



**UNIFOR**

**FUNDAÇÃO EDSON QUEIROZ**

**UNIVERSIDADE DE FORTALEZA- UNIFOR**

**VICE-REITORIA DE ENSINO DE GRADUAÇÃO E PÓS-GRADUAÇÃO - VRE**

**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA**

**DOUTORADO EM INFORMÁTICA APLICADA**

**RAFAEL PONTES BOMFIM**

**MOBILIDADE HUMANA EM GRANDES CIDADES A NÍVEL DE BAIRRO E SEUS  
IMPACTOS NA EPIDEMIOLOGIA DA DENGUE EM FORTALEZA**

**FORTALEZA-CE**

**Agosto, 2023**

RAFAEL PONTES BOMFIM

MOBILIDADE HUMANA EM GRANDES CIDADES A NÍVEL DE BAIRRO E SEUS  
IMPACTOS NA EPIDEMIOLOGIA DA DENGUE EM FORTALEZA

Tese apresentada ao Curso de Doutorado em Informática Aplicada do Programa de Pós-Graduação em Informática Aplicada do VICE-REITORIA DE ENSINO DE GRADUAÇÃO E PÓS-GRADUAÇÃO - VRE da Universidade de Fortaleza- UNIFOR, como requisito parcial à obtenção do título de doutor em Informática Aplicada. Área de Concentração: Informática Aplicada

Orientador: Prof. João José Vasco Peixoto Furtado, Docteur.

FORTALEZA-CE

Agosto, 2023

Ficha catalográfica da obra elaborada pelo autor através do programa de geração automática da Biblioteca Central da Universidade de Fortaleza

---

Bomfim, Rafael Pontes.

Mobilidade humana em grandes cidades a nível de bairro e seus impactos na epidemiologia da dengue em Fortaleza / Rafael Pontes Bomfim. - 2023  
114 f.

Tese (Doutorado) - Universidade de Fortaleza. Programa de Doutorado Em Informática Aplicada, Fortaleza, 2023.

Orientação: João José Vasco Peixoto Furtado.

Coorientação: Jorge Luiz Bezerra de Araújo.

I. Furtado, João José Vasco Peixoto. II. de Araújo, Jorge Luiz Bezerra. III. Título.

---

RAFAEL PONTES BOMFIM

MOBILIDADE HUMANA EM GRANDES CIDADES A NÍVEL DE BAIRRO E SEUS  
IMPACTOS NA EPIDEMIOLOGIA DA DENGUE EM FORTALEZA

Tese apresentada ao Curso de Doutorado em Informática Aplicada do Programa de Pós-Graduação em Informática Aplicada do VICE-REITORIA DE ENSINO DE GRADUAÇÃO E PÓS-GRADUAÇÃO - VRE da Universidade de Fortaleza- UNIFOR, como requisito parcial à obtenção do título de doutor em Informática Aplicada. Área de Concentração: Informática Aplicada

Aprovada em: 17 de agosto de 2023

BANCA EXAMINADORA

---

Prof. João José Vasco Peixoto Furtado, Docteur.  
(Orientador)  
Universidade de Fortaleza - UNIFOR

---

Prof. Jorge Luiz Bezerra de Araújo, D.Sc.  
(Coorientador)  
Universidade de Fortaleza - UNIFOR

---

Prof. José Soares de Andrade Júnior, D.Sc.  
Universidade Federal do Ceará - UFC

---

Prof. Antônio Silva Lima Neto, D.Sc.  
Universidade de Fortaleza - UNIFOR

---

Prof. José Antônio Fernandes de Macêdo, D.Sc.  
Universidade Federal do Ceará - UFC

---

Prof. Carlos de Oliveira Caminha Neto, D.Sc.  
Universidade Federal do Ceará - UFC

## RESUMO

A dengue é um problema significativo de saúde pública em áreas urbanas, onde a densidade populacional e a mobilidade das pessoas contribuem para sua transmissão. O movimento das pessoas, particularmente através do transporte público de ônibus, desempenha um papel crucial na propagação da dengue. No entanto, a intrincada interação entre esses aspectos ainda é pouco estudada. Esta pesquisa visa explorar e entender essa interação e alavancar conhecimento para aumentar a precisão dos modelos preditivos para prever casos de dengue, em especial prever o mais rápido possível quando um ano será epidêmico.

O estudo tem como foco Fortaleza, uma cidade com aproximadamente 2,6 milhões de habitantes no nordeste do Brasil que, nas últimas décadas, vem sofrendo com a transmissão endêmica da dengue, intercalada com epidemias maiores. Esta tese fornece uma caracterização dos casos de dengue em Fortaleza de 2007 a 2020, bem como dados de mobilidade humana de 2015 e 2020. Os dados de casos de dengue, incluindo o número de casos e sua distribuição espacial, são obtidos das autoridades de saúde locais ou sistemas de vigilância em saúde. Os padrões de movimento das pessoas são derivados de dados de transporte público de ônibus, como contagem de passageiros, rotas e paradas de ônibus. Esses conjuntos de dados são integrados e novos modelos de redes neurais são desenvolvidos para prever o comportamento da doença considerando diferentes métricas. Uma arquitetura de rede neural recorrente foi proposta para aprender o impacto de cada bairro na transmissão de casos de dengue na cidade de Fortaleza como um todo. Versões mais elaboradas do modelo levaram em conta a população residente de cada bairro como informação heurística de forma a incrementar a precisão do modelo nas previsões para cada bairro. Essa arquitetura proposta teve um desempenho melhor do que outras abordagens usando apenas informações isoladas sobre incidência ou exclusivamente número de casos. Ademais, uma análise sobre o quanto os dados de mobilidade dentro da cidade poderiam aumentar a acurácia de modelos preditivos foi realizada. Um sistema de previsão de casos de dengue em Fortaleza usando redes neurais recorrentes do tipo LSTM foi proposto e comparado com um modelo tradicional mecanicista de transmissão da doença. Os resultados mostram que tanto as redes neurais artificiais quanto os modelos mecânicos podem prever com precisão os casos de dengue e que os dados de mobilidade humana podem melhorar substancialmente o desempenho de ambos os sistemas em prever antecipadamente os picos de caso de Dengue.

Os resultados da pesquisa destacam uma associação significativa entre os casos de dengue e a movimentação de pessoas no transporte público de ônibus. A análise espacial, particularmente no

ano de 2020, quando a epidemia de COVID-19 obrigou a proibição completa de deslocamentos na cidade (*lockdown*), revela o impacto da movimentação de pessoas na taxa de contágio da Dengue. Além disso, a incorporação de dados de transporte de ônibus nos modelos preditivos melhora substancialmente a precisão das previsões indicando que os modelos capturam a intrincada relação entre os padrões de mobilidade das pessoas e a dinâmica da transmissão da dengue. O conhecimento da interação entre dados de mobilidade em uma grande cidade e modelos preditivos de Dengue que essa tese desvela tem implicações importantes para a previsão de doenças e intervenções proativas. Abre-se a possibilidade para estratégias mais eficazes no combate a surtos de dengue e minimização de seu impacto nas populações urbanas. A maior precisão dos modelos permite a identificação precoce de potenciais focos de dengue e facilita intervenções de saúde pública direcionadas. Ao considerar a influência do transporte público de ônibus na transmissão da dengue, as autoridades podem implementar medidas oportunas de controle de vetores, alocar recursos de forma eficiente e desenvolver estratégias eficazes de prevenção.

**Palavras-chave:** dengue, previsão, rede neural, mobilidade humana

## ABSTRACT

Dengue is a significant public health problem in urban areas, where population density and people's mobility contribute to its transmission. The movement of people, particularly through public bus transportation, plays a crucial role in the spread of dengue. However, the intricate interaction between these aspects is still poorly understood. This research aims to explore and understand this interaction, leveraging knowledge to improve the accuracy of predictive models for forecasting dengue cases, especially predicting as early as possible when a year will be epidemic.

The study focuses on Fortaleza, a city with approximately 2.6 million inhabitants in northeastern Brazil, which has been experiencing endemic dengue transmission over the last decades, interspersed with larger epidemics. This thesis provides a characterization of dengue cases in Fortaleza from 2007 to 2020, along with human mobility data from 2015 and 2020. Dengue case data, including the number of cases and their spatial distribution, are obtained from local health authorities or health surveillance systems. People's movement patterns are derived from public bus transportation data, such as passenger counts, routes, and bus stops. These datasets are integrated, and new neural network models are developed to predict the disease's behavior considering different metrics.

A recurrent neural network architecture was proposed to learn the impact of each neighborhood on the transmission of dengue cases in the city of Fortaleza as a whole. More elaborate versions of the model took into account the resident population of each neighborhood as heuristic information to enhance the model's accuracy in predictions for each neighborhood. This proposed architecture outperformed other approaches using only isolated information on incidence or solely the number of cases. Furthermore, an analysis of how much mobility data within the city could increase the accuracy of predictive models was conducted.

A dengue case forecasting system in Fortaleza using LSTM-type recurrent neural networks was proposed and compared with a traditional mechanistic model of disease transmission. The results show that both artificial neural networks and mechanistic models can accurately predict dengue cases and that human mobility data can substantially improve the performance of both systems in early forecasting of Dengue case peaks.

The research results highlight a significant association between dengue cases and people's movement in public bus transportation. Spatial analysis, particularly in the year 2020, when the COVID-19 epidemic forced a complete ban on travel in the city (lockdown), reveals the

impact of people's movement on the Dengue transmission rate. Furthermore, the incorporation of bus transportation data into predictive models substantially improves the accuracy of forecasts, indicating that the models capture the intricate relationship between people's mobility patterns and the dynamics of dengue transmission.

The knowledge of the interaction between mobility data in a large city and predictive models of Dengue revealed by this thesis has important implications for disease forecasting and proactive interventions. It opens up the possibility for more effective strategies in combating dengue outbreaks and minimizing their impact on urban populations. The increased accuracy of the models allows for early identification of potential dengue hotspots and facilitates targeted public health interventions. By considering the influence of public bus transportation on dengue transmission, authorities can implement timely vector control measures, allocate resources efficiently, and develop effective prevention strategies.

**Keywords:** dengue, forecasting, neural network, human mobility



## LISTA DE ILUSTRAÇÕES

- Figura 1** – O ciclo de transmissão do vírus da dengue: As origens silvestres do vírus da dengue e a 'zona de emergência', onde os ciclos silvestres entram em contato com populações humanas em áreas rurais da África Ocidental e do Sudeste Asiático. Além disso, o vírus da dengue pode persistir nas populações de mosquitos por transmissão transovariana (TOT), na qual mosquitos infectados pelo vírus transferem o vírus para seus ovos (isso foi demonstrado em algumas espécies, mas não em todas). . . . . 29
- Figura 2** – A figura (a) mostra exemplo de classificação. O modelo supervisionado classifica os clientes como de alto risco ou baixo risco de acordo com dois atributos: idade e salário. A figura (b) mostra exemplo de agrupamento. O modelo agrupo as pessoas de acordo com os dois atributos. . . . . 34
- Figura 3** – Neurônio e perceptron: na figura (a) é ilustrada a transformação linear realizada por um neurônio, onde  $\hat{y}$  é a saída do neurônio,  $x$  é a entrada e  $w$  é o parâmetro a ser treinado. Na figura (b), é mostrado um perceptron com camadas de entrada, escondida e saída. Neste exemplo, existe apenas uma camada escondida, mas um perceptron pode ter várias camadas escondidas. . . . . 34
- Figura 4** – Fluxo temporal da rede neural recorrente.  $U, V$  e  $W$  são parâmetros da rede.  $X$  e  $Y$  são a entrada e a predição do modelo, respectivamente.  $S$  é o estado do modelo no tempo  $t$ . . . . . 35
- Figura 5** – Uma unidade LSTM possui uma estrutura em cadeia que contém quatro redes neurais e diferentes blocos de memória chamados células. A informação é retida pelas células e as manipulações da memória são feitas através das portas de entrada ( $i_t$ ), esquecer ( $f_t$ ) e saída ( $o_t$ ). A memória de longo prazo é geralmente chamada de estado da célula em que as informações de intervalos anteriores são armazenadas. . . . . 37
- Figura 6** – Arquitetura encoder-decoder. . . . . 38
- Figura 7** – Temporal Graph Neural Network. . . . . 40
- Figura 8** – Fluxo da população entre os três estados do modelo SIR. . . . . 40
- Figura 9** – Evolução da curva de pessoas suscetíveis (azul), infectadas (laranja) e recuperadas (verde). . . . . 42

<b>Figura 10 – Fluxo da população entre os três estados do modelo Suscetível, Exposto, Infectado e Removido. . . . .</b>	42
<b>Figura 11 – Evolução da curva de pessoas suscetíveis (azul), infectadas (laranja), recuperadas (verde) e expostas (vermelha). . . . .</b>	44
<b>Figura 12 – Visão geral dos registro de dengue na cidade de Fortaleza. É apresentado a série temporal semanal do número de casos de dengue observados na cidade de Fortaleza no intervalo de 2007 a 2020. No gráfico interno destacamos a evolução temporal durante o período de 2020. Em 2020 restrições a mobilidade urbana nomeadas como isolamento social e <i>lockdown</i> foram implementadas para mitigar o processo de proliferação do COVID-19 gerando assim um perfil atípico de na série temporal. . . . .</b>	47
<b>Figura 13 – Correlograma do número total de casos de dengue para cada bairro em diversos anos. Nós calculamos o coeficiente de correlação de Spearman <math>\rho</math> (Equação 3.1) entre os pares de anos compreendidos no intervalo [2011,2000]. O cálculo de <math>\rho</math> consiste em comparar o ranqueamento do número total de casos de dengue em cada bairro para um par de anos analisados. Assim, a correlação será alta quando observações tiverem uma classificação semelhante. As análises demonstram que há, efetivamente, uma tendência em manter a ordem de infecção por bairros. A distribuição dos valores de correlação, <math>g(\rho)</math>, pode ser vista através do <i>inset</i> da figura onde <math>\langle \rho \rangle = 0.77</math> identifica a média de todos os pares de anos analisados. . . . .</b>	48
<b>Figura 14 – Distribuições dos valores correlação de Pearson (Equação 3.2) entre as séries temporais para cada bairro em diversos anos. Em (a) é apresenta a distribuição para anos epidêmicos <math>Y_{epi} = (2008, 2011, 2012, 2015, 2016, 2017)</math> enquanto (b) apresenta resultados para anos não-epidêmicos <math>Y_{non-epi} = (2007, 2009, 2010, 2013, 2014, 2019, 2020)</math>. De modo geral, anos epidêmicos geram séries síncronas e semelhantes. . . . .</b>	49

- Figura 15** – Em (a), é mostrado os o número total de casos observados na cidade de Fortaleza em diferentes momentos de contágio. Os anos de 2011, 2012, 2015, 2016 e 2017 são chamados de epidêmicos, pois apresentam uma grande quantidade de casos confirmados. Os anos restantes são considerados não epidêmicos. Em (b), analisamos o número total de casos apenas para as primeiras 10 semanas de infecção por dengue, o período logo antes da aplicação das restrições em 2020. Durante este período, embora 2020 tenha apresentado um padrão de contaminação semelhante aos anos epidêmicos, sugerindo que altas taxas de contágio possam ter ocorrido nas semanas restantes do ano. No entanto, essa estimativa não foi confirmada e, no final do ano, 2020 foi caracterizado como um ano não epidêmico. . . . . 51
- Figura 16** – Relação exponencial entre a incidência de casos  $\left(\frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)$  e o tempo de chegada ( $T_a^i$ ) do vírus da dengue para cada bairro na cidade de Fortaleza. Em (a)-(j), são apresentadas as relações  $\left(T_a^i \times \frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)$ . As linhas tracejadas representam ajustes pela Equação 3.3. As figuras com o quadro em vermelho representam anos epidêmicos. . . . . 52
- Figura 17** – Arquitetura isolada de rede neural - o modelo  $RNN_i$ . A arquitetura isolada prevê, com uma semana de antecedência, os casos de dengue para um bairro específico com base nas últimos 5 semanas de casos de dengue e em um vetor de índice que rotula cada semana de entrada. A camada de *embedding* aprende uma representação para o rótulo de cada semana, que é concatenado com os dados de dengue. Assim, a evolução dos casos de dengue é aprendida por meio dos dados concatenados e, em seguida, a camada densa gera a previsão para a semana seguinte. Os valores entre parênteses no *dropout* representam a fração de neurônios a serem ignorados em cada etapa de treinamento. Os valores entre colchetes representam as dimensões de entrada e saída de dados para cada camada. . . . . 73

**Figura 18** – O fluxo semanal de pessoas entre bairros. As setas amarelas desenhadas no mapa destacam um exemplo de fluxo entre 7 bairros.  $flowrate_{4,5,1}$  representa o fluxo do bairro 4 para o bairro 5 na semana 1. O modelo completo inclui taxas de fluxo para e de cada um dos 119 bairros da cidade para cada semana do período. . . . . 74

**Figura 19** – Arquitetura de rede neural artificial para o modelo RNNc com dados de mobilidade humana. A arquitetura assume a entrada de três conjuntos de dados: uma série temporal de cinco semanas de 119 bairros, um vetor com um identificador para cada semana e uma representação vetorial (embedding) descrevendo as informações de transporte para cada bairro. Os valores entre parênteses representam a dimensão de cada conjunto de dados. A dimensão das entradas são: [5, 119], representando a dimensão da matriz contendo cinco semanas de casos de dengue para todos os 119 bairros; [5, 1], representando um vetor com os índices para cada semana; e [5, 1785], representando 5 semanas e o embedding de tamanho 15 que representa o fluxo entre os 119 bairros ( $15 \times 119 = 1785$ ). A dimensão da camada oculta *concatenate* é [5, 271], representando a concatenação das colunas das três camadas de entrada. Uma pilha de LSTMs (a figura mostra as LSTMs desenroladas) é alimentada com os resultados da concatenação e cada camada da pilha propaga informações (dimensão 300) para a próxima camada. A última camada da pilha é totalmente conectada com uma camada densa (camada de saída) para prever o caso de dengue de cada um dos 119 bairros. . . . . 75

<b>Figura 20</b> – Para todos os gráficos, o eixo x indica a semana em que a previsão é gerada. O eixo y mostra os valores para precisão, revocação e F-score. Por exemplo, em (a), o valor de y para $x = 12$ (tempo de previsão = 12) mostra a precisão calculada a partir das previsões feitas a partir da semana 12 até a última semana do ano (semana 52). (a) a precisão da classificação para temporadas com casos totais de dengue menores ou iguais a 200, (b) a revocação da classificação para temporadas com casos totais de dengue menores ou iguais a 200, (c) F-score da classificação para temporadas com casos de dengue maiores que 200 (d) a precisão da classificação para temporadas com casos totais de dengue maiores que 200, (e) a revocação da classificação para temporadas com casos totais de dengue maiores que 200 e (f) F-score da classificação para temporadas com casos totais de dengue menores ou iguais a 200. . . . .	76
<b>Figura 21</b> – Erro médio absoluto (MAE) para a intensidade de pico (Eq. 4.10) e o momento do pico (Eq. 4.11) para surtos intensos ( <i>dengue</i> ; $casos \geq 200$ ). (a) MAE entre a incidência real e prevista de dengue nas semanas com maior incidência de dengue na temporada. (b) MAE entre o índice da semana real e previsto nas semanas com maior incidência de dengue na temporada. . . . .	77
<b>Figura 22</b> – Erro médio quadrático (RMSE) entre as séries temporais previstas e reais da incidência de dengue para surtos intensos ( <i>dengue</i> ; $casos \geq 200$ ). ARIMAI, RNNi e MECHi são os resultados para modelos que usam apenas dados locais de dengue, os modelos base, e ARIMAc, RNNc e MECHc são os resultados para as previsões geradas incluindo dados de mobilidade humana. . . . .	77
<b>Figura 23</b> – <i>Log score</i> para intensidade de pico (a), tempo do pico (b) e para a série temporal como um todo (c) para grandes surtos de casos de dengue com mais de 200 casos no ano. . . . .	78
<b>Figura 24</b> – Resultados de log-score para intensidade de pico (a), tempo do pico (b) e série temporal de dengue (c) para surtos com um número total de casos de dengue maior que 200. . . . .	79

- Figura 25** – Resultados de log-score para intensidade de pico (a), tempo de pico (b) e série temporal de dengue (c) para surtos com um número total de casos de dengue maior que 200. . . . . 80
- Figura 26** – A relação entre a quantidade total de casos observados no final de cada ano  $C_T^{(i)}$  em cada bairro  $i$  como função da população  $N_i$ . A linha preta tracejada é um guia para os olhos. . . . . 81
- Figura 27** – O comportamento da Equação 4.16 em função do número acumulado de casos ( $C(t)$ ) para dois cenários:  $L = C_T$  e  $L = N$ , para os anos de 2015, 2016, 2019 e 2020 (a-d, respectivamente). O eixo x foi normalizado em relação ao número total de casos observados ao final de cada período ( $C_T$ ). Adicionalmente, uma linha vertical tracejada indica o momento de máximo de contaminação semanal ( $C(t = t_{cri}) = C_T/2$ ). Antes de atingir o pico de infecção ( $t < t_{cri}$ ), ambas as curvas  $h(L = C_T)$  e  $h(L = N)$  apresentam taxas de crescimento semelhantes. Observei que  $\frac{dh(L=C_T)}{dC} \approx \frac{dh(L=N)}{dC}$  (representado pelas setas azuis), então essa similaridade implica que a população local pode atuar como regulador para a taxa de crescimento dos casos. . . . . 82
- Figura 28** – As duas redes neurais usadas neste artigo. Entre colchetes estão as dimensões dos dados passados entre cada camada da rede neural. O número de neurônios (400) da LSTM e o valor de dropout (0.2) apresentados na figura representam os valores ótimos encontrados pelo *Grid Search*. Além desses valores a rede neural foi treinada com um *epoch* de 300 e *batch size* de 64. Em (a) é apresentada a arquitetura da primeira rede neural, que recebe como entrada os casos de dengue ( $W^{(i)}(t)$ ), a incidência de casos de dengue ( $w^{(i)}(t)$ ) ou apenas os casos de dengue transformados por *ICF* (conforme mostrado na Equação 4.17). Em (b) é apresentada a arquitetura da segunda rede neural, que recebe como entrada tanto a série semanal de casos de dengue ( $W^{(i)}(t)$ ) quanto suas modificações através do *ICF*. . . . . 83
- Figura 29** – Séries temporais reais (Dado) e previstas para os 119 bairros acumulados semanalmente para os anos de 2015, 2016, 2019 e 2020, considerando as quatro variações de teste (*ICF + RE*, *RE*, *INC* e *ICF*). . . . . 84

- Figura 30** – (a) Root Mean Squared Log Error (*RMSLE*) entre dados preditos e reais (Equação 4.21), (b) Mean Absolute Log Error (*MALE*) entre o valor de casos de dengue do pico predito e real (Equação 4.19), (b) Mean Absolute Error (*MAE*) entre o índice real e predito da semana que ocorre o pico de casos de dengue (Equação 4.20) e (d) a correlação de Pearson entre a série temporal predita e real (Equação 3.2). . . . . 85
- Figura 31** – Em (a) Comparação de  $F_1$  entre testes *RE* e *ICF + RE* para os anos de 2015, 2016, 2019 e 2020 com a predição iniciando na semana 9. Em (b), (c), (d) e (e) mostram para os anos de 2015, 2016, 2019 e 2020, respectivamente, a comparação entre os valores de  $F_1^{(RE)}$  e  $F_1^{(ICF+RE)}$ , medido pela Equação 4.24 dos *RE* e *ICF + RE*, respectivamente. *Week* representa a semana do começo da predição que foi variada de 9 a 12. O tamanho do ponto representa o valor de  $C_{cri}$  utilizado no experimento com  $C_{cri} \in [10, 200]$ . A estrela vermelha identifica o experimento feito na semana 9 (Semana = 9) com  $C_{cri} = 100$ . . . . . 86
- Figura 32** – Ilustração da arquitetura de rede neural. A arquitetura tem como entrada uma matriz de tamanho [119,52], contendo as séries temporais de casos semanais de dengue de tamanho 52 para cada um dos 119 bairros. A dependência temporal e espacial entre esses valores é aprendida na camada LSTM com 150 neurônios, e a previsão para os 119 bairros com um passo à frente é feita na camada Dense com 119 neurônios. É aplicado um dropout de 10% antes das camadas LSTM e Dense para evitar o overfitting. . . . . 88
- Figura 33** – Série temporal do número de casos de dengue (reais e previstos) para a cidade de Fortaleza nos anos de 2020 (a), 2012 (b) e 2013 (c). As curvas azuis representam a série temporal real, e a curva vermelha representa a média dos casos totais das 5 previsões feitas para cada ano, juntamente com o intervalo de confiança de 90% (C.I. 90%). Em 2012, houve 38.448 casos de dengue, sendo que o modelo subestimou o número de casos com uma média de 24.139,90. Em 2013, houve 8.687 casos, e o valor médio das previsões foi de 5.794 casos. No total, houve 7.720 casos ao longo de 2020. A série temporal prevista pelo modelo de rede neural é superestimada, prevendo um ano com 27.792 casos. . . . . 90

**Figura 34 – Previsão do tempo efetivo de transmissão da dengue em 2020. Utilizando um modelo de rede neural, foi possível estimar o número total de casos de dengue para cada bairro  $i$  da cidade de Fortaleza em 2020, em um cenário hipotético em que as restrições de mobilidade não tivessem sido implementadas. Essa medida é observada no final do ano e armazenada na variável  $Pred^{(i)}$ . Observamos a série temporal que foi prevista para cada bairro durante o ano de 2020 e calculamos o tempo de chegada da doença prevista pela rede ( $T_a^{(i)-Pred}$ ). A relação  $T_a^{(i)-Pred} \times \left( \frac{Pred_T^{(i)} \times 10^4}{N^{(i)}} \right)$  segue um comportamento exponencial na forma da Equação 3.3. Assim, é possível estimar o tempo efetivo de transmissão da doença, onde  $\tau \approx 71$  dias. Portanto, nossas previsões indicam que em 2020 a dengue contaminaria os bairros de Fortaleza de forma mais rápida se não houvesse a implementação de medidas restritivas de mobilidade urbana.** 91

**Figura 35 – Impacto das medidas restritivas no número de casos de dengue. Aqui, observamos a quantidade de casos previstos ( $Pred^{(i)}$ ) para cada bairro na cidade de Fortaleza pelo modelo de rede neural em um cenário hipotético em que nenhuma medida de restrição de mobilidade foi implementada. Quantificamos o impacto das medidas restritivas como a diferença entre o número total de casos previstos pelo modelo de rede neural em relação ao número total de casos observados (veja a Equação 5.1). A relação entre  $T_a^{(i)} \times \Delta^{(i)}$  indica que bairros inicialmente infectados experimentaram maiores reduções no número total de casos devido à implementação de medidas restritivas. Curiosamente, as OLs iniciais são geralmente conhecidas como geradoras de grandes quantidades de viagens de saída ( $F^{(i)}$ ) durante o período imediatamente antes das restrições de mobilidade. A combinação dessas observações sugere que a mobilidade urbana é importante para o processo de disseminação da dengue. As linhas tracejadas indicam ajustes lineares como um guia a seguir.** . . . . . 92



<b>Figura 36 – E relação entre a média dos coeficientes R, representado por <math>\rho</math>, e a incidência de casos <math>\left(\frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)</math>. O coeficiente R é extraído pelo <i>fit</i> da relação <math>\left(T_a^i \times \frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)</math>. Observa-se um mínimo local para incidência sendo 7</b>	104
<b>Figura 37 – Resultados para previsões com períodos e bairros com mais de 100 casos de dengue.</b>	106
<b>Figura 38 – Resultados para previsões com períodos e bairros com mais de 300 casos de dengue.</b>	107
<b>Figura 39 – Exemplo de previsões com modelo RNNc para 8 bairros começando a partir da semana 11.</b>	108
<b>Figura 40 – Exemplo de previsões com modelo MECHc para 8 bairros começando a partir da semana 11.</b>	109
<b>Figura 41 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2008</b>	110
<b>Figura 42 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2011</b>	111
<b>Figura 43 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2012</b>	112
<b>Figura 44 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2015</b>	113
<b>Figura 45 – Predições de casos de dengue a partir da semana 11 para os anos de 2011 a 2020</b>	114

## LISTA DE TABELAS

<b>Tabela 2 – Divisão de dados para validação cruzada. Os dados de validação são usados para ajustar os parâmetros do modelo durante a fase de treinamento, enquanto os dados de teste são usados para avaliar o desempenho final do modelo. . . . .</b>	<b>57</b>
<b>Tabela 3 – separação de dados para treinamento, validação e testes das arquiteturas de rede neural propostas. . . . .</b>	<b>69</b>
<b>Tabela 4 – Divisão de Dados: Os dados de treinamento são utilizados para aprender o padrão histórico da evolução dos casos de dengue. Os dados de validação são empregados para ajustar os parâmetros do modelo durante a fase de treinamento, enquanto os dados de teste são utilizados para fins de previsão e análise. . . . .</b>	<b>88</b>

## LISTA DE ABREVIATURAS E SIGLAS

AR	modelo auto-regressivo
ARIMA	modelo auto-regressivo integrado de médias móveis
ARMA	modelo auto-regressivo de médias móveis
CNN	Convolutional Neural Networks
EAKF	ensemble adjustment Kalman filter
GCN	Graph Convolutional Neural Networks
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MA	modelo de médias móveis
OL	outbreak location
RMSE	erro médio quadrático
RMSLE	erro quadrático médio e logarítmico
RNN	Recurrent Neural Networks
S.E.I.R.	Suscetível, Exposto, Infectado e Removido
S.I.R.	Suscetível, Infectado e Removido

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	21
1.1	CONTEXTO . . . . .	21
1.2	MOTIVAÇÃO . . . . .	22
1.3	OBJETIVOS . . . . .	23
1.4	CONTRIBUIÇÕES DA TESE . . . . .	23
1.5	ESTRUTURAÇÃO DA TESE . . . . .	23
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	25
2.1	COMPREENSÃO DA TRANSMISSÃO DE DOENÇAS EM ÁREAS URBANAS . . . . .	25
2.2	DENV: CARACTERÍSTICAS E DIFICULDADES NA COMPREENSÃO DA DINÂMICA DE TRANSMISSÃO EM ÁREAS URBANAS . . . . .	28
2.3	MODELOS PREDITIVOS PARA O APRENDIZADO DE PROBLEMAS TEMPORAIS . . . . .	30
2.3.1	<b>Métodos estatísticos</b> . . . . .	31
2.3.2	<b>Aprendizado de séries temporais com redes neurais</b> . . . . .	32
2.3.3	<b>Rede neural em séries temporais</b> . . . . .	33
2.3.4	<b>Modelo Encoder-Decoder</b> . . . . .	36
2.3.5	<b>Graph Convolutional Neural Network</b> . . . . .	37
2.3.6	<b>Modelos epidemiológicos comportamentais</b> . . . . .	39
<b>3</b>	<b>O IMPACTO DA MOBILIDADE HUMANA NA TRANSMISSÃO DO VÍRUS DA DENGUE</b> . . . . .	45
3.1	DESCRIÇÃO DOS DADOS E ANÁLISES PRELIMINARES . . . . .	45
3.2	A IMPORTÂNCIA DA MOBILIDADE HUMANA NA INTENSIDADE DA FORÇA DE CONTÁGIO DO VÍRUS . . . . .	49
<b>4</b>	<b>PREVISÃO DE CASOS DE DENGUE</b> . . . . .	54
4.1	DADOS DE MOBILIDADE HUMANA COMO ATRIBUTO DE UMA REDE NEURAL . . . . .	54
4.1.1	<b>A arquitetura de rede neural proposta</b> . . . . .	54
4.1.2	<b>Modelo mecanístico</b> . . . . .	56
4.1.3	<b>Métodos de avaliação</b> . . . . .	58
4.1.4	<b>Predições e comparações</b> . . . . .	61

4.2	A POPULAÇÃO RESIDENTE COMO FATOR DE IMPACTO NA PREDIÇÃO DE CASOS A NÍVEL DE BAIRRO . . . . .	64
4.2.1	<b>Função de controle de intensidade como atributos de uma rede neural para predição de casos de dengue . . . . .</b>	65
4.2.2	<b>Modelo preditivo . . . . .</b>	67
4.2.3	<b>Predições e comparações . . . . .</b>	70
5	<b>O ANO DE 2020: UM LABORATÓRIO PARA COMPREENSÃO DE DOENÇAS CONTAGIOSAS . . . . .</b>	87
5.1	MODELO PREDITIVO, TREINAMENTO E DADOS UTILIZADOS . . . . .	87
6	<b>CONCLUSÕES E TRABALHOS FUTUROS . . . . .</b>	93
6.1	CONTRIBUIÇÕES DO TRABALHO . . . . .	93
6.2	LIMITAÇÕES . . . . .	94
6.3	TRABALHOS FUTUROS . . . . .	95
	<b>REFERÊNCIAS . . . . .</b>	96
	<b>APÊNDICES . . . . .</b>	103
	APÊNDICE A – ANÁLISE DA VARIAÇÃO DO NÚMERO DE CASOS DE DENGUE MÍNIMO PARA A CHEGADA DA DOENÇA SER CONSIDERADA EM UM BAIRRO . . . . .	104
	APÊNDICE B – ANÁLISE DE LIMIAR EPIDÊMICO DE PREVISÃO DE SURTOS DE DENGUE EM NÍVEL DE BAIRRO USANDO MOBILIDADE HUMANA EM ÁREAS URBANAS . . . . .	105
	APÊNDICE C – EXEMPLOS DE PREDIÇÕES UTILIZANDO DADOS DE MOBILIDADE HUMANA COM REDES NEURAIIS E MODELOS COMPORTAMENTAIS . . . . .	108
	APÊNDICE D – MAPAS COM A EVOLUÇÃO SEMANAIS DE CASOS DE DENGUE PARA OS ANOS COM GRANDES SURTOS DE CASOS DE DENGUE . . . . .	110
	APÊNDICE E – PREDIÇÕES DE CASOS DE DENGUE PARA OS ANOS DE 2011 A 2020 . . . . .	114

# 1 INTRODUÇÃO

## 1.1 CONTEXTO

A mobilidade humana, seja entre países, estados ou cidades, desempenha um papel fundamental na formação e interação de sociedades. Quanto mais densa a aglomeração populacional, maior é a movimentação e interação entre os indivíduos, seja de maneira intencional (por exemplo, em reuniões) ou não intencional (por exemplo, em transportes públicos). Em grandes cidades, que são centros urbanos densamente povoados, essa interação decorrente da mobilidade humana demanda a implementação de estruturas governamentais e privadas para lidar com os efeitos que causa, como aumento da criminalidade em áreas movimentadas. O aumento nos gastos com saúde também é consequência disso, devido à maior transmissão de doenças em razão da interação presencial entre as pessoas.

Pesquisadores em todo o mundo dedicam esforços para compreender o impacto da mobilidade humana na transmissão de doenças, seja rastreando as pessoas em deslocamentos entre países por meio de veículos para longo alcance (e.g. aviões, navios, etc.) (GRAIS; ELLIS; GLASS, 2003; MONTECUBIO, 2010; BAJARDI *et al.*, 2011), ou dentro de regiões urbanas utilizando dados de transporte público (PONTE *et al.*, 2021) ou até mesmo com rastreamentos individualizados por meio dispositivos móveis com acesso à internet e localização por GPS (VIGFUSSON *et al.*, 2021). Esses esforços se intensificaram em 2020 devido ao início da pandemia de COVID-19, que teve início na China e se espalhou rapidamente pelo continente europeu e, posteriormente, pelo mundo, levando à implementação de medidas de isolamento social por vários governos.

No campo das ciências exatas, como física estatística, ciência de dados e inteligência artificial, entre outros, existem numerosos estudos com o objetivo de desenvolver modelos matemáticos capazes de representar o impacto da mobilidade humana na transmissão de doenças em grandes cidades, a fim de compreender esse fenômeno e prever futuras epidemias com base na locomoção da população como o artigo (STODDARD *et al.*, 2009), os autores investigaram o impacto da mobilidade humana na propagação de doenças transmitidas por mosquitos. Os autores desenvolveram um modelo utilizando o vírus da dengue como exemplo para ilustrar como a mobilidade humana pode influenciar a dinâmica de transmissão do vírus. Uma das contribuições destacadas nesse artigo, que se alinha com a motivação desta tese, é o de que a densidade do mosquito transmissor não é necessariamente o fator primordial para a transmissão

de doenças, mas sim a mobilidade humana.

Na área de aprendizado de máquina e redes neurais, é comum encontrar estudos que analisam a evolução temporal de dados, incluindo a incidência de doenças. No trabalho de (WANG *et al.*, 2020), foram treinados modelos de redes neurais, especificamente Graph Neural Network (SCARSELLI *et al.*, 2008), para aprender a evolução temporal e espacial da COVID-19. Os resultados demonstraram que o uso de dados de mobilidade humana pode aprimorar a previsão de casos da COVID-19, e o *framework* proposto baseado em GNN apresentou maior acurácia em comparação com outros modelos de aprendizado de máquina e redes neurais.

Quanto ao vírus influenza, uma doença que gera custos significativos para governos em todo o mundo e é especialmente fatal para certos grupos populacionais, como idosos, é comum estudar o impacto da mobilidade humana em sua transmissão. Por exemplo, em (CAI *et al.*, 2020), um modelo preditivo baseado em equações diferenciais estocásticas foi desenvolvido, enquanto em (KHATUA *et al.*, 2020), foram utilizados modelos do tipo S.I.R..

## 1.2 MOTIVAÇÃO

Compreender o impacto da mobilidade humana na transmissão do vírus da dengue torna-se ainda mais desafiador, considerando que se trata de uma doença arboviral, ou seja, causada por um vírus transmitido por mosquitos, especificamente o mosquito *Aedes aegypti*. Vários estudos têm abordado o impacto da mobilidade humana na transmissão da dengue, considerando o deslocamento de pessoas entre locais distantes, como países, estados ou cidades. Nestes casos, o papel do mosquito é atenuado, uma vez que ele não é capaz de percorrer grandes distâncias. No entanto, ao tentar compreender o envolvimento da mobilidade em distâncias menores, como dentro das cidades, o impacto da interação constante e, por vezes, intensa entre seres humanos e mosquitos se torna evidente.

A motivação deste trabalho é o de investigar de forma quantitativa a contribuição da mobilidade humana na transmissão do vírus da dengue em grandes cidades. Buscar-se-á evidências empíricas quantitativas de que esse fator, em especial devido à complexa dinâmica de transmissão entre o mosquito vetor e os seres humanos, é fundamental para a transmissão do vírus. Além disso, visa-se investigar formas de considerar informações espaciais como população residente a nível de bairro e movimentação das pessoas entre bairros para aumentar acurácia de modelos preditivos de casos da doença.

### 1.3 OBJETIVOS

O objetivo central da tese é fornecer elementos para aumentar o conhecimento sobre o impacto da dinâmica de deslocamento de pessoas dentro de grandes cidades para a transmissão da Dengue. O ano de 2020, atípico devido a pandemia de COVID-19 e as restrições de mobilidade implementadas nas cidades, foi de especial importância nesse estudo, pois serviu como laboratório para compreender esse impacto. Análises comparativas com anos sem restrições de mobilidade mostram a diferença do comportamento da doença.

Após uma análise sobre a influência da mobilidade humana na transmissão da dengue, foi conduzido um estudo para avaliar o impacto da mobilidade humana e da população residente em modelos preditivos. A capacidade de modelar o papel da mobilidade na propagação do vírus e prever novos casos de dengue é de extrema importância. Nesse contexto, este estudo também tem como objetivo identificar as melhores práticas e modelos para essa finalidade. Para alcançar esse objetivo, propõe-se um *framework* baseado em redes neurais, capazes de aprender a relação espacial e temporal entre a mobilidade humana e a evolução da transmissão do vírus da dengue, assim como estudar alternativas de capturar esse impacto mesmo quando não há dados de mobilidade humana disponível.

### 1.4 CONTRIBUIÇÕES DA TESE

a) Realização de um estudo abrangente sobre o impacto da mobilidade humana na transmissão do vírus da dengue. b) Identificação e comparação de modelos preditivos capazes de aprender aspectos espaciais e temporais, aplicando-os ao domínio da previsão de casos de dengue em áreas urbanas. c) Proposição de um *framework* baseado em redes neurais capaz de prever a ocorrência da dengue com a maior antecedência em relação ao pico de casos.

### 1.5 ESTRUTURAÇÃO DA TESE

No Capítulo 2, serão descritas as bases teóricas para a compreensão da transmissão de doenças em áreas urbanas, incluindo a participação da mobilidade humana nesse processo e a dinâmica de transmissão do vírus da dengue. No Capítulo 3, será realizada uma análise do impacto da mobilidade humana na transmissão do vírus da dengue, utilizando dados de dengue da cidade de Fortaleza e dados de transporte público. Também será analisado o impacto do *lockdown* aplicado em 2020 durante a pandemia de COVID-19 na transmissão do vírus da



dengue.

O Capítulo 4 revê os modelos preditivos capazes de aprender os aspectos espaciais e temporais da transmissão do vírus da dengue para prever novos casos de dengue em nível de bairro e semanal. Modelos estatísticos, mecanicistas e modelos de rede neural são estudados. Um *framework* utilizando redes neurais capaz de aprender a relação espacial e temporal entre mobilidade humana e evolução da transmissão do vírus da dengue será proposto. Além disso, também será estudado o impacto da população residente em nível de bairro como atributos para modelos preditivos de casos de dengue.

No Capítulo 5, os dados do ano de 2020 servirão como um laboratório para a melhor compreensão de doenças contagiosas, tendo em vista as características atípicas deste ano devido as restrições de mobilidade humana aplicadas por conta da pandemia de COVID-19.

O Capítulo 6 conclui a tese com uma discussão que envolve as limitações do trabalho e sugestões para trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo descrever os principais temas relacionados ao assunto abordado neste trabalho. Na seção 2.1, serão apresentados estudos que investigaram a transmissão e a dinâmica de doenças infecciosas em áreas urbanas. Na seção 2.2, serão abordados trabalhos específicos sobre a transmissão do vírus da dengue, que é o foco deste estudo, com o intuito de compreender suas características e a dinâmica de transmissão entre o mosquito transmissor (*Aedes aegypti*) e os seres humanos. Por fim, na seção 2.3, serão descritos estudos que exploram modelos preditivos capazes de aprender dados com características temporais e espaciais, concentrando-se na previsão de casos de doenças infecciosas.

### 2.1 COMPREENSÃO DA TRANSMISSÃO DE DOENÇAS EM ÁREAS URBANAS

Durante anos, a urbanização tem sido vista como um símbolo de progresso e bem-estar pelas sociedades modernas, proporcionando à população acesso mais fácil e compartilhado a diversos recursos e bens de consumo. O mundo vem se tornando cada vez mais urbano desde a Revolução Industrial, e de acordo com as Nações Unidas (CENGİZ; GÖRMÜŞ; OĞUZ, 2022), 55,3% da população mundial vive em áreas urbanas, com estimativas de que esse número chegue a 68% até 2050. No Brasil, a população urbana representa 84,72% do total de habitantes, segundo o IBGE (IBGE, 2023b). Apesar dos pontos positivos de viver em áreas urbanas, ter uma grande parcela da população habitando um pequeno espaço traz inúmeros desafios para o poder público, especialmente na área da saúde pública. Grandes cidades podem funcionar como centros de transmissão de doenças devido à intensa interação de pessoas em locais de trabalho, shopping centers e aglomerações no transporte público. Além disso, a alta densidade populacional, principalmente em países com pouca infraestrutura de saneamento básico, favorece a proliferação de insetos e roedores, aumentando o risco de transmissão de doenças zoonóticas, além de facilitar a transmissão de doenças infecciosas, como a gripe.

De forma mais genérica, o trabalho de (NEIDERUD, 2015) aborda o impacto da rápida urbanização no mundo na epidemiologia de doenças infecciosas emergentes. O estudo argumenta que a urbanização pode resultar em cidades desestruturadas, com favelas e acesso limitado a saneamento básico, água limpa e infraestrutura habitacional inadequada, criando um ambiente propício para a proliferação de doenças infecciosas. O trabalho resume diversos fatores que favorecem a transmissão dessas doenças, como as condições de moradia, em que indivíduos que vivem em favelas têm maior probabilidade de se contaminar com determinados

tipos de doenças; o planejamento urbano, em que as áreas urbanas frequentemente crescem mais rapidamente do que o governo consegue fornecer infraestrutura adequada; a expansão de grandes cidades em áreas de mata, aproximando os seres humanos da vida silvestre e aumentando as chances de doenças zoonóticas; e o impacto da mobilidade humana na transmissão de doenças, já que a mobilidade humana é mais intensa e dinâmica em regiões urbanas.

Do ponto de vista da mobilidade humana, diversos estudos têm sido realizados para compreender o impacto dessa mobilidade na transmissão de doenças. A dinâmica e a compreensão desse impacto variam de acordo com o nível de deslocamento das pessoas, seja entre países, estados, cidades ou bairros. Em (HÂNCEAN; SLAVINEC; PERC, 2020), é estudado o impacto da mobilidade humana na transmissão do COVID-19 entre países, com um enfoque não apenas nas viagens aéreas globais, mas também na compreensão do tipo de viagens realizadas. Os autores investigam duas formas de mobilidade humana global: o fluxo de entrada de migração em um país e o fluxo de turistas. Eles demonstram que esses dois fatores têm um impacto significativo na transmissão do COVID-19 durante o estágio inicial da pandemia. Ao analisarem os primeiros casos de COVID-19 em todos os países do mundo, observaram que 64% dos infectados eram viajantes que retornavam de viagens internacionais, e 24% eram visitantes. Além disso, constataram que 71% desses casos envolviam transporte aéreo internacional. Isso reforça a importância da mobilidade humana na disseminação do vírus. Os resultados também evidenciam que a rede de transmissão do COVID-19 entre países segue uma distribuição de lei de potência, com cinco principais países atuando como transmissores globais: China, França, Grã-Bretanha, Irã, Itália e Estados Unidos.

No trabalho de (COSTA *et al.*, 2018), os autores investigaram a transmissão do vírus da gripe entre cidades do estado do Paraná, Brasil. O objetivo foi caracterizar a distribuição espaço-temporal do vírus da gripe desde a introdução do primeiro caso do vírus H1N1 até o final da epidemia em 2009. O estudo identificou alguns fatores importantes que favoreceram a transmissão da doença, tais como cidades com maior Índice de Desenvolvimento Humano (IDH), maior densidade populacional e temperaturas mais baixas. Além disso, os autores constataram que as cidades com maior fluxo de pessoas por meio do transporte público municipal apresentaram as maiores taxas de transmissão do vírus da gripe, destacando a relevância da mobilidade humana na disseminação de doenças entre as cidades.

O trabalho de (AGUILAR-SÁNCHEZ *et al.*, 2021) investiga como a estrutura organizacional da mobilidade humana em nível de cidade pode influenciar a transmissão de doenças. Além disso, analisa a eficácia de medidas de contenção. O estudo examina os padrões de

mobilidade humana nas 22 maiores cidades dos Estados Unidos, utilizando dados de mobilidade do Google COVID-19 Aggregated Mobility Research Dataset (GOOGLE, 2022). Os autores classificam as cidades em dois tipos: mobilidade hierárquica e mobilidade espalhada. Cidades com mobilidade hierárquica apresentam fluxos intensos para regiões específicas, como grandes polos comerciais (por exemplo, Nova Iorque e Chicago). Já cidades com mobilidade espalhada não possuem essas regiões bem definidas em uma única parte da cidade (como foi o caso de Atlanta). Os resultados mostraram que cidades com estrutura de mobilidade hierárquica facilitam a transmissão do COVID-19. No entanto, essas cidades também apresentaram melhor resposta às medidas de contenção, como o *lockdown*, que envolve o bloqueio da mobilidade humana.

Em (PONTE *et al.*, 2021), os autores propõem dois modelos matemáticos baseados em redes complexas e dinâmicas não lineares para investigar o papel do transporte público na transmissão do COVID-19 em Fortaleza, Brasil, ao longo do ano de 2020. O estudo analisa especificamente a contribuição dos profissionais de saúde na disseminação do vírus durante a primeira e a segunda onda de casos. Os dados de movimentação das pessoas por meio de ônibus, o principal meio de transporte na cidade, bem como os dados de casos de COVID-19 de março a dezembro de 2020, foram utilizados na análise.

O trabalho realiza uma análise da transmissão do COVID-19 no transporte público, empregando um modelo de rastreamento de contato e uma análise macroscópica de transmissão com base em modelos comportamentais, incluindo um modelo inspirado nos modelos *SEIR*, que considera as categorias de população suscetível, infectada, removida e exposta, para estimar a transmissão em nível da cidade. A contribuição do estudo reside na comparação entre os dois modelos mencionados, utilizando dados de mobilidade e casos de COVID-19 durante o período pandêmico de 2020, abrangendo diferentes regimes de isolamento social aplicados ao longo do período.

Os resultados do estudo mostram que ambos os modelos produzem resultados semelhantes, com exceção de um período de três meses entre as duas ondas de transmissão. Além disso, os autores demonstraram que os profissionais de saúde foram responsáveis por uma maior disseminação da doença durante a primeira onda, em comparação com outros usuários do transporte público. Também foi observado que profissionais de outras áreas, como policiais, desempenharam um papel importante na transmissão do vírus. Com base nesses achados, o estudo sugere a necessidade de considerar a priorização da vacinação para grupos altamente expostos ao vírus que dependem do uso de transporte público.

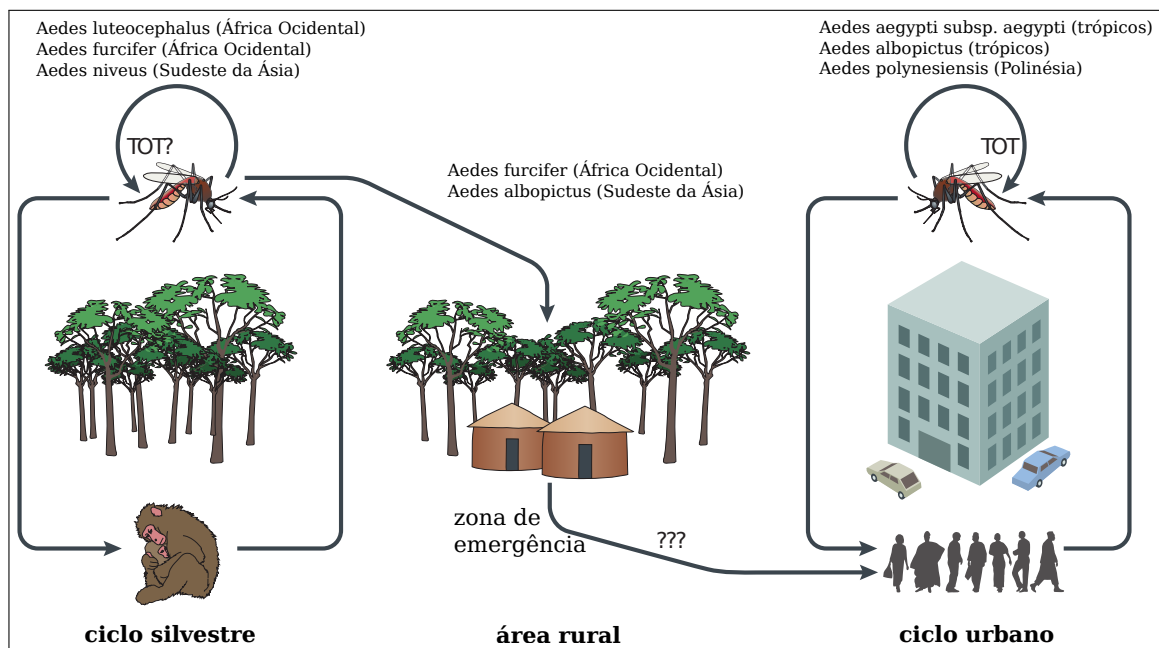
## 2.2 DENV: CARACTERÍSTICAS E DIFICULDADES NA COMPREENSÃO DA DINÂMICA DE TRANSMISSÃO EM ÁREAS URBANAS

O Dengue (DENV) é um arbovírus transmitido pelo mosquito *Aedes*, com quatro sorotipos distintos (DENV-1, DENV-2, DENV-3 e DENV-4), encontrado em todo o mundo. Estima-se anualmente a ocorrência de 390 milhões de casos de dengue (BHATT *et al.*, 2013), com 40% da população mundial vivendo em áreas de risco (WHO, 2019). As regiões tropicais são as mais afetadas pelo vírus, sendo que as Américas e a Ásia apresentam as maiores taxas de transmissão. Os sintomas comuns do DENV incluem febre, dor de cabeça, náuseas e vômitos. Embora a maioria dos casos de dengue seja leve, aproximadamente 0,5% dos casos evoluem para dengue hemorrágica, uma forma grave da doença que pode levar à morte em cerca de 5% dos casos (GIBBONS; VAUGHN, 2002).

Estudar a dinâmica de transmissão do vírus da dengue não é uma tarefa trivial devido à quantidade de variáveis envolvidas. Além dos quatro sorotipos que o vírus da dengue apresenta, a presença do mosquito é um dos principais fatores na transmissão da doença. A complexidade é elevada ainda mais pelo envolvimento de um terceiro agente na dinâmica: o próprio ser humano. No artigo de Vasilakis *et al.* (2011), é feita uma revisão detalhada sobre a evolução da transmissão do vírus da dengue. Os autores explicam que cada um dos sorotipos do DENV ocorre em dois ciclos de transmissão distintos: o ciclo silvestre e o ciclo humano. O ciclo silvestre ocorre entre primatas não humanos e o mosquito *Aedes*, tendo sido documentados casos na África e na Malásia. Já o ciclo humano do DENV ocorre entre o mosquito *Aedes* e os seres humanos, e esse sorotipo do vírus está presente em todo o mundo. Os autores descrevem a evolução temporal do vírus da dengue, desde seu ciclo silvestre na África e Malásia, passando de forma desconhecida do ambiente silvestre para os seres humanos através das áreas rurais, até chegar ao ciclo humano nas grandes cidades. O objetivo principal do trabalho é compreender o impacto da mobilidade humana na transmissão do vírus em áreas urbanas, o que é desafiador, considerando que a participação do mosquito *Aedes* é essencial na transmissão do vírus, independentemente do ambiente em que ocorre. A Figura 1 ilustra ambos os ciclos e a provável transição do vírus do ambiente silvestre para o ambiente humano (Vasilakis *et al.*, 2011).

O processo de urbanização e o aumento da densidade populacional são fatores significativos na transmissão de doenças infecciosas, especialmente o vírus da dengue, devido à proliferação de vetores como o mosquito *Aedes*, responsável pela disseminação de arboviroses. Com o crescimento da população em áreas urbanas, é crucial adotar medidas de controle para

**Figura 1 – O ciclo de transmissão do vírus da dengue: As origens silvestres do vírus da dengue e a 'zona de emergência', onde os ciclos silvestres entram em contato com populações humanas em áreas rurais da África Ocidental e do Sudeste Asiático. Além disso, o vírus da dengue pode persistir nas populações de mosquitos por transmissão transovariana (TOT), na qual mosquitos infectados pelo vírus transferem o vírus para seus ovos (isso foi demonstrado em algumas espécies, mas não em todas).**



Fonte: (VASILAKIS *et al.*, 2011)

prevenir a propagação dessas doenças, incluindo melhorias no saneamento básico e o uso de pesticidas, a fim de evitar epidemias. O estudo de referência (KOLIMENAKIS *et al.*, 2021) investiga o impacto dos fatores relacionados à urbanização na proliferação do mosquito *Aedes* e, conseqüentemente, nas doenças transmitidas por ele. Além disso, aborda o efeito da urbanização no ecossistema do *Aedes*, a influência da urbanização na dinâmica de transmissão, o papel da densidade populacional e a relação entre urbanização e doenças transmitidas pelo *Aedes*, sob uma perspectiva socioecológica. O artigo em questão revisa a presença do mosquito *Aedes* em áreas urbanas, evidenciando que nessas regiões a quantidade de mosquitos é maior do que em áreas rurais. Isso ocorre devido à alta densidade populacional humana, que proporciona mais oportunidades para a alimentação sanguínea dos mosquitos e o surgimento de mais recipientes de água artificiais. Por fim, os autores concluem que a rápida urbanização, com grandes populações vivendo em áreas urbanas não planejadas, cria condições favoráveis para o mosquito *Aedes* e aumenta a transmissão de doenças.

Tendo em vista a importância da presença humana na proliferação do mosquito transmissor da dengue, é crucial considerar o aspecto climático. Muitas vezes, a combinação de chuva com a presença humana favorece o acúmulo de água próximo a estabelecimentos comerciais e áreas residenciais, proporcionando um ambiente propício para a proliferação de mosquitos. O mosquito *Aedes* é sensível às variações de temperatura e depende de água parada para colocar seus ovos e se reproduzir, tornando o período chuvoso especialmente favorável para sua proliferação.

No estudo realizado por (FARUK; JANNAT; RAHMAN, 2022), foi analisado o impacto das variações climáticas na transmissão da dengue no Sri Lanka. Os dados de 26 cidades no período de janeiro de 2015 a dezembro de 2019 foram considerados. Considerando que o Sri Lanka é uma ilha de clima tropical, os autores dividiram o país em três regiões de acordo com o nível de precipitação: região molhada (chuva anual de 2000 mm), região intermediária (chuva entre 1150 e 2000 mm) e região seca (chuva abaixo de 1100 mm por ano).

Os principais fatores climáticos analisados neste estudo foram: temperatura, umidade relativa, precipitação, clareza do céu, pressão atmosférica e velocidade do vento. Observou-se que, durante o período analisado, a região molhada apresentou o maior número de casos de dengue, seguida pela região intermediária. Entre os fatores climáticos analisados, a precipitação e a umidade do ar apresentaram a maior correlação positiva com a transmissão do vírus da dengue, seguidas pela velocidade do vento e pressão atmosférica. Por outro lado, a temperatura apresentou uma correlação negativa, indicando que quanto maior a temperatura e pressão atmosférica, menor o número de casos de dengue. Quanto à clareza do céu, não foi observado impacto significativo na transmissão do vírus. Embora os autores não tenham realizado análises específicas sobre a mobilidade humana, concluem a importância de considerá-la nesse contexto.

### 2.3 MODELOS PREDITIVOS PARA O APRENDIZADO DE PROBLEMAS TEMPORAIS

Além de compreender o impacto da mobilidade na transmissão do vírus da dengue, este trabalho também tem como objetivo estudar modelos de predição capazes de aprender e prever casos de dengue, além de propor um framework para a predição de casos da doença. Nesta subseção, serão descritos trabalhos e métodos que se concentram no aprendizado de dados temporais e espaciais em diferentes domínios.

### 2.3.1 Métodos estatísticos

Na área das ciências exatas, diversos modelos são utilizados para o aprendizado de séries temporais. Na estatística, existem modelos amplamente conhecidos e utilizados, tais como o modelo auto-regressivo (AR), o modelo de médias móveis (MA), o modelo auto-regressivo de médias móveis (ARMA) e o modelo auto-regressivo integrado de médias móveis (ARIMA) (BISGAARD; KULAHCI, 2011).

Os modelos auto-regressivos são modelos estatísticos que assumem que o valor de uma variável depende linearmente do seu próprio valor em um momento anterior. Assim, o modelo é definido pela equação 2.1.

$$Y_{t+1} = Y_t + X + b \quad (2.1)$$

onde  $Y_{t+1}$  é o valor da variável no tempo  $t + 1$ ,  $Y_t$  é o valor conhecido da variável no tempo  $t$ ,  $X$  é o parâmetro aprendido e  $b$  é o viés. Modelos de auto regressão podem ter  $N$  ordens: AR(1) indica ordem 1, em que o valor da variável a ser previsto ( $Y_{t+1}$ ) depende apenas do seu valor em um momento anterior; AR(2) indica ordem 2, em que o valor da variável a ser previsto depende tanto de seu valor em um passo anterior quanto em 2 passos anteriores, e assim por diante para ordens 3, 4 e  $N$ . O modelo AR é comumente utilizado na análise de auto-regressão. Esse estudo determina qual valor anterior daquela variável apresenta maior correlação com a variável a ser prevista, seja o valor um passo antes, dois passos antes, três passos antes ou  $N$  passos antes.

Diferentemente do modelo de auto regressão, o modelo de médias móveis considera que a variável dependente ( $\hat{Y}_{t+1}$ ) possui uma correlação cruzada com outra variável. Em outras palavras,  $\hat{Y}_{t+1}$  depende dos valores de outra variável em momentos anteriores (variável independente), e não de seus próprios valores. Os modelos de médias móveis são representados pela sigla MA(q) e pela equação 2.2, onde q representa a ordem do modelo, semelhante aos modelos de auto regressão. A equação do modelo de médias móveis é semelhante à equação 2.1, mas  $Y_t$  é substituído por valores de variáveis diferentes daquela que se deseja prever. A análise de auto-regressão também pode ser realizada com modelos de médias móveis, com o objetivo de determinar qual instante anterior da variável independente melhor explica a variável dependente.

$$Y_{t+1} = \beta_0 + \varepsilon_{t+1} + \theta_1 \varepsilon_t \quad (2.2)$$



O modelo auto-regressivo de médias móveis é a combinação dos modelos AR e MA. Nesse modelo, é realizada uma regressão utilizando os valores passados da própria variável a ser prevista, além de outras variáveis que possuam correlação cruzada com a variável em questão. A equação 2.3 representa o modelo ARMA, que é a junção das equações 2.1 e 2.2.

$$Y_{t+1} = \beta_0 + \beta_1 Y_t + \varepsilon_{t+1} + \theta_1 \varepsilon_t \quad (2.3)$$

No artigo (VOYANT *et al.*, 2011), os autores utilizaram o modelo ARMA para prever as taxas de radiação solar e demonstraram que, em certos cenários, esse modelo apresentou resultados similares aos obtidos por modelos baseados em redes neurais. Já no artigo (CHO *et al.*, 2021), o modelo ARMA foi combinado com redes neurais para prever o rendimento de uma plantação de tomate na agricultura.

Os modelos AR, MA e ARMA pressupõem que a série temporal a ser analisada seja estacionária. Uma série temporal é considerada estacionária quando suas propriedades estatísticas, como média, variância e autocorrelação, permanecem constantes ao longo do tempo. No entanto, nem todas as séries temporais são estacionárias. Portanto, é necessário aplicar um processo chamado "integração" para torná-las estacionárias, o qual é incorporado no modelo ARMA para lidar com séries temporais não estacionárias. Dessa forma, o modelo auto-regressivo integrado de médias móveis (ARIMA) é especificado como ARIMA(p, d, q), em que p é a ordem do termo auto-regressivo, q é a ordem do termo de médias móveis e d é o número de diferenciações necessárias para alcançar a estacionariedade da variável.

### 2.3.2 Aprendizado de séries temporais com redes neurais

O Aprendizado de Máquina é uma área da Inteligência Artificial que tem como objetivo fazer com que computadores aprendam a partir de exemplos e dados, sem a necessidade de serem explicitamente programados. Ele é dividido em três categorias principais de aprendizado, que agrupam características sobre como esses tipos de aprendizado lidam com diferentes tarefas e como os dados devem ser preparados para esse fim. Essas três categorias são: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

- a) **Aprendizado supervisionado:** é uma área do aprendizado de máquina em que modelos de aprendizagem são treinados usando dados rotulados, ou seja, dados que associam determinadas entradas com determinadas saídas conhecidas. Esses modelos são usados para resolver problemas de classificação, como detecção

de spam em e-mails ou análise de sentimento em textos, e problemas de regressão, como previsão de incidência de casos de dengue ou estimativa de valores imobiliários em uma região.

- b) **Aprendizado não supervisionado:** esse tipo de aprendizado refere-se a dados não rotulados ou não estruturados, ou seja, dados que não possuem uma informação de saída associada a eles. O objetivo é aprender padrões entre esses dados e identificar estruturas e tendências ocultas. O aprendizado não supervisionado é utilizado em algoritmos de agrupamento, análise de componentes principais, redução de dimensionalidade e em atividades como sugestão de produtos em e-commerces ou de filmes em empresas de streaming.
- c) **Aprendizado por reforço:** é um tipo de aprendizado em que um agente de aprendizagem interage com um ambiente virtual, no qual uma determinada situação é modelada, e o objetivo é encontrar a melhor ação a ser tomada em cada estado do ambiente, visando maximizar uma medida de recompensa. O aprendizado é realizado por meio de tentativa e erro, no qual o agente aprende a partir do feedback fornecido pelo ambiente, que pode ser positivo ou negativo, dependendo das ações tomadas. Esse tipo de aprendizado é utilizado em diversas áreas, como robótica, jogos, controle de processos industriais, entre outros.

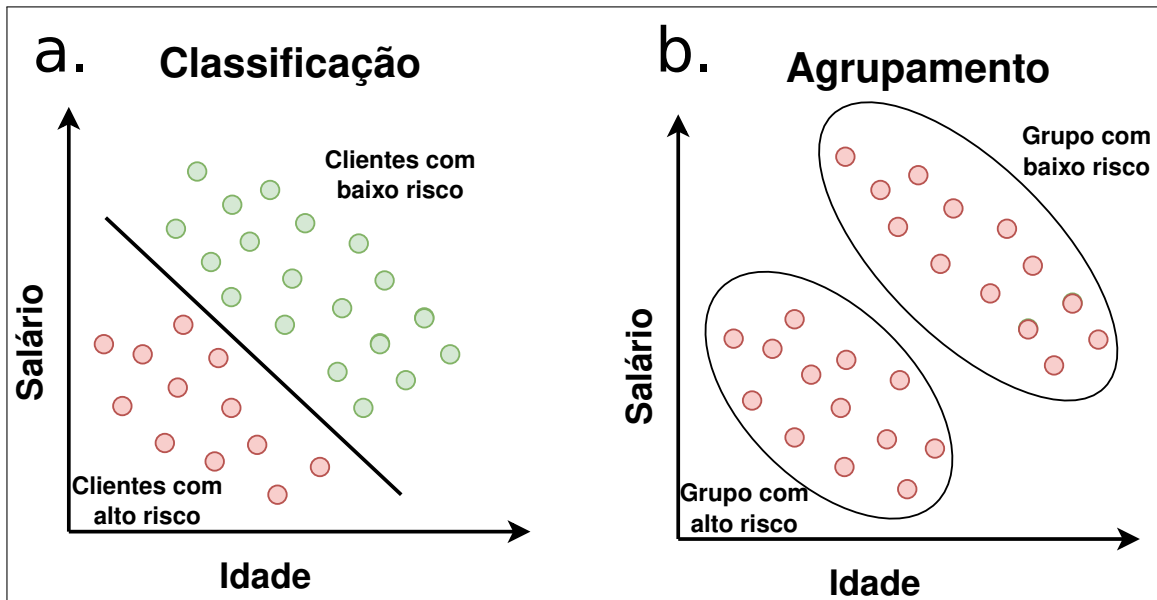
### 2.3.3 Rede neural em séries temporais

A rede neural é uma área de aprendizado de máquina capaz de realizar tarefas de aprendizado supervisionado, não supervisionado e aprendizado por reforço, imitando o funcionamento do cérebro humano. A ideia é processar as informações de entrada em um modelo de redes neurais em várias camadas, onde cada camada aprende uma característica específica do problema. As redes neurais são capazes de lidar com problemas de maior complexidade do que os algoritmos convencionais de aprendizado de máquina.

Cada camada de uma rede neural é composta por neurônios que realizam uma transformação linear nos dados de entrada. A organização desses neurônios em camadas é chamada de arquitetura de rede neural. A arquitetura mais simples estudada é chamada de perceptron, ilustrada na Figura 3.

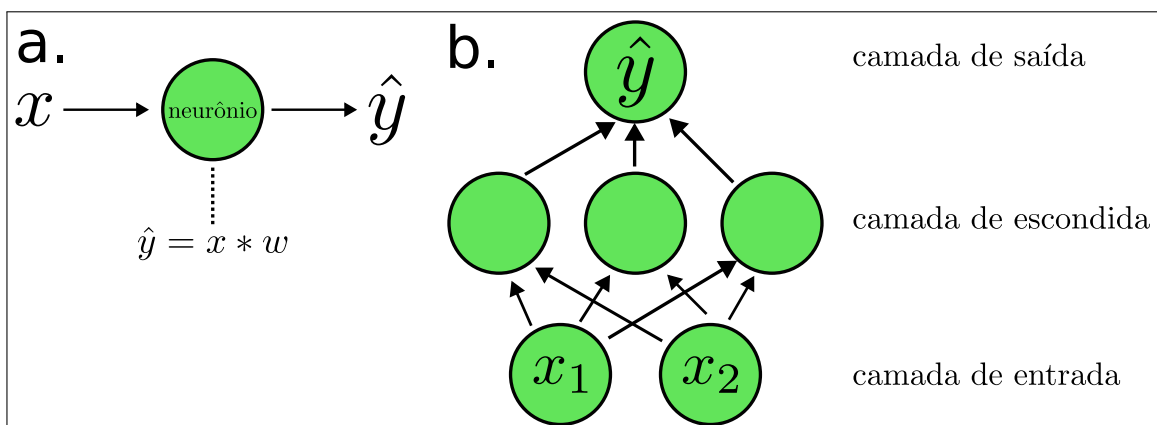
Perceptrons são a forma mais simples de definição de redes neurais e, consequentemente, não são comuns no aprendizado de séries temporais. Embora sejam utilizados na

**Figura 2 – A figura (a) mostra exemplo de classificação. O modelo supervisionado classifica os clientes como de alto risco ou baixo risco de acordo com dois atributos: idade e salário. A figura (b) mostra exemplo de agrupamento. O modelo agrupo as pessoas de acordo com os dois atributos.**



Fonte: Elaborado pelo autor

**Figura 3 – Neurônio e perceptron: na figura (a) é ilustrada a transformação linear realizada por um neurônio, onde  $\hat{y}$  é a saída do neurônio,  $x$  é a entrada e  $w$  é o parâmetro a ser treinado. Na figura (b), é mostrado um perceptron com camadas de entrada, escondida e saída. Neste exemplo, existe apenas uma camada escondida, mas um perceptron pode ter várias camadas escondidas.**

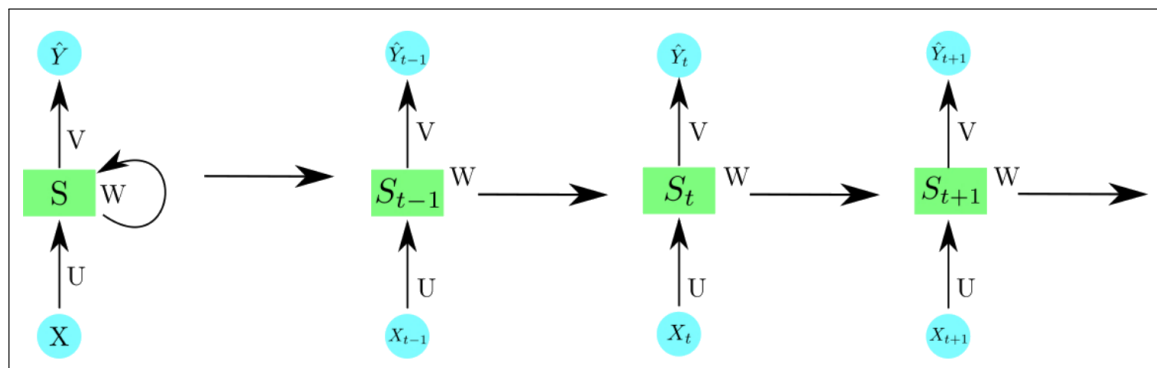


Fonte: Elaborado pelo autor

resolução de certos problemas (FREUND; SCHAPIRE, 1998; RUCK; ROGERS; KABRISKY, 1990; ORHAN; HEKIM; OZER, 2011), e seja possível definir séries temporais como a entrada de uma arquitetura de rede neural baseada em perceptron, esse tipo de arquitetura não é capaz de

tratar os dados como uma sequência. Em uma sequência, o valor em um tempo  $t$  pode afetar o valor no tempo  $t + 1$  e assim por diante. Em 1990, Jeffrey L. Elman propôs as redes neurais recorrentes ou Recurrent Neural Networks (RNN) em seu artigo (ELMAN, 1990). As RNNs são um tipo de rede neural capaz de aprender sequências temporais. Para prever o valor de saída  $\hat{Y}$  em um tempo  $t$ , a arquitetura considera como entrada o valor  $X_t$  e também o valor previsto em um tempo anterior,  $Y$  em  $t - 1$ . Essa arquitetura é capaz de aprender a dependência temporal da variável de entrada devido à aplicação do algoritmo de aprendizado chamado backpropagation (RUMELHART; HINTON; WILLIAMS, 1985). O princípio dessa técnica é que é possível calcular o gradiente do erro em relação aos parâmetros da rede ( $U$ ,  $V$  e  $W$ , ilustrados na Figura 4), e assim aprender parâmetros aprimorados usando o gradiente descendente estocástico. Durante o treinamento, os erros e gradientes são somados em cada passo de tempo. A Figura 4 ilustra uma rede neural recorrente.

**Figura 4 – Fluxo temporal da rede neural recorrente.  $U$ ,  $V$  e  $W$  são parâmetros da rede.  $X$  e  $Y$  são a entrada e a previsão do modelo, respectivamente.  $S$  é o estado do modelo no tempo  $t$ .**



Fonte: Elaborado pelo autor

Especificamente no contexto do aprendizado de séries temporais, o uso de backpropagation em redes neurais recorrentes (RNNs) pode levar ao problema do *vanishing gradient*, tornando difícil aprender dependências de tempo de longo prazo. Esse problema ocorre quando uma RNN lida com sequências extensas, resultando em gradientes muito pequenos durante o processo de retropropagação e, conseqüentemente, pouca atualização nos pesos da RNN.

Para contornar o problema do *vanishing gradient*, foi proposta a arquitetura conhecida como Long Short-Term Memory (LSTM) (GRAVES; GRAVES, 2012). O LSTM é um tipo de RNN que incorpora uma unidade de memória celular e um conjunto de portas (portas de entrada, saída e esquecimento) que regulam o fluxo de informações para dentro e para fora da

memória celular.

A Figura 5 e as equações a seguir ilustram o funcionamento das unidades LSTM.

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \quad (2.4)$$

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \quad (2.5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \quad (2.6)$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_{t-1} + b_o) \quad (2.7)$$

$$m_t = o_t \odot h(c_t) \quad (2.8)$$

$$y_t = \phi(W_{ym}m_t + b_y) \quad (2.9)$$

onde  $W_{ix}$ ,  $W_{ic}$ ,  $W_{fc}$  e  $W_{oc}$  são as matrizes de pesos,  $b$  são os vetores de viés ( $b_i$ ,  $b_f$ ,  $b_c$ ,  $b_o$  e  $b_y$ ),  $\sigma$  é a função logística sigmoideal e  $i$ ,  $f$ ,  $o$ ,  $c$  representam, respectivamente, a porta de entrada, a porta de esquecimento, a porta de saída e o vetor de ativação da célula.  $m$  é o vetor de saída,  $\odot$  é o produto elementar de vetores, e  $g$  e  $h$  são as funções de ativação da entrada e da saída da célula.

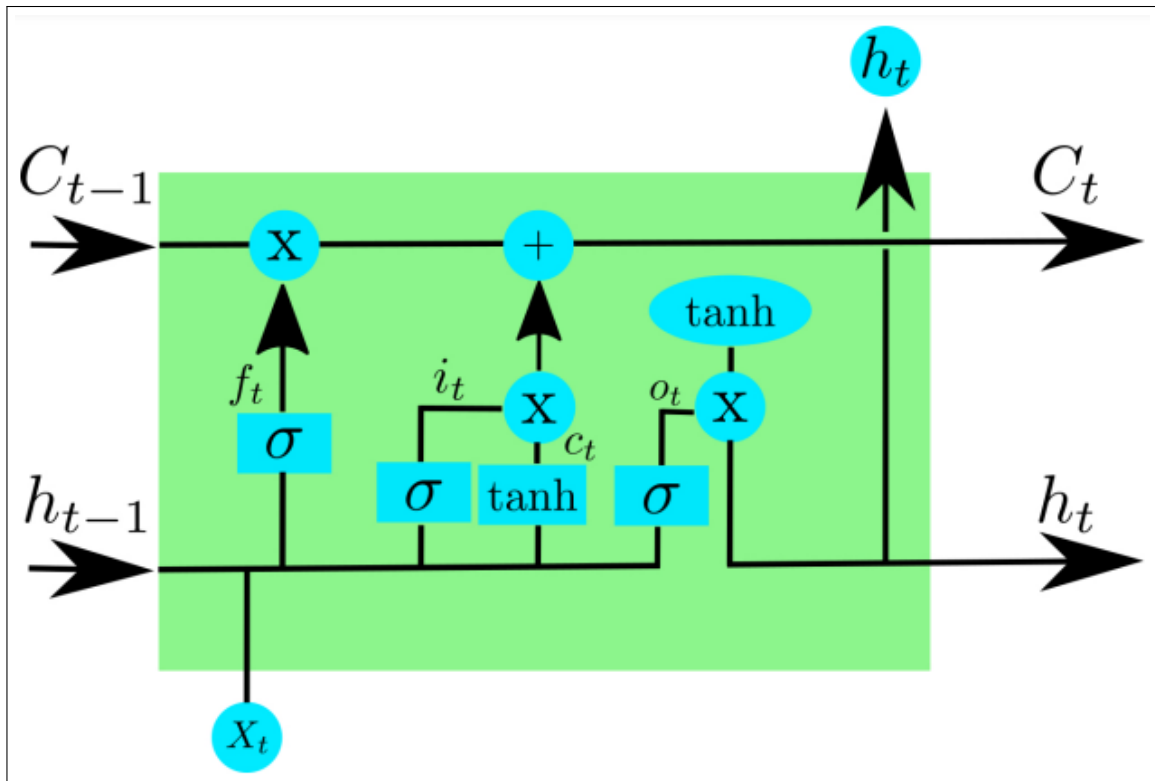
LSTMs foram utilizadas na resolução de diversos problemas como na previsão de chuva (SALEHIN *et al.*, 2020; POORNIMA; PUSHPALATHA, 2019; FENTE; SINGH, 2018) e na previsão de doenças como na influenza (LIU *et al.*, 2018) e dengue (SALEH; BAIWEI, 2021).

### 2.3.4 Modelo Encoder-Decoder

Uma evolução das redes neurais baseadas em RNN ou LSTM é a arquitetura *Encoder-Decoder*, proposta em 2014 por Cho et al. (CHO *et al.*, 2014b), com o objetivo de prever múltiplos passos de uma série temporal. Inicialmente desenvolvida para a tarefa de tradução automática de textos, essa arquitetura é agora amplamente aplicada em várias outras atividades, como resumo de texto, processamento de imagem, previsão de séries temporais e sistemas de chatbots.

As arquiteturas de *Encoder-Decoder* são compostas por três elementos principais: um componente chamado *encoder*, um vetor de contexto e outro componente chamado *decoder*. Tanto o *encoder* quanto o *decoder* podem ser implementados usando uma *stack* de LSTMs,

**Figura 5 – Uma unidade LSTM possui uma estrutura em cadeia que contém quatro redes neurais e diferentes blocos de memória chamados células. A informação é retida pelas células e as manipulações da memória são feitas através das portas de entrada ( $i_t$ ), esquecer ( $f_t$ ) e saída ( $o_t$ ). A memória de longo prazo é geralmente chamada de estado da célula em que as informações de intervalos anteriores são armazenadas.**



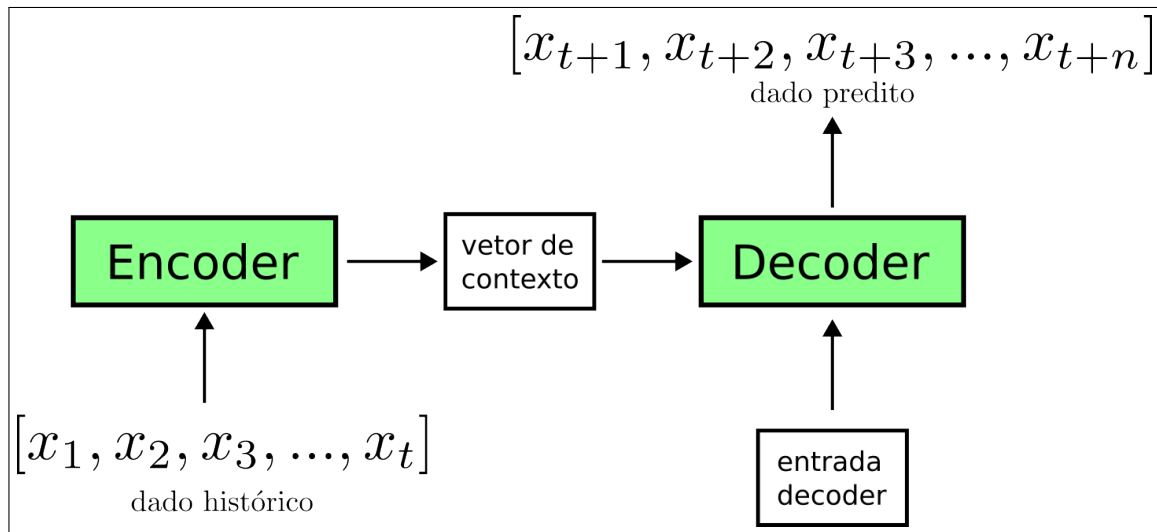
Fonte: Elaborado pelo autor

RNNs ou qualquer outro algoritmo de rede neural. O *encoder* tem a função de aprender as características do histórico da série temporal de entrada da arquitetura e transformá-las em uma representação vetorial de tamanho fixo, conhecida como contexto. Por sua vez, o *decoder* é responsável pela geração das previsões, ou seja, pela produção dos valores de saída da arquitetura com base nos dados de entrada em tempos anteriores, levando em consideração as características do histórico armazenado no vetor de contexto. A Figura 6 ilustra a arquitetura *Encoder-Decoder*.

### 2.3.5 Graph Convolutional Neural Network

Hoje em dia, muitos problemas no mundo envolvem o processamento de dados no formato de grafos. No entanto, os principais algoritmos de aprendizado de séries temporais mencionados anteriormente (LSTMs, RNNs, perceptrons, AR, MA, ARMA e ARIMA) não são

**Figura 6 – Arquitetura encoder-decoder.**



Fonte: Elaborado pelo autor

adequados para lidar com esse tipo de estrutura. Nos últimos anos, o problema de aprendizado em grafos tem ganhado destaque, e um modelo em particular tem se destacado: as Graph Convolutional Neural Networks (GCN) (KIPF; WELLING, 2016). Essas redes neurais convolucionais em grafos têm como objetivo resolver problemas como classificação de grafos, classificação de nós, predição de conexões entre nós, detecção de comunidades e embedding de grafos.

As GCNs são uma arquitetura de rede neural que funciona de maneira semelhante às redes neurais convolucionais (CNNs) (LECUN *et al.*, 1989), amplamente utilizadas no processamento de imagens. As CNNs multiplicam os neurônios de entrada da rede por um conjunto de pesos conhecidos como filtros. O filtro é um parâmetro aprendido da rede neural convolucional e funciona como uma matriz de tamanho  $N \times N$  que desliza sobre todos os neurônios de entrada, sendo capaz de aprender características do dado de entrada. As CNNs são amplamente utilizadas em problemas de classificação de imagens (SPANHOL *et al.*, 2016; DHRUV; NASKAR, 2020) e detecção de objetos em imagens (GALVEZ *et al.*, 2018; SAQIB *et al.*, 2018). A diferença entre as CNNs e as GCNs é que os filtros das CNNs aprendem características entre pontos vizinhos na imagem, ou seja, a relação entre pontos próximos na imagem, permitindo assim aprender características como formas e intensidades de cores dentro de uma imagem. As GCNs funcionam de maneira semelhante, mas são capazes de trabalhar com dados organizados em um formato não euclidiano, permitindo aprender grafos com diferentes números de nós e conexões entre eles.

A Graph Convolutional Neural Network é definida por uma função que aprende

atributos em grafos,  $G = (V, E)$ , onde  $V$  é o conjunto de nós e  $E$  o conjunto de arestas. A entrada dessa função é uma matriz de atributos  $X_i$  dos nós, no formato  $N \times D$ , onde  $N$  é o número de nós e  $D$  é o número total de atributos que cada nó pode ter. A função também recebe uma matriz  $A$  no formato  $N \times N$  que contém os valores que representam os pesos das conexões entre os nós. Caso um valor seja zero, significa que não há conexão entre os dois nós correspondentes. A função retorna novos valores de atributos para cada nó.

$$f(H^{(l)}, A) = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (2.10)$$

onde  $X_{l+1}$  é o estado dos atributos dos nós em um momento posterior,  $\hat{A} = A + I$ ,  $A$  é a matriz de adjacência que representa as conexões entre os nós e  $I$  é a matriz identidade, utilizada para representar loops entre os nós.  $\hat{D}$  representa a matriz de graus utilizada para a normalização da matriz de adjacência.  $H^{(l)}$  representa a matriz de atributos dos nós no momento  $l$ , e  $W$  é o parâmetro aprendido da rede neural.

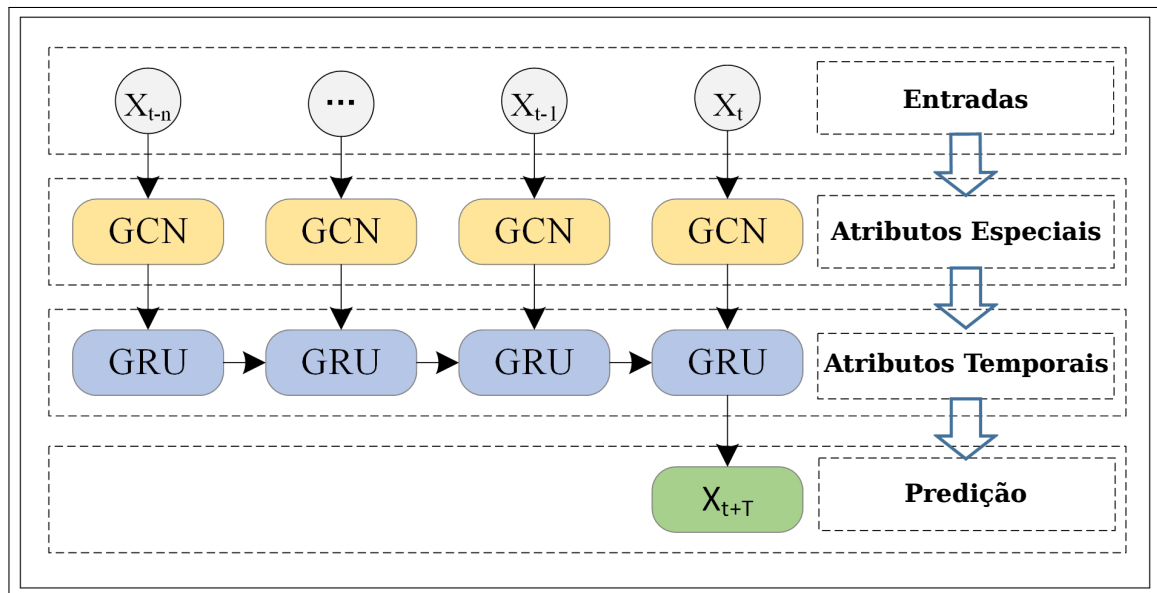
As GCNs são capazes de aprender dados no formato de grafos por meio do processo de convolução. No entanto, surge um problema quando essa estrutura de grafo é temporal. Para abordar essa questão, (ZHAO *et al.*, 2019) propôs uma arquitetura baseada em algoritmos de redes neurais temporais, mais especificamente *Gated Recurrent Unit* (GRU) (CHO *et al.*, 2014a), capaz de aprender a estrutura temporal de saída de múltiplas GCNs. A ideia é desenvolver uma arquitetura de rede neural capaz de aprender dados espaciais e temporais. O aprendizado espacial é realizado por meio da estrutura de grafos do GCN, enquanto o aprendizado temporal é realizado por meio da GRU. A Figura 7 ilustra a arquitetura proposta para a predição de dados temporais e espaciais.

### 2.3.6 Modelos epidemiológicos comportamentais

Na área da teoria matemática das epidemias, vários modelos comportamentais foram propostos para compreender a transmissão de doenças entre populações. Um dos modelos mais conhecidos é o modelo S.I.R., que representa os estados Suscetível (S), Infectado (I) e Recuperado (R). Esse modelo foi proposto em 1927 por W. O. Kermack e A. G. McKendrick no artigo (KERMACK; MCKENDRICK, 1927). Os modelos do tipo S.I.R. consideram que os indivíduos de uma população podem estar em três estados diferentes em um determinado momento: suscetíveis a uma doença, infectados ou recuperados.

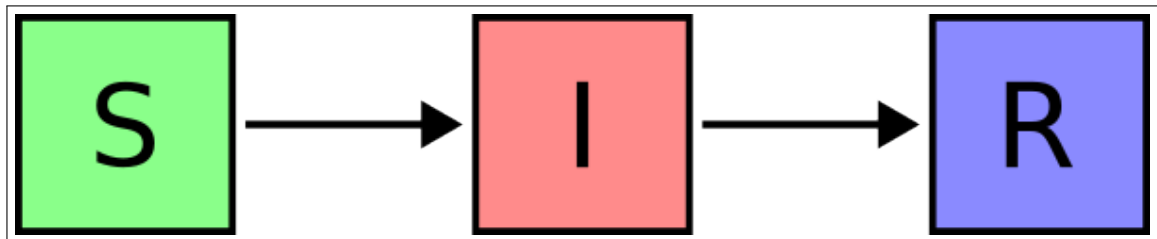


**Figura 7 – Temporal Graph Neural Network.**



Fonte: (ZHAO *et al.*, 2019)

**Figura 8 – Fluxo da população entre os três estados do modelo SIR.**



Fonte: (ZHAO *et al.*, 2019)

$$N = S(t) + I(t) + R(t) \quad (2.11)$$

A Equação (2.11) mostra que o tamanho da população  $N$  é a soma dos indivíduos suscetíveis, infectados e recuperados.

A intuição por trás do modelo é a seguinte: um indivíduo infectado entra em contato com uma comunidade composta por indivíduos suscetíveis. Quando um indivíduo suscetível entra em contato com uma pessoa infectada, há uma probabilidade de se infectar, dependendo das características da doença. Uma vez infectado, o indivíduo permanece com a doença por um período, durante o qual pode infectar outras pessoas, até se recuperar. O modelo assume que o período de uma epidemia é curto em comparação com a vida útil da população, e considera o tamanho da população como fixo, ou seja, não há adição de novos indivíduos ao modelo. Dessa forma, o modelo é capaz de inferir quantas pessoas estão infectadas, suscetíveis ou recuperadas

em um determinado momento, e considera que a epidemia acaba quando todos os indivíduos da comunidade estão recuperados.

Para cada um dos estados, os autores definiram equações que descrevem a dinâmica desses estados no modelo durante a execução.

$$\frac{dS}{dT} = -BSI \quad (2.12)$$

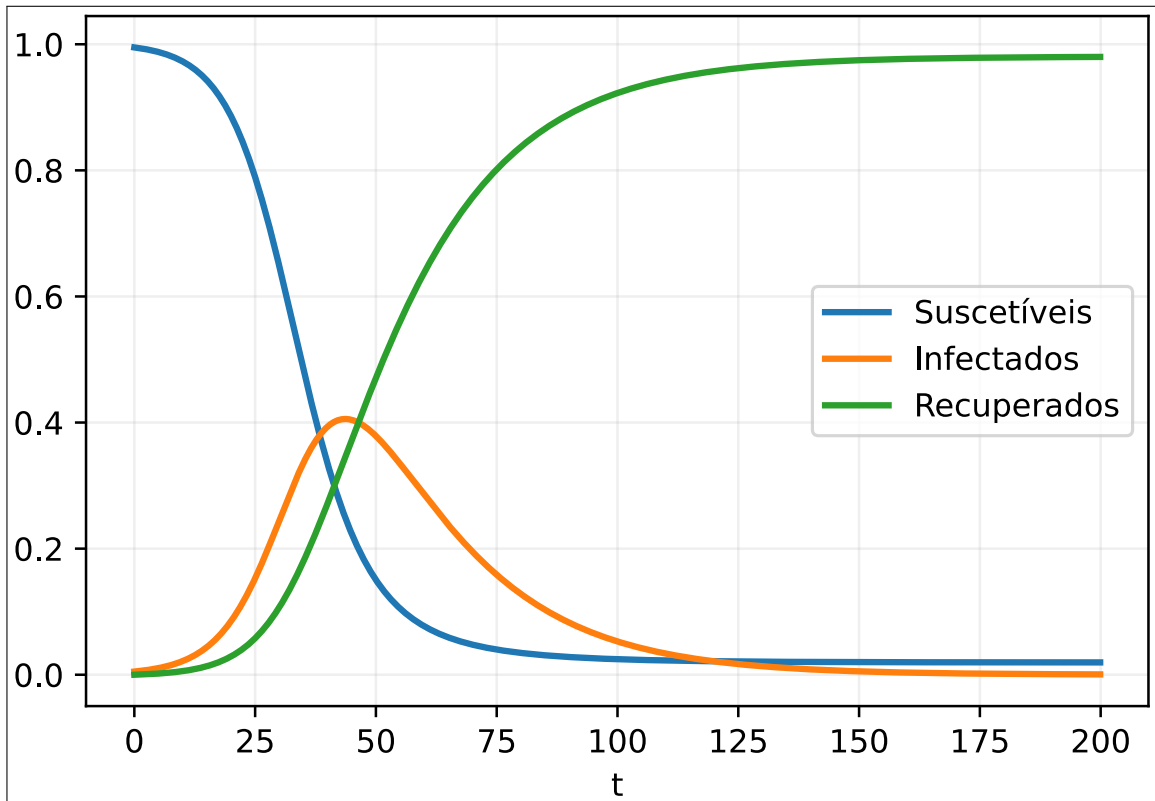
$$\frac{dI}{dT} = BSI - yI \quad (2.13)$$

$$\frac{dR}{dT} = yI \quad (2.14)$$

Na Equação 2.12,  $\frac{dS}{dT}$  representa a variação da quantidade de pessoas suscetíveis à doença ao longo do tempo  $T$ . O  $B$  denota a taxa de transmissão do vírus em questão,  $S$  representa a quantidade de pessoas suscetíveis, e  $I$  representa a quantidade de pessoas infectadas que entraram em contato com indivíduos suscetíveis. Por sua vez, a Equação 2.13 descreve a variação da quantidade de pessoas infectadas ao longo do tempo  $T$ . Mais uma vez,  $B$ ,  $S$ , e  $I$  têm os mesmos significados de taxa de transmissão, população suscetível e população infectada, respectivamente, enquanto  $y$  representa a taxa de pessoas recuperadas. Assim, a Equação 2.13 indica a diferença entre a quantidade de pessoas infectadas no momento e a quantidade de pessoas que se recuperaram e não estão mais infectadas. Por fim, a Equação 2.14 representa a variação da quantidade de pessoas recuperadas ao longo do tempo  $T$ . A Figura 9 ilustra a evolução temporal dos três estados da população: suscetíveis, infectados e recuperados.

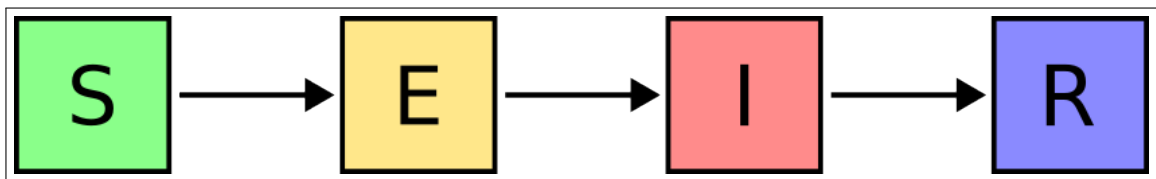
Apesar do modelo S.I.R. ser amplamente utilizado, existem variações que podem representar outros comportamentos além dos três estados suscetível, infectado e recuperado. Durante o ciclo de transmissão de um vírus, um indivíduo em uma população pode entrar em contato com o vírus e passar por um período de incubação, no qual não apresenta sintomas e não está infectado. Além disso, há casos em que um indivíduo pode evoluir para um estado assintomático. Para modelar esses comportamentos de exposição e incubação, o modelo S.E.I.R. inclui um novo estado entre os estados suscetível e infectado, representado pela letra E. O estado exposto engloba os indivíduos que estão em período de exposição ao vírus e ainda não foram infectados.

**Figura 9 – Evolução da curva de pessoas suscetíveis (azul), infectadas (laranja) e recuperadas (verde).**



Fonte: Elaborado pelo autor

**Figura 10 – Fluxo da população entre os três estados do modelo Suscetível, Exposto, Infectado e Removido.**



Fonte: Elaborado pelo autor

No modelo Suscetível, Exposto, Infectado e Removido, a população é representada pela soma de todos os indivíduos suscetíveis, expostos, infectados e recuperados. As equações que definem cada um dos quatro estados são as seguintes:

$$\frac{dS}{dT} = -BS\left(\frac{I}{N}\right) \quad (2.15)$$

$$\frac{dE}{dT} = BS\left(\frac{I}{N}\right) \quad (2.16)$$

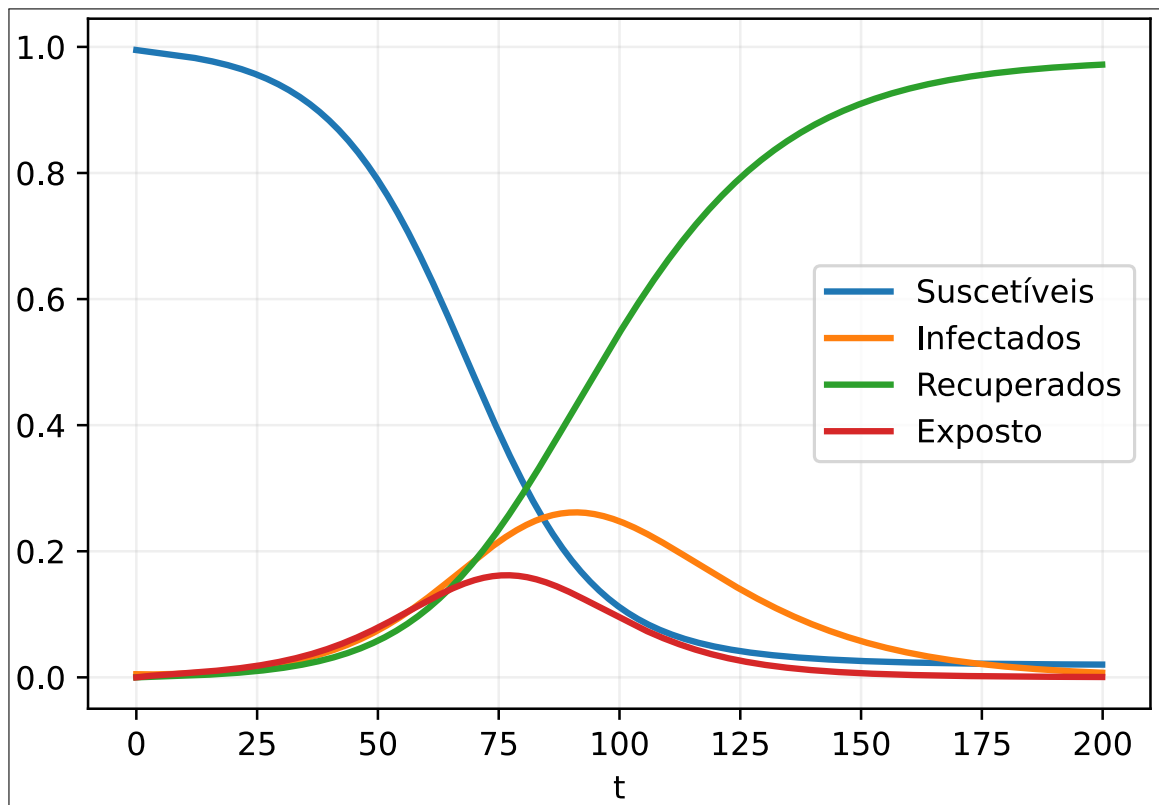
$$\frac{dI}{dT} = kE - yI \quad (2.17)$$

$$\frac{dR}{dT} = yI \quad (2.18)$$

as variáveis  $S$ ,  $E$ ,  $I$  e  $R$  representam a quantidade de indivíduos suscetíveis, expostos, infectados e recuperados, respectivamente. As taxas  $b$ ,  $k$  e  $y$  correspondem à taxa de transmissão, à taxa de transição de exposto para infectado e à taxa de recuperação, respectivamente. A Figura 11 ilustra a evolução temporal dos quatro estados da população: suscetíveis, infectados, recuperados e expostos.

Embora esses modelos tenham sido propostos há muito tempo, eles ainda são amplamente utilizados no estudo da transmissão de doenças. Durante a pandemia de COVID-19, diversos trabalhos utilizaram o modelo SIR para prever casos (MA; KITSACK; MIEGHEM, 2022; COOPER; MONDAL; ANTONOPOULOS, 2020; CHEN *et al.*, 2020). Além disso, o modelo SIR também foi aplicado na previsão de casos de dengue (CHANPRASOPCHAI; TANG; PONGSUMPUN, 2018; SIDE; NOORANI, 2013) e influenza (OSTHUS *et al.*, 2017; CASAGRANDE *et al.*, 2006).

**Figura 11 – Evolução da curva de pessoas suscetíveis (azul), infectadas (laranja), recuperadas (verde) e expostas (vermelha).**



Fonte: Elaborado pelo autor

### 3 O IMPACTO DA MOBILIDADE HUMANA NA TRANSMISSÃO DO VÍRUS DA DENGUE

Conforme mencionado na introdução desta tese, a mobilidade humana desempenha um papel crucial na transmissão de doenças virais. Estudos anteriores já demonstraram uma correlação entre a mobilidade humana e a propagação de doenças em diversos níveis, incluindo o transporte aéreo entre países e a mobilidade entre estados e cidades. No entanto, entender a transmissão do vírus da dengue e o impacto da mobilidade humana nesse processo é particularmente desafiador devido à complexa interação entre seres humanos e mosquitos transmissores.

Um dos principais objetivos desta tese é compreender o impacto da mobilidade humana em nível de bairro na transmissão do vírus da dengue em uma grande cidade. Para atingir esse objetivo, foram realizadas análises utilizando dados da cidade de Fortaleza, localizada no estado do Ceará, Brasil. Fortaleza é a quinta maior capital do país, com uma população de 2,687 milhões de habitantes, de acordo com o censo do IBGE de 2020 (IBGE, 2023a).

Na Subseção 3.1, é feita a descrição dos dados utilizados neste trabalho. Nas subseções seguintes, serão apresentados os resultados da análise desses dados.

#### 3.1 DESCRIÇÃO DOS DADOS E ANÁLISES PRELIMINARES

A cidade de Fortaleza é a capital do estado do Ceará localizada no nordeste brasileiro, com uma população de 2,6 milhões de habitantes. Nesta tese, foram utilizados dois tipos de dados para compreender a dinâmica da transmissão de casos de dengue em Fortaleza: dados de mobilidade humana e de casos de dengue registrados em hospitais e postos de saúde.

Para capturar o movimento humano em Fortaleza, foram utilizados dados de transporte de ônibus do sistema público de transporte nos anos de 2015 e 2020. O sistema de ônibus é o principal meio de transporte em Fortaleza, contando, em 2015, com 356 linhas de ônibus, 2034 ônibus e 700 mil usuários distintos por dia, resultando em uma média de 1 milhão de viagens por dia (CAMINHA *et al.*, 2017). Neste estudo, a mobilidade das pessoas na cidade foi representada por dados fornecidos em (CAMINHA *et al.*, 2016), que incluem os pontos de origem e destino dos usuários do sistema de ônibus. A origem é registrada no momento em que o usuário entra no ônibus, através do passe eletrônico utilizado. O ponto de destino é estimado com base na segunda entrada do mesmo usuário no mesmo dia. Por exemplo, se um usuário embarcou em um ônibus na parada  $BS_1$  pela manhã e entrou em um ônibus na parada  $BS_2$  no final do dia, presume-se que houve uma viagem da origem  $BS_1$  para o destino  $BS_2$ . Se um usuário embarcou em 3 ônibus em

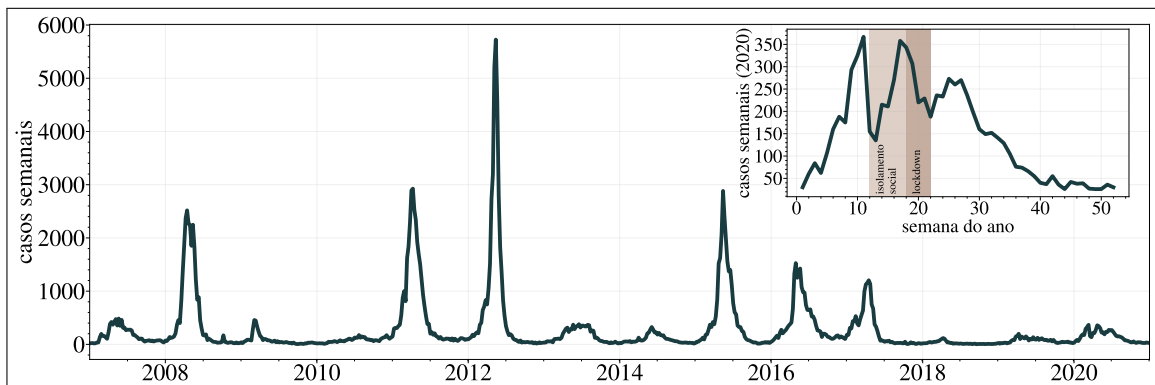
um dia, usando as paradas  $BS_1$ ,  $BS_2$  e  $BS_3$ , são geradas 2 viagens para esse usuário. A primeira viagem começa em  $BS_1$  e termina em  $BS_2$ , e a segunda viagem começa em  $BS_2$  e termina em  $BS_3$ . Mais detalhes sobre a metodologia podem ser encontrados em (CAMINHA *et al.*, 2016). No geral, os autores demonstraram, analisando as viagens realizadas por cerca de 500.000 pessoas por ano e percorrendo 4.783 paradas de ônibus georreferenciadas, que os dados não apresentam viés espacial e representam de forma significativa o movimento da população da cidade. Esses dados têm sido utilizados em estudos recentes com diferentes propósitos, como compreender a criminalidade e identificar problemas de tráfego (CAMINHA; FURTADO, 2017; SULLIVAN *et al.*, 2017; PONTE; CAMINHA; FURTADO, 2016; FURTADO *et al.*, 2017; CAMINHA *et al.*, 2018). Neste estudo, foram utilizadas as coordenadas de cada parada de ônibus para agregar espacialmente os dados de ônibus em nível de bairro. Os dados diários de ônibus foram calculados como uma média ao longo de uma semana. O conjunto de dados resultante, contendo as taxas de fluxo semanais entre os 119 bairros, foi utilizado como entrada para os sistemas de previsão geograficamente vinculados.

Através da Secretaria Municipal de Saúde e do sistema de Monitoramento Diário de Agravos (SIMDA) (SIMDA, 2023), são disponibilizados dados sobre registros de doenças infecciosas em hospitais e postos de saúde da cidade. Para esta tese, foram utilizados os dados de casos de dengue dos anos de 2007 a 2020. O SIMDA fornece os casos agregados semanalmente para os 119 bairros da cidade de Fortaleza.

Na Figura 12, é possível observar o comportamento da série temporal do número de casos semanais de dengue ao longo de um intervalo de quatorze anos. Normalmente, o aumento no número de casos ocorre no início do mês de março, coincidindo com a estação chuvosa. Os meses de abril a junho são geralmente associados à rápida proliferação do mosquito, resultando em altas taxas de contágio. Após esse período, a infestação tende a diminuir. Os anos de 2009, 2011, 2012, 2015, 2016 e 2017 apresentaram endemias, com um maior número de casos de dengue, variando de 13 mil a 37 mil casos. Outros momentos endêmicos foram caracterizados por um menor número de casos de dengue. Em particular, destaca-se o ano de 2018, com 1310 casos de dengue.

O perfil temporal da curva de casos geralmente apresenta uma forma sinuosa com sazonalidade anual, onde um pico de máxima infecção é observado durante o período endêmico. No entanto, em 2020, esse perfil apresentou uma diferença, com o pico máximo fracionado em dois momentos distintos. O primeiro ocorreu após o final de fevereiro, e o segundo durante o período de *lockdown* (ver *inset* da Figura 12). Além disso, a distribuição geográfica dos casos de

**Figura 12 – Visão geral dos registro de dengue na cidade de Fortaleza. É apresentado a série temporal semanal do número de casos de dengue observados na cidade de Fortaleza no intervalo de 2007 a 2020. No gráfico interno destacamos a evolução temporal durante o período de 2020. Em 2020 restrições a mobilidade urbana nomeadas como isolamento social e *lockdown* foram implementadas para mitigar o processo de proliferação do COVID-19 gerando assim um perfil atípico de na série temporal.**



Fonte: Elaborado pelo autor

dengue na cidade é heterogênea.

Ao analisar a distribuição geográfica, observa-se que os casos de dengue estão dispostos de forma heterogênea na cidade. Ao examinar o ranqueamento dos casos por bairros ao longo dos últimos dez anos, nota-se que a ordem não varia significativamente. Para quantificar essa relação, foi utilizado o coeficiente de correlação de *Spearman* (Equação 3.1) (MYERS; SIROIS, 2004). O coeficiente de correlação  $\rho_{x,y}$  é calculado considerando o ranqueamento do número de casos totais nos anos  $x$  e  $y$ , a covariância  $cov(R_x, R_y)$  e o desvio padrão  $\sigma(R_x)$ . Um valor próximo de  $\rho_{x,y} \approx 1$  indica que o ranqueamento do número de casos nos anos  $x$  e  $y$  não sofre alterações significativas, enquanto um valor próximo de  $\rho \approx 0$  indica anos não correlacionados. Na Figura 13, são comparados todos os pares de anos e verificamos que, em média, há uma alta correlação entre os anos analisados, resultando na persistência do ranqueamento dos casos observados para cada bairro. Mesmo em 2020, quando foram implementadas restrições rigorosas de mobilidade, o ranqueamento não se modificou. A fórmula da correlação de Spearman é apresentada abaixo:

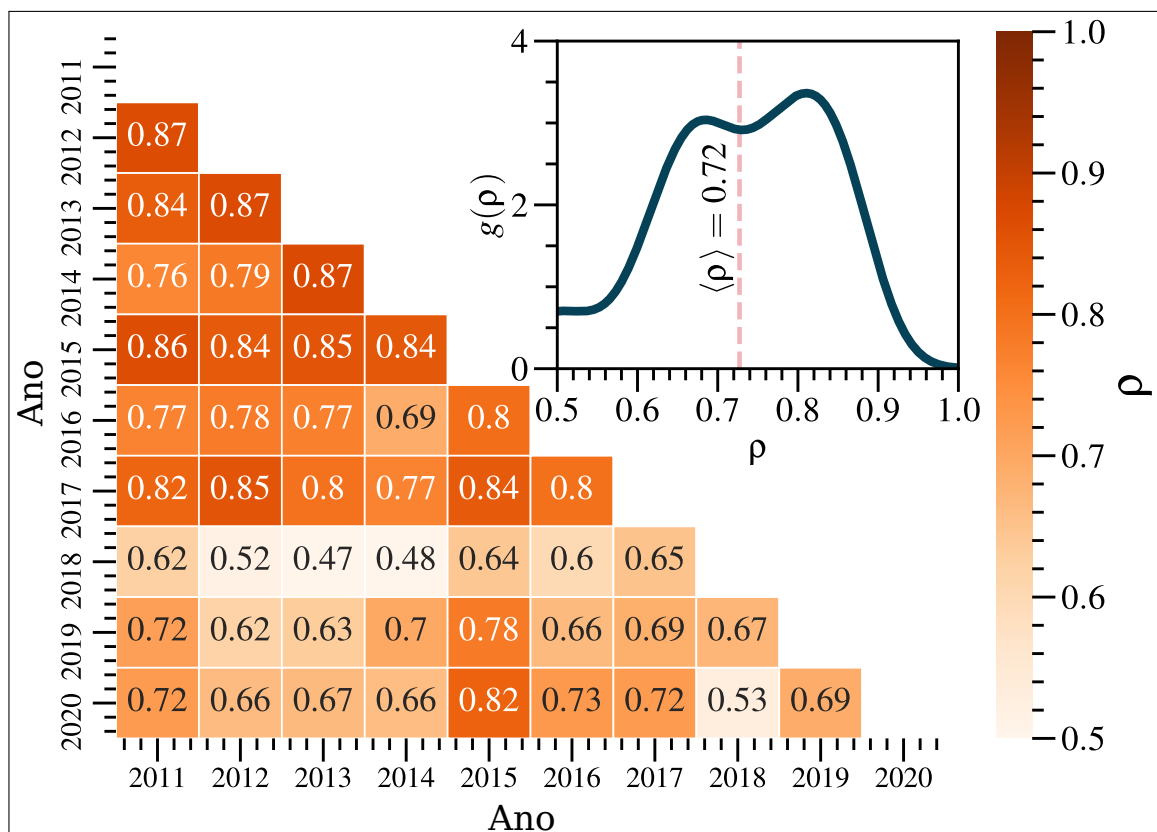
$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (3.1)$$

onde  $\rho$  representa o coeficiente de correlação de Spearman,  $d_i$  são as diferenças dos ranquea-



mentos dos pares de observações e  $n$  é o número total de pares de observações.

**Figura 13 – Correlograma do número total de casos de dengue para cada bairro em diversos anos. Nós calculamos o coeficiente de correlação de Spearman  $\rho$  (Equação 3.1) entre os pares de anos compreendidos no intervalo [2011,2020]. O cálculo de  $\rho$  consiste em comparar o ranqueamento do número total de casos de dengue em cada bairro para um par de anos analisados. Assim, a correlação será alta quando observações tiverem uma classificação semelhante. As análises demonstram que há, efetivamente, uma tendência em manter a ordem de infecção por bairros. A distribuição dos valores de correlação,  $g(\rho)$ , pode ser vista através do *inset* da figura onde  $\langle \rho \rangle = 0.77$  identifica a média de todos os pares de anos analisados.**



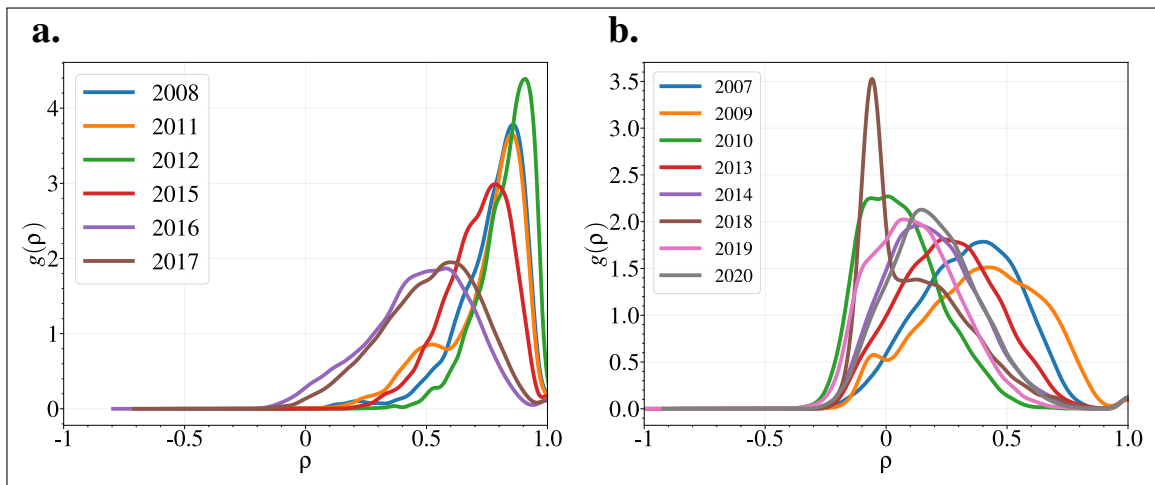
Fonte: Elaborado pelo autor

Já foi constatado que o período de maior incidência de casos em Fortaleza ocorre entre os meses de abril e junho. No entanto, durante os anos epidêmicos, observou-se um crescimento síncrono de casos entre os bairros, ou seja, os aumentos ocorrem em períodos muito próximos uns dos outros. Essa constatação pode ser verificada por meio da correlação de Pearson (Equação 3.2) entre as séries temporais semanais em nível de bairro.

$$\rho_{ij} = \frac{\text{cov}(W_c^{(i)}, W_c^{(j)})}{\text{var}(W_c^{(i)})\text{var}(W_c^{(j)})} \quad (3.2)$$

Assim, foi calculada a correlação de Pearson entre as séries temporais semanais de casos em todos os bairros para cada ano epidêmico e não epidêmico. A Figura 14 ilustra a distribuição dessas correlações. Observa-se que nos anos epidêmicos (2008, 2011, 2012, 2015, 2016 e 2017), o pico ocorre com valores de  $\rho$  mais próximos de 1.0, enquanto nos anos não epidêmicos (2007, 2009, 2010, 2013, 2014, 2019, 2020), os picos de correlação variam entre 0 e 0.5.

**Figura 14 – Distribuições dos valores correlação de Pearson (Equação 3.2) entre as séries temporais para cada bairro em diversos anos. Em (a) é apresentada a distribuição para anos epidêmicos  $Y_{epi} = (2008, 2011, 2012, 2015, 2016, 2017)$  enquanto (b) apresenta resultados para anos não-epidêmicos  $Y_{non-epi} = (2007, 2009, 2010, 2013, 2014, 2019, 2020)$ . De modo geral, anos epidêmicos geram séries síncronas e semelhantes.**



Fonte: Elaborado pelo autor

### 3.2 A IMPORTÂNCIA DA MOBILIDADE HUMANA NA INTENSIDADE DA FORÇA DE CONTÁGIO DO VÍRUS

O ano de 2020 foi atípico devido à pandemia de COVID-19, que levou governos de todo o mundo a impor restrições à locomoção humana nas cidades com o objetivo de conter a transmissão do vírus. No entanto, a redução da mobilidade não afetou apenas a transmissão do vírus da COVID-19, mas também a do vírus da dengue. Durante o ano de 2020, houve uma

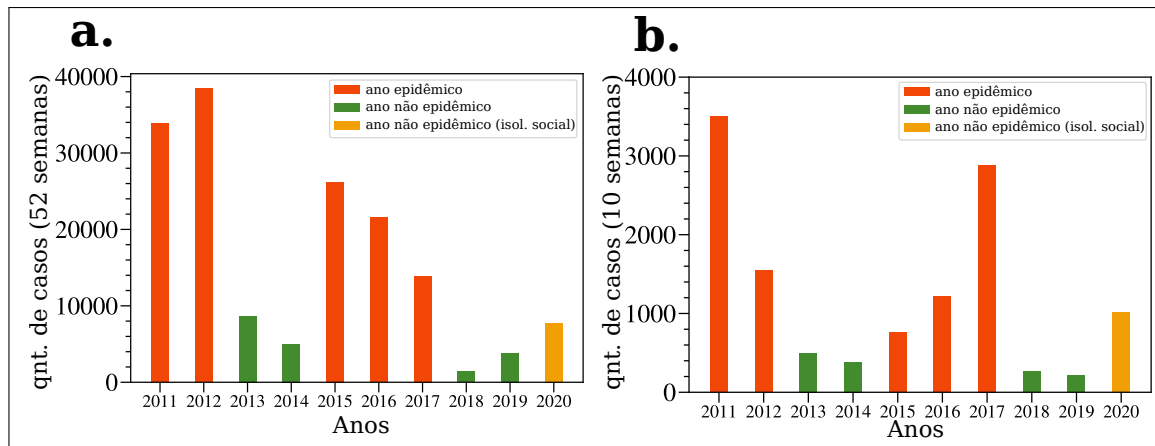
modificação no padrão de evolução dos casos de dengue em comparação com anos anteriores. O *inset* na Figura 12 mostra que nos três primeiros meses do ano houve um rápido aumento nos casos de dengue, comparável aos anos epidêmicos. No entanto, logo após a implementação do isolamento social pelas autoridades públicas, a transmissão de casos de dengue perdeu força, voltando a aumentar logo em seguida e diminuindo novamente após a implementação do *lockdown*. A Figura 15(a) mostra que em 2020 houve um número total de casos de dengue comparável aos anos não epidêmicos. No entanto, ao analisarmos a Figura 15(b), que mostra a quantidade de casos de dengue até a semana anterior à implementação do isolamento social, percebemos que o número de casos é comparável aos anos epidêmicos, indicando que 2020 deveria ter tido quantidade de casos de dengue total similar a anos epidêmicos. Essa característica identificada em 2020 motivou a análise da intensidade de transmissão de casos de dengue em anos epidêmicos e não epidêmicos, a fim de compreender melhor o impacto das restrições de mobilidade durante o ano de 2020.

O impacto das restrições de mobilidade na disseminação interurbana de doenças originárias de um local de surto inicial OL foi estudado por Brockmann e Helbing (BROCKMANN; HELBING, 2013). Eles demonstram que o tempo de chegada da doença  $T_a$  em uma determinada região, a partir de um OL, está negativamente correlacionado com a intensidade do deslocamento do OL para essa região. Consequentemente, a doença tende a chegar mais rapidamente em regiões com maior grau de conectividade efetiva devido à mobilidade. Assim, reduções no número de deslocamentos entre regiões levam a um aumento em  $T_a$  e, como resultado, uma diminuição no número total de casos devido à redução na velocidade de disseminação. Para investigar se esse fenômeno também está presente dentro da mesma cidade, especificamente na escala do bairro, foi examinado o comportamento entre a incidência acumulada de casos em cada bairro e o tempo de chegada da doença. As Figuras 16 (a)-(i) mostram que, de fato, a incidência decai exponencialmente com  $T_a^{(i)}$ , expresso como:

$$\frac{C_T^{(i)} \times 10^4}{N^{(i)}} \propto \exp(-T_a^{(i)} / \tau_y) \quad (3.3)$$

onde  $C_T^{(i)}$  representa o número de casos acumulados para o bairro  $i$ ,  $N^{(i)}$  a população residente do bairro  $i$  e  $\tau_y$  denota um tempo característico para o ano  $y$ . Em cada ano analisado,  $T_a^{(i)}$  representa o número de dias, a partir do primeiro dia do ano, necessários para que um bairro  $i$  atinja  $C^{(*)} = 7$  casos acumulados a cada 10.000 habitantes. A definição para a escolha do valor de  $C^{(*)}$  está presente no Apêndice A. Variações no valor de  $C^{(*)}$  não alteram significativamente

**Figura 15 – Em (a), é mostrado os o número total de casos observados na cidade de Fortaleza em diferentes momentos de contágio. Os anos de 2011, 2012, 2015, 2016 e 2017 são chamados de epidêmicos, pois apresentam uma grande quantidade de casos confirmados. Os anos restantes são considerados não epidêmicos. Em (b), analisamos o número total de casos apenas para as primeiras 10 semanas de infecção por dengue, o período logo antes da aplicação das restrições em 2020. Durante este período, embora 2020 tenha apresentado um padrão de contaminação semelhante aos anos epidêmicos, sugerindo que altas taxas de contágio possam ter ocorrido nas semanas restantes do ano. No entanto, essa estimativa não foi confirmada e, no final do ano, 2020 foi caracterizado como um ano não epidêmico.**

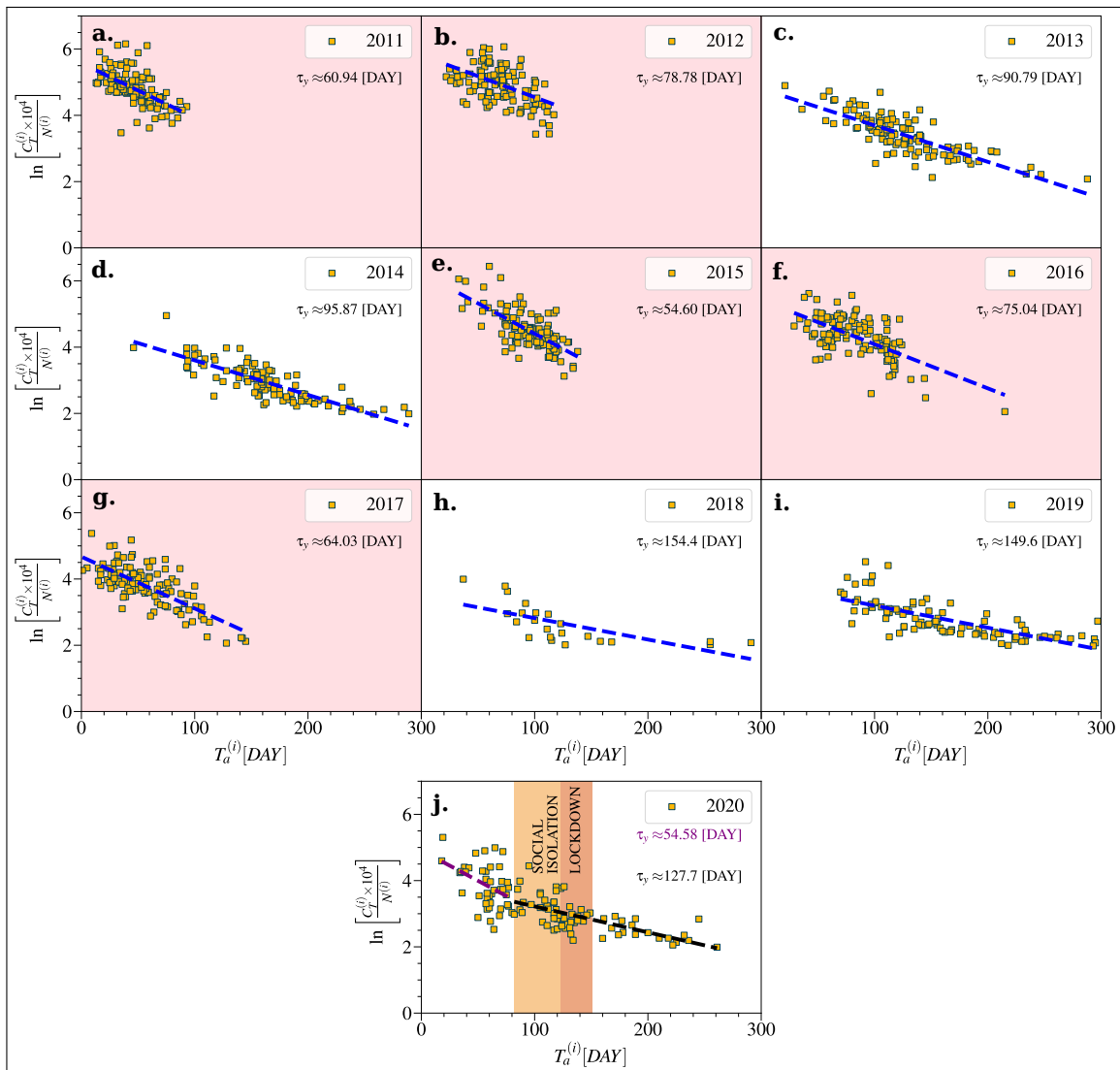


Fonte: Elaborado pelo autor

o comportamento exponencial das curvas plotadas nas Figuras 16 (a)-(i). Desta forma, bairros inicialmente infectados apresentam um número maior de casos por população residente. Esse padrão também é observado em todos os anos anteriores a 2020. Anos epidêmicos são caracterizados por  $\tau_y < 90$  dias, enquanto anos não epidêmicos são bem representados por  $\tau_y > 90$  dias. Em anos epidêmicos, a contaminação ocorre rapidamente e se intensifica nos bairros até o final dos primeiros três meses do ano. Anos não epidêmicos apresentam contaminação tardia em diferentes bairros da cidade. Durante esses anos, alguns surtos ocorrem após a estação chuvosa, e a contaminação em todos os bairros ocorre em julho.

Admitindo que  $\tau_y$  está relacionado efetivamente ao tempo característico de propagação da dengue dentro da cidade. Em anos epidêmicos, quando a doença se espalha rapidamente, são observados valores menores de  $\tau_y$ . A rápida propagação atinge mais pessoas suscetíveis e produz um maior número de casos. A queda exponencial observada na Equação 3.3 para o ano

**Figura 16 – Relação exponencial entre a incidência de casos  $\left(\frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)$  e o tempo de chegada ( $T_a^i$ ) do vírus da dengue para cada bairro na cidade de Fortaleza. Em (a)-(j), são apresentadas as relações  $\left(T_a^i \times \frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)$ . As linhas tracejadas representam ajustes pela Equação 3.3. As figuras com o quadro em vermelho representam anos epidêmicos.**



Fonte: Elaborado pelo autor

de 2020 é diferente das anteriores. A Figura 16(j) mostra zonas de diferentes regimes na relação  $T_a^i \times \log\left(\frac{C_T^{(i)}}{N^{(i)}}\right)$ . Os bairros que foram infectados antes do período de restrição apresentam uma propagação mais rápida em comparação aos infectados posteriormente, o que implica na redução do número total de incidência. A interrupção na taxa de contágio, como visto na Figura 16(j), é uma forte evidência de que a mobilidade urbana desempenha um papel relevante nos níveis de contaminação de dengue na cidade.

No estudo realizado, constatou-se que a implementação das medidas de isolamento

social na cidade de Fortaleza teve impacto significativo na propagação do vírus da dengue. Os resultados indicam que o número de casos registrados no ano de 2020 foi menor do que o esperado, levando em consideração o padrão histórico da doença.

## 4 PREVISÃO DE CASOS DE DENGUE

Caracterizar os dados de casos de dengue e mobilidade humana é uma etapa essencial, pois isso auxilia na compreensão da dinâmica da transmissão do vírus em ambientes urbanos. No entanto, prever tais casos é igualmente importante para permitir que as autoridades adotem medidas de mitigação diante das possíveis consequências de uma epidemia.

Nesta seção, utilizaremos métodos de redes neurais para realizar previsões semanais de casos de dengue em nível de bairro. Essas previsões serão comparadas com os resultados com métodos tradicionais para a previsão de dados epidemiológicos.

### 4.1 DADOS DE MOBILIDADE HUMANA COMO ATRIBUTO DE UMA REDE NEURAL

Na seção 3.2, foi concluído que as restrições de mobilidade decorrentes da pandemia de COVID-19 foram capazes de reduzir a velocidade de transmissão dos casos de dengue em Fortaleza, resultando em uma diminuição no número total de casos. Esta conclusão enfatiza a importância da utilização do dado de mobilidade humana em modelos preditivos para a previsão de casos, porém incluir esses dados em modelos preditivos não é uma atividade trivial.

Nesta seção, Serão propostas arquiteturas de rede neural capaz de aprender o impacto da mobilidade humana na transmissão de casos de dengue e de prever os casos semanais de dengue. Essa arquitetura será comparada com outras duas abordagens: o modelo SIR, um modelo comportamental amplamente utilizado para estudar a dinâmica de transmissão de doenças infecciosas, e o modelo ARIMA, um modelo estatístico de referência.

#### 4.1.1 A arquitetura de rede neural proposta

Estou propondo duas arquiteturas de Recurrent Neural Networks (RNN) para determinar o impacto do transporte público na previsão da propagação do vírus da dengue: a primeira, RNNi, considera os casos de dengue em cada bairro isoladamente, enquanto a segunda, RNNc, utiliza os dados de transporte de ônibus para conectar os bairros.

A arquitetura RNNi prevê casos de dengue para um bairro específico com base apenas em seus próprios dados de casos de dengue passados, utilizando a técnica de Long short-term memory (LSTM). Esse modelo foi desenvolvido para cada um dos 119 bairros de Fortaleza, utilizando duas entradas: uma série temporal com o número semanal de casos de dengue para um bairro (com uma janela de entrada de 5 semanas) e um vetor índice, variando de

1 a 52, que rotula cada uma dessas cinco semanas. Esse vetor é utilizado como entrada para uma camada de *embedding*, que gera uma saída a ser concatenada com os casos de dengue. Os dados concatenados são então alimentados em uma camada LSTM, que se conecta a uma camada de saída (camada densa) com uma função de ativação sigmooidal, gerando a previsão dos casos de dengue para a semana seguinte.

Camadas de *dropout* foram adicionadas para evitar o *overfitting*. Essas camadas forçam, de forma aleatória, uma porcentagem determinada de neurônios da camada anterior a serem ignorados durante cada etapa de treinamento, a fim de evitar o *overfitting* por meio de um único caminho de neurônios. A Figura 17 ilustra a arquitetura do modelo RNNi, com suas camadas de entrada e saída. As camadas de *dropout* permitem avaliar a variabilidade das previsões. Para gerar uma previsão probabilística, utiliza-se o método de Monte Carlo *Dropout* (MC *Dropout*) (GAL; GHAHRAMANI, 2016), gerando  $n = 100$  realizações de previsões, atribuindo aleatoriamente as probabilidades dos *dropouts* (taxa de *dropout*) após cada camada.

O modelo RNNi deve ser capaz de prever várias semanas a frente para fornecer valores de previsões com antecedência. Existem várias estratégias para lidar com o problema de previsão de múltiplos passos, como recursividade, direta, direta-recursiva e estratégia de saída múltipla (BONTEMPI; TAIEB; BORGNE, 2012; TAIEB; HYNDMAN *et al.*, 2012). Aqui, utilizamos a estratégia recursiva, em que a primeira previsão é baseada nas incidências observadas e os valores previstos são usados recursivamente para prever o restante da série temporal.

A segunda arquitetura, RNNc, prevê os casos de dengue para todos os bairros simultaneamente, utilizando dados históricos de dengue, conforme descrito na arquitetura isolada, juntamente com a mobilidade humana. A mobilidade humana entre bairros foi modelada como um grafo (Figura 18), onde os nós são os bairros e os pesos das arestas representam uma taxa de fluxo (Eq. 4.1) representando o fluxo de pessoas ( $flowrate_{i,k,w}$ ) do bairro  $i$  para o bairro  $k$  na semana  $w$ , calculado como:

$$flowrate_{i,k,w} = \frac{count(flow_{i,k,w})}{\sum_{i=0}^{TN} count(flow_{i,k,w})} \quad (4.1)$$

onde  $count(flow_{i,k,w})$  é o número de viagens de ônibus do bairro  $i$  para o bairro  $k$  na semana  $w$ .  $TN$  é o número total de bairros na cidade.

Uma representação de cada nó do grafo foi gerada utilizando técnicas de *embedding*



de nós chamada Node2Vec e, juntamente com os dados históricos de dengue, foi utilizada como entrada para uma arquitetura LSTM hierárquica.

O algoritmo de *embedding* de nós, Node2Vec (GROVER; LESKOVEC, 2016; MIKOLOV *et al.*, 2013), foi utilizado para aprender as características de cada nó no grafo de mobilidade humana. O Node2Vec utiliza uma rede neural para capturar o contexto em torno do nó, gerando representações vetoriais unidimensionais dos nós do grafo, chamadas de *embedding*. No caso deste trabalho, cada vetor produzido pelo Node2Vec é uma sequência de bairros de comprimento pré-determinado. Essa sequência é gerada por meio de passeios aleatórios ponderados para nós adjacentes. A ideia fundamental por trás das técnicas de *embedding* baseadas em passeios aleatórios é transformar o grafo em uma coleção de sequências de nós, em que a frequência de ocorrência de um par nó-contexto reflete a distância estrutural entre eles.

Diversos tamanhos para a representação vetorial do *embedding* foram testados, e o tamanho 15 foi selecionado para o *embedding* de cada bairro. Esse *embedding* foi utilizado como entrada para o modelo, juntamente com os casos semanais de dengue de todos os bairros, além do *embedding* representando a semana (utilizando o vetor de índice descrito anteriormente no modelo RNNi). Esses três componentes alimentam três camadas LSTM, que visam capturar a relação espaço-temporal dos casos de dengue. Dessa forma, o modelo RNNc (Figura 19) realiza previsões dos casos semanais de dengue para todos os bairros simultaneamente, com base nos dados de incidência de dengue em nível de bairro em toda a cidade, juntamente com os dados de mobilidade humana.

A validação cruzada foi aplicada a 9 anos de dados, de 2007 a 2015. Para cada ano a ser previsto, dois anos foram utilizados para avaliar o modelo: um ano epidêmico e outro sem epidemia. Os dados restantes foram usados para treinar o modelo. A distribuição dos anos utilizados para definir a base de validação e treinamento para cada ano de teste está apresentada na Tabela 2.

#### **4.1.2 Modelo mecanístico**

Para avaliar ainda mais o impacto da mobilidade humana na previsão de dengue, foi realizado um segundo conjunto de comparações. Geramos previsões usando um sistema desenvolvido com um modelo mecanístico, com e sem a inclusão de dados de mobilidade humana. Devido à interpretabilidade dos modelos mecanísticos, sistemas de previsão baseados em processos têm sido amplamente utilizados na previsão de surtos de doenças infecciosas (SHA-

Test	Evaluation	Training
2007	2010, 2012	2008, 2011, 2012, 2013, 2014, 2015
2008	2010, 2012	2007, 2011, 2012, 2013, 2014, 2015
2009	2010, 2012	2007, 2008, 2011, 2012, 2013, 2014, 2015
2010	2007, 2012	2008, 2011, 2012, 2013, 2014, 2015
2011	2010, 2012	2007, 2008, 2012, 2013, 2014, 2015
2012	2007, 2015	2007, 2008, 2011, 2013, 2014, 2015
2013	2010, 2012	2007, 2008, 2011, 2012, 2014, 2015
2014	2010, 2012	2007, 2008, 2011, 2012, 2013, 2015
2015	2010, 2012	2007, 2008, 2011, 2012, 2013, 2014

**Tabela 2 – Divisão de dados para validação cruzada. Os dados de validação são usados para ajustar os parâmetros do modelo durante a fase de treinamento, enquanto os dados de teste são usados para avaliar o desempenho final do modelo.**

MAN; KARSPECK, 2012; SHAMAN *et al.*, 2013; PEI; SHAMAN, 2017; YANG; LIPSITCH; SHAMAN, 2015; PEI; CANE; SHAMAN, 2019), incluindo dengue (YAMANA; KANDULA; SHAMAN, 2016), Ebola (YANG *et al.*, 2015), doenças semelhantes à influenza (KANDULA *et al.*, 2018) e patógenos resistentes a antibióticos (PEI *et al.*, 2018b).

O sistema de previsão local, MECHi, utiliza dados de incidência de cada bairro e prevê as curvas epidêmicas de forma independente para diferentes locais. Devido à falta de dados detalhados sobre mosquitos, modelamos a transmissão do vírus da dengue em cada bairro usando um modelo suscetível-exposto-infectado-recuperado (S.E.I.R.) parcimonioso:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \quad (4.2)$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \frac{E}{Z}, \quad (4.3)$$

$$\frac{dI}{dt} = \frac{E}{Z} - \frac{I}{D}, \quad (4.4)$$

onde  $N$ ,  $S$ ,  $E$  e  $I$  são o número total de pessoas, suscetíveis, expostas e infectadas;  $\beta$  é a taxa de transmissão;  $Z$  é o tempo de incubação; e  $D$  é o período infeccioso. Para o sistema de previsão em rede, MECHc, conectamos bairros considerando a mistura da população devido ao transporte de ônibus. A transmissão entre 119 bairros é descrita por um modelo S.E.I.R. em rede (YANG *et al.*, 2015):

$$\frac{dS_i}{dt} = -S_i \sum_j \frac{c_{ji} \beta_j I_j}{\hat{N}_j}, \quad (4.5)$$

$$\frac{dE_i}{dt} = S_i \sum_j \frac{c_{ji} \beta_j I_j}{\hat{N}_j} - \frac{E_i}{Z}, \quad (4.6)$$

$$\frac{dI_i}{dt} = \frac{E_i}{Z} - \frac{I_i}{D}. \quad (4.7)$$

Aqui  $S_i$ ,  $E_i$  e  $I_i$  são o número de pessoas suscetíveis, expostas e infectadas no bairro  $i$ ;  $\beta_j$  é a taxa de transmissão no local  $j$ ;  $c_{ji}$  é a fração da população que viaja do bairro  $i$  para  $j$ ; e  $\hat{N}_j = \sum_i c_{ji} N_i$  é a população total presente no local  $j$ . As informações de mobilidade humana  $c_{ji}$  são obtidas a partir dos dados de transporte de ônibus. Em ambos os modelos, os intervalos iniciais das variáveis e parâmetros são definidos da seguinte forma:  $S \in [0.3, 0.55]N$ ,  $E \in [0, 0.0005]N$ ,  $I \in [0, 0.0005]N$ ,  $\beta \in [0.3, 0.5]$ ,  $Z \in [3, 7]$  dias,  $D \in [3, 7]$  dias, e  $N$  é definido como a população em cada bairro.

Para estimar as variáveis e parâmetros do modelo usando incidências observadas, utilizamos um algoritmo de assimilação de dados - o *ensemble adjustment Kalman filter* (EAKF) (ANDERSON, 2001). Especificamente, as variáveis e parâmetros do modelo são ajustados iterativamente usando observações desde o início da temporada até o momento atual, e o modelo otimizado é integrado ao futuro para gerar previsões. Com o EAKF, a distribuição do estado do modelo é representada por um conjunto de vetores de estado. Como resultado, o sistema de previsão é capaz de gerar previsões probabilísticas. Para nossa implementação, usamos 300 membros do conjunto. Estruturas semelhantes de assimilação de dados-modelo têm sido aplicadas com sucesso na previsão e inferência de influenza (SHAMAN; KARSPECK, 2012; SHAMAN *et al.*, 2013; PEI; SHAMAN, 2017; YANG; LIPSITCH; SHAMAN, 2015; PEI; CANE; SHAMAN, 2019), dengue (YAMANA; KANDULA; SHAMAN, 2016), Ebola (YANG *et al.*, 2015), doenças semelhantes à influenza (KANDULA *et al.*, 2018) e patógenos resistentes a antibióticos (PEI *et al.*, 2018b). Em particular, um sistema de previsão em rede semelhante foi usado recentemente para prever a transmissão espacial de influenza nos Estados Unidos (PEI *et al.*, 2018a). No material suplementar, há exemplos de previsões para os modelos mecanístico e de rede neural.

### 4.1.3 Métodos de avaliação

Para examinar quantas semanas com antecedência os modelos podem prever os picos de casos de dengue, as previsões foram geradas a partir da semana 6 até a semana 25 do ano,

considerando que a partir da semana 25 o período de maior contaminação de dengue passou. Para cada semana de previsão, foram geradas previsões para o restante do ano para cada bairro, com três objetivos: classificar anos e bairros com e sem surtos intensos, prever a intensidade e o tempo de pico e prever toda a série temporal de casos de dengue para surtos intensos. Bairros com mais de 200 casos de dengue em um ano foram classificados como bairros com quantidade intensa de casos de dengue, ou seja grande surto de casos. Além disso, testamos a sensibilidade desse valor de limite. Para medir a classificação de surtos como intensos ou não, foi utilizada métricas de precisão e revocação da seguinte forma:

$$preciso = \frac{VP}{VP + FP}, \quad (4.8)$$

$$revocacao = \frac{VP}{VP + FN}. \quad (4.9)$$

onde ( $VP$ ) é verdadeiro positivo, ( $VN$ ) é verdadeiro negativo, ( $FP$ ) representa falso positivo, e ( $FN$ ) falso negativo.

Para avaliar as previsões do ponto de vista da intensidade do pico ( $PI$ ) e tempo do pico ( $PT$ ), foi calculado o erro absoluto médio ( $MAE$ ) como a diferença entre o  $PI$  e  $PT$  predito e real respectivamente:

$$PI = \frac{\sum_{k=0}^K |y_k - \hat{y}_k|}{K}, \quad (4.10)$$

$$PT = \frac{\sum_{k=0}^K |r_k - \hat{r}_k|}{K}. \quad (4.11)$$

onde, para  $PI$ ,  $y$  é o número real de casos de dengue durante a semana de pico e  $\hat{y}$  é o número previsto de casos de dengue durante a semana de pico. Para  $PT$ ,  $r$  é o índice da semana de pico da série temporal real,  $\hat{r}$  é o índice da semana de pico da série temporal prevista e  $K$  é o número total de surtos com casos anuais de infecções superiores a 200. Para medir o erro da previsão da série temporal, o erro médio quadrático ( $RMSE$ ) foi calculado entre a série temporal real e prevista de casos de dengue para cada grande surto. Em geral, a série temporal de casos de dengue em Fortaleza mostra poucos picos no ano. Portanto, o aspecto mais importante para um modelo de previsão é identificar o pico e suas variações.  $RMSE$  (Equação.4.12) é a métrica

mais apropriada para capturar isso, pois a diferença entre o valor previsto e o real é elevada ao quadrado antes que a média seja retida. Isso faz com que grandes diferenças sejam destacadas como previsões ruins durante períodos críticos, quando a incidência é alta, sejam penalizadas.

$$RMSE = \frac{\sum_{w=f}^W (x_w - \hat{x}_w)^2}{W - f} \quad (4.12)$$

$x$  é o número total de casos de dengue no surto real,  $\hat{x}$  é o número total de casos de dengue no surto previsto,  $W$  é o número total de semanas no ano e  $f$  é o tempo de previsão, a semana do ano em que a previsão foi iniciada. O RMSE é medido para cada grande surto e, finalmente, a média de todos os surtos é calculada.

Como comparação final, os resultados dos modelos RNNi e MECHi foram comparados com um modelo ARIMA (Autoregressive Integrated Moving Average) (BOX *et al.*, 2015). O ARIMA é um modelo estatístico univariado amplamente estudado para previsão de séries temporais e usado em várias áreas. (ARIYO; ADEWUMI; AYO, 2014; RAHMAN *et al.*, 2013; HE; TAO, 2018).

Por meio da análise de auto-correlação (detalhada no material suplementar de auto-correlação), observou-se que o número de casos de dengue apresenta alta correlação com valores anteriores, ou seja, com defasagens pequenas. Assim, foi realizada uma análise ARIMA para o valor de auto-correlação 5 (AR(5)) e valor de média móvel 1 (MA(1)).

Avaliamos as previsões probabilísticas calculando o seu escore de log, um método comumente usado para avaliar previsões (JOHANSSON *et al.*, 2019; REICH *et al.*, 2019; BIGGERSTAFF *et al.*, 2018). Para a intensidade do pico, o escore de log da previsão de um surto intenso é o valor logarítmico da porcentagem de previsões que estão dentro de um intervalo de  $\pm 10$  casos em torno do número real de casos. Para o tempo de pico, o escore de log é o valor logarítmico da porcentagem de previsões que estão dentro de um intervalo de  $\pm 1$  semana do pico real, e para a previsão da série temporal, a média dos valores logarítmicos da porcentagem de previsões de casos semanais de dengue para as semanas restantes do período que estão dentro de  $\pm 10$  casos dos casos observados. Cada medida foi calculada com base nas previsões semanais para todos os surtos intensos em todos os bairros. Por definição, um escore de log maior indica um melhor desempenho da previsão. Quando a porcentagem média de previsões é zero, o valor padrão de penalização de -5 é usado.

#### 4.1.4 Predições e comparações

Como referência, foram comparadas as previsões mecánísticas isoladas e de RNN com uma previsão padrão de ARIMA,  $ARIMA_i$  para previsão local e  $ARIMA_c$  (usando ARIMAX) para previsão considerando todos os bairros. A Figura 20 mostra os resultados de precisão e revocação para anos e bairros com e sem epidemias intensas obtidas de todos os modelos. Precisão e revocação são métricas de avaliação para problemas de classificação. Neste caso, precisão mede a fração de classificação correta de epidemias intensas, casos de dengue  $> 200$ , de todos os casos previstos como epidemias intensas, e o revocação mede as epidemias intensas corretamente classificadas de todas as epidemias intensas que realmente aconteceram. O mesmo vale para classificações de não epidemia, casos de dengue  $\leq 200$ .

No problema de classificação de epidemias de dengue, ambos os modelos obtiveram melhor desempenho ao considerar dados de transporte ( $RNN_c$  e  $MECH_c$ ). Embora o modelo  $MECH_c$  tenha tido menos precisão do que o modelo  $RNN_c$ , o modelo mecánístico exibiu um revocação mais alto em tempos de previsão mais próximos do início do ano, indicando que o modelo  $MECH_c$  prevê melhor grandes surtos (alta revocação para casos de dengue  $> 200$ ). Isso implica que o modelo  $MECH_c$  em rede tende a prever grandes surtos, mas superestimar alguns pequenos surtos. Em geral, tanto o modelo  $RNN_c$  quanto o modelo  $MECH_c$  podem classificar o tamanho do surto de dengue com um desempenho satisfatório. Quanto ao modelo gerado por  $ARIMA_i$ , há uma alta revocação para pequenos surtos (Figura 20b). Isso ocorre porque, nesses casos, não há requisitos para o modelo capturar variações importantes e, portanto, o modelo acaba tendo um bom desempenho em períodos em que não houve pico de casos de dengue. No entanto, quando analisamos a revocação desse modelo em situações com grandes picos (Figura 20e), os resultados não são bons. O comportamento é particularmente ruim para o tempo de previsão inicial. O mesmo comportamento é observado ao analisar os valores de precisão. Como a maioria das previsões feitas por  $ARIMA_i$  foi definida como períodos sem picos de casos de dengue, então os valores de precisão esperados são baixos para períodos com casos de dengue  $\leq 200$  e para períodos com casos de dengue  $> 200$ , os valores altos, dado que poucos períodos foram previstos como grandes surtos. Os resultados para o modelo  $ARIMA_c$  são ainda piores do que os resultados do  $ARIMA_i$ , indicando que, quando a complexidade do problema é aumentada, modelos do tipo ARIMA falham em apresentar resultados razoáveis. A análise quanto à precisão e à revocação pode ser resumida nos gráficos de f-score, mostrando que o uso de dados de transporte para grandes e pequenos surtos permitiu uma melhoria nos resultados dos

modelos RNNc e MECHc, quando comparados com dados de grandes surtos, o que não ocorreu com os modelos ARIMA.

É de interesse prático examinar se esses modelos podem prever o momento e a intensidade dos surtos. Dado que nosso objetivo é prever picos de surto, a próxima análise se concentra em surtos em bairros com um número total de casos de dengue superior a 200, já que surtos abaixo desse limiar muitas vezes não tiveram um pico claramente definido. A Figura 21 mostra o erro médio quadrático (MAE) da intensidade e do momento de pico para anos e bairros com epidemias intensas, casos de dengue  $> 200$ . Tanto para a intensidade quanto para o momento de pico, a previsão RNNi apresentou MAE maior do que a previsão MECHi durante a primeira metade da temporada. A adição de dados de mobilidade reduziu significativamente o MAE na previsão mecanística para ambos os objetivos. Para a previsão RNNc, os dados de mobilidade levaram a uma pequena diminuição no MAE para a intensidade de pico e nenhuma vantagem clara no MAE do tempo de pico. A previsão RNNc teve menor MAE do que a previsão MECHc para o objetivo de intensidade de pico, mas os dois métodos apresentaram MAE semelhantes para o objetivo de tempo de pico. Mesmo com resultados semelhantes aos de MECHi para a intensidade de pico, ARIMAi e ARIMAc apresentaram os piores resultados para a intensidade de pico e resultados muito piores para o tempo de pico.

A Figura 22 mostra o RMSE para as previsões de séries temporais. Novamente, é mostrado que a mobilidade melhorou as previsões, levando a um RMSE menor tanto para as previsões RNN quanto para os modelos mecanísticos em comparação com as previsões isoladas. Comparando a versão vinculada dos dois métodos de previsão, mostra que o RNNc tem um RMSE menor do que as previsões MECHc. Esses resultados reforçam a suposição apresentada na avaliação de classificação: os modelos mecanísticos são inclinados a superestimar pequenos surtos, mas capturam melhor os surtos intensos. No entanto, ao prever toda a série temporal, modelos mecanísticos, com menos parâmetros, são menos flexíveis e, portanto, menos capazes de simular curvas epidêmicas não lineares. Os resultados ruins de RMSE para ARIMAc reforçam que os modelos ARIMA não são capazes de aprender problemas complexos como a relação entre fluxo de pessoas e casos de dengue.

A Figura 23 mostra os resultados dos escores médios de log para a intensidade de pico, tempo de pico e previsão de toda a série temporal. Com essa métrica, que avalia a precisão da previsão probabilística, o efeito dos dados de mobilidade foi menos direto. Ambos os tipos de modelos geralmente se beneficiaram dos dados de mobilidade na previsão do tempo de pico. No entanto, a inclusão dos dados de mobilidade levou a pontuações de log mais baixas para

o tempo de pico e, em determinados momentos, para a série temporal de incidência futura de dengue. Para a intensidade de pico, o modelo mecânico mostra resultados mais consistentes, apresentando valores mais altos tanto para a previsão local quanto para a rede. Os modelos de rede neural mostram claramente melhores resultados para a previsão de tempo de pico com base em MAE e pontuação de log. Esse resultado também se mantém para a previsão da série temporal como um todo.

Foi testada a sensibilidade da análise à escolha de 200 casos como o limiar de epidemia em nível de bairro. Avaliamos a capacidade dos modelos de prever casos de dengue em um período com casos de dengue maiores que 100 e maiores que 300. Os resultados reforçam as descobertas de nossa análise principal: o modelo mecânico prevê melhor grandes epidemias, mas superestima períodos com menos casos de dengue, e o modelo de rede neural continua a prever melhor o tempo de pico em todas as análises. Os resultados para os testes com valores de 100 e 300 estão disponíveis no Apêndice B.

Para entender as razões pelas quais o modelo de rede neural que usa representações vetoriais da rede de fluxo de pessoas obteve melhores resultados na previsão da série temporal como um todo (Figura 22) e na previsão da intensidade de pico de casos de dengue (Figura 21a), investiguei mais profundamente o cenário espaço-temporal.

A Figura 24 mostra a distribuição espacial de RMSE para as previsões feitas usando o modelo RNNc durante anos epidêmicos. Nestes anos, o RMSE para bairros no oeste da cidade (lado esquerdo) é menor. Esta região é composta pelos bairros mais populosos com os indicadores de pobreza mais elevados. As figuras apresentadas no Apêndice D indicam que o oeste da cidade é a região onde os casos de dengue aparecem primeiro nos anos de pico. Isso sugere que o uso de dados de mobilidade humana é importante para capturar o fluxo do vírus da dengue do oeste para o leste da cidade. A Figura 25 mostra que ao usar dados de mobilidade humana houve uma redução do erro de previsão em bairros do leste, especialmente para a intensidade de pico e previsão da série temporal como um todo.

Os bairros com previsões mais precisas estão localizados no oeste da cidade, a região mais populosa e onde ocorreram os primeiros casos de dengue em anos de pico. Isso indica que o modelo de rede neural teve melhor desempenho na previsão da incidência durante o estágio inicial de um surto com uma tendência clara de crescimento. Esse resultado sugere a necessidade de fortalecer ainda mais a compreensão da propagação da dengue por meio de dados de mobilidade humana e de prever melhor os casos de dengue em bairros que não apresentem tendências tão claras. Ao mesmo tempo, no modelo RNNc houve uma melhoria na precisão das



previsões para bairros que não estavam na região oeste, indicando que o modelo de rede neural foi capaz de capturar os aspectos espaço-temporais da transmissão da dengue.

Na comparação entre modelos baseados em rede neural e comportamentais conclui-se que as redes neurais, que possuem mais parâmetros e uma estrutura mais complexa, são capazes de capturar uma ampla gama de dinâmicas de transmissão não lineares. Essa diferença pode explicar por que as redes neurais superam os modelos comportamentais na previsão da intensidade máxima do surto de dengue e na série temporal de incidência. Por fim, exemplos de comparações entre dados previstos e reais para os modelos RNNc e MECHc pode ser visualizados no Apêndice C.

#### 4.2 A POPULAÇÃO RESIDENTE COMO FATOR DE IMPACTO NA PREDIÇÃO DE CASOS A NÍVEL DE BAIRRO

O aparecimento de casos de doenças infecciosas em uma grande cidade, normalmente não ocorre de forma homogênea no espaço e no tempo, pois varia em função de características sócio-ambientais de cada parte da cidade. A tarefa de construir modelos preditivos nesse contexto torna-se mais complexa devido à dificuldade de mapear os impactos dessas diferentes características. Não é só a movimentação das pessoas em uma cidade que é importante na transmissão de doenças infecciosas, mas a população de um bairro é um outro exemplo de fator social que pode influenciar, por exemplo, na velocidade de contágio da doença.

Modelos preditivos de doenças infecciosas, baseados em dados históricos, devem levar em conta as particularidades das diferentes regiões da cidade e as interações que ocorrem entre elas para poder capturar a heterogeneidade típica do fenômeno. No entanto, a multifatoriedade inerente a esse processo é difícil de ser capturada, principalmente pelo fato de que nem todos os dados desses diferentes fatores são registrados com precisão e frequência altas. No caso do contágio da dengue, por exemplo, uma doença que se transmite por meio de um vetor (e.g. *Aedes aegypti*), o índice de infestação do mosquito em cada setor censitário ou mesmo bairro da cidade não está sempre disponível (MANRIQUE-SAIDE *et al.*, 2014; PILGER *et al.*, 2011).

Apesar dos resultados promissores apresentados na seção anterior utilizando dados de mobilidade humana em modelos baseados em redes neurais, a dinâmica de movimentação das pessoas não está sempre disponível também, o que força os modelos preditivos de casos de doenças infecciosas a serem treinados apenas com base no histórico de casos passado. Porém, isso inclui um novo problema: quando o treinamento desses modelos é feito em granularidade

espaço-temporal pequena (por bairros, por exemplo) e apenas com dados de histórico de casos de dengue, a interação entre as regiões tende a enviesar os resultados, fazendo com que as previsões para cada bairro sejam homogêneas, obtendo tendências a valores médios. Isso ocorre principalmente para os dados de dengue utilizados neste trabalho, que apresentam forte correlação entre as séries temporais dos bairros, situação já demonstrada na Figura 14.

Assim, um dos desafios que este trabalho aborda também é o de criar modelos preditivos epidemiológicos onde essa natureza espaço-temporal precisa ser levada em conta, mas pouca informação sobre a mesma é disponível. Tendo em vista que a população residente dos bairros está correlacionada com a quantidade total de casos de dengue de um bairro em um ano, é importante estudar formas de utilizar essa informação, como atributo para modelos preditivos, com o objetivo de incorporar ao modelo a importância de cada bairro na transmissão geral de casos de dengue em uma cidade.

Visando concretizar a ideia levantada no parágrafo anterior, será utilizado um conjunto de dados georreferenciados de mais de dez anos de ocorrências de casos de dengue na cidade de Fortaleza. A tarefa consiste em prever, a partir dos dados das primeiras semanas do ano, para cada um dos 119 bairros da cidade e para a cidade como um todo, a quantidade de casos de dengue do restante do ano, alertando os bairros que provavelmente serão epidêmicos. Propõe-se aqui usar uma informação heurística que captura a intensidade da propagação da doença em cada bairro normalizada pela população do bairro num modelo de rede neural baseado em LSTM.

#### **4.2.1 Função de controle de intensidade como atributos de uma rede neural para predição de casos de dengue**

A dengue é uma doença viral transmitida por mosquitos que representa uma ameaça significativa à saúde pública em muitas regiões do mundo. Sua transmissão é influenciada por uma série de fatores, incluindo a presença de vetores de mosquitos, fatores ambientais e climáticos, além da mobilidade urbana. Como resultado, a propagação da dengue em áreas urbanas pode ser particularmente desafiadora devido à complexidade do ambiente urbano e à interação entre os fatores que influenciam a transmissão da doença. No entanto, esses fatores estão tipicamente relacionados à população total do ambiente urbano.

É sabido que a quantidade de casos de dengue em cada bairro escala-se com a população local. Observa-se na Figura 26 uma correlação linear entre o número total de casos ( $C_T^{(i)}$ ) de cada ano  $T$ , em cada bairro  $i$  com a população residente do bairro  $N_i$ . Sendo assim, de

modo geral, a população local pondera a importância de cada bairro no valor da quantidade total de casos e regula a taxa de crescimento de casos per-capita observada na cidade.

Assim a população local torna-se uma variável relevante para controlar a intensidade de propagação da dengue, uma vez que a série temporal de casos de dengue semanal  $W_c(t)$  pode ser integrada da seguinte forma:

$$C(t) = \sum_{t_0 \leq t} W_c(t_0), \quad (4.13)$$

onde  $C(t)$  é a série temporal do número de casos acumulados na semana  $t = 1, 2, 3, \dots, t_f$  e  $C(t = t_f) = C_T$ . A Equação 4.13 gera um comportamento similar a uma função logística dado por:

$$C(t) = \frac{L}{1 + \exp(-k(t - t_{cri}))}, \quad (4.14)$$

que pode ser modificado para a seguinte forma:

$$\ln\left(\frac{C(t)}{L - C(t)}\right) = k(t - t_{cri}). \quad (4.15)$$

Considerando apenas o lado esquerdo da Equação 4.15, temos que:

$$h(L) = \ln\left(\frac{C(t)}{L - C(t)}\right) \quad (4.16)$$

onde  $L$ ,  $k$  e  $t_{cri}$  são constantes.  $t_{cri}$  representa o momento efetivo de máximo de infecção semanal, com  $C(t = t_{cri}) = C_T/2$ . Para ajustar os dados à equação, é necessário conhecer a série temporal completa, tornando impraticável utilizá-la na previsão de casos de dengue nas primeiras semanas do ano. Em particular, a constante  $L$  governa a quantidade efetiva de casos totais observados no final do ano ( $L \approx C_T \propto N$ ).

Na Figura 27, é mostrado o comportamento da Equação 4.16 para  $L = C_T$  e  $L = N$  para a cidade de Fortaleza nos anos de 2015, 2016, 2019 e 2020. Para cada ano, o momento  $t_{cri}$  é destacado através da igualdade  $C = C_T/2$  (Equação 4.14). Observa-se que  $h(L = C_T)$  e  $h(L = N)$  diferem em escala para o mesmo número de casos observados. No entanto, para os períodos anteriores ao pico de contaminação ( $t < t_{cri}$ ), as funções apresentam taxas de crescimento semelhantes ( $\frac{dh}{dC}$ ). Para avaliar a hipótese de que as distribuições dos valores de  $\frac{dh(L=C_T)}{dC}$  e

$\frac{dh(L=N)}{dC}$  são iguais, foi realizado o teste de Kolmogorov-Smirnov. O resultado do teste indica a distância máxima entre as duas CDFs. Considerando um valor de significância de 0.05, as análises mostraram que a distância é de 0.43 ( $p = 0.14$ ), indicando que as distribuições são estatisticamente iguais. Assim, decidimos definir a Função de Controle de Intensidade (*ICF*, do inglês Intensity Control Function) da propagação da dengue como:

$$\begin{aligned} f(N, C(t)) &= \frac{dh(L=N)}{dC} \\ f(N, C(t)) &= \frac{N}{C(t)[N-C(t)]} \end{aligned} \quad (4.17)$$

A Equação 4.17 pode ser calculada para todos os bairros ao longo do período epidêmico. Valores altos de  $f(N, C(t))$  são encontrados nos primeiros momentos de um surto de dengue, enquanto valores baixos de  $f(N, C(t))$  são observados durante processos de grande infecção. Além disso,  $f(N_1, C(t)) > f(N_2, C(t))$  quando  $N_1 < N_2$ . Dessa forma,  $f(N, C(t))$  mitiga o processo de propagação da dengue, o que significa que valores menores em  $f(N, C(t))$  regulam velocidades de propagação efetivas mais altas da doença em áreas urbanas. A hipótese é que a informação produzida por *ICF* contribui para uma melhor previsão da série temporal, uma vez que também captura o tamanho da população na taxa de crescimento dos casos.

#### 4.2.2 Modelo preditivo

Foram desenvolvidas duas arquiteturas de redes neurais para determinar se o *ICF* afeta a previsão de casos de dengue. Ambas as arquiteturas utilizam uma pilha de *Long short-term memory* (LSTM) e têm como objetivo prever casos semanais de dengue em nível de bairro, especificamente para os 119 bairros de Fortaleza. A primeira arquitetura permite a entrada de dados na forma de uma matriz com dimensões  $52 \times 119$ , fornecendo à rede neural uma série temporal de tamanho 52 como entrada. Isso representa um histórico de 52 semanas para cada um dos 119 bairros em Fortaleza, denotados como  $(X_{b1_{t51}}, X_{b1_{t50}}, \dots, X_{b1_t}, X_{b2_{t51}}, X_{b2_{t50}}, \dots, X_{b2_t}, \dots, X_{bn_{t51}}, X_{bn_{t50}}, \dots, X_{bn_t})$ . A série temporal de entrada pode ser o histórico de casos de dengue, dados de incidência ou dados gerados pelo *ICF*. A saída da arquitetura é uma previsão de uma semana à frente para todos os bairros simultaneamente, representada como  $(Y_{b1_{t+1}}, Y_{b2_{t+1}}, \dots, Y_{bn_{t+1}})$  e as predições são realizadas de forma recursiva, a predição para o tempo  $t + 1$  é utilizada também para prever valores no tempo  $t + 2$ . Foi selecionado um histórico de 52 semanas como entrada para a rede neural devido à natureza sazonal da dengue em Fortaleza.

A segunda arquitetura é semelhante à primeira, mas inclui uma série temporal transformada usando o *ICF*, além da série temporal de casos de dengue nos bairros. Assim, a entrada da segunda arquitetura é uma matriz com dimensões  $52 \times 238$ . A Figura 28 ilustra as duas arquiteturas de redes neurais, incluindo suas entradas e saídas.

Aplicamos quatro testes que utilizam dados de entrada e arquiteturas de redes neurais diferentes para prever os casos semanais de dengue em nível de bairro. A lista a seguir descreve as variações que serão comparadas:

- **RE** - teste realizado com a arquitetura representada na Figura 28a utilizando a série temporal de casos de dengue nos 119 distritos como entrada para a rede neural.
- **ICF** - teste realizado com a arquitetura representada na Figura 28a, tendo como entrada para a rede neural a série temporal de casos de dengue nos 119 bairros modificada pela Equação 4.17.
- **ICF + RE** - teste realizado com a arquitetura representada na Figura 28b utilizando duas matrizes como entrada para a rede neural, uma contendo a série temporal de casos de dengue nos 119 bairros e outra contendo a série temporal modificada pela Equação 4.17.
- **INC** - teste realizado com a arquitetura representada na Figura 28a utilizando a série temporal de incidências de casos de dengue nos 119 distritos como entrada para a rede neural. Aqui, incidência ( $w^{(i)}$ ) é definida pela relação:

$$w^{(i)}(t) = \frac{W^{(i)}(t)}{N_i} \quad (4.18)$$

onde  $W^{(i)}$  é a série temporal semanal de casos semanais e  $N_i$  é a população do bairro  $i$ .

#### Treinamento e testes

Para validar as variações dos testes e prever os casos de dengue para os 119 bairros, os modelos de rede neural foram treinados utilizando dados de 2007 a 2014. Os parâmetros da rede neural, incluindo épocas, método de otimização, número de neurônios, *dropout* e tamanho da pilha de LSTM, foram otimizados usando o método *Grid Search* fornecido pela biblioteca (PEDREGOSA *et al.*, 2011). Esse método pesquisa por todos os valores pré-definidos para cada parâmetro a fim de identificar a combinação ideal. Para o processo de treinamento e validação os dados foram agrupados utilizando a técnica de *sliding window*.

Durante o processo de treinamento da rede neural, dados de dois anos foram separados para validação: um ano epidêmico e um ano não epidêmico. Para testar o desempenho dos modelos, foram utilizados dois anos epidêmicos (2015 e 2016) e dois anos não epidêmicos

(2019 e 2020). A divisão dos dados para treinamento, validação e teste da rede neural para a previsão de casos de dengue nos quatro anos é apresentada na Tabela 3.

Treinamento	Validação	Teste
2007, 2008, 2011, 2012, 2013, 2014	2010, 2016	2015
2007, 2008, 2011, 2012, 2013, 2014	2010, 2015	2016
2007, 2008, 2011, 2012, 2013, 2014	2010, 2016	2019
2007, 2008, 2011, 2012, 2013, 2014	2010, 2016	2020

**Tabela 3 – separação de dados para treinamento, validação e testes das arquiteturas de rede neural propostas.**

Sabendo que o pico de casos de dengue em Fortaleza ocorre em abril ou maio e que é de interesse público saber antecipadamente se ocorrerá ou não uma epidemia de casos de dengue, as previsões feitas nos testes foram realizadas a partir da semana 9, considerando que, para todos os 4 anos de teste, a semana 9 é antes do pico de casos de dengue.

### Métodos de avaliação

As previsões para os 119 bairros foram avaliadas com base em quatro objetivos: prever a intensidade máxima de casos, prever o tempo de pico, prever a série temporal completa de casos de dengue e classificar os anos e bairros como epidêmicos ou não epidêmicos.

Para avaliar as previsões de intensidade máxima, calculamos o erro médio absoluto logarítmico ( $MALE_p$ ), entre os casos de dengue reais e previstos no pico:

$$MALE_p = \frac{\sum_{k=0}^n |\log(y_k + 1) - \log(\hat{y}_k + 1)|}{n}, \quad (4.19)$$

onde  $y$  é o número real de casos de dengue durante a semana de pico e  $\hat{y}$  é o número previsto de casos de dengue durante a semana de pico.

O erro absoluto médio ( $MAE_t$ ) representa a diferença entre os casos reais e previstos de dengue no pico, em termos da semana em que o pico ocorre:

$$MAE_t = \frac{\sum_{k=0}^n |r_k - \hat{r}_k|}{n}, \quad (4.20)$$

onde  $r$  é o índice da semana de pico da série temporal real,  $\hat{r}$  é o índice da semana de pico da série temporal prevista, e  $n$  é o número total de bairros previstos.

O erro logarítmico médio-quadrático ( $RMSLE$ ) mede a diferença entre as séries temporais reais e previstas como um todo, sendo calculada como:

$$RMSLE = \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{\sum_{w=p}^{52} (\log(x_w^{(i)} + 1) - \log(\hat{x}_w^{(i)} + 1))^2}{52 - p}}, \quad (4.21)$$

onde  $x$  representa o número total de casos registrados de dengue,  $\hat{x}$  representa o número total de casos de dengue previstos,  $n$  representa o número total de bairros, e  $p$  representa o tempo em que a previsão começa. Neste caso,  $p = 9$ , o que corresponde à semana do ano em que a previsão foi iniciada. Como a série histórica de casos de dengue em Fortaleza normalmente exibe apenas alguns picos por ano, o objetivo principal de um modelo de previsão é identificar o pico e seus altos e baixos. *RMSLE* é a métrica mais apropriada para capturar isso, pois penaliza grandes diferenças entre os valores previstos e reais em períodos críticos, quando a incidência é alta.

Consideramos que bairros epidêmicos são bairros que atingem ao final do anos (semana 52) o total de casos superior a  $C_{cri}$ . Para medir a performance de classificação de bairros epidêmicos ou não-epidêmicos, usamos a métrica *f1 – score* (F1) derivado por:

$$precisao = \frac{TP}{TP + FP}, \quad (4.22)$$

$$revocacao = \frac{TP}{TP + FN}. \quad (4.23)$$

$$F1 = 2 * \frac{precisao * revocacao}{precisao + revocacao}. \quad (4.24)$$

onde  $TP$  é verdadeiro positivo,  $TN$  é verdadeiro negativo,  $FP$  é falso positivo e  $FN$  é falso negativo.

### 4.2.3 Predições e comparações

Dado o interesse do governo em identificar picos nos casos de dengue com antecedência e o fato de que esses picos geralmente ocorrem em torno das semanas  $16 \leq \tau \leq 20$  do ano, é crucial que um modelo de rede neural que preveja casos de dengue para um ano completo possa prever com precisão os casos com algum aviso prévio para o período da semana  $\tau$ . Portanto, em nossos testes, as previsões começam a partir da semana 9.

As Figuras 29(a)-(d) mostram os resultados das nossas previsões para os anos de 2015, 2016, 2019 e 2020, que foram obtidos utilizando os testes *RE*, *ICF + RE*, *INC* e *ICF*. As

curvas ilustram o perfil temporal para a cidade de Fortaleza, onde a linha preta representa os registros reais de casos de dengue, e as outras linhas representam as previsões obtidas a partir dos diferentes testes.

Os primeiros surtos de dengue são observados nas primeiras semanas do ano, e os casos aumentam até períodos próximos a  $\tau$  com o perfil de intensidade apropriado. Após o pico, os casos semanais diminuem com intensidade semelhante. No entanto, esse comportamento não é claramente capturado pela maioria dos testes, especialmente o teste *INC*. Os resultados de previsão do teste *INC* indicaram que os dados modulados pela incidência (conforme mostrado na Equação 4.18) não são relevantes o suficiente para que o modelo de rede neural aprenda a evolução dos casos de dengue. Uma situação semelhante é observada quando apenas os dados de *ICF* são utilizados. No entanto, vale ressaltar que a variação de *ICF* antecede o momento do pico de casos de dengue.

Quando analisamos apenas o modelo *RE*, observamos que para o ano de 2019 (um ano não epidêmico), o modelo foi capaz de acompanhar a curva de evolução dos casos reais de dengue. No entanto, o mesmo não ocorreu com os outros anos (2015, 2016 e 2020). Por outro lado, quando as informações de controle de intensidade são adicionadas, formando o teste *ICF + RE*, o modelo foi capaz de prever melhor a evolução dos casos de dengue.

O painel da Figura 30 mostra os resultados para as 3 métricas detalhadas na subseção "Métodos de avaliação",  $MALE_p$ ,  $MAE_t$  e  $RMSLE$ . O painel mostra que para as 3 métricas de avaliação a variação de O resultados apontam que o teste *ICF + RE* apresentou melhores índices de performance para quase todas as métricas avaliadas, sendo capaz de melhor prever a série temporal como um todo, o momento do pico e a intensidade do pico de casos de dengue. O teste *RE* apresentou piores resultados de  $MALE_p$ ,  $MAE_t$  e  $RMSLE$  para 2015 quando comparado com todos os outros testes, porém para 2016, os resultados foram similares ao teste *INC*. Para 2019 e 2020, o *RE* apresentou melhores resultados quando comparado com o *INC* e *ICF*. Apesar da variação *INC* ter apresentado bons resultados para prever a intensidade do pico e tempo do pico para os anos de 2015, 2016 e 2020, em 2019 esse teste mostrou resultados péssimos para todas as variações, demonstrando uma grande instabilidade na predição. Por fim a variação *RE* mostrou ter resultados bem estáveis, mostrando ser uma variação de teste com performance com resultados bons para todos os anos quando comparada aos testes *ICF* e *INC*.

Os resultados apresentados para os testes nas Figuras 29 e Figuras 30(a)-(c) são conflitantes, apesar dos resultados para  $MALE_p$ ,  $MAE_t$  e  $RMSLE$  apresentarem valores competitivos. O que aconteceu de fato com as predições do teste *INC* foram resultados constantes,



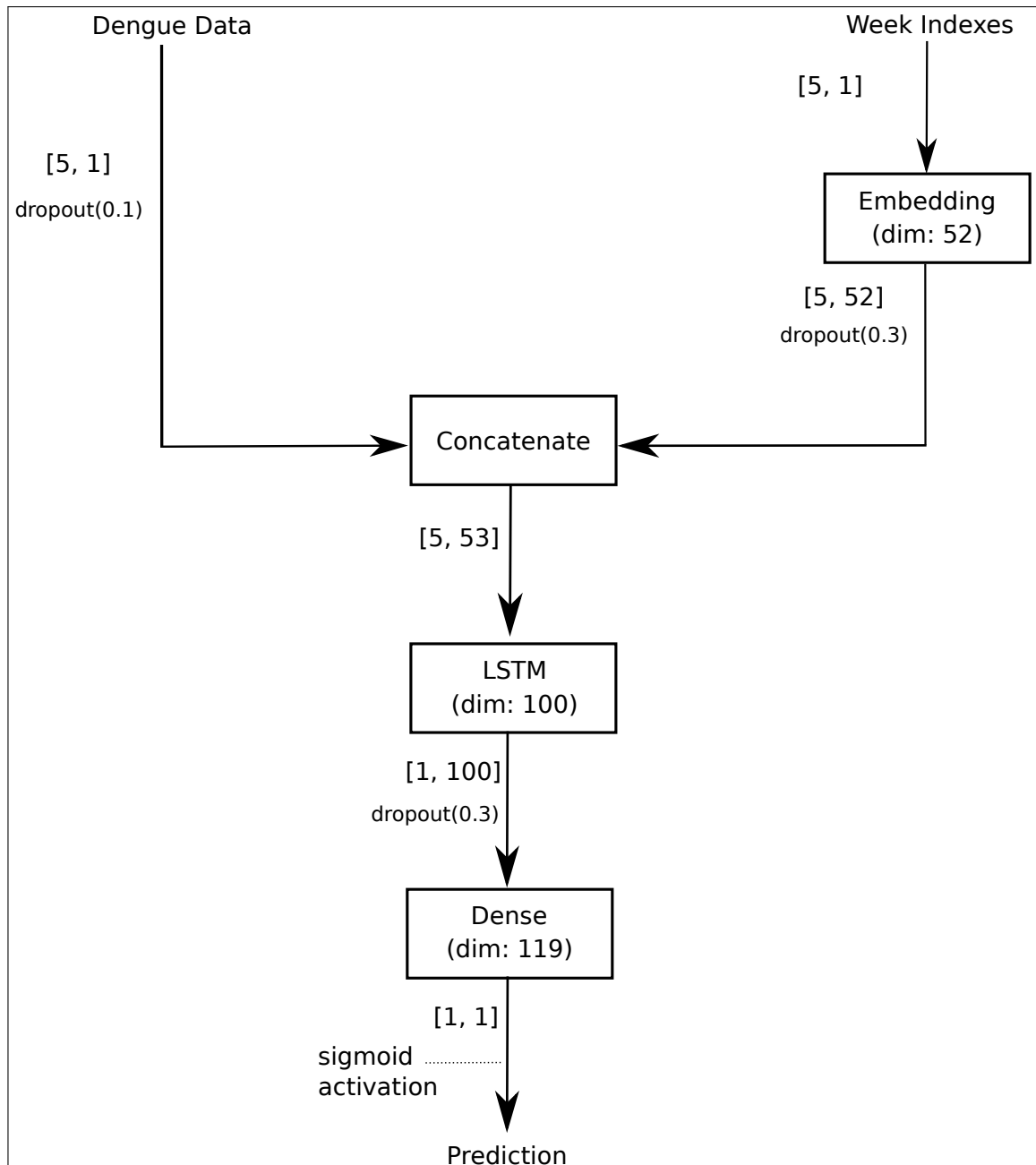
sendo assim as métricas de  $MALE_p$ ,  $MAE_t$  e  $RMSLE$  nos levam a análises equivocadas para esse caso específico. Tendo conhecimento deste fato, a Figura 29(d) é apresentada a média da correlação de Pearson entre os séries temporais preditas e reais a nível de bairro para os 4 anos em análise. Esta correlação nos indica que o teste com maior capacidade de prever a evolução temporal de casos de dengue é a  $ICF + RE$ , seguida do  $RE$ . Os testes  $INC$  e  $ICF$  tiveram os piores resultados.

Os testes  $RE$  e  $RE + ICF$  geraram melhores resultados nas predições. Contudo, é importante identificar quais testes garantem melhores performances nas classificações de bairros epidêmicos. A Figura 31(a) mostra o valor de  $F1$  (Equação 4.24) para a classificação de bairros epidêmicos ou não-epidêmicos para os testes  $ICF + RE$  e  $RE$ . Observa-se que o teste  $ICF + RE$  possui melhor performance para todos os anos analisados.

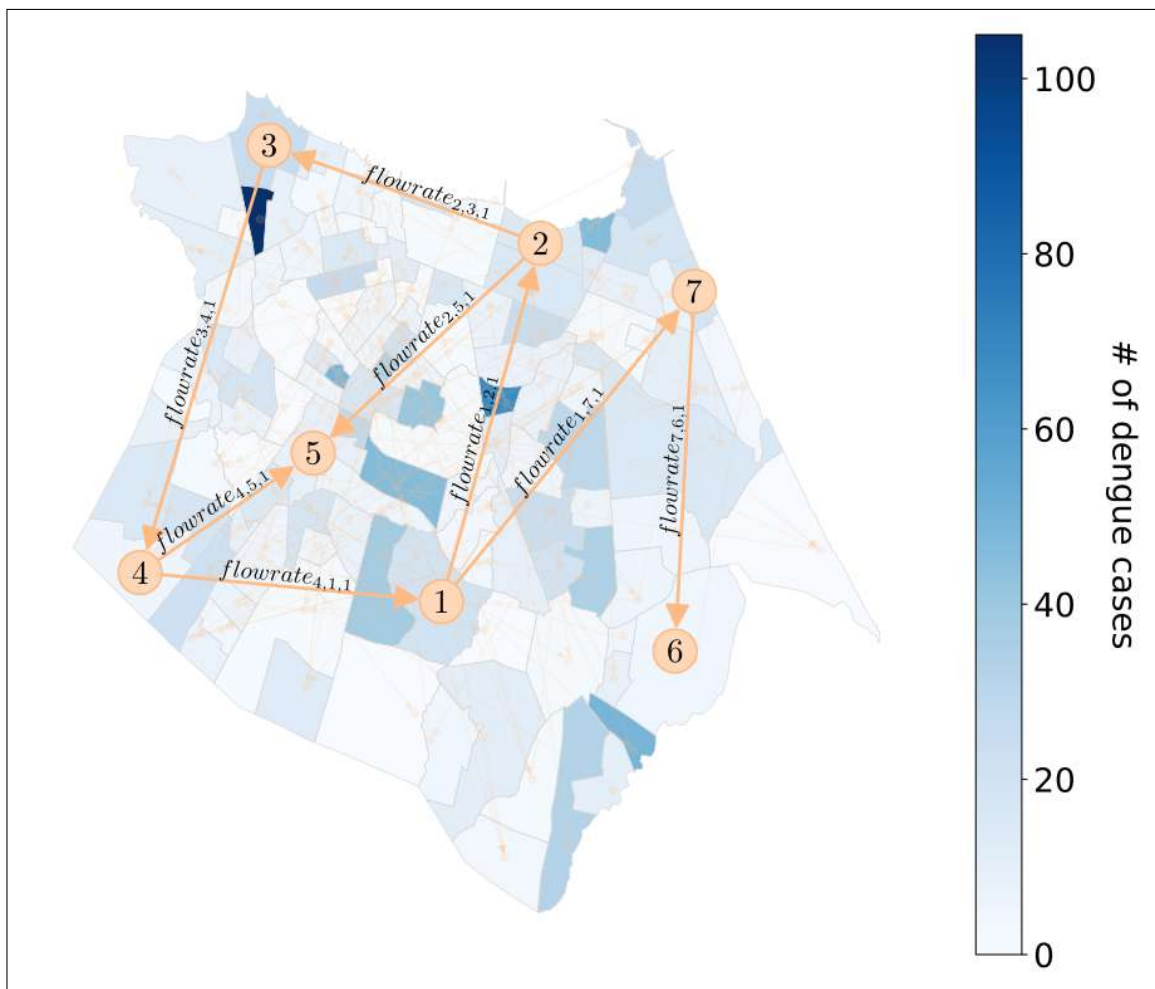
Para melhor analisar a capacidade de generalização dos testes  $ICF + RE$  e  $RE$ , nós modificamos a semana do começo das predições nos intervalos 9, 10, 11 e 12. Além disso, variações em  $C_{cri}$  foram atribuídas no intervalo  $[10, 200]$ . As Figuras 31(b)-(d) mostram, para os anos de 2015, 2016, 2019 e 2020, respectivamente, a relação entre os valores de  $F1$  medidos pelos testes  $RE$  e  $ICF + RE$ . Para ponto nos gráficos representam um experimento controlando a semana de predição e o valor de  $C_{cri}$ . É notório que para predições do ano 2015 o teste  $ICF + RE$  previu melhor os testes começando nas semanas 9, 11 e 12. O mesmo ocorre quase de forma hegemônica para todas as semanas de começo de predição para ano de 2016. Para os anos não-epidêmicos, 2019 e 2020, os resultados de  $F1$  entre os dois testes são semelhantes.

Concluo que a utilização de uma função controladora de intensidade ( $ICF$ ) traz benefícios para a predição de casos de dengue onde a rede neural precisa ser capaz de aprender a evolução da doenças em varias localidades ao mesmo tempo e aprender o impacto que cada uma dessas localidades tem na predição do todo. Utilizar apenas os casos de dengue como forma de aprender a evolução dos casos é uma estratégia competitiva, porém devido a alta correlação de casos de dengue entre os diferentes bairros, um modelo que apenas considera essa informação, tende a prever picos de casos em um mesmo momento do ano. Assim, a  $ICF$  incorpora ao modelo de rede neural uma informação limitadora de casos, a população residente de cada bairro, permitindo que a rede neural possa ponderar a importância de cada bairro na transmissão de dengue em outros bairros, evitando predições realizadas para valores médios.

**Figura 17 – Arquitetura isolada de rede neural - o modelo  $RNN_i$ . A arquitetura isolada prevê, com uma semana de antecedência, os casos de dengue para um bairro específico com base nas últimas 5 semanas de casos de dengue e em um vetor de índice que rotula cada semana de entrada. A camada de *embedding* aprende uma representação para o rótulo de cada semana, que é concatenado com os dados de dengue. Assim, a evolução dos casos de dengue é aprendida por meio dos dados concatenados e, em seguida, a camada densa gera a previsão para a semana seguinte. Os valores entre parênteses no *dropout* representam a fração de neurônios a serem ignorados em cada etapa de treinamento. Os valores entre colchetes representam as dimensões de entrada e saída de dados para cada camada.**

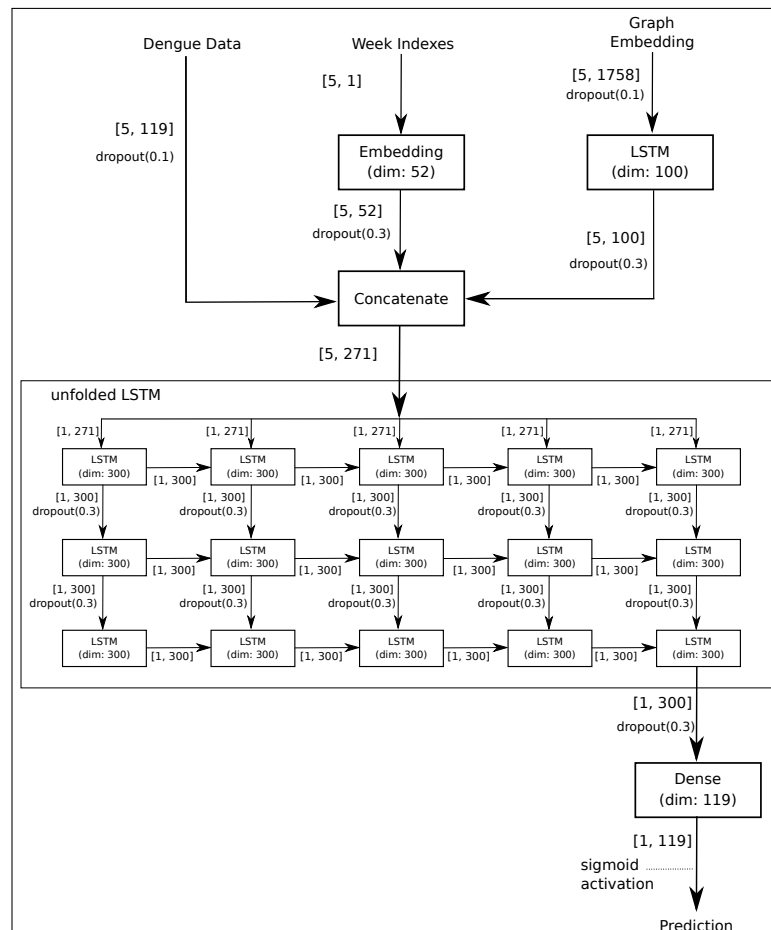


**Figura 18 – O fluxo semanal de pessoas entre bairros. As setas amarelas desenhadas no mapa destacam um exemplo de fluxo entre 7 bairros.  $flowrate_{4,5,1}$  representa o fluxo do bairro 4 para o bairro 5 na semana 1. O modelo completo inclui taxas de fluxo para e de cada um dos 119 bairros da cidade para cada semana do período.**



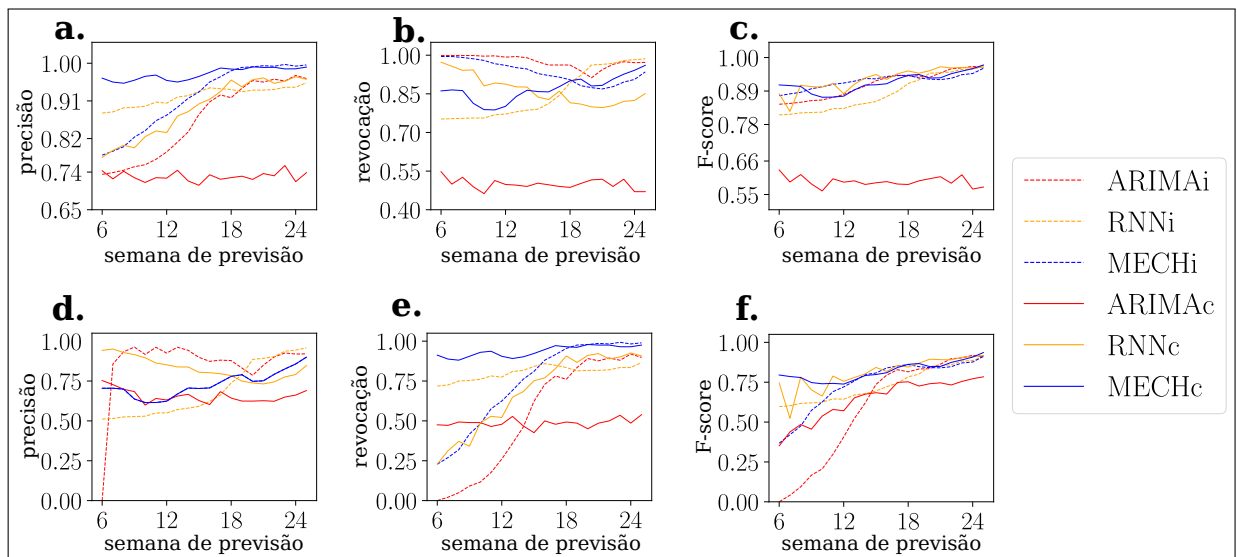
Fonte: Elaborado pelo autor

**Figura 19 – Arquitetura de rede neural artificial para o modelo RNNc com dados de mobilidade humana. A arquitetura assume a entrada de três conjuntos de dados: uma série temporal de cinco semanas de 119 bairros, um vetor com um identificador para cada semana e uma representação vetorial (embedding) descrevendo as informações de transporte para cada bairro. Os valores entre parênteses representam a dimensão de cada conjunto de dados. A dimensão das entradas são: [5, 119], representando a dimensão da matriz contendo cinco semanas de casos de dengue para todos os 119 bairros; [5, 1], representando um vetor com os índices para cada semana; e [5, 1785], representando 5 semanas e o embedding de tamanho 15 que representa o fluxo entre os 119 bairros ( $15 \times 119 = 1785$ ). A dimensão da camada oculta *concatenate* é [5, 271], representando a concatenação das colunas das três camadas de entrada. Uma pilha de LSTMs (a figura mostra as LSTMs desenroladas) é alimentada com os resultados da concatenação e cada camada da pilha propaga informações (dimensão 300) para a próxima camada. A última camada da pilha é totalmente conectada com uma camada densa (camada de saída) para prever o caso de dengue de cada um dos 119 bairros.**



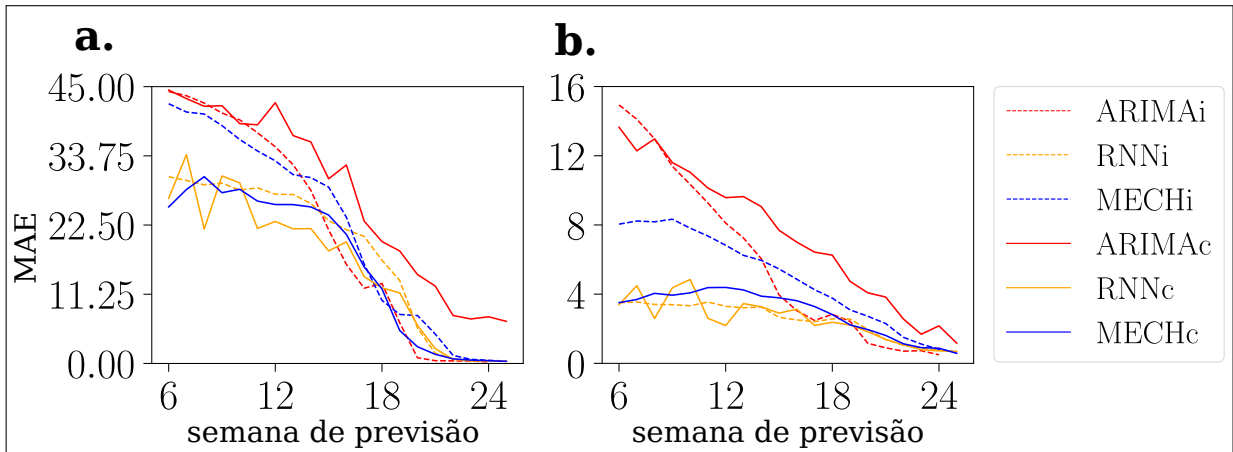
Fonte: Elaborado pelo autor

**Figura 20 –** Para todos os gráficos, o eixo x indica a semana em que a previsão é gerada. O eixo y mostra os valores para precisão, revocação e F-score. Por exemplo, em (a), o valor de y para  $x = 12$  (tempo de previsão = 12) mostra a precisão calculada a partir das previsões feitas a partir da semana 12 até a última semana do ano (semana 52). (a) a precisão da classificação para temporadas com casos totais de dengue menores ou iguais a 200, (b) a revocação da classificação para temporadas com casos totais de dengue menores ou iguais a 200, (c) F-score da classificação para temporadas com casos de dengue maiores que 200 (d) a precisão da classificação para temporadas com casos totais de dengue maiores que 200, (e) a revocação da classificação para temporadas com casos totais de dengue maiores que 200 e (f) F-score da classificação para temporadas com casos totais de dengue menores ou iguais a 200.



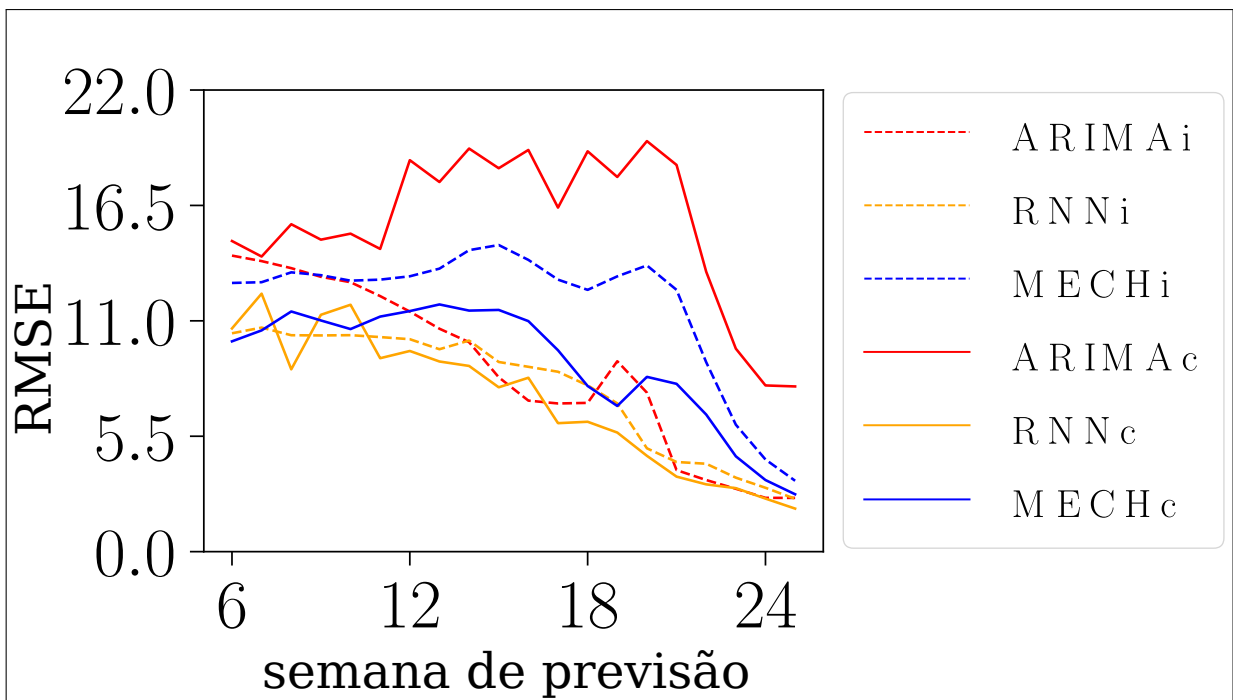
Fonte: Elaborado pelo autor

**Figura 21 – Erro médio absoluto (MAE) para a intensidade de pico (Eq. 4.10) e o momento do pico (Eq. 4.11) para surtos intensos (*dengue*;  $\text{casos} \geq 200$ ). (a) MAE entre a incidência real e prevista de dengue nas semanas com maior incidência de dengue na temporada. (b) MAE entre o índice da semana real e previsto nas semanas com maior incidência de dengue na temporada.**



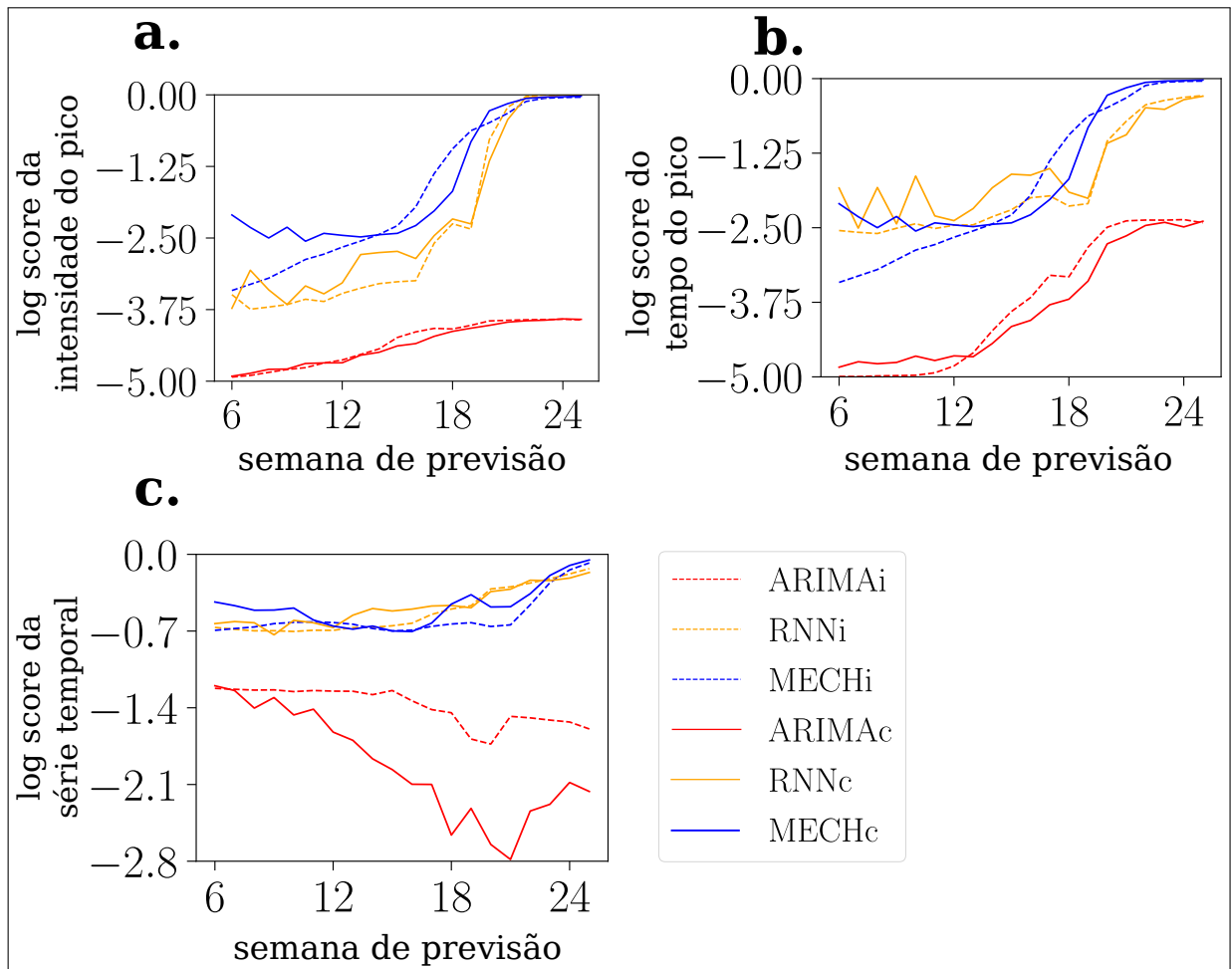
Fonte: Elaborado pelo autor

**Figura 22 – Erro médio quadrático (RMSE) entre as séries temporais previstas e reais da incidência de dengue para surtos intensos (*dengue*;  $\text{casos} \geq 200$ ). ARIMAi, RNNi e MECHi são os resultados para modelos que usam apenas dados locais de dengue, os modelos base, e ARIMAc, RNNc e MECHc são os resultados para as previsões geradas incluindo dados de mobilidade humana.**



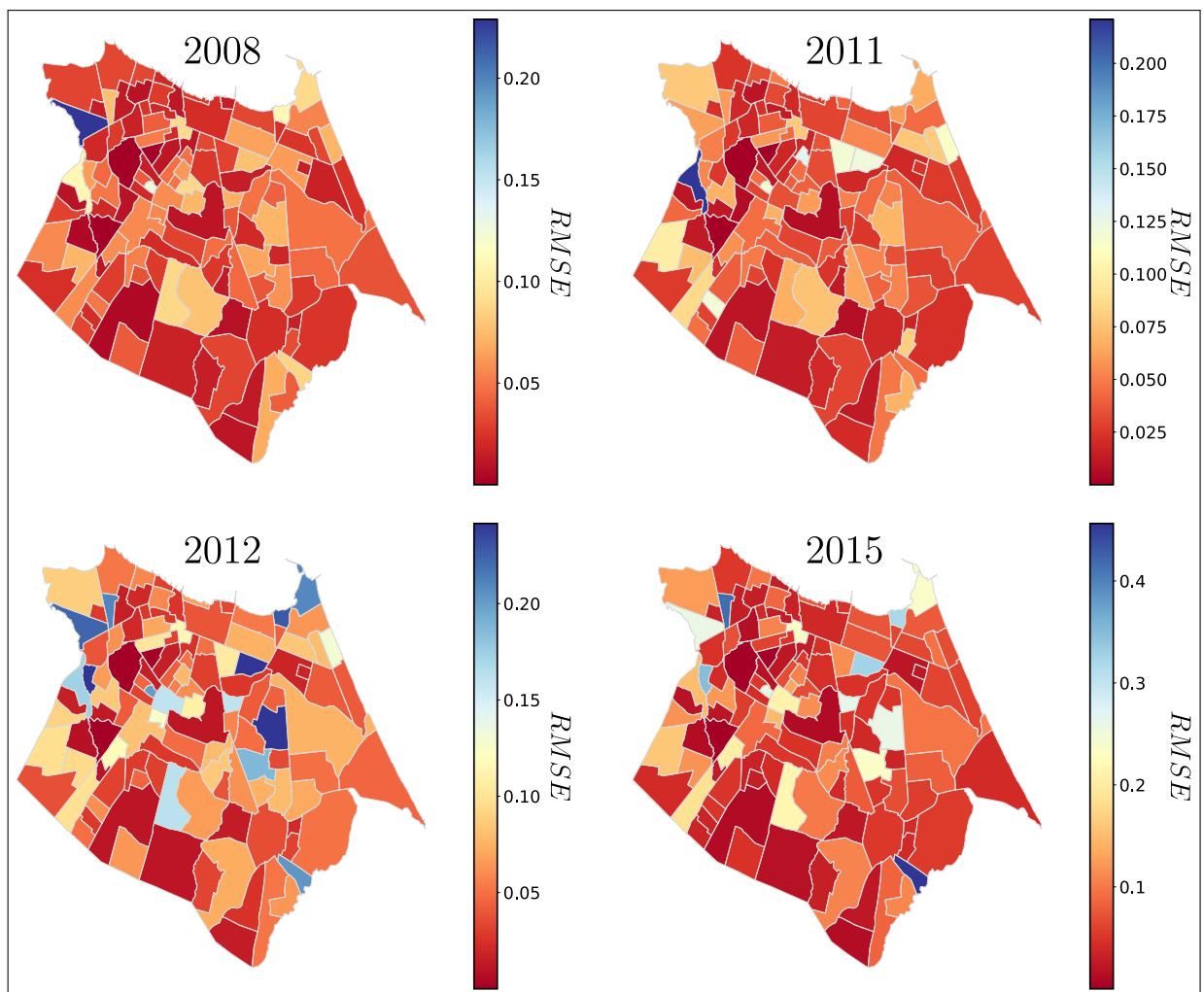
Fonte: Elaborado pelo autor

**Figura 23 – Log score para intensidade de pico (a), tempo do pico (b) e para a série temporal como um todo (c) para grandes surtos de casos de dengue com mais de 200 casos no ano.**



Fonte: Elaborado pelo autor

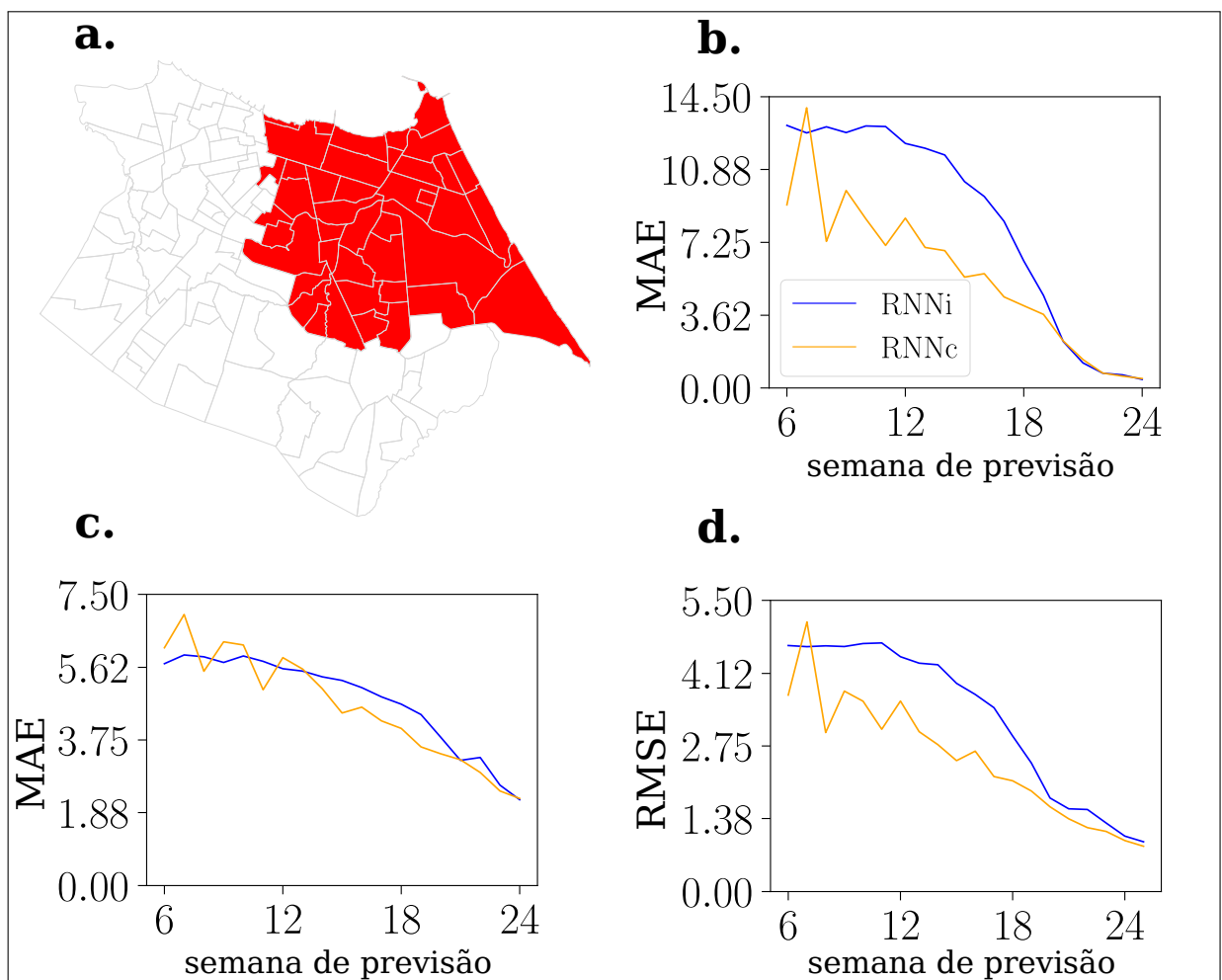
**Figura 24 – Resultados de log-score para intensidade de pico (a), tempo do pico (b) e série temporal de dengue (c) para surtos com um número total de casos de dengue maior que 200.**



Fonte: Elaborado pelo autor

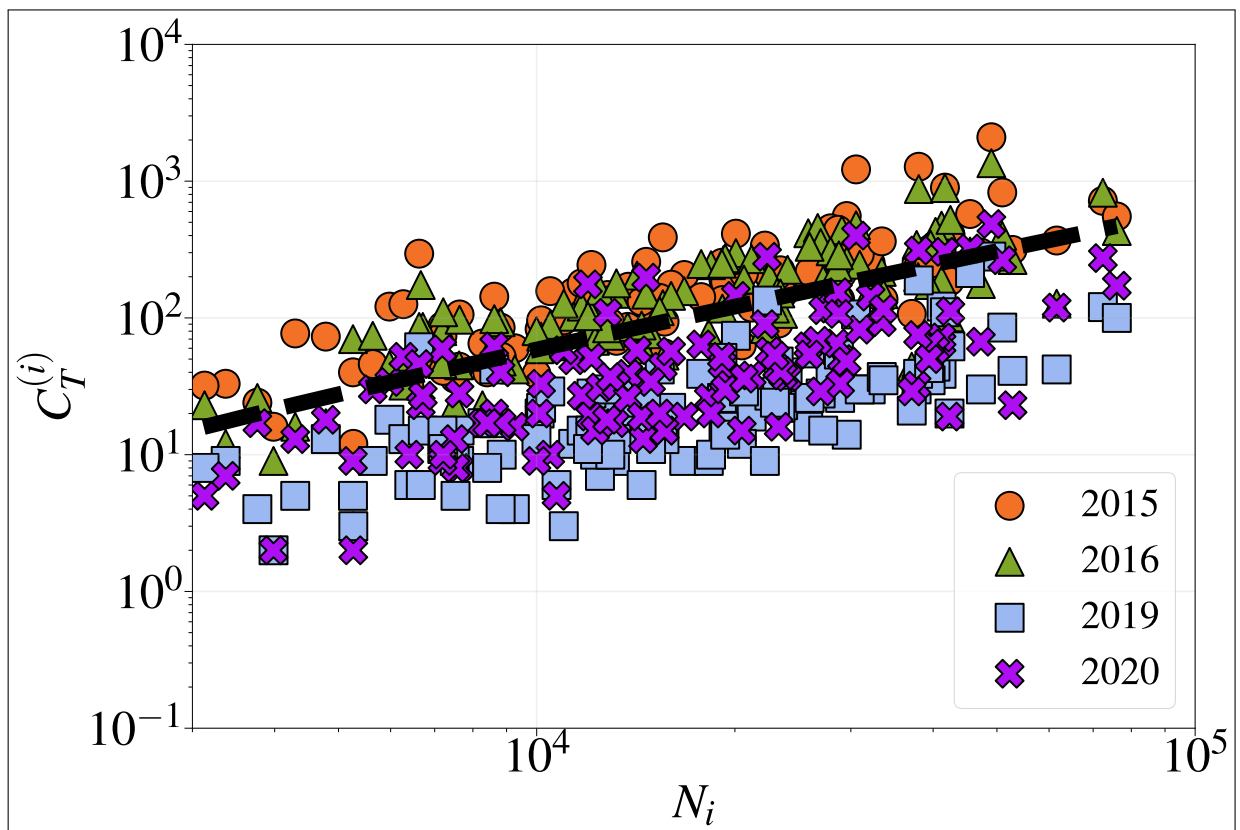


**Figura 25 – Resultados de log-score para intensidade de pico (a), tempo de pico (b) e série temporal de dengue (c) para surtos com um número total de casos de dengue maior que 200.**



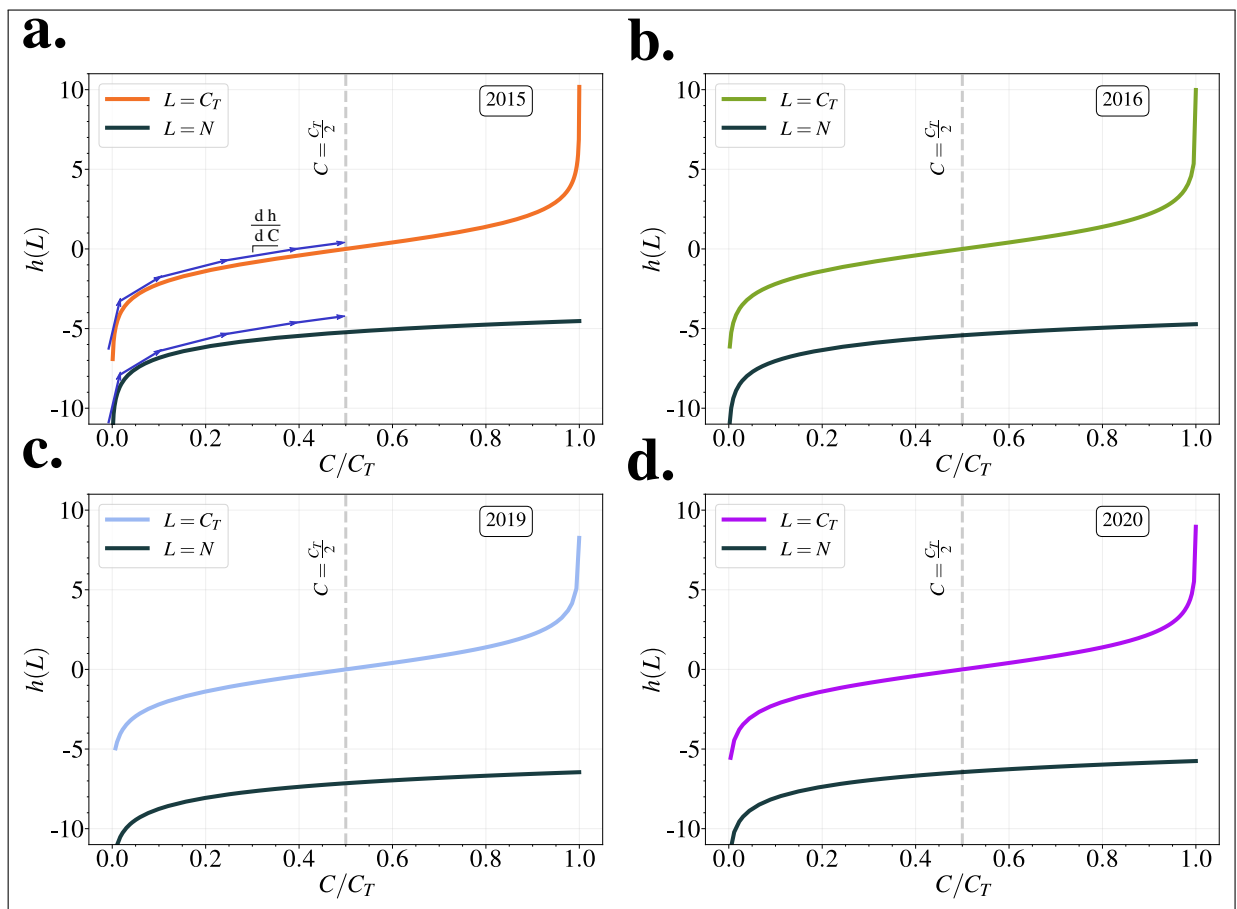
Fonte: Elaborado pelo autor

**Figura 26 – A relação entre a quantidade total de casos observados no final de cada ano  $C_T^{(i)}$  em cada bairro  $i$  como função da população  $N_i$ . A linha preta tracejada é um guia para os olhos.**



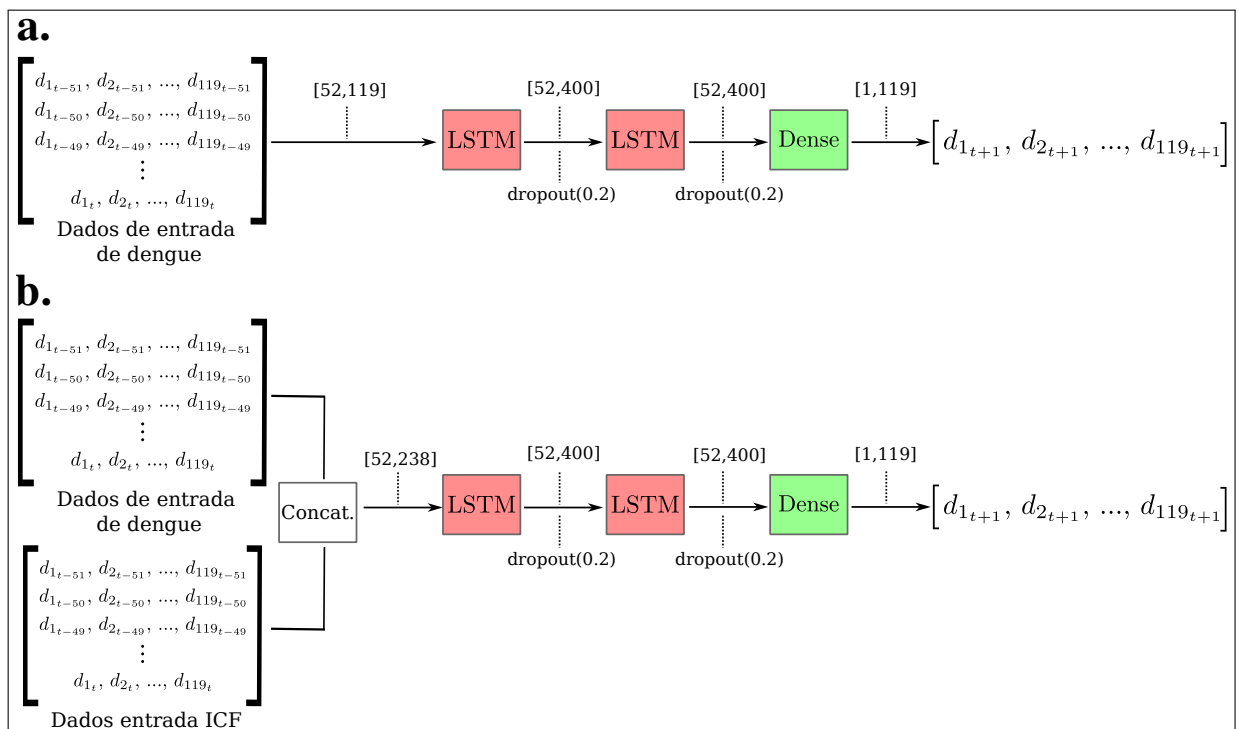
Fonte: Elaborado pelo autor

**Figura 27 – O comportamento da Equação 4.16 em função do número acumulado de casos ( $C(t)$ ) para dois cenários:  $L = C_T$  e  $L = N$ , para os anos de 2015, 2016, 2019 e 2020 (a-d, respectivamente). O eixo x foi normalizado em relação ao número total de casos observados ao final de cada período ( $C_T$ ). Adicionalmente, uma linha vertical tracejada indica o momento de máximo de contaminação semanal ( $C(t = t_{cri}) = C_T/2$ ). Antes de atingir o pico de infecção ( $t < t_{cri}$ ), ambas as curvas  $h(L = C_T)$  e  $h(L = N)$  apresentam taxas de crescimento semelhantes. Observei que  $\frac{dh(L=C_T)}{dC} \approx \frac{dh(L=N)}{dC}$  (representado pelas setas azuis), então essa similaridade implica que a população local pode atuar como regulador para a taxa de crescimento dos casos.**



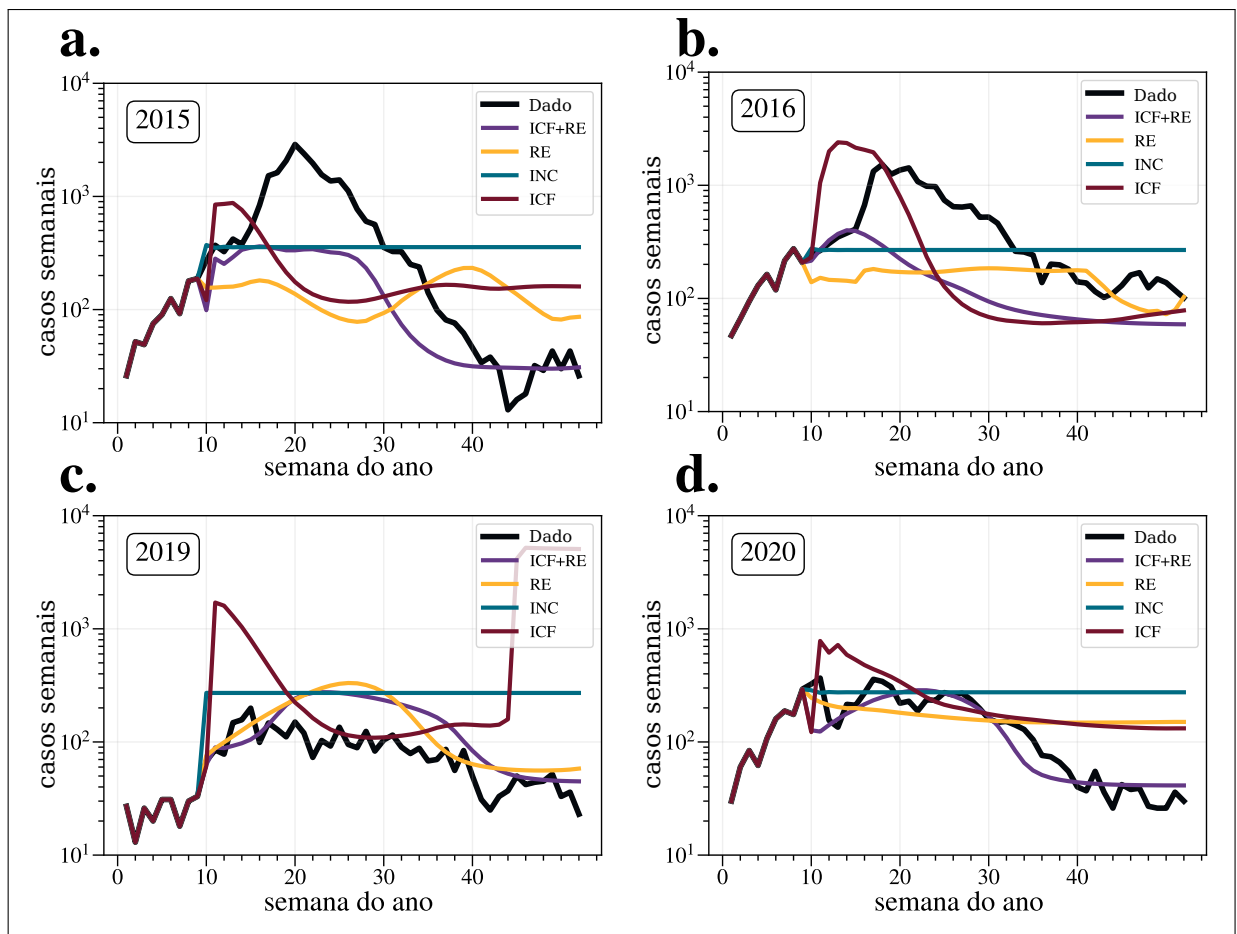
Fonte: Elaborado pelo autor

**Figura 28** – As duas redes neurais usadas neste artigo. Entre colchetes estão as dimensões dos dados passados entre cada camada da rede neural. O número de neurônios (400) da LSTM e o valor de dropout (0.2) apresentados na figura representam os valores ótimos encontrados pelo *Grid Search*. Além desses valores a rede neural foi treinada com um *epoch* de 300 e *batch size* de 64. Em (a) é apresentada a arquitetura da primeira rede neural, que recebe como entrada os casos de dengue ( $W^{(i)}(t)$ ), a incidência de casos de dengue ( $w^{(i)}(t)$ ) ou apenas os casos de dengue transformados por *ICF* (conforme mostrado na Equação 4.17). Em (b) é apresentada a arquitetura da segunda rede neural, que recebe como entrada tanto a série semanal de casos de dengue ( $W^{(i)}(t)$ ) quanto suas modificações através do *ICF*.



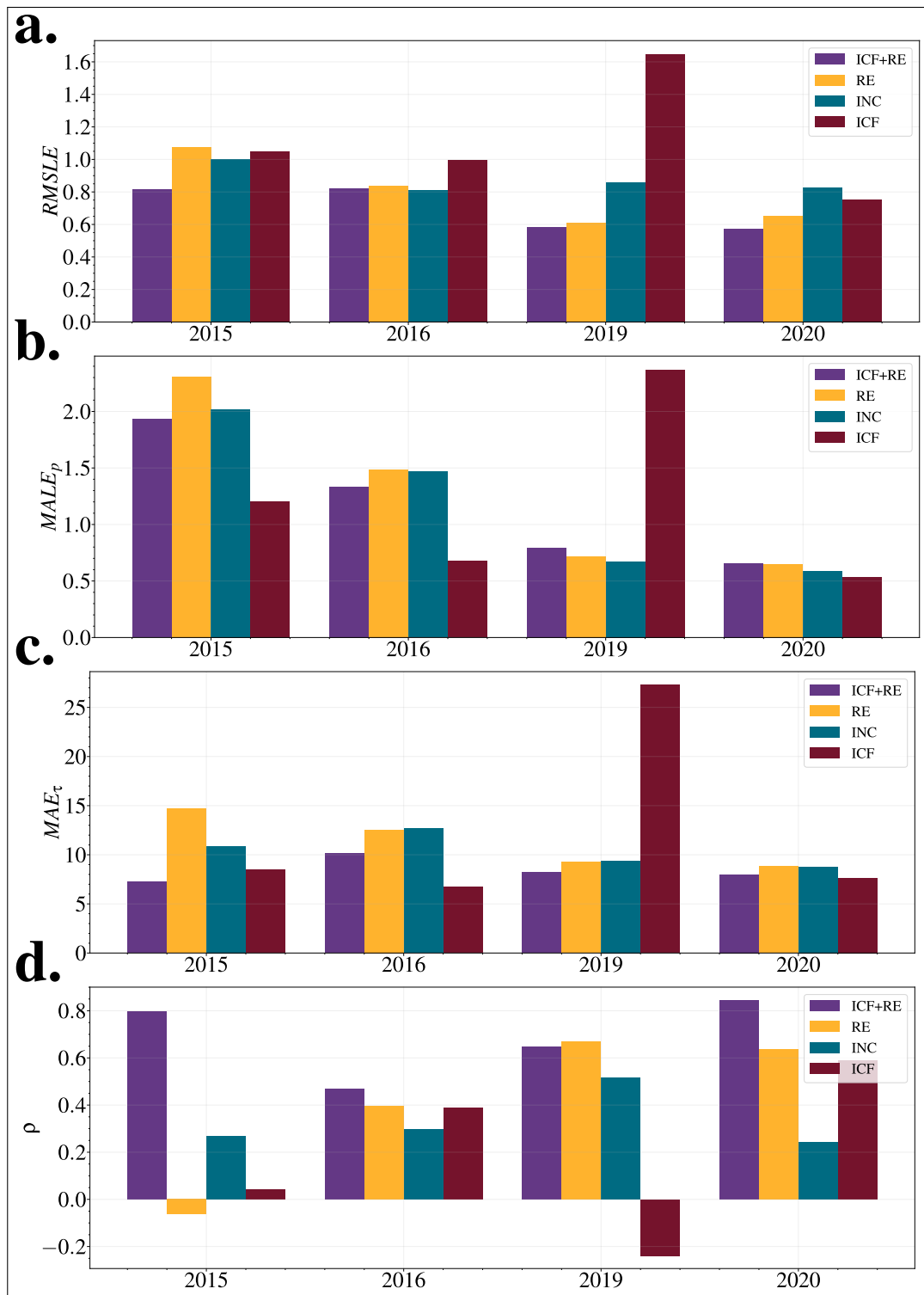
Fonte: Elaborado pelo autor

**Figura 29 – Séries temporais reais (Dado) e previstas para os 119 bairros acumulados semanalmente para os anos de 2015, 2016, 2019 e 2020, considerando as quatro variações de teste (ICF + RE, RE, INC e ICF).**

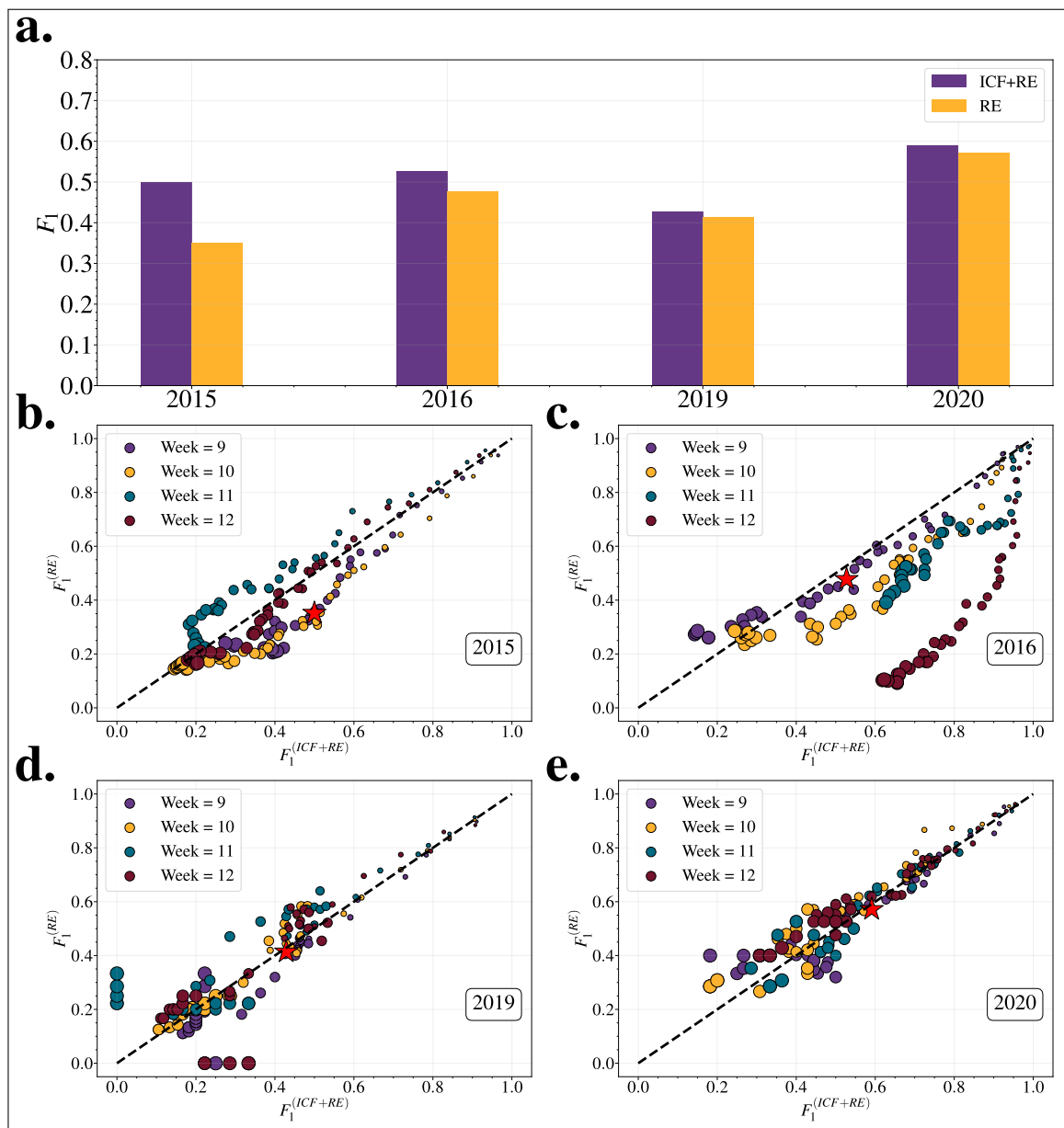


Fonte: Elaborado pelo autor

**Figura 30 – (a) Root Mean Squared Log Error (*RMSLE*) entre dados preditos e reais (Equação 4.21), (b) Mean Absolute Log Error (*MALE*) entre o valor de casos de dengue do pico predito e real (Equação 4.19), (b) Mean Absolute Error (*MAE*) entre o índice real e predito da semana que ocorre o pico de casos de dengue (Equação 4.20) e (d) a correlação de Pearson entre a série temporal predita e real (Equação 3.2).**



**Figura 31** – Em (a) Comparação de  $F_1$  entre testes  $RE$  e  $ICF + RE$  para os anos de 2015, 2016, 2019 e 2020 com a predição iniciando na semana 9. Em (b), (c), (d) e (e) mostram para os anos de 2015, 2016, 2019 e 2020, respectivamente, a comparação entre os valores de  $F_1^{(RE)}$  e  $F_1^{(ICF+RE)}$ , medido pela Equação 4.24 dos  $RE$  e  $ICF + RE$ , respectivamente.  $Week$  representa a semana do começo da predição que foi variada de 9 a 12. O tamanho do ponto representa o valor de  $C_{cri}$  utilizado no experimento com  $C_{cri} \in [10, 200]$ . A estrela vermelha identifica o experimento feito na semana 9 (Semana = 9) com  $C_{cri} = 100$ .



Fonte: Elaborado pelo autor

## 5 O ANO DE 2020: UM LABORATÓRIO PARA COMPREENSÃO DE DOENÇAS CONTAGIOSAS

No capítulo 3, foi demonstrado que o regime de isolamento social aplicado na cidade de Fortaleza, devido à pandemia de COVID-19 no ano de 2020, foi capaz de reduzir a quantidade total de casos de dengue no mesmo ano. Isso evidencia a importância da interação populacional também para a evolução de arboviroses, como a dengue.

Já no capítulo 4, foram desenvolvidos modelos preditivos para demonstrar, na prática, a relevância da interação populacional entre os bairros de uma grande cidade na previsão de casos de dengue. Essa interação foi estudada por meio da mobilidade humana e através de informações heurísticas capazes de ponderar a importância de cada bairro na evolução dos casos na cidade. Concluiu-se que ambas as informações, quando adicionadas a modelos preditivos baseados em redes neurais, têm o potencial de melhorar a previsão da evolução de casos de dengue em nível de bairro.

No entanto, apesar das conclusões obtidas até o momento, é importante determinar como a evolução dos casos teria sido se nenhuma restrição de mobilidade tivesse sido implementada. Esse esforço visa reforçar as conclusões obtidas e avançar ainda mais nas análises epidemiológicas da dengue em um ano atípico como 2020. Para tal propósito, foi desenvolvido um modelo de rede neural baseado em Long Short-Term Memory, seguindo o padrão dos modelos previamente desenvolvidos e testados. O objetivo desse modelo é realizar previsões dos casos de dengue para o ano de 2020 e analisar a evolução desses casos, desconsiderando a mobilidade humana e quaisquer outras informações heurísticas, focando exclusivamente nos dados históricos de casos de dengue por bairro.

### 5.1 MODELO PREDITIVO, TREINAMENTO E DADOS UTILIZADOS

Um modelo neural preditivo, baseado em dados históricos de casos de dengue, foi utilizado para estimar os casos de dengue do ano de 2020 e verificar o número de casos efetivamente prevenidos devido às medidas de restrição de mobilidade humana. Para isso, foi treinada uma arquitetura baseada em *Long Short-Term Memory* (LSTM) (BOMFIM *et al.*, 2020) ilustrada na Figura 32. Foram utilizados dados de casos de dengue dos anos de 2011 a 2020 para o processo de treinamento, validação e teste da rede neural. Os dados foram divididos de acordo com a Tabela 4.

Foram geradas previsões semanais de casos no nível do bairro a partir da décima

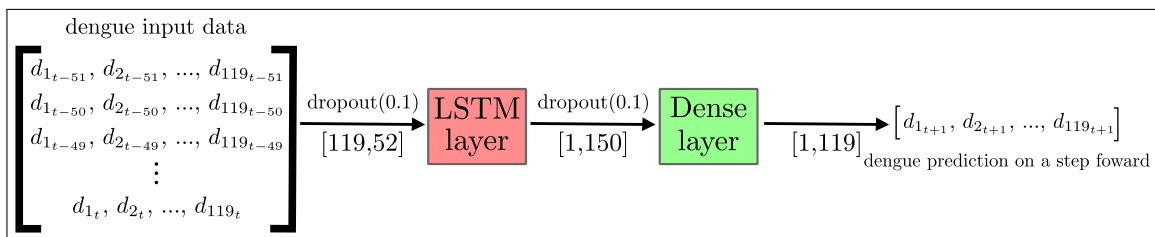


Test	Evaluation	Training
2012	2016, 2018	2011,2013,2014,2015,2017,2019
2013	2016, 2018	2011,2012,2014,2015,2017,2019
2020	2016, 2018	2011,2012,2013,2014,2015,2017,2019

**Tabela 4 – Divisão de Dados: Os dados de treinamento são utilizados para aprender o padrão histórico da evolução dos casos de dengue. Os dados de validação são empregados para ajustar os parâmetros do modelo durante a fase de treinamento, enquanto os dados de teste são utilizados para fins de previsão e análise.**

primeira semana de 2020, tendo em vista que esta foi a semana com maior quantidade de casos de dengue na cidade de Fortaleza antes do início do período de isolamento social. Para validar a generalização do modelo, também foram feitas previsões para outros dois anos, 2012 e 2013, respectivamente, um ano com casos endêmicos e outro ano com poucos casos.

**Figura 32 – Ilustração da arquitetura de rede neural. A arquitetura tem como entrada uma matriz de tamanho [119,52], contendo as séries temporais de casos semanais de dengue de tamanho 52 para cada um dos 119 bairros. A dependência temporal e espacial entre esses valores é aprendida na camada LSTM com 150 neurônios, e a previsão para os 119 bairros com um passo à frente é feita na camada Dense com 119 neurônios. É aplicado um dropout de 10% antes das camadas LSTM e Dense para evitar o overfitting.**



Fonte: Elaborado pelo autor

## A EVOLUÇÃO DE CASOS DE DENGUE SEM RESTRIÇÕES DE MOBILIDADE

A Figura 33 exibe a curva de casos previstos pelo modelo, comparada aos dados reais coletados ao longo do ano de 2020. Os resultados previstos pelo modelo neural indicam um aumento contínuo no número de casos por semana a partir da décima primeira semana, atingindo um valor máximo de 2101 na semana 23. Após essa data, o número de casos diminui e alcança valores mais baixos por volta da metade da semana 45. A partir da semana 26, o comportamento das curvas torna-se semelhante devido à baixa taxa de infecção por dengue durante esse período.

Pode-se observar que o modelo capturou com sucesso a tendência de crescimento dos casos de dengue tanto no ano endêmico de 2012 quanto no ano não endêmico de 2013. No entanto, ele subestimou o número de casos durante esses dois anos, enquanto os casos previstos para 2020 foram superestimados. Predições para outros anos foram realizadas e estão disponíveis no Apêndice E.

O número total de casos ( $Pred^{(i)}$ ), a partir da décima primeira semana, para cada bairro e o tempo de chegada da dengue ( $T_a^{(pred-i)}$ ) foram estimados e podem ser vistos na Figura 34. Existe uma relação exponencial entre a incidência de casos de dengue e o tempo de chegada da doença. Tal relação exponencial é similar àquelas observadas na seção 3.2 em anos epidêmicos com  $\tau_{pred} = 71,24$  dias.

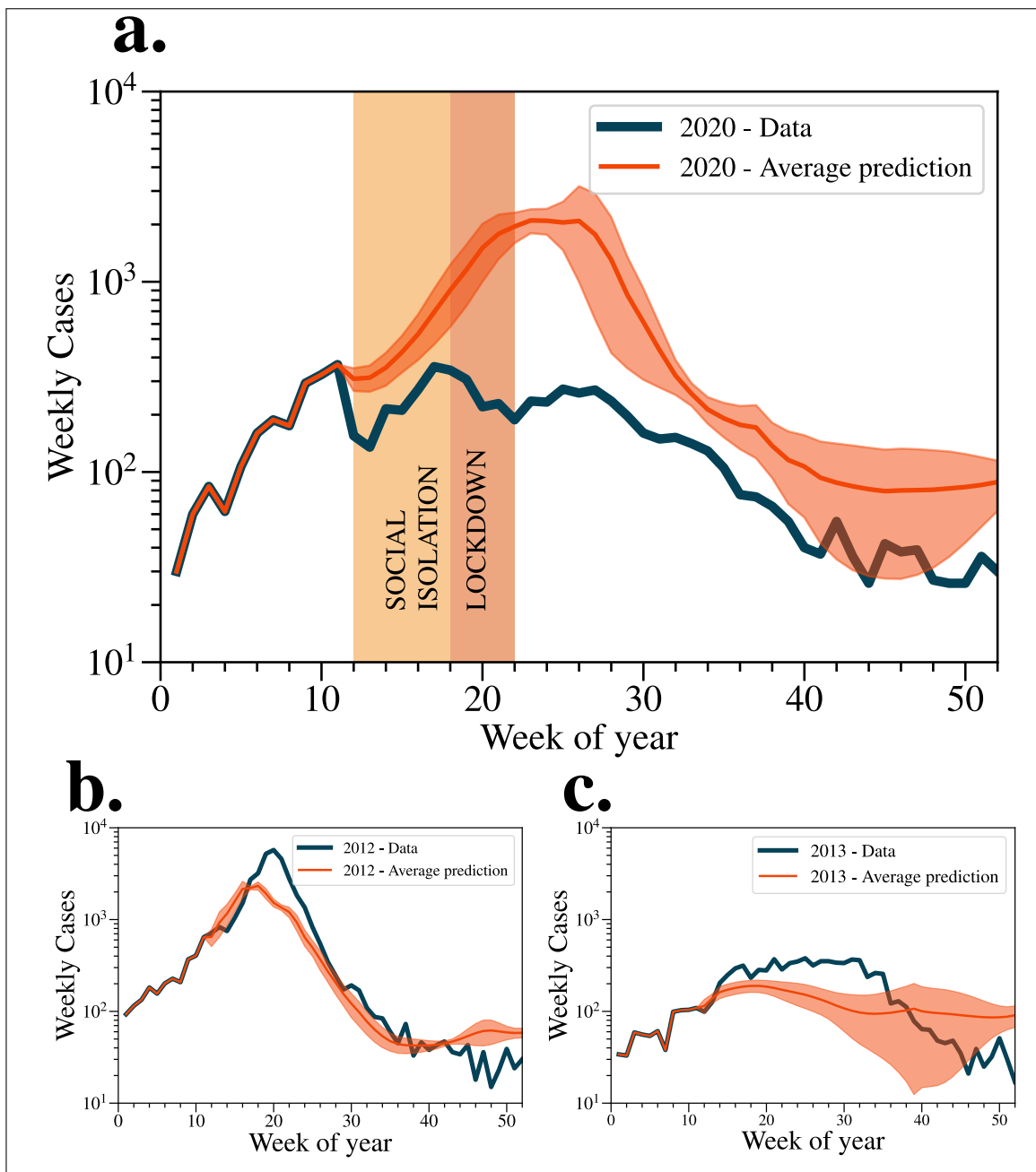
O impacto das restrições de mobilidade em relação ao número de casos é medido por:

$$\Delta^{(i)} = Pred^{(i)} - Data^{(i)}. \quad (5.1)$$

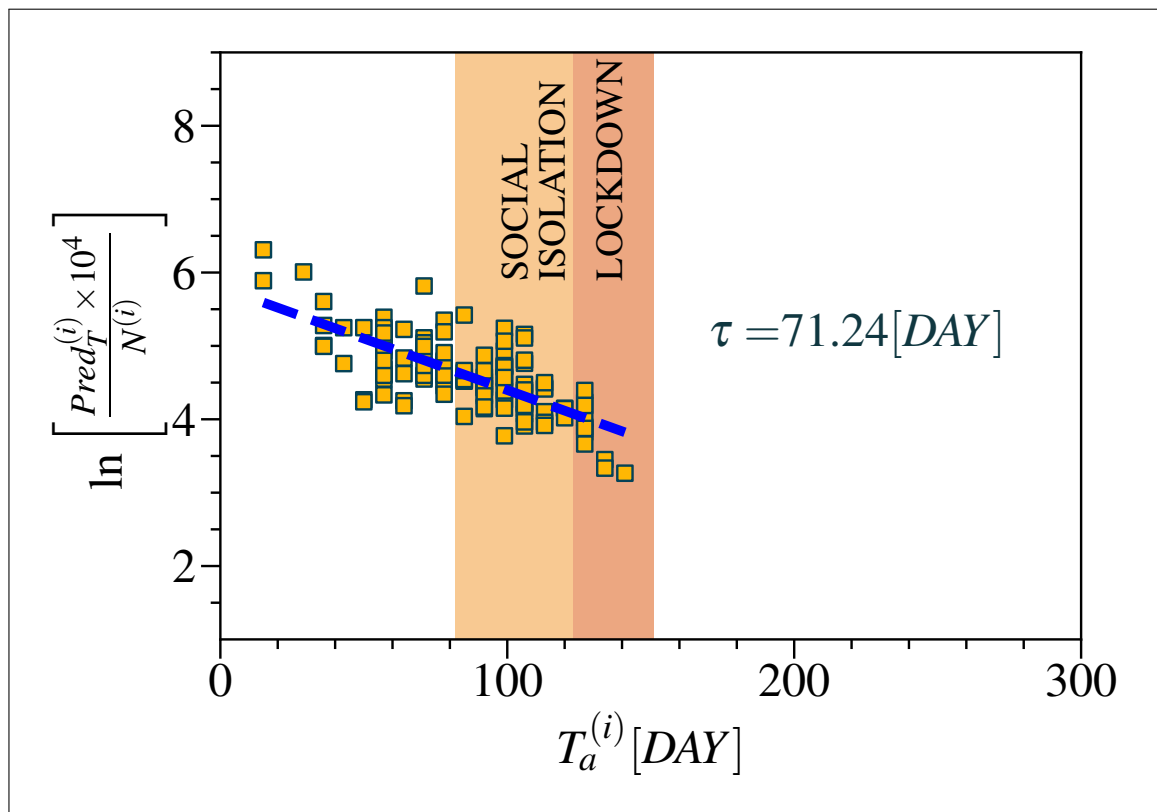
$\Delta^{(i)}$  é a diferença entre os valores previstos ( $Pred^{(i)}$ ) e reais ( $Data^{(i)}$ ) de casos obtidos para cada bairro  $i$ . Assim,  $Pred^{(city)} = \sum_i Pred^{(i)}$  e  $Data^{(city)} = \sum_i C_T^{(i)}$  se referem, respectivamente, ao número de casos previstos e reais para toda a cidade.  $\Delta^{(city)} = Pred^{(city)} - Data^{(city)}$  quantifica o impacto efetivo das restrições de mobilidade em relação ao número de casos em toda a cidade. Note que  $\Delta^{(city)} \approx 23.730$  para os casos de dengue observados ao longo de 2020 enquanto  $Data^{(city)} = 7.720$  casos, indicando que as medidas restritivas de mobilidade urbana reduziram em aproximadamente 300% os casos de dengue na cidade de Fortaleza.

A Figura 35 mostra a relação entre  $\Delta^{(i)}$  e o tempo de chegada da doença  $T_a^{(i)}$  em cada bairro. A forte correlação indica que os bairros inicialmente infectados foram mais fortemente impactados pela redução dos casos de dengue. O número de casos também se correlaciona com a quantidade de viagens de saída  $F^{(i)}$  de cada bairro no período anterior às restrições de mobilidade (semana 11). É maior em bairros com uma maior quantidade de viagens. Historicamente, esses são os bairros que sempre têm um alto número de casos de dengue, o que pode ser consequência do alto nível de contato entre indivíduos.

**Figura 33 – Série temporal do número de casos de dengue (reais e previstos) para a cidade de Fortaleza nos anos de 2020 (a), 2012 (b) e 2013 (c). As curvas azuis representam a série temporal real, e a curva vermelha representa a média dos casos totais das 5 previsões feitas para cada ano, juntamente com o intervalo de confiança de 90% (C.I. 90%). Em 2012, houve 38.448 casos de dengue, sendo que o modelo subestimou o número de casos com uma média de 24.139,90. Em 2013, houve 8.687 casos, e o valor médio das previsões foi de 5.794 casos. No total, houve 7.720 casos ao longo de 2020. A série temporal prevista pelo modelo de rede neural é superestimada, prevendo um ano com 27.792 casos.**

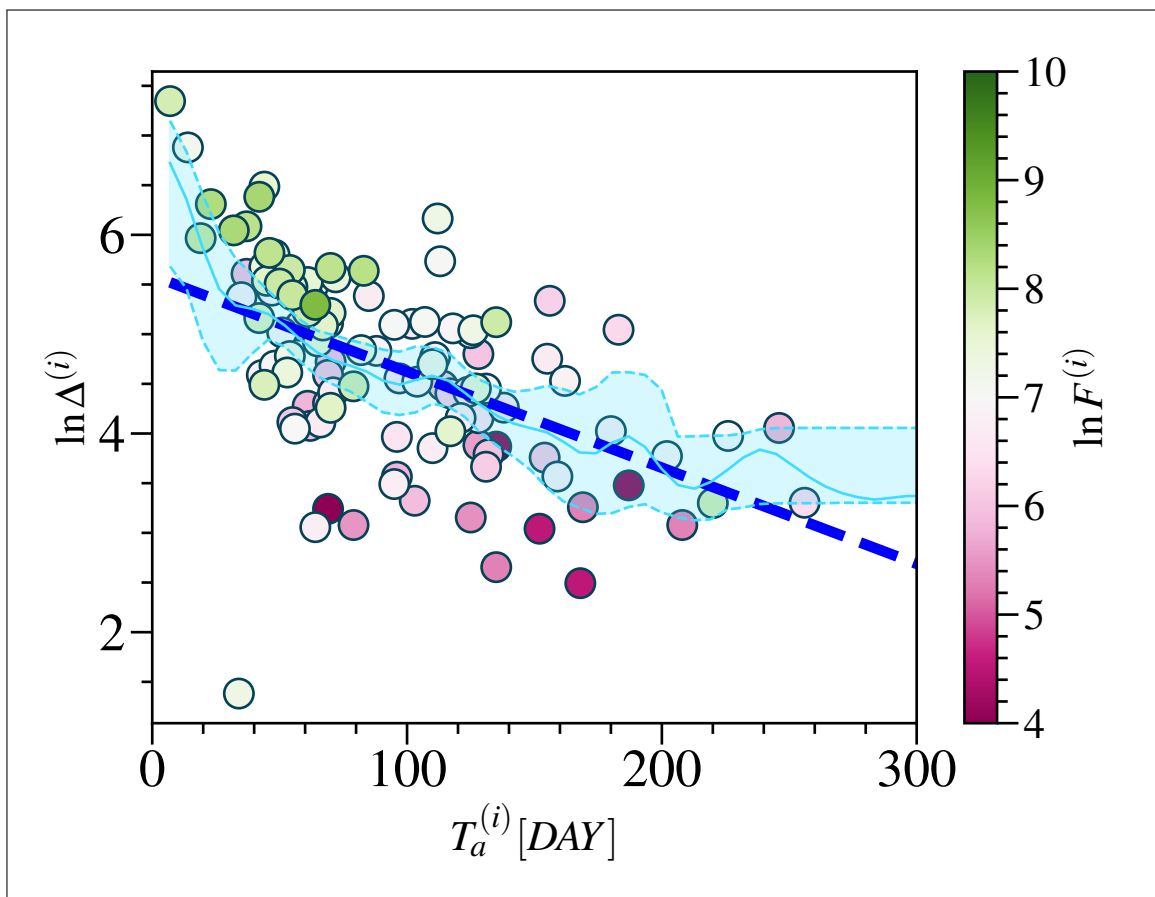


**Figura 34 – Previsão do tempo efetivo de transmissão da dengue em 2020. Utilizando um modelo de rede neural, foi possível estimar o número total de casos de dengue para cada bairro  $i$  da cidade de Fortaleza em 2020, em um cenário hipotético em que as restrições de mobilidade não tivessem sido implementadas. Essa medida é observada no final do ano e armazenada na variável  $Pred^{(i)}$ . Observamos a série temporal que foi prevista para cada bairro durante o ano de 2020 e calculamos o tempo de chegada da doença prevista pela rede ( $T_a^{(i)-Pred}$ ). A relação  $T_a^{(i)-Pred} \times \left( \frac{Pred_T^{(i)} \times 10^4}{N^{(i)}} \right)$  segue um comportamento exponencial na forma da Equação 3.3. Assim, é possível estimar o tempo efetivo de transmissão da doença, onde  $\tau \approx 71$  dias. Portanto, nossas previsões indicam que em 2020 a dengue contaminaria os bairros de Fortaleza de forma mais rápida se não houvesse a implementação de medidas restritivas de mobilidade urbana.**



Fonte: Elaborado pelo autor

**Figura 35 – Impacto das medidas restritivas no número de casos de dengue. Aqui, observamos a quantidade de casos previstos ( $Pred^{(i)}$ ) para cada bairro na cidade de Fortaleza pelo modelo de rede neural em um cenário hipotético em que nenhuma medida de restrição de mobilidade foi implementada. Quantificamos o impacto das medidas restritivas como a diferença entre o número total de casos previstos pelo modelo de rede neural em relação ao número total de casos observados (veja a Equação 5.1). A relação entre  $T_a^{(i)} \times \Delta^{(i)}$  indica que bairros inicialmente infectados experimentaram maiores reduções no número total de casos devido à implementação de medidas restritivas. Curiosamente, as OLs iniciais são geralmente conhecidas como geradoras de grandes quantidades de viagens de saída ( $F^{(i)}$ ) durante o período imediatamente antes das restrições de mobilidade. A combinação dessas observações sugere que a mobilidade urbana é importante para o processo de disseminação da dengue. As linhas tracejadas indicam ajustes lineares como um guia a seguir.**



Fonte: Elaborado pelo autor

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Este capítulo é dedicado às considerações finais, como conclusão desta tese. Na seção 6.1, apresentam-se os resultados alcançados com a execução deste trabalho, bem como as principais contribuições desta pesquisa. A seção 6.2 descreve as limitações deste trabalho e realiza uma discussão sobre as principais dificuldades encontradas para ampliar o escopo da pesquisa. Por fim, na seção 6.3, são sugeridas proposições para pesquisas futuras.

### 6.1 CONTRIBUIÇÕES DO TRABALHO

Neste trabalho, identificou-se que as restrições de mobilidade aplicadas pelo poder público de Fortaleza, com o objetivo de frear a transmissão de COVID-19 durante a pandemia em 2020, contribuíram também para reduzir a velocidade de transmissão de casos de dengue. Considerando que a dengue é uma doença sazonal e que sua transmissão é intensificada durante o período chuvoso, os regimes de isolamento social de 2020 contribuíram para a redução do número de casos de dengue daquele ano para níveis não epidêmicos.

Ao demonstrar que a restrição da mobilidade humana por meio do isolamento social é capaz de reduzir também a propagação do vírus da dengue, e que a população residente é um fator importante que funciona como limitador da intensidade de transmissão nos bairros, foi natural considerar que os dados de mobilidade humana podem ser incorporados a um modelo de rede neural, fornecendo informações sobre o fluxo de pessoas infectadas e sua capacidade de transmitir a doença, bem como a participação de cada bairro nessa transmissão. Dessa forma, esta tese propõe uma arquitetura de rede neural capaz de aprender o impacto da mobilidade humana na transmissão do vírus da dengue. A arquitetura de rede neural, utilizando dados de mobilidade através do Node2Vec, foi capaz de prever com maior antecedência os casos de dengue a nível de bairro. Ela previu de forma mais precisa o momento do pico de casos, a intensidade do pico e a série temporal como um todo, quando comparada com os modelos *baseline*, S.E.I.R. e o modelo ARIMA.

Tendo em vista que não é sempre que dados de mobilidade humana estão a disposição, principalmente com a granularidade que os dados desta tese apresentam é importância estudar outras formas de incorporar aos modelos preditivos informações que permitam o aprendizado do impacto da evolução de casos de dengue de um bairro em outros bairros. Então, devido à existência de uma correlação linear entre a população residente de um bairro e o número de casos, foi proposta uma função de controle de intensidade capaz de fornecer informações heurísticas

a uma arquitetura de rede neural sobre a participação da transmissão de casos de dengue de cada bairro na transmissão total de casos na cidade, além de servir como um limitador superior de casos para cada bairro. Observou-se que a rede neural que utilizou essa informação na predição de casos semanais, a nível de bairro, apresentou melhores resultados quando comparada a arquiteturas de redes neurais que utilizaram apenas dados de casos de dengue e/ou incidência de casos.

Por fim, por meio da previsão de casos de dengue em 2020 utilizando uma arquitetura de rede neural baseada em LSTM, constatou-se que o isolamento social foi responsável por uma redução de 300% no número de casos de dengue durante o período de *lockdown*.

Conclui-se neste trabalho que, mesmo existindo diversos fatores responsáveis pela evolução da transmissão do vírus da dengue em uma grande cidade, o fator humano é determinante na transmissão do vírus em nível de bairro, sendo capaz de reduzir ou aumentar a força de transmissão. Também é concluído que modelos baseados em redes neurais são capazes de aprender a evolução desta transmissão e incorporar o fator humano como atributo para este aprendizado, considerando a população residente e a mobilidade humana.

## 6.2 LIMITAÇÕES

As limitações deste estudo abrangem as características relacionadas à abordagem da evolução dos casos de dengue em áreas urbanas, bem como à perspectiva do algoritmo de predição.

Para uma análise completa e para auxiliar nas medidas de combate à propagação de casos de dengue em grandes cidades, é crucial compreender a disseminação do vírus dentro das regiões urbanas. Isso permitirá que as autoridades públicas identifiquem com precisão qual região da cidade será afetada por um surto de dengue, possibilitando a tomada de medidas preventivas para interromper o surto local antes que se espalhe por toda a cidade. Foi identificada a importância da mobilidade humana na transmissão do vírus da dengue, porém não foi possível compreender o caminho da evolução do vírus dentro da cidade.

O segundo aspecto limitante deste estudo está relacionado à implementação de um modelo utilizando algoritmos baseados em Transformers (VASWANI *et al.*, 2017). O Transformer é uma arquitetura de rede neural que representa o estado da arte na predição de séries temporais, com ênfase em modelos linguísticos. Portanto, é essencial investigar o impacto dessa arquitetura para compreender sua eficiência em diferentes domínios.

### 6.3 TRABALHOS FUTUROS

Neste trabalho, propôs-se uma informação heurística a ser utilizada como entrada em uma rede neural, com o objetivo de fornecer ao modelo as informações necessárias para aprender a participação e o impacto de cada bairro na evolução dos casos de dengue na cidade de Fortaleza, além de funcionar como um limitador de casos de dengue para cada bairro. Também foi proposta uma arquitetura de rede neural capaz de aprender o impacto da mobilidade humana na transmissão do vírus da dengue a nível de bairro. Embora, de maneira geral, todas as abordagens propostas tenham apresentado resultados melhores em comparação com abordagens tradicionais, nem todas se destacaram em todas as métricas. Houve situações em que a informação heurística gerada pela função de controle de intensidade não obteve bons resultados, assim como os modelos S.E.I.R. apresentaram um desempenho superior na seção de previsão de casos de dengue do que os modelos de redes neurais. Considerando que as soluções propostas nesta tese não apresentaram resultados superiores em todas as métricas, é importante considerar a possibilidade de combinar os resultados de todos os modelos para gerar um super-ensemble.

Uma abordagem de *super-ensemble*, tem por objetivo combinar resultados produzidos por diferentes métodos por meio de médias ponderadas com base em seu desempenho histórico. Além disso, uma abordagem de super-ensemble pode se beneficiar da capacidade de cada sistema de considerar diferentes tipos de informações, como informações espaciais, temporais, etc., e de representar essas informações de maneiras distintas. Por exemplo, um modelo pode representar a dependência espacial usando *embeddings* gerados pelo algoritmo *Node2Vec*, enquanto outro modelo pode representá-la por meio de uma rede de convolução em grafo (KIPF; WELLING, 2016). Além disso, um conjunto de modelos pode permitir a ponderação dos modelos em relação às características básicas da série temporal e/ou às propriedades espaciais. Por exemplo, um modelo LSTM sem dados espaciais ou de transporte pode produzir melhores resultados para bairros com baixa quantidade de movimento.



## REFERÊNCIAS

- AGUILAR-SÁNCHEZ, J.; BASSOLAS, A.; GHOSHAL, G.; HAZARIE, S. A.; KIRKLEY, A.; MAZZOLI, M.; MELONI, S.; MIMAR, S.; NICOSIA, V.; RAMASCO, J. J. *et al.* Impact of urban structure on infectious disease spreading. 2021.
- ANDERSON, J. L. An ensemble adjustment kalman filter for data assimilation. **Monthly weather review**, v. 129, n. 12, p. 2884–2903, 2001.
- ARIYO, A. A.; ADEWUMI, A. O.; AYO, C. K. Stock price prediction using the arima model. In: IEEE. **2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation**. [S.l.], 2014. p. 106–112.
- BAJARDI, P.; POLETTO, C.; RAMASCO, J. J.; TIZZONI, M.; COLIZZA, V.; VESPIGNANI, A. Human mobility networks, travel restrictions, and the global spread of 2009 h1n1 pandemic. **PloS one**, Public Library of Science San Francisco, USA, v. 6, n. 1, p. e16591, 2011.
- BHATT, S.; GETHING, P. W.; BRADY, O. J.; MESSINA, J. P.; FARLOW, A. W.; MOYES, C. L.; DRAKE, J. M.; BROWNSTEIN, J. S.; HOEN, A. G.; SANKOH, O. *et al.* The global distribution and burden of dengue. **Nature**, Nature Publishing Group UK London, v. 496, n. 7446, p. 504–507, 2013.
- BIGGERSTAFF, M.; JOHANSSON, M.; ALPER, D.; BROOKS, L. C.; CHAKRABORTY, P.; FARROW, D. C.; HYUN, S.; KANDULA, S.; MCGOWAN, C.; RAMAKRISHNAN, N. *et al.* Results from the second year of a collaborative effort to forecast influenza seasons in the united states. **Epidemics**, Elsevier, v. 24, p. 26–33, 2018.
- BISGAARD, S.; KULAHCI, M. **Time series analysis and forecasting by example**. [S.l.]: John Wiley & Sons, 2011.
- BOMFIM, R.; PEI, S.; SHAMAN, J.; YAMANA, T.; MAKSE, H. A.; JR, J. S. A.; NETO, A. S. L.; FURTADO, V. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. **Journal of the Royal Society Interface**, The Royal Society, v. 17, n. 171, p. 20200691, 2020.
- BONTEMPI, G.; TAIEB, S. B.; BORGNE, Y.-A. L. Machine learning strategies for time series forecasting. In: SPRINGER. **European business intelligence summer school**. [S.l.], 2012. p. 62–77.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BROCKMANN, D.; HELBING, D. The hidden geometry of complex, network-driven contagion phenomena. **science**, American Association for the Advancement of Science, v. 342, n. 6164, p. 1337–1342, 2013.
- CAI, Y.; LI, J.; KANG, Y.; WANG, K.; WANG, W. The fluctuation impact of human mobility on the influenza transmission. **Journal of the Franklin Institute**, Elsevier, v. 357, n. 13, p. 8899–8924, 2020.
- CAMINHA, C.; FURTADO, V. Impact of human mobility on police allocation. In: IEEE. **2017 IEEE International Conference on Intelligence and Security Informatics (ISI)**. [S.l.], 2017. p. 125–127.

CAMINHA, C.; FURTADO, V.; PEQUENO, T. H.; PONTE, C.; MELO, H. P.; OLIVEIRA, E. A.; JR, J. S. A. Human mobility in large cities as a proxy for crime. **PloS one**, Public Library of Science San Francisco, CA USA, v. 12, n. 2, p. e0171609, 2017.

CAMINHA, C.; FURTADO, V.; PINHEIRO, V.; SILVA, C. Micro-interventions in urban transportation from pattern discovery on the flow of passengers and on the bus network. In: **IEEE. 2016 IEEE International Smart Cities Conference (ISC2)**. [S.l.], 2016. p. 1–6.

CAMINHA, C.; FURTADO, V.; PINHEIRO, V.; PONTE, C. Graph mining for the detection of overcrowding and waste of resources in public transport. **Journal of Internet Services and Applications**, SpringerOpen, v. 9, n. 1, p. 22, 2018.

CASAGRANDE, R.; BOLZONI, L.; LEVIN, S. A.; ANDREASEN, V. The sirc model and influenza a. **Mathematical biosciences**, Elsevier, v. 200, n. 2, p. 152–169, 2006.

CENGİZ, S.; GÖRMÜŞ, S.; OĞUZ, D. Analysis of the urban growth pattern through spatial metrics; ankara city. **Land Use Policy**, Elsevier, v. 112, p. 105812, 2022.

CHANPRASOPCHAI, P.; TANG, I. M.; PONGSUMPUN, P. Sir model for dengue disease with effect of dengue vaccination. **Computational and mathematical methods in medicine**, Hindawi, v. 2018, 2018.

CHEN, Y.-C.; LU, P.-E.; CHANG, C.-S.; LIU, T.-H. A time-dependent sir model for covid-19 with undetectable infected persons. **Ieee transactions on network science and engineering**, IEEE, v. 7, n. 4, p. 3279–3294, 2020.

CHO, K.; MERRIËNBOER, B. V.; BAHDANAU, D.; BENGIO, Y. On the properties of neural machine translation: Encoder-decoder approaches. **arXiv preprint arXiv:1409.1259**, 2014.

CHO, K.; MERRIËNBOER, B. V.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. **arXiv preprint arXiv:1406.1078**, 2014.

CHO, W.; KIM, S.; NA, M.; NA, I. Forecasting of tomato yields using attention-based lstm network and arma model. **Electronics**, MDPI, v. 10, n. 13, p. 1576, 2021.

COOPER, I.; MONDAL, A.; ANTONOPOULOS, C. G. A sir model assumption for the spread of covid-19 in different communities. **Chaos, Solitons & Fractals**, Elsevier, v. 139, p. 110057, 2020.

COSTA, A. C. C. da; CODEÇO, C. T.; KRAINSKI, E. T.; GOMES, M. F. d. C.; NOBRE, A. A. Spatiotemporal diffusion of influenza a (h1n1): Starting point and risk factors. **PloS one**, Public Library of Science San Francisco, CA USA, v. 13, n. 9, p. e0202832, 2018.

DHRUV, P.; NASKAR, S. Image classification using convolutional neural network (cnn) and recurrent neural network (rnn): a review. **Machine Learning and Information Processing: Proceedings of ICMLIP 2019**, Springer, p. 367–381, 2020.

ELMAN, J. L. Finding structure in time. **Cognitive science**, Wiley Online Library, v. 14, n. 2, p. 179–211, 1990.

FARUK, M.; JANNAT, S.; RAHMAN, M. S. Impact of environmental factors on the spread of dengue fever in sri lanka. **International Journal of Environmental Science and Technology**, Springer, v. 19, n. 11, p. 10637–10648, 2022.

FENTE, D. N.; SINGH, D. K. Weather forecasting using artificial neural network. In: IEEE. **2018 second international conference on inventive communication and computational technologies (ICICCT)**. [S.l.], 2018. p. 1757–1761.

FREUND, Y.; SCHAPIRE, R. E. Large margin classification using the perceptron algorithm. In: **Proceedings of the eleventh annual conference on Computational learning theory**. [S.l.: s.n.], 1998. p. 209–217.

FURTADO, V.; FURTADO, E.; CAMINHA, C.; LOPES, A.; DANTAS, V.; PONTE, C.; CAVALCANTE, S. A data-driven approach to help understanding the preferences of public transport users. In: IEEE. **2017 IEEE International Conference on Big Data (Big Data)**. [S.l.], 2017. p. 1926–1935.

GAL, Y.; GHAHRAMANI, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: PMLR. **international conference on machine learning**. [S.l.], 2016. p. 1050–1059.

GALVEZ, R. L.; BANDALA, A. A.; DADIOS, E. P.; VICERRA, R. R. P.; MANINGO, J. M. Z. Object detection using convolutional neural networks. In: IEEE. **TENCON 2018-2018 IEEE Region 10 Conference**. [S.l.], 2018. p. 2023–2027.

GIBBONS, R. V.; VAUGHN, D. W. Dengue: an escalating problem. **Bmj**, British Medical Journal Publishing Group, v. 324, n. 7353, p. 1563–1566, 2002.

GOOGLE. **Google - COVID-19 Community Mobility Reports**. 2022. Disponível em: <<https://www.google.com/covid19/mobility/>>.

GRAIS, R. F.; ELLIS, J. H.; GLASS, G. E. Assessing the impact of airline travel on the geographic spread of pandemic influenza. **European journal of epidemiology**, Springer, v. 18, p. 1065–1072, 2003.

GRAVES, A.; GRAVES, A. Long short-term memory. **Supervised sequence labelling with recurrent neural networks**, Springer, p. 37–45, 2012.

GROVER, A.; LESKOVEC, J. node2vec: Scalable feature learning for networks. In: ACM. **Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2016. p. 855–864.

HÂNCEAN, M.-G.; SLAVINEC, M.; PERC, M. The impact of human mobility networks on the global spread of covid-19. **Journal of Complex Networks**, Oxford University Press, v. 8, n. 6, p. cnaa041, 2020.

HE, Z.; TAO, H. Epidemiology and arima model of positive-rate of influenza viruses among children in wuhan, china: A nine-year retrospective study. **International Journal of Infectious Diseases**, Elsevier, v. 74, p. 61–70, 2018.

IBGE. **Instituto Brasileiro de Geografia e Estatística (IBGE)**. 2023. Disponível em: <<http://www.ibge.gov.br>>.

IBGE. **POPULAÇÃO RURAL E URBANA**. 2023. Disponível em: <<https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18313-populacao-rural-e-urbana.html>>.

JOHANSSON, M. A.; APFELDORF, K. M.; DOBSON, S.; DEVITA, J.; BUCZAK, A. L.; BAUGHER, B.; MONIZ, L. J.; BAGLEY, T.; BABIN, S. M.; GUVEN, E. *et al.* An open challenge to advance probabilistic forecasting for dengue epidemics. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 116, n. 48, p. 24268–24274, 2019.

KANDULA, S.; YAMANA, T.; PEI, S.; YANG, W.; MORITA, H.; SHAMAN, J. Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. **Journal of The Royal Society Interface**, The Royal Society, v. 15, n. 144, p. 20180174, 2018.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. **Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character**, The Royal Society London, v. 115, n. 772, p. 700–721, 1927.

KHATUA, A.; KAR, T. K.; NANDI, S. K.; JANA, S.; KANG, Y. Impact of human mobility on the transmission dynamics of infectious diseases. **Energy, Ecology and Environment**, Springer, v. 5, p. 389–406, 2020.

KIPF, T. N.; WELLING, M. Semi-supervised classification with graph convolutional networks. **arXiv preprint arXiv:1609.02907**, 2016.

KOLIMENAKIS, A.; HEINZ, S.; WILSON, M. L.; WINKLER, V.; YAKOB, L.; MICHAELAKIS, A.; PAPACHRISTOS, D.; RICHARDSON, C.; HORSTICK, O. The role of urbanisation in the spread of aedes mosquitoes and the diseases they transmit—a systematic review. **PLoS neglected tropical diseases**, Public Library of Science San Francisco, CA USA, v. 15, n. 9, p. e0009631, 2021.

LECUN, Y.; BOSER, B.; DENKER, J.; HENDERSON, D.; HOWARD, R.; HUBBARD, W.; JACKEL, L. Handwritten digit recognition with a back-propagation network. **Advances in neural information processing systems**, v. 2, 1989.

LIU, L.; HAN, M.; ZHOU, Y.; WANG, Y. Lstm recurrent neural networks for influenza trends prediction. In: SPRINGER. **Bioinformatics Research and Applications: 14th International Symposium, ISBRA 2018, Beijing, China, June 8-11, 2018, Proceedings 14**. [S.l.], 2018. p. 259–264.

MA, L.; KITSACK, M.; MIEGHEM, P. V. Two-population sir model and strategies to reduce mortality in pandemics. In: SPRINGER. **Complex Networks & Their Applications X: Volume 2, Proceedings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021 10**. [S.l.], 2022. p. 265–276.

MANRIQUE-SAIDE, P.; COLEMAN, P.; MCCALL, P.; LENHART, A.; VÁZQUEZ-PROKOPEC, G.; DAVIES, C. Multi-scale analysis of the associations among egg, larval and pupal surveys and the presence and abundance of adult female aedes aegypti (*stegomyia aegypti*) in the city of merida, mexico. **Medical and veterinary entomology**, Wiley Online Library, v. 28, n. 3, p. 264–272, 2014.

MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. Efficient estimation of word representations in vector space. **arXiv preprint arXiv:1301.3781**, 2013.

MONTERRUBIO, J. C. Short-term economic impacts of influenza a (h1n1) and government reaction on the mexican tourism industry: an analysis of the media. **International journal of tourism policy**, Inderscience Publishers, v. 3, n. 1, p. 1–15, 2010.

MYERS, L.; SIROIS, M. J. Spearman correlation coefficients, differences between. **Encyclopedia of statistical sciences**, Wiley Online Library, v. 12, 2004.

NEIDERUD, C.-J. How urbanization affects the epidemiology of emerging infectious diseases. **Infection ecology & epidemiology**, Taylor & Francis, v. 5, n. 1, p. 27060, 2015.

ORHAN, U.; HEKIM, M.; OZER, M. Eeg signals classification using the k-means clustering and a multilayer perceptron neural network model. **Expert Systems with Applications**, Elsevier, v. 38, n. 10, p. 13475–13481, 2011.

OSTHUS, D.; HICKMANN, K. S.; CARAGEA, P. C.; HIGDON, D.; VALLE, S. Y. D. Forecasting seasonal influenza with a state-space sir model. **The annals of applied statistics**, NIH Public Access, v. 11, n. 1, p. 202, 2017.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. *et al.* Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011.

PEI, S.; CANE, M. A.; SHAMAN, J. Predictability in process-based ensemble forecast of influenza. **PLoS computational biology**, Public Library of Science, v. 15, n. 2, p. e1006783, 2019.

PEI, S.; KANDULA, S.; YANG, W.; SHAMAN, J. Forecasting the spatial transmission of influenza in the united states. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 115, n. 11, p. 2752–2757, 2018.

PEI, S.; MORONE, F.; LILJEROS, F.; MAKSE, H.; SHAMAN, J. L. Inference and control of the nosocomial transmission of methicillin-resistant staphylococcus aureus. **eLife**, eLife Sciences Publications Limited, v. 7, p. e40977, 2018.

PEI, S.; SHAMAN, J. Counteracting structural errors in ensemble forecast of influenza outbreaks. **Nature communications**, Nature Publishing Group, v. 8, n. 1, p. 925, 2017.

PILGER, D.; LENHART, A.; MANRIQUE-SAIDE, P.; SIQUEIRA, J. B.; ROCHA, W. T. D.; KROEGER, A. Is routine dengue vector surveillance in central brazil able to accurately monitor the aedes aegypti population? results from a pupal productivity survey. **Tropical Medicine & International Health**, Wiley Online Library, v. 16, n. 9, p. 1143–1150, 2011.

PONTE, C.; CAMINHA, C.; FURTADO, V. Busca de melhor caminho entre dois pontos quando múltiplas origens e múltiplos destinos são possíveis. **Recife: ENIAC**, 2016.

PONTE, C.; CARMONA, H. A.; OLIVEIRA, E. A.; CAMINHA, C.; LIMA, A. S.; JR, J. S. A.; FURTADO, V. Tracing contacts to evaluate the transmission of covid-19 from highly exposed individuals in public transportation. **Scientific Reports**, Nature Publishing Group UK London, v. 11, n. 1, p. 24443, 2021.

POORNIMA, S.; PUSHPALATHA, M. Prediction of rainfall using intensified lstm based recurrent neural network with weighted linear units. **Atmosphere**, MDPI, v. 10, n. 11, p. 668, 2019.

RAHMAN, M.; ISLAM, A. S.; NADVI, S. Y. M.; RAHMAN, R. M. Comparative study of anfis and arima model for weather forecasting in dhaka. In: IEEE. **2013 International Conference on Informatics, Electronics and Vision (ICIEV)**. [S.l.], 2013. p. 1–6.

REICH, N. G.; BROOKS, L. C.; FOX, S. J.; KANDULA, S.; MCGOWAN, C. J.; MOORE, E.; OSTHUS, D.; RAY, E. L.; TUSHAR, A.; YAMANA, T. K. *et al.* A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the united states. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 116, n. 8, p. 3146–3154, 2019.

RUCK, D. W.; ROGERS, S. K.; KABRISKY, M. Feature selection using a multilayer perceptron. **Journal of Neural Network Computing**, Citeseer, v. 2, n. 2, p. 40–48, 1990.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning internal representations by error propagation**. [S.l.], 1985.

SALEH, A. Y.; BAIWEI, L. Dengue prediction using deep learning with long short-term memory. In: IEEE. **2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)**. [S.l.], 2021. p. 1–5.

SALEHIN, I.; TALHA, I. M.; HASAN, M. M.; DIP, S. T.; SAIFUZZAMAN, M.; MOON, N. N. An artificial intelligence based rainfall prediction using lstm and neural network. In: IEEE. **2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)**. [S.l.], 2020. p. 5–8.

SAQIB, M.; KHAN, S. D.; SHARMA, N.; BLUMENSTEIN, M. Person head detection in multiple scales using deep convolutional neural networks. In: IEEE. **2018 International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2018. p. 1–7.

SCARSELLI, F.; GORI, M.; TSOI, A. C.; HAGENBUCHNER, M.; MONFARDINI, G. The graph neural network model. **IEEE transactions on neural networks**, IEEE, v. 20, n. 1, p. 61–80, 2008.

SHAMAN, J.; KARSPECK, A. Forecasting seasonal outbreaks of influenza. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 109, n. 50, p. 20425–20430, 2012.

SHAMAN, J.; KARSPECK, A.; YANG, W.; TAMERIUS, J.; LIPSITCH, M. Real-time influenza forecasts during the 2012–2013 season. **Nature communications**, Nature Publishing Group, v. 4, p. 2837, 2013.

SIDE, S.; NOORANI, S. M. A sir model for spread of dengue fever disease (simulation for south sulawesi, indonesia and selangor, malaysia). **World Journal of Modelling and Simulation**, Citeseer, v. 9, n. 2, p. 96–105, 2013.

SIMDA. **Sistema de Monitoramento Diário de Agravos (SIMDA)**. 2023. Disponível em: <<http://tc1.sms.fortaleza.ce.gov.br/simda>>.

SPANHOL, F. A.; OLIVEIRA, L. S.; PETITJEAN, C.; HEUTTE, L. Breast cancer histopathological image classification using convolutional neural networks. In: IEEE. **2016 international joint conference on neural networks (IJCNN)**. [S.l.], 2016. p. 2560–2567.

STODDARD, S. T.; MORRISON, A. C.; VAZQUEZ-PROKOPEC, G. M.; SOLDAN, V. P.; KOHEL, T. J.; KITRON, U.; ELDER, J. P.; SCOTT, T. W. The role of human movement in the transmission of vector-borne pathogens. **PLoS neglected tropical diseases**, Public Library of Science San Francisco, USA, v. 3, n. 7, p. e481, 2009.

SULLIVAN, D.; CAMINHA, C.; MELO, H. P.; FURTADO, V. Towards understanding the impact of crime on the choice of route by a bus passenger. In: SPRINGER. **EPIA Conference on Artificial Intelligence**. [S.l.], 2017. p. 41–50.

TAIEB, S. B.; HYNDMAN, R. J. *et al.* **Recursive and direct multi-step forecasting: the best of both worlds.** [S.l.]: Citeseer, 2012. v. 19.

VASILAKIS, N.; CARDOSA, J.; HANLEY, K. A.; HOLMES, E. C.; WEAVER, S. C. Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. **Nature Reviews Microbiology**, Nature Publishing Group UK London, v. 9, n. 7, p. 532–541, 2011.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

VIGFUSSON, Y.; KARLSSON, T. A.; ONKEN, D.; SONG, C.; EINARSSON, A. F.; KISHORE, N.; MITCHELL, R. M.; BROOKS-POLLOCK, E.; SIGMUNSDOTTIR, G.; DANON, L. Cell-phone traces reveal infection-associated behavioral change. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 118, n. 6, p. e2005241118, 2021.

VOYANT, C.; MUSELLI, M.; PAOLI, C.; NIVET, M.-L. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. **Energy**, Elsevier, v. 36, n. 1, p. 348–359, 2011.

WANG, L.; BEN, X.; ADIGA, A.; SADILEK, A.; TENDULKAR, A.; VENKATRAMANAN, S.; VULLIKANTI, A.; AGGARWAL, G.; TALEKAR, A.; CHEN, J. *et al.* Using mobility data to understand and forecast covid19 dynamics. **medRxiv**, Cold Spring Harbor Laboratory Preprints, 2020.

WHO. **World Health Organization - Dengue and severe dengue.** 2019. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>>.

YAMANA, T. K.; KANDULA, S.; SHAMAN, J. Superensemble forecasts of dengue outbreaks. **Journal of The Royal Society Interface**, The Royal Society, v. 13, n. 123, p. 20160410, 2016.

YANG, W.; LIPSITCH, M.; SHAMAN, J. Inference of seasonal and pandemic influenza transmission dynamics. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 112, n. 9, p. 2723–2728, 2015.

YANG, W.; ZHANG, W.; KARGBO, D.; YANG, R.; CHEN, Y.; CHEN, Z.; KAMARA, A.; KARGBO, B.; KANDULA, S.; KARSPECK, A. *et al.* Transmission network of the 2014–2015 ebola epidemic in sierra leone. **Journal of The Royal Society Interface**, The Royal Society, v. 12, n. 112, p. 20150536, 2015.

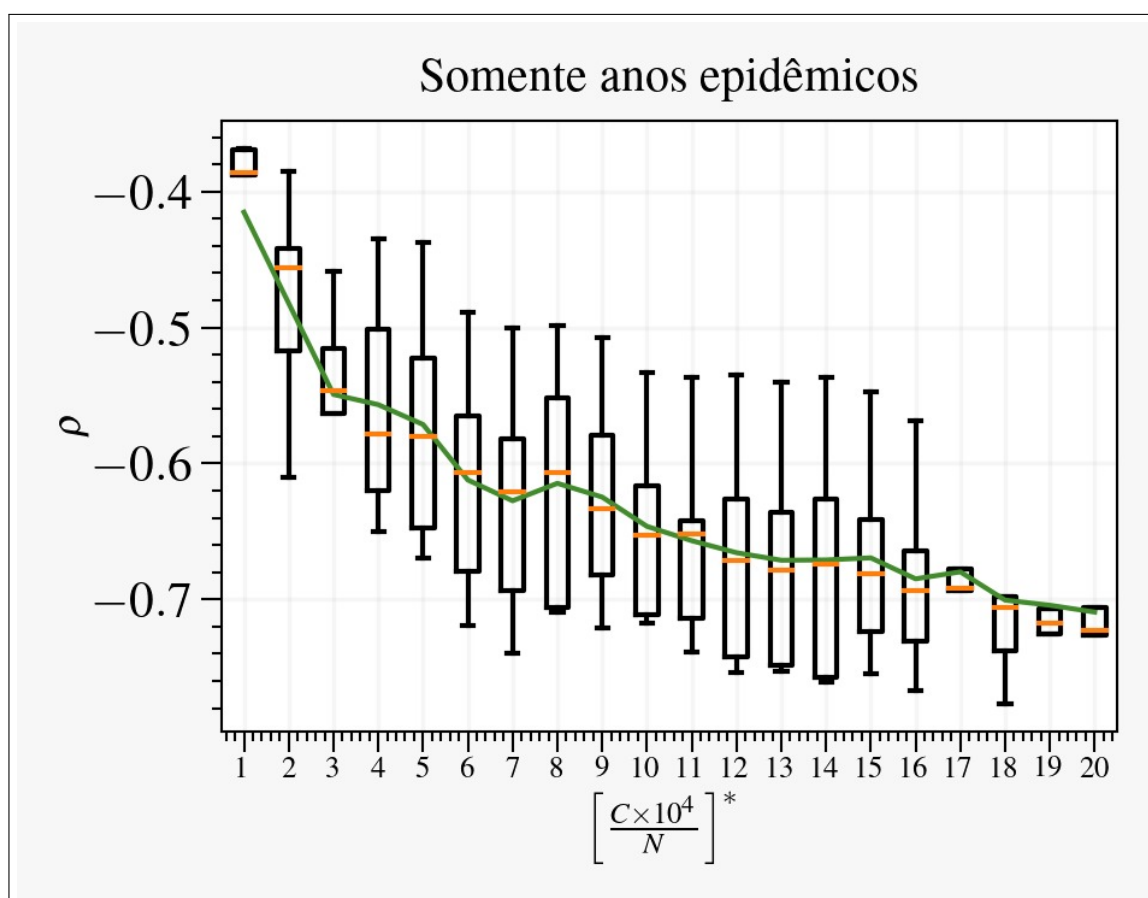
ZHAO, L.; SONG, Y.; ZHANG, C.; LIU, Y.; WANG, P.; LIN, T.; DENG, M.; LI, H. T-gcn: A temporal graph convolutional network for traffic prediction. **IEEE transactions on intelligent transportation systems**, IEEE, v. 21, n. 9, p. 3848–3858, 2019.

## **APÊNDICES**



APÊNDICE A – ANÁLISE DA VARIAÇÃO DO NÚMERO DE CASOS DE DENGUE MÍNIMO PARA A CHEGADA DA DOENÇA SER CONSIDERADA EM UM BAIRRO

**Figura 36 – E relação entre a média dos coeficientes R, representado por  $\rho$ , e a incidência de casos  $\left(\frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)$ . O coeficiente R é extraído pelo *fit* da relação  $\left(T_a^i \times \frac{C_T^{(i)} \times 10^4}{N^{(i)}}\right)$ . Observa-se um mínimo local para incidência sendo 7**

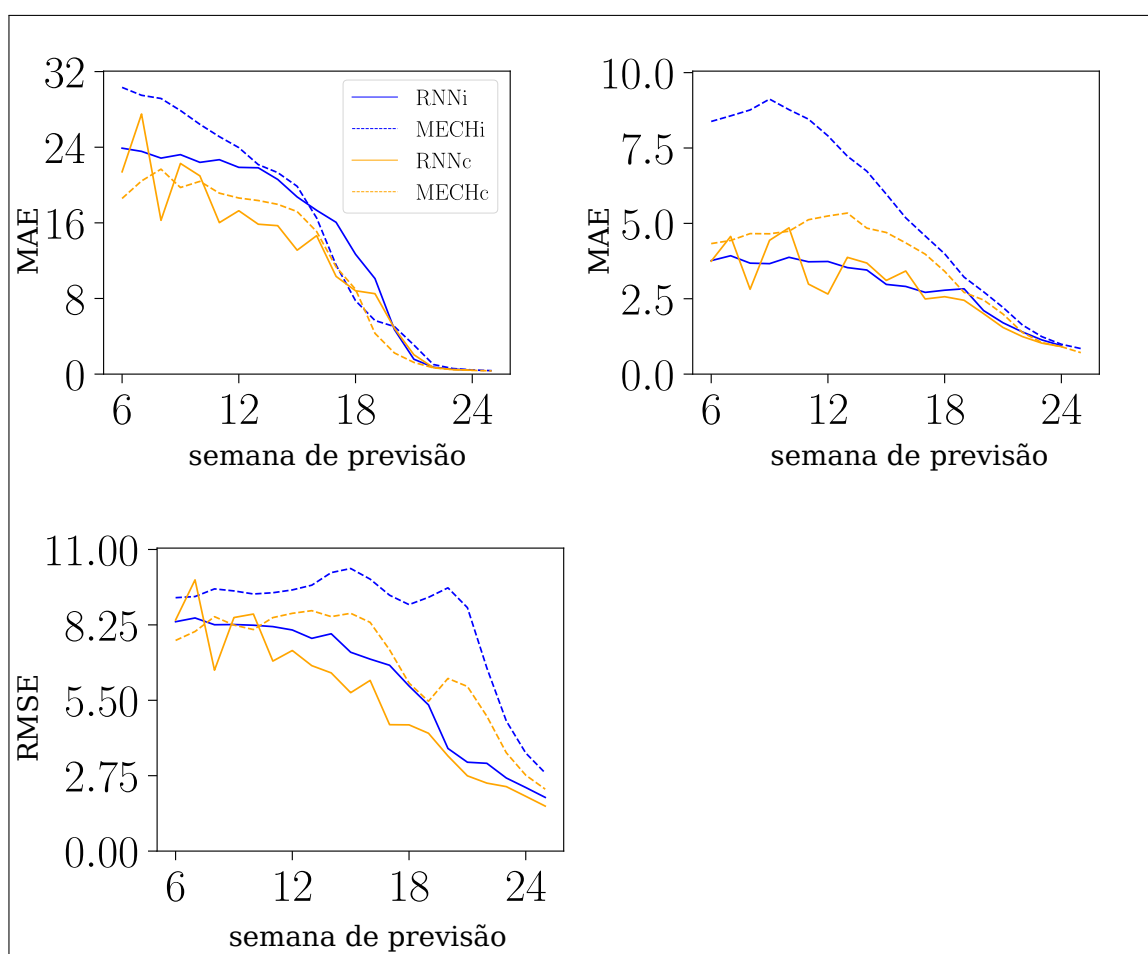


Fonte: Elaborado pelo autor

## APÊNDICE B – ANÁLISE DE LIMIAR EPIDÊMICO DE PREVISÃO DE SURTOS DE DENGUE EM NÍVEL DE BAIRRO USANDO MOBILIDADE HUMANA EM ÁREAS URBANAS

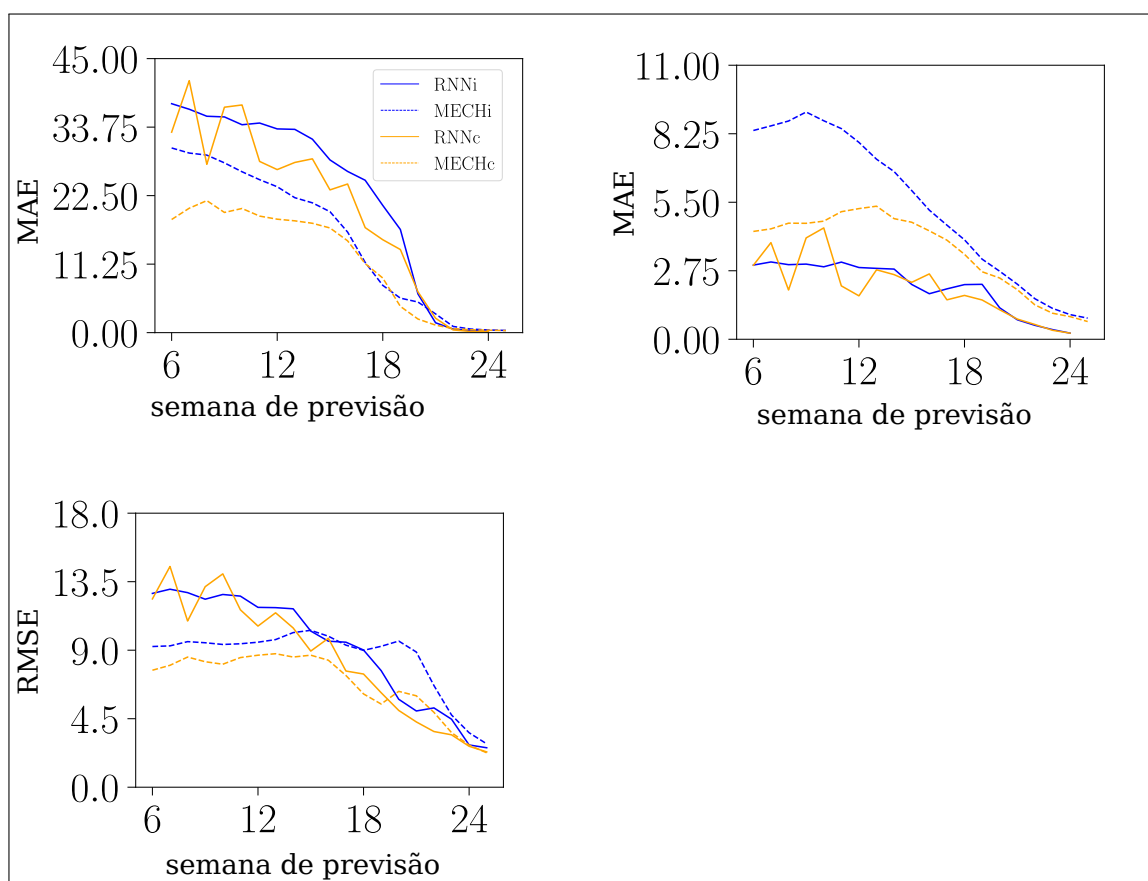
Foi testada a sensibilidade das análises em relação à escolha de 200 casos como limite epidêmico em nível de bairro. Avaliamos a capacidade dos modelos de prever casos de dengue em um período com mais de 100 e mais de 300 casos de dengue. As Figuras 37 e 38 mostram, respectivamente, os valores de MAE (Erro Médio Absoluto) para a intensidade máxima e o tempo, e o RMSE (Erro Médio Quadrático) para a curva completa da série temporal. Ambos os modelos continuam apresentando melhores resultados quando as informações de mobilidade são usadas para prever casos de dengue, demonstrando que a utilidade das informações de deslocamento não é sensível à escolha do limite epidêmico. Com um limite de 100 casos, os modelos RNN têm um desempenho melhor. Quando usamos 300 ou mais casos de dengue, os modelos mecanicistas têm um desempenho melhor para a previsão da intensidade máxima e para o RMSE da série temporal como um todo, e o modelo RNNc tem um desempenho melhor para o MAE do tempo de pico. Esses resultados reforçam as conclusões da análise principal: o modelo mecanicista prevê melhor grandes epidemias, mas superestima períodos com menos casos de dengue, enquanto o modelo de rede neural continua a prever melhor o tempo de pico em todas as análises.

**Figura 37 – Resultados para previsões com períodos e bairros com mais de 100 casos de dengue.**



Fonte: Elaborado pelo autor

**Figura 38 – Resultados para previsões com períodos e bairros com mais de 300 casos de dengue.**

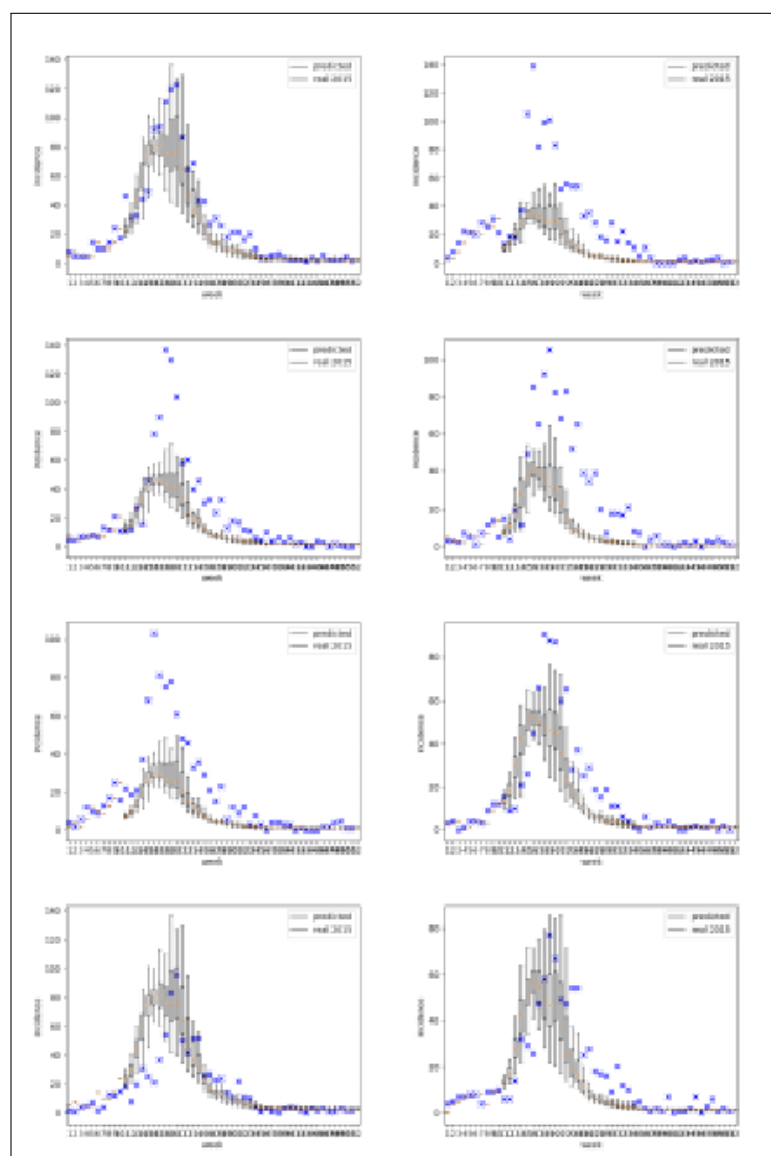


Fonte: Elaborado pelo autor

## APÊNDICE C – EXEMPLOS DE PREDIÇÕES UTILIZANDO DADOS DE MOBILIDADE HUMANA COM REDES NEURAIS E MODELOS COMPORTAMENTAIS

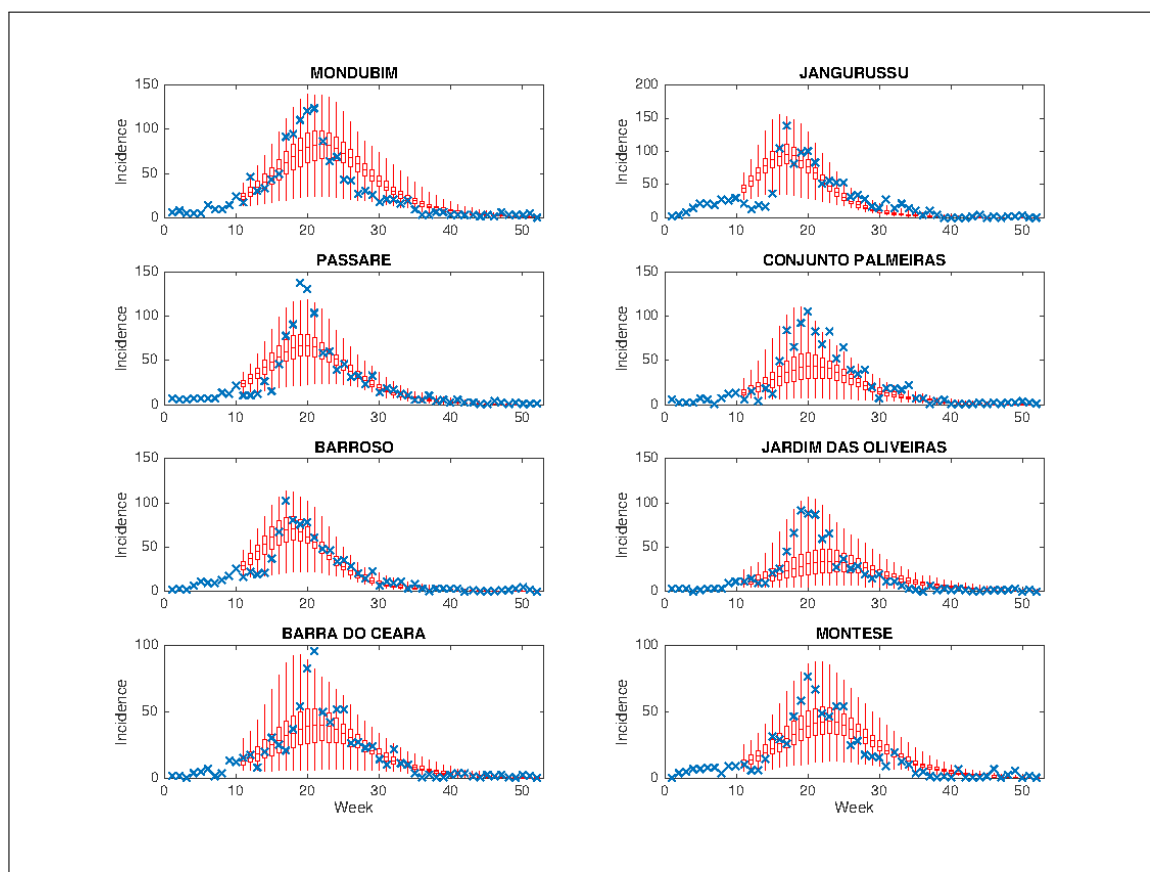
Figura 39 e Figura 40 mostram, respectivamente, exemplos de previsões para o modelo RNNc e MECHc.

**Figura 39 – Exemplo de previsões com modelo RNNc para 8 bairros começando a partir da semana 11.**



Fonte: Elaborado pelo autor

**Figura 40 – Exemplo de previsões com modelo MECHc para 8 bairros começando a partir da semana 11.**

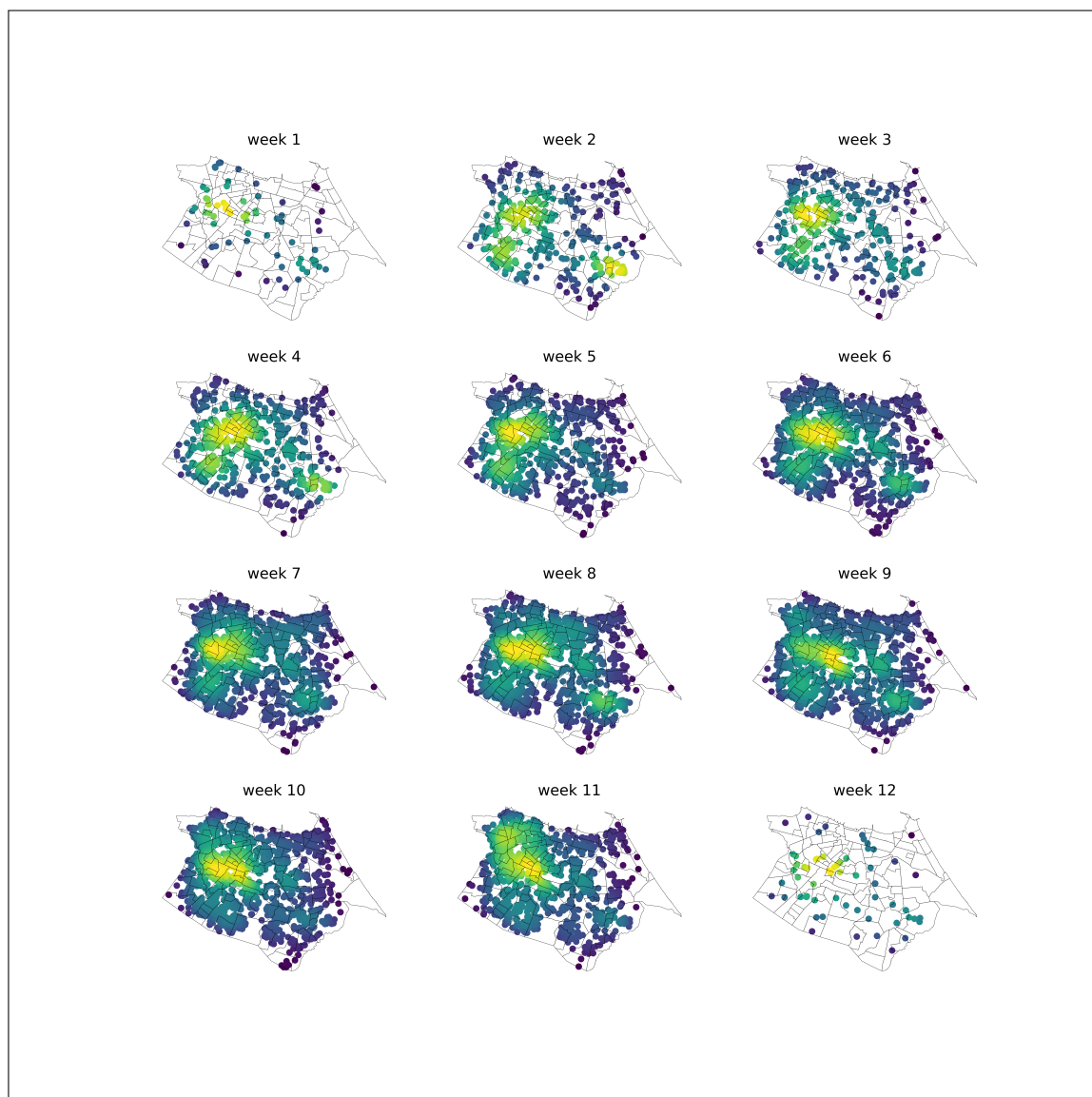


Fonte: Elaborado pelo autor

APÊNDICE D – MAPAS COM A EVOLUÇÃO SEMANAIS DE CASOS DE DENGUE  
PARA OS ANOS COM GRANDES SURTOS DE CASOS DE DENGUE

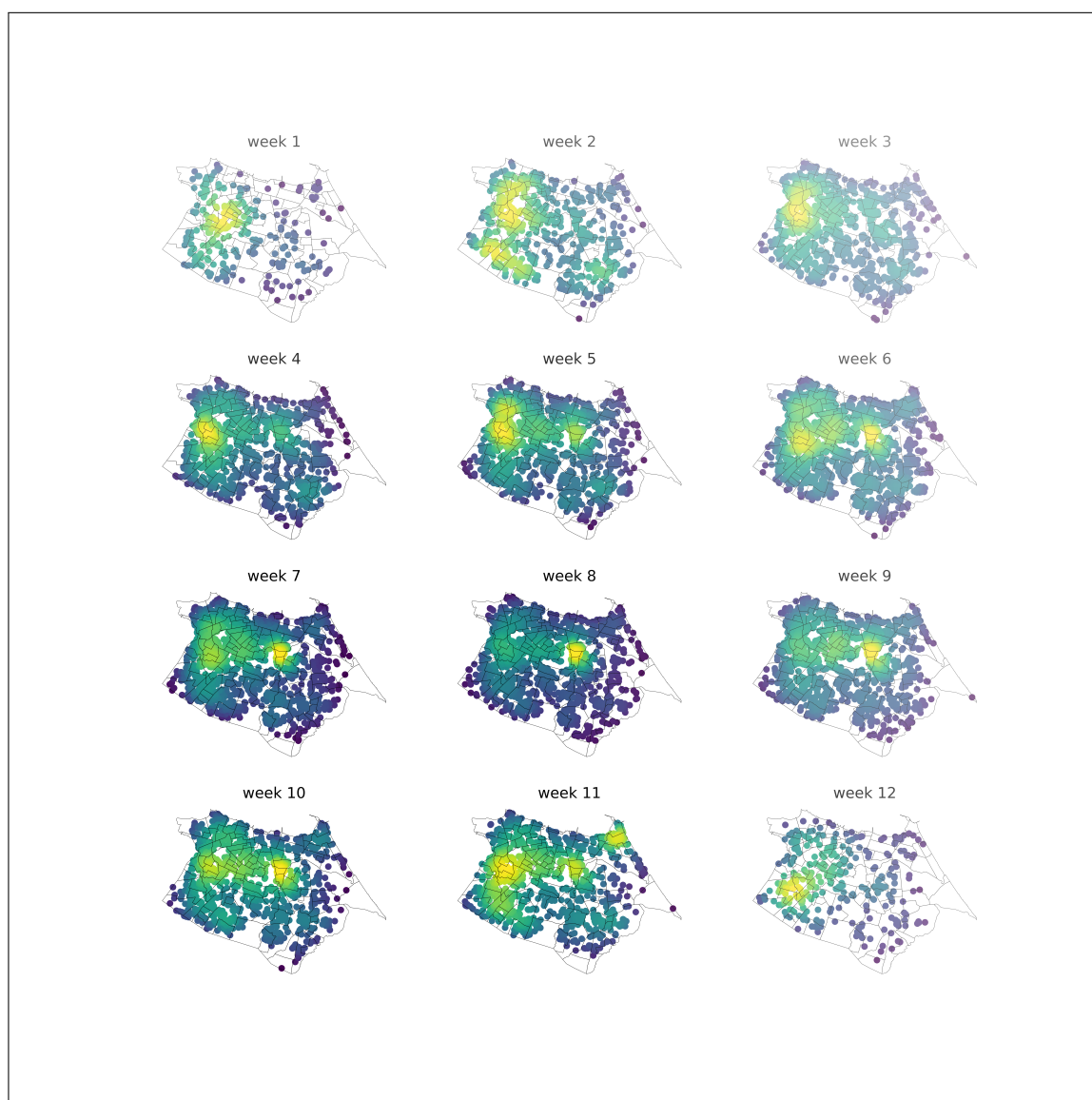
As figuras a seguir mostram a evolução dos casos de dengue para alguns anos de pico, 2008, 2011, 2012 e 2015

**Figura 41 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2008**



Fonte: Elaborado pelo autor

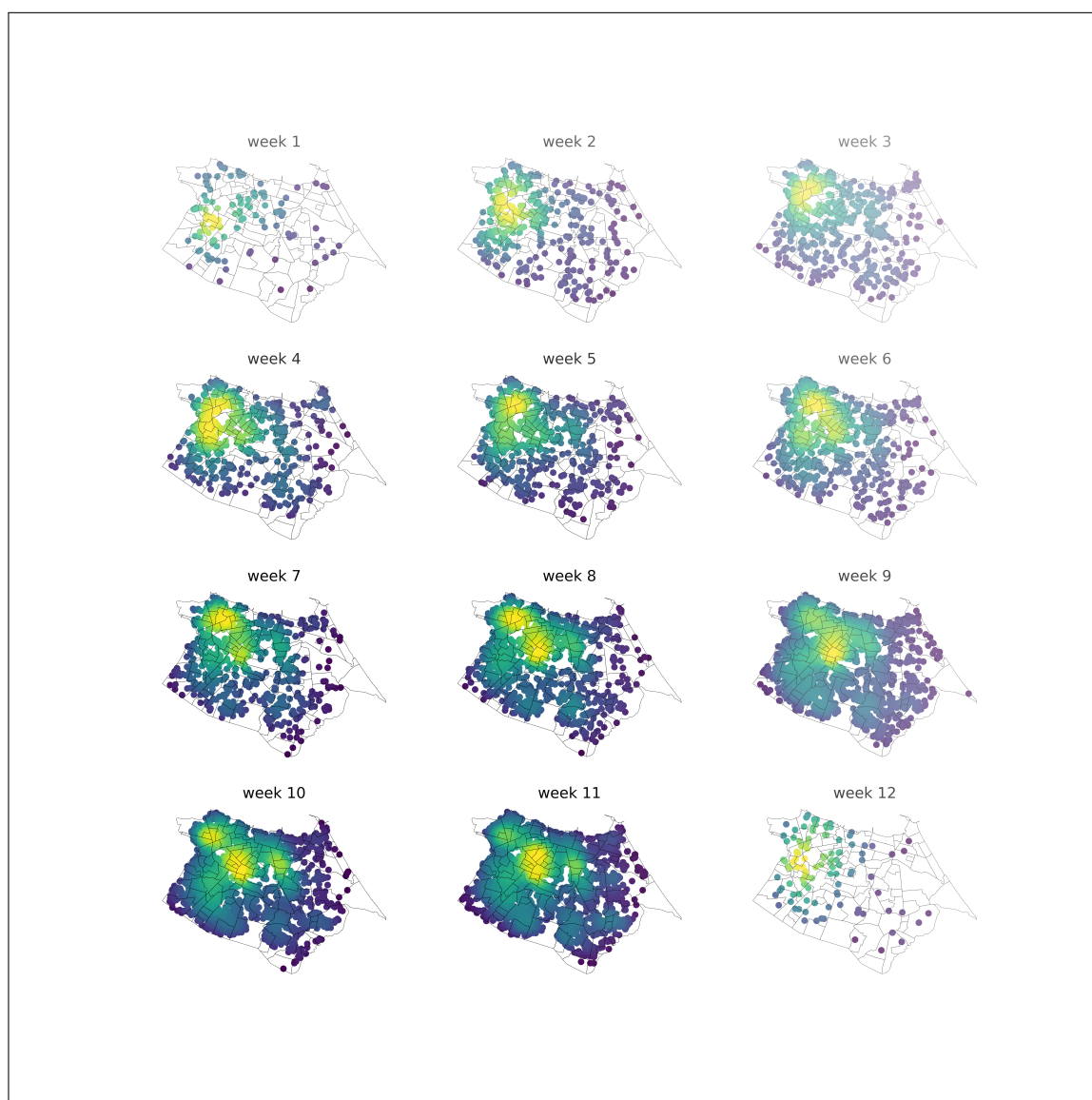
**Figura 42 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2011**



Fonte: Elaborado pelo autor

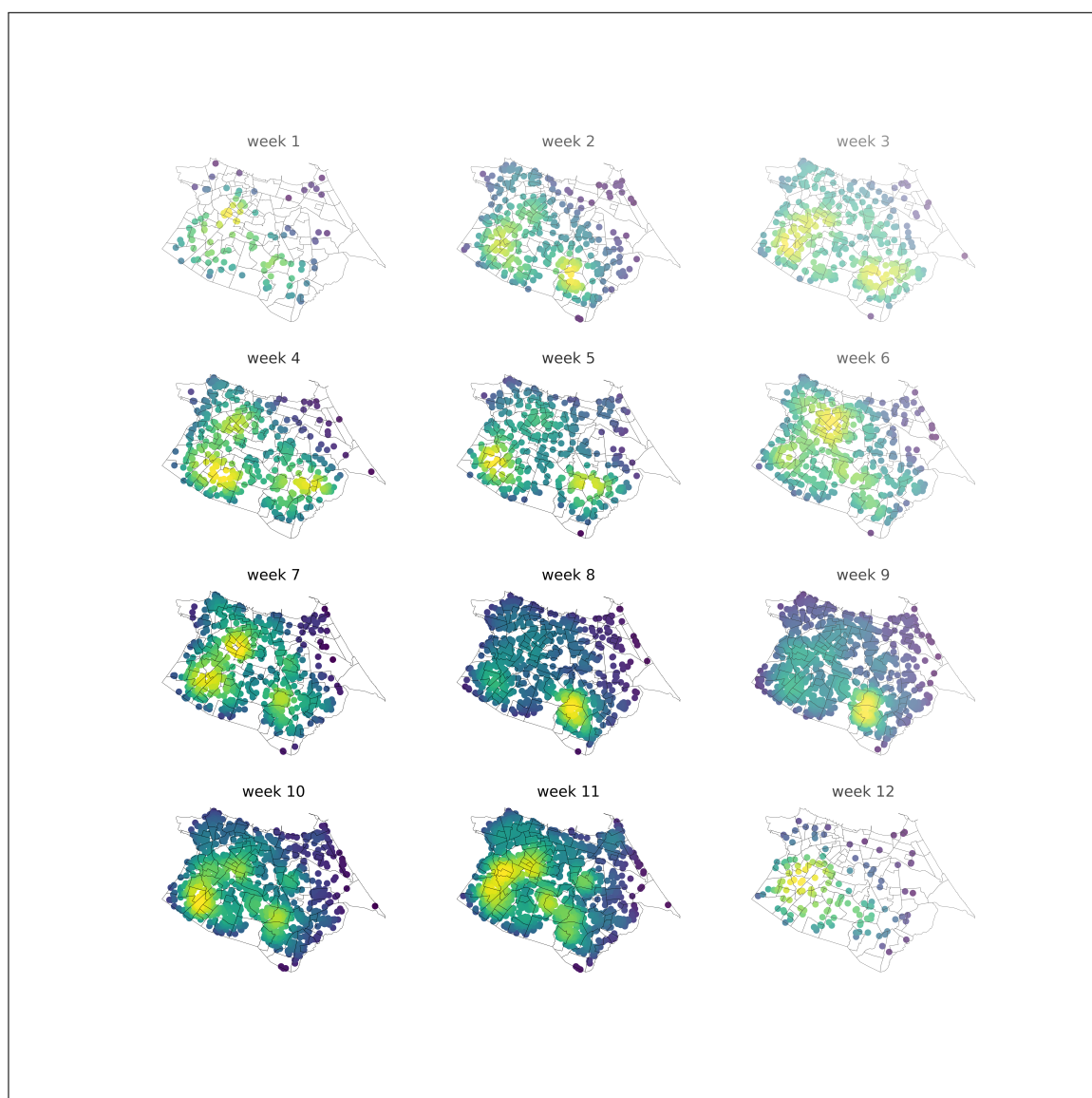


**Figura 43 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2012**



Fonte: Elaborado pelo autor

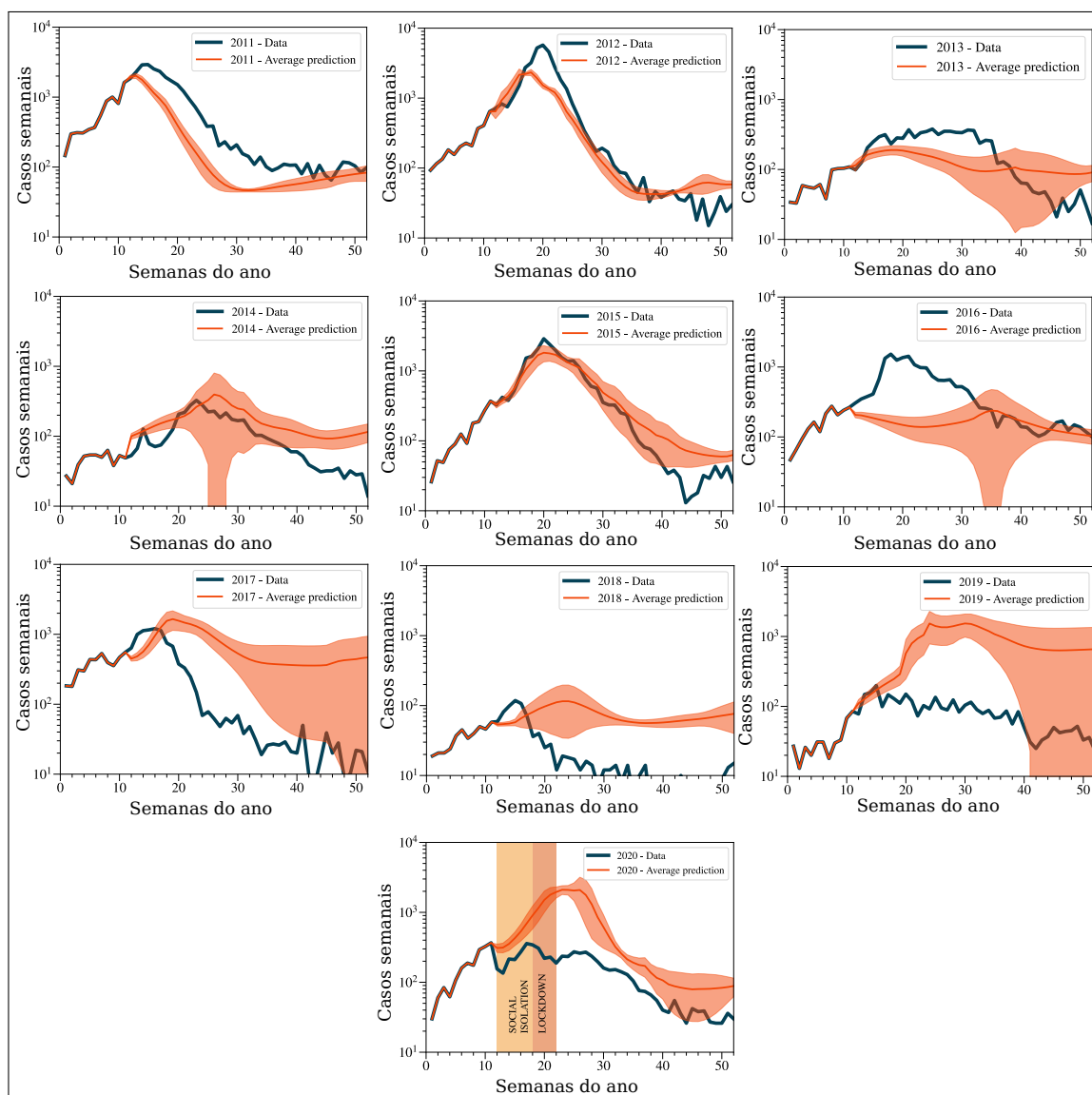
**Figura 44 – Evolução de casos de dengue da semana 1 a semana 12 para o ano de 2015**



Fonte: Elaborado pelo autor

## APÊNDICE E – PREDIÇÕES DE CASOS DE DENGUE PARA OS ANOS DE 2011 A 2020

**Figura 45 – Predições de casos de dengue a partir da semana 11 para os anos de 2011 a 2020**



Fonte: Elaborado pelo autor