

## AVALIAÇÃO DE PADRÕES CLÍNICOS DE CASOS DE DENGUE UTILIZANDO REDES NEURAIIS NÃO-SUPERVISIONADAS (SOM)

G. A. Macedo\*, P. Brasil\*\*, e F.F. Nobre\*

\*Laboratório de Engenharia de Sistemas de Saúde, PEB/ COPPE / UFRJ, Rio de Janeiro-RJ, Brasil.

\*\*Laboratório de Doenças Febris Agudas, Instituto Nacional de Infectologia-INI/FIOCRUZ, Rio de Janeiro-RJ, Brasil.

e-mail: gleicy.amorim@gmail.com

**Resumo:** Este estudo teve como objetivo analisar os perfis clínicos de 908 pacientes com suspeita de dengue por meio de mapas auto-organizáveis (SOM) e algoritmo de análise de aglomerados K-médias para a identificação de grupos de pacientes com padrões similares. Três grupos naturais foram identificados, sendo um grupo com características intermediárias e casos não classificados, um com casos maior frequência de graves e sinais de alarme e outro com características de dengue clássico. Não houve concordância entre a classificação feita por especialistas e o SOM ( $Kappa < 0$ ), pois o SOM foi capaz de agrupar todos os casos inclusive os casos que não foram classificados por falta de dados ou casos descartados.

**Palavras-chave:** dengue, SOM, classificação.

**Abstract:** This study aimed to analyze the clinical profiles of 908 patients with suspected of dengue fever using self-organizing maps (SOM) algorithm and K-means cluster analysis in order to identify groups of patients with similar patterns. Three natural groups were identified, one with intermediate characteristics and discarded cases, another with potential severe cases with warning signs and one group with characteristics of classic dengue. There was no agreement between the classification made by experts and the SOM ( $Kappa < 0$ ), because the SOM was able to group all cases including cases that were not classified by experts due to missing data or that cases which were discarded.

**Keywords:** classification, dengue, Self-Organized map (SOM).

### Introdução

A dengue é uma doença aguda, sistêmica e de etiologia viral, ocasionada pelo vírus do dengue ou DENV, que apresenta um amplo espectro clínico, variando desde infecções assintomáticas a formas graves [1]. Grande parte dos indivíduos infectados apresenta a forma branda da doença conhecida como febre do dengue, dengue sem sinais de alarme ou clássico, porém uma proporção dos pacientes evolui para formas mais graves [2]. Em períodos epidêmicos, o diagnóstico pode ser feito através de critério clínico-epidemiológico, porém a sensibilidade para identificar pacientes com dengue e com maior

probabilidade de desenvolver doença grave ainda é muito baixa [1].

De acordo com a nova classificação proposta pela Organização Mundial da Saúde (OMS) em 2009 [2], a dengue pode ser classificada em caso suspeito de dengue ou dengue sem sinais de alarme, dengue com sinais de alarme e dengue grave. Um estudo multicêntrico conduzido por especialistas de 18 países para avaliar esta nova classificação demonstrou uma maior sensibilidade quando aplicada em regiões endêmicas como o Brasil. Porém, a necessidade de investigação sobre sinais de alarme e uma melhor definição clínica de caso de dengue na ausência de testes laboratoriais, em especial, foi destacada [3].

Alguns trabalhos vêm discutindo o uso em conjunto do conhecimento médico e de técnicas de inteligência computacional como as redes neurais como uma possível ferramenta para auxiliar os clínicos para prever e investigar o risco em pacientes suspeitos de dengue [4, 5]. Redes Neurais Artificiais são técnicas computacionais não-lineares que utilizam um modelo matemático para a modelagem de problemas complexos, concebidos como um sistema integrado de entradas e saídas. Uma abordagem no uso de redes neurais seria o uso de mapas auto-organizáveis (SOM do inglês *Self Organizing Map*) que visa encontrar padrões naturais em um determinado conjunto de dados, sendo esta uma técnica de escolha para visualização e compreensão da complexidade de dados com altas dimensões [6]. Nesse contexto, esse estudo tem como objetivo analisar o perfil clínico-laboratorial de casos suspeitos de dengue usando mapas auto-organizáveis (SOM) e K-médias para reconhecer padrões que melhor caracterizem o perfil clínico dos casos suspeitos de dengue. Os grupos encontrados foram avaliados quanto à frequência das variáveis clínicas e laboratoriais presentes em cada grupo e comparados quanto à concordância com a nova classificação proposta pela OMS em 2009 e usada pelos especialistas.

### Materiais e métodos

**Dados** Os dados utilizados se referem à primeira consulta dos casos suspeitos de dengue do Laboratório de Doenças Febris Agudas (DFA) do Instituto de Pesquisa Clínica Evandro Chagas (IPEC/FIOCRUZ). Para este estudo foram selecionados 908 pacientes entre 2007-2013

e 30 variáveis clínico-laboratoriais. As variáveis clínicas dicotômicas (0 para a ausência e 1 para a presença da característica em questão) foram normalizadas usando a função *som\_normalize* do pacote *somtoolbox* e as variáveis laboratoriais contínuas foram normalizadas por *Zscore*.

**Metodologia** O Mapa Auto-organizável de Kohonen (SOM do inglês *Self Organizing Map*) é uma rede neural com aprendizado não supervisionado, capaz de mapear um conjunto de dados de entrada em um conjunto de neurônios normalmente representados em uma ou duas dimensões [6]. Cada neurônio  $mi$  no mapa representa um conjunto de exemplos vizinhos entre si, formando grupos de acordo com a distância euclidiana entre eles. O tamanho do mapa foi escolhido com base no menor erro de quantização, no erro topológico e no número de neurônios ativados. O erro topológico ou  $Te$  (Equação 1) representa a proporção de vetores de dados na qual os neurônios vizinhos e o neurônio vencedor não são unidades adjacentes, enquanto o erro de quantização ou  $Qe$  (Equação 2) expressa a distância média entre cada vetor de entrada e o neurônio para  $k$  fatores.

$$Te = \frac{1}{n} \sum_{k=1}^n u(x_k) \quad (1)$$

Onde  $u(x_k)$  corresponde a 1 se o neurônio vencedor (*Best Match Unit*- BMU1) e o segundo neurônio vencedor (BMU2) são vizinhos e 0 se eles não o são.

$$Qe = \frac{\sum_{k=1}^n \|x_k - \omega_{BMU}\|}{n} \quad (2)$$

Onde o erro médio correspondente à diferença entre o vetor de entrada ( $x_k$ ) e vetor de pesos ( $\omega_{BMU}$ ).

O algoritmo de partição em  $k$ -médias, conforme descrito por Macqueen [7] foi utilizado para agrupar os neurônios de acordo com suas similaridades, possibilitando a definição do número de grupos entre eles. O método pode ainda ser avaliado por metodologias quanto à qualidade dos exemplos distribuídos pelos  $k$  grupos através de técnicas de validação, sendo que nesse estudo foi utilizado o índice de Davies-Bouldin [8].

O índice Davies-Bouldin é baseado na medida de similaridade ( $R_{ij}$ ) dos clusters. O  $R_{ij}$  baseia-se na medida de dispersão nos clusters  $s_i$  e medida de distância  $d_{ij}$ , independente do número de agrupamentos e do método de partição utilizada. A medida de similaridade do cluster  $R_{ij}$  é definida na Equação 3 abaixo.

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \quad (3)$$

Onde  $d_{ij}=d(v_i,v_j)$  e  $s_i = \frac{1}{|c_i|} \sum_{x \in c_i} d(x,v_i)$ .

O índice de Davies-Bouldin é descrito na Equação 4 abaixo.

$$DB = \frac{1}{n_c} \sum_{i=1}^{n_c} R_i \quad (4)$$

Onde  $R_i = \max_{j=1, \dots, n, i \neq j} (R_{ij})$  e  $n_c$  é o número de clusters.

Um bom agrupamento dos dados é definido pelo valor mais baixo, o qual indica uma boa partição.

Os grupos definidos pelo SOM e a classificação do especialista (OMS 2009) foram comparados utilizando o índice *Kappa*. O índice *Kappa* mede o grau de concordância entre diferentes técnicas, calculando a divisão da diferença entre a concordância esperada e a concordância observada e a diferença entre a concordância absoluta e a concordância esperada como descritos na Equação 5 abaixo.

$$K = \frac{C_o - C_e}{1 - C_e} \quad (5)$$

Onde  $K$  indica o índice *Kappa*,  $C_o$  indica a concordância observada e  $C_e$  a concordância esperada.

A interpretação do índice *Kappa* se dá por teste de hipótese onde a hipótese nula ou  $H_0$  é que o  $K=0$ , o que indicaria concordância nula, ou se ele é maior que zero (hipótese alternativa,  $H_1>0$ ), o que indica uma concordância maior que o acaso (teste monocaudal:  $H_0: K=0; H_1: K>0$ ). Um valor de *Kappa* com valor máximo (um) 1 indica total concordância, enquanto valores negativos sugerem que a concordância encontrada foi inferior àquela esperada ao acaso.

### Aspectos Éticos

O estudo atendeu às diretrizes éticas em pesquisa envolvendo seres humanos (Portaria CNS no 466/12), com registro prévio no Comitê de Ética em Pesquisa do Instituto Nacional de Infectologia (INI) de número CAAE: 0026.0.009.000-07.

### Resultados

De acordo com a Figura 1 o menor erro topográfico (0,03), e de quantização (1,15) avaliados conjuntamente com o maior número de neurônios ativados definiu o tamanho do mapa sendo composto por 130 neurônios distribuídos na forma retangular.

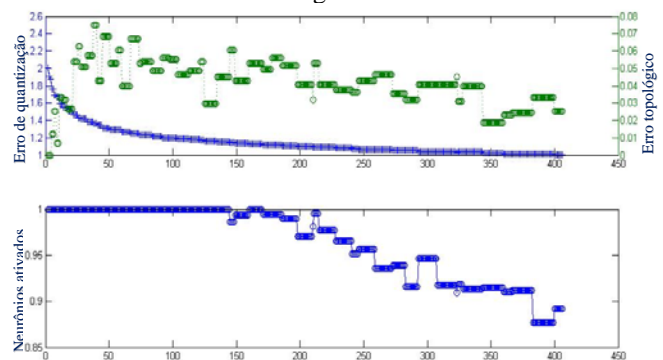


Figura 1. Erro topográfico e de quantização de acordo com o número de neurônios ativados.

A matriz-U e número de grupos encontrados pelo algoritmo de  $K$ -médias podem ser visualizados na Figura 2. As tonalidades das cores indicam a proximidade das unidades sendo que mais próximo da cor vermelha maior distância e quanto mais próxima da cor azul menor distância.

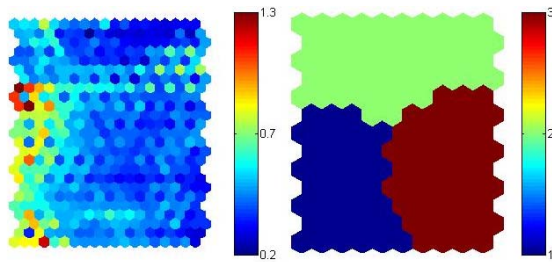


Figura 2. Matriz U e os três grupos encontrados pelo K-médias.

Após fazer o SOM seguido do algoritmo de K-médias, o índice de Davies-Bouldin avaliou o melhor número de clusters como sendo três grupos (valor mínimo).

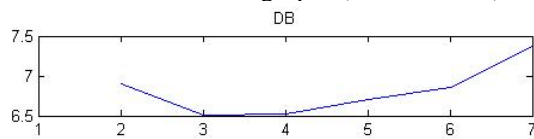


Figura 3. Melhor número de grupos definido índice de Davies-Bouldin.

Os três grupos encontrados foram avaliados quanto à frequência de sinais e sintomas, características clínicas por grupo e comparados com a classificação da OMS 2009 utilizada pelo especialista em dengue (Tabela 1, 2 e 3).

Tabela 1. Características clínicas dos grupos encontrados.

Características	Grupo 1		Grupo 2		Grupo 3	
	220	(%)	341	(%)	347	(%)
Cefaléia	179	(81,4)	227	(66,5)	294	(84,7)
Exantema	65	(29,5)	134	(39,3)	151	(43,5)
Petéquiás	31	(14)	135	(39,6)	37	(10,6)
Paladar alterado	84	(38,2)	43	(12,6)	163	(47,0)
Disúria	76	(34,5)	103	(30,2)	107	(30,8)
Tosse	61	(27,7)	59	(17,3)	74	(21,3)
Dispnéia	19	(8,6)	42	(12,3)	26	(7,5)
Náusea	18	(8,2)	44	(12,9)	25	(7,2)
Sonolência	16	(7,2)	54	(15,8)	28	(8,0)
Dor abdominal*	71	(32,3)	220	(64,5)	68	(19,6)
Tonteira	44	(20,0)	57	(16,7)	91	(26,2)
Desidratação	41	(18,6)	100	(29,3)	67	(19,3)
Eritema	35	(15,9)	59	(17,3)	96	(27,6)
Adenomegalia	23	(10,4)	30	(8,8)	44	(12,7)
Derrame Pleural*	12	(5,4)	107	(31,4)	8	(2,3)
Sangramentos	14	(6,3)	59	(17,3)	27	(7,8)
Ausculta pulmonar	12	(5,4)	81	(23,7)	7	(2,0)
Hepatomegalia*	15	(6,8)	81	(23,7)	10	(2,8)
Vômitos persistentes*	18	(8,2)	47	(13,8)	4	(1,1)
Hemorragias	12	(5,5)	61	(17,9)	6	(1,7)
Sono/Agitação*	9	(4,1)	39	(11,4)	7	(2,0)
Aumento do Hematócrito*	20	(9,1)	140	(41,0)	4	(1,1)
Queda de Plaquetas*	27	(12,3)	234	(68,6)	15	(4,3)
Desconforto respiratório	12	(5,4)	58	(17,0)	5	(1,4)
Enchimento capilar	7	(3,2)	53	(15,5)	2	(0,6)

\*Sinais de alarme conforme OMS 2009.

Tabela 2. Distribuição do sexo e mediana das características laboratoriais entre os grupos encontrados.

Características	Grupo 1	Grupo 2	Grupo 3
Sexo M	103	173	176
Sexo F	117	168	171
Idade (anos)	28	10	43
Dias após o início da febre	4	4	3
Leucograma*	7.695	4.270	3.850
Plaquetas#	241.500	51.900	164.000

\*VR.: 4.000 a 11.000/mm<sup>3</sup> # VR.: 140.000 a 450.000/mm<sup>3</sup>.

Tabela 3. Comparação entre a classificação da OMS (2009) e grupos definidos pelo SOM.

Classificação	SOM					
	Grupo 1	(%)	Grupo 2	(%)	Grupo 3	(%)
Especialistas						
DSSA*	82	(37,3)	71	(20,8)	218	(62,8)
DCSA#	41	(18,6)	143	(42,0)	69	(19,8)
DG**	23	(10,4)	105	(30,8)	11	(3,2)
NC~	74	(33,7)	22	(6,4)	49	(14,2)
Confirmados+	86	(39,0)	246	(72,1)	281	(81,0)
Total	220	(100)	341	(100)	347	(100)

\*Dengue sem sinais de alarme, #Dengue com sinais de alarme, \*\*Dengue Grave, ~Casos não classificados por falta de dados. + Casos confirmados por exame laboratorial específico.

De acordo com os resultados apresentados na Tabela 1 os grupos naturais encontrados apresentaram diferentes perfis clínicos. No grupo 2 todos os sinais de alarme foram encontrados com maior predominância que nos grupos restantes.

Os grupos 1 e 3 apresentaram perfis próximos caracterizados pela forma mais branda da doença, embora no grupo 3, os sintomas paladar alterado (47%), exantema (43,5%) e eritema (27,6%) foram predominantes. Nota-se também que neste grupo 62,8% dos casos foram classificados como sendo dengue sem sinais de alarme enquanto que no grupo 1, 33,7% foram casos não classificados por falta de informação ou porque o resultado laboratorial foi negativo (Tabela 3). Além disso, 81% dos pacientes do grupo 3 obtiveram a confirmação laboratorial comparados a apenas 39% do grupo 1.

O grupo 2 se caracterizou como grave e potencialmente grave (30,8% dos pacientes classificados como dengue grave e 42% dos pacientes como dengue com sinais de alarme). Este grupo apresentou a menor faixa etária entre os grupos (Tabela 2), o que estabelece um perfil peculiar a menores de 15 aos (mediana de 10 anos). Este grupo também apresentou uma menor contagem de plaquetas (mediana de 51.900/mm<sup>3</sup>) enquanto o grupo 3 apresentou a menor contagem de leucócitos (mediana de 3.850/mm<sup>3</sup>).

O SOM seguido do algoritmo de K-médias obteve uma acurácia de 26% (IC:23%-28%) quando comparado com a classificação do especialista. Não houve concordância entre os grupos definido pelo SOM e a classificação do especialista (Kappa <0, p>0.05).

## Discussão

Em alguns países tropicais e subtropicais onde além da dengue, outras viroses são doenças predominantes, vê-se a necessidade de uma melhor definição do caso dengue a fim de diferenciá-lo de outras doenças febris. Em períodos

epidêmicos muitas vezes o diagnóstico é clínico, não havendo tempo hábil ou recursos para o diagnóstico laboratorial específico. Pacientes com dengue são frequentemente hospitalizados para monitorização e tratamento de suporte, pois não há ferramentas diagnósticas ou prognósticas eficazes capazes de distinguir pacientes com dengue de outras doenças febris e os pacientes clássicos dos pacientes que podem desenvolver a forma grave da doença. Esse procedimento resulta em um grande número de internações como aconteceu no Rio de Janeiro em 2008 [9].

Neste estudo, ao analisar casos suspeitos de dengue em pacientes de fase aguda, notamos que o grupo que se caracterizou por ter um perfil mais próximo do quadro clássico, apresentou maior frequência de sintomas como exantema, alteração do paladar e eritema quando comparados com o grupo onde houve um maior número de casos descartados (Tabela 1). Os resultados também mostram que a baixa contagem de leucócitos (leucopenia) e contagem de plaquetas (plaquetopenia), são mais frequentes nos casos de dengue e casos com risco gravidade como observados nos grupos 3 e 2, respectivamente (Tabela 2). Estudos adicionais poderiam avaliar melhor essas características clínicas como possíveis marcadores para definição de caso dengue.

Desde a epidemia mais grave de dengue ocorrida em 2007-2008 na qual houve um aumento significativo no número de crianças acometidas pela forma mais grave da doença, que este cenário vem prevalecendo no Brasil. [9]. Notou-se que o grupo com menor faixa etária possui um perfil peculiar associado à gravidade. A gravidade em menores de 18 anos no Brasil precisa ainda ser melhor explorada principalmente no que diz respeito à definição dos sinais de alarme associados ao perfil mais grave da doença.

Estudos recentes avaliaram a nova classificação proposta pela OMS em 2009 como sendo muito abrangente, com alta sensibilidade e especificidade intermediária [1, 10]. Ao comparar os resultados do SOM com a classificação utilizada pelos especialistas, observou-se que os casos agrupados pelo SOM se aproximam do perfil proposto pela OMS (2009) no entanto, o SOM foi capaz de agrupar todos os casos, inclusive os casos não classificados pelos especialistas por falta de dados o que justifica a acurácia de 26% e índice  $Kappa < 0$ . Estes casos foram mais frequentes (33,7%) no grupo 1, o que leva a crer que esse grupo reflete a dificuldade dos médicos em definir qual paciente poderia ser dengue ou outra doença febril.

## Conclusão

A metodologia do SOM aplicada em pacientes na fase aguda da doença foi capaz de agrupar distintamente os casos que foram alocados como não classificados pelos especialistas, dos grupos onde houve maior confirmação laboratorial. O presente estudo demonstrou que o uso de tais ferramentas computacionais poderia auxiliar os clínicos a identificar e tratar adequadamente pacientes com

suspeita de dengue, prevenindo assim as formas grave da doença e conduzindo-os para uma boa evolução clínica.

## Agradecimentos

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e a Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro – FAPERJ pelo o apoio financeiro.

## Referências

- [1] Simmons CP, Farrar JJ, Nguyen van VC, Wills B. Dengue. *N Engl J Med*. 2012 Apr 12; 366(15):1423.
- [2] WHO | Dengue: guidelines for diagnosis, treatment, prevention and control [Internet]. WHO. [cited 2013 Jun 10]. Available from: [http://www.who.int/csr/resources/publications/dengue\\_9789241547871/en/](http://www.who.int/csr/resources/publications/dengue_9789241547871/en/)
- [3] Barniol J, Gaczkowski R, Barbato EV, da Cunha RV, Salgado D, Martínez E, et al. Usefulness and applicability of the revised dengue case classification by disease: multi-centre study in 18 countries. *BMC Infect Dis*. 2011;11:106.
- [4] Faisal T, Taib MN, Ibrahim F. Reexamination of risk criteria in dengue patients using the self-organizing map. *Med Biol Eng Comput*. 2010 Mar;48(3):293–301.
- [5] Ibrahim F, Faisal T, Salim MIM, Taib MN. Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network. *Med Biol Eng Comput*. 2010 Nov; 48(11):1141–8.
- [6] Kohonen T. The self-organizing map. *Proceedings of the IEEE*. 1990; 78(9):1464–80.
- [7] MacQueen, JB. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*; 1967; University of California Press.
- [8] Pétróvic, S. A comparison between the Silhouette Index and the Davies-Bouldin Index in labelling IDS clusters. 2006. p. 53–64.
- [9] Gibson G, Souza-Santos R, Brasil P, Pacheco AG, Cruz OG, Honório NA, et al. From primary care to hospitalization: clinical warning signs of severe dengue fever in children and adolescents during an outbreak in Rio de Janeiro, Brazil. *Cad Sad. Pub*. 2013 Jan; 29(1):82–90.
- [10] Horstick O, Farrar J, Lum L, Martinez E, San Martin JL, Ehrenberg J, et al. Reviewing the development, evidence base, and application of the revised dengue case classification. *Pathog Glob Health*. 2012 May;106(2):94–101.