

Aplicação de Técnicas Computacionais na Análise e Predição da Incidência do Mosquito da Dengue: Uma Revisão Sistemática da Literatura

Lucas de Sousa Wanderley¹, José Trajano M. Neto¹, Pablo Roberto Fernandes de Oliveira¹, Keila Lucas dos Santos¹, Wellington Candeia de Araújo¹

¹Centro de Ciências exatas e Sociais Aplicadas (CCEA) - Universidade Estadual da Paraíba (UEPB)

Campus VII – Governador Antônio Mariz
Cep 58707 - 040 - Patos – PB – Brasil

lucaswanderley4@gmail.com, jose_netto_mendes@hotmail.com,
(pablrobertofernando, keila.luk.8, wcandeia}@gmail.com

Abstract. *Dengue fever is a viral disease that has spread rapidly in all regions of the world in recent years. This article presents data from a literature review on dengue and data prediction methods on dengue. We searched for research that described or used the computational techniques of data analysis on dengue. We searched too for methods that use Algorithmic, Optimization and / or Data Mining techniques in the Prediction and Control of Aedes aegypti disease vector, as well as the tools used in the analysis and display of spatial data distribution. As a result of the review, 10 of 214 papers were selected from the IEEE, ACM, CAPES and ScienceDirect databases.*

Keywords. *dengue fever; mosquito; systematic review; data science.*

Resumo. *A dengue é uma doença viral que se espalhou rapidamente em todas as regiões do mundo nos últimos anos. Este artigo apresenta os dados resultantes de uma Revisão Sistemática da Literatura (RSL) sobre a doença dengue e métodos de predição de dados sobre a dengue. Buscou-se trabalhos sobre pesquisas que descreviam ou utilizavam técnicas computacionais na análise de dados sobre dengue. Procurou-se também por trabalhos que utilizaram técnicas de Algoritmos, Otimização e/ou Mineração de Dados na predição e controle da incidência do mosquito Aedes aegypti, vetor da doença, bem como as ferramentas utilizadas na análise e exibição da distribuição espacial desses dados. Como resultado da revisão foram selecionados 10 de 214 trabalhos presente nas bases IEEE, ACM, CAPES e ScienceDirect.*

Palavras-Chave. *dengue; mosquito; revisão sistemática; ciência de dados.*

1. Introdução

A Organização Mundial da Saúde (OMS) classifica a dengue como uma das principais questões de saúde pública. Avalia-se que 2,5 bilhões de habitantes estão sob risco de contrair dengue, ocorrendo cerca de 50 milhões de casos anualmente. Do total avaliado, em torno de 550 mil precisam de hospitalização e pelo menos 20 mil morrem em decorrência da doença. (BRASIL, 2009).

No Brasil, os casos de dengue têm preocupado as autoridades, que por meio da conscientização tentam evitar a proliferação do mosquito vetor da doença, o *Aedes aegypti*. Alguns fatores podem favorecer a proliferação do mosquito transmissor, como água parada, períodos de chuva, entre outros. Estuda-se, então, maneiras de evitar o surto da dengue, identificando possíveis causas, e quais condições favorecem a proliferação do mosquito.

A ciência de dados, que utiliza técnicas de inteligência computacional para análise e predição de dados e extração de conhecimento, tem demonstrado importante contribuição em áreas com pouco ou grande volume de dados, em que há a necessidade da interpretação e visualização inteligente da informação.

Pensando nisso, este trabalho apresenta uma pesquisa bibliográfica, em forma de revisão sistemática da literatura, cujo objetivo foi o levantamento do estado da arte sobre as técnicas e algoritmos aplicados para predição e análise de dados da dengue, a fim de nortear pesquisas futuras aplicadas a esta finalidade, auxiliando na escolha das técnicas utilizadas bem como das ferramentas disponíveis.

2. Processo da Revisão Sistemática

Revisão Sistemática da Literatura (RSL), segundo Kitchenham, (2004), trata-se de uma metodologia que permite ao pesquisador identificar, avaliar e interpretar todas as pesquisas disponíveis relacionadas a um tema ou assunto específico. Esta metodologia difere-se da revisão de literatura tradicional quando oferece métodos formais para coletar e interpretar os estudos, descreve todo o percurso da pesquisa desde a busca até a seleção e avaliação da qualidade dos estudos, além de apoiar pesquisas como o Estado da Arte. A RSL tem sido amplamente utilizada em diversas áreas, como Medicina e Engenharia de Software, por exemplo.

Este trabalho utilizou-se da RSL para levantar o Estado da Arte sobre pesquisas relacionadas ao tratamento dos dados sobre a proliferação do mosquito da Dengue. A fim de identificar quais técnicas, algoritmos, otimização e/ou mineração de dados na predição e controle da incidência do mosquito *Aedes Aegypti*, bem como as ferramentas utilizadas na análise e exibição da distribuição espacial desses dados.

A RSL deste trabalho seguiu o processo de Kitchenham (2007), e estruturamos os resultados com base no trabalho de Queiroz e Braga (2014). Quatro pesquisadores participaram da revisão, sendo o revisor principal estudante de mestrado, responsável pela elaboração do protocolo seguido para a pesquisa, dois alunos participantes de iniciação científica, e o revisor especialista responsável por validar as revisões.

2.1. Objetivos da Revisão Sistemática

A motivação para esta RSL partiu da necessidade do tratamento dos dados da dengue disponibilizados de forma pública, para predizer o surto de dengue nas regiões do Brasil. Como resultado esperou-se obter uma visão geral das técnicas já utilizadas e as ferramentas que apoiam essas técnicas. O objetivo assim da RSL é identificar os algoritmos utilizados, ferramentas e formas de seleção, qualificação e exibição dos dados geográficos da dengue, como também analisar os resultados dos estudos quanto a predição do surto da dengue.

2.2. Questões de Pesquisa

Para alcançarmos os objetivos desta revisão, buscamos responder às seguintes questões de pesquisa:

- **Questão Primária (QP1):** Como a computação e seus algoritmos podem contribuir com a atenuação da propagação espacial do mosquito *Aedes Aegypti* e suas consequências na saúde pública?
- **Questão Primária (QP2):** como as técnicas de computação Natural, otimização e mineração de Dados podem ser utilizadas para auxiliar a previsão da distribuição espacial do mosquito no ambiente?

Com bases nesses questionamentos, elaborou-se para a revisão as seguintes questões secundárias:

- **Questão Secundária (QS1):** O problema do artigo concentra-se na visualização, predição ou análise de dados?
- **Questão Secundária (QS2):** Qual a técnica utilizada?
- **Questão Secundária (QS3):** Os dados estão relacionados com qual doença: dengue, zika, chikungunya ou outro?
- **Questão Secundária (QS4):** Qual a ferramenta utilizada no processo?
- **Questão Secundária (QS5):** Qual a forma exibição de visualização dos dados, utilizou alguma ferramenta?

2.3. Estratégia de Pesquisa

Utilizou-se as bibliotecas digitais IEEE Xplore, ACM *Digital Library*, *Science Direct* e Periódicos da CAPES. Considerou-se apenas artigos escritos em inglês e que haviam sido publicados a partir do ano de 2010.

A pesquisa nas bases informadas foi feita utilizando a combinação de palavras-chave inseridas nos buscadores das bibliotecas digitais. Para realizar a busca nas bases utilizamos a pesquisa avançada das mesmas, e antes de tudo analisamos a *String* na base IEEE, para verificar se a mesma retornava trabalhos significativos.

A *String* utilizada nas buscas em cada base foi refinada para se adaptar a pesquisa avançada das bases citadas anteriormente, como apresentado na Tabela 1 a seguir.

Tabela 1 - *String* de busca nas bases de pesquisa

Biblioteca Digital	<i>String</i> de Busca
IEEE	(dengue OR "aedes aegypti") AND ("data mining" OR algorithm OR optimization OR "natural computing") AND (prediction OR analysis OR visualization OR "spatial data" OR "spatial distribution" OR "propagation" OR incidence)
ACM	(dengue OR "aedes aegypti") AND ("data mining" OR algorithm OR optimization OR "natural computing") AND (prediction OR analysis OR visualization OR "spatial data" OR "spatial distribution" OR "propagation" OR incidence)
CAPES	Opção marcada: Título + Contém: (dengue OR "aedes aegypti") AND Opção marcada: Qualquer + Contém: ("data mining" OR algorithm OR optimization OR "natural computing") AND (prediction OR analysis OR visualization OR "spatial data" OR "spatial distribution" OR "propagation" OR incidence)
Science Direct	TITLE-ABSTR-KEY ((dengue OR "aedes aegypti") AND TITLE-ABSTR-KEY(("data mining" OR algorithm OR optimization OR "natural computing") AND (prediction OR analysis OR visualization OR "spatial data" OR "spatial distribution" OR "propagation" OR incidence)))

2.4. Estratégias de Seleção de Estudos

Para a seleção dos estudos, seguimos os seguintes critérios de inclusão e exclusão:

- **Critério de inclusão** - Artigos completos e resumos estendidos que apresentem a utilização de Técnicas de Computação Natural, Mineração de Dados ou Otimização na predição, análise e visualização de dados sobre a proliferação e incidência da dengue.
- **Critério de exclusão 1** - Tiverem sido publicados antes de 2010;
- **Critério de exclusão 2** - Não sejam artigos escritos na língua inglesa;
- **Critério de exclusão 3** - É um resumo ou Revisão Sistemática de Literatura ou pesquisa bibliográfica;
- **Critério de exclusão 4** - Artigos iguais ou que possuam versão atualizada (deve ser mantido apenas o mais recente).
- **Critério de exclusão 5** - Não apresenta aplicação de nenhuma técnica de computação natural, mineração de dados, otimização no tratamento dos dados da dengue
- **Critério de exclusão 6** - Não apresenta a incidência da dengue ou predição do surto da doença ou mosquito.
- **Critério de exclusão 7** - Não disponível para visualização ou *download*

Os critérios foram aplicados cumprindo alguns passos. A partir da leitura dos resumos, mas também introdução, resultados e considerações finais dos trabalhos quando necessário, foram aplicados os critérios de inclusão e exclusão.

3. Resultado da Revisão Sistemática

Com a realização do processo exposto na seção anterior, obtivemos como resultado da *String* de busca 214 artigos, somando as quatro bases de pesquisa. Neste momento foram lidos os resumos de cada artigo para a aplicação dos critérios de Exclusão EX1, EX2, EX3 e EX4. Na próxima etapa e aplicação dos critérios exclusão EX5, EX6, EX7, para os quais, foram lidos os resumos e quando necessário a Introdução e Conclusão dos mesmos. Permaneceram ao final desta fase, 10 artigos para avaliação do Especialista. A tabela a seguir (Tabela 2) refere-se de forma breve os resultados da aplicação dos critérios em cada etapa.

Tabela 2 - Resultados da aplicação dos critérios

Base	String	EX1	EX2	EX3	EX4	EX5/EX6	EX7	Incluídos
CAPES	35	9	0	1	2	17	2	4
ACM	43	7	0	0	0	36	0	0
ScienceDirect	102	23	0	2	1	75	0	1
IEEE	34	9	0	0	0	20	0	5
Total	214	48	0	3	3	148	2	10
Total Subtraído	214	166	166	163	160	12	10	10

Todos os artigos incluídos, após o critério de inclusão, foram analisados e aceitos pelo Especialista e seguiram para a avaliação de qualidade e leitura na íntegra. Os estudos foram analisados de acordo com as questões de pesquisa e os critérios de qualidade. A distribuição de artigos por ano pode ser observada no gráfico a seguir (Gráfico 1).

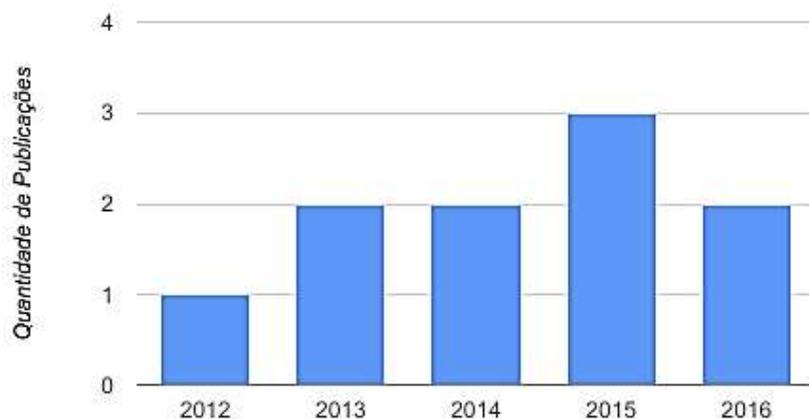


Gráfico 1: Distribuição por ano dos artigos selecionados

3.1. Questões primárias da pesquisa

Através da análise dos artigos foi observado que a maioria das pesquisas tiveram como objetivo a utilização de algoritmos ou modelos criados computacionalmente para análise e predição, como também visualização. Viu-se que foi possível desenvolver métodos capazes de entender o comportamento dos dados da epidemia dengue em uma região, e que também foi possível realizar um controle mais eficiente por parte dos responsáveis pelo controle do vetor.

Foi observado que, embora alguns artigos apresentassem de forma predominante o uso de Inteligência Artificial (IA) e Mineração de Dados, eles também associaram modelagens matemáticas e estatísticas. O número de artigos que foi permitido identificar isso é apresentado no gráfico 2:

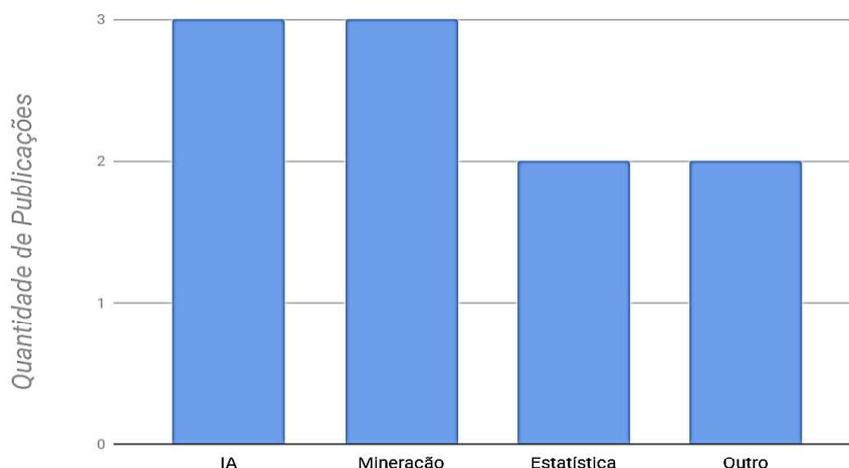


Gráfico 2 - Quantidade de artigos por métodos

A próxima pergunta da pesquisa foi em relação a qual domínio o artigo esteve direcionado, mesmo que tenha trazido em seu tema dominante visualização e análise dos dados. Avaliamos se o foco também estava relacionado a predição ou otimização. A subseção seguinte apresenta esses resultados.

3.2. Resposta às questões secundárias de pesquisa

Os artigos em sua maioria tiveram como principal foco o desenvolvimento de um modelo preditivo, no entanto para a criação desse modelo foi utilizado, na grande maioria dos modelos,

técnicas de visualização, análise e mineração de dados. O número de artigos que foi permitido identificar isso é apresentado no gráfico 3:

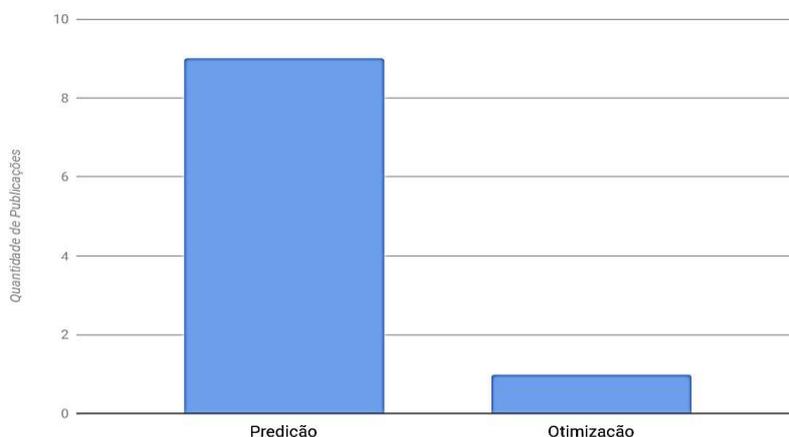


Gráfico 3 - Quantidade de artigos por objetivo

Nota-se a partir dos dados do gráfico 3 que, a maioria dos estudos observados tiveram como destaque a predição de dados. Foram identificadas várias técnicas utilizadas nos estudos selecionados, tais como o uso do algoritmo de classificação *Naive Bayes*, pesquisas exaustivas, modelos híbridos baseados em algoritmos genéticos para determinar o peso em um modelo de rede neural e algoritmos genéticos multiobjetivo.

Dentre as ferramentas apresentadas nos artigos foram destacadas WEKA (*Waikato Environment for Knowledge Analysis*), GIS (*Geographic Information System*), Scilab, Matlab, Pacote R, Redes Neurais e Algoritmos Genéticos. Com maior destaque, observou-se a utilização das ferramentas WEKA, GIS, Matlab e Scilab.

4. Discussão

Nesta seção são apresentados os estudos selecionados de interesse deste trabalho, listados na tabela 3. Os artigos foram organizados a partir de um identificador, da base utilizada onde foram encontrados, autores, título e o ano de publicação.

Tabela 4: Estudos Fundamentais Selecionados

ID	Base	Autor(es)	Título	Ano
5	SD	Claudia Torresa, Samier Barguila, Miguel Melgarejoa, Andrés Olarteb	<i>Fuzzy model identification of dengue epidemic in Colombia based on multiresolution analysis</i>	2014
80	IEEE	Vanny Narita, Anto Satriyo Nugroho	<i>Dengue outbreak prediction for GIS based Early Warning System</i>	2015
82	IEEE	Arthi Balasundaram, P. T. V. Bhuvaneswari	<i>Comparative study on decision tree based data mining algorithm to assess risk of epidemic</i>	2013
86	IEEE	Nor Azura Husin, Norwati Mustapha, Md. Nasir Sulaiman, Razali Yaakob	<i>A hybrid model using genetic algorithm and neural network for predicting dengue outbreak</i>	2012
94	IEEE	Dini Rahmawati, Yo-Ping Huang	<i>Using C-support vector classification to forecast dengue</i>	2016

			<i>fever epidemics in Taiwan</i>	
100	IEEE	Nirbhay Mathur, Vijanth S. Asirvadam, Sarat C. Dass	<i>Generating vulnerability maps of dengue incidences for Petaling district in Malaysia</i>	2016
143	CAP ES	Marat Rafikov, Elvira Rafikova, and Hyun Mo Yang	<i>Optimization of the Aedes aegypti control strategies for integrated vector management</i>	2015
146	CAP ES	Kashish Ara Shakil, Shadma Anis, Mansaf Alam	<i>Dengue disease prediction using weka data mining tool</i>	2015
150	CAP ES	Florentino, Helenice O. ; Cantane, Daniela R. ; Santos, Fernando L.P. ; Bannwart, Bettina F.	<i>Multiobjective Genetic Algorithm applied to dengue control</i>	2014
152	CAP ES	Chen, Chia-Chern ; Chang, Hsien-Chang	<i>Predicting dengue outbreaks using approximate entropy algorithm and pattern recognition</i>	2013

Como indagado nas questões primárias de pesquisa, os artigos tiveram como objetivo predominante a utilização de técnicas computacionais como Inteligência Artificial, mineração de Dados e modelagens matemáticas, para também compreender o comportamento da epidemia, auxiliando no controle do vetor da dengue. As técnicas computacionais e os algoritmos de mineração de dados e inteligência artificial puderam auxiliar no acompanhamento e predição da propagação da doença.

5. Considerações Finais

Tendo em vista os aspectos observados, os dados apresentados no processo da Revisão Sistemática nos permitiram perceber que estudos utilizando técnicas computacionais envolvendo algoritmos de otimização e análise espacial dos dados, com enfoque em predição de casos de dengue, foram poucos discutidos até o momento. Logo, conclui-se que a área de algoritmos de otimização e mineração de dados, com foco em predição de dados é uma área em ascensão, todavia necessitando de um maior enfoque técnico-científico.

Como ficou evidenciado nos estudos, as técnicas computacionais como a mineração de dados, entre outras, têm sido utilizadas para tentar combater epidemias de dengue, por meio também de predições a partir da distribuição espacial dos dados. Percebeu-se também que as análises de dados presentes nos estudos observados utilizaram-se de ferramentas, como é o caso da ferramenta WEKA, com maior utilização, para as análises e mineração de dados.

Os resultados desta pesquisa proporcionam aos futuros trabalhos científicos no Brasil suporte na escolha das técnicas e ferramentas disponíveis, assim como algoritmos e inteligência artificial para o processo de predição e análise espacial dos dados referentes à dengue.

Referências

Balasundaram, A. & Bhuvaneswari, P. T. V. "Comparative study on decision tree based data mining algorithm to assess risk of epidemic," In IET Chennai Fourth International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2013), 2013, Chennai on (pp. 390-396). IEEE.

Brasil, Ministério da Saúde (2009) (A) "Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue" Disponível em:

- <http://bvsmis.saude.gov.br/bvs/publicacoes/diretrizes_nacionais_prevencao_controle_dengue.pdf>. Acesso em: 10 nov. 2017.
- Chen, C. & Chang, H. "Predicting dengue outbreaks using approximate entropy algorithm and pattern recognition." In *Journal of Infection*, Elsevier, 2013
- Florentino, H. O., Cantane, D. R., Santos, F. L. P. & Bannwart, B. F. "Multiobjective Genetic Algorithm applied to dengue control." In *Mathematical Biosciences*, Elsevier, 2014.
- Husin, N. A., Mustapha, N., Sulaiman, M. N. & Yaakob, R. "A hybrid model using genetic algorithm and neural network for predicting dengue outbreak," In *4th Conference on Data Mining and Optimization (DMO)*, 2012, Langkawi on (pp. 23-27). IEEE.
- Kitchenham, B. (2007) "Guidelines for performing Systematic Literature Reviews" In *Software Engineering*, version 2.3. Technical Report EBSE. Software Engineering Group. School of Computer Science and Mathematics Keele University.
- Kitchenham, B. (2004) "Procedures for Performing Systematic Reviews", Staffordshire: Keele University. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3308&rep=rep1&type=pdf>>. Acesso em: 2 out. 2017.
- Mathur, N., Asirvadam, V. S., Dass, S. C. & Gill, B. S. "Generating vulnerability maps of dengue incidences for Petaling district in Malaysia," In *IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA)*, 2016, Malacca City on (pp. 227-232). IEEE.
- Queiroz, P. G. G., & Braga, R. T. V. (2014). "Development of critical embedded systems using model-driven and product lines techniques: A systematic review." In *Software Components, Architectures and Reuse (SBCARS)*, 2014 Eighth Brazilian Symposium on (pp. 74-83). IEEE.
- Rafikov, M., Rafikova, E., & Yang, H. M. "Optimization of the Aedes aegypti Control Strategies for Integrated Vector Management," In *Journal of Applied Mathematics*, vol. 2015, 2015.
- Rahmawati, D. & Huang, Y. "Using C-support vector classification to forecast dengue fever epidemics in Taiwan," In *International Conference on System Science and Engineering (ICSSE)*, 2016, Puli on (pp. 1-4). IEEE.
- Shakil, K. A., Anis, S. & Alam, M. "Dengue disease prediction using weka data mining tool.", arXiv preprint arXiv:1502.05167, 2015.
- Tazkia, R. A. K., Narita, V. & Nugroho, A. S. "Dengue outbreak prediction for GIS based Early Warning System," In *International Conference on Science in Information Technology (ICSITech)*, 2015, Yogyakarta on (pp.121-125). IEEE.
- Torres, C., Barguil, S., Melgarejo, M. & Olarte, A. "Fuzzy model identification of dengue epidemic in Colombia based on multiresolution analysis." In *Artificial Intelligence in Medicine*, Elsevier, 2014.
- WHO. Revision of the International Health Regulations. World Health Assembly Resolution WHA58.3, adopted by the 58th World Health Assembly, 2005.