

## Spatial Distribution of Seropositive Dogs to *Leptospira* spp. and Evaluation of Leptospirosis Risk Factors Using a Decision Tree\*

Daniele Bier<sup>1</sup>, Flávia Toledo Martins-Bedê<sup>2</sup>, Vivien Midori Morikawa<sup>1</sup>, Leila Sabrina Ullmann<sup>3</sup>, Mariana Kikuti<sup>4</sup>, Helio Langoni<sup>4</sup>, Ricardo José Canever<sup>1</sup>, Alexander Welker Biondo<sup>1</sup> & Marcelo Beltrão Molento<sup>1</sup>

### ABSTRACT

**Background:** Leptospirosis is an important zoonotic disease associated with poor areas of urban settings of developing countries and early diagnosis and prompt treatment may prevent disease. Although rodents are reportedly considered the main reservoirs of leptospirosis, dogs may develop the disease, may become asymptomatic carriers and may be used as sentinels for disease epidemiology. The use of Geographical Information Systems (GIS) combined with spatial analysis techniques allows the mapping of the disease and the identification and assessment of health risk factors. Besides the use of GIS and spatial analysis, the technique of data mining, decision tree, can provide a great potential to find a pattern in the behavior of the variables that determine the occurrence of leptospirosis. The objective of the present study was to apply Geographical Information Systems and data prospection (decision tree) to evaluate the risk factors for canine leptospirosis in an area of Curitiba, PR.

**Materials, Methods & Results:** The present study was performed on the Vila Pantanal, a urban poor community in the city of Curitiba. A total of 287 dog blood samples were randomly obtained house-by-house in a two-day sampling on January 2010. In addition, a questionnaire was applied to owners at the time of sampling. Geographical coordinates related to each household of tested dog were obtained using a Global Positioning System (GPS) for mapping the spatial distribution of reagent and non-reagent dogs to leptospirosis. For the decision tree, risk factors included results of microagglutination test (MAT) from the serum of dogs, previous disease on the household, contact with rats or other dogs, dog breed, outdoors access, feeding, trash around house or backyard, open sewer proximity and flooding. A total of 189 samples (about 2/3 of overall samples) were randomly selected for the training file and consequent decision rules. The remained 98 samples were used for the testing file. The seroprevalence showed a pattern of spatial distribution that involved all the Pantanal area, without agglomeration of reagent animals. In relation to data mining, from 189 samples used in decision tree, a total of 165 (87.3%) animal samples were correctly classified, generating a Kappa index of 0.413. A total of 154 out of 159 (96.8%) samples were considered non-reagent and were correctly classified and only 5/159 (3.2%) were wrongly identified. On the other hand, only 11 (36.7%) reagent samples were correctly classified, with 19 (63.3%) samples failing diagnosis.

**Discussion:** The spatial distribution that involved all the Pantanal area showed that all the animals in the area are at risk of contamination by *Leptospira* spp. Although most samples had been classified correctly by the decision tree, a degree of difficulty of separability related to seropositive animals was observed, with only 36.7% of the samples classified correctly. This can occur due to the fact of seronegative animals number is superior to the number of seropositive ones, taking the differences in the pattern of variable behavior. The data mining helped to evaluate the most important risk factors for leptospirosis in an urban poor community of Curitiba. The variables selected by decision tree reflected the important factors about the existence of the disease (default of sewer, presence of rats and rubbish and dogs with free access to street). The analyses showed the multifactorial character of the epidemiology of canine leptospirosis.

**Keywords:** decision tree, *Leptospira* spp., leptospirosis, dogs, risk factors, spatial analysis.

## INTRODUCTION

Leptospirosis is an emerging zoonosis, which has been spread out through urban areas particularly in low income regions of developing countries [17]. Although rodents have been reportedly shown as the main reservoirs for the disease, dogs may develop the disease and become asymptomatic carriers [4]. Since dogs may be over exposed in endemic areas when compared to humans, they may play a major role as sentinels in serological surveys to a particular infecting serovar [5].

Since leptospirosis is a geographical and seasonal disease determined by environmental and social factors, geographical information systems (GIS) may be useful to identify these factors and to determine potential risk area. When applied in this context, GIS may serve as basis for the planning of intervention measures on exposed populations to risk factors. GIS has been reportedly employed on human and animal leptospirosis in Brazil and other countries [3,8,10,23]

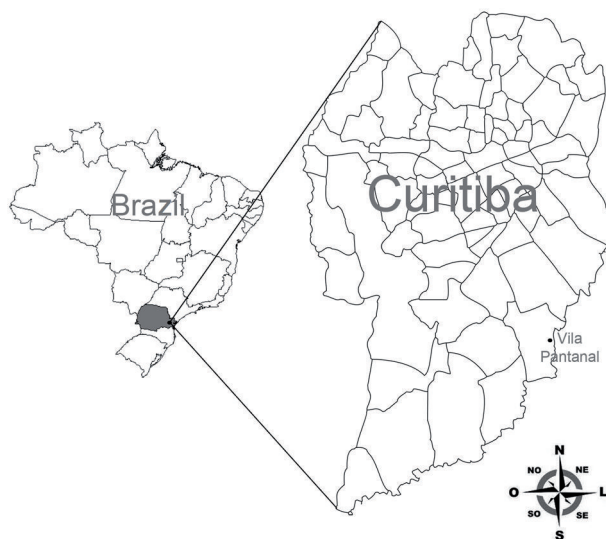
The use of GIS combined with spatial analysis techniques allows, in addition to mapping the disease, the identification and assessment of health risk factors [2]. Besides the use of GIS and spatial analysis, the technique of data mining, decision tree, is often used to determine the behavior pattern of data sets in different forms of representations [1,14]. This technique can provide high potential to find a pattern in the behavior of the variables that determine the occurrence of leptospirosis. Accordingly, the aim of the present study was to apply the technique of data mining (decision tree) to evaluate risk factors to canine leptospirosis and to delimit areas of risk of transmission of leptospirosis, using variables related to the owner, environment and animals, using GIS tools.

## MATERIALS AND METHODS

### Area

The present study was performed on the Vila Pantanal, a low-income and riverside neighborhood of Curitiba (Figure 1), which is the capital of Paraná state and the eighth most populated city of Brazil, with 1.8 million inhabitants [11]. The area is considered an environmental preserved land bordered by the Iguaçu River, characterized by seasonal flooding during rainy season. Original human settlement occurred on late 80s' and today the population is of 765 families (2,322 people)

[16]. Although at the time of study residents had access to electric power and partially to treated water, sanitary sewage was mostly released to the environment. Organic garbage was dumped in open containers and weekly collected; however, since recycling material was the main source of income, large quantities of residual material (plastic, cardboard boxes, aluminum cans) could be found in the majority of households.



**Figure 1.** Localization of Vila Pantanal in the City of Curitiba, Paraná State, Brazil.

### Dogs

According to the local Health Service, a total of 845 dogs and 128 cats were registered at the time. Only dogs with a minimum of 4 months of age and signed permission by owners were included in the study. A total of 287 dog blood samples were randomly obtained house by house in a two-day sampling on January 2010. In addition, a questionnaire was applied to owners at the time of sampling.

Dog blood samples (10 mL) were collected by venopuncture of jugular vein in tubes without anti-coagulant and kept at room temperature (25°C) until visible clot retraction, centrifuged at 1200 g for 5 min and serum separated and kept at -20°C until testing. Leptospirosis serology was performed using the gold-standard serological test microscopic agglutination test (MAT) with live antigens for antibody detection [25], and included serovars for Australis, Bratislava, Autumnalis, Canicola, Cynopteri, Djasiman, Grippityphosa, Copenhageni, Icterohaemorrhagiae, Pomona, Pyrogenes and Hardjo. Regardless, samples with titer equal or superior to 100 to a specific serovar were all considered reagent.

Geographical coordinates related to each household of tested dog were obtained using a Global Positioning System (GPS)<sup>1</sup>. Spatial distribution map of reagent and non-reagent dogs to leptospirosis was generated using a commercially available software<sup>2</sup>.

#### Variables and Decision Tree

An overall of 57 questions were obtained, of which 11 Boolean-type questions (yes-no) were selected for the present study (Table 1). Explanative variables were divided in three groups: first group related to the household and second to the dogs, obtained by a questionnaire answered by owners as mentioned above and a third group related to the environment, filled by the evaluator.

**Table 1.** Description of 11 explanatory variables used to the generation of the decision tree. Vila Pantanal, City of Curitiba, Paraná State, Brazil.

1. Variables related to the household and dog owners	
KnowLepto	Anyone at the household have knowledge about leptospirosis?
LeptoHouse	Anyone at the household was diagnosed with leptospirosis?
2. Variables related to the dogs	
ContDogs	Dog have contact with other dogs?
Breed	Dog have a specific breed?
OutAccess	Dog have outdoor access?
FeedBowl	Are water and food bowls stored overnight?
3. Variables related to the environment	
Rats	Rats nearby the household?
Rubble	Rubble or garbage nearby the household?
Sewer	Open sewer nearby the household?
Flooding	Flooding nearby the household?
CitySewer	Household with city or closed sewer?

A publicly available software was used for data prospection (*Waikato Environment for Knowledge Analysis - Weka*)<sup>3</sup>, using the algorithm for decision tree C4.5 developed by Quinlan [20] and improved in its version to Java language (in *Weka*) named J4.8 [24]. Starting with the 11 explicative questions (Table 2) in association to MAT results (predictive variable), database was generated, data analyzed, results obtained and presented. A total of 189 samples (about 2/3 of overall samples) were randomly selected for the training file and consequent decision rules. The remained 98 samples were used for the testing file.

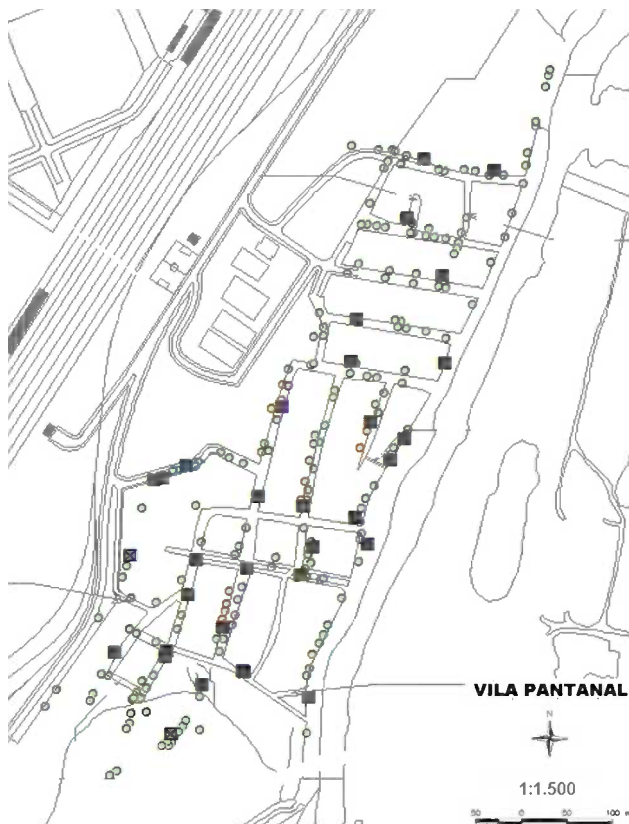
**Table 2.** Confusion matrix of the decision tree generated through the algorithm J4.8 from *Weka*.

Class \ Classification	Classification	
	Positive (POS)	Non Reagent (NR)
POSITIVE	<b>11</b> (36.7%)	19 (63.3%)
NON REAGENT	5 (3.2%)	<b>154</b> (96.8%)

#### RESULTS

Note in Figure 2 the spatial distribution of sampled dogs. In Figure 2, geographic localization of negative dogs at MAT is represented by gray circles and positive ones by gray squares. A spatial distribution that involved all area, without agglomeration of reagent animals can also be observed in this figure. Out of 287 samples tested by MAT, 47 (16.4%) were reagent to at least one serovar, considering titers above 100 and the occurrence of co-infection, more than one serovar involved. All dogs were georeferenced in their households.

Both serology results and 11 explanatory variables were used to the generation of the decision tree. According to the tree, the main risk factors for canine leptospirosis are related to the variables: OutAccess, KnowLepto, ContDogs, Sewer, Rats, LeptoHouse, CitySewer and Rubble (explained on Table 1). Figure 3 shows the graphic representation of the selected decision tree.



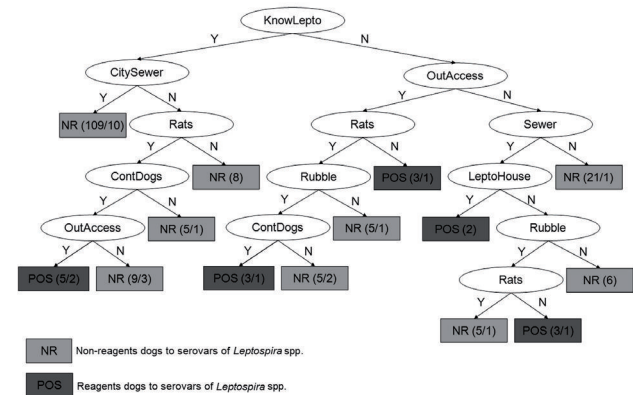
**Figure 2.** Spatial distribution of sampled dogs in Vila Pantanal, Curitiba, Paraná State, Brazil. Negative dogs at MAT is represented by gray circles and positive ones by gray squares.

In Figure 3, for each rule of classification, inside the parenthesis there is the number of samples classified in the leaf/and the number of samples wrongly classified in this leaf. The variable placed in the source of the tree was KnowLepto, corresponding to a division into two groups: when someone of the dog's house knew the disease (136 samples, about 72%) and when did not know (53 samples, about 28%).

Follows the division of the tree in variable function always in function of two possibilities, classifying samples as non reagent (NR) or reagent dogs (POS) or samples are classified as a new rule of classification in function of another variable.

This tree (Figure 3) reached 87.3% (165/189) of correct classified samples, generating a Kappa index of 0.413. Table 2 shows the confusion matrix of the tree that was generated through the algorithm J4.8 from *Weka*. Analyzing the matrix it was possible to identify problems in the classification and also the separability among classes. Values in bold are the numbers of correctly classified samples.

In confusion matrix, from 189 samples, 165 were classified correctly. The 159 samples from non-reagent animals, 154 (96.8%) were correctly classified and only 5 (3.2%) were classified as positive. From the 30 seropositive animals, 11 (36.7%) of the samples were correctly classified and 19 (63.3%) were classified as non-reagent.



**Figure 3.** Graphic representation of the decision tree. Vila Pantanal in the City of Curitiba, Paraná State, Brazil.\*

\* - Note: Y - Yes; N - No; KnowLepto - Anyone at the household have knowledge about leptospirosis?; CitySewer - Household with city or closed sewer?; OutAccess - Dog have outdoor access?; Rats - Rats nearby the household?; Sewer - Open sewer nearby the household?; ContDogs - Dog have contact with other dogs?; Rubble - Rubble or garbage nearby the household?; LeptoHouse - Anyone at the household was diagnosed with leptospirosis?

## DISCUSSION

Seroprevalence showed a pattern of spatial distribution that involved all the Pantanal area, without agglomeration of reagent animals, as observed in Figure 2. It can be deducted that in the addressed scale, there were spatial correlation between epidemiological characteristics, showing that all the animals in the area are at risk of contamination by *Leptospira* spp., differently from others authors [13] that verified seroreagent dog concentration in localized areas, but also with overlap in lacking sanitation.

To determine the spatial analysis of leptospirosis in a municipal level three, one must check for the social conditions of the population, the risk of flooding and the population concentration [18]. Analyzing these three aspects together makes clear the reason why Vila Pantanal had this random distribution to reagent animals to leptospirosis. The high demographic density and the critical poverty situation allied to the fact that many regions of Vila Pantanal are under flooding risk makes the area very favorable for disease dissemination.



The decision tree classified dogs as non-reagent if the owner knew the disease and had sewer or cesspool in the household. This information corroborates with what we were expecting, when there was sewer or fossa in the property, the fewer the local probability of rats and environment contamination. This would in turn reduce the chances of dogs that live in this condition do get infected, having less contact with contaminated water from cesspool and gutters [9].

On the other hand, when owners did not know about the disease, the division of the tree was in function if the dog had or not access to the street. The decision tree classified a seropositive animal to leptospirosis when the dog had access to street, there were rats inside or near the household, there was rubbish at the house and if the dog had contact with other dogs. Breeding management can be a relevant factor, once the index of reagent dogs that had access to the street is higher from those that did not have [22]. The same way, in other study, dogs that did not remained confined had 2.61 times more chances to acquire the disease, probably because these animals would get in contact with other dogs and other risk factors [9]. This shows that dog domestication is fundamental to decrease the risk of contamination.

The increase of the disease risk in the local might be attributed to the fact that about 15.9% of families work with collection of recyclable material, activity that involves beyond the informality, unhealthy and dangerous work conditions. This fact is aggravated once 79% of these families use domestic space to collected material separation or storage [16], which is favorable to the presence of rodents.

The dogs were classified as seropositive by the tree when there was not the presence of rats near or inside the household. Higher number of reagent dogs when there was no contact with rodents was also observed in Botucatu, Brazil [12], differently from other studies [9,15].

The risk factor can be considered as non accordingly evaluated, due to the fact of being completely against the expected, once the presence of rats are closely related to disease appearance of disease. They report yet that many times, owners hide important facts that consider undesirable [9,15].

The analysis of the risk factors through the decision tree showed that variables sewer, rats, rubbish and access to street, possibly explain the seropositivity of dogs in Vila Pantanal. The main risk factors in acquire the disease observed in previous studies include walk in areas with stagnant water in case of people [7] and the habit of hunting rodents, presence of flooding areas near households and the access to street to dogs [19].

Such factors are aggravated due to the geographic localization of Vila Pantanal, located on the banks of the channel of Iguaçu River, with many houses at flood and flooding risks and by the allowed dog access to the street. The same way, researchers using GIS concluded that stagnant water, rubbish and sewer (all associated with the presence of rodents) contributed to the occurrence of leptospirosis in human in different points of Rio de Janeiro, associated to these variables [3].

Generally, the decision tree classified dogs as reagent if the owner knew the disease, but did not have sewage in households, there were rats near or inside the houses and dogs had access to street. And, when the owner did not know the disease and there were rats and rubbish in the household and dogs had contact with other dogs. So the result of the classification can be considered coherent as the evaluated risk factors in relation to reality: living condition of people, animals and poor sanitation conditions are important factors to the existence of the disease. The fact that population is more exposed to risk factors as their animals was also verified in many studies due to the deficiency of basic infrastructure [3,6,8,21,23].

Despite 87.3% of samples had been classified correctly by the decision tree, it can be shown the difficulty of separability related to seropositive animals, and only 36.7% of samples were classified correctly. This can occur due to the fact of seronegative animals number is superior to the number of seropositive ones, taking the differences in the pattern of variable behavior and because of the relatively low sample size in use for WEKA. However, 96.8% of samples from non-reagent animals were classified correctly, indicating thus, the meaningful use of this tree to diagnosis discard of canine leptospirosis.

## CONCLUSIONS

Geoprocessing techniques and data mining helped to better understand the distribution of serological results in an urban community from Curitiba and to identify and to evaluate the most important risk factors. The variables selected by decision tree reflect the important factors about the existence of the disease (default of sewer, presence of rats and rubbish and dogs with free access to street). The spatial distribution of seroreagent dogs and the evaluation of the main risk factors may also allow local preventive measures to be taken. Besides, the accuracy in the classification of non-reagent animals found in the decision tree indicates the meaningful use of the tree to discard the diagnosis of canine leptospirosis.

## SOURCES AND MANUFACTURERS

<sup>1</sup>Geoexplorer, Trimble, Sunnyvale, CA, USA.

<sup>2</sup>ArcView GIS 3.2, ESRI, Redlands, CA, USA.

<sup>3</sup>Waikato Environment for Knowledge Analysis -Weka, University of Waikato, New Zealand.

**Ethical approval.** The study was approved by the Ethics Committee in Animal Use and Welfare at the Federal University of Paraná, Curitiba, Brazil (Protocol 007/2009).

**Declaration of interest.** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## REFERENCES

- 1 **Araki H. 2005.** Fusão de informações espectrais, altimétricas e de dados auxiliares na classificação de imagens de alta resolução espacial. 136p. Curitiba, PR. Tese (Doutorado em Ciências Geodésicas). Programa de Pós-graduação em Ciências Geodésicas, Universidade Federal do Paraná.
- 2 **Barcellos C. & Bastos F.I. 1996.** Geoprocessamento, ambiente e saúde: uma união possível? *Cadernos de Saúde Pública*. 12(3): 389-397.
- 3 **Barcellos C. & Sabroza P. 2000.** Socio-environmental determinants of the leptospirosis outbreak of 1996 in western Rio de Janeiro: a geographical approach. *International Journal Health Research*. 10(4): 301-313.
- 4 **Batista C.S.A., Azevedo S.S., Alves C.B., Vasconcellos A.S., Moraes Z.M., Clementino I.J., Lima F.S. & Araújo Neto J.O. 2004.** Soroprevalência de leptospirose em cães errantes da cidade de Patos, Estado da Paraíba, Brasil. *Brazilian Journal of Veterinary Research and Animal Science*. 41(2): 131-136.
- 5 **Blazius R.D., Romão P.R.T., Blazius E.M.C.G. & Silva O.S. 2005.** Ocorrência de cães errantes soropositivos para *Leptospira* spp. na cidade de Itapema, Santa Catarina, Brasil. *Cadernos de Saúde Pública*. 21(6): 1952-1956.
- 6 **Dias J.P., Teixeira M.G., Costa M.C.N., Mendes C.M., Guimarães P., Reis M.G., Ko A.I. & Barreto M.L. 2007.** Factors associated with *Leptospira* sp. infection in large urban center in northeastern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*. 40(5): 499-504.
- 7 **Douglin C.P., Jordan C., Rock R., Hurley A. & Levett P.N. 1997.** Risk factors for severe leptospirosis in the parish of St. Andrew, Barbados. *Emerging Infectious Diseases*. 3(1): 78-80.
- 8 **Figueiredo C.M., Mourão A.C., Oliveira M.A.A., Alves W.R., Ooteman M.C., Chamone C.B. & Koury M.C. 2001.** Leptospirose humana no município de Belo Horizonte, Minas Gerais, Brasil: uma abordagem geográfica. *Revista da Sociedade Brasileira de Medicina Tropical*. 34(4): 331-338.
- 9 **Furtado L.R.I., Ávila M.O., Fehlberg M.F.B., Teixeira M.M., Rosado R.L.I., Martins L.F.S. & Brod C.S. 1997.** Prevalência e avaliação de fatores de risco à leptospirose canina no Município de Pelotas-RS. *Arquivos do Instituto Biológico*. 64(1): 57-61.
- 10 **George S.G., Joshua H.V., Bruno B.C., Philip H.K., Daphne A.D. & Michael L.J. 2007.** Use of a case-control study and geographic information systems to determine environmental and demographic risk factors for canine leptospirosis. *Veterinary Research*. 38: 37-50.
- 11 **Instituto Brasileiro de Geografia e Estatística (IBGE). 2010.** Censo Demográfico: Cidades - Curitiba. IBGE. Available from: < <http://www.ibge.gov.br/cidadesat/topwindow.htm?1>>. Cited: 2010 April 20.
- 12 **Lopes A.L.S., Silva W.B., Padovani C.R., Langoni H. & Modolo J.R. 2005.** Frequência sorológica antileptospírica em cães: sua correlação com roedores e fatores ambientais, em área territorial urbana. *Arquivos do Instituto Biológico*. 72(3): 289-296.

- 13 Magalhães D.F., Silva J.A., Moreira E.C., Wilke V.M.L., Haddad J.P.A. & Meneses J.N.C. 2006. Prevalência de aglutininas anti-*Leptospira interrogans* em cães de Belo Horizonte, Minas Gerais, 2001 a 2002. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*. 58(2): 167-174.
- 14 Marcelino I.P.V.O. 2003. Análise de episódios de tornados em Santa Catarina: caracterização sinótica e mineração de dados. 222 p. São José dos Campos, SP. Dissertação (Mestrado em Sensoriamento Remoto) - Programa de Pós-graduação em Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais.
- 15 Masculli R., Pinheiro S.R., Vasconcellos A.S., Ferreira F., Moraes Z.M., Pinto C.O., Sucupira M.C.A., Dias R.A., Miraglia F., Cortez A., Costa S.S., Tabata R. & Marcondes A.G. 2002. Inquérito sorológico para leptospirose em cães do Município de Santana de Parnaíba, São Paulo, utilizando a campanha de vacinação anti-rábica do ano de 1999. *Arquivos do Instituto Biológico*. 69(2): 25-32.
- 16 Morato M., Lange M.F.L. & Farion R.J. 2009. Diagnóstico da área de abrangência da Unidade Municipal de Saúde Pantanal. 27f. Curitiba, PR. Trabalho de conclusão de curso (Especialização em Saúde Coletiva) - Curso de Especialização em Saúde Coletiva, Universidade Positivo.
- 17 Pappas G., Papadimitriou P., Siozopoulou V., Christou L. & Akritidis N. 2008. The globalization of leptospirosis: worldwide incidence trends. *International Journal Infectious Diseases*. 12(4): 351-357.
- 18 Paula E.V. 2005. Leptospirose Humana: uma análise climato-geográfica de manifestação no Brasil, Paraná e Curitiba. In: *Anais do XII Simpósio Brasileiro de Sensoriamento Remoto* (Goiânia, Brasil). pp.2301-2308.
- 19 Querino A.M.V., Delbem A.C.B., Oliveira R.C., Silva F.G., Müller E.E., Freire R.L. & de Freitas J.C. 2003. Fatores de risco associados à leptospirose em cães do município de Londrina-PR. *Semina: Ciências Agrárias*. 24(1): 27-34.
- 20 Quinlan J.R. 1993. *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann Publishers, 303p.
- 21 Romero E.C., Bernardo C.C.M. & Yasuda P.H. 2003. Human leptospirosis: a twenty-nine-year serological study in São Paulo, Brazil. *Revista do Instituto de Medicina Tropical de São Paulo*. 45(5): 245-248.
- 22 Silva W.B., Simões L.B., Lopes A.L.S., Padovani C.R., Langoni H. & Modolo J.R. 2006. Avaliação de fatores de risco de cães sororreagentes à *Leptospira* spp. e sua distribuição espacial, em área territorial urbana. *Brazilian Journal of Veterinary Research and Animal Science*. 43(6): 783-792.
- 23 Soares T.S., Latorre M.R.D.O., Laporta G.Z. & Buzzar M.R. 2010. Spatial and seasonal analysis on leptospirosis in the municipality of São Paulo, Southeastern Brazil, 1998 to 2006. *Revista de Saúde Pública*. 44(2): 283-291.
- 24 Waikato Environment for Knowledge Analysis (WEKA). 2006. *WEKA: Data Mining Software in Java*. University of Waikato. Available from: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Cited: 2011 Mar 21.
- 25 World Health Organization (WHO). 2003. Human leptospirosis: guidance for diagnosis, surveillance and control. Geneva: WHO-Technical Report Series, 109p.