

Caracterização Socioespacial de Áreas Vulneráveis à Chikungunya em Pernambuco

Bianka Uiara Cândido da Silva¹, Andrêza Leite de Alencar²

¹Departamento de Informática e Estatística,
Universidade Federal Rural de Pernambuco (UFRPE)
Caixa Postal 52171-900 – Recife – PE – Brasil

²Departamento de Computação,
Universidade Federal Rural de Pernambuco (UFRPE)
Caixa Postal 52171-900 – Recife – PE – Brasil

{bianka.uiara, andreza.leite}@ufrpe.br

Abstract. *This paper presents partial results of a research aimed at developing a predictive risk model for chikungunya in Pernambuco, integrating social, territorial, and climatic determinants. In this initial stage, 106,223 microdata records from SINAN/DATASUS. Exploratory analysis revealed epidemic concentration in 2021–2022, first-semester seasonality with peak frequency in May, and female and mixed-race predominance, consistent with socioeconomically vulnerable populations. The integrated database will support georeferenced cluster mapping and predictive modeling of risk areas.*

Resumo. *Este trabalho apresenta resultados parciais de uma pesquisa que visa desenvolver um modelo preditivo de risco à chikungunya em Pernambuco, integrando determinantes sociais, territoriais e climáticos. Foram integrados nesta etapa inicial 106.223 microdados do SINAN/DATASUS. A análise exploratória revelou concentração epidêmica no biênio 2021–2022, sazonalidade com pico predominante no primeiro semestre com maior frequência em maio e predominância feminina e parda, consistente com populações socioeconomicamente vulneráveis. A base integrada subsidiará mapeamento georreferenciado de clusters e modelagem preditiva de áreas de risco.*

1. Introdução

As doenças transmitidas por vetores, em especial as arboviroses, representam um crescente problema de saúde pública no Brasil e na América Latina. A chikungunya, causada pelo vírus CHIKV e transmitida pelo mosquito *Aedes aegypti*, emergiu como epidemia no Brasil a partir de 2014, com padrão endêmico-epidêmico nas regiões Nordeste e Sudeste. Segundo o Boletim Epidemiológico do Ministério da Saúde, foram notificados 102.259 casos prováveis somente da 1ª à 23ª semana epidemiológica de 2025, dos quais Pernambuco concentrou parcela expressiva [Brasil. Ministério da Saúde 2025].

As doenças tropicais negligenciadas (DTNs) afetam predominantemente populações em situação de pobreza, com infraestrutura sanitária precária e limitado acesso a serviços de saúde [Organization et al. 2022]. No Brasil, as condições climáticas tropicais, a rápida urbanização desordenada e as iniquidades socioeconômicas estruturais

criam ambiente propício para a proliferação do vetor. Indicadores como o coeficiente de Gini, a proporção de domicílios sem coleta de lixo e a densidade populacional estão associados à maior incidência de arboviroses [Kikuti et al. 2015, Aguiar-Santos et al. 2023].

A vigilância epidemiológica baseada em dados (Data-Driven Surveillance) permite integrar registros de notificação compulsória, censos demográficos e dados de mobilidade humana para identificar padrões de transmissão e estratificar populações de risco [Coelho and Carvalho 2015]. O presente trabalho propõe a construção de uma base de dados integrada para a caracterização socioespacial de áreas vulneráveis à chikungunya em Pernambuco, com potencial de subsidiar políticas públicas de vigilância vetorial com maior precisão territorial.

2. Objetivos

O objetivo geral é estruturar uma base de dados integrada para análise de vulnerabilidade socioespacial à chikungunya em Pernambuco, a partir de fontes públicas brasileiras, para subsidiar o desenvolvimento de um modelo preditivo de machine learning para identificação de áreas de risco. Os objetivos específicos são: (i) identificar preditores de vulnerabilidade; (ii) treinar algoritmos de classificação supervisionada para mapeamento de hotspots; (iii) validar o modelo com dados históricos; e (iv) desenvolver um dashboard georreferenciado para apoio à decisão em saúde pública.

3. Metodologia

A pesquisa adota abordagem quantitativa exploratório-descritiva com dados secundários, seguindo o framework CRISP-DM [Wirth and Hipp 2000] adaptado à saúde pública. O pipeline foi implementado em Python versão 3.11.

3.1. Fontes de Dados

Utilizou-se o SINAN/Chikungunya (CID-10: A92.0) [Brasil. Ministério da Saúde. DATASUS 2020], acessado via biblioteca PySUS [Coelho et al. 2026] pelo FTP do DATASUS. Como fontes complementares: SIDRA/IBGE [IBGE 2023] para indicadores socioeconômicos do Censo 2022; Atlas Brasil [PNUD et al. 2013] para IDH municipal; SIVEP-Gripe [Brasil. Ministério da Saúde. DATASUS 2021] para co-circulação viral; e malha municipal IBGE [IBGE 2022] para análise espacial.

3.2. Pipeline de Pré-processamento

Os dados foram extraídos para 2017–2023 e filtrados: (i) agravo CHIK na extração; (ii) restrição a Pernambuco por SG_UF_NOT = 26; (iii) verificação de integridade (n=0 excluídos). O campo de data foi padronizado para datetime64. Das 122 variáveis originais, 49 foram retidas por completude 70% e relevância epidemiológica — as demais correspondiam a resultados laboratoriais e sinais de gravidade, sistematicamente subnotificados. CS_ESCOL_N (28,9% nulos) e EVOLUCAO (20,2% nulos) foram mantidas com ressalva.

A integração com SIDRA/IBGE será realizada pelo código IBGE de município (7 dígitos), compatível com ID_MUNICIP no SINAN. A análise espacial utilizará geopandas com left join sobre a malha municipal do IBGE, preservando municípios sem registro como missing explícito. Os arquivos DBC foram convertidos para Parquet via PySUS, reduzindo o tempo de leitura nas iterações subsequentes.

4. Resultados e Discussão

4.1. Bases Coletadas e Volume de Dados

Conforme a Tabela 1, após extração e filtragem, a base do SINAN resultou em 106.223 registros de notificação de chikungunya para o período de 2017–2023, abrangendo todos os municípios de Pernambuco. Os dados socioeconômicos do IBGE encontram-se em extração via SIDRA, a partir do Censo Demográfico 2022, e serão integrados na etapa subsequente.

Tabela 1. Dimensões das bases de dados obtidas na fase inicial.

Base	Registros	Variáveis	Cobertura
SINAN – Chikungunya	106.223	122 (brutas), 49 (sel.)	PE – 2017–2023

Nota: As 49 variáveis selecionadas (completude $\geq 70\%$) abrangem identificação da notificação, temporalidade, dados demográficos, sintomas clínicos, comorbidades e desfecho. Nenhum registro foi excluído no pré-processamento ($n=0$).

A presente análise utiliza exclusivamente a base do SINAN, já pré-processada. A integração com SIDRA/IBGE e a malha municipal encontra-se em andamento e será incorporada nas etapas subsequentes, permitindo análises espaciais e de determinantes socioeconômicos.

4.2. Análise Exploratória Preliminar

A análise exploratória investigou a distribuição temporal dos casos de 2017 a 2023. Observou-se variabilidade interanual no mês de pico de transmissão: março (2023), maio (2017, 2018, 2019 e 2022), junho (2021) e julho (2020). A proporção de casos entre março e junho variou de 24,3% em 2020 a 73,4% em 2022. A Figura 1 evidencia esse contraste, com a curva de 2022 apresentando pico acentuado em maio, compatível com o período chuvoso de Pernambuco.

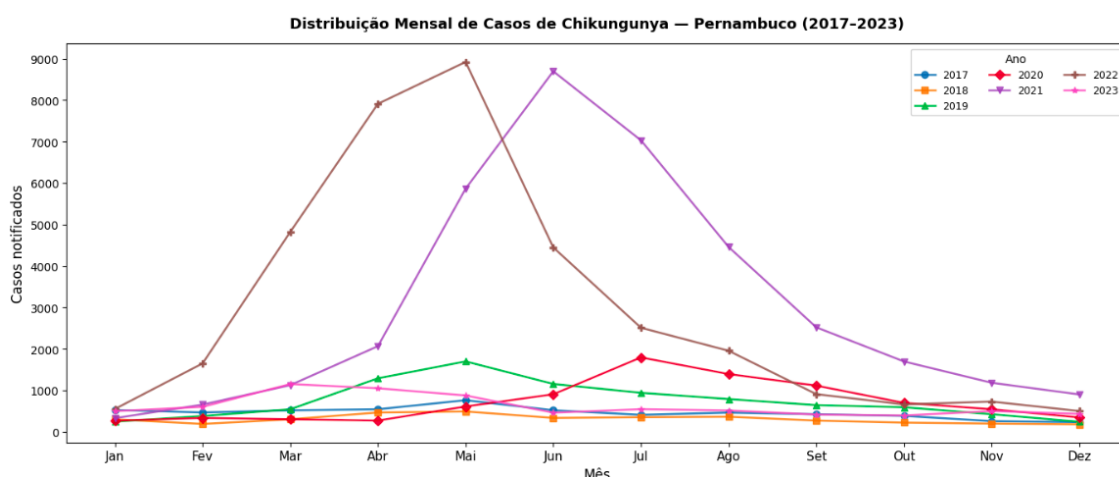


Figura 1. Distribuição mensal de casos de chikungunya em Pernambuco, 2017–2023.

O biênio 2021–2022 concentrou 72.053 notificações (67,9% do total), consistente com o enfraquecimento das ações de controle vetorial durante a pandemia de COVID-19.

A queda em 2023 (7.421 casos; 7,0%) pode refletir redução da suscetibilidade populacional e retomada da vigilância vetorial. A distribuição por sexo indicou predominância feminina: 61.875 casos (58,3%), contra 43.969 masculinos (41,4%) e 379 com sexo ignorado (0,4%). Entre os casos com raça/cor informada, houve prevalência de pardos: 56.822 registros (53,5%), seguidos de brancos (10,3%), pretos (2,4%), amarelos (0,5%) e indígenas (0,5%). A alta incompletude de raça/cor (32,7%) limita interpretações conclusivas (Figura 2).

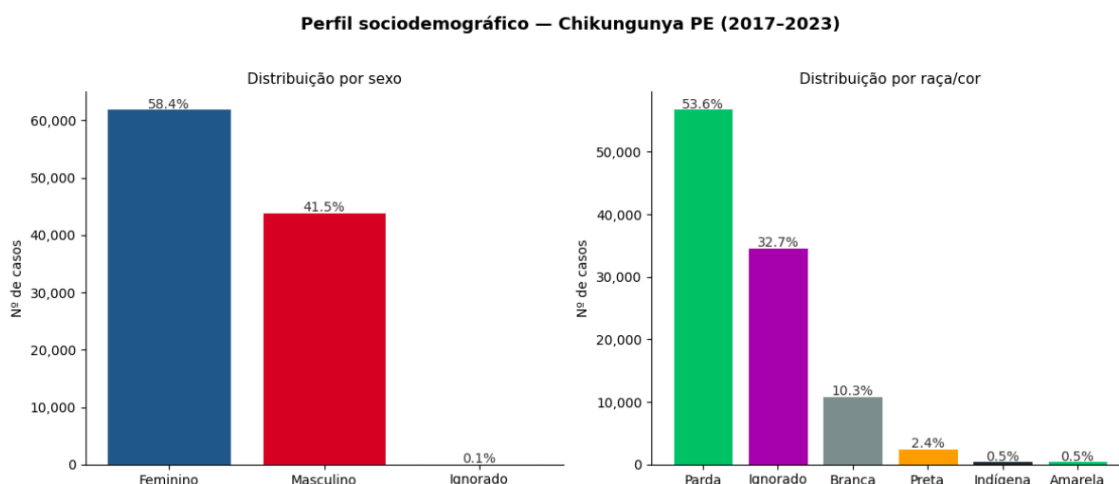
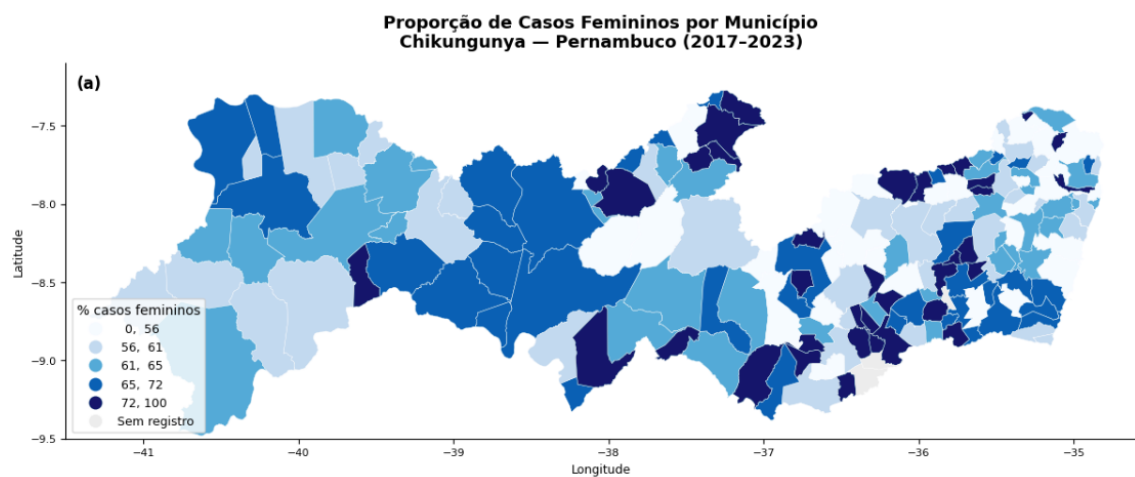


Figura 2. Distribuição de casos de chikungunya por sexo e raça/cor, Pernambuco, 2017-2023.

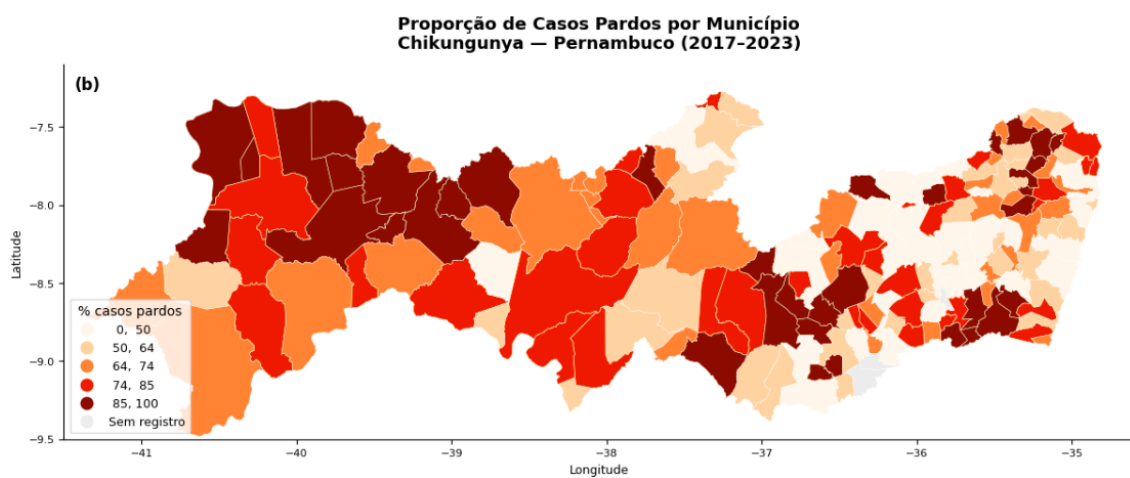
As Figuras 3a e 3b apresentam a distribuição espacial da proporção de casos femininos e pardos por município, em quintis. Observa-se tendência de maior proporção feminina na porção oeste do estado, padrão a ser verificado estatisticamente. A proporção de casos pardos é expressiva em praticamente todo o território, refletindo o perfil demográfico pernambucano e sua associação com vulnerabilidade socioeconômica.

5. Considerações Finais

As atividades da etapa inicial permitiram estruturar a base metodológica e de dados necessária para as análises subsequentes do projeto. A extração dos microdados do SINAN foi concluída com êxito, resultando em 106.223 registros de notificação de chikungunya para o estado de Pernambuco no período de 2017 a 2023. A extração dos indicadores socioeconômicos do IBGE para todos os municípios pernambucanos encontra-se em andamento, com integração à base epidemiológica prevista para a etapa seguinte. A análise exploratória revelou variabilidade interanual significativa, com concentração expressiva de casos no biênio 2021-2022 (67,9% do total), padrão consistente com o enfraquecimento das ações de controle vetorial durante a pandemia de COVID-19 e a dinâmica de recorrência do CHIKV. A distribuição mensal indicou predominância no primeiro semestre, com pico em maio, compatível com o período chuvoso de Pernambuco. O perfil sociodemográfico revelou predominância feminina (58,4%) e prevalência de indivíduos pardos (53,6%), achados consistentes com a literatura epidemiológica. A análise espacial desta etapa limitou-se à visualização coroplética por quintis, sem validação estatística dos padrões identificados. A próxima fase incluirá o cálculo do Índice de Moran Global e



(a)



(b)

Figura 3. Distribuição espacial da proporção de casos por município, Pernambuco, 2017–2023. (a) Proporção de casos femininos; (b) Proporção de casos pardos. Classificação por quintis. Fonte: SINAN/DATASUS; Malha municipal: IBGE (2022).

estatísticas LISA para confirmação dos clusters, bem como o uso de taxas de incidência padronizadas por 100.000 habitantes para controle da densidade populacional. A próxima fase compreenderá a integração dos dados censitários em nível de setor censitário, o cálculo de taxas de incidência padronizadas por 100.000 habitantes, a incorporação de indicadores de desenvolvimento humano via Atlas Brasil e a construção de índices de vulnerabilidade socioespacial para caracterização das áreas de risco à chikungunya em Pernambuco.

Referências

- Aguiar-Santos, M., Mendes, L. G. D. C., Passos, D. F. D., Santos, T. G. D. S., Lins, R. H. F. B., and Monte, A. C. P. D. (2023). Análise espacial da incidência da febre de Chikungunya e dos fatores socioeconômicos, demográficos e de infestação vetorial associados, em municípios de Pernambuco, Brasil, 2015–2021. *Revista Brasileira de Epidemiologia*, 26:e230018.
- Brasil. Ministério da Saúde (2025). Informe Semanal nº 17 - COE Dengue e outras Arboviroses - SE 1 a 23 | 17 de junho — Brasil. Secretaria de Vigilância em Saúde e Ambiente. Ministério da Saúde.
- Brasil. Ministério da Saúde. DATASUS (2020). Sistema de informação de agravos de notificação (sinan). <https://portalsinan.saude.gov.br/>. Acesso em: 21 abr. 2026.
- Brasil. Ministério da Saúde. DATASUS (2021). Sistema de informação de vigilância epidemiológica da gripe (sivep-gripe). <https://sivepgripe.saude.gov.br/>. Acesso em: 21 abr. 2026.
- Coelho, F. C. and Carvalho, L. M. D. (2015). Estimating the Attack Ratio of Dengue Epidemics under Time-varying Force of Infection using Aggregated Notification Data. *Scientific Reports*, 5(1):18455.
- Coelho, F. C., Vacaro, L., Semantic Release Bot, Sandro Loch, Bernardo Chrispim Baron, Gabriel Machado de Castro Fonseca, Ana Paula Gomes, Pedro Reck, Felipe Adeildo, Daidson Alves, Daniela Palumbo, Justin Poehnelt, Luis Novo, and shotaf (2026). *Aler-taDengue/PySUS: 2.2.0*.
- IBGE (2022). Malha municipal do brasil. <https://geoftp.ibge.gov.br/>. Acesso em: 17 abr. 2026.
- IBGE (2023). Sistema ibge de recuperação automática (sidra). <https://sidra.ibge.gov.br>. Acesso em: 21 abr. 2026.
- Kikuti, M., Cunha, G. M., Paploski, I. A. D., Kasper, A. M., Silva, M. M. O., Tavares, A. S., Cruz, J. S., Queiroz, T. L., Rodrigues, M. S., Santana, P. M., Lima, H. C. A. V., Calcagno, J., Takahashi, D., Gonçalves, A. H. O., Araújo, J. M. G., Gauthier, K., Diuk-Wasser, M. A., Kitron, U., Ko, A. I., Reis, M. G., and Ribeiro, G. S. (2015). Spatial distribution of dengue in a brazilian urban slum setting: Role of socioeconomic gradient in disease risk. *PLOS Neglected Tropical Diseases*, 9(7):1–18.
- Organization, W. H. et al. (2022). *Ending the neglect to attain the sustainable development goals: a rationale for continued investment in tackling neglected tropical diseases 2021–2030*. World Health Organization.

PNUD, IPEA, and FJP (2013). Atlas do desenvolvimento humano no brasil. <http://www.atlasbrasil.org.br>. Acesso em: 17 abr. 2026.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester.