

DINÂMICA DA DENGUE INFLUENCIADA PELO CLIMA EM UM AMBIENTE URBANO TROPICAL: UMA ANÁLISE COMPARATIVA ENTRE A REGRESSÃO BINOMIAL NEGATIVA E O MODELO RANDOM FOREST

José Wanderson Ferreira de Freitas¹, Romário de Sousa Almeida^{1,2}, Ítalo Gabriel da Silva¹, Elizeu Silva de Medeiros¹, Morramed da Silva Souto¹, Arthur Victor Ramos Vidal Negreiros¹

¹Universidade Estadual da Paraíba, Campina Grande, Brasil
(jose.wanderson.freitas@aluno.uepb.edu.br)

²Universidade Federal de Lavras - UFLA, , Brasil

Resumo:

A dengue é uma arbovirose de grande impacto em regiões tropicais, cuja dinâmica está fortemente associada às condições climáticas. Este estudo teve como objetivo modelar e prever a incidência de dengue no município do Recife utilizando variáveis meteorológicas, comparando o desempenho da Regressão Binomial Negativa e do algoritmo Random Forest. Foram analisados dados mensais de casos de dengue entre 2018 e 2021, juntamente com variáveis climáticas (precipitação, temperatura, umidade relativa e radiação solar). Os modelos foram avaliados por meio das métricas RMSE, MAE, MAPE e coeficiente de determinação (R^2). A Regressão Binomial Negativa apresentou melhor desempenho preditivo, explicando 82,5% da variabilidade dos dados ($R^2 = 0,825$), enquanto o Random Forest apresentou maior capacidade de capturar relações não lineares. A análise evidenciou que a combinação de temperatura elevada e precipitação moderada favorece o aumento dos casos, enquanto eventos pluviométricos extremos podem reduzir a incidência, sugerindo a ocorrência do efeito *washout* (lavagem). Além disso, foi identificada defasagem temporal de 1 a 2 meses entre eventos climáticos e notificações de casos. Os resultados destacam o potencial do uso de variáveis climáticas como ferramentas de alerta precoce, contribuindo para o aprimoramento de estratégias de vigilância epidemiológica e controle da dengue.

Palavras-chave: Arboviroses; Modelagem epidemiológica; Variabilidade climática; Saúde pública; Previsão de surtos.

INTRODUÇÃO

A dengue é uma doença viral sistêmica de grande impacto global, sendo classificada como uma das principais arboviroses, caracterizadas por enfermidades causadas por vírus mantidos e transmitidos por artrópodes hematófagos (SANABANI, 2025).

O vírus da dengue (DENV), pertencente à família *Flaviviridae*, é transmitido primariamente pelas fêmeas infectadas de *Aedes aegypti* e, secundariamente, por *Aedes albopictus* (BUCZAK *et al.*, 2018). Essa patologia apresenta elevada carga de morbidade, sendo endemicamente mais expressiva em países em desenvolvimento localizados em regiões de clima tropical e subtropical (ZHOU *et al.*, 2025).

Nesses locais, a rápida urbanização, aliada às deficiências na infraestrutura de saneamento básico e às mudanças climáticas, cria macroambientes urbanos altamente favoráveis à fixação e adaptação desses vetores, tornando o controle epidemiológico um desafio contínuo para a saúde pública (BAGCCHI, 2023)

No contexto epidemiológico brasileiro, o estado de Pernambuco destaca-se historicamente por apresentar índices elevados e recorrentes de focos do vetor, resultando em epidemias cíclicas de dengue. Esse aumento expressivo na incidência de casos não ocorre de forma aleatória, mas está frequentemente e fortemente associado à variabilidade de fatores climáticos sazonais locais. A elevação das temperaturas médias diárias e a intensificação dos regimes pluviométricos atuam de forma sinérgica,



favorecendo diretamente a proliferação em larga escala do mosquito *Aedes aegypti* (GOMES *et al.*, 2024).

Essa dinâmica climática, caracterizada por períodos úmidos bem definidos, não apenas multiplica os habitats aquáticos, mas também prolonga a sobrevivência dos mosquitos adultos, contribuindo substancialmente para a ocorrência de surtos explosivos da doença nas zonas urbanas intensamente povoadas (SILVA *et al.*, 2025).

O clima exerce, portanto, papel ecológico determinante na dinâmica de transmissão da dengue, configurando-se como uma das variáveis preditivas mais relevantes. Temperaturas elevadas aceleram o metabolismo do vetor, encurtam o ciclo de desenvolvimento das fases imaturas e reduzem o Período de Incubação Extrínseco (PIE), aumentando a eficiência da transmissão viral. Paralelamente, a elevada pluviosidade amplia a disponibilidade de criadouros com água estagnada, essenciais para o desenvolvimento das fases iniciais do mosquito (VÁSQUEZ *et al.*, 2023).

Como consequência, há um aumento exponencial tanto na densidade populacional do vetor quanto na sua capacidade de transmissão do vírus para humanos suscetíveis, exigindo ferramentas estatísticas rigorosas para modelar esse comportamento (LIMA *et al.*, 2024).

Além disso, as variáveis climáticas influenciam diretamente todas as etapas do ciclo de vida do *Aedes aegypti*, impactando a incidência da dengue (SILVA *et al.*, 2024). O aumento da temperatura acelera o metabolismo do inseto e reduz o tempo necessário para que o vírus se torne infectante no mosquito (Período de Incubação Extrínseco), enquanto encurta o ciclo das fases imaturas, resultando em maior taxa de renovação populacional (MAGALHÃES *et al.*, 2020).

A combinação de alta umidade e pluviosidade elevada também contribui para a manutenção de criadouros ativos e aumento da longevidade das fêmeas, ampliando a janela de transmissão viral (SANABANI, 2025).

Do ponto de vista metodológico, a modelagem da incidência de dengue tem sido amplamente realizada por meio de abordagens estatísticas. Considerando que os dados epidemiológicos são, em geral, discretos e não negativos, os Modelos Lineares Generalizados (MLG) apresentam-se como uma estrutura adequada de análise.

Nesse cenário, a regressão de Poisson destaca-se como uma técnica clássica para estimar taxas de incidência. No entanto, devido à frequente superdispersão presente nos registros de arboviroses, onde a variância dos casos excede significativamente a média durante os surtos, a regressão Binomial Negativa apresenta-se como a formulação matemática mais robusta. Essa abordagem permite quantificar com maior precisão o risco relativo associado a cada alteração climática, ajustando as incertezas inerentes ao comportamento dinâmico da doença (XAVIER *et al.*, 2024).

Para continuar auxiliando o estudo, também foi utilizado o algoritmo Random Forest (RF). Esse modelo tem a capacidade de mostrar quais variáveis mais contribuem para a tomada de decisão, sendo frequentemente utilizado como uma técnica de seleção de características. Essa seleção é considerada uma etapa essencial para a análise de dados, pois reduz a complexidade e a dimensionalidade do problema, além disso, o Random Forest se destaca por criar múltiplas árvores de decisão durante o seu processo, o que melhora a precisão dos resultados e ajuda a lidar muito bem com as variações complexas dos dados climáticos e epidemiológicos (LIMA *et al.*, 2021).

Apesar dos avanços na literatura, ainda existe uma lacuna significativa no que se refere à integração e comparação sistemática entre modelos estatísticos interpretáveis e técnicas de aprendizado de máquina em ambientes urbanos tropicais. Estudos que avaliem simultaneamente a capacidade explicativa e o desempenho preditivo dessas abordagens ainda são limitados, especialmente em contextos regionais específicos.

A escolha de contrapor um modelo paramétrico clássico com uma técnica não linear de aprendizado de máquina justifica-se pela necessidade de encontrar a ferramenta de maior precisão explicativa para o ecossistema local, otimizando o direcionamento de políticas públicas de controle vetorial (SEPÚLVEDA-SALCEDO *et al.*, 2024).

Dessa forma, o objetivo da presente pesquisa foi modelar e prever a incidência de dengue no Recife a partir de variáveis climáticas, comparando o desempenho da Regressão Binomial Negativa e do Random Forest.

MATERIAL E MÉTODOS

Área de estudo

O presente estudo foi realizado no município do Recife, capital do estado de Pernambuco, Brasil. A escolha da área justifica-se pela sua elevada densidade demográfica, características de infraestrutura urbana e

condições climáticas tropicais, que em conjunto criam um ambiente altamente propício para a proliferação do *Aedes aegypti*, vetor transmissor da dengue.

A pesquisa caracteriza-se como um estudo retrospectivo de séries temporais, compreendendo o período de janeiro de 2018 a dezembro de 2021. Foram analisados os registros mensais de casos de dengue, com o intuito de investigar a dinâmica temporal da doença e sua relação com variáveis climáticas.

Base de dados e pré-processamento

Os dados epidemiológicos foram obtidos a partir dos registros oficiais da Secretaria de Saúde, disponibilizados no Portal de Dados Abertos da Prefeitura do Recife. A evolução temporal dos casos de dengue é ilustrada na Figura 1.

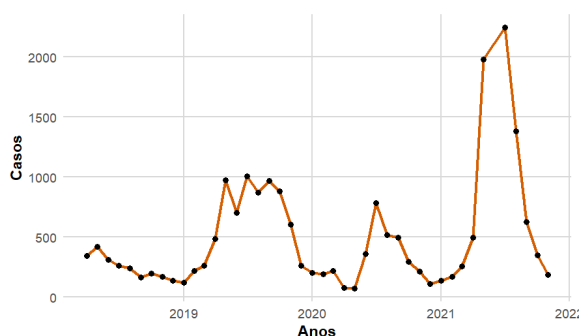


Figura 1. Série temporal dos casos mensais de dengue no Recife (2018-2021).

A observação visual da série demonstra uma repetição anual no aumento de infecções, com explosões de casos ocorrendo sistematicamente em janelas temporais específicas. Este padrão cíclico e sazonal evidencia uma forte dependência das condições ambientais, sugerindo que o vetor *Aedes aegypti* encontra, em meses cruciais do ano, o cenário meteorológico ideal para a sua proliferação exponencial.

As variáveis climáticas precipitação total, temperatura média, umidade relativa e radiação solar, foram obtidas do Instituto Nacional de Meteorologia (INMET).

Inicialmente, os dados foram organizados em formato de série temporal, sendo convertidos para uma base mensal, a fim de garantir compatibilidade entre as frequências de registro (dados climáticos diários e epidemiológicos semanais). O processamento, limpeza e integração dos dados foram realizados no ambiente de programação R.

Variáveis auxiliares relacionadas à sazonalidade foram incluídas, como mês e ano, permitindo capturar padrões temporais recorrentes. Adicionalmente, foram geradas variáveis defasadas (*lags*) das variáveis climáticas, considerando atrasos de até três meses, com o objetivo de capturar efeitos retardados das condições meteorológicas sobre a incidência da dengue.

Outliers foram identificados e removidos com base no escore Z padronizado ($|z| > 3$), garantindo maior robustez na modelagem estatística e reduzindo a influência de valores extremos.

Análise exploratória de séries temporais

A estrutura temporal dos dados foi investigada por meio de um conjunto integrado de análises exploratórias, com o objetivo de identificar padrões sazonais, dependência temporal e possíveis relações entre variáveis climáticas e a incidência de dengue. Inicialmente, aplicou-se a decomposição sazonal da série temporal (STL), permitindo separar os componentes de tendência, sazonalidade e ruído, o que possibilitou uma compreensão mais detalhada da dinâmica subjacente dos dados. Em seguida, foram utilizadas as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) para avaliar a presença de dependência temporal e a estrutura de memória da série, contribuindo para a identificação de padrões recorrentes ao longo do tempo.

Adicionalmente, realizou-se a análise de correlação entre as variáveis climáticas e o número de casos de dengue, visando quantificar a intensidade das associações lineares. Para complementar essa abordagem, foi empregada a análise de correlação cruzada (*cross-correlation*), permitindo investigar relações defasadas entre variáveis meteorológicas e a incidência da doença, identificando possíveis atrasos temporais entre mudanças climáticas e seus efeitos epidemiológicos. Em conjunto, essas análises forneceram subsídios fundamentais para a compreensão da dinâmica temporal da dengue e para a definição das estruturas preditivas adotadas na modelagem.

Divisão dos dados e estratégia de validação

Para avaliação do desempenho preditivo, os dados foram divididos em conjuntos de treinamento (80%) e teste (20%), respeitando a ordem temporal da série, evitando vazamento de informação.

Adicionalmente, foi aplicada validação do tipo *walk-forward*, na qual o modelo é iterativamente treinado com dados históricos acumulados e testado em observações futuras. Essa abordagem preserva a



estrutura temporal da série e fornece uma avaliação mais realista da capacidade preditiva dos modelos em cenários reais.

Modelagem estatística: Regressão Binomial Negativa

A relação entre as variáveis climáticas e a incidência de dengue foi inicialmente modelada por meio da regressão Binomial Negativa. Essa abordagem foi adotada devido à natureza discreta e não negativa dos dados de contagem, bem como à presença de superdispersão, característica comum em séries epidemiológicas.

O modelo utiliza uma função de ligação logarítmica para relacionar a média esperada do número de casos às variáveis explicativas:

$$\ln(\mu_t) = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} \quad (1)$$

Onde μ_t representa o número esperado de casos de dengue no tempo t , β_0 é o intercepto e β_i são os coeficientes associados às variáveis climáticas X_i no respectivo mês t .

A adequação do modelo foi avaliada por meio da análise dos resíduos, incluindo a função de autocorrelação (ACF), visando verificar a independência temporal e a qualidade do ajuste.

Modelo de *Machine Learning*: Random Forest

Com o intuito de estabelecer um termo de comparação robusto em relação ao modelo estatístico clássico, foi implementado o algoritmo Random Forest, uma abordagem de aprendizado de máquina (*Machine Learning*) caracterizada por metodologias recentes e de alto impacto para a previsão de surtos de dengue (SEBASTIANELLI *et al.*, 2024).

Trata-se de um método do tipo *ensemble* (conjunto), baseado na construção de múltiplas árvores de decisão independentes durante a fase de treinamento e combina os seus resultados para gerar uma previsão final mais robusta e menos suscetível a erros (*overfitting*).

No contexto de problemas de regressão, como a predição do número de casos de dengue, a saída do modelo é obtida por meio da média das previsões individuais de todas as árvores que compõem a floresta, conforme expresso pela seguinte equação:

$$\hat{Y} = \frac{1}{B} \sum_{b=1}^B T_b(X) \quad (2)$$

Onde $\hat{Y}$ representa a previsão final do número de casos de dengue, B é o número total de árvores de decisão geradas, e $T_b(X)$ indica a previsão feita pela árvore b a partir do conjunto de variáveis climáticas preditoras X , incluindo precipitação, temperatura, umidade relativa e Radiação solar.

A principal vantagem da utilização desta abordagem computacional, é a sua capacidade inerente de capturar relações não-lineares e interações complexas entre as variáveis climáticas sem a necessidade de assumir premissas rígidas sobre a distribuição dos dados originais (SEBASTIANELLI *et al.*, 2024). Dessa forma, o algoritmo consegue identificar limiares climáticos específicos (por exemplo, combinações exatas de alta temperatura e umidade) que engatilham a explosão de focos do vetor na cidade.

A importância das variáveis foi avaliada com base na contribuição de cada preditor para a redução do erro de predição, permitindo identificar os fatores climáticos mais relevantes para a incidência da dengue.

Avaliação do desempenho dos modelos

O desempenho preditivo dos modelos de Regressão Binomial Negativa e Random Forest foi avaliado com base na comparação entre os valores observados (Y_i) e previstos ($\hat{Y}_i$), utilizando métricas estatísticas amplamente consolidadas na literatura. A Raiz do Erro Quadrático Médio (RMSE) foi empregada para quantificar a magnitude dos erros, com maior penalização para desvios elevados:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (3)$$

Complementarmente, o Erro Absoluto Médio (MAE) foi utilizado por sua interpretação direta na mesma unidade da variável resposta:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (4)$$

O Erro Percentual Absoluto Médio (MAPE) foi aplicado para avaliar o erro relativo em termos percentuais:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (5)$$

Adicionalmente, a capacidade explicativa dos modelos foi mensurada por meio do coeficiente de determinação (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (6)$$

Onde $\hat{y}_i$ é o valor predito pelo modelo, y_i é o valor observado, n é o número de observações e $\bar{y}$ é a média dos valores observados.

A seleção do modelo mais adequado baseou-se na combinação dessas métricas, sendo considerado superior aquele que apresentou menores valores de erro (RMSE, MAE e MAPE) e maior capacidade explicativa (R^2).

Todas as análises foram realizadas utilizando o pacote de *software* R versão 4.4.1 (R CORE TEAM, 2025) com o RStudio IDE.

RESULTADOS E DISCUSSÃO

Os modelos de Regressão Binomial Negativa e Random Forest foram avaliados quanto à sua capacidade de prever a dinâmica temporal da dengue no Recife, utilizando dados climáticos como variáveis preditoras. Os resultados são apresentados na Tabela 1.

Tabela 1. Desempenho preditivo dos modelos de Regressão Binomial Negativa e Random Forest na previsão de casos de dengue no Recife.

Modelo	RMSE	MAE	MAPE (%)	R^2
Binomial Negativa	636,58	384,61	29,91	0,825
Random Forest	748,53	470,31	38	0,686

RMSE: raiz do erro quadrático médio; MAE: erro absoluto médio; MAPE (%): erro percentual absoluto médio; R^2 : coeficiente de determinação.

A análise evidencia uma clara superioridade preditiva da abordagem estatística paramétrica em relação ao algoritmo de aprendizado de máquina neste cenário específico. A Regressão Binomial Negativa apresentou menor Erro Quadrático Médio (RMSE = 636,58) em comparação ao Random Forest (RMSE = 748,53), indicando maior capacidade do modelo estatístico tradicional em capturar e lidar com os "picos" extremos e abruptos de incidência da doença, penalizando menos as previsões de surtos epidêmicos.

Essa abordagem está de acordo com Hamdi (2025) e Wu *et al.* (2025), que destacam os modelos de regressão para dados de contagem, especialmente a

binomial negativa, indicando que apresentam melhor desempenho em cenários com superdispersão e alta variabilidade, superando nesse quesito o modelo de aprendizado de máquina.

No que se refere aos erros absolutos e relativos, a Binomial Negativa consolidou a sua vantagem ao registrar um MAE = 384,61 e uma margem percentual de erro (MAPE) de apenas 29,91%, enquanto o algoritmo Random Forest distanciou-se da realidade em cerca de 38%. Conforme discutido por Li *et al.* (2025), modelos estatísticos aplicados a dados de contagem tentem a apresentar melhor desempenho em cenários com alta variabilidade, a exemplo dos epidemiológicos.

A capacidade explicativa dos modelos, mensurada pelo coeficiente de determinação, também favoreceu a abordagem paramétrica. A Regressão Binomial Negativa explicou 82,5% da variabilidade dos casos ($R^2 = 0,825$), enquanto o Random Forest apresentou $R^2 = 0,686$. Esse resultado sugere que a estrutura probabilística do modelo estatístico foi mais eficiente em representar a dinâmica da doença a partir de variáveis climáticas, corroborando com os achados de Sánchez *et al.* (2025).

Apesar do melhor desempenho preditivo da Regressão Binomial Negativa, o Random Forest apresentou vantagens relevantes na interpretação das relações não lineares entre variáveis climáticas e incidência da dengue. A análise de importância das variáveis indicou que a radiação solar média foi o fator mais influente, seguida pela temperatura média e precipitação, enquanto a umidade relativa apresentou menor contribuição (Figura 2).

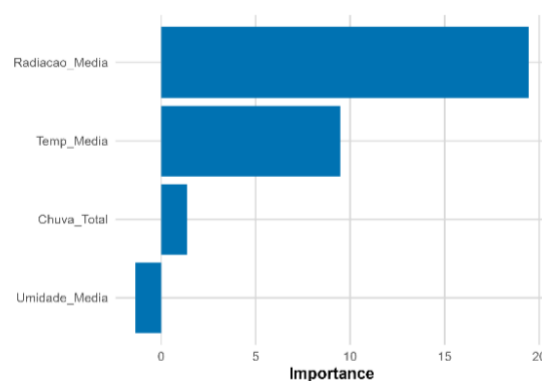


Figura 2. Importância das variáveis climáticas na previsão da dengue segundo o modelo Random Forest.

Esse resultado sugere que a radiação solar pode atuar como um regulador microclimático crítico, influenciando diretamente a temperatura dos criadouros e o metabolismo do vetor, conforme discutido na literatura recente (RMASAY, 2025). A

predominância desse fator sobre a precipitação indica que a disponibilidade de água, embora essencial, não é o único determinante da dinâmica populacional do *Aedes aegypti*.

Para compreender o impacto individual da precipitação, a Análise de Dependência Parcial (*Partial Dependence Plot*) forneceu evidências robustas sobre os limiares de risco na região (Figura 3).

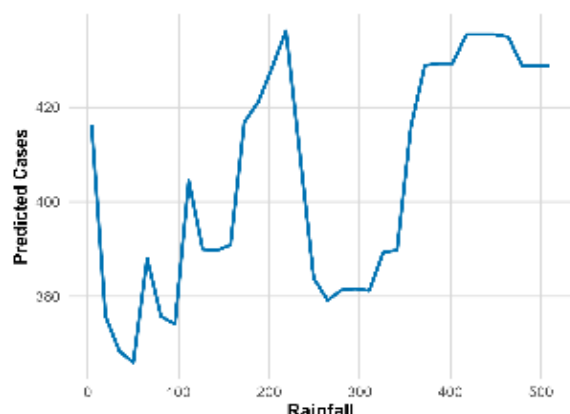


Figura 3. Análise de dependência parcial evidenciando o efeito da precipitação na incidência de dengue.

A análise de dependência parcial evidenciou uma relação não linear entre precipitação e incidência da doença, com aumento expressivo no número de casos em volumes de chuva entre 200 e 220 mm mensais. Entretanto, volumes extremos (acima de 250 mm) estiveram associados à redução da incidência, sugerindo a ocorrência do chamado *washout effect*, no qual chuvas intensas podem eliminar criadouros larvais, impedindo mecanicamente o desenvolvimento das larvas e pupas antes que emergjam como vetores adultos.

A interação entre variáveis climáticas foi explorada por meio da superfície de resposta (Figura 4) que ilustra visualmente a "tempestade perfeita" para os surtos de dengue na região metropolitana do Recife.

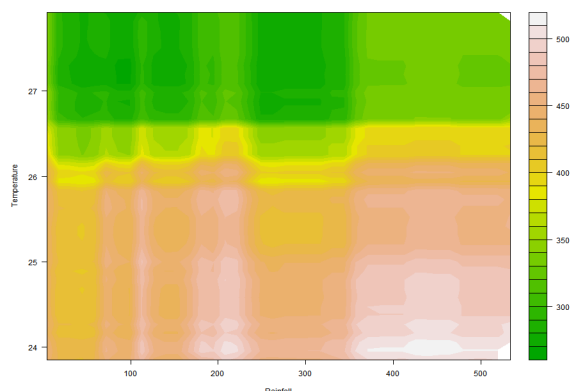


Figura 4. Superfície de resposta climática representando a interação entre precipitação e temperatura na previsão de casos.

A análise evidencia que a doença não responde de forma linear. Surpreendente, as zonas de maior risco epidemiológico (representadas pelas cores mais claras/rosadas, indicando previsões acima de 450 casos) não ocorrem nos picos absolutos de calor. Pelo contrário, o cenário crítico é engatilhado pela convergência de altas precipitações (acima de 300 mm) associadas a temperaturas mais amenas para a região, gravitando entre 24 e 25°C. À medida que a temperatura se aproxima dos 27 °C, a previsão de casos cai drasticamente (zonas verdes da superfície), indicando que o calor extremo na capital pernambucana pode atuar como um fator de estresse térmico para o mosquito, reduzindo a sobrevivência das fêmeas adultas e limitando a transmissão.

Um dos desafios mais complexos na modelagem epidemiológica de doenças transmitidas por vetores é a não-simultaneidade dos eventos climáticos e do adocimento humano. A análise de correlação cruzada (Figura 5) evidenciou a presença de defasagem temporal entre eventos climáticos e aumento de casos de dengue. Observou-se que o pico pluviométrico não reflete em um aumento imediato das notificações no mesmo mês.

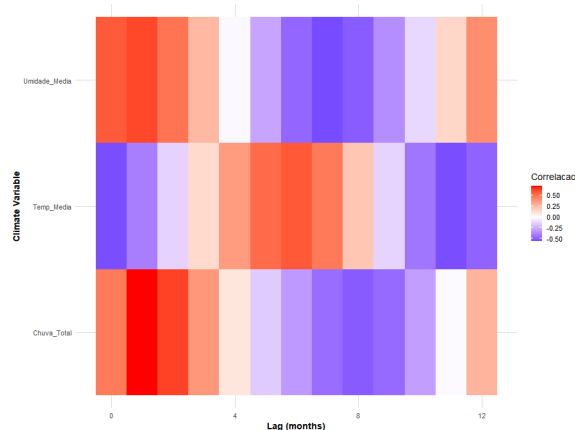


Figura 5. Correlação cruzada entre variáveis climáticas e casos de dengue com diferentes defasagens temporais.

A análise evidencia uma forte correlação positiva (tons de vermelho intenso para a Chuva Total) quando se aplica um *lag* (atraso) de 1 a 2 meses entre o evento de precipitação e a variável de saúde. Esse atraso temporal reflete com precisão as etapas biológicas obrigatórias do ciclo de transmissão. Após a chuva preencher os reservatórios, há a eclosão, desenvolvimento imaturo e emergência do mosquito adulto. A fêmea infectada necessita do Período de Incubação Extrínseco (PIE) para que o vírus alcance

suas glândulas salivares e, após picar um humano, soma-se o período de incubação intrínseco até a notificação do sintoma no hospital.

O modelo indica, portanto, que a vigilância em saúde no Recife não deve utilizar a chuva do mês corrente para dimensionar os leitos, mas sim usá-la como um alerta precoce e preditivo para os surtos epidêmicos que estourarão num horizonte de 30 a 60 dias.

A aplicação de técnicas de interpretabilidade baseadas em valores SHAP (*SHapley Additive exPlanations*) permitiu compreender a contribuição individual das variáveis no modelo Random Forest (Figura 6).

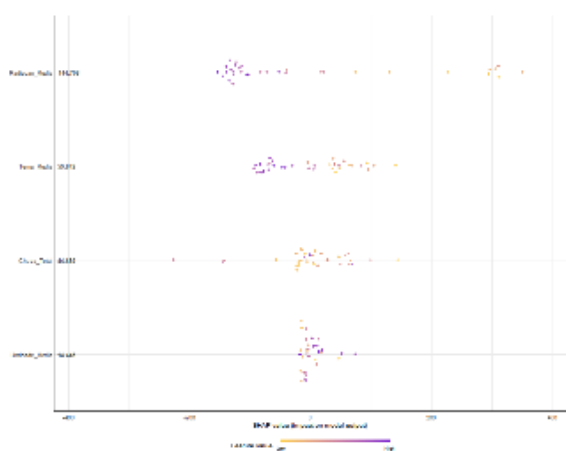


Figura 6. Valores SHAP mostrando a contribuição das variáveis climáticas nas previsões do modelo Random Forest.

Observa-se que valores elevados de radiação solar e temperatura média (representados pelos pontos de coloração mais quente/vermelha à direita do eixo central) estão associados a impactos estritamente positivos na saída do modelo. Esses resultados indicam que condições de maior insolação e temperatura favorecem o aumento previsto no número de casos de dengue, possivelmente devido à aceleração de desenvolvimento e da atividade do vetor *Aedes aegypti*.

Em contrapartida, a precipitação (chuva total) apresenta um padrão de impacto mais complexo. A distribuição dos valores SHAP evidencia uma relação não linear, na qual volumes moderados de precipitação estão associados ao aumento do risco de surtos, enquanto eventos pluviométricos extremos apresentam menor contribuição positiva, refletindo o chamado *washout effect*, que pode reduzir a disponibilidade de criadouros ao remover larvas e pupas.

Por sua vez, a umidade relativa apresentou menor influência preditiva, com valores SHAP concentrados próximos ao eixo central, indicando contribuição

limitada para variações significativas nas previsões do modelo.

Em conjunto, a análise baseada em valores SHAP contribui para a interpretação do modelo Random Forest, permitindo não apenas avaliar sua capacidade preditiva, mas também compreender a direção e magnitude dos efeitos das variáveis climáticas, fornecendo subsídios interpretáveis e consistentes para aplicações em vigilância epidemiológica e tomada de decisão em saúde pública.

CONCLUSÃO

Este estudo evidenciou a forte dependência da incidência de dengue às variáveis climáticas no município do Recife, destacando o papel do clima como um preditor relevante para a vigilância epidemiológica. A Regressão Binomial Negativa apresentou desempenho superior ao Random Forest em todas as métricas avaliadas, indicando maior adequação para modelagem de dados de contagem caracterizados por superdispersão.

Por outro lado, o Random Forest contribuiu de forma complementar ao permitir a identificação de relações não lineares e limiares críticos entre variáveis climáticas e a incidência da doença. Os resultados indicam que a combinação de temperatura elevada e precipitação moderada favorece o aumento dos casos, enquanto eventos pluviométricos extremos podem reduzir a proliferação do vetor, sugerindo a ocorrência do *washout effect* (efeito lavagem).

Adicionalmente, a identificação de defasagens temporais entre eventos climáticos e notificações de casos (1 a 2 meses) reforça o potencial do uso de variáveis meteorológicas como ferramentas de alerta precoce. Esses achados fornecem subsídios científicos relevantes para o desenvolvimento de estratégias preditivas e para o aprimoramento de políticas públicas voltadas ao controle da dengue em ambientes urbanos tropicais.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsas de pesquisa.

À Universidade Estadual da Paraíba (UEPB), pelo suporte institucional.

REFERÊNCIAS

BAGCCHI, S. Dengue outbreak in Brazil. *The Lancet Infectious Diseases*, v. 23, n. 2, p. 152, 2023.



- BU CZAK, A. L. et al. Ensemble method for dengue prediction. *PLoS ONE*, v. 13, n. 1, e0189988, 2018.
- GOMES, A. F. et al. Climate variability and dengue incidence in northeastern Brazil. *Revista Brasileira de Epidemiologia*, v. 27, e240012, 2024.
- HAMDI, R. Statistical modeling for infectious disease counts under overdispersion. *Journal of Applied Statistics*, v. 52, n. 1, p. 88–102, 2025.
- LI, X. et al. Comparative performance of statistical and machine learning models in epidemiological prediction. *Scientific Reports*, v. 15, 2025.
- LIMA, T. P. F. et al. Previsão de óbito e importância de características clínicas em idosos com COVID-19 utilizando o Algoritmo Random Forest. **Revista Brasileira de Saúde Materno Infantil**, v. 21, 2021.
- LIMA-CAMARA, T. N. A dengue é produto do meio: uma abordagem sobre os impactos do ambiente no mosquito *Aedes aegypti* e nos casos da doença. **Revista Brasileira de Epidemiologia**, v. 27, e240048, 2024.
- MORENO SÁNCHEZ, J. C. et al. Improving wheat yield prediction through variable selection using Support Vector Regression, Random Forest, and Extreme Gradient Boosting. **Smart Agricultural Technology**, v. 10, p. 100791, 2025.
- MAGALHÃES, T. et al. The influence of temperature on dengue transmission dynamics. *Parasites & Vectors*, v. 13, p. 1–10, 2020.
- OLIVEIRA, M. R. Climate change and public health impacts in Brazil. *Revista de Saúde Pública*, v. 58, p. 1–12, 2024.
- RAMSAY, E. E. et al. Causal drivers of mosquito abundance in urban informal settlements. *Environmental Research Letters*, [S.l.], v. 20, n. 3, p. 1-12, 2025.
- R Core Team. *_R: A Language and Environment for Statistical Computing_*. R Foundation for statistical Computing, Vienna, Austria, 2025. Disponível em: <https://www.R-project.org/>. Acesso em: 27 jan. 2026.
- SEPÚLVEDA-SALCEDO, L. S. et al. Effects of climatic variables on dengue incidence in Cali. **medRxiv**, p. 2024.05.01, 2024.
- SILVA, M. R. Gestão de desastres climáticos e o impacto nas emergências de saúde pública. **Revista Emergência**, Novo Hamburgo, v. 16, 2024.
- SANABANI, S. S. Epidemiologia da dengue no Brasil: tendências recentes e resposta de saúde pública. *Discover Public Health*, v. 22, p. 518, 2025.
- SEBASTIANELLI, M. et al. Machine learning approaches for dengue outbreak prediction. *Environmental Modelling & Software*, v. 180, 2024.
- VÁSQUEZ, C. et al. Climate-driven transmission of dengue: a global perspective. *The Lancet Planetary Health*, v. 7, n. 3, p. e210–e220, 2023.
- WU, P. et al. Advances in statistical modeling for infectious disease prediction. *Nature Communications*, v. 16, 2025.
- XAVIER, F. et al. Big Data-Planetary Health approach for evaluating the Brazilian Dengue Control Program. **Revista de Saúde Pública**, v. 58, p. 17, 2024.
- ZHOU, Y. et al. Global patterns of dengue transmission under climate change scenarios. *Nature Medicine*, v. 31, 2025.