Spatial and Risk Analysis of Cutaneous Leishmaniasis in an Integration Region of Pará, Eastern Brazilian Amazon

Bruna C. de Souza[1,2], **André C.P.L.F. de Carvalho**[1], **Robson P. Bonidia**[3], **Nelson V. Gonçalves**[1,4], **Claudia S.C. Miranda**[1,4]

[1]Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP) São Carlos – SP – Brazil

[2]Instituto de Ciências da Saúde – Universidade Federal do Pará (UFPA) Belém – PA – Brazil

[3]Departamento de Ciência da Computação– Universidade Tecnológica Federal do Paraná (UTFPR) Cornélio Procópio – PR – Brazil

[4]Laboratório de Epidemiologia e Geoprocessamento da Amazônia – Universidade do Estado do Pará (UEPA) Belém – PA – Brazil

souza.bruna@usp.br, andre@icmc.usp.br, bonidia@utfpr.edu.br,

nelsonveiga@uepa.br, claudia.miranda@uepa.br

Abstract. *This paper analyzed the occurrence of Cutaneous Leishmaniasis (CL) and its relationship with risk factors in Pará, Brazil. The ecological and experimental study was carried out using official data from the Brazilian government. The results showed that the disease still has a significant impact on Amazonian populations, reflected by the epidemiological scenario, whose most affected individuals are systematically in a situation of socioeconomic and environmental vulnerability. The use of AI tools enables information synthesis to facilitate analysis of the situations. Ultimately, the occurrence of CL establishes the need to create and implement more efficient methods for eliminating the disease to achieve the third Sustainable Development Goal.*

1. Introduction

Cutaneous Leishmaniasis (CL) is a neglected anthropozoonosis that is transmitted to humans through the bite of female mosquitoes of the Lutzomyia and Phlebotomus genera infected with Leishmania [Brasil 2017]. The main clinical evidence of this disease is the existence of localized or disseminated lesions on the skin or mucous membranes, which can cause permanent deformities, as well as a huge psychological impact on those affected by the illness [WHO 2010, Bailey et al. 2019].

Currently, CL persists as an endemic disease in various regions of the world with an incidence of approximately 1.2 million cases per year in 85 countries [Brasil 2017]. The Pan American Health Organization (PAHO) considers the disease to be one of the six most important neglected diseases today [WHO 2010]. In the last six years, the Americas have recorded 252,998 cases of CL, with an annual average of approximately 42,000 cases. Around 97% of these cases come from the following territories: the Andean region (41%), Brazil (37%), and Central America (19%), with a linear downward trend in the occurrence of this anthropozoonosis in all the countries on these continents except Guatemala, Mexico, and Panama [PAHO 2023].

In recent decades, the Brazilian northern region has concentrated the highest number of cases, with 40% of these notifications coming from Pará. During this period, the Xingu Integration Region (XIR) recorded the highest prevalence in the state, a situation that is possibly related to the characteristics of this territory. In this context, this region has undergone intense environmental transformations, such as the construction of the Belo Monte hydroelectric power station, which has had repercussions on the socio-economic and public health aspects of the resident and migratory populations, establishing a favorable environment for the disease risk factors [Gauthier and Moran 2018].

As a result, the simultaneous occurrence and interrelationship of socio-economic, environmental and public policy factors create a huge variety of epidemiological scenarios, which are difficult to observe due to the different spatial scales they appear on [Souza et al. 2023]. In this context, constructing and analyzing these scenarios becomes a challenge for public health, especially when considering that despite the expansion of AI use in healthcare, recent technologies are concentrated in the global north and are rarely developed with a focus on neglected diseases.

Therefore, this study enabled a more detailed analysis of the occurrence of the disease, not only by observing its prevalence in time and space, but also by considering the combination of various risk factors present in this territory, such as social, economic, and environmental issues, which are varied and imprecise. In light of the above, the aim of this study was to analyze the occurrence of CL and its relationship with socioeconomic, environmental, and public health policy risk factors in the Xingu Integration Region, in the state of Pará, Brazil, from 2014 to 2024.

2. Materials and Methods

2.1. Ethical aspects

As far as health research is concerned, this study complied with the precepts of Resolution 466 of 2012 of the Brazilian National Health Council, which regulates research with human beings. Furthermore, as this paper used secondary databases with unrestricted public access, it was not necessary to undergo approval by the Research Ethics Committee. In terms of the use of artificial intelligence (AI), this project has adopted the guidelines suggested in the literature, such as the development and responsible use of AI, AI for all, Guidelines for Trustworthy AI, among others. The principles of Data-Centric AI were also applied, making data the central object in the process of creating this system.

2.2. Study Design

An ecological and experimental study was carried out covering all the notified cases of CL that occurred in the municipalities that make up the Xingu Integration Region. This territory, located in the southwest mesoregion of the state of Pará, contains ten municipalities: Altamira, Anapu, Brasil Novo, Medicilândia, Pacajá, Placas, Porto de Moz, Senador José Porfírio, Uruará, and Vitória do Xingu.

2.3. Data acquisition

Epidemiological data (cases) were obtained from the Ministry of Health's Notifiable Diseases Information System. Socio-economic data (Municipal Human Development Index

- MHDI, income and schooling) and cartographic data (territorial boundaries) were collected from the databases of the National Household Sample Survey and the 2022 Census, both from the Brazilian Institute of Geography and Statistics. Environmental data (deforestation increment) was obtained from the National Institute for Space Research's Project for Monitoring deforestation in the Legal Amazon by Satellite. The data related to public health policies (Coverage of the Family Health Strategy - CFHS) was acquired from the Ministry of Health's E-GestorAB.

2.4. Spatial Analysis

The prevalence was calculated (cases/population x 1,000 inhabitants), and the cases and other cartographic data were georeferenced for indexing in the Geographic Database. MHDI was used for socioeconomic variables. Prevalence was divided into five categories and the MHDI into four, using natural breaks. The Bivariate Global Moran's Index (I) was measured between the prevalence of CL and the socioeconomic variable, considering values close to the limits of variation (-1 and 1) to be a strong relationship, inverse for negative values and direct for positive ones. These results were obtained using the ArcGIS 10.5 program.

2.5. Fuzzy analysis

The system was developed using the scikit-fuzzy, scikit-learn, Pandas, Seaborn, and Matplotlib libraries of the Python programming language using the Jupyter Notebook interactive computing platform. The arithmetic mean and standard deviation of the data were calculated to define the parameters of the membership function. The rule database used was elaborated with the aid of a heat map made through Pearson's Correlation. The input variables were classified as epidemiological (prevalence); socioeconomic (income: between 1/2 and 1 minimum wage and low-education: no schooling or schooling up to elementary school); environmental (deforestation increment) and public health policy (CFHS), which went through the fuzzification process and were processed by a rule database and a Mamdani inference engine. Then, after defuzzification using the area center method, the output variable obtained was the Risk of Occurrence of CL (ROCL). Finally, the algorithm accuracy was assessed through the Receiver Operating Characteristic Curve, considering the area under the curve greater than 0.5 or 50% to be significant.

3. Results

The prevalence of CL was very high in Medicilândia (45.47 cases/1,000 inhab.), high in Uruará (28.35 cases/1,000 inhab.), moderate in Placas (21.32 cases/1,000 inhab.), Brasil Novo (20.63 cases/1,000 inhab.) and Anapu (19.09 cases/1,000 inhab.), low in Senador José Porfírio (16.17 cases/1,000 inhab.), Vitória do Xingu (15.83 cases/1,000 inhab.) and Pacajá (13.07 cases/1,000 inhab.) and very low in Altamira (7.75 cases/1,000 inhab.) and Porto de Moz (4.02 cases/1,000 inhab.). Regarding the MHDI, the municipalities had different values, all of which were below 0.8. Moran's index was direct and weak I = 0.075 (Figure 1).

According to the assessment of the model developed, the XIR municipalities had the following ROCL: high in Placas and Medicilândia, moderate in Senador José Porfírio and Uruará, low in Altamira, Anapu, Brasil Novo, Pacajá, and Porto de Moz, and very low in Vitória do Xingu. The defuzification of the variables studied for the municipality

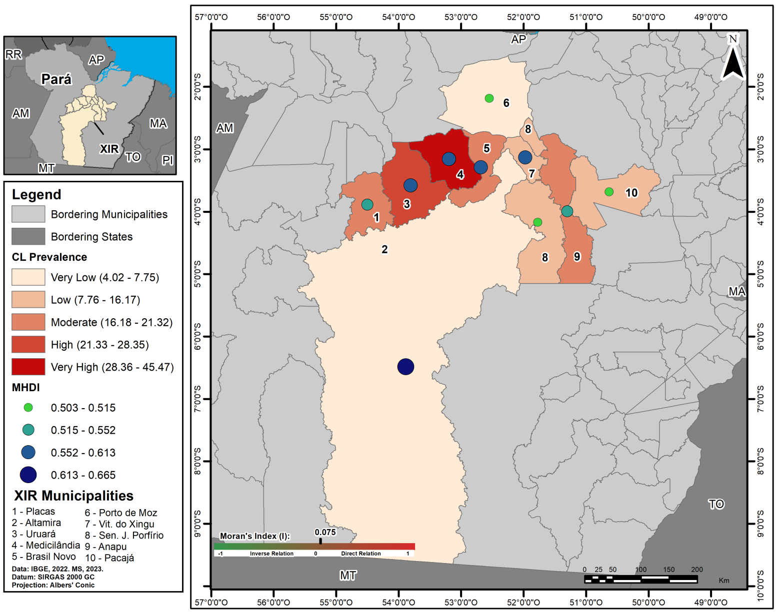

Figure 1. Prevalence of CL and MHDI in the municipalities of the XIR, state of Pará, from 2014 to 2024.

of Placas, which presented the highest risk, showed high low-education, low income, a very low deforestation increment, a very low CFHS, and low prevalence. The ROCL obtained after processing the data using the inference engine was high, at around 68.42% (Figure 2).

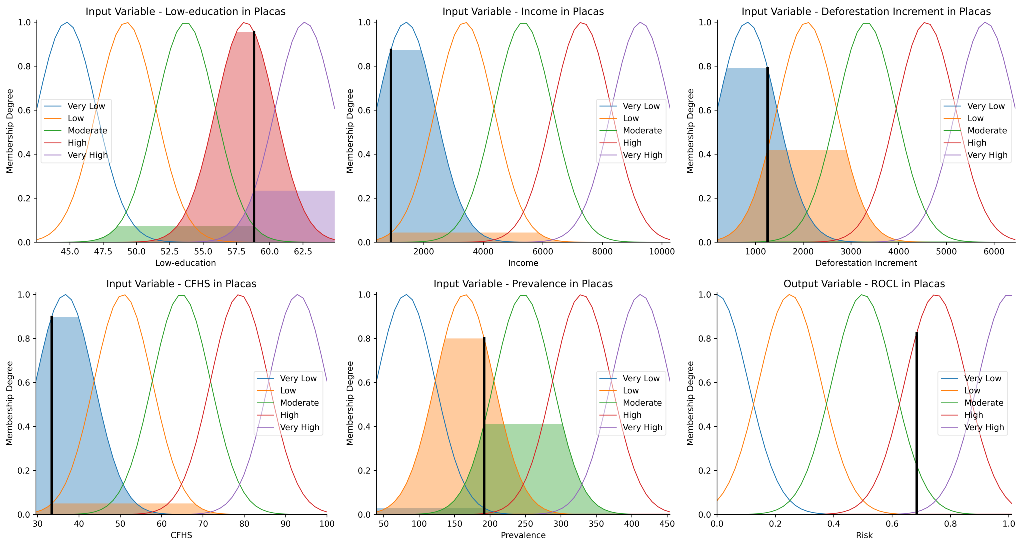

Figure 2. Input variables and ROCL for Placas, state of Pará, from 2014 to 2024.

For the municipality of Vitória do Xingu, which presented the lowest risk, the defuzzification of the variables showed a high rate of low-education, low income, a very low deforestation increment, a very high CFHS, and low prevalence. The risk obtained after processing the data by the inference engine was low, around 13.54% (Figure 3). The ROC curve for the model developed had an area under the curve of 64%.

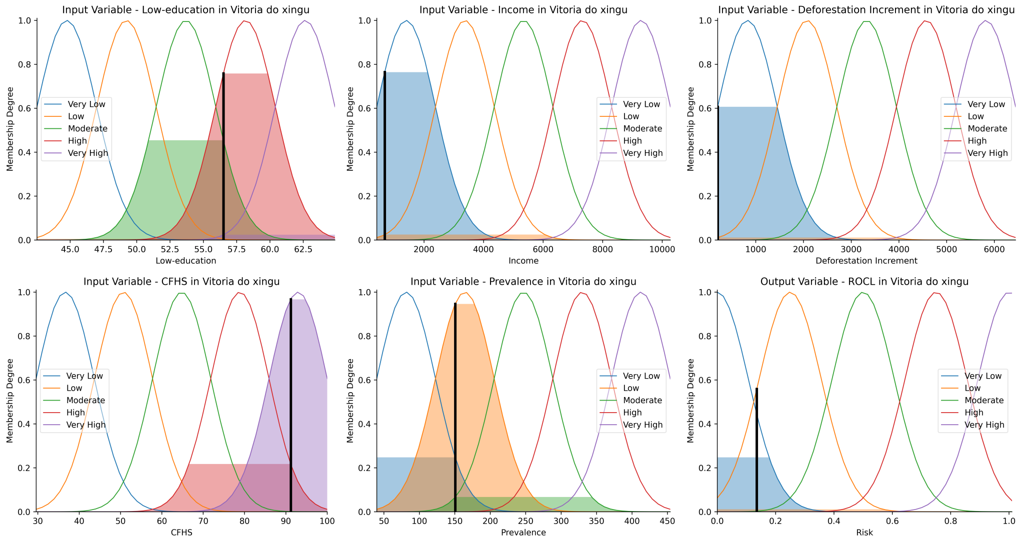

Figure 3. Input variables and ROCL for Vitória do Xingu, state of Pará, from 2014 to 2024.

4. Discussion and Final Considerations

In general, the results showed that although there has been a reduction in its occurrence, CL remains a major public health problem in the XIR, mainly due to the complex epidemiological scenario of the disease in its municipalities, which have different environmental and socioeconomic risk factors that enable the establishment of cycles of transmission of the disease [Sousa-júnior et al. 2020, Souza et al. 2023].

The very high and high prevalence of the disease in municipalities with low and moderate MHDI reflects the impact of the socioeconomic factors of the region on the occurrence of CL, whose impoverished populations are more exposed to the disease. This situation shows how the economic model adopted in the region directly affects its inhabitants, especially those who are socioeconomically vulnerable, and reinforces the neglected nature of CL [Sousa-júnior et al. 2020, Souza et al. 2023].

The algorithm developed was able to show the impact on the overall risk assessment of a municipality, considering the interrelationship between the risk factors, with good accuracy. In this study, a fuzzy logic system was developed based on a multiple input, single output model with a Mamdani inference machine for CL risk, using information collected on the values of 5 risk factors.

The factors identified in this study as being associated with CL were all noninvasive, as they consisted of socioeconomic, environmental, and public health policy information. The results showed that the values of the linguistic variables of the risk factors related to the disease have a wide range of levels of membership, which were assigned based on precise values. Thus, the degree of impact of each linguistic variable on ROCL was identified as a function of the crisp value, which fell between the intervals 0 to 1.

This study analyzed the occurrence of CL and its relationship with socioeconomic, environmental, and public health policy risk factors in the XIR, Pará, Brazil. The epidemiological profile observed follows the trend of disease in other territories of the Amazon region. Analysis of the spatial distribution of the disease and its relationship with the MHDI showed a high prevalence of the disease in municipalities with indicators that point to precarious socio-economic levels.

The fuzzy algorithm developed was satisfactory for observing the interrelationship between the variables studied and was highly accurate. This initial analysis showed that it is possible to use AI tools to analyze infectious diseases such as CL. The combination of risk factors with the result in a single variable facilitates the observation of priority areas and enables better visualization of the scenario of occurrence of this anthropozoonosis, especially for health managers, consequently leading to better management of human, financial, and infrastructure resources to combat this disease.

Overall, the results obtained in this study showed that despite the reduction in the number of cases, CL still has a significant impact on human populations in the Amazon, especially forest communities in the region, a situation that is reflected in the epidemiological scenario of XIR, whose most affected individuals are systematically in a situation of socioeconomic and environmental vulnerability. In addition, it is also possible to conclude that the use of AI tools enables information synthesis for easier observation of the aforementioned situations. Finally, the status of CL in this territory establishes the need to create and implement more efficient methods for monitoring and controlling the disease, such as the use of AI-based tools, with a view to eliminating the disease, in accordance with item 3 of the Sustainable Development Goals.

5. Acknowledgments

To EPIGEO/UEPA and BioCom/USP. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

Bailey, F. et al. (2019). Cutaneous leishmaniasis and co-morbid major depressive disorder: A systematic review with burden estimates. *PLoS Neglected Tropical Diseases*, 13.

Brasil (2017). Manual de vigilância da leishmaniose tegumentar. Technical report, Brazillian Ministry of Health.

Gauthier, C. and Moran, E. (2018). Public policy implementation and basic sanitation issues associated with hydroelectric projects in the brazilian amazon: Altamira and the belo monte dam. 97.

PAHO (2023). Leishmanioses: Informe epidemiológico das américas, no. 12 (dezembro 2023). Technical report, Pan American Health Organization.

Sousa-júnior, A. et al. (2020). Cutaneous leishmaniasis spatial distribution and epidemiological and environmental risk factors in cametá, state of pará, brazil. *Brazilian Journal of Infectious Diseases*, 24.

Souza, B. et al. (2023). Leishmaniose cutânea no estado do pará, brasil: Análise baseada em lógica fuzzy. *Jounal of Health Informatics*, 15.

WHO (2010). Control of the leishmaniases: report of a meeting of the who expert committee on the control of leishmaniasis. Technical report, World Health Organization.