



Ana Clara Gomes da Silva

**Modelagem e simulação da transmissão de
doenças baseada em aprendizado de máquina
para a predição de dengue, zika, chikungunya e
Covid-19**

Dissertação de Mestrado

Recife

2022

Ana Clara Gomes da Silva

**Modelagem e simulação da transmissão de doenças
baseada em aprendizado de máquina para a predição de
dengue, zika, chikungunya e Covid-19**

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

Orientador: Prof. Dr. Wellington Pinheiro dos Santos

Coorientador: Prof. Dr. César Augusto Rodrigues Castilho

Recife

2022

Catalogação na fonte:
Bibliotecário Josias Machado, CRB-4 / 1690

S586m Silva, Ana Clara Gomes da.
Modelagem e simulação da transmissão de doenças baseada em aprendizado de máquina para a predição de dengue, zica, chikungunya e covid 19 / Ana Clara Gomes Silva. – 2022.
94 f.: il., figs., tabs., abrev. e siglas.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos.
Coorientador: Prof. César Augusto Rodrigues Castilho
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG.
Programa de Pós-Graduação em Engenharia Biomédica. Recife, 2022.
Inclui referências e índice.

1. Engenharia biomédica. 2. Arboviroses. 3. Covid 19. 4. Epidemiologia digital. 5. Aprendizagem de máquina. 6. Predição de arboviroses. I. Santos, Wellington Pinheiro dos (Orientador). II. Castilho, César Augusto Rodrigues (Coorientador). III. Título.

UFPE

610.28 CDD (22. ed.)

BCTG/2022-431

Ana Clara Gomes da Silva

Modelagem e simulação da transmissão de doenças baseada em aprendizado de máquina para a predição de dengue, zika, chikungunya e Covid-19

Trabalho submetido ao Programa de Pós-Graduação em Engenharia Biomédica, na área de concentração Bioengenharia - Computação 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 11 de fevereiro de 2022:

Prof. Dr. Wellington Pinheiro dos Santos
Orientador

Prof. Dra. Cristine Martins Gomes de Gusmão
Examinadora Interna

Prof. Dr. Sidney Marlon Lopes de Lima
Examinador Externo

Recife
2022

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

“A educação é a arma mais poderosa para mudar o mundo.”

Nelson Mandela.

Resumo

A prevenção e o controle das arboviroses, especialmente da dengue, da febre chikungunya e da zika, no caso do Brasil, têm sido um grande desafio de saúde pública para muitos países, especialmente a partir de 2015, uma vez que outras arboviroses passaram a interagir com o vírus da dengue. A situação se agravou a partir de 2016, com o surgimento do zika vírus e de sua ação sobre a gravidez, estando relacionado em um certo grau com casos de microcefalia mas, principalmente, com a síndrome de Guillain-Barret, uma doença autoimune que afeta o sistema nervoso, provocando desde fraqueza muscular até a paralisia. Em dezembro de 2019 começou, na cidade de Wuhan, na China, a epidemia de Covid-19, provocada pelo coronavírus SARS-CoV-2. Rapidamente o vírus se espalhou pelo mundo, dando origem à pandemia de Covid-19, o maior problema de saúde do século XXI até o momento. No seu começo considerada como uma doença do trato respiratório, como as penumonias virais, a Covid-19 se mostrou uma doença do sistema cardiovascular que afeta não somente os pulmões, mas também os rins e o sistema nervoso, podendo causar sequelas que podem ser permanentes. A letalidade da doença é relativamente baixa, mas como o contágio é rápido, principalmente por conta das variantes, o baixo percentual de casos graves acaba resultando em milhões de mortes. 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. Neste trabalho temos como objetivo investigar modelos baseados em aprendizado de máquina para predição da distribuição espacial e temporal de casos de arboviroses e de Covid-19, buscando lançar as bases para a construção de sistemas de predição espaço-temporal para fins epidemiológicos. Neste trabalho, utilizamos a base de dados de casos e locais de arboviroses LIRAA, do Sistema Único de Saúde da Cidade do Recife, de 2016 a 2019, para a predição de arbovírus; para a predição de Covid-19, utilizamos as bases de dados do Sistema Nacional de Notificações fornecidas pelas secretarias estaduais de saúde. E para a Covid-19 usaremos o Portal Brasil.io como base de dados para casos e óbitos de cada um dos estados e cidades brasileiras.

Palavras-chaves: arboviroses; covid-19; epidemiologia digital; aprendizagem de máquina; predição de arboviroses.

Abstract

The prevention and control of arboviruses, especially dengue, chikungunya fever and Zika, in the case of Brazil, has been a major public health challenge for many countries, especially from 2015, since other arboviruses began to interact with the dengue virus. The situation worsened from 2016, with the emergence of Zika virus and its action on pregnancy, being related to a certain degree with cases of microcephaly, but mainly with Guillain-Barre syndrome, an autoimmune disease that affects the nervous system, causing everything from muscle weakness to paralysis. In December 2019, the Covid-19 epidemic, caused by the SARS-CoV-2 coronavirus, began in the city of Wuhan, China. The virus quickly spread around the world, giving rise to the Covid-19 pandemic, the biggest health problem of the 21st century to date. At its beginning, considered as a disease of the respiratory tract, like viral pneumonia, Covid-19 proved to be a disease of the cardiovascular system that affects not only the lungs, but also the kidneys and the nervous system, and can cause sequelae that can be permanent. The lethality of the disease is relatively low, but as the contagion is fast, mainly due to the variants, the low percentage of severe cases ends up resulting in millions of deaths. The advancement of Digital Epidemiology and geoprocessing technologies, together with the development of Data Mining and Machine Learning techniques, have provided 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. In this work we aim to investigate models based on machine learning to predict the spatial and temporal distribution of arboviruses and Covid-19 cases, seeking to lay the foundations for the construction of spatiotemporal prediction systems for epidemiological purposes. In this work, we used the database of cases and locations of arboviruses LIRAA, from the Unified Health System of the City of Recife, from 2016 to 2019, to predict arborivores; to predict Covid-19, we used the National Notification System databases provided by the state health departments.

Keywords: arboviruses; covid-19; digital epidemiology; machine learning; arboviruses prediction.

Listas de ilustrações

Figura 1 – Modelo compartimentado SIR. Fonte (LIMA et al., 2020)	40
Figura 2 – Metodologia usada para gerar os mapas de predição da distribuição de criadouros na Cidade do Recife.	45
Figura 3 – Metodologia do sistema de predição COVID-SGIS com abordagem temporal	47
Figura 4 – Tabela de Regressores e suas configurações.	55
Figura 5 – Resultados do coeficiente de correlação, root relative squared error e tempo médio de treino referente à regressão linear, Random Forest (10, 20 ,30 e 40 árvores), SVM (kernels polinomiais de graus 1, 2 e 3, e kernel RBF) e MLP(10, 20 e 30 neurônio).	56
Figura 6 – Comparação dos resultados obtidos para regressão linear e SVM com kernel polinomial de grau 3.	57
Figura 7 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o primeiro (a), segundo (b), terceiro (c) e quarto (d) bimestres de 2014, nessa ordem.	58
Figura 8 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o quinto (a) e sexto (b) bimestres de 2014, nessa ordem.	59
Figura 9 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o primeiro (a), segundo (b), terceiro (c) e quarto (d) bimestres de 2015, nessa ordem.	61
Figura 10 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o quinto (a) e sexto (b) bimestres de 2015, nessa ordem.	62
Figura 11 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o primeiro (a), segundo (b), terceiro (c) e quarto (d) bimestres de 2016, nessa ordem.	63
Figura 12 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o quinto (a) e sexto (b) bimestres de 2016, nessa ordem.	64
Figura 13 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Acre, (b) Alagoas, (c) Amazonas, (d) Amapá, (e) Bahia, (f) Ceará, (g) Distrito Federal, e (h) Espírito Santo. . .	66
Figura 14 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Goiás, (b) Maranhão, (c) Minas Gerais, (d) Mato Grosso do Sul, (e) Mato Grosso, (f) Pará, (g) Paraíba, e (h) Pernambuco.	67

Figura 15 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Piauí, (b) Paraná, (c) Rio de Janeiro, (d) Rio Grande do Norte, (e) Rondônia, (f) Roraima, (g) Rio Grande do Sul, e (h) Santa Catarina.	68
Figura 16 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Sergipe, (b) São Paulo, (c) Tocantins, e (d) todo o país.	69
Figura 17 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Região Nordeste. Fonte: (LIMA et al., 2020)	70
Figura 18 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Região Norte. Fonte: (LIMA et al., 2020)	71
Figura 19 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Região Centro-Oeste. Fonte: (LIMA et al., 2020)	72
Figura 20 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Sudeste. Fonte: (LIMA et al., 2020)	73
Figura 21 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Sul. Fonte: (LIMA et al., 2020) .	73
Figura 22 – Resultado das projeções dos casos confirmados acumulados de Covid-19 para o Brasil entre 6 a 11 de maio de 2020. Fonte: (LIMA et al., 2020) .	74
Figura 23 – Resultados do desempenho dos algoritmos de regressão linear, do perceptron multicamadas (MLP) e do regressor vetorial de suporte (SVR) para o conjunto de dados do Brasil.Fonte: (SILVA et al., 2021b)	75
Figura 24 – Resultados do desempenho dos algoritmos de regressão linear, do perceptron multicamadas (MLP) e do regressor vetorial de suporte (SVR) para o conjunto de dados de Pernambuco.fonte: (SILVA et al., 2021b) . .	75
Figura 25 – Resultados da validação dos modelos criados por meio de regressão linear e SVR, kernel RBF, para Brasil e Pernambuco. Fonte: (SILVA et al., 2021b)	76

Figura 26 – (A) Tela inicial do COVID-SGIS. (B) Tela de seleção do mapa, onde é possível visualizar o mapa dos casos cumulativos e os mapas de distribuição previstos. (C) Tela de previsão dos mapas de distribuição dos casos cumulativos de Covid-19. (D) Casos acumulados do gráfico de previsão Covid-19 com ARIMA. A linha verde representa a previsão ARIMA, enquanto as linhas vermelha e azul representam o pior e o melhor cenário. Tela dos gráficos de casos confirmados acumulados e diários (E) e óbitos acumulados e confirmados (F). Telas do painel personalizado (G, H). O painel personalizado fornece informações detalhadas - sobre os casos confirmados cumulativos e casos diários, e mortes cumulativas e mortes diárias - para cada estado em nível de município). Fonte: (LIMA et al., 2020)

78

Figura 27 – (A) Tela inicial do aplicativo da web COVID SGIS. (B) Casos acumulados do gráfico de previsão Covid-19. A previsão com ARIMA é representada pela linha verde. O pior cenário (indicado pelo limite superior da previsão) é representado pela linha em vermelho. O melhor cenário (indicado pelo limite inferior da previsão) é representado pela linha azul. (C) Tela do gráfico de distribuição de casos confirmados e óbitos por Covid-19. Neste gráfico o usuário pode ter uma visão geral dos casos confirmados e óbitos acumulados em todos os estados do Brasil e no Distrito Federal. No COVID SGIS, o usuário pode acompanhar os casos confirmados diários e acumulados (D) e óbitos (E) da Covid-19 para cada estado brasileiro e Distrito Federal, separadamente. Fonte: (LIMA et al., 2020) .

79

Listas de tabelas

Tabela 1 – Resultados do teste de Dickey-Fuller para séries históricas do número de casos acumulados de covid-19 no Brasil e seus 27 estados. Quando t-estatística for melhor que τ , isto indica que a série possui uma única raiz (que não é estacionária). E quando a t-estatística for inferior a τ a série é estacionária. Fonte: (LIMA et al., 2020)	50
Tabela 2 – Resultados do teste de Dickey-Fuller para séries históricas do número de casos acumulados de covid-19 no Brasil e seus 27 estados. Quando t-estatística for melhor que τ , isto indica que a série possui uma única raiz (que não é estacionária). E quando a t-estatística for inferior a τ a série é estacionária. Fonte: (LIMA et al., 2020)	51
Tabela 3 – Resultados dos coeficientes de correlação de Pearson, Spearman e Kendal, e do erro quadrático relativo (RMSE%) para os modelos ARIMA construídos para o Brasil e seus 27 estados.	69

Sumário

1	INTRODUÇÃO	15
1.1	Contexto e motivação	15
1.2	Objetivos	18
1.3	Organização do trabalho	19
2	TRABALHOS RELACIONADOS	20
2.1	Introdução	20
2.2	Predição de arboviroses	20
2.3	Predição de covid-19	25
2.4	Considerações	30
3	FUNDAMENTAÇÃO TEÓRICA	32
3.1	Epidemiologia Digital e Geografia da Saúde	32
3.2	Dengue, Febre <i>chinkungunya</i> e Zika	33
3.2.1	Dengue	34
3.2.2	Febre <i>chikungunya</i>	35
3.2.3	Zika	37
3.3	Doença do Coronavírus	38
3.4	Predição de doenças infecciosas	39
3.4.1	Modelo estatístico	39
3.4.2	Modelos compartimentais	39
3.4.3	Modelos baseados em aprendizado de máquina	41
3.5	Régressores	41
3.5.1	Régressão linear	41
3.5.2	Máquina de Vetor de Suporte	42
3.5.3	Rede Neural Perceptron Multicamadas	42
3.5.4	Random Forest	43
3.6	Considerações	43
4	PROPOSTA	44
4.1	Introdução	44
4.2	Modelos propostos	44
4.2.1	Modelo espaço-temporal - arboviroses	44
4.2.2	Modelo temporal - Covid-19	46
4.2.3	Modelo espaço-temporal - Covid-19	49
4.3	Base de dados de arboviroses	51

4.4	Base de dados de Covid-19 - abordagem temporal	51
4.5	Base de dados de Covid-19 - abordagem espaço-temporal	52
5	RESULTADOS E DISCUSSÕES	53
5.1	Predição espaço-temporal de arboviroses	53
5.2	Predição temporal de Covid-19	65
5.2.1	Predição ARIMA	65
5.3	Predição espaço-temporal de Covid-19	74
5.4	Aplicação web	76
6	CONCLUSÃO	80
6.1	Dificuldades apresentadas	80
6.2	Conclusões gerais	81
6.3	Contribuições	81
	REFERÊNCIAS	83
	Índice	96

1 Introdução

Nesta seção introdutória serão apresentados o contexto e a motivação, o objetivo e como este trabalho está organizado.

1.1 Contexto e motivação

A prevenção e o controle da dengue, da febre chikungunya e da zika têm sido um grande desafio de saúde pública para muitos países, especialmente a partir de 2015, uma vez que outras arboviroses passaram a interagir com o vírus da dengue, que se disseminou rapidamente nas últimas duas décadas (LIMA et al., 2016; BHATT et al., 2013). Estima-se que por volta de 390 milhões de novos casos de dengue ocorram a cada ano. Entretanto problemas como diagnósticos incorretos e registro impreciso 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). A emergência de outras arboviroses, como a febre chikungunya e a zika, em especial na América do Sul, levanta novos desafios ao monitoramento e controle do vetor. Essa situação passa por um agravamento a partir de 2015 e 2016, com a rápida difusão da chikungunya, provocando febre e fraqueza muscular, dentre outros sintomas, e com o surgimento do zika vírus, parcialmente relacionado com casos de microcefalia em recém-nascidos e diretamente relacionado à ocorrência da síndrome de Guillain-Barre, doença autoimune que afeta o sistema nervoso, provocando desde fraqueza muscular até a paralisia (CAO-LORMEAU et al., 2016).

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, mas também se faz cada vez mais presente na América do Norte e na Europa, devido à presença dos mosquitos *Aedes* e da introdução dos vírus (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 se dão essas interações 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; BHATT et al., 2013; 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). Condições climáticas locais, como temperatura, pluviometria e umidade, interferem no desenvolvimento do vetor, da 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).

No Brasil, as arboviroses têm recebido uma atenção especial do Sistema Único de Saúde por meio de políticas públicas de saúde e campanhas. 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 não somente os dados do LIRAA e do sistema de ovitrampas, como também o mapeamento das doenças e dos sintomas por unidade de saúde e pelo bairro de origem do paciente desde 2015. A Agência Pernambucana de Águas e Clima, APAC, também disponibiliza um sistema de informações geográficas onde são publicizadas as séries temporais pluviométricas diárias e mensais desde o ano de 2006, por cidade e, no caso do município do Recife, por bairro.

Fazendo uso da mineração dos dados de sistemas abertos, é possível inferir regras e expressões que podem servir para alimentar simuladores de epidemias. Esses simuladores podem ser baseados em sistemas de equações diferenciais ou em redes bayesianas, que por suas vez formam uma área da Inteligência Artificial que se concentra na modelagem de sistemas compostos por agentes autônomos. Diversos trabalhos utilizam o modelo geral de transmissão de doenças por mosquitos, baseado em equações diferenciais, e os modelos baseados em agentes (PADMANABHAN; SESHAIYER; CASTILLO-CHAVEZ, 2017; JINDAL; RAO, 2017; LIMA et al., 2016; LEGROS et al., 2016).

Em dezembro de 2019, o surto de uma doença nova semelhante a uma pneumonia deixou autoridades em alerta na cidade Wuhan, na China. A pneumonia atípica foi atribuída ao novo coronavírus, o SARS-CoV-2, e a doença denominada Doença do Coronavírus 2019, ou Covid-19 (SILVA et al., 2021b). Com a chegada do Festival de Primavera da China, período no qual muitos chineses tiram férias e viajam pelo país e pelo mundo, o que facilitou a propagação da doença. E em 30 de janeiro de 2020 a doença foi declarada uma Emergência de Saúde Pública de Interesse Internacional (PHEIC), em março de 2020 foi declarada uma pandemia pela Organização Mundial de Saúde(OMS) (LIMA et al., 2020; SILVA et al., 2021b). A Covid-19 apresenta alta taxa de transmissão, e nos primeiros 5 meses desde o primeiro caso da doença (dezembro 2019 a maio de 2020) já tinham sido contabilizados mais de 4,7 milhões de pessoas diagnosticadas em 216 países (WHO, 2020). Hoje, após quase um ano e oito meses já são mais de 202 milhões de casos confirmados e mais de 4 milhões de mortes em todo o mundo (WHO, 2021). Dentre os sintomas mais frequentes de Covid-19 se destacam tosse, dor de garganta, febre, fadiga e falta de ar, e em alguns casos, se não houver apoio médico precoce, a doença pode evoluir para uma pneumonia grave, levando a condições críticas, por exemplo, como sepse e síndrome respiratória aguda grave, e nesses casos o indivíduo pode vir a óbito (LIMA et al., 2020). Para detectar a infecção por Covid-19, dois exames são mais comuns: os testes rápidos e o RT-PCR (transcrição reversa seguida de reação em cadeia da polimerase); esse último é o padrão ouro. Contudo o RT-PCR precisa de horas para dar um diagnóstico. A fim de apoiar o diagnóstico da Covid-19, os pesquisadores Xie et al. (2020), Jin et al. (2020), Meng et al. (2020) e Barbosa et al. (2021) usaram modelos com métodos relativamente novos, como inteligência artificial e aprendizado de máquina.

Até hoje não se tem tratamento específico para a doença do coronavírus, e no início do surto não se tinha desenvolvido uma vacina. E, considerando a fácil locomoção do mundo globalizado, a estratégia para controlar a dispersão do vírus foram medidas de isolamento social e físico em grandes escalas adotados por governos nacionais e locais, como diminuição da capacidade máxima de pessoas em estabelecimentos e fechamento de centros comerciais, parques e restaurantes, bem como adiamento ou cancelamento de grandes eventos públicos. Além de colocar as pessoas infectadas em quarentena (LIMA et al., 2020). Os trabalhos de Day (2020), Salathé et al. (2020), Oliveira et al. (2020) mostraram

resultados positivos do isolamento social na Itália, Suíça e Brasil, respectivamente. Contudo uma recente revisão sistemática da literatura e meta-análise conduzida pelo professor Jonas Herby do Johns Hopkins Institute for Applied Economics, Global Health and the Study of Business Enterprise mostrou que a política de lockdown não tiveram efeitos na saúde pública, todavia esse tipo de política trouxe grandes impactos na economia e na dinâmica da sociedade (HERBY et al., 2022).

Não obstante, o número de casos aumentava, bem como o número de óbitos pela doença, com o intuito de diminuir os efeitos e evitar o colapso dos sistemas públicos de saúde, vários pesquisadores começaram a buscar formas de prever o número de casos, dentre eles Lima et al. (2020), Silva et al. (2021b), que estudaram predições de casos no Brasil, Tomar e Gupta (2020) usaram esses métodos para predição de casos na Índia. Também, para modelar a predição de casos e óbitos de Covid-19 na Índia, Sarkar, Khajanchi e Nieto (2020) e Khajanchi e Sarkar (2020) construíram novos modelos compartimentais, mas foram os dados de Wuhan, China, que foram utilizados (SILVA et al., 2021b).

1.2 Objetivos

Este trabalho teve como objetivo avaliar e construir uma ferramenta de predição da distribuição das arboviroses transmitidas pelo mosquito *Aedes aegypti* por meio da implementação de simuladores da transmissão da dengue, da zika e da febre Chikungunya baseados em sistemas de equações diferenciais e técnicas de aprendizado de máquina. Para tanto, também foi construída uma ferramenta para mineração de dados para extrair informações de interesse epidemiológico e geográficas do Portal de Dados Abertos da Prefeitura do Recife, cujos dados foram utilizados não somente para modelagem, mas também para validação dos modelos computacionais implementados. Com informações de longitude e latitude, os dados epidemiológicos e climáticos obtidos, foram gerados mapas de distribuição de criadouros no software QGIS. O software Weka foi utilizado para avaliar os modelos de regressão linear, random forest, máquinas de vetor de suporte (SVM) e redes neurais multilayer perceptron, além de fazer a predição dos mapas de criadouros.

Nossa hipótese foi de que, com os mapas de distribuição de seis semestres consecutivos, pudéssemos predizer o mapa de distribuição do bimestre seguinte. As ferramentas adicionais foram construídas em software livre, usando as linguagens de programação Matlab/Octave e Python. Os resultados das simulações foram também exportados em tabelas, para uso em sistemas de informações geográficas, com a finalidade de visualizar os dados simulados e contrapor os resultados das simulações com os dados reais de disseminação dos ovos do mosquito e das arboviroses e seus sintomas.

Para Covid-19, o objetivo foi criar uma metodologia para monitoramento e predição de casos e óbitos de Covid-19 para cada um dos estados brasileiros. Foram criados modelos

com duas abordagens: uma temporal e uma espaço-temporal. Para a abordagem temporal, modelos ARIMA são interligados com os dados do número de casos confirmados e óbitos de todos os estados brasileiros obtidos no portal Brasil.io, sendo mostrados os padrões temporais crescentes dos casos de Covid-19. Já para a abordagem espaço-temporal, a partir de dados de casos confirmados e óbitos de cada uma das cidades brasileiras, obtidos no portal Brasil.io, e suas respectivas informações de latitude e longitude, foram criados mapas de distribuição da doença. No software Weka, foram avaliados os algoritmos regressores de máquina de vetor de suporte (SVM) e redes neurais multilayer perceptron, para avaliação e criação dos mapas de predição da distribuição espacial dos casos de Covid-19 nas cidades brasileiras. Em ambas as abordagens foi possível realizar as tarefas de vigilância e predição em tempo real.

Essas abordagens relacionam saúde digital, modelagem estatística, sistema de informações geográficas (SIG) e ciência da computação, a fim de criar um sistema que consiste em uma aplicação web que seja dinâmica, com interface amigável e com múltiplos banco de dados para relatar casos confirmados e óbitos pela doença do coronavírus. Com o protótipo da ferramenta assim construído, foi possível obter informações de predição de casos a nível nacional, estadual e municipal.

1.3 Organização do trabalho

Esta dissertação está estruturada da seguinte forma:

No Capítulo 2, é apresentado o estado da arte dos métodos de predição utilizados, das bases de dados utilizadas, bem como as técnicas computacionais existentes. Nesse capítulo também é realizada uma análise comparativa do estado da arte com a metodologia proposta neste trabalho.

No Capítulo 3, são apresentados os conceitos sobre arboviroses, epidemiologia, epidemiologia digital e geoprocessamento. Também são apresentadas os métodos de predição de arboviroses, bem como uma breve apresentação das técnicas de aprendizado de máquina.

O Capítulo 4 apresenta a metodologia adotada durante o estudo. Neste capítulo, também são detalhadas as bases de dados que foram utilizadas e os regressores usados.

No Capítulo 5, são apresentados e discutidos os resultados experimentais 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 gerais a respeito dos trabalhos desenvolvidos até o presente momento, as dificuldades apresentadas e as principais contribuições desse trabalho.

2 Trabalhos relacionados

2.1 Introdução

Com o avanço do conhecimento médico no século XIX e a fim de entender como se dá a evolução das doenças contagiosas, muitos cientistas voltaram seus esforços à modelagem de surtos de epidemias. Várias abordagens foram criadas, dentre elas modelos matemáticos, modelos estatísticos, e mais recentemente, modelos baseados em aprendizado de máquina e modelos híbridos. Neste capítulo serão apresentados alguns estudos que modelam as arboviroses como dengue, zika e chikungunya e a Covid-19.

2.2 Predição de arboviroses

As arboviroses são um problema de saúde pública de vários países, em especial países de clima tropical e subtropical, e estão entre as doenças tropicais negligenciadas (DTNs). A fim de entender como as dengue, zika e Chikungunya se desenvolvem e, claro, para controle e predição de surtos dessas doenças, alguns pesquisadores criaram sistemas baseados em modelos matemáticos, estatísticos e de aprendizado de máquina para auxiliar na tomada de decisão de saúde pública. Nesta seção serão apresentados alguns estudos que fizeram modelagens para controle e predição de dengue, zika e chikungunya em alguns países do mundo.

O uso de modelos compartimentais baseados em equações diferenciais são bastante usuais, na grande maioria se baseia no modelo SIR clássico e suas variações. Agora serão apresentados os trabalhos que fizeram uso desse tipo de abordagem.

Um modelo compartimentado SIR determinístico baseado em equações diferenciais ordinárias (EDOs) foi proposto por Subramanian et al. (2020). Foram desenvolvidas linhas de evidência para investigar a incerteza e a possibilidade do tempo de reemergência para doenças com número de reprodução básico R_0 baixo. Como dados de entrada, foram usadas as estimativas mensais de casos de dengue na cidade do Rio de Janeiro, no Brasil, no período de 1986 a 1990. Esse trabalho propôs um modelo SIR com sazonalidade e forçado com surtos intermitentes, e rotatividade populacional, para considerar as características gerais de reemergência para baixo número de reprodução R_0 . Estendendo esse trabalho, Stone, Olinky e Huppert (2007) incluíram o crescimento da população, a sazonalidade sinusoidal (contínua) e os casos de subnotificação. Foi usada uma função cosseno para ajustar o modelo SIR estocástico uma simplificação para representação da força sazonal que seria criada pela variação climática pela mudança nos mosquitos infectados. Os

pesquisadores trazem duas abordagens. Numa delas, eles trazem uma comparação entre a média sazonal resultante do modelo com valores do R_0 estimado diretamente apoiado nos dados de série temporais na literatura para DENV1 e DENV4 no Rio de Janeiro de 2010 a 2016. A outra abordagem considera um modelo SIR baseado na temperatura e feita uma simulação estocástica usando os dados e predição numérica dos tempos esperados para a reemergência.

Os modelos $S_h E_h A_h I_{h_1} I_{h_2} R_h$ e $S_m E_h I_h$ descritos por sistema de EDOs foram construídos por Musa et al. (2019) para entender a epidemia em Kaohsiung e Tainan, Taiwan, na China. Foi analisado o equilíbrio livre, endêmico e global da doença, equilíbrio do sistema e bifurcação e da sensibilidade do sistema. As estruturas de inferências plug-and-play foram usadas para a simulação numérica. E os dados de séries temporais de casos de dengue entre 2014 e 2016 nas cidades Kaohsiung e Tainan foram usados para o ajuste do modelo. Os autores constataram que a bifurcação reversa pode ser removida usando $R_0 = 1$. Por meio do Critério Akaike (AIC), o melhor desempenho de ajuste foi feito usando a abordagem de onda quadrada não sincronizada comparada à abordagem de spline cúbica. Também com a proposta de um modelo compartimentado SIR modificado representado por um sistema de EDOs foi a proposta de Abidemi, Aziz e Ahmad (2020) para descrever a relação vetor-hospedeiro na presença de duas cepas do vírus da dengue na Ilha da Madeira, Portugal. A vantagem desse modelo foi integrar a vacina e adulticida para estudar a estratégia de controle. Assim, o número básico de reprodução, a existência de equilíbrio, a determinação da constante de controle e as análises de sensibilidade são feitas e analisadas matematicamente. As simulações numéricas foram feitas usando a função ode45 do MATLAB no intervalo de 365 dias.

Para estudar a transmissão de dengue na Ilha do Governador, Rio de Janeiro, Brasil, os pesquisadores Lima et al. (2016) propuseram uma ferramenta de código aberto para apoiar o desenvolvimento de modelos com abordagem espaço-temporal para simular dengue e seu vetor usando Sistema de Informações Geográficas (GIS). A ferramenta se chama DengueME (Dengue Modeling Meio Ambiente). O modelo clássico SIR SI foi um dos vários desenvolvidos por esse estudo. E com o DengueME é possível, por meio do VDE, criar e analisar o cenário do modelo. O modelo matemático-computacional é usado para explicar a dinâmica espaço-temporal da ecologia dos vetores e da transmissão da doença. A ferramenta conta com interface amigável, permitindo que os usuários criem e personalizem cenários por meio de modelos integrados. Conta com uma arquitetura extensível permitindo a adição de novos modelos. É possível fazer testes, adaptações, extensões e integrações e comparações de diferentes modelos e abordagens de modelagem para doença e seus vetores, assim os professores podem usá-lo como recurso pedagógico. Contudo não há soluções voltadas para a modelagem da dengue com dinâmica local em escalas intra-urbanas. Kao e Eisenberg (2018) trouxeram uma variação do modelo compartmental clássico, o ASEIR SEIR, que conta com a classe A de larvas e pulpas. É

analizada a capacidade (ou não) de fornecer insights (identificabilidade) sobre os parâmetros desconhecidos, estrutural e prática dos parâmetros, além de uma estimativa do R_0 . Essa identificabilidade foi analisada por meio da implementação numérica que usou vários cenários e condições para o problema. A implementação foi feita usando o Nelder-Mead de NumPy em Python 2.7.10. Constatou-se que, para que o modelo seja consistente com os resultados de álgebra diferencial (estruturalmente identificável) e abordagens FIM (Fisher information matrix), o modelo não é praticamente identificável.

Suparit, Wiratsudakul e Modchang (2018) analisaram a propagação do Zika vírus entre 2015 e 2016 na Bahia, Brasil. Foi proposto o modelo compartimental SEIR SEI (humano e vetor) que busca um valor de mordida semanal mais adequado ajuste entre o resultado da simulação e dos dados (reais) semanais relatados correspondentes. Os dados do número total de casos notificados durante o surto na Bahia foram utilizados para calibrar o modelo. Entre 2015-2016, os dados foram estaduais. Após maio de 2016, foram usados nacionais (junho a dezembro de 2016). Os coeficientes de correlação entre as taxas de picada de mosquitos e temperatura foram calculados usando o MATLAB R2017a. Como resultado, verificou-se que a dinâmica do Zika vírus é menos sensível à mudança do tamanho da população do que a mudança na taxa semanal de picada de mosquitos. O estudo também se concentrou na predição dos casos de zika vírus semanais, além da simulação de dois picos da doença na Bahia. Os autores usaram como hipótese nesse modelo que o ZIKV se comporta como os vírus da dengue e da chikungunya por exemplo, que há homogeneidade espacial no contato e na transmissão e não considera a transmissão sexual da doença.

Seguindo o estudo anterior, Ndaïrou et al. (2018) fizeram da modelagem e predição de casos de Zika em gestantes e recém nascidos no Brasil seu objeto de estudo. O trabalho propôs um modelo SEI SEIR com 11 compartimentos descritos por um sistema de EDOs. Foram feitas estimativas dos parâmetros epidemiológicos e do R_0 . O método de coeficientes de correlação foi usado para classificação parcial para sensibilidade global e para análise de identificabilidade. O princípio do máximo de Pontryagin e a teoria do controle ótimo foram usados para propor o modelo. Bem como o estudo anterior, o MATLAB foi usado para as simulações numéricas. O método usado foi o Runge-Kutta. Para tal implementação, foram usados os dados semanais de casos no Brasil em 2016, ou seja, 52 semanas. O modelo não é afetado por pequenas variações, portanto é robusto, e é multifacetado. Assim, ele se mostra útil para apoiar a tomada de decisões pelos gestores de saúde pública.

Kumar et al. (2017) desenvolveram um estudo da progressão dinâmica do surto de dengue nos anos de 2015 e 2016 em El Salvador. Foi proposto um modelo determinístico SEI SEIR representado por um sistema de equações. E para este sistema foi analisado o equilíbrio livre de doença e o R_0 . O método Runge-Kutta de quarta ordem (RK4) foi usado para fazer a simulação numérica no MATLAB. Os pesquisadores concluíram que houve uma

boa correlação entre as simulações e os dados reais de casos de El Salvador. Os parâmetros que afetam a disseminação do surto também foram apresentados, sendo que o parâmetro mais importante que foi a proporção do mosquito-humano. A implementação numérica foi feita no MATLAB e usada para prever a progressão do ZIKV em regiões geográficas distintas. Assim, é possível usar essas informações pra predição de surtos e medidas de controle da doença. No entanto o estudo não leva em consideração a transmissão sexual e da gestante para o feto e a heterogeneidade da infecção do ZIKV.

O modelo compartimental SEI SVEIR, descrito por equações diferenciais, foi desenvolvido por Usman, Adamu e Tahir (2017) para compreender e definir os parâmetros importantes na transmissão do Zika vírus. Esse modelo introduz um compartimento de indivíduos vacinados no clássico SEI SEIR. Uma análise matemática foi feita para analisar o equilíbrio livre de doença e o número de reprodução básica R_0 . Também foi calculado o índice de estabilidade, e assim foi possível estudar a importância de cada parâmetro do modelo da disseminação da infecção por ZIKV. Esses parâmetros aumentam ou diminuem à medida em que aparecem novos casos de indivíduos infectados na população hospedeira. Portanto, esse estudo pode auxiliar na tomada de decisão a cerca do controle da propagação do vírus a partir desses parâmetros.

Já para entender a propagação da doença na China, Bonyah et al. (2017) propuseram um modelo SEI SEIR baseado em um sistema de equações diferenciais. O princípio do Máximo de Pontryagin foi usado para definir quais são as condições de controle da Zika. Uma análise matemática do equilíbrio (local) e estabilidade livre e endêmico da doença, e análise e existência da bifurcação, além do equilíbrio endêmico de estabilidade local e global. Também foi feita a análise de sensibilidade do número de reprodução básico R_0 e controle ideal via Máximo de Pontryagin. Foi usado o MATLAB para implementações numéricas. A teoria da variedade central foi usada no estudo da estabilidade do equilíbrio endêmico, e concluíram que o equilíbrio é assintoticamente estável.

Para modelar estratégias para surto de Chikungunya no Rio de Janeiro, Dodero-Rojas et al. (2020) foi estabelecido um modelo compartimental SEIR. As análises matemáticas foram feitas com dados de 2016, 2017 e 2019. O número de reprodução básico (R_0) foi estimado, usando o método da matriz de próxima geração foi empregado para o número de reprodução R_T foi constado. Também foi feita uma análise de sensibilidade, e nessa análise foram feitas mil simulações com o método Cadeia de Markov de Monte Carlo com amostragem em torno de 5% do valor do melhor ajuste dos dados da epidemia de Chikungunya de 2018.

Jing et al. (2018) desenvolveram um estudo avaliando variáveis como número de casos importados e temperatura mínima na transmissão local da dengue em Guangzhou, na China, por meio de um modelo estatístico. O modelo, apesar de bons resultados, unicamente identifica a correlação dos casos de dengue e as variáveis de risco, contudo

não há identificação das relações causais entre as variáveis. Para mais, o modelo não leva em consideração fatores socioeconômicos e os efeitos da subnotificação de casos.

A temperatura, pressão atmosférica e umidade relativa foram analisadas pelos pesquisadores Zhu et al. (2019). Eles observaram a relação entre os casos de dengue e a média, máxima e mínima temperatura, a média da pressão atmosférica e a média da umidade relativa do ar. Chegaram a conclusão que a temperatura mínima adequada para a transmissão da doença é acima de 18°C. Ainda como resultado foi obtido $R^2 = 0.6$ e 0.8104 de coeficiente de correlação, que é um bom resultado. Porém o ajuste do modelo com os casos observados foi ruim, e não foram apontadas as métricas para avaliação dos erros de predição. E o modelo tem algumas limitações como não apresentar fatores socioeconômicos e espaciais da região em estudo.

Com o objetivo de auxiliar com alertas antecipados, alocação de recursos, comunicação de risco e controle de doença, os autores Chan, Hu e Hwang (2015) trabalharam na predição do risco local de casos de dengue. Como base de dados de casos confirmados foram coletados do National Notifiable Diseases Surveillance System, entre 2009 e 2012. Dados de densidade populacional e variáveis ambientais também foram usados. O modelo foi desenvolvido usando regressão logística para predição da probabilidade de incidência de surtos de dengue na região determinada. Como resultado foram obtidos 83% de sensibilidade média para presença do surto, com 23% de média de falso positivo. Esse estudo trouxe um modelo robusto e simples, com bom desempenho de predição do risco de casos de dengue em uma determinada área.

Árvore de decisão (*Decision Tree*, DT), k vizinhos mais próximos (*k-Nearest Neighbours*, kNN), rede neural perceptron multicamadas (MLP) e máquinas de vetor de suporte (SVMs) foram usados para criar um sistema de predição da taxa de morbidade da Febre Hemorágica da Dengue (DHF) e taxa de infecção por *Aedes aegypti* em Akhon Pathom, Ratchaburi e Samut Sakhon, na Tailândia, por Kesorn et al. (2015). O problema, que originalmente seria feito usando regressão, foi transformado em um problema de classificação da taxa de mortalidade é alta ou baixa. A sensibilidade, especificidade e a acurácia foram usadas como métricas para essa classificação. Foram usadas quatro bases de dados (mensais) para predição: distribuição de densidade populacional, fornecida pelo Ministério do Interior da Tailândia; taxa de infecção por *Aedes aegypti* (larvas, população de mosquitos machos e de fêmeas, em percentual), fornecida pela Universidade de Chulalongkorn; fatores climáticos (temperatura, umidade, velocidade dos ventos e pluviometria), fornecida pelo Departamento de Meteorologia Tailândês; e distribuição espaço-temporal dos casos de dengue, fornecida pelo Birô de Epidemiologia. O período de estudo foi de 2007 a 2013. Foi calculada a taxa de mortalidade com a distribuição de casos de dengue. E foi determinado o limiar da taxa que deve ser considerada alta. A implementação foi feita usando o software Weka. Foi usada a rede neural supervisionada MLP com uma camada escondida (acredita-se que foram

utilizados 11 neurônios na camada escondida. Também foi usada a máquina de vetor de suporte (SVMs) com configurações de kernel linear, polinomial (não foi especificado o grau do polinômio) e de função de base radial(RBF). E testes com validação cruzada. Foram obtidos melhores resultados com SVM com kernel RBF, sendo 0,882 de sensibilidade, 0,789 de especificidade e 90,740% de acurácia. Porém o modelo traz algumas limitações, como a imprecisão de classificar apenas a taxa de mortalidade entre alta ou baixa. Será que isso poderia ser melhorado a fim de apoiar a tomada de decisões por gestores de saúde? Além disso também não foi levado em consideração o peso de cada fator na construção do modelo de predição.

Uma predição do surto de dengue feita por Shaukat et al. (2015) meio da mineração de dados de informações clínicas fornecidas por unidades de saúde do Paquistão. O estudo usou febre, fadiga, mialgia (dor muscular), gripe muito forte hemorragia e outros sintomas, e o diagnóstico de presença ou não de dengue como atributos. Esses dados foram coletados do Sistema de Saúde Pública do Paquistão e corresponde a dados do município de Jhelum, e corresponde a registros de 95 pacientes, e são informações binárias. O problema é modelado como uma classificação, e foram usados os seguintes classificadores: classificador ingênuo de Bayes, árvores de decisão (J48, Random Tree e Rep Tree) e SVM linear. Taxa de falsos positivos, taxa de falsos negativos, recall, acurácia, medida F e área da curva ROC foram usados para avaliar os classificadores. O classificador ingênuo de Bayes apresentou os melhores resultados, e como esse classificador considera que os atributos são estatisticamente independentes, a modelagem é simples, o que não ocorre na maioria dos problemas reais. O estudo apresenta algumas limitações como não considerar a localização geográfica do paciente e os fatores climáticos e ambientais da região de estudo. Além disso, o sistema se mostra muito mais um sistema de apoio ao diagnóstico do que um sistema para predição de surtos, que foi a proposta inicial.

2.3 Predição de covid-19

Com o surto de Covid-19 a partir de 2019, os pesquisadores iniciaram esforços propondo diversos sistemas e soluções de monitoramento e predição de casos confirmados e óbitos desta doença com o intuito de assistir a tomada de decisões por meio dos gestores de saúde de todo o mundo. Os modelos matemáticos contribuem para análise e compreensão de como a doença se comporta, mas esse tipo de modelo traz uma abordagem qualitativa, mas era imprescindível uma abordagem quantitativa para guiar os próximos empenhos.

Os modelos compartmentais são modelos baseados em equações diferenciais, e podem ser de dois tipos de equações, as equações diferenciais ordinárias (EDOs) e as equações diferenciais parciais (EDPs). Para resolução de cada um desses modelos

são usadas técnicas e métodos de simulações numéricas diferentes. A seguir, serão apresentados alguns estudos que fizeram uso de modelagem matemática para estudar o comportamento e fazer previsões usando técnicas computacionais associadas a cada tipo de modelo.

Sarkar, Khajanchi e Nieto (2020) propuseram um modelo compartimentado baseado em EDOs dispostos em 6 compartimentos que estende o SEIR clássico para previsão. Para encontrar o parâmetro relativo mais sensível à população infectada foi feita uma análise de sensibilidade, e se usou a técnica do coeficiente de correlação (PRCC) para os parâmetros referentes aos indivíduos infectados. Para implementação numérica foi usado o método dos mínimos quadrados (MMQ) no programa FORTRAN visando ajuste dos casos confirmados diários da doença. Segundo a mesma linha, Suba et al. (2020) desenvolveram sete modelos, além da análise de sensibilidade, foi feita a implementação numérica e também usaram o método MMQ além de planilhas excel e gráficos plotados no MATLAB. Porém, se houver uma mudança no número de reprodução básico R_0 todo o sistema muda automaticamente, isto é, é um sistema sensível. Também usando o software MATLAB para os cálculos numéricos, os pesquisadores Zhong et al. (2020) trouxeram um sistema com previsões de da epidemia com diferentes cenários, diferentes níveis de medidas anti-epidêmica e de assistência médica e fazendo uso de dados reais. Contudo a quantidade de dados (foram usados dados até 18 de janeiro de 2020) e sua credibilidade limitam esse estudo.

Mandal et al. (2020) propuseram um modelo SEIQR, e as simulações numéricas foram feitas com o método Runge-Kutta de quarta ordem (RK4) por meio do software MATLAB, e um análise de estabilidade e estimativa do R_0 . E a previsão feita é de curto prazo. Com o mesmo software, porém usando o método de Netlog para construir uma simulação da transmissão do vírus e a função ode45 para a solução numérica e ajuste de curvas, foi criado por Jiang X e Zhao (2020). Os valores obtidos pela função foram bem pertinentes com os dados reais.

Massonis, Banga e Villaverde (2020) fizeram uma revisão com 255 artigos recentes, sendo 98 com modelos compartimentais SIR e 157 com modelos compartimentais SEIR, a fim de avaliar a identificabilidade estrutural e a observabilidade (estados não medidos), ou seja, avaliar a capacidade de fornecer informações confiáveis, usando conceitos teóricos de identificabilidade estrutural e confiabilidade. O STRIKE-GOLDD, GenSSI2 do MATLAB, código *Observability Test* no Maple, *Identifiability Analysis* no Mathematica, SIAN no Maple e outros. Os estudiosos relataram que a maioria dos modelos têm parâmetros identificáveis. Essa revisão traz uma análise em detalhes da identificabilidade estrutural e observabilidade de um grande conjunto de modelos compartimentados SIR e SEIR de Covid-19 apresentados na literatura recente.

Com a intenção de modelar e predizer a evolução da Covid-19 no Brasil, Bastos e

Cajueiro (2020) propuseram dois modelos SIRD e SIRASD descritos por sistema de EDOs. As soluções numéricas para o sistema de equações foi feita no método Runge-Kutta de ordem 5 (RK5) com implementação *solve_ivp* da biblioteca *scipy Python*. Foram utilizados dados do Instituto Brasileiro de Geografia e Estatística (IBGE) para condição inicial e os dados do Ministério da Saúde (25 de fevereiro a 30 de março de 2020). Foi percebido que o número de infectados que o modelo SIRAD prevê que é maior do que o estimado pelo modelo SIRD, além disto este último prevê um pico mais baixo para os infectados com sintomas, e estes são justamente os que requerem cuidados médicos. Esse estudo fez uso de informações de dados demográficos, levando em conta seus efeitos para dispersão do vírus, e também pressumiu que não havia reinfecção. Foi verificada também a sensibilidade do modelo SIRADS a condição inicial de indivíduos assintomáticos. E por fim, o estudo trabalhou com hipóteses, visto que, a testagem durante o período de estudo foi pequeno.

A fim de trazer uma visão do número de casos confirmados da doença, Dong, Du e Gardner (2020) propuseram um painel interativo online que apresentasse os números de casos confirmados e mortes pelo doença do coronavírus em tempo real. Já com a abordagem de classificação de pacientes infectados e análise de indivíduos que estão mais vulneráveis, os cientistas de Xie et al. (2020) proporam um modelo para predição clínica. Para isso foi usado regressão logística multivariável para prever a mortalidade de pacientes, e assim poder usar os recursos da saúde, já que estes são limitados, e calcular a taxa de sobrevivência dos pacientes. Também seguindo essa linha de uma calculadora online S-COVID-19-P foi construído por Feng et al. (2020) com o objetivo de auxiliar o diagnóstico precoce pacientes com febre com suspeita de pneumonia por Covid-19, e feito usando a técnica regressão LASSO. Um sistema para diagnóstico rápido de algumas doenças, dentre elas Covid-19, foi proposto por Jin et al. (2020). Esse sistema foi baseado em aprendizado profundo, LASSO, para encontrar as 12 características mais diferenciam Covid-19 de outras doenças, e conta com precisão comparável a de radiologistas experientes, sendo possível fazer a classificação com precisão de penseumonia, CAP(Pneumonia Adquirida na Comunidade), influenza A e B e Covid-19. Tal como esse estudo, Gomes et al. (2020) desenvolveram uma ferramenta inteligente que por meio análise de imagens de raio-X de tórax pode apoiar o diagnóstico de Covid-19. Esses pesquisadores desenvolveram esse sistema favorável para diferenciar a pneumonia de Covid-19 e pneumonia bacteriana ou viral por meio do método Random Forest para diferenciação de imagens baseadas em textura e classificação. Além dessa, outros estudos usando análise e extração de atributos de raio-X foram feitos nos seguintes estudos: Ismael e Şengür (2021), Basu, Mitra e Saha (2020), Apostolopoulos e Mpesiana (2020), Apostolopoulos, Aznaouridis e Tzani (2020), Luz et al. (2020), Khan, Shah e Bhat (2020), Civit-Masot et al. (2020), Wang, Lin e Wong (2020), Jain et al. (2020), Maghdid et al. (2020). Contudo soluções baseadas em aprendizagem profunda e recursos de textura foram evitados em Gomes et al. (2020), para oferecer um ambiente computacional de baixo custo, online e que suporte vários usuários simultâneos

sem sobrecarregar os recursos da rede.

O Covid-19 Diagnostic Aid APP foi a proposta de Meng et al. (2020) para calcular a probabilidade de infecção por Covid-19 e feito com uso de testes laboratoriais simples e fáceis. Assim, é possível auxiliar na realização do controle e seleção dos pacientes que mais precisam de cuidados médicos, otimizando o processo de diagnóstico e economizando os recursos médicos. Também fazendo uso de análise sanguínea de 24 exames de sangue. Barbosa et al. (2021) criaram o HegIA, um sistema inteligente baseado em Redes bayesianas e Random Forest para apoiar o diagnóstico de Covid-19. Esse sistema é completamente funcional, tem bom desempenho e se assemelha ao do teste RT-PCR (Reverse Transcription Polymerase Chain Reaction) para indivíduos sintomáticos, está disponível gratuitamente e fornece testes rápidos com custo baixo.

No decorrer da pandemia também foram construídos modelos híbridos como, por exemplo, Putra e Mu'tamar (2019), Mbuvha e Marwala (2020), Qi et al. (2020), Salgotra, Gandomi e Gandomi (2020b), para criar ajustes de modelos compartimentais usando Computação Evolutiva e Método de Inteligência de Enxames. Putra e Mu'tamar (2019) utilizaram o algoritmo de Otimização de Enxames de Partículas (PSO) para fazer estimativas do modelo SIR, e assim é possível ajustar este modelo com precisão comparado a outras abordagens analíticas. Já Mbuvha e Marwala (2020) usaram dados de casos de Covid-19 da África do Sul e analisando vários cenários diferentes do número de reprodução básico R_0 para relatar infecções e estimativa dos recursos de saúde locais. Os pesquisadores assumiram algumas hipóteses como, por exemplo assumiram que os casos confirmados representam cerca de 0,2% a 1% da população infectada, parâmetros fixos do modelo compartimental SIR. E para trabalhos futuros usar técnicas bayesianas usando Cadeia de Markov Monte carlo para tratar as incertezas em torno dos parâmetros do SIR que foram verificadas durante o estudo. Usando parâmetros de temperatura diária (AT) e umidade relativa (ARH) para investigar a influência desses parâmetros na ocorrência de Covid-19 em 31 províncias chinesas, sobretudo em Hubei que foi a cidade com a primeira notificação da doença, foi feito o estudo Qi et al. (2020). Apesar desse estudo ser feito levando em consideração dados meteorológicos como temperatura média diária (AT) e umidade relativa média do ar (ARH) de cada capital provincial, a associação com os casos de Covid-19 não foram considerados consistentes, visto que há uma heterogeneidade espacial na Covid-19, além do que a China é um país de dimensões continentais não tem uma homogeneidade de AT e ARH diários.

A programação genética (GP) foi usada por Salgotra, Gandomi e Gandomi (2020b) na construção de modelos de predição para casos confirmados e óbitos nos três estados da Índia que mais foram afetados: Maharashtra, Gujarat e Delhi. Esse estudo traz uma análise da impotência das variáveis de predição e para avaliar e validar esses modelos foram usados parâmetros e métricas estatísticas. O método proposto se mostrou altamente confiável na

predição de casos da doença do coronavírus 19 na Índia. A fim de entender e analisar artigos de pesquisa técnica usando algoritmos de Inteligência Computacional para previsão de Covid-19, Rahimi, Chen e Gandomi (2021) fizeram uma revisão sistemática da literatura. Foi feita uma pesquisa por publicações usando as palavras chaves “forecasting”, “prediction”, “Covid-19” e “coronavirus” nas bases Web of Science(WoS) e na Scopus. Nessa busca foram selecionados 920 artigos apresentando apenas descrições algorítmicas, artigos de revisão, artigos de conferências, estudos de caso e capazes de fornecer insights gerenciais, publicados até 10 de outubro de 2020. E como resultado dessa revisão os artigos de revisão foram classificados com relação aos algoritmos usados para criação dos modelos:

- Média móvel simples
- *Auto-Regressive Integrated Moving Average (ARIMA)* (MOFTAKHAR; MOZHGAN; SAFE, 2020; ROY; BHUNIA; SHIT, 2020; ALZAHRANI; ALJAMAAN; AL-FAKIH, 2020; KUFEL, 2020; SINGH et al., 2020)
- Distribuições de duas peças com base na escala (MALEKI et al., 2020)
- Função logística: funções me forma de S para modelar curvas epidemiológicas (CHEN; CHEN; CHEN, 2020; LI; FENG; QUAN, 2020; QEADAN et al., 2020)
- Método de Regressão (JI et al., 2020; RIBEIRO et al., 2020; SUJATH; CHATTERJEE; HASSANIEN, 2020; VELÁSQUEZ; LARA, 2020; ALMESHAL et al., 2020)
- Redes Neurais Canônica (MOFTAKHAR; MOZHGAN; SAFE, 2020; FONG et al., 2020; TAMANG; SINGH; DATTA, 2020)
- Métodos de aprendizado profundo baseado em Redes Neurais Convolucionais (CNN) (SCHMIDHUBER, 2015; BENGIO; COURVILLE; VINCENT, 2013; LECUN; BENGIO; HINTON, 2015)
- Métodos de aprendizado profundo baseado em Redes neurais de curto prazo (LSTM) (AYYOUZADEH et al., 2020; CHIMMULA; ZHANG, 2020)
- Programação genética (SALGOTRA; GANDOMI; GANDOMI, 2020b; SALGOTRA; GANDOMI; GANDOMI, 2020a)
- Modelos compartimentais modificados: SIR, SEIR e SIRD. (KERMACK; MCKENDRICK, 1927; CAPASSO; SERIO, 1978; PENG et al., 2020; AHMAR; VAL, 2020)

Para entender as tendências da curva de evolução da doença, Tamang, Singh e Datta (2020) usaram redes neurais artificiais para ajustar e prever casos de infecção e óbitos por Covid-19 nos países Índia, EUA, França e Reino Unido, considerando as tendências progressivas da China e da Coreia do Sul. Nesse estudo, foram analisados três cenários

diferentes. As técnicas usadas foram redes neurais artificiais multilayer perceptron, com dados de fontes confiáveis. E como resultado foi visto que a tendência da China poderia melhorar se tomadas medidas mais rígidas de isolamento, confinamento e distanciamento social. Para a tendência da Coreia do Sul seria melhor um controle mais lento, visto que tem menos casos notificados. Assim, o estudo conclui, por meio de resultados experimentais, que redes neurais artificiais preveem com erros pequenos os casos futuros de Covid-19 em praticamente qualquer país.

A ferramenta COVID-SGIS foi apresentada por Lima et al. (2020) para monitoramento, predição e visualização de casos e óbitos de Covid-19 em tempo real no Brasil e suas unidades federativas. O sistema se baseia em um aplicativo web que usa banco de dados Brazil.io, do qual ele captura informações de casos confirmados e óbitos relatados diariamente das 27 unidades federativas brasileiras. Foram criados modelos ARIMA de séries temporais foram integrados pra predição do cumulativos de casos confirmados e óbitos de Covid-19. Esse modelo se baseia em uma predição dinâmica, e que conta com predição de 6 dias para o Brasil e para cada um dos 27 estados brasileiros. São apresentados resultados gráficos de 3 tipos diferentes, que mostra o melhor e pior cenários com um intervalo de confiança de 95%. A variação do erro quadrático médio percentual entre os valores previstos e os valores reais ficou entre 2,56 e 6,50%. E quando os valores saiam do intervalo de precisão os erros percentuais ficaram abaixo de 5% com relação ao pior cenário. E os resultados obtidos foram bons e esse método de predição dinâmica pode orientar a tomada de decisões por gestores de saúde pública, bem como, basear o planejamento de futuras ações que interferem na saúde e bem estar dos cidadãos.

2.4 Considerações

As arboviroses como dengue, Zika e chikungunya são doenças que provocam danos graves, em especial em países tropicais, com surtos que ocorrem sazonalmente, provocando mortes e sequelas nem sempre reversíveis. A Covid-19 é uma doença nova, que teve seu primeiro caso relatado em dezembro de 2019 e que em março de 2020 foi declarado com pandemia e que dura até hoje. Essa doença provocou mortes em todo o mundo, além de sequelas muitas vezes irreversíveis, e começou uma busca para compreender e minimizar seus efeitos. O controle e a predição de casos dessas doenças se mostram cada vez mais importantes, além da predição de localização (real ou potencial) de criadouros do mosquito, no casos das arboviroses, e estudar as medidas de distanciamento e isolamento social, além do controle sanitário e vacinação, já que as vacinas para Covid-19 já estão disponíveis.

Assim, o desenvolvimento de ferramentas baseadas em modelos compartimentais, estatísticos e de aprendizado de máquina, como os apresentados neste capítulo podem auxiliar a tomada de decisão de forma rápida e a fim de reduzir o possível número de

danos causados, além de contribuir para futuras medidas de políticas de saúde pública para prevenção e para um melhor direcionamento dos recursos.

3 Fundamentação Teórica

Neste capítulo é apresentada a fundamentação teórica com alguns conceitos de epidemiologia digital, fundamentos básicos das arboviroses transmitidas pelo *Aedes aegypti* que foram estudadas neste trabalho e alguns métodos de predição de arboviroses. Os métodos de predição se dividem em modelos de aprendizado estatístico, modelos compartimentados e modelos baseado em apredizados de máquina e seus princípios serão apresentados nesse capítulo também.

3.1 Epidemiologia Digital e Geografia da Saúde

O estudo de doenças que ocorrem com frequência, suas causas e o motivo pelo qual ocorrem é definido como Epidemiologia (COGGON; BARKER; ROSE, 2009; SALATHÈ, 2018). Com esse estudo se busca entender qual a dinâmica da doença, seus padrões e como ela se dissemina numa população. Com esses padrões determinados focar na prevenção da doença e promoção da saúde para a população (SALATHÈ, 2018). Desde os anos 90, os dados que eram armazenados em documentos físicos passaram a ser armazenados digitalmente, em formas de planilhas em hospitais e clínicas médicas. Nos anos 2000 com os avanços da tecnologia e com a chegada da Revolução 4.0, também conhecida como Revolução dos Dados, o fluxo de dados não só teve um aumento considerável como começaram a ser gerados fora da saúde pública (SALATHÈ, 2018). Essa Revolução Industrial 4.0 marcada por grandes avanços das tecnologias e técnicas computacionais, dentre elas as técnicas de mineração de dados. Assim, essa revolução digital tornou possível uma nova forma de estudar as doenças por meio de dados, é chamado Epidemiologia Digital (SALATHE et al., 2012). Ou seja, a epidemiologia que usa dados que não foram gerados para esse controle e reconhecimento dos padrões da doença para fins de fazer epidemiologia (SALATHÈ, 2018; PARK et al., 2018; ECKHOFF; TATEM, 2015; SALATHE et al., 2012).

Como os ambientes sociais, sejam os naturais ou os que foram construídos pelo homem, tem relação direta com a saúde e o bem-estar de uma população. Portanto, é importante considerar estes fatores ao criar políticas públicas para promoção de saúde. O estudo dos conceitos da sociedade e do espaço, descrevendo o papel do local e geografia na saúde, bem-estar e na doença é chamado de geografia da saúde (DUMMER, 2008). Essa área se concentra em entender as relações e padrões espaciais (DUMMER, 2008). Assim, a geografia da saúde pode ser usada para entender a interação do ambiente com a doença, e facilitando o desenvolvimento de políticas públicas eficazes para a promoção da saúde de uma população (DUMMER, 2008).

O uso de Sistema de Informações Geográficas (GIS) é um grande aliado a vigilância epidemiológica, contribuindo para o controle e a tomada de decisão. Essa tecnologia fornece dados de saúde (PORCASI et al., 2012; DOM et al., 2017). Com o GIS é possível incorporar dados físicos e sociais, com uma análise espacial. Assim é possível criar ferramentas para projetar e implementar sistema para a predição de grandes mudanças endêmicas e incidência de epidemias (DOM et al., 2017).

Ao conjunto de tecnologias de coleta, exposição e tratamento de dados produzindo informações demográficas para reconhecimento das condições de risco no território dá-se o nome de geoprocessamento (BARCELLOS; BASTOS, 1996; NARDI et al., 2013). São usadas variáveis para processamento de imagens e manipulação de dados para análise de saúde (BARCELLOS; BASTOS, 1996). Com softwares de geoprocessamento é possível analisar como se dar a distribuição de casos de doença de uma região e estudar como acontece o espalhamento da mesma. Possibilitando uma melhor avaliação e tomada de decisão dos gestores de saúde, facilitando o trabalho dos profissionais, a distribuição de recursos e ações educativas para controle, prevenção de doenças e manutenção do bem-estar da população (NARDI et al., 2013). Deste modo, o geoprocessamento é um conjunto de técnicas importantes para construção de mapas, auxílio no planejamento e monitoramento de políticas para promoção de saúde (NARDI et al., 2013).

3.2 Dengue, Febre chinkungunya e Zika

Define-se como arbovirose as doenças causadas por arbovírus, como por exemplo a Dengue, a febre *chinkungunya* e a Zika (FIGUEIREDO et al., 2017). Essas doenças são transmitidas por artrópodes, que podem ser insetos e aracnídeos. No caso das doenças deste estudo, o *Aedes aegypti* é o hospedeiro não humano, ele é importante para a transmissão das doenças (PAIVA et al., 2018; MUSSO; GUBLER, 2016). E essas doenças se mantém através do ciclo mosquito-humano-mosquito (MUSSO; GUBLER, 2016). Assim, há uma independência do vírus com relação ao hospedeiro não-humano (MUSSO; GUBLER, 2016).

O mosquito *Aedes aegypti* também é conhecido popularmente como mosquito-dengue, é um mosquito pequeno, tem o corpo negro com listas brancas (NATAL, 2002). É de origem africana e foi introduzido no Brasil após a colonização portuguesa (NATAL, 2002; MUSSO; GUBLER, 2016; VALOTTO et al., 2010). O macho dessa espécie se alimenta de frutas, já a fêmea precisa de sangue para amadurecer os ovos (VALOTTO et al., 2010). Fatores como luminosidade, temperatura, quantidade de chuva, podem influenciar no desenvolvimento das larvas. A combinação da globalização, mudanças climáticas e da adaptação do mosquito a ambientes urbanos facilitou a dissipação da doença, fazendo crescer o número de casos visto que esse mosquito se reproduz em água parada, seja ela

limpa ou poluída (VALOTTO et al., 2010). A transmissão biológica vetorial das arboviroses se dar por meio da picada da fêmea (VIRUSES, 1961). Toda vez que o mosquito infectado picar o hospedeiro humano, é injetado o vírus junto com a saliva do animal (MUSSO; GUBLER, 2016). Além disso, há a transmissão não vetorial que ocorre de humano para humano e pode ser de mãe para filho durante a gravidez ou no momento do parto, por transfusão de sangue, nosocomialmente, via medula óssea ou transplante de órgãos e por via sexual (MUSSO; GUBLER, 2016). As arboviroses que foram estudadas neste trabalho serão apresentadas nas subseções 3.2.1, 3.2.2 e 3.2.3.

3.2.1 Dengue

A dengue é uma doença causada pelo vírus DENV, do gênero *Flavivirus* e da família *Flaviviridae* (MARTÍNEZ-BELLO; LÓPEZ-QUÍLEZ; PRIETO, 2018; VALOTTO et al., 2010). Esse vírus tem quatro sorotipos com抗ígenos diferentes são eles DEN-1, DEN-2, DEN-3 e DEN-4 (MARTÍNEZ-BELLO; LÓPEZ-QUÍLEZ; PRIETO, 2018; FIGUEIREDO, 2006; SIRIYASATIEN et al., 2018). Ao se infectar com um dos sorotipos, o indivíduo ganha imunidade para aquele sorotipo, mas não para os outros. Após a fêmea do *Aedes aegypti* picar o indivíduo ele é infectado e passa por um período de incubação de 5 a 6 dias (MALAGON,). As manifestações clínicas mais comuns da dengue são dores musculares e articulares, febre alta, dor de cabeça, dor nos olhos, náuseas, vômito e perda de apetite, e aparecem após o tempo de incubação (SIRIYASATIEN et al., 2018; GOMES; THOMÉ, 2015). Contudo esses sintomas podem se agravar, levando a falha dos órgãos e vazamento de plaquetas, em ambos os casos pode ser fatal para o indivíduo infectado (SIRIYASATIEN et al., 2018; GOMES; THOMÉ, 2015).

A dengue é conhecida a mais de 200 anos (século XVIII), contudo só se apresentou como um problema de saúde pública após a Segunda Guerra Mundial, quando houve um aumento de 30 vezes na incidência da infecção (VALLE; PIMENTA; CUNHA, 2015; SIRIYASATIEN et al., 2018). A dengue se espalhou para mais de 128 países no mundo desde a década de 1970 (SIRIYASATIEN et al., 2018). No Brasil, o primeiro registro de dengue foi no século XIX. Em 1846, epidemias no Rio de Janeiro. Em 1852 e 1916, epidemia em São Paulo. Em 1923, epidemia em Niterói, Rio de Janeiro (VALLE; PIMENTA; CUNHA, 2015). Mas no período de 1923 a 1981 o Brasil viveu um silêncio epidemiológico, que ocorreu (possivelmente) em virtude da boa campanha de eliminação do *Aedes aegypti* (VALLE; PIMENTA; CUNHA, 2015). Mas a doença voltou a aparecer em 1981 com um surto na cidade de Boa Vista, Roraima. A dengue também é conhecida com outros nomes como polca, patuleia, urucubaca, febre eruptiforme, febre quebra-ossos (VALLE; PIMENTA; CUNHA, 2015).

Essa doença é um sério problema de saúde pública em todo o mundo, em particular em países de clima tropical e subtropical. Até o presente momento não há vacinas nem

tratamento específico para a dengue, daí a importância de um trabalho de predição de casos dessa doença.

3.2.2 Febre chikungunya

A febre chikungunya é uma doença causada pelo *Alphavirus CHIKV*. O CHIKV é um arbovírus da família do *Togaviridae* e do complexo antigênico *Semliki forest*. Esse completo antígeno é formado pelos alfavírus Ross River, O'nyong-nyong, Getah, Bebaru, *Semliki forest* e Mayaro que podem causar distúrbios articulares em humanos (DAS et al., 2010; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). São conhecidas quatro linhagens distintas do vírus chikungunya, são elas Oeste Africano, Leste-Centro-Sul Africano (ECSA), Asiático e Oceano Índico (IOL) (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a).

Este vírus é originário da África. Seu nome significa "curvar-se ou tornar-se contorcido" na língua Makonde (Kimakonde), um dos idiomas da Tanzânia (LUMSDEN, 1955; DAS et al., 2010). Em 1953, na Tanzânia que o CHIKV foi isolado pela primeira vez durante um surto de doença febril (ROSS, 1956; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a; DAS et al., 2010). O vírus da chikungunya também foi isolado em Benin, Costa do Marfim, Nigéria, República da Guiné e Senegal (DAS et al., 2010).

A doença febre chikungunya acontece após picada de um mosquito infectado. Após a infecção, há o período de incubação do vírus varia entre 2 a 6 dias (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). As manifestações clínicas mais comuns da infecção por CHIKV são febre alta e exantema (erupção aguda generalizada da pele) e artralgia (dores articulares) (DAS et al., 2010; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). Além desses sintomas também foram relatados mialgia, dor de cabeça, edema nas extremidades e problemas gastrointestinais (DAS et al., 2010). Foi observada a capacidade de reprodução do vírus em diferentes tecidos corporais nos pacientes com febre chikungunya. Assim, havia essa reprodução no sistema nervoso central, músculo cardíaco, articulações, fígado etc, causando uma grande variabilidade de sintomas (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a).

As manifestações clínicas agudas tendem a diminuir após duas semanas, contudo as dores nas articulações podem durar semanas ou meses (DAS et al., 2010). Em pessoas com mais de 65 anos os sintomas de artrite aguda são muito profundas, podendo evoluir para artrite crônica (DAS et al., 2010; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). Também foi relatada uma artrite febril que afeta as extremidades como tornozelos, punhos e falanges. Além disso, foram relatadas lesões de pele como descamação, úlceras aftosas e lesões de vasculite. E como o vírus se reproduz em outros tecidos, também tiveram relatos de hepatite, problemas cardíacos (miocardite, pericardite e miocardiopatia dilatada), manifestações oculares (iritociclite e retinite, as mais frequentes) e doenças neurológicas (mieloneuropatia, encefalites, Síndrome de Guillain-Barré, paralisia flácida e as neuropatias.)

(DAS et al., 2010; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a).

Alguns trabalhos estudaram a relação entre a infecção pelo CHIKV durante a gravidez, e foram mostradas evidências de grande risco de aborto no primeiro trimestre além da transmissão materno-fetal no último trimestre de gravidez (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). A transmissão vertical foi verificada e houve infecção severa em 10 dos 7504 avaliados. Sendo encefalopatia em 9 desses casos e febre hemorrágica em um (GÉRARDIN et al., 2008; AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a).

O vírus da febre chikungunya foi frequentemente isolado entre 1960 e 1990 em países da África Central e do Sul, e em países da África Ocidental. Entre 1960 e anos 2000 o vírus também foi isolado no Sudeste Asiático (DAS et al., 2010). Este vírus foi visto inicialmente na Tanzânia, África. Reapareceu no Quênia em 2004 e entre os anos de 2004 e 2007 houve casos em países da Europa, como Itália, Reino Unido, Bélgica, Alemanha, Noruega, Espanha e França. Seguido de relatos em Hong Kong, Canadá, Taiwan e Sri Lanka (DAS et al., 2010).

O fato do mosquito *Ae. aegypti* está muito adaptado ao ambiente urbano e com a globalização, facilita a dispersão do vírus em vários países do mundo. O CHIKV já era conhecido na África e na Ásia, mas se espalhou em ilhas do sudoeste do Oceano Índico (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). Em 2007, houve um número grande de casos em países ocidentais não-tropicais, como Itália. Em 2013, houve relatos de casos na região do Caribe, e em 2014, a doença se espalhou para áreas continentais do continente americano (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015b).

Em 2014, foram relatados os primeiros casos de febre chikungunya no Brasil. Em setembro de 2014, no estado do Amapá foi detectado o primeiro caso de transmissão local de febre chikungunya (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a). Um surto na cidade de Feira de Santana, na Bahia, foi relatado também em setembro de 2014. E foi detectado que o genótipo do vírus circulante no Amapá é o Asiático, já o da Bahia, é o do Leste-Centro-Sul Africano (AZEVEDO; OLIVEIRA; VASCONCELOS, 2015a; SCHUFFENECKER et al., 2006; HONÓRIO et al., 2015). Durante o ano de 2014 foram confirmados casos nos estados do amapá, Bahia, Distrito Federal, Mato Grosso do Sul, Roraima e Goiás. O território brasileiro é abundantemente infestado pelo *Aedes aegypti* que é o vetor do CHIKV. Após 2015, houve a circulação simultânea dos vírus DENV, ZIKV e CHIKV, dificultando o diagnóstico e a abordagem terapêutica para cada uma das doenças causadas por esses vírus. Além da maior proporção de casos sintomáticos de chikungunya comparados a dengue, o tempo do vírus no corpo ser maior (cerca de 8 dias), abundância de espécies primatas e grande extensão territorial do país fizeram com que aumentasse a probabilidade de grandes epidemias (HONÓRIO et al., 2015).

3.2.3 Zika

A Zika é uma doença causada pelo zika vírus (ZIKV). ZIKV é um arbovírus do gênero *Flavivirus* e da família *Flaviridade*, que são os mesmos gênero e família do DENV (JUNIOR et al., 2015). Essa arbovirose é originária do Oeste da África, e teria surgido por volta de 1920.

Na década de 1940, o vírus teria migrado para Ásia e originado a linhagem asiática (JUNIOR et al., 2015). Assim, esse vírus tem duas linhagens, uma africana e uma asiática. E essa última linhagem foi a responsável pela transmissão desse vírus dos casos ocorridos no Brasil (JUNIOR et al., 2015). Em 1948, o ZIKV foi isolado pela primeira vez na floresta de Zika, no Uganda (MUSSO; GUBLER, 2016; JUNIOR et al., 2015). Nos anos de 2013 e 2014 houve um grande surto na Polinésia Francesa, com relatos de complicações neurológicas graves (MUSSO; GUBLER, 2016). E após 2015, no Brasil também foram relatadas complicações do mesmo tipo, além de um aumento dramático em malformações congênitas graves, como microcefalia que foram associadas ao ZIKV (MUSSO; GUBLER, 2016).

Com a globalização este vírus está presente em grandes centros urbanos e, assim como o DENV, tem o ciclo de transmissão mosquito-humano-mosquito (MUSSO; GUBLER, 2016). Na maioria das vezes, a transmissão do ZIKV se dá pela picada do vetor, no caso o *Aedes aegypti* (MUSSO; GUBLER, 2016; JUNIOR et al., 2015). Contudo, o ZIKV pode ser transmitido via sexual, perinatal(materno-fetal) e por transfusão de sangue (JUNIOR et al., 2015; MUSSO; GUBLER, 2016). Após a incubação o vírus passa pelo tempo de incubação de 3 a 12 dias (MUSSO; GUBLER, 2016). Dependendo da localidade os sintomas desta doença podem mudar, contudo a manifestação clínica mais frequente é uma síndrome 'tipo-dengue' (JUNIOR et al., 2015). Os sintomas mais comuns são exanterma (manchas vermelhas), conjutivite, febre (habitualmente baixa, chegando a 39°C), cefaleia, mialgia, artralgia, e as vezes plaquetopenia (nível extremamente baixo de plaquetas), dor lombar discreta (JUNIOR et al., 2015; MUSSO; GUBLER, 2016). Esses sintomas desaparecem em uma semana, com duração média entre 3 e 5 dias (JUNIOR et al., 2015).

A zika pode ser descrita como uma infecção que normalmente não tem uma evolução grave e dura um curto período, contudo no Brasil foram registrados casos de Síndrome de Guillain-Barré (SGB) após o aparecimento dos sintomas da zika (JUNIOR et al., 2015). Estes casos foram registrados em sua maioria em estados do Nordeste brasileiro, 121 de 205 casos relatados entre janeiro e julho de 2015 (MUSSO; GUBLER, 2016). Em 11 de novembro de 2015, o Ministério da Saúde do Brasil declarou emergência nacional de saúde pública, (MUSSO; GUBLER, 2016). E com os crescentes casos de microcefalia, após a infecção por ZIKV durante a gravidez, se iniciou uma força-tarefa para apurar essa relação (SCHULER-FACCINI et al., 2016). Há uma associação da infecção por ZIKV e danos neurológicos graves em neonatos. A Organização Mundial de Saúde (OMS) declarou uma emergência de saúde global em fevereiro de 2016 (GULLAND, 2016; MUSSO; GUBLER,

2016). Além disso, outros problemas neurológicos, oftalmológicos e auditivos em recém-nascidos (SCHULER-FACCINI et al., 2016; MELO et al., 2016; MUSSO; GUBLER, 2016).

3.3 Doença do Coronavírus

A doença do Coronavírus é causada pelo vírus SARS-CoV-2. O SARS-CoV-2 é da família *Coronaviridae*, da ordem *Nidovirales*, do gênero *Betacoronavirus* e subgênero *Sarbecovirus* (KHALIL; KHALIL, 2020). A saber, o gênero desse vírus infecta apenas mamíferos. Esse vírus é originário da China (KHALIL; KHALIL, 2020). O nome desse vírus está diretamente ligado ao principal sintoma da infecção, a síndrome respiratória aguda grave (SARS), então SARS-CoV-2 é a sigla do inglês *Severe Acute Respiratory Syndrome CoronaVirus* (KHALIL; KHALIL, 2020).

A transmissão do vírus se dar pela inalação ou contato com gotículas de indivíduos infectados pelo vírus, sejam eles sintomáticos ou assintomáticos (SILVA; SANTOS; MELO, 2020). E o vírus é passado principalmente via espirros ou tosse, assim há dispersão das gotículas num espaço de 1 a 2 metros de distância, sendo inaladas ou contaminando superfícies (SINGHAL, 2020). Caso os indivíduos toquem essa superfície e entre em contato com olhos, boca e nariz dá-se transmissão desse vírus (SILVA; SANTOS; MELO, 2020). Na literatura há relatos diferentes para o tempo de incubação. Contudo a menor média de tempo relatada foi de 3 (embora a variação seja de 0 a 24 dias) e o maior 5,2 dias, e esse período pode levar de 2 a 14 dias (SILVA; SANTOS; MELO, 2020; GUAN et al., 2020; PUNG et al., 2020; SINGHAL, 2020). O tempo médio do início dos sintomas até a dificuldade de respirar é de 5 dias, até a hospitalização é de 7 dias e até a síndrome respiratória aguda é de 8 dias (SINGHAL, 2020). Apenas 25-30% dos casos de infecção pelo SARS-CoV-2 necessitam de internação na unidade de terapia intensiva (UTI). Dentre as manifestações graves relatadas estão lesão pulmonar aguda, síndrome respiratória aguda grave (SARS), choque e lesão renal aguda. Todavia a recuperação é de 10 dias em média (SINGHAL, 2020). Sendo de 4 a 11% a taxa de mortalidade em adultos hospitalizados e de 2 a 3% a taxa de letalidade geral da doença (SINGHAL, 2020).

Em dezembro de 2019, aconteceu um surto alarmante de uma doença infecciosa parecida com pneumonia na cidade de Wuhan, na China (SILVA et al., 2021b; LIMA et al., 2020). Então a Organização mundial de Saúde (OMS) foi informada que era um vírus novo e que, até então, não tinha sido identificado em humanos (PRADO et al., 2020). Esse vírus teve uma rápida disseminação, visto que ocorreu durante o festival da Primavera. Neste período é comum que os chineses viagem pela China e pelo resto do mundo (SILVA et al., 2021b). Em janeiro de 2020 o Covid-19 já era a segunda principal causa de resfriado comum (PRADO et al., 2020) e já tinha sido declarado uma Emergência de Saúde Pública de Interesse Internacional (PHEIC) (SILVA et al., 2021b). Na América Latina foi um dos

últimos continentes que o Covid-19 chegou. Em 25 de fevereiro de 2020, foi registrado o primeiro caso no Brasil (PRADO et al., 2020).

3.4 Predição de doenças infecciosas

As doenças infecciosas são modeladas a alguns anos por vários métodos, aqui serão comentados alguns desses métodos, como modelos estatísticos, modelos compartimentais e modelos baseados em aprendizado de máquina.

3.4.1 Modelo estatístico

A modelagem estatística de doenças tem como métodos mais comuns são os modelos autoregressivos e modelos lineares generalizados. Esse primeiro grupo modelam usando uma série temporal, que é uma sequência de dados obtidos em intervalos regulares de tempo em um período específico, para descrever a dinâmica do fenômeno em estudo (LATORRE; CARDOSO, 2001). Os modelos auto-regressivos (AR) e auto-regressivo incorporando média móvel (ARMA) fazem a análise de como a série temporal livre de tendências e de sazonalidade. Já quando a série temporal tem uma tendência são usados modelos modelos auto-regressivos integrados de médias móveis (ARIMA), e quando há sazonalidade se usa os modelos SARIMA (LATORRE; CARDOSO, 2001). Estes modelos podem capturar as tendências de mudanças, mudanças periódicas e distorções aleatórias dessas séries temporais. Já os modelos lineares generalizados, a variável resposta (y) é um processo de contagem e o comportamento da série ao longo do tempo é descrito pelas variáveis independentes (x_1, x_2, \dots, x_n) (LATORRE; CARDOSO, 2001). São modelos usados quando não é possível usar a distribuição normal. Esses modelos compõem a família exponencial de distribuições e contém várias funções como a regressão linear, regressão de Poisson, regressão logística etc.(LATORRE; CARDOSO, 2001).

3.4.2 Modelos compartimentais

Os modelos matemáticos contextualiza a doença através de parâmetros sociais por meio de sistema de equações para poder entender como funciona a propagação, evolução e poder desenvolver um controle da doença.

A modelagem matemática data do ano de 1760, foi iniciada por Daniel Bernoulli para modelar a varíola durante a segunda metade do século XVIII na Europa (BERNOULLI, 1760; MATEUS, 2020; CAMILO; MOURA; SALLES, 2020). Tal trabalho desenvolveu um modelo temporal para evolução da doença e influenciou políticas públicas de saúde pública (CAMILO; MOURA; SALLES, 2020). Em seguida, Hamer postula o princípio da Lei de Ação das Massas para um modelo determinístico, ou seja, o princípio de mistura homogênea, no qual a epidemia depende do número de indivíduos susceptíveis, número de infectados e

da taxa de contato entre eles (CAMILO; MOURA; SALLES, 2020). O estudo se concentrou na investigação da curva epidêmica do sarampo e é usado como base a trabalhos na área de epidemiologia matemática desde então (CAMILO; MOURA; SALLES, 2020). Em 1927, Kermack e McKendrick criaram um modelo para estudar a peste bulbônica baseados nas pesquisas de Ronald Ross e Hilda Hudson (KUCHARSKI, 2020; KERMACK; MCKENDRICK, 1927). E nesse modelo SIR fez uma predição temporal do número de casos da peste bulbônica na Índia entre os anos de 1905 e 1906 (KERMACK; MCKENDRICK, 1927). Hoje em dia os modelos compartimentais baseados em sistemas de equações diferenciais ordinárias (EDOs) usam como base o modelo SIR (Susceptible-Infected-Removed) criado por Kermarck e McKendrick e o adaptam para descrever outras doenças infecciosas (CAMILO; MOURA; SALLES, 2020; HETHCOTE, 2000).

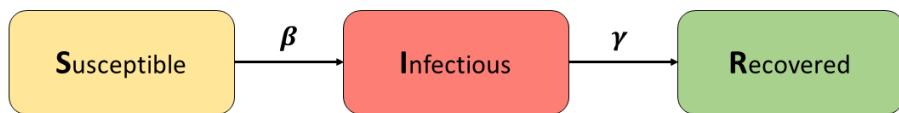


Figura 1 – Modelo compartmentado SIR. Fonte (LIMA et al., 2020)

O modelo SIR divide a população em três compartimentos: Susceptíveis, Infectados e Recuperados (HETHCOTE, 2000). No modelo original não foi considerada a dinâmica vital (KERMACK; MCKENDRICK, 1927). Baseados nesses foram criados outros modelos compartimentados, ou seja, modelos que dividem a população em classes. Os modelos matemáticos mais comuns são os baseados em sistemas de equações diferenciais ordinárias (EDOs) e em sistemas de equações diferenciais parciais (EDPs).

O modelo compartmental SIR é descrito pelo sistema de EDOs

$$\begin{cases} \frac{d}{dt}S = -\beta SI, \\ \frac{d}{dt}I = \beta SI - \gamma I, \\ \frac{d}{dt}R = \gamma I \end{cases}$$

onde $S(t)$ é a população de indivíduos suscetíveis, $I(t)$ são os indivíduos infectados sintomáticos, $R(t)$ são os indivíduos recuperados com imunidade, β é a taxa de contato entre indivíduos suscetíveis e infectados, e γ é a taxa de transferência de I para R . O diagrama na Figura 1 ilustra os compartimentos do modelo SIR. Esse tipo de modelo é resolvido usando algumas técnicas numéricas como mostram os trabalhos (SUBRAMANIAN et al., 2020), (STONE; OLINKY; HUPPERT, 2007), (MUSA et al., 2019), (SUPARIT; WIRATSUDAKUL; MODCHANG, 2018), (KUMAR et al., 2017) e (JING et al., 2018).

Assim, desde de Daniel Bernoulli, a representação da doença por meio de modelos matemáticos facilita o monitoramento e controle de doenças, podendo ajudar como a doença evolui e trazendo informações úteis que podem impactar no desenvolvimento de programas de promoção de saúde e qualidade de vida para população (NEPOMUCENO, 2005).

3.4.3 Modelos baseados em aprendizado de máquina

A aprendizagem de máquina utiliza técnicas computacionais para que a máquina seja capaz de identificar padrões de um conjunto de dados (WITTEN; FRANK, 2005). O aprendizado de máquina é uma aplicação da Inteligência artificial (IA) que fornece a capacidade de aprender a partir dos dados, ou seja, com respeito a alguma classe de tarefas e medida de desempenho (HAYKIN, 2001). O aprendizado de máquina pode ser de dois tipos: supervisionado e não supervisionado (SIRIYASATIEN et al., 2018). No aprendizado supervisionado o professor ou especialista pré determina classes ou rótulos. Já no aprendizado não-supervisionado, não existe uma pré-classificação, a caracterização se dá por analogia, agrupa os dados automaticamente. O objetivo é entender a estrutura do conjunto de dados (HAYKIN, 2001; SIRIYASATIEN et al., 2018).

Na próxima seção serão apresentados os regressores de aprendizado de máquina foram usados nesta pesquisa.

3.5 Regressores

Para os modelos de predição espaço-temporal de Covid-19 e de criadouros de arboviroses foram usados alguns regressores que serão descritos nesta seção.

3.5.1 Regressão linear

Para prever valores numéricos o método mais simples é a regressão linear. A regressão linear é um dos conceitos de análise de dados mais utilizados em aprendizado de máquina. Este método é originário da correlação linear, portanto assume que os dados tem um comportamento linear, e que a variável a ser predita pode ser representada por meio de uma combinação linear dos atributos com seus pesos pré-determinados (WITTEN; FRANK, 2002). Assim, essa técnica utiliza os pontos de dados para encontrar a melhor linha de ajuste (por meio de um diagrama de dispersão) para modelar a relação. E essa relação pode ser representada pela equação 3.1

$$y = C_0 + C_1x_1 + C_2x_2 + \dots + C_nx_n \quad (3.1)$$

onde y é a variável de predição; x_1, x_2, \dots, x_n são os valores dos atributos e $C_0, C_1, C_2, \dots, C_n$ são os pesos de cada atributo. Por meio dessa técnica é possível otimizar os pesos e encontrar o melhor ajuste possível.

3.5.2 Máquina de Vetor de Suporte

A técnica de máquina de vetor de suporte é bastante usada para problemas de classificação de padrões e estimativa de regressão, como categorização de textos, análise de imagens e bioinformática. Proposta por Vladimir Vapnik em 1979, mas ficou famosa após os seus trabalhos (BOSER; GUYON; VAPNIK, 1992) e (VAPNIK, 1998), nos anos de 1992 e 1998, respectivamente. É um método de aprendizado de máquina supervisionado, originalmente focado em problemas linearmente separáveis, e que mostra resultados muitas vezes melhores que outros algoritmos como redes neurais, por exemplo.

A ideia principal de uma máquina de vetor de suporte é construir um hiperplano como superfície de decisão de tal forma que a margem de separação entre exemplos positivos e negativos seja máxima. (HAYKIN, 2001).

Portanto, pode-se dizer que a máquina de vetor de suporte implementa o método de minimização estrutural. Dentre os tipos de máquina de aprendizagem que pode ser construída, destacam-se as máquinas de aprendizagem polinomial (kernel polinomial) e as redes de função de base radial (kernel RBF).

Esse tipo de algoritmo é mais genérico e por essa característica tem uma ampla aplicabilidade. Além disso, possui boa capacidade de generalização, robustez em grandes dimensões, a função objetivo é conexa e tem teoria fundamentada na matemática e estatística. Porém é sensível a escolha dos parâmetros. Os SVMs são utilizados com sucesso em diversas aplicações biomédicas (BARBOSA et al., 2021; LIMA; SILVA-FILHO; SANTOS, 2016; SANTOS; SOUZA; FILHO, 2007; COMMOWICK et al., 2018; AZEVEDO et al., 2015; CRUZ; CRUZ; SANTOS, 2018; CORDEIRO et al., 2012; VASCONCELOS; SANTOS; LIMA, 2018; RODRIGUES et al., 2019; FREITAS et al., 2019; JUNIOR et al., 2019).

3.5.3 Rede Neural Perceptron Multicamadas

As Redes Neural Perceptron Multicamadas é uma técnica que tem como objetivo imitar o comportamento de tomada de decisão do cérebro humano. Essas redes são formadas por elementos chamados de neurônios, dos quais estão interconectados e que formam uma rede. Essa rede pode ter uma ou várias camadas. Este tipo de técnica forma um sistema dotado de arquiteturas paralelizáveis (multiprocessamento de dados). A rede neural de camada única é a mais simples, e tem esse nome por ter apenas uma camada de nós computacionais (neurônios) (HAYKIN, 2007). Foi criada por Frank Rosenblatt em 1958 (ROSENBLATT, 1958), após usar o neurônio MCP, criado por McCulloch e Pitts em 1943 (MCCULLOCH; PITTS, 1943), e implementado no computador IBM740 no Laboratório da Aeronáutica de Cornell em 1957 (HAYKIN, 2007). As redes neurais artificiais são utilizadas com sucesso em diversas aplicações (BARBOSA et al., 2021; LIMA; SILVA-FILHO; SANTOS,

2016; SANTOS; SOUZA; FILHO, 2007; COMMOWICK et al., 2018; AZEVEDO et al., 2015; CRUZ; CRUZ; SANTOS, 2018; CORDEIRO et al., 2012; VASCONCELOS; SANTOS; LIMA, 2018; RODRIGUES et al., 2019; FREITAS et al., 2019; JUNIOR et al., 2019; SANTOS; SOUZA; FILHO, 2007; SANTOS et al., 2008; SANTOS et al., 2009b; SANTOS et al., 2009a; SANTOS; FILHO; SANTOS, 2019).

3.5.4 Random Forest

Árvore de decisão é uma técnica de aprendizado de máquina não supervisionado e não-paramétrico com ampla aplicação em classificação e regressão, como diagnóstico médicos por imagens e análise de risco de crédito. Pode-se dizer que este método é a combinação de simplicidade das árvores de decisão com a flexibilidade e aleatoriedade proporcionando um modelo mais preciso. Nesse método as instâncias são descritas por partes atributo-valor, a função objetivo é discreta e pode ser necessária hipóteses disjuntas.

As árvores de decisão são estruturas de dados formadas por um conjunto de elementos que armazenam informações chamadas nós. Toda árvore possui um nó chamado de raiz, que é o ponto de partida e liga os outros denominados filhos. Esses filhos podem possuir seus próprios filhos que por sua vez possuem os seus. Os nós folha são os nós que não possuem filhos. Ou seja, a árvore de decisão é uma árvore que armazena regras em seus nós, e a decisão a ser tomada é dada pelos nós folhas(DUDA; HART; STORK, 2000).

3.6 Considerações

Nesta seção, foram apresentados alguns conceitos básicos de Epidemiologia Digital, Geografia da Saúde e arboviroses. Na seção de arboviroses foram detalhadas informações das doenças dengue, zika e febre chikungunya. Além disso foram apresentados também modelos de predição com abordagem estatística, matemática e de aprendizado de máquina, e detalhados os regressores usados nos experimentos desta dissertação.

4 Proposta

4.1 Introdução

Como já foi comentado no capítulo 1, as arboviroses e Covid-19 precisam de uma atenção especial. Com base nisso, neste trabalho, foram propostos sistemas de controle e predição para essas doenças. Neste capítulo serão apresentadas as propostas desses sistemas em detalhes.

4.2 Modelos propostos

Para arboviroses foi proposto um protótipo de um sistema preditor da disseminação das arboviroses transmitidas pelo Aedes aegypti, a saber, a dengue, a febre chikungunya e a zika, com abordagem espaço-temporal e utilizando modelagem baseada em algoritmos de aprendizado de máquina.

Foram feitos dois modelos para predição de covid-19, um com abordagem temporal e outro com abordagem espaço-temporal. Para a abordagem temporal, foi proposto um sistema de predição de casos cumulados de Covid-19 em tempo real para o Brasil e seus 27 estados. E para a abordagem espaço-temporal, foi proposto um sistema de predição de casos acumulados de Covid-19 em tempo real para cada município do Brasil.

4.2.1 Modelo espaço-temporal - arboviroses

O modelo proposto para controle da distribuição e disseminação do Aedes aegypti foi baseado em aprendizado de máquina com uma análise espaço temporal por meio de mapas, esses mapas indicam onde estão as maiores concentrações de criadouros do vetor.

O primeiro passo da pesquisa foi organizar a planilha de levantamento rápido de índices de infestação de criadouros da prefeitura, obtido pelo LIRAA. Nesses dados estavam as informações dos tipos de recipientes para Aedes aegypti, as mesmas estavam divididas por distritos sanitários (I ao VI) de cada bimestre dos anos de 2014, 2015 e 2016. Montamos uma tabela com o nome do bairro, sua respectiva latitude e longitude e suas informações com a quantidade de cada tipo de criadouro do vetor.

Para criar o sistema de predição da distribuição espaço-temporal dos criadouros de Aedes aegypti no Recife, foram usados mapas de criadouros e a distribuição das variáveis climáticas (temperatura, chuva e velocidade dos ventos) nos anos de 2014 a 2016.

No software QGIS foram gerados arquivos shapefile (.shp) com as informações de

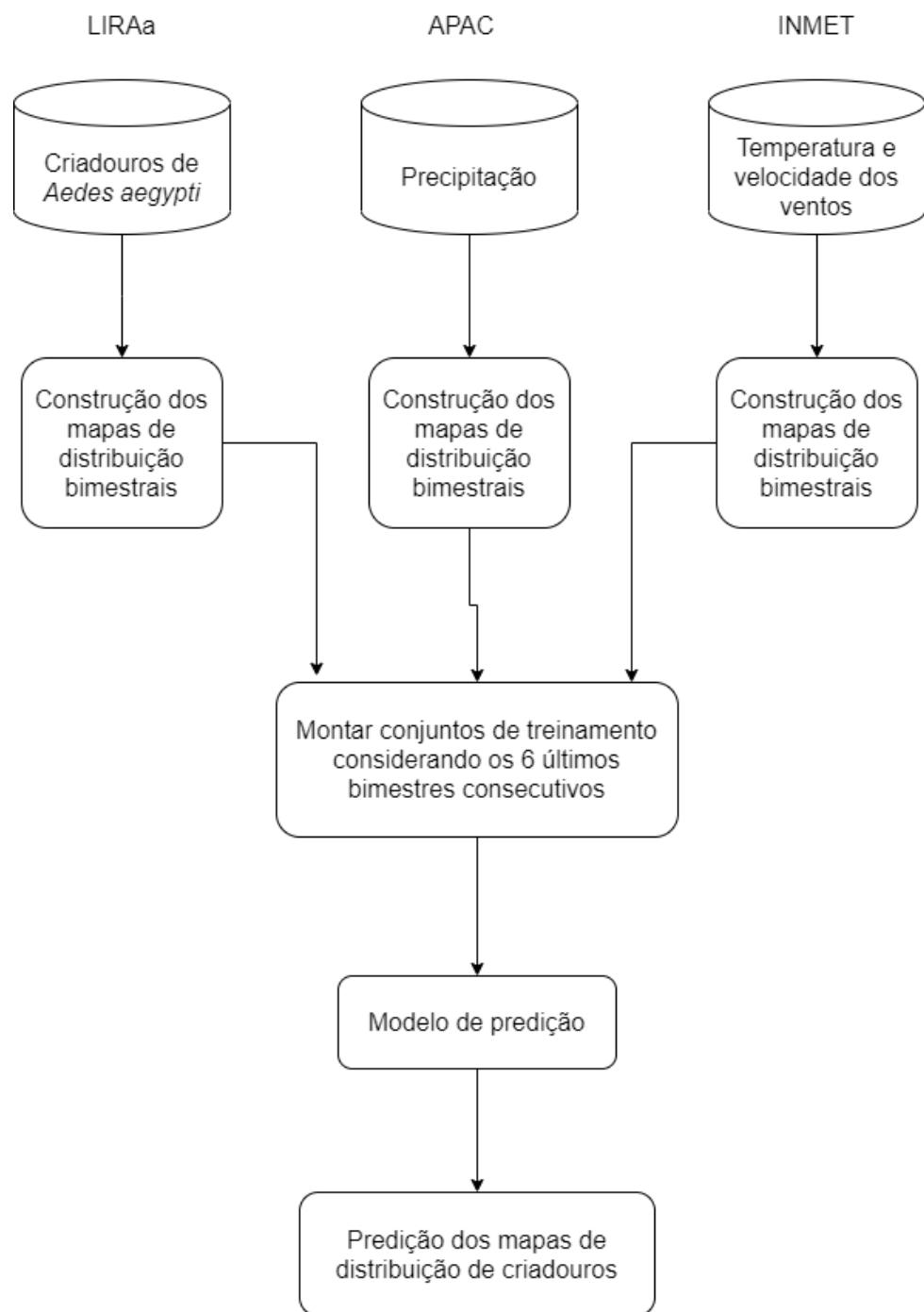


Figura 2 – Metodologia usada para gerar os mapas de predição da distribuição de criadouros na Cidade do Recife.

criadouros, temperatura, chuva e velocidade dos ventos no período escolhido, e com esse arquivo traz a localização geográfica de criadouros e variáveis climáticas para cada bairro da cidade do Recife. Para o passo seguinte é feita uma interpolação no QGIS usando o método de ponderação da distância inversa (IDW) para gerar os mapas de distribuição espacial. Para os criadouros esses mapas são bimestrais e para as variáveis climáticas os mapas são mensais. Como passo seguinte é feito o agrupamento dos mapas resultantes da etapa anterior com um script Python, e assim são armazenadas informações na devida ordem: latitude, longitude, distribuição dos criadouros e - para cada bimestre - a distribuição de temperatura, chuvas e velocidade dos ventos. Dessarte, também com um script Python, foram montados os conjuntos de predição com mapas de seis bimestres consecutivos de modo a prever o mapa de distribuição do bimestre seguinte. Feito isso, é usado o software Weka para avaliar os regressores e fazer a predição da distribuição dos criadouros. E no QGIS é feita a visualização dos mapas de predição da distribuição de cruzamento.

A ideia inicial era codificar cada caso com relação aos endereços da plataforma de dados abertos, mas alguns eram digitados com erros e não tinham padronização, então não foi possível trabalhar com esses dados. Então que foi feito o somatório de casos por bairros. Os pontos de coleta de dados climáticos são estimados usando uma distribuição gaussiana, pois em Recife existem três bairros nos quais são observadas estas informações de chuva da APAC. E pelo INMET são fornecidos dados de dois pontos de coleta, e também é usada uma distribuição gaussiana ao se construir os shapefile. Por essas informações não serem de cada bairro do Recife, existe um erro associado, o qual no momento não é possível mensurá-lo. Na base de criadouros estima-se que em cada bairro tenha a mesma quantidade de casos, visto que a base inicial era dividida os estratos, e se considera que cada bairro do estrato tenha a mesma quantidade de recipientes. Esse tipo de processo pode criar alterações, como por exemplo o bairro de Monteiro, que é um bairro de classe média mas foge a regra de que bairros de classe média tem menos casos, o que talvez seja um erro causado pelo fato de ter fixado um número de recipientes.

4.2.2 Modelo temporal - Covid-19

O sistema funciona da seguinte forma: cada estado alimenta um banco de dados com as notificações de casos diários e acumulados de Covid-19. O portal Brasil.io reúne essas informações em um banco de dados. A seguir, o software que foi desenvolvido, COVID-SGIS, adere diariamente os dados que estão no banco de dados do portal Brasil.io por meio do rastreador web SGIS. Após esses dados serem acumulados são reunidos em um arquivo de valores e separados por vírgulas em um arquivo (.csv). Por conseguinte, são derivados os conjuntos de dados de treinamento para os quais os modelos são criados para o número acumulado de mortes confirmadas e relatadas de Covid-19.

Os modelos criados são modelos clássicos amplamente usados para prever com-

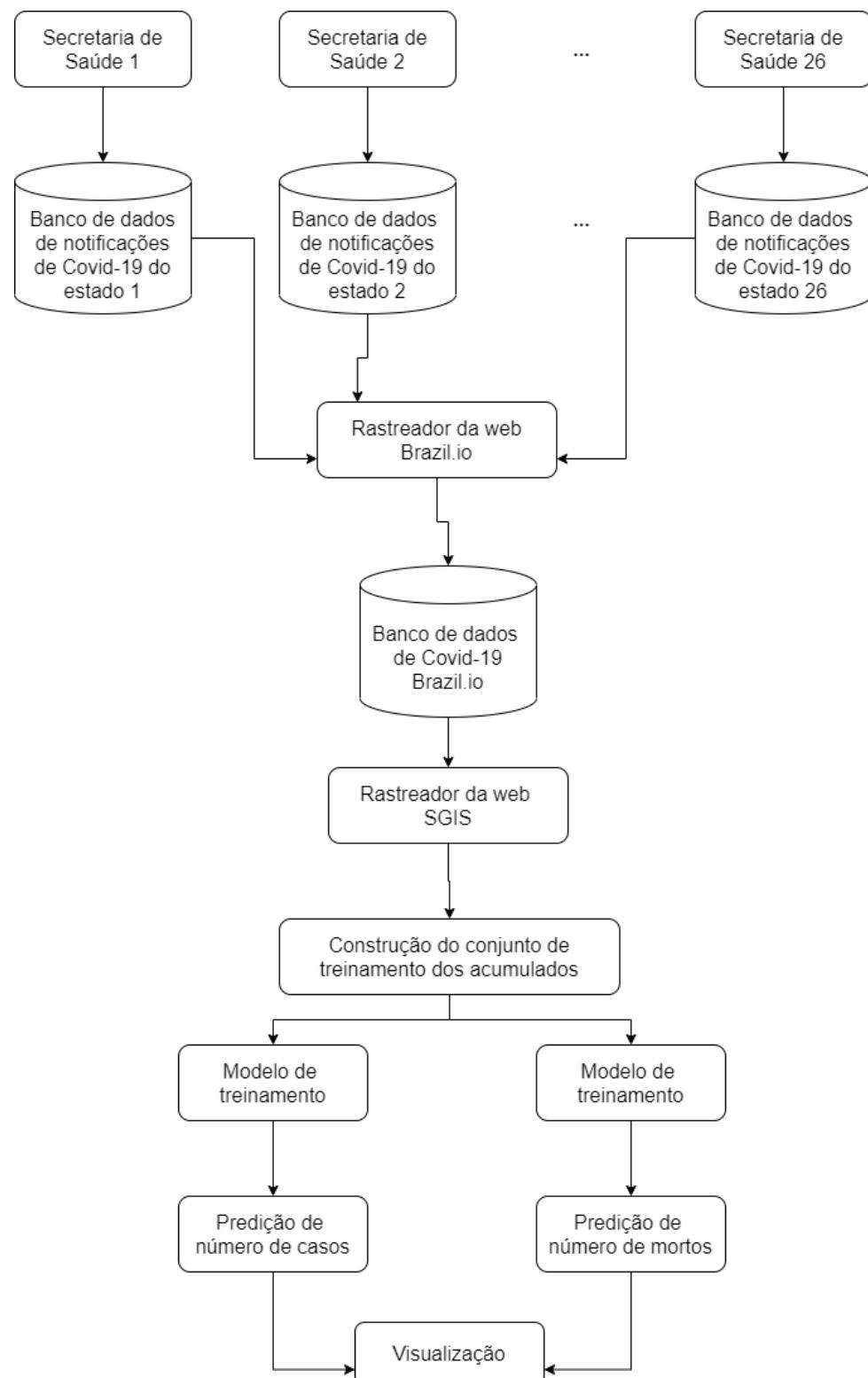


Figura 3 – Metologia do sistema de predição COVID-SGIS com abordagem temporal

portamentos futuros de séries temporais estacionárias, chamado ARIMA. Esse modelo é uma generalização do modelo autorregressivo de médias móveis ARMA. O ARIMA é uma combinação dos modelos autoregressivo (AR) e de média móvel (MA) com série original indiferenciada do modelo, que foi diferenciada d vezes até se tornar uma série estacionária antes de ajustar o modelo ARMA. Assim, p e q correspondem às ordens dos modelos AR e MA, respectivamente, no modelo ARIMA(p,d,q), e d corresponde ao nível de diferenciação. A expressão matemática do modelo é dada pela equação 4.1, e nela a série diferenciada é representada por y'_t , os parâmetros do modelo são c e ϕ , e ϵ é o erro aleatório no tempo t (CHAKRABORTY; GOSH, 2020).

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (4.1)$$

Como metodologia foi feito o ajuste do modelo ARIMA(p,d,q) a um conjunto de dados. E um passo a passo guia a construção dos modelos propostos.

1. Avaliação da estacionariedade da série, ou seja, se a série é estacionária ou não. E caso não seja, a diferenciação é aplicada até que a série se torne estacionária. Os resultados do teste Dickey-Fuller estão na Tabela 1.
2. Estimativa dos parâmetros p e q baseado nos gráficos de função de autocorrelação (ACF) e da função de autocorrelação parcial (PACF).
3. Avaliação do modelo de predição bem mais ajustado usando Akaike Information Criterion (AIC) e o Bayesian Information Criterion (BIC).

O Akaike Information Criterion (AIC), é um estimador de erro da previsão fora da amostra. Formalmente, temos que dada uma coleção de modelos, o critério AIC irá estimar a qualidade de um modelo da coleção ao compará-lo com outro da mesma coleção, estimando o erro de previsão fora da amostra. Assim, é criada uma pontuação que determina qual dos modelos da coleção tem a maior probabilidade de ser o melhor modelo para a coleção em questão, e a melhor pontuação é sempre a mais baixa. O AIC é usado quando não se tem acesso aos dados fora da coleção e se quer analisar a qualidade dos modelos dentre eles, além do uso por conveniência de tempo, isto é, o AIC mede apenas a qualidade relativa dos modelos a uma certa variável que está em análise, o bom disso é ter uma forma de mostrar que os modelos estão em um padrão absoluto aceitável. Este critério é aplicado para séries temporais e regressões. A expressão para este estimador é dada pela equação

$$AIC = 2 \ln(L) + 2k \quad (4.2)$$

onde L é a probabilidade de verossimilhança e k é o número de parâmetros.

O Bayesian Information Criterion (BIC) é um estimador baseado em máxima verossimilhança, que mede a eficiência do modelo parametrizado em termos da precisão dos dados. E funciona para comparação de modelos dentro de um conjunto finito. Esse critério introduz um termo de penalidade para o número de parâmetros no modelo. Ele está intimamente relacionado ao AIC. Com ele é possível aumentar a probabilidade dos modelos ao adicionar parâmetros ao ajustá-los modelos, o que pode causar sobreajuste (overfitting). Comumente usado para identificação em série temporal e regressão, mas pode ser usado para qualquer conjunto de modelos baseado em verossimilhança. É usado para comparar modelos estimados somente se os valores números da variável independente são idênticos para todos os modelos que estão sendo comparados, e esses modelos não precisam necessariamente estarem alinhados. Os testes dos modelos são feitos usando os valores de BIC correspondente, e o valor mais baixo indica termos de penalidade mais baixos, logo o melhor modelo. Porém, esse estimador não lida com coleções complexas, como por exemplo seleção de variável (ou de recursos) em alta dimensão. A expressão para este estimador é dada pela equação

$$BIC = \ln(n)k + 2 \ln(L_M) \quad (4.3)$$

Algumas considerações sobre esses critérios: o BIC considera o número de observações na fórmula, o prazo de penalidade no BIC é maior que no AIC; e o BIC introduz um termo de penalidade para o número de parâmetros do modelo. E o valor do BIC é sempre maior que o do AIC, e quanto menor esses valores, melhor será o modelo.

Foram usadas duas funções do pacote de predição no software R para estes modelos: para calcular automaticamente os parâmetros p , d e q e retornar o modelo ajustado foi usada a função *auto.arima()* e em seguida a função *forecast()* para fazer as previsões com base no modelo ajustado pela função *auto.arima()*. Assim, com os dados coletados do número de casos confirmados acumulados de cada estado brasileiro, foi criado um banco de dados que corresponde à série histórica dos casos acumulados da doença Covid-19 para o Brasil e cada um dos estados. O período para essa série histórica foi da data da primeira notificação de casos em cada estado e no Brasil até 5 de maio de 2020 (LIMA et al., 2020). Então foram construídos três tipos de modelos ARIMA: (1) um que foi estratificado por cada estado.

4.2.3 Modelo espaço-temporal - Covid-19

Para a abordagem espaço-temporal, foi proposto um sistema que prever a distribuição da Covid-19 no Brasil e no estado de Pernambuco. O sistema proposto funciona para cada município no Brasil e em Pernambuco. Essas informações são obtidas através do Portal Brasil.io, que obtém estes dados através das secretarias de saúde de cada cidade e alimenta o banco de dados desse portal. E com isso é calculado o número de casos acu-

estado	ordem de diferencial	t-estatística	τ
Acre	$d = 0$	4.4121	-1.95
	$d = 1$	1.874	-1.95
	$d = 2$	-3.983	-1.95
Alagoas	$d = 0$	1.6772	-1.95
	$d = 1$	-0.7374	-1.95
	$d = 2$	-7.8526	-1.95
Amazonas	$d = 0$	5.8866	-1.95
	$d = 1$	1.6766	-1.95
	$d = 2$	-5.5631	-1.95
Amapá	$d = 0$	4.674	-1.95
	$d = 1$	-0.2412	-1.95
	$d = 2$	-6.2019	-1.95
Bahia	$d = 0$	7.9516	-1.95
	$d = 1$	1.4916	-1.95
	$d = 2$	-8.651	-1.95
Ceará	$d = 0$	6.1006	-1.95
	$d = 1$	0.7138	-1.95
	$d = 2$	-2.7642	-1.95
Distrito Federal	$d = 0$	5.481	-1.95
	$d = 1$	-1.0392	-1.95
	$d = 2$	-8.9064	-1.95
Espírito Santo	$d = 0$	4.7973	-1.95
	$d = 1$	-1.1292	-1.95
	$d = 2$	-9.1055	-1.95
Goiás	$d = 0$	5.0383	-1.95
	$d = 1$	-0.9949	-1.95
	$d = 2$	-7.6139	-1.95
Maranhão	$d = 0$	4.1517	-1.95
	$d = 1$	0.674	-1.95
	$d = 2$	-5.2222	-1.95
Minas Gerais	$d = 0$	6.9037	-1.95
	$d = 1$	0.62	-1.95
	$d = 2$	-9.5418	-1.95
Mato Grosso do Sul	$d = 0$	3.988	-1.95
	$d = 1$	-1.9052	-1.95
	$d = 2$	-7.5919	-1.95
Mato Grosso	$d = 0$	4.9929	-1.95
	$d = 1$	-1.1216	-1.95
	$d = 2$	-6.0585	-1.95
Pará	$d = 0$	5.8845	-1.95
	$d = 1$	0.5428	-1.95
	$d = 2$	-5.5288	-1.95
Paraíba	$d = 0$	7.6861	-1.95
	$d = 1$	2.0185	-1.95
	$d = 2$	-7.0357	-1.95

Tabela 1 – Resultados do teste de Dickey-Fuller para séries históricas do número de casos acumulados de covid-19 no Brasil e seus 27 estados. Quando t-estatística for melhor que τ , isto indica que a série possui uma única raiz (que não é estacionária). E quando a t-estatística for inferior a τ a série é estacionária. Fonte: (LIMA et al., 2020)

mulados e as coordendas de cada município. São enviados os bancos de dados da etapa anterior para o módulo de interpolação para que os conjuntos de treino sejam montados. No módulo de interpolação o número de casos é distribuido, usando latitude e a longitude do centro do município, em uma grade não homogênea. Posteriormente é gerada uma grade regular calculada pelo método de ponderação de distância inversa (IDW), e estimada a distribuição de casos confirmados e óbitos. Logo após, os mapas de predição são gerados usando mapas de 3 dias consecutivos para prever o mapa de distribuição do dia seguinte. Enfim, é apresentado o mapa de predição dos casos acumulados de Covid-19 para o Brasil e para Pernambuco, e com o modelo proposto são feitas previsões espaciais de 3 dias.

estado	ordem de diferencial	t-estatística	τ
Pernambuco	$d = 0$	3.5281	-1.95
	$d = 1$	-0.2507	-1.95
	$d = 2$	-8.4942	-1.95
Piauí	$d = 0$	5.693	-1.95
	$d = 1$	1.7041	-1.95
	$d = 2$	-5.3715	-1.95
Paraná	$d = 0$	2.5805	-1.95
	$d = 1$	-1.4351	-1.95
	$d = 2$	-7.2115	-1.95
Rio de Janeiro	$d = 0$	8.1792	-1.95
	$d = 1$	0.3303	-1.95
	$d = 2$	-9.0729	-1.95
Rio Grande do Norte	$d = 0$	4.6418	-1.95
	$d = 1$	-1.7051	-1.95
	$d = 2$	-6.9444	-1.95
Rondônia	$d = 0$	4.8637	-1.95
	$d = 1$	0.3534	-1.95
	$d = 2$	-5.8114	-1.95
Roraima	$d = 0$	3.0461	-1.95
	$d = 1$	0.1819	-1.95
	$d = 2$	-2.3616	-1.95
Rio Grande do Sul	$d = 0$	4.1479	-1.95
	$d = 1$	0.3525	-1.95
	$d = 2$	-3.7048	-1.95
Santa Catarina	$d = 0$	3.7958	-1.95
	$d = 1$	-1.6657	-1.95
	$d = 2$	-8.1044	-1.95
Sergipe	$d = 0$	8.3063	-1.95
	$d = 1$	-0.2861	-1.95
	$d = 2$	-12.0538	-1.95
São Paulo	$d = 0$	4.3266	-1.95
	$d = 1$	-1.4522	-1.95
	$d = 2$	-7.4511	-1.95
Tocantins	$d = 0$	5.9875	-1.95
	$d = 1$	1.1544	-1.95
	$d = 2$	-7.4744	-1.95
Brazil	$d = 0$	5.1648	-1.95
	$d = 1$	1.4947	-1.95
	$d = 2$	-5.7441	-1.95

Tabela 2 – Resultados do teste de Dickey-Fuller para séries históricas do número de casos acumulados de covid-19 no Brasil e seus 27 estados. Quando t-estatística for melhor que τ , isto indica que a série possui uma única raiz (que não é estacionária). E quando a t-estatística for inferior a τ a série é estacionária. Fonte: (LIMA et al., 2020)

4.3 Base de dados de arboviroses

Foram utilizadas as bases de dados de localização de focos de Aedes aegypti e da distribuição das arboviroses e de sintomas a partir da Plataforma de Dados Abertos da Prefeitura do Recife ¹, e as informações de distribuição de temperatura e umidade ao longo dos anos e meses a partir da base de dados da Agência Pernambucana de Águas e Clima ², especialmente do seu Sistema de Informação Geográfica SIGHPE ³, onde as sequências temporais de informações pluviométricas mensais estão armazenadas a partir do ano de 2006.

4.4 Base de dados de Covid-19 - abordagem temporal

Para o sistema COVID-SGIS das abordagens temporal e espaço-temporal, foram usados como atributos do modelo os dados de casos confirmados e óbitos do banco de

¹ <http://dados.recife.pe.gov.br/>

² <http://www.apac.pe.gov.br/>

³ <http://www.apac.pe.gov.br/sighpe>

dados do portal Brasil.io⁴.

Este portal disponibiliza os registros de casos confirmados e óbitos da Covid-19, estes dados são obtidos através dos boletins das Secretarias de Saúde de cada estado brasileiro. São registrados os casos desde a primeira data de confirmação de casos até a última atualização, feita em 5 de maio de 2020, para cada estado.

4.5 Base de dados de Covid-19 - abordagem espaço-temporal

Para o sistema com abordagem espaço-temporal, também foi usado o banco de dados do portal Brasil.io⁵ para o número de casos confirmados e óbitos dos municípios do Brasil e de Pernambuco que foram os atributos usados no modelo.

No Brasil.io são disponibilizados boletins das secretarias de Saúde de cada uma das unidades federativas do Brasil, e em especial de cada um dos municípios. E para cada município, os dados obtidos vão do primeiro dia de notificação até o dia 6 de junho de 2020. E por fim, foram calculados os casos acumulados para os municípios do Brasil e de Pernambuco para cada dia de notificação.

⁴ <https://brasil.io/home/>

⁵ <https://brasil.io/home/>

5 Resultados e discussões

Neste capítulo serão apresentados os resultados parciais obtidos pelos algoritmos usados para modelagem da predição temporal e espaço-temporal de casos de Covid-19 para o Brasil e Pernambuco. Sendo a abordagem temporal a nível estadual e Brasil, e a abordagem espaço-temporal a nível municipal. Ainda serão detalhados e discutidos os gráficos de predição e os mapas de distribuição de casos obtidos pelos modelos propostos.

5.1 Predição espaço-temporal de arboviroses

Foram coletados dados de criadouros para cada bairro da cidade do Recife através do Sistema LIRAA, sendo usado o número total de criadouros do mosquito, armazenados em um arquivo .csv, com a latitude e longitude de cada bairro. Os dados foram separados por bimestre para os anos de 2013 a 2016.

E já que os fatores climáticos interferem no comportamento do mosquito, também foram coletados dados de variáveis climáticas por meio da base de dados do Instituto Nacional de Meteorologia¹, INMET. Foram coletadas séries históricas mensais no período de 2013 a 2016. E dados pluviométricos foram obtidos através do Sistema de Informação Geográfica², SIGHPE, da Agência Pernambucana de Águas e Climas³, APAC, com séries históricas mensais no período de 2013 a 2016. Essas informações climáticas que foram coletadas (com latitude e longitude) para cada bairro foram armazenadas em arquivos .csv.

Com o software QGIS e utilizando os arquivos .csv, foram gerados arquivos vetoriais de pontos (.shp) bimestrais para a localização geográfica dos criadouros para cada bairro para os anos de 2013 a 2016. O mesmo foi feito para os fatores climáticos, porém em intervalos mensais do mesmo período de tempo. E gerados mapas de distribuição espacial de temperatura, chuva e velocidade dos ventos para estimar a distribuição dos criadouros na cidade do Recife. E para tal foi usado o software QGIS, com o processo Interpolação pelo Inverso da Distância (IDW) com potência igual a 2.

A partir dos mapas de distribuição de criadouros e das variáveis climáticas para cada bimestre, foram criados conjuntos de predição. Como o Sistema Único de Saúde (SUS) usa o ciclo bimestral para o planejamento de políticas de combate às arboviroses, o modelo escolhido foi bimestral. Os vetores de predição foram criados para realizar uma varredura simultânea nos mapas de distribuição pixel a pixel, e relacionando a latitude e longitude com a distribuição bimestral de criadouro e a distribuição de temperatura, chuva e

¹ <https://portal.inmet.gov.br/>

² <http://www.sirh.srh.pe.gov.br/apac/sighpe/>

³ <https://www.apac.pe.gov.br/>

velocidade dos ventos para os meses do respectivo bimestre. E assim, foram construídos 18 conjuntos de predição com 15.446 instâncias cada, ou seja, um pacote de treinamento para cada bimestre entre os anos de 2014 e 2016.

Usando o software Weka, com a ferramenta “resample”, foi montado o conjunto de treinamento a partir do conjunto com 15.446 instâncias. Com valores aleatórios para as instâncias, a ferramenta *resample* permite criar uma nova base de dados mantendo as mesmas características da base original. O número de instâncias deve ser especificado pelo usuário, visto que o conjunto de treinamento foi gerado pela ferramenta do Weka para cada conjunto de predição e com o número de instâncias equivalente a 30% do conjunto original. E estes conjuntos de 15.446 instâncias foram utilizados para validação dos modelos gerados pelos regressores.

O objetivo inicial era geocodificar os endereços (parciais) dos pacientes disponíveis no portal de dados de Recife, ou seja, transformar os endereços em latitude e longitude (para poder usar uma grade no QGIS). Contudo não foi possível fazer, então foi decidido trabalhar com os bairros dos endereços. E foi definido com o centroide do bairro. E a partir do centroide do bairro é usado o método de interpolação (IDW) para espalhar artificialmente. Mas ao espalhar a base cresce bastante com dados artificiais, então foi decidido usar 30% do conjunto original para minimizar o efeito do crescimento da base.

Como métricas para avaliar o modelo criado foram usados o coeficiente de correlação de Pearson e o Root Relative Squared Error (RRSE percentual). Esse primeiro foi usado para avaliação global do modelo de predição e avalia a relação entre duas variáveis contínuas, os valores reais e os valores preditos pelo modelo. E varia entre -1 e 1, próximo de 1 apresenta uma alta correlação, próximo de -1 uma correlação negativa entre as variáveis, próximo de 0 as variáveis têm uma fraca ou nenhuma correlação (WITTEN; FRANK, 2005). Já o Relative Quadratic Error, foi usado como métrica de avaliação local, para evitar uma avaliação superficial do modelo. e assim conseguimos mensurar o erro dos modelos. Com a combinação de uma métrica global e local é possível obter uma melhor avaliação dos modelos de predição que foram criados.

Além dos modelos serem avaliados com o coeficiente de correlação de Pearson e o RRSE(%), também foi computado o tempo de treinamento na análise dos modelos. Neste estudo, consideramos um ótimo regressor quando possui um R acima de 0.9 e um RRSE(%) menor que 5%.

A baixo temos a figura 4 que apresenta a tabela de regressores e os resultados da investigação desses regressores é mostrado na figura 5. Para o algoritmo de regressão linear obteve coeficiente de correlação médio de 0.8488, com desvio padrão de 0.05, mas apresentou um RRSE(%) médio de 51.09% alto. Esse algoritmo apresenta bom resultado no tempo de treinamento (menos que 1 segundo), todavia os resultados das métricas usadas na avaliação dos modelos estão fora do que foi estabelecido como bom para este estudo.

Regressor	Parameters
Linear Regressor	-
SVR	Linear Kernel, C=0.1 Polynomial kernel, degree 1, C=0.1 Polynomial kernel, degree 2, C=0.1 Polynomial kernel, degree 3, C=0.1 RBF kernel
Random Forest	Number of trees: 10, 20, 30, 40.
MLP	Number of neurons in the hidden layer: 10, 20, 30

Figura 4 – Tabela de Regressores e suas configurações.

Para o algoritmo Random Forest foram usadas quatro configurações, e obtivemos como média do coeficiente de correlação 0.9687, 0.9745, 0.9764 e 0,9773 para as configurações de 10, 20, 30 e 40 árvores, nesta ordem. Para esse algoritmo a média do RRSE(%) variam de 22.53% a 25.47% e diminui quando aumenta o número de árvores, e em todos o desvio padrão associado variando entre 4.9 e 5.07, ou seja, não apresentam grande variação.

Para os experimentos com SVM apenas as configurações com kernel polinomial de graus 2 e 3, com média de coeficientes de correlação de 0.9488 e 0.9875, mutuamente. Sendo a configuração com kernel polinomial de grau 3 com o menor erro RRSE(%) de 14.60%. No entanto, a configuração kernel RBF obteve o menor desempenho, sendo o coeficiente de correlação médio de 0.8002 e RRSE(%) de 63.84%. Já para os experimentos com o algoritmo redes neurais MLP, teve resultados aceitáveis, sendo as médias dos coeficientes de correlação 0.9750, 0.9770 e 0.9809, nesta ordem para as configurações de 10, 20 e 30 neurônios. E os valores para o RRSE(%) médio variando entre 23.29% e 20.61% e esses valores decrescem quando o número de neurônios aumenta.

A tabela da figura 5 mostra que o algoritmo SVM com kernel polinomial de grau 3 foi o que apresentou a melhor performance, contudo ele foi o que apresentou o maior tempo de treinamento. Já o pior desempenho foi apresentado pelos algoritmos regressão linear e SVM com kernel RBF, os quais tiveram maus resultados para o coeficiente de correlação e para o RRSE(%). Os erros RRSE(%) da regressão linear e do SVM com kernel RBF foram de 51.09% e 63.84% , respectivamente, que são valores considerados bem altos. Random Forest e Multilayer Perceptron tiveram desempenho semelhantes tanto com relação do coeficiente de correlação (sempre maior que 0.96) quanto ao erro associado (maior que 20%), contudo no tempo de treinamento o Random Forest tem vantagem.

As figuras 7 e 8 mostram os mapas de predição com algoritmo SVM (kernel, p=3) para o primeiro bimestre de 2014. A região sul do mapa indicam uma alta concentração de

		RRSE(%)		R		Training time (s)	
Regression method	Configuration	average	standard deviation	average	standard deviation	average	standard deviation
Linear Regression	-	51.09	13.86	0.8488	0.05	0.09	0.03
Random Forest	10 trees	25.47	5.07	0.9687	0.01	0.31	0.06
	20 trees	23.54	4.93	0.9745	0.01	0.62	0.11
	30 trees	22.87	4.91	0.9764	0.01	0.93	0.16
	40 trees	22.53	4.91	0.9773	0.01	1.27	0.20
Support Vector Regressor	polynomial kernel, p=1	53.34	14.50	0.8343	0.06	18.72	6.23
	polynomial kernel, p=2	30.30	9.98	0.9488	0.03	191.17	65.32
	polynomial kernel, p=3	14.60	6.41	0.9875	0.01	3121.20	1083.96
	RBF kernel	63.84	7.44	0.8002	0.06	19.49	4.48
MLP, one hidden layer	10 neurons	23.29	9.44	0.9750	0.02	62.68	7.95
	20 neurons	22.63	8.44	0.9770	0.01	74.50	9.50
	30 neurons	20.61	7.61	0.9809	0.01	86.76	10.49

Figura 5 – Resultados do coeficiente de correlação, root relative squared error e tempo médio de treino referente à regressão linear, Random Forest (10, 20 ,30 e 40 árvores), SVM (kernels polinomiais de graus 1, 2 e 3, e kernel RBF) e MLP(10, 20 e 30 neurônio).

criadouros de mosquitos, com destaque para os bairros do Jordão (região 1) e Ibura (região 2). Altas concentrações de criadouros foram mostradas nos bairros Monteiro (região 8), Alto do Mandu (região 9) e Casa Amarela (região 10) na região norte da Cidade do Recife. Na região oeste os bairros mais afetados foram Curado (região 3), Jardim São Paulo (região 4), Torrões (região 5) e Engenho do Meio (região 6) e parte do bairro de Tejipió (região 7). Do primeiro para o segundo bimestre de 2014 (7(b)), alguns bairros tiveram um aumento significativos com por exemplo Guabiraba (região 11), Várzea (região 12), Cohab (região 13), Macaxeira (região 14), Nova Descoberta (região 15), Vasco da Gama (região 16) e Alto José do Pinho (região 17).

Do segundo para o terceiro bimestre de 2014 (figura 7(c)) houve um aumento de concentração de criadouros na região norte da cidade, com destaque para Água Fria (região 48), Alto José do Pinho (região 17), Arruda (região 64), Bomba do Hemetério (região 21) e Parnamirim (região 55).

As figuras 7(c) e 7(d) mostram as predições para os terceiro e quarto bimestres

Table 3: Comparison of the results obtained for linear regression and SVM 3-degree polynomial kernel

Year	Bimester	SVM, kernel = RBF		SVM – 3-degree poly-kernel	
		CC	RRSE%	CC	RRSE%
2014	1	0.7671	68.26	0.9953	9.70
	2	0.7664	68.23	0.9968	8.03
	3	0.7649	68.35	0.9964	8.53
	4	0.8296	58.73	0.9987	5.05
	5	0.8142	61.46	0.9978	6.69
	6	0.7935	64.33	0.9971	7.61
2015	1	0.8278	60.86	0.9898	14.25
	2	0.8131	61.82	0.9917	12.99
	3	0.8492	56.54	0.9903	14.06
	4	0.7781	66.44	0.9781	21.40
	5	0.8044	61.69	0.9944	10.55
	6	0.7739	67.70	0.9768	21.62
2016	1	0.7675	66.32	0.9684	25.19
	2	0.8271	59.05	0.9907	13.64
	3	0.7522	74.75	0.9903	14.13
	4	0.8352	58.23	0.9962	8.76
	5	0.8210	63.53	0.9901	14.22
	6	0.9424	41.47	0.9999	1.04

Figura 6 – Comparação dos resultados obtidos para regressão linear e SVM com kernel polinomial de grau 3.

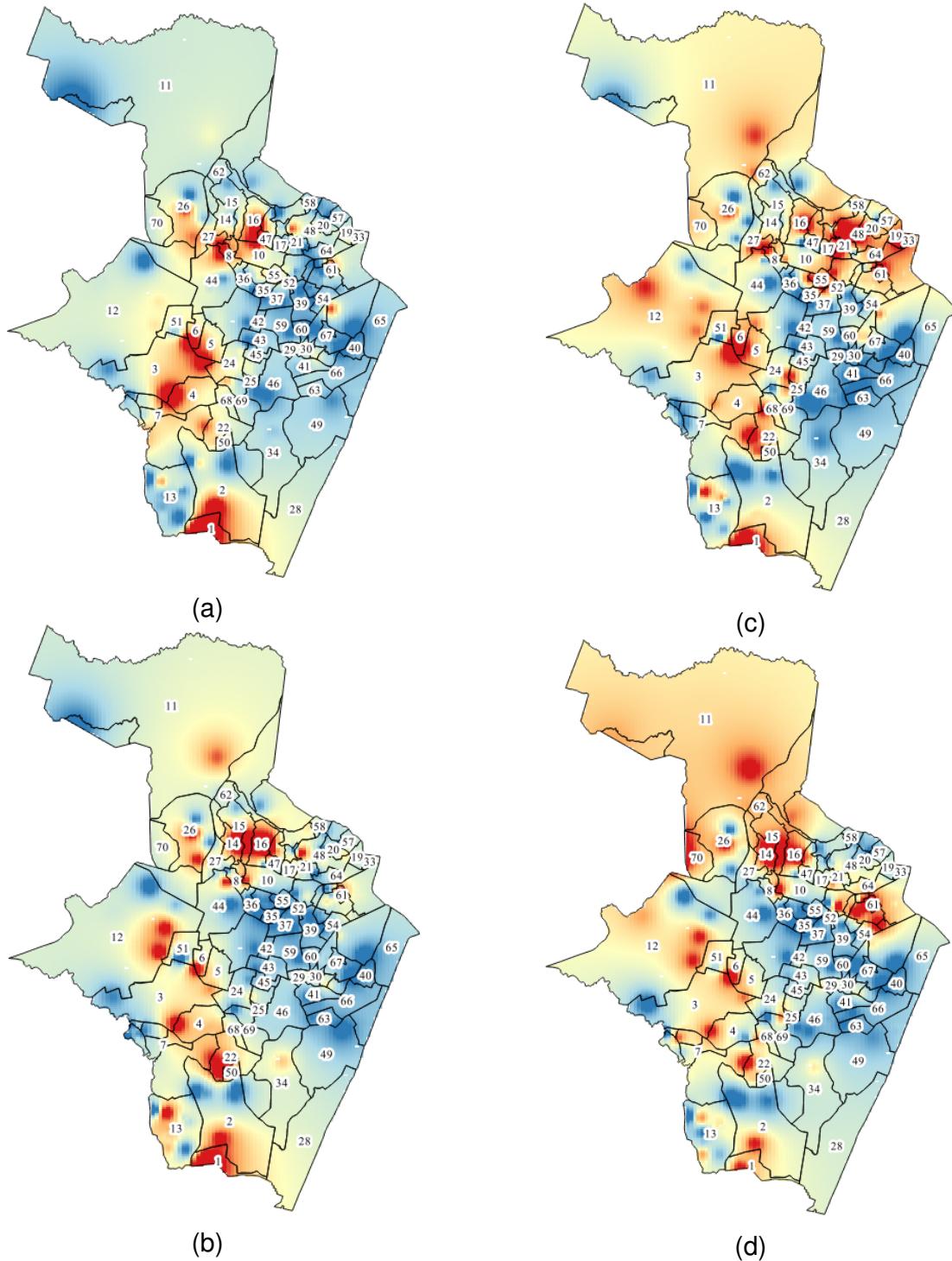


Figura 7 – Visualização dos mapas de predição com algoritmo SVM (kernel, $p=3$) para o primeiro (a), segundo (b), terceiro (c) e quarto (d) bimestres de 2014, nessa ordem.

de 2014, e apresentam uma concentração alta a moderada de criadouros de mosquito em grande parte da cidade. Do terceiro para o quarto bimestre de 2014 (figura 7(d)), alguns bairros da região norte e oeste da cidade tiveram um aumento da concentração de criadouros. Na região norte, temos Água Fria (região 48), Alto José do Pinho (região 17), Mangabeira (região 18), Campina do Barreto (região 19), Fundão (região 20) e Bomba do

Hemetério (região 21). Na região oeste, Várzea (região 12), Areias (região 22), Engenho do Meio (região 6) e parte dos bairros da Mustardinha (região 23), San Martin (região 24) e Mangueira (regiões 25). Contudo regiões como Casa amarela (região 10), Parnamirim (região 55) e Jaqueira (região 52) diminuiram a concentração de criadouros.

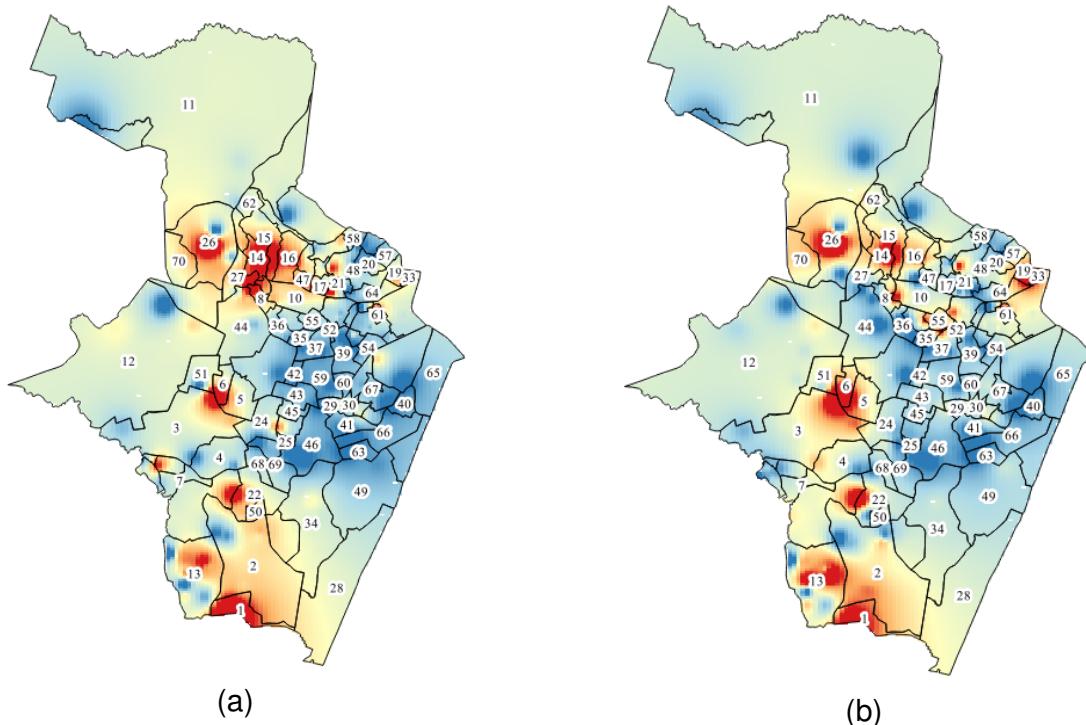


Figura 8 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o quinto (a) e sexto (b) bimestres de 2014, nessa ordem.

Do quarto (figura 7(d)) para o quinto bimestre de 2014 (figura 8(a)), o mapa de distribuição ficou mais azul, ou seja, boa parte dos bairros teve uma diminuição da concentração de criadouros. Na região norte, destaque para os bairros Guabiraba (região 11), Casa Amarela (região 10) e Hipódromo (região 61). Na zona oeste, destaque para os bairros da Várzea (região 12), Jardim São Paulo (região 4) e partes do Curado (região 3). Contudo bairros como Jordão (região 1) e Cohab (região 13) aumentaram suas concentrações de criadouros de mosquitos.

Para o quinto e o sexto bimestre de 2014 (figura 8(b)), a concentração de criadouros foi maior nos bairros das regiões norte e sul do Recife. Ao norte se destacam os bairros de Dois Irmãos (região 26), Macaxeira (região 14), Alto José do Pinho (região 17), Mangabeira (região 18), Nova Descoberta (região 15), Apipucos (região 27), Monteiro (região 8), Vasco da Gama (região 16) e Alto do Mandu (região 9).

Para o primeiro bimestre de 2015 (figura 9(a)) houve uma diminuição de criadouros na região oeste, exceto no bairro da Várzea (regiões12) que teve um discreto aumento. Para a região sul houve uma pequena diminuição da concentração de criadouros nos bairros da Cohab (região 13), Boa Viagem (região 28) e Ibura (região 2), excetuando o bairro do

Jordão (regiões 1). A região norte teve um aumento do número de criadouros do mosquito. Esse aumento foi apresentado por bairros da região leste do mapa, como Ilha do Retiro (região 29), Paissandu (região 30), Ilha do Leite (região 31) e Coelhos (região 32). Os bairros de Campina do Barreto (região 19) e Peixinhos (região 33) tiveram um alto nível de concentração se opondo ao comportamento da região nordeste da cidade que teve moderados níveis de concentração. No segundo bimestre 2015 (figura 9(b)), os bairros mais ao norte e nordeste da cidade mostraram uma alta concentração de criadouros, com destaque para os bairros de Dois Irmãos (região 26) e Sítio dos Pintos (região 70). Já na região sul também houve um aumento na concentração, destacando os bairros da Cohab (região 13) e Ibura (regiões 2). Entretanto o bairro do Jordão (região 1) se manteve com uma alta concentração.

Na figura 9(c) foi observada uma diminuição significativa na concentração de criadouros da predição para cidade do Recife. Contudo os bairros da Cohab (região 13) e do Ibura (região 2) contrastam com o comportamento geral da predição, esse primeiro manteve o alto índice de criadouros. E o bairro da Várzea (região 12) se destacou por apresentar uma alta concentração de criadouros no terceiro bimestre quando comparado ao primeiro e ao segundo.

Do terceiro para o quarto bimestre de 2015 (figura 9(d)), os bairros que tiveram aumento de concentração de criadouros foram Peixinhos (região 33), Jordão (região 1), Vasco da Gama (região 16), Macaxeira (região 14), Nova Descoberta (região 15), Alto José Bonifácio e Imbiribeira (região 34). Por outro lado, alguns bairros tiveram uma menor concentração de criadouros. Foram eles Santo Antônio (região 40), Ilha Joana Bezerra (região 41) e Paissandu (região 30), na região central da cidade, Poço da Panela (região 36), Torre (região 37), Casa Forte (região 38) e Graças (região 39), da região norte, e Zumbi (região 42), Prado (região 43), Iputinga (região 44), Bongi (região 45) e Afogados (região 46), da região oeste.

A concentração de criadouros do quarto para o quinto bimestre de 2015 (figura 10(a)) voltou a aumentar em vários bairros do Recife. Um aumento intenso foi apresentado nos bairros do Pina (região 49), Cohab (região 13) e Dois Irmãos (região 26). Contudo, bairros que ainda apresentaram altas ou moderadas concentrações de criadouros tiveram uma queda visível nessas concentrações. São eles Nova Descoberta (região 15), Vasco da Gama (região 16) e Campina do Barreto (região 19), nas regiões norte e nordeste do mapa. O bairro da Imbiribeira (região 34) teve um leve decrescimento.

Do quinto para o último bimestre de 2015 (figura 10(b)) nas regiões de centro e norte da cidade tiveram uma diminuição na concentrações de criadouros. Com destaque para os bairros do Caçote, Várzea (região 12), Cidade Universitária, Jaqueira (região 52), Graças (região 39), Aflitos, Espinheiro (região 54), Parnamirim (região 35), Porto da Madeira , Cajueiro e Beberibe. Contudo houve aumento de criadouros nos bairros de Vasco da

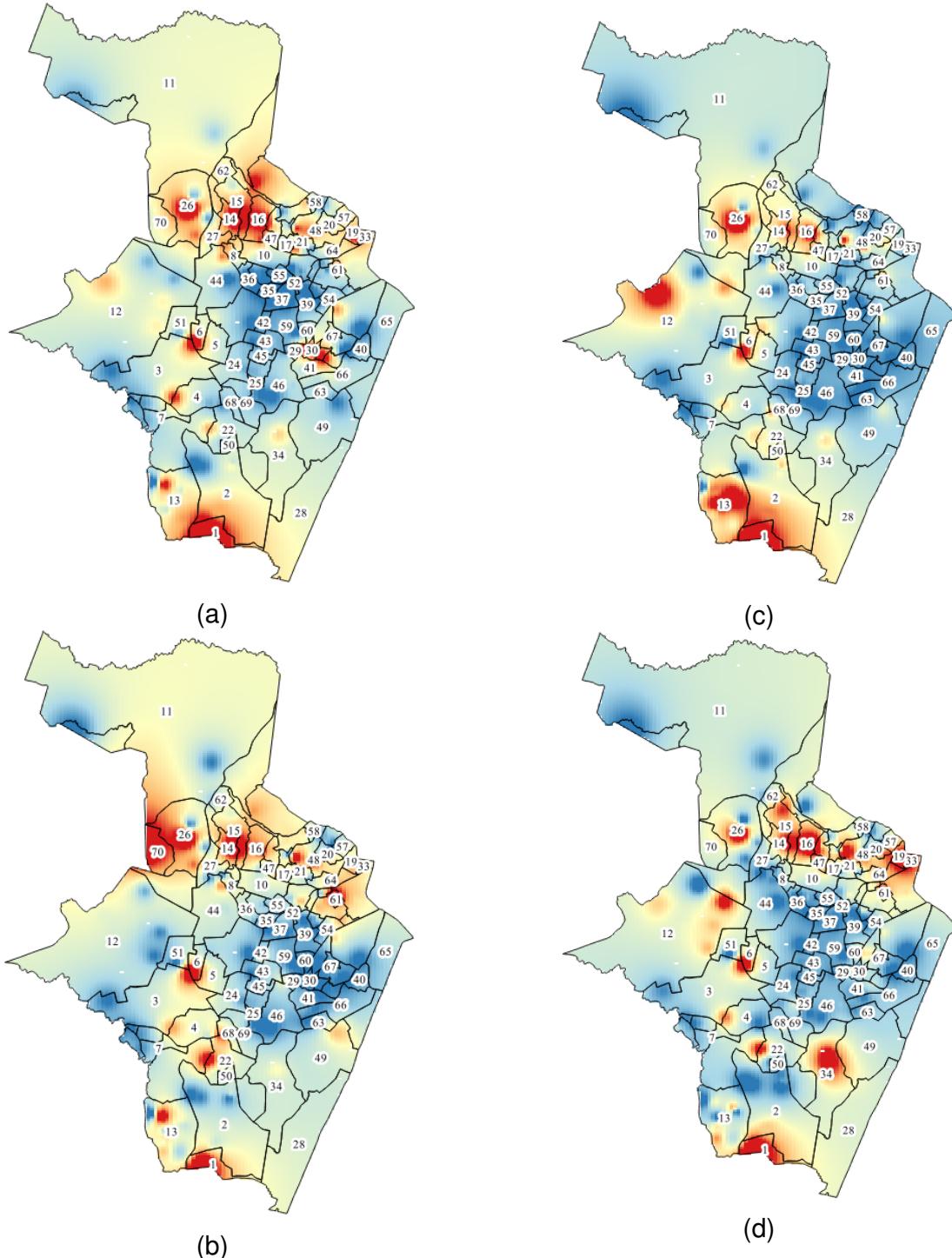


Figura 9 – Visualização dos mapas de predição com algoritmo SVM (kernel, $p=3$) para o primeiro (a), segundo (b), terceiro (c) e quarto (d) bimestres de 2015, nessa ordem.

Gama (região 16), Macaxeira (região 14) e Alto do Mandu (região 9).

O mapa de predição do primeiro bimestre de 2016 (figura 11(a)) teve a menor concentração de criadouro nas regiões centro e norte da cidade. Destaque para o bairro do Zumbi (região 42), Prado (região 43), Bongi (região 44), Mustardinha (região 45), Afogados (região 46), Ilha Joana Bezerra (região 41), Aflitos, Santo Antônio (região 40), Ilha do Leite

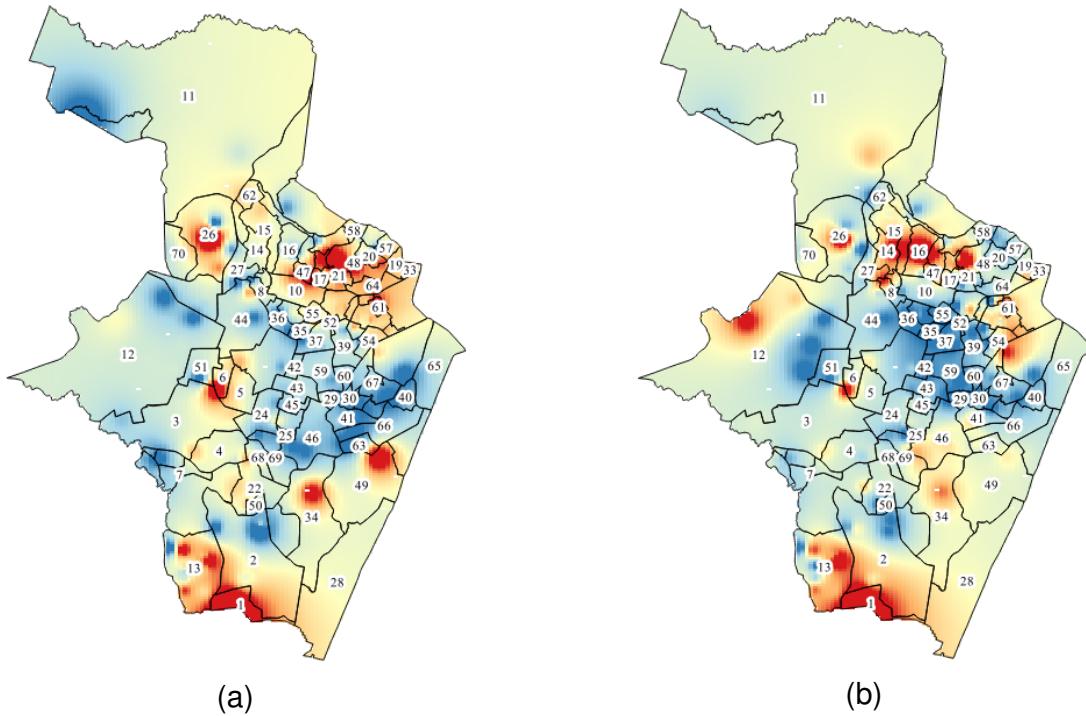


Figura 10 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o quinto (a) e sexto (b) bimestres de 2015, nessa ordem.

(região 31), Madalena (região 59) e Derby (região 60). Já os bairros de Dois Irmãos (região 26), Imbiribeira, Jordão (região 1), Engenho do Meio (região 6), Várzea (região 12), Jardim São Paulo (região 4) e Vasco da Gama (região 16) mostraram uma maior concentração dos criadouros de mosquitos. E a região de centro foi a que apresentou menor concentração de criadouros.

Do mapa do primeiro para o mapa do segundo bimestre de 2016 (figura 11(b)), as maiores concentrações de criadouros foram apresentadas pelos bairros de Dois Irmãos (região 26), Vasco da Gama (região 16), Hipódromo (região 61), Brejo da Guabiraba, Alto José Bonifácio (região 17) e Guabiraba (região 11), na região norte. E na região oeste e sul as altas concentrações foram observadas nos bairros de Jordão (região 1), Engenho do Meio (região 6), Imbiribeira (região 34) e Jardim São Paulo (região 4). Já nos bairros de Casa Forte (região, Parnamirim (região 55), Santana (região 35), Torre (região 37), Jaqueira (região 52) e Madalena (região 59) na região central da cidade tiveram as menores concentrações de criadouros de *Aedes aegypti*.

Do segundo para o terceiro bimestre de 2016 (figura 11(a)) as concentrações de criadouros aumentaram em algumas regiões. Na região oeste se destacam a Vázea (região 12) e Engenho do Meio (região 6). Na região sul, o bairro da Imbiribeira (região 34). E na região norte os bairros de Vasco da Gama (região 16) e Campina do Barreto (região 19). E a região central manteve as concentrações aparentemente constantes, como nos bairros Ilha do Leite (região 31), Coelhos (região 32), São José (região 66), Ilha Joana Bezerra

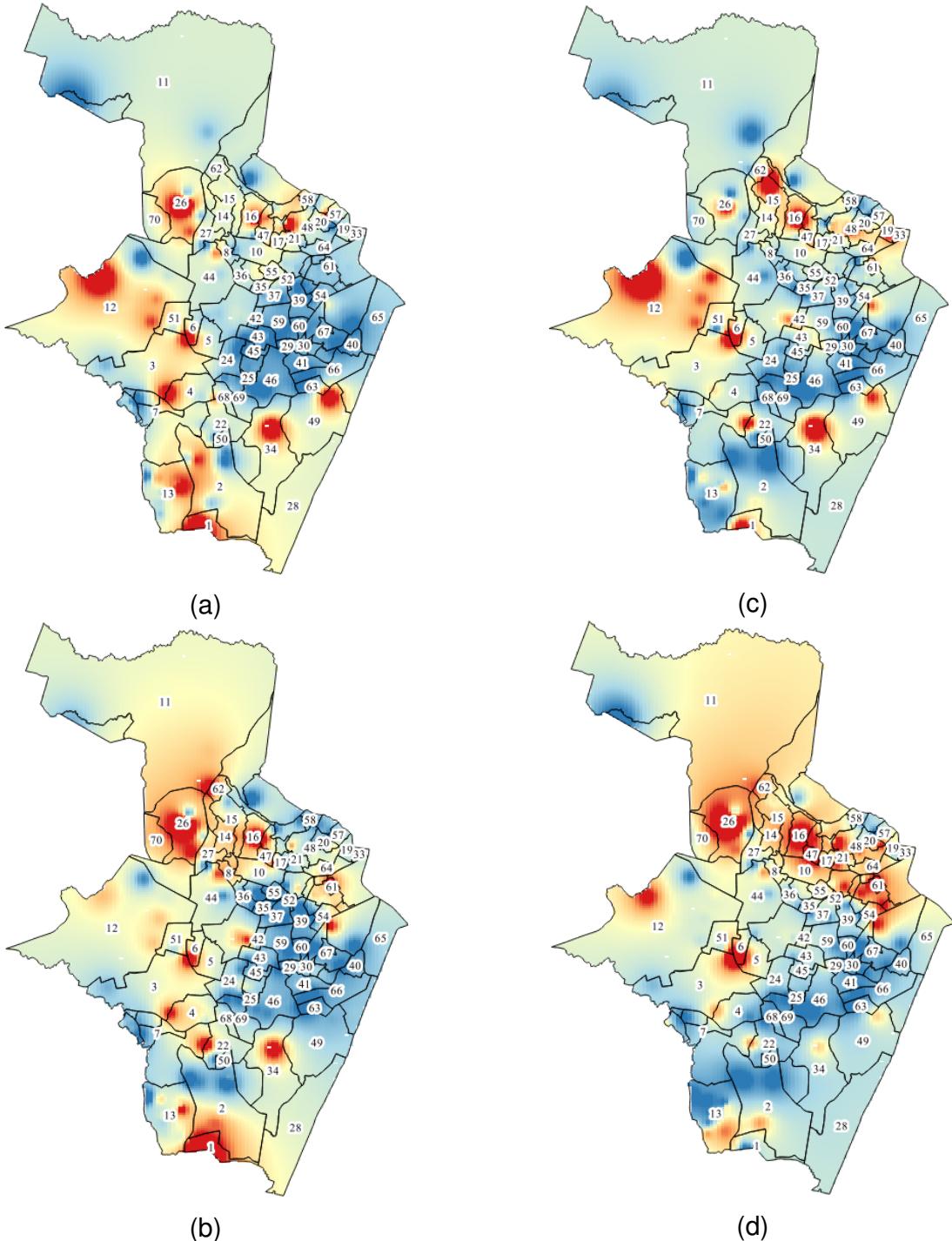


Figura 11 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o primeiro (a), segundo (b), terceiro (c) e quarto (d) bimestres de 2016, nessa ordem.

(região 41) e Boa Vista (região 67). Na zona oeste os bairros da Madalena (região 59) e Afogados (região 46) também tiveram baixas concentrações de criadouros.

Os bairros da região norte tiveram um aumento considerável na concentração de criadouros, se sobressaindo com relação aos bairros das outras regiões. Com destaque para os bairros Vasco da Gama (região 16), Morro da Conceição (região 47) e Hipódromo

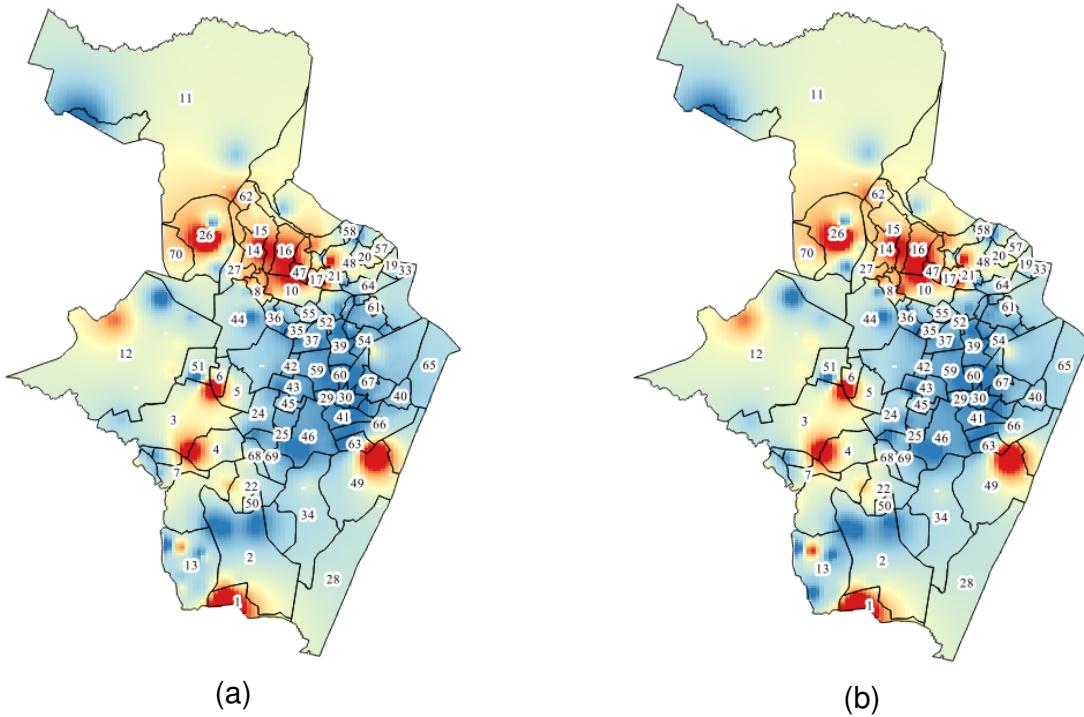


Figura 12 – Visualização dos mapas de predição com algoritmo SVM (kernel, p=3) para o quinto (a) e sexto (b) bimestres de 2016, nessa ordem.

(região 61). Já as regiões sul e oeste tiveram uma diminuição na concentração de criadouros, como mostra o mapa da figura 11(d). Destacando os bairros da Imbiribeira (região 34), Areias (região 22) e Pina (região 49) na região sul do mapa. E na região oeste, os bairros da Várzea (região 12) e Curado (região 3). E as menores concentrações foram observadas na região central da cidade, ênfase para os bairros Paissandu (região 30), Derby (região 60), Ilha do Leite (região 31) e Ilha Joana Bezerra (região 41).

Do quarto para o quinto bimestre de 2016 (figura 12(a))houve uma leve diminuição na distribuição da concentração de criadouros nos bairros da zona norte, contudo no Vasco da Gama (região 16), Morro da Conceição (região 47) e Hipódromo (região 61) as concentrações se mantiveram moderadas ou altas. O aumento da concentração foi observado nos bairros de Engenho do Meio (região 6), Jordão (região 1), Pina (região 49) e Jardim São Paulo (região 4). E nos bairros do Derby (região 60), Ilha do Leite (região 31) e Ilha Joana Bezerra (região 41) houve a menor concentração de criadouros.

E por fim da a evolução do quinto para o sexto bimestre de 2016 (figura 12(b)), a região central do mapa manteve a baixa concentração de criadouros. Destacando os bairros de Paissandu (região 30), Boa vista (região 67) e Santo Antônio (região 40). Bairros da região norte como Vasco da Gama (região 16), Morro da Conceição (região 47) e Nova Descoberta (região 15). Na região sul os bairros de Pina (região 49) e Jordão (região 1) da zona sul e Dois Irmãos (região 26) e Engenho do Meio (região 6) da região oeste apresentaram altas concentrações de criadouros.

5.2 Predição temporal de Covid-19

A autora participou do estudo (LIMA et al., 2020) para predição de casos de Covid-19 com abordagem temporal dos 27 estados brasileiros e do Brasil, todos os resultados do estudo serão apresentados em detalhes nesta seção.

As figuras 13, 14, 15, e 16 mostram as predições (de 6 dias) do número de casos de Covid-19 para todos os estados brasileiros e do Brasil no período de 05 a 11/06/2020. A janela de predição é curta, pois no início do projeto eram usados os dados da Secretaria Estadual de Pernambuco. E inicialmente os dados disponíveis era uma planilha do SINAN com dados de duas semanas, e o governo do Estado de Pernambuco estava juntando um comitê multidisciplinar de especialistas para tomada de decisão da semana todo domingo a noite. Então foi pensado em uma predição semanal para ser mais conveniente e ajudar em tal tomada de decisão.

5.2.1 Predição ARIMA

Os coeficientes de correlação de Pearson, Spearman e Kendall e o RMSE(%) foram usados para avaliação dos modelos construídos, sendo o primeiro para avaliação global e o segundo para avaliação local. A predição foi feita usando a série temporal, a contar do primeiro dia de notificação de casos e óbitos de cada uma das unidades federativas brasileiras. A tabela 3 mostra os resultados dos coeficientes de correlação de Pearson, Spearman e Kendall, e do erro quadrático relativo (RMSE%) para os modelos ARIMA construídos para o Brasil e seus 27 estados.

As tabelas 17, 18, 19, 20 e 21 mostram predições feitas com o método ARIMA para o Brasil e suas 27 unidades federativas. Para avaliar o desempenho dos modelos foi usado o erro quadrático médio entre os valores previstos e os casos reais dos casos acumulados de Covid-19, bem como o erro absoluto entre os casos reais e o mínimo ou limite da previsão. Para as situações em que os valores reais estão fora do intervalo predito os valores estão em vermelho, e o erro absoluto entre o limite inferior ou superior é menor do que 5%.

Os resultados dos modelos ARIMA para região Nordeste do Brasil são exibidos na tabela da figura 17. No período de 06 a 11 de maio de 2020, os casos reais da Bahia, Maranhão, Piauí e Rio Grande do Norte ficaram dentro da faixa de predição. O estado do Piauí obteve o melhor desempenho, com erros entre os casos reais e as predições entre 0,54% e 3,36%. Já o estado de Sergipe teve o pior desempenho, foi observado que para os 6 dias de predição apenas os casos para 6 de maio estavam dentro da predição e nos outros dias os casos reais foi superior ao limite máximo das predições, com erros de 8,52%, 12,39%, 14,16%, 15,09% e 8% nos dias seguintes (na ordem).

Os resultados obtidos para os estados da região Norte estão mostrados na tabela da figura 18. Amapá e Rondônia foram os estados que obtiveram os melhores desempenhos,

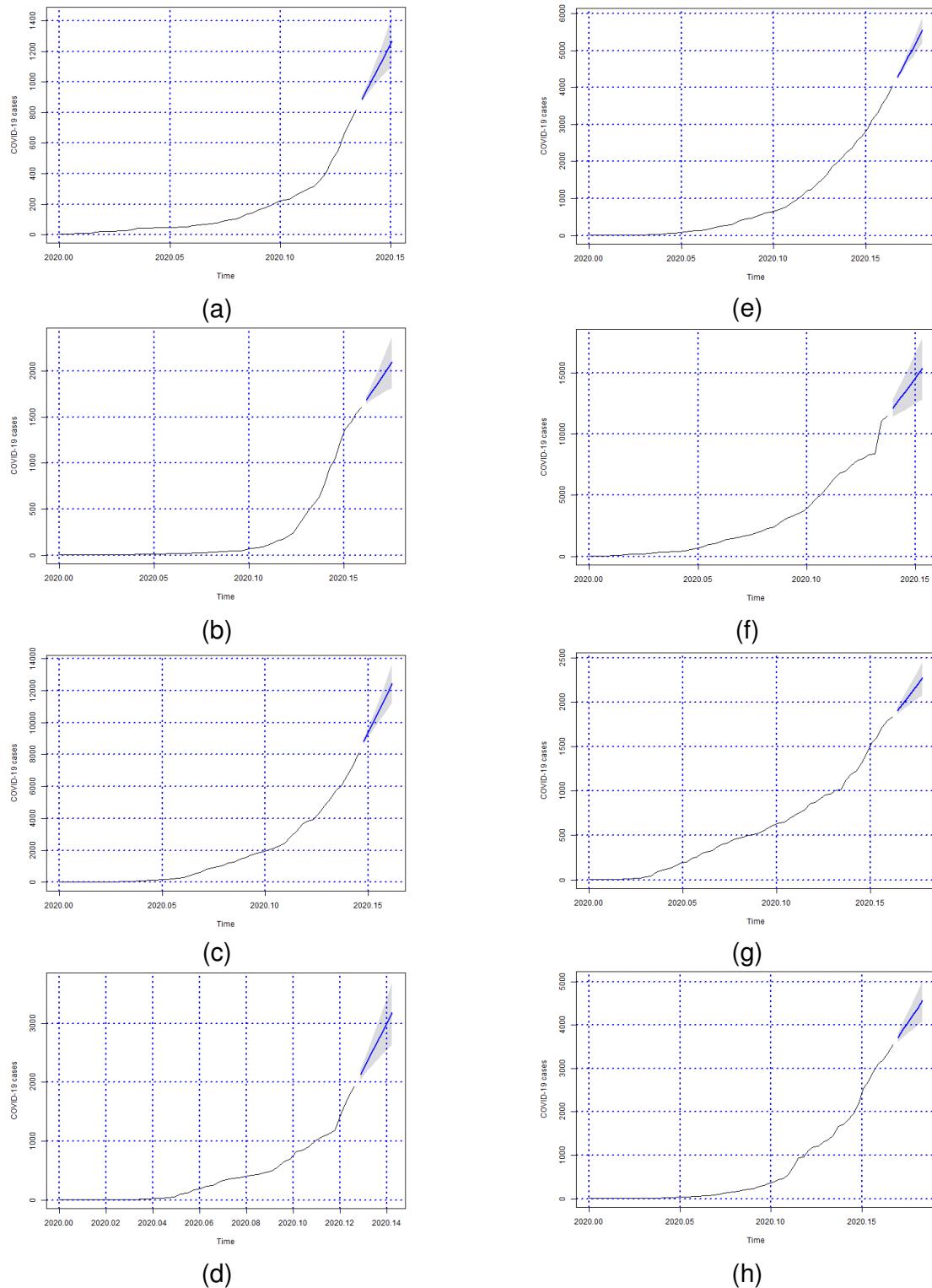


Figura 13 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Acre, (b) Alagoas, (c) Amazonas, (d) Amapá, (e) Bahia, (f) Ceará, (g) Distrito Federal, e (h) Espírito Santo.

e para todos os 6 dias de predição os valores se mantiveram dentro dos limites de predição. Em contrapartida, Roraima e Tocantis tiveram os piores desempenhos, com erros que atingiram mais de 55 e 25%.

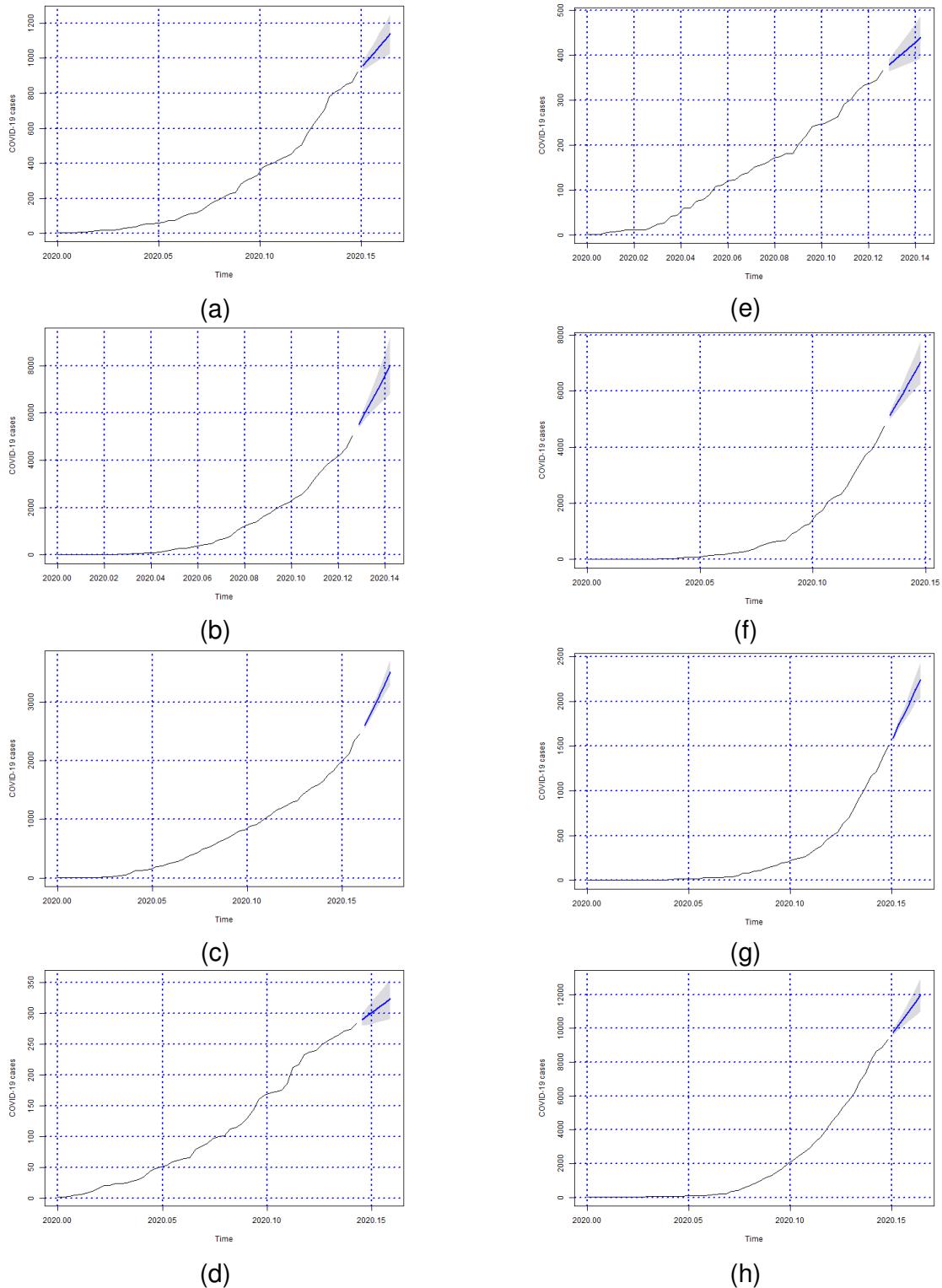


Figura 14 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Goiás, (b) Maranhão, (c) Minas Gerais, (d) Mato Grosso do Sul, (e) Mato Grosso, (f) Pará, (g) Paraíba, e (h) Pernambuco.

Os resultados obtidos para os estados da região Centro-Oeste e Distrito Federal são exibidos na tabela da figura 19. O Mato Grosso do Sul e Goiás foram os modelos que apresentaram os melhores desempenhos. O modelo do Mato Grosso do Sul não obteve os

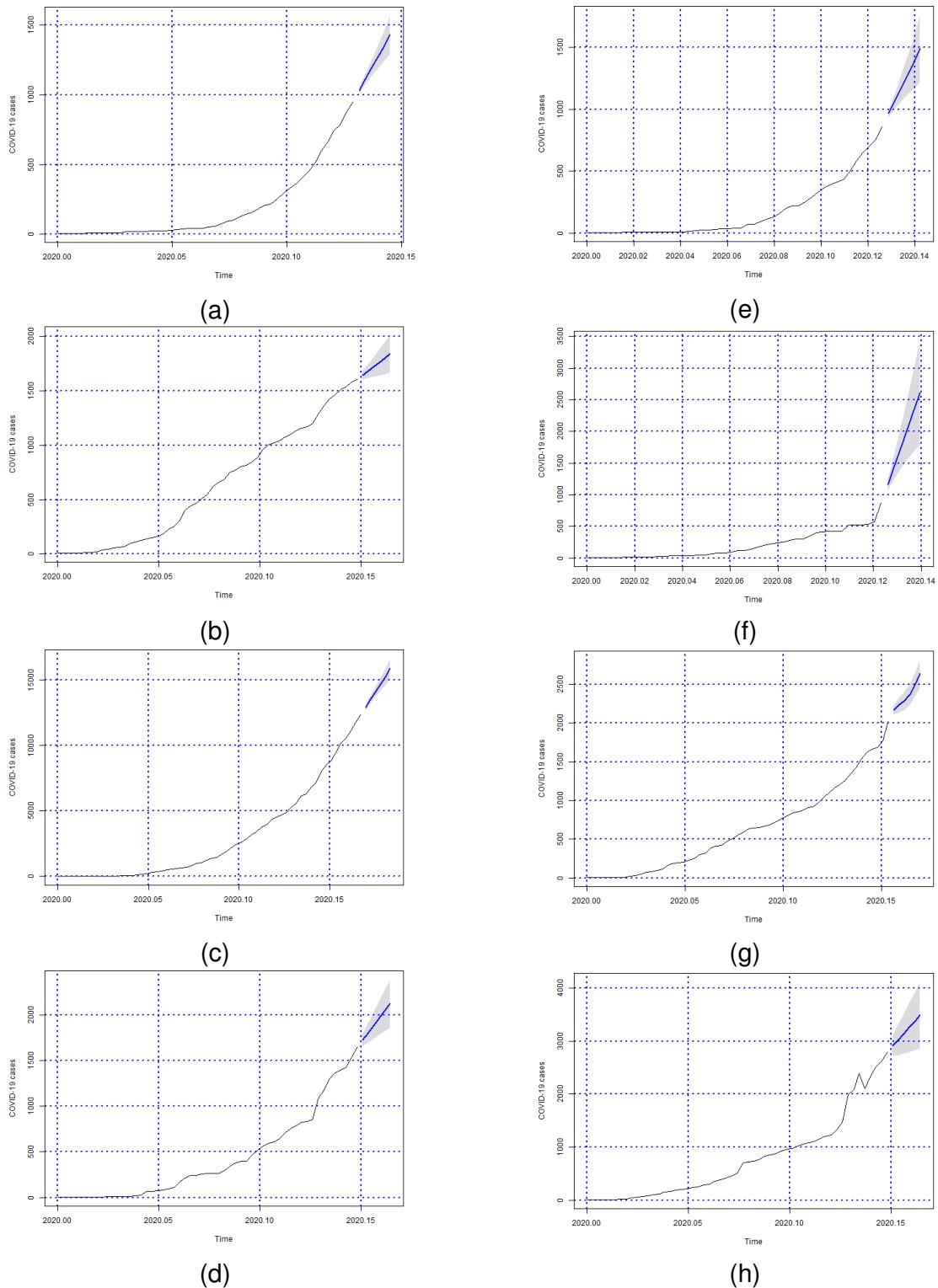


Figura 15 – Predição do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Piauí, (b) Paraná, (c) Rio de Janeiro, (d) Rio Grande do Norte, (e) Rondônia, (f) Roraima, (g) Rio Grande do Sul, e (h) Santa Catarina.

valores dentro da predição para todo os dias do período de avaliação, mas o erro entre os casos reais e o limite superior foi menor do que 5%. Já para o estado de Goiás, apenas

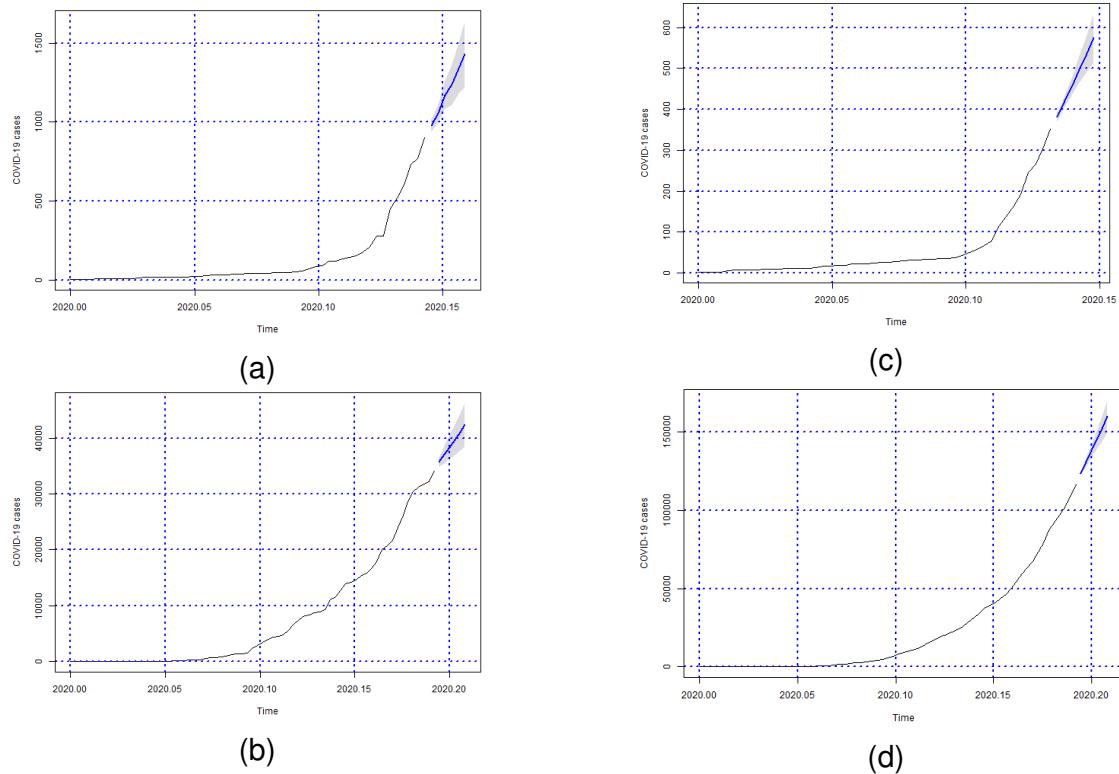


Figura 16 – Previsão do número de casos acumulados de Covid-19 no período de 06 a 11/05/2020 para os estados de (a) Sergipe, (b) São Paulo, (c) Tocantins, e (d) todo o país.

state	model	pearson	spearman	kendall	rmse	rmse (%)	MAPE	MPE	MAE	ME	number of observations
Acre	ARIMA(2,2,0)	0.99890	0.99515	0.96645	9.30	4.78	7.50	1.48	6.34	1.56	50
Alagoas	ARIMA(0,2,1)	0.99873	0.99371	0.96337	22.21	5.13	6.93	2.36	10.83	2.50	59
Amazonas	ARIMA(0,2,1)	0.99928	0.99899	0.99194	87.65	4.10	6.56	3.77	54.04	20.92	54
Amapá	ARIMA(0,2,1)	0.99647	0.99826	0.98327	43.04	8.78	10.59	3.77	21.64	8.31	47
Bahia	ARIMA(2,2,1)	0.99971	0.99892	0.98795	28.59	2.56	7.78	1.30	18.40	7.36	61
Ceará	ARIMA(0,2,1)	0.99408	1.00000	1.00000	351.5	11.31	5.69	4.65	122.61	75.44	51
Distrito Federal	ARIMA(0,2,1)	0.99922	0.99937	0.99151	20.70	4.02	7.00	3.56	14.29	3.87	60
Espírito Santo	ARIMA(0,2,1)	0.99874	0.99870	0.98565	51.50	5.09	7.82	3.70	26.9	8.60	62
Goiás	ARIMA(0,2,1)	0.99884	0.99971	0.99596	13.55	4.91	5.50	2.62	8.54	2.67	55
Maranhão	ARIMA(0,2,0)	0.99908	0.99867	0.98841	63.78	4.42	10.00	1.20	41.77	10.57	47
Minas Gerais	ARIMA(1,2,2)	0.99945	0.99982	0.99766	24.16	3.46	6.88	3.40	15.64	4.57	59
Mato Grosso do Sul	ARIMA(0,2,1)	0.99860	0.99966	0.99600	4.88	5.40	5.06	2.37	3.36	0.78	53
Mato Grosso	ARIMA(0,2,1)	0.99828	0.99841	0.98654	6.76	6.03	7.75	3.54	4.96	1.58	47
Pará	ARIMA(0,2,1)	0.99868	0.99941	0.99191	71.56	5.52	9.5	4.44	41.76	19.46	49
Paraíba	ARIMA(3,2,0)	0.99931	0.99586	0.97577	15.11	3.91	13.92	3.37	8.52	3.42	55
Pernambuco	ARIMA(0,2,1)	0.99949	0.99955	0.99360	89.79	3.28	7.58	3.08	48.25	19.18	55
Piauí	ARIMA(1,2,0)	0.99928	0.99813	0.98266	10.28	4.00	7.2	1.12	6.41	2.55	48
Paraná	ARIMA(0,2,2)	0.99942	0.99949	0.99392	18.18	3.47	4.96	2.90	12.96	2.67	55
Rio de Janeiro	ARIMA(0,2,2)	0.99960	0.99918	0.99258	105.29	2.96	8.78	2.44	67.51	27.06	62
Rio Grande do Norte	ARIMA(0,2,1)	0.99741	0.99727	0.97657	34.52	7.37	8.24	4.36	18.31	6.99	55
Rondônia	ARIMA(0,2,0)	0.99819	0.99485	0.96633	14.40	6.20	10.61	-0.09	7.98	2.19	47
Roraima	ARIMA(0,2,0)	0.97919	0.99321	0.94769	43.73	21.08	12.08	1.56	19.02	6.33	46
Rio Grande do Sul	ARIMA(2,2,2)	0.99882	0.99940	0.99278	27.95	5.10	7.56	2.71	17.73	6.35	57
Santa Catarina	ARIMA(0,2,1)	0.99254	0.99955	0.99562	96.88	12.41	6.72	4.25	38.95	17.18	55
Sergipe	ARIMA(3,2,0)	0.99598	0.99631	0.97041	18.97	9.11	9.84	1.39	7.93	3.22	53
São Paulo	ARIMA(0,2,2)	0.99917	0.99914	0.98688	411.73	4.14	9.96	4.26	243.34	74.83	71
Tocantins	ARIMA(0,2,2)	0.99785	0.99308	0.95766	5.52	6.76	10.02	2.40	3.07	1.20	49
Brazil	ARIMA(0,2,1)	0.99978	0.99718	0.97963	688.24	2.20	15.65	-0.06	401.29	142.02	71

Tabela 3 – Resultados dos coeficientes de correlação de Pearson, Spearman e Kendall, e do erro quadrático relativo (RMSE%) para os modelos ARIMA construídos para o Brasil e seus 27 estados.

a previsão do dia 6 de maio ficou fora do intervalo de previsão, mas o erro percentual absoluto não passou de 5%. Os estados do Mato Grosso e Distrito Federal apresentaram

State	Date	Forecasted value	Lower limit	Upper limit	Actual cases	Absolute percentage error*	Absolute percentage error**	MAPE
Alagoas	2020-05-06	1,687	1,642	1,732	1,703	0.95	—	
	2020-05-07	1,769	1,686	1,851	1,867	5.27	0.84	
	2020-05-08	1,851	1,725	1,976	2,033	8.97	2.81	7.91
	2020-05-09	1,932	1,759	2,105	2,170	10.95	2.98	
	2020-05-10	2,014	1,789	2,239	2,258	10.80	0.83	
	2020-05-11	2,096	1,815	2,377	2,343	10.54	—	
Bahia	2020-05-06	4,288	4,230	4,347	4,301	0.29	—	
	2020-05-07	4,515	4,416	4,614	4,528	0.29	—	
	2020-05-08	4,821	4,672	4,969	4,818	0.05	—	2.24
	2020-05-09	5,023	4,807	5,239	5,174	2.91	—	
	2020-05-10	5,320	5,044	5,596	5,558	4.28	—	
	2020-05-11	5,555	5,200	5,910	5,808	4.35	—	
Ceará	2020-05-06	12,119	11,409	12,829	12,310	1.55	—	
	2020-05-07	12,768	11,677	13,858	13,888	8.07	0.21	
	2020-05-08	13,416	11,973	14,860	15,134	11.35	1.81	9.66
	2020-05-09	14,065	12,272	15,858	15,879	11.42	0.13	
	2020-05-10	14,714	12,567	16,861	16,692	11.85	—	
	2020-05-11	15,363	12,854	17,872	17,599	12.71	—	
Maranhão	2020-05-06	5,526	5,398	5,654	5,389	2.54	—	
	2020-05-07	6,024	5,738	6,310	5,909	1.95	—	
	2020-05-08	6,522	6,044	7,000	6,765	3.59	—	5.18
	2020-05-09	7,020	6,320	7,720	7,599	7.62	—	
	2020-05-10	7,518	6,571	8,465	8,144	7.69	—	
	2020-05-11	8,016	6,797	9,235	8,526	5.98	—	
Paraíba	2020-05-06	1,586	1,555	1,617	1,657	4.26	2.39	
	2020-05-07	1,750	1,701	1,799	1,849	5.35	2.68	
	2020-05-08	1,860	1,782	1,938	2,030	8.39	4.55	7.11
	2020-05-09	1,987	1,870	2,104	2,156	7.85	2.43	
	2020-05-10	2,136	1,983	2,289	2,341	8.75	2.21	
	2020-05-11	2,243	2,044	2,443	2,525	11.15	3.26	
Pernambuco	2020-05-06	9,767	9,586	9,948	9,881	1.16	—	
	2020-05-07	10,208	9,894	10,522	10,824	5.69	2.79	
	2020-05-08	10,650	10,193	11,107	11,587	8.09	4.14	8.69
	2020-05-09	11,092	10,479	11,705	12,470	11.05	6.14	
	2020-05-10	11,534	10,753	12,314	13,275	13.12	7.24	
	2020-05-11	11,975	11,015	12,935	13,768	13.02	6.05	
Piauí	2020-05-06	1,032	1,012	1,053	1,051	1.78	—	
	2020-05-07	1,111	1,074	1,147	1,131	1.81	—	
	2020-05-08	1,192	1,133	1,250	1,233	3.36	—	1.47
	2020-05-09	1,271	1,188	1,354	1,278	0.54	—	
	2020-05-10	1,351	1,242	1,461	1,332	1.45	—	
	2020-05-11	1,431	1,292	1,570	1,443	0.82	—	
Rio Grande do Norte	2020-05-06	1,724	1,654	1,794	1,739	0.86	—	
	2020-05-07	1,804	1,695	1,913	1,821	0.93	—	
	2020-05-08	1,884	1,737	2,031	1,919	1.82	—	3.42
	2020-05-09	1,964	1,779	2,149	1,919	2.35	—	
	2020-05-10	2,044	1,819	2,269	1,919	6.52	—	
	2020-05-11	2,124	1,858	2,390	1,989	6.79	—	
Sergipe	2020-05-06	977	938	1,016	998	2.08	—	
	2020-05-07	1,055	999	1,111	1,214	13.12	8.52	
	2020-05-08	1,171	1,083	1,260	1,438	18.55	12.39	17.64
	2020-05-09	1,236	1,108	1,363	1,588	22.18	14.16	
	2020-05-10	1,341	1,178	1,504	1,771	24.29	15.09	
	2020-05-11	1,431	1,220	1,642	1,800	20.49	8.77	

The * symbol is the error between the forecasted values and the current cases, while ** indicates the percentage error between the current value and one of the limits (lower or higher).

Figura 17 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Região Nordeste. Fonte: (LIMA et al., 2020)

State	Date	Forecasted value	Lower limit	Upper limit	Actual cases	Absolute percentage error*	Absolute percentage error**	MAPE
Acre	2020-05-06	889	870	908	943	5.73	3.69	11.48
	2020-05-07	968	931	1,005	1,014	4.54	0.90	
	2020-05-08	1,041	977	1,105	1,177	11.56	6.13	
	2020-05-09	1,118	1,025	1,211	1,335	16.26	9.27	
	2020-05-10	1,192	1,064	1,320	1,447	17.63	8.77	
	2020-05-11	1,268	1,102	1,434	1,460	13.16	1.81	
Amazonas	2020-05-06	8,833	8,656	9,010	9,243	4.43	—	5.38
	2020-05-07	9,557	9,218	9,897	10,099	5.36	2.00	
	2020-05-08	10,282	9,754	10,809	10,727	4.15	—	
	2020-05-09	11,006	10,267	11,744	11,925	7.71	1.52	
	2020-05-10	11,730	10,759	12,701	12,599	6.90	—	
	2020-05-11	12,454	11,231	13,677	12,919	3.60	—	
Amapá	2020-05-06	2,140	2,052	2,227	2,046	4.58	—	11.31
	2020-05-07	2,348	2,189	2,508	2,199	6.79	—	
	2020-05-08	2,557	2,316	2,798	2,322	10.11	—	
	2020-05-09	2,765	2,435	3,096	2,493	10.93	—	
	2020-05-10	2,974	2,545	3,403	2,613	13.82	—	
	2020-05-11	3,183	2,647	3,718	2,671	19.16	—	
Pará	2020-05-06	5,136	4,991	5,281	5,524	7.03	4.40	8.65
	2020-05-07	5,516	5,267	5,765	5,935	7.06	2.87	
	2020-05-08	5,896	5,536	6,256	6,519	9.56	4.04	
	2020-05-09	6,276	5,796	6,756	7,018	10.58	3.74	
	2020-05-10	6,656	6,046	7,265	7,348	9.42	1.13	
	2020-05-11	7,035	6,288	7,783	8,069	12.81	3.54	
Rondônia	2020-05-06	966	937	995	943	2.44	—	7.54
	2020-05-07	1,071	1,007	1,135	1,098	2.46	—	
	2020-05-08	1,176	1,068	1,284	1,222	3.76	—	
	2020-05-09	1,281	1,123	1,439	1,263	1.43	—	
	2020-05-10	1,386	1,172	1,600	1,302	6.45	—	
	2020-05-11	1,491	1,216	1,766	1,398	6.65	—	
Roraima	2020-05-06	1,160	1,072	1,248	932	24.46	15.06	62.15
	2020-05-07	1,451	1,255	1,647	1,020	42.25	23.04	
	2020-05-08	1,742	1,414	2,070	1,124	54.98	25.81	
	2020-05-09	2,033	1,553	2,513	1,202	69.13	29.20	
	2020-05-10	2,324	1,674	2,974	1,290	80.16	29.77	
	2020-05-11	2,615	1,779	3,451	1,295	101.93	37.37	
Tocantins	2020-05-06	381	370	392	423	9.90	7.24	23.19
	2020-05-07	420	405	434	494	15.00	12.07	
	2020-05-08	459	437	481	572	19.81	15.97	
	2020-05-09	497	465	530	688	27.70	22.97	
	2020-05-10	536	491	581	747	28.21	22.17	
	2020-05-11	575	516	634	828	30.55	23.37	

The * symbol is the error between the forecasted values and the current cases, while ** indicates the percentage error between the current value and one of the limits (lower or higher).

Figura 18 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Região Norte. Fonte: (LIMA et al., 2020)

State	Date	Forecasted value	Lower limit	Upper limit	Actual cases	Absolute percentage error*	Absolute percentage error**	MAPE
Distrito Federal	2020-05-06	1,910	1,868	1,951	2,046	6.66	4.63	14.81
	2020-05-07	1,982	1,914	2,051	2,268	12.21	9.17	
	2020-05-08	2,055	1,959	2,151	2,442	15.85	11.90	
	2020-05-09	2,128	2,002	2,253	2,576	17.40	12.52	
	2020-05-10	2,200	2,044	2,357	2,682	17.96	12.11	
	2020-05-11	2,273	2,083	2,463	2,799	18.79	12.01	
Goiás	2020-05-06	958	931	986	1,024	6.43	3.76	1.77
	2020-05-07	994	951	1,038	1,027	3.18	—	
	2020-05-08	1,031	971	1,090	1,053	2.13	—	
	2020-05-09	1,067	991	1,143	1,069	0.21	—	
	2020-05-10	1,103	1,010	1,196	1,093	0.91	—	
	2020-05-11	1,139	1,028	1,250	1,100	3.56	—	
Mato Grosso do Sul	2020-05-06	290	280	300	288	0.62	—	8.68
	2020-05-07	297	282	311	311	4.64	—	
	2020-05-08	303	284	323	326	6.95	1.00	
	2020-05-09	310	286	334	346	10.37	3.50	
	2020-05-10	317	289	345	362	12.46	4.69	
	2020-05-11	324	291	356	385	15.93	—	
Mato Grosso	2020-05-06	378	365	392	385	1.74	—	14.84
	2020-05-07	391	370	411	419	6.78	—	
	2020-05-08	403	375	430	464	13.17	7.26	
	2020-05-09	415	381	449	502	17.29	10.55	
	2020-05-10	427	387	468	519	17.63	9.87	
	2020-05-11	440	393	487	545	19.31	10.70	

The * symbol is the error between the forecasted values and the current cases, while ** indicates the percentage error between the current value and one of the limits (lower or higher).

Figura 19 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Região Centro-Oeste. Fonte: (LIMA et al., 2020)

baixo desempenho, com valores fora do intervalo para os 6 dias de predição.

Os resultados obtidos para os estados da região Sudeste são exibidos na tabela da figura 20. Apenas o modelo do estado do Rio de Janeiro não apresentou bom desempenho, apenas 2 dos 6 dias de predição atenderam aos critérios de bom desempenho. Os modelos dos estados do Espírito Santo e Minas Gerais ficaram dentro das projeções para os 6 dias do período de avaliação. Os erros percentuais absolutos dos casos reais e das previsões variaram entre 0,10 e 5,36% e 0,03 e 5,74% para Espírito Santo e Minas Gerais, respectivamente. O modelo para o estado de São Paulo obteve valores fora dos limites inferior e superior do intervalo, com erros variando entre 2,40 a 3,57%.

Os resultados obtidos para os estados da região Sul são apresentados na tabela da figura 21. Os estados dessa região obtiveram bons resultados da soma acumulada de casos de Covid-19. Para os modelos de Santa Catarina e Paraná, os casos reais ficaram dentro do intervalo de predição e muito próximos dos valores estimados, com erros percentuais absolutos variando entre 0,26 e 3,21% e 0,19 e 3,30%, nesta ordem. O modelo para o estado do Rio Grande do Sul os casos reais de 8 e 9 de maio ficaram fora do intervalo de predição. Mas com baixos erros entre o valor real e o limite superior, variando entre 3,17 e 1,66%, para dia 8 e 9 de maio de 2020.

State	Date	Forecasted value	Lower limit	Upper limit	Actual cases	Absolute percentage error*	Absolute percentage error**	MAPE
Espírito Santo	2020-05-06	3,710	3,607	3,814	3,714	0.10	–	3.59
	2020-05-07	3,880	3,709	4,052	3,988	2.70	–	
	2020-05-08	4,050	3,809	4,291	4,242	4.52	–	
	2020-05-09	4,221	3,906	4,535	4,412	4.34	–	
	2020-05-10	4,391	3,998	4,784	4,599	4.53	–	
	2020-05-11	4,561	4,085	5,037	4,819	5.36	–	
Mines Gerais	2020-05-06	2,596	2,547	2,646	2,605	0.34	–	1.93
	2020-05-07	2,760	2,691	2,830	2,770	0.35	–	
	2020-05-08	2,938	2,845	3,031	2,943	0.17	–	
	2020-05-09	3,124	3,000	3,248	3,123	0.03	–	
	2020-05-10	3,315	3,150	3,481	3,237	2.43	–	
	2020-05-11	3,511	3,285	3,727	3,320	5.74	–	
Rio de Janeiro	2020-05-06	12,911	12,697	13,124	13,295	2.89	–	9.30
	2020-05-07	13,506	13,230	13,782	14,156	4.59	2.64	
	2020-05-08	14,101	13,728	14,474	15,741	10.42	8.05	
	2020-05-09	14,696	14,200	15,191	16,929	13.19	10.26	
	2020-05-10	15,291	14,652	15,929	17,062	10.38	6.64	
	2020-05-11	15,885	15,885	15,088	16,683	17.939	7.00	
São Paulo	2020-05-06	35,672	34,841	36,503	37,853	5.76	3.57	8.83
	2020-05-07	37,010	35,483	38,537	39,928	7.31	3.48	
	2020-05-08	38,348	36,214	40,482	41,830	8.32	3.22	
	2020-05-09	39,686	36,958	42,414	44,411	10.64	4.50	
	2020-05-10	41,024	37,694	44,353	45,444	9.73	2.40	
	2020-05-11	42,362	38,416	46,608	46,131	8.17	–	

The * symbol is the error between the forecasted values and the current cases, while ** indicates the percentage error between the current value and one of the limits (lower or higher).

Figura 20 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Sudeste. Fonte: (LIMA et al., 2020)

State	Date	Forecasted cases	Lower limit	Upper limit	Actual cases	Absolute percentage error*	Absolute percentage error**	MAPE
Paraná	2020-05-06	1,643	1,606	1,680	1,647	0.26	–	1.50
	2020-05-07	1,682	1,616	1,748	1,678	0.23	–	
	2020-05-08	1,721	1,629	1,813	1,734	0.75	–	
	2020-05-09	1,760	1,642	1,879	1,809	2.70	–	
	2020-05-10	1,799	1,654	1,944	1,859	3.21	–	
	2020-05-11	1,838	1,666	2,011	1,873	1.85	–	
Santa Catarina	2020-05-06	2,911	2,716	3,107	2,917	0.19	–	2.10
	2020-05-07	3,028	2,735	3,321	3,082	1.75	–	
	2020-05-08	3,144	2,764	3,524	3,205	1.89	–	
	2020-05-09	3,261	2,797	3,724	3,372	3.30	–	
	2020-05-10	3,377	2,831	3,923	3,429	1.51	–	
	2020-05-11	3,494	2,865	4,123	3,529	1.00	–	
Rio Grande do Sul	2020-05-06	2,167	2,109	2,224	2,100	3.17	–	4.94
	2020-05-07	2,233	2,138	2,328	2,182	2.35	–	
	2020-05-08	2,283	2,166	2,401	2,493	8.41	3.71	
	2020-05-09	2,367	2,233	2,500	2,542	6.90	1.66	
	2020-05-10	2,492	2,340	2,645	2,576	3.25	–	
	2020-05-11	2,640	2,458	2,821	2,808	6.00	–	

The * symbol is the error between the forecasted values and the current cases, while ** indicates the percentage error between the current value and one of the limits (lower or higher).

Figura 21 – Resultados das projeções de casos confirmados de Covid-19 entre 6 e 11 de maio de 2020 para os estados da Sul. Fonte: (LIMA et al., 2020)

State	Date	Forecasted value	Lower limit	Upper limit	Actual cases	Absolute percentage error*	Absolute percentage error**	MAPE
Brazil	2020-05-06	123,706	122,328	125,085	126,957	2.56	1.47	5.18
	2020-05-07	131,001	128,154	133,747	136,689	4.16	2.15	
	2020-05-08	138,295	133,946	142,644	147,093	5.98	3.02	
	2020-05-09	145,589	139,424	151,755	155,329	6.27	2.30	
	2020-05-10	152,883	144,709	161,057	163,509	6.50	1.50	
	2020-05-11	160,178	149,819	170,537	169,733	5.63	0.47	

The * symbol is the error between the forecasted values and the current cases, while ** indicates the percentage error between the current value and one of the limits (lower or higher).

Figura 22 – Resultado das projeções dos casos confirmados acumulados de Covid-19 para o Brasil entre 6 a 11 de maio de 2020. Fonte: (LIMA et al., 2020)

Os resultados do modelo para o Brasil são apresentados na tabela da figura 22. Os casos reais ficaram acima do limite superior entre 6 e 10 de maio. Contudo o erro percentual absoluto entre o limite superio e os casos reais foi menor do que 5%. Com um erro de 5,63% entre o valor predito e os casos reais, no dia 11 de maio o valor dos casos reais estavam dentro do intervalo de predição.

5.3 Predição espaço-temporal de Covid-19

A autora também participou do estudo (SILVA et al., 2021b) para distribuição e predição de casos de Covid-19 nos municípios de Pernambuco e do Brasil, e todos os resultados obtidos com esse estudo foram detalhados nesta seção.

Foi usada uma janela de 3 dias para a predição da distribuição espacial do dia seguinte. Logo depois, foi considerado quatro dias de notificação para o próximo passo. Com isso, para obter a localização geográfica dos casos confirmados de cada município, inicialmente foram geradas as camadas vetoriais de pontos (.shp), o shapefile. Em sequência, usando o pacote *sf* do R, com a função *sfwrite()* é gerado com arquivo .shp para cada dia de predição. E então foram gerados mapas de distribuição geográfica dos dados de casos acumulados para se estimar como se dava a distribuição da doença no território de estudo.

Os modelos foram avaliados usando o coeficiente de correlação de e o erro quadrático relativo (RMSE percentual). Sendo este último uma métrica de qualidade local e o primeiro, métrica de qualidade global. Os resultados da avaliação das métricas foram colocados nas tabelas.

Os resultados dos experimentos no território brasileiro estão na tabela da figura 23. Para cada regressor, a média e o desvio padrão de cada métrica são mostrados na tabela 1. O regressor regressão linear obteve 0,9826 e 0,143 como média e desvio padrão do coeficiente de correlação, e este foi o método com o valor mais alto para esta métrica. Contudo, o erro RMSE(%) foi de 11,42% e 22,94% para média e desvio padrão, respectivamente. No método Random Forest, a média e o desvio padrão do coeficiente de correlação variaram de 0,9737 a 0,9759, e 0,072 a 0,068, nesta ordem. Sendo a

Regression method	Configuration	Correlation coefficient		RMSE (%)		Training time (s)	
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Linear regression	-	0.9826	0.143	11.42	22.94	0.003	0.006
	10 trees	0.9737	0.072	22.35	24.08	0.06	0.01
	20 trees	0.9752	0.070	21.59	23.03	0.20	0.04
	30 trees	0.9756	0.069	21.32	22.68	0.29	0.06
Random forest	40 trees	0.9759	0.068	21.17	22.33	0.22	0.04
	20 neurons	0.9948	0.006	11.29	7.12	13.91	0.19
MLP	30 neurons	0.9867	0.011	17.9	4.12	6.48	0.30
	Polynomial kernel, $p = 1$	0.9670	0.109	18.39	24.08	3.82	2.14
	Polynomial kernel, $p = 2$	0.9818	0.006	31.70	10.62	7.99	10.57
	Polynomial kernel, $p = 3$	0.9623	0.004	45.15	17.35	7.55	9.45
SVR	RBF kernel	0.5341	0.375	87.19	14.94	39.97	56.17

For each regressor, we calculated the correlation coefficient, the relative square error [RMSE (%)], and the training time.

Figura 23 – Resultados do desempenho dos algoritmos de regressão linear, do perceptron multicamadas (MLP) e do regressor vetorial de suporte (SVR) para o conjunto de dados do Brasil. Fonte: (SILVA et al., 2021b)

Regression method	Configuration	Correlation coefficient		RMSE (%)		Training time(s)	
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Linear regression	-	0.9991	0.006	1.92	3.81	0.02	0.01
	10 trees	0.9983	0.003	6.14	4.14	0.27	0.05
	20 trees	0.9985	0.003	5.78	4.03	0.52	0.09
	30 trees	0.9985	0.003	5.65	4.01	0.78	0.11
Random forest	40 trees	0.9985	0.003	5.57	3.99	1.21	0.14
	20 neurons	0.9990	0.005	4.07	4.95	45.82	1.56
MLP	30 neurons	0.9991	0.005	3.81	4.95	62.69	3.60
	Polynomial kernel, $p = 1$	0.9988	0.006	3.28	4.05	18.61	22.15
	Polynomial kernel, $p = 2$	0.9984	0.011	6.48	5.85	82.79	34.22
	Polynomial kernel, $p = 3$	0.9989	0.001	4.48	1.99	89.64	8.34
SVR	RBF kernel	0.9387	0.064	76.86	13.82	112.78	122.81

For each regressor, we calculated the correlation coefficient, the relative square error [RMSE (%)], and the training time.

Figura 24 – Resultados do desempenho dos algoritmos de regressão linear, do perceptron multicamadas (MLP) e do regressor vetorial de suporte (SVR) para o conjunto de dados de Pernambuco. fonte: (SILVA et al., 2021b)

configuração com 40 árvores a que obteve o menor RMSE(%), 21,17% de média, e as outras configurações não obtiveram grandes variações no erro. O regressor perceptron multicamadas obteve ótimo desempenho de coeficiente de correlação, atingindo valores maiores que 0,9. E os erros médios para as configuração de 20 e 30 neurônios foram de 11,29% e 17,90%, na devida ordem. Não foram obtidos bons desempenhos pelos modelos criados usando máquina de vetor de suporte. A configuração de kernel polinomial de grau 1 obteve melhor desempenho e o kernel RBF apresentou o pior, sendo os valores da média do coeficiente de correlação e RMSE(%) foram de 0,9670 e 18,39% e 0,5341 e 87%, respectivamente.

Na tabela 2 são mostrados os resultados dos desempenhos para os mapas de distribuição dos casos acumulados no estado de Pernambuco. O resultado do método regressão linear foi bom, com coeficiente de correlação maior que 0,9 e RMSE(%) de 1,92%. O método Random Forest obteve alto valor de coeficiente de correlação, sendo maior que 0,9 em todas as configurações. E o RMSE variaram de 5,57 a 6,14%, sendo o

	Regression method	Prediction date	Correlation coefficient	RMSE (%)
PERNAMBUCO	Linear regression	25/05/2020	0.999994	1.32
		26/05/2020	0.999990	3.27
		27/05/2020	0.999970	3.84
BRAZIL	Linear regression	25/05/2020	0.998447	5.68
		26/05/2020	0.998840	5.03
		27/05/2020	0.978811	20.86
PERNAMBUCO	SVR, kernel = RBF	25/05/2020	0.949280	86.55
		26/05/2020	0.869430	99.33
		27/05/2020	0.511389	101.2
BRAZIL	SVR, kernel = RBF	25/05/2020	0.993423	42.15
		26/05/2020	0.951116	71.23
		27/05/2020	0.751036	87.91

We forecasted the distribution of accumulated cases of Covid-19 between May 25 and 27, 2020.

Figura 25 – Resultados da validação dos modelos criados por meio de regressão linear e SVR, kernel RBF, para Brasil e Pernambuco. Fonte: (SILVA et al., 2021b)

maior valor com configuração de 10 árvores e o menor com configuração de 40 árvores. As configurações do perceptron multicamadas apresentaram ótimos resultados, sendo a média do coeficiente de correlação de 0,9990 e 0,9991 para 20 e 30 neurônios. E os resultados do RMSE foram ótimos em torno de 4%.

Foram geradas três distribuições de grade regular a partir de distribuições diárias irregulares por meio da interpolação IDW para cada predição espaço-temporal. É preciso pontuar que o Brasil é um país de dimensão continental, assim diversidade de biomassa natural, os estados do Amazonas e Pará são compostos por municípios muito grandes, dentre eles vários com densidade reduzida. O que gera uma região relativamente espaçada na malha irregular por os centróides dos municípios tenderem a ficar distantes um dos outros. Assim, o método de interpolação IDW tende a gerar estimativas com altos erros locais, e para todo o Brasil, são esperados RMSE(%) maiores do que o esperado, e o comportamento se mostra diferente para o estado de Pernambuco mesmo usando a mesma métrica.

Já o método máquina de vetor de suporte, obteve o melhor desempenho com a configuração de kernel polinomial de grau 1 e a pior com kernel RBF. Sendo a média do coeficiente de correlação e RMSE(%) de 0,9988 e 3,28 e nesta ordem, para kernel polinomial de grau 1. E o RMSE(%) foi de aproximadamente 77% apesar da média do coeficiente de correlação ter sido alto.

5.4 Aplicação web

Agora será apresentado um pouco da aplicação web que foi desenvolvida. Na figura 27(A) pode-se ver a tela inicial do aplicativo, onde o usuário pode visualizar o monitoramento de Covid-19. Esta seção disponibiliza o número de casos acumulados e diários de Covid-19 para cada uma das 27 unidades federativas do Brasil. Na opção "Alterar mapa" do canto

inferior direito o usuário pode selecionar os mapas de predição espaço-temporal (figura 27(B)). E na figura 27(C) é possível visualizar mapas que apresentam a densidade (alta, média e baixa) dos casos acumulados de Covid-19 para cada região do Brasil.

A opção "Mais informações" no canto inferior direito da tela inicial o usuário é direcionado para tela de gráficos. E são apresentados os gráficos da predição temporal dos casos acumulados e de óbitos para cada um dos 27 estados (27 (D)). As figuras 27 (E) e (F) mostram a seleção de gráficos de casos acumulados, casos diários, óbitos acumulados e diários, podem ser selecionados os estados ou cidades brasileiras. Por fim, nas telas 27 (G, H) os usuários também podem selecionar painéis personalizados para cada uma das 27 unidades federativas do Brasil para visualizar informações sobre Covid-19 a nível municipal.

No repositório Github:⁴ está disponível gratuitamente o backend do software para fins não comerciais.

⁴ <https://github.com/Biomedical-Computing-UFPE/Covid-SGIS>.

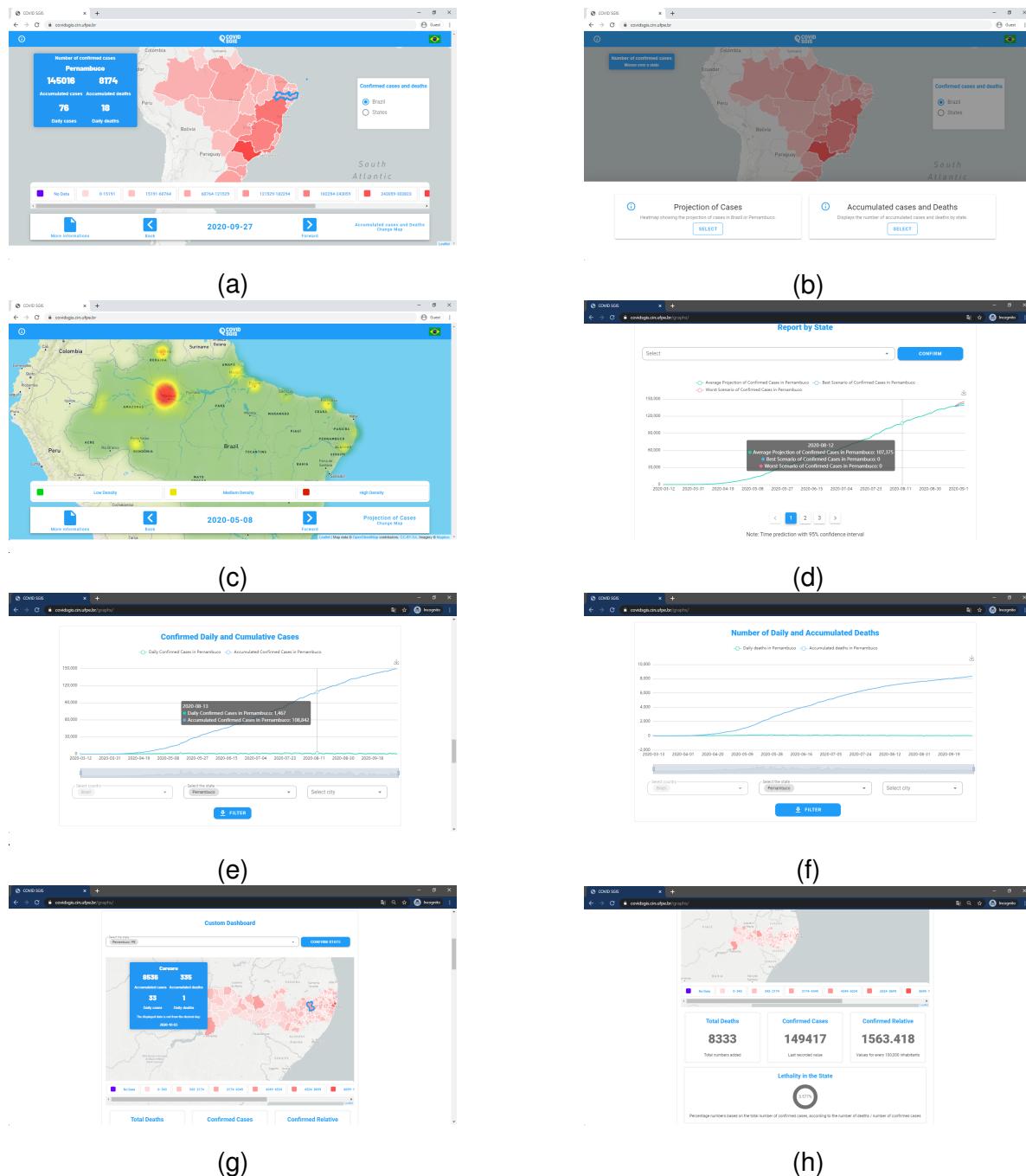


Figura 26 – (A) Tela inicial do COVID-SGIS. (B) Tela de seleção do mapa, onde é possível visualizar o mapa dos casos cumulativos e os mapas de distribuição previstos. (C) Tela de previsão dos mapas de distribuição dos casos cumulativos de Covid-19. (D) Casos acumulados do gráfico de previsão Covid-19 com ARIMA. A linha verde representa a previsão ARIMA, enquanto as linhas vermelha e azul representam o pior e o melhor cenário. Tela dos gráficos de casos confirmados acumulados e diários (E) e óbitos acumulados e confirmados (F). Telas do painel personalizado (G, H). O painel personalizado fornece informações detalhadas - sobre os casos confirmados cumulativos e casos diários, e mortes cumulativas e mortes diárias - para cada estado em nível de município). Fonte: (LIMA et al., 2020)

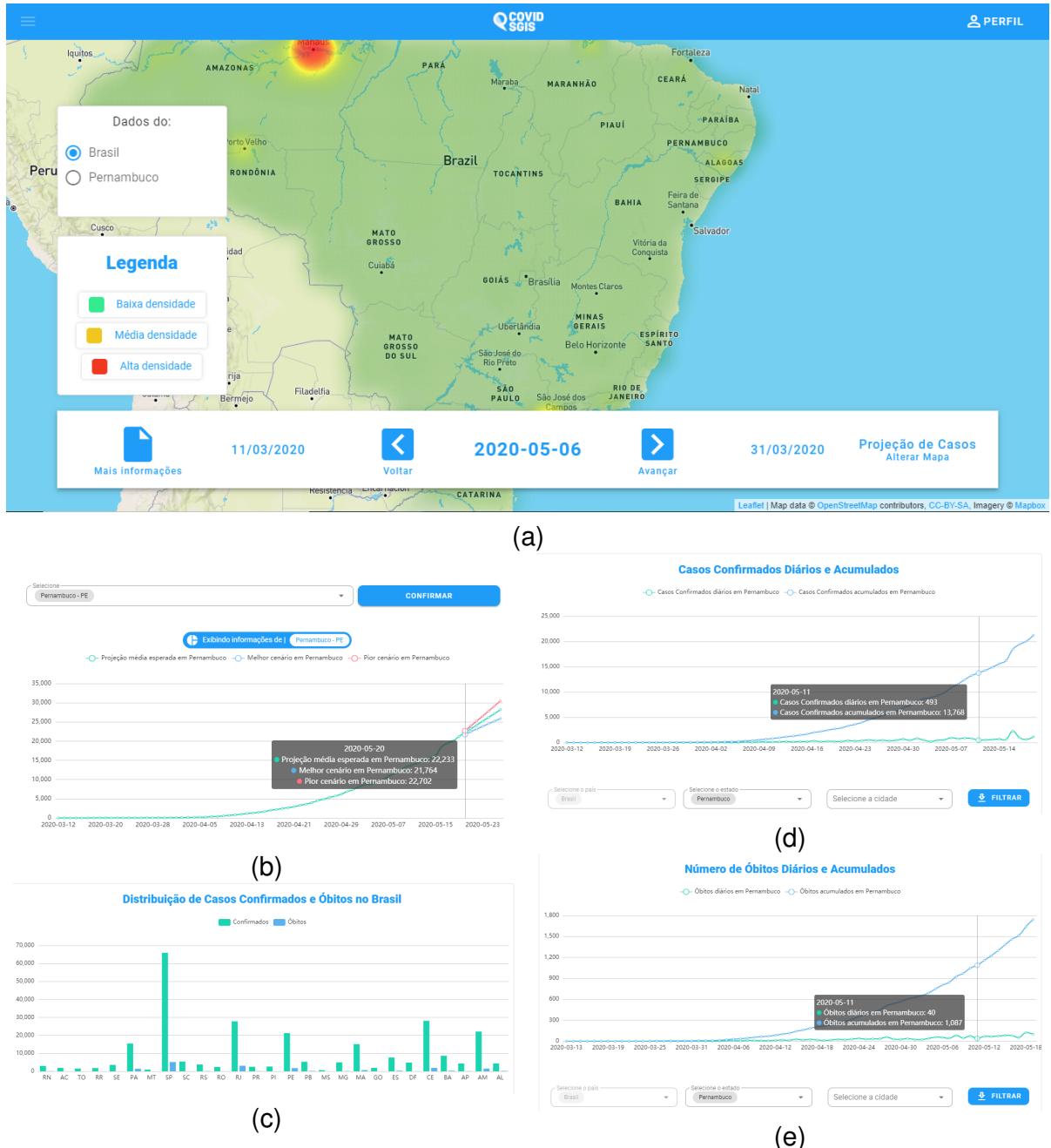


Figura 27 – (A) Tela inicial do aplicativo da web COVID SGIS. (B) Casos acumulados do gráfico de previsão Covid-19. A previsão com ARIMA é representada pela linha verde. O pior cenário (indicado pelo limite superior da previsão) é representado pela linha em vermelho. O melhor cenário (indicado pelo limite inferior da previsão) é representado pela linha azul. (C) Tela do gráfico de distribuição de casos confirmados e óbitos por Covid-19. Neste gráfico o usuário pode ter uma visão geral dos casos confirmados e óbitos acumulados em todos os estados do Brasil e no Distrito Federal. No COVID SGIS, o usuário pode acompanhar os casos confirmados diários e acumulados (D) e óbitos (E) da Covid-19 para cada estado brasileiro e Distrito Federal, separadamente. Fonte: (LIMA et al., 2020)

6 Conclusão

Neste capítulo serão apresentadas as dificuldades apresentadas durante o estudo, as conclusões gerais e as contribuições científicas da autora.

6.1 Dificuldades apresentadas

No desenvolvimento do estudo (LIMA et al., 2020) houve algumas limitações, dentre elas a subnotificação de casos. A grande demanda populacional somada à baixa disponibilidade de exames, além de pessoas assintomáticas ou com sintomas da doença, e não testadas, não são contabilizadas. Tornando os dados substancialmente subestimados. E isso pode ter impactado no baixo desempenho para algumas unidades federativas brasileiras. E como os modelos ARIMA levam em consideração séries temporais univariadas, o modelo é sensível à quantidade.

O tempo de atualização do número de casos das bases de dados das Secretarias de Saúde também pode estar associado a maiores erros das previsões. O que acaba afetando a série histórica de dados, corroborando para erros na previsão. E além disso, temos que o dia do primeiro caso notificado muda para cada estado brasileiro.

Em específico, quanto ao Covid-19 estudo também não levou em consideração a distribuição populacional e a influência de políticas públicas de isolamento e distanciamento social. Informações climáticas e geográficas das regiões de estudo não foram usadas. E a autora tem ciência que isso pode ter impactado os elevados erros nos estados do Acre, Tocantins e Rondônia.

O estudo (SILVA et al., 2021b) não levou em consideração as variáveis climáticas na geração dos modelos, bem como os fatores sociodemográficos que podem influenciar na dinâmica e propagação da doença.

Com relação a previsão de criadouros, é preciso levar em consideração alguns pontos. Como o fato de termos estimado o número de criadouros por bairros no mesmo estrato de um distrito sanitário. Os estratos sanitários são áreas com mesmas características que são definidas por uma equipe de epidemiologistas da Cidade do Recife, e essa definição varia e esses estratos podem sofrer mudanças e agrupar diferentes bairros de um mesmo distrito sanitário. Isso impacta diretamente nos resultados, como por exemplo no bairro de Monteiro. Esse bairro apresentou uma grande concentração de criadouros, que é uma ressalva ao comportamento dos bairros que possuem renda média e alta. Além disso, no Recife existem apenas três estações meteorológicas, portanto foi preciso criar um estimador

baseado na distribuição gaussiana para poder estimar a distribuição da temperatura e da velocidade do vento para cada bairro. O mesmo acontece com os dados do INMET, que são coletados de uma estação e também é feita uma distribuição gaussiana para construção do shapefile para cada bairro. Assim, pode haver um erro associado por essas informações não são de cada bairro, contudo não foi possível medir tal detalhe.

6.2 Conclusões gerais

Para modelagem de Covid-19, os sistemas criados são robustos, e com as abordagens temporal e espaço temporal é possível se ter uma avaliação ampla das regiões onde os casos acumulados se concentram. Assim os dois sistemas são bastante relevantes para apoiar a tomada de decisões pelos gestores de saúde e aos epidemiologistas no planejamento de ações para controle e prevenção da doença.

Para modelagem de criadouros, o sistema criado também foi robusto e a abordagem espaço-temporal é bastante importante e útil, já que assim é possível classificar as regiões que precisam de maior atenção. Assim, esse sistema se torna vantajoso para controle de surtos de arboviroses e criação de políticas públicas de promoção de saúde e qualidade de vida.

6.3 Contribuições

Este trabalho resultou nas seguintes publicações:

1. Artigo "COVID-SGIS: A Smart Tool for Dynamic Monitoring and Temporal Forecasting of Covid-19", publicado na revista *Frontiers in Public Health*, sobre modelagem temporal de Covid-19 (LIMA et al., 2020);
2. Artigo "Covid-19 Dynamic Monitoring and Real-Time Spatio-Temporal Forecasting", publicado na revista *Frontiers in Public Health*, sobre modelagem espaço-temporal de Covid-19 (SILVA et al., 2021b);
3. Artigo "A review exploring the overarching burden of Zika virus with emphasis on epidemiological case studies from Brazil", publicada na revista *Environmental Science and Pollution Research*, uma revisão da literatura que explora os principais fatores para transmissão da zika com ênfase nos riscos ambientais, antrópicos e sociais (TUNALI et al., 2021);
4. Artigo "Forecasting Dengue, Chikungunya and Zika cases in Recife, Brazil: a spatio-temporal approach based on climate conditions, health notifications and machine learning", publicado na revista *Research, Society and Development*, que é uma predição da distribuição de arboviroses na Cidade do Recife (SILVA et al., 2021a);

5. Capítulo de livro com título "Machine learning approaches for temporal and spatio-temporal Covid-19 forecasting: a brief review and a contribution", publicado no livro "Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis" da editora Springer (SILVA et al., 2022; PANI et al., 2021);
6. Capítulo de livro com título "Intelligent systems for dengue, chikungunya and zika temporal and spatio-temporal forecasting: a contribution and a brief review", publicado no livro "Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis" da editora Springer (LIMA et al., 2022; PANI et al., 2021).
7. Artigo "Spatiotemporal forecasting for dengue, chikungunya fever and zika using machine learning and artificial expert committees based on meta-heuristics", publicado na revista Research on Biomedical Engineering (SILVA et al., 2022);
8. Artigo "Prediction of Aedes aegypti breeding distribution through spatio-temporal analysis and machine learning: A case study in Recife Pernambuco", submetido à revista Research on Biomedical Engineering;
9. Artigo "On arboviruses forecasting machine learning models - A review", submetido à revista Frontiers in Public Health;
10. Trabalho completo "COVID-SGIS: Uma ferramenta inteligente para monitoramento dinâmico previsão temporal da Covid-19", publicado nos Anais do IV SABIO: Simpósio de Inovação em Engenharia Biomédica (SILVA et al., 2020);
11. Trabalho completo "Monitoramento dinâmico e predição espaço-temporal da Covid-19 usando aprendizagem de máquina", publicado nos Anais do IV SABIO: Simpósio de Inovação em Engenharia Biomédica (LIMA et al., 2020);
12. Resumo expandido "Modelagem e simulação da transmissão de arboviroses baseada em modelos bayesianos para predição de casos de dengue, chikungunya e zika", publicado nos anais do IV SABIO: Simpósio de Inovação em Engenharia Biomédica (SILVA; SANTOS, 2020).

Referências

- ABIDEMI, A.; AZIZ, M. A.; AHMAD, R. Vaccination and vector control effect on dengue virus transmission dynamics: Modelling and simulation. *Chaos, Solitons & Fractals*, Elsevier, p. 109648, 2020.
- AHMAR, A. S.; VAL, E. B. D. SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. *Science of The Total Environment*, Elsevier, v. 729, p. 138883, 2020.
- ALMESHAL, A. M.; ALMAZROUEE, A. I.; ALENIZI, M. R.; ALHAJERI, S. N. Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 10, p. 3402, 2020.
- ALZahrani, S. I.; Aljamaan, I. A.; Al-fakih, E. A. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of Infection and Public Health*, Elsevier, v. 13, n. 7, p. 914–919, 2020.
- APOSTOLOPOULOS, I.; AZNAOURIDIS, S.; TZANI, M. Extracting possibly representative covid-19 biomarkers from x-ray images with deep learning approach and image data related to pulmonary diseases. *arXiv preprint arXiv:2004.00338*, 2020.
- APOSTOLOPOULOS, I. D.; MPESIANA, T. A. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, Nature Publishing Group, p. 1, 2020.
- AYYOUNBZADEH, S. M.; AYYOUNBZADEH, S. M.; ZAHEDI, H.; AHMADI, M.; KALHORI, S. R. N. Predicting COVID-19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study. *JMIR Public Health and Surveillance*, JMIR, v. 6, n. 2, p. e18828, 2020.
- AZEVEDO, R. d. S. d. S.; OLIVEIRA, C. S.; VASCONCELOS, P. F. d. C. Risco do chikungunya para o brasil. *Revista de Saúde Pública*, SciELO Public Health, p. 58, 2015.
- 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.I.], 2015. p. 1–8.
- BARBOSA, V. A. d. F.; GOMES, J. C.; SANTANA, M. A. de; JENIFFER, E. d. A.; SOUZA, R. G. de; SOUZA, R. E. de; SANTOS, W. P. dos. Heg.IA: An intelligent system to support diagnosis of Covid-19 based on blood tests. *Research on Biomedical Engineering*, Springer-Nature, p. 1–18, 2021.
- BARCELLOS, C.; BASTOS, F. I. Geoprocessamento, ambiente e saúde: uma união possível? *Cadernos de Saúde Pública*, SciELO Brasil, p. 389–397, 1996.

- BASTOS, S. B.; CAJUEIRO, D. O. Modeling and forecasting the early evolution of the covid-19 pandemic in brazil. *Scientific Reports*, Nature Publishing Group, v. 10, n. 1, p. 1–10, 2020.
- BASU, S.; MITRA, S.; SAHA, N. Deep learning for screening covid-19 using chest x-ray images. In: IEEE. *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. [S.I.], 2020. p. 2521–2527.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 35, n. 8, p. 1798–1828, 2013.
- BERNOULLI, D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem, Des Math. And Phis., Mem*, p. 1–45, 1760.
- BHATT, S.; GETHING, P. W.; BRADY, O. J.; MESSINA, J. P.; FARLOW, A. W.; MOYES, C. L.; DRAKE, J. M.; BROWNSTEIN, J. S.; HOEN, A. G.; SANKOH, O. et al. The global distribution and burden of dengue. *Nature*, Nature Publishing Group, v. 496, n. 7446, p. 504–507, 2013.
- BONYAH, E.; KHAN, M. A.; OKOSUN, K.; ISLAM, S. A theoretical model for zika virus transmission. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 12, n. 10, p. e0185540, 2017.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*. [S.I.: s.n.], 1992. p. 144–152.
- CAMILO, M. J.; MOURA, D. F. C.; SALLES, R. M. Modelagem matemática para epidemia de covid-19 no brasil. *Revista Militar de Ciência e Tecnologia*, v. 37, n. 2, 2020.
- 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.
- CAPASSO, V.; SERIO, G. A generalization of the Kermack-McKendrick deterministic epidemic model. *Mathematical Biosciences*, Elsevier, v. 42, n. 1-2, p. 43–61, 1978.
- CHAKRABORTY, T.; GOSH, I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solition & Fractals*, Elsevier, 2020. Disponível em: <<https://doi.org/10.1016/j.chaos.2020.109850>>.
- CHAN, T.-C.; HU, T.-H.; HWANG, J.-S. Daily forecast of dengue fever incidents for urban villages in a city. *International Journal of Health Geographics*, v. 14, 2015.
- CHEN, D.-G.; CHEN, X.; CHEN, J. K. Reconstructing and forecasting the COVID-19 epidemic in the United States using a 5-parameter logistic growth model. *Global Health Research and Policy*, Springer, v. 5, p. 1–7, 2020.
- CHIMMULA, V. K. R.; ZHANG, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, Elsevier, v. 135, p. 109864, 2020.

- CIVIT-MASOT, J.; LUNA-PEREJÓN, F.; MORALES, M. D.; CIVIT, A. Deep learning system for covid-19 diagnosis aid using x-ray pulmonary images. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 13, p. 4640, 2020.
- COGGON, D.; BARKER, D.; ROSE, G. *Epidemiology for the Uninitiated*. [S.I.]: John Wiley & Sons, 2009.
- 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.I.], 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.
- DAS, T.; JAFFAR-BANDJEE, M. C.; HOARAU, J. J.; TROTOT, P. K.; DENIZOT, M.; LEE-PAT-YUEN, G.; SAHOO, R.; GUIRAUD, P.; RAMFUL, D.; ROBIN, S. et al. Chikungunya fever: Cns infection and pathologies of a re-emerging arbovirus. *Progress in Neurobiology*, Elsevier, n. 2, p. 121–129, 2010.
- DAY, M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ: British Medical Journal (Online)*, BMJ Publishing Group LTD, v. 368, 2020.
- DODERO-ROJAS, E.; FERREIRA, L. G.; LEITE, V. B.; ONUCHIC, J. N.; CONTESSOTO, V. G. Modeling chikungunya control strategies and mayaro potential outbreak in the city of rio de janeiro. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 15, n. 1, p. e0222900, 2020.
- 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. *EnvironmentAsia*, n. 2, 2017.
- DONG, E.; DU, H.; GARDNER, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, Elsevier Ltd, v. 20, n. 5, p. 533–534, 2020. ISSN 14744457. Disponível em: <[http://dx.doi.org/10.1016/S1473-3099\(20\)30120-1](http://dx.doi.org/10.1016/S1473-3099(20)30120-1)>.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification by Richard O. Duda, David G. Stork, Peter E.Hart .pdf*. [S.I.: s.n.], 2000. 738 p.
- DUMMER, T. J. Health geography: supporting public health policy and planning. *Cmaj*, Can Med Assoc, n. 9, p. 1177–1180, 2008.
- ECKHOFF, P. A.; TATEM, A. J. *Digital methods in epidemiology can transform disease control*. [S.I.]: Oxford University Press, 2015.

- FENG, C.; HUANG, Z.; WANG, L.; CHEN, X.; ZHAI, Y.; ZHU, F.; CHEN, H.; WANG, Y.; SU, X.; HUANG, S.; TIAN, L.; ZHU, W.; SUN, W.; ZHANG, L.; HAN, Q.; ZHANG, J.; PAN, F.; CHEN, L.; ZHU, Z.; XIAO, H.; LIU, Y.; LIU, G.; CHEN, W.; LI, T. A Novel Triage Tool of Artificial Intelligence Assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics. *medRxiv*, p. 1–68, 2020. Disponível em: <<https://ssrn.com/abstract=3551355>>.
- FIGUEIREDO, L. T. M. Febres hemorrágicas por vírus no brasil. *Revista da Sociedade Brasileira de Medicina Tropical*, SciELO Brasil, v. 39, p. 203–210, 2006.
- FIGUEIREDO, R.; PAIVA, C.; MORATO, M. et al. Arboviroses. 2017.
- FONG, S. J.; LI, G.; DEY, N.; CRESPO, R. G.; HERRERA-VIEDMA, E. Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. *arXiv preprint arXiv:2003.10776*, v. 2020, 2020.
- 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.
- GÉRARDIN, P.; BARAU, G.; MICHAULT, A.; BINTNER, M.; RANDRIANAIVO, H.; CHOKER, G.; LENGLLET, Y.; TOURET, Y.; BOUVERET, A.; GRIVARD, P. et al. Multidisciplinary prospective study of mother-to-child chikungunya virus infections on the island of la reunion. *PLoS medicine*, Public Library of Science San Francisco, USA, n. 3, p. e60, 2008.
- GOMES, J. C.; BARBOSA, V. A. de F.; SANTANA, M. A. de; BANDEIRA, J.; VALENCA, M. J. S.; SOUZA, R. E. de; ISMAEL, A. M.; SANTOS, W. P. dos. IKONOS: An intelligent tool to support diagnosis of Covid-19 by texture analysis of x-ray images. *Research on Biomedical Engineering*, Springer-Nature, v. 2020, p. 1–14, 2020.
- GOMES, L. L. F.; THOMÉ, R. C. Modelagem matemática do aedes aegypti com a utilização de mecanismos de combate à dengue. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 3, n. 1, 2015.
- GUAN, W.-j.; NI, Z.-y.; HU, Y.; LIANG, W.-h.; OU, C.-q.; HE, J.-x.; LIU, L.; SHAN, H.; LEI, C.-l.; HUI, D. S. et al. Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine*, Mass Medical Soc, n. 18, p. 1708–1720, 2020.
- 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.
- GULLAND, A. *Zika virus is a global public health emergency, declares WHO*. [S.I.]: British Medical Journal Publishing Group, 2016.
- HAYKIN, S. *Redes neurais: princípios e prática*. Porto Alegre, RS, Brasil: Bookman, 2001.
- HAYKIN, S. *Redes neurais: princípios e prática*. [S.I.]: Bookman Editora, 2007.
- HERBY, J.; JONUNG, L.; HANKE, S. et al. A literature review and meta-analysis of the effects of lockdowns on covid-19 mortality. *Studies in Applied Economics*, The Johns Hopkins Institute for Applied Economics, Global Health, and the . . . , n. 200, 2022.

- HETHCOTE, H. W. The mathematics of infectious diseases. *SIAM review*, SIAM, n. 4, p. 599–653, 2000.
- 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.
- ISMAEL, A. M.; ŞENGÜR, A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications*, Elsevier, v. 164, p. 114054, 2021.
- JAIN, G.; MITTAL, D.; THAKUR, D.; MITTAL, M. K. A deep learning approach to detect covid-19 coronavirus with x-ray images. *Biocybernetics and Biomedical Engineering*, Elsevier, v. 40, n. 4, p. 1391–1405, 2020.
- JI, D.; ZHANG, D.; XU, J.; CHEN, Z.; YANG, T.; ZHAO, P.; CHEN, G.; CHENG, G.; WANG, Y.; BI, J.; TAN, L.; LAU, G.; QIN, E. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clinical Infectious Diseases*, Oxford University Press US, v. 71, n. 6, p. 1393–1399, 2020.
- JIANG X E ZHAO, B. Modeling on the epidemic of coronavirus disease 2019. *J Bio Med Open Access*, v. 1, n. 1, p. 103, 2020.
- JIN, C.; CHEN, W.; CAO, Y.; XU, Z.; ZHANG, X.; DENG, L.; ZHENG, C.; ZHOU, J.; SHI, H.; FENG, J. Development and Evaluation of an AI System for COVID-19 Diagnosis. *medRxiv*, 2020. Disponível em: <<http://medrxiv.org/content/early/2020/03/27/2020.03.20.20039834.abstract>>.
- JINDAL, A.; RAO, S. Agent-based modeling and simulation of mosquito-borne disease transmission. In: *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. [S.I.: s.n.], 2017. p. 426–435.
- JING, Q. L.; CHENG, Q.; MARSHALL, J. M.; HU, W. B.; YANG, Z. C.; LU, J. H. Imported cases and minimum temperature drive dengue transmission in guangzhou, china: evidence from arimax model. *Epidemiology and Infection*, PubMed, v. 146, p. 1226–1235, 2018.
- 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. et al. Vírus zika: revisão para clínicos. Ordem dos Médicos, 2015.
- KAO, Y.-H.; EISENBERG, M. C. Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment. *Epidemics*, Elsevier, v. 25, p. 89–100, 2018.
- KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, The Royal Society London, v. 115, n. 772, p. 700–721, 1927.

- KESORN, K.; ONGRUK, P.; CHOMPOOSRI, J.; PHUMEE, A.; THAVARA, U.; TAWATSIN, A.; SIRIYASATIEN, P. Morbidity rate prediction of dengue hemorrhagic fever (DHF) using the support vector machine and the Aedes aegypti infection rate in similar climates and geographical areas. *PloS One*, Public Library of Science, v. 10, n. 5, p. e0125049, 2015.
- KHAJANCHI, S.; SARKAR, K. Forecasting the daily and cumulative number of cases for the covid-19 pandemic in india. *Chaos: An interdisciplinary journal of nonlinear science*, AIP Publishing LLC, v. 30, n. 7, p. 071101, 2020.
- KHALIL, O. A. K.; KHALIL, S. da S. SARS-CoV-2: taxonomia, origem e constituição. *Revista de Medicina*, n. 5, p. 473–479, 2020.
- KHAN, A. I.; SHAH, J. L.; BHAT, M. M. Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 196, p. 105581, 2020.
- KUCHARSKI, A. *As regras do contágio: Por que as coisas se disseminam—e por que param de se propagar*. Kucharski, Adam. [S.I.]: Editora Record, 2020.
- KUFEL, T. ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, Instytut Badań Gospodarczych, v. 15, n. 2, p. 181–204, 2020.
- KUMAR, N.; ABDULLAH, M.; FAIZAN, M. I.; AHMED, A.; ALSENAIDY, H. A.; DOHARE, R.; PARVEEN, S. Progression dynamics of zika fever outbreak in el salvador during 2015–2016: a mathematical modeling approach. *Future Virology*, Future Medicine, v. 12, n. 5, p. 271–281, 2017.
- LATORRE, M. d. R. D. d. O.; CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 4, p. 145–152, 2001.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature, v. 521, n. 7553, p. 436–444, 2015.
- LEGROS, M.; OTERO, M.; AZNAR, V. R.; SOLARI, H.; GOULD, F.; LLOYD, A. L. Comparison of two detailed models of aedes aegypti population dynamics. *Ecosphere*, Wiley Online Library, v. 7, n. 10, p. e01515, 2016.
- LI, Q.; FENG, W.; QUAN, Y.-H. Trend and forecasting of the COVID-19 outbreak in China. *Journal of Infection*, Elsevier, v. 80, n. 4, p. 469–496, 2020.
- LIMA, C. L. de; SILVA, A. C. G. da; SILVA, C. C. da; SILVA, E. L.; MARQUES, G. S.; ARAÚJO, L. J. B. de; JÚNIOR, L. A. A.; SOUZA, S. B. J. de; SANTANA, M. A. de; GOMES, J. C. et al. Monitoramento dinâmico e predição espaço-temporal da covid-19 usando aprendizagem de máquina. In: *Anais do IV SABIO: Simpósio de Inovação em Engenharia Biomédica*. [S.I.: s.n.], 2020.
- LIMA, C. L. de; SILVA, A. C. G. da; SILVA, C. C. da; MORENO, G. M. M.; FILHO, A. G. da S.; MUSAH, A.; ALDOSERY, A.; DUTRA, L.; AMBRIZZI, T.; BORGES, I. V. G. et al. Intelligent systems for dengue, chikungunya, and zika temporal and spatio-temporal forecasting: A contribution and a brief review. In: *Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis*. [S.I.]: Springer, 2022. p. 299–331.

- LIMA, C. L. de; SILVA, C. C. da; SILVA, A. C. G. da; SILVA, E. L.; MARQUES, G. S.; ARAÚJO, L. J. B. de; JÚNIOR, L. A. A.; SOUZA, S. B. J. de; SANTANA, M. A. de; GOMES, J. C. et al. COVID-SGIS: A smart tool for dynamic monitoring and temporal forecasting of Covid-19. *Frontiers in Public Health*, Frontiers, v. 8, p. 761, 2020.
- 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.
- LUMSDEN, W. H. An epidemic of virus disease in southern province, tanganyika territory, in 1952–1953 ii. general description and epidemiology. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Royal Society of Tropical Medicine and Hygiene, n. 1, p. 33–57, 1955.
- LUZ, E.; SILVA, P. L.; SILVA, R.; MOREIRA, G. Towards an efficient deep learning model for covid-19 patterns detection in x-ray images. *arXiv preprint arXiv:2004.05717*, 2020.
- MAGHDID, H. S.; ASAAD, A. T.; GHAFOOR, K. Z.; SADIQ, A. S.; KHAN, M. K. Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms. *arXiv preprint arXiv:2004.00038*, 2020.
- MALAGON, O. L. M. Fatores de risco que incidem sobre a dengue na população da ubs marinheiro marcilio dias, no município de iperó/sp.
- MALEKI, M.; MAHMOUDI, M. R.; WRAITH, D.; PHO, K.-H. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease*, Elsevier, v. 37, p. 101742, 2020.
- MANDAL, M.; JANA, S.; NANDI, S. K.; KHATUA, A.; ADAK, S.; KAR, T. A model based study on the dynamics of covid-19: Prediction and control. *Chaos, Solitons & Fractals*, Elsevier, v. 136, p. 109889, 2020.
- MARTÍNEZ-BELLO, D.; LÓPEZ-QUÍLEZ, A.; PRIETO, A. T. Spatiotemporal modeling of relative risk of dengue disease in colombia. *Stochastic environmental research and risk assessment*, Springer, v. 32, n. 6, p. 1587–1601, 2018.
- MASSONIS, G.; BANGA, J. R.; VILLAVERDE, A. F. Structural identifiability and observability of compartmental models of the covid-19 pandemic. *Annual reviews in control*, Elsevier, 2020.
- MATEUS, L. A matemática e as epidemias. *Revista de Ciência Elementar*, Casa das Ciências, v. 8, n. 3, 2020.
- MBUVHA, R. R.; MARWALA, T. On data-driven management of the Covid-19 outbreak in South Africa. *medRxiv*, Cold Spring Harbor Laboratory Press, v. 2020, 2020.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.

- MELO, A. S. O.; MALINGER, G.; XIMENES, R.; SZEJNFELD, P.; SAMPAIO, S. A.; FILIPPIS, A. Bispo de. Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? *Ultrasound in Obstetrics & Gynecology*, Wiley Online Library, n. 1, p. 6–7, 2016.
- MENG, Z.; WANG, M.; SONG, H.; GUO, S.; ZHOU, Y.; LI, W.; ZHOU, Y.; LI, M.; SONG, X.; ZHOU, Y.; LI, Q.; LU, X.; YING, B. Development and utilization of an intelligent application for aiding COVID-19 diagnosis. *medRxiv*, n. 37, 2020.
- MOFTAKHAR, L.; MOZHGAN, S.; SAFE, M. S. Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iranian Journal of Public Health*, v. 2020, 2020.
- MOHAMMED, A.; CHADEE, D. D. Effects of different temperature regimens on the development of aedes aegypti (I.) (diptera: Culicidae) mosquitoes. *Acta tropica*, Elsevier, v. 119, n. 1, p. 38–43, 2011.
- MUSA, S. S.; ZHAO, S.; CHAN, H.-S.; JIN, Z.; HE, D. et al. A mathematical model to study the 2014–2015 large-scale dengue epidemics in kaohsiung and tainan cities in taiwan, china. *Mathematical Biosciences and Engineering*, American Institute of Mathematical Sciences, 2019.
- MUSSO, D.; GUBLER, D. J. Zika virus. *Clinical Microbiology Reviews*, American Society for Microbiology, v. 29, p. 487–524, 2016.
- NARDI, S. M. T.; PASCHOAL, J. A. A.; PEDRO, H. d. S. P.; PASCHOAL, V. D. A.; SICHERI, E. P. Geoprocessamento em saúde pública: fundamentos e aplicações. *Revista do Instituto Adolfo Lutz*, n. 3, p. 185–191, 2013.
- NATAL, D. Bioecologia do aedes aegypti. *Biológico*, n. 2, p. 205–207, 2002.
- NDAÏROU, F.; AREA, I.; NIETO, J. J.; SILVA, C. J.; TORRES, D. F. Mathematical modeling of zika disease in pregnant women and newborns with microcephaly in brazil. *Mathematical Methods in the Applied Sciences*, Wiley Online Library, v. 41, n. 18, p. 8929–8941, 2018.
- NEPOMUCENO, E. G. Dinâmica, modelagem e controle de epidemias. *UFMG. Tese de Doutorado*. <http://www.cpdee.ufmg.br/defesas/D>, v. 534, 2005.
- OLIVEIRA, W. K. d.; DUARTE, E.; FRANÇA, G. V. A. d.; GARCIA, L. P. How Brazil can hold back COVID-19. *Epidemiologia e Serviços de Saúde*, SciELO Brasil, v. 29, 2020.
- 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*, v. 4, n. 1, p. 148–166, 2017.
- PAIVA, C.; MORATO, M.; FIGUEIREDO, R. et al. Arboviroses. 2018.
- PANI, S. K.; DASH, S.; SANTOS, W. dos; BUKHARI, S. A. C.; FLAMMINI, F. *Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis*. [S.I.]: Springer, 2021.
- 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.

- PENG, L.; YANG, W.; ZHANG, D.; ZHUGE, C.; HONG, L. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv preprint arXiv:2002.06563*, v. 2020, 2020.
- 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.
- PORCASI, X.; ROTELA, C. H.; INTROINI, M. V.; FRUTOS, N.; LANFRI, S.; PERALTA, G.; ELIA, E. A. D.; LANFRI, M. A.; SCAVUZZO, C. M. An operative dengue risk stratification system in argentina based on geospatial technology. *Geospatial health*, n. 3, p. S31–S42, 2012.
- PRADO, A. D.; PEIXOTO, B. C.; SILVA, A. M. B. da; SCALIA, L. A. M. A saúde mental dos profissionais de saúde frente à pandemia do covid-19: uma revisão integrativa. *Revista Eletrônica Acervo Saúde*, n. 46, p. e4128–e4128, 2020.
- PUNG, R.; CHIEW, C. J.; YOUNG, B. E.; CHIN, S.; CHEN, M. I.; CLAPHAM, H. E.; COOK, A. R.; MAURER-STROH, S.; TOH, M. P.; POH, C. et al. Investigation of three clusters of covid-19 in singapore: implications for surveillance and response measures. *The Lancet*, Elsevier, v. 395, n. 10229, p. 1039–1046, 2020.
- PUTRA, S.; MU'TAMAR, Z. K. Estimation of Parameters in the SIR Epidemic Model Using Particle Swarm Optimization. *American Journal of Mathematical and Computer Modelling*, Science Publishing Group, v. 4, n. 4, p. 83–93, 2019.
- QEADAN, F.; HONDA, T.; GREN, L. H.; DAILEY-PROVOST, J.; BENSON, L. S.; VANDERSLICE, J. A.; PORUCZNIK, C. A.; WATERS, A. B.; LACEY, S.; SHOAF, K. Naive forecast for COVID-19 in Utah based on the South Korea and Italy models-the fluctuation between two extremes. *International Journal of Environmental Research and Public Health*, Multidisciplinary Digital Publishing Institute, v. 17, n. 8, p. 2750, 2020.
- QI, H.; XIAO, S.; SHI, R.; WARD, M. P.; CHEN, Y.; TU, W.; SU, Q.; WANG, W.; WANG, X.; ZHANG, Z. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. *Science of the Total Environment*, Elsevier, v. 728, p. 138778, 2020.
- RAHIMI, I.; CHEN, F.; GANDOMI, A. H. A review on COVID-19 forecasting models. *Neural Computing and Applications*, Springer, v. 2020, p. 1–11, 2021.
- RIBEIRO, M. H. D. M.; SILVA, R. G. da; MARIANI, V. C.; COELHO, L. dos S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, Elsevier, v. 135, p. 109853, 2020.
- 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.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

- ROSS, R. The newala epidemic: iii. the virus: isolation, pathogenic properties and relationship to the epidemic. *Epidemiology & Infection*, Cambridge University Press, n. 2, p. 177–191, 1956.
- ROY, S.; BHUNIA, G. S.; SHIT, P. K. Spatial prediction of covid-19 epidemic using arima techniques in india. *Modeling Earth Systems and Environment*, Springer, v. 2020, p. 1–7, 2020.
- SALATHÈ, M. Digital epidemiology: what is it, and where is it going? *Life Sci Policy*, v. 14, n. 1, 2018.
- SALATHÉ, M.; ALTHAUS, C. L.; NEHER, R.; STRINGHINI, S.; HODCROFT, E.; FELLAY, J.; ZWAHLEN, M.; SENTI, G.; BATTEGAY, M.; WILDER-SMITH, A. et al. Covid-19 epidemic in switzerland: on the importance of testing, contact tracing and isolation. *Swiss medical weekly*, EMH Media, v. 150, n. 1112, 2020.
- 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.
- SALGOTRA, R.; GANDOMI, M.; GANDOMI, A. H. Evolutionary modelling of the COVID-19 pandemic in fifteen most affected countries. *Chaos, Solitons & Fractals*, Elsevier, v. 140, p. 110118, 2020.
- SALGOTRA, R.; GANDOMI, M.; GANDOMI, A. H. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos, Solitons & Fractals*, Elsevier, v. 138, p. 109945, 2020.
- 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.I.]: 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.I.], 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.I.], 2008. p. 1728–1735.

- SARKAR, K.; KHAJANCHI, S.; NIETO, J. J. Modeling and forecasting the covid-19 pandemic in india. *Chaos, Solitons & Fractals*, Elsevier, v. 139, p. 110049, 2020.
- SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural Networks*, Elsevier, v. 61, p. 85–117, 2015.
- 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.
- SCHULER-FACCINI, L.; RIBEIRO, E. M.; FEITOSA, I. M.; HOROVITZ, D. D.; CAVALCANTI, D. P.; PESSOA, A.; DORQUI, M. J. R.; NERI, J. I.; NETO, J. M. de P.; WANDERLEY, H. Y. et al. Possible association between zika virus infection and microcephaly—brazil, 2015. *Morbidity and Mortality Weekly Report*, JSTOR, n. 3, p. 59–62, 2016.
- SHAUkat, K.; MASOOD, N.; MEHREEN, S.; AZMEEN, U. Dengue fever prediction: A data mining problem. *Journal of Data Mining in Genomics & Proteomics*, v. 2015, 2015.
- SILVA, A. C. G. da; LIMA, C. L. de; SILVA, C. C. da; SILVA, E. L.; MARQUES, G. S.; ARAÚJO, L. J. B. de; JÚNIOR, L. A. A.; SOUZA, S. B. J. de; SANTANA, M. A. de; GOMES, J. C. et al. Covid-sgis: Uma ferramenta inteligente para monitoramento dinâmico previsão temporal da covid-19. In: *Anais do IV SABIO: Simpósio de Inovação em Engenharia Biomédica*. [S.I.: s.n.], 2020.
- SILVA, A. C. G. da; LIMA, C. L. de; SILVA, C. C. da; MORENO, G. M. M.; SILVA, E. L.; MARQUES, G. S.; ARAÚJO, L. J. B. de; JÚNIOR, L. A. A.; SOUZA, S. B. J. de; SANTANA, M. A. de et al. Machine learning approaches for temporal and spatio-temporal covid-19 forecasting: A brief review and a contribution. In: *Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis*. [S.I.]: Springer, 2022. p. 333–357.
- SILVA, A. C. G. da; SANTOS, W. P. dos. Modelagem e simulação da transmissão de arboviroses baseada em modelos bayesianos para predição de casos de dengue, chikungunya e zika. In: *Anais do IV SABIO: Simpósio de Inovação em Engenharia Biomédica*. [S.I.: s.n.], 2020.
- SILVA, C.; LIMA, C.; SILVA, A.; MORENO, G. M. M.; MUSAH, A.; ALDOSERY, A.; DUTRA, L.; AMBRIZZI, T.; BORGES, I. V. G.; TUNALI, M. et al. Forecasting Dengue, Chikungunya and Zika cases in Recife, Brazil: a spatio-temporal approach based on climate conditions, health notifications and machine learning. *Research, Society and Development*, CDRR Editors, v. 10, n. 12, 2021.
- SILVA, C. C. da; LIMA, C. L. de; SILVA, A. C. G. da; SILVA, E. L.; MARQUES, G. S.; ARAÚJO, L. J. B. de; JÚNIOR, L. A. A.; SOUZA, S. B. J. de; SANTANA, M. A. de; GOMES, J. C. et al. Covid-19 dynamic monitoring and real-time spatio-temporal forecasting. *Frontiers in Public Health*, Frontiers Media SA, v. 9, 2021.
- SILVA, C. C. da; LIMA, C. L. de; SILVA, A. C. G. da; MORENO, G. M. M.; MUSAH, A.; ALDOSERY, A.; DUTRA, L.; AMBRIZZI, T.; BORGES, I. V.; TUNALI, M. et al. Spatiotemporal

- forecasting for dengue, chikungunya fever and zika using machine learning and artificial expert committees based on meta-heuristics. *Research on Biomedical Engineering*, Springer, p. 1–39, 2022.
- SILVA, D. P. da; SANTOS, I. M. R. dos; MELO, V. dos S. Aspectos da infecção ocasionada pelo coronavírus da síndrome respiratória aguda grave 2 (sars-cov-2). *Brazilian Journal of Health Review*, v. 3, n. 2, p. 3763–3779, 2020.
- SINGH, S.; PARMAR, K. S.; KUMAR, J.; MAKKHAN, S. J. S. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. *Chaos, Solitons & Fractals*, Elsevier, v. 135, p. 109866, 2020.
- SINGHAL, T. A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*, Springer, n. 4, p. 281–286, 2020.
- 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.
- STONE, L.; OLINKY, R.; HUPPERT, A. Seasonal dynamics of recurrent epidemics. *Nature*, Nature Publishing Group, v. 446, n. 7135, p. 533–536, 2007.
- SUBA, M.; SHANMUGAPRIYA, R.; BALAMURALITHARAN, S.; JOSEPH, G. A. Current mathematical models and numerical simulation of sir model for coronavirus disease-2019 (covid-19). *European Journal of Molecular & Clinical Medicine*, n. 05, 2020.
- SUBRAMANIAN, R.; ROMEO-AZNAR, V.; IONIDES, E.; CODEÇO, C. T.; PASCUAL, M. Predicting re-emergence times of dengue epidemics at low reproductive numbers: Denv1 in rio de janeiro, 1986–1990. *Journal of the Royal Society Interface*, The Royal Society, v. 17, n. 167, p. 20200273, 2020.
- SUJATH, R.; CHATTERJEE, J. M.; HASSANIEN, A. E. A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 34, p. 959–972, 2020.
- SUPARIT, P.; WIRATSUDAKUL, A.; MODCHANG, C. A mathematical model for zika virus transmission dynamics with a time-dependent mosquito biting rate. *Theoretical Biology and Medical Modelling*, BioMed Central, v. 15, n. 1, p. 1–11, 2018.
- TAMANG, S.; SINGH, P.; DATTA, B. Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique. *Global Journal of Environmental Science and Management*, GJESM Publisher, v. 6, n. Special Issue (Covid-19), p. 53–64, 2020.
- TOMAR, A.; GUPTA, N. Prediction for the spread of covid-19 in india and effectiveness of preventive measures. *Science of The Total Environment*, Elsevier, v. 728, p. 138762, 2020.
- TUNALI, M.; RADIN, A. A.; BAŞIBÜYÜK, S.; MUSAH, A.; BORGES, I. V. G.; YENIGUN, O.; ALDOSERY, A.; KOSTKOVA, P.; SANTOS, W. P. dos; MASSONI, T. et al. A review exploring the overarching burden of zika virus with emphasis on epidemiological case studies from brazil. *Environmental Science and Pollution Research*, Springer, p. 1–15, 2021.

- USMAN, S.; ADAMU, I. I.; TAHIR, A. Determining important parameters in the transmission of zika virus infection using sensitivity analysis of a compartmental model. 2017.
- VALLE, D.; PIMENTA, D. N.; CUNHA, R. V. da. *Dengue: teorias e práticas*. [S.I.]: SciELO-Editora FIOCRUZ, 2015.
- VALOTTO, C. F. B.; CARVASIN, G.; SILVA, H. H. G.; GERIS, R.; SILVA, I. G. d. Alterações morfo-histológicas em larvas de aedes aegypti (*linaeus*, 1762)(diptera, culicidae) causadas pelo tanino catéquico isolado da planta do cerrado *magonia pubescens* (sapindaceae). 2010.
- VAPNIK, V. The support vector method of function estimation. Springer, p. 55–85, 1998.
- 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.
- VELÁSQUEZ, R. M. A.; LARA, J. V. M. Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression. *Chaos, Solitons & Fractals*, Elsevier, v. 136, p. 109924, 2020.
- VIRUSES, W. H. O. S. G. on A.-B. *Arthropod-borne viruses: report of a study group*. [S.I.]: World Health Organization, 1961.
- WANG, L.; LIN, Z. Q.; WONG, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, Nature Publishing Group, v. 10, n. 1, p. 1–12, 2020.
- WHO. *Coronavirus disease (COVID-19) pandemic*. [S.I.], 2020. Last accessed: 2020 May 20. Disponível em: <www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- WHO. *Coronavirus disease (COVID-19) pandemic*. [S.I.], 2021. Last accessed: 2021 august 09. Disponível em: <<https://covid19.who.int/>>.
- WITTEN, I. H.; FRANK, E. *Data mining: practical machine learning tools and techniques with Java implementations*. [S.I.]: ACM New York, NY, USA, 2002. v. 31. 76–77 p.
- WITTEN, I. H.; FRANK, E. *Data Mining: Pratical Machine Learning Tools and Technique*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 2005.
- XIE, J.; HUNGERFORD, D.; CHEN, H.; ABRAMS, S. T.; LI, S.; WANG, G.; WANG, Y.; KANG, H.; BONNETT, L.; ZHENG, R.; LI, X.; TONG, Z.; DU, B.; QIU, H.; TOH, C.-H. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *The Lancet*, p. 1–29, 2020.
- ZHONG, L.; MU, L.; LI, J.; WANG, J.; YIN, Z.; LIU, D. Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model. *IEEE Access*, IEEE, v. 8, p. 51761–51769, 2020.
- ZHU, B.; WANG, L.; WANG, H.; CAO, Z.; ZHA, L.; LI, Z.; YE, Z.; ZHANG, J.; SONG, H.; SUN, Y. Prediction model for dengue fever based on interactive effects between multiple meteorological factors in guangdong, china (2008–2016). *PLoS One*, Public Library of Science, v. 14, p. 1–12, 12 2019.

Índice

- Conclusão, 80
- Fundamentação teórica, 32
- Introdução, 15
- Proposta, 44
- Resultados Experimentais, 53
- Trabalhos relacionados, 20