

RESEARCH

Open Access



Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects

Xiang Chen^{1*} and Paula Moraga¹

Abstract

Background Dengue fever is a mosquito-borne viral disease that poses significant health risks and socioeconomic challenges in Brazil, necessitating accurate forecasting across its 27 federal states. With the country's diverse climate and geographical spread, effective dengue prediction requires models that can account for both climate variations and spatial dynamics. This study addresses these needs by using Long Short-Term Memory (LSTM) neural networks enhanced with SHapley Additive exPlanations (SHAP) integrating optimal lagged climate variables and spatial influence from neighboring states.

Method An LSTM-based model was developed to forecast dengue cases across Brazil's 27 federal states, incorporating a comprehensive set of climate and spatial variables. SHAP was used to identify and select the most important lagged climate predictors. Additionally, lagged dengue cases from neighboring states were included to capture spatial dependencies. Model performance was evaluated using MAE, MAPE, and CRPS, with comparisons to baseline models.

Results The LSTM-Climate-Spatial model consistently demonstrated superior performance, effectively integrating temporal, climatic, and spatial information to capture the complex dynamics of dengue transmission. SHAP-enhanced variable selection improved accuracy by focusing on key drivers such as temperature, precipitation and humidity. The inclusion of spatial effects further strengthened forecasts in highly connected states showcasing the model's adaptability and robustness.

Conclusion This study presents a scalable and robust framework for dengue forecasting across Brazil, effectively integrating temporal, climatic, and spatial information into an LSTM-based model. The model's successful application across Brazil's diverse regions demonstrates its generalizability to other dengue-endemic areas with varying climatic and epidemiological conditions. By integrating diverse data sources, the framework captures key transmission drivers, demonstrating the potential of LSTM neural networks for robust predictions. These findings provide valuable insights to enhance public health strategies and outbreak preparedness in Brazil.

Keywords Dengue forecasting, Machine learning, Climate, Spatial, LSTM, SHAP

Introduction

Dengue fever, a mosquito-borne viral disease caused by the dengue virus (DENV), remains one of the most pressing public health issues globally, posing a major health threat to half of the world's population [1, 2]. The dengue virus is primarily transmitted to humans by the *Aedes aegypti* mosquito, which has adapted well to urban

*Correspondence:

Xiang Chen
xiang.chen@kaust.edu.sa

¹ Computer, Electrical and Mathematical Sciences and Engineering
Division, King Abdullah University of Science and Technology (KAUST),
Thuwal 23955-6900, Saudi Arabia



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

environments, making the virus easily transmissible in densely populated areas [3, 4]. The disease, endemic to more than 100 countries, has seen a dramatic surge in recent decades, with incidence rates increasing thirty-fold over the last 50 years [5]. The World Health Organization (WHO) estimates that dengue infections now number around 390 million annually, with approximately 3.9 billion people worldwide at risk of infection [1].

In Brazil, dengue is a persistent public health challenge, with outbreaks occurring regularly across different regions. This country's diverse climate and rapid urbanization create ideal conditions for mosquito proliferation, making Brazil one of the most affected countries globally [6, 7]. Despite eradication efforts targeting *Aedes aegypti* in the 1950s, the mosquito was reintroduced in the 1970s, leading to regular epidemics throughout the country [8]. The year 2024 marked a historic dengue outbreak in Brazil, with approximately 6.6 million probable cases and 6,199 deaths reported by the end of the year 2024 [9]. This unprecedented epidemic reflects an expansion of dengue into previously unaffected regions, such as Brazil's southern municipalities, which reported significant increases in cases for the first time [10].

Given the high disease burden and the geographic spread of dengue across Brazil, there is an urgent need for accurate, localized prediction models to support public health planning [11]. Current prevention strategies primarily target the mosquito vector, as there remains no effective antiviral treatment for dengue. In recent years, efforts to predict dengue outbreaks have increasingly incorporated climate and spatial data to improve prediction accuracy. However, the scope of these studies often remains limited, either focusing on specific cities or states or employing models that do not adequately capture the complex spatial-temporal dynamics at play.

Forecasting dengue cases has seen the application of diverse methodologies, blending statistical models, machine learning techniques, and the integration of external factors like climate and spatial data. Traditional models such as ARIMA and SARIMA have long been utilized for their ability to capture trends and seasonal patterns in time-series data [12–15], but they struggle with non-linear relationships and fail to adapt to sudden changes in dengue transmission driven by external environmental and socio-economic factors. To address these challenges, machine learning methods [16–20] have emerged as powerful alternatives, leveraging complex temporal dependencies to improve prediction accuracy. However, many studies applying these models lack systematic approaches for selecting relevant climate variables and often ignore spatial dependencies, limiting their ability to capture broader disease transmission dynamics.

The inclusion of climate and environmental data has been pivotal in improving the accuracy of both statistical and machine learning approaches. Variables like temperature, rainfall, and humidity, which directly influence mosquito breeding and survival, are critical predictors of dengue transmission. Both traditional statistical and machine learning models have benefited from this integration, as these climatic factors provide critical context for understanding disease spread. For instance, statistical models incorporating climate data have been shown to outperform purely case-based models in regions like Guadeloupe [21], Mexico [22], Brazil [23], and Myanmar [24], while machine learning models integrating environmental variables have delivered improved long-term predictions in Malaysia [25], and Colombia [26].

The temporal relationship between climatic factors and dengue occurrence is critical, as it affects mosquito vector dynamics and virus transmission over time, leading researchers to apply various lags when predicting dengue. Descoux et al. [27] observed that dengue outbreaks in Noumea, New Caledonia, lagged 1–2 months behind peak temperatures, aligning with maximum precipitation and humidity. Sang et al. [28] found that monthly minimum temperature with a one-month lag and cumulative precipitation with a three-month lag were effective predictors of dengue in Guangzhou, China. Gharbi et al. [21] identified minimum temperature at a lag of five weeks as a significant predictor in Guadeloupe, French West Indies. Luz et al. [12] showed that lag-0 maximum temperature and lag-1 rainfall correlated with dengue cases in Rio de Janeiro, Brazil. Wu et al. [29] reported that temperature and humidity at a two-month lag were significant predictors of dengue trends in Taiwan.

In addition to climate variables, spatial factors have been widely recognized as crucial in improving the accuracy of dengue prediction models. Studies have shown that incorporating neighboring regions' dengue case data significantly enhances model performance [30]. For instance, ensemble ARIMA models integrating spatial effects outperformed traditional ARIMA models by accounting for dengue patterns in adjacent districts, as demonstrated in Selangor, Malaysia [31]. Similarly, research in southern Vietnam revealed significant spatial correlations in epidemic timing and magnitude within districts located 50–100 km apart, emphasizing the role of local drivers like microclimatic conditions and human mobility [32]. At finer scales, cluster analyses in Thai villages identified focal dengue transmission within 100 meters of index cases, where localized factors such as mosquito density and human movement patterns played a pivotal role [33]. Moreover, temporal correlations across neighboring areas in Taiwan have been effectively utilized to predict outbreak risks [34].

These findings collectively demonstrate that incorporating spatial dependencies into predictive models provides a more comprehensive understanding of dengue spread, enabling a more comprehensive understanding of disease dynamics and enhancing the precision of predictive frameworks.

In this study, we propose an enhanced Long Short-Term Memory (LSTM) neural network model [35, 36] to predict dengue cases at the state level across Brazil. The model integrates a diverse set of lagged climate variables, including temperature, relative humidity, precipitation, atmospheric pressure, thermal range, and rainy days, to account for environmental factors influencing dengue transmission. To optimize model performance, we utilize SHapley Additive exPlanations (SHAP) [37], which identify and select the most critical variables and their lag structures, ensuring that the model focuses on the most impactful predictors.

Incorporating a spatial component, the model also leverages lagged dengue cases from neighboring states, capturing the spatial dependencies and cross-regional influences critical to understanding disease dynamics. Additionally, seasonal patterns are included to reflect the cyclic nature of dengue outbreaks. Our approach employs a moving window strategy, using fixed 7-year periods (2016–2022) to predict dengue cases for 2023, with forecasts generated for 4 weeks (1 month) and 12 weeks (3 months) ahead. This scalable framework offers robust, adaptable state-level predictions across diverse geographical and climatic contexts in Brazil (Fig. 1).

To evaluate the model's performance, we compare it against three alternative models, including simpler LSTM models and a Bayesian hierarchical model baseline. Specifically, we use a LSTM using only dengue case data; a LSTM with lagged climate variables and seasonal patterns (without neighbor spatial effects); and a Bayesian random effects model using the same covariates as our proposed model. Model performance is assessed using three metrics - Mean Absolute Error (MAE), Mean

Absolute Percentage Error (MAPE), and Continuous Ranked Probability Score (CRPS) - to comprehensively evaluate accuracy, reliability, and uncertainty in the forecasts. Unlike previous studies that focus on specific cities or regions, our model is evaluated across all 27 federal states in Brazil, demonstrating its scalability and robustness.

Methods

Study design

We develop an LSTM model for dengue forecasting in Brazil by combining SHAP-selected climate variables, spatial effects from neighboring states, cyclic patterns, and historical dengue case data. Forecasts are made at the state (federation unit) level, enabling the generation of insights for all 27 federal units, with the flexibility to adapt the approach to specific regional contexts. Focusing on medium-term forecasting horizons, our model generates forecasts for two horizons: 1 month (4 weeks) and 3 months (12 weeks) ahead. Very shorter-term forecasts, such as 1-week predictions, often lack sufficient lead time to enable effective public health interventions. By concentrating on 1- and 3-month forecasts, the study aims to provide actionable insights, equipping public health authorities with timely information to prepare for and mitigate potential outbreaks more effectively.

We implemented a moving window strategy with a fixed window size of 7 years to predict dengue cases at the state level across Brazil (Fig. 2). This approach was chosen to balance capturing long-term trends and seasonal patterns of dengue incidence while ensuring a sufficiently large dataset for robust model training. The initial training window spanned 2016-01-03 (the first epidemiological week of 2016) to 2022-12-25 (the last epidemiological week of 2022), comprising 364 weeks (52 weeks per year \times 7 years). The window was then moved forward by one week at a time to generate predictions for 2023, with the process repeated iteratively until the last epidemiological week of 2023. By continuously updating the training

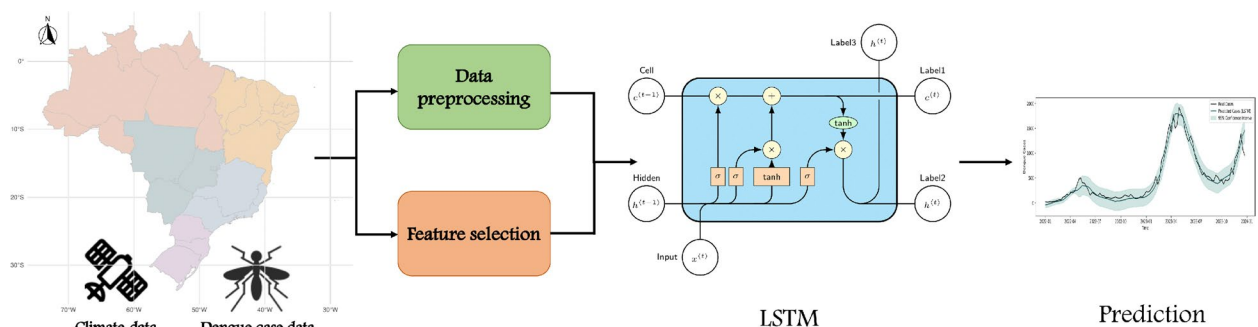


Fig. 1 Data processing pipeline for the forecasting of dengue cases in Brazil

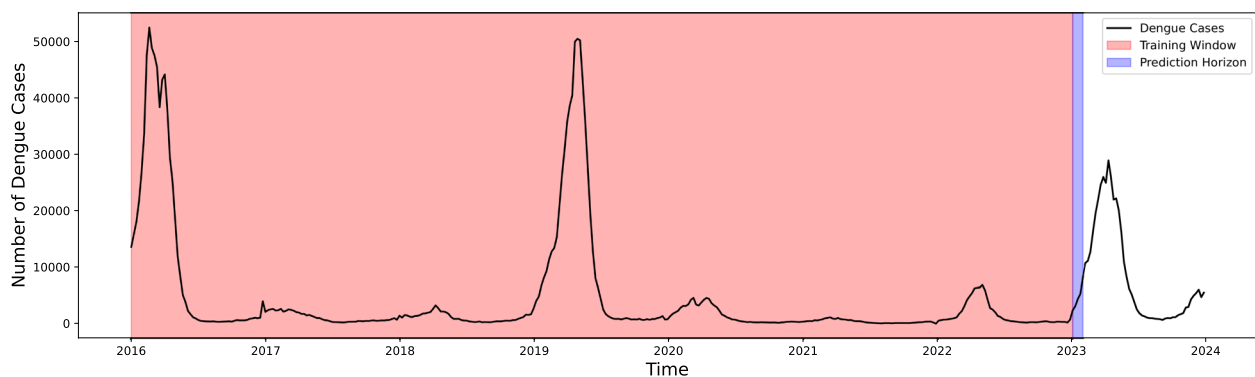


Fig. 2 Illustration of the moving window strategy for dengue case prediction

data to include the most recent observations, this strategy enables the model to adapt dynamically to evolving temporal patterns and maintain predictive accuracy.

Study site

Brazil, the largest country in South America, is geographically and ecologically diverse. It is bordered by ten countries and the Atlantic Ocean, covering an area of over 8.5 million square kilometers. Brazil's population reached an estimated 212.6 million people on July 1, 2024, according to the Brazilian Institute of Geography and Statistics (IBGE), making it the most populous country in Latin America. Over 30% of the population resides in 48 cities with more than 500,000 inhabitants, with São Paulo being the most populous municipality [38].

Brazil's diverse geography creates significant variations in climate, which directly influence dengue transmission dynamics. Figure 3 shows a state-level map of Brazil, highlighting the regional divisions. The country is divided into five macro-regions: North, Northeast, Central-West, Southeast, and South, each presenting distinct climatic, demographic, and epidemiological profiles.

The population density varies significantly across regions, as shown in Fig. 4. The Southeast region, home to São Paulo and Rio de Janeiro, is the most densely populated and economically developed area, whereas the North region, dominated by the Amazon rainforest, has sparse population density. Urbanization and human mobility patterns in densely populated regions amplify the risk of dengue transmission.

Brazil's climate varies from tropical in the North to temperate in the South, with significant regional differences captured by the Köppen climate classification [40] (Fig. 5). The North and Northeast regions predominantly experience tropical rainforest and savanna climates, characterized by high temperatures and seasonal rainfall. The Central-West region, with its tropical savanna climate, has distinct wet and dry seasons, influencing the

temporal dynamics of dengue outbreaks. The Southeast and South regions, with their subtropical and temperate climates, have historically reported fewer dengue cases, though recent outbreaks suggest climate change and urbanization are expanding transmission zones [10].

This diversity in climate and population density underscores the importance of developing dengue forecasting models tailored to Brazil's unique geographical and ecological contexts.

Dengue data

The dengue data used in this study is sourced from InfoDengue [41], a public health surveillance system designed to monitor dengue and other arboviruses across Brazil. InfoDengue integrates official dengue case reports with meteorological data, providing a comprehensive overview of the factors driving disease transmission. The system operates in 788 cities across Brazil, offering timely and up-to-date epidemiological information to health agents, local decision-makers, and the public.

The dataset provides weekly dengue case counts for each epidemiological week, allowing for high temporal resolution in forecasting efforts. This granularity ensures that seasonal patterns and short-term trends in dengue transmission are accurately captured, forming the foundation for our medium-term prediction horizons.

Figure 6 illustrates the monthly dengue incidence rate (per 100,000 people) for each federal unit from 2016 to 2023, highlighting the geographical and temporal variability of dengue across Brazil. The states are ordered by their geographical location, emphasizing the regional heterogeneity in dengue dynamics. Northern states like Amazonas and Pará show consistent transmission patterns with peaks during the rainy season, while southern states such as Santa Catarina and Rio Grande do Sul exhibit lower and more sporadic transmission. Additionally, states like Goiás, Minas Gerais, São Paulo, Ceará, and Paraná stand out with very high incidence rates (dark



Fig. 3 State-level map of Brazil highlighting regional divisions

red), emphasizing their vulnerability to severe outbreaks. These patterns underscore the regional heterogeneity of dengue dynamics and the necessity of incorporating climatic and spatial factors into forecasting models.

Climate data

The climate data used in this study is derived from the Copernicus ERA5 reanalysis dataset [42], which provides comprehensive global weather and climate information with high spatial and temporal resolution. This dataset includes hourly estimates of atmospheric, oceanic, and land-surface variables, making it an invaluable resource for understanding environmental factors influencing dengue transmission. To align with our epidemiological analysis, we considered ERA5 data on a weekly basis by utilizing weekly averages for each variable.

At the state level, climate data were aggregated using a population-weighted average to ensure that the derived values accurately reflect the conditions experienced by the majority of the population in each state.

The population data used for weighting were obtained from the Brazilian Institute of Geography and Statistics (IBGE) [43]. This approach accounts for population density variations, providing a more representative measure of climatic influences on dengue cases across diverse regions.

Table 1 shows the climate variables considered in this study, including minimum, median, and maximum values for temperature, precipitation, and atmospheric pressure, as well as thermal range, relative humidity, total precipitation, and the number of rainy days per week. These variables capture key environmental conditions linked to mosquito survival, breeding, and dengue transmission.

Climate variables are considered with a lag of 1–3 months to predict dengue cases one month (4 weeks) ahead, and a lag of 3 months for predicting three months (12 weeks) ahead. This choice aligns with findings from previous studies, and is also driven by data availability, as climate information is typically available only up to the current time, creating a natural lag when

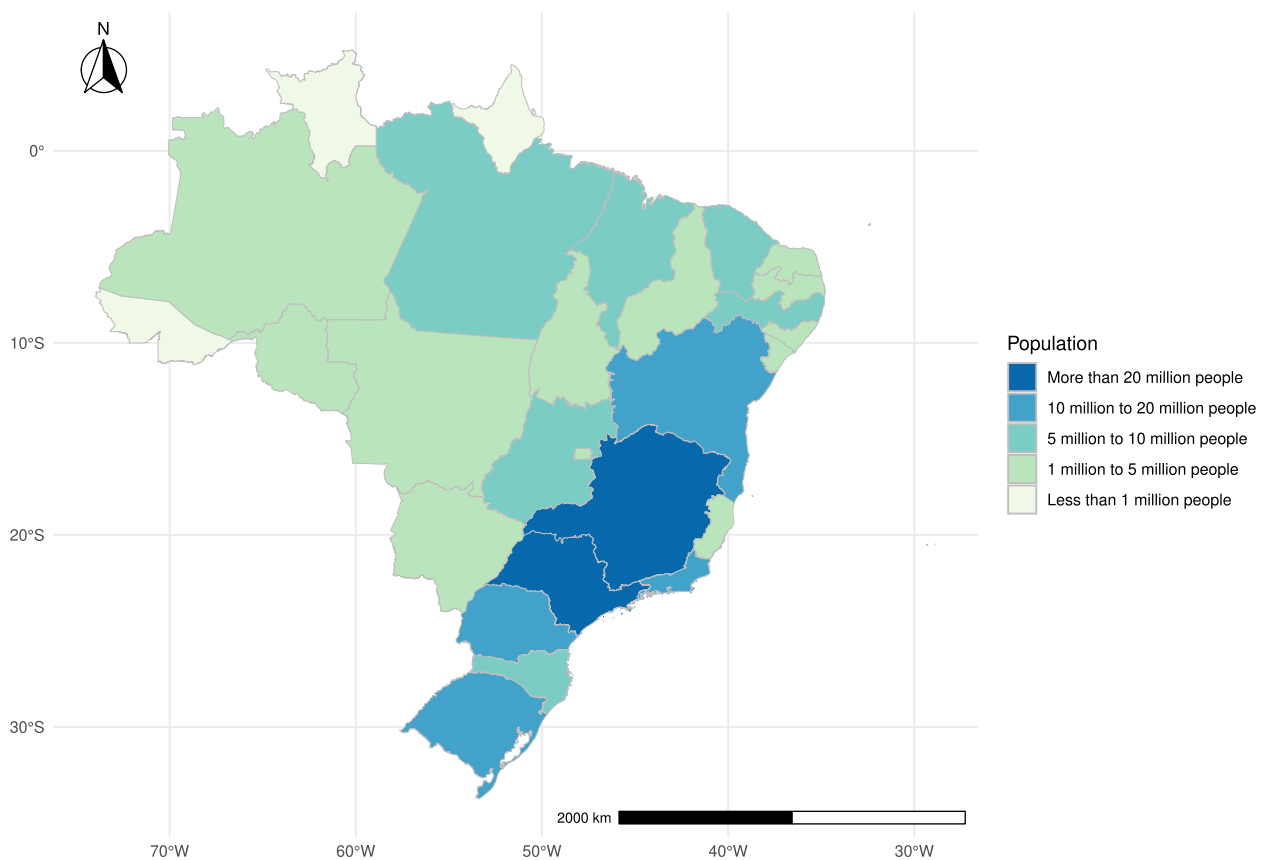


Fig. 4 Population map in each state of Brazil from the 2022 census [39]

predicting future dengue cases. The selected lags effectively capture both immediate and cumulative climatic influences on dengue transmission by accounting for the temporal relationship between environmental conditions and mosquito vector dynamics.

Feature selection with SHAP

To enhance the interpretability and efficiency of the model, we employ SHapley Additive exPlanations (SHAP) [37] for feature selection. SHAP is a game-theoretic approach that quantifies the contribution of each feature to the model's predictions, based on the concept of Shapley values from cooperative game theory. This method provides a principled and consistent way to attribute the prediction of the model to individual input features.

SHAP explains the output $f(x)$ of a model for a specific input $x = \{x_1, x_2, \dots, x_n\}$ by computing the contribution of each feature x_i . The contribution, or Shapley value, is calculated as the average marginal contribution of x_i across all possible feature subsets $S \subseteq \{x_1, x_2, \dots, x_n\}$:

$$\phi_i = \sum_{S \subseteq \{x_1, \dots, x_n\} \setminus \{x_i\}} \frac{|S|!(n - |S| - 1)!}{n!} [f(S \cup \{x_i\}) - f(S)],$$

where ϕ_i represents Shapley value for feature i ; S is a subset of features excluding x_i ; $f(S)$ is the model prediction using only features in subset S ; and n is the total number of features. It does so by evaluating how the model's prediction changes when including or excluding each feature, averaging these effects across all possible feature orderings to ensure a fair distribution of importance scores. The more a feature impacts the model's predictions, the higher its SHAP value. This approach ensures that the contributions of all features sum to the total model output, offering a fair and consistent attribution of importance.

In this study, SHAP is used to rank the importance of climate variables derived from the Copernicus ERA5 reanalysis dataset [42]. Climate variables such as temperature, precipitation, atmospheric pressure, relative humidity, thermal range, and rainy days often show high correlations within their respective groups (e.g., minimum, median, and maximum values). To ensure an optimal balance between model interpretability and

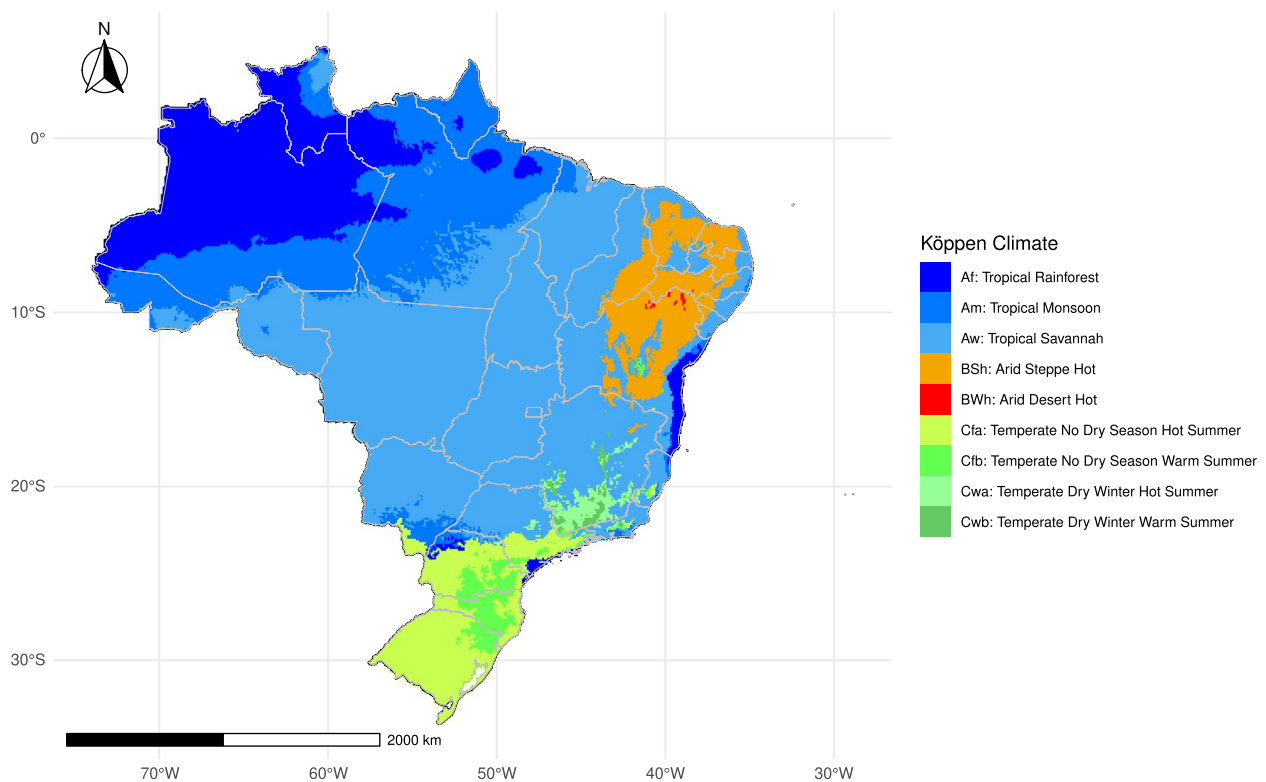


Fig. 5 Köppen climate classification map of Brazil [40]

predictive performance, we selected the top five most important climate variables based on their SHAP values. SHAP identifies the most influential variables among these correlated groups, potentially selecting more than one if their contributions to the model are equally significant. This reduces redundancy while preserving predictive power and computational efficiency. For example, SHAP may retain both median and maximum temperature if they provide complementary predictive insights.

To implement SHAP-based feature selection, an initial LSTM model is trained with all candidate climate variables. This process is carried out separately for each state to account for regional differences in dengue transmission dynamics. Rather than manually selecting climate features, we apply SHapley Additive exPlanations (SHAP) to rank and identify the most influential climate variables. SHAP quantifies the contribution of each feature to the model's predictions, ensuring a principled and consistent selection process. Shapley values are then computed for each feature, and variables with the highest average absolute Shapley values are retained. By focusing on the most impactful variables, SHAP ensures that the LSTM model captures the critical relationships between climate factors, spatial effects, and dengue transmission. This systematic approach reduces the feature space while

enhancing interpretability by clearly identifying the climate variables driving dengue transmission.

Cyclic seasonal component

Dengue transmission exhibits strong seasonal patterns, primarily influenced by climatic conditions such as temperature, precipitation, and humidity, which affect mosquito breeding and virus transmission cycles. While climatic variables are included as predictors in the model, the model also incorporates a seasonal component that captures residual cyclic trends that are not fully explained by climate data alone. These trends may arise from non-climatic factors, such as human behavior, vector control efforts, and reporting practices, which tend to repeat annually.

The seasonal component is encoded using trigonometric transformations of the epidemiological week, which effectively model the cyclical nature of dengue incidence over the course of a year. Specifically, we use $\text{Seasonality}_1 = \sin(2\pi \times \text{week}/52)$ and $\text{Seasonality}_2 = \cos(2\pi \times \text{week}/52)$, where week represents the epidemiological week within a year (i.e., 1 to 52). By incorporating both climate variables and the seasonal component, the model ensures that the cyclical structure of dengue incidence is captured comprehensively,

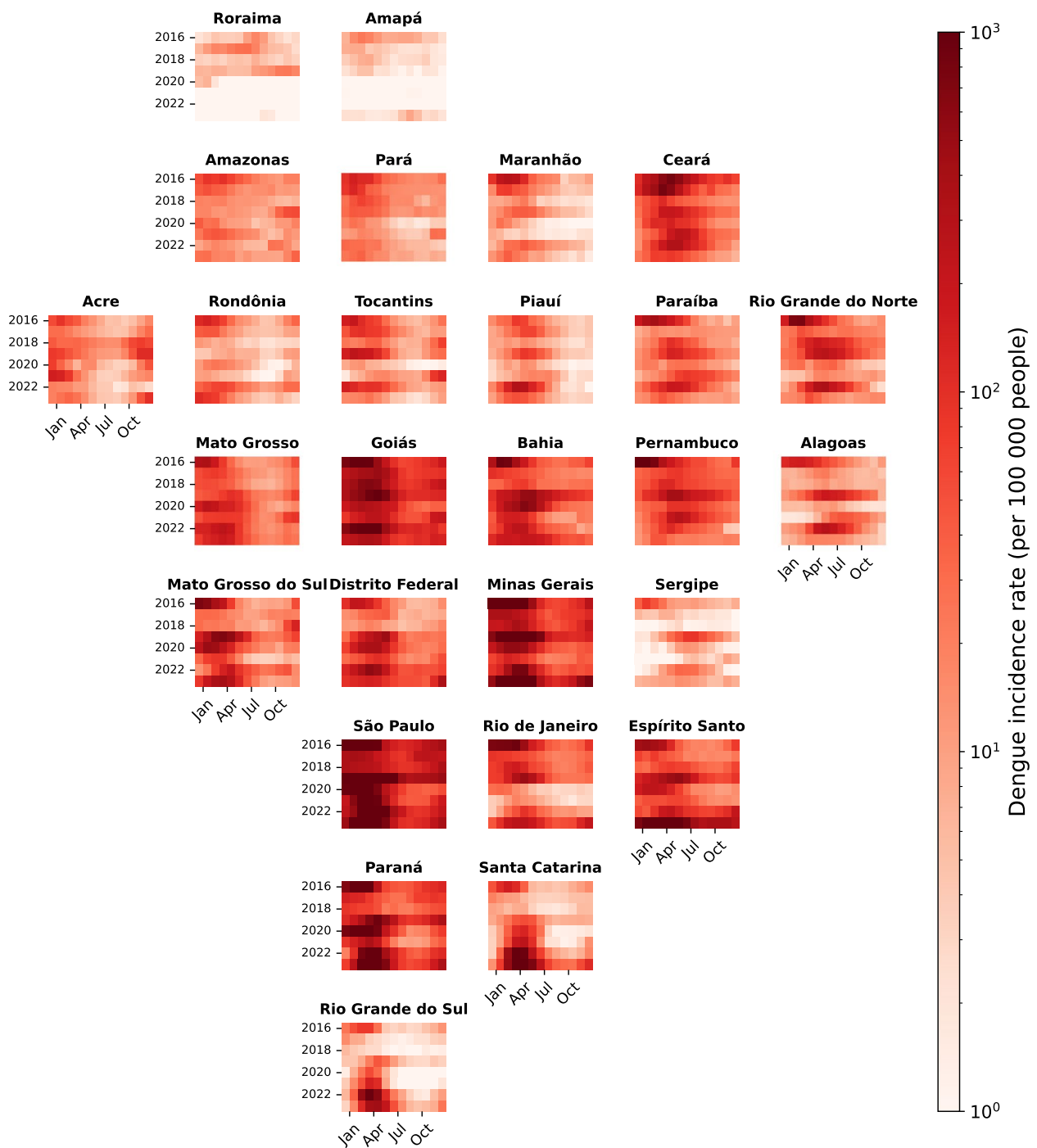


Fig. 6 Monthly dengue incidence rate (per 100 000 people) for each federal unit between 2016 and 2023. States are ordered by their geographical location

aligning peaks with favorable conditions while accounting for patterns beyond the direct influence of climate.

Neighbouring effect

Dengue transmission is influenced not only by local factors but also by spatial dependencies, as neighboring

regions often share similar environmental conditions and human mobility patterns that facilitate disease spread. Therefore, in addition to climate variables and temporal patterns, we incorporate the effect of neighboring states' dengue cases to account for spatial dependencies to improve prediction accuracy. Specifically, the model

Table 1 Climate variables derived from Copernicus ERA5 Reanalysis Data, summarized by week

Variable	Unit	Description
Minimum Temperature	°C	Lowest temperature recorded within the week.
Median Temperature	°C	Median temperature across the week.
Maximum Temperature	°C	Highest temperature recorded within the week.
Minimum Precipitation Rate	mm/h	Lowest hourly precipitation rate recorded during the week.
Median Precipitation Rate	mm/h	Median hourly precipitation rate recorded during the week.
Maximum Precipitation Rate	mm/h	Highest hourly precipitation rate recorded during the week.
Total Precipitation	mm	Cumulative precipitation over the week.
Minimum Atmospheric Pressure	atm	Lowest atmospheric pressure measured at sea level during the week.
Median Atmospheric Pressure	atm	Median atmospheric pressure measured at sea level during the week.
Maximum Atmospheric Pressure	atm	Highest atmospheric pressure measured at sea level during the week.
Minimum Relative Humidity	%	Lowest relative humidity value recorded during the week.
Median Relative Humidity	%	Median relative humidity recorded during the week.
Maximum Relative Humidity	%	Highest relative humidity recorded during the week.
Thermal Range	°C	Difference between the daily maximum and minimum temperatures.
Rainy Days	Days	Number of days within a week where the total precipitation exceeded 0.03 mm.

includes the lagged dengue case counts from each state's neighbors as predictors. Direct neighbors were chosen because dengue transmission often follows a spatially contiguous pattern, where outbreaks spread gradually from one region to adjacent areas. This approach provides a straightforward and widely applicable method for modeling spatial effects, ensuring consistency across all states, particularly in contexts where detailed human mobility data may not be readily available. By focusing on direct neighbors, the model captures the primary regional transmission dynamics, making it a reasonable and interpretable first approximation of spatial dependencies.

Table 2 lists the neighboring states for each federal unit, where neighbors are defined as states that share a common geographical border. For instance, as illustrated in Fig. 7, the state of Minas Gerais (MG) is influenced by seven neighboring states, namely, Bahia (BA), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Mato Grosso do Sul (MS), Goiás (GO), and Distrito Federal (DF). This information ensures that the spatial context of dengue transmission is effectively captured, enabling the model to reflect the interconnected dynamics across regions.

LSTM model architecture

The forecasting framework developed in this study integrates temporal, climatic, and spatial information into a Long Short-Term Memory (LSTM) neural network. The entire process, from data preprocessing to prediction, is illustrated in the pipeline shown in Fig. 1. This pipeline begins with the collection of dengue case and climate data, followed by integration of the data at the state level. Once integrated, the data undergo feature selection using SHAP, which identifies the most relevant climate

Table 2 Neighbors of each of the Brazilian states

Federal Unit (FU)	Neighbors
Acre (AC)	AM RO
Alagoas (AL)	BA PE SE
Amapá (AP)	PA
Amazonas (AM)	AC MT PA RO RR
Bahia (BA)	AL ES GO MG PE PI SE TO
Ceará (CE)	PB PE PI RN
Distrito Federal (DF)	GO MG
Espírito Santo (ES)	BA MG RJ
Goiás (GO)	BA DF MG MS MT TO
Maranhão (MA)	PA PI TO
Mato Grosso (MT)	AM RO PA TO MS GO
Mato Grosso do Sul (MS)	MG SP PR MT GO
Minas Gerais (MG)	BA ES RJ SP MS GO DF
Pará (PA)	AM RR AP TO MA MT
Paraíba (PB)	CE RN PE
Paraná (PR)	SP SC MS
Pernambuco (PE)	PI CE PB AL BA
Piauí (PI)	TO MA CE PE BA
Rio de Janeiro (RJ)	MG ES SP
Rio Grande do Norte (RN)	CE PB
Rio Grande do Sul (RS)	SC
Rondônia (RO)	AC AM MT
Roraima (RR)	AM PA
Santa Catarina (SC)	PR RS
São Paulo (SP)	MG RJ PR MS
Sergipe (SE)	AL BA
Tocantins (TO)	BA MA PI PA MT GO

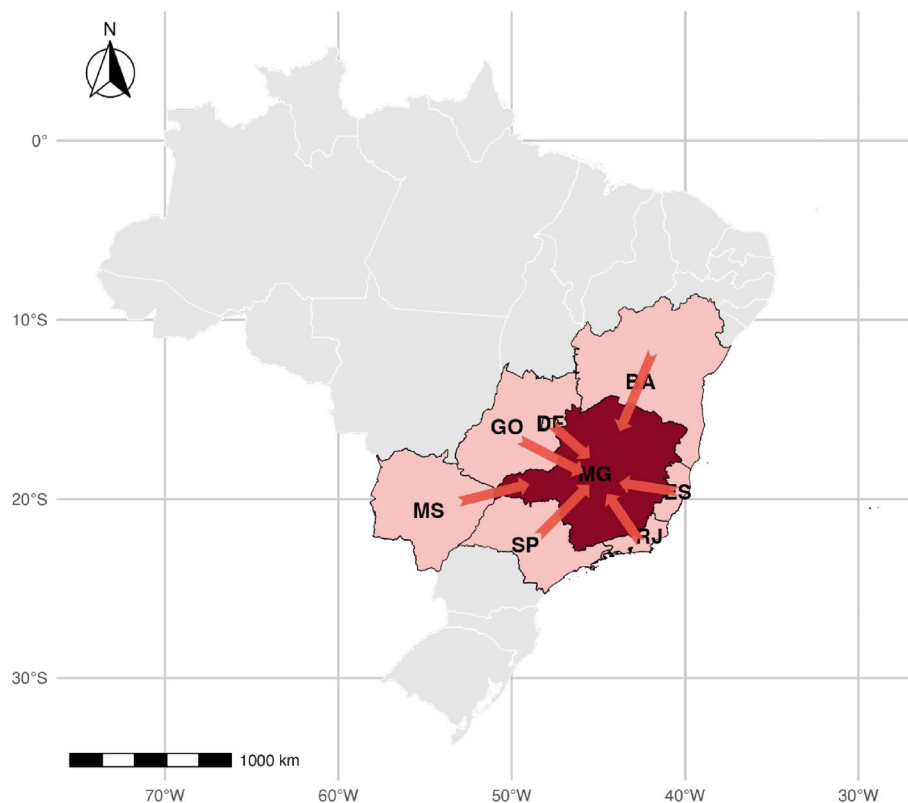


Fig. 7 Minas Gerais (MG) and its neighboring states. Arrows indicate the spatial influence of neighbors

variables for prediction. The refined dataset, including lagged climate variables, historical dengue cases, and neighboring states' dengue data, is then fed into the LSTM model, which learns complex temporal patterns and spatial dependencies to generate dengue case forecasts.

The pipeline is designed to handle the intricate relationships between dengue transmission and its influencing factors, ensuring that the model captures both the immediate and broader drivers of outbreaks. This architecture combines state-level data integration, lagged variable preparation, sequential processing, and feature selection, optimizing model interpretability and accuracy.

The LSTM was chosen for its effectiveness in capturing temporal dependencies in sequential data, such as dengue case time series. Figure 1 provides a visual overview of the LSTM architecture. The architecture consists of multiple layers designed to process both climate variables and spatial information from neighboring states described as follows.

Input layer

The input layer receives a multivariate time series, represented as:

$$X_t = \{x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(n)}\},$$

where X_t is the input at time t , and $x_t^{(i)}$ represents the i -th variable at time t . This set includes lagged dengue cases, SHAP-selected lagged climate variables, lagged dengue cases from neighboring states and cyclic seasonal patterns. Each time series input is standardized to ensure consistent scaling and convergence during training. The input dimensions are defined as (L, n) , where L represents the number of previous time steps used as input for the model, and n is the number of features. To ensure consistent scaling and convergence during training, all inputs are standardized by transforming each input feature to have a mean of 0 and a standard deviation of 1. This step ensures that all variables are on a similar scale, preventing features with larger numerical ranges from dominating the learning process.

LSTM layers

The core of the model comprises stacked LSTM layers, which allow the network to learn complex temporal dependencies in the data. These layers process sequential data by selectively updating their internal states through forget, input, and output gates:

1. Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)$$

2. Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, X_t] + b_C)$$

3. Cell State Update

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

4. Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$

5. Hidden State Update

$$h_t = o_t \odot \tanh(C_t)$$

Here, W and b represent the weight matrices and bias vectors for each gate, respectively, σ denotes the sigmoid activation function, and \odot represents element-wise multiplication. The term h_{t-1} refers to the hidden state from the previous time step $t - 1$, which encodes information from earlier inputs in the sequence. These mechanisms allow the LSTM layers to retain relevant information over long time horizons while filtering out less critical details, making them highly effective for time-series data such as dengue cases. The LSTM layers are optimized to capture non-linear relationships between dengue cases, climate covariates, and spatial effects from neighboring states.

Dense layers

Following the LSTM layers, dense (fully connected) layers are used to refine the extracted features. The output from the last LSTM cell, h_T , is passed through the dense layers:

$$y_t = \phi(W_d \cdot h_T + b_d),$$

where y_t represents the predicted dengue cases, W_d and b_d are the weights and biases of the dense layer, and ϕ is the activation function (e.g., ReLU). These layers enable the model to map the learned features to the target dengue case counts.

Output layer

The output layer is a single node that predicts the number of dengue cases for the target time step. It uses a linear activation function suitable for continuous predictions:

$$\hat{y}_{t+1} = W_{out} \cdot y_t + b_{out},$$

where \hat{y}_{t+1} is the predicted dengue case count for the next time step.

Regularization and optimization

To prevent overfitting, dropout is applied to the LSTM and dense layers during training:

$$h_t^{\text{drop}} = \text{Dropout}(h_t, p),$$

where p is the dropout rate. Early stopping is also used to halt training when validation performance ceases to improve. The model is trained using the Adam optimizer with the mean squared error (MSE) loss function:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2,$$

where N is the number of samples, \hat{y}_i is the predicted value, and y_i is the actual dengue case count.

Model implementation

Our study focuses on introducing methods for dengue forecasting at the national level across all 27 Brazilian states, emphasizing model comparison rather than hyperparameter optimization. Given that hyperparameter selection is highly dataset-dependent, we opted for a standard LSTM architecture instead of performing exhaustive tuning. This ensures consistency across all states and allows for a fair comparison between different modeling approaches without overfitting to specific regions.

For reproducibility, our LSTM model consists of a single LSTM layer with 1,000 units, followed by a dense output layer with one unit. The model uses ReLU activation, the Adam optimizer, and mean squared error (MSE) loss function. We trained the model for 2,500 epochs and scaled all input features using MinMaxScaler. The model architecture was chosen based on commonly used configurations in time-series forecasting rather than extensive tuning.

While different hyperparameter choices could yield varying results, our goal is to provide a generalizable framework rather than an optimized model for a specific region. In real-world applications, public health practitioners and researchers may further fine-tune hyperparameters to improve accuracy for specific states or outbreak conditions.

Baseline: Bayesian baseline random effects model

As a baseline for comparison, we employ a Bayesian random effects model to predict dengue cases. This

model captures temporal and spatial variability in dengue incidence by incorporating weekly counts, climate variables, cyclic seasonal patterns, and spatial effects from neighboring regions. The model is implemented using the integrated nested Laplace approximations (INLA) framework [44]. The model assumes the number of dengue cases at time t and region r , denoted as $Y_{t,r}$, follows a negative binomial distribution to handle overdispersion:

$$Y_{t,r} \sim \text{NegBin}(\theta_{t,r}, \phi_r)$$

Here, $\theta_{t,r}$ is the mean number of cases, and ϕ_r is the dispersion parameter specific to each region r . The linear predictor $\log(\theta_{t,r})$ includes terms for baseline effects, seasonal and weekly variations, climate influences, and spatial dependencies as follows:

$$\log(\theta_{t,r}) = \alpha_r + \beta_{S[t],r} + \gamma_{wS[t],r} + \sum_k \delta_k X_{k,t,r} + \eta_{N_r}$$

Here, the intercept α_r represents region-specific intercepts in region r , capturing baseline differences in dengue incidence across states. Seasonal variations, $\beta_{S[t],r}$, are included to capture region-specific seasonal patterns across epidemiological seasons $S[t]$. The terms $\gamma_{wS[t],r}$ capture weekly cyclic variations within each season using Bayesian splines. Climate variables $X_{k,t,r}$, such as temperature, precipitation, and humidity, consistent with those used in the LSTM model, are included to account for environmental influences on dengue transmission. These variables are incorporated as fixed effects to quantify their direct impact on case counts. Finally, the spatial effects η_{N_r} , account for spatial correlations among neighboring regions and are modeled using an Intrinsic Conditional Autoregressive (ICAR) structure [45]. The ICAR prior assumes that the risk in a given region is conditionally dependent on its neighboring regions, ensuring spatial smoothing and reducing extreme variability in localized estimates.

For each state, posterior predictive distributions are generated to forecast weekly dengue case counts. This baseline serves as a robust benchmark, incorporating the same covariates and spatial effects to evaluate the added value of our proposed model.

Adaptive conformal prediction

To quantify the uncertainty of dengue case forecasts, we implement an adaptive conformal prediction framework [46–49]. Conformal prediction is a robust statistical method that provides prediction intervals with a predefined confidence level, ensuring that the true value falls within the interval with a specified probability. This

approach is particularly useful in epidemiological forecasting, where understanding prediction uncertainty is critical for informed public health decision-making.

For each time t and forecast horizon h , the nonconformity score is defined as the residual at time t for h steps ahead calculated as the difference between the actual and predicted values from the time series model:

$$\text{residual}_{t,h} = \text{actual}_{t+h} - \text{predicted}_{t,h}.$$

For a window size W and time t , residuals are considered from $t - W$ to t . Thus, the set of nonconformity scores within the window at time t for a forecast horizon h is

$$\{\text{residual}_{t-W,h}, \text{residual}_{t-W+1,h}, \dots, \text{residual}_{t,h}\}.$$

The quantiles of these nonconformity scores are then computed to determine the prediction intervals. Specifically, the lower and upper bounds of the prediction interval at each time step t for forecast horizon h are given by

$$\text{lower_bound}_{t,h} = \text{predicted}_{t,h} - Q_{1-\alpha/2,h},$$

$$\text{upper_bound}_{t,h} = \text{predicted}_{t,h} + Q_{\alpha/2,h},$$

where $Q_{1-\alpha/2,h}$ and $Q_{\alpha/2,h}$ are the quantiles of the residuals within the window for the h -step ahead forecast, and α is the significance level (e.g., 0.05 for a 95% prediction interval). For the k -th quantile at forecast horizon h , the corresponding value is given by

$$Q_{k,h} = \inf \left\{ q \in \mathbb{R} : \frac{1}{W} \sum_{i=t-W}^t \mathbb{I}\{\text{nonconformity_score}_{i,h} \leq q\} \geq k \right\},$$

where \mathbb{I} is the indicator function.

The prediction interval is then adjusted dynamically at each time step based on these scores. The data is divided into training and calibration sets. The LSTM model is trained on the training set, while the calibration set is used to compute nonconformity scores. As predictions are made, the calibration set is dynamically updated to incorporate new data, allowing the prediction intervals to adapt to temporal changes in dengue transmission.

Forecasting accuracy

To evaluate the predictive performance of the proposed model, we use three key metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Continuous Ranked Probability Score (CRPS). These metrics collectively provide a comprehensive assessment of both the accuracy and the reliability of the forecasts.

Let y_i and \hat{y}_i represent, respectively, the actual and forecast number of dengue cases, and let n be the number of observations. The Mean Absolute Error (MAE)

measures the average magnitude of the errors in a set of predictions, without considering their direction. It is calculated as the average of the absolute differences between the forecasted values and the actual values:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$

The Mean Absolute Percentage Error (MAPE) expresses the accuracy as a percentage, which provides a relative measure of the errors. It is calculated as the average of the absolute percentage differences between the predicted and actual values:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|.$$

The Continuous Ranked Probability Score (CRPS) is a commonly used metric to evaluate the accuracy of probabilistic forecasts, especially when assessing both the sharpness and calibration of prediction intervals. CRPS measures the difference between the predicted cumulative distribution function and the observed values, providing a comprehensive assessment of interval reliability. CRPS is particularly useful for evaluating confidence intervals, as it quantifies how well the predicted distribution captures the

observed values. A lower CRPS indicates better-calibrated and more informative prediction intervals, while higher values reflect poorer alignment. CRPS is calculated as follows:

$$\text{CRPS}(\mathcal{N}(\mu_i, \sigma_i^2), y_i) = \sigma_i \left\{ \omega_i [\Phi(\omega_i) - 1] + 2\phi(\omega_i) - \frac{1}{\sqrt{\pi}} \right\},$$

where $\Phi(\omega_i)$ and $\phi(\omega_i)$ are the cumulative distribution function (CDF) and the probability density function (PDF) of the standard normal distribution, respectively, evaluated at the normalized prediction error $\omega_i = \frac{y_i - \mu_i}{\sigma_i}$. Additionally, y_i represents the cases observed in week i , μ_i is the predicted mean dengue case value for week i , and σ_i is the standard deviation of the forecast on week i , approximated as the length of the adaptive confidence intervals derived from the residuals between observed and predicted values.

Results

Feature importance

Correlation analysis among climate variables revealed similar patterns across states, with high correlations observed within variable groups, such as minimum, median, and maximum temperature, and among precipitation metrics. Figure 8 shows a heatmap for the state of Minas Gerais as an example. These correlations highlight

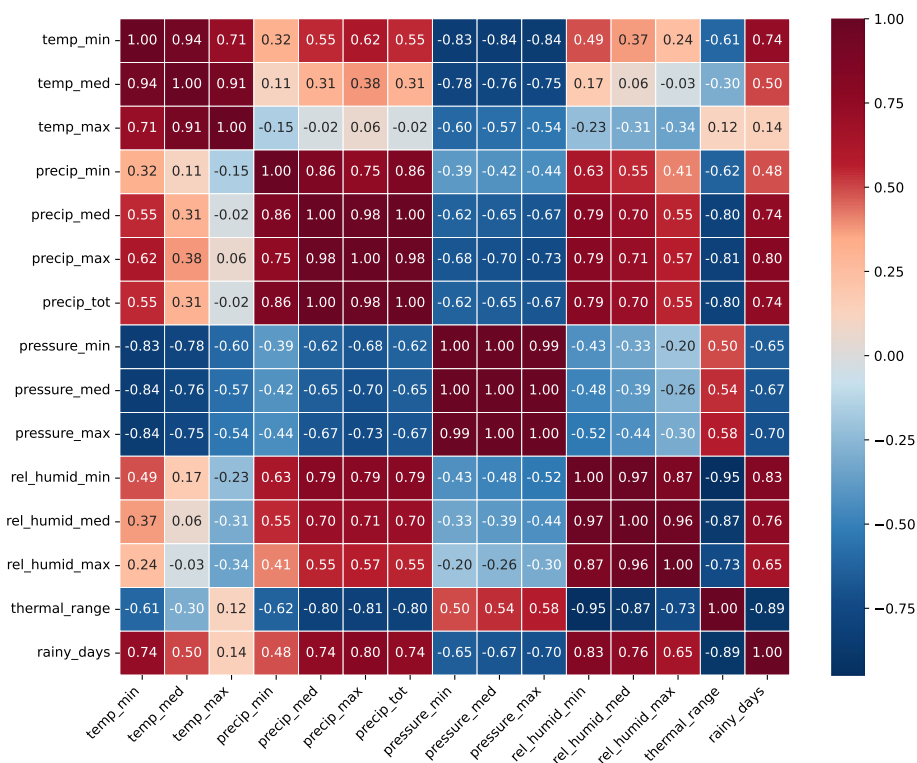


Fig. 8 Correlation heatmap of climate variables for Minas Gerais, Brazil

the potential redundancy and potential multicollinearity in climate features, as strongly correlated variables may provide overlapping information. Given these findings, we employed SHapley Additive exPlanations (SHAP) to identify and select the most relevant variables, ensuring that only the most impactful predictors were included in the model, reducing redundancy and enhancing computational efficiency. The feature importance analysis is conducted independently for all 27 federal units of Brazil, as each state has unique climatic and epidemiological characteristics, resulting in distinct sets of influential variables.

Table 3 shows the top five most important climate variables identified by SHAP for predicting dengue cases across Brazilian states. The selection of these variables is purely data-driven, ensuring that the most relevant climatic factors for each state are incorporated into the model. While certain variables, such as relative humidity, frequently appear among the top-ranked features, their dominance varies across states due to differences in local climate patterns and epidemiological conditions. For specific analysis, we choose

Minas Gerais (MG) as an example to illustrate how SHAP provides interpretability in identifying key drivers of dengue transmission. Figure 9 illustrates the SHAP summary plot for Minas Gerais as an example. Each point in the plot corresponds to a weekly prediction, colored by the magnitude of the feature values (red for high values and blue for low values). The variables are ranked by their mean absolute SHAP values, indicating their average contribution to the model's predictions for this state. For Minas Gerais, the most influential variables include maximum relative humidity, minimum precipitation rate, and minimum temperature, suggesting that humidity, precipitation and temperature are key drivers in the dynamics of dengue transmission in this state. Other states exhibit different rankings and influential variables due to their unique climatic and epidemiological contexts. This tailored feature selection process ensures that the model captures region-specific drivers of dengue transmission, improving prediction accuracy across diverse geographical and environmental conditions in Brazil.

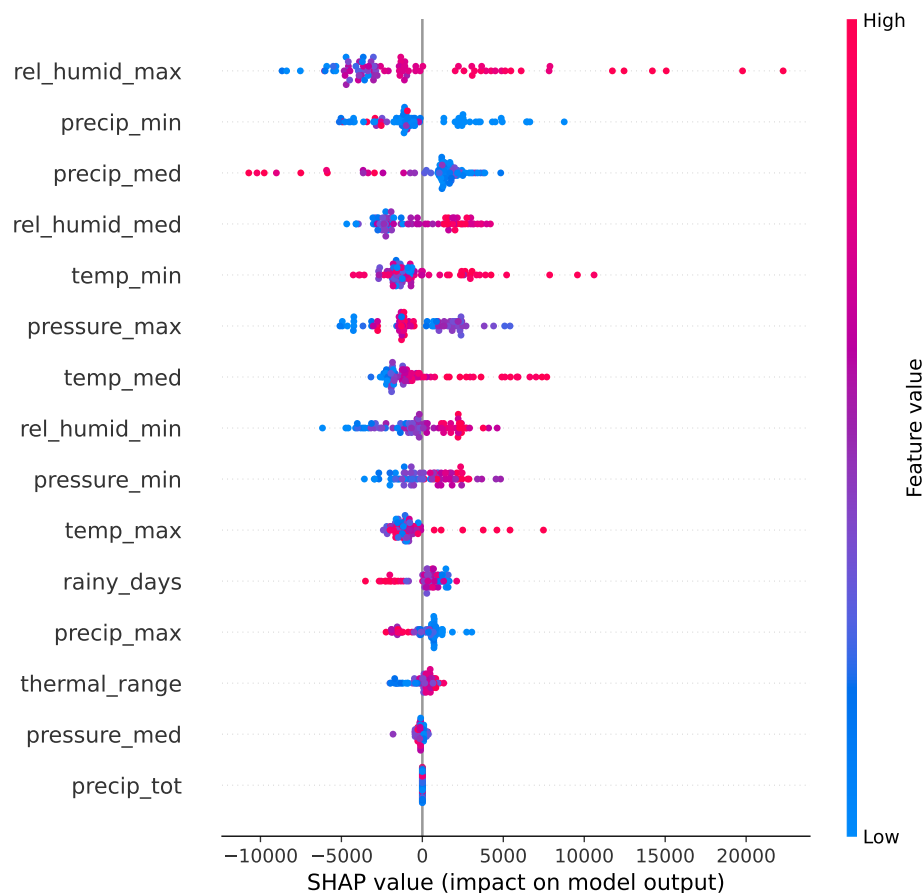


Fig. 9 SHAP feature importance for dengue case prediction in Minas Gerais, Brazil

Table 3 Top five climate variables identified by SHAP for dengue prediction across Brazilian states

Federal Unit (FU)	Code	Top five climate variables
Acre (AC)	12	pressure_max, thermal_range, pressure_min, pressure_med, rel_humid_med
Alagoas (AL)	27	rel_humid_max, rel_humid_min, precip_min, temp_med, rel_humid_med
Amapá (AP)	16	rel_humid_max, temp_min, thermal_range, rel_humid_med, precip_min
Amazonas (AM)	13	precip_med, pressure_min, rel_humid_min, temp_med, precip_min
Bahia (BA)	29	rel_humid_max, temp_max, temp_med, pressure_min, rel_humid_med
Ceará (CE)	23	rel_humid_max, temp_med, rel_humid_med, rel_humid_min, temp_min
Distrito Federal (DF)	53	rel_humid_med, precip_max, rel_humid_max, thermal_range, pressure_max
Espírito Santo (ES)	32	precip_max, rel_humid_min, precip_med, rel_humid_max, pressure_med
Goiás (GO)	52	rel_humid_max, temp_med, precip_max, temp_max, pressure_min
Maranhão(MA)	21	rel_humid_max, temp_med, precip_max, temp_max, thermal_range
Minas Gerais (MG)	51	rel_humid_max, precip_min, precip_med, temp_min, pressure_med
Mato Grosso do Sul (MS)	50	rel_humid_med, rel_humid_max, thermal_range, pressure_min, precip_max
Mato Grosso (MT)	31	rel_humid_max, precip_min, precip_med, rel_humid_med, temp_min
Pará (PA)	15	precip_min, rel_humid_min, rel_humid_max, temp_med, temp_min
Paraíba (PB)	25	thermal_range, rel_humid_min, precip_max, rel_humid_med, temp_max
Pernambuco (PE)	41	temp_med, pressure_min, pressure_max, rel_humid_max, temp_max
Piauí (PI)	26	precip_max, precip_min, rel_humid_min, rel_humid_max, thermal_range
Paraná (PR)	22	temp_med, rel_humid_min, rainy_days, pressure_min, pressure_max
Rio de Janeiro (RJ)	33	thermal_range, temp_min, rel_humid_max, temp_med, rel_humid_min
Rio Grande do Norte (RN)	24	rel_humid_med, thermal_range, rel_humid_max, temp_max, rel_humid_min
Rondônia (RO)	43	pressure_min, rel_humid_min, pressure_max, rel_humid_med, rainy_days
Roraima (RR)	11	rainy_days, temp_med, rel_humid_max, pressure_max, precip_max
Rio Grande do Sul (RS)	14	rel_humid_max, temp_med, rel_humid_med, temp_min, pressure_med
Santa Catarina (SC)	35	rel_humid_max, rainy_days, pressure_min, precip_med, temp_min
Sergipe (SE)	42	pressure_min, pressure_med, pressure_max, rainy_days, thermal_range
São Paulo (SP)	28	rel_humid_min, pressure_min, precip_med, temp_max, rel_humid_med
Tocantins (TO)	17	rel_humid_max, temp_med, precip_min, thermal_range, rel_humid_med

Model performance

In this section, we present the performance of the four models evaluated:

1. LSTM-Cases: LSTM using only dengue case data.
2. LSTM-Climate: LSTM with lagged climate variables and seasonal patterns.
3. LSTM-Climate-Spatial: LSTM with lagged climate variables, seasonal patterns, and neighboring spatial effects.
4. Bayesian Baseline: Bayesian random effects model using the same covariates as Model 3.

Tables 4 and 5 present the forecasting performance of the models for 1-month and 3-month prediction horizons, respectively. As shown in these tables, in general our proposed LSTM-Climate-Spatial model achieves superior performance across the majority of states, outperforming the other models in terms of MAE, MAPE, and CRPS. This underscores the effectiveness of combining lagged climate variables, seasonal patterns, and

spatial effects in capturing the dynamics of dengue transmission. Additionally, Figures S1 and S2 in the supplementary materials show the time series plots of the forecasts for all 27 states across the four models for the 1-month and 3-month horizons, respectively. These figures provide a detailed visualization of the models' performance over time, highlighting the differences in prediction accuracy across the states. They illustrate how well each model captures fluctuations in dengue cases, offering further validation of the quantitative results presented in the tables. By comparing observed and predicted values, these figures provide insight into the stability and adaptability of each approach across different epidemiological scenarios.

In most states, the inclusion of spatial dependencies through neighboring state dengue data improves model accuracy, particularly in states with high connectivity or where regional influences play a significant role in disease spread. For instance, in Minas Gerais (MG), the integration of spatial effects significantly reduces prediction errors. For 1-month forecasts, the LSTM-Climate-Spatial

Table 4 Forecasting performance for 1-month horizon across Brazilian states using different models

Federal Unit (FU)	Code	LSTM-Cases			LSTM-Climate			LSTM-Climate-Spatial			Bayesian Baseline		
		MAE	MAPE	CRPS	MAE	MAPE	CRPS	MAE	MAPE	CRPS	MAE	MAPE	CRPS
Acre (AC)	12	305.19	45.50%	90.91	129.76	22.30%	35.68	136.83	24.89%	37.34	382.77	47.23%	96.13
Alagoas (AL)	27	177.96	43.29%	38.14	79.24	30.54%	16.27	61.08	23.17%	12.98	69.39	24.28%	13.41
Amapá (AP)	16	51.21	47.90%	34.05	22.45	23.49%	5.35	27.45	26.98%	6.02	30.53	34.09%	7.12
Amazonas (AM)	13	188.17	41.56%	32.14	100.21	19.63%	19.23	111.60	21.64%	22.44	143.57	28.79%	31.40
Bahia (BA)	29	886.64	29.94%	165.30	639.44	23.20%	123.86	532.46	17.13%	120.50	718.63	22.84%	137.74
Ceará (CE)	23	562.67	46.52%	108.09	245.17	27.54%	52.99	187.56	15.51%	35.01	315.69	30.16%	60.26
Distrito Federal (DF)	53	1040.21	26.69%	244.60	926.73	23.24%	219.97	767.30	16.72%	211.70	997.42	24.57%	249.25
Espírito Santo (ES)	32	8431.94	30.90%	1713.56	7262.14	30.35%	1310.95	6300.78	23.06%	1308.43	6967.74	25.93%	1552.96
Goiás (GO)	52	1708.00	30.34%	310.44	1277.24	27.36%	226.00	1195.70	19.87%	222.87	1722.08	29.75%	321.36
Maranhão (MA)	21	143.59	56.31%	26.44	102.87	38.07%	18.93	59.27	23.91%	10.88	147.31	53.05%	28.14
Mato Grosso (MT)	51	657.81	34.65%	189.42	563.40	26.97%	142.56	340.72	16.69%	72.73	624.21	28.36%	125.27
Mato Grosso do Sul (MS)	50	1711.05	75.23%	342.47	568.10	59.97%	108.71	404.48	40.11%	81.94	1646.17	50.03%	344.61
Minas Gerais (MG)	31	15099.46	52.28%	3253.80	7730.47	33.33%	1648.53	5088.71	24.52%	1035.86	14220.67	40.19%	3472.85
Pará (PA)	15	178.13	35.24%	39.22	98.64	23.21%	18.35	101.22	23.95%	20.38	174.33	35.36%	42.92
Paraíba (PB)	25	190.60	29.09%	48.24	129.42	25.04%	31.52	116.80	17.72%	24.68	141.34	26.51%	29.61
Paraná (PR)	41	9083.89	78.92%	2147.66	6730.21	67.56%	941.83	4450.76	31.37%	606.80	8025.29	54.30%	1909.77
Pernambuco (PE)	26	180.75	24.52%	39.93	138.94	19.21%	32.38	116.29	15.84%	20.96	157.04	22.33%	35.10
Piauí (PI)	22	261.43	44.46%	116.97	154.99	33.94%	66.48	140.58	23.39%	30.21	143.47	29.98%	32.34
Rio de Janeiro (RJ)	33	1599.72	38.27%	332.21	737.28	27.17%	128.86	532.26	16.31%	122.86	1524.53	38.43%	271.95
Rio Grande do Norte (RN)	24	237.30	39.26%	48.96	151.58	19.37%	39.75	100.87	17.10%	18.27	214.70	28.15%	51.01
Rio Grande do Sul (RS)	43	1983.01	160.62%	505.22	979.15	78.99%	249.31	396.54	56.30%	80.54	1354.97	113.82%	326.95
Rondônia (RO)	11	371.50	79.49%	103.44	300.93	42.08%	79.71	285.61	40.93%	69.17	323.33	43.03%	104.34
Roraima (RR)	14	9.33	40.63%	6.03	6.52	43.27%	5.02	6.36	44.31%	5.93	8.66	52.55%	2.55
Santa Catarina (SC)	42	7381.10	79.16%	1765.31	1585.71	56.28%	395.28	1556.58	15.21%	294.09	3028.00	40.68%	831.25
São Paulo (SP)	35	9544.39	49.61%	2196.34	4088.67	31.97%	921.95	3068.46	17.28%	612.34	8468.53	31.64%	1961.50
Sergipe (SE)	28	54.35	20.82%	13.07	45.92	17.48%	9.43	41.52	16.50%	8.11	86.08	31.38%	18.57
Tocantins (TO)	17	124.25	49.07%	31.51	103.21	37.89%	28.46	92.55	29.12%	20.44	128.97	45.07%	27.20

Table 5 Forecasting performance for 3-month horizon across Brazilian states using different models

Federal Unit (FU)	Code	LSTM-Cases			LSTM-Climate			LSTM-Climate-Spatial			Bayesian Baseline		
		MAE	MAPE	CRPS	MAE	MAPE	CRPS	MAE	MAPE	CRPS	MAE	MAPE	CRPS
Acre (AC)	12	752.39	196.70%	144.35	360.28	96.57%	77.95	473.87	152.33%	91.31	737.51	157.52%	146.92
Alagoas (AL)	27	186.93	83.76%	31.76	119.99	67.12%	21.07	87.26	41.45%	13.22	230.37	71.56%	41.24
Amapá (AP)	16	63.52	59.09%	13.35	28.44	29.11%	5.97	37.99	44.55%	8.63	73.23	68.21%	14.43
Amazonas (AM)	13	243.53	58.67%	44.88	121.88	33.79%	23.49	153.30	42.84%	24.28	263.66	61.00%	46.45
Bahia (BA)	29	2932.49	123.10%	488.73	1822.88	84.79%	281.81	1577.97	61.53%	251.77	2245.48	94.66%	380.28
Ceará (CE)	23	1091.28	169.44%	174.46	741.25	135.17%	90.75	553.31	89.30%	75.73	892.62	104.49%	171.03
Distrito Federal (DF)	53	2154.51	67.95%	450.12	1128.11	53.25%	299.34	1107.71	42.02%	196.99	2302.85	65.83%	498.63
Espírito Santo (ES)	32	34345.03	171.74%	6034.83	19454.42	94.16%	2780.90	20177.49	104.72%	3644.69	34625.67	157.42%	6479.79
Goiás (GO)	52	3861.00	96.09%	672.20	2657.15	58.21%	361.41	1930.93	55.77%	320.78	4371.14	97.44%	933.24
Maranhão (MA)	21	383.87	182.60%	67.56	259.98	124.16%	41.58	201.98	90.98%	33.54	341.11	131.34%	65.09
Mato Grosso (MT)	51	2337.15	206.86%	418.48	1398.12	159.67%	253.62	1328.79	103.76%	192.74	2253.75	132.46%	466.69
Mato Grosso do Sul (MS)	50	5706.37	691.33%	1023.65	3584.34	559.51%	573.65	2897.14	329.40%	485.26	6107.55	376.26%	1219.84
Minas Gerais (MG)	31	52866.08	616.96%	9337.94	29911.05	389.95%	4831.86	28714.48	307.06%	4809.34	54301.97	300.83%	11290.86
Pará (PA)	15	368.34	100.94%	65.99	202.42	52.07%	27.24	213.98	73.29%	40.61	359.90	75.50%	80.77
Paraíba (PB)	25	383.65	72.85%	68.27	230.07	49.60%	40.20	228.13	36.36%	30.30	409.14	78.68%	72.17
Paraná (PR)	41	30897.03	754.90%	5645.43	17908.14	608.19%	3574.21	14474.89	365.68%	2879.87	46440.78	549.20%	10150.45
Pernambuco (PE)	26	418.48	46.85%	86.49	260.08	36.48%	50.93	247.33	22.94%	46.38	391.87	44.99%	87.64
Piauí (PI)	22	772.83	300.11%	129.42	409.20	233.42%	83.75	403.98	163.89%	62.27	512.70	147.07%	91.26
Rio de Janeiro (RJ)	33	4470.28	145.66%	731.96	3095.96	101.89%	435.09	2172.39	73.40%	354.31	5024.65	130.49%	893.55
Rio Grande do Norte (RN)	24	365.27	66.23%	72.67	253.76	45.57%	43.31	189.04	35.11%	30.73	497.00	91.80%	115.72
Rio Grande do Sul (RS)	43	5974.48	1493.76%	1117.43	3837.54	983.06%	727.93	3581.22	826.28%	527.29	8950.27	934.24%	1974.41
Rondônia (RO)	11	1018.95	372.19%	189.96	624.46	227.83%	107.96	532.71	217.42%	91.97	691.08	161.25%	141.96
Roraima (RR)	14	11.46	43.59%	3.60	6.55	34.15%	1.88	5.70	20.76%	1.59	13.92	60.99%	3.84
Santa Catarina (SC)	42	42907.90	694.70%	7933.01	28463.14	536.58%	3986.80	23381.95	418.37%	3540.09	45467.00	337.38%	10867.04
São Paulo (SP)	35	23468.90	993.41%	4361.10	15017.27	806.61%	2747.93	11338.25	570.01%	2013.90	32842.88	559.51%	6656.80
Sergipe (SE)	28	97.58	38.53%	18.57	53.17	31.26%	11.26	44.62	18.62%	7.62	191.96	78.06%	48.12
Tocantins (TO)	17	271.10	196.45%	45.61	184.62	141.02%	28.33	126.23	120.72%	20.70	436.49	213.06%	77.62

model achieves an MAE of 5088.71, MAPE of 24.52%, and CRPS of 1035.86, compared to the LSTM-Climate model's MAE of 7730.47, MAPE of 33.33%, and CRPS of 1648.53. A similar trend is observed for 3-month forecasts, where the LSTM-Climate-Spatial model demonstrates its superiority across all metrics. This makes the LSTM-Climate-Spatial model the most accurate option for forecasting in this geographically diverse state.

Similar improvements are observed in Paraná (PR) and Ceará (CE). In Paraná, the LSTM-Climate-Spatial model achieves a significant reduction in MAE (4450.76 vs. 6730.21 in the LSTM-Climate model for 1-month predictions), as well as improvements in MAPE and CRPS. In Ceará, the inclusion of spatial effects enhances the model's ability to account for regional transmission dynamics, driven by population movement and climatic factors, with substantial gains in prediction accuracy across all metrics. Goiás (GO) also presents a compelling case where the spatial effects contribute substantially to the model's success. For 1-month predictions, the LSTM-Climate-Spatial model achieves an MAE of 1195.70, MAPE of 19.87%, and CRPS of 222.87, compared to the LSTM-Climate model's MAE of 1277.24, MAPE of 27.36%, and CRPS of 226.00. This improvement reflects the considerable connectivity of Goiás with its neighboring states,

allowing the model to capture the influence of dengue incidence in surrounding regions. For 3-month forecasts, the trend persists, highlighting the adaptability of the LSTM-Climate-Spatial model to varying spatial structures.

Compared to the Bayesian baseline model, the LSTM-Climate-Spatial model consistently demonstrates better accuracy across most states. For example, for 1-month predictions, in Ceará (CE), the Bayesian baseline model achieves an MAE of 315.69, MAPE of 30.16%, and CRPS of 60.26, while the LSTM-Climate-Spatial model outperforms it with an MAE of 187.56, MAPE of 15.51%, and CRPS of 35.01. Similarly, in Paraná (PR), the Bayesian baseline model shows an MAE of 8025.29 and CRPS of 1909.77, significantly higher than the LSTM-Climate-Spatial's MAE of 4450.76 and CRPS of 606.80. This highlights the limitations of the Bayesian approach, which struggles to account for the complex temporal and spatial patterns captured effectively by the LSTM-based models.

However, not all states benefit equally from the inclusion of spatial effects. In some cases, the addition of neighboring state data leads to a slight decline in model performance. These states, highlighted in Fig. 10, are predominantly located in the northern region of Brazil, including Acre (AC), Amazonas (AM), Amapá (AP),

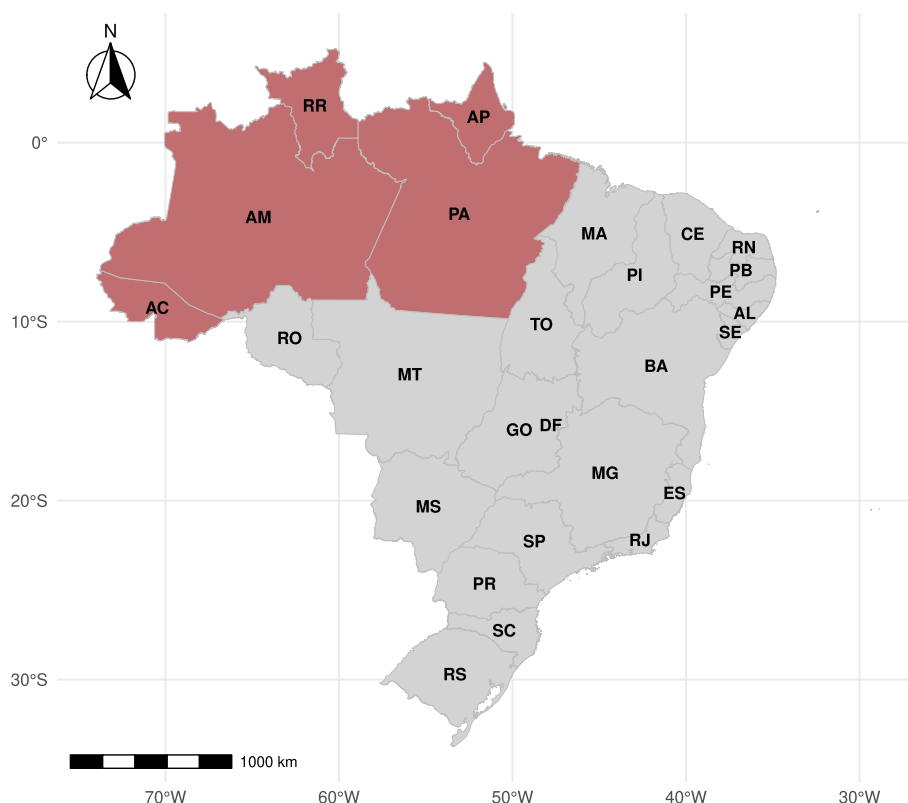


Fig. 10 Northern states in Brazil without neighbor influence for dengue forecasting marked in red

Rondônia (RO), Roraima (RR), and Pará (PA). These states are characterized by vast tropical rainforests and tropical monsoons (Fig. 5), which limit population mobility and regional connectivity compared to the more densely populated and urbanized central-west and south-east regions. The relatively sparse population distribution and lower human movement across these areas reduce the influence of spatial dependencies, thereby diminishing the utility of neighboring state dengue case data for improving forecasts.

For example, in Acre (AC), the LSTM-Climate-Spatial model shows a slightly higher MAE (136.83) compared to the LSTM-Climate model (129.76) for 1-month predictions. Similarly, in Pará (PA), the MAE increases from 98.64 in the LSTM-Climate model to 101.22 in the LSTM-Climate-Spatial model. This pattern is consistent across CRPS values as well, with AC increasing from 35.68 to 37.34 and PA increasing from 18.35 to 20.38. These results suggest that the inclusion of spatial effects introduces additional noise rather than improving the accuracy of the forecasts in these northern regions.

Despite these limitations, incorporating climate variables still significantly improves forecasting accuracy over models that rely solely on dengue case data. Even in states where spatial effects fail to enhance performance, the LSTM-Climate model consistently outperforms the LSTM-Cases model. For instance, for the 1-month forecast in Amazonas (AM), the MAE for the LSTM-Climate model is 100.21, compared to 188.17 for the LSTM-Cases model, and the CRPS equal to 19.23 improves from 32.14. Similarly, in Pará (PA), the MAE drops dramatically from 178.13 in the LSTM-Cases model to 98.64 in the LSTM-Climate model, and the CRPS improves from 39.22 to 18.35. These findings highlight the robustness of using lagged climate variables to enhance dengue forecasting, even in geographically isolated states where spatial interactions may be less relevant. By relying on climatic conditions that directly affect mosquito breeding and transmission dynamics, the LSTM-Climate model ensures a significant improvement in forecasting accuracy with respect to the model that just uses cases data, reinforcing the critical role of environmental factors in dengue prediction.

In summary, the results emphasize the effectiveness of our proposed LSTM-Climate-Spatial model, which achieves the best performance across most Brazilian states. By effectively integrating temporal patterns, lagged climate variables, and spatial dependencies, this model captures the complex dynamics of dengue transmission across diverse regions. The careful selection of climate variables, guided by SHAP, ensures that the model captures critical climatic factors driving dengue transmission. For the majority of states, the integration of spatial

effects further enhances forecasting accuracy, making the LSTM-Climate-Spatial model the most robust and reliable approach. Even in a few isolated states, such as Acre and Roraima, where spatial effects provide limited benefits, the climate component still significantly outperforms models relying solely on dengue case data. These results underscore the robustness and adaptability of the LSTM-Climate-Spatial model, reinforcing its critical role in improving dengue forecasting accuracy across diverse epidemiological and geographic contexts in Brazil, and establishing it as a reliable tool for medium-term dengue forecasting and a valuable asset for public health planning across Brazil.

Discussion

This study addresses the challenges of dengue forecasting at a national scale by developing a robust model capable of integrating climatic, temporal, and spatial information. Unlike many previous studies that focus on specific cities or states or employ models that fail to fully capture spatial-temporal dynamics, our work presents a comprehensive LSTM-based framework applied to all 27 Brazilian states. By incorporating lagged climate variables, seasonal patterns, and spatial effects from neighboring states, our approach provides a scalable and adaptable solution for forecasting dengue cases across diverse geographic and climatic contexts.

The proposed LSTM-Climate-Spatial model consistently outperforms baseline models, demonstrating its ability to capture the multifaceted dynamics driving dengue transmission. This superiority stems from several key innovations. First, the inclusion of SHAP-selected climate variables ensures that only the most relevant climatic factors are utilized, enhancing the model's predictive power. Second, the integration of spatial dependencies leverages cross-regional influences, effectively capturing transmission patterns in states such as Minas Gerais, Paraná, Ceará, and Goiás. These states, located predominantly in the Southeast and Midwest regions, are characterized by higher population densities and greater human mobility, compared to the northern states. Furthermore, the climates prevalent in these regions facilitate favorable conditions for dengue transmission, increasing the interconnectedness of dengue dynamics across state borders. These features make the model not only highly accurate but also adaptable to the heterogeneous transmission dynamics observed across Brazil.

Additionally, the use of lagged climate data allows the model to account for delayed effects of climate factors on dengue transmission, further improving its predictive reliability. The incorporation of temporal patterns, such as seasonal cycles, enables the model to adapt to recurring trends, ensuring robust performance across

short- and medium-term forecasting horizons. These elements collectively make the LSTM-Climate-Spatial model a highly effective tool for understanding and forecasting dengue, providing valuable insights for public health planning and intervention strategies. The selected lags (1–3 months for predicting dengue cases 4 weeks ahead and 3 months for predicting 12 weeks ahead) align with previous studies, and also reflect data availability since climate data is typically available only up to the current time. These lags capture both immediate and cumulative climatic effects on dengue transmission by accounting for the timing of environmental influences on mosquito dynamics. In specific public health settings, the lag duration can be appropriately adjusted based on local conditions and the availability of real-time climate data to ensure the most accurate and actionable forecasts.

Beyond Brazil, our approach is highly adaptable to other regions with similar dengue transmission dynamics. Because the model has been evaluated across all 27 Brazilian states-spanning diverse climatic, geographic, and epidemiological contexts-it demonstrates robustness and flexibility that could be extended to other dengue-endemic areas. The reliance on lagged climate variables and spatial dependencies ensures that the model remains applicable to different locations with varying environmental conditions.

Additionally, our approach of incorporating direct neighboring states' dengue cases as spatial effects ensures that the model does not rely on complex or hard-to-acquire datasets, such as high-resolution human mobility data. This makes it more broadly applicable to regions where data accessibility is a challenge. Future research could explore fine-tuning the model for specific countries by incorporating local epidemiological data, refining spatial connectivity measures, and integrating socio-economic or urbanization factors that influence dengue transmission.

While our approach shows significant promise, challenges remain. One notable limitation is the potential biases and uncertainties in the data, particularly under-reporting of dengue cases. Passive surveillance systems often suffer from inconsistent reporting across regions, with variations in healthcare access, diagnostic capacity, and case definitions affecting data reliability. Additionally, climate data from the Copernicus ERA5 reanalysis dataset, while widely used, represents gridded estimates rather than direct ground measurements, which may introduce discrepancies at finer spatial resolutions. Despite these limitations, our modeling approach aims to mitigate such biases by incorporating long-term trends, spatial dependencies, and climate-driven patterns to enhance robustness.

Another challenge is the variability in spatial effects across states, which underscores the need for a deeper exploration of region-specific factors. The influence of spatial dependencies can be shaped by diverse elements, such as human mobility patterns, socio-economic factors, vector ecology, and local epidemiology [50–52]. These factors are often highly heterogeneous and can vary not only between states but also within regions, affecting the ability of spatial models to capture dengue transmission dynamics accurately.

Our current model accounts for spatial effects by incorporating dengue cases from neighboring states, where neighboring is defined as sharing a physical border. While this approach provides a useful baseline, it is somewhat simplistic and may not fully represent the complex inter-connectivity between regions [53]. For example, human mobility does not always align with geographic borders, and highly connected urban centers or transportation hubs might influence dengue transmission across greater distances or through indirect pathways. Future work could consider more advanced methods to capture spatial effects, such as integrating mobility data, air travel patterns, or broader regional interactions beyond direct neighbors [54]. These enhancements would allow the model to better reflect the intricate spatial dynamics that influence dengue outbreaks.

Conclusion

This study presents a comprehensive and scalable framework for dengue forecasting across Brazil, integrating temporal, climatic, and spatial information into an LSTM-based model. Our results reveal that the proposed LSTM-Climate-Spatial model consistently outperforms baseline models in most states, underscoring its ability to capture complex dynamics in dengue transmission.

The incorporation of SHAP-selected climate variables proves to be a pivotal improvement, ensuring that the model focuses on the most influential predictors, which directly reflect climatic drivers of dengue outbreaks. The inclusion of spatial effects also significantly enhances model performance in states with strong inter-regional connectivity, such as Minas Gerais, Paraná, Ceará, and Goiás. This demonstrates the utility of leveraging neighboring state data to capture cross-regional transmission patterns. However, in geographically isolated states like Acre and Roraima, where inter-regional interactions are limited, spatial effects introduced noise and slightly reduced model accuracy. Nonetheless, even in these cases, the climate-enhanced models consistently outperformed case-only models, reaffirming the critical role of climate variables in dengue forecasting.

This work represents a significant step toward national-scale dengue forecasting in Brazil, addressing gaps in

previous studies that often focus on limited regions or lack the integration of spatial-temporal dynamics. Future research should explore more sophisticated methods to model spatial dependencies, such as incorporating human mobility data or broader regional interactions, to further enhance prediction accuracy. Despite these challenges, the results underscore the utility of combining climate, spatial, and temporal data for robust dengue forecasting, providing valuable insights to inform public health strategies and outbreak preparedness in Brazil.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-025-22106-7>.

Supplementary Material 1: Figure S1. Observed dengue cases and model forecasts for 1-month horizon across 27 Brazilian states. Figure S2. Observed dengue cases and model forecasts for 3-month horizon across 27 Brazilian states.

Authors' contributions

PM and XC conceived and designed the study. XC collected and processed the data, and performed the experiments. All the figures and maps are drawn by XC. XC and PM were contributors in writing and revising the manuscript. All authors read and approved the final manuscript.

Funding

This research received financial support from The Letten Prize (<https://letten-prize.com/>), with a personal award to Paula Moraga. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

All data used in this study is open access and freely available on the internet, see the "Methods" section for details. For reproducibility purposes, we provide a permanent GitHub repository with the codes, which can be found at <https://github.com/ChenXiang1998/LSTM-Based-Dengue-Prediction-Across-Brazil>.

Declarations

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 January 2025 Accepted: 26 February 2025

Published online: 12 March 2025

References

- World Health Organization. Dengue and Severe Dengue. 2023. <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>. Accessed 2 Nov 2023.
- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504–7.
- Kraemer MU, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, et al. The global compendium of *Aedes aegypti* and *Aedes albopictus* occurrence. *Sci Data*. 2015;2(1):1–8.
- Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol*. 2014;22(3):138–46.
- Guzman MG, Harris E. Dengue Lancet. 2015;385(9966):453–65.
- Paz-Bailey G, Adams LE, Deen J, Anderson KB, Katzelnick LC. Dengue Lancet. 2024;403(10427):667–82.
- Gurgel-Gonçalves R, Oliveira WKd, Croda J. The greatest Dengue epidemic in Brazil: Surveillance, Prevention, and Control. *Rev Soc Bras Med Trop*. 2024;57:e00203–2024.
- Braga IA, Valle D. *Aedes aegypti*: histórico do controle no Brasil. *Epidemiol Serviços Saúde*. 2007;16(2):113–8.
- Brazil. Atualização de Casos de Arboviroses. 2024. <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/aedes-aegypti/monitoramento-das-arboviroses>. Accessed 1 Dec 2024.
- Souza CDFd, Nascimento RPDs, Bezerra-Santos M, Armstrong AdC, Gomes OV, Nicácio JM, et al. Space-time dynamics of the dengue epidemic in Brazil, 2024: an insight for decision making. *BMC Infect Dis*. 2024;24(1):1056.
- Paixão ES, Teixeira MG, Rodrigues LC. Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases. *BMJ Glob Health*. 2018;3(Suppl 1):e000530.
- Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP, et al. Time series analysis of dengue incidence in Rio de Janeiro. Brazil: American Society of Tropical Medicine and Hygiene; 2008.
- Cortes F, Martelli CMT, de Alencar Ximenes RA, Montarroyos UR, Junior JBS, Cruz OG, et al. Time series analysis of dengue surveillance data in two Brazilian cities. *Acta Trop*. 2018;182:190–7.
- Silawan T, Singhasivanon P, Kaewkungwal J, Nimmanitya S, Suwonkerd W, et al. Temporal patterns and forecast of dengue infection in Northeastern Thailand. *Southeast Asian J Trop Med Public Health*. 2008;39(1):90.
- Buczak AL, Baugher B, Moniz LJ, Bagley T, Babin SM, Guven E. Ensemble method for dengue prediction. *PLoS ONE*. 2018;13(1):e0189988.
- Chen X, Moraga P. Assessing dengue forecasting methods: A comparative study of statistical models and machine learning techniques in Rio de Janeiro, Brazil. *medRxiv*. 2024:2024–06.
- Roster K, Connaughton C, Rodrigues FA. Machine-learning-based forecasting of dengue fever in Brazilian cities using epidemiologic and meteorological variables. *Am J Epidemiol*. 2022;191(10):1803–12.
- Kakarla SG, Kondeti PK, Vavilala HP, Boddada GSB, Mopuri R, Kumaraswamy S, et al. Weather integrated multiple machine learning models for prediction of dengue prevalence in India. *Int J Biometeorol*. 2023;67(2):285–97.
- Zhao N, Charland K, Carabali M, Nsoesie EO, Maheu-Giroux M, Rees E, et al. Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. *PLoS Negl Trop Dis*. 2020;14(9):1–16. <https://doi.org/10.1371/journal.pntd.0008056>.
- Majeed MA, Shafri HZM, Zulkafli Z, Wayayok A. A deep learning approach for dengue fever prediction in Malaysia using LSTM with spatial attention. *Int J Environ Res Public Health*. 2023;20(5):4130. <https://doi.org/10.3390/ijerph20054130>.
- Gharbi M, Quenel P, Gustave J, Cassadou S, Ruche GL, Girdary L, et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC Infect Dis*. 2011;11:1–13.
- Johansson MA, Reich NG, Hota A, Brownstein JS, Santillana M. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. *Sci Rep*. 2016;6(1):33707.
- Pavani J, Bastos LS, Moraga P. Joint spatial modeling of the risks of co-circulating mosquito-borne diseases in Ceará. Brazil *Spat Spatio-Temporal Epidemiol*. 2023;47:100616.
- Zaw W, Lin Z, Ko KJ, Rotejanaprasert C, Pantanilla N, Ebener S, et al. Dengue in Myanmar: Spatiotemporal epidemiology, association with climate and short-term prediction. *PLoS Negl Trop Dis*. 2023;17(6):e0011331.
- Salim NAM, Wah YB, Reeves C, Smith M, Yaacob WFW, Mudin RN, et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Sci Rep*. 2021;11(1):939.
- Zhao X, Li K, Ang CKE, Cheong KH. A deep learning based hybrid architecture for weekly dengue incidences forecasting. *Chaos, Solitons Fractals*. 2023;168:113170.

27. Descloux E, Mangeas M, Menkes CE, Lengaigne M, Leroy A, Tehei T, et al. Climate-based models for understanding and forecasting dengue epidemics. *PLoS Negl Trop Dis*. 2012;6(2):e1470.
28. Sang S, Gu S, Bi P, Yang W, Yang Z, Xu L, et al. Predicting unprecedented dengue outbreak using imported cases and climatic factors in Guangzhou, 2014. *PLoS Negl Trop Dis*. 2015;9(5):e0003808.
29. Wu PC, Guo HR, Lung SC, Lin CY, Su HJ. Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Trop*. 2007;103(1):50–7.
30. Moraga P. Geospatial health data: Modeling and visualization with R-INLA and Shiny. *Biostatistics series*. Boca Raton: Chapman & Hall/CRC; 2019.
31. Thiruchelvam L, Dass SC, Asirvadam VS, Daud H, Gill BS. Determine neighboring region spatial effect on dengue cases using ensemble ARIMA models. *Sci Rep*. 2021;11(1):5873.
32. Cuong HQ, Vu NT, Cazelles B, Boni MF, Thai KT, Rabaa MA, et al. Spatiotemporal dynamics of dengue epidemics, southern Vietnam. *Emerg Infect Dis*. 2013;19(6):945.
33. Mammen MP Jr, Pimgate C, Koenraadt CJM, Rothman AL, Aldstadt J, Nisalak A, et al. Spatial and temporal clustering of dengue virus transmission in Thai villages. *PLoS Med*. 2008;5(11):e205.
34. Lai WT, Chen CH, Hung H, Chen RB, Shete S, Wu CC. Recognizing spatial and temporal clustering patterns of dengue outbreaks in Taiwan. *BMC Infect Dis*. 2018;18:1–11.
35. Hochreiter S. Long Short-term Memory. *Neural Computation* MIT-Press; 1997.
36. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. *IEEE Trans Neural Netw Learn Syst*. 2016;28(10):2222–32.
37. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems* 30. Curran Associates, Inc.; 2017. p. 4765–74. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>. Accessed 01 Jan 2025.
38. Instituto Brasileiro de Geografia e Estatística (IBGE). Brazil's population reaches 212.6 million. 2025. <https://agenciadenoticias.ibge.gov.br/en/agencia-news/2184-news-agency/news/41122-populacao-estimada-do-pais-chega-a-212-6-milhoes-de-habitantes-em-2025>. Accessed 3 Dec 2024.
39. IBGE. Panorama - Censo 2022. 2022. <https://censo2022.ibge.gov.br/panorama/>. Accessed 10 Dec 2024.
40. Beck HE, McVicar TR, Vergopolan N, Berg A, Lutsko NJ, Dufour A, et al. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Sci Data*. 2023;10(1):724.
41. Codeco C, Coelho F, Cruz O, Oliveira S, Castro T, Bastos L. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. *Rev Epidemiol Sante Publique*. 2018;66:S386.
42. (C3S) CCCS. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. 2017. <https://doi.org/10.24381/cds.adbb2d47>.
43. Brazilian Institute of Geography and Statistics (IBGE). Population | IBGE. 2022. <https://www.ibge.gov.br/en/statistics/social/population.html>. Accessed 3 Dec 2024.
44. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol*. 2009;71(2):319–92.
45. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;43:1–20.
46. Vovk V, Gammerman A, Shafer G. *Algorithmic learning in a random world*, vol. 29. Springer; 2005.
47. Balasubramanian V, Ho SS, Vovk V. Conformal prediction for reliable machine learning: theory, adaptations and applications. *Newnes*; 2014.
48. Gibbs I, Candes E. Adaptive conformal inference under distribution shift. *Adv Neural Inf Process Syst*. 2021;34:1660–72.
49. Zaffran M, Féron O, Goude Y, Josse J, Dieuleveut A. Adaptive conformal predictions for time series. In: *International Conference on Machine Learning*. PMLR; 2022. pp. 25834–25866.
50. Moraga P, Dorigatti I, Kamvar ZN, Piatkowski P, Toikkanen SE, Nagraj V, et al. epiflows: an R package for risk assessment of travel-related spread of disease. *F1000Research*. 2019;7:1374.
51. Kraemer MU, Golding N, Bisanzio D, Bhatt S, Pigott DM, Ray S, et al. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Sci Rep*. 2019;9(1):5151.
52. Oliveira JF, Alencar AL, Cunha MCL, Vasconcelos AO, Cunha GG, Miranda RB, et al. Human mobility patterns in Brazil to inform sampling sites for early pathogen detection and routes of spread: a network modelling and validation study. *Lancet Digit Health*. 2024;6(8):e570–9.
53. Moraga P. *Spatial statistics for data science: theory and practice with R*. Data Science series. Boca Raton: Chapman & Hall/CRC; 2023.
54. Chen X, Moraga P. Dengue forecasting and outbreak detection in Brazil using LSTM: integrating human mobility and climate factors. *medRxiv*. 2025:2025–03. <https://doi.org/10.1101/2025.03.02.25323168>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.