

# Classificação do risco de prevalência da esquistossomose no Estado de Minas Gerais

Flávia de Toledo Martins<sup>1</sup>, Luciano Vieira Dutra<sup>1</sup>, Corina da Costa Freitas<sup>1</sup>

<sup>1</sup>Instituto Nacional de Pesquisas Espaciais – INPE  
Caixa Postal 515 – 12201-970 – São José dos Campos – SP – Brasil

{flavinha, dutra, corina}@dpi.inpe.br

**Abstract.** *The aim of this work is to determine the behavior pattern within data derived from remote sensing, social- economic variables and climatic variables, through the decision tree analysis by choosing the more appropriate ones to explain the prevalence of schistosomiasis in State of Minas Gerais, Brazil.*

**Resumo.** *O objetivo deste trabalho é determinar o padrão de comportamento dos dados derivados de sensoriamento remoto, das variáveis socioeconômicas e das variáveis climáticas, através de árvore de decisão, escolhendo as variáveis mais apropriadas explicar a prevalência da esquistossomose no estado de Minas Gerais, Brasil.*

## 1. Introdução

A esquistossomose é uma doença causada por um trematódeo do gênero *Schistosoma* presente em cerca de 74 países. No Brasil a esquistossomose encontra-se distribuída do nordeste até a região sul, principalmente nos estados da Bahia, Espírito Santo e Minas Gerais (WHO, 1993).

Existem também outros focos da doença disseminados pelo país. Por isso é necessário investir no mapeamento e modelagem espacial do risco de prevalência da doença como ferramenta de formação de políticas públicas de controle da endemia.

Este trabalho descreve a utilização de um Sistema de Informações Geográficas (SIG) para mapear 3 categorias de risco estimadas por árvores de decisão que usam como atributos de entrada dados de Sensoriamento Remoto (SR) obtidos de diversas fontes. Árvore de decisão é uma das técnicas de reconhecimento de padrões disponíveis e tem como atrativo um resultado de fácil interpretação (Guimarães *et al.*, 2006).

## 2. Materiais e Métodos

A área de estudo é o Estado de Minas Gerais, com uma área de aproximadamente 590.000 km<sup>2</sup>, dividido politicamente em 853 municípios, com aproximadamente 18 milhões de habitantes (IBGE, 2006).

### 2.1. Variáveis utilizadas

Para identificar as classes de risco da doença, foram usados dados de prevalência da esquistossomose em 197 municípios do Estado de Minas Gerais, divididas em 3 classes: alta, média e baixa, juntamente com variáveis de SR, derivadas dos sensores MODIS

(Moderate Resolution Imaging Spectroradiometer), dados altimétricos gerados pela missão SRTM (Shuttle Radar Topography Mission), variáveis climáticas e socioeconômicas.

## 2.2. Metodologia

Para aplicar a técnica de reconhecimento de padrões foi utilizado o software de domínio público, denominado Waikato Environment for Knowledge Analysis (Weka), da Universidade de Waikato, Nova Zelândia (Weka, 2006).

A árvore de decisão foi gerada usando as 197 amostras foram divididas em dois conjuntos, conjunto de treinamento, com 142 amostras e conjunto de teste, com 55 amostras. Após a geração da árvore de decisão, a classificação foi extrapolada, via o SIG TerraView, distribuído pelo INPE, para os outros 656 municípios do Estado.

## 3. Resultados e discussões

A árvore de decisão gerada alcançou, em seu melhor caso, 71,8% de amostras classificadas corretamente, o que corresponde a um índice de concordância Kappa na ordem de 0,565.

As variáveis selecionadas para esse caso foram: EVI\_I (índice de vegetação melhorada no inverno), NDVI\_I (índice de vegetação da diferença normalizada no inverno), AH1 (média da acumulação hídrica), Tmin\_V (temperatura mínima no verão), IDH\_00 (índice de desenvolvimento humano de 2000) e SAN1 (porcentagem de domicílios com esgoto ligado a rio ou lago). A Figura 1 mostra a representação gráfica desta árvore.

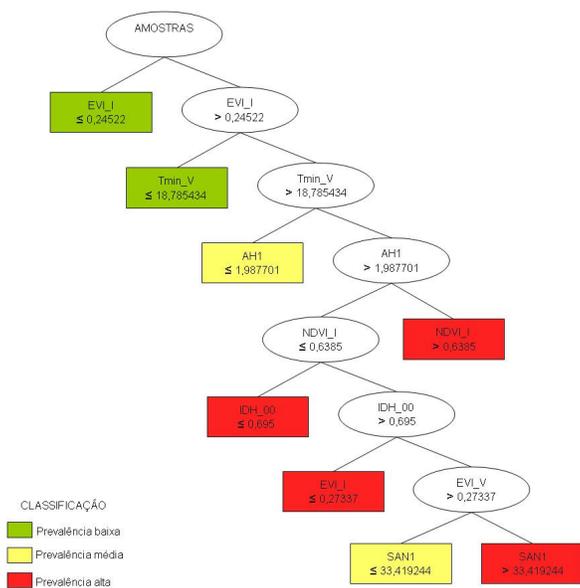


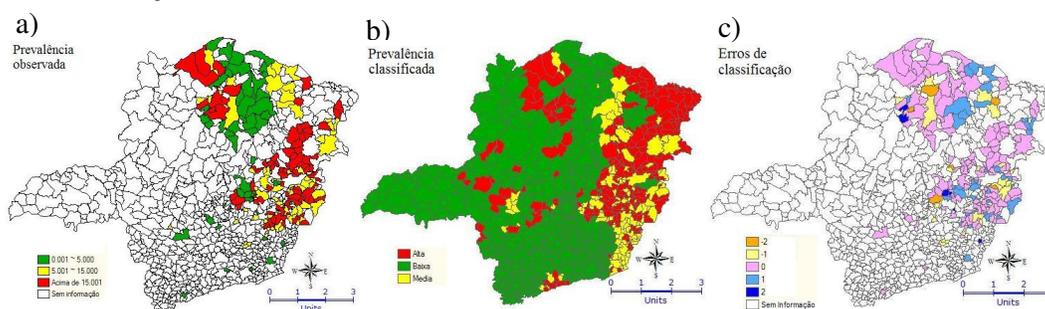
Figura 1. Representação gráfica da árvore de decisão

O resultado da classificação pode ser considerado satisfatório uma vez que as variáveis selecionadas refletem fatores considerados importantes para a existência da doença. A Tabela 1 apresenta a matriz de confusão da classificação.

**Tabela 1. Matriz de confusão**

Classe	Classificação		
	Baixa	Média	Alta
Baixa (34)	28	5	1
Média (48)	13	18	17
Alta (60)	2	2	56

Observou-se maior confusão na classe de prevalência média, já que 35,4% das amostras dessa classe foram classificadas como alta e 27,1% como baixa. Para as classes de prevalência alta e baixa, a classificação atingiu, respectivamente, 93,3% e 82,4% de amostras classificadas corretamente. A Figura 2 mostra a o mapa temático com da prevalência estimada da doença, a classificação da árvore de decisão e os erros da classificação.



**Figura 2 (a)Prevalência observada da doença; (b)Prevalência da esquistossomose classificada através da árvore de decisão; (c) Erros de classificação**

Na Figura 2 (b) os municípios representados em laranja, vermelho e amarelo a classificação é subestimada em relação a 2 classe e em 1 classe observada, respectivamente (por exemplo, os dois municípios em laranja, foram classificados como prevalência baixa, porém pertenciam à classe de prevalência alta). Já nos municípios representados em azul escuro e claro a classificação é superestimada em 2 e em 1 classe, respectivamente. Os municípios classificados corretamente estão representados em rosa.

## Referência

Guimarães, R. J. d. P. S.; Freitas, C. C.; Dutra, L. V.; Moura, A. C. d. M.; Amaral, R. S.; Drummond, S. C.; Scholte, R. G. C.; Freitas, C. R.; Carvalho, O. d. S., 2006, **Analysis and estimative of schistosomiasis prevalence for Minas Gerais state, Brazil, using multiple regression with social and environmental spatial data**, Mem Inst Oswaldo Cruz, p. 91-96.

IBGE. Instituto Brasileiro de Geografia e Estatística. Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=mg>>. Acesso em: abril, 2006.

Weka. **Data Mining Software in Java The University of Waikato**. 2006. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: jun, 2006.

WHO. **The Control of Schistosomiasis**. Geneva: World Health Organization, 1993. 94 p.