

Previsão de Casos de Dengue em Capitais Brasileiras com Redes Neurais Espaço-Temporais em Grafos

Arthur V. C. Silva¹, Giovanni E. Zanardo¹,
Fabiola G. Nakamura¹, Eduardo F. Nakamura¹

¹Instituto de Computação (IComp) – Universidade Federal do Amazonas (UFAM)
Manaus – AM – Brasil

{arthur.vinicio, gio.zanardo, fabiola, nakamura}@icomp.ufam.edu.br

Abstract. *Dengue fever is a challenge for epidemiological surveillance in Brazil, requiring predictive tools capable of anticipating disease case numbers and supporting decision-making by public health managers. This study presents the development and training of a Spatio-Temporal Graph Convolutional Network (STGCN) model for forecasting weekly dengue cases in the 26 Brazilian state capitals and the Federal District, using data from the InfoDengue platform over the period from 2010 to 2025. The approach integrates epidemiological and climatic variables into a spatio-temporal representation, in which relationships between cities are modeled through an adjacency matrix constructed from the correlation between time series.*

Resumo. *A dengue é um desafio para a vigilância epidemiológica no Brasil, exigindo ferramentas preditivas capazes de antecipar o número de casos da doença e apoiar a tomada de decisão de gestores de saúde pública. Este trabalho apresenta a construção e treino de um modelo baseado em Redes Neurais Convolucionais Espaço-Temporais em Grafos (STGCN) para previsão de casos semanais de dengue nas 26 capitais brasileiras e no Distrito Federal, utilizando dados da plataforma InfoDengue no período de 2010 a 2025. A abordagem integra variáveis epidemiológicas e climáticas em uma representação espaço-temporal, na qual as relações entre cidades são modeladas por meio de uma matriz de adjacência construída a partir da correlação entre séries temporais.*

1. Introdução

A dengue é um desafio contínuo para a saúde pública global, especialmente em regiões tropicais e subtropicais como o Brasil, onde a incidência de casos impactam significativamente a morbidade e mortalidade populacional [Bomfim 2023]. Nesse contexto, aprimorar a capacidade preditiva de modelos epidemiológicos é fundamental para antecipar a dinâmica de transmissão e apoiar a gestão eficiente de recursos e a implementação de medidas de controle [Siabi et al. 2026].

Modelos baseados em aprendizado de máquina, como redes neurais LSTM, têm demonstrado bons resultados na previsão da incidência de dengue em relação a abordagens tradicionais ao incorporar séries temporais multivariadas, dados climáticos e componentes espaciais [Chen and Moraga 2025]. Nesse sentido, arquiteturas baseadas em grafos, como as *Spatio-Temporal Graph Convolutional Networks* (STGCN), destacam-se por

integrar simultaneamente dependências temporais e relações espaciais [Siabi et al. 2026]. Essas abordagens modelam cidades como nós de um grafo, cuja matriz de adjacência pode ser derivada da correlação de Pearson entre séries temporais, capturando padrões de interação entre localidades [Chen and Moraga 2025].

Este estudo em andamento desenvolve um modelo STGCN para previsão de casos semanais de dengue nas 26 capitais brasileiras e no Distrito Federal, usando dados da plataforma InfoDengue no período de 2010 a 2025. A abordagem integra variáveis epidemiológicas (casos de dengue) e climáticas (temperatura e umidade médias) em uma representação espaço-temporal, com grafo construído a partir da correlação entre as cidades, filtrado por limiar e normalizado por grau. O modelo, composto por blocos de convolução temporal com *Gated Linear Units* (GLU) e convolução espacial via redes convolucionais em grafos de Chebyshev como proposto por [Yu et al. 2018] e [Bhandari et al. 2024], é treinado para horizontes de 1 a 8 semanas, apresentando boa capacidade de capturar tendências de curto prazo, embora ainda com limitações na previsão de picos abruptos de incidência e o desempenho em horizontes mais longos.

2. Revisão da literatura

A predição de dengue no Brasil tem evoluído a partir de modelos clássicos de séries temporais e técnicas de aprendizado de máquina supervisionado. Métodos como *Auto-Regressive Integrated Moving Average* (ARIMA) e *Random Forest* são amplamente utilizados, enquanto redes neurais como *Long Short-Term Memory* (LSTM) têm apresentado melhor desempenho em previsões semanais, especialmente quando incorporam variáveis climáticas relevantes [Chen and Moraga 2025].

Apesar dos avanços na modelagem temporal, esses métodos apresentam limitações na captura de dependências espaciais. Nesse contexto, modelos baseados em grafos, como as STGCN, surgem ao integrar relações espaciais e temporais, representando cidades como nós e usando correlações para modelar interações [Siabi et al. 2026].

3. Metodologia

Esta sessão descreve o *pipeline* expresso na figura 1, adotado para a construção, treinamento e avaliação do modelo.

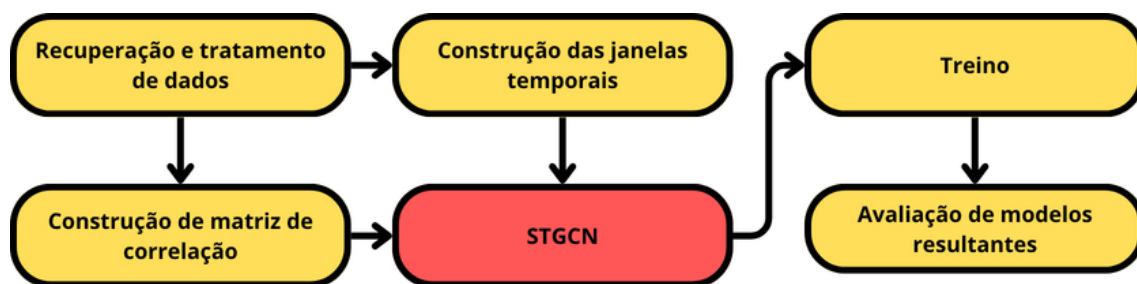


Figura 1. Metodologia

3.1. Conjunto de dados

Os dados utilizados foram obtidos na plataforma InfoDengue [Codeco et al. 2018], que disponibiliza registros epidemiológicos semanais para municípios brasileiros. Os dados são organizados em séries temporais semanais, com um arquivo por município. Cada série contém variáveis climáticas, sociais e de incidência da doença.

3.2. Pré-processamento

Inicialmente, foi realizada uma análise de disponibilidade das variáveis, com verificação da cobertura de dados por cidade. Essa etapa permitiu identificar inconsistências, como valores ausentes, tratados com interpolação temporal, ou séries incompletas, indicando a remoção de variáveis inadequadas.

Em seguida, as séries temporais das cidades alvo foram selecionadas e padronizadas conforme as especificações do experimento como o período correspondente dos dados a ser usado no treinamento, as variáveis a serem usadas e a garantia da ordenação cronológica em cada cidade. Posteriormente, realizou-se um alinhamento entre as cidades por meio de um eixo temporal comum, resultando em cada semana representar um estado global do sistema de forma sincronizada, formando um tensor tridimensional com as dimensões de tempo, cidade e variáveis, compatível com a arquitetura do modelo STGCN.

A matriz de adjacência (grafo) foi construída a partir da correlação de Pearson entre as séries temporais das cidades a fim de alinhar cidades com comportamento epidemiológico parecidos, utilizando apenas os dados de treino. A matriz de adjacência resultante foi filtrada por um limiar para manter apenas conexões relevantes, reduzindo a densidade do grafo. Em seguida, a matriz foi normalizada com base no grau dos nós, garantindo uma propagação de informação mais estável durante o treinamento do modelo.

3.3. Arquitetura do modelo

Para modelar a dinâmica espaço-temporal, foi utilizada uma arquitetura do tipo *Spatio-Temporal Graph Convolutional Network* (STGCN), ilustrada na Figura 2. A figura apresenta uma visão geral do fluxo do modelo, no qual janelas temporais e a matriz de adjacência alimentam blocos compostos por operações temporais e espaciais.

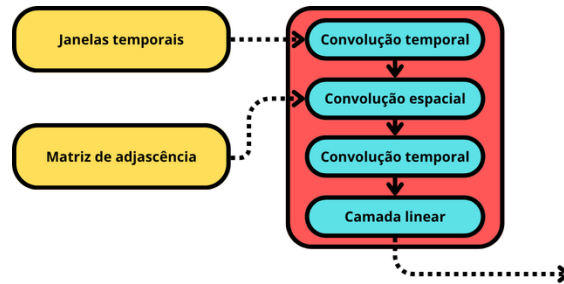


Figura 2. Arquitetura do modelo STGCN

Na implementação proposta, a entrada é um tensor $X \in R^{B \times C \times T \times V}$, representando séries temporais multivariadas para múltiplas cidades. Nesse contexto, B corresponde ao tamanho do *batch*, C ao número de variáveis (*features*), T ao comprimento da janela temporal de entrada e V ao número de cidades.

Cada bloco STGCN segue a sequência mostrada na figura: uma convolução temporal, uma convolução em grafo e uma nova etapa temporal. As convoluções temporais utilizam mecanismo de *Gated Linear Units* (GLU) conforme a arquitetura proposta por [Yu et al. 2018], permitindo capturar padrões locais no tempo enquanto filtra informações irrelevantes. A etapa espacial é implementada via redes convolucionais em grafos de Chebyshev, que realiza a agregação de informação entre cidades com base na

matriz de adjacência, considerando vizinhanças de até K saltos, conforme proposto por [Bhandari et al. 2024].

O modelo pode ser formado por um ou mais blocos, seguido de uma camada de projeção 1×1 , responsável por mapear as representações aprendidas pelo modelo para o horizonte de previsão, que é o número de passos futuros estimados pelo modelo. A saída final produz previsões no formato (B, V, H) , onde H representa o horizonte de previsão.

3.4. Treinamento

O modelo foi treinado usando treino supervisionado utilizando a função de perda de erro quadrático médio (MSE). A otimização foi feita com o algoritmo Adam, incluindo termo de regularização *weight decay*. Além disso, o treinamento foi feito em *batches*, com embaralhamento somente no conjunto de treino, e monitorado por meio da perda no conjunto de validação. Para evitar sobreajuste, foi utilizado o mecanismo de parada antecipada *early stopping*, interrompendo o treinamento quando não houve melhoria na perda de validação por um número fixo de épocas consecutivas. O modelo final corresponde àquele com melhor desempenho no conjunto de validação.

3.5. Avaliação

O desempenho do STGCN foi avaliado no conjunto de teste, com dados desnormalizados, para horizontes de previsão de 1 a 8 semanas, usando as métricas *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Scaled Error* (MASE) e *Root Mean Squared Scaled Error* (RMSSE) [Hyndman and Koehler 2006]. Os resultados foram analisados por horizonte e em média global. Além disso, o modelo foi comparado a uma arquitetura LSTM usando as mesmas variáveis e cidades de entrada. Por fim, agregou-se previsões *multistep* para simular projeções contínuas do modelo.

4. Resultados

Neste trabalho, foram consideradas três variáveis: número de casos de dengue (também servindo como alvo), temperatura média e umidade média. As séries temporais foram organizadas em janelas de 50 semanas, com divisão temporal de 70% para treino, 15% para validação e 15% para teste. Foi usado normalização robusta por nó (cidade). O grafo foi construído com a correlação de Pearson usando limiar de 0.3. O STGCN foi configurado com camadas ocultas de 16, 32 e 64 canais e treinado com *batches* de tamanho 128.

Os resultados da Tabela 1 mostram comportamentos distintos entre os modelos ao longo dos horizontes de previsão. Em horizontes curtos (1 e 2 semanas), o STGCN apresentou melhor desempenho que o LSTM em todas as métricas analisadas, sugerindo que a modelagem explícita das dependências espaço-temporais entre cidades contribui positivamente para previsões epidemiológicas de curto prazo.

Porém, o desempenho do STGCN degrada conforme o horizonte aumenta. No horizonte de 4 semanas, apesar do RMSE inferior ao LSTM, os valores de MAE e métricas escaladas tornam-se superiores. No horizonte de 8 semanas, o modelo alcança MAE de 661.06 e RMSSE de 26.13, indicando perda significativa de capacidade preditiva.

Do ponto de vista epidemiológico, os resultados sugerem que relações espaciais entre cidades são mais relevantes no curto prazo, quando padrões recentes de transmissão

Tabela 1. Comparação entre os modelos LSTM e STGCN para diferentes horizontes de previsão.

Modelo	Horizonte	MAE	RMSE	MASE	RMSSE
LSTM	1	246.26	1395.99	16.65	14.72
STGCN	1	119.74	408.87	7.26	6.35
LSTM	2	248.40	1251.36	16.70	14.01
STGCN	2	196.51	599.69	11.23	8.74
LSTM	4	257.22	1181.45	17.27	14.13
STGCN	4	357.19	984.62	19.55	15.24
LSTM	8	307.73	1345.81	20.80	16.84
STGCN	8	661.06	1656.07	34.75	26.13

ainda permanecem correlacionados. Em horizontes maiores, a variabilidade temporal da dengue e aumentos abruptos de casos dificultam a generalização do modelo.

O crescimento contínuo de MASE e RMSSE sugere que o STGCN tende a fazer previsões mais conservadoras em horizontes longos, aproximando-se de padrões médios das séries e reduzindo a capacidade de detectar aumento de casos abruptos.

Os resultados indicam potencial de arquiteturas espaço-temporais baseadas em grafos para aplicações de vigilância epidemiológica de curto prazo.

A figura 3 apresenta a comparação entre valores reais e previstos para um determinado conjunto, incluindo previsões de curto e longo prazo.

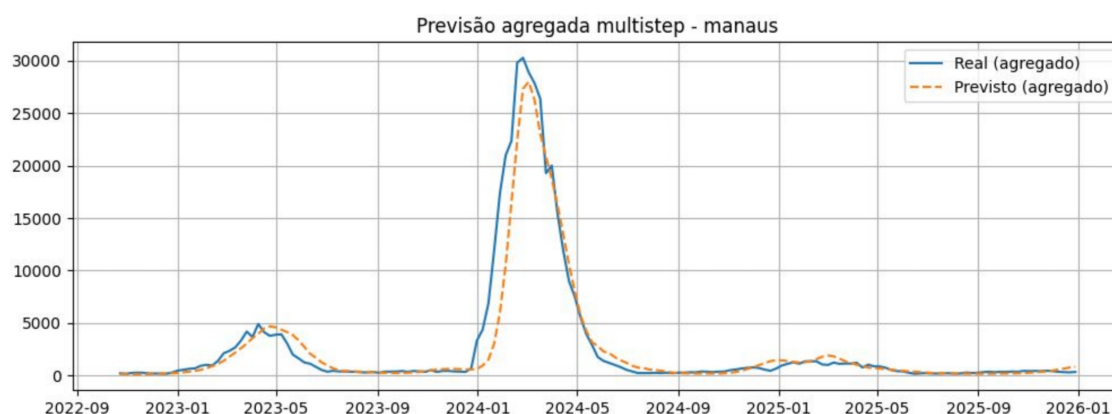


Figura 3. Fonte: Próprio autor

A análise visual mostra que o modelo é capaz de capturar o comportamento geral da série temporal, onde as previsões acompanham de forma mais próxima os valores reais.

No entanto, duas limitações principais são observadas. A primeira é a dificuldade em estimar corretamente a magnitude do aumento do número de casos, frequentemente resultando em sub ou superestimação. A segunda é a presença de atraso (*lag*) nas previsões, onde o modelo reage às mudanças na série com algumas semanas de defasagem.

5. Conclusão

Este trabalho apresentou o desenvolvimento inicial de um modelo STGCN para previsão de casos semanais de dengue nas capitais brasileiras e no Distrito Federal, explorando dependências temporais e espaciais entre cidades simultaneamente.

Os resultados mostram que o modelo apresenta desempenho competitivo em horizontes curtos (uma e duas semanas), superando o LSTM em métricas como RMSE e RMSSE. Porém, observou-se degradação progressiva do desempenho conforme o horizonte aumenta, além da presença de atraso (*lag*) nas previsões e dificuldade em capturar aumentos abruptos de casos. As métricas escaladas também indicam limitações da abordagem em horizontes longos, mostrando que a dinâmica epidemiológica da dengue ainda representa um desafio relevante para modelos espaço-temporais baseados em grafos.

Como trabalhos futuros, pretende-se investigar estratégias diferentes para construção da matriz de adjacência, uma vez que a correlação de Pearson utilizada considera apenas relações lineares e simultâneas entre cidades, podendo contribuir para o atraso (*lag*) observado nas previsões. Nesse contexto, serão avaliadas abordagens baseadas em mobilidade humana, similaridade climática e correlações temporais defasadas.

Por fim, pretende-se explorar modificações arquiteturais e a inclusão de novas variáveis epidemiológicas e climáticas, buscando melhorar o desempenho em horizontes longos e a capacidade do modelo em representar variações abruptas nos casos e fazer comparações com arquiteturas espaço-temporais alternativas em avaliações futuras.

Referências

- Bhandari, H., Pandey, H., Pandeya, Y. R., and Jha, K. (2024). Dynamic insights into dengue: Leveraging spatio-temporal graph convolution networks. *Journal of Nepal Mathematical Society*, 7:30–39.
- Bomfim, R. P. (2023). *Mobilidade humana em grandes cidades a nível de bairro e seus impactos na epidemiologia da dengue em Fortaleza*. Tese (doutorado em informática aplicada), Universidade de Fortaleza, Fortaleza. Acesso em: 28 abr. 2026.
- Chen, X. and Moraga, P. (2025). Dengue forecasting and outbreak detection in brazil using lstm: integrating human mobility and climate factors. *medRxiv*.
- Codeco, C., Coelho, F., Cruz, O., Oliveira, S., Castro, T., and Bastos, L. (2018). Infodengue: A nowcasting system for the surveillance of arboviruses in brazil. *Revue d'Épidémiologie et de Santé Publique*, 66:S386. European Congress of Epidemiology “Crises, epidemiological transitions and the role of epidemiologists”.
- Hyndman, R. and Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688.
- Siabi, N., Son, R., Thomas, M., Irrgang, C., and Saynisch-Wagner, J. (2026). Innovative application of a traffic-prediction spatio-temporal graph convolutional network for dengue disease forecasting. *Scientific Reports*, 16(1):2594.
- Yu, B., Yin, H., and Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-2018*, page 3634–3640. International Joint Conferences on Artificial Intelligence Organization.