



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

CLARISSE LINS DE LIMA

**MODELAGEM E SIMULAÇÃO DA TRANSMISSÃO DE ARBOVIROSES BASEADOS EM  
AGENTES INTELIGENTES PARA PREDIÇÃO DE CASOS DE DENGUE, *CHIKUNGUNYA*  
E ZIKA**

Recife

2020

CLARISSE LINS DE LIMA

**MODELAGEM E SIMULAÇÃO DA TRANSMISSÃO DE ARBOVIROSES BASEADOS EM AGENTES INTELIGENTES PARA PREDIÇÃO DE CASOS DE DENGUE, *CHIKUNGUNYA* E ZIKA**

Trabalho submetido ao Programa de Pós-Graduação em Engenharia Biomédica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestra em Engenharia Biomédica.

**Área de concentração:** Computação Biomédica.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos.

Recife

2020

Catálogo na fonte  
Bibliotecária Margareth Malta, CRB-4 / 1198

L732m Lima, Clarisse Lins de.  
Modelagem e simulação da transmissão de arboviroses baseados em agentes inteligentes para predição de casos de dengue, *chikungunya* e zika / Clarisse Lins de Lima - 2020.  
88 folhas, il., gráfs., tabs.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos.

Dissertação (Mestrado) – Universidade Federal de Pernambuco.  
CTG. Programa de Pós-Graduação em Engenharia Biomédica, 2020.

Inclui Referências e Índice.

1. Engenharia Biomédica. 2. Arboviroses. 3. Epidemiologia digital. 4. Aprendizagem de máquina. 5. Predição de arboviroses.  
I. Santos, Wellington Pinheiro dos (Orientador). II. Título.

UFPE

610.28 CDD (22. ed.)

BCTG/2021-98

CLARISSE LINS DE LIMA

**MODELAGEM E SIMULAÇÃO DA TRANSMISSÃO DE ARBOVIROSES BASEADOS EM AGENTES INTELIGENTES PARA PREDIÇÃO DE CASOS DE DENGUE, *CHIKUNGUNYA* E ZIKA**

Trabalho submetido ao Programa de Pós-Graduação em Engenharia Biomédica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestra em Engenharia Biomédica.

Aprovada em: 19/02/2020.

**BANCA EXAMINADORA**

---

Prof. Dr. Wellington Pinheiro dos Santos (Orientador)  
Universidade Federal de Pernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup> Cristine Martins Gomes de Gusmão (Examinador Interno)  
Universidade Federal de Pernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup> Ana Lúcia Bezerra Candeias (Examinador Externo)  
Universidade Federal de Pernambuco

À minha família, que sempre foi meu apoio e minha fortaleza.

## AGRADECIMENTOS

A Deus, por me guiar e me dar forças em todos os momentos da minha vida.

Ao meu orientador, o Prof. Dr. Wellington Pinheiro dos Santos, pela confiança, paciência, disponibilidade, dedicação e pelos ensinamentos durante este percurso da minha vida acadêmica.

Às componentes da banca examinadora, Prof<sup>a</sup>. Dr<sup>a</sup> Ana Lúcia e Prof<sup>a</sup>. Dr<sup>a</sup> Cristine Gusmão, pelas contribuições nesta dissertação.

Aos professores do Departamento de Engenharia Biomédica, por todo conhecimento compartilhado; e aos servidores, Rangel e Juliana, pela disponibilidade e apoio nos processos burocráticos.

Aos meus pais, Ana e Ricardo, por todo amor, carinho e incentivo ao longo da minha trajetória. Sem o apoio incondicional de vocês, não seria possível chegar tão longe.

Ao meu irmão, Artur, pelos seus ensinamentos diários por meio de nossas conversas, nossos momentos de diversão e por meio de seus questionamentos.

Aos meus amigos, Mamá, Manu, Gabi, Laila, Gabi Alves, Carol e Paulo, que me deram apoio e incentivo nas horas mais difíceis de desânimo e cansaço.

A todos e todas que, direta ou indiretamente, contribuíram para a realização deste trabalho, o meu sincero agradecimento.

*"Se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda"*  
(FREIRE, 2000)

## RESUMO

A prevenção e o controle das arboviroses têm sido um grande desafio de saúde pública para muitos países. No Brasil, especialmente a partir de 2015, a situação se agravou, uma vez que o vírus zika e o vírus chikungunya foram introduzidos no país. No intuito de auxiliar na prevenção de surtos de casos de arboviroses, vários trabalhos têm sido direcionados para o estudo de modelos de predição a partir de séries históricas das doenças e das variáveis climáticas, visto que fatores climáticos e ambientais interferem desde a eclosão dos ovos até a dispersão do mosquito. Apesar dos resultados positivos desse tipo de abordagem, a modelagem temporal não leva em consideração a heterogeneidade espacial na transmissão das arboviroses, o que se dá por meio de uma modelagem espaço-temporal. No entanto, esta abordagem tem sido pouco explorada devido à complexidade da estimativa dos parâmetros para a modelagem espaço-temporal e das variações inerentes ao processo epidemiológico. Adicionalmente, o avanço das tecnologias de Epidemiologia Digital e técnicas de geoprocessamento, juntamente ao desenvolvimento de técnicas de Mineração de Dados e Aprendizado de Máquina, podem proporcionar um rápido monitoramento, controle e simulação da propagação de doenças, auxiliando os sistemas de saúde pública no controle de epidemias e fatores ambientais e comportamentais que favorecem os vetores dessas doenças. Este trabalho tem, portanto, como objetivo construir uma ferramenta de predição da distribuição dos casos de arboviroses transmitidas pelo mosquito *Aedes aegypti* utilizando séries históricas de mapas de distribuição dos casos das doenças e das variáveis climáticas (pluviometria, temperatura e velocidade dos ventos). Para tanto, foram coletadas informações de interesse climático a partir da base de dados fornecida pelo Instituto Nacional de Meteorologia (INMET) e da Agência Pernambucana de Águas e Climas (APAC), bem como das informações de interesse epidemiológico e geográfico obtidas através do Portal de Dados Abertos da Cidade do Recife. Para a modelagem da distribuição dos casos das doenças foram testadas as técnicas de Regressão Linear, Máquina de Vetor de Suporte para regressão (SVMreg) e as Redes Neurais Artificiais (RNAs). Os resultados mostraram que a modelagem espaço-temporal utilizando técnicas de aprendizagem de máquina mostrou-se bastante eficaz na predição da distribuição dos casos de arboviroses na cidade do Recife, tornando-se relevante no apoio aos gestores e epidemiologistas da saúde no planejamento de ações de curto e médio prazo para prevenir surtos de arboviroses transmitidos pelo *Aedes aegypti*.

Palavras-chave: arboviroses, epidemiologia digital, aprendizagem de máquina, predição de arboviroses.

## ABSTRACT

The prevention and control of arboviruses has been a major public health challenge for many countries. In Brazil, especially since 2015, the situation has worsened, since the zika virus and the chikungunya virus were introduced in the country. In order to prevent outbreaks of cases of arboviruses, several studies have been directed to the study of prediction models from historical series of diseases and climatic variables, since climatic and environmental factors interfere from the hatching of eggs to the dispersion of the mosquito. Despite the positive results of this type of approach, the temporal modeling does not take into account the spatial heterogeneity in the transmission of arboviruses, which occurs through a space-temporal modeling. However, this type of approach has been little explored due to the complexity of estimating parameters for spatio-temporal modeling and the variations inherent to the epidemiological process. In addition, the advancement of Digital Epidemiology technologies and geoprocessing techniques, together with the development of Data Mining and Machine Learning techniques, can provide rapid monitoring, control and simulation of the spread of diseases, helping public health systems to control epidemics and environmental and behavioral factors that favor the vectors of these diseases. This work, therefore, aims to build a tool for predicting the distribution of cases of arboviruses transmitted by the *Aedes aegypti* mosquito using historical series of distribution maps of cases of diseases and climatic variables (pluviometry, temperature and wind speed). For this purpose, information of climatic interest was collected from the database provided by the National Institute of Meteorology (INMET) and the Pernambuco and Water and Climate Agency (APAC), as well as information of epidemiological and geographic interest obtained through the Information Portal. Recife Open Data. For modeling the distribution of disease cases, the techniques of Linear Regression, Support Vector Machine for regression (SVMreg) and also Artificial Neural Networks (ANNs) were tested. The results showed that space-time modeling using machine learning techniques proved to be quite effective in predicting the distribution of cases of arboviruses in the city of Recife, becoming relevant in supporting health managers and epidemiologists in planning short-term actions and medium term to prevent outbreaks of arboviruses transmitted from *Aedes aegypti*.

Keywords: arboviruses, digital epidemiology, machine learning, arbovirus prediction.

## LISTA DE FIGURAS

Figura 1 – Representação do sistema nervoso em diagrama de blocos. . . . .	29
Figura 2 – Representação do neurônio. . . . .	30
Figura 3 – Modelo do neurônio artificial proposto por McCulloch-Pits. . . . .	31
Figura 4 – Rede Neural Artificial de Camada Única. . . . .	32
Figura 5 – Rede Neural Artificial de Camada Múltipla. . . . .	32
Figura 6 – Redes recorrentes. . . . .	32
Figura 7 – Localização geográfica da Cidade do Recife, capital do Estado de Pernambuco, Nordeste do Brasil. . . . .	39
Figura 8 – Gráfico da série histórica da distribuição da temperatura na cidade do Recife ao longo do ano de 2013 correspondente à estação A301, localizada na latitude -8.05928 e longitude -34.959239. . . . .	40
Figura 9 – Gráfico da série histórica da distribuição da umidade relativa do ar na cidade do Recife ao longo do ano de 2013 correspondente à estação A301, localizada na latitude -8.05928 e longitude -34.959239. . . . .	40
Figura 10 – Distribuição da precipitação na cidade do Recife ao longo do ano de 2013 correspondente à estação A301, localizada na latitude -8.05928 e longitude -34.959239. . . . .	41
Figura 11 – Interface do mapa da rede de estações do Instituto Nacional de Meteorologia para o monitoramento das variáveis climáticas. . . . .	43
Figura 12 – Interface do banco de dados do SIGHPE para a seleção dos dados pluviométricos. . . . .	44
Figura 13 – Estação pluviométrica da Agência Pernambucana de Águas e Climas, localizada no Alto da Brasileira, Nova Descoberta, Recife-PE. . . . .	44
Figura 14 – Estação pluviométrica da Agência Pernambucana de Águas e Climas, localizada no bairro de Santo Amaro, Recife-PE. . . . .	45
Figura 15 – Estação pluviométrica da Agência Pernambucana de Águas e Climas, localizada no bairro da Várzea, Recife-PE. . . . .	45
Figura 16 – Fluxograma da realização da contagem dos casos confirmados, de cada tipo de arboviroses, por bairro, para cada bimestre de cada ano. . . . .	48
Figura 17 – Camada vetorial de pontos dos casos de arbovirose na cidade do Recife para o primeiro bimestre de 2014. . . . .	49
Figura 18 – Mapa de distribuição dos casos de arboviroses para a cidade do Recife para o primeiro bimestre de 2014. . . . .	49

Figura 19 – Resultados da predição da distribuição dos casos de arboviroses do primeiro bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 19a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 19b). . . . .	58
Figura 20 – Resultados da predição da distribuição dos casos de arboviroses do segundo bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 20a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 20b). . . . .	59
Figura 21 – Resultados da predição da distribuição dos casos de arboviroses do terceiro bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 21a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 21b). . . . .	60
Figura 22 – Resultados da predição da distribuição dos casos de arboviroses do quarto bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 22a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 22b). . . . .	61
Figura 23 – Resultados da predição da distribuição dos casos de arboviroses do quinto bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 23a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 23b). . . . .	62
Figura 24 – Resultados da predição da distribuição dos casos de arboviroses do sexto bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 24a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 24b). . . . .	63

Figura 25 – Resultados da predição da distribuição dos casos de arboviroses do primeiro bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 25a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 25b). . . . .	64
Figura 26 – Resultados da predição da distribuição dos casos de arboviroses do segundo bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 26a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 26b). . . . .	65
Figura 27 – Resultados da predição da distribuição dos casos de arboviroses do terceiro bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 27a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 27b). . . . .	66
Figura 28 – Resultados da predição da distribuição dos casos de arboviroses do quarto bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 28a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 28b). . . . .	67
Figura 29 – Resultados da predição da distribuição dos casos de arboviroses do quinto bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 29a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 29b). . . . .	68
Figura 30 – Resultados da predição da distribuição dos casos de arboviroses do primeiro bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 30a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com <i>kernel</i> RBF (Figura 30b). . . . .	69

Figura 31 – Resultados da predição da distribuição dos casos de arbovirozes do segundo bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arbovirozes utilizando MLP com 30 neurônios na última camada escondida (Figura 31a). À direita, tem-se o mapa da predição dos casos de arbovirozes utilizando SVM com <i>kernel</i> RBF (Figura 31b). . . . .	70
Figura 32 – Resultados da predição da distribuição dos casos de arbovirozes do terceiro bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arbovirozes utilizando MLP com 30 neurônios na última camada escondida (Figura 32a). À direita, tem-se o mapa da predição dos casos de arbovirozes utilizando SVM com <i>kernel</i> RBF (Figura 32b). . . . .	71
Figura 33 – Resultados da predição da distribuição dos casos de arbovirozes do quarto bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arbovirozes utilizando MLP com 30 neurônios na última camada escondida (Figura 33a). À direita, tem-se o mapa da predição dos casos de arbovirozes utilizando SVM com <i>kernel</i> RBF (Figura 33b). . . . .	72
Figura 34 – Resultados da predição da distribuição dos casos de arbovirozes do quinto bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arbovirozes utilizando MLP com 30 neurônios na última camada escondida (Figura 34a). À direita, tem-se o mapa da predição dos casos de arbovirozes utilizando SVM com <i>kernel</i> RBF (Figura 34b). . . . .	73
Figura 35 – Resultados da predição da distribuição dos casos de arbovirozes do sexto bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arbovirozes utilizando MLP com 30 neurônios na última camada escondida (Figura 34a). À direita, tem-se o mapa da predição dos casos de arbovirozes utilizando SVM com <i>kernel</i> RBF (Figura 34b).	74

## LISTA DE TABELAS

Tabela 1 – Resultados de coeficiente de correlação, erro quadrático relativo e tempo de treinamento para a Regressão Linear. . . . .	53
Tabela 2 – Resultados do coeficiente de correlação, erro quadrático médio e tempo de treinamento para o perceptron multicamadas, MLP, com 10, 20, 30 e 40 neurônios na única camada escondida. . . . .	54
Tabela 3 – Resultados do coeficiente de correlação, erro quadrático médio e tempo de treinamento para o SVM, com <i>kernel</i> s linear (ou grau 1), polinomial de graus 2 e 3, e RBF. . . . .	55
Tabela 4 – Resultados para a regressão linear e perceptron multicamadas com 30 neurônios na camada interna, considerando os conjuntos de treino com 4665 instâncias e os conjuntos de teste com 15553 instâncias. Como métricas de avaliação foram levadas em consideração o índice de correlação R e o erro RMSE%. Em vermelho, foram destacados os piores resultados. . . . .	56
Tabela 5 – Rótulos numéricos dos bairros do Recife. . . . .	57

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
1.1	Motivação e justificativa	16
1.2	Objetivos	19
1.3	Organização do trabalho	19
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>21</b>
2.1	Epidemiologia Digital e Geografia da Saúde	21
2.2	Arboviroses	22
2.2.1	Dengue	22
2.2.2	<i>Chikungunya</i>	23
2.2.3	Zika	26
2.3	Métodos de predição de arboviroses	27
2.4	Aprendizagem de máquina	28
2.4.1	Regressão linear	28
2.4.2	Redes Neurais Artificiais	29
2.4.3	Máquina de Vetores de Suporte	33
2.5	Considerações	33
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>34</b>
3.1	Estado da Arte	34
3.1.1	Abordagem temporal	34
3.1.2	Abordagem espaço-temporal	37
3.2	Considerações	38
<b>4</b>	<b>PROPOSTA</b>	<b>39</b>
4.1	Área sob estudo	39
4.2	Base de dados	40
4.2.1	Portal de Dados Abertos da Prefeitura de Recife	40
4.2.2	Banco de Dados Meteorológicos para Ensino e Pesquisa - BDMEP	41
4.2.3	Sistema de Geoinformação Hidrometeorológico de Pernambuco - SIGHPE	42
4.3	Softwares utilizados	44
4.3.1	Python	44
4.3.2	QGIS	45
4.3.3	Weka	46
4.4	Metodologia	46
4.4.1	Mapas de distribuição dos casos	46

4.4.2	Mapa de distribuição das variáveis climáticas . . . . .	47
4.4.3	Conjuntos de predição . . . . .	50
4.5	<b>Considerações</b> . . . . .	<b>51</b>
5	<b>RESULTADOS EXPERIMENTAIS</b> . . . . .	<b>52</b>
5.1	<b>Resultados quantitativos</b> . . . . .	<b>52</b>
5.2	<b>Resultados qualitativos</b> . . . . .	<b>56</b>
6	<b>CONCLUSÃO</b> . . . . .	<b>75</b>
6.1	<b>Conclusão geral</b> . . . . .	<b>75</b>
6.2	<b>Dificuldades apresentadas</b> . . . . .	<b>76</b>
6.3	<b>Contribuições</b> . . . . .	<b>76</b>
6.4	<b>Trabalhos futuros</b> . . . . .	<b>77</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>78</b>
	<b>Índice</b> . . . . .	<b>87</b>

## 1 INTRODUÇÃO

Este capítulo introdutório à Dissertação de Mestrado apresenta o contexto, a motivação e a justificativa para a realização desta pesquisa. Além disso, neste capítulo, também são apresentados o objetivo geral e objetivos específicos deste trabalho.

### 1.1 Motivação e justificativa

A prevenção e o controle da dengue, da *chikungunya* e da zika têm sido um grande desafio de saúde pública para muitos países. Estima-se que, só de dengue, cerca de 390 milhões de novos casos ocorram a cada ano. Entretanto, problemas como diagnósticos incorretos e registros imprecisos ou ausência do registro dos casos, em muitas regiões, podem contribuir para que o impacto da dengue e de outras arboviroses transmitidas pelo mosquito *Aedes aegypti* seja subestimado (LIMA et al., 2016; BHATT et al., 2013). No Brasil, onde a dengue já é endêmica, a partir de 2015, outras arboviroses passaram a interagir com o *Aedes aegypti*, seu principal vetor transmissor. Essa situação passa por um agravamento devido à rápida difusão da *chikungunya*, provocando febre e fraqueza muscular, dentre outros sintomas, e ao surgimento do vírus zika que está parcialmente relacionado com casos de microcefalia em recém-nascidos e diretamente relacionado à ocorrência da síndrome de Guillain-Barret, doença autoimune que afeta o sistema nervoso, provocando desde fraqueza muscular até a paralisia (CAO-LORMEAU et al., 2016). Ou seja, a emergência de outras arboviroses, como a *chikungunya* e a zika, em especial na América do Sul, levanta novos desafios ao monitoramento e controle do vetor.

A dengue é uma infecção viral transmitida aos humanos por meio de mosquitos, e está se espalhando rapidamente pelo mundo. Seu vetor primário é o mosquito *Aedes aegypti*, uma espécie bem adaptada às áreas urbanas e distribuída, principalmente, em regiões tropicais e subtropicais, mas também com atuação na América do Norte e na Europa. Evidências indicam que um vetor secundário, o mosquito *Aedes albopictus*, também vem expandindo sua área de atuação geográfica (LIMA et al., 2016; BHATT et al., 2013). O risco de surtos de arboviroses e de sua presença endêmica é maior em regiões tropicais e subtropicais, porém pode ser identificado um aumento no registro de casos na América do Norte e na Europa, devido à presença dos mosquitos *Aedes* e da introdução dos vírus nessas regiões (LIMA et al., 2016; BHATT et al., 2013).

A transmissão das arboviroses é um processo complexo que envolve a interação de múltiplos agentes: populações humanas, mosquitos e vírus condicionados por fatores climáticos e ambientais em um espaço muito heterogêneo. O espaço em que essas interações ocorrem é complexo o suficiente para que o estudo da transmissão das arboviroses

seja carregado de desafios. As pandemias de arboviroses têm sido favorecidas por uma combinação de diversos fatores: a mobilidade global das populações humanas e circulação de mosquitos; o inchaço das áreas urbanas superpovoadas; a dificuldade de acesso das populações urbanas, especialmente dos setores menos favorecidos economicamente, ao saneamento básico, ao fornecimento regular de água, e ao sistema de saúde pública; fatores ambientais e climáticos, como a temperatura e o índice pluviométrico, que mede a densidade e a ocorrência das chuvas; e, por fim, a ineficiência de estratégias de controle do vetor (LIMA et al., 2016; GUBLER, 2011; MOHAMMED; CHADEE, 2011).

Diversos grupos de pesquisa têm se dedicado a construir mapas de risco e estimar a distribuição global das arboviroses e sua correlação com dados ambientais. Apesar da importância desses esforços de mapear a distribuição dessas doenças, também é importante entender a dinâmica das arboviroses em uma escala local, o que se dá por meio de modelos matemáticos e computacionais (PADMANABHAN; SESHAIYER; CASTILLO-CHAVEZ, 2017; JINDAL; RAO, 2017; LIMA et al., 2016; KRAEMER et al., 2015; PAUL et al., 2018; PHUNG et al., 2015). Condições climáticas locais, como temperatura, pluviometria e umidade, interferem no desenvolvimento do vetor, desde a eclosão dos ovos até o tempo de vida dos mosquitos e sua dispersão, além de outros aspectos da transmissão das arboviroses (LIMA et al., 2016; GUBLER, 2011). O avanço da Epidemiologia Digital e das tecnologias de geoprocessamento, aliados ao desenvolvimento das técnicas de Mineração de Dados e Aprendizado de Máquina, têm proporcionado o rápido acompanhamento, controle e simulação da disseminação de doenças - auxiliando os sistemas públicos de saúde no controle de epidemias - e dos fatores ambientais e comportamentais que favorecem os vetores dessas doenças (SALATHE et al., 2012; ALLEN et al., 2016; BAKKER et al., 2016; GUO et al., 2017; SALATHÈ, 2018; THIÉBAUT; THIESSARD, 2017).

Uma das abordagens usadas para o combate de doenças é a predição do número de casos da doença baseada no comportamento de eventos passados. No caso das arboviroses, além da série histórica do número de casos da doença, também são utilizadas séries históricas de variáveis climáticas para a predição de comportamentos futuros. Os modelos mais comuns são os que se baseiam no modelo de média móvel integrada autorregressiva sazonal (SARIMA), no modelo linear generalizado (GAM), no algoritmo da árvore de regressão com gradiente de impulso (GBM), no modelo de regressão binomial negativo (NBM) e no modelo de regressão linear com mínimo operador absoluto de encolhimento e seleção (LASSO). Porém, técnicas de aprendizagem de máquina como, por exemplo, as redes neurais artificiais e a máquina de vetores de suporte também têm sido aplicadas na modelagem baseada na análise temporal dos casos (GUO et al., 2017; SIRIYASATIEN et al., 2018; BAQUERO; SANTANA; CHIARAVALLOTI-NETO, 2018).

Apesar dos bons resultados trazidos pela modelagem temporal, esses modelos não descrevem a variabilidade entre populações (vetor e hospedeiro), principalmente, a

heterogeneidade espacial e temporal (LIMA et al., 2016). A análise espacial é uma técnica que busca descrever os padrões existentes nos dados espaciais e estabelecer, de forma quantitativa, os relacionamentos entre as diferentes variáveis geográficas (SILVA et al., 2015). Poucos estudos exploram a predição de casos de arboviroses baseada na modelagem espaço-temporal devido à complexidade da estimativa dos parâmetros da modelagem espaço-temporal, pois variações inerentes ao processo epidemiológico são difíceis de analisar (Ana Carolina C. Costa et al., 2015; MARTÍNEZ-BELLO; LÓPEZ-QUÍLER; PRIETO, 2018).

No Brasil, as arboviroses têm recebido uma atenção especial do Sistema Único de Saúde por meio de políticas públicas de saúde e campanhas educativas. Agentes públicos diversos, como a Secretaria Municipal de Saúde do Recife, a Secretaria Estadual de Saúde de Pernambuco, e o Ministério da Saúde, têm disponibilizado dados geográficos abertos de infecção pelo mosquito *Aedes aegypti* desde 2013, por meio do sistema LIRAA, que mapeia a distribuição geográfica do Índice de Infestação Predial (PESSANHA et al., 2009). A Secretaria Municipal de Saúde do Recife, em parceria com a Fundação Oswaldo Cruz, busca controlar a difusão da dengue e, mais recentemente, da *chikungunya* e da zika, por meio do uso de ovitrampas, armadilhas utilizadas tanto para contar quanto para destruir os ovos do mosquito. Adicionalmente, a Secretaria Municipal de Saúde do Recife, por meio do seu Portal de Dados Abertos, distribui o mapeamento das doenças e dos sintomas por unidade de saúde e pelo bairro de origem do paciente desde 2015 (RECIFE, 2018). A Agência Pernambucana de Águas e Clima (APAC) também disponibiliza um sistema de informações geográficas no qual são publicizadas as séries temporais pluviométricas diárias e mensais desde o ano de 2006 (APAC, 2018).

Fazendo uso da mineração dos dados de sistemas abertos, combinando com o uso de algoritmos de aprendizagem de máquina, é possível inferir regras e expressões que podem servir para alimentar simuladores de epidemias. Esses simuladores podem auxiliar os gestores de saúde em situações críticas em que sejam necessárias as tomadas de decisões rápidas, principalmente, em situações nas quais há deficiência tanto de recursos financeiros quanto de recursos humanos (FERNANDES; FILHO, 2019).

Este trabalho é orientado pela seguinte pergunta: como construir modelos e preditores de casos de arboviroses eficientes, capazes de fazer predições com baixo erro, alimentados com bases de dados de informações geográficas de diversas fontes e que possam ser utilizados como apoio aos gestores de saúde no planejamento, vigilância e combate ao vetor?

Portanto, neste trabalho, é proposta uma metodologia para prever a distribuição dos casos de arbovírus transmitidos pelo mosquito *Aedes aegypti* através da implementação de simuladores de transmissão da dengue, zika e *chikungunya* baseados em sistemas multiagentes, utilizando uma abordagem espaço-temporal. Para tanto, foram coletadas

informações de interesse climático a partir da base de dados fornecida pelo Instituto Nacional de Meteorologia (INMET) e da Agência Pernambucana de Águas e Climas (APAC), bem como das informações de interesse epidemiológico e geográfico obtidas através do Portal de Dados Abertos da Prefeitura de Recife. Algoritmos de aprendizagem de máquina tais como regressão linear, redes neurais e máquinas de vetores de suporte foram empregados para criar um modelo espaço-temporal, a fim de avaliar a distribuição de casos de arboviroses na cidade do Recife.

## 1.2 Objetivos

Este trabalho teve como objetivo construir um sistema preditor da disseminação das arboviroses transmitidas pelo vetor *Aedes aegypti*, a saber, a dengue, a febre chikungunya e a zika, utilizando séries históricas georreferenciadas dos casos de arboviroses e das variáveis climáticas como temperatura e velocidade dos ventos. Para isso, foram obtidas informações dos casos de arboviroses, por meio do Portal de Dados Abertos do Recife, e as informações climáticas através Instituto Nacional de Meteorologia e pela Agência Pernambucana de Águas Climas. A partir dessas informações, foram gerados mapas de distribuição, os quais foram agrupados de modo que para a predição da distribuição espacial da doença em um bimestre são utilizados dados de um ciclo de 6 bimestres anteriores. Como objetivos específicos, têm-se:

- Elaborar uma metodologia para completar as bases de dados estimando os valores faltantes;
- Elaborar e organizar as bases de dados de predição em séries históricas de mapas para a elaboração das bases de predição;
- Investigar técnicas de aprendizado de máquina para a construção de um preditor dinâmico dos casos de arboviroses para modelagem da dinâmica populacional humana na área sob estudo, no tempo e no espaço;
- Validar os modelos criados utilizando informações geográficas, climáticas e de casos confirmados na cidade do Recife, de 2013 a 2016;
- Gerar os mapas de predição da distribuição dos casos de arboviroses para a melhor e a pior técnica avaliada.

## 1.3 Organização do trabalho

Esta dissertação está estruturada da seguinte forma:

No capítulo 2, são apresentados os conceitos sobre arboviroses, epidemiologia, epidemiologia digital e geoprocessamento. Também são apresentados os métodos de

predição de arboviroses, bem como uma breve apresentação das técnicas de aprendizagem de máquina.

No capítulo 3, é apresentado o estado da arte dos métodos de predição utilizados, das bases de dados utilizadas, bem como as técnicas computacionais existentes. Neste capítulo também é realizada uma análise comparativa do estado da arte com a metodologia proposta no presente trabalho.

O capítulo 4 apresenta a metodologia adotada durante o estudo. Neste capítulo, também são detalhadas as bases de dados que foram utilizadas e as limitações da pesquisa.

No capítulo 5, são apresentados os resultados quantitativos de cada modelo avaliado e os resultados qualitativos dos modelos que apresentaram os melhores e os piores resultados. Por fim, no capítulo 6, são apresentadas as conclusões a respeito do trabalho, as dificuldades apresentadas, as contribuições e trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo apresentar a fundamentação teórica de princípios de epidemiologia digital, os conceitos básicos das arboviroses transmitidas pelo mosquito *Aedes aegypti* e os métodos de predição de arboviroses. Também são apresentadas algumas técnicas de aprendizagem de máquina: a regressão linear, as redes neurais artificiais e a máquina de vetores de suporte.

### 2.1 Epidemiologia Digital e Geografia da Saúde

A Epidemiologia pode ser definida como estudo da distribuição das doenças e lesões nas populações humanas. Ou seja, a epidemiologia tem como objetivo compreender os padrões das doenças, o que causa esses padrões e, a partir dessas informações, auxiliar na prevenção de doenças e promoção de saúde (SANTANA, 2014; SALATHÈ, 2018). Os avanços da na área da computação têm possibilitado a criação de modelos matemáticos complexos, que podem ser construídos para obtenção de respostas que ainda estão em aberto na epidemiologia e controle de doenças. Com o aumento do grande volume de dados, devido à internet e o grande uso de celulares, bem como o avanço de técnicas computacionais e dos métodos de análise de dados, surge um novo tipo de Epidemiologia conhecida como Epidemiologia Digital (PARK et al., 2018; ECKHOFF; TATEM, 2015). A Epidemiologia Digital é definida, portanto, como a epidemiologia que utiliza métodos digitais para a coleta e análise dos dados (PARK et al., 2018).

A epidemiologia sempre esteve extremamente relacionada com a geografia como um núcleo central do conhecimento em Saúde Pública (GUIMARÃES, 2015). A Geografia da Saúde no Brasil busca compreender a relação entre as condições socioeconômicas e ambientais com a saúde da população (SANTANA, 2014). O uso de Sistemas de Informações Geográficas (SIG) com as técnicas de geoprocessamento permitem a integração de várias camadas que representam dados físicos e sociais em uma plataforma para análise espacial (DOM et al., 2017).

Define-se geoprocessamento como o conjunto de tecnologias de coleta de dados por meios de técnicas matemáticas e computacionais para a realização do tratamento da informação geográfica (NARDI et al., 2013; NETO et al., 2014). O geoprocessamento proporciona a identificação, localização, acompanhamento e o monitoramento de populações. O processamento de informações georreferenciadas possibilitam a elaboração de mapas temáticos em formatos digitais para a representação do fenômeno em estudo (NARDI et al., 2013).

Na área da saúde, o geoprocessamento permite a análise da distribuição espacial

dos casos de uma determinada doença, em uma determinada região. Isto é importante pois, a partir da distribuição espacial, é possível realizar estudos a respeito da transmissibilidade da doença na região de abrangência, fornecendo assim, informações necessárias para realizar a vigilância dos casos. Os mapas temáticos possibilitam a identificação de regiões onde a população é mais atingida, o que é essencial para a promoção de políticas públicas preventivas, distribuição de medicamentos e realocação de recursos humanos e físicos (NARDI et al., 2013; NETO et al., 2014).

## 2.2 Arboviroses

As arboviroses (de *arbovirus*, abreviação para *arthropode-borne virus*) são doenças transmitidas por vírus que se mantêm na natureza por meio de um hospedeiro vertebrado e um artrópode hematófago como, por exemplo, um mosquito (MUSSO; GUBLER, 2016). A maioria dos arbovírus causam zoonoses que, geralmente, necessitam de um hospedeiro não-humano para manutenção da espécie. Os humanos muitas vezes funcionam como segundo hospedeiro para a finalização do ciclo do vírus. Porém, alguns arbovírus como, por exemplo, o da zika, o da dengue e o da *chikungunya*, adaptaram-se de tal forma aos humanos que são capazes de se manter na natureza por meio do ciclo mosquito-humano-mosquito (MUSSO; GUBLER, 2016). Neste ciclo, os vírus tornam-se independente do hospedeiro primata não-humano.

O vetor transmissor de uma arbovirose é um artrópode que transmite o vírus de um vertebrado para outro através de uma picada. A transmissão biológica comumente ocorre quando o artrópode hematófago se alimenta de um vertebrado virêmico e deposita a saliva infecciosa durante a alimentação do sangue de outro vertebrado (MUSSO; GUBLER, 2016). No entanto, outras formas de transmissão já foram reportadas tais como: a transmissão de mãe para filho durante a gravidez, a transmissão durante uma transfusão de sangue, a transmissão por meio da relação sexual (MUSSO; GUBLER, 2016). Nas seções 2.2.1, 2.2.2 e 2.2.3 serão apresentadas as arboviroses transmitidas pelo *Ae. aegypti*.

### 2.2.1 Dengue

O vírus da dengue, DENV, é um vírus pertencente ao gênero *Flavivirus* e à família *Flaviviridae*. Atualmente, são conhecidos quatro sorotipos distintos: DEN-1, DEN-2, DEN-3 e DEN-4 (MARTÍNEZ-BELLO; LÓPEZ-QUÍLER; PRIETO, 2018). Geralmente, quando um indivíduo é infectado por um sorotipo, ganha imunidade vitalícia para aquele sorotipo, porém, para os outros sorotipos esta imunidade é apenas temporária. Após a picada do mosquito, o vírus passa por um período de incubação de 5 a 6 dias. Passado o período de incubação, as manifestações clínicas começam a aparecer: febre alta, dor de cabeça, dores musculares e articulares, dor nos olhos, vômito, náusea, perda de apetite (SIRIYASATIEN et al., 2018). Caso os sintomas agravem-se, pode haver vazamento de plaqueta, causando

falha dos órgãos, o que pode ser fatal tanto para adultos quanto para crianças. Até a presente data, não existe nem vacina e nem drogas disponíveis para o tratamento da doença (SIRIYASATIEN et al., 2018).

O vírus da dengue foi isolado pela primeira no ano de 1950, muito embora a sua existência já era conhecida há mais de 20 anos antes de seu isolamento. Desde o ano de 1970, as epidemias de dengue afetaram mais de 128 países ao redor do mundo e possuem grandes possibilidades de se espalharem em novas regiões (SIRIYASATIEN et al., 2018). Portanto, a dengue é um grande desafio nacional e internacional para as autoridades de saúde nas tomadas de ações efetivas na prevenção da doença (SIRIYASATIEN et al., 2018).

### 2.2.2 *Chikungunya*

O vírus *chikungunya*, CHIKV (do inglês, *chikungunya virus*), é um alfavírus originário na África, pertencente à família *Togaviridae* e do complexo antigênico *Semliki forest*, que também é composto pelos alfavírus Ross River, O'nyong-nyong, Getah, Bebaru, *Semliki forest* e Mayaro (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015). Atualmente, são encontradas quatro linhagens do vírus: Oeste Africano, Leste-Centro-Sul-Africano (ESCSA), Asiático e Oceano Índico (IOL) (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015).

O nome *chikungunya* é derivado do Makonde, um dos idiomas falados no sudoeste da Tanzânia, que pode ser traduzido como "curvar-se ou torna-se contorcido". Este nome refere-se à posição do paciente devido às graves dores articulares causadas pelo CHIKV (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015). Após a infecção pelo CHIKV, o vírus passa por um período de incubação, que pode durar de 2 a 7 dias (a média é de 12 dias), antes do indivíduo apresentar os sintomas da doença (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015; POWERS; LOGUE, 2007). Durante a fase aguda (que pode durar 10 dias), há um início súbito dos sintomas, caracterizado por febre alta, exantema e artralgia - a qual se apresenta, geralmente, de forma simétrica e em articulações periféricas como tornozelos, punhos, joelhos e falanges (GOUPIL; MORES, 2016; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015; PIALOUX et al., 2007; POWERS; LOGUE, 2007). Durante a fase crônica da doença, há persistência da poliartralgia, que pode afetar por meses ou até mesmo anos. Este quadro clínico clássico tem sido descrito durante as epidemias mais recentes em cerca de até 95% dos pacientes. Apenas de 3,0% a 5,0% das pessoas infectadas não apresentam sintomas (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015).

Apesar de, na grande maioria dos casos, a infecção apresentar um curso benigno, no curso das epidemias foram observadas manifestações atípicas da doença. O vírus *chikungunya* é capaz de se reproduzir em diversos tipos de tecidos corporais, desde o tegumento até o sistema nervoso central. Também pode afetar o músculo cardíaco, articulações e fígado; podendo resultar em uma grande variabilidade de manifestações clínicas. Os relatos de manifestações atípicas da doença incluem: alterações neurológicas,

cardíacas, renais e oculares - mais incidente em indivíduos com mais de 65 anos e com doenças preexistentes - podendo apresentar complicações e resultar em óbitos. No caso das complicações neurológicas, estas se apresentam de forma similar tanto em crianças quanto em adultos e compreendem: a mielo-neuropatia, encefalites, Síndrome de Guillain-Barré, paralisia flácida e as neuropatias. As manifestações oculares mostram-se, com mais frequência, como iridociclite e a retinite, na maioria dos casos, com resolução e preservação da visão após cerca de seis a oito semanas. Nas manifestações cardíacas, observa-se quadros de miocardite, pericardite e miocardiopatia dilatada (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015).

Os registros das epidemias de *chikungunya* têm ocorrido ainda antes do século XX, no entanto, como não havia conhecimento a respeito do vírus, foram documentadas como epidemias de dengue pelo fato de ambas as doenças apresentarem sintomas bastante semelhantes (YACTAYO et al., 2016; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015). A primeira documentação do isolamento do CHIKV ocorreu no ano de 1953, no que hoje corresponde ao território da Tanzânia, durante uma epidemia de uma doença com características semelhantes à dengue, hoje conhecida como *chikungunya* (HONÓRIO et al., 2015; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015; POWERS; LOGUE, 2007). No ano de 1958, o CHIKV foi isolado pela primeira vez no continente asiático, em Bangkok, na Tailândia, com o vírus provavelmente originário da África. Essa trajetória do vírus foi analisada como consequência de um segundo surto de CHIKV no ano de 1958 na África do Sul (MORRISON, 2014)

Durante a década de 60 até a década de 90, o vírus foi isolado diversas vezes na África e os surtos esporádicos de *chikungunya* ficaram limitados apenas aos países deste continente (Tanzânia, Uganda, Zimbábue, África do Sul, Senegal, Nigéria, República da África Central e República Democrática do Congo) e no sudoeste do continente asiático, compreendendo a Tailândia, Cambodja, Vietnam, Laos, Malásia, Myanmar, Filipinas e Indonésia (SCHUFFENECKER et al., 2006; DIOP; MESENZNIKOV; SANICAS, 2015; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015). Na África, acredita-se que o vírus era mantido por meio de ciclos silvestres envolvendo primatas não-humanos e mosquitos silvestres da família *Aedes*. As espécies mais comuns do ciclo silvestre são: *Ae. furcifer*, *Ae. taylori*, *Ae. luteocephalus*, *Ae. africanus* e *Ae. neoafricanus*. Por outro lado, na Ásia, o vetor transmissor mais comum era o *Ae. aegypti*, que são mosquitos urbanos, peri-domésticos e antropofílicos (PIALOUX et al., 2007).

O agravamento da situação ocorreu a partir do ano de 2005, quando o vírus foi detectado no sudoeste do Oceano Índico, mais precisamente na ilha Comoros, introduzido, provavelmente, através de viajantes virêmicos vindos do Quênia, onde havia sido registrado um surto em junho de 2004. O vírus rapidamente se dispersou pelas ilhas do Oceano Índico, dentre elas: Comores, Madagascar, Mayotte, Ilhas Seychelles, Maurícias, Ilhas da

Reunião (BEENSOON et al., 2008; JOSSERAN et al., 2008; SCHUFFENECKER et al., 2006; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015). Até o final do ano de 2005, a Índia reportou casos da doença em inúmeras províncias, onde o número de casos suspeitos chegou a mais de 1,3 milhões de pessoas. O CHIKV continuou a se dispersar, causando grandes surtos no Sri Lanka e em vários países do sudoeste asiático (MORRISON, 2014; YERGOLKAR et al., 2006). Acredita-se que a epidemia de Ilhas da Reunião, em 2005, esteve associada a uma mutação do vírus - descoberta por Schuffenecker et al. (2006). Esta mutação pode ter favorecido a adaptação do vírus à espécie *Aedes albopictus* - o qual possui características zoofílicas e antropofílicas e consegue se adaptar a ambientes naturais e urbanos (PIALOUX et al., 2007; POWERS; LOGUE, 2007) - que está presente em países de zona temperada (REZZA et al., 2007).

Durante a epidemia de 2005-2006, o vírus foi introduzido a países não endêmicos por meio de indivíduos virêmicos, e a transmissão autóctone foi observada pela primeira vez na Itália (BELTRAME et al., 2007; REZZA et al., 2007; BONILAURI et al., 2008), Nova Caledônia, Papua Nova Guiné, Butão, Iêmen (MORRISON, 2014) e França (GRAMDADAM et al., 2011). O CHIKV chegou ao Caribe em outubro de 2013, onde foi identificada a transmissão autóctone da doença na Guiana Francesa, São Martinho, Martinica, Guadalupe, República Dominicana, São Bartolomeu e Ilhas Virgens Britânicas. Em janeiro do ano seguinte, foram identificados casos em países da área continental da América Central.

No Brasil, os casos importados, começaram a aparecer a partir de junho de 2014 e a transmissão autóctone do vírus foi detectada em setembro de 2014, no Oiapoque, Amapá (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015; HONÓRIO et al., 2015). Naquele mesmo ano, foram confirmados 2.772 casos de CHIKV distribuídos em seis estados: Amapá (1.554 casos), Bahia (1.214 casos), Distrito Federal (2), Mato Grosso do Sul (1), Roraima (1) e Goiás (1) (HONÓRIO et al., 2015).

O Brasil é uma região bastante propícia para grandes epidemias de *chikungunya*, devido a vários fatores dos quais podemos citar: (1) o fato de o país ser amplamente infestado pelos dois vetores transmissores; (2) a circulação tanto do DENV quanto do CHIKV - o que dificulta o diagnóstico e abordagem terapêutica; (3) a possibilidade da adaptação do CHIKV ao *Ae. albopictus* (como foi observado em outros países); (4) os sintomas aparecem com maior frequência - ao contrário da dengue que pode ser assintomática; (5) o período de viremia é maior, durando até 8 dias após o início da febre; (6) os seres humanos estão mais suscetíveis à doença - o que favorece sua disseminação; (7) a presença de um grande número de primatas e culicídeos que nunca foram expostos ao CHIKV - podendo assim possibilitar o estabelecimento do ciclo silvestre observado até então, apenas, na África; por fim, a grande extensão territorial do Brasil, o que torna difícil a vigilância e o acesso dos serviços de saúde a testes laboratoriais (HONÓRIO et al., 2015).

### 2.2.3 Zika

O vírus zika (ZIKV), é um arbovírus que pertence à família *Flaviridae* e ao gênero *Flavivirus*, transmitido, principalmente, pela picada de mosquitos do gênero *Aedes* (JUNIOR et al., 2015). No Brasil, o principal vetor transmissor é o mosquito da espécie *Aedes aegypti* (BRITO et al., 2016; MELLO et al., 2008; PORTELA et al., 2009). No entanto, existem evidências da possibilidade de outras formas de transmissão desse arbovírus. Já foram descritas transmissões por via sexual, de transfusão de sangue e, também, de mãe para filho - como foi observado em outros tipos de zoonoses causadas por flavivírus (como, por exemplo, o vírus dengue) e alfavirus (como, por exemplo, o vírus *chikungunya*) (JUNIOR et al., 2015; MUSSO; GUBLER, 2016).

O primeiro isolamento do ZIKV foi em abril de 1940, na Floresta de Zika, Uganda, a partir da amostra sorológica de um macaco *Rhesus*. No ano de 1954, o zika foi isolado novamente, desta vez, a partir de amostra sorológica de uma criança nigeriana. Durante a década de 60, na Nigéria, foram identificados indivíduos com sorologia positiva para o ZIKV e em indivíduos com quadros febris durante uma epidemia de febre amarela (JUNIOR et al., 2015). No ano de 1969, o ZIKV foi isolado pela primeira vez fora da África - em mosquitos da espécie *Aedes aegypti* - na Malásia. As primeiras infecções na Ásia foram reportadas no Java, na Indonésia, no ano de 1977 (MUSSO; GUBLER, 2016).

No entanto, o primeiro surto de zika foi documentado pela primeira vez no ano de 2007, na Federação dos Estados da Micronésia, mais especificamente, na Ilha de Yap, configurando-se o primeiro foco fora do continente africano e asiático. Estima-se que cerca de 5005 habitantes da ilha foram infectados pelo ZIKV (cerca de 72.6% da população) (JUNIOR et al., 2015; MUSSO; GUBLER, 2016). O vírus zika espalhou-se rapidamente pelas ilhas do Oceano Pacífico e, em outubro de 2013, os primeiros casos da doença foram reportados na Polinésia Francesa. A situação, nesta região, evoluiu para uma situação de epidemia, na qual houve 19.000 casos suspeitos e 289 casos confirmados. No ano de 2014, o vírus chegou nas Américas, onde foram reportados casos da doença na Ilha de Páscoa - pertencente ao Chile - , provavelmente relacionado com os surtos na Micronésia e Polinésia Francesa (JUNIOR et al., 2015). No ano seguinte, a circulação do ZIKV em território brasileiro foi detectada na Região Nordeste. O surto da doença afetou milhares de pessoas em toda a região, cujos estados mais afetados foram: Bahia, Maranhão, Pernambuco, Paraíba, Sergipe e Rio Grande do Norte (MUSSO; GUBLER, 2016; CAMPOS; BANDEIRA; SARDI, 2016; VARGAS et al., 2016).

Após a picada do mosquito infectado, o vírus passa por um período de incubação de aproximadamente 12 dias. No geral, os sintomas são descritos como febre baixa - muito embora no Brasil, houve relatos de febre alta -, mialgia, e artralgia. O quadro de artralgia apresenta-se com menor intensidade do que nos casos de *chikungunya* e acometem as mãos, tornozelos e joelhos. Há, também, relatos de conjuntivite, porém, sem secreção

purulenta; vertigem, náusea, vômito e dor retro-orbital (JUNIOR et al., 2015).

Apesar da doença apresentar uma evolução benigna, há evidências que o ZIKV possa estar associado ao aumento dos casos da Síndrome de Guillain-Barret (tanto no surto da Polinésia, quanto no surto no Brasil), a qual ocorre poucos dias após a infecção. Esta síndrome é caracterizada por evolução ascendente de fraqueza dos membros, hipo ou arreflexia e dissociação celulo-proteica no líquido cefalorraquidiano (JUNIOR et al., 2015; MUSSO; GUBLER, 2016; NÓBREGA et al., 2018). Durante a epidemia do ZIKV no Brasil, levantou-se a suspeita da relação com os casos de microcefalia com o ZIKV após o Estado de Pernambuco relatar o aumento do número de casos. A microcefalia é um déficit no crescimento cerebral que causa deficiência intelectual, paralisia cerebral, epilepsia, dificuldade de deglutição, anomalias no sistema visual e auditivo, além de distúrbios do desenvolvimento, como autismo e TDAH (MUSSO; GUBLER, 2016; EICKMANN et al., 2016). Avaliando o histórico das epidemias de zika, foi observado que durante o surto ocorrido na Polinésia Francesa, também houve casos de má formação do sistema nervoso central, incluindo casos de microcefalia, em recém-nascidos (MUSSO; GUBLER, 2016).

### 2.3 Métodos de predição de arboviroses

Na elaboração de modelos de predição, o conjunto de dados é um elemento importante para que resultados obtidos sejam relevantes e utilizáveis. Além disso, no processo de predição, o modelo é construído a partir de observações históricas de um determinado evento. Sendo assim, faz-se necessário uma grande quantidade de dados para a elaboração do modelo de predição. O volume de dados deve ser o suficiente de modo que seja possível tanto a elaboração do modelo a partir do conjunto de treinamento, quanto a predição de novos eventos baseados no que foi observado historicamente (SIRIYASATIEN et al., 2018). Sendo assim, a quantidade de dados disponíveis é tão importante quanto a sua confiabilidade. Geralmente, os dados utilizados para predição são obtidos através de instituições governamentais como, por exemplo, Ministério da Saúde, Ministério da Agricultura, entre outros (SIRIYASATIEN et al., 2018). No entanto, novas fontes de dados vêm sendo exploradas para a criação de novos modelos de predição. Na era da informação, a internet tem sido responsável pela geração e pela disponibilização de um grande volume de dados de interesse epidemiológico. Esse grande volume de dados é o que chamamos de *Big Data* (SIRIYASATIEN et al., 2018).

As fontes dos dados podem ser classificadas em convencionais e modernas. Os dados convencionais incluem informações médicas e epidemiológicas obtidas através de centros de saúde (como hospitais); dados climáticos e ambientais obtidos através de departamentos meteorológicos; e dados demográficos e geográficos, obtidos através de fontes governamentais (MARQUES-TOLEDO et al., 2017; MORSY et al., 2018). Os dados modernos, por sua vez, são obtidos de informações disponíveis em redes sociais como

Facebook, Twitter, Myspace; através de busca na internet (SIRIYASATIEN et al., 2018; LI et al., 2017). Nas regiões onde a internet não está bem difundida, foram utilizados registros de chamadas telefônicas, direcionadas ao departamento de monitoramento de surtos e de ocorrência de casos da doença, para obtenção dos dados da predição (SIRIYASATIEN et al., 2018).

Após a aquisição dos dados, eles são pré-processados antes da aplicação de um modelo de predição. O pré-processamento é importante para a redução de ruídos e aumentar a acurácia e consistência dos dados. Caso o conjunto de dados tenha uma grande quantidade de dados incompletos ou incorretos, a eficácia do modelo preditor pode ser comprometida (SIRIYASATIEN et al., 2018).

Para os problemas de predição de arboviroses, existem dois tipos de abordagem: a abordagem estatística e a modelagem dinâmica. No caso da abordagem estatística, os métodos mais comuns são: ARIMA, regressão multivariada de Poisson e SARIMA. No entanto, esses métodos não permitem a compreensão das características dinâmicas da transmissão das doenças (CAO et al., 2018). Os trabalhos recentes de predição de arboviroses têm se concentrado em modelos populacionais, isto é, a modelagem da transmissão de doenças entre indivíduos em uma população. Para isto, são aplicadas técnicas de aprendizagem de máquina para a criação de um modelo que será utilizado na predição. Os algoritmos de aprendizagem de máquina mais utilizados são as redes neurais artificiais e a máquina de vetores de suporte (CAO et al., 2018; RUBIO-SOLIS et al., 2019; KOSTKOVA; SANTOS; MASSONI, 2019; MUSAH et al., 2019; BELTRÁN et al., 2018).

## **2.4 Aprendizagem de máquina**

A aprendizagem de máquina é um campo de estudo na qual são utilizadas técnicas computacionais para que a máquina seja capaz de identificar e de descrever padrões estruturais em um conjunto de dados (WITTEN; FRANK, 2005). O tipo do aprendizado pode ser classificado em duas categorias: supervisionado e não-supervisionado (SIRIYASATIEN et al., 2018). No aprendizado supervisionado, a figura do especialista fornecerá um rótulo como resposta esperada para o padrão. No aprendizado não-supervisionado, a figura do especialista não é necessária e o próprio algoritmo é responsável por agrupar, em diferentes categorias, de acordo com os dados de entrada (SIRIYASATIEN et al., 2018). Nas seções 2.4.1, 2.4.2 e 2.4.3 são apresentadas as técnicas de aprendizagem de máquina que foram exploradas neste trabalho.

### **2.4.1 Regressão linear**

A regressão linear é um método de análise estatística para avaliar a correlação entre duas ou mais variáveis. Neste método, assume-se que os dados possuem um comportamento linear. A ideia da regressão linear é expressar a classe como uma combinação linear

dos atributos com seus pesos predeterminados (WITTEN; FRANK, 2005), como mostra a Equação 2.1

$$y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (2.1)$$

em que  $y$ , representa o valor esperado da predição;  $x_1, x_2, \dots, x_n$  representam os valores dos atributos e  $w_0, w_1, w_2, \dots, w_n$  representam os pesos dos atributos, os quais são calculados pelos dados de treinamento. No método da regressão linear, o algoritmo calcula os pesos (coeficientes da equação) em busca da melhor representação dos dados, de modo a minimizar a soma dos quadrados da diferença de todos os dados de treinamento. A soma dos quadrados da diferença é dada pela Equação 2.2, cujo valor dentro dos parênteses corresponde à diferença entre o valor real do se deseja prever e o valor previsto pelo modelo (WITTEN; FRANK, 2005).

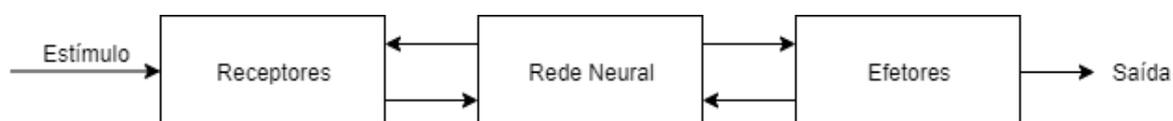
$$S = \sum_{i=1}^n (y^{(i)} - \sum_{j=0}^k w_j x_j^{(i)})^2 \quad (2.2)$$

#### 2.4.2 Redes Neurais Artificiais

A rede neural artificial (RNA), que é comumente chamada de rede neural, consiste em uma técnica computacional baseada no comportamento do cérebro humano (SIRIYASATIEN et al., 2018). O cérebro humano é capaz de organizar seus menores constituintes, os neurônios, para realizar diferentes tarefas como, por exemplo, reconhecimento de padrão e controle motor. Redes neurais artificiais têm sido utilizadas com sucesso em diversas aplicações biomédicas e de uso geral (LIMA; SILVA-FILHO; SANTOS, 2016; COMMOWICK et al., 2018; SANTOS; SOUZA; FILHO, 2007; SANTOS et al., 2008; SANTOS et al., 2009b; AZEVEDO et al., 2015; SANTOS et al., 2009a; CRUZ; CRUZ; SANTOS, 2018; CORDEIRO et al., 2012; VASCONCELOS; SANTOS; LIMA, 2018; SANTOS; FILHO; SANTOS, 2019; RODRIGUES et al., 2019; FREITAS et al., 2019; RODRIGUES et al., 2019; JUNIOR et al., 2019).

O sistema nervoso pode ser visto como um sistema de três estágios. O cérebro representa a figura central do sistema e está sempre recebendo estímulos do meio exterior, processando e dando respostas apropriadas a estes estímulos. Os receptores são responsáveis por converter os estímulos externos ao corpo humano em corrente elétrica e transmitir a informação para a rede neural (ver Figura 1). Os efetores são responsáveis por converter os impulsos elétricos em respostas, representando a saída do sistema (HAYKIN, 2001).

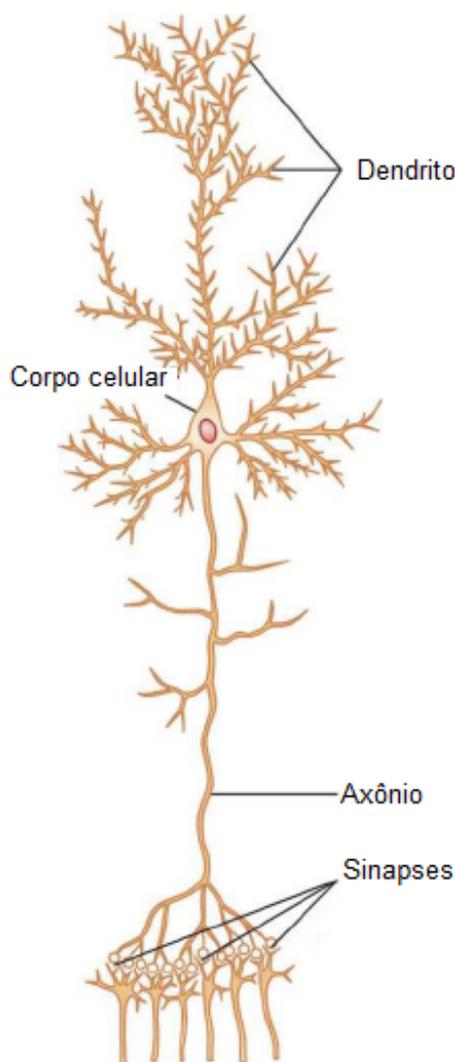
Figura 1 – Representação do sistema nervoso em diagrama de blocos.



Fonte: Haykin (2001)

O neurônio é constituído pelo axônio que é responsável por transmitir os sinais para outras células; os dendritos, que são extensões ramificadas responsáveis por receber os sinais químicos de outros neurônios e o corpo celular, no qual está localizado o núcleo do neurônio e suas organelas, conforme Figura 2 (HALL, 2011).

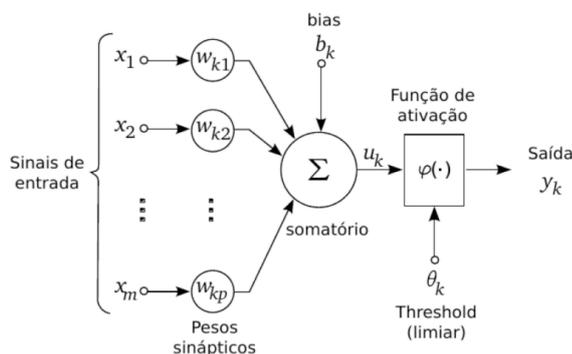
Figura 2 – Representação do neurônio.



Fonte: Hall (2011)

Um elemento importante no funcionamento do cérebro são as sinapses. As sinapses são os pontos de contato entre um neurônio e o neurônio seguinte, ou seja, elas mediam a interação entre neurônios (HAYKIN, 2001; HALL, 2011). O tipo mais comum de sinapse é a química. O processo pré-sináptico libera uma substância, o neurotransmissor, que vai se difundir na junção sináptica entre os neurônios podendo desencadear o potencial de ação do neurônio seguinte (excitando-o) ou não (inibindo-o) (HALL, 2011). As redes neurais artificiais são, portanto, inspiradas neste mecanismo, porém são bastante primitivas quando comparadas as redes neurais encontradas no cérebro (HAYKIN, 2001).

Figura 3 – Modelo do neurônio artificial proposto por McCulloch-Pits.



Fonte: Haykin (2001)

As RNAs são constituídas por unidades menores, os neurônios artificiais, que são fundamentais para a operação da rede (HAYKIN, 2001). Os elementos de um neurônio artificial são: um conjunto de *sinapses* ou *conectores* que são caracterizadas por um peso  $w$  - em que um sinal  $x_i$  na entrada da sinapse  $j$  conectada ao neurônio  $k$  é multiplicado pelo peso sináptico  $w_{k,j}$  -; um *somador* para adicionar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio; e uma função de ativação para limitar a saída de um neurônio (HAYKIN, 2001). A Figura 3 ilustra o modelo de um neurônio artificial.

Matematicamente, um neurônio é representado por:

$$u_k = \sum_{j=1}^m w_{k,j} x_i \quad (2.3)$$

e pela equação

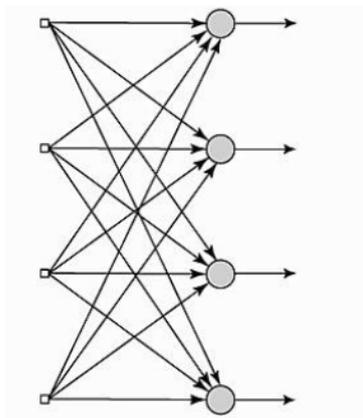
$$y_k = \varphi(u_k + b_k) \quad (2.4)$$

em que  $x_1, x_2, \dots, x_n$  representam os sinais de entrada;  $w_0, w_1, w_2, \dots, w_n$  representam os pesos sinápticos dos sinais de entrada;  $b_k$ , é o termo *bias* e  $\varphi$  é função de ativação do neurônio.

Quanto a sua arquitetura, a rede neural pode ser classificada como:

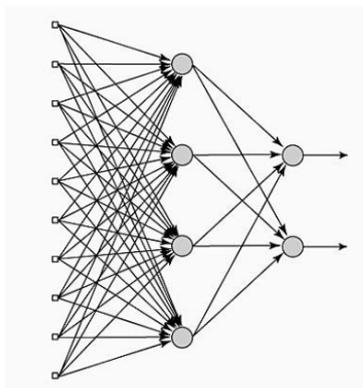
- Rede Neural de Camada Única: nesta arquitetura de rede, o neurônio possui uma única camada de entrada que projeta o sinal para uma camada de saída (Figura 4);
- Rede Neural Multicamadas: nesta configuração, a rede possui uma ou mais de uma camada escondida, na qual os neurônios têm como entrada os valores de saída do sinal anterior (Figura 5);
- Rede recorrente: na rede recorrente, cada neurônio é alimentado pelo sinal de saída de cada neurônio da rede (Figura 6).

Figura 4 – Rede Neural Artificial de Camada Única.



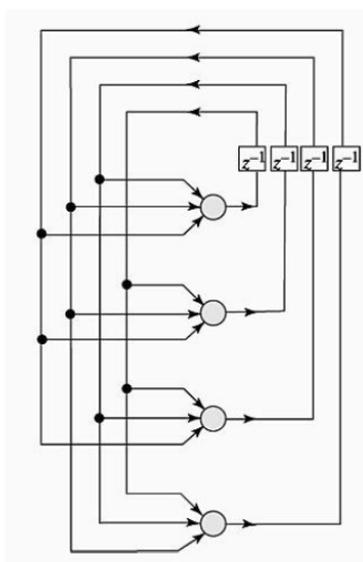
Fonte: Haykin (2001)

Figura 5 – Rede Neural Artificial de Camada Múltipla.



Fonte: Haykin (2001)

Figura 6 – Redes recorrentes.



Fonte: Haykin (2001)

As redes neurais têm sido bastante utilizadas na predição de doenças. Baquero, Santana e Chiaravalloti-Neto (2018) utilizaram as redes neurais artificiais para a predição de casos de dengue na cidade de São Paulo, Brasil. Laureano-Rosario et al. (2018), também fizeram uso das redes neurais para a predição de surtos de dengue na costa do nordeste de Yucatán, México, e em San Juan, Porto Rico. No trabalho de Akil e Ahmad (2016), as redes neurais foram utilizadas para a modelagem de casos de infecção pela bactéria *Salmonella* no estado do Mississippi, EUA. Além disso, as redes neurais têm sido aplicadas para diagnóstico de infarto do miocárdio, embolia pulmonar, hemorragia gastrointestinal, entre outros (AKIL; AHMAD, 2016). Sendo assim, as RNAs estão bastante difundidas na área da saúde.

### 2.4.3 Máquina de Vetores de Suporte

A máquina de vetores de suporte (SVM, do inglês *support vector machine*) é uma técnica de aprendizagem de máquina supervisionado para análise de dados e reconhecimento de padrão utilizada para classificação e regressão (SHARMA et al., 2015). A ideia do SVM para regressão é encontrar o melhor hiperplano  $H$  que represente os dados  $x \in \mathbb{R}^n$  como um problema linearmente separável, os quais são mapeados através de uma função matemática  $\varphi$  (também conhecida como *kernel*) - sendo esta uma função polinomial, sigmoideal, ou gaussiana - e uma regressão linear é aplicada no hiperplano correspondente (LIU et al., 2019).

A expressão matemática que representa esta ideia é dada pela equação 2.5, onde  $w$  representa o vetor da regressão linear do hiperplano.

$$\begin{cases} (x) = w \cdot \varphi(x) + b, \\ \varphi : x \rightarrow H \end{cases} \quad (2.5)$$

## 2.5 Considerações

Neste capítulo, foi apresentada a fundamentação teórica deste trabalho. Foram mostrados os conceitos básicos a respeito da Epidemiologia e Geografia da Saúde. Também foram explorados os aspectos gerais da dengue, zika e *chikungunya*; doenças transmitidas pelo *Ae. aegypti*, bem como os métodos mais utilizados na predição de arboviroses. Além disso, foram apresentadas, brevemente, as técnicas de aprendizagem de máquina utilizadas nesta pesquisa.

### 3 TRABALHOS RELACIONADOS

Este capítulo tem como objetivo apresentar os trabalhos relacionados a esta dissertação e que foram publicados nos últimos cinco anos. A revisão é dividida entre os trabalhos que utilizaram uma abordagem temporal para a modelagem dos casos de arboviroses e os trabalhos que utilizaram uma abordagem espaço-temporal.

#### 3.1 Estado da Arte

##### 3.1.1 Abordagem temporal

O aumento de ameaças de surtos de arboviroses e sua dispersão em várias regiões do mundo é a principal razão dos pesquisadores buscarem desenvolver ferramentas eficazes para a predição de doenças (SIRIYASATIEN et al., 2018).

Baquero, Santana e Chiaravalloti-Neto (2018) realizaram um estudo comparativo de predição dos casos de dengue na cidade de São Paulo, Brasil, para avaliar o desempenho de modelos de média móvel integrada autorregressiva sazonal (SARIMA), modelos aditivos generalizados (GAM) e redes neurais artificiais (RNAs). Os dados referentes aos casos de dengue foram obtidos pelo Centro de Vigilância Epidemiológica de São Paulo. Os casos suspeitos, mas com resultados laboratoriais negativos, ou casos sem resultados laboratoriais e com manifestações clínicas compatíveis com outra doença, foram excluídos do modelo. Os dados brutos utilizados foram referentes aos meses de janeiro a abril entre os anos de 2000 e 2016. As variáveis climáticas utilizadas no modelo foram obtidas no Instituto Nacional de Meteorologia. Foram utilizadas séries mensais, de janeiro a abril de 2000 a 2016, da temperatura (mínima, média e máxima), precipitação e umidade relativa. No pré-processamento dos dados, os valores faltantes de temperatura média e máxima, foram estimados por interpolação linear. Para o conjunto de treinamento, Baquero, Santana e Chiaravalloti-Neto (2018) utilizaram séries mensais entre janeiro de 2000 a dezembro de 2014, enquanto que o restante foi utilizado para teste. Como métrica de avaliação Baquero, Santana e Chiaravalloti-Neto (2018) utilizaram o RMSE (*root mean square error*) e uma medida de desempenho relativo (razão entre o RMSE do modelo e o RMSE do naïve model). Dentre os modelos avaliados por Baquero, Santana e Chiaravalloti-Neto (2018), o que obteve melhor resultado foi o GAM, com RMSE de 2152. Nos métodos avaliados por Baquero, Santana e Chiaravalloti-Neto (2018), as informações referentes ao posicionamento geográfico não foram consideradas, sendo assim, não retornam uma distribuição espacial dos casos da doença.

Atualmente, novas fontes de dados têm sido utilizadas para realizar a predição de casos de dengue e muitas dessas informações têm sido buscadas em redes sociais

(SIRIYASATIEN et al., 2018). No trabalho de Marques-Toledo et al. (2017), os autores investigaram os *tweets*, com indicação pessoal do conteúdo da dengue, como fonte de dados para a criação de um modelo preciso, para a detecção e o monitoramento precoces das epidemias da dengue nos níveis nacional e local. Os dados referentes aos casos da doença foram obtidos em mais de uma fonte. Os registros oficiais dos casos de dengue foram obtidos no Sistema de Informação de Agravos de Notificação do Ministério da Saúde do Brasil, enquanto que os indicativos dos casos de dengue foram obtidos por meio do Twitter, Google Trends e os acessos aos artigos da Wikipédia. Os dados sociodemográficos foram obtidos por meio do censo de 2010 realizado pelo Instituto Nacional de Geografia e Estatística. Foram usados os seguintes índices: população, PIB per capita, IDH (Índice de Desenvolvimento Humano), índice de desenvolvimento humano de renda (IDHM-Renda), índice de desenvolvimento humano de educação (IDHM-Educação) e índice de desenvolvimento humano de longevidade (IDHM-Longevidade). Também se avaliou o percentual de casas com acesso à internet e computadores. No total foram avaliadas 238 cidades brasileiras com população acima dos 40.000 habitantes.

Para avaliar a correlação entre os casos oficiais de dengue e os dados da web - a nível nacional -, Marques-Toledo et al. (2017) aplicaram uma regressão linear. Os resultados obtidos mostraram que os dados do Google Trends obtiveram uma maior associação com os casos oficiais de dengue ( $r = 0.92$ ,  $p < 0.001$ ), seguido pelo Twitter ( $r = 0.87$ ,  $p < 0.001$ ) e, com menor correlação, o Wikipédia ( $r = 0.71$ ,  $p < 0,01$ ). Apesar da alta correlação dos dados do Google Trends, apenas os dados do Twitter foram utilizados para a predição de casos a nível de cidade, por ser a única fonte que possibilita a obtenção da localização geográfica. Na predição dos casos, Marques-Toledo et al. (2017) utilizaram modelos aditivos generalizados (GAM), nos quais 170 semanas dos dados foram estabelecidos como conjunto de treinamento e o restante (39 semanas) foi utilizado para validação. As métricas de avaliação foram o desvio explicado,  $R^2$  e erro médio relativo. Das 238 cidades, o método utilizado por Marques-Toledo et al. (2017) ajustou um modelo para 199, pois algumas cidades possuíam poucos dados para a convergência da estimativa do modelo. As cidades utilizadas para validação do modelo foram: Belo Horizonte ( $r = 0.93$ ,  $r^2 = 90.3$ ), Fortaleza ( $r = 0.41$ ,  $r^2 = 90.0$ ), Manaus ( $r = 0.78$ ,  $r^2 = 83.5$ ), Porto Alegre ( $r = 0.71$ ,  $r^2 = 76.1$ ), Rio de Janeiro ( $r = 0.80$ ,  $r^2 = 82.6$ ) e São Paulo ( $r = 0.47$ ,  $r^2 = 89.0$ ). A pesquisa mostrou que o Twitter pode ser uma boa fonte de dados para a vigilância em tempo real, para regiões em que a dengue é endêmica, principalmente, em lugares nos quais os serviços de vigilância em saúde sejam fracos ou lentos na notificação dos casos. Um dos limites desse método é o fato de que, de acordo com os achados dos autores, os melhores modelos foram gerados nas cidades com o maior número populacional e melhores índices socioeconômicos com acesso à internet e computador. Ou seja, o modelo pode subestimar os casos da doença nas populações mais pobres, visto que para esse recorte populacional, o acesso a essas ferramentas é limitado. Um outro obstáculo está relacionado à distribuição espacial dos casos, pois, apesar do

Twitter fornecer a localização geográfica, esta não foi utilizada no modelo de predição.

Morsy et al. (2018) estudaram um modelo de predição de casos de zika no Brasil e na Colômbia. Os casos oficiais da doença tanto para o Brasil quanto na Colômbia foram obtidos através da Organização de Saúde Pan Americana. No caso do Brasil, os dados foram obtidos de janeiro de 2016 até julho de 2016. Para a Colômbia, foram obtidos dados de 9 de agosto de 2015 a 21 de maio de 2016. Os dados oriundos da internet foram obtidos através do Google Trends por meio da busca da palavra "Zika". Os métodos de predição adotados foram: a regressão múltipla e o modelo de média móvel integrado autorregressivo sazonal. Os resultados obtidos por Morsy et al. (2018) mostraram que a consulta na web pode ser um método eficaz na predição de casos da doença. No entanto, as informações geográficas não foram consideradas na predição dos casos, portanto, o método proposto não retorna uma distribuição espacial local aproximada dos casos.

Guo et al. (2017) propuseram cinco modelos de predição de casos da dengue na China: o modelo baseado na máquina de vetores de suporte para regressão (SVR), modelo de regressão linear *step-down*, algoritmo de árvore de regressão reforçada com gradiente (GBM, do inglês *gradient boosted regression tree algorithm*), o modelo da regressão binomial negativa (NBM, do inglês, *negative binomial regression model*), modelo de regressão linear LASSO (um acrônimo para *least absolute shrinkage and selection operator*) e modelo aditivo generalizado (GAM, do inglês *generalized additive model*). Os casos semanais oficiais de dengue de Guangdong foram obtidos através de fontes governamentais da China. Os dados meteorológicos foram coletados no Sistema de Serviço de Compartilhamento de Dados Meteorológicos da China, com informações da temperatura média, umidade relativa do ar e pluviometria. Também foram coletados dados de vigilância de consultas de pesquisas referentes a dengue - de forma semelhante a Marques-Toledo et al. (2017) - no site Baidu. Para o treinamento dos modelos foram utilizados os dados da primeira semana de 2011 até a quadragésima primeira semana de 2014. Os dados das semanas restantes foram utilizados para a validação. Os modelos foram avaliados através do cálculo do RMSE. De acordo com os resultados de Guo et al. (2017), o melhor modelo de regressão foi aquele baseado na máquina de vetores de suporte. Assim como as abordagens de Baquero, Santana e Chiaravalloti-Neto (2018) e Marques-Toledo et al. (2017), as informações geográficas não foram consideradas na predição dos casos, portanto, o método proposto não retorna uma distribuição espacial aproximada dos casos, o que pode dificultar o planejamento local de medidas preventivas.

A dinâmica populacional já foi identificada como um dos fatores importantes para a dinâmica de transmissão dos casos de dengue. Viagens para regiões endêmicas podem levar a infecções individuais e, quando esses indivíduos infectados retornam ao seu local de origem, podem contribuir na transmissão local (SANG et al., 2015). Levando em consideração este cenário, Sang et al. (2015) estudaram um modelo de predição levando

em consideração os pacientes infectados fora da região de estudo. A área de estudo selecionada foi a área de Guangzhou, China. Os dados referentes aos casos importados de dengue, entre os anos de 2006 a 2014, foram obtidos pelo Sistema de Vigilância de Doenças Notificáveis na China. Os dados meteorológicos entre março de 2006 a setembro de 2014. O preditor foi modelado utilizando-se funções *spline* cúbicas suaves, com 3 graus de liberdade. Como resultado, os autores obtiveram um coeficiente de determinação de 0,98 e um desvio explicado de 95,4%. Assim como os métodos de predição utilizados por outros autores, Sang et al. (2015) não utilizaram informações geográficas para a elaboração de um modelo de predição. Desta forma, o modelo proposto não retorna informações a respeito da distribuição espacial dos casos.

### 3.1.2 Abordagem espaço-temporal

A disseminação espacial das arboviroses é regida por complexas interações dinâmicas entre humanos, mosquitos e os diferentes sorotipos do vírus (LIMA et al., 2016). A análise espacial das arboviroses têm como principal objetivo mensurar propriedades e relacionamentos levando-se em consideração a localização geográfica do fenômeno em estudo (SILVA et al., 2015).

Neste sentido, Albrieu-Llinás et al. (2018) focaram em examinar a evolução temporal do número de casas infestadas com estágios imaturos do *Aedes aegypti* em cada bairro de Clorinda, Argentina, e explorar se as diferentes características poderiam ser atribuídas a áreas que compartilham as mesmas condições ambientais. As pesquisas entomológicas em Clorinda foram realizadas entre outubro de 2011 a novembro de 2013, em que se registrou o número de casas com criadouros do mosquito. Um total de 10.981 residências foram avaliadas aleatoriamente. Os dados foram organizados por bairro, por tempo (2011-2012, 2012-2013) e de acordo com as estações do ano. Os dados ambientais foram obtidos por meio de uma imagem de satélite SPOT 5, nas quais foram obtidas as seguintes variáveis de cobertura da terra: solo descoberto, águas superficiais, zonas úmidas, vegetação baixa (grama), vegetação alta (árvores e arbustos), edifícios urbanos e pastagens ou culturas. Essas variáveis foram submetidas a um particionamento utilizando o algoritmo de k-médias e em seguida, um modelo geral de regressão linear foi ajustado para poder avaliar a contribuição de cada agrupamento ambiental para explicar a mudança nos locais de reprodução do *Aedes aegypti*. Os resultados de Albrieu-Llinás et al. (2018) mostraram que o número de casas infestadas pelo *Ae. aegypti* é afetado por uma interação entre os grupos ambientais e o ano da amostra. Um agrupamento ambiental simples e robusto pode permitir a identificação fácil e rápida das áreas mais problemáticas da cidade. O método de Albrieu-Llinás et al. (2018) não mostra a predição de novos focos.

Tariq e Zaidi (2019) utilizaram a regressão geograficamente ponderada para modelar os casos de dengue em Lahore, um distrito da província de Punjab, Paquistão. Os casos da

doença foram obtidos por meio do Ministério da Saúde e da Autoridade de Gerenciamento de Desastre de Punjab. Além dos dados dos casos da doença, foram utilizados Índice de Vegetação da Diferença Normalizada (NDVI), Índice de Diferença Normalizada da Água (NDWI), Temperatura da Superfície Terrestre (LST) - máxima e média -, em graus Celsius; obtidos por imagens de satélite SPOT-5, área construída (metros quadrados); obtidos pela Unidade Urbana de Lahore, densidade populacional e população (em milhares) no nível do Conselho Sindical (UC); obtidos do relatório de pré-investimento do governo de Punjab. Para modelagem dos dados, os autores utilizaram a regressão dos mínimos quadrados ordinários (OLS) e a regressão geográfica ponderada (GWR) do ArcGIS. O modelo que representou melhores resultados estatísticos foi o que utilizou áreas de construção e densidade populacional com  $R^2 = 0.805$ . Os mapas da distribuição espacial da predição da incidência de dengue também foram apresentados estudo. Por outro lado, os autores não exploraram outros métodos de regressão. Os trabalhos de Ren et al. (2017) e Silva et al. (2015) também utilizaram a modelagem baseada na regressão geográfica ponderada (GWR) para estudar a influência do espaço na distribuição dos casos de arboviroses. No entanto, não são utilizados para realizar predições dos casos.

Até o momento, a modelagem espaço-temporal para a predição da distribuição dos casos de arboviroses ainda é pouco explorada. Também não foram encontrados trabalhos que utilizem as redes neurais artificiais e as máquinas de vetores de suporte para a modelagem espaço-temporal para a predição de casos das arboviroses transmitidas pelo *Aedes aegypti*. Isto condiz com o levantamento do estado da arte realizado por Siriyasatien et al. (2018), em 2018, que grande parte das pesquisas envolvendo predição de arboviroses, adota apenas a modelagem temporal para a predição dos casos.

### 3.2 Considerações

Neste capítulo, foram apresentadas as diferentes propostas de modelagem dos casos de arboviroses utilizando a abordagem temporal e a abordagem espaço-temporal. No caso da abordagem temporal, destacam-se os trabalhos focados em criar modelos de predição baseados em modelos estatísticos e em algoritmos de aprendizagem de máquina. Na abordagem espaço-temporal, foram apresentados os trabalhos que relacionam os aspectos espaciais com a dinâmica da doença. Um outro ponto a se destacar é que, dentre as três doenças abordadas neste trabalho, as pesquisas que envolvem predição de arboviroses estão, em sua maioria, voltados para a predição da dengue e não no conjunto de doenças transmitidas por um único vetor.

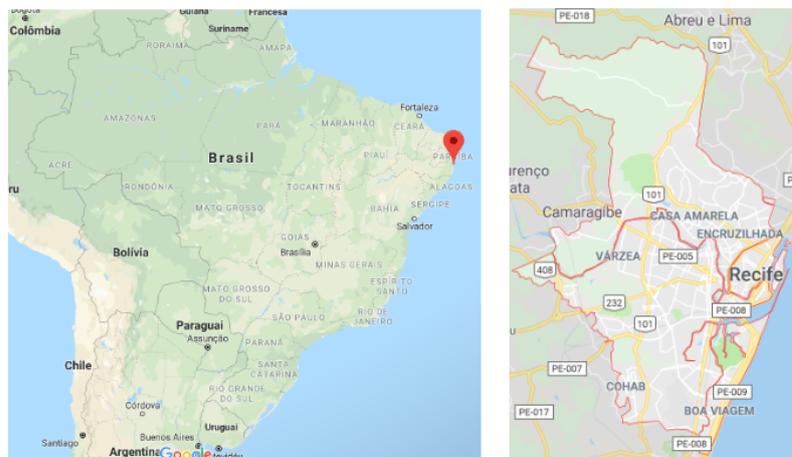
## 4 PROPOSTA

Neste capítulo, é apresentada a metodologia utilizada neste trabalho para a predição de casos de arboviroses transmitidas pelo *Aedes aegypti* na Cidade do Recife, bem como as bases de dados que foram utilizadas ao longo da pesquisa, a saber: a base de dados dos casos de arboviroses (2013-2016), disponibilizada pela Prefeitura da Cidade do Recife, e as bases de dados das variáveis climáticas (temperatura, velocidade dos ventos e pluviometria), que são disponibilizadas pelo Instituto Nacional de Meteorologia e pela Agência Pernambucana de Águas e Climas. Também foram apresentados os *softwares* utilizados ao longo da pesquisa, além da elaboração dos conjuntos de dados das predições dos casos.

### 4.1 Área sob estudo

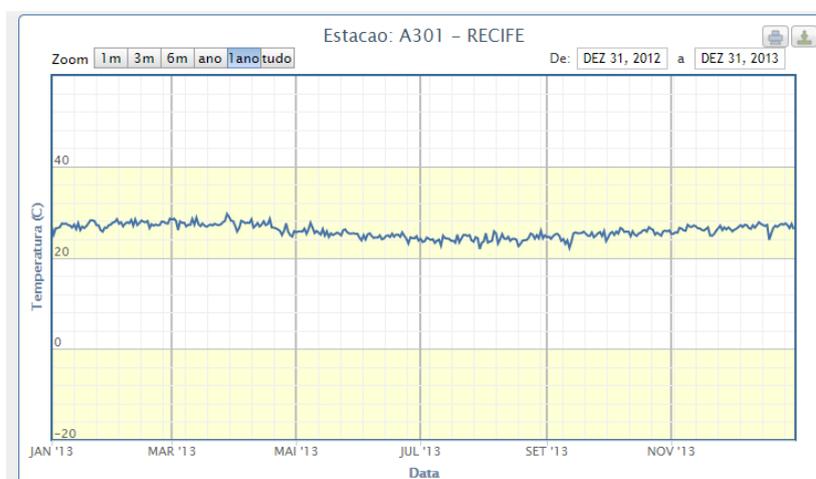
A área delimitada para o presente trabalho foi a Cidade do Recife ( $8^{\circ}03'14''S$ ,  $34^{\circ}52'51''W$ ), capital do Estado de Pernambuco o qual está localizado na Região Nordeste do Brasil (Figura 7). O Recife, de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE) possui uma extensão territorial de aproximadamente  $218 \text{ km}^2$  e cerca de 1.637.834 milhões de habitantes, além de ser a cidade nordestina com maior Índice de Desenvolvimento Humano (IDH), de acordo com o censo de 2010 (IBGE, 2018). O clima da Cidade do Recife é caracterizado como tropical úmido, com temperaturas médias mensais superiores a  $18^{\circ}\text{C}$ , alta umidade relativa do ar e precipitação elevada ao longo do ano, de acordo com o Instituto Nacional de Meteorologia, como pode visto nas Figuras 8, 9 e 10.

Figura 7 – Localização geográfica da Cidade do Recife, capital do Estado de Pernambuco, Nordeste do Brasil.



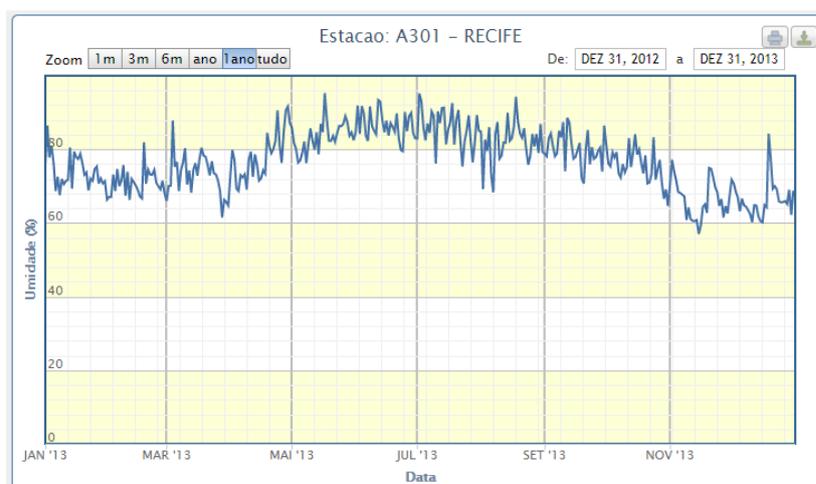
Fonte: Google Maps.

Figura 8 – Gráfico da série histórica da distribuição da temperatura na cidade do Recife ao longo do ano de 2013 correspondente à estação A301, localizada na latitude -8.05928 e longitude -34.959239.



Fonte: Instituto Nacional de Meteorologia

Figura 9 – Gráfico da série histórica da distribuição da umidade relativa do ar na cidade do Recife ao longo do ano de 2013 correspondente à estação A301, localizada na latitude -8.05928 e longitude -34.959239.



Fonte: Instituto Nacional de Meteorologia.

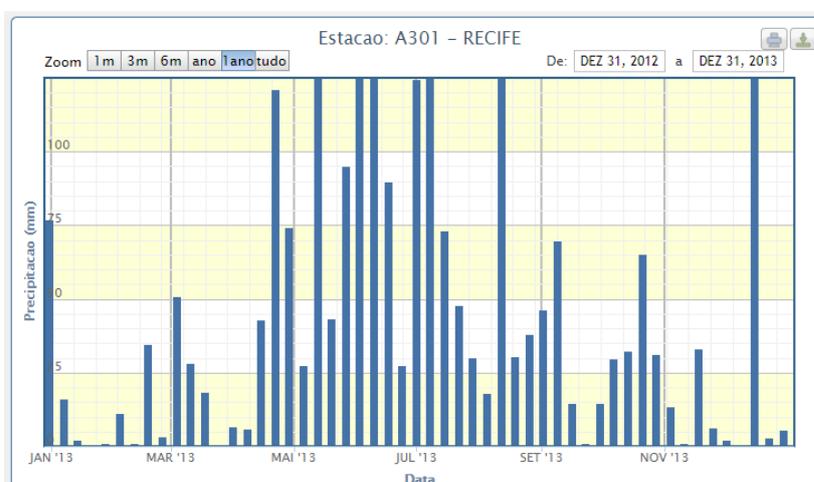
## 4.2 Base de dados

### 4.2.1 Portal de Dados Abertos da Prefeitura de Recife

O Portal de Dados Abertos da Prefeitura de Recife <sup>1</sup> é uma plataforma desenvolvida pela EMPREL - Empresa Municipal de Informática, cujo objetivo é a disponibilização, de forma pública e fácil, da busca de dados governamentais gerados por secretarias e órgãos da gestão municipal. O acesso a esses dados permite que qualquer um desenvolva

<sup>1</sup> Portal de Dados Abertos da Prefeitura de Recife, disponível em <<http://dados.recife.pe.gov.br/>>, acessado em 15 de novembro de 2019.

Figura 10 – Distribuição da precipitação na cidade do Recife ao longo do ano de 2013 correspondente à estação A301, localizada na latitude -8.05928 e longitude -34.959239.



Fonte: Instituto Nacional de Meteorologia

aplicações ou visualizações no intuito de facilitar a análise de dados, promovendo assim, a melhoria de serviços por meio da inovação e da criatividade, contribuindo para uma maior participação da sociedade junto ao governo municipal. Os dados estão disponibilizados nos formatos CSV e PDF, mas também é possível realizar buscas utilizando protocolo JSON (RECIFE, 2018).

A Secretaria de Saúde disponibiliza 29 (vinte e nove) conjuntos de dados, nos quais se encontram dados referentes a Unidades Básicas de Saúde, Unidades Saúde da Família, Ambulatórios, Policlínicas, incluindo os casos confirmados de dengue (2013-2016), zika (2015 e 2016) e *chikungunya* (2015-2016).

Para este trabalho, foram coletados os dados referentes aos casos de dengue, zika e *chikungunya*. O tratamento dado a esses dados é descrito mais adiante, ainda neste capítulo. É importante frisar que as informações relacionadas à identidade dos pacientes são ocultadas pela própria Secretaria de Saúde, garantindo assim, o anonimato dos mesmos.

#### 4.2.2 Banco de Dados Meteorológicos para Ensino e Pesquisa - BDMEP

O Instituto Nacional de Meteorologia <sup>2</sup> (INMET) foi criado pelo Decreto 7.672 do presidente Nilo Peçanha, em 18 de novembro de 1909, sob o nome de Diretoria de Meteorologia e Astronomia, vinculado ao Ministério da Agricultura Indústria e Comércio. Esse instituto tem como principal missão fornecer informações meteorológicas à sociedade brasileira e influir, construtivamente, no processo de tomada de decisão, contribuindo para o desenvolvimento sustentável do país (INMET, 2018). Sendo assim, o INMET realiza o monitoramento, análise

<sup>2</sup> Instituto Nacional de Meteorologia - INMET, disponível em <[http://www.inmet.gov.br/portal/index.php?r=home/page&page=sobre\\_inmet](http://www.inmet.gov.br/portal/index.php?r=home/page&page=sobre_inmet)>, acessado em 15 de novembro de 2019.

e previsão de tempo e de clima.

São atribuições do INMET: elaborar e divulgar, diariamente, em nível nacional, a previsão do tempo, avisos e boletins meteorológicos especiais; promover a execução de estudos e levantamentos meteorológicos e climatológicos aplicados à agricultura e outras atividades correlatas; coordenar, elaborar e executar programas e projetos de pesquisas agrometeorológicas e de acompanhamento das modificações climáticas e ambientais; estabelecer, coordenar e operar as redes de observações meteorológicas e de transmissão de dados, inclusive aquelas integradas à rede internacional; propor a programação e acompanhar a implementação de capacitação e treinamento de recursos humanos, em atendimento a demandas técnicas específicas (INMET, 2018).

As informações meteorológicas monitoradas pelo INMET são armazenadas no Banco de Dados Meteorológicos para Ensino e Pesquisa <sup>3</sup> (BDMEP). Esse banco de dados contém os registros meteorológicos diários, em formato digital, de séries históricas das diversas estações meteorológicas convencionais da rede de estações, referentes às medições diárias, conforme as normas técnicas internacionais da Organização Meteorológica Mundial (OMM). No BDMEP, estão disponibilizados registros diários a partir do ano de 1961, para os quais se dispõem, em formato digital, de cerca de 80% dos dados que foram registrados naquele período. Para os anos anteriores a 1961, os dados históricos ainda não estão disponíveis no BDMEP. As variáveis atmosféricas armazenadas no BPMED são (INMET, 2018): precipitação ocorrida nas últimas 24h, temperatura do bulbo seco, temperatura do bulbo úmido, temperatura máxima, temperatura mínima, umidade relativa do ar, direção e velocidade do vento, pressão atmosférica ao nível da estação e insolação.

Neste trabalho, foram utilizados os dados mensais da temperatura máxima e velocidade dos ventos para os anos de 2013 a 2014. No caso da cidade do Recife, essas variáveis são registradas pela estação A301, de acordo com o código do INMET (ou 81958, de acordo com o código da OMM), localizadas na latitude -8.05928 e longitude -34.959239 (Figura 11).

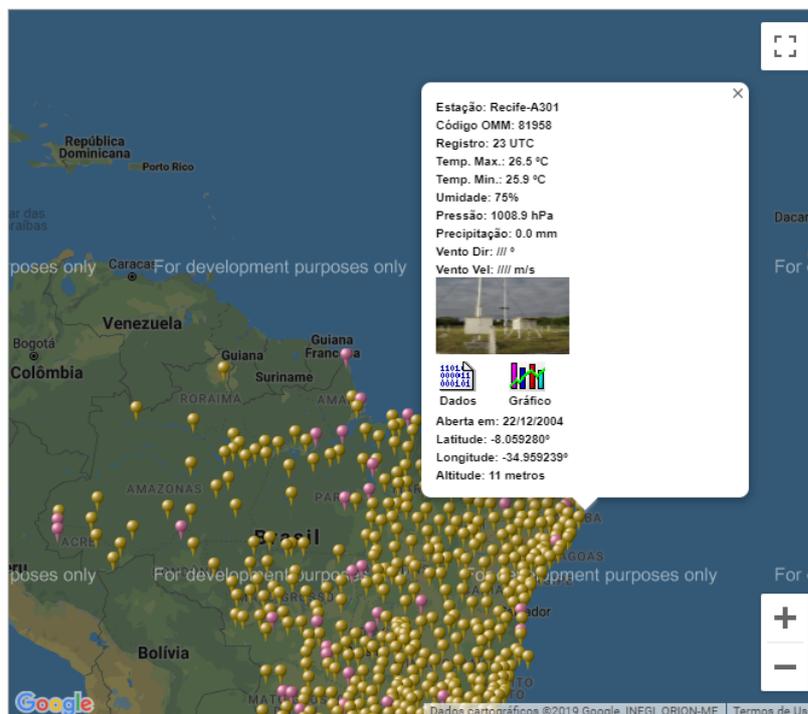
#### 4.2.3 Sistema de Geoinformação Hidrometeorológico de Pernambuco - SIGHPE

A Agência Pernambucana de Águas e Climas <sup>4</sup> foi criada através da Lei Estadual N° 14.028 de 26 de março de 2010 para a execução da Política Estadual de Recursos Hídricos (Lei n°12.984/2005) e complementação do Sistema Integrado de Gerenciamento dos Recursos Hídricos (SIGRH), bem como fortalecer o planejamento e regulação dos múltiplos

<sup>3</sup> Banco de Dados Meteorológicos para Ensino e Pesquisa - BDMEP, disponível em <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>, acessado em 15 de novembro de 2018.

<sup>4</sup> Agência Pernambucana de Águas e Clima - APAC, disponível em <<http://www.apac.pe.gov.br/>>, acessado em 10 de novembro de 2018.

Figura 11 – Interface do mapa da rede de estações do Instituto Nacional de Meteorologia para o monitoramento das variáveis climáticas.



Fonte: Instituto Nacional de Meteorologia.

recursos hídricos no Estado, realizar monitoramento hidrometeorológicos e previsões de tempo e clima no Estado (APAC, 2018).

O monitoramento hidrometeorológico é realizado, em tempo real, por meio de Plataformas de Coleta de Dados (PCDs) que estão distribuídas em todo estado de Pernambuco. As PCDs fornecem informações dos postos pluviométricos (que medem a precipitação em mm), das barragens e da qualidade da água dos reservatórios, bem como informações das bacias hidrográficas do Estado. As informações referentes aos dados pluviométricos estão armazenados no banco de dados do Sistema de Geoinformação Hidrometeorológica de Pernambuco<sup>5</sup> (SIGHPE), no qual constam as informações das medições diárias ou os acumulados mensais medido por um determinado posto pluviométrico (Figura 12). Como pode ser visto nas Figuras 13, 14 e 15, a cidade do Recife possui três postos pluviométricos que são o Alto da Brasileira, Várzea e Codecipe/Santo Amaro, respectivamente. Neste trabalho foram coletados, dos três postos, os acumulados mensais dos anos de 2013 a 2016.

<sup>5</sup> Sistema de Geoinformação Hidrometeorológica de Pernambuco - SIGHPE, disponível em <<http://www.apac.pe.gov.br/sighpe/>>, acessado em 10 de novembro de 2018.

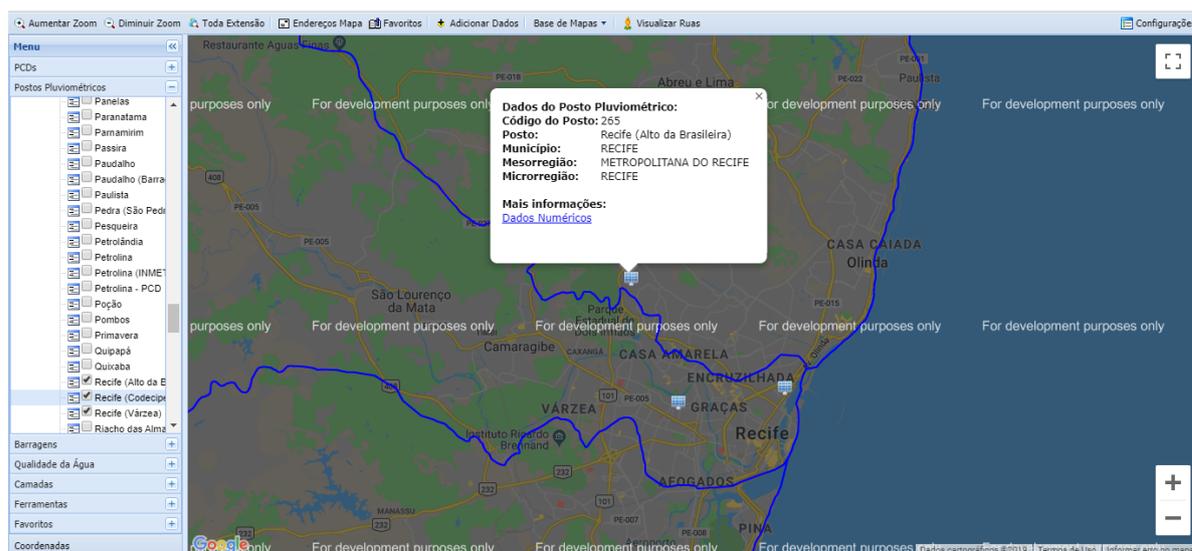
Figura 12 – Interface do banco de dados do SIGHPE para a seleção dos dados pluviométricos.



The screenshot shows the web interface of the SIGHPE database. At the top, there are logos for 'Apac Agência Pernambucana de Águas e Clima' and 'GOVERNO DO ESTADO DE PERNAMBUCO'. Below the logos, a breadcrumb trail reads: 'Você está aqui: Página Inicial > Serviços > Meteorologia > Monitoramento Pluviométrico'. A message states: 'Escolha as opções de filtro abaixo para exibir os resultados.' The main form includes: 'Visualização' with radio buttons for 'Diária' and 'Mensal' (selected); 'Data Inicial' set to '01/11/2019'; 'Data Final' set to '10/11/2019'; 'Mesorregião' dropdown set to 'Selecione'; 'Posto' dropdown set to '1 selecionados'; a 'Pesquisar' button; and an 'Exportar para excel' button. At the bottom, 'Ordenação por' has radio buttons for 'Data' and 'Posto'.

Fonte: Agência Pernambucana de Águas e Climas.

Figura 13 – Estação pluviométrica da Agência Pernambucana de Águas e Climas, localizada no Alto da Brasileira, Nova Descoberta, Recife-PE.



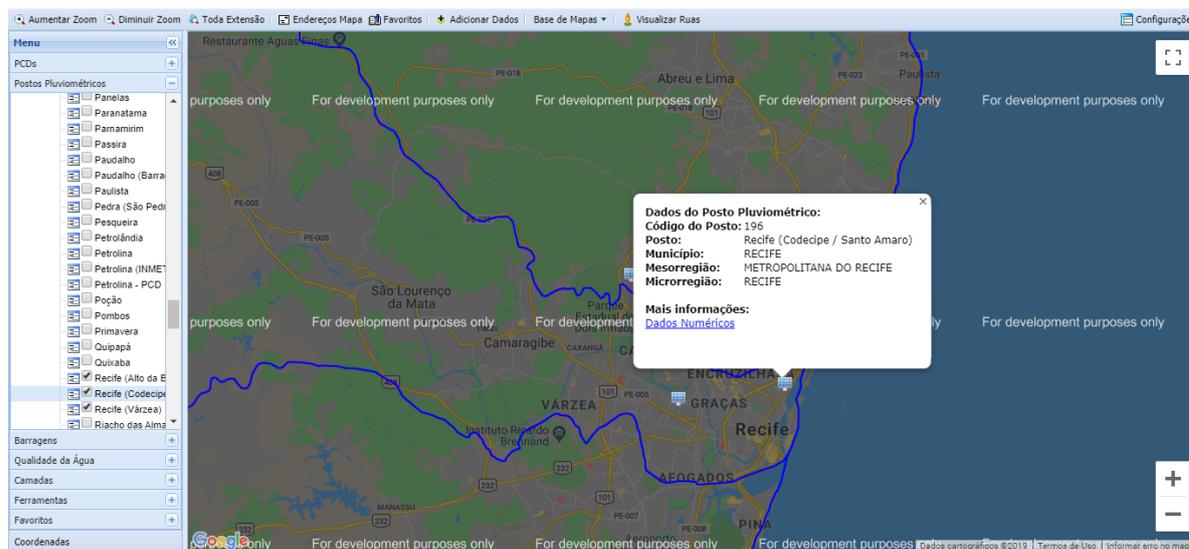
Fonte: Agência Pernambucana de Águas e Climas.

## 4.3 Softwares utilizados

### 4.3.1 Python

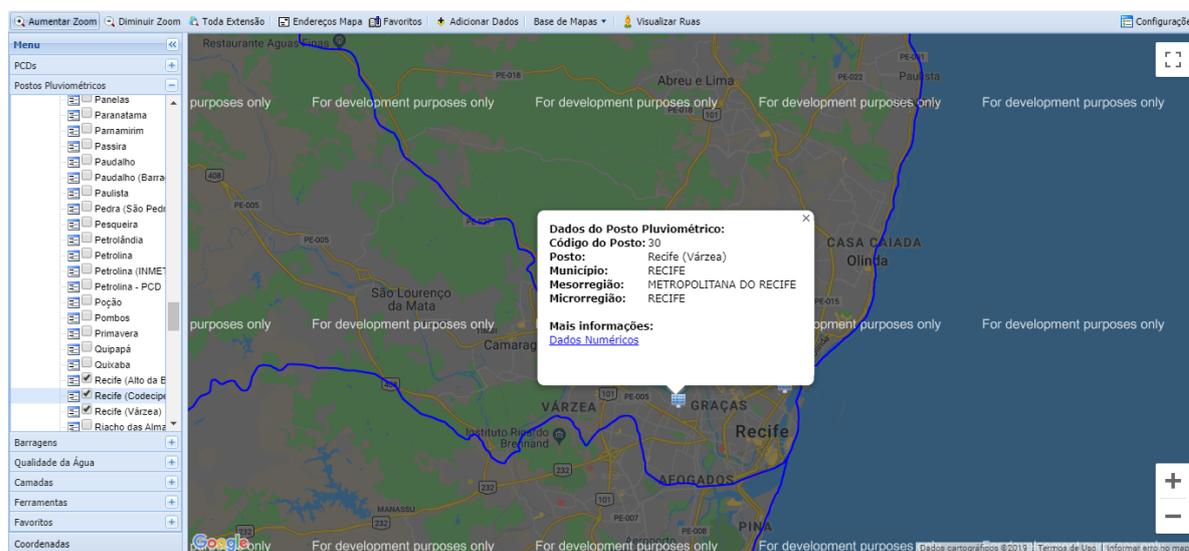
Python é uma linguagem de programação interativa e orientada a objetos. Ela incorpora módulos, exceções, digitação dinâmica, tipos de dados dinâmicos de nível muito alto e classes. O Python pode ser executado em várias plataformas, por exemplo: Unix, Mac, Windows 2000 e as versões posteriores (Python Software Foundation, 2019). No presente trabalho, foi utilizada a sua versão 3.7.4 para a elaboração do código da contagem do número de casos de arboviroses nos bairros do Recife e do código da estimativa dos valores faltantes das bases de dados das variáveis climáticas. Além disso, foi por meio

Figura 14 – Estação pluviométrica da Agência Pernambucana de Águas e Climas, localizada no bairro de Santo Amaro, Recife-PE.



Fonte: Agência Pernambucana de Águas e Climas.

Figura 15 – Estação pluviométrica da Agência Pernambucana de Águas e Climas, localizada no bairro da Várzea, Recife-PE.



Fonte: Agência Pernambucana de Águas e Climas.

dessa ferramenta que também foram elaborados os *scripts* para a geração das imagens dos modelos validados.

#### 4.3.2 QGIS

O QGIS é um sistema de informações geográficas de código aberto e foi estabelecido como projeto no SourceForge em junho de 2002. Esse *software* foi desenvolvido utilizando as ferramentas Qt (<https://www.qt.io>) e C++, podendo ser executado na maioria das plataformas Unix, Windows e macOS. Atualmente, a ferramenta suporta vários formatos

de dados rasterizados e vetoriais e outros novos formatos que podem ser adicionados por meio de *plug-in*. Por ser um *software* livre, o QGIS é liberado sob a GNU General Public License (GPL), o que significa que é possível realizar, livremente, inspeções e/ou modificações no código-fonte (QGIS Development Team, 2018). Neste trabalho, foi utilizada a sua versão 2.18.25 para gerar os mapas de distribuição dos casos das arboviroses e das variáveis climáticas.

### 4.3.3 Weka

O Weka (*Waikato Environment for Knowledge Analysis*), desenvolvido pela Universidade de Waikato na Nova Zelândia, é um *software* que apresenta uma coleção de algoritmos de aprendizagem de máquina e ferramentas para pré processamento de dados. É uma ferramenta desenvolvida em Java e, por ser um *software* de código aberto, está sob os termos da GNU (General Public License). Além disso, o Weka pode ser executado nos sistemas operacionais Windows, Linux e Macintosh (WITTEN; FRANK, 2005). A versão utilizada nesta pesquisa foi a 3.8, tanto para gerar as bases de dados de treinamento, na avaliação dos regressores para a predição em diversas configurações, quanto na validação dos modelos de predição.

## 4.4 Metodologia

Neste trabalho, é proposta uma metodologia para a predição de casos de arboviroses na cidade do Recife utilizando-se uma abordagem espaço-temporal a partir dos mapas de distribuição dos casos e variáveis climáticas dos anos de 2013 a 2016. A discussão abordada nessa seção é a realização dos mapeamentos bimestrais das arboviroses e das variáveis climáticas, bem como a elaboração das bases de predição.

### 4.4.1 Mapas de distribuição dos casos

Os dados referentes aos casos de arboviroses foram obtidos a partir do Portal de Dados Abertos da Prefeitura de Recife, do qual foram coletados casos de dengue (2013-2016), zika (2015-2016) e *chikungunya* (2015-2016). Os dados dos casos confirmados de cada tipo de arbovirose foram organizados por bimestre e por ano, para a realização da contagem dos casos em cada bairro. A contagem dos casos de cada tipo de arbovirose foi realizado por meio de um código em Python, o qual recebe como entrada o arquivo .csv dos casos confirmados da doença e gera um arquivo .xlsx com o número de casos em cada bairro. Como nos arquivos (.csv) originais possuem erros de digitação, neste código, quando o(s) bairro(s) não é(ão) identificado(s), um novo arquivo é gerado com a lista do(s) bairro(s) não identificado(s). Sendo assim, a cada vez que a lista de bairros não identificados era gerada, corrigia-se manualmente o .csv e este mesmo arquivo era submetido para a recontagem de casos (Figura 16).

Como este trabalho não tem intuito de realizar a predição de casos da doença para cada tipo de arbovirose de forma separada e, sim, casos de doenças transmitidas pelo *Aedes aegypti*, foram somados todos os casos confirmados das três arboviroses para cada bimestre, para cada ano, nos bairros do Recife no período de 2013 a 2016. Em seguida, os 93 bairros da cidade do Recife, juntamente com o somatório dos casos de arboviroses, foram geocodificados por meio da ferramenta *geocoding* do QGIS e, para cada bimestre dos anos de 2013 a 2016, foi gerada uma camada vetorial de pontos, *shapefile* (.shp), localizando, geograficamente, no mapa da cidade do Recife, a quantidade de casos em cada bairro da cidade. Esta camada vetorial de pontos também foi gerada por meio do *software* QGIS e o resultado do processo descrito pode ser visualizado na Figura 17.

Para poder estimar a distribuição de casos de arboviroses em todo território do Recife, foi utilizada a ferramenta de interpolação do QGIS. O método selecionado foi a interpolação pelo inverso da distância (com distância para coeficiente  $P$  igual a 2). Como resultado da interpolação, obteve-se uma imagem *raster* (.tif) que foi recortada utilizando o *shapefile* da cidade do Recife como máscara, o que pode ser observada na Figura 18. Sendo assim, foram gerados *rasters* dos casos confirmados das arboviroses para cada bimestre dos anos de 2013 a 2014.

#### 4.4.2 Mapa de distribuição das variáveis climáticas

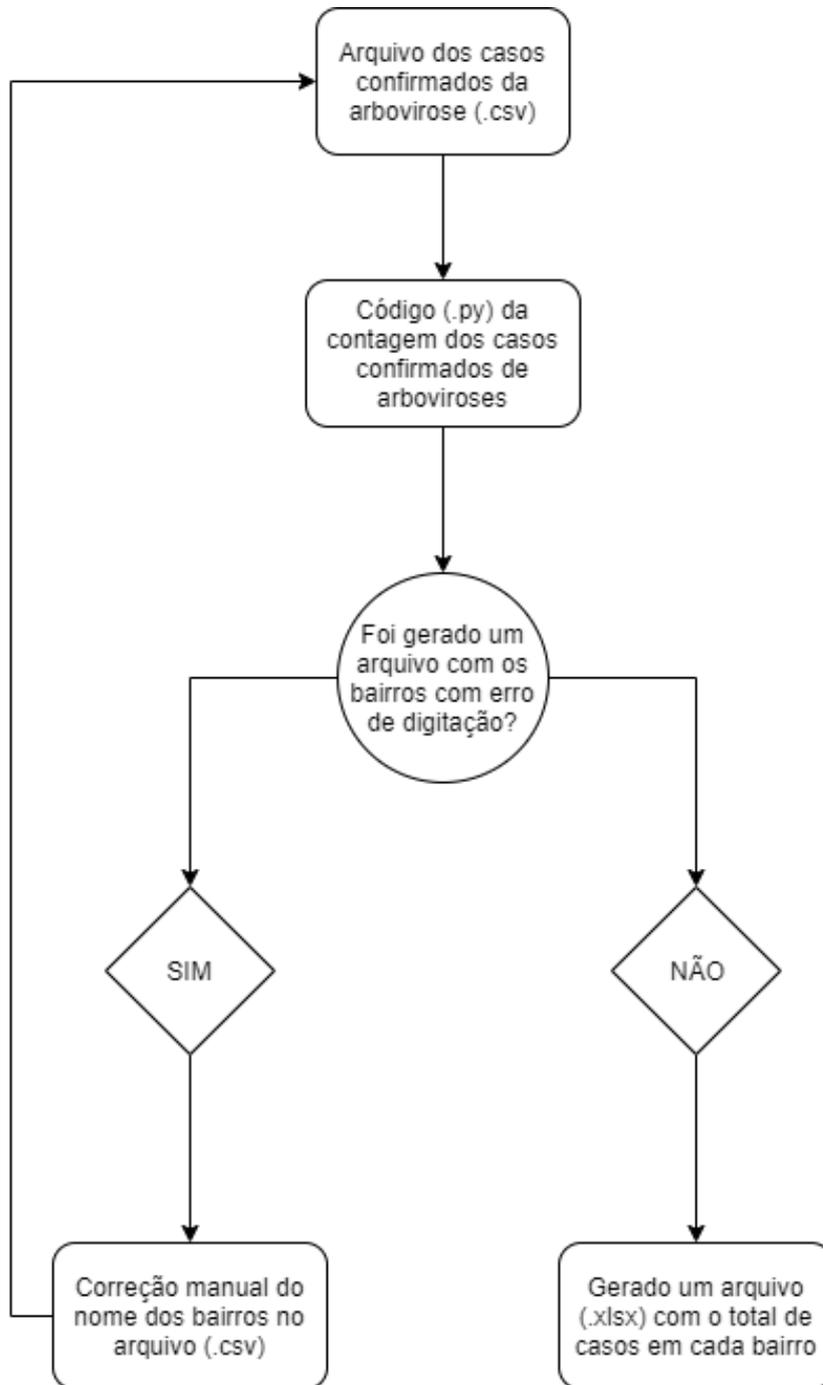
Os fatores climáticos como precipitação, temperatura e umidade estão entre as causas de um aumento no número de casos de arboviroses. O comportamento do vetor transmissor é determinado pelas condições climáticas. Isso ocorre porque chuvas, temperatura e umidade influenciam na interação de vetores biológicos e vetores virais, ao longo da vida, idade de acasalamento, disseminação, alimentação e replicação viral mais rápida (TOSEPU et al., 2018; SIRIYASATIEN et al., 2018).

Como mencionado na seção 4.2.2, o monitoramento das variáveis climáticas, no Brasil, é realizado pelo Instituto Nacional de Meteorologia que faz este monitoramento através de estações meteorológicas distribuída em todo território nacional, dentre eles, o território recifense. Os dados referentes à temperatura e à velocidade dos ventos foram coletados na base de dados do INMET, da qual foram obtidas as séries históricas dos anos de 2013 a 2016 (para ser compatível com o intervalo de tempo dos casos de arboviroses).

A APAC, no Recife, realiza o monitoramento dos índices hidrometeorológicos e distribui essas informações por meio da plataforma do SIGHPE (como foi detalhado na seção 4.2.3). O SIGHPE disponibiliza as séries históricas dos índices pluviométricos desde o ano de 2006. Para este trabalho, foram coletados apenas os acumulados mensais de chuvas dos anos de 2013 até o ano de 2016.

Na cidade do Recife, os registros de temperatura e velocidade dos ventos são monitorados por uma única estação, enquanto que os registros dos acumulados mensais

Figura 16 – Fluxograma da realização da contagem dos casos confirmados, de cada tipo de arboviroses, por bairro, para cada bimestre de cada ano.



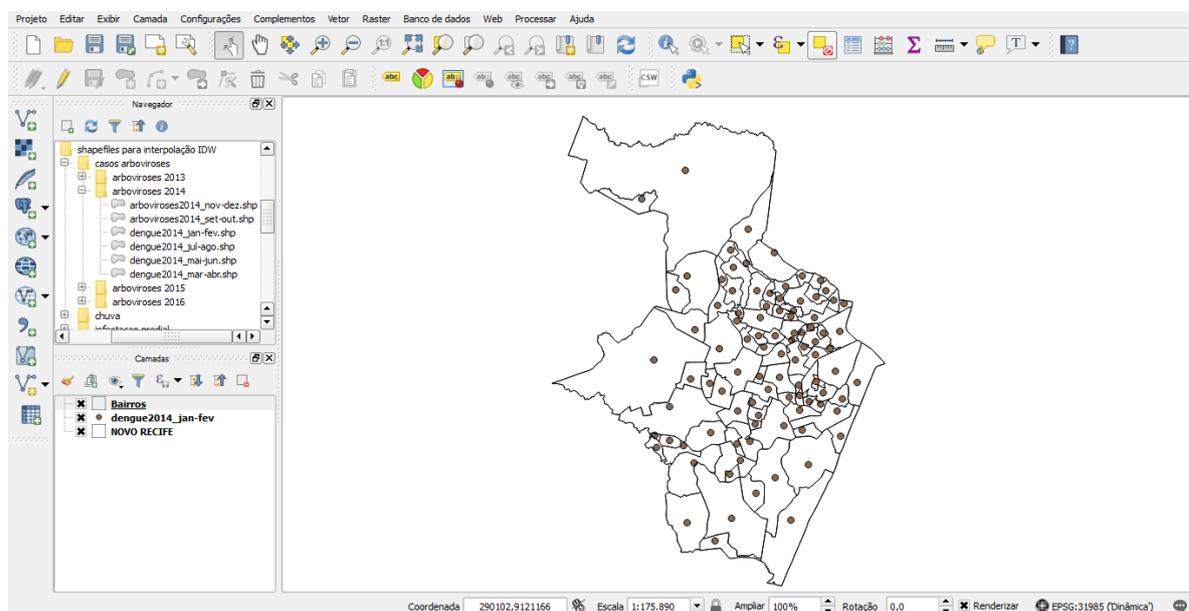
Fonte: A autora

de chuvas são realizados por três estações em bairros distintos. Os registros nos demais bairros do Recife foram estimados por meio de uma distribuição gaussiana (Equação 4.1), em que,  $\sigma^2$  representa a variância da distribuição e  $\mu$  representa a média da distribuição.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}} \quad (4.1)$$

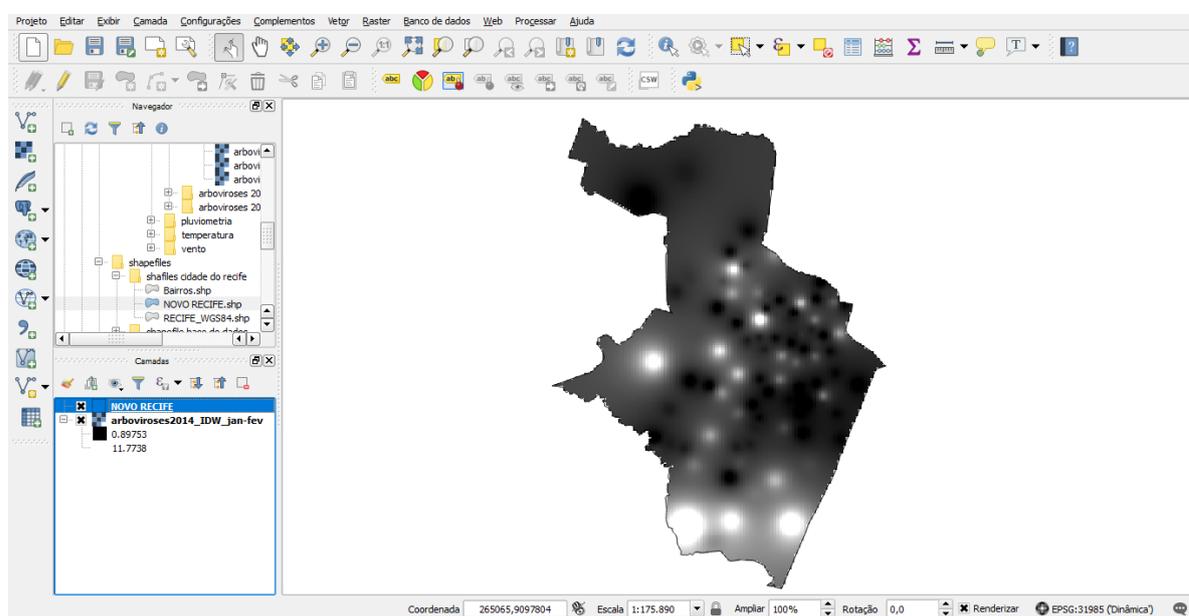
Os valores do desvio padrão da amostra foram calculados a partir da Equação 4.2,

Figura 17 – Camada vetorial de pontos dos casos de arbovirose na cidade do Recife para o primeiro bimestre de 2014.



Fonte: A autora.

Figura 18 – Mapa de distribuição dos casos de arbovirose para a cidade do Recife para o primeiro bimestre de 2014.



Fonte: A autora.

em que  $x_{max}$  representa o valor máximo e  $\mu$  representa a média mensal dos valores de temperatura e velocidade dos ventos.

$$\sigma = \frac{x_{max} - \mu}{4} \quad (4.2)$$

Para os acumulados mensais das chuvas, o valor máximo considerado foi o valor máximo registrado entre as três estações de monitoramento, enquanto que a média con-

siderada foi a média do acumulado de chuva entre as três estações de monitoramento. A estimativa dos valores das variáveis climáticas para cada bairro do Recife foi gerada por meio de um arquivo (.py), no qual foi utilizado a função *random.normal* do módulo *numpy* (versão 1.17.0) para calcular a distribuição normal dessas variáveis, utilizando a média e o desvio padrão, e salvá-las em formato .xlsx. Com as informações das variáveis climáticas em todos os bairros do Recife, foram gerados os *shapefiles* para cada uma das variáveis, para cada mês dos anos de 2013 a 2016. Por fim, a ferramenta de interpolação pelo inverso da distância (com distância para coeficiente  $P$  igual a 2) foi utilizada para estimar a distribuição espacial das variáveis climáticas ao longo do território do Recife, de modo semelhante ao que foi feito com os casos das doenças.

#### 4.4.3 Conjuntos de predição

Os conjuntos de predição foram montados a partir dos mapas de distribuição dos casos de arboviroses e das variáveis climáticas para cada bimestre. O modelo de predição bimestral foi escolhido devido ao fato do Sistema Único de Saúde (SUS) realizar o planejamento de combate aos surtos de arboviroses considerando o ciclo de bimestres. Para a leitura dos mapas de distribuição dos casos e das variáveis climáticas, foi elaborado um código em Python, utilizando-se a biblioteca GDAL (versão 2.4.1) para a manipulação dos *rasters*. Neste código, os mapas de distribuição espacial foram varridos pixel a pixel, simultaneamente, concatenando latitude e longitude com as seguintes informações, na seguinte ordem: distribuição dos casos de arboviroses e, para cada mês do bimestre, a distribuição de temperatura, pluviometria e velocidade dos ventos. Como saída, foram gerados 4 arquivos .xlsx, contendo as informações dos casos e variáveis climáticas para cada ano. A partir desses arquivos .xlsx foram montados, manualmente, os conjuntos de predição bimestral, com cada conjunto contendo informações dos seis bimestres que antecedem o bimestre de predição. Após a elaboração dos conjuntos de predição, os conjuntos de predição com extensão .xlsx foram convertidos para arquivos de extensão .arff com o auxílio da biblioteca *liac-arff* (Python). Desta forma, foram montados 18 conjuntos de predição, cada um com 15553 instâncias e 44 atributos, no qual a saída de cada conjunto de predição é valor do pixel da distribuição dos casos de arboviroses na coordenada correspondente. Os conjuntos de 15553 instâncias foram estabelecidos como conjuntos de teste.

O conjunto de treino foi montado utilizando a ferramenta *resample* do *software* Weka. Esta ferramenta permite com que uma nova base de dados seja criada com valores aleatórios para instâncias, porém, com as mesmas características estatísticas da base original (WITTEN; FRANK, 2005). O número de instâncias da nova base, no entanto, deve ser especificado. Neste caso, os conjuntos de treinamento foram gerados aplicando o *resample* em cada um dos conjuntos de predição com o número de instâncias equivalente a 30% o conjunto original. Os conjuntos contendo 15553 instâncias foram utilizados para testar os modelos criados pelo melhor regressor.

A partir do conjunto de treinamento, foram investigadas as melhores arquiteturas de regressores para a predição da distribuição dos casos de arboviroses, sendo elas:

- Regressão Linear;
- Máquina de Vetor de Suporte (SVM):  $C = 0.1$ , *kernels* polinomiais de grau 1 (ou *kernel* linear), graus 2 e 3, e RBF;
- Perceptron Multicamadas (MLP) com uma única camada escondida com 10, 20, 30 e 40 neurônios.

#### 4.5 Considerações

Neste capítulo, foi detalhada a metodologia proposta neste trabalho. Foram descritas a região geográfica sob a qual este estudo foi realizado e as bases de dados utilizadas para a obtenção dos casos confirmados de dengue, *chikungunya* e zika; e as variáveis climáticas. Também foram apresentados os *softwares* utilizados ao longo da pesquisa e como foi realizado o tratamento dos dados coletados. Além disso, foram apresentados os métodos de regressão investigados, a saber: a regressão linear, as redes neurais e as máquinas de vetores de suporte para regressão.

## 5 RESULTADOS EXPERIMENTAIS

Neste capítulo, são detalhados os resultados obtidos pelos algoritmos de regressão utilizados para a modelagem da distribuição dos casos de arboviroses para os bimestres dos anos de 2014 a 2016. Também são apresentados e discutidos os mapas de distribuição de casos obtidos por meio da aplicação dos modelos gerados pelo melhor e pelo pior regressor.

### 5.1 Resultados quantitativos

Cada regressor foi avaliado em 30 rodadas utilizando-se a validação cruzada com 10 *folds*. Para a avaliação quantitativa dos regressores, foram calculados o coeficiente de correlação (R), o erro médio absoluto (MAE), o erro médio quadrático (RMSE) e o erro quadrático relativo (RMSE percentual). No entanto, os dados foram avaliados levando-se em consideração apenas o coeficiente de correlação, como qualidade global, e o erro quadrático relativo, como métrica de qualidade local.

O coeficiente de correlação é uma medida estatística entre os valores esperados e os valores dos preditos. Este valor varia de -1 a 1, em que quando se aproxima de 1, indica uma forte correlação positiva. Quando o coeficiente de correlação se aproxima-se de -1, indica que as variáveis possuem uma forte correlação negativa, e quando o coeficiente de correlação é próximo de zero, indica que não há correlação entre os resultados (WITTEN; FRANK, 2005). O valor do coeficiente de correlação serve como avaliador geral do regressor, sendo assim, é possível obter um alto valor para o coeficiente de correlação e, ao mesmo tempo, obter altos valores para os erros locais. Por esta razão, o R não pode ser a única métrica para avaliar o desempenho dos modelos. Para evitar que houvesse uma avaliação superficial dos regressores, o erro quadrático médio, que é uma métrica de erro local, foi escolhido como uma segunda métrica de avaliação. A Equação 5.1 mostra a expressão do cálculo do erro médio quadrático percentual, em que  $p$  é o valor predito, e  $a$  corresponde ao valor real.

$$\text{RMSE}(\%) = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - a_m)^2 + \dots + (a_n - a_m)^2}} \times 100\% \quad (5.1)$$

Quanto mais próximo de 0% for o valor do RMSE%, melhor regressor. Desta forma, um regressor com um bom desempenho é aquele cujo valor de R é um valor considerado alto e o valor de RMSE% é um valor considerado baixo. Neste trabalho, considera-se um alto coeficiente de correlação como sendo acima de 0,9 e um baixo RMSE% como sendo abaixo de 5%. Apesar do tempo de treinamento não ser algo crítico para este trabalho, essa

grandeza foi utilizada para auxiliar na tomada de decisão de qual regressor obteve o melhor desempenho dentro de um mesmo conjunto de métodos. Afinal, o tempo de treinamento está relacionado com a complexidade do método escolhido, ou seja, um alto tempo de treinamento acarreta um maior custo computacional. Os resultados detalhados dos valores de R, RMSE% e tempo de treinamento, de cada método avaliado, são exibidos nas tabelas 1, 2 e 3. Em vermelho, estão destacados os melhores resultados.

Na Tabela 1, os resultados mostram que a regressão linear apresenta um valor satisfatório para o coeficiente de correlação R, com média de 0,97 e desvio padrão de 0,03, e para o tempo de treinamento, com média de 0,05 e desvio padrão de 0,03. Este método de regressão mostrou-se bastante rápido e estável (devido ao baixo valor do desvio padrão). Por outro lado, o erro quadrático relativo RMSE%, apresenta um valor consideravelmente alto, com média 21,73% e desvio padrão de 12,11%. De acordo com os critérios estabelecidos para este trabalho, podemos afirmar que a regressão linear não apresentou um bom desempenho.

Tabela 1 – Resultados de coeficiente de correlação, erro quadrático relativo e tempo de treinamento para a Regressão Linear.

Método de regressão	Configuração	R		RMSE (%)		Tempo de treinamento(s)	
		Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
Regressão Linear	Única	0,97	0,03	21,73	12,11	0,05	0,03

Fonte: A autora.

Na Tabela 2, são apresentados os resultados para os perceptrons multicamadas, com uma única camada escondida, nas configurações com 10, 20, 30 e 40 neurônios. Os resultados apontam que, para todas as arquiteturas avaliadas, os coeficientes de correlação R apresentaram valores bastante elevados com médias em torno de 0,999 e 1, e desvios padrão de 0,001. Quanto ao erro quadrático relativo, observa-se que a configuração com 10 neurônios apresenta um RMSE% razoavelmente baixo, com média de 4,15% e desvio padrão de 2,12%. Na configuração com 20 neurônios, o valor do RMSE% caiu para 3,66%, com um desvio padrão de 1,80%. Para a configuração com 30 neurônios, o RMSE% apresentou um valor de 3,29% e mostrou-se bastante estável. Na última configuração de MLP avaliada, o RMSE% apresentou um valor razoavelmente baixo, com média de 3,69% e desvio padrão de 1,89%. Quanto ao tempo de treinamento, observa-se (na Tabela 2) que este se eleva com o aumento da quantidade de neurônios na única camada escondida. No caso dos perceptron multicamadas, o aumento do número de neurônios na camada escondida influencia na complexidade da rede (e também na complexidade da fronteira de decisão) (HAYKIN, 2001). Aumentando a complexidade da rede, aumenta-se o tempo de treinamento dos modelos e, conseqüentemente, acarretando um maior custo computacional. Analisando o tempo de treinamento e, considerando as métricas de avaliação, conclui-se que a melhor configuração de rede, dentre as que foram avaliadas, foi a configuração de MLP com 30 neurônios na única camada escondida. Pois, dentre as configurações testadas,

apresentou o RMSE% com o valor mais próximo de zero e com um tempo de treinamento razoavelmente baixo.

Tabela 2 – Resultados do coeficiente de correlação, erro quadrático médio e tempo de treinamento para o perceptron multicamadas, MLP, com 10, 20, 30 e 40 neurônios na única camada escondida.

Método de regressão	Configuração	R		RMSE (%)		Tempo de treinamento(s)	
		Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
MLP, única camada escondida	10 neurônios	0,999	0,001	4,15	2,12	46,59	10,30
	20 neurônios	0,999	0,001	3,66	1,80	57,44	6,55
	30 neurônios	1,000	0,001	3,29	1,57	71,28	5,88
	40 neurônios	0,999	0,001	3,67	1,89	78,72	10,08

Fonte: A autora.

A Tabela 3 mostra os resultados para o SVM nas configurações com *kernels* polinomiais de grau 1 (ou linear), graus 2, 3 e RBF. Os valores dos coeficientes de correlação para todas as configurações de *kernel* foram consideravelmente altos e bastante estáveis (com baixos desvios padrão). Destacam-se, principalmente, os coeficientes de correlação dos *kernels* polinomiais de graus 2 e 3, os quais apresentaram médias de 0,999 e 1, respectivamente, e desvios padrão de 0,001 e  $6.818 \times 10^{-5}$ , nesta ordem. Quanto ao RMSE%, os *kernels* polinomiais de graus 2 e 3 apresentaram valores satisfatórios de 3,09% e 1,20%, respectivamente. Estas configurações de SVM também mostraram ser bastante estáveis, apresentando baixos desvios padrão de 1,98% e 0,40%, respectivamente. Em contrapartida, as configurações utilizando *kernels* linear e RBF apresentaram valores altíssimos de RMSE%, com médias de 26,49%, para a configuração de SVM com *kernel* linear, e 40,49%, para a configuração de SVM com *kernel* RBF. Esses estão bem acima do que o estabelecido para este tipo de erro (5%). Além disso, apresentaram altos valores de desvio padrão, indicando que essas duas configurações mostram-se bastante instáveis para a resolução do problema apresentado neste trabalho. Em relação ao tempo de treinamento, o SVM mostrou-se bastante lento para os *kernels* polinomiais de graus 2 e 3, com uma grande desvantagem para o de polinomial de grau 3, tanto em relação a sua rapidez quanto a sua estabilidade. Em contrapartida, para os *kernels* linear e RBF, os resultados mostraram que esses métodos mostraram-se ser relativamente rápidos, quando comparados com o tempo de treinamento das redes neurais. Entretanto, são consideravelmente lentos quando comparados à regressão linear. Levando-se em consideração as métricas de avaliação, o *kernel* polinomial de grau 2 foi a melhor configuração de SVM, por apresentar alto coeficiente de correlação R, baixo RMSE% e possuir menor tempo de treinamento (e mais estável) dentre as configurações que atendem os requisitos de R e RMSE% estabelecidos para este trabalho.

No geral, realizando uma avaliação global das arquiteturas testadas, levando-se

Tabela 3 – Resultados do coeficiente de correlação, erro quadrático médio e tempo de treinamento para o SVM, com *kernels* linear (ou grau 1), polinomial de graus 2 e 3, e RBF.

Método de regressão	Configuração	R		RMSE (%)		Tempo de treinamento(s)	
		Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
SVM	kernel polinomial, p=1	0,95	0,05	26,49	16,47	14,83	5,05
	kernel polinomial, p=2	0,999	0,001	3,09	1,98	174,41	72,15
	kernel polinomial, p=3	1,000	6,818E-05	1,20	0,40	587,02	484,65
	kernel RBF	0,92	0,07	40,49	15,54	36,80	12,02

Fonte: A autora.

em consideração as métricas estabelecidas, observa-se que o MLP com 30 neurônios na única camada escondida foi a arquitetura que obteve melhor desempenho, atendendo às necessidades do problema em questão. Como mencionado nesta seção, a configuração de MLP com 30 neurônios na camada escondida foi a mais bem avaliada, por apresentar alto coeficiente de correlação R, RMSE% abaixo de 5% e tempo de treinamento razoavelmente baixo quando comparado a outros regressores. Apesar do tempo de treinamento não ser um parâmetro crítico neste tipo de problema, ele foi essencial para a escolha do melhor regressor. Afinal, apesar de possuir ótimos coeficientes de correlação e baixíssimos RMSE%, os SVM nas configurações de *kernels* polinomiais de graus 2 e 3, possuem valores muito elevados para o tempo de treinamento, o que leva a um alto custo computacional. Quanto ao regressor que se obteve o pior desempenho, destaca-se a configuração de SVM com *kernel* RBF. Muito embora o tempo de treinamento para esta configuração tenha sido consideravelmente mais baixo do que o tempo de treinamento para o MLP com 30 neurônios na camada escondida, o RMSE% associado a esta configuração apresentou um valor altíssimo, muito além do limite estabelecido de 5%.

Como mencionado no capítulo 4, os conjuntos de predição contendo 15553 instâncias foram utilizados para a validação dos modelos criados a partir dos conjuntos com 4665 instâncias. Os testes dos modelos foram realizados para o melhor e o pior regressor a fim que fosse possível realizar uma comparação qualitativa dos dois métodos de regressão. Na Tabela 4, são apresentados os resultados para o MLP com 30 neurônios na camada escondida, considerando o conjunto de treinamento com 4665 instâncias, e o conjunto de teste, com 15553 instâncias. Também são mostrados os resultados para a configuração do SVM com *kernel* RBF, considerando os mesmos conjuntos. A partir da Tabela 4, podemos verificar que os testes dos modelos gerados pelas redes neurais obtiveram bons desempenhos. Das 18 predições realizadas, apenas duas apresentaram o RMSE% acima de 5%. Sendo estas as predições do primeiro bimestre do ano de 2015 e o quinto bimestre do ano de 2016. Quanto ao coeficiente de correlação, para todos os testes utilizando-se este método, os valores de R foram acima de 0,9. Os testes de validação dos

modelos gerados pelo MLP mostraram-se, portanto, coerentes com o que foi apresentado na Tabela 2. Por outro lado, os testes com os modelos gerados pelo SVM com *kernel* RBF apresentaram um desempenho ruim. Para esses modelos, os coeficientes de correlação, no geral, apresentaram valores acima de 0,9, com exceção das predições do quarto e do quinto bimestre do ano de 2015. Por outro lado, os erros associados a esses modelos atingiram valores altíssimos, obtendo até 73,24% de erro na predição da distribuição dos casos do quarto bimestre do ano de 2014. Esses resultados estão coerentes com o que foi apresentado na Tabela 3, configurando o SVM com *kernel* como um péssimo preditor para o problema em questão.

Tabela 4 – Resultados para a regressão linear e perceptron multicamadas com 30 neurônios na camada interna, considerando os conjuntos de treino com 4665 instâncias e os conjuntos de teste com 15553 instâncias. Como métricas de avaliação foram levadas em consideração o índice de correlação R e o erro RMSE%. Em vermelho, foram destacados os piores resultados.

		MLP, 30 neurons		SVM, kernel = RBF	
		R	RMSE%	R	RMSE%
2014	1	0,9995	4,62%	0,9704	24,58%
	2	0,9994	3,69%	0,9301	38,66%
	3	0,9996	3,30%	0,9246	42,82%
	4	0,9994	4,32%	0,8543	59,87%
	5	1	1,21%	0,9543	38,92%
	6	1	1,04%	0,9862	23,97%
2015	1	0,999	5,87%	0,7966	63,58%
	2	0,9994	4,34%	0,9346	37,68%
	3	0,9998	2,63%	0,9593	32,07%
	4	0,9998	2,00%	0,7674	73,24%
	5	0,9993	4,10%	0,8233	65,89%
	6	0,9992	3,95%	0,9056	43,98%
2016	1	0,9996	3,53%	0,9433	35,94%
	2	0,9998	2,87%	0,9545	34,10%
	3	0,9997	2,73%	0,966	28,56%
	4	0,9997	2,34%	0,9638	28,66%
	5	0,9997	5,40%	0,9777	23,11%
	6	0,9998	1,94%	0,9797	25,04%

Fonte: A autora.

## 5.2 Resultados qualitativos

Os resultados qualitativos foram obtidos através da geração das imagens a partir dos valores dos resultados das validações dos modelos gerados pelos regressores com melhor e pior desempenho. As imagens das predições bimestrais dos anos de 2014 a 2016 foram geradas utilizando-se MLP com 30 neurônios na camada escondida e SVM com *kernel* RBF. As regiões mais quentes das imagens correspondem às áreas com as maiores

concentrações de casos de arboviroses. Por outro lado, as regiões mais frias representam baixos índices dos casos. Os rótulos numéricos dos bairros são apresentados na Tabela 5.

Tabela 5 – Rótulos numéricos dos bairros do Recife.

<b>Bairro</b>	<b>Rótulo</b>	<b>Bairro</b>	<b>Rótulo</b>
Cohab	1	Madalena	17
Ibura	2	Prado	18
Boa Viagem	3	Campo Grande	19
Várzea	4	Alto José Bonifácio	20
Dois Unidos	5	Morro da Conceição	21
Casa Amarela	6	Afogados	22
Imbiribeira	7	Torrões	23
Linha do Tiro	8	Iputinga	24
Macaxeira	9	Areias	25
Nova Descoberta	10	Água Fria	26
Vasco da Gama	11	Torre	27
Ipsep	12	Guabiraba	28
Jordão	13	Brejo de Beberibe	29
Córrego do Jenipapo	14	Caxangá	30
Jardim São Paulo	15	Caçote	31
Cordeiro	16	Pina	32

Fonte: A autora.

A Figura 19 mostra as predições de casos de arboviroses do primeiro bimestre de 2014 utilizando-se o MLP com 30 neurônios na única camada escondida (Figura 19a) e o SVM com *kernel* RBF (Figura 19b). De acordo com a predição utilizando o MLP, observa-se que houve uma maior concentração dos casos de arboviroses na região sul da cidade do Recife. Os bairros mais afetados dessa área foram: o bairro da Cohab, seguido pelo Ibura e Boa Viagem. Na região oeste da cidade, o bairro com maior índice de casos foi o bairro da Várzea. Já na região mais ao norte da cidade, os bairros mais afetados foram Dois Unidos e Casa Amarela. No caso da predição com o SVM, observa-se que a maioria dos casos de arboviroses concentraram-se na região sul da cidade do Recife, compreendendo os bairros da Cohab, Boa Viagem e Ibura. Na região oeste da cidade, o bairro com maior concentração do número de casos foi o bairro da Várzea. Indo em direção à região norte da cidade, a predição indicou que o bairro mais afetado foi Nova Descoberta.

No segundo bimestre desse mesmo ano (Figura 20), o mapa da predição da distribuição dos casos utilizando MLP, Figura 20a, mostra que a região sul da cidade continua sendo a região mais afetada. Porém, nota-se um aumento dos casos de arboviroses no bairro de Boa Viagem e uma diminuição dos casos na Cohab. É possível, também, notar um aumento dos casos no bairro da Imbiribeira em direção ao Pina. Na região norte da cidade, identifica-se um aumento significativo nos bairros de Dois Unidos, Linha do Tiro, Macaxeira, Nova Descoberta e Vasco da Gama. Em relação à predição utilizando SVM com *kernel* RBF

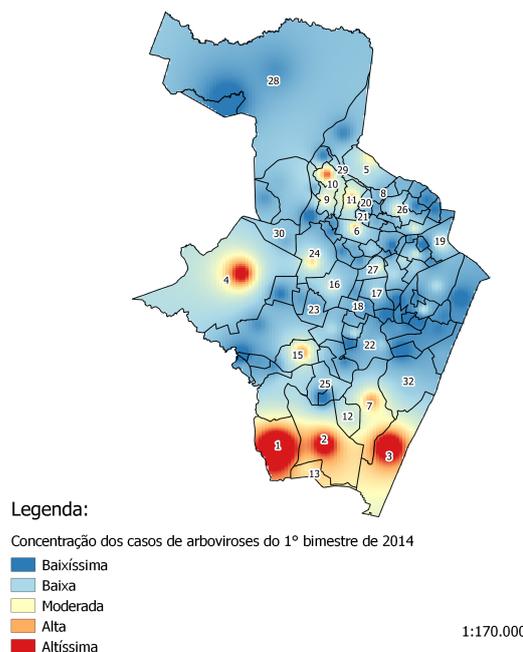
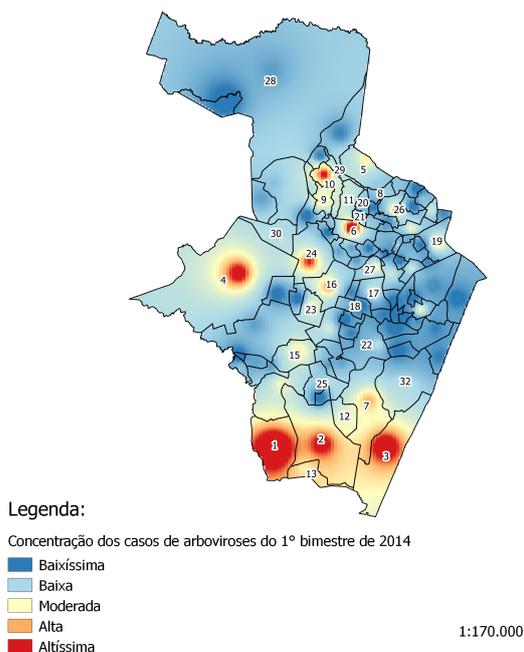
Figura 19 – Resultados da predição da distribuição dos casos de arboviroses do primeiro bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 19a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 19b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição de casos de arboviroses do primeiro bimestre de 2014

Mapa da distribuição dos casos de arboviroses do primeiro bimestre de 2014



(Figura 20b), observa-se uma pequena queda no número de casos nos bairros da Cohab e Jordão, porém, o oposto ocorre nos bairros da Imbiribeira, Ibura e Boa Viagem. Saindo da região sul, indo em direção à região oeste, os bairros mais afetados foram Jardim São Paulo, Várzea, Cordeiro e Iputinga. Na região norte, o número de casos de arboviroses também aumentaram nos bairros da Torre, Vasco da Gama e Nova Descoberta. Mas, também, houve um aumento nos bairros da Macaxeira, Dois Unidos e Água Fria. A predição usando SVM com *kernel* RBF também mostrou um leve aumento nos bairros da Guabiraba e Dois Irmãos (este localizado logo abaixo do bairro da Guabiraba).

No terceiro bimestre de 2014, a predição com MLP (Figura 21a) mostra que na zona sul da cidade há uma diminuição considerável nos casos em Boa Viagem, Ipsep e Imbiribeira. No entanto, o oposto ocorre no bairro do Ibura e da Cohab. Na zona oeste da cidade, a Várzea também mostra uma notável redução dos casos. Por outro lado, na predição usando SVM, do segundo bimestre de 2014 para o terceiro bimestre deste mesmo ano (Figura 21b), as predições usando SVM com *kernel* RBF mostram que a distribuição dos casos de arboviroses apresenta uma redução significativa no bairro da Várzea e Jardim São Paulo e um aumento do número de casos na Iputinga. No bairro da Imbiribeira, localizado na região sul da cidade, também houve uma redução dos casos.

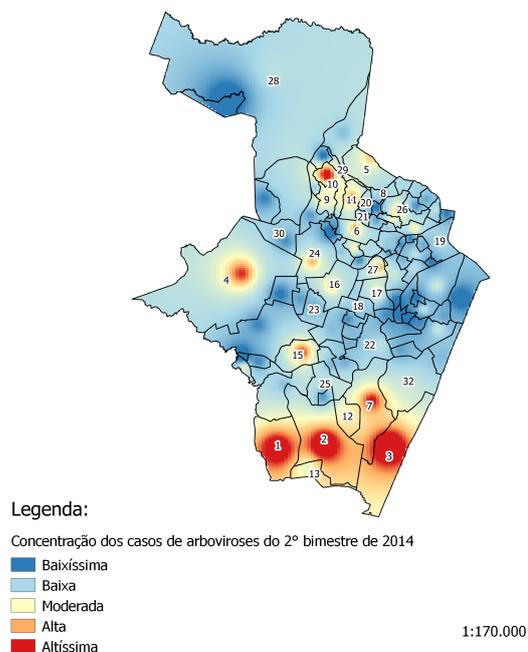
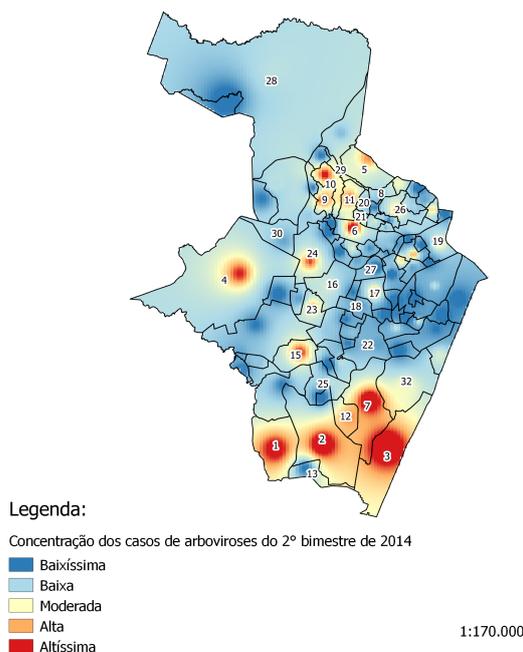
Figura 20 – Resultados da predição da distribuição dos casos de arboviroses do segundo bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 20a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 20b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição de casos de arboviroses do segundo bimestre de 2014

Mapa da distribuição dos casos de arboviroses do segundo bimestre de 2014



Do terceiro bimestre para o quarto bimestre (Figura 22) deste mesmo ano, a predição com SVM mostra uma intensificação e maior concentração do número de casos na região sul da cidade (Figura 22b). Com destaque, principalmente, para os bairros da Cohab, Ibura e Jordão. Na Várzea, o aumento de casos também foi bastante considerável de acordo com este método de predição. Já na predição utilizando-se MLP, o mapa de predição da distribuição dos casos (Figura 22a) mostra, no geral, uma redução significativa em relação ao bimestre anterior, principalmente, na região sul da cidade, havendo um agravamento dos casos apenas no bairro do Jordão. No quinto bimestre de 2014 (Figura 23), a predição com a máquina de vetores de suporte mostra que houve uma diminuição dos casos em relação ao bimestre anterior. Os bairros que tiveram grandes concentrações do número de casos de arboviroses foram os bairros do Ibura e Jordão. Quanto à predição utilizando as redes neurais, os resultados apontaram que a distribuição dos casos de arboviroses exibiram comportamento semelhantes ao comportamento do bimestre anterior, conforme aponta o mapa da predição dos casos de arboviroses do quinto bimestre de 2014 (Figura 23a).

No último bimestre de 2014, comparando com os mapas da Figura 23b, nota-se apenas uma pequena diminuição no número de casos no bairro da Cohab, no entanto, mesmo com essa diminuição, este bairro ainda teve um número elevado do número de

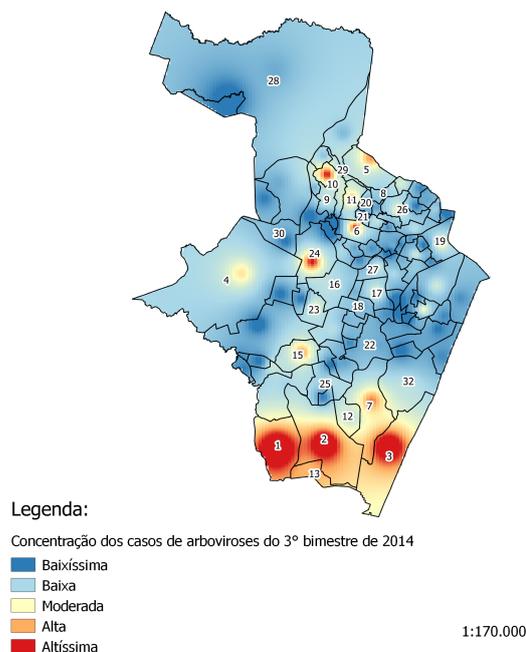
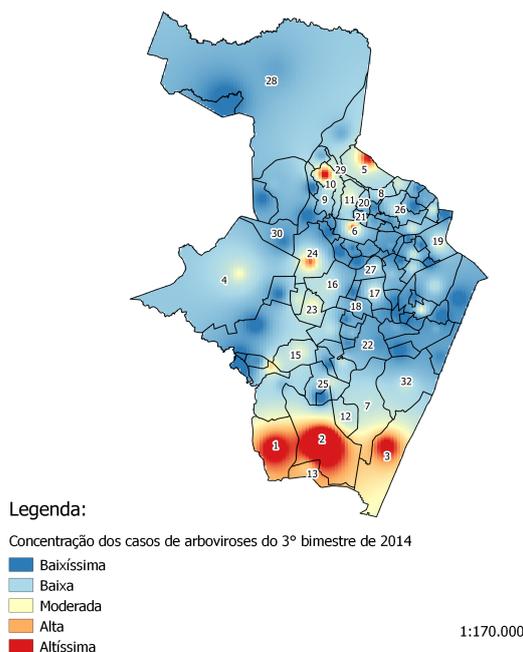
Figura 21 – Resultados da predição da distribuição dos casos de arboviroses do terceiro bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 21a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 21b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição de casos de arboviroses do terceiro bimestre de 2014

Mapa da distribuição dos casos de arboviroses do terceiro bimestre de 2014



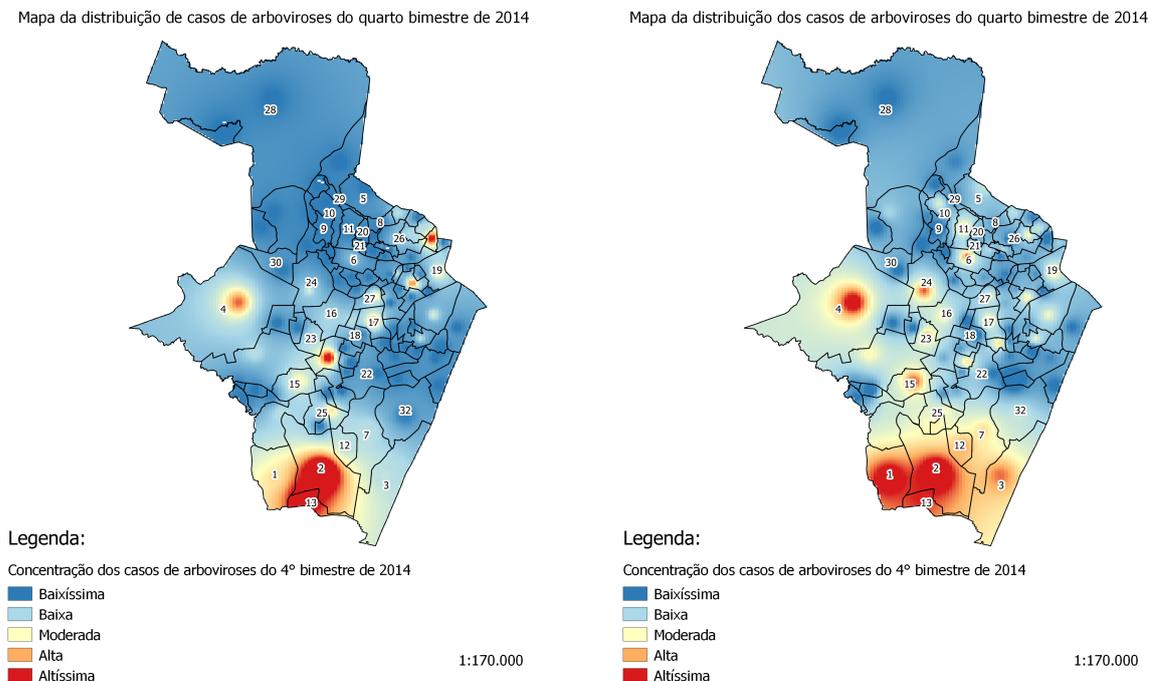
casos, como mostra a predição com SVM (Figura 24b). Já o comportamento para o bairro da Várzea foi bastante semelhante para os dois últimos bimestres deste mesmo ano, quando se compara os mapas 23b e 24b. Em contrapartida, a predição com MLP mostrou que a distribuição dos casos exibiu um comportamento bastante semelhante ao bimestre anterior, em que os bairros da Cohab e Jordão foram os que mais sofreram com os casos das doenças.

Do último bimestre de 2014 (Figura 24a) para o primeiro bimestre de 2015 (Figura 25a), a predição utilizando MLP com 30 neurônios na camada escondida mostrou um grande aumento dos casos de arboviroses na região norte do Recife. Os bairros afetados foram, principalmente, Córrego do Jenipapo, Casa Amarela, Vasco da Gama, Nova Descoberta e Dois Unidos. A zona sul apresentou uma considerável diminuição nos casos de arboviroses, tendo uma altíssima incidência apenas no bairro da Cohab. No bairro da Várzea, nota-se uma pequena diminuição da quantidade de casos quando comparado ao bimestre anterior. Observando a Figura 25b, percebe-se que a predição utilizando SVM com *kernel* RBF também indicou que houve um agravamento da situação, principalmente, nas regiões norte da cidade do Recife. Quanto à região norte, para o bimestre em questão, as populações dos bairros da Macaxeira, Brejo de Beberibe, Dois Unidos e Casa Amarela foram as mais

Figura 22 – Resultados da predição da distribuição dos casos de arboviroses do quarto bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 22a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 22b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

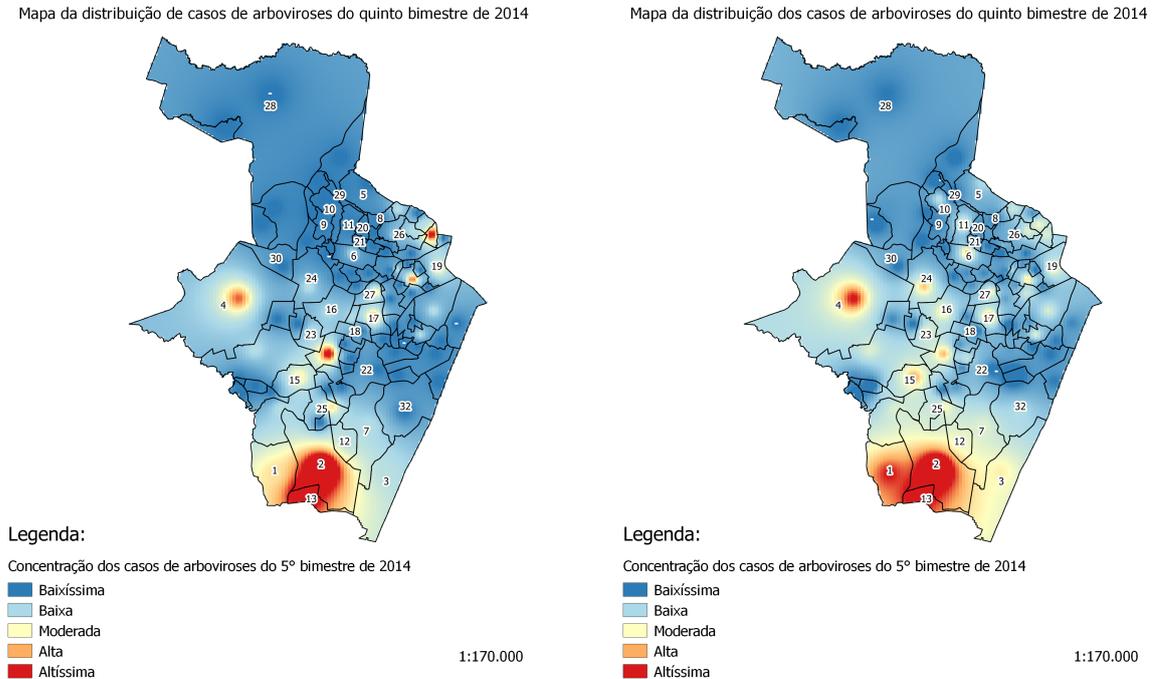


afetadas pelas arboviroses. Mas, também, é possível notar um aumento significativo na quantidade de casos no bairro da Guabiraba e Dois Irmãos. Saindo da região norte e descendo em direção à região oeste da cidade, a predição aponta que houve um aumento significativo no número de casos nos bairros da Iputinga, Torrões, Caxangá e Várzea. Partindo do bairro da Várzea em direção à região sul da cidade, o bairro mais afetado foi o de Jardim São Paulo. No caso da região sul, as populações dos bairros do Ibura, Jordão, Ipsep e Cohab foram as mais afetadas. No segundo bimestre deste mesmo ano, Figura 26a, a predição com MLP aponta uma redução bastante significativa na quantidade de casos na região norte da cidade. Nesse bimestre, a grande maioria dos casos ficou concentrada nos bairros da zona sul, onde houve um aumento dos casos de arboviroses na Cohab, Boa Viagem, Ibura, Jordão e Imbiribeira. Na região oeste, houve um aumento significativo de casos nos bairros da Várzea e Iputinga. Partindo da Várzea, descendo em direção à região sul da cidade, destaca-se o bairro de Areias. O resultado obtido com o SVM, exibe uma notável diminuição na quantidade de casos nas regiões oeste e norte da cidade. No entanto, na região sul, a quantidade de casos manteve-se elevada, havendo um considerável aumento nos bairros da Cohab, Ibura, Jordão e Boa Viagem e uma pequena diminuição no bairro da Imbiribeira.

Figura 23 – Resultados da predição da distribuição dos casos de arboviroses do quinto bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 23a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 23b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.



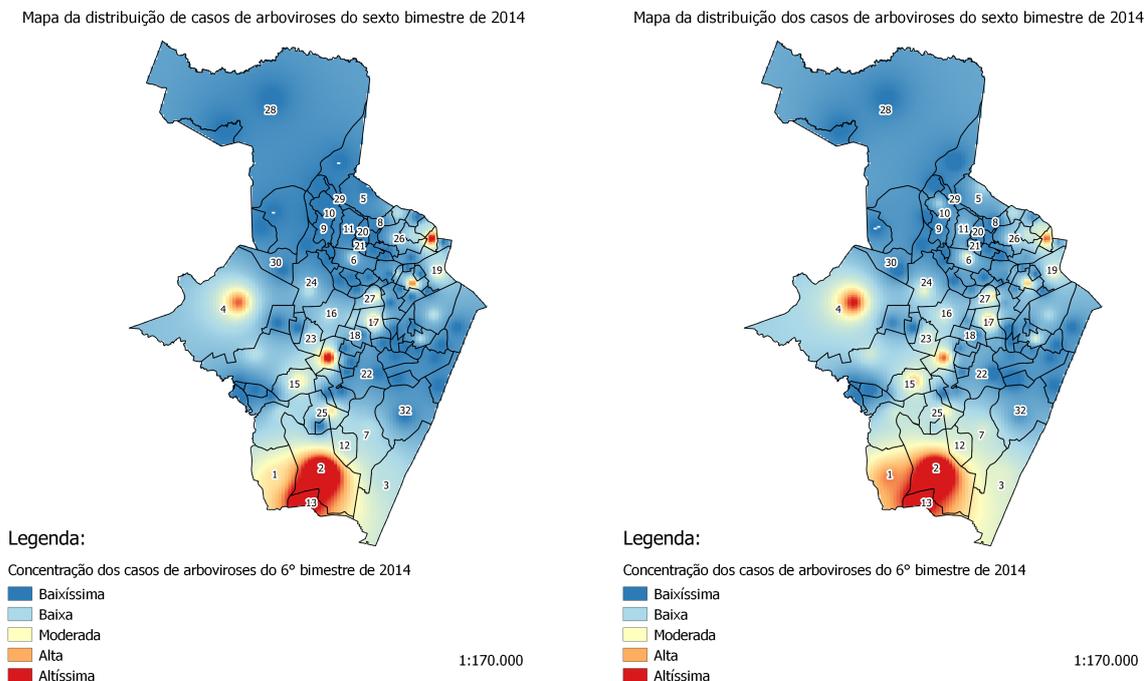
A Figura 27 exhibe os mapas das predições das distribuições dos casos de arboviroses para o terceiro bimestre do ano de 2015 com MLP e SVM. Neste bimestre, comparando o mapa da distribuição do segundo bimestre deste mesmo ano, a predição utilizando MLP aponta que a distribuição dos casos apresentou um comportamento semelhante ao bimestre anterior, como pode ser observado na Figura 26a. Existe, no entanto, um aumento na concentração de casos no bairro de Jardim São Paulo. Já os bairros de Areias, Imbiribeira e Iputinga, apresentaram uma diminuição bastante significativa na quantidade de casos. Quanto à predição da distribuição dos casos utilizando o SVM com *kernel* RBF, comparando os mapas 26b e 27b, percebe-se que houve uma diminuição dos casos nos bairros da Imbiribeira, Boa Viagem e Ipsep. Do terceiro para o quarto bimestre, ambas as predições exibem que a situação fica mais controlada com uma grande diminuição da quantidade de casos em grande parte do território recifense. Para a predição com MLP (Figura 28a), o mapa aponta que os casos de arboviroses ficaram concentrados nos bairros do Ibura e da Cohab. Entretanto, na predição com SVM, os casos de arboviroses também se concentraram na zona sul da cidade, nos bairros do Jordão, Ibura e Cohab, conforme apresentado no mapa da Figura 28b.

No bimestre seguinte, o quinto bimestre de 2015, comparando os mapas da Figura

Figura 24 – Resultados da predição da distribuição dos casos de arboviroses do sexto bimestre de 2014. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 24a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 24b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.



28 e da Figura 29, percebe-se que os casos de arboviroses voltam a aumentar. De acordo com a predição usando MLP, o aumento dos casos deu-se em alguns pontos específicos da cidade. No caso do bimestre 5, os bairros localizados mais ao centro da cidade tiveram um aumento significativo dos casos, com destaque para o bairro do Cordeiro, Madalena e Prado. Na região nordeste da cidade, a maior concentração de casos foi no bairro de Campo Grande. A predição com SVM apresentou um agravamento mais intenso em relação ao aumento da quantidade dos casos de arboviroses, com um aumento significativo do número de casos nos bairros da Imbiribeira, Boa Viagem, Pina, Várzea, Afogados e nos bairros próximos ao bairro de Areias. Já no bairro do Jordão, nota-se uma pequena diminuição da concentração dos casos, muito embora esta permaneça, ainda, muito elevada, conforme o mapa exibido na Figura 29b.

As predições dos dois primeiros bimestres de 2016 (Figuras 30a e 31a), usando MLP com 30 neurônios na camada escondida, apresentaram comportamentos bastante semelhantes. De acordo com as imagens, as regiões mais afetadas nesses dois bimestres foram as regiões sul e oeste da cidade. Na região sul, as maiores concentrações dos casos foram detectados na Cohab e Ibura, e com menos intensidade nos bairros do Ipsep, Boa Viagem e Imbiribeira. Na zona oeste da cidade, os bairros mais afetados foram a Várzea

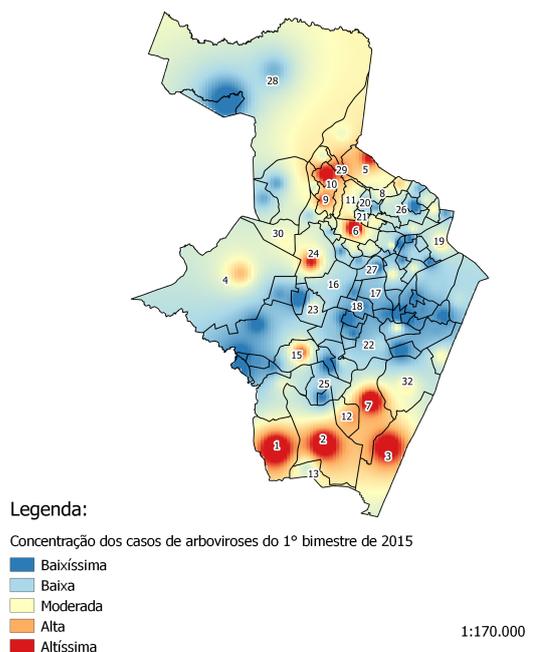
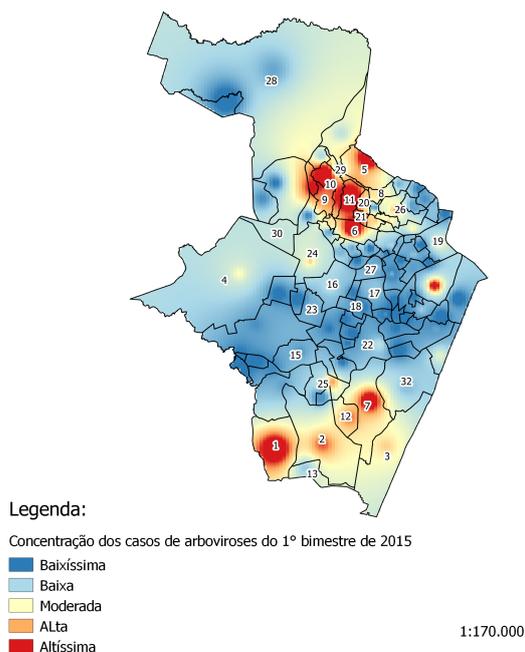
Figura 25 – Resultados da predição da distribuição dos casos de arboviroses do primeiro bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 25a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 25b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do primeiro bimestre de 2015

Mapa da distribuição dos casos de arboviroses do primeiro bimestre de 2015



e Iputinga. Os dois primeiros bimestres do ano de 2016 usando SVM, Figuras 30b e 31b, foram marcados por um aumento nos casos de arboviroses na região oeste; nos bairros da Várzea e Iputinga; e uma diminuição razoável nos bairros do Ibura, Imbiribeira, Jordão e Boa Viagem, Caçote e Afogados. Por outro lado, nesses dois bimestres, a distribuição dos casos no bairro da Cohab permaneceu quase constante. De acordo com a predição usando a rede neural, no terceiro bimestre de 2016 (Figura 32a), houve uma considerável diminuição dos casos na região sul da cidade, concentrando-se apenas nos bairros da Cohab e Ibura. Na zona oeste da cidade, também é possível perceber uma redução dos casos no bairro da Várzea. No entanto, nota-se que os casos nos bairros da Iputinga, Cordeiro e Torrões intensificaram-se. Para a predição com a máquina de vetor de suporte, Figura 32b, tantos nos bairros da região sul quanto nos bairros da região oeste houve uma redução da concentração dos casos, porém, tanto para o segundo bimestre quanto para o terceiro bimestre a concentração de casos das doenças apresentou um comportamento semelhante.

No quarto bimestre de 2016, a situação em relação aos casos de arboviroses volta a se agravar como mostra a Figura 33. De acordo com o que foi predito pelo MLP, os casos concentraram-se, principalmente, na região sul em direção à região oeste da

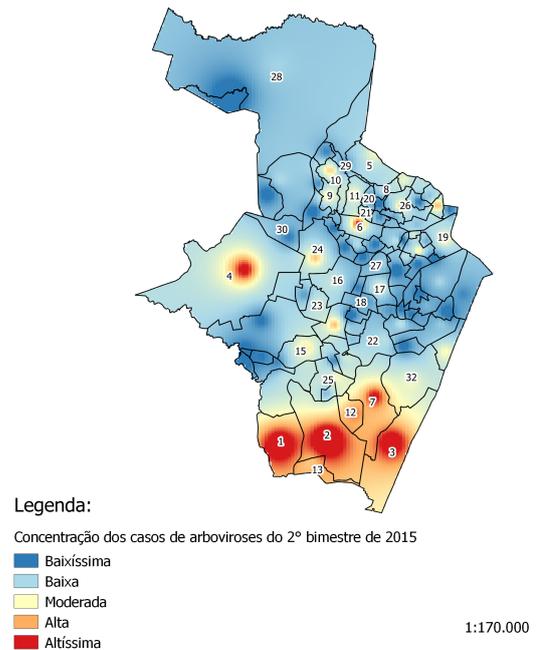
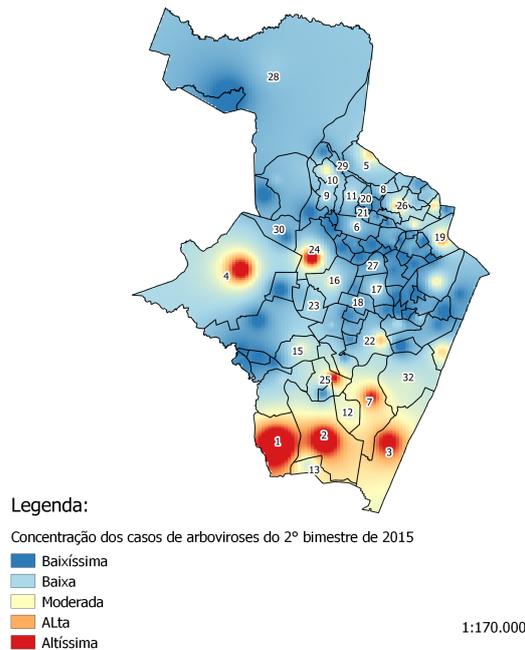
Figura 26 – Resultados da predição da distribuição dos casos de arboviroses do segundo bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 26a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 26b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do segundo bimestre de 2015

Mapa da distribuição dos casos de arboviroses do segundo bimestre de 2015



cidade. Os bairros mais afetados foram Cohab, Imbiribeira, Jordão, Boa Viagem, Areias, Várzea, Iputinga e Cordeiro. A predição com SVM indica que houve um aumento bastante significativo na região sul, compreendendo os bairros de Boa Viagem, Imbiribeira, Cohab, Ibura, Ipsep. Partindo da Imbiribeira, indo em direção ao norte, os bairros do Pina e Afogados também exibiram um aumento na concentração dos casos de arboviroses quando comparado ao bimestre anterior. No quinto bimestre de 2016 (Figura 34a), o comportamento da distribuição dos casos de arboviroses foi bastante semelhante ao bimestre anterior, como previsto pelo perceptron multicamadas com 30 neurônios na única camada escondida. Há uma diminuição dos casos em alguns bairros da região sul, como é o caso dos bairros do Ibura e Imbiribeira. Em contraponto, o bairro de Boa Viagem apresentou um aumento significativo da quantidade de casos do quarto para o quinto bimestre. De acordo com a predição usando SVM, o quinto bimestre mostrou uma redução bastante significativa para os bairros de Boa Viagem, Imbiribeira e Ibura. Indo em direção a região oeste da cidade, a predição mostrou um aumento razoável para os bairros desta região. Por fim, no último bimestre de 2016 (Figura 35a), a predição com MLP mostrou que houve uma melhora na situação referente à concentração dos casos. A região sul apresentou uma redução significativa nos casos de arboviroses, com exceção do bairro da Cohab. Na região oeste

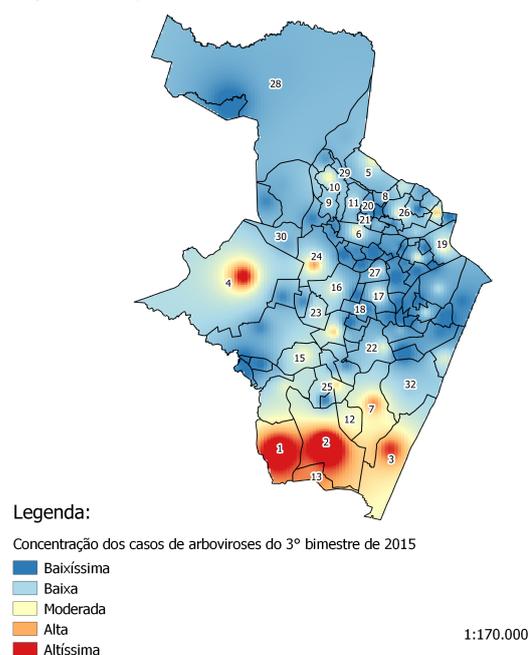
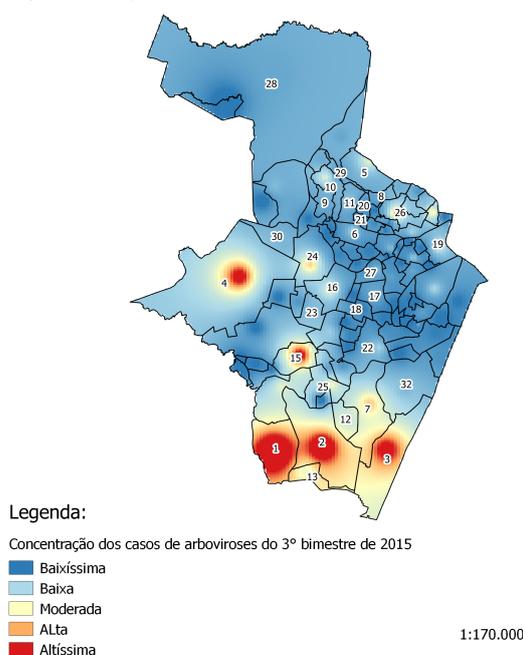
Figura 27 – Resultados da predição da distribuição dos casos de arboviroses do terceiro bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 27a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 27b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do terceiro bimestre de 2015

Mapa da distribuição dos casos de arboviroses do terceiro bimestre de 2015



da cidade também se observa uma diminuição significativa dos casos, com exceção dos bairros da Várzea e Iputinga. Neste último bimestre de 2016, a predição com SVM também indicou uma diminuição, no geral, dos casos de arboviroses (como pode ser observado na Figura 35b). Houve redução do número de casos na grande maioria dos bairros com exceção dos bairros da Várzea, Iputinga, Cordeiro e Torrões, localizados na região oeste da cidade, e do bairro da Cohab, cuja quantidade de casos de arboviroses permaneceu praticamente constante ao longo de todo ano de 2016.

No geral, os mapas de predição utilizando MLP com 30 neurônios, mostraram que as principais regiões do Recife com grande concentração casos são as regiões oeste e sul. Na região oeste, o bairro que aparece com maior frequência quanto à altíssima concentração dos casos é o bairro da Várzea. Já na região sul, os bairros que aparecem com maior frequência à altíssima concentração dos casos são os bairros da Cohab, Iburá, Imbiribeira e Boa Viagem. Os mapas de predição apresentados assemelham-se aos mapas das distribuições reais dos casos de arboviroses. Isto corrobora com as métricas de análise quantitativa deste método de regressão. As predições usando SVM com *kernel* mostraram que as regiões do Recife que mais sofrem com os casos de arboviroses são as regiões sul e oeste. Na região sul, aparecem com grande frequência os bairros da Cohab, Imbiribeira,

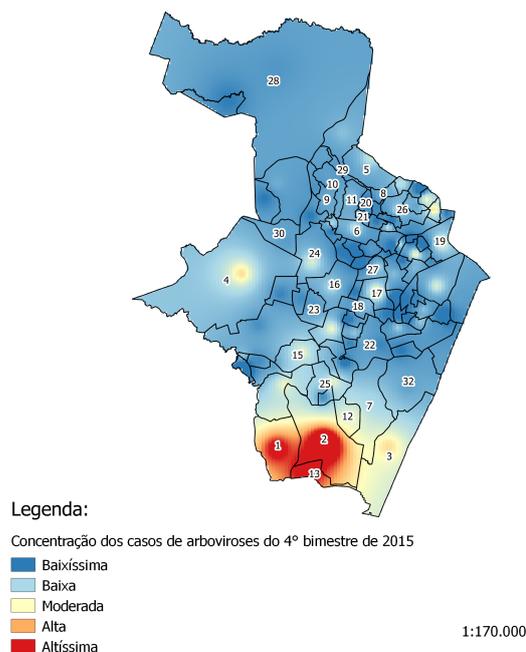
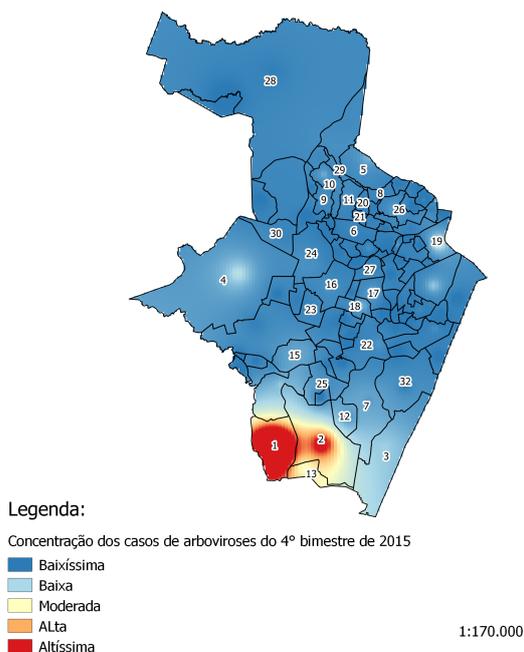
Figura 28 – Resultados da predição da distribuição dos casos de arboviroses do quarto bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 28a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 28b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do quarto bimestre de 2015

Mapa da distribuição dos casos de arboviroses do quarto bimestre de 2015



Ibura, Boa Viagem e Jordão. Na região oeste, o bairro da Várzea é que aparece com maior frequência, apresentando uma elevada concentração de casos de arboviroses. No entanto, os mapas das predições usando SVM com *kernel* RBF, quando comparados com as distribuições reais, muitas vezes acentuam regiões em que, para aquele bimestre não foram alarmantes. Como, por exemplo, no quarto bimestre de 2014 (Figura 22b), as predições com SVM (*kernel* RBF) mostraram que houve uma grande quantidade de casos para quase todos os bairros da região sul, quando, na verdade, os bairros cujas situações foram bastante alarmante foram os bairros do Jordão e Ibura. O contrário também ocorre, ou seja, em alguns bairros as predições com SVM (*kernel* RBF) mostraram que não houve muitos casos de arboviroses, quando, no mapa de distribuição original, indica que a quantidade de casos foi elevada. Os resultados qualitativos estão, portanto, de acordo com os resultados quantitativos apresentados na seção 5.1, a qual apresenta este regressor como o pior regressor avaliado (dentro um conjunto de regressores testados) para a resolução deste problema.

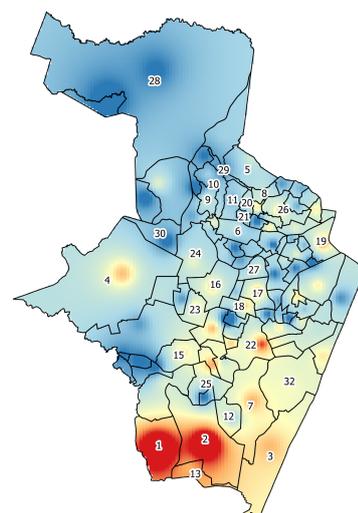
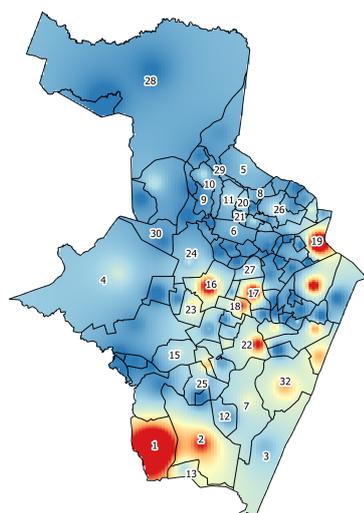
Figura 29 – Resultados da predição da distribuição dos casos de arboviroses do quinto bimestre de 2015. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 29a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 29b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do quinto bimestre de 2015

Mapa da distribuição dos casos de arboviroses do quinto bimestre de 2015



Legenda:

Concentração dos casos de arboviroses do 5º bimestre de 2015

- Baixíssima
- Baixa
- Moderada
- ALta
- Altíssima

1:170.000

Legenda:

Concentração dos casos de arboviroses do 5º bimestre de 2015

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

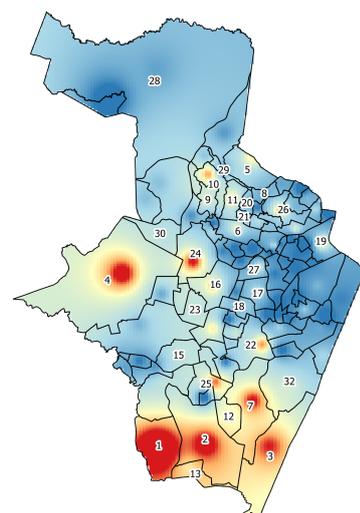
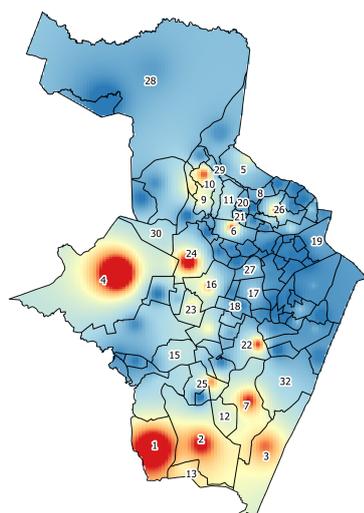
Figura 30 – Resultados da predição da distribuição dos casos de arboviroses do primeiro bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 30a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 30b).

(a) Predição utilizando MLP.

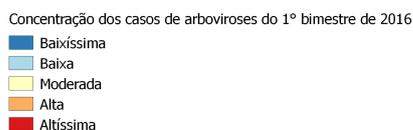
(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do primeiro bimestre de 2016

Mapa da distribuição dos casos de arboviroses do primeiro bimestre de 2016

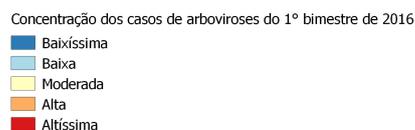


Legenda:



1:170.000

Legenda:



1:170.000

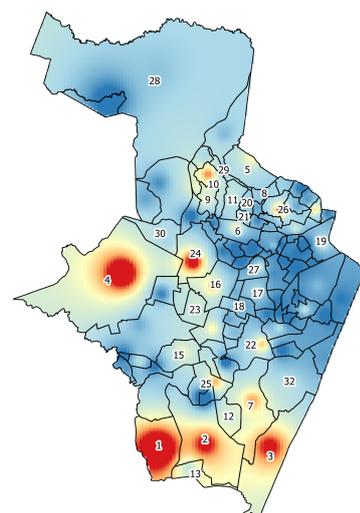
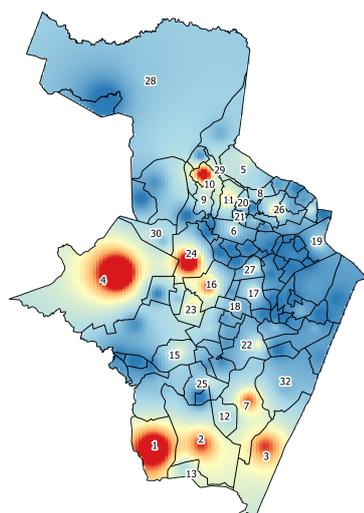
Figura 31 – Resultados da predição da distribuição dos casos de arboviroses do segundo bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 31a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 31b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do segundo bimestre de 2016

Mapa da distribuição dos casos de arboviroses do segundo bimestre de 2016



Legenda:

Concentração dos casos de arboviroses do 2º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

Legenda:

Concentração dos casos de arboviroses do 2º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

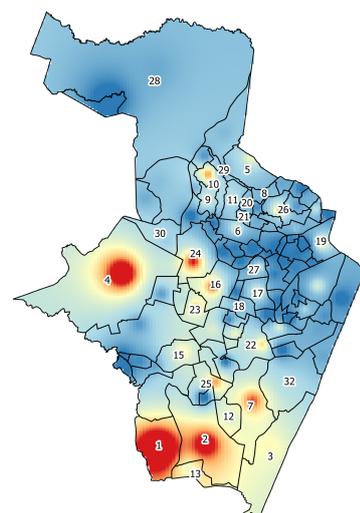
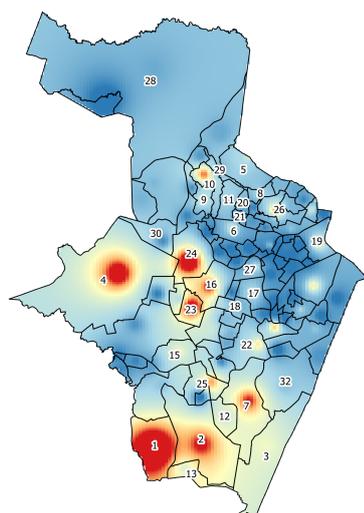
Figura 32 – Resultados da predição da distribuição dos casos de arboviroses do terceiro bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 32a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 32b).

(a) Predição utilizando MLP.

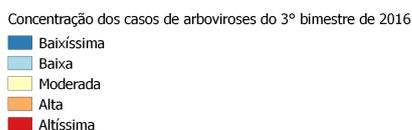
(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do terceiro bimestre de 2016

Mapa da distribuição dos casos de arboviroses do terceiro bimestre de 2016

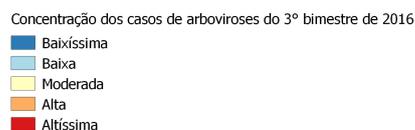


Legenda:



1:170.000

Legenda:



1:170.000

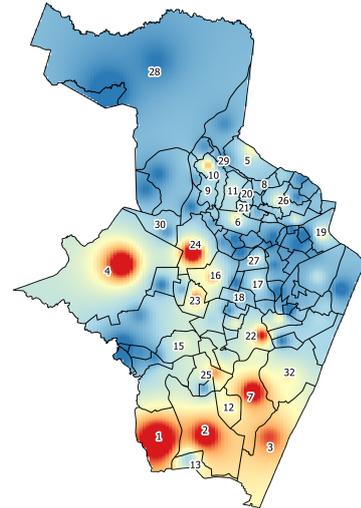
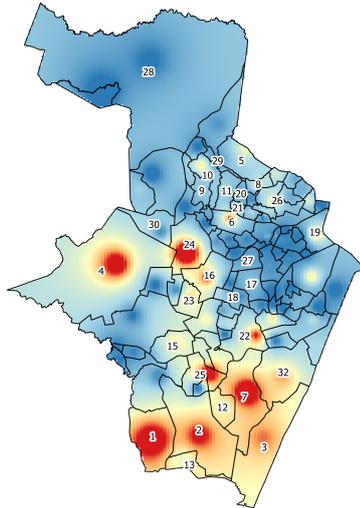
Figura 33 – Resultados da predição da distribuição dos casos de arboviroses do quarto bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 33a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 33b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do quarto bimestre de 2016

Mapa da distribuição dos casos de arboviroses do quarto bimestre de 2016



Legenda:

Concentração dos casos de arboviroses do 4º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

Legenda:

Concentração dos casos de arboviroses do 4º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

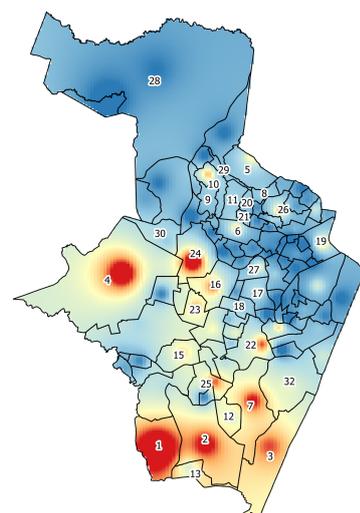
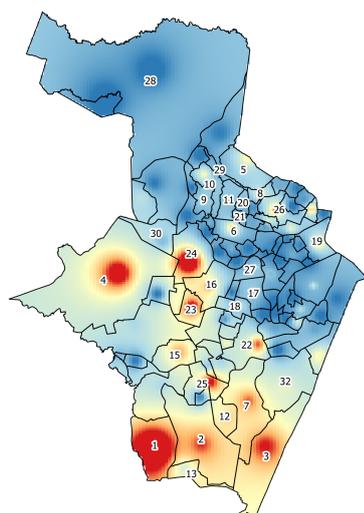
Figura 34 – Resultados da predição da distribuição dos casos de arboviroses do quinto bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 34a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 34b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do quinto bimestre de 2016

Mapa da distribuição dos casos de arboviroses do quinto bimestre de 2016



Legenda:

Concentração dos casos de arboviroses do 5º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

Legenda:

Concentração dos casos de arboviroses do 5º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

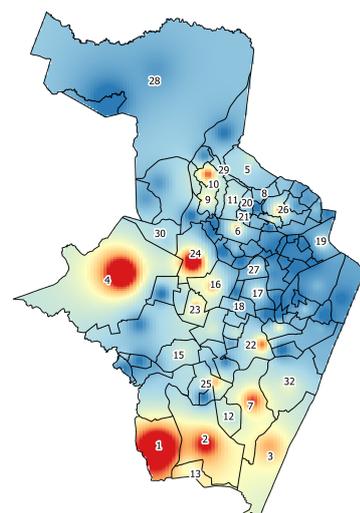
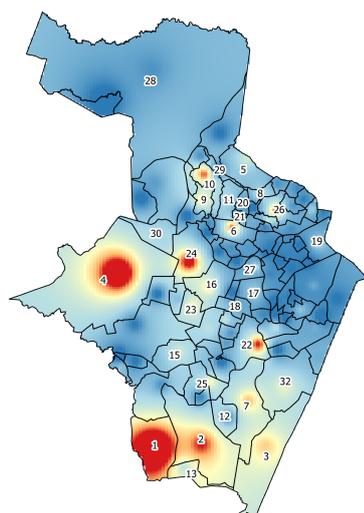
Figura 35 – Resultados da predição da distribuição dos casos de arboviroses do sexto bimestre de 2016. À esquerda, tem-se o mapa da predição da distribuição dos casos de arboviroses utilizando MLP com 30 neurônios na última camada escondida (Figura 34a). À direita, tem-se o mapa da predição dos casos de arboviroses utilizando SVM com *kernel* RBF (Figura 34b).

(a) Predição utilizando MLP.

(b) Predição utilizando SVM.

Mapa da distribuição dos casos de arboviroses do sexto bimestre de 2016

Mapa da distribuição dos casos de arboviroses do sexto bimestre de 2016



Legenda:

Concentração dos casos de arboviroses do 6º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

Legenda:

Concentração dos casos de arboviroses do 6º bimestre de 2016

- Baixíssima
- Baixa
- Moderada
- Alta
- Altíssima

1:170.000

## 6 CONCLUSÃO

### 6.1 Conclusão geral

As arboviroses são zoonoses transmitidas por arbovírus através de artrópodes hematófagos. O Brasil é um país que, historicamente, sofre com ciclos de epidemias de arboviroses em grande parte do seu território e, a partir de 2014, este cenário vem se agravando com a introdução do vírus *chikungunya* e zika.

As técnicas de geoprocessamento permitem avaliar a distribuição espacial de uma determinada doença em uma determinada região, possibilitando estudos sobre sua transmissibilidade nas regiões em que são abrangidas, fornecendo informações essenciais para a realização da vigilância dos casos. Essas técnicas associadas a técnicas de aprendizagem de máquina podem auxiliar na predição de casos das doenças, o que podem servir de apoio para os agentes de saúde na tomada de decisão de medidas preventivas.

Este trabalho teve como proposta o desenvolvimento de uma ferramenta de predição de casos de arboviroses transmitidas pelo *Aedes aegypti*. Para tanto, foram obtidos dados a respeito dos casos confirmados de dengue, *chikungunya* e zika, através do Portal de Dados Abertos da Prefeitura do Recife; dados a respeito das variáveis climáticas - temperatura, velocidade dos ventos - por meio da base de dados do Instituto Nacional de Meteorologia; e os dados referentes aos índices pluviométricos da cidade do Recife, por meio do banco de dados do Sistema de Geoinformação Hidrometeorológica de Pernambuco. Para a predição da distribuição dos casos, foram avaliados os seguintes métodos de regressão: a regressão linear, as redes neurais artificiais e máquina de vetores de suporte para regressão.

O uso de preditores baseados em algoritmos de aprendizagem de máquina provou ser muito eficaz na predição da distribuição de casos de zoonoses transmitidas por arbovírus. Como mostraram os resultados quantitativos, apresentados no capítulo 5, o método de regressão que apresentou o melhor desempenho foi o MLP, com 30 neurônios na camada escondida, resultando em um alto coeficiente de correlação, baixo RMSE% (em média 3,29%) e um tempo de treinamento razoavelmente baixo, quando comparado a outros regressores avaliados. De acordo com os resultados qualitativos, as regiões em que predominam os surtos de arbovírus transmitidos pelo *Aedes aegypti* são as regiões sul e oeste do Recife. Na região oeste, o bairro que aparece com mais frequência em relação à maior concentração de casos é o bairro da Várzea. Na região sul, os bairros que aparecem com mais frequência com alta/altíssima concentração de casos são os bairros da Cohab, Ibura, Imbiribeira e Boa Viagem. Já a região norte da cidade aparece com alta concentração de casos nos dois primeiros trimestres do ano. Embora existam casos ao longo do ano, também foi observado que os casos arboviroses, em geral, ocorrem, predominantemente,

nos meses mais quentes do ano (outubro a março).

Além disso, a abordagem usando análise espaço-temporal forneceu uma avaliação mais ampla de regiões onde estão localizadas as maiores e menores concentrações de casos de arboviroses. A partir dos resultados qualitativos, foi possível diferenciar nos mapas de calor as regiões com alta concentração de casos das regiões com baixa concentração e as regiões que estão no intervalo de transição. Esse tipo de abordagem é muito relevante no apoio a gestores de saúde e epidemiologistas no planejamento de ações de curto e médio prazo para prevenir surtos de arbovírus transmitidos pelo *Aedes aegypti*, além de apoiar o desenvolvimento de políticas públicas de promoção de saneamento e da saúde.

## 6.2 Dificuldades apresentadas

As dificuldades encontradas neste trabalho foram relacionadas à organização dos dados para a elaboração dos conjuntos de treinamento e teste. Nos dados referentes aos casos confirmados de dengue, *chikungunya* e zika, houve uma dificuldade no georreferenciamento dos endereços dos casos. Os campos de latitude e longitude não são fornecidos na base de dados do Portal de Dados Abertos da Prefeitura do Recife e, no momento de preencher esses campos, ou os logradouros não eram reconhecidos ou continham muito erro de digitação, ou abreviações, o que dificultou bastante o uso de ferramentas de georreferenciamento. Por isso, optou-se por trabalhar com a quantidade de casos em cada bairro da Cidade do Recife.

Uma outra dificuldade encontrada na realização deste trabalho foi o fato de haver poucas estações de medição, tanto para temperatura quanto para velocidade do vento e pluviometria, ao longo do território recifense. Esta pouca quantidade de estações resulta em uma pequena quantidade de pontos para a elaboração de seus respectivos mapas de distribuição. A utilização da distribuição gaussiana para estimar os valores das variáveis climáticas para cada bairro da Cidade do Recife foi uma alternativa para contornar este obstáculo, abrindo possibilidades de erros inerentes ao processo. A instalação de estações de medição das variáveis climáticas em outros pontos da cidade pode trazer resultados mais fidedignos para a elaboração dos mapas de distribuição e, possivelmente, melhorar a performance de alguns preditores.

## 6.3 Contribuições

Como principais contribuições deste trabalho, podemos citar:

- A elaboração de uma ferramenta para a predição da distribuição de casos de arboviroses, transmitidas pelo mosquito *Aedes aegypti*, a partir de mapas de distribuição dos casos das doenças e fatores ambientais (temperatura, chuva e velocidade dos ventos).

- Desenvolvimento de um sistema para mineração dos dados da Plataforma de Dados Abertos da Prefeitura do Recife e extração de informações geográficas e de interesse epidemiológico úteis para a modelagem e validação dos modelos.
- A investigação do desempenho de técnicas de aprendizagem de máquina para a construção de um preditor dinâmico dos casos de arboviroses para a modelagem dinâmica populacional humana na área sob estudo, no tempo e no espaço.

A ferramenta desenvolvida, portanto, faz uso de técnicas de mineração de dados abertos aliado às técnicas de geoprocessamento e aprendizado de máquina para a simulação da distribuição espacial das arboviroses. O presente trabalho pode ter um grande impacto para o Sistema Único de Saúde, visto que pode ser utilizado como ferramenta de apoio para os gestores de saúde e epidemiologistas, no planejamento de políticas públicas de curto, médio e longo prazo para a promoção da saúde da população. Além de auxiliar nas tomadas de decisão para que os recursos humanos e financeiros sejam distribuídos de forma eficaz para as regiões mais necessitadas. E, como consequência, promover a melhoria da qualidade de vida da população nessas regiões.

#### 6.4 Trabalhos futuros

Como visto nos resultados, as técnicas de aprendizagem de máquina mostraram-se eficientes na predição de casos de arboviroses na cidade do Recife. No entanto, os modelos utilizados não levam em consideração os índices socioeconômicos e nem a população de mosquitos da espécie *Aedes aegypti*. Recomenda-se, portanto, estudos mais profundos levando em consideração esses dois fatores.

Além disso, como visto, o Brasil é um país vasto e com uma grande população de mosquitos que transmitem vários tipos de arboviroses, estando sujeito a surtos não só de doenças transmitidas pelo *Aedes aegypti*, mas também por doenças transmitidas por outros tipos de vetores. Desta forma, devido a sua grande extensão territorial a vigilância e o planejamento do controle dos vetores transmissores de arboviroses torna-se um grande desafio. Sendo assim, recomenda-se que este estudo seja estendido para a predição da distribuição de casos de arboviroses transmitidas por outras espécies de vetores.

## REFERÊNCIAS

- AKIL, L.; AHMAD, H. A. Salmonella infections modelling in mississippi using neural network and geographical information system (gis). *BMJ Open*, British Medical Journal Publishing Group, v. 6, n. 3, 2016. ISSN 2044-6055. Disponível em: <https://bmjopen.bmj.com/content/6/3/e009255>.
- ALBRIEU-LLINÁS, G.; ESPINOSA, M. O.; QUAGLIA, A.; ABRIL, M.; SCAVUZZO, C. M. Urban environmental clustering to assess the spatial dynamics of aedes aegypti breeding sites. *Geospatial Health*, v. 13, n. 1, p. 135–142, 2018.
- ALLEN, C.; TSOU, M.-H.; ASLAM, A.; NAGEL, A.; GAWRON, J.-M. Applying gis and machine learning methods to twitter data for multiscale surveillance of influenza. *PLOS ONE*, Public Library of Science, v. 11, n. 7, p. 1–10, 07 2016. Disponível em: <https://doi.org/10.1371/journal.pone.0157734>.
- Ana Carolina C. Costa; Cláudia T. Codeço; Nildimar A. Honório; Gláucio R. Pereira; AND Aline A. Nobre, C. Surveillance of dengue vectors using spatio-temporal bayesian modeling. *BCM Medical Informatics an Decision Making*, Springer, v. 15, 2015.
- APAC. *Agência Pernambucana de Águas e Climas. Sobre o Portal*. 2018. Disponível em: <https://www.apac.pe.gov.br/institucional>. Acesso em: 10 de nov. de 2018.
- AZEVEDO, R. do Socorro da S.; OLIVEIRA, C. S.; VASCONCELOS, P. F. da C. Risco do chikungunya para o brasil. *Revista de Saúde Pública*, SciELO, p. 49–51, 2015.
- AZEVEDO, W. W.; LIMA, S. M.; FERNANDES, I. M.; ROCHA, A. D.; CORDEIRO, F. R.; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. Fuzzy morphological extreme learning machines to detect and classify masses in mammograms. In: IEEE. *2015 IEEE international conference on fuzzy systems (fuzz-IEEE)*. [S.l.], 2015. p. 1–8.
- BAKKER, K. M.; MARTINEZ-BAKKER, M. E.; HELM, B.; STEVENSON, T. J. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *PNAS*, n. 24, p. 6689–6694, 06 2016. Disponível em: [www.pnas.org/cgi/doi/10.1073/pnas.1523941113](http://www.pnas.org/cgi/doi/10.1073/pnas.1523941113).
- BAQUERO, O. S.; SANTANA, L. M. R.; CHIARAVALLOTI-NETO, F. Dengue forecasting in são paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLOS ONE*, Public Library of Science, v. 13, n. 4, p. 1–12, 04 2018. Disponível em: <https://doi.org/10.1371/journal.pone.0195065>.
- BEENSOON, S.; FUNKHOUSER, E.; KOTEA, N.; SPIELMAN, A.; ROBICH, R. M. Chikungunya fever, mauritius, 2006. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 14, n. 2, p. 337–338, 2008.
- BELTRAME, A.; ANGHEN, A.; BISOFFI, Z.; MONTEIRO, G.; MAROCCO, S.; CARELLI, G.; LIPANI, F.; GOBBI, F.; CANTA, F.; CASTELLI, F.; GULLETTA, M.; BIGONI, S.; PUNTA, V. D.; IACOVAZZI, T.; ROMI, R.; NICOLETTI, L.; CIUFOLINI, M. G.; RORATO, G.; NEGRI, C.; VIALE, P. Imported chikungunya infection, italy. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 13, 2007.

BELTRÁN, J. D.; BOSCOR, A.; SANTOS, W. P. dos; MASSONI, T.; KOSTKOVA, P. Zika: A new system to empower health workers and local communities to improve surveillance protocols by e-learning and to forecast zika virus in real time in brazil. In: *Proceedings of the 2018 International Conference on Digital Health*. [S.l.: s.n.], 2018. p. 90–94.

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 Research, v. 496, n. 7446, p. 504–507, 2013.

BONILAURI, P.; BELLINI, R.; CALZOLARI, M.; ANGELINI, R.; VENTURI, L.; FALLACARA, F.; CORDIOLI, P.; ANGELINI, P.; VENTURELLI, C.; MERIALDI, G.; DOTTORI, M. Imported chikungunya infection, italy. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 14, 2008.

BRITO, A. F. de; RIBEIRO, I. P.; MIRANDA, R. M. de; FERNANDES, R. S.; CAMPOS, S. S.; SILVA, K. A. B. da; CASTRO, M. G. de; BONALDO, M. C.; BRASIL, P.; OLIVEIRA, R. L. de. First detection of natural infection of aedes aegypti with zika virus in brazil and throughout south america. *Memória do Instituto Oswaldo Cruz*, SciELO, v. 110, p. 655–658, 2016.

CAMPOS, G. S.; BANDEIRA, A. C.; SARDI, S. I. Zika virus outbreak, bahia, brazil. *Saúde e Sociedade*, SciELO, v. 25, 2016.

CAO-LORMEAU, V.-M.; BLAKE, A.; MONS, S.; LASTÈRE, S.; ROCHE, C.; VANHOMWEGEN, J.; DUB, T.; BAUDOUIN, L.; TEISSIER, A.; LARRE, P. et al. Guillain-barré syndrome outbreak associated with zika virus infection in french polynesia: a case-control study. *The Lancet*, Elsevier, v. 387, n. 10027, p. 1531–1539, 2016.

CAO, T. H.; NGUYEN, A. D.; DINH, T. Q.; LUONG, Q. C.; DIEP, H. T. Forecasting dengue incidence: Statistical and dynamic models. *Oxford Journal of Intelligent Decision and Data Science*, International Scientific Publications and Consulting Services, v. 2018, p. 1–13, 2018.

COMMOWICK, O.; ISTACE, A.; KAIN, M.; LAURENT, B.; LERAY, F.; SIMON, M.; POP, S. C.; GIRARD, P.; AMELI, R.; FERRÉ, J.-C. et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–17, 2018.

CORDEIRO, F. R.; LIMA, S. M.; SILVA-FILHO, A. G.; SANTOS, W. Segmentation of mammography by applying extreme learning machine in tumor detection. In: SPRINGER. *International Conference on Intelligent Data Engineering and Automated Learning*. [S.l.], 2012. p. 92–100.

CRUZ, T.; CRUZ, T.; SANTOS, W. Detection and classification of lesions in mammographies using neural networks and morphological wavelets. *IEEE Latin America Transactions*, v. 16, n. 3, p. 926–932, 2018.

DIOP, D.; MESENZNIKOV, G.; SANICAS, M. Chikungunya outbreaks from 2000 to 2015: a review. *MOJ Public Health*, MedCrave, v. 2, n. 6, p. 181–187, 2015.

DOM, N. C.; AHMAD, A. H.; LATIF, Z. A.; ISMAIL, R. Integration of gis-based model with epidemiological data as a tool for dengue surveillance. *Environmental Asia*, v. 10, p. 135–146, 2017.

- ECKHOFF, P. A.; TATEM, A. J. *Digital methods in epidemiology can transform disease control*. [S.l.]: Oxford University Press, 2015.
- EICKMANN, S. H.; CARVALHO, M. D. C. G.; RAMOS, R. C. F.; ROCHA, M. Ângela W.; LINDEN, V. van der; SILVA, P. F. S. da. Síndrome da infecção congênita pelo vírus zika. *Caderno de Saúde Pública*, SciELO, v. 32, p. :e00047716, 2016.
- FERNANDES, F. T.; FILHO, A. D. P. C. Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. *Revista Brasileira de Saúde Ocupacional*, SciELO, v. 44, 2019. Disponível em: <<http://dx.doi.org/10.1590/2317-6369000019418>>.
- FREIRE, P. *Pedagogia da indignação: cartas pedagógicas e outros escritos*. 1. ed. São Paulo: UNESP, 2000.
- FREITAS, R. C. de; ALVES, R.; FILHO, A. G. da S.; SOUZA, R. E. de; BEZERRA, B. L.; SANTOS, W. P. dos. Electromyography-controlled car: A proof of concept based on surface electromyography, extreme learning machines and low-cost open hardware. *Computers & Electrical Engineering*, Elsevier, v. 73, p. 167–179, 2019.
- GOUPIL, B. A.; MORES, C. N. A review of chikungunya virus-induced arthralgia: Clinical manifestations, therapeutics, and pathogenesis. *The Open Rheumatology Journal*, Bentham Open, v. 10, p. 129–140, 2016.
- GRAMDADAM, M.; CARO, V.; PLUMENT, S.; THIBERGE, J.-M.; SOUARÈS, Y.; FAILLOUX, A.-B.; TOLOU, H. J.; BUDELLOT, M.; COSSERAT, D.; LEPARC-GOFFART, I.; DESPRÈS, P. Chikungunya virus, southeastern france. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 17, n. 5, p. 910–913, 2011.
- GUBLER, D. J. Dengue, urbanization and globalization: the unholy trinity of the 21st century. *Tropical medicine and health*, Japanese Society of Tropical Medicine, v. 39, n. 4SUPPLEMENT, p. S3–S11, 2011.
- GUIMARÃES, R. B. Geography and public health in brazil. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 21, 2015.
- GUO, P.; LIU, T.; ZHANG, Q.; WANG, L.; XIAO, J.; ZHANG, Q.; LUO, G.; LI, Z.; HE, J.; ZHANG, Y.; MA, W. Developing a dengue forecast model using machine learning: A case study in china. *PLOS Neglected Tropical Diseases*, Public Library of Science, v. 11, n. 10, p. 1–22, 10 2017. Disponível em: <<https://doi.org/10.1371/journal.pntd.0005973>>.
- HALL, J. E. *Tratado de fisiologia médica*. Rio de Janeiro, RJ, Brasil: Elsevier, 2011.
- HAYKIN, S. *Redes neurais: princípios e prática*. Porto Alegre, RS, Brasil: Bookman, 2001.
- HONÓRIO, N. A.; CÂMARA, D. C. P.; CALVET, G. A.; BRASIL, P. Chikungunya: uma arbovirose em estabelecimento e expansão no brasil. *Caderno de Saúde Pública*, SciELO, v. 31, n. 5, p. 906–908, 2015.
- IBGE. *Instituto Brasileiro de Geografia e Estatística*. 2018. Disponível em: <<https://www.ibge.gov.br/pt/inicio.html>>. Acesso em: 15 de nov. de 2018.
- INMET. *Instituto Nacional de Meteorologia. Sobre o INMET*. 2018. Disponível em: <<https://portal.inmet.gov.br/sobre>>. Acesso em: 15 de nov. de 2018.

- JINDAL, A.; RAO, S. Agent-based modeling and simulation of mosquito-borne disease transmission. In: INTERNATIONAL FOUNDATION FOR AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. [S.l.], 2017. p. 426–435.
- JOSSERAN, L.; PAQUET, C.; ZEHGNOUN, A.; CAILLERE, N.; TERTRE, A. L.; SOLET, J.-L.; LEDRANS, M. Chikungunya disease outbreak, reunion island. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 12, n. 12, p. 1994–1995, 2008.
- JUNIOR, M. da S.; FREITAS, R. C. de; SANTOS, W. P. dos; SILVA, W. W. A. da; RODRIGUES, M. C. A.; CONDE, E. F. Q. Exploratory study of the effect of binaural beat stimulation on the eeg activity pattern in resting state using artificial neural networks. *Cognitive Systems Research*, Elsevier, v. 54, p. 1–20, 2019.
- JUNIOR, V. L. P.; LUZ, K.; PARREIRA, R.; FERRINHO, P. Vírus zika: Revisão para clínicos. *Acta Médica Portuguesa*, Revista Científica da Ordem dos Médicos, v. 28, p. 760–765, 2015.
- KOSTKOVA, P.; SANTOS, W. P. dos; MASSONI, T. Zika: improved surveillance and forecast of zika virus in brazil: Patty kostkova. *European Journal of Public Health*, Oxford University Press, v. 29, n. Supplement\_4, p. ckz186–085, 2019.
- KRAEMER, M. U.; SINKA, M. E.; DUDA, K. A.; MYLNE, A. Q.; SHEARER, F. M.; BARKER, C. M.; MOORE, C. G.; CARVALHO, R. G.; COELHO, G. E.; BORTEL, W. V.; HENDRICKX, G.; SCHAFFNER, F.; ELYAZAR, I. R.; TENG, H.-J.; BRADY, O. J.; MESSINA, J. P.; PIGOTT, D. M.; SCOTT, T. W.; SMITH, D. L.; WINT, G. W.; GOLDING, N.; HAY, S. I. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife*, eLife Sciences Publications, Ltd, v. 4, p. e08347, jun 2015. ISSN 2050-084X. Disponível em: <<https://doi.org/10.7554/eLife.08347>>.
- LAUREANO-ROSARIO, A. E.; DUNCAN, A. P.; MENDEZ-LAZARO, P. A.; GARCIA-REJON, J. E.; GOMEZ-CARRO, S.; FARFAN-ALE, J.; SAVIC, D. A.; MULLER-KARGER, F. E. Application of artificial neural networks for dengue fever outbreak predictions in the northwest coast of yucatan, mexico and san juan, puerto rico. *Tropical medicine and infectious disease*, Multidisciplinary Digital Publishing Institute, v. 3, n. 1, p. 5, 2018.
- LI, Z.; LIU, T.; ZHU, G.; LIN, H.; ZHANG, Y.; HE, J.; DENG, A.; PENG, Z.; XIAO, J.; RUTHERFORD, S.; XIE, R.; ZENG, W.; LI, X.; MA, W. Dengue baidu search index data can improve the prediction of local dengue epidemic: A case study in guangzhou, china. *PLOS Neglected Tropical Diseases*, Public Library of Science, v. 11, n. 3, p. 1–13, 03 2017. Disponível em: <<https://doi.org/10.1371/journal.pntd.0005354>>.
- LIMA, S. M. de; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. Detection and classification of masses in mammographic images in a multi-kernel approach. *Computer methods and programs in biomedicine*, Elsevier, v. 134, p. 11–29, 2016.
- LIMA, T. F. M. de; LANA, R. M.; CARNEIRO, T. G. de S.; CODEÇO, C. T.; MACHADO, G. S.; FERREIRA, L. S.; MEDEIROS, L. C. de C.; JUNIOR, C. A. D. Dengueme: A tool for the modeling and simulation of dengue spatiotemporal dynamics. *International Journal of Environmental Research and Public Health*, Multidisciplinary Digital Publishing Institute, v. 13, n. 9, p. 920, 2016.

- LIU, G.; HU, X.; WANG, E.; ZHOU, G.; CAI, J.; ZHANG, S. Svr-eemd: An improved eemd method based on support vector regression extension in ppg signal denoising. *Computational and Mathematical Methods in Medicine*, Hindawi, 2019. Disponível em: <<https://doi.org/10.1155/2019/5363712>>.
- MARQUES-TOLEDO, C. d. A.; DEGENER, C. M.; VINHAL, L.; COELHO, G.; MEIRA, W.; CODEÇO, C. T.; TEIXEIRA, M. M. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLOS Neglected Tropical Diseases*, Public Library of Science, v. 11, n. 7, p. 1–20, 07 2017. Disponível em: <<https://doi.org/10.1371/journal.pntd.0005729>>.
- MARTÍNEZ-BELLO, D.; LÓPEZ-QUÍLER, A.; PRIETO, A. T. Spatiotemporal modeling of relative risk os dengue disease in colombia. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 32, p. 1587–1601, 2018.
- MELLO, C. A.; SANTOS, W. P. D.; RODRIGUES, M. A.; CANDEIAS, A. L. B.; GUSMAO, C. M. Image segmentation of ovitraps for automatic counting of aedes aegypti eggs. In: IEEE. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.], 2008. p. 3103–3106.
- MOHAMMED, A.; CHADEE, D. D. Effects of different temperature regimens on the development of aedes aegypti (l.)(diptera: Culicidae) mosquitoes. *Acta tropica*, Elsevier, v. 119, n. 1, p. 38–43, 2011.
- MORRISON, T. E. Reemergence of chikungunya virus. *Journal of Virology*, American Society for Microbiology Journals, v. 88, n. 20, p. 11644–11647, 2014.
- MORSY, S.; DANG, T.; KAMEL, M.; ZAYAN, A.; MAKRAM, O.; ELHADY, M.; HIRAYAMA, K.; HUY, N. Prediction of zika-confirmed cases in brazil and colombia using google trends. *Epidemiology and Infection*, Cambridge University Press, v. 146, n. 13, p. 1625–1627, 2018.
- MUSAH, A.; RUBIO-SOLIS, A.; BIRJOVANU, G.; SANTOS, W. P. dos; MASSONI, T.; KOSTKOVA, P. Assessing the relationship between various climatic risk factors & mosquito abundance in recife, brazil. In: *Proceedings of the 9th International Conference on Digital Public Health*. [S.l.: s.n.], 2019. p. 97–100.
- MUSSO, D.; GUBLER, D. J. Zika virus. *Clinical Microbiology Rewies*, American Society for Microbiology, v. 29, p. 487–524, 2016.
- NARDI, S. M. T.; PASCHOAL, J. A. A.; PEDRO, H. da S. P.; PASCHOAL, V. D.; SICHIERI, E. P. Geoprocessamento em saúde pública: fundamentos e aplicações. *Revista Instituto Adolfo Lutz*, v. 72, p. 185–191, 2013.
- NETO, V. C.; CHIARI, N. S.; CARVALHO, I.; PISA, I. T.; ALVES, D. Desenvolvimento e integração de mapas dinâmicos georreferenciados para o gerenciamento e vigilância em saúde. *Journal of Health Informatics*, v. 6, p. 3–9, 2014.
- NÓBREGA, M. E. B. da; ARAÚJO, E. L. de L.; WADA, M. Y.; LEITE, P. L. e; DIMECH, G. S.; PÉRCIO, J. First detection of natural infection of aedes aegypti with zika virus in brazil and throughout south america. *Epidemiologia e Serviços da Saúde*, SciELO, v. 27, p. e2017039, 2018.

- PADMANABHAN, P.; SESHAIYER, P.; CASTILLO-CHAVEZ, C. Mathematical modeling, analysis and simulation of the spread of zika with influence of sexual transmission and preventive measures. *Letters in Biomathematics*, Taylor & Francis, v. 4, n. 1, p. 148–166, 2017.
- PARK, H.; JUNG, H.; ON, J.; PARK, S. K.; KANG, H. Digital epidemiology: Use of digital data collected for non-epidemiological purposes in epidemiological studies. *Healthcare informatics research*, v. 24, n. 4, p. 253–262, 2018.
- PAUL, K. K.; DHAR-CHOWDHURY, P.; HAQUE, C. E.; AL-AMIN, H. M.; GOSWAMI, D. R.; KAFI, M. A. H.; DREBOT, M. A.; LINDSAY, L. R.; AHSAN, G. U.; BROOKS, W. A. Risk factors for the presence of dengue vector mosquitoes, and determinants of their prevalence and larval site selection in dhaka, bangladesh. *PLOS ONE*, Public Library of Science, v. 13, n. 6, p. 1–19, 06 2018. Disponível em: <<https://doi.org/10.1371/journal.pone.0199457>>.
- PESSANHA, J. E. M.; CAIAFFA, W. T.; CÉSAR, C. C.; PROIETTI, F. A. Avaliação do plano nacional de controle da dengue evaluation of the brazilian national dengue control plan. *Cad. saúde pública*, v. 25, n. 7, p. 1637–1641, 2009.
- PHUNG, D.; HUANG, C.; RUTHERFORD, S.; CHU, C.; WANG, X.; NGUYEN, M.; NGUYEN, N. H.; MANH, C. D. Identification of the prediction model for dengue incidence in can tho city, a mekong delta area in vietnam. *Acta Tropica*, v. 141, p. 88 – 96, 2015. ISSN 0001-706X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0001706X14003179>>.
- PIALOUX, G.; Bernard-Alex Gaüzère; Stéphane Jauréguiberry; STROBEL, M. Chikungunya, an epidemic arbovirosis. *The Lancet Infectious Diseases*, PubMed, v. 7, n. 5, p. 319–327, 2007.
- PORTELA, N. M.; MELLO, C. A. B. de; SANTOS, W. P. dos; CANDEIAS, A. L. B.; GUSMÃO, C. M. G. de; MACHADO, S. C.; RODRIGUES, M. A. B. A new algorithm for segmenting and counting aedes aegypti eggs in ovitraps. In: IEEE. *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.], 2009. p. 6714–6717.
- POWERS, A. M.; LOGUE, C. H. Changing patterns of chikungunya virus: re-emergence of zoonotic arbovirus. *Journal of General Virology*, Microbiology Society, v. 88, p. 2363–2377, 2007.
- Python Software Foundation. *Python Language Reference*. [S.l.], 2019. Disponível em: <<https://www.python.org/downloads/>>.
- QGIS Development Team. *QGIS Geographic Information System*. [S.l.], 2018. Disponível em: <<http://qgis.org>>.
- RECIFE. *Portal de Dados Abertos da Prefeitura de Recife.Sobre o INMET*. 2018. Disponível em: <<https://portal.inmet.gov.br/sobre>>. Acesso em: 15 de nov. de 2018.
- REN, H.; ZHENG, L.; LI, Q.; YUAN, W.; LU, L. Exploring determinants of spatial variations in the dengue fever epidemic using geographically weighted regression model: A case study in the joint guangzhou-foshan area, china, 2014. *International Journal of Environmental Research and Public Health*, MDPI, v. 14, n. 12, 2017.

REZZA, G.; NICOLETTI, L.; ANGELINI, R.; ROMI, R.; FINARELLI, A. C.; PANNING, M.; CORDIOLI, P.; FORTUNA, C.; BOROS, S.; MAGURANO, F.; SILVI, G.; ANGELINI, P.; DOTTORI, M.; CIUFOLINI, M. G.; MAJORI, G. C.; CASSONE, A. Infection with chikungunya virus in italy: an outbreak in a temperate region. *The Lancet*, Elsevier, v. 370, p. 1840–1846, 2007.

RODRIGUES, A. L.; SANTANA, M. A. de; AZEVEDO, W. W.; BEZERRA, R. S.; BARBOSA, V. A.; LIMA, R. C. de; SANTOS, W. P. dos. Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. *Research on Biomedical Engineering*, Springer, v. 35, n. 3, p. 213–222, 2019.

RUBIO-SOLIS, A.; MUSAH, A.; SANTOS, W. P. D.; MASSONI, T.; BIRJOVANU, G.; KOSTKOVA, P. Zika virus: Prediction of aedes mosquito larvae occurrence in recife (brazil) using online extreme learning machine and neural networks. In: *Proceedings of the 9th International Conference on Digital Public Health*. [S.l.: s.n.], 2019. p. 101–110.

SALATHÈ, M. Digital epidemiology: what is it, and where is it going? *Life Sci Policy*, v. 14, n. 1, 2018.

SALATHE, M.; BENGTSSON, L.; BODNAR, T. J.; BREWER, D. D.; BROWNSTEIN, J. S.; BUCKEE, C.; CAMPBELL, E. M.; CATTUTO, C.; KHANDELWAL, S.; MABRY, P. L. et al. Digital epidemiology. *PLoS computational biology*, Public Library of Science, v. 8, n. 7, p. e1002616, 2012.

SANG, S.; GU, S.; BI, P.; YANG, W.; YANG, Z.; XU, L.; YANG, J.; LIU, X.; JIANG, T.; WU, H.; CHU, C.; LIU, Q. Predicting unprecedented dengue outbreak using imported cases and climatic factors in guangzhou, 2014. *PLOS Neglected Tropical Diseases*, Public Library of Science, v. 9, n. 5, p. 1–12, 05 2015. Disponível em: <<https://doi.org/10.1371/journal.pntd.0003808>>.

SANTANA, P. *Introdução a Geografia da Saúde*. Rio de Janeiro, RJ, Brasil: Imprensa da Universidade de Coimbra, 2014.

SANTOS, M. M. dos; FILHO, A. G. da S.; SANTOS, W. P. dos. Deep convolutional extreme learning machines: filters combination and error model validation. *Neurocomputing*, Elsevier, v. 329, p. 359–369, 2019.

SANTOS, W. P. d.; ASSIS, F.; SOUZA, R.; FILHO, P. B. S.; NETO, F. L. Dialectical multispectral classification of diffusion-weighted magnetic resonance images as an alternative to apparent diffusion coefficients maps to perform anatomical analysis. *Computerized Medical Imaging and Graphics*, Elsevier, v. 33, n. 6, p. 442–460, 2009.

SANTOS, W. P. dos; ASSIS, F. M. de; SOUZA, R. E. de; MENDES, P. B.; MONTEIRO, H. S. de S.; ALVES, H. D. A dialectical method to classify alzheimer's magnetic resonance images. In: *Evolutionary Computation*. [S.l.]: IntechOpen, 2009.

SANTOS, W. P. dos; SOUZA, R. E. de; FILHO, P. B. dos S. Evaluation of alzheimer's disease by analysis of mr images using multilayer perceptrons and kohonen som classifiers as an alternative to the adc maps. In: IEEE. *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.], 2007. p. 2118–2121.

SANTOS, W. P. dos; SOUZA, R. E. de; FILHO, P. B. S.; NETO, F. B. L.; ASSIS, F. M. de. A dialectical approach for classification of dw-mr alzheimer's images. In: IEEE. *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. [S.l.], 2008. p. 1728–1735.

SCHUFFENECKER, I.; ITEMAN, I.; MICHAULT, A.; MURRI, S.; FRANGEUL, L.; VANEY, M.-C.; LAVENIR, R.; PARDIGON, N.; REYNES, J.-M.; PETTINELLI, F.; BISCORNET, L.; DIANCOURT, L.; MICHEL, S.; DUQUERROY, S.; GUIGON, G.; FRENKIEL, M.-P.; BRÉHIN, A.-C.; CUBITO, N.; DESPRÈS, P.; KUNST, F.; REY, F. A.; ZELLER, H.; BRISSE, S. Genome microevolution of chikungunya viruses causing the indian ocean outbreak. *PLOS Medicine*, Public Library of Science, v. 3, n. 7, 05 2006.

SHARMA, V.; KUMAR, A.; PANAT, L.; KARAJKHEDE, D. G.; LELE, A. Malaria outbreak prediction model using machine learning. *International Journal of Advanced Research in Computer Engineering and Technology*, Revista Científica da Ordem dos Médicos, v. 4, p. 4415–4419, 2015.

SILVA, A. M.; SILVA, R. M. da; ALMEIDA, C. A. P. de; CHAVES, J. J. da S. Modelagem geoestatística dos casos de dengue e da variação termopluiométrica em João Pessoa, Brasil. *Sociedade e Natureza*, SciELO, v. 27, n. 1, p. 157–169, 2015.

SIRIYASATIEN, P.; CHADSUTHI, S.; JAMPACHAISRI, K.; KESORN, K. Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes. *IEEE Access*, v. 6, p. 53757–53795, 2018. ISSN 2169-3536.

TARIQ, B.; ZAIDI, A. Z. Geostatistical modeling of dengue disease in Lahore, Pakistan. *SN Applied Science*, 2019. Disponível em: <<https://doi.org/10.1007/s42452-019-0428-1>>.

THIÉBAUT, R.; THIESSARD, F. Public health and epidemiology informatics. *Yearbook of Medical Informatics*, v. 26, n. 01, p. 248–251, 2017.

TOSEPU, R.; TANTRAKARNAPA, K.; WORKHUNPISET, S.; NAKHAPAKORN, K. Climatic factors influencing dengue hemorrhagic fever in Kolaka district, Indonesia. *Environment and Natural Resources Journal*, v. 16, n. 2, p. 1–10, 2018.

VARGAS, A.; SAAD, E.; DIMECH, G. S.; SANTOS, R. H.; SIVINI, M. A. V. C.; ALBUQUERQUE, L. C.; LIMA, P. M. S.; BARRETO, I. de C.; ANDRADE, M. E. de; ESTIMA, N. M.; CARVALHO, P. I. de; AZEVEDO, R. S. de A.; VASCONCELOS, R. de Cássia de O.; ASSUNÇÃO, R. S.; FRUTUOSO, L. C. V.; CARMO, G. M. I. do; SOUZA, P. B. de; WADA, M. Y.; OLIVEIRA, W. K. de; HENRIQUES, C. M. P.; PERCIO, J. Características dos primeiros casos de microcefalia possivelmente relacionados ao vírus Zika notificados na região metropolitana de Recife, Pernambuco. *Epidemiologia e Serviços de Saúde*, SciELO, v. 25, p. 691–700, 2016.

VASCONCELOS, J. de; SANTOS, W. D.; LIMA, R. D. Analysis of methods of classification of breast thermographic images to determine their viability in the early breast cancer detection. *IEEE Latin America Transactions*, IEEE, v. 16, n. 6, p. 1631–1637, 2018.

WITTEN, I. H.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 2005.

YACTAYO, S.; STAPLES, J.; MILLOT, V.; RAMON-PARDO, L. P. Epidemiology of chikungunya in the Americas. *The Journal of Infectious Diseases*, PCM, v. 214, p. S441–S445, 2016.

---

YERGOLKAR, P. N.; TANDALE, B. V.; ARANKALLE, V. A.; SATHE, P. S.; B., S. A.; GANDHE, S. S.; GOKHLE, M. D.; JACOB, G. P.; HUNDEKAR, S. L.; MISHRA, A. C. Chikungunya outbreaks caused by african genotype, india. *Emerging Infectious Diseases*, Centers for Disease Control and Prevention, v. 12, n. 10, p. 1580–1583, 2006.

## Índice

Conclusão, 75

Fundamentação Teórica, 21

Proposta, 39

Resultados Experimentais, 52

Trabalhos relacionados, 34