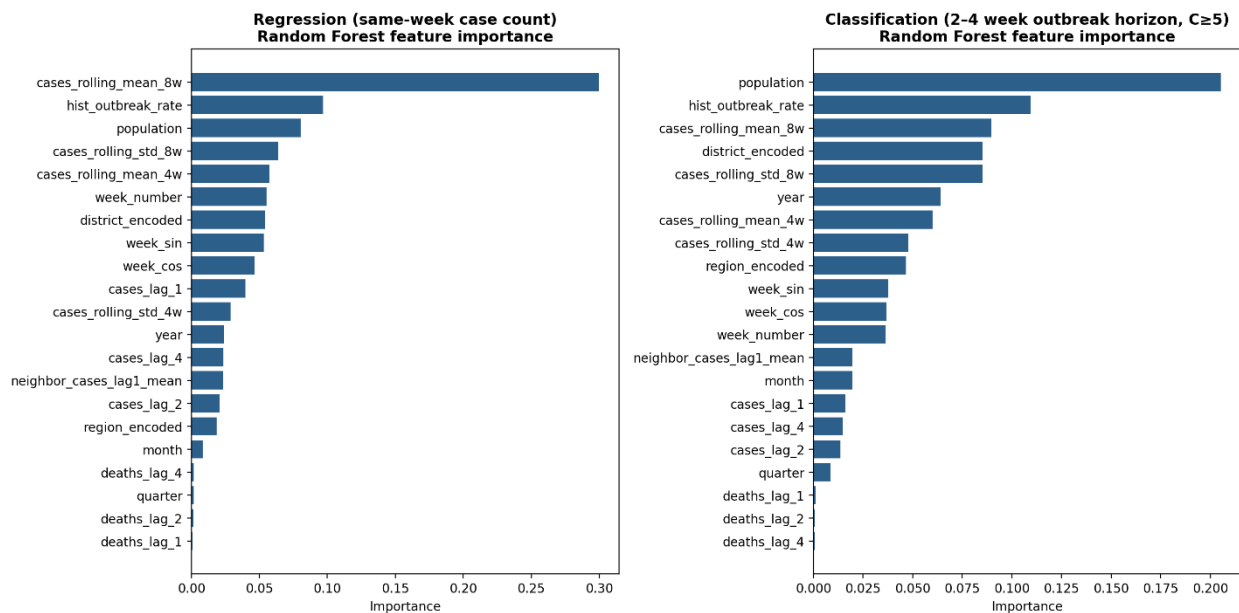


Figure 4: Feature importance plot