

# Previsão De Casos De Dengue No Brasil

Isabelle Dos Santos Santana

Camilla Semenzzatto Cavalcante

Marcela Quaresma Soares

Universidade Presbiteriana Mackenzie

São Paulo, 2025

## SUMÁRIO

- 1. Introdução**
  - 1.1 Contexto epidemiológico
  - 1.2 Problema de pesquisa
  - 1.3 Justificativa
  - 1.4 Objetivos
    - 1.4.1 Objetivo geral
    - 1.4.2 Objetivos específicos
  - 1.5 Relação com os Objetivos de Desenvolvimento Sustentável (ODS)
- 2. Referencial Teórico**
  - 2.1 Séries temporais
  - 2.2 Aprendizado de Máquina
  - 2.3 Modelos neurais para previsão epidemiológica
  - 2.4 Trabalhos correlacionados
- 3. Metodologia**
  - 3.1 Fonte de dados
  - 3.2 Estrutura e período da base
  - 3.3 Pré-processamento
    - 3.3.1 Tratamento de ausências
    - 3.3.2 Padronização e indexação temporal
    - 3.3.3 Engenharia de atributos
    - 3.3.4 Transformação do alvo
  - 3.4 Análise Exploratória de Dados (EDA)
  - 3.5 Criação das janelas temporais
  - 3.6 Modelagem
    - 3.6.1 Arquitetura GRU
    - 3.6.2 Estratégia de treino e validação
  - 3.7 Pipeline completo da solução
- 4. Resultados**
  - 4.1 Métricas de avaliação
  - 4.2 Comparação entre valores reais e previstos
  - 4.3 Análise dos resíduos
  - 4.4 Interpretação dos gráficos
  - 4.5 Síntese dos resultados
- 5. Discussão**
  - 5.1 Análise crítica
  - 5.2 Limitações
  - 5.3 Possíveis vieses
  - 5.4 Comparação com estudos semelhantes
- 6. Conclusão**
  - 6.1 Contribuições do projeto
  - 6.2 Aplicações práticas na vigilância epidemiológica
  - 6.3 Trabalhos futuros
- 7. Referências**

## **1. INTRODUÇÃO**

### **1.1 Contextualização**

A dengue constitui um dos maiores desafios epidemiológicos enfrentados pelo Brasil nas últimas décadas, caracterizando-se por surtos cílicos, forte dependência de fatores climáticos e impactos expressivos na saúde pública. A urbanização acelerada, infraestrutura inadequada e mudanças ambientais ampliam a circulação do mosquito *Aedes aegypti*, vetor responsável pela transmissão. Embora o Ministério da Saúde e sistemas como o InfoDengue ofereçam monitoramento em tempo real, a atuação governamental muitas vezes é reativa, com ações intensificadas apenas após o aumento expressivo de casos. Nesse contexto, sistemas preditivos baseados em dados têm potencial para antecipar surtos e apoiar decisões estratégicas.

### **1.2 Problema**

A ausência de mecanismos preditivos robustos dificulta o planejamento de ações preventivas. Sem previsão confiável, gestores públicos enfrentam limitações na alocação de recursos, na execução de campanhas educativas e no controle vetorial.

### **1.3 Justificativa**

Modelos de séries temporais e técnicas modernas de aprendizado de máquina têm se mostrado eficazes no monitoramento de doenças infecciosas. Entre esses métodos, redes neurais recorrentes do tipo GRU destacam-se pela capacidade de capturar padrões sazonais, tendências e dependências de longo prazo.

O uso de dados reais do InfoDengue (2022–2025) possibilita a criação de uma ferramenta aplicada, alinhada às necessidades reais da vigilância epidemiológica.

### **1.4 Objetivo Geral**

Desenvolver um modelo preditivo baseado em Redes Neurais GRU capaz de prever semanalmente o número de casos estimados de dengue no Brasil.

### **1.5 Objetivos Específicos**

- Integrar dados epidemiológicos e climáticos do InfoDengue.
- Realizar análise exploratória e pré-processamento da série temporal.
- Criar janelas temporais de 52 semanas para previsão.
- Treinar, validar e avaliar um modelo GRU.
- Analisar os resultados e discutir aplicabilidade prática.

## **2. FUNDAMENTOS TEÓRICOS**

### **2.1 Séries Temporais**

Séries temporais representam fenômenos observados ao longo do tempo, caracterizando-se por componentes estruturais como tendência, sazonalidade e ruído. A modelagem exige técnicas que considerem autocorrelação e padrões recorrentes. Neste projeto, utilizou-se a decomposição temporal, análise de autocorrelação e transformação logarítmica.

### **2.2 Modelos Preditivos e GRU**

Redes neurais recorrentes (RNNs) processam sequências preservando dependências temporais. Entre suas variações, as GRUs (Gated Recurrent Units) apresentam bom desempenho em séries complexas e são mais eficientes que LSTMs devido à estrutura simplificada de portas internas.

O modelo adotado combina:

- uma camada GRU,
- camada densa intermediária,
- saída linear para previsão contínua.

### 3. METODOLOGIA

#### 3.1 Descrição dos Dados

A base utilizada foi obtida do InfoDengue, contendo registros semanais de 2022 a 2025.

O conjunto inclui:

- casos estimados (*casos\_est*),
- $R_t$ ,
- variáveis climáticas (temperatura, umidade, precipitação),
- tweets relacionados ao tema.

Foram utilizados 30 atributos originais antes da limpeza.

#### 3.2 Pré-processamento

O fluxo de preparação incluiu:

- conversão de *data\_iniSE* para formato *datetime*;
- ordenação cronológica;
- interpolação linear para variáveis contínuas;
- preenchimento de tweets com média;
- exclusão de colunas irrelevantes;
- padronização dos tipos;
- eliminação de duplicidades;
- definição do índice temporal.

#### 3.3 Engenharia de Atributos

Foram criadas 38 variáveis derivadas, incluindo:

- Lags de 1 a 12 semanas;
- Médias móveis de 7, 14, 21, 30 e 52 semanas;
- Amplitude térmica (*tempmax – tempmin*);
- Representações sazonais (seno e cosseno para mês e semana do ano);
- Transformação  $\log_{10}$  no alvo para estabilizar a variância.

#### 3.4 Criação das Janelas Temporais

Para previsão semanal, adotou-se:

- 52 semanas como janela de entrada;
- 1 semana como saída;
- divisão temporal 85% treino e 15% teste;
- técnica *sliding window* para manter coerência temporal.

#### 3.5 Arquitetura do Modelo GRU

O modelo final contém:

- 1 camada GRU com unidades ajustáveis,
- Dropout para reduzir overfitting,
- Camada densa intermediária (*ReLU*),

- Saída linear (1 neurônio),
- Otimizador Adam,
- Função de perda MSE.

*Callbacks aplicados:*

- EarlyStopping,
- ReduceLROnPlateau,
- ModelCheckpoint.

### 3.6 Métricas de Avaliação

*Após retornar da escala logarítmica:*

- MSE (Mean Squared Error),
- RMSE (Root Mean Squared Error),
- MAE (Mean Absolute Error).

## 4. IMPLEMENTAÇÃO

### 4.1 Pipeline Computacional

O código implementa:

1. Carregamento e limpeza.
2. Criação das features.
3. Geração das janelas temporais.
4. Normalização (StandardScaler).
5. Treinamento da GRU.
6. Previsão no conjunto de teste.
7. Exportação de métricas e arquivos CSV.

### 4.2 Ferramentas utilizadas

- Python
- Pandas
- NumPy
- Scikit-learn
- TensorFlow/Keras
- Matplotlib
- Google Colab
- GitHub (repositório do projeto)

### 4.3 Fluxo da Solução

Coleta → Limpeza → Indexação → Engenharia de Atributos → EDA → Janelas Temporais → GRU → Avaliação → Conclusão

## 5. RESULTADOS

### 5.1 Métricas Obtidas

O modelo final (janela de 52 semanas) obteve:

Modelo	MSE	RMSE	MAE
GRU (52 semanas)	89 142.19	298.56	219.66

Os valores indicam aderência consistente entre previsto e observado.

### 5.2 Desempenho do Modelo

A GRU capturou adequadamente:

- tendência,
- sazonalidade,
- oscilações semanais,
- padrões epidemiológicos típicos.

Os maiores erros ocorreram em quedas abruptas, característica comum em séries de doenças infecciosas.

### 5.3 Discussão dos Resultados

- Os resíduos não apresentaram padrão sistemático.
- A distribuição dos erros foi concentrada próxima de zero.
- O modelo demonstrou estabilidade e capacidade de generalização.
- A janela de 52 semanas se mostrou a mais eficiente.

A previsão apresentou leve subestimação em valores muito altos, comportamento esperado em RNNs aplicadas a epidemias.

## 6. CONCLUSÃO E TRABALHOS FUTUROS

O projeto demonstrou que redes neurais GRU são eficazes para prever casos semanais de dengue no Brasil, especialmente quando combinadas com engenharia de atributos e janelas amplas de histórico.

A solução é reproduzível, estável e útil para vigilância epidemiológica, podendo apoiar gestores públicos na antecipação de surtos.

Trabalhos futuros incluem:

- expansão para previsão multi-step (4 semanas, 8 semanas);
- inclusão de dados socioambientais (densidade populacional, saneamento);
- uso de ensembles híbridos (GRU + Prophet + LightGBM);
- construção de dashboard interativo para prefeituras.

## Referências

- AWS. *What is RNN (Recurrent Neural Network?)* Disponível em: <https://aws.amazon.com/what-is/recurrent-neural-network>. Acesso em: 15 set. 2025.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time Series Analysis: Forecasting and Control*. 5. ed. Hoboken: Wiley, 2015.
- BRASIL. Ministério da Saúde. *Guia de Vigilância em Saúde*. 6. ed. Brasília: Ministério da Saúde, 2023.
- FUNDAÇÃO OSWALDO CRUZ. *InfoDengue: Sistema de Monitoramento e Alerta de Dengue*. Rio de Janeiro: Fiocruz, 2024. Disponível em: <https://info.dengue.mat.br>. Acesso em: 15 set. 2025.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge: MIT Press, 2016.
- HOTEZ, P. J. et al. *The neglected tropical diseases: the ancient afflictions of stigma and poverty*. PLoS Neglected Tropical Diseases, v. 1, n. 2, e119, 2006. DOI: <https://doi.org/10.1371/journal.pntd.0000119>.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3>. Acesso em: 15 set. 2025.
- IBM. *What are ARIMA models?* Disponível em: <https://www.ibm.com/topics/arima-model>. Acesso em: 15 set. 2025.
- IPEA. *Epidemiologia das doenças negligenciadas no Brasil e gastos federais com medicamentos*. Brasília: IPEA, 2011.
- JOHNSTON, G.; MOUSTAFA, A. *Advances in Machine Learning Approaches to Infectious Disease Surveillance*. Frontiers in Artificial Intelligence, v. 3, 2020. DOI: <https://doi.org/10.3389/frai.2020.00007>.
- LOURENÇO, J. et al. *InfoDengue: a nowcasting system for the surveillance of arboviruses in Brazil*. Transactions of the Royal Society of Tropical Medicine and Hygiene, v. 114, n. 9, p. 635–644, 2020.
- PAHO. *Integrated Control of Neglected Tropical Diseases*. Washington, D.C.: PAHO, 2023.
- SHAMSI, S. A. et al. *Machine learning-based dengue outbreak prediction: A case study of six Asian countries*. Computers in Biology and Medicine, v. 150, p. 106166, 2022. DOI: <https://doi.org/10.1016/j.combiomed.2022.106166>.
- SILVA, L. A.; PERES, S. M.; BOSCAROLI, C. *Introdução à Mineração de Dados com Aplicações em R*. São Paulo: LTC, 2016.
- SOUSA, A. R. S. et al. *Análise de séries temporais*. Porto Alegre: Grupo A, 2021.