



Artigo/Article

A utilização de uma rede neural artificial para previsão da incidência da malária no Município de Cantá, Estado de Roraima

Use of an artificial neural network to predict the incidence of malaria in the City of Cantá, State of Roraima

Guilherme Bernardino da Cunha^{1,2}, José Francisco Luitgards-Moura^{2,3,4}, Eduardo Lázaro Martins Naves⁵, Adriano Oliveira Andrade⁵, Adriano Alves Pereira⁵ e Selma Terezinha Milagre⁵

RESUMO

Introdução: A malária é uma doença endêmica na Amazônia Legal Brasileira, apresentando riscos diferentes para cada região. O Município de Cantá, no Estado de Roraima, apresentou para todo o período estudado, um dos maiores índices parasitários anuais do Brasil, com valor sempre maior que 50. O presente estudo visa à utilização de uma rede neural artificial para previsão da incidência da malária nesse município, a fim de auxiliar os coordenadores de saúde no planejamento e gestão dos recursos. **Métodos:** Os dados foram coletados no site do Ministério da Saúde, SIVEP - Malária entre 2003 e 2009. Estruturou-se uma rede neural artificial com três neurônios na camada de entrada, duas camadas intermediárias e uma camada de saída com um neurônio. A função de ativação foi à sigmoide. No treinamento, utilizou-se o método *backpropagation*, com taxa de aprendizado de 0,05 e momentum 0,01. O critério de parada foi atingir 20.000 ciclos ou uma meta de 0,001. Os dados de 2003 a 2008 foram utilizados para treinamento e validação. Compararam-se os resultados com os de um modelo de regressão logística. **Resultados:** Os resultados para todos os períodos previstos mostraram-se que as redes neurais artificiais obtiveram um menor erro quadrático médio e erro absoluto quando comparado com o modelo de regressão para o ano de 2009. **Conclusões:** A rede neural artificial se mostrou adequada para um sistema de previsão de malária no município estudado, determinando com pequenos erros absolutos os valores preditivos, quando comparados ao modelo de regressão logística e aos valores reais.

Palavras-chaves: Predição. Malária. Rede neural artificial. Cantá. Estado de Roraima. Regressão logística.

ABSTRACT

Introduction: Malaria is endemic in the Brazilian Amazon region, with different risks for each region. The City of Cantá, State of Roraima, presented one of the largest annual parasite indices in Brazil for the entire study period, with a value always greater than 50. The present study aimed to use an artificial neural network to predict the incidence of malaria in this city in order to assist health coordinators in planning and managing resources. **Methods:** Data were collected on the website of the Ministry of Health, SIVEP - Malaria between 2003 and 2009. An artificial neural network was structured with three neurons in the input layer, two intermediate layers and an output layer with one neuron. A sigmoid activation function was used. In training, the backpropagation method was used, with a learning rate of 0.05 and momentum of 0.01. The stopping criterion was to reach 20,000 cycles or a target of 0.001. The data from 2003 to 2008 were used for training and validation. The results were compared with those from a logistic regression model. **Results:** The results for all periods provided showed that the artificial neural network had a smaller mean square error and absolute error compared with the regression model for the year 2009. **Conclusions:** The artificial neural network proved to be adequate for a malaria forecasting system in the city studied, determining smaller predictive values with absolute errors compared to the logistic regression model and the actual values.

Key-words: Prediction. Malaria. Artificial neural network. Cantá. State of Roraima. Logistic regression.

1. Curso de Ciência da Computação, Universidade Federal de Roraima, Boa Vista, RR. 2. Núcleo Avançado de Vetores, Universidade Federal de Roraima, Boa Vista, RR. 3. Curso de Medicina, Universidade Federal de Roraima, Boa Vista, RR. 4. Projeto Observatório de Doenças Tropicais, Universidade Federal de Roraima, Boa Vista, RR. 5. Curso de Engenharia Biomédica, Universidade Federal de Uberlândia, Uberlândia, MG.

Endereço para correspondência: Dr. Guilherme Bernardino da Cunha. DCC/UFRR. Av. Ene Garcez 2413, Aeroporto, 69305-000, Boa Vista, RR.
e-mail: guilherme.ufrr@gmail.com

Recebido para publicação em 20/04/2010

Aceito em 13/07/2010

INTRODUÇÃO

A malária representa um dos sérios problemas de saúde humana no século XXI, de acordo com a Organização Mundial da Saúde¹. Atualmente, a doença está confinada às áreas pobres da África, Ásia e América Latina. No Brasil, aproximadamente 99% dos casos de malária se concentram na região compreendida como Amazônia Legal, composta pelos Estados do Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins². Mesmo que a Amazônia Legal seja uma área endêmica, o risco de contrair a infecção não é uniforme para toda a região e localidade. Este risco é medido pelo índice parasitário anual (IPA) que classifica as áreas de transmissão em: baixo (IPA < 10), médio (IPA ≥ 10 e < 50) e alto (IPA ≥ 50)^{1,2}.

O Estado de Roraima apresentou um dos maiores IPA do Brasil, no período compreendido entre 1990 e 2005. Dentre os Municípios do Estado de Roraima, o Cantá se destaca como um dos maiores IPAs do estado e do Brasil, apresentando sempre um alto risco de infecção, ou seja, com IPA ≥ 50, com um pico máximo em 2005 com IPA igual a 615,9 casos por mil habitantes.

A vigilância epidemiológica do Estado de Roraima tem como propósito, fornecer orientação técnica permanente para os responsáveis pela decisão e execução de ações de controle de doenças e agravos. Para subsidiar esta atividade, deve tornar disponíveis informações atualizadas sobre a ocorrência dessas doenças ou agravos, bem como dos seus fatores condicionantes, em uma área geográfica ou população determinada. A vigilância epidemiológica constitui-se, ainda, em importante instrumento para o planejamento, a organização e a operacionalização dos serviços de saúde, como também para a normatização de atividades técnicas correlatas. Do ponto de vista estatístico, detectar uma epidemia é perceber a presença de picos (valores) fora do padrão de uma série histórica. A vigilância epidemiológica adota como prática rotineira

em alguns países, a detecção desses valores com base na notificação dos últimos cinco anos.

A técnica de previsão (predição para alguns autores) em vigilância epidemiológica tem como o principal objetivo a projeção das necessidades futuras para a saúde pública, ou a detecção de algum comportamento de uma série temporal indicando a ocorrência da incidência da malária. Algumas das técnicas empregadas para a previsão e predição incluem o ajuste de funções como as redes neurais artificiais (RNA), os modelos de regressão, de suavização exponencial e os modelos ARIMA³⁻⁵.

As redes neurais artificiais (RNAs) foram concebidas nos últimos anos para realizar tarefas complexas em diferentes áreas e como uma estratégia de modelagem matemática de problemas, concebidos como sistemas de entradas e saídas. Ao contrário de outras estratégias de modelagem, não é necessário conhecer a relação matemática entre as entradas e saídas. Assim, ao contrário do modelo de regressão utilizado como comparativo neste trabalho, não é preciso propor uma função para o modelo, uma vez que as RNAs são aproximadores universais de funções. A RNA solucionaria qualquer problema que ela possa representar⁶. Exemplos incluem o reconhecimento de padrões de som e imagem, a previsão de séries temporais, simulação, controle e otimização de sistemas. Especificamente, em relação às séries temporais, as aplicações de redes neurais podem ser divididas em dois grupos: aquelas que são predominantemente determinísticas e, as séries temporais com uma forte componente aleatória, com no caso dos dados epidemiológicos. A aplicação das RNAs para o segundo grupo ainda é muito recente no âmbito de vigilância epidemiológica^{7,8}.

O presente estudo visa à aplicação de uma rede neural artificial para a previsão da incidência da malária no Município de Cantá, Estado de Roraima, em três períodos distintos, curto (3 meses), médio (6 meses) e longo (12 meses). Os resultados serão comparados ao modelo de regressão logística utilizando como parâmetros, os valores preditivos, o erro absoluto e o erro quadrático médio (EQM) para os três períodos.

MÉTODOS

Os dados da série histórica foram extraídos do sistema do Ministério da Saúde chamado de SIVEP - Malária, entre os anos de 2003 e 2009. Esse sistema foi criado em 2003 com o intuito de oferecer uma maior confiabilidade dos dados perante o sistema anterior².

O Município de Cantá no Estado de Roraima foi escolhido devido a sua representação no cenário atual quando se trata de municípios brasileiros com maiores taxas de incidências de malária. Essas taxas são calculadas através do índice parasitário anual (IPA). Cantá apresentou entre todos os anos analisados, um IPA maior que 50, ou seja, um alto risco de contrair a enfermidade, conforme a classificação da Organização Mundial de Saúde¹.

Como técnica estatística para comparação, foi utilizado o modelo de regressão exponencial, bastante comum em aplicações com séries temporais expressas por dados epidemiológicos. O modelo de regressão logística é muito versátil, permitindo a sua aplicação em diversas áreas, principalmente na modelagem de sistemas epidemiológicos^{5,9}.

A RNA escolhida para representar o modelo de previsão foi uma multicamada (*feedforward*), com três neurônios na camada de entrada, correspondente aos três meses sucessivos, duas camadas

intermediárias ou ocultas, sendo a primeira, com oito neurônios e a segunda com cinco neurônios e a camada de saída com somente um neurônio. A função de ativação para todas as camadas foi a função sigmoide. O treinamento foi realizado pelo método *backpropagation*, com taxa de aprendizado de 0,05 e o *momentum* de 0,01. O critério para o fim do treinamento foi atingir 20.000 *epochs* ou a meta (*goal*) de 0,001. Utilizou-se como ferramenta de treinamento, validação e previsão, o software Matlab R2007a[®].

Antes de iniciar o treinamento, fez-se necessário normalizar os dados de entrada na RNA. Essa normalização é muito importante para adequar os dados ao padrão estabelecido pela RNA. A normalização utilizada foi no intervalo entre 0 e 1¹⁰. Os pesos são inicializados com valores pequenos e randômicos. Para o treinamento e ajustamento dos modelos, foram utilizados os dados de janeiro de 2003 a dezembro de 2008, correspondendo a 72 observações. Os dados de 2008 foram utilizados para validar o treinamento da RNA e calcular o erro quadrático médio (EQM). Se o EQM for maior que 50 para o longo prazo (12 meses), o treinamento será repetido até que o EQM seja menor que o valor definido, ou seja, menor que 50. Os valores referentes a 2009 foram obtidos para validar o modelo como ferramenta preditiva.

Para o modelo de regressão logística, os dados utilizados são do período de janeiro de 2003 a dezembro de 2008, sem a necessidade de nenhuma adequação, ou seja, pré-processamento¹¹⁻¹³. Para obter o modelo de regressão logística, utilizou-se a ferramenta SPSS *Statistics* 17.0[®]. Os dados de 2009 foram obtidos a partir do modelo de regressão encontrado.

RESULTADOS

A **Figura 1** mostra a evolução do IPA em Cantá e no Estado de Roraima, no período compreendido entre 2003 e 2008. Tanto em Cantá quanto em Roraima, o IPA apresentou uma queda a partir de 2005 até 2008. Em 2009, houve um aumento. No Município de Cantá, o menor IPA registrado foi de 132,4 no ano de 2003 e o maior de 615,9 em 2005, com IPA sempre maior que 100 para todo o período analisado.

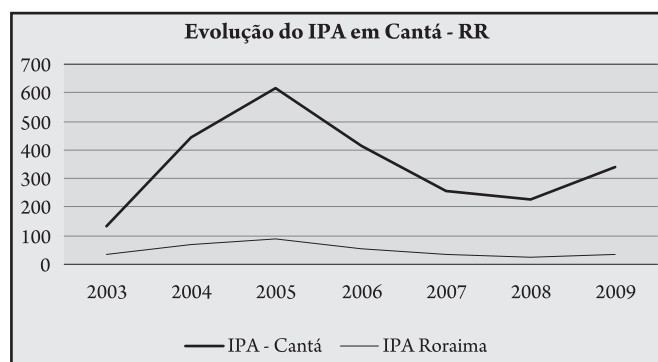


Figura 1 - Evolução do IPA no Município de Cantá (RR), em relação a Roraima.

Na **Figura 2**, encontra-se a curva do modelo de regressão logística para o Município de Cantá utilizando os dados de incidência da série histórica. Observa-se que a curva possui um comportamento crescente quanto aos valores de incidência futuros. Esses valores sempre estiveram acima de 100, exceto pelos primeiros meses de 2003.

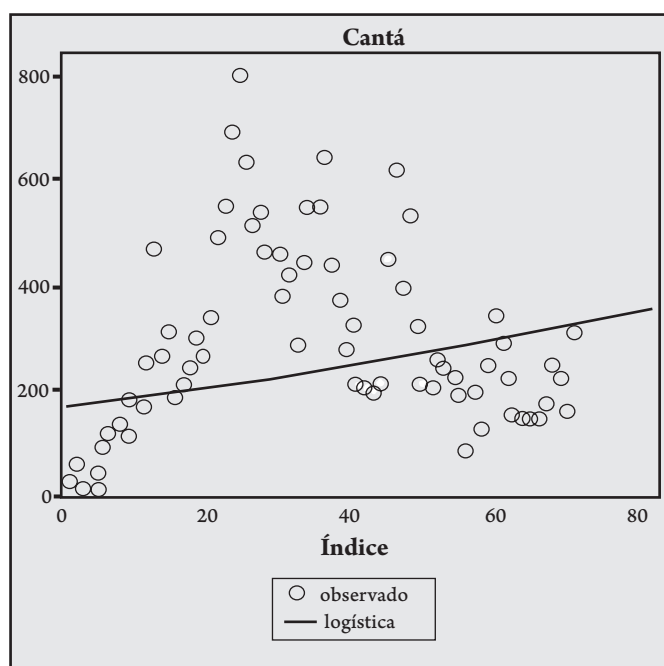


FIGURA 2 - Curva do modelo de regressão logística para o Município de Cantá.

A equação encontrada para representar os dados referentes ao Município de Cantá no período entre 2003 e 2008 é mostrada na equação 1 abaixo:

$$Y = 1 / (0 + 0,00567 * 0,9915^{**} X) \quad \text{Equação 1}$$

Os resultados encontrados para os três períodos distintos juntamente com o erro absoluto e o EQM utilizando as RNAs e o modelo de regressão, são mostrados na **Tabela 1**.

Por meio da **Tabela 1**, verifica-se que para os três períodos, que tanto o EQM quanto o erro absoluto foi menor quando se utiliza as RNAs em relação ao modelo de regressão logística. A diferença entre os erros absolutos gerados por meio das RNAs e do modelo de regressão é significativa. À medida que o prazo de previsão foi aumentando, o EQM e o erro absoluto também aumenta nos dois modelos estudados.

TABELA 1 - Resultados encontrados utilizando as RNAs e o modelo de regressão para os três períodos.

Período	Curto (3 meses)	Médio (6 meses)	Longo (12 meses)
EQM RNA	104,33	146,17	611,92
EQM regressão	11.897,13	14.485,20	20.816,41
Erro absoluto RNA	29	57	255
Erro absoluto Regressão	301,17	649,54	1576,07

EQM: erro quadrático médio. RNA: redes neurais artificiais.

A **Figura 3** mostra a previsão da incidência da malária utilizando a RNA e o modelo de regressão logística para um período longo, ou seja, 12 meses.

Conforme a **Figura 3**, observa-se que para a previsão utilizando RNA num longo período, o maior erro absoluto foi de 47 (quarenta e sete), encontrado no nono mês, ou seja, setembro de 2009 e, o menor foi de 5 (cinco), para o quarto mês, ou seja, abril de 2009. Os casos notificados no município sofrem grande aumento a partir de setembro.

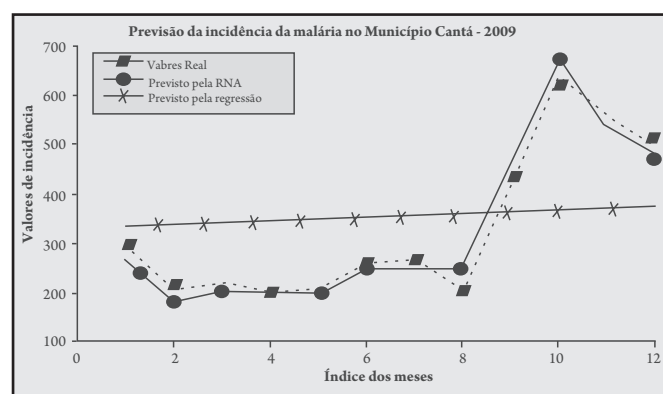


FIGURA 3 - Previsão a longo prazo da incidência da malária em Cantá no ano de 2009.

DISCUSSÃO

O objetivo final deste estudo foi investigar viabilidade de um modelo baseado em redes neurais artificiais para previsões séries temporais epidemiológicas. Embora, importantes critérios que definem uma estrutura de uma RNA como a quantidade de entradas, camadas intermediárias e parâmetros de treinamento são importantes pré-requisitos para uma boa avaliação e ajuste do modelo.

A escolha da arquitetura da RNA foi uma rede neural multicamadas (*feedforward*) devido a sua grande aplicabilidade em problemas que envolvem aproximações funcionais como a série temporal estudada.

A maior dificuldade da RNA foi obter os parâmetros que melhor se comportam com os dados apresentados. Para treinar os dados, utilizaram-se entradas com 3, 5, 7 e 9 neurônios. As taxas de aprendizagem, momentum, meta, quantidade de neurônios na camada escondida, quantidade de ciclos e o algoritmo de treinamento foram obtidos após exaustivos testes^{7,11}.

Os resultados obtidos pela RNA através dos três períodos, curto, médio e longo, mostraram-se adequados para o sistema de previsão de incidência da malária conforme ilustra a **Figura 3** e a **Tabela 1**. As redes neurais apresentaram um maior poder de previsão ao modelo de regressão logística, diante dos dados da série histórica da malária no Município de Cantá, Estado de Roraima.

O modelo de regressão apontou que para o Município de Cantá, os valores da incidência da malária seriam crescentes para todo o período estudado e para os valores futuros. Observando os dados reais, entre 2003 e 2009, constata-se que a partir de 2003 os valores foram crescentes até o ano de 2005, onde o município apresentou a maior incidência de malária com 6.041 casos confirmados. A partir de 2006, os valores foram decrescendo até o ano de 2009, onde apresentou um leve acréscimo com 4.058 casos confirmados de malária.

Os erros quadráticos médios (EQM) obtidos através das RNAs foram bem menores que as obtidas pelo método de regressão exponencial mostrando a capacidade de aprendizagem e previsão de uma rede neural artificial, baseado em séries temporais epidemiológicas.

A RNA é uma área a ser melhorada e desenvolvida dentro da previsão em sistemas epidemiológicos devido a sua grande aplicabilidade em aproximações funcionais. A principal dificuldade no seu uso é a pouca familiaridade dos pesquisadores em geral, já que se trata de um método recente em relação aos métodos estatísticos^{12,14}. O critério de seleção de uma rede é obtido de modo pragmático, ou seja, a que melhor atinge os resultados esperados.

AGRADECIMENTOS

Os autores agradecem a todos os colaboradores da Secretaria de Saúde do Estado de Roraima - SESAU/RR pela disponibilização do acesso a base de dados do SIVEP - Malária e a Fundação Estadual do Meio Ambiente, Ciência e Tecnologia do Estado de Roraima, ao apoio financeiro do projeto PPSUS/SVS - "Identificação e mapeamento de áreas de riscos de incidência da malária utilizando um SIG".

CONFLITO DE INTERESSE

Os autores declaram não haver nenhum tipo de conflito de interesse no desenvolvimento do estudo.

SUPORTE FINANCEIRO

FEMACT/RR - Fundação Estadual do Meio Ambiente, Ciência e Tecnologia

REFERÊNCIAS

1. Report on a technical consultation on Research in Support of malaria Control in the Amazon Brazil. TDR/FIELDMAL/SC/AMAZ 88.3. Geneva: World Health Organization (WHO); 2008.
2. Ministério da Saúde. Malária. Guia de Vigilância Epidemiológica. 6ª. ed. Brasília, Secretaria de Vigilância em Saúde; 2008. p. 521-531.
3. Alves MT, Silva AAM, Nemes MIB, Brito GO. Tendência da incidência e da mortalidade por Aids no Maranhão, 1985 a 1998. *Rev Saude Publica* 2003; 37:177-182.
4. Choi K, Thacher SB. An evaluation of influenza mortality surveillance, 1962-1979. *Am J Epidemiol* 1981; 113:215-222.
5. Barros AJD. Modelagem estatística em estudos epidemiológicos. O modelo logístico. [Dissertação de Mestrado]. [Campinas, BR]: Universidade Estadual de Campinas; 1990.
6. Haykin S. Redes Neurais: princípios e prática. Tradução Paulo Martins Engel. 2ª Ed. Porto Alegre: Bookman; 2001.
7. Bustamante-Sá C, Nobre FF. Forecasting Epidemiological Time Series With Backpropagation Neural Networks. *In: Proceedings of the 38th Midwest Symposium on Circuits and Systems*, Rio de Janeiro; 1996. p.1365-1368.
8. Penna MLF. Rede Neural Artificial para Detecção de Sobremortalidade Atribuível à Cólera no Ceará. *Rev Saude Publica* 2004; 38:351-357.
9. Santos AM, Seixas JM, Pereira BB, Medronho RA. Usando Redes Neurais Artificiais e Regressão Logística na Predição da hepatite A. *Rev Bras Epidemiol* 2005; 8:117-126.
10. Cortez PAR. Algoritmos Genéticos e Redes Neuronais na Previsão de Séries Temporais. [Dissertação de Mestrado]. [Braga, PT]: Universidade do Minho; 1997. 105 p.
11. Duh M, Walker AM, Pagano M, Kronlund K. Prediction and cross-validation of neural networks versus logistic regression: using hepatic disorders as na examples. *Am J Epidemiol* 1998; 147:407-412.
12. Kattan MW, Hess KR. Experiments to determine whether recursive partitioning (CART) or na artificial neural network overcomes theoretical limitations of Cox proportional hazards regression. *Comput Biomed Res* 1998; 31:363-373.
13. Nobre FF, Stroup DF A monitoring system to detect changes in public health surveillance data. *Int J Epidemiol* 1994; 23:408-418.
14. Portugal MS. Neural networks versus time series methods: a forecasting exercise. *In: Presented at the 14th International Symposium on Forecasting*, Stockholm, Sweden; 1994.