

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WILSON ROGÉRIO SOARES E SILVA

**METODOLOGIA DE MONITORAMENTO DE
EPIDEMIAS: UMA ABORDAGEM BASEADA
EM REDES NEURAS ARTIFICIAIS**

DM: 04/2018

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2018

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WILSON ROGÉRIO SOARES E SILVA

METODOLOGIA DE MONITORAMENTO DE
EPIDEMIAS: UMA ABORDAGEM BASEADA
EM REDES NEURAS ARTIFICIAIS

DM: 04/2018

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2018

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WILSON ROGÉRIO SOARES E SILVA

**METODOLOGIA DE MONITORAMENTO DE
EPIDEMIAS: UMA ABORDAGEM BASEADA
EM REDES NEURAS ARTIFICIAIS**

Dissertação de Mestrado submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica como requisito para obtenção do Grau de Mestre em Engenharia Elétrica, na Área de Concentração de Computação Aplicada.

UFPA/ITEC/PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2018

**Dados Internacionais de Catalogação na Publicação (CIP)
Sistemas de Biblioteca da UFPA**

Silva, Wilson Rogério Soares e, 1982-
Metodologia de monitoramento de epidemias: uma
abordagem baseada em redes neurais artificiais/ Wilson
Rogério Soares e Silva.- 2018.

Orientador: Carlos Renato Lisboa Francês

Dissertação (Mestrado) - Universidade Federal do Pará.
Instituto de Tecnologia. Programa de Pós-Graduação em
Engenharia Elétrica, Belém, 2018.

1. Redes neurais (computação) 2. Dengue- Pará- Métodos
estatísticos I. Título

CDD 22.ed.006.3

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**"METODOLOGIA DE MONITORAMENTO DE EPIDEMIAS NA AMAZÔNIA:
UM ABORDAGEM BASEADA EM REDES NEURAIIS ARTIFICIAIS"**

AUTOR: WILSON ROGÉRIO SOARES E SILVA

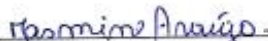
DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 26/01/2018

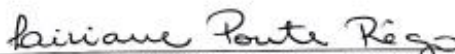
BANCA EXAMINADORA:



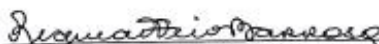
Prof. Dr. Carlos Renato Lisboa Francês
(Orientador – PPGE/UFPA)



Prof.ª Dr.ª Jasmine Priscyla Leite de Araújo
(Avaliadora Interna – PPGE/UFPA)



Prof.ª Dr.ª Liviane Ponte Régio
(Avaliadora Externa ao Programa – CCAST/UFPA)



Prof.ª Dr.ª Regina Fátima Feio Barroso
(Avaliadora Externa ao Programa – ICS/UFPA)

VISTO:

Prof.ª Dr.ª Maria Emilia de Lima Tostes
(Coordenadora do PPGE/ITEC/UFPA)

"Dedico especialmente à minha amada família"

AGRADECIMENTOS

A Deus por proporcionar mais essa vitória e a Santa Rita de Cássia pelo amparo diário.

À minha família pelo carinho, minha esposa Natalle pelo companheirismo e aos meus filhos Gustavo e nosso bebê Henrique que está por vir.

Ao meu orientador Renato Francês pela oportunidade e orientações ao longo de toda a pesquisa.

Aos meus amigos que sempre me ajudaram: Érica, Kleise.

Aos professores do LPRAD pelo profissionalismo e conhecimento repassados.

Aos amigos que fiz no LPRAD (Seu Zé Maria, Alfaia, Anderson, Priscila) que me ajudaram nessa caminhada.

Ao Instituto Federal do Pará pelo apoio a minha qualificação.

A Universidade Federal do Pará e ao Programa de Pós-Graduação em Engenharia Elétrica pela oportunidade e aprendizado.

Muito obrigado!

SUMÁRIO

Capítulo 1 – Introdução.....	1
1.1. Contextualização	1
1.2. Objetivos.....	5
1.3. Organização da Dissertação	6
Capítulo 2 – Trabalhos Correlatos e Revisão Bibliográfica	7
2.1. Considerações Iniciais	7
2.2. Trabalhos correlatos acerca de técnicas computacionais usadas em previsões	7
2.3. Redes Neurais Artificiais	16
2.3.1 Neurônio biológico.....	16
2.3.2 Modelo matemático de um neurônio.....	17
2.3.3 Redes Feed-Forward	19
2.3.4 Redes Recorrentes.....	20
2.3.5 Redes Multilayer Perceptron (MLP)	21
2.3.6 Algoritmos de treinamento e aprendizagem	23
2.3.6.1 Algoritmo Backpropagation	24
2.3.7 Métricas para Desempenho de rede	25
2.3.8 Predição de séries temporais com RNA	26
2.4. Considerações Finais	27
Capítulo 3 – Metodologia de monitoramento de epidemias: uma abordagem usando redes neurais artificiais.....	28
3.1 Considerações Iniciais	28
3.2 Metodologia desenvolvida.....	28
3.2.1 Área de estudo	28
3.2.2 Dados coletados	29
3.2.3 Desenvolvimento do Sistema de monitoramento e alertas de epidemias.....	30
3.2.3.1 Arquitetura do sistema de monitoramento e alertas.....	30
3.2.3.2 Modelagem do sistema – Caso de uso	31
3.2.3.3 Modelagem do sistema – Diagrama de classe	32
3.2.3.4 Modelagem do sistema – Diagrama de atividade	33
3.2.3.5 Protótipo do sistema – Interface do sistema desenvolvido.....	34
3.2.3.6 Módulo de alertas.....	34
3.2.4 Modelo da Rede Neural Artificial.....	35
3.2.4.1 Definição dos parâmetros da RNA	36
3.2.4.2 Entrada da RNA.....	36
3.2.4.3 Treinamento, validação e testes da RNA	37
3.2.4.4 Avaliação da RNA	37
3.3 Considerações finais.....	38
Capítulo 4 - Estudo de Caso: Predição de ocorrência de dengue no estado do Pará....	39
4.1 Considerações iniciais	39
4.2 Dados Analisados.....	39
4.2.1 Série Histórica dos Municípios.....	40
4.2.2 Estatística descritiva dos municípios	43
4.2.3 Distribuição dos dados	44
4.2.4 Séries Temporais das amostras	46
4.3 Modelo de Rede Neural Artificial.....	52
4.3.1 Experimento I	52

4.3.2	Experimento II	56
4.3.3	Avaliação dos experimentos I e II.....	59
4.4	Predição para novos casos de Dengue.....	60
4.4.1	Cenário com alto índice pluviométrico	60
4.4.2	Cenário com baixo índice pluviométrico	63
4.5	Avaliação das previsões dos casos de dengue	65
4.6	Análise dos Resultados.....	66
4.7	Considerações Finais	67
Conclusão	68
5.1	Contribuições	68
5.2	Trabalhos Futuros.....	69
5.3	Dificuldades Encontradas.....	70
REFERÊNCIAS	71

LISTA DE FIGURAS

- Figura 1.1** Distribuição global de dengue no mundo
- Figura 1.2** Situação epidemiológica no Brasil até 2015
- Figura 1.3** Óbitos por dengue no Brasil até 2015
- Figura 1.4** Casos registrados de dengue no Brasil 2015 a 2017
- Figura 2.1.** Estrutura do neurônio biológico
- Figura 2.2** Representação de um neurônio artificial
- Figura 2.3** Modelo matemático de um neurônio
- Figura 2.4** Arquitetura é do tipo Feed-forward.
- Figura 2.5** Arquitetura é do tipo recorrentes.
- Figura 2.6** Representação gráfica de RNA do tipo MLP
- Figura 2.7** Funções de ativação
- Figura 2.8** Fórmulas matemáticas das funções de ativação
- Figura 2.9** Predição ao longo do tempo $T+h$.
- Figura 3.1** Metodologia para predição de novos casos de dengue
- Figura 3.2** Região Metropolitana de Belém-Pa-Brasil.
- Figura 3.3** Estratégia de escolha da melhor configuração da rede
- Figura 3.4** Caso de uso do sistema desenvolvido
- Figura 3.5** Diagrama de classe sistema de monitoramento
- Figura 3.6** Diagrama de Atividade sistema desenvolvido
- Figura 3.7** Interface sistema de monitoramento
- Figura 3.8** Visualização do alerta para novos casos
- Figura 3.9** Estratégias para redes neurais
- Figura 4.1** Lista dos municípios com maior ocorrência de dengue
- Figura 4.2** Casos notificados de dengue do município de Belém
- Figura 4.3** Casos notificados de dengue de Parauapebas
- Figura 4.4** Casos notificados de dengue de Santarém
- Figura 4.5** Casos notificados de dengue de Altamira
- Figura 4.6** Índice pluviométrico x casos notificados
- Figura 4.7** Distribuição dos casos no município de Altamira
- Figura 4.8** Distribuição dos casos no município de Parauapebas
- Figura 4.9** Distribuição dos casos no município de Altamira

- Figura 4.10** Distribuição dos casos no município de Santarém
- Figura 4.11** Box-plot dos dados
- Figura 4.12** FAC Função de Auto Correlação das séries analisadas
- Figura 4.13** Tendência da Série
- Figura 4.14** Performance da RNA ajustada ao município de Belém
- Figura 4.15.** Performance da RNA ajustada ao município de Parauapebas
- Figura 4.16** Performance da RNA ajustada ao município de Altamira
- Figura 4.17** Performance da RNA ajustada ao município de Santarém
- Figura 4.18** Resultado em 3 épocas – Parauapebas
- Figura 4.19** Resultado em 3 épocas – Altamira
- Figura 4.20** Resultado em 3 épocas – Santarém
- Figura 4.21** Belém no primeiro semestre de 2006 a 2016
- Figura 4.22** Parauapebas no primeiro semestre de 2006 a 2016
- Figura 4.23** Altamira no primeiro semestre de 2006 a 2016
- Figura 4.24** Santarém no primeiro semestre de 2006 a 2016
- Figura 4.25** Belém no segundo semestre de 2006 a 2016
- Figura 4.26** Parauapebas no segundo semestre de 2006 a 2016
- Figura 4.27** Altamira no segundo semestre de 2006 a 2016
- Figura 4.28** Santarém no segundo semestre de 2006 a 2016
- Figura 4.29** Predição de casos de dengue de Belém-Pa
- Figura 4.30** Predição de casos de dengue de Parauapebas -Pa
- Figura 4.31** Predição de casos de dengue de Altamira-Pa
- Figura 4.32** Predição de casos de dengue de Santarém-Pa

LISTA DE TABELAS

Tabela 1.1 Número de casos prováveis e incidência de dengue (/100mil hab.)

Tabela 2.1 Resumo das principais técnicas usados na predição de epidemias de dengue.

Tabela 3.1. Lista dos municípios do estado

Tabela 3.2 Índice pluviométrico x casos notificados

Tabela 3.3 Estatística descritiva dos dados

Tabela 3.4 Teste de Estacionaridade

Tabela 3.5 RNA sazonalidade-PA

Tabela 3.6 RNA município Belém-PA

Tabela 3.7 RNA Resumo das RNAs

Tabela 3.8 RNA município Parauapebas-PA

Tabela 3.9 RNA município Altamira-PA

Tabela 3.10 RNA município Santarém-PA

Tabela 3.11 RNA Resumo das RNAs

Tabela 3.12 RNA Comparação das predições

LISTA DE SIGLAS E ABREVIATURAS

RNA – Redes Neurais Artificiais

OMS – Organização Mundial De Saúde

DATASUS - Departamento de Informática do Sistema Único de Saúde

SINAN - Sistema de Nacional e Agravos e Notificação

ARIMA - Auto-Regressivo De Médias Móveis

MLP – Perceptron Multi Camadas

RESUMO

A dengue é uma doença infecciosa viral presente em mais de 100 países no mundo. Nos países subdesenvolvidos como o Brasil essa patologia apresenta contornos dramáticos quando se acrescentam fatores socioeconômicos preponderantes como as condições precárias de saneamento básico características das grandes cidades. Ao associarmos esse cenário à Amazônia percebemos que a localização geográfica e as condições climáticas desse espaço contribuem para que a ocorrência dessa doença seja dimensionada. O Ministério da Saúde, disponibilizou dados resultantes de uma pesquisa que constata que dos 409.073 casos notificados na região Norte, 106.433 ocorreram no estado do Pará, em que os municípios com maiores notificações de casos de dengue são: Belém, Parauapebas, Altamira e Santarém. Este trabalho propõe uma metodologia para monitorar epidemias com base na utilização de Redes Neurais Artificiais, a partir de um estudo de caso de predição de casos de dengue no estado do Pará. Para isso, desenvolveu-se um sistema que usa base de dados públicos de casos da doença, de ocorrência semanal dos municípios já mencionados. Em adição, realiza a análise estatística das séries dos municípios constando complexidade, e justificando o uso de redes neurais para esse tipo de problema. Realiza os ajustes das camadas, janela de tempo do modelo neural treinado que nesse caso é uma variação conhecida como rede neural recorrente. E implementa um módulo de emissão de alertas, visando à detecção de um aumento repentino de novos casos da doença, contribuindo para tomada de decisão dos órgãos de saúde pública e suas respectivas ações de controle das epidemias nos municípios em estudo. A partir de nossas análises podemos concluir que a metodologia descrita na pesquisa tem validade para realizar previsões de casos de dengue, usando redes neurais, antecipando ações de combate e contribuindo para a tomada de decisão, que poderá ser usado por gestores públicos da área da saúde. E que o uso de redes neurais recorrentes consegue se ajusta a complexidade das séries estudadas. Os resultados demonstraram que o modelo de RNA, para o cenário em voga, obteve um bom desempenho na predição epidemiológica, alcançando acurácia satisfatória.

Palavras - Chave: Predição, Dengue, Redes Neurais Artificiais.

ABSTRACT

Dengue fever is a viral infectious disease that is present in more than 100 countries worldwide. In underdeveloped countries such as Brazil, this pathology presents dramatic contours when prevailing socioeconomic factors are added, such as the precarious basic sanitation conditions characteristic of large cities. When we associate this scenario with the Amazon we perceive that the geographic location and climatic conditions of this space contribute to the occurrence of this disease is dimensioned. The Ministry of Health provided data from a survey that found that of the 409,073 reported cases in the North, 106,433 occurred in the state of Pará, where the municipalities with the highest reports of dengue cases are: Belém, Parauapebas, Altamira and Santarém. This work proposes a methodology to monitor epidemics based on the use of Artificial Neural Networks, based on a case study of prediction of dengue cases in the state of Pará. To this end, a system was developed that uses a public database of cases of the disease, of weekly occurrence of the municipalities already mentioned. In addition, it performs the statistical analysis of the series of municipalities showing complexity, and justifying the use of neural networks for this type of problem. It performs the layer adjustments, time window of the trained neural model which in this case is a variation known as recurrent neural network. It implements a module for issuing alerts to detect a sudden increase in new cases of the disease, contributing to the decision-making of public health agencies and their respective actions to control epidemics in the municipalities under study. From our analysis we can conclude that the methodology described in the research is valid for predicting dengue cases using neural networks, anticipating combat actions and contributing to decision making, which can be used by public health managers . And that the use of recurrent neural networks can adjust to the complexity of the series studied. The results demonstrated that the RNA model, for the current scenario, performed well in the epidemiological prediction, reaching satisfactory accuracy.

Keywords: Forecast, Dengue, Artificial Network Neural.

Capítulo 1 – Introdução

1.1. Contextualização

O aumento alarmante dos casos de dengue, zica e chikungunya nos últimos anos vem preocupando diversos países em todo o mundo. A dengue é uma doença infecciosa viral que afeta mais de 100 países no mundo. Dados da Organização Mundial da Saúde (OMS) mostram mais de 50 milhões de indivíduos infectados anualmente, com 500 mil casos de dengue hemorrágica e 20 mil mortos [1]. Ainda de acordo com a OMS, cerca de 2,5 bilhões de pessoas (aproximadamente metade da população mundial) estão atualmente em risco de contrair dengue.

Todos os anos, estima-se que cerca de 50 milhões de pessoas sofrem com sintomas do vírus da dengue, com aproximadamente 55 mil indivíduos que necessitam de hospitalização [2]. A OMS, ao reconhecer os perigos dessa doença, incluiu a dengue em sua campanha de 2014 para o controle de doenças transmitidas por vetores. Dos 96 milhões de casos estimados globalmente de dengue por ano, 70% ocorrem na Ásia, sendo que a Índia contribui em 34% de todos os casos [3], como observado na Figura 1.1:

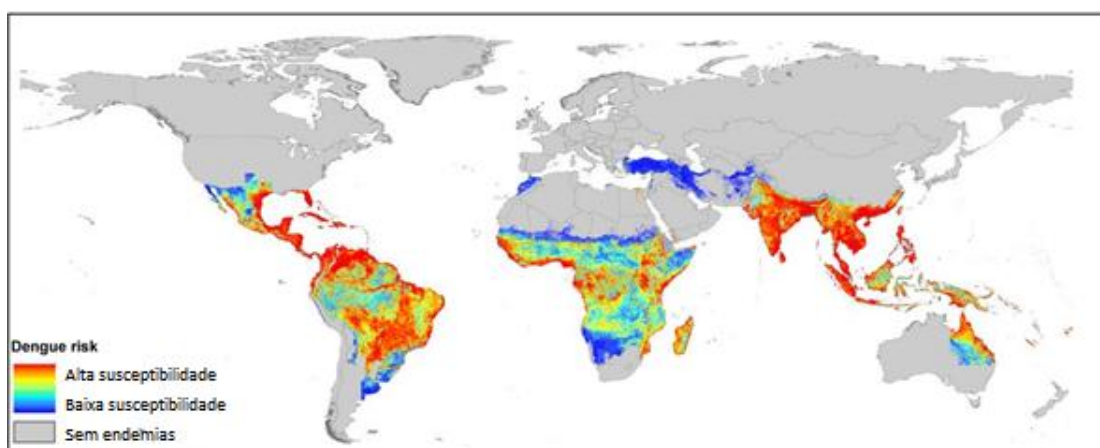


Figura 1.1. Distribuição global de dengue no mundo. [57]

Segundo [4], o Brasil possui um dos mais abrangentes sistemas de vigilância de dengue, com mais de 200 milhões de pessoas sob observação e dados mensalmente tabulados, ao

longo de dez anos ou mais, em 5570 municípios. O país já dedica recursos significativos para o controle de vetores, incluindo os transmissores da dengue, além de ter uma comunidade de pesquisadores dedicada a estudar o avanço de epidemias, suas causas e consequências. Houve aumento nos últimos anos com internações hospitalares causados pelo aumento de casos de dengue, zica e chikungunya em todo território nacional, agravando-se nos últimos anos da série histórica, como podemos observar na a seguir Figura 1.2:

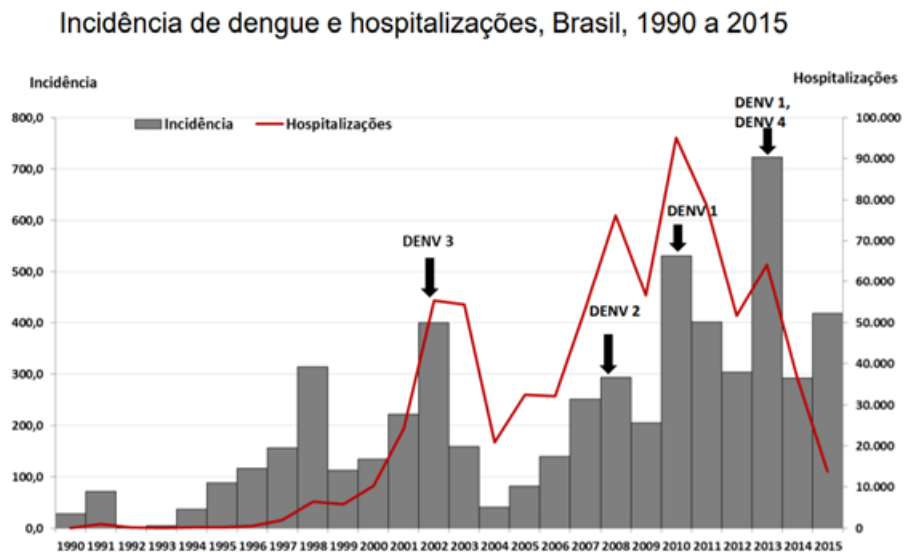


Figura 1.2 Situação epidemiológica no Brasil até 2015. Fonte: [82]

O Ministério da Saúde chama a atenção para o aumento de casos de óbitos por dengue ao longo da última década. O relatório propõe um plano de contingência nacional para epidemias de dengue, como observado na Figura 1.3:

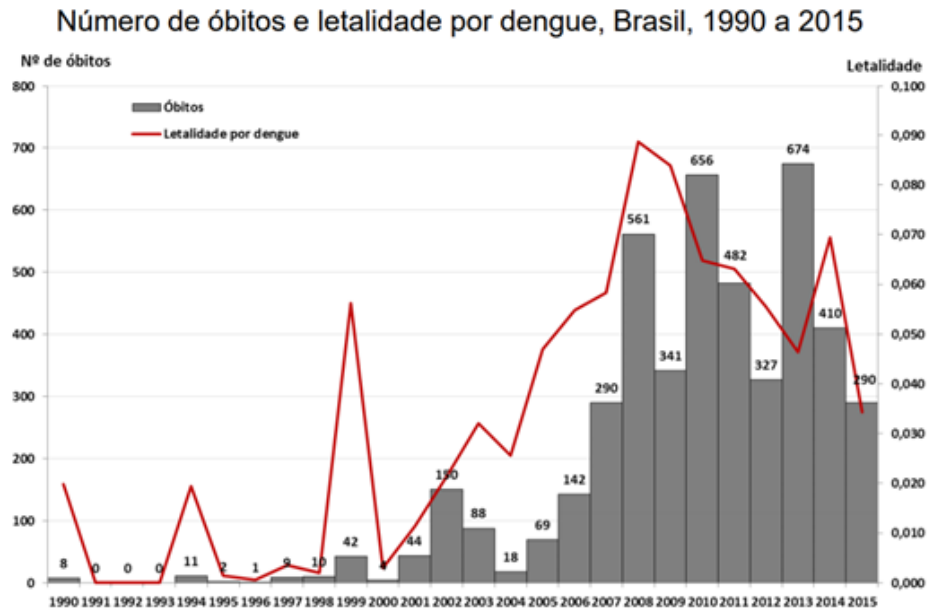


Figura 1.3 Óbitos por dengue no Brasil até 2015. Fonte: [82]

Dengue, febre de Chikungunya e febre pelo vírus Zika são doenças de notificação compulsória e estão presentes na Lista Nacional de Notificação Compulsória de Doenças, Agravos e Eventos de Saúde Pública, sendo que a febre pelo vírus Zika foi acrescentada a essa lista apenas pela Portaria nº 204, de 17 de fevereiro de 2016, do Ministério da Saúde. Em 2015, 1.688.688 casos prováveis de dengue foram registrados, em 2016 esse número diminuiu para 1.483.623 casos prováveis e no ano 2017, de (1/1/2017 a 19/08/2017), foram registrados 214.990 casos prováveis de dengue no país, com uma incidência de 104,3 casos/100 mil habitantes, como demonstrado na Figura 1.4. apresentada a seguir:

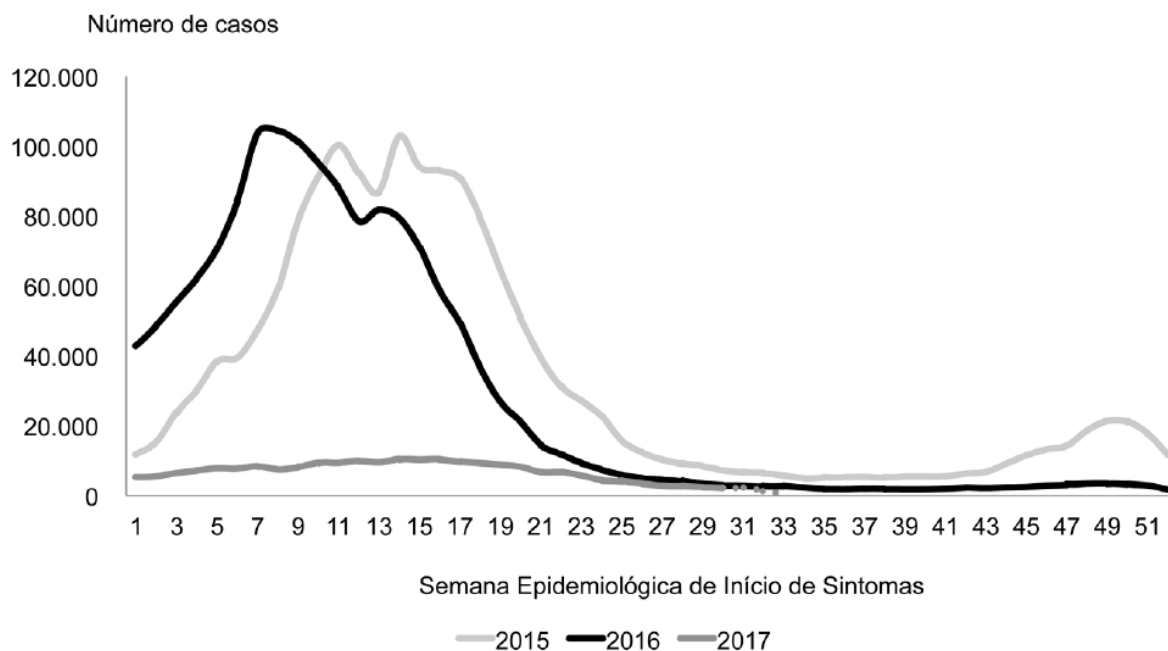


Figura 1.4 Casos registrados de dengue no Brasil 2015 a 2017. Fonte: [56]

O Ministério da Saúde afirma que dos 409.073 casos notificados na região Norte, no período de 2000 a 2010, 106.433 ocorreram no Estado do Pará, o restante distribuído nos demais Estados da região Norte, ou seja, cerca e 20% em um dos sete estados existentes. Os primeiros casos notificados de dengue ocorreram em 1995, nos municípios de Redenção e Rondon do Pará, região Sudeste do Pará.

Observa-se que há incidência da dengue na maioria dos municípios do estado do Pará. Atualmente 32 municípios paraenses desenvolvem ações prioritárias para a prevenção da doença a partir do Programa Nacional de Controle da Dengue, dentre os quais estão Altamira, Belém, Cametá, Conceição do Araguaia, Itaituba, Marabá, Monte Alegre, Santarém, Soure e Tucuruí.

Em 2017, até a SE 33, a região Nordeste apresentou o maior número de casos prováveis (80.447 casos; 37,4%) em relação ao total do país. Em seguida, aparecem as regiões Centro-Oeste (64.370 casos; 29,9%), Sudeste (47.290 casos; 22,0%), Norte (20.428 casos; 9,5%) e Sul (2.455 casos; 1,1%), como demonstrado na Tabela 1 a seguir.

Tabela 1. Número de casos prováveis e incidência de dengue (/100mil hab.), até a Semana Epidemiológica 33, por região e Unidade da Federação, Brasil, 2016 e 2017 Fonte: [56]

Região/Unidade da Federação	Casos (n)		Incidência (/100 mil hab.)	
	2016	2017	2016	2017
Norte	33.752	20.428	190,6	115,4
Rondônia	6.916	2.185	387,0	122,3
Acre	1.912	1.284	234,1	157,2
Amazonas	6.687	3.272	167,1	81,8
Roraima	144	683	28,0	132,8
Pará	9.385	6.887	113,4	83,2
Amapá	1.523	762	194,7	97,4
Tocantins	7.185	5.355	468,7	349,3

Segundo as notificações registradas do SINAN no estado do Pará os 10 (dez) municípios com maiores incidência de casos de dengue notificados por município são: Belém, Parauapebas, Altamira, Santarém, Marabá, Itaituba, Ananindeua, Conceição do Araguaia, Marituba e São Félix do Xingu [].

Assim, este trabalho propõe uma estratégia de monitoramento de epidemias, baseada em Redes Neurais Artificiais (RNAs), para prever ocorrências na Amazônia. Para comprovar a eficácia da proposta foi realizado um estudo de casos de dengue, com municípios com maior incidência de casos, que utiliza bases de dados públicos do SINAN (Sistema de Informação de Agravos de Notificação) de casos semanais de dengue nos municípios de Belém, Parauapebas, Altamira e Santarém. Posteriormente, os dados foram submetidos ao modelo computacional RNA para predição de novos casos epidemiológicos e emissão de alertas, que poderão ser utilizados pelos órgãos competentes da área da saúde em tomadas de decisão e no planejamento de ações para o controle das epidemias, nos seus municípios.

A partir do contexto exposto é imperioso que tomadores de decisão no âmbito de políticas públicas possuam minimamente dados e mecanismos de apoio à eficiência de ações em saúde pública em geral e, em particular, em doenças epidêmicas como a dengue, zica e chikungunya.

1.2 Objetivos

Nesta pesquisa é proposta, como objetivo precípua, uma estratégia de monitoramento de epidemias, baseada em RNA, para prever ocorrências na Amazônia. A fim de validar tal estratégia, um estudo para efetivar a predição de casos de dengue na Amazônia é realizado, com dados provenientes de dados públicos do SINAN nos municípios de Belém,

Parauapebas, Altamira e Santarém.

Como objetivos específicos, têm-se, entre outros:

- Analisar o modelo computacional neural capaz de realizar previsões de epidemias.
- Uso de arquiteturas de RNA para o problema específico de previsões em séries temporais.
- Implementar um sistema de monitoramento de epidemias.
- Manipular bases oficiais do Ministério da Saúde, obtidas a partir do SINAN;

1.3 Organização da Dissertação

Esta dissertação está organizada em 05 capítulos, além desta introdução, conforme descrevemos a seguir:

Capítulo 2: Apresenta o alicerce teórico, que trata sobre a predição de epidemias usando modelos computacionais e RNAs, no qual a proposta foi construída.

Capítulo 3: Expõe a metodologia proposta para predição de epidemias na Amazônia.

Capítulo 4: Neste capítulo demonstra-se o estudo de casos realizado a partir da utilização dados de casos de dengue nos 04 municípios do Pará observados.

Capítulo 5: Apresenta a conclusão da pesquisa, as contribuições deste trabalho, apresentando as dificuldades encontradas e os possíveis desdobramentos em trabalhos futuros.

Capítulo 2 – Trabalhos Correlatos e Revisão Bibliográfica

2.1. Considerações Iniciais

Este capítulo apresenta inicialmente um levantamento de trabalhos correlatos, pesquisados a partir da literatura especializada, com a finalidade de traçar o panorama de eventuais consensos e tendências. Em seguida, são abordados os principais conceitos relacionados a técnicas computacionais usadas para predição e, em especial, as RNAs. Tal compreensão dos termos e sua implementação é essencial para o melhor entendimento do escopo desta pesquisa.

2.2. Trabalhos correlatos acerca de técnicas computacionais usadas em previsões

A detecção precoce de surtos da doença e ações de controle rápido são essenciais para prevenir e conter a propagação de doenças infecciosas e reduzir a morbidade. Assim sendo, diversas abordagens são encontradas na bibliografia focadas no estudo de propagação de diferentes doenças epidemiológicas. Ao longo das últimas décadas, numerosos métodos de predição foram propostos no campo da predição epidemiológica. Tais métodos podem ser classificados em diferentes categorias, como determinísticos *versus* probabilísticos [5].

Conforme o trabalho de [6], os casos diários de dengue nas 457 aldeias urbanas da cidade de Kaohsiung, Taiwan, de 2009 a 2012, foram utilizados para desenvolvimento e avaliação de modelos de predição. Esta pesquisa tem como objetivo prever os casos confirmados de dengue usando RNA. Houve no total 2,997 casos confirmados de dengue durante esse período, um modelo de regressão foi ajustado aos incidentes diários ocorridos nas aldeias nos últimos 30 dias. O modelo foi usado para prever as probabilidades de incidência do surto de dengue para as aldeias no dia seguinte.

A pesquisa de [7] tem como objetivo prever os casos confirmados por dengue usando RNA. Nesse estudo os autores utilizaram dados fornecidos pela Agência Nacional de Meio Ambiente de Singapura para modelar o comportamento dos casos de dengue com base nos parâmetros físicos da temperatura, da umidade relativa e da precipitação total. Os resultados mostraram que as quatro características importantes, como a temperatura média, a umidade relativa média, a precipitação total e o número total de casos confirmados por dengue foram muito efetivas na predição do número de casos confirmados por dengue.

Nesse contexto o trabalho de [1] tem como objetivo auxiliar a Secretaria de Saúde do Estado do Rio de Janeiro a reduzir os riscos de epidemia de dengue no Estado. Os aumentos dos casos epidemiológicos vêm exigindo atenção dos gestores públicos de saúde, seja em ações, em serviços de saúde para a sua prevenção e/ou no seu controle.

No trabalho de [8] usa um Modelo de autômato celular com vizinhança é apresentada para simular uma surto da doença infecciosa. O modelo facilita o estudo da dinâmica de Epidemias de diferentes doenças infecciosas. Utilizando um sistema de vigilância, [8] afirmam que é possível aplicar métodos estatísticos, média móvel integrada auto regressiva (*ARIMA*), a fim de prever uma curva epidêmica ou para detectar aberrações na propagação da doença . A predição de casos de dengue em uma população usando modelos de séries temporais pode fornecer informações úteis que podem ser usadas para facilitar o planejamento de intervenções de saúde pública. O modelo de predição para incidência de dengue foi realizado com software R usando o modelo de média móvel integrada auto regressiva sazonal (*SARIMA*), como afirma [9].

Pesquisadores do Laboratório de Análise Espacial de Dados Epidemiológicos (epiGeo - UNICAMP) realizaram estudos sobre o risco de dengue na região sul da cidade de Campinas (São Paulo, Brasil) entre 2006 e 2007, por meio de métodos estatísticos e um modelo *Fuzzy* do tipo *Takagi-Sugeno* [10]. Em [11] estima-se que os efeitos climáticos de dengue, em Selangor, Kuala Lumpur e Putrajaya, Malásia, de 2008 a 2010. Selecionaram-se os parâmetros meteorológicos com um modelo aditivo generalizado de Poisson, e depois avaliou-se os efeitos de temperatura mínima, precipitação acumulada semanalmente e velocidade do vento sobre casos de dengue usando um modelo de desfasamento não-linear distribuído. Descobriu-se que o risco relativo de casos de dengue é positivamente associado ao aumento da temperatura mínima em uma porcentagem acumulada. Em [12] apresenta-se a influência da incidência de dengue com variável de clima duplo. Técnicas de aprendizado de máquina, como agrupamento e regressão, técnicas não supervisionadas de agrupamento de *K-means* são feitas para agrupar os dados de acordo.

O trabalho de [6] aponta que casos diários de dengue nas 457 aldeias urbanas da cidade de Kaohsiung, Taiwan, de 2009 a 2012 foram utilizados para desenvolvimento e avaliação de modelos de predição. Houve no total 2.997 casos confirmados de dengue. Durante esse período um modelo de regressão foi ajustado aos incidentes diários ocorridos nas aldeias nos últimos 30 dias. O modelo ajustado foi usado para prever as probabilidades de incidência do surto de dengue para as aldeias no dia seguinte nessa pesquisa o objetivo

era prever os casos confirmados por dengue usando *RNA*. Os dados fornecidos pela Agência Nacional de Meio Ambiente de Singapura (NEA) foram utilizados para modelar o comportamento de casos de dengue com base nos parâmetros físicos da temperatura, da umidade relativa e da precipitação total. O conjunto de dados registrados consiste em 14.209 casos confirmados por dengue foram analisados usando *RNAs*. Foram produzidos resultados muito encorajadores neste estudo [7]. Modelos do trabalho de [13] foram desenvolvidos para prever alta e baixa incidência de dengue para fornecer avisos oportunos nas Filipinas. As entradas do modelo foram escolhidas com base em estudos que indicam variáveis que podem afetar a incidência da dengue. Primeiro usa técnicas de Mineração de Regras de Associação *Fuzzy* para extrair as regras de associação desses dados epidemiológicos históricos, dados ambientais e socioeconômicos, bem como dados climáticos que indicam padrões climáticos futuros. Os modelos previam alta ou baixa incidência de dengue na província de Filipinas com quatro semanas de antecedência.

Em [14] desenvolveu modelos preditivos para detecção de surto de dengue na Malásia, usando classificadores baseados em regras múltiplas. Os classificadores baseados em regras usados: árvore de decisão, *Naive Bayes* e Classificador Associativo. Em [15] mostra que o surto de dengue no sul de Taiwan coloca a nação inteira em alerta. O artigo aborda a Otimização linear e vetor de suporte para prever o surto de dengue que tem associação com localização, temperatura do ar e precipitação diária.

A dengue continua a ser a doença transmitida pelo vírus mais importante no mundo, particularmente na Ásia e América Latina, e é significativamente afetada pela variabilidade climática. A influência do clima em uma região endêmica como a Colômbia serviu como fonte para análise importante nos estudos sobre a dengue. Foram coletados dados de vigilância epidemiológica (casos semanais) de 2010 a 2011 e as taxas de incidência foram calculadas. Os modelos de regressão de Poisson foram utilizados para avaliar a influência da variável climática [16].

O trabalho de [12] centra-se na identificação de pontos quentes de dengue usando o algoritmo de agrupamento *KNN*. Os dados são coletados do estado de Selangor na Malásia de 2013 a 2014. A vulnerabilidade é obtida através de modelos que usam o algoritmo *K-means*. Os resultados demonstram a capacidade de visualização de renderização para a vulnerabilidade das incidências de dengue de cluster de alta densidade e baixa densidade. Em [17] mostra que os testes relevantes na análise de séries temporais de janeiro de 2010 a dezembro de 2014 são realizado para desenvolver dois modelos Auto Regressivo Médias

Móveis (*ARMA*), publicado pela Ministério da Saúde do Sri Lanka.

O trabalho de [18] tem como objetivo examinar a associação entre os fatores climáticos e a incidência da dengue e identificar a melhor predisposição climática em Mekong Delta no Vietnã. Utilizaram três modelos de regressão diferentes, que incluem: modelo de regressão múltipla padrão (*SMR*), modelo médio vertical auto regressivo sazonal (*SARIMA*) e Poisson distribuição modelo (*PDLM*) para examinar a associação entre os fatores climáticos e a incidência da dengue no período 2003-2010. Os resultados indicam que a temperatura e a umidade relativa estão significativamente associadas às mudanças na incidência de dengue consistentemente em todos os métodos do modelo utilizados, mas não nas chuvas cumulativas. Em [19] afirma que apesar de todo o esforço feito pelo Ministério da Saúde na luta contra a dengue na Malásia, o número de casos aumentou anualmente e tornou-se mais grave no início do século XXI. Os dados foram coletados desde o início do 2005 até 2010 e foram analisados usando o pacote estatístico de Ciências Sociais (SPSS). Os resultados da análise mostraram que a incidência dos casos é alta (mais de 100 casos) quando há a presença de precipitação e temperatura no intervalo relativo acima de 28,5°C. Em [20] afirma que os *drones* transmitem e recebem pacotes uns com os outros em uma configuração de rede sem fio. A transmissão de pacotes entre *drones* falha por vários motivos. O padrão de propagação da informação através da transmissão de pacotes em uma rede de *drones* pode ser considerado semelhante ao padrão de transmissão de doenças infecciosas em uma rede de interação humana. Neste trabalho, utiliza-se uma abordagem de cadeia de *Markov microscópica (MMCA)*, que foi aplicada para modelar os padrões de epidemias de doenças em uma rede humana, para investigar o padrão de transmissão de pacotes em uma escala microscópica.

No trabalho de [9] tem como objetivo desenvolver um modelo de predição para a incidência de dengue em Ribeirão Preto, Estado de São Paulo, utilizando análise de séries temporais. O modelo foi realizado utilizando a *Média Móvel Integrada Auto regressiva Sazonal (SARIMA)*. Criaram um modelo considerando notificações mensais de casos de dengue registrados de 2000 a 2008 em Ribeirão Preto. Em seguida, extraíram os valores previstos para 2009 do modelo ajustado e os comparamos com o número de casos observados nesse ano. O modelo *SARIMA* oferece o melhor ajuste para os dados de indenização da dengue. Os resultados mostraram que o modelo *ARIMA* de época prevê o número de casos de dengue de forma muito eficaz e confiável. É uma ferramenta útil para o controle e prevenção de doenças. A Indonésia tem o maior número casos de dengue no

Sudeste Asiático.

Em [21] a detecção precoce da doença é necessária para preparar medidas preventivas contra a dengue. Pesquisas anteriores mostraram que certas consultas relacionadas a doenças transmissíveis no *Google Trends* estão altamente correlacionadas com o número de casos de doenças transmissíveis na Coreia do Sul. Todos os dados abrangem dezembro de 2010 até agosto de 2015. As técnicas de interpolação e extrapolação são usadas para lidar com os dados faltantes. O modelo *ARIMA* e *ARIMAX* com dados do *Google Trends* é implementado para prever o número de casos de dengue. A pesquisa mostra que a adição do *Google Trends* ao modelo *ARIMAX* melhora o desempenho de predição. O melhor modelo *ARIMAX* com *Google Trends* melhora o valor MAPE em 3%.

Em [15] afirma que o surto de dengue no sul de Taiwan coloca a nação inteira em alerta. Este artigo aborda o *C-Support Vector Classification (C-SVC)* para prever o surto estatístico de febre de dengue que tem associação com a localização, temperatura do ar e precipitação diária. Em [22] afirma que mapas de risco que identificam locais adequados para transmissão de infecção são importantes para o planejamento da saúde pública. Os dados sobre as taxas de infecção por dengue não estão prontamente disponíveis na maioria dos lugares onde a doença é conhecida. Um novo complemento disponível no pacote de software *ArcGIS*, foi usado para identificar locais em África com características ambientais suscetíveis de serem adequadas para a transmissão do vírus da dengue. Em [23] afirma que há evidências de que a temperatura influenciará a transmissão da dengue. No entanto, as projeções com base em termos de longo prazo ainda não foram desenvolvidas. Avaliou-se o impacto da variabilidade climática na transmissão da dengue em Dhaka, Bangladesh e projetou o futuro risco de dengue atribuível à mudança climática. Os resultados mostram que as variáveis climáticas particularmente a temperatura e a umidade foram positivamente associadas à transmissão da dengue.

Em [24] explora as associações entre o número mensal de casos de febre de dengue e possíveis fatores de risco em Guangzhou, uma cidade subtropical da China. O número mensal de casos e medidas meteorológicas durante 2006-2014 registrado em Guangzhou, China, foram avaliados. Um modelo de regressão binomial foi utilizado para avaliar as relações entre fatores meteorológicos e o número mensal de casos. Em [25] afirma que o primeiro surto de dengue ocorrido na Ilha da Madeira em 2012, com o sorotipo de vírus do vírus *Aedes aegypti* foi o vetor da doença e é improvável que seja eliminado da ilha. Portanto, um novo surto de dengue pode ocorrer e, se isso acontecer, o risco para a

população aumenta quando dois sorotipos coexistem. Neste trabalho a modelagem matemática e as simulações numéricas são realizadas para prever o que pode acontecer na Ilha da Madeira. Em [26] afirma que muitas técnicas populares de predição de dengue têm sido utilizadas por vários pesquisadores para extrapolar taxas de incidência de dengue. A metodologia de análise de séries temporais, particularmente *ARIMA* e *SARIMA*, tem sido cada vez mais aplicada no campo da pesquisa epidemiológica para dengue, dengue hemorrágico e outras doenças infecciosas. Novos fatores correlacionados com a doença são necessários para aumentar a precisão da predição do modelo quando aplicado em áreas de climas similares, onde fatores climáticos como temperatura, precipitação total e umidade não são substancialmente diferentes.

Em [27] afirma-se que o clima é um importante fator contribuinte para o surto e a propagação da dengue porque afeta fortemente a densidade e a distribuição dos mosquitos que carregam dengue. No entanto, os modelos de predição de dengue baseados exclusivamente em fatores climáticos locais tiveram sucesso limitado. Assim, o artigo propõe um novo método de detecção de surto de dengue que depende de características de entropia de permutação de atraso (DPE) extraídas dos dados meteorológicos diários. Usando dados de Hong Kong entre 2004 e 2015, 4383 funções DPE diárias foram extraídas e analisadas usando técnicas de aprendizado de máquina. Os resultados da análise mostram que existe uma forte correlação entre os casos de dengue e os recursos DPE de precipitação.

No trabalho de [28] afirma que na área urbana de Guadalajara, no sudoeste da Colômbia, os locais de criação potenciais dentro e fora das casas foram caracterizados pela primeira vez; o pessoal local foi treinado para avaliar a produtividade com base no índice das pessoas. Simultaneamente, treinamento e monitoramento foram implementados para melhorar o sistema de vigilância de casos de dengue. Foram utilizados dados de entomologia e estudos de resistência a inseticidas para definir a intervenção direcionada. Em seguida, utilizou-se um delineamento quase experimental para avaliar a eficácia da intervenção em termos de índice de positividade dos locais de reprodução alvo e não-alvo e o impacto nos casos de dengue. Em [29] desenvolveram um modelo de regressão multivariada de Poisson com séries temporais utilizando a temperatura média semanal e as chuvas cumulativas durante o período 2000-2010. Analisamos vários tempos de atraso entre a dengue e as variáveis climáticas para identificar o ótimo período de predição da dengue. Auto regressão, sazonalidade e tendência foram considerados no modelo. Validou-se o modelo prevendo casos de dengue para a semana 1 de 2011 até a semana 16 de 2012

usando dados meteorológicos sozinhos.

Em [2] faz-se uma revisão sobre a situação da dengue desde o início da década de 1960 até as epidemias mais recentes no Paquistão, e também descreve o vetor primário desta doença (*Aedes aegypti*) no Paquistão. [30] apresenta uma abordagem metodológica que combina análise multi-resolução e sistemas difusos para representar casos de dengue e dengue grave na Colômbia. O desempenho desta proposta foi comparado com o obtido pela aplicação de técnicas tradicionais de modelagem difusa no mesmo conjunto de dados. Essa comparação foi obtida por duas medidas de desempenho que avaliam a semelhança entre os dados originais e o sinal aproximado: o erro quadrático médio e a variância representada. Finalmente, a capacidade preditiva da técnica proposta foi avaliada para prever o número de casos de dengue e dengue grave em um horizonte de três anos (2012-2015). Em [31] apresenta a *Dengue Active Surveillance System(DASS)* para um sistema de alerta precoce do surto. A dengue e a dengue hemorrágica estão surgindo como grandes problemas de saúde pública na maioria dos países asiáticos, como a Malásia. Programas eficazes de prevenção e controle dependerão de uma melhor vigilância. Uma nova abordagem de vigilância ativa delineada com ênfase no período da epidemia. O objetivo é desenvolver um sistema de vigilância de alerta precoce (estrutura) que possa prever a dengue epidêmica para melhorar o sistema de vigilância passiva atual disponível na Malásia.

Em [32] um protótipo de sistema de alerta precoce de dengue foi desenvolvido para produzir previsões probabilísticas de risco de dengue três meses antes da Copa do Mundo 2014, no Brasil. Aqui, avaliou-se as previsões categóricas de dengue em todas as microrregiões no Brasil, utilizando casos de dengue relatados em junho de 2014 para validar o modelo. Também comparamos a estrutura do modelo de predição com um modelo nulo, com base nas médias sazonais da incidência de dengue observada anteriormente. Este modelo de modelo de alerta precoce pode ser útil para os serviços de saúde pública, não só antes dos encontros de massa, mas também antes da estação máxima de dengue a cada ano, para controlar epidemias potencialmente explosivas de dengue. [33] mostra que os casos de dengue aumentaram durante as últimas décadas, particularmente em áreas não endêmicas, e a Argentina não foi exceção. A contribuição relativa das variáveis geográficas, demográficas e climáticas da ocorrência de casos de dengue foi avaliada. De acordo com a história da dengue no país, o estudo foi dividido em duas décadas, uma primeira década correspondente ao ressurgimento da doença e a segunda, incluindo várias epidemias. A distribuição espacial da ocorrência de dengue foi modelada como uma função da produção

do modelo mecanicista, variáveis climáticas, geográficas e demográficas durante as duas décadas. [34] afirma que o sistema de vigilância em Guadalupe (Antilhas Francesas) é atualmente operacional para a detecção de surtos precoce de dengue. O objetivo do estudo foi melhorar esse sistema de vigilância, avaliando uma ferramenta de modelagem para prever a ocorrência de epidemias de dengue poucos meses antes e assim ajudar um controle eficiente da dengue.

Em [35] analisa os surtos de dengue na cidade de Ahamedadbad na Índia, no período de 2005-2012. Usaram simulação de Monte Carlo para distribuição dos parâmetros. [36] afirma que o sistema de informação geográfica e o método analítico espacial poderiam ser aplicados para aumentar as estratégias de vigilância e outras doenças transmissíveis em um esforço para promover ações de prevenção e controle. Os dados de vigilância da doença obtidos poderiam ser integrados com uma base de dados ambiental de forma sinérgica, o que, por sua vez, fornecerá contribuições adicionais para o desenvolvimento de modelos de predição de epidemias. Em [37] destaca a necessidade de incorporar o impacto do meio ambiente urbano em tentativas de modelar a distribuição potencial de *A. aegypti* e discute-se brevemente o potencial de tecnologia futura para auxiliar a gestão e controle dessas espécies de vetores generalizados.

A pesquisa de [38] afirma que a dengue ressurgiu em Cingapura durante 2013-14, causando um surto com números sem precedentes de casos no país. No presente estudo, resumem-se os achados epidemiológicos, virológicos e entomológicos coletados através do programa de vigilância da dengue e destacam-se os impulsionadores da epidemia. Também descreve como o sistema de vigilância facilitou a preparação para moderar a transmissão epidêmica de dengue no país. [39] investigou fatores meteorológicos e demográficos que afetam a duração da epidemia de dengue e o período de tempo entre epidemias em Barbados, Brasil e Tailândia. Dados meteorológicos e demográficos específicos da região foram coletados em 104 locais de fontes públicas. A análise de história de eventos logísticos por parte dos efeitos corrigidos foi utilizada para quantificar os efeitos de variáveis no tempo sobre a duração das epidemias. [40] faz uma revisão com o intuito de avaliar a precisão do Google em comparação com os sistemas tradicionais de vigilância na Venezuela.

O trabalho de [41] comenta que apesar dos grandes esforços para controlar a transmissão, vetores competentes agora infestam a maioria das regiões tropicais e subtropicais; *Aedes*, também um vetor capaz de sobreviver em áreas temperadas, colocando

partes da Europa e da América do Norte em risco de transmissão local. O artigo de [42] tenta quantificar o risco da dengue na Europa, após a chegada e propagação de uma das espécies de vetores de dengue. Um mapa de risco global para a dengue é apresentado, com base em um banco de dados global da ocorrência desta doença, derivado de pesquisas de literatura. O trabalho de [1] tem como objetivo auxiliar a Secretaria de Saúde do Estado do Rio de Janeiro a reduzir os riscos de epidemia de dengue no Estado. As epidemias de dengue vêm cada vez mais exigindo atenção dos gestores públicos de saúde, seja em ações, em serviços de saúde para a sua prevenção e/ou no seu controle. Por conta da magnitude desse problema, é crucial adotar medidas que impeçam e/ou se antecipem às epidemias.

O trabalho de [43] utiliza dados de casos de dengue diariamente confirmados, ocorridos nas últimas duas décadas em Kaohsiung (Taiwan) em 2002, 2014 e 2015. Em [44] apresentam um artigo que faz uma comparação de três algoritmos de mineração de dados baseados em árvore de decisão de renome, como *C4.5* para prever o risco de dengue. O trabalho apresentado é simulado em "*weka*", uma ferramenta de mineração de dados. A análise de desempenho dos algoritmos é comparada em termos de taxa de sucesso e tempo de processamento.

A Tabela 3.1 mostra o resumo das principais modelos usados para predição de epidemias, seus respectivos autores e as principais técnicas de inteligência computacional usada para estudo das epidemias de dengue.

Tabela 3.1 Resumo das principais técnicas computacionais usados na predição de epidemias de dengue.

N	AUTORES	MODELO
1	[6]	RNA
2	[7]	RNA
3	[1]	RNA
4	[13]	FUZZY
5	[14]	NAIVE BAYES
6	[15]	OTIMIZAÇÃO LINEAR
7	[16]	POISON
8	[12]	KMEANS
9	[17]	ARIMA
10	[18]	SARIMA
11	[19]	MINERAÇÃO
12	[20]	MARKOV
13	[9]	SARMA
14	[21]	ARIMAX

15	[15]	SVM
16	[23]	TIME SERIES MODEL POISSON
17	[24]	MODELO REGRESSÃO BINOMIAL
18	[25]	SIMULAÇÃO
19	[26]	SARIMA/ARIMA
20	[27]	SVM
21	[34]	SARIMA
22	[38]	PROCESSAMENTO IMAGEM
23	[44]	ÁRVORE DE DECISÃO

De acordo com o levantamento de trabalhos correlatos, percebe-se o interesse de diversos grupos de pesquisas em propagação de epidemias ao redor do mundo. Dentre os estudos realizados, está o uso de modelos computacionais que permitem desenvolver estudos afim de prever novos casos da doença baseados em séries históricas.

Por conta da magnitude desse problema, é crucial adotar medidas que impeçam e/ou se antecipem os surtos da doença. Neste contexto, este trabalho apresenta um estudo aplicado a epidemias na região amazônica usando uma das abordagens citadas anteriormente que são RNAs.

2.3. Redes Neurais Artificiais

2.3.1 Neurônio biológico

A estrutura neural de um cérebro biológico é basicamente formada por neurônios, cuja composição é ilustrada na Figura 2.1. Um neurônio biológico é basicamente constituído por dendritos que são finos prolongamentos e representam os terminais de entrada de sinais recebidos de outros neurônios, pelo corpo central que é responsável por coletar e combinar as informações recebidas a partir dos dendritos; e pelo axônio, que é constituído de uma fibra tubular e representa o terminal de saída. Eles se comunicam através de sinapses, que consistem numa região onde dois neurônios entram em contato e através da qual os impulsos nervosos são transmitidos entre eles. Os impulsos recebidos por um neurônio não são imediatamente retransmitidos, mas sim processados e, caso atinja um dado limiar de ação (*threshold*) o neurônio produz uma substância neurotransmissora que flui do corpo central para o axônio, que pode estar conectado a um dendrito de um outro neurônio [1].

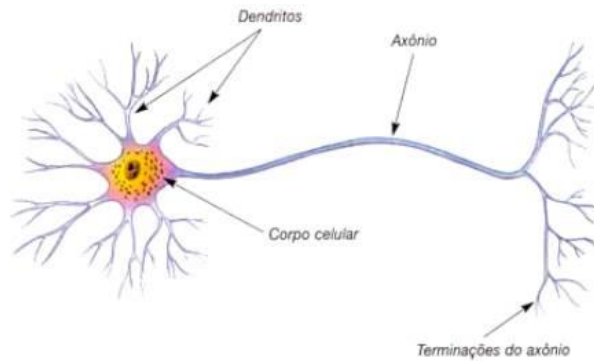


Figura 2.1. Estrutura do neurônio biológico. Fonte: [51]

2.3.2 Modelo matemático de um neurônio

As RNAs é um modelo matemático ou computacional que tenta simular a estrutura e os aspectos funcionais de uma rede neural biológica. Consiste num grupo interconectado de neurônios artificiais. Os primeiros estudos das RNAs são do neurofisiologista e psicólogo Warren McCulloch e do matemático Walter Pitts e foram desenvolvidos, originalmente, no ano de 1943. Esses cientistas apresentaram o primeiro modelo de neurônio artificial inspirado no neurônio orgânico; o neurônio artificial nada mais é do que uma estrutura lógico-matemática que simula o comportamento e as funções do biológico [1].

Neste modelo o nodo possui n terminais de entrada (x_1, x_2, \dots, x_n), por onde é alimentado com sinais provenientes das saídas de outros neurônio ou por informações de entrada da rede. Esses sinais ou informações ponderados por pesos fixos (w_1, w_2, \dots, w_n) das sinapses. Caso o somatório (Σ) dos produtos $x_i w_i$, apresente valor igual ou superior a um limiar μ (*threshold*), o sinal de saída (Y) do nodo recebe o valor igual a 1 (nodo excitado). Por outro lado, se o somatório apresenta resultado inferior ao limiar, a saída do nodo é zero [45],

Figura 2.2

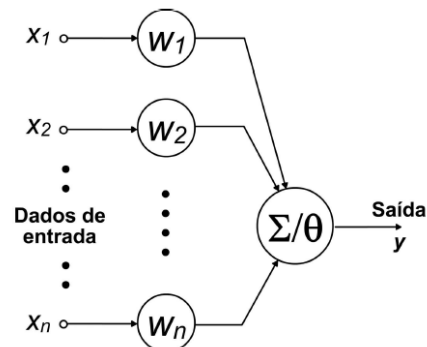


Figura 2.2 Representação de um neurônio artificial (Nodo) de acordo com a proposta apresentada por McCulloch e Pitts. Fonte: [45]

Em 1949 Donal Hebb, discutiu o aprendizado das redes biológicas e artificiais, observando que o aprendizado dos neurônios biológicos se dava pelo reforço das ligações sinápticas excitadas. Em 1958 Frank Rosemblat, propõe o modelo neural Perceptron de uma camada e demonstrou que o modelo poderia ser treinado para classificar alguns tipos de padrões. A arquitetura Perceptron Multicamadas(PMC) é a mais famosa RNA. Nela os neurônios são agrupados em camadas (uma camada de entrada, uma de saída e uma ou mais camadas escondidas) e existem apenas conexões para frente (*feedforward*) [7]. Em 1960 Widroff e Hoff propuseram o modelo *Adaline* (*Adaptive Linear Neuron*) que se diferenciavam dos modelos *perceptron* por possuírem saídas bipolares (-1 e 1), onde o ajuste dos erros era feito com método gradiente conhecido como regra delta. Em 1970 Minsky e Papert apontaram limitações nos modelos de redes *perceptron*, que seriam usadas somente para resolver problemas lineares. Em 1982, *Hopfield* demonstrou o algoritmo de treinamento *back propagation* que possibilitou redes múltiplas camadas solucionarem problemas não lineares [45].

A partir dessas características de um neurônio biológico pode-se definir um modelo matemático bem simples, formado por um conjunto de entradas, uma unidade de processamento e uma ou mais unidades de saídas, que comparando com a estrutura do neurônio biológico, correspondem, respectivamente, aos dendritos, corpo celular e os axônios.

A representação do neurônio matemático (neurônio artificial) é mostrada na Figura 2.3. Em regra, uma rede neural artificial é composta por camadas nas quais se define uma camada de entrada onde estão os neurônios de entrada; a camada de saída, no qual estão os neurônios de saída; e uma camada oculta onde encontram-se neurônios internos ou ocultos que recebem os sinais de saída dos neurônios da camada de entrada e após processá-los transmite-os através de suas saídas para a entrada dos neurônios da camada de saída. Porém, é muito comum a rede neural possuir mais de uma camada oculta. [1]

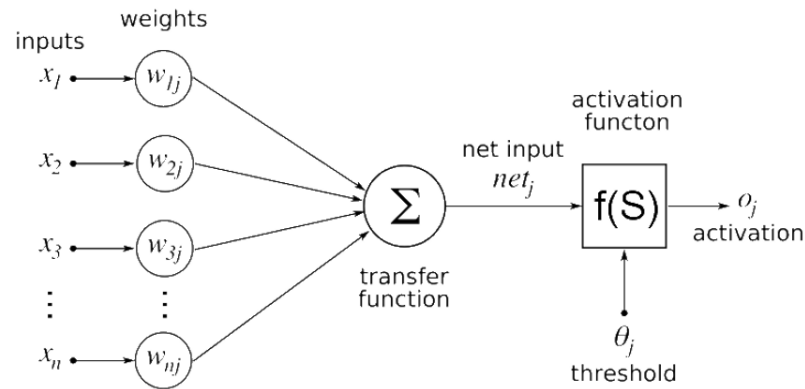


Figura 2.3. Modelo matemático de um neurônio. Fonte: [46]

Com o objetivo de simular um neurônio biológico, MacCulloch e Pitts (1943) propuseram o modelo de neurônio artificial, que representa o corpo do neurônio, com suas entradas E_i e sua saída Y . O corpo do neurônio é dividido em duas partes, sendo a primeira responsável por somar (Σ) o produto das entradas E_i pelos seus respectivos pesos W_i . Na segunda parte, tem-se a função de ativação ou transferência f , que irá controlar o valor a ser enviado pela saída Y . Além das entradas E , também se prevê uma entrada extra, chamada entrada de viés (bias), que sempre tem como entrada um valor 1. Esta entrada, não identificada nos neurônios biológicos, é opcional nas RNAs e tem se mostrado muito útil em várias situações (ARTERO, 2009). A primeira parte do processamento, geralmente, indicada com uma variável chamada net , como demonstrado na equação (1), em que X_i são entradas da rede, W_i são os pesos associados às entradas.

:

$$net = \sum_{i=0}^n x_i w_i \quad (1)$$

↑ entrada
↓ peso

2.3.3 Redes Feed-Forward

RNAs foram idealizadas como uma forma de criar uma representação matemática análoga ao que se acreditava ser a estrutura de um cérebro biológico. De forma que a estrutura possui unidades que são equivalentes a neurônios e às reações químicas

responsáveis por transportar informações entre os neurônios são representadas por funções de ativação. Cada unidade (ou neurônio) de uma rede neural possui um determinado valor, conhecido como peso. E esse peso é atribuído através de um processo conhecido como treinamento, em que vários exemplos são abastecidos à rede e “aprende” os melhores pesos em um processo de otimização, tentando ao máximo generalizar e errar o menos possível, quando exposta a exemplos desconhecidos. Inicialmente os dados a serem computados são passados às unidades de entrada, os valores são multiplicados pelos pesos destas e passados em sequência às unidades com as quais possuem conexões diretas. Redes neurais são organizadas em camadas, com cada camada possuindo uma ou mais unidades, e as saídas das unidades são combinadas, utilizando a função de ativação e usadas como entradas das unidades da camada seguinte. Redes neurais possuem um fluxo de informação unidirecional, não existindo nenhum tipo de ciclo, ou seja, um grafo acíclico como pode ser observado na Figura 2.4. Esse tipo de arquitetura é denominado como *feed-forward*, segundo [47].

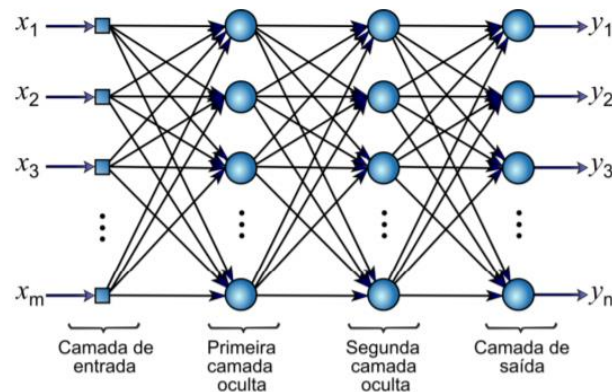


Figura 2.4 Arquitetura é do tipo Feed-forward. Fonte: [72]

2.3.4 Redes Recorrentes

Diferentemente das redes neurais feed-forward, redes neurais recorrentes possuem ciclos entre suas unidades. Em outras palavras, unidades podem ter conexões com unidades de camadas anteriores, ou da mesma camada, como vemos na Figura 2.5. Essa topologia de rede mais complexa é a recorrente, na qual existe ao menos uma conexão de retroalimentação, que tem um impacto significativo no desempenho e na capacidade de aprendizagem da rede, e fornece à rede um comportamento dinâmico não-linear [68]

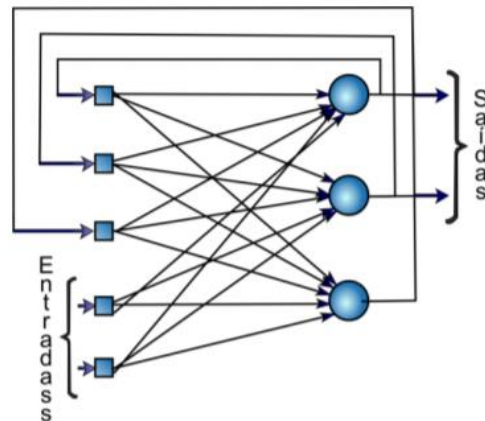


Figura 2.5 Arquitetura recorrentes. Fonte: [72]

As redes recorrentes possuem laços de realimentação, ou seja, a saída de um neurônio pode ser entrada para outro de uma camada precedente ou, no caso de auto-realimentação, para o próprio neurônio (PASCHOALINO, 2007). Desta forma, a informação não flui em um único sentido, e a saída da rede não depende mais apenas da entrada corrente, mas também das entradas anteriores. O efeito prático disso é a existência de memória de curto prazo na rede. Se consideramos o aprendizado por treinamento uma espécie de memória de longo prazo, então adicionamos à nova capacidade manter uma memória recente, redes neurais recorrentes, as quais podem criar modelos muito mais complexos, que apesar de serem de compreensão mais difícil apresentam capacidade de resolver uma gama maior de problemas [47].

As redes recorrentes, chamadas de redes com memória não possuem organização rígida e seus neurônios têm liberdade para se ligar a qualquer outro neurônio, são exemplos: redes de Elman, redes de Jordan, redes de Hopfield, NAR e NARX (PASCHOALINO, 2007). Dependendo do arranjo, as redes podem ser classificadas como total ou parcialmente recorrente Hopfield e as redes parcialmente recorrentes como a Elman e Jordan [68].

2.3.5 Redes Multilayer Perceptron (MLP)

As redes neurais MLP são as redes constituídas de uma ou mais camadas intermediárias ou escondidas, além da camada de entrada e da camada de saída, Figura 2.6. As camadas intermediárias fornecem à rede MLP a capacidade de resolver problemas não lineares, uma vez que os primeiros tipos de RNA, como as redes Perceptron e Adaline, por exemplo, não possuem esse poder de generalização. Tal diferença acontece devido ao fato das camadas

ocultas funcionarem como extratoras de características na qual os pesos apresentam uma codificação de tais características que são apresentadas nos padrões de entrada e essas características funcionam de modo a permitir que a rede neural crie sua própria representação, mais complexa e rica. Com efeito, a camada de saída tem como finalidade construir o padrão de saída a partir dos dados fornecidos pelas camadas intermediárias, escondidas ou ocultas.

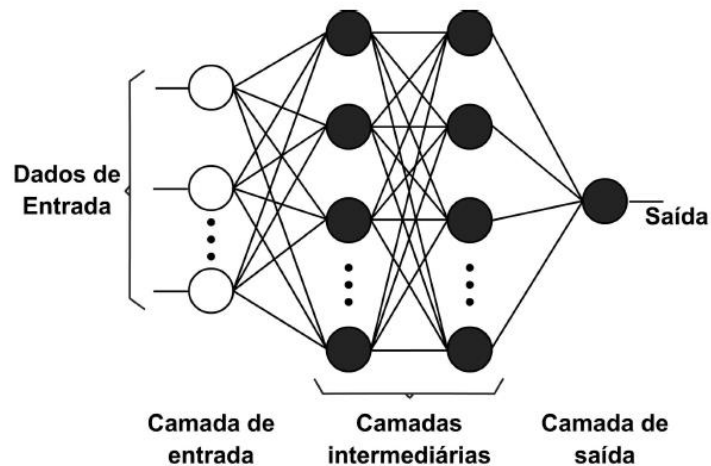


Figura 2.6 Representação gráfica de RNA do tipo MLP com duas camadas escondidas. Fonte: [45]

Uma rede neural MLP, possui três características principais: Os neurônios das camadas ocultas, possuem uma função de ativação não-linear do tipo sigmoide, função logística ou tangente hiperbólica e os neurônios da camada de saída possui uma função de ativação linear, Figura 2.7. A rede possui uma ou mais camadas escondidas e a rede possui um alto grau de conexão [46].

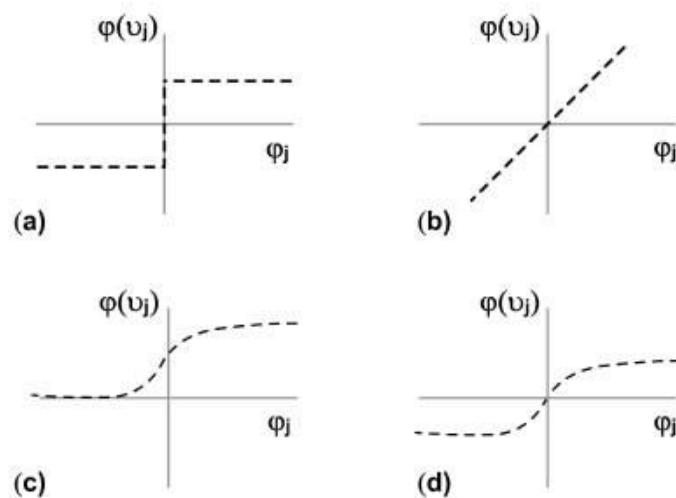


Figura 2.7. Funções de ativação. (a) função degrau. (b). função linear. (c) função logística. (d) função tangente hiperbólica. Fonte: [45]

Para função de transferência ou função de ativação f , existem várias possibilidades, sendo as mais comuns: função limiar, função linear, sigmóide e tangente hiperbólica como mostrado abaixo:

Função limiar

$$\varphi(v_j) = \begin{cases} +\gamma & \text{se } v_j \geq 0 \\ -\gamma & \text{se } v_j < 0 \end{cases} \quad (2)$$

Função linear

$$\varphi(V_j) = \alpha \cdot V_j \quad (3)$$

Sigmóide logística

$$\varphi(v_j) = \frac{1}{1 + \exp(-\alpha \cdot v_j)} \quad (4)$$

Sigmóide Tg.hip.:

$$\varphi(v_j) = \operatorname{tgh} \frac{v_j}{2} = \frac{1 - \exp(-v_j)}{1 + \exp(-v_j)} \quad (5)$$

2.3.6 Algoritmos de treinamento e aprendizagem

Os algoritmos de treinamento são usados baseados em dois paradigmas: os supervisionados e os não-supervisionados. No supervisionado é apresentada a rede os valores de entrada e os valores esperados, já no tipo não-supervisionados são apresentados apenas os valores de entrada.

Existem vários algoritmos computacionais para treinamento de redes neurais. Eles são geralmente do tipo supervisionado. Eles são assim chamados porque a entrada e a saída desejadas para a rede neural são dadas por um supervisor externo, cujo objetivo é ajustar os parâmetros para encontrar uma relação entre os pares de entrada e saída fornecidos [46]. O aprendizado supervisionado é o mais usado no treinamento de RNA. O que ocorre nessa fase de aprendizagem é que um conjunto de dados são apresentados para a rede, que ajusta seus pesos sinápticos de forma interativa até que extraia as características para representar as classes do problema. Após isso, esses valores ajustados são então usados para gerar soluções

de novos valores dados como entrada para a RNA.

O treinamento supervisionado compara os valores de saídas calculados com os valores esperados e avalia o erro, assim os pesos sinápticos são modificados de forma a minimizar o erro entre saída e entrada da RNA. Matematicamente, o termo $e(t)$ relativo ao erro é escrito como a diferença entre o valor $y(t)$ calculado pela RNA no instante t e o valor da saída esperado [45]:

$$\mathbf{d}(t) \quad e_j(t) = \mathbf{d}_j(t) - y_j(t) \quad (5)$$

A cada etapa de treinamento são feitas pequenas modificações nos pesos provocando uma minimização incremental dos erros, convergindo em direção do valor esperado. A forma genérica de alteração dos pesos é dada por:

$$\mathbf{W}_{j,i}(t + 1) = \mathbf{w}_{j,i}(t) + \mathbf{ne}(t)\mathbf{x}_i(t) \quad (6)$$

N é a taxa de aprendizado que determina a velocidade com que os pesos serão ajustados em direção ao menor erro e $x_i(t)$ é a entrada para o nodo i no tempo t [45].

2.3.6.1 Algoritmo Backpropagation

O algoritmo de aprendizado mais conhecido e utilizado para treinar RNAs é o algoritmo *backpropagation*. Seu método de aprendizado é do tipo supervisionado, que utiliza pares de entrada e saída desejados para ajustar os pesos da rede através de um mecanismo de correção de erros.

O treinamento da rede neural MLP utilizando o algoritmo *backpropagation* ocorre em duas fases: A fase *forward* e a fase *backward*. A primeira é utilizada para definir a saída a partir de um dado padrão de entrada. A segunda atualiza os pesos de suas conexões a partir da saída desejada e fornecida pela rede neural. A finalidade do algoritmo *backpropagation* é de minimizar o erro obtido pela rede neural, ajustando os pesos e limiares para que fiquem nos pontos mais inferiores das coordenadas da superfície de erro, utilizando o método de gradiente descendente, em que a função de ativação utilizada precisa ser contínua, diferenciável e não-decrescente e a mais utilizada é a *sigmoidal*, de acordo com [46].

O ajuste dos pesos é baseado no método do gradiente para minimização do erro, isto é,

a correção aplicada a cada peso do neurônio é dada pelo produto da taxa de aprendizado adotada e a derivada parcial da função custo em relação ao peso em questão.

$$\begin{aligned}\delta_j(t) &= e_j \cdot \varphi_j'(v_j(t)), \\ \Delta w_{ji}(t) &= \eta \cdot \delta_j(t) \cdot x_i(t).\end{aligned}\quad (6)$$

Ao aplicar a regra da cadeia, (Δw_{ij}) tornam-se dependentes do valor do gradiente local $\delta_j(t)$, onde o termo $\varphi_j'(v_j(t))$ é a derivada da função de ativação aplicada ao nível de atividade interna do neurônio, conforme [46].

Existem diversos algoritmos usados para o treinamento de retropropagação. Um deles é o algoritmo *Levenberg-Marquardt*, o mais encontrado na literatura computacional sobre redes neurais. O gradiente é calculado com uma matriz transposta dos erros de custos.

2.3.7 Métricas para Desempenho de rede

As métricas de desempenho são usadas para averiguar o quanto a rede neural consegue representar os dados aprendidos ao longo do treinamento e testes. Os mais usados são demonstrados a seguir, de acordo com o que descreve [53]:

- *Mean Squared Error – MSE*

Realiza a média dos erros do modelo ao quadrado, ou seja, a média do quadrado do desvio entre o valor de predição e o valor desejado. Assim o melhor modelo é o que tiver menos MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

- *Mean Absolute Error – MAE*

Realiza o valor absoluto dos erros, atribuindo um peso a todas as diferenças lineares:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

- *Mean Absolute Percentage Error – MAPE*

Dentre as diversas técnicas utilizadas para validar o desempenho obtido com RNAs,

podemos citar: comparação gráfica, diagramas de dispersão, raiz do erro médio quadrático, erro percentual médio absoluto (MAPE) entre outros. O modelo com menor valor MAPE indica o melhor ajuste [48]. Este erro calcula a média percentual do desvio absoluto entre as previsões e a realidade. Quanto maior for o MAPE mais incorreta é a predição, por outro lado, quanto menor mais correta será a predição.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

- *Coefficiente de Correlação de Pearson - R*

Corresponde a valores entre 0 e 1, o valor 1 significa correlação positiva entre duas variáveis e o valor 0 determina a ausência de correlação, assim como postula [46].

$$R = \sqrt{\frac{\sum_{i=1}^n (\hat{y}(t) - \bar{y}(t))^2}{\sum_{i=1}^n (y(t) - \bar{y}(t))^2}} \quad (10)$$

2.3.8 Predição de séries temporais com RNA

Série Temporal é um conjunto de observações sobre uma variável, ordenado no tempo e registrado em períodos regulares. Segundo o modelo clássico todas as séries temporais são compostas de quatro padrões: tendência (T), que é o comportamento de longo prazo da série. Ciclos (C), flutuações nos valores da variável com duração superior a um ano e que se repetem com certa periodicidade. Sazonalidade (S), flutuações nos valores da variável com duração inferior a um ano e que se repetem todos os anos, geralmente em função das estações do ano, como nos d (MENEZES, 2014). Estacionariedade (E) significa que os dados oscilam sobre uma média constante, independentemente do tempo, com a variância das flutuações permanecendo essencialmente a [64].

São métodos de previsões em que uma sequência de dados ordenado no tempo consiste em saídas geradas por um determinado sistema o qual se deseja analisar, podendo ser usado quando existem informações sobre o passado. Essa informação pode ser quantificada e ou assume algum padrão. Dado uma série temporal de instante de tempo t e um modelo a predição da série no tempo $t+h$ pode ser obtida, como mostrado na Figura 2.8. Essa predição é usada em áreas de economia, controle, saúde, entre outros (JÚNIOR, 2010).

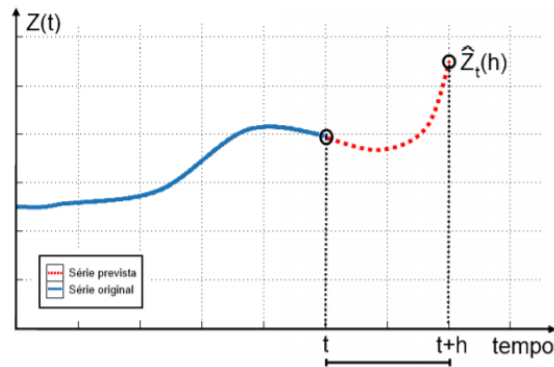


Figura 2.8 Predição ao longo do tempo. Fonte: [72]

Para esse tipo de série pode existir duas formas de prever valores futuros: predição de passo simples e predição de múltiplos passos. Na predição de passo simples a predição é realizada imediatamente após o atual, tendo considerado o horizonte de curto prazo. Na predição de múltiplos passos, um conjunto de valores é usado para determinar o valor futuro, tendo horizonte de médio e longo prazos, como afirma (JÚNIOR, 2010). De acordo com o autor, um modelo ajustado de Rede Neural é considerado não paramétrico, uma vez que não é necessário conhecer os parâmetros que geram a série e também são capazes de lidar com séries não lineares e não estacionárias (JÚNIOR, 2010).

2.4. Considerações Finais

Neste capítulo foram discutidos os principais conceitos encontrados na literatura sobre métodos computacionais usados em predição de epidemias e RNAs. Sendo abordadas definições elementares de arquitetura de rede neural, tipos de treinamento, entre outros. Obviamente, não se pretendeu esgotar o assunto ao longo do capítulo em questão. Para maiores detalhes acerca do assunto, sugere-se a leitura de [76], [77], [78] e [79] entre outros.

Capítulo 3 – Metodologia de monitoramento de epidemias: uma abordagem usando redes neurais artificiais

3.1 Considerações Iniciais

O objetivo deste capítulo é descrever a metodologia utilizada para analisar o uso de redes neurais na predição de epidemias de dengue nos municípios estudados: Belém, Parauapebas, Altamira e Santarém.

3.2 Metodologia desenvolvida

A metodologia foi desenvolvida nas seguintes etapas: (A) banco de dados, (B) pré-processamento, (C) o sistema de predição e alertas desenvolvido, (D) o modelo neural, (E) avaliação dos resultados e (F) as predições realizadas, conforme mostra a Figura 3.1 abaixo:

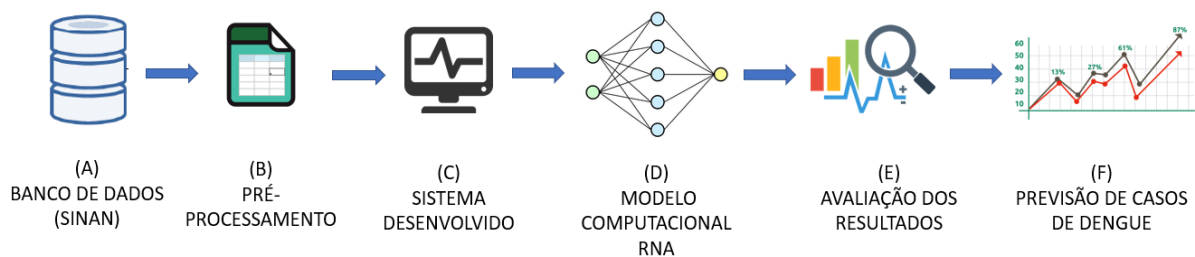


Figura 3.1. Metodologia para predição de novos casos de dengue.

Em (A) é realizada a busca das bases de dados públicas disponíveis no SINAN e são selecionadas as variáveis de interesse (que é a de casos notificados de dengue na região Amazônica). Em (B) consideram-se os aspectos de pré-processamento, com intuito de filtrar as informações pertinentes, para selecionar as variáveis usadas na pesquisa. Em (C) é desenvolvido um protótipo que implementa o modelo neural. Em (D) é avaliada a melhor configuração da rede neural usada e são detalhadas as etapas de treinamento, validação e testes do modelo desenvolvido. Em (E) são avaliados os resultados dos cenários de testes, com intuito de definir o melhor RNA para cada série temporal. E por fim em (F), os dados são processados e exibidos em gráficos relacionados a predição de novos casos e possíveis alertas sobre aumento para o período avaliado.

3.2.1 Área de estudo

O estudo foi realizado no estado do Pará, região norte do Brasil. Com população estimada em 2017 de 8.366.628 hab, com área de 1.247.955 km² e 144 municípios, Figura 3.2. [83].

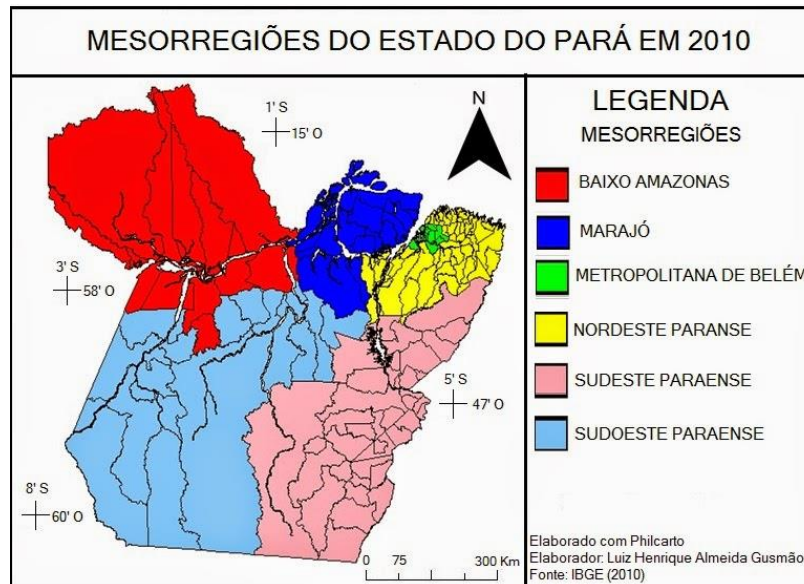


Figura 3.2. Região Metropolitana de Belém-Pa-Brasil. Fonte: [83]

3.2.2 Dados coletados

Esse trabalho utilizou duas bases de dados públicas para estudo: casos de dengue e índice pluviométrico. Também foi usado técnicas de pré-processamento para escolha dos dados estudados segundo o que propõe [58]. Essa etapa possui fundamental importância no processo de descoberta de conhecimento permitindo organizar as funções de pré-processamento em de acordo com: extração de dados, seleção de atributos, limpeza dos dados, discretização, transformação de variáveis.

A primeira base de dados usada foi proveniente de dados públicos disponível no Portal SINAN, que um órgão responsável por disponibilizar informações relacionados a saúde pública, vinculado ao Ministério da Saúde. A escolha da base de dados se justifica por conter uma série histórica sobre casos de dengue no país, mostrando informações de casos confirmados de dengue em todas as cidades do estado do Pará, extraindo assim os municípios para estudo.

Para dados relacionados com índices pluviométricos usou-se dados provenientes de Instituto Nacional de Meteorologia (INMET) [74]. A série temporal é multivalorada e esta

relacionada a dados climáticos de 10 anos (2006 à 2016). As variáveis atmosféricas disponibilizadas para consultas no Banco de Dados para Apoiar as Atividades de Ensino e Pesquisa (BDMEP) são: precipitação ocorrida nas últimas 24 horas; temperatura do bulbo seco; temperatura do bulbo úmido; temperatura máxima; temperatura mínima; umidade relativa do ar; pressão atmosférica ao nível da estação; insolação; direção e velocidade do vento. Em seguida filtrou-se somente a variável relacionada a precipitação do estado do Pará.

Na primeira etapa, os dados foram extraídos do SINAM usando o framework *Jupyter Notebook Python*, apresentado em [62] que permite selecionar e visualizar os dados pertinentes para estudo. Essa extração gerou um *dataset* contendo a série histórica de notificação de casos de dengue e índices pluviométricos desses municípios de interesse da pesquisa.

3.2.3 Desenvolvimento do Sistema de monitoramento e alertas de epidemias

Para realizar os experimentos da pesquisa foi desenvolvido um protótipo de sistema que implementa a rede neural artificial proposta anteriormente. De modo que seja capaz de avaliar, em condições reais, o desempenho do modelo computacional proposto.

3.2.3.1 Arquitetura do sistema de monitoramento e alertas

Foi desenvolvido um protótipo de monitoramento e alerta contendo uma arquitetura dividida em três módulos: um módulo para importar dados públicos, um módulo que implementa o modelo de ANN e um módulo que emite alertas sobre o avanço da epidemia de dengue, informando assim os órgãos públicos de saúde responsáveis, conforme arquitetura apresentada na Figura 3.3.

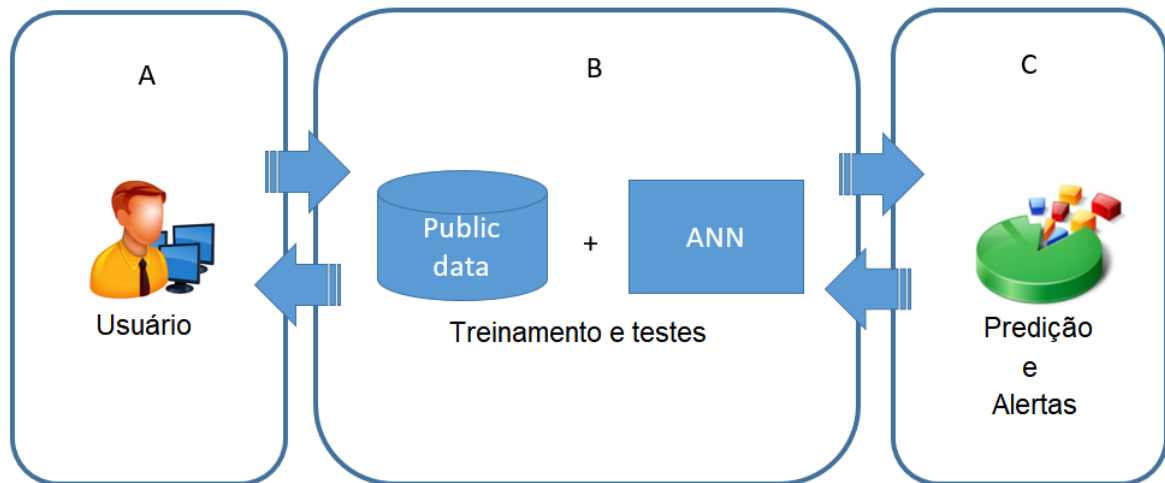


Figura 3.3 Proposta de arquitetura do sistema. Em A o módulo para importação das bases públicas. Em B o modelo ajustado de RNAs para prever novos casos. Em C os gráficos e alertas do sistema com novos avanços da doença.

Em A, o usuário importa as bases públicas de diversas fontes: município, estado ou federação. Tais dados são disponibilizados em bases públicas em diversos formatos, essa arquitetura usa extensão CSV (*Comma Separated Values*).

Em B, o módulo de predição usa um modelo de ANN ajustado anteriormente. Esse modelo é capaz de receber como entradas para os neurônios, o histórico de casos mensais das epidemias, aprender o comportamento da série temporal e realizar a predição de novos casos de dengue.

Em C, são mostradas as previsões de novos casos. Essa predição informa aos gestores de saúde, sobre o aumento de novos casos, emitindo alertas sobre possíveis avanços das doenças; podendo usar tais informações para tomada de decisão e ações de controle das epidemias em seu município.

3.2.3.2 Modelagem do sistema – Caso de uso

Na modelagem de caso de uso percebe-se que o usuário pode realizar diversas operações no sistema, dentre elas: importar dados, pré-processados, treinar o modelo de rede neural com a melhor configuração, realizar a predição de novos casos e visualizar os alertas quando o número de casos alcançar um valor médio e gerar gráficos, como demonstra a Figura 3.4.

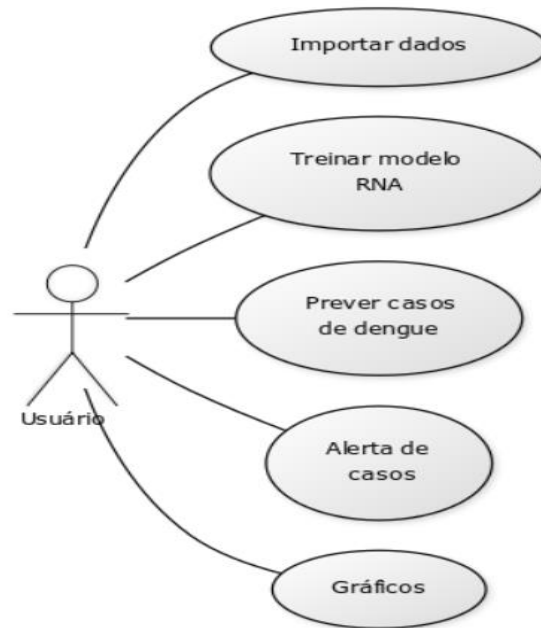


Figura 3.4 Caso de uso do sistema desenvolvido.

3.2.3.3 Modelagem do sistema – Diagrama de classe

O diagrama de classe do sistema proposto tem por objetivo mostrar o conjunto de classes e seus relacionamentos. Observa-se que a classe principal System se relaciona com outras classes: a classe importar tem a função de importar dados, a classe modelo RNA cuja finalidade é treinar e testar o modelo de rede neural, a classe gráficos que exibe as previsões do sistema e a classe Alertas que exibe alertas quando o sistema detecta aumento de casos de dengue previsto, como apresentamos na Figura 3.5.

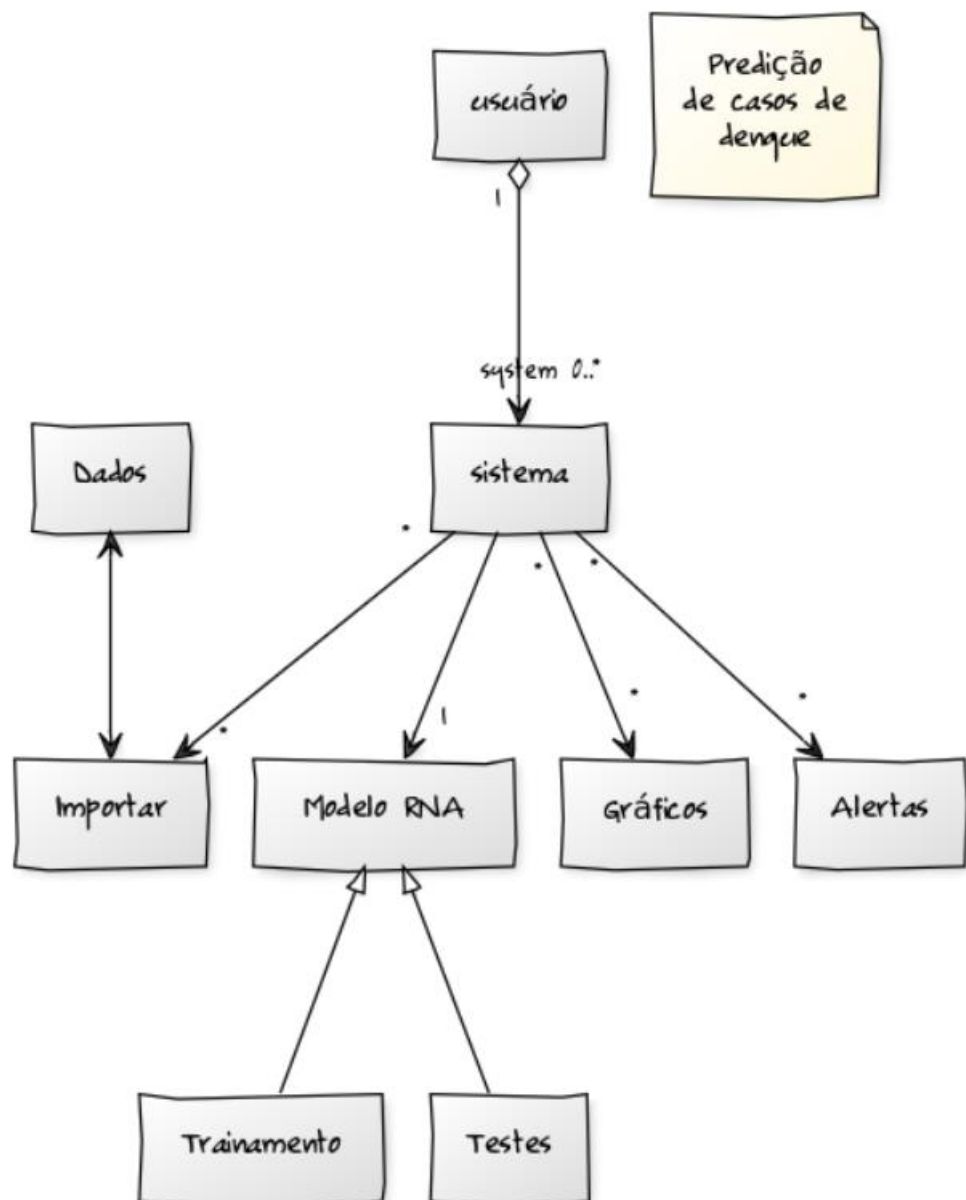


Figura 3.5 Diagrama de classe sistema de monitoramento e alertas de dengue.

3.2.3.4 Modelagem do sistema – Diagrama de atividade

O diagrama de atividade desenvolvido mostra, como exemplo, uma etapa de ajuste do modelo de RNA aplicados aos dados. Se o modelo já foi ajustado ele é testado com os dados novos, caso o modelo ainda não tenha sido ajustado é necessário realizar o treinamento e só em seguida testar o modelo, que observamos na Figura 3.6.

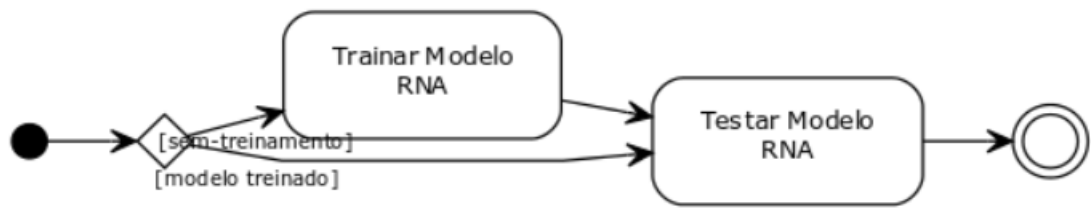


Figura 3.6 Diagrama de Atividade sistema desenvolvido para o treinamento do modelo ANN.

O sistema foi desenvolvido usando o *Matlab R2015a* [81], do tipo *Graphical User Interface (GUI)*, como pode ser visualizado na Figura 3.7:

3.2.3.5 Protótipo do sistema – Interface do sistema desenvolvido

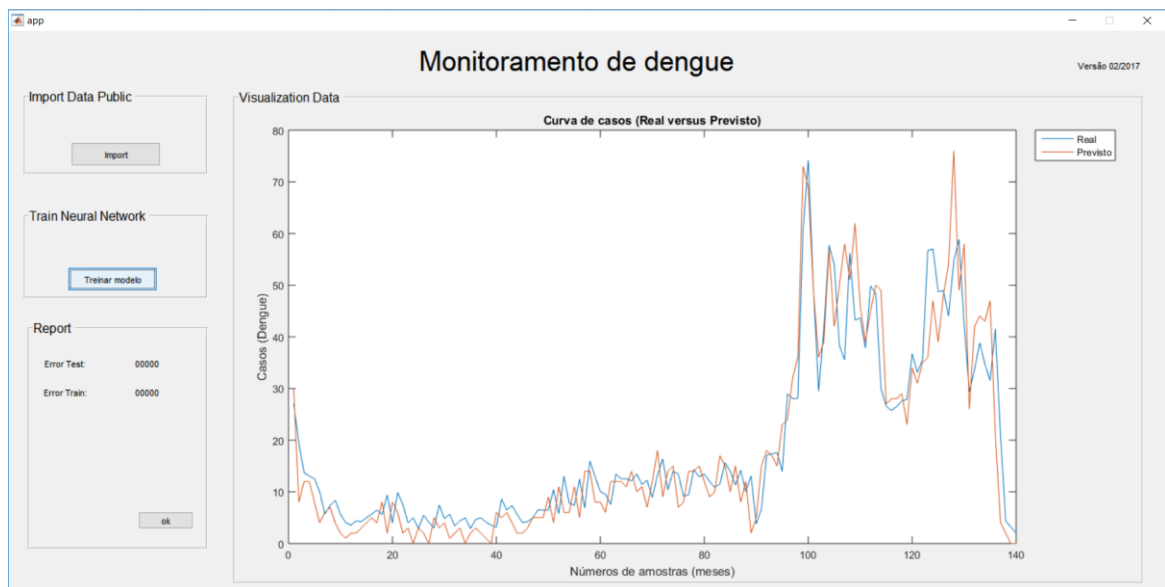


Figura 3.7 Interface sistema de monitoramento e alertas de dengue

O programa foi utilizado por ser familiar no meio acadêmico e ser de fácil aprendizagem. É um programa de fácil manuseio e atualizações futuras dos sistemas.

3.2.3.6 Módulo de alertas

Foi implementado um módulo de alerta baseado na quantidade de casos semanais previstos na fase de testes. Ao considerar a distribuição semanal desses novos casos, quando o número extrapolar o valor determinado, que em nosso estudo foi de 30 casos previstos, o sistema dispara um alerta, como podemos observar na Figura 3.8. O algoritmo

realiza a comparação do valor previsto pela RNA com o limiar, caso seja maior o alerta é emitido. Uma caixa de mensagem informa o usuário sobre o aumento de casos na predição.

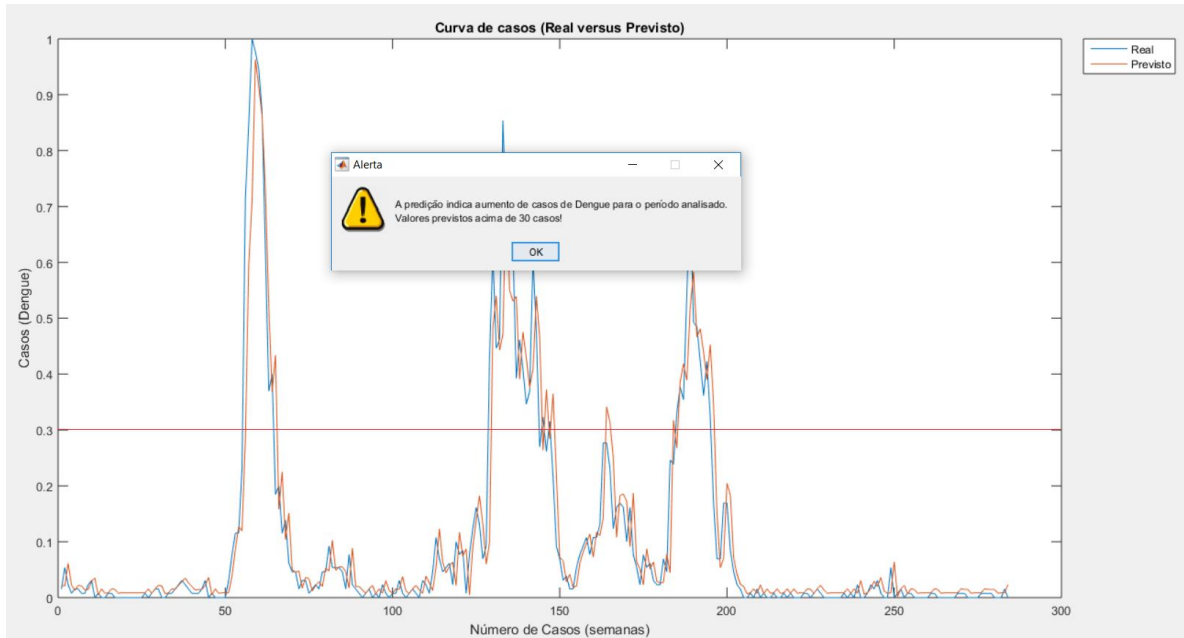


Figura 3.8 Visualização do alerta para novos casos previstos no sistema.

3.2.4 Modelo da Rede Neural Artificial

Nessa etapa usou-se a estratégia de observar o menor erro de predição de dengue, com intuito de escolher a melhor configuração da rede neural, como pode ser observado na Figura 3.9.:

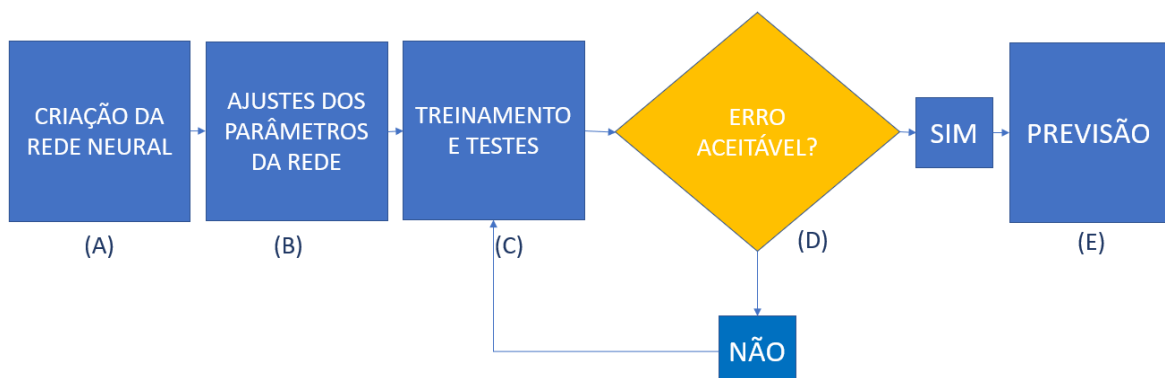


Figura 3.9 Estratégia de escolha da melhor configuração da RNA

Conforme visto na figura 3.9, as etapas de escolha da melhor arquitetura: (A)

inicialmente a rede neural foi criada de forma aleatória, em seguida, (B) variou-se então os ajustes dos parâmetros da rede, (C) realizou-se então o treinamento e testes do modelo, (D) observou-se a saída do erro da rede, caso não seja aceitável o modelo retorna a etapa anterior, caso contrário, sendo aceitável segue então para predição de novos casos de dengue (E).

As etapas foram definidas, como propõe [72], em 70% para treinamento, 15% para validação e 15% para testes. Na fase de treinamento os dados são apresentados para rede e ajustados de acordo com o erro. A fase de validação é usada para medir a generalização das redes e interromper o treinamento quando a rede não consegue melhorar o aprendizado e, por fim, na fase de testes, novos dados são apresentados para rede, afim de verificar a performance para realizar previsões.

3.2.4.1 Definição dos parâmetros da RNA

Durante essa etapa diversos parâmetros foram ajustados: janela de tempo, número de neurônios das camadas intermediárias (oculta). Todas as arquiteturas foram testadas com a mesma configuração: para camadas de entrada e escondida foi usada a função de ativação tangente hiperbólica função *tansig* (*Matlab*) e para camada de saída a função linear *purelin* (*Matlab*); para o critério de parada do algoritmo foi considerado o tempo máximo de épocas determinadas, que nesse caso foi de 1000.

Foi considerado, o uso de janela de tempo ($t-1$), em que para prever a próxima semana t ($t+1$) se faz uso da semana atual t , bem como os dois tempos anteriores $t-1$, $t-2$. Nesse trabalho foram variadas as janelas de tempos ($t-n$), que representam a quantidade de valores anteriores usados para prever a próxima semana.

3.2.4.2 Entrada da RNA

Durante essa etapa foram identificadas e definidas as variáveis usadas na predição para as séries históricas semanais de cada município estudado. Os municípios forneceram uma quantidade de dados diferente para a camada de entrada da RNA.

Os valores de entrada foram normalizados entre 0 e 1. Essa etapa tem como objetivo adaptar os dados de entrada à faixas das funções de ativação (camada intermediária) e na função linear (camada de saída). Os valores de normalização foram usados inicialmente ao

conjunto de treinamento, em seguida estendido ao conjunto de teste.

3.2.4.3 Treinamento, validação e testes da RNA

Conforme (Conde, 2007 et al), a rede neural usada é do tipo MLP (Multilayer Perceptron) e o algoritmo de treinamento é o *Levenberg-Marquardt*. Durante essa etapa o número de neurônios da camada escondida, a taxa de aprendizagem da camada escondida e a taxa de aprendizado da camada de saída são atualizados constantemente, criando assim uma MLP para cada série temporal (uma para cada município) apresentada à RNA.

Conforme afirma [81] para cada tipo de problema de predição de séries temporais dinâmicas, foi usado a rede neural artificial do tipo Auto regressiva não linear (*NOR*), a qual prever séries $y(t)$ dados d valores passados de $y(t)$, cujo algoritmo conhecido é o *NARNET*. A retroalimentação possibilita o processo dinâmico de informação, o que a torna recomendável em aplicações de problemas variantes com o tempo, como predição de séries temporais. Um exemplo desse tipo de redes retroalimentação, como propõe [72].

Foram definidos 02 (dois) experimentos para estudo: o experimento I usa o mesmo modelo MLP, treinado a partir da série do município de Belém para prever os demais municípios; e o experimento II com o intuito de melhorar a predição, repete o processo de treinamento para cada uma das quatro séries apresentadas. Dessa vez, variando-se as camadas e quantidade de neurônios, escolhendo assim, a melhor MLP para cada série. Ao final dessa etapa de treinamento, tem-se 04 (quatro) melhores MLPs ajustadas para cada série temporal.

3.2.4.4 Avaliação da RNA

Par avaliação do modelo computacional utilizou-se nessa pesquisa o erro médio de porcentagem absoluta (MAPE) [49], para validar a capacidade de predição do modelo, já apresentado na capítulo 2.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (11)$$

3.3 Considerações finais

Esse capítulo demonstrou a metodologia desenvolvida para predição de epidemias na Amazônia, o desenvolvimento de um sistema de monitoramento e alertas de epidemias, os parâmetros ajustados e o tipo no modelo neural utilizado na pesquisa.

Capítulo 4 - Estudo de Caso: Predição de ocorrência de dengue no estado do Pará

4.1 Considerações iniciais

Neste capítulo serão apresentados os resultados de desempenho dos modelos preditivos para casos de dengue. Usou-se o sistema de monitoramento e alertas de dengue apresentado anteriormente para os experimentos aqui demonstrados. Na primeira etapa, os dados foram analisados estatisticamente e em seguida foram realizados 2 (dois) experimentos com intuito de encontrar o melhor modelo RNA para predição em cada município estudado. Logo em seguida é realizado a predição de novos casos, usando 2 (dois) cenários que apresentam alto índice pluviométrico (primeiro semestre), baixo índice pluviométrico (segundo semestre).

4.2 Dados Analisados

Foram selecionados os 04 (quatro) municípios com maiores incidências de casos de dengue no estado do Pará, entre os anos de 2006 a 2016, como demonstra a tabela 4.1. Percebe-se que o município de Belém tem a maior quantidade de casos de dengue notificados, seguido dos municípios Parauapebas, Santarém e Altamira, como observa-se na Figura 4.1.

Tabela 4.1. Lista dos municípios do estado. Fonte: [82]

Casos notificados de dengue por ano nos 04 municípios												
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	TOTAL
Belém	1092	2800	1221	1427	3064	2422	1957	457	475	1459	758	17.132
Parauapebas	54	644	1324	203	158	921	1292	743	520	449	327	6.635
Altamira	5	8	13	409	1342	1340	973	272	244	293	106	5.005
Santarém	42	66	1140	103	445	1267	479	830	18	36	12	4.438

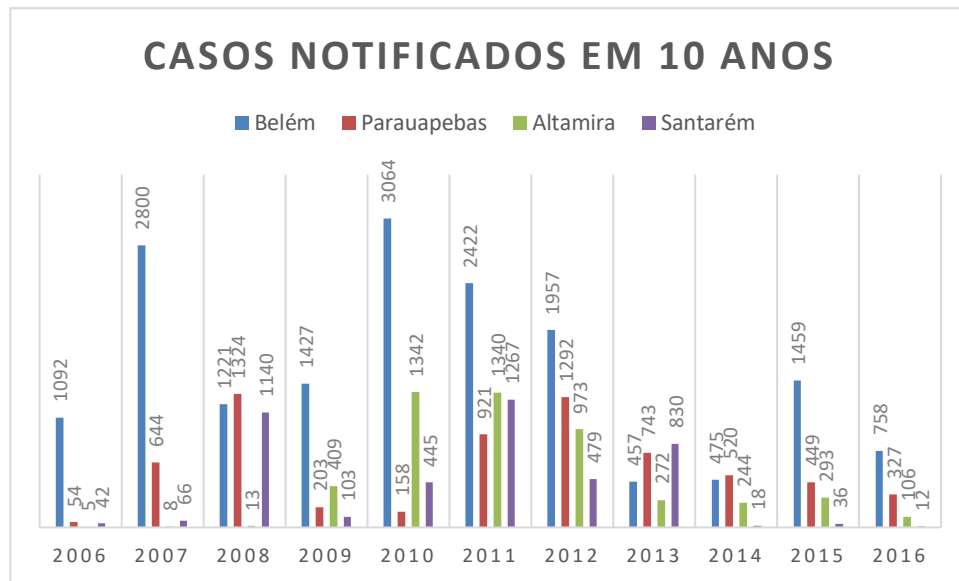


Figura 4.1 Lista dos municípios com maior ocorrência de dengue no estado. Fonte: [82]

4.2.1 Série Histórica dos Municípios

Para realizar os experimentos o modelo neural usou a série histórica de 10 (dez) anos. Essas séries têm como referência 562 semanas no período de janeiro de 2006 a dezembro de 2016. São apresentadas as séries temporais de cada município estudado, conforme mostra a Figura 4.2, 4.3, 4.4 e 4.5:

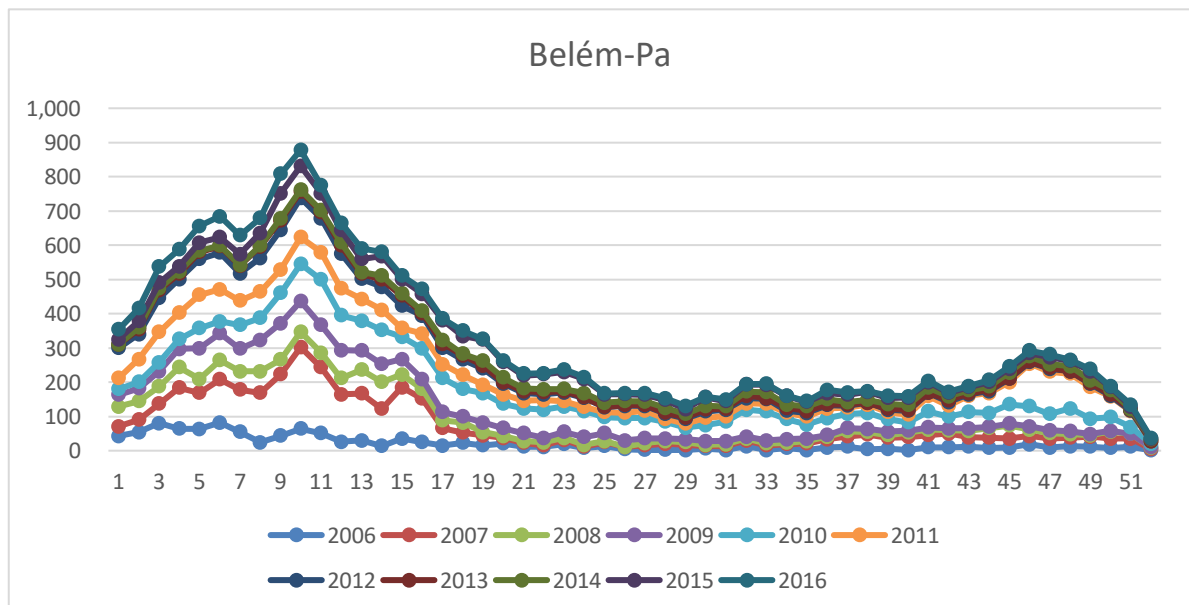


Figura 4.2 Casos notificados de dengue na região Metropolitana de Belém-Pa.

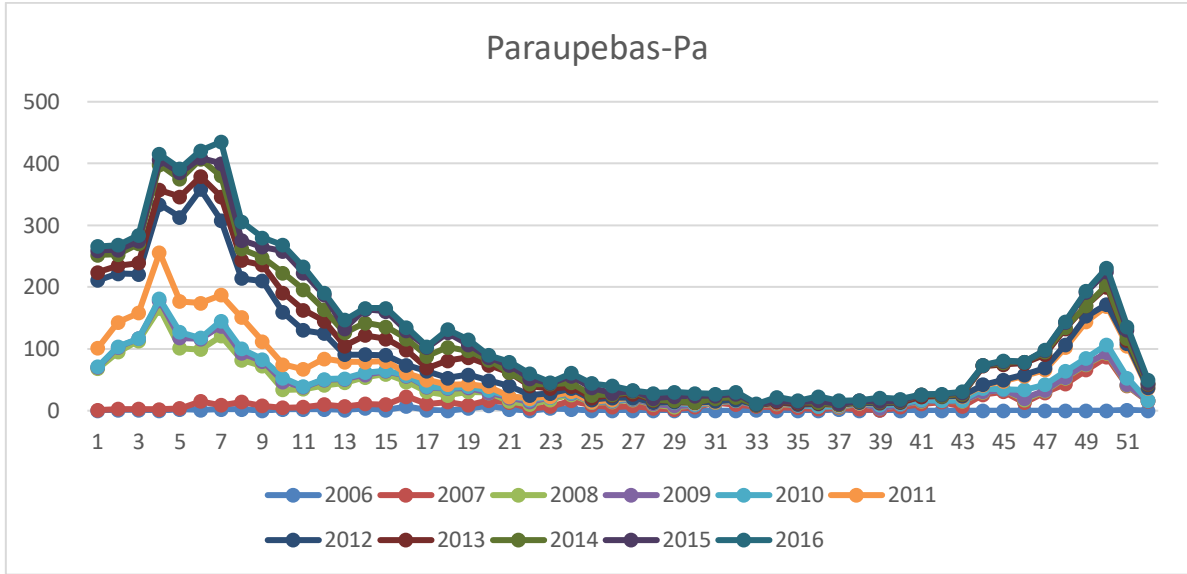


Figura 4.3 Casos notificados de dengue do município de Paraupibas-Pa

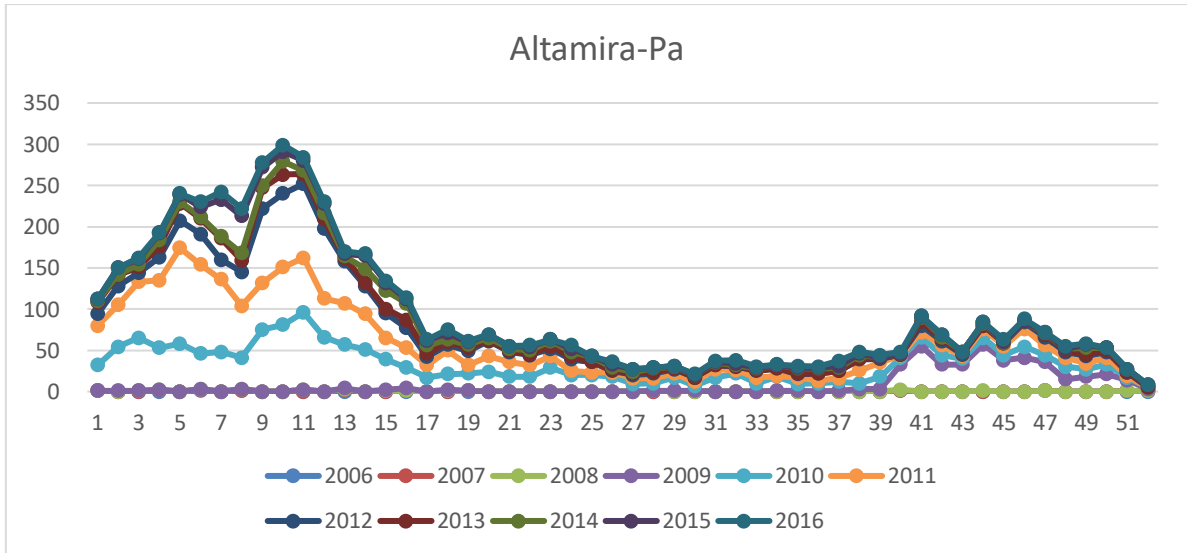


Figura 4.4. Casos notificados de dengue do município de Santarém-Pa

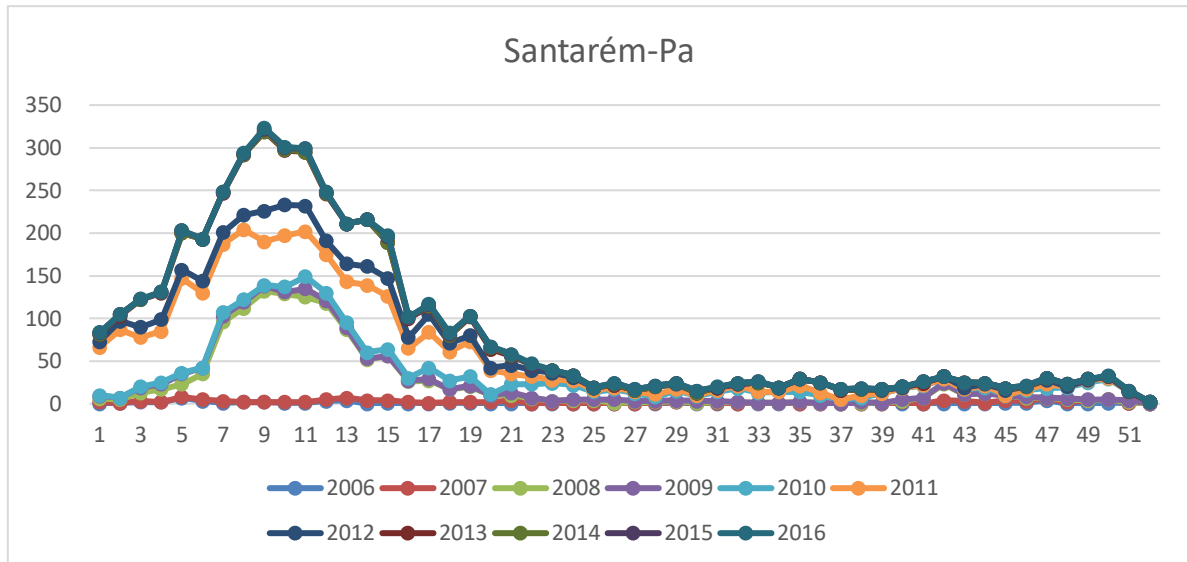


Figura 4.5 Casos notificados de dengue do município de Altamira-Pa

Percebe-se que todas as séries dos municípios apresentam crescimento no primeiro semestre de quase todos os anos. Esse fato está relacionado a outros fatores climáticos como precipitação, segundo afirma [69]. A densidade larvária de *Aedes aegypti* flutua de acordo com as variações climáticas sazonais, elevando-se nas estações de maior pluviosidade, em função do número de potenciais criadouros disponíveis, o que predispõe ao aumento da incidência de dengue, segundo o que nos diz (SOUSA, 2010). Considerando essas postulações, percebeu-se que as médias dos casos notificados e os índices pluviométricos tem correlação conforme a tabela 4.2 e a Figura 4.6.

Tabela 4.2 Correlação de Casos de dengue x Índice pluviométrico

MUNICÍPIOS	CORRELAÇÃO R
Belém	0,9413333
Parauapebas	0,8238194
Altamira	0,8712455
Santarém	0,9063295

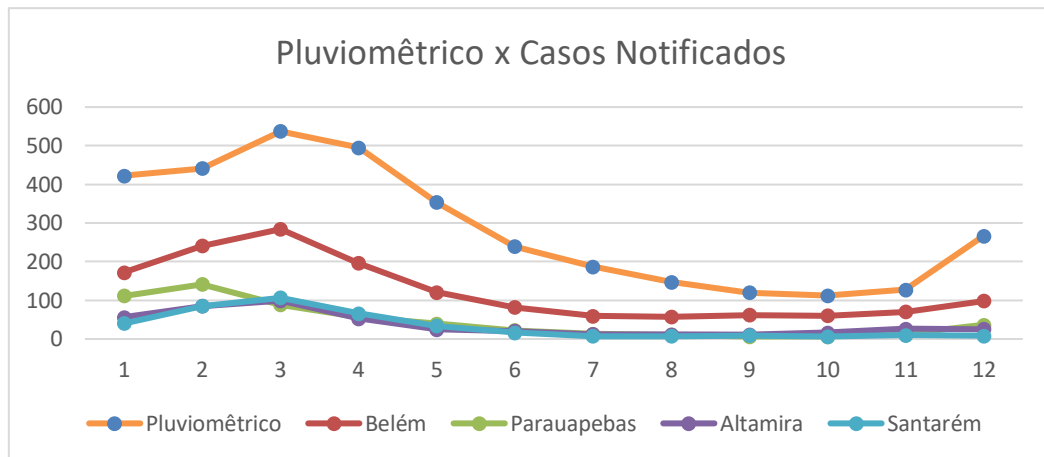


Figura 4.6 Índice pluviométrico x casos de dengue no período de 2006 a 2016. Fonte: do Instituto Nacional de Meteorologia [74]

A tabela acima mostra a correlação positiva entre em todas as séries analisadas, o que demonstra que os casos de dengue notificados seguem a curva dos índices pluviométricos. Observa-se também que há um aumento de casos de dengue nos meses em que se registram os maiores índices pluviométricos, o que é percebido no primeiro semestre de cada ano em cada município estudado. Ao longo do segundo semestre ocorre a diminuição dos casos de dengue acompanhando a curva de índices pluviométricos. Essa correlação será usada como instrumento de predição de novos casos.

4.2.2 Estatística descritiva dos municípios

A estatística descritiva é a etapa inicial da análise, a qual é utilizada para descrever e resumir os dados. A disponibilidade de uma grande quantidade de dados e de métodos estatísticos adequados a sua compreensão, de acordo com o que nos afirma [59] descreve e compreende os dados de uma distribuição usando medidas de tendência central (média, mediana e moda), medidas de dispersão (valores mínimo e máximo, desvio padrão e variância), percentis, quartis, e medidas de distribuição (achatamento e simetria da curva de distribuição).

Como nos diz [60] realizou-se a estatística descritiva dos dados relacionados a pesquisa, no período dos 10 (dez) anos selecionados para a análise, que resultou em 567 amostras, como mostra a tabela 3.16 e figura 4.7.

Tabela 4.3 Estatística descritiva dos dados

	Belém	Parauapebas	Altamira	Santarém
Tamanho da amostra =	567	567	567	567
Mínimo	0.0000	0.0000	0.0000	0.0000
Máximo	238.0000	184.0000	116.0000	238.0000
Amplitude Total	238.0000	184.0000	116.0000	238.0000
Mediana	17.0000	4.0000	2.0000	16.0000
Primeiro quartil (25%)	8.0000	1.0000	0.0000	7.0000
Terceiro Quartil (75%)	44.0000	12.0000	9.0000	44.0000
Desvio Interquartilico	36.0000	11.0000	9.0000	37.0000
Média Aritmética	29.6578	11.6631	8.8166	29.1235
Variância	1051.2926	443.4535	294.7967	1048.7338
Desvio Padrão	32.4236	21.0583	17.1696	32.3842
Erro Padrão	1.3617	0.8844	0.7211	1.3600
Coefficiente de Variação	109.33%	180.55%	194.74%	111.20%
Assimetria (g1)	2.0062	3.8714	3.1685	1.9973
Curtose (g2)	5.4660	19.5109	11.1731	5.5195
Média Harmônica =	8.7599	3.0509	2.9638	7.9460
N (média harmônica) =	560	473	370	546
Média geométrica =	16.9402	6.0672	5.8840	16.7125
N (média geométrica) =	560	473	370	546
Variância (geom.) =	1.7669	2.0673	2.0870	1.8511
Desvio Padrão (geom.) =	3.1419	3.6443	3.6750	3.2895

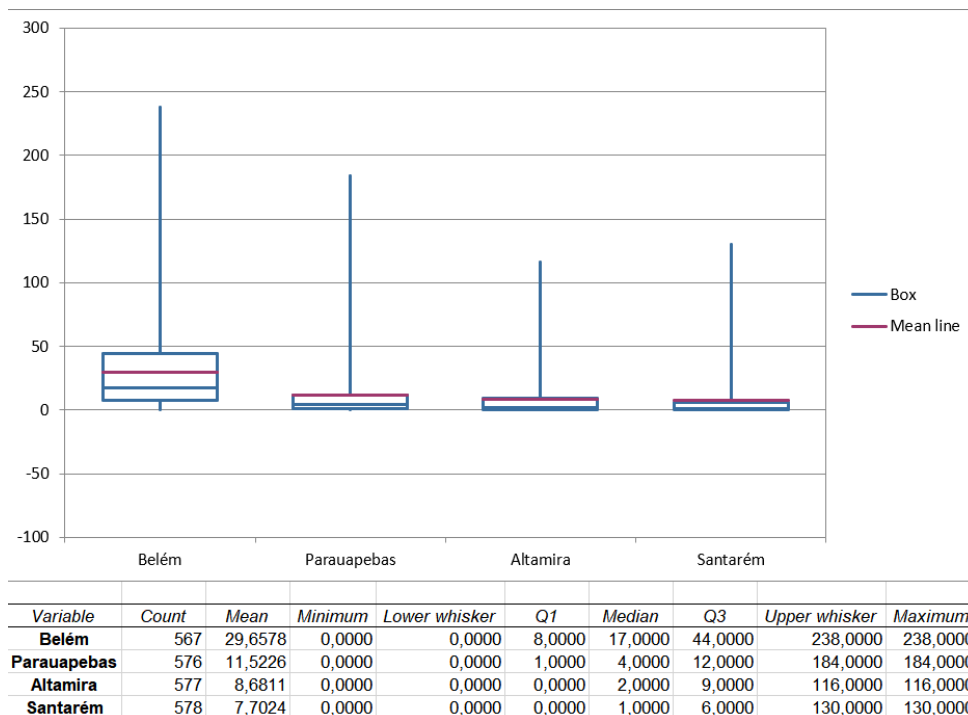


Figura 4.7 Box-plot dos dados

4.2.3 Distribuição dos dados

Dada a série histórica dos casos de dengue é interessante visualizar a histogramas das

séries, com intuito de entender como os mesmos estão distribuídos. Os histogramas relevam como os dados estão agrupadas em relação a frequência absoluta.

A série de Belém mantém uma quantidade de dados agrupados entre os valores 0 a 50 casos de dengue por 400 instancias analisadas, Figura 4.8

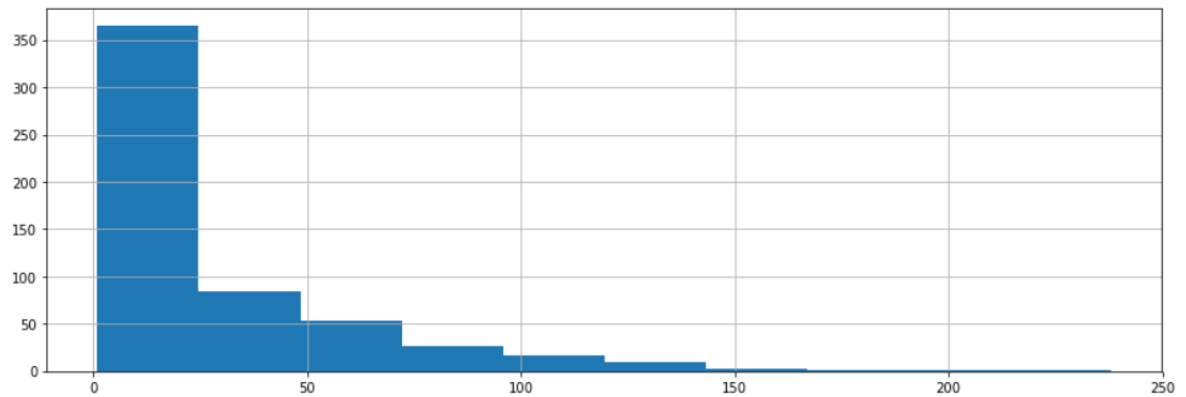


Figura 4.8 Distribuição dos casos no município de Belém-Pa

A série de Parauapebas mantém uma quantidade de casos agrupados entre 0 a 25, conforme Figura 4.9:

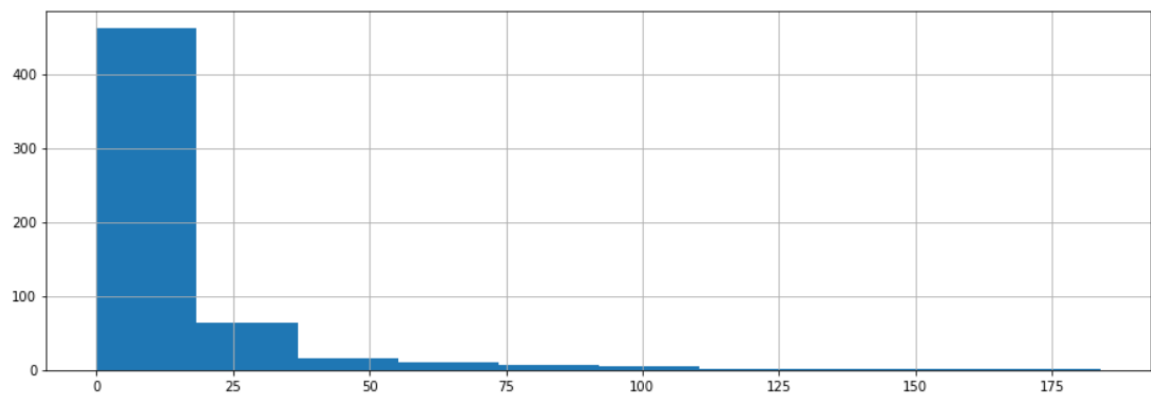


Figura 4.9 Distribuição dos casos no município de Parauapebas-Pa

A série de Altamira mantém uma quantidade de casos de casos agrupados entre 0 a 20, conforme Figura 4.10:

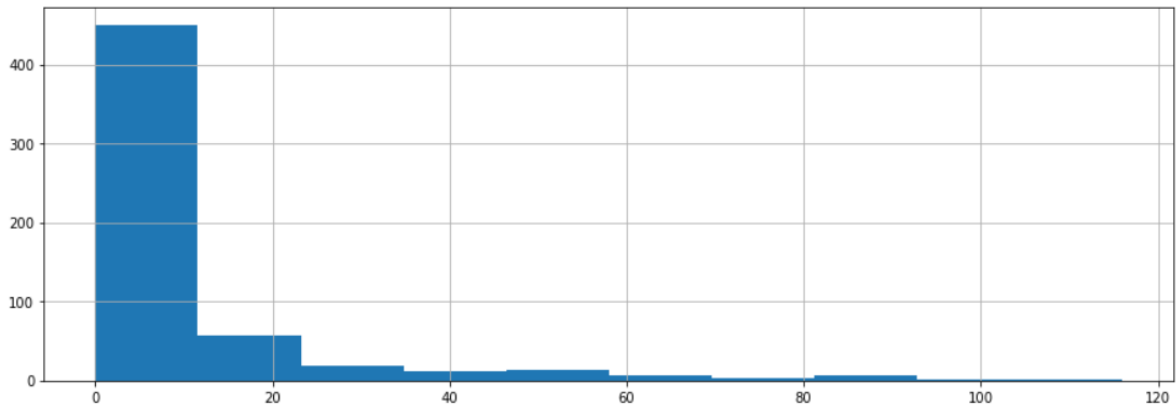


Figura 4.10 Distribuição dos casos no município de Altamira-Pa

A série de Santarém mantém os casos agrupadas entre 0 a 10. Conforme Figura 4.11:

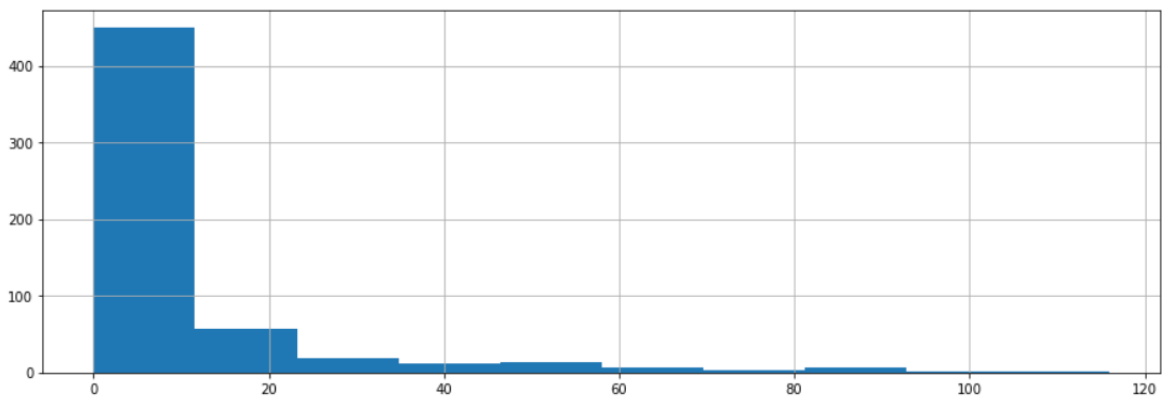


Figura 4.11 Distribuição dos casos no município de Santarém-Pa

4.2.4 Séries Temporais das amostras

Os dados foram avaliados usando a ferramenta computacional [61] e *Gretl* 2017d. Foram definidos métodos estatísticos para avaliar, de forma individual, se as séries temporais dos municípios estudados apresentam comportamento estacionário ou não-estacionário, tendência e/ou sazonalidade. O que implica na correta interpretação dos dados, metodologia empregada e resultados da pesquisa.

Para comportamento estacionário foi usado o teste de *Dickey-Fuller Aumentado*, que, na prática testa a hipótese do modelo possuir ou não uma raiz unitária, o que aponta estacionaridade ou não-estacionaridade da série, segundo [10]. De acordo com a tabela 4.4 o $pvalor=0$ implica em séries não-estacionárias. Essa afirmação é válida para todas as séries de todos os municípios.

Tabela 4.4. Testes de Estacionaridade

<i>Teste de Dickey-Fuller Aumentado</i>	
<i>BELÉM</i>	
Estatística	-4,74139175
PValor	0,01
Tamanho da Amostra	567
Hipótese Nula	Existe pelo menos uma raiz unitária
Hipótese Alternativa	Não existe raiz unitária
<i>SANTARÉM</i>	
Estatística	-4,894895323
PValor	0,01
Tamanho da Amostra	567
Hipótese Nula	Existe pelo menos uma raiz unitária
Hipótese Alternativa	Não existe raiz unitária
<i>ALTAMIRA</i>	
Estatística	-5,136822185
PValor	0,01
Tamanho da Amostra	567
Hipótese Nula	Existe pelo menos uma raiz unitária
Hipótese Alternativa	Não existe raiz unitária
<i>PARAUPEBAS</i>	
Estatística	-5,738590362
PValor	0,01
Tamanho da Amostra	567
Hipótese Nula	Existe pelo menos uma raiz unitária
Hipótese Alternativa	Não existe raiz unitária

Também é possível verificar estacionaridade ou não-estacionaridade das séries usando a função de auto correlação (FAC) das séries, como afirma [66] A Figura 4.12 indica que as séries apresentam comportamento não estacionário por apresentarem decaimento lento de auto correlação. O comportamento senoidal ao longo da correlação indica sazonalidade das séries.

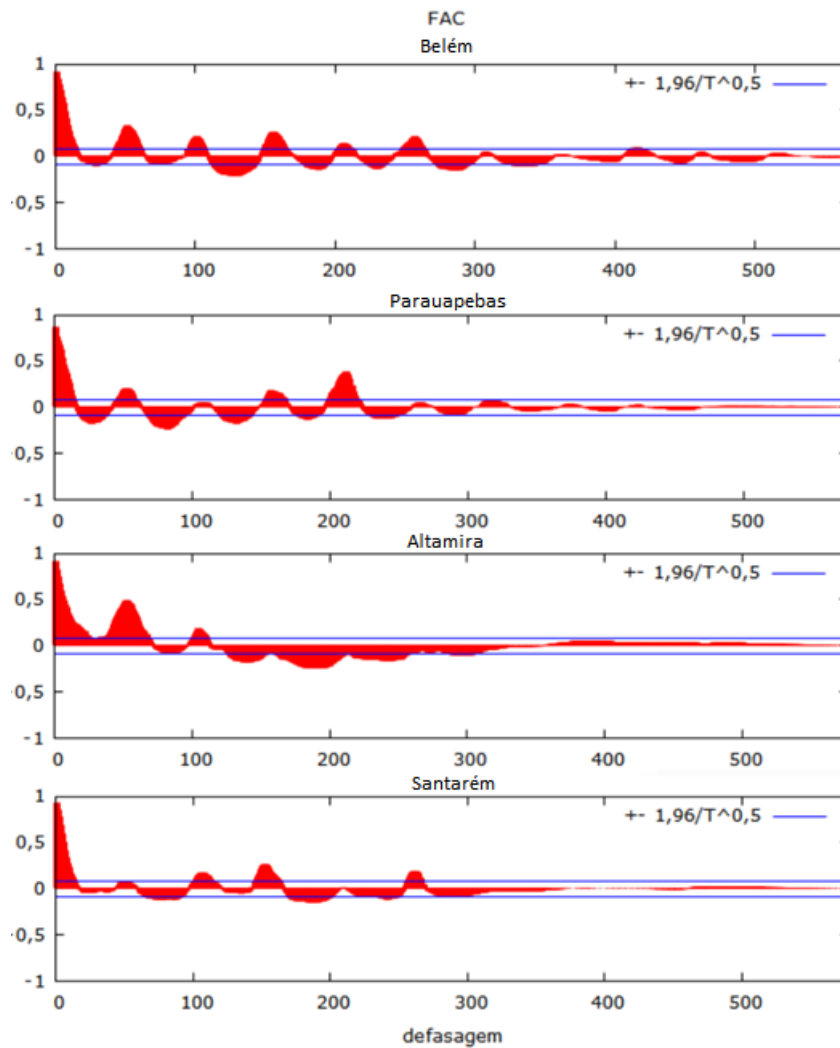


Figura 4.12 FAC – Função de Auto Correlação das séries analisadas

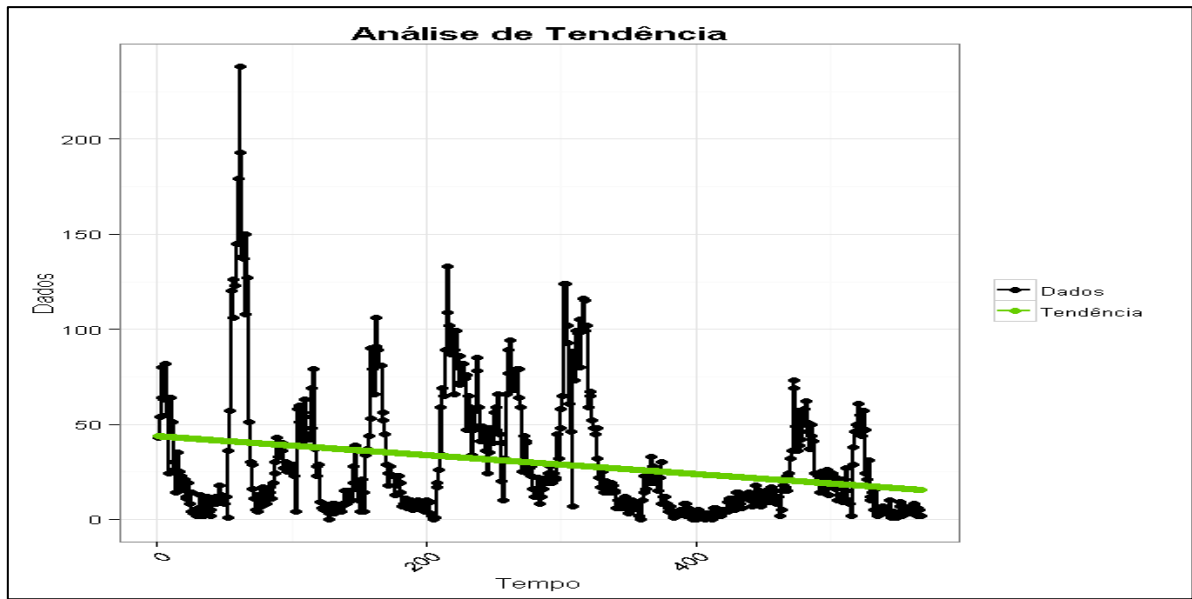
Para a sazonalidade das séries estudadas, usou-se o modelo *Kruskal-Wallis* que avalia a hipótese nula ou alternativa de existir ou não sazonalidade na série [80]. Percebe-se que o $p\text{Valor} = 0$ aponta hipótese de sazonalidade de todas as séries estudadas, de acordo com tabela 4.5.

Tabela 4.5. Testes de Sazonalidade

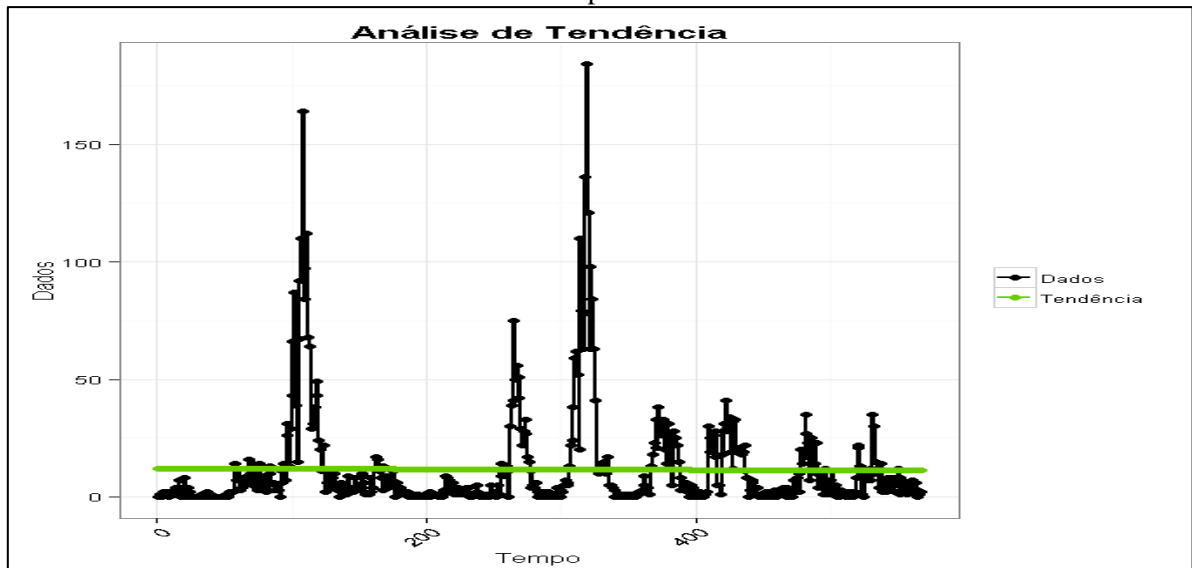
Teste Kruskal-Wallis	
<i>BELÉM</i>	
Estatística	4.000000801
P-Valor	0.969917
Tamanho da Amostra	567

Comprimento Sazonal	12
Hipótese Nula	Não há sazonalidade determinística
Hipótese Alternativa	Há sazonalidade determinística
SANTARÉM	
Estatística	18.17660576
P-Valor	0.077572167
Tamanho da Amostra	566
Comp. Sazonal	12
Hipótese Nula	Não há sazonalidade determinística
Hipótese Alternativa	Há sazonalidade determinística
ALTAMIRA	
Estatística	13.45319189
P-Valor	0.264739159
Tamanho da Amostra	566
Comp. Sazonal	12
Hipótese Nula	Não há sazonalidade determinística
Hipótese Alternativa	Há sazonalidade determinística
PARAUPEBAS	
Estatística	18.35810141
P-Valor	0.073634765
Tamanho da Amostra	566
Comp. Sazonal	12
Hipótese Nula	Não há sazonalidade determinística
Hipótese Alternativa	Há sazonalidade determinística

Para as tendências usou-se o modelo *Mann-Kendall* que avalia hipótese crescente, decrescente ou constante da série ao longo do tempo [80]. A Figura 4.13, mostra que as séries temporais dos municípios Belém e Santarém apresentam tendência decrescente ao longo do tempo. O município de Parauapebas demonstra ser constante ao longo do tempo e o município de Altamira apresenta tendência crescente ao longo do tempo.



Parauapebas-PA



Atamira-Pa

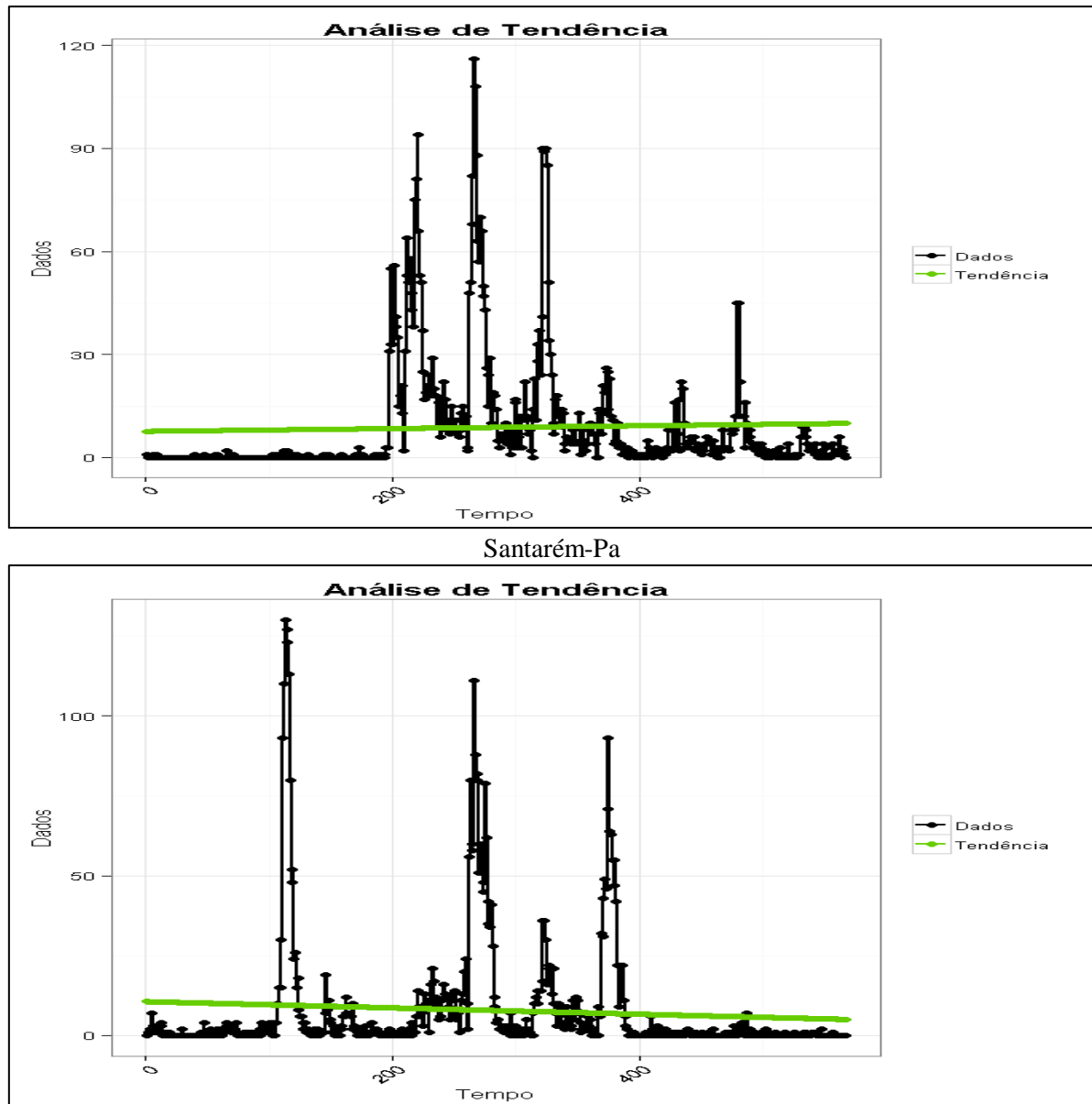


Figura 4.13 Tendência das séries

Após análise estatística das séries temporais de casos de dengue dos municípios observa-se que as séries temporais apresentam complexidade, pois apresentam tendências, sazonalidades e são não-estacionárias, não sendo possível usar métodos de previsões tradicionais como *ARIMA*. No entanto, para usar tais modelos, seria necessário retirar a complexidade das séries (sazonalidade, tendência e não-estacionaridade). O que ocorre nesse tipo de preparo é que ao se ajustar a série para esses modelos, perde-se dados relevantes, acabando em não representação real da série.

Por esse motivo, nessa pesquisa, será usado RNA para realizar as previsões. São

modelos computacionais mais apropriados para esse tipo de séries temporais, cujo comportamento é instável ao longo do tempo. As RNAs são do tipo não-paramétricas, ou seja, não precisam conhecer o comportamento da série antecipadamente e têm a capacidade de se ajustar a séries com tendências, sazonalidade e não-estacionárias.

4.3 Modelo de Rede Neural Artificial

Nessa seção serão demonstrados os cenários e as técnicas para o estudo investigativo das previsões de casos de dengue já discutidos no capítulo anterior. Para comparar os resultados obtidos utilizou-se um conjunto de 32 execuções para cada RNA testada, o que gera erros para cada execução.

Uma vez armazenado o erro em questão é feita o histograma de distribuição dos erros, assim é possível observar os erros e em torno da média, como ocorre na distribuição normal. Assim o erro MAPE da RNA que mais se aproximar da média, será o escolhido para representar a série. Além do modelo que obteve o menor erro percentual médio absoluto (MAPE) e o maior valor Regressão (R). O MAPE é a diferença média quadrada entre saídas e alvos, em que os valores mais baixos são melhores, enquanto zero significa que não há erro. Por outro lado, enquanto os valores de regressão R medem a correlação entre resultados e metas, o valor R igual a 1 significa uma relação íntima e 0 uma relação aleatória, de acordo com [81]

Foram definidos 02 (dois) experimentos para estudo: o experimento I usa o mesmo modelo RNA treinado a partir da série do município de Belém para prever os demais municípios. E o experimento II realizar testes com diferentes arquiteturas de RNA para cada município e escolhe aquele que melhor se ajusta ao modelo. Por fim é avaliado qual experimento retorna o melhor resultado para predição de dengue usando RNAs.

4.3.1 Experimento I

Inicialmente os dados históricos do município de Belém foram usados para treinamento, validade e testes. Logo em seguida esse modelo ajustado foi usado para prever os demais municípios.

A RNA da série do município de Belém obteve erro percentual médio absoluto MAPE

de 0.0347, usando as configurações de RNA com 02 neurônios e 01 camada escondida, de acordo com tabela 4.6.

Tabela 4.6. RNA município Belém-PA

MUNICÍPIO BELÉM			
ESTRUTURA NEURAL	Atraso (t)	MAPE TREINAMENTO	MAPE VALIDAÇÃO
1-2-1	t-2	0.0019	0.0020
1-10-1	t-2	0.0061	0.0028
1-20-1	t-2	0.0035	0.0027
1-301	t-2	0.0044	0.0030

A RNA usada para prever o município de Belém foi aplicada na predição dos demais municípios: Altamira, Parauapebas e Santarém, conforme tabela 4.7 a seguir:

Tabela 4.7 Resumo das RNAs

RNAs				
MUNICÍPIOS	ESTRUTURA NEURAL	Atraso (t)	MAPE TREINAMENTO	MAPE VALIDAÇÃO
Belém	1-2-1	t-2	0.0019	0.0020
Parauapebas	1-2-1	t-2	0.0044	0.0030
Altamira	1-2-1	t-2	0.0028	0.0035
Santarém	1-2-1	t-2	0.0071	0.0030

O modelo de RNA do município de Belém convergiu para o resultado em 15 épocas, e apresentou as seguintes curvas de ajustes conforme a Figura 4.14.

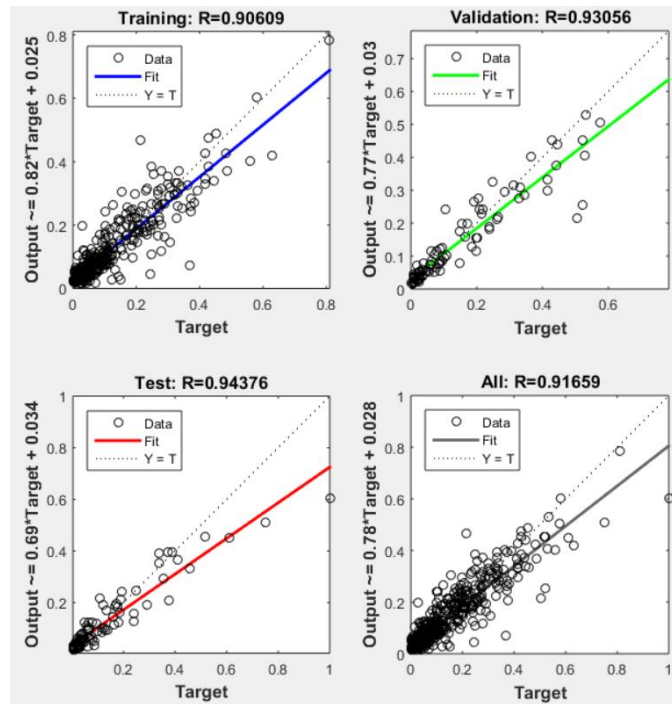


Figura 4.14. Performance da RNA ajustada ao município de Belém

O município de Parauapebas obteve erro percentual médio absoluto (MAPE) de 0.0030, o modelo convergiu para o resultado em 08 épocas, com as seguintes curvas de ajustes descrita na Figura 4.15.

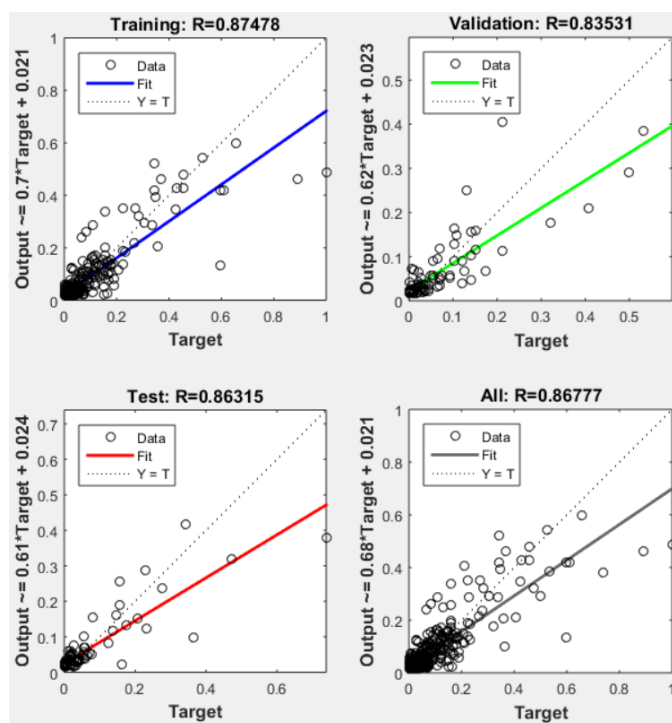


Figura 4.15. Performance da RNA ajustada ao município de Parauapebas

O município de Altamira obteve erro percentual médio absoluto (MAPE) de 0.0035. O modelo convergiu para o resultado em 14 épocas, conforme as curvas de ajustes da Figura 4.16.

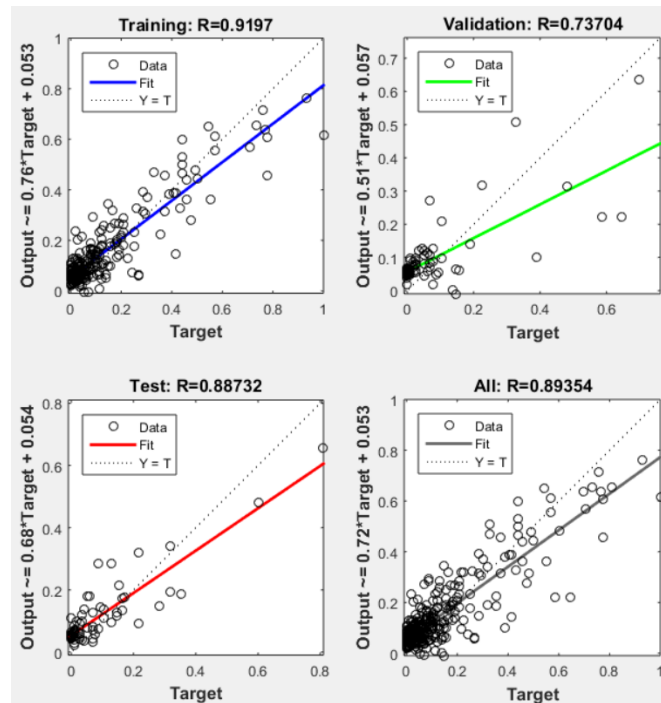


Figura 4.16. Performance da RNA ajustada ao município de Altamira

O município de Santarém obteve erro percentual médio absoluto (MAPE) de 0.0030. O modelo convergiu para o resultado em 11 época, conforme as curvas de ajustes da Figura 4.17.

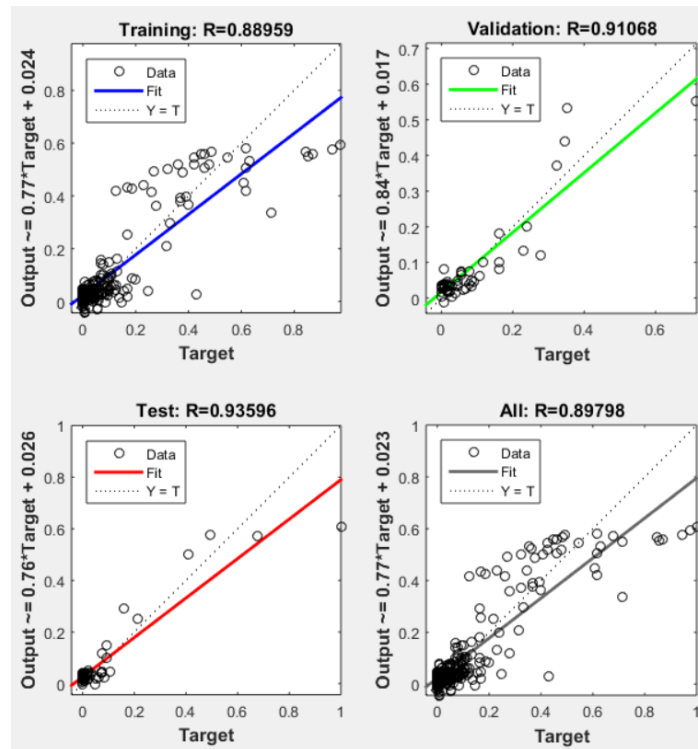


Figura 4.17. Performance da RNA ajustada ao município de Santarém

As figuras 4.14, 4.15, 4.16, 4.17 demonstram que a etapa de treinamento, validação e testes ficaram acima de 0.91, 0.86, 0.89, 0.89 o que mostra boa acurácia no modelo neural ajustado para Belém e valores foram ficando ruins para as demais séries.

4.3.2 Experimento II

Com intuito de melhorar as previsões anteriores, realizou-se testes com diferentes arquiteturas de RNA para cada município, escolhendo assim a RNA que melhor representa a série, conforme tabela 4.8, 4.9, 4.10 e 4.11 a seguir:

Tabela 4.8 RNA município Parauapebas-PA

MUNICÍPIO PARAUPEBAS			
ESTRUTURA NEURAL	ATRASSO (t)	MAPE TREINAMENTO	MAPE VALIDAÇÃO
1-2-1	t-1	0.0033	0.0035
1-4-1	t-2	0.0074	0.0029
1-6-1	t-4	0.0042	0.0028
1-8-1	t-6	0.0030	0.0019

Tabela 4.9 RNA município Altamira-PA

MUNICÍPIO ALTAMIRA			
ESTRUTURA NEURAL	ATRASSO (t)	MAPE TREINAMENTO	MAPE VALIDAÇÃO
1-2-1	t-1	0.0073	0.0035
1-4-1	t-2	0.0042	0.0032
1-6-1	t-4	0.0037	0.0036
1-8-1	t-6	0.0054	0.0038

Tabela 4.10 RNA município Santarém-PA

MUNICÍPIO SANTARÉM			
ESTRUTURA NEURAL	ATRASSO (t)	MAPE TREINAMENTO	MAPE VALIDAÇÃO
1-2-1	t-1	0.0025	0.0030
1-4-1	t-2	0.0027	0.0025
1-6-1	t-4	0.0052	0.0024
1-8-1	t-6	0.0024	0.0036

Tabela 4.11 Resumo das RNAs

RESUMO DAS RNAs				
MUNICÍPIOS	ESTRUTURA NEURAL	ATRASSO (t)	MAPE TREINAMENTO	MAPE VALIDAÇÃO
Belém	1-2-1	t-2	0.0019	0.0020
Parauapebas	1-8-1	t-6	0.0030	0.0019
Altamira	1-4-1	t-2	0.0042	0.0032
Santarém	1-6-1	t-4	0.0052	0.0024

O município de Belém obteve o menor erro percentual médio absoluto (MAPE) de 0.0020, já apresentado na Tabela 4.07.

O município de Parauapebas obteve o menor erro percentual médio absoluto (MAPE) de 0.0019, usando as configurações de RNA com 08 neurônios. O modelo convergiu para o resultado em 13 épocas, conforme as curvas de ajustes da Figura 4.18.

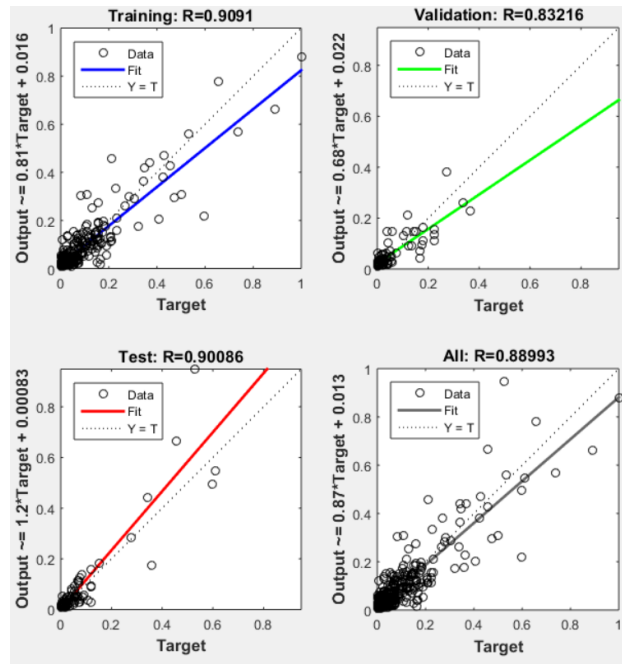


Figura 4.18. Resultado em 3 épocas – Parauapebas

O município de Altamira obteve o menor erro percentual médio absoluto (MAPE) de 0.0032, usando as configurações de RNA com 04 neurônios. O modelo convergiu para o resultado em 11 épocas, conforme as curvas de ajustes da Figura 4.19.

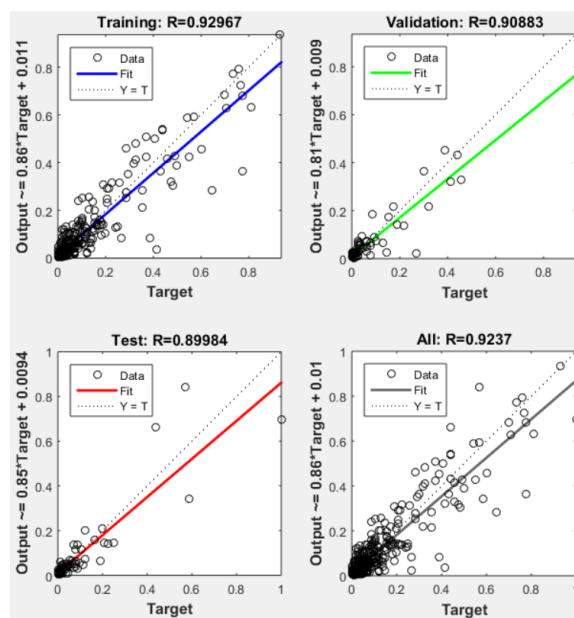


Figura 4.19. Resultado em 7 épocas - Altamira

O município de Santarém obteve o menor erro percentual médio absoluto (MAPE) de 0.0024, usando as configurações de RNA com 6 neurônios. O modelo convergiu para o

resultado em 12 épocas, conforme as curvas de ajustes da Figura 4.20.

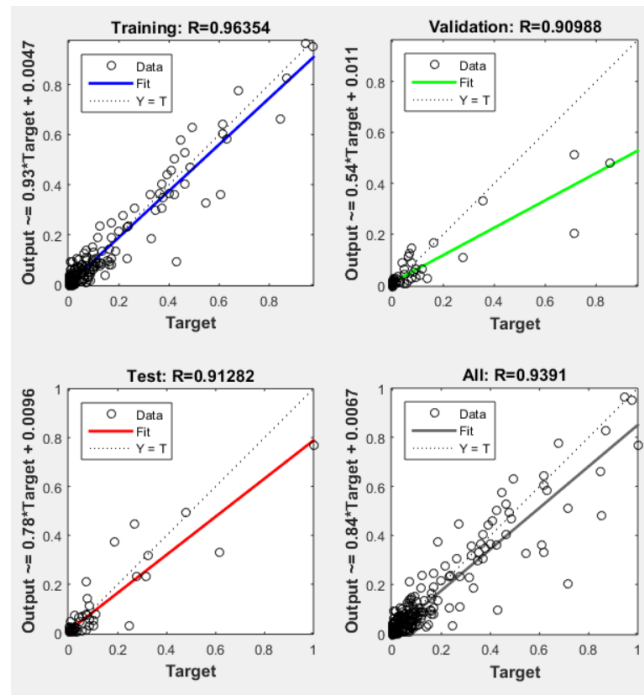


Figura 4.20 Resultado em 2 épocas – Santarém

Com ajustes diferentes para cada RNA, as figuras 4.18, 4.19, 4.20 demonstram que a etapa de treinamento, validação e testes ficaram com 0.92, 0.88, 0.92, 0.93 o que mostra melhor ajuste ao modelo neural usado.

4.3.3 Avaliação dos experimentos I e II

Na análise do experimento I é perceptível que as séries dos demais municípios não obtiveram o desempenho semelhante ao do município de Belém. Isso ocorreu porque o modelo usado para predição do município de Belém se ajustou aos seus dados, não conseguindo realizar a predição dos outros municípios. Para melhorar os resultados é realizado um novo cenário com as séries individuais.

No experimento II foi necessário treinar uma RNA para cada município. Então executou-se a RNA com diferentes arquiteturas, variando a janela de tempo, a quantidade de neurônio e assim obteve-se o melhor resultado em cada etapa. Com isso, foi possível escolher a melhor configuração neural para cada município estudado.

Houve melhora significativa comparando o experimento I com o experimento II. Fato

esse detectado na diminuição do erro MAPE e o R em todos os municípios avaliados, o que demonstra a necessidade de encontrar um modelo neural que represente cada série temporal avaliada. Isso ocorreu devido o comportamento diferente de cada série temporal (tendência, sazonalidade e não-estacionaridade).

Após os experimentos observou-se que os modelos RNAs que tiveram melhor desempenho são os que foram ajustados individualmente por município. Assim as melhores RNAs testadas foram usadas para prever cada município, melhorando as previsões de casos de dengue. Após encontrar a melhor RNA para cada série estudada, foi realizado a predição de novos casos de dengue para cada município.

4.4 Predição para novos casos de Dengue

Levando em consideração o que nos diz [69] nas correlações existentes nas séries temporais de casos de dengue e índice pluviométrico na região, os modelos neurais ajustados foram usados para realizar as previsões em dois cenários: cenário com maior índice pluviométrico e o cenário com o menor índice pluviométrico.

As séries são divididas em 06 meses, 283 instancias de um total de 566, distribuídas em 10 anos, ou seja, metade de instancias para cada cenário estudado. Sendo apresentadas para as RNAs configuradas anteriormente e avaliadas com erro MAPE.

4.4.1 Cenário com alto índice pluviométrico

A primeira série prevista foi o município de Belém que obteve erro MAPE 0.0045 para teste e erro MAPE de predição de 0.0042, como observado na Figura 4.21.

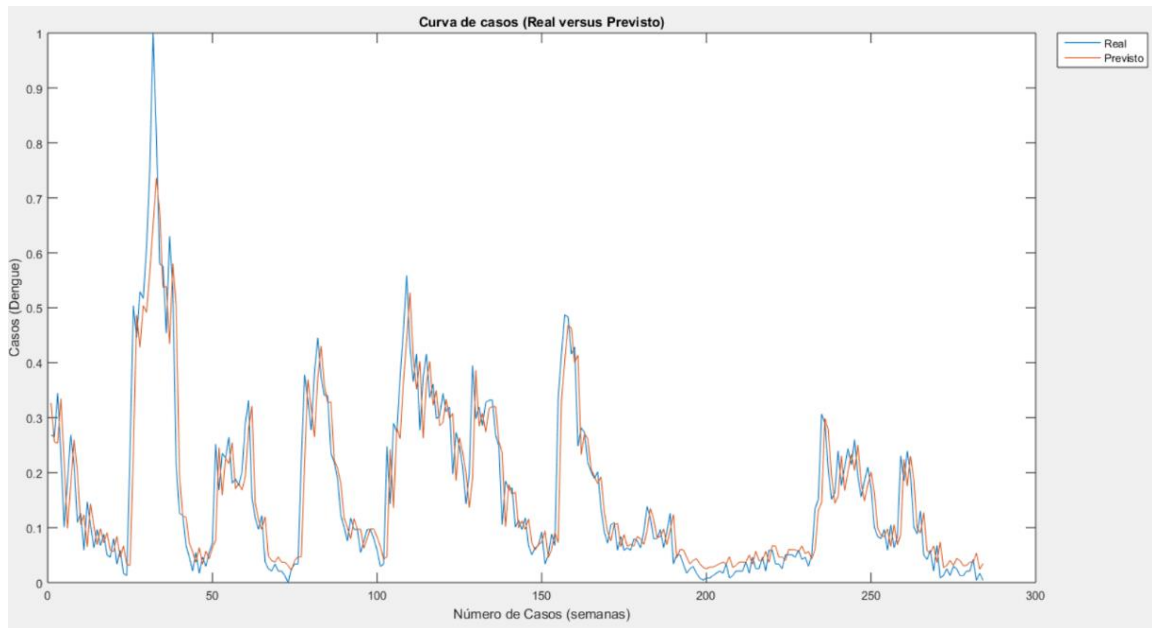


Figura 4.21 Belém no primeiro semestre de 2006 a 2016

Em seguida avaliou-se o município de Parauapebas que obteve o erro MAPE 00.81 para teste e predição de 0.0045, segundo a Figura 4.22.

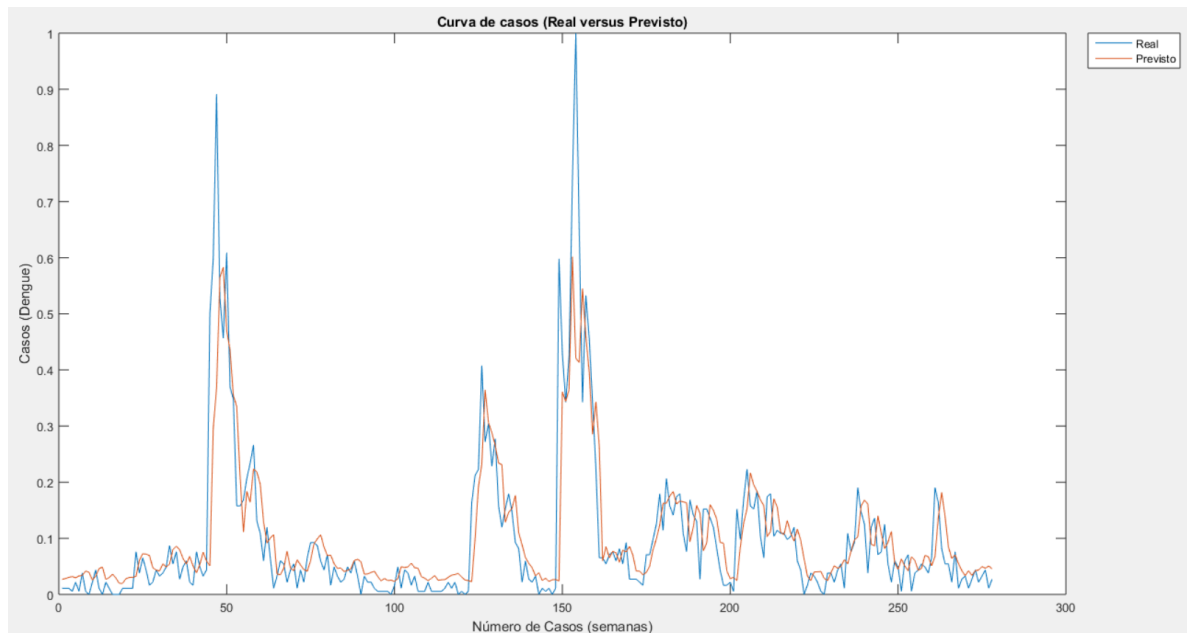


Figura 4.22 Parauapebas no primeiro semestre de 2006 a 2016

Para município de Altamira o erro MAPE foi de 0.0047 para teste e predição de 0.0048, segundo a Figura 4.23.

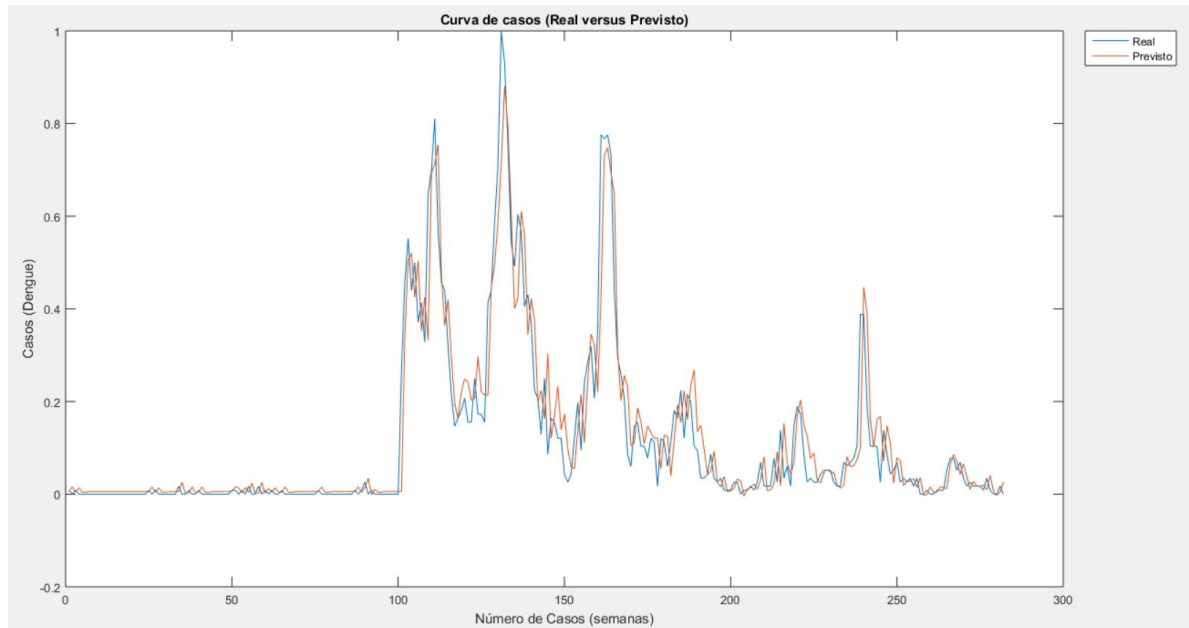


Figura 4.23 Altamira no primeiro semestre de 2006 a 2016

E por fim o município de Santarém apresentou erro MAPE de 0.0030 e predição de 0.0054, conforme a Figura 4.24.

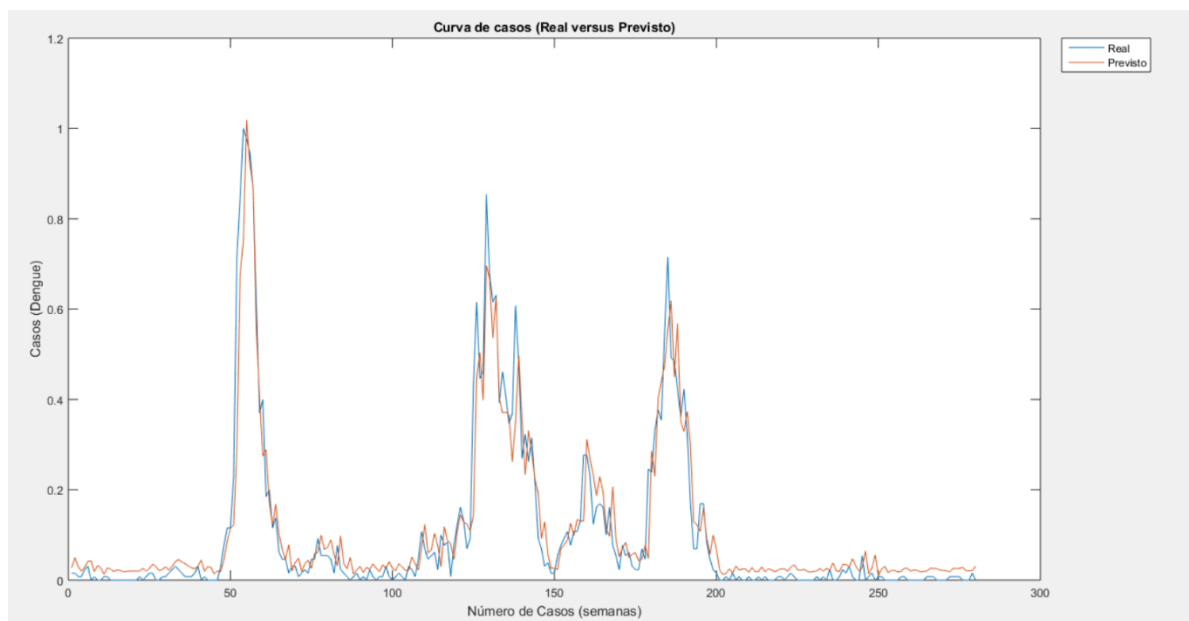


Figura 4.24 Santarém no primeiro semestre de 2006 a 2016

Os gráficos apresentados anteriormente mostram a capacidade das RNAs ajustadas na fase de treinamento e validade em prever casos de dengue. Quanto a sazonalidade apresentada pela série, o modelo ajusta a série de Belém apresenta uma dificuldade em acompanhar a série, o mesmo acontece com as demais séries. Quanto as tendências das séries, as RNAs conseguem acompanhar tanto decrescente, como é o caso de Belém e Santarém, quanto crescente como Altamira, a série de Parauapebas é constante.

4.4.2 Cenário com baixo índice pluviométrico

O município de Belém obteve erro MAPE 0.0029 para teste e erro MAPE de predição de 0.0044, como observado na Figura 4.25.

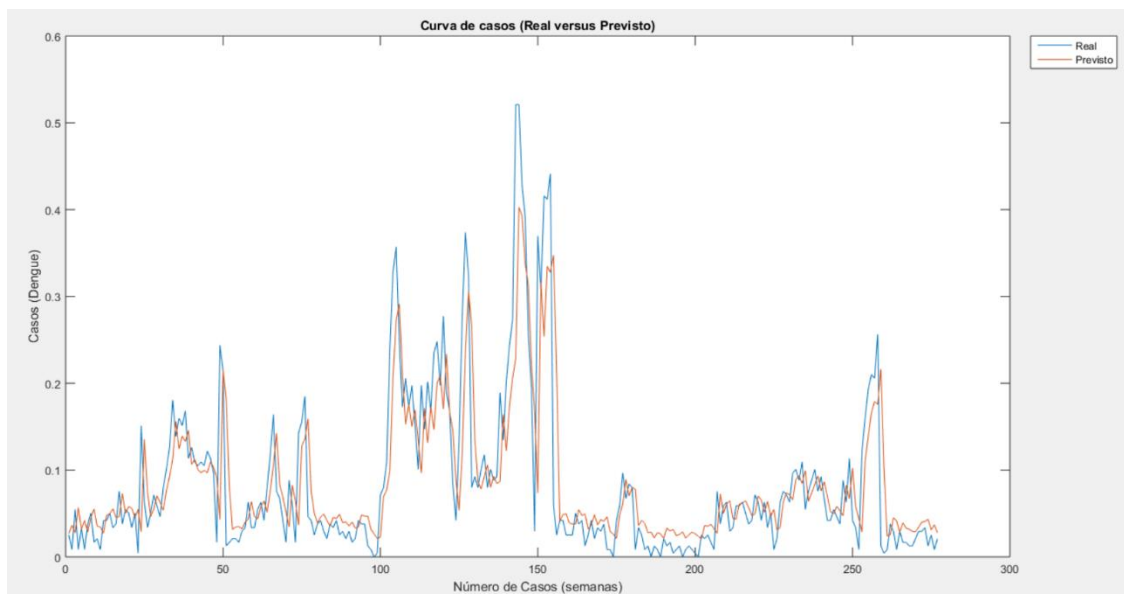


Figura 4.25 Belém no segundo semestre de 2006 a 2016

O município de Parauapebas obteve o erro MAPE 0.0047 de para teste e predição de 0.0058, como podemos observar na Figura 4.26.

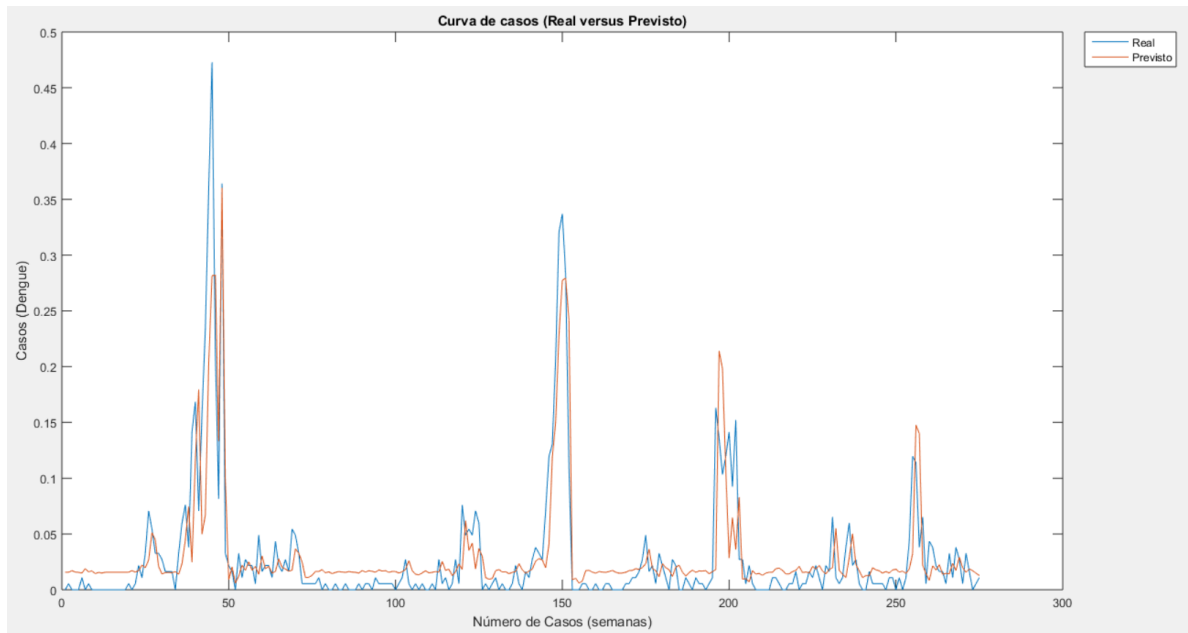


Figura 4.26 Parauapebas no segundo semestre de 2006 a 2016

Para município de Altamira o erro MAPE foi de 0.0020 para teste e predição de 0.0064, de acordo com a Figura 4.27.

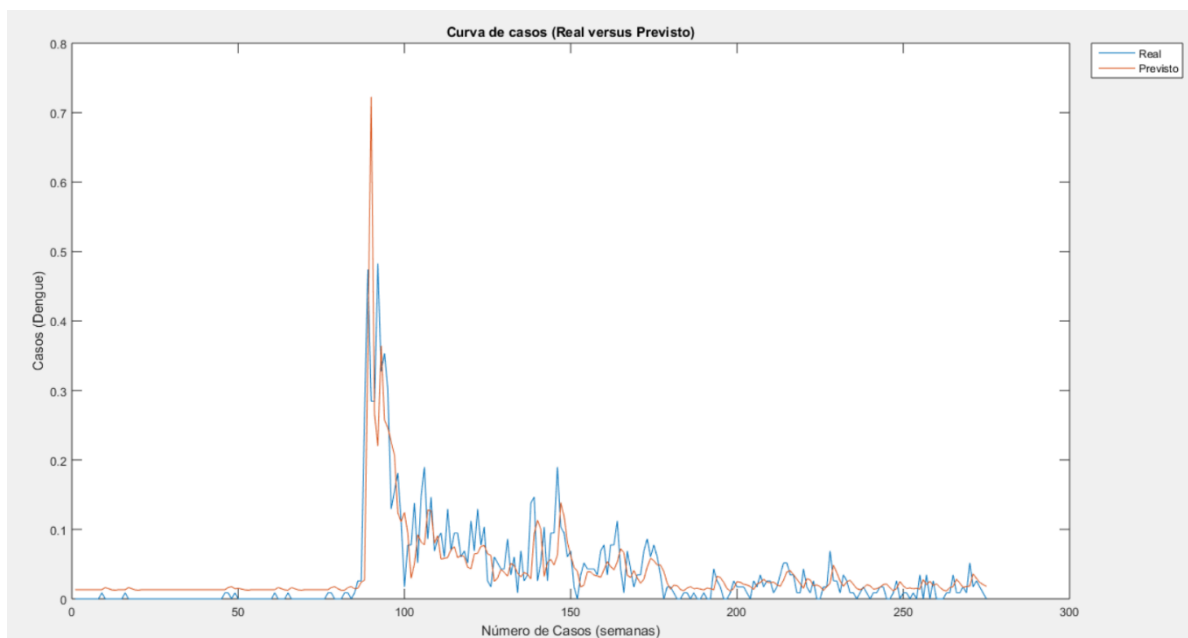


Figura 4.27 Altamira no segundo semestre de 2006 a 2016

E por fim o município de Santarém apresentou erro MAPE de 0.0042 e predição de 0.0072, como descrito na Figura 4.28.

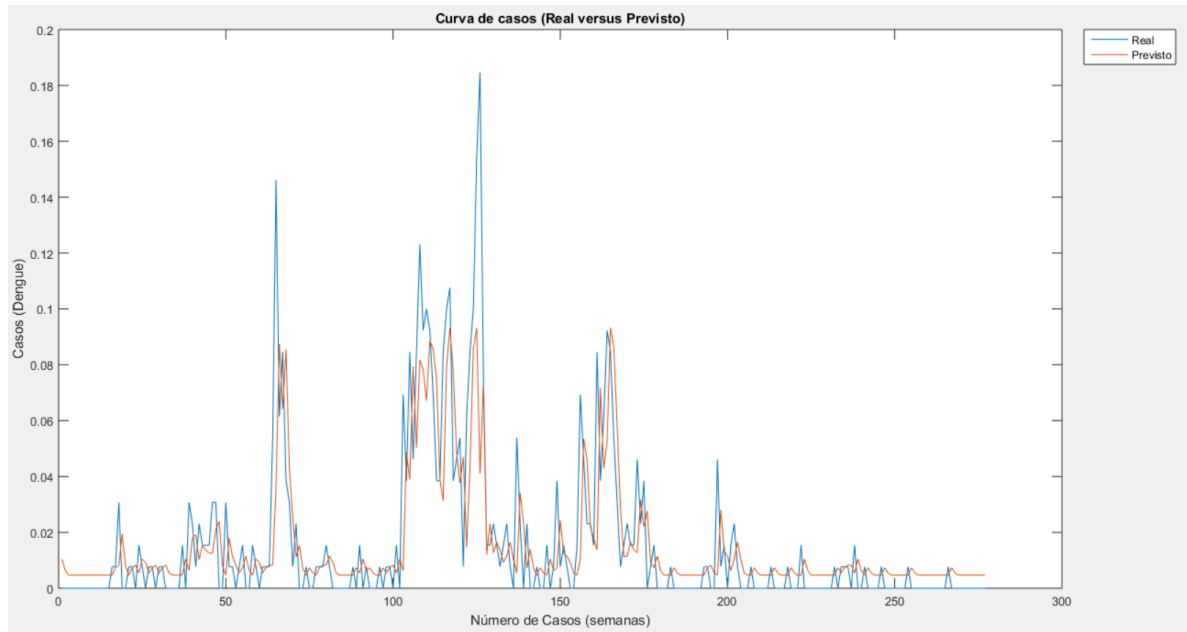


Figura 4.28 Santarém no segundo semestre de 2006 a 2016

Os gráficos demonstram a dificuldade do modelo RNA ajustado para prever casos de dengue no período de baixo índice pluviométrico, uma vez que, os casos de dengue são menores e alterna de forma abrupta no início de cada período, esse comportamento é visível nas séries de Parauapebas e Altamira. Os demais municípios apresentam resultados com maiores erros.

4.5 Avaliação das previsões dos casos de dengue

Foram analisados 2 (dois) cenários de predição em que perceber-se a capacidade da RNA de realizar a predição dos casos de dengue. Esses cenários foram baseados na forte correlação entre os casos de dengue e os índices pluviométricos, já que as séries dos municípios apresentam similaridade, pois tendem a aumentar no primeiro semestre e diminuir no segundo semestre de cada ano.

O cenário com alto índice pluviométrico mostrou um melhor desempenho comparado com o cenário de baixo índice pluviométrico. Isso deve ao fato da RNA se ajustar melhor as tendências das séries dos municípios. Como observado na tabela 4.12 todos as séries do primeiro semestre obtiveram menor erro comparado com os do segundo semestre. Embora

ambos os conjuntos de dados sejam iguais a 283 estâncias (equivalente a 06 meses), o primeiro semestre registra os maiores casos notificados comparado com o segundo, o que explica o melhor ajuste do modelo a esses dados e conseqüentemente melhores resultados.

Tabela 4.12 Comparação das predições

MUNICÍPIOS	Baixo Índice Pluviométrico	Alto Índice Pluviométrico
Belém	0.0042	0.0044
Parauapebas	0.0045	0.0058
Altamira	0.0048	0.0062
Santarém	0.0054	0.0072

4.6 Análise dos Resultados

Este estudo investigou uso de RNA na predição de epidemias de dengue. Os resultados indicaram que o modelo de rede neural, dentro do cenário adequado (tamanho da base de dados, pré-processamento, bases do SINAN, entre outros abordados anteriormente) consegue prever novos casos usando séries históricas.

A estatística usada na pesquisa revela comportamento interessante sobre as séries temporais estudadas, uma vez que são complexas e apresentam tendências, sazonalidade e características não-estacionárias, o que leva a busca de modelos computacionais capazes de prever séries com essas características.

Para os municípios analisados o modelo RNA se mostra eficiente na predição de novos casos, baseados em séries históricas. Porém não foi possível encontrar uma única configuração de rede que pudesse prever todos os municípios, como mostrado no experimento I; sendo necessário realizar o treinamento novamente para cada nova base de dados afim de encontrar uma que melhor represente o comportamento dos dados, como descrito no experimento II.

Baseado na forte correlação entre casos de dengue e índice pluviométricos na região estudada, utilizou-se para predição de novos casos dois cenários para avaliação do modelo

neural: com alto índice pluviométrico e baixo índice pluviométrico. Os testes demonstraram capacidade da RNA em prever novos casos de dengue. Também foi possível observar que as predições dos municípios do primeiro semestre se mostraram melhor as comparadas com o segundo semestre. O que faz o modelo melhor ajustado para prever dados notificados no primeiro semestre, que tem maior número de casos notificados.

Na literatura, de acordo com [7], as RNAs são constantemente usadas para prever epidemias, em especial a de dengue, em diversos países. Em [6] são apontados os casos diários de dengue nas 457 aldeias urbanas da cidade de Kaohsiung, Taiwan, de 2009 até 2012, totalizando 2.997 casos confirmados, realizando predição apenas com cenário de curto prazo de até 30 dias. No entanto, o estudo atual obteve melhores resultados, com relação ao erro da MLP quando comparado ao modelo preditivo desenvolvido pelo autor.

A predição de epidemias usando séries históricas desenvolvido em nosso estudo corrobora trabalhos anteriores. O sistema monitoramento e alertas, desenvolvido nesse estudo, é um passo chave na adição de valor aos dados epidemiológicos coletados por sistemas de vigilância. Outros trabalhos usam sistemas de monitoramento para prever surtos de dengue, como por exemplo [50] que implementa uma ferramenta para prever casos de dengue, mas não prevê a emissão de alerta, que é funcional e demonstra o avanço de nosso estudo.

Também é possível observar trabalhos que usam sistemas inteligentes para auxiliar a tomada de decisão de gestores públicos de saúde, com intuito de melhorar o sistema de controle e prevenção de doenças. Em [1] desenvolve-se um modelo de rede neural capaz de prever casos de dengue no estado do Rio de Janeiro usando séries temporais de casos de dengue, sem utilizar características de redes neurais realimentadas, como é proposto em nosso estudo.

4.7 Considerações Finais

Esse capítulo mostrou o estudo de caso relacionado aos experimentos com séries temporais dos municípios em análise. Foram avaliadas as métricas para cada série e encontrado um modelo para ser usado na predição. Por fim, realizamos a predição de casos de dengue para os municípios estudados.

Conclusão

Este trabalho demonstra o estudo de Redes Neurais Artificiais (RNA) na predição de novos casos de dengue na região amazônica, no estado do Pará, nos seguintes municípios: Belém, Parauapebas, Altamira e Santarém.

A análise estatística revelou a complexidade das séries temporais, tendências, sazonalidades e não-estacionaridade, o que levou a busca de modelos computacionais capazes de serem usados nesse tipo de problema, em nosso caso, a rede neural artificial auto regressiva não linear.

O uso de redes neurais recorrentes, a configuração da arquitetura da rede em treinamento, validação e testes e a variação de janela de tempo, permite um melhor ajuste para predição de todas as séries estudadas, alcançando um erro MAPE adequado a esse tipo de problema.

Os experimentos da pesquisa mostraram que as redes neurais conseguem aprender o comportamento de cada série individual, mas são necessários diversos ajustes para encontrar o melhor modelo que represente a série. Diante do exposto, foi usada uma rede neural adequada para séries temporais que utiliza janela de tempo para melhorar o erro no treinamento, validação e testes.

As previsões do modelo neural demonstraram capacidade em prever o comportamento dos casos em cenários diferentes. Cenário com maior incidência de chuvas e menor incidência foram usados para avaliar a acurácia do modelo.

A metodologia descrita no estudo se difere de outros trabalhos correlatos, pois em sua maioria descreve apenas o uso de redes neurais aplicados a epidemias, não abordando as etapas usadas para realizar os experimentos. Nesse estudo é possível visualizar o as etapas bem como compreender suas particularidades que envolvem predição de epidemias.

A partir de nossas análises podemos concluir que a metodologia descrita na pesquisa tem validade para realizar previsões de casos de dengue, usando redes neurais, antecipando ações de combate e contribuindo para a tomada de decisão, que poderá ser usado por gestores públicos da área da saúde.

5.1 Contribuições

Destacamos as seguintes contribuições desta dissertação:

- Apresentação de uma proposta metodológica para predição de epidemias, usando séries históricas de casos notificados.
- Caracterização da problemática que envolve 04 municípios dentre os que apresentam os maiores índices de dengue no estado do Pará: Belém, Parauapebas, Altamira e Santarém.;
- Desenvolvimento de uma ferramenta computacional que poderá auxiliar na tomada de decisão dos gestores públicos da área da saúde, no que se refere ao planejamento de ações de controle e combate de epidemias.
- A divulgação dos resultados desta pesquisa aconteceu, inicialmente, a partir da publicação do artigo intitulado *Monitoramento de epidemia de dengue na Amazônia usando Redes Neurais Artificiais*, apresentado no *CBIC-2017: XIII Brazilian Congress on Computational Intelligence*, conforme [75] e consolidar-se á nesta dissertação.

5.2 Trabalhos Futuros

Para pesquisas futuras pretende-se utilizar bases de dados públicas relacionando outras epidemias características da região amazônica, como febre amarela, para gerar novos modelos neurais preditivos.

Nesse sentido, integrar bases de dados de epidemias com outros dados climáticos (temperatura, umidade, índices pluviométricos, etc) possibilita a realização de novos experimentos multivalorados.

Pretende-se usar outros modelos computacionais de aprendizagem computacional (como *Máquina Vetor Suporte*, *K-means*) para realizar previsões de epidemias de dengue. Assim, poderá ser possível comparar os resultados atuais com os de outros modelos que utilizam a mesma base de dados.

Isso poderá aperfeiçoar as funcionalidades do sistema proposto para prever casos epidemiológicos com diferentes bases de dados e diferentes modelos computacionais.

5.3 Dificuldades Encontradas

Durante o desenvolvimento desta dissertação foram encontradas diversas dificuldades. Entre as principais dificuldades elencamos:

- Escolha das variáveis que representam o problema em questão;
- O pré-processamento dos dados utilizados e preparação do *dataset*;
- Algoritmos para escolha da melhor configuração da arquitetura de rede neural escolhida;
- As métricas usadas para problemas de predição usando redes neurais.
- O estudo aprofundado de séries temporais, seus métodos e conceitos.

REFERÊNCIAS

- [1] E. El, “Charles Sodré Schramm dos Santos Previsão da Incidência de Dengue no Estado do Rio de Janeiro através de Redes Neurais Artificiais Monografia de Final de Curso Especialização em Business Intelligence,” 2013.
- [2] S. B. Rasheed, R. K. Butlin, and M. Boots, “A review of dengue as an emerging disease in Pakistan,” *Public Health*, vol. 127, no. 1, pp. 11–17, 2013.
- [3] S. Mukherji and S. K. Kaushik, “Dengue: A runaway epidemic and a bewildered public health worker,” *Med. J. Armed Forces India*, vol. 71, no. 1, pp. 3–4, 2015.
- [4] J. P. Messina et al., “Global spread of dengue virus types: Mapping the 70 year history,” *Trends Microbiol.*, vol. 22, no. 3, pp. 138–146, 2014.
- [5] F. S. Tabataba et al., “A Framework for Evaluating Epidemic Forecasts,” pp. 1–27, 2017.
- [6] T.-C. Chan, T.-H. Hu, and J.-S. Hwang, “Daily forecast of dengue fever incidents for urban villages in a city,” *Int. J. Health Geogr.*, vol. 14, no. 1, p. 9, 2015.
- [7] H. M. Aburas, B. G. Cetiner, and M. Sari, “Dengue confirmed-cases prediction: A neural network model,” *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4256–4260, 2010.
- [8] C. Guan, W. Yuan, and Y. Peng, “A cellular automaton model with extended neighborhood for epidemic propagation,” *Proc. - 4th Int. Jt. Conf. Comput. Sci. Optim. CSO 2011*, pp. 623–627, 2011.
- [9] E. Z. Martinez and E. A. S. da Silva, “Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model,” *Cad. Saude Publica*, vol. 27, no. 9, pp. 1809–1818, 2011.
- [10] G. P. Silveira and L. C. De Barros, “Numerical methods integrated with fuzzy logic and stochastic method for solving PDEs: An application to dengue,” *Fuzzy Sets Syst.*, vol. 225, pp. 39–57, 2013.
- [11] Y. L. Cheong, K. Burkart, P. J. Leitão, and T. Lakes, “Assessing weather effects on dengue disease in Malaysia,” *Int. J. Environ. Res. Public Health*, vol. 10, no. 12, pp. 6319–6334, 2013.
- [12] N. Mathur, V. S. Asirvadam, S. C. Dass, and B. S. Gill, “Visualization of dengue incidences for vulnerability using K-means,” *IEEE 2015 Int. Conf. Signal Image Process. Appl. ICSIPA 2015 - Proc.*, pp. 569–573, 2016.
- [13] A. L. Buczak et al., “Prediction of High Incidence of Dengue in the Philippines,” vol. 8, no. 4, 2014.
- [14] A. A. Bakar, Z. Kefli, S. Abdullah, and M. Sahani, “Predictive models for dengue outbreak using multiple rulebase classifiers,” *Proc. 2011 Int. Conf. Electr. Eng. Informatics, ICEEI 2011*, no. July, 2011.
- [15] D. Rahmawati and Y. P. Huang, “Using C-support vector classification to forecast dengue fever epidemics in Taiwan,” *2016 IEEE Int. Conf. Syst. Sci. Eng. ICSSE 2016*, 2016.
- [16] L. L. Quintero-Herrera et al., “Potential impact of climatic variability on the epidemiology of dengue in Risaralda, Colombia, 2010-2011,” *J. Infect. Public Health*, vol. 8, no. 3, pp. 291–297, 2015.
- [17] S. R. Gnanapragasam and T. Cooray, “Time Series Models to Forecast Dengue Fever Incidences in Western Province of Sri Lanka,” no. November, pp. 50–56, 2015.
- [18] L. T. N. Anh, H. X. Dau, and N. H. Phuong, “Cholera Forecast based on Association Rule Mining,” vol. 5, no. 8, pp. 138–143, 2015.

- [19] C. D. Nazri, A. Abu Hassan, Z. Abd Latif, and R. Ismail, "Impact of climate and landuse variability based on dengue epidemic outbreak in Subang Jaya," 2011 IEEE Colloq. Humanit. Sci. Eng. CHUSER 2011, no. Chuser, pp. 907–912, 2011.
- [20] Y. Kim, J. Kim, J. Seok, and B. Du Kim, "Information Propagation Modeling in a Drone Network using Disease Epidemic Models," pp. 79–81, 2016.
- [21] W. Anggraeni and L. Aristiani, "Using Google Trend Data in Forecasting Number of Dengue Fever Cases with ARIMAX Method Case Study: Surabaya, Indonesia," pp. 114–118, 2016.
- [22] D. F. Attaway, K. H. Jacobsen, A. Falconer, G. Manca, and N. M. Waters, "Risk analysis for dengue suitability in Africa using the ArcGIS predictive analysis tools (PA tools)," *Acta Trop.*, vol. 158, pp. 248–257, 2016.
- [23] S. Banu, H. WenBiao, G. YuMing, C. Hurst, and T. ShiLu, "Projecting the impact of climate change on dengue transmission in Dhaka, Bangladesh," *Environ. Int.*, vol. 63, pp. 137–142, 2014.
- [24] J. C. Shen et al., "The Impacts of Mosquito Density and Meteorological Factors on Dengue Fever Epidemics in Guangzhou, China, 2006-2014: a Time-series Analysis," *Biomed. Environ. Sci.*, vol. 28, no. 5, pp. 321–329, 2015.
- [25] F. P. Rocha, H. S. Rodrigues, M. T. T. Monteiro, and D. F. M. Torres, "Coexistence of two dengue virus serotypes and forecasting for Madeira Island," *Oper. Res. Heal. Care*, vol. 7, pp. 122–131, 2015.
- [26] P. Siriyasatien, A. Phumee, P. Ongruk, K. Jampachaisri, and K. Kesorn, "Analysis of significant factors for dengue fever incidence prediction," *BMC Bioinformatics*, vol. 17, no. 1, p. 166, 2016.
- [27] G. Zhu, J. Hunter, and Y. Jiang, "Improved Prediction of Dengue Outbreak Using the Delay Permutation Entropy," 2016 IEEE Int. Conf. Internet Things IEEE Green Comput. Commun. IEEE Cyber, Phys. Soc. Comput. IEEE Smart Data, pp. 828–832, 2016.
- [28] C. B. Ocampo, N. J. Mina, M. Carabalí, N. Alexander, and L. Osorio, "Reduction in dengue cases observed during mass control of *Aedes (Stegomyia)* in street catch basins in an endemic urban area in Colombia," *Acta Trop.*, vol. 132, no. 1, pp. 15–22, 2014.
- [29] Y. L. Hii, H. Zhu, N. Ng, L. C. Ng, and J. Rocklöv, "Forecast of Dengue Incidence Using Temperature and Rainfall," *PLoS Negl. Trop. Dis.*, vol. 6, no. 11, pp. 1–10, 2012.
- [30] C. Torres, S. Barguil, M. Melgarejo, and A. Olarte, "Fuzzy model identification of dengue epidemic in Colombia based on multiresolution analysis," *Artif. Intell. Med.*, vol. 60, no. 1, pp. 41–51, 2014.
- [31] M. K. Othman and M. S. N. M. Danuri, "Proposed conceptual framework of Dengue Active Surveillance System (DASS) in Malaysia," 2016 Int. Conf. Inf. Commun. Technol., no. May, pp. 90–96, 2016.
- [32] R. Lowe et al., "Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil," *Elife*, vol. 5, no. FEBRUARY2016, pp. 1–18, 2016.
- [33] A. E. Carbajo, M. V Cardo, and D. Vezzani, "Is temperature the main cause of dengue rise in non-endemic countries? The case of Argentina," *Int. J. Health Geogr.*, vol. 11, no. 1, p. 26, 2012.
- [34] M. Gharbi et al., "Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors," *BMC Infect. Dis.*, vol. 11, no. 1, p. 166, 2011.
- [35] M. K. Enduri and S. Jolad, "Estimation of reproduction number and non stationary

- spectral analysis of dengue epidemic,” *Math. Biosci.*, vol. 288, pp. 140–148, 2017.
- [36] N. C. Dom, A. H. Ahmad, Z. A. Latif, and R. Ismail, “Application of geographical information system-based analytical hierarchy process as a tool for dengue risk assessment,” *Asian Pacific J. Trop. Dis.*, vol. 6, no. 12, pp. 928–935, 2016.
- [37] C. C. Jansen and N. W. Beebe, “The dengue vector *Aedes aegypti*: what comes next,” *Microbes Infect.*, vol. 12, no. 4, pp. 272–279, 2010.
- [38] H. C. Hapuarachchi et al., “Epidemic resurgence of dengue fever in Singapore in 2013–2014: A virological and entomological perspective,” *BMC Infect. Dis.*, vol. 16, no. 1, p. 300, 2016.
- [39] D. Parker and D. Holman, “International Journal of Infectious Diseases Event history analysis of dengue fever epidemic and inter-epidemic spells in,” *Int. J. Infect. Dis.*, vol. 16, no. 11, pp. e783–e788, 2012.
- [40] R. A. Strauss, J. S. Castro, R. Reintjes, and J. R. Torres, “Google dengue trends: An indicator of epidemic behavior. The Venezuelan Case,” *Int. J. Med. Inform.*, vol. 104, no. May, pp. 26–30, 2017.
- [41] M. E. Wilson and L. H. Chen, “Dengue: Update on Epidemiology,” *Curr. Infect. Dis. Rep.*, vol. 17, no. 1, pp. 1–8, 2015.
- [42] D. J. Rogers, J. E. Suk, and J. C. Semenza, “Using global maps to predict the risk of dengue in Europe,” *Acta Trop.*, vol. 129, no. 1, pp. 1–14, 2014.
- [43] T. Farkhondeh, S. Samarghandian, and F. Samini, “Asian Pacific Journal of Tropical Medicine,” *Asian Pac. J. Trop. Med.*, vol. 9, no. 10, pp. 947–953, 2016.
- [44] A. Balasundaram and P. T. V. Bhuvaneshwari, “Comparative Study on Decision Tree Based Data Mining Algorithm To Assess Risk of Epidemic.”
- [45] V. Antonio, M. Xavier, and D. A. L. P. Jr, “REDES NEURAIAS APLICADAS À PREVISÃO DE SÉRIES TEMPORAIS BATTAGELLO, Vinícius Antonio * , CORDEIRO, Marciso Xavier, DAL PINO Jr, Arnaldo.”
- [46] R. C. Costa, “Modelos Preditivos de Velocidade de Vento para Sistemas Eólicos Baseados em Redes Neurais Artificiais,” 2016.
- [47] D. M. Q. Nelson, “Uso de redes neurais recorrentes para previsão de séries temporais financeiras,” 2017.
- [48] F. Saberian, A. Zamani, M. M. Gooya, P. Hemmati, M. A. Shoorehdeli, and M. Teshnehlab, “Prediction of Seasonal Influenza Epidemics in Tehran Using Artificial Neural Networks,” *Ieee*, no. Ieee, pp. 1921–1923, 2014.
- [49] D. Phung et al., “Identification of the prediction model for dengue incidence in Can Tho city, a Mekong Delta area in Vietnam,” *Acta Trop.*, vol. 141, no. Part A, pp. 88–96, 2015.
- [50] F. I. Eljorje, D. S. Clarite, B. D. Gerardo, and Y. Byun, “Tracking and prediction of dengue outbreak using cloud-based services and artificial neural network,” *Int. J. Multimed. Ubiquitous Eng.*, vol. 11, no. 5, pp. 355–366, 2016.
- [51] LMTECH. Disponível em: <http://www.gsigma.ufsc.br/~popov/aulas/rna/neuronio_implementation/>. Acesso em: 20 de nov. 2017
- [52] GSIGMA. Disponível em: <<http://www.lmtech.info/index.php/tecnologia/inteligencia-artificial/129-introducao-a-redes-neurais-artificiais>>. Acesso em: 21 de nov. 2017
- [53] FILHO, M. Disponível em: <<http://mariofilho.com/as-metricas-mais-populares-para-avaliar-modelos-de-machine-learning/>>. Acesso em: 22 de nov. 2017.
- [54] CAMARA LEGISLATIVA. Disponível em: <<http://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-permanentes/cssf/audiencias-publicas/audiencia->

- publica-2015/audiencia-28-05-dengue/apresentacao-3/>. Acesso em: 22 de nov. 2017.
- [55] MONITORAMENTO. <http://portalarquivos.saude.gov.br/images/pdf/2017/agosto/29/2017-026-Monitoramento-dos-casos-de-dengue-febre-de-chikungunya-e-febre-pelo-virus-Zika-ate-a-Semana-Epidemiologica-33-de-2017.pdf>>. Acesso em: 22 de nov. 2017
- [56] MINISTÉRIO DA SAÚDE. Disponível em: <<http://portalsaude.saude.gov.br/index.php/cidadao/principal/>>. Acesso em: 22 de nov. 2017
- [57] SCIENTIFIC. Disponível em: <<http://magazine.scientificmalaysian.com/issue-12-2016/dengue-vaccine-dilemma-route-prevention-yet/>>. Acesso em: 10 de ago. 2017
- [58] UEM. Disponível em: <<http://www.din.uem.br/~gpea/linhas-de-pesquisa/mineracao-de-dados/pre-processamento/pre-processamento-em-data-mining/>>. Acesso em: 22 de nov. 2017
- [59] DAVILA, 2015. Disponível em: <<http://www.ime.unicamp.br/~hlauchos/estdescr1.pdf>>. Acesso em: 20 de nov. 2017
- [60] SONDAGENS, 2015. Disponível em: <<https://sondagenseestudosdeopiniao.wordpress.com/estatistica/estatistica-descritiva/>>. Acesso em: 20 de nov. 2017
- [61] PORTAL ACTION. Disponível em: <<http://portalaction.com.br/>>. Acesso em: 22 de nov. 2017.
- [62] JUPYTER NOTEBOOK. Disponível em: <<http://jupyter.org/>>. Acesso em: 10 de ago. 2017.
- [63] MENESES M. Disponível em: <<http://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf/>>. Acesso em: 10 de ago. 2017.
- [64] LUIS, J. Análise De Séries Temporais. PUC RIO, 2015
- [65] JÚNIO L. C. C. et al. Estudo do comportamento da Dengue na cidade de Araguari-MG por meio de séries temporais. 2010
- [66] BATISTA A. M. et al. Comparação entre modelos de séries temporais para predição de lucro de uma companhia brasileira de alimentos. Universidade Federal do Paraná. 2007.
- [67] PASCHOALINO F. F. PREDIÇÃO DE DEMANDA DE ENERGIA ELÉTRICA NO BRASIL UTILIZANDO REDES NEURAIAS DE ELMAN. 2007. SPOLM.
- [68] DUARTE, A. R. et al. USO DE REDES NEURAIAS PARA PREDIÇÃO DO CONSUMO DE ENERGIA EM HOTÉIS. 2010 CONEM.
- [69] SIQUEIRA, I. S. Análise de incidência de dengue na área urbana de Belém. Uma aplicação modelos espaciais e temporais. 2013. UFPA.
- [70] MACIEL, L, T. PRECIFICAÇÃO DE OPÇÕES DE TAXA DE JUROSNO BRASIL: UMA ANÁLISE DOS MODELOS DE BLACK, VASICEK, CIR E REDESNEURAIAS RECORRENTES.
- [71] SEGATTO, E. N. et al. Redes neurais aplicadas a relés diferenciais para transformadores de potência. 2008. Sba: Controle & Automação Sociedade Brasileira de Automatica.
- [72] OLIVEIRA, T. P. Predição de tráfego usando redes neurais artificiais, para gerenciamento adaptativo em largura de banda em roteadores. 2014. UFU.

- [73] OLIVEIRA, E. Análise e classificação de redes de colaboração utilizando a plataforma lattes e redes neurais artificiais. 2014. II Encontro Mineiro de Modelagem Computacional
- [74] INMET. Disponível em: < <http://www.inmet.gov.br>>. Acesso em: 20 de dezembro. 2017.
- [75] SOUZA, Sócrates Siqueira de. Associação entre incidência de dengue, pluviosidade e densidade larvária de *Aedes aegypti*, no Estado de Goiás. 2010. Revista da Sociedade Brasileira de Medicina Tropical.
- [75] SILVA, W. R. S.; FRANCES, C. R. L. Monitoramento de epidemia de dengue na Amazônia usando Redes Neurais Artificiais. Aceito para publicação no CBIC-2017: XIII Brazilian Congress on Computational Intelligence. Niterói-RJ, 2017.
- [76] SCHMIDHUBER, Jurgen. Deep learning in neural networks: An overview. *Neural networks*, v. 61, p. 85-117. Elsevier, 2015.
- [77] DING, Shifei et al. Evolutionary artificial neural networks: a review. *Artificial Intelligence Review*, p. 1-10. Springer Netherlands, 2013.
- [78] KARAYIANNIS, Nicolaos; VENETSANOPOULOS, Anastasios N. Artificial neural networks: learning algorithms, performance evaluation, and applications. Springer Science & Business Media, 2013.
- [79] SILVA, I. N et al. Redes Neurais Artificiais para engenharia e ciências aplicadas. São Paulo. Artliber, 2010.
- [80] TRINDADE A. L. C. et al. Tendências temporais e espaciais da qualidade das águas superficiais da sub-bacia do Rio das Velhas, estado de Minas Gerais. 2013.
- [81] MATLAB. Disponível em: < <http://www.mathworks.com>>. Acesso em: 20 de dezembro. 2017
- [82] SINAN. Disponível em: < <http://www.portalsinan.saude.gov.br.com>>. Acesso em: 20 de agosto. 2017
- [83] IBGE. Disponível em: < <http://www.ibge.gov.br>>. Acesso em: 20 de agosto. 2017
- [84] DATASUS. Disponível em: <<http://tabnet.datasus.gov.br/>>. Acesso em: 20 de agosto. 2017