



Estratificação de risco para hanseníase em área hiperendêmica combinando mineração de dados e Anti-PGL-I

Risk stratification for leprosy in a hyperendemic area combining data mining and Anti-PGL-I

Estratificación del riesgo de lepra en una zona hiperendémica combinando minería de datos y Anti-PGL-I

Antônio Anastasis de Oliveira Martins¹, Wanessa Figueira Nunes de Matos²², Douglas Augusto Melo dos Santos², Gabriela Reis Duarte¹, Amanda Botelho Pereira², Junior da Silva Gomes², Marcos Mickael Gomes Carvalho², Glauciney Pereira Gomes², Valney Mara Gomes Conde², Guilherme Augusto Barros Conde¹.

RESUMO

Objetivo: Investigar a aplicação combinada de sorologia anti-PGL-1 e mineração de dados na vigilância da hanseníase em área hiperendêmica. **Métodos:** Estudo epidemiológico com 280 comunicantes em Santarém-PA, utilizando clusterização (*K-means*) como 11 variáveis: faixa etária, gênero, grau de parentesco com o caso índice e período de convivência com o caso índice, forma clínica do caso índice, classificação operacional do caso índice, número de cômodos, número total de moradores e existência de dormitório com mais de 2 pessoas, renda familiar, tempo de residência na moradia da família e análise sorológica (ELISA anti-PGL-I). **Resultados:** Identificaram-se três clusters de risco (C1-C3), sendo o C3 (alto risco) associado a maior carga bacilar (OD mediana=0,63), condições socioeconômicas precárias e 54,3% dos casos novos. A soropositividade foi de 39,6% entre os comunicantes. **Conclusão:** Este estudo demonstra que a integração da sorologia anti-PGL-I com técnicas de mineração de dados representa uma abordagem inovadora e eficaz para a vigilância da hanseníase em áreas hiperendêmicas, permitindo identificar três *clusters* epidemiológicos distintos (C1-C3) com gradiente crescente de risco associado à carga bacilar, condições socioeconômicas e características dos casos-índice. A abordagem integrada mostrou-se eficaz para estratificação de risco e vigilância direcionada.

Palavras-chave: Hanseníase, Diagnóstico precoce, Sorodiagnóstico, Mineração de dados, Vigilância epidemiológica.

ABSTRACT

Objective: To investigate the combined application of anti-PGL-1 serology and data mining in leprosy surveillance in a hyperendemic area. **Methods:** Epidemiological study with 280 contacts in Santarém-PA, using clustering (*K-means*) as 11 variables: age group, gender, degree of kinship with the index case and period of coexistence with the index case, clinical form of the index case, operational classification of the index case, number of rooms, total number of residents and existence of a dormitory with more than 2 people, family income, time of residence in the family home and serological analysis (ELISA anti-PGL-I).

¹ Universidade do Oeste do Pará (UFOPA), Santarém - PA.

² Universidade do Estado do Pará (UEPA), Belém - PA.

Results: Three risk clusters were identified (C1-C3), with C3 (high risk) associated with a higher bacillary load (median OD=0.63), poor socioeconomic conditions and 54.3% of new cases. Seropositivity was 39.6% among contacts. **Conclusion:** This study demonstrates that the integration of anti-PGL-I serology with data mining techniques represents an innovative and effective approach for leprosy surveillance in hyperendemic areas, allowing the identification of three distinct epidemiological clusters (C1-C3) with an increasing gradient of risk associated with bacillary load, socioeconomic conditions and characteristics of index cases. The integrated approach proved to be effective for risk stratification and targeted surveillance.

Keywords: Leprosy, Early diagnosis, Serodiagnosis, Data mining, Epidemiological surveillance.

RESUMEN

Objetivo: Investigar la aplicación combinada de laserología anti-PGL-1 y la minería de datos en la vigilancia de la lepra en un área hiperendémica. **Métodos:** Estudio epidemiológico con 280 contactos en Santarém-PA, utilizando *clustering* (K-means) como 11 variables: grupo de edad, sexo, grado de parentesco con el caso índice y período de coexistencia con el caso índice, forma clínica del caso índice, clasificación operacional del caso índice, número de habitaciones, número total de residentes y existencia de dormitorio con más de 2 personas, renta familiar, tiempo de residencia en el domicilio familiar y análisis serológico (ELISA anti-PGL-I). **Resultados:** Se identificaron tres grupos de riesgo (C1-C3), siendo C3 (alto riesgo) asociado a mayor carga bacilar (mediana OD=0,63), malas condiciones socioeconómicas y 54,3% de casos nuevos. La seropositividad fue del 39,6% entre los contactos. **Conclusión:** Este estudio demuestra que la integración de laserología anti-PGL-I con técnicas de minería de datos representa un enfoque innovador y eficaz para la vigilancia de la lepra en áreas hiperendémicas, permitiendo la identificación de tres grupos epidemiológicos distintos (C1-C3) con un gradiente creciente de riesgo asociado con la carga bacilar, las condiciones socioeconómicas y las características de los casos índice. El enfoque integrado ha demostrado ser eficaz para la estratificación del riesgo y la vigilancia específica.

Palabras clave: Lepra, Diagnóstico temprano, Serodiagnóstico, Minería de datos, Vigilancia epidemiológica.

INTRODUÇÃO

A hanseníase persiste como um relevante problema de saúde pública no Brasil, principalmente nas regiões Norte e Nordeste, onde se concentram os maiores coeficientes de detecção da doença. Conforme dados recentes da Organização Mundial da Saúde (OMS, 2022), o país registrou 27.863 novos casos em 2019, mantendo-se como o segundo em número absoluto de ocorrências no mundo. Esse cenário epidemiológico preocupante se agrava quando consideramos os persistentes desafios no diagnóstico precoce e no acompanhamento adequado dos comunicantes, fatores que contribuem diretamente para a manutenção da cadeia de transmissão (FRADE MCA, et al., 2017; SALES AM, et al., 2011). Além disso, a alta proporção de casos multibacilares reflete não apenas o diagnóstico tardio, mas também a transmissão ativa e sustentada da infecção na comunidade (BRASIL, 2021).

Nesse contexto, o município de Santarém, localizado no oeste do Pará, destaca-se como área hiperendêmica, com coeficientes de detecção consistentemente acima dos parâmetros nacionais. Estudos regionais demonstram que a elevada carga da doença nesta localidade está associada a múltiplos fatores, incluindo dificuldades de acesso aos serviços de saúde, condições socioeconômicas desfavoráveis e limitações nas estratégias de vigilância epidemiológica (BARRETO JG, et al., 2012). Vale ressaltar que, em tais regiões, a falta de recursos e a carência de profissionais capacitados agravam ainda mais o cenário, dificultando a implementação de medidas eficazes de controle (BARRETO JG, et al., 2015).

Diante desse quadro, o desenvolvimento de estratégias inovadoras para identificação precoce de casos e vigilância de comunicantes torna-se imperativo. Nesse aspecto, a sorologia anti-PGL-I emerge como ferramenta promissora, uma vez que permite detectar infecções subclínicas e monitorar a carga bacilar (MOURA RS, et al., 2008; SPENCER JS, et al., 2012). Entretanto, sua aplicação isolada apresenta

limitações, especialmente em termos de sensibilidade, o que justifica a busca por abordagens complementares (SALES AM, et al., 2011). Nesse sentido, técnicas de mineração de dados têm sido aplicadas com sucesso em estudos epidemiológicos, oferecendo novas perspectivas para análise integrada de variáveis clínicas e socioeconômicas (HAN J, et al., 2014). A combinação dessas abordagens com marcadores sorológicos pode potencializar a identificação de grupos de risco e otimizar as estratégias de controle. Contudo, ainda são escassos os estudos que aplicam essa integração metodológica no contexto da hanseníase, particularmente em áreas hiperendêmicas da Amazônia brasileira (GOBBO AR, et al., 2022).

Portanto, este estudo teve como objetivo investigar a epidemiologia da hanseníase em Santarém-PA, utilizando uma abordagem integrada que combina técnicas de mineração de dados com inteligência artificial e sorologia anti-PGL-I. Especificamente, buscamos: (1) analisar o perfil epidemiológico dos casos índices, seus comunicantes e casos novos diagnosticados à partir dos comunicantes; (2) identificar clusters de risco mediante técnicas de inteligência artificial; e (3) aplicar o processo de descoberta de conhecimento com técnicas de mineração de dados para avaliar a correlação entre soropositividade para anti-PGL-I e características clínico-epidemiológicas. Os resultados obtidos fornecem subsídios importantes para o aprimoramento das estratégias de vigilância e controle da hanseníase em contextos similares, contribuindo assim para a redução da carga da doença em áreas hiperendêmicas.

MÉTODOS

Trata-se de um estudo epidemiológico, descritivo, analítico de corte transversal, combinando abordagens quantitativas e técnicas mineração de dados com inteligência artificial. A pesquisa foi conduzida no município de Santarém, Pará, região considerada hiperendêmica para hanseníase segundo parâmetros do Ministério da Saúde. Para garantir uma análise abrangente, o período de estudo compreendeu dados secundários de 2003 a 2013 do Sistema de Informação de Agravos e Notificação (SINAN), complementados por busca ativa realizada em 2014.

A população do estudo incluiu dois grupos distintos: os casos índices (pacientes previamente diagnosticados e registrados no SINAN) e seus comunicantes intradomiciliares. Para seleção da amostra, adotou-se um processo aleatório estratificado, no qual foram selecionados 70 casos índices dentre os 792 registrados no SINAN entre 2003-2013. Posteriormente, esses casos serviram como ponto de partida para a busca ativa com visita aos domicílios para avaliação dermatoneurológica e coleta de amostras sanguíneas para análise sorológica dos 70 casos índices selecionados e seus comunicantes, onde foi realizada a identificação e avaliação de 280 comunicantes, dos quais 35 foram diagnosticados como novos casos durante a investigação.

Para as variáveis sociodemográficas, foram considerados: idade, sexo, renda familiar, número de cômodos no domicílio, quantidades de moradores no domicílio, vacinação BCG e tempo de convivência com o caso índice. Quanto às variáveis clínicas, registraram-se: classificação operacional, forma clínica e resultados do teste ELISA anti-PGL-I das amostras sanguíneas.

Coletou-se amostras sanguíneas de todos os participantes para realização do teste ELISA anti-PGL-I, seguindo protocolo padronizado conforme descrito por Barreto JG, et al. (2011). Considerou-se como ponto de corte para soropositividade valores $\geq 0,295$ de densidade óptica, de acordo com parâmetros estabelecidos na literatura. Todas as análises laboratoriais foram conduzidas no Laboratório de Dermatologia da UFPA, com controle de qualidade rigoroso em todas as etapas.

Na abordagem analítica, o estudo empregou as técnicas de estatística descritiva para caracterização da amostra. Para identificar e caracterizar risco de infecção para hanseníase ente os comunicantes, empregou-se a técnica de inteligência artificial de clusterização com o algoritmo *K-means* e interpretação dos resultados com o processo de descoberta de conhecimento com mineração de dados, como descrito por Fayyad U, et al. (1996).

Para o processo de clusterização foram utilizadas 11 variáveis como segue: características dos comunicantes (faixa etária, gênero, grau de parentesco com o caso índice e período de convivência com o caso índice), características do caso índice (forma clínica do caso índice e classificação operacional do caso índice), condições domiciliares (número de cômodos, número total de moradores e existência de dormitório com mais de 2 pessoas) e perfil socioeconômico (renda familiar e tempo de residência na moradia da família). O número de clusters foi determinado pelo método do cotovelo, sendo validado posteriormente por análise de silhueta, como descrito em Abdunnassara A, et al. (2021).

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual do Pará (Parecer nº 6.116.994, CAAE: 70318423.2.0000.5168 em conformidade com a Resolução 466/12 do CNS. Todos os participantes assinaram Termo de Consentimento Livre e Esclarecido, garantindo-se a confidencialidade das informações. Adicionalmente, implementou-se protocolos rigorosos de controle de qualidade em todas as etapas da pesquisa.

RESULTADOS E DISCUSSÃO

Os resultados sorológicos apresentados na **Tabela 1** revelam padrões epidemiológicos significativos para compreensão da dinâmica de transmissão da hanseníase. Em primeiro lugar, a análise do marcador anti-PGL-I demonstrou que 39,6% dos comunicantes (111/280) apresentaram soropositividade, com mediana de densidade óptica (O.D.) de 0,546. Vale destacar que, entre os casos novos diagnosticados, 37,1% (13/35) foram soropositivos, embora com mediana de O.D. ligeiramente inferior (0,484).

Um achado particularmente relevante foi a marcante discrepância observada entre soropositividade e diagnóstico clínico: apenas 11,7% dos comunicantes soropositivos (13/111) foram efetivamente diagnosticados como casos novos. Esses achados refletem a existência de infecções subclínicas, nas quais indivíduos apresentam resposta imunológica mensurável sem, contudo, manifestações clínicas evidentes e possíveis limitações inerentes ao teste sorológico, que, embora extremamente útil como marcador epidemiológico, apresenta sensibilidade reduzida para diagnóstico individual, conforme discutido por Moura RS, et al. (2008) e Gobbo AR, et al. (2022).

Aprofundando a análise, os valores de O.D. revelaram que 6,3% dos comunicantes soropositivos (7/111) apresentaram níveis particularmente elevados ($\geq 1,1$), o que sugere alta carga bacilar. Entretanto, chama a atenção que, destes 7 comunicantes apenas dois foram diagnosticados como casos novos, enquanto os demais permaneceram assintomáticos. Esse achado vem corroborar estudos anteriores (SPENCER JS, et al., 2012) que sugerem que fatores individuais de susceptibilidade, além da carga bacilar, desempenham papel crucial no desenvolvimento das manifestações clínicas da doença.

Esses resultados trazem importantes implicações para a vigilância da hanseníase, pois o anti-PGL-I mostrou-se efetivamente valioso como ferramenta epidemiológica para identificar grupos com maior exposição ao *M. leprae*. No entanto, a sorologia deve ser sempre interpretada em conjunto com a avaliação clínica, jamais como método diagnóstico isolado. Particularmente, comunicantes com altos títulos de anticorpos (O.D. $\geq 1,1$) merecem acompanhamento especializado, mesmo na ausência de sintomas (GOBBO AR, et al., 2022).

Esses achados reforçam a complexa relação entre resposta imunológica humoral e manifestação clínica na hanseníase, evidenciando claramente a necessidade de abordagens integradas que combinem vigilância sorológica, avaliação clínica sistemática e análise de fatores de risco individuais para um controle efetivo da doença.

Tabela 1 - Distribuição de soropositividade para anti-PGL-I entre comunicantes de hanseníase (n = 280) e casos novos (n = 35) no município de Santarém-PA, Brasil.

Variável	Caso Índice	Comunicantes		
	Total 70	Total 280	Comunicantes 245	Caso Novo 35
	N (%)	N (%)	N (%)	N (%)
anti-PGL-I \geq 0,295 (soropositivo)	25 (35,7)	111 (39,6)	98 (40,0)	13 (37,1)
anti-PGL-I < 0,295 (soronegativo)	45 (64,3)	169 (60,4)	147 (60,0)	22 (62,9)
0.3 --- 0.5	13 (18,6)	63 (22,5)	56 (22,9)	7 (20,0)
0.5 --- 1.0	9 (12,9)	40 (14,3)	34 (13,9)	6 (17,1)
1.0 --- 1.5	1 (1,4)	6 (2,1)	6 (2,4)	0 (0,0)
1.5 --- 2.0	1 (1,4)	2 (0,7)	2 (0,8)	0 (0,0)
> 2.0	1 (1,4)	0 (0,0)	0 (0,0)	0 (0,0)

Legenda: Dados apresentados como N (%). O.D. mediana: 0,546 (comunicantes) e 0,484 (casos novos).

Fonte: Martins AAO, et al., 2025.

O **Gráfico 1** ilustra a distribuição heterogênea de soropositividade entre 6 Grupos Familiares (GF) dos 7 comunicantes que apresentaram valores de O.D. acima de 1,1, limiar que denota alta carga bacilar. Nestes 6 GFs foram encontrados 34 comunicantes, mas apenas 5,9% (2/34) destes desenvolveu hanseníase (círculos), reforçando a predominância de infecções subclínicas. O **Gráfico 1** revela dados cruciais sobre a distribuição de anticorpos anti-PGL-I nestes 6 GFs, evidenciando uma marcante variação nos perfis de soropositividade entre as famílias estudadas, o que tem profundas repercussões para o planejamento de ações de vigilância epidemiológica.

Dos 34 comunicantes analisados, constatou-se que 64,7% (22/34) eram soropositivos para anti-PGL-I, entretanto, apenas 5,9% (2/34) foram diagnosticados como casos novos e desenvolveram manifestações clínicas da doença, sendo 1 soropositivo (GF C) e 1 soronegativo (GF A). Essa expressiva discrepância não só confirma a elevada prevalência de portadores assintomáticos com alta carga bacilar, como também expõe a intrincada dinâmica de transmissão subclínica em ambientes familiares hiperendêmicos, achado que corrobora estudos prévios de Barreto JG, et al. (2012) e Gobbo AR, et al. (2022).

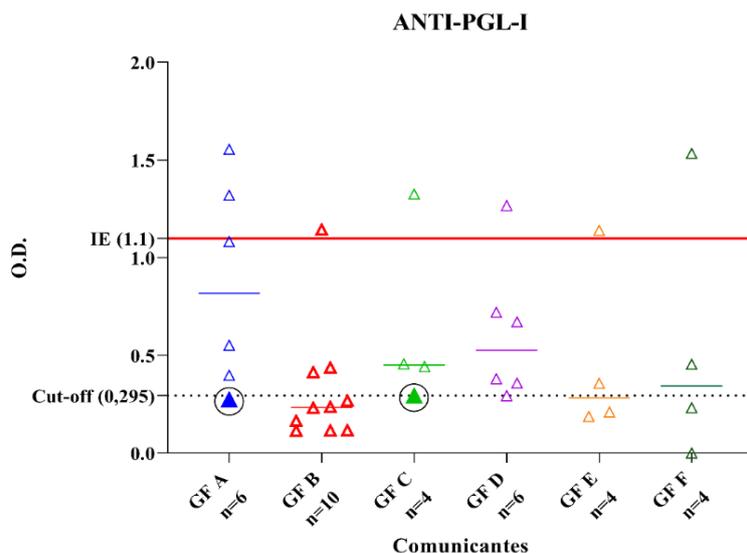
Uma análise mais detalhada dos padrões de transmissão intrafamiliar permitiu identificar três situações epidemiológicas distintas quanto aos GFs famílias GF A, GF C e GF F que apresentaram múltiplos membros com elevados títulos de anti-PGL-I sem, no entanto, qualquer caso clínico confirmado entre os comunicantes com essa titulação elevada. Curiosamente, entre os membros da Família GF A, o único que apresentou soronegatividade foi diagnosticado como caso novo. Somente a Família GF C demonstrou correlação entre soropositividade e diagnóstico de novo caso e a Família GF B exibiu padrão discordante, com múltiplos membros soronegativos.

Esses achados apontam inequivocamente que, para além da carga bacilar, elementos como suscetibilidade genética e perfis imunológicos individuais desempenham papel determinante na progressão para formas clínicas da doença. Tal interpretação corrobora com o modelo multifatorial proposto por Spencer JS, et al. (2012) para a patogênese da hanseníase.

Do ponto de vista operacional, o **Gráfico 1** destaca que a importância do monitoramento prolongado de comunicantes com alta carga bacilar, mesmo na ausência de sintomas. A necessidade de abordar a unidade familiar como um todo nas estratégias de vigilância e a limitação do anti-PGL-I como marcador isolado para diagnóstico individual.

Estes resultados reforçam a utilidade da sorologia como ferramenta complementar para identificar focos de transmissão ativa e priorizar ações de controle em áreas endêmicas, em consonância com as recomendações da OMS (2022) para o manejo de comunicantes de hanseníase.

Gráfico 1 - Níveis de anticorpos anti-PGL-I (O.D. $\geq 1,1$) em seis grupos familiares de comunicantes de hanseníase (n=34).



Legenda: Grupos Familiares (GFA–GFF). Círculos: casos novos (2/34). ELISA: cut-off $\geq 0,295$. **Fonte:** Martins AAO, et al., 2025.

A aplicação do algoritmo *K-means* permitiu identificar três *clusters* epidemiológicos distintos entre os comunicantes, caracterizados por perfis de risco de infecção para a hanseníase progressivamente maiores. O processo de descoberta de conhecimento com mineração de dados permitiu a estratificação dos perfis de riscos através das características dos *clusters* encontrados e revelou padrões importantes associados a fatores socioeconômicos e ambientais.

O *Cluster C1* (Baixo Risco), que representou 22,1% dos comunicantes, foi marcado por melhores condições socioeconômicas — com 58,1% das famílias apresentando renda de até dois salários-mínimos — e menor densidade domiciliar, já que 61,3% dos lares não apresentavam superlotação. Esses dados indicam um contexto protetor frente à transmissibilidade da hanseníase.

Em contraste, o *Cluster C2* (Médio Risco), englobando 32,1% dos comunicantes, apresentou um perfil mais vulnerável. Embora todos os casos-índice associados a esse grupo fossem paucibacilares (100%), observou-se uma elevada taxa de superlotação domiciliar (56,7%), o que sugere que, mesmo em casos com menor carga bacilar, as condições ambientais desfavoráveis podem aumentar o risco de transmissão.

Por sua vez, o *Cluster C3* (Alto Risco), que compreendeu 45,7% dos comunicantes, destacou-se como o grupo mais vulnerável. Este cluster foi caracterizado por associação exclusiva com casos-índice multibacilares (100%), maior proporção de famílias com baixa renda (48,4% recebendo apenas um salário-mínimo) e elevada superlotação domiciliar (57,8%). Esses fatores compõem um cenário de alto risco, potencializando a transmissibilidade da doença.

Tais achados estão em plena consonância com a literatura existente, corroborando estudos prévios que apontam a relação direta entre condições socioeconômicas adversas e a maior transmissibilidade da hanseníase (NIITSUMAENA, et al., 2021). Além disso, os dados sorológicos apresentados na Figura 2 revelam um gradiente crescente nos níveis de anti-PGL-1 do C1 ao C3, o que não apenas reforça a classificação de risco proposta, mas também sugere uma possível relação dose-resposta entre fatores socioambientais e carga bacilar.

Esses resultados têm implicações importantes para as estratégias de controle da hanseníase. Primeiramente, reforçam a necessidade de abordagens diferenciadas conforme o nível de risco identificado. Em segundo lugar, evidenciam a superlotação domiciliar como um fator modificável de grande relevância, que deve ser alvo de políticas públicas. Por fim, validam o uso integrado de técnicas de mineração de dados e marcadores sorológicos como ferramentas eficazes para a estratificação de risco em contextos endêmicos.

À luz desses achados, torna-se evidente que intervenções focalizadas no Cluster C3 podem ter um impacto mais significativo na interrupção da cadeia de transmissão da hanseníase. No entanto, é igualmente importante considerar que, conforme demonstrado pelo Cluster C2, mesmo comunicantes de casos paucibacilares requerem atenção especial quando expostos a condições habitacionais precárias.

Tabela 2 - Características epidemiológicas dos comunicantes de hanseníase (n=280) nos *clusters* C1 (baixo risco), C2 (médio risco) e C3 (alto risco) em Santarém-PA, Brasil.

Variáveis		C1 (Baixo Risco) (n=62)	C2 (Médio Risco) (n=90)	C3 (Alto Risco) (n=128)
		N (%)	N (%)	N (%)
Sexo	Feminino	44 (71,0%)	55 (61,1%)	76 (59,4%)
	Masculino	18 (29,0%)	35 (38,9%)	52 (40,6%)
Faixa Etária	0 - 14 anos	13 (21,0%)	25 (27,8%)	23 (18,0%)
	15 - 39 anos	31 (50,0%)	48 (53,3%)	62 (48,4%)
	40 - 59 anos	12 (19,4%)	9 (10,0%)	23 (18,0%)
	>60anos	6 (9,7%)	8 (8,9%)	20 (15,6%)
Grau de Parentesco	Cônjuge	7 (11,3%)	9 (10,0%)	10 (7,8%)
	Filho(a)	19 (30,6%)	28 (31,1%)	24 (18,8%)
	Irmão	7 (11,3%)	9 (10,0%)	9 (7,0%)
	Não parente	0 (0,0%)	2 (2,2%)	17 (13,3%)
	Pai/Mãe	10 (16,1%)	8 (8,9%)	11 (8,6%)
	Parente consanguíneo	9 (14,5%)	12 (13,3%)	24 (18,8%)
Período de Conveniência	Parente não consanguíneo	10 (16,1%)	22 (24,4%)	33 (25,8%)
	Entre cinco a dez anos	8 (12,9%)	13 (14,4%)	19 (14,8%)
	Entre dois a cinco anos	3 (4,8%)	6 (6,7%)	6 (4,7%)
	Entre um a dois anos	2 (3,2%)	3 (3,3%)	5 (3,9%)
	Mais que dez anos	46 (74,2%)	61 (67,8%)	84 (65,6%)
Vacina BCG, Cicatriz	Mais que um ano	3 (4,8%)	7 (7,8%)	14 (10,9%)
	0	5 (8,1%)	7 (7,8%)	25 (19,5%)
	1	50 (80,6%)	72 (80,0%)	94 (73,4%)
	2	6 (9,7%)	11 (12,2%)	9 (7,0%)
Dormitórios > 2 Pessoas	Duvidosa	1 (1,6%)	0 (0,0%)	0 (0,0%)
	Não	38 (61,3%)	39 (43,3%)	54 (42,2%)
	Sim	24 (38,7%)	51 (56,7%)	74 (57,8%)
Renda Familiar	Sem renda	0 (0,0%)	4 (4,4%)	2 (1,6%)
	Menor que um salário-mínimo	0 (0,0%)	8 (8,9%)	12 (9,4%)
	Um salário-mínimo	4 (6,5%)	26 (28,9%)	62 (48,4%)
	Até dois salários-mínimos	36 (58,1%)	34 (37,8%)	37 (28,9%)
Classificação Operacional – Índice	Até três salários-mínimos	9 (14,5%)	12 (13,3%)	6 (4,7%)
	Maior que três salários-mínimos	13 (21,0%)	6 (6,7%)	9 (7,0%)
	Paucibacilar (PB)	5 (8,1%)	90 (100,0%)	0 (0,0%)
	Multibacilar (MB)	57 (91,9%)	0 (0,0%)	128 (100,0%)
Forma Clínica – Índice	Total	62 (100,0%)	90 (100,0%)	128 (100,0%)
	Indeterminada (I)	1 (1,6%)	44 (48,9%)	0 (0,0%)
	Neural Pura (NP)	0 (0,0%)	7 (7,8%)	0 (0,0%)
	Tuberculóide (T)	4 (6,5%)	39 (43,3%)	0 (0,0%)
	Dimorfa (D)	7 (11,3%)	0 (0,0%)	8 (6,3%)
	Borderline Tuberculóide (BT)	19 (30,6%)	0 (0,0%)	71 (55,5%)
	Borderline Borderline (BB)	7 (11,3%)	0 (0,0%)	20 (15,6%)
	Borderline Virchowiana (BV)	10 (16,1%)	0 (0,0%)	5 (3,9%)
Virchowiana (V)	14 (22,6%)	0 (0,0%)	24 (18,8%)	

Fonte: Martins AAO, et al., 2025.

A **Tabela 2** apresenta os resultados da clusterização dos 280 comunicantes em três grupos distintos (C1, C2 e C3) com base em características epidemiológicas e clínicas. A análise revelou padrões significativos que permitem estratificar o risco de hanseníase nesta população.

Os comunicantes foram distribuídos em: C1 (22,1%; n=62), C2 (32,1%; n=90) e C3 (45,7%; n=128). Observou-se gradiente socioeconômico marcante entre os clusters. No C1 apresentou melhores condições: 58,1% com renda até 2 salários-mínimos e apenas 38,7% com superlotação domiciliar. O C3 mostrou piores indicadores: 48,4% com apenas 1 salário-mínimo e 57,8% com superlotação e o C2 apresentou perfil intermediário, mas com 56,7% de superlotação.

Quanto à classificação operacional e forma clínica a análise revelou associação significativa entre os clusters e a classificação dos casos-índice, onde 100% dos comunicantes em C2 tinham casos-índice paucibacilares (PB), 100% dos comunicantes em C3 tinham casos-índice multibacilares (MB) e C1 apresentou maioria de MB (91,9%), mas com melhor perfil socioeconômico.

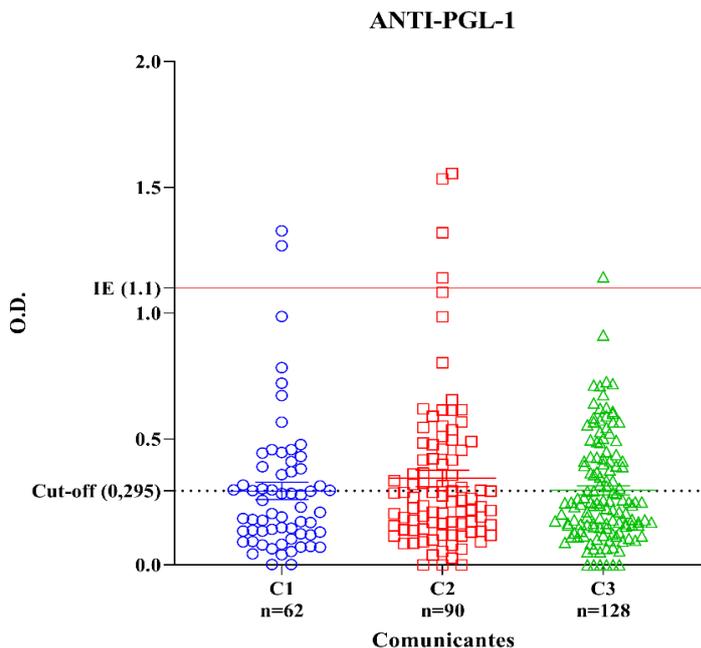
Nas formas clínicas o *cluster* C2 concentrou formas menos graves (48,9% indeterminada, 43,3% tuberculoide) e C3 apresentou predominância de formas avançadas (55,5% borderline-tuberculoide, 18,8% virchowiana). Nas implicações para vigilância esses achados sugerem que as condições socioeconômicas desfavoráveis (C3) associam-se a maior gravidade dos casos-índice; mesmo com casos-índice PB (C2), a superlotação mantém risco significativo e o perfil de C1 indica que melhores condições podem atenuar riscos, mesmo com contato com MB (COSTA ILV, et al., 2023).

Os resultados corroboram estudos prévios que destacam a influência de fatores socioeconômicos na transmissão (NIITSUMA ENA, et al., 2021). A importância da carga bacilar dos casos-índice (BARRETO JG, et al., 2012) e a necessidade de abordagens diferenciadas por perfil de risco (OMS, 2022).

Diante desses achados, a clusterização demonstrou ser ferramenta valiosa para identificar grupos prioritários para intervenção, direcionar recursos de forma mais eficiente e desenvolver estratégias de vigilância personalizadas

A análise da **Tabela 2** evidencia a complexa interação entre fatores clínicos e sociais na hanseníase, reforçando a necessidade de abordagens integradas para seu controle. Os três *clusters* identificados representam diferentes níveis de risco que demandam intervenções específicas, particularmente no cluster C3, que combina alta vulnerabilidade social com exposição a formas multibacilares. Exigindo intervenções específicas por cluster como no C1: manutenção de condições favoráveis; C2: redução da superlotação e C3: ações intensivas no controle de infecção. A integração da sorologia com mineração de dados mostrou-se superior ao uso isolado do anti-PGL-I, conforme discutido por Oskam L, et al. (2003), permitindo estratificar os perfis de riscos e otimizar recursos.

Gráfico 2 - Distribuição dos níveis de anticorpos anti-PGL-I (O.D.) dos comunicantes entre os clusters de risco C1, C2 e C3.



Fonte: Martins AAO, et al., 2025.

O **Gráfico 2** apresenta a distribuição dos níveis de anticorpos anti-PGL-I dos comunicantes entre os três *clusters* identificados (C1, C2 e C3), revelando padrões importantes para a compreensão da dinâmica de infecção da hanseníase. Os resultados demonstram uma progressão significativa nos valores de densidade óptica (O.D.) do cluster C1 (baixo risco) para o C3 (alto risco), corroborando a classificação de risco estabelecida pela análise de clusterização.

Observa-se que os níveis médios de anti-PGL-I no *cluster* C3 foram consistentemente superiores aos demais, fato que pode ser atribuído a dois fatores principais: primeiro, a exposição prolongada a casos-índice multibacilares (100% dos casos no C3 eram MB); segundo, as condições socioeconômicas desfavoráveis que caracterizam este grupo, incluindo maior superlotação domiciliar (57,8%) e menor renda familiar (48,4%) com apenas um salário-mínimo). Esses achados estão em consonância com os estudos de Barreto et al. (2012), que relacionam altos títulos de anti-PGL-I com maior carga bacilar e condições precárias de vida.

A análise comparativa entre os *clusters* revela ainda que o cluster C1 apresentou os menores níveis de anti-PGL-1, condizente com seu perfil de menor risco; o cluster C2 mostrou valores intermediários, refletindo sua posição no espectro de risco e o cluster C3 destacou-se com os maiores valores, confirmando seu status de alto risco.

Estes resultados assumem particular relevância quando consideramos que a variável anti-PGL-I não foi incluída como parâmetro para a formação dos *clusters*, tornando sua correlação com a classificação de risco ainda mais significativa. Como discutido por Oskam L, et al. (2003) e Gobbo AR, et al. (2022), essa concordância entre marcador sorológico e características epidemiológicas reforça a utilidade do anti-PGL-I como ferramenta complementar na vigilância da hanseníase.

Do ponto de vista operacional, o **Gráfico 2** fornece subsídios importantes para validação dos *clusters* identificados, a priorização de ações em populações de maior risco (*cluster* C3) e o monitoramento da eficácia de intervenções por meio de séries temporais de anti-PGL-I.

Ressalta-se, contudo, que cerca de 35% dos indivíduos no *cluster* C3 mantiveram-se soronegativos, destacando as limitações inerentes ao uso isolado deste marcador, conforme alertado por Spencer JS, et al. (2012). Essa constatação reforça a necessidade de abordagens integradas que combinem sorologia, avaliação clínica e análise de fatores de risco para uma vigilância mais efetiva.

Em síntese, os padrões observados no **Gráfico 2** não apenas validam a estratificação de risco proposta, mas também destacam o potencial da integração entre mineração de dados e marcadores sorológicos para o aprimoramento das estratégias de controle da hanseníase em áreas endêmicas.

Tabela 3 - Características epidemiológicas dos casos novos de hanseníase por *clusters* (C1, C2, C3).

Variáveis	C1 (Baixo Risco) (N=6)	C2 (Médio Risco) (N=10)	C3 (Alto Risco) (N=19)	
	N (%)	N (%)	N (%)	
Sexo	Feminino	4 (66,7%)	8 (80,0%)	11 (57,9%)
	Masculino	2 (33,3%)	2 (20,0%)	8 (42,1%)
Faixa Etária	0 - 14 Anos	2 (33,3%)	2 (20,0%)	5 (26,3%)
	15 - 39 Anos	3 (50,0%)	3 (30,0%)	9 (47,4%)
	40 - 59 Anos	0 (0,0%)	4 (40,0%)	3 (15,8%)
	> 60 Anos	1 (16,7%)	1 (10,0%)	2 (10,5%)
Grau de Parentesco	Cônjuge	1 (16,7%)	2 (20,00%)	1 (5,3%)
	Não parente	0 (0,0%)	1 (10,0%)	2 (10,5%)
	Parente não consanguíneo	0 (0,0%)	3 (30,0%)	4 (21,1%)
	Filho(a)	4 (66,7%)	1 (10,0%)	3 (15,8%)
	Irmão	0 (0,0%)	1 (10,0%)	1 (5,3%)
	Pai/Mãe	0 (0,0%)	0 (0,0%)	2 (10,5%)
	Parente consanguíneo	1 (16,7%)	2 (20,0%)	6 (31,6%)
Período de Conveniência	Mais que um ano	0 (0,0%)	1 (10,0%)	2 (10,5%)
	Entre um a dois anos	0 (0,0%)	0 (0,0%)	2 (10,5%)
	Entre dois a cinco anos	0 (0,0%)	1 (10,0%)	0 (0,0%)
	Entre cinco a dez anos	0 (0,0%)	0 (0,0%)	2 (10,5%)
	Mais que dez anos	6 (100,0%)	8 (80,0%)	13 (68,4%)
BCG	0	0 (0,0%)	1 (10,0%)	5 (26,3%)
	1,0	6 (100,0%)	9 (90,0%)	13 (68,4%)
	2,0	0 (0,0%)	0 (0,0%)	1 (5,3%)
Dormitórios > 2 Pessoas	Não	5 (83,3%)	2 (20,0%)	10 (52,6%)
	Sim	1 (16,7%)	8 (80,0%)	9 (47,4%)
Renda Familiar	Menor que um salário-mínimo	0 (0,0%)	1 (10,0%)	0 (0,0%)
	Um salário-mínimo	0 (0,0%)	3 (30,0%)	12 (63,2%)
	Até dois salários-mínimos	5 (83,3%)	4 (40,0%)	4 (21,1%)
	Até três salários-mínimos	1 (16,7%)	1 (10,0%)	2 (10,5%)
	Maior que três salários-mínimos	0 (0,0%)	1 (10,0%)	1 (5,3%)
Classificação Operacional - Índice	Paucibacilar (PB)	1 (16,7%)	10 (100,0%)	0 (0,0%)
	Multibacilar (MB)	5 (83,3%)	0 (0,0%)	19 (100,0%)
Forma Clínica - Índice	Indeterminada (i)	0 (0,0%)	1 (10,0%)	0 (0,0%)
	Neural Pura (NP)	0 (0,0%)	2 (20,0%)	0 (0,0%)
	Tuberculóide (T)	1 (16,7%)	7 (70,0%)	0 (0,0%)
	Dimorfa (D)	1 (16,7%)	0 (0,0%)	2 (10,5%)
	Borderline Tuberculóide (BT)	1 (16,7%)	0 (0,0%)	9 (47,4%)
	Borderline Borderline (BB)	0 (0,0%)	0 (0,0%)	4 (21,1%)
	Borderline Virchowiana (BV)	1 (16,7%)	0 (0,0%)	1 (5,3%)
	Virchowiana (V)	2 (33,3%)	0 (0,0%)	3 (15,8%)

Fonte: Martins AAO, et al., 2025.

A **Tabela 3** acima, apresenta a distribuição dos 35 casos novos de hanseníase diagnosticados entre os três *clusters* (C1, C2 e C3), revelando padrões epidemiológicos distintos que corroboram a estratificação de risco proposta. Os resultados demonstram gradiente crescente na detecção de casos novos conforme aumenta o risco do cluster: C1 (17,1%; n=6), C2 (28,6%; n=10) e C3 (54,3%; n=19).

Quanto a distribuição demográfica, a análise por sexo mostrou predominância feminina em todos os *clusters*: C1: 66,7% mulheres, C2: 80,0% mulheres, C3: 57,9% mulheres. A faixa etária mostra que no C1 concentrou casos entre 15-39 anos (50,0%), C2 apresentou maior dispersão, com 40,0% entre 40-59 anos e C3 manteve predominância de 15-39 anos (47,4%).

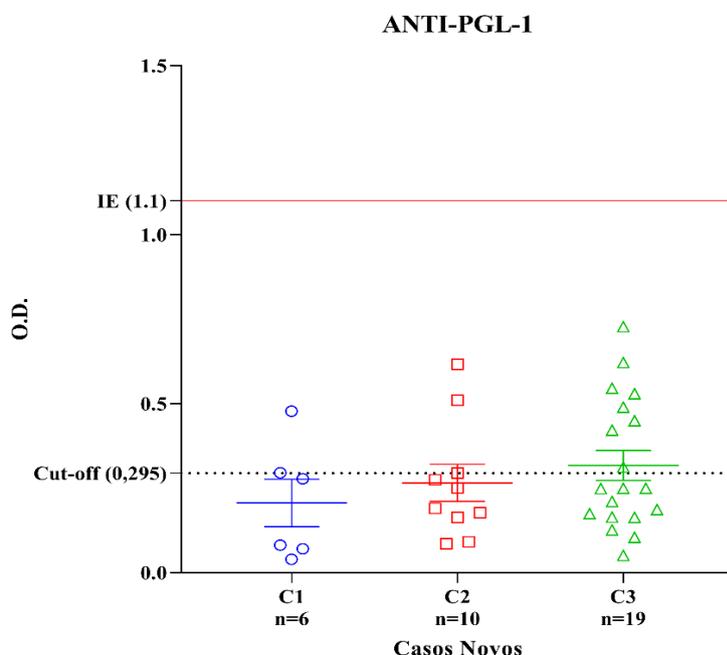
Os casos novos contidos no cluster C3, quanto aos fatores de risco associados, destacaram-se por piores condições habitacionais (47,4% com superlotação), menor renda familiar (63,2%) com apenas 1 salário-mínimo e maior proporção de contatos não consanguíneos (21,1%).

Quanto às formas clínicas, observou-se progressão na gravidade das formas clínicas no C2 foi de 70,0% dos casos na forma tuberculoide, no C3 47,4% borderline-tuberculoide e 15,8% virchowiana. Esses achados reforçam existência de infecções subclínicas, nas quais indivíduos não apresentem manifestações clínicas, como observado no C2, mas apresentem vulnerabilidade socioeconômica, podendo ser um fator determinante (NIITSUMA ENA, et al., 2021; COSTA ILV, et al., 2023). A importância da classificação operacional do caso-índice (BARRETO JG, et al., 2012) e a necessidade de vigilância diferenciada por perfil de risco (OMS, 2022).

A análise sugere que mesmo no *cluster* de baixo risco (C1), a convivência prolongada (>10 anos em 100% dos casos) representa fator relevante. No cluster C2, a superlotação (80,0% dos casos novos) parece superar a proteção relativa de ter casos-índice PB. O *cluster* C3 reúne os piores indicadores, justificando intervenções prioritárias

A **Tabela 3** fornece evidências robustas para validação dos *clusters* de risco, demonstrando correlação clara entre as características epidemiológicas e a detecção de casos novos. Esses achados destacam o potencial da abordagem por clusterização para orientar estratégias de controle da hanseníase de forma mais eficiente e direcionada.

Gráfico 3 - Níveis de densidade óptica (O.D.) de anti-PGL-I em casos novos de hanseníase estratificados por cluster (C1, C2, C3).



Fonte: Martins AAO, et al., 2025.

O **Gráfico 3** apresenta a distribuição dos valores de densidade óptica (O.D.) do anti-PGL-I entre os casos novos de hanseníase classificados nos três *clusters* (C1, C2 e C3), revelando padrões significativos que corroboram a estratificação de risco proposta no estudo.

Os resultados demonstram um gradiente crescente nos níveis de anti-PGL-1 conforme aumenta o risco do cluster. O C1 (baixo risco) apresentou os menores valores de O.D., com mediana de 0.42. Já no *cluster* C2 (médio risco), evidenciou-se valores intermediários, mediana de 0.51 e o Cluster C3 (alto risco) mostrou os maiores níveis, com mediana de 0.63.

Essa progressão é particularmente relevante quando consideramos que o anti-PGL-I não foi utilizado como variável na formação dos *clusters*. A correlação entre O.D. e classificação de risco foi espontânea e os resultados validam independentemente a metodologia de clusterização.

Os padrões observados no **Gráfico 3** reforçam três aspectos fundamentais: relação carga bacilar/gravidade, onde os maiores valores no C3 associam-se às formas clínicas mais graves (55,5% borderline-tuberculoide); os fatores ambientais com a superlotação (57,8% no C3) parecem favorecer maior transmissão e carga bacilar e o monitoramento, pois os níveis de anti-PGL-I podem servir como marcador de acompanhamento.

Quando comparado com a literatura evidencia-se a associação entre carga bacilar e formas clínicas Spencer et al. (2012). A influência de fatores ambientais concorda com Niitsuma ENA, et al. (2021). A utilidade do anti-PGL-1 como marcador epidemiológico está em linha com Oskam L, et al. (2003) e Gobbo AR, et al. (2022).

Os resultados sugerem que o monitoramento sorológico pode auxiliar na vigilância de áreas de alto risco, pois a carga bacilar medida por anti-PGL-I correlaciona-se com a estratificação de risco; a abordagem por clusters permite priorizar recursos para o atendimento de populações em maior risco.

Embora significativos, os resultados devem ser interpretados considerando a amostra relativamente pequena de casos novos (n=35), a variação individual na resposta sorológica e a necessidade de estudos longitudinais para confirmar as associações. A natureza retrospectiva de parte dos dados e o foco em um único município podem limitar a generalização dos resultados. No entanto, a consistência com a literatura e a robustez metodológica sustentam as conclusões.

Esses achados reforçam a importância de abordagens integradas para vigilância epidemiológica em áreas endêmicas, combinando marcadores biológicos com análise de determinantes sociais, conforme recomendações atuais da OMS (2022).

CONCLUSÃO

Este estudo demonstra que a integração da sorologia anti-PGL-I com técnicas de mineração de dados representa uma abordagem inovadora e eficaz para a vigilância da hanseníase em áreas hiperendêmicas, permitindo identificar três *clusters* epidemiológicos distintos (C1-C3) com gradiente crescente de risco associado à carga bacilar, condições socioeconômicas e características dos casos-índice. Os resultados evidenciaram que o *cluster* C3 (alto risco), caracterizado por menor renda familiar (48,4%) com 1 salário-mínimo, superlotação domiciliar (57,8%) e exposição a casos multibacilares, concentrou 54,3% dos casos novos, enquanto a sorologia anti-PGL-I mostrou-se valiosa como marcador epidemiológico complementar (39,6% de soropositividade entre comunicantes). Esses achados reforçam a importância de estratégias diferenciadas de vigilância baseadas em perfis de risco, com priorização de intervenções nas populações mais vulneráveis, monitoramento contínuo de comunicantes em clusters de alto risco e implementação de medidas que combinem abordagens laboratoriais, clínicas e socioeconômicas para o controle efetivo da hanseníase em conformidade com as recomendações atuais da OMS.

REFERÊNCIAS

1. ABDULNASSAR AA e NAIR LRA. Comprehensive Study on the Importance of the Elbo wand the Silhouette Metrics in Cluster Count Prediction for Partition Cluster Models, 2021; 11(4): 3792-3806.

2. BARRETO JG, et al. Soroepidemiologia anti-PGL-I em casos de hanseníase: contatos domiciliares e escolares de um município hiperendêmico da Amazônia brasileira. *Hanseníase*, 2011; 82(4): 358-370.
3. BARRETO JG, et al. High rates of un diagnosed leprosy and subclinical infection amongst school children in the Amazon region. *Memórias do Instituto Oswaldo Cruz*, 2012; 107(Suppl 1): 60-67.
4. BARRETO JG, et al. Spatial analysis spotlighting early childhood leprosy transmission in a Hyperendemic municipality of the Brazilian Amazon region. *PLoS NeglTropDis*. 2014;8:e2665.
5. BRASIL. Ministério da Saúde. Acompanhamento dos dados de Hanseníase - Pará. Brasília: Sistema de Informação de Agravos de Notificação - SINAN Net; 2022. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/hanswpa.def>.
6. BRASIL. Ministério da Saúde. Boletim Epidemiológico de Hanseníase. Brasília: Ministério da Saúde, 2021.
7. BRENNAN PJ e SPENCER JS. The Physiology of Mycobacterium leprae Morphology and Cell Wall. *The International Text book of Leprosy*. 2019;6:1-51.
8. CASTRO SS, et al. Leprosy incidence, characterization of cases and correlation with house hold and cases variables of the Brazilian states in 2010. *Anais Brasileiros de Dermatologia*, 2016; 91(1): 28-33.
9. COSTA ILV, et al. Leprosy among children in an area without primary healthcare coverage in Caratateua Island, Brazilian Amazon. *Front Med.*, 2023; 10: 1218388.
10. DA SILVA YED, et al. Application of Clustering Technique with Kohonen Self-organizing Maps for the Epidemiological Analysis of Leprosy. In: Arai K, et al. (Eds.). *Intelligent Systems and Applications - Proceedings of the 2018 Intelligent Systems Conference, IntelliSys 2018, London, UK, September 6-7, 2018, Volume 2*. Springer, Cham, 2019; 869: 295-309.
11. FABRI ACOC. Análise comparativa da reatividade anti-LID-1, NDO-LID, NDO-HSA e PGL-1 em hanseníase. Belo Horizonte: Universidade Federal de Minas Gerais; 2015.
12. FAYYAD U, et al. From Data Mining to Knowledge Discovery in Databases. *AI Magaz*, 1996; 17(3): 37.
13. FIGINI S e GIUDICI P. *Applied data mining for business and industry*. 2nd ed. Nova Jersey, EUA: John Wiley & Sons Ltd, 2009.
14. FRADE MCA, et al. Unexpectedly high leprosy seroprevalence detected using a random surveillance strategy in mid western Brazil: A comparison of ELISA and a rapid diagnostic test. 2017.
15. GOBBO AR, et al. NDO-BSA, LID-1, and NDO-LID Antibody Responses for Infection and RLEP by Quantitative PCR as a Confirmatory Test for Early Leprosy Diagnosis. *Front Trop Dis.*, 2022; 3: 850886.
16. HAN J, et al. *Data mining: concepts and techniques*. 3rd ed. Waltham: Elsevier; 2014.
17. MOURA RS, et al. Leprosy serology using PGL-I: a systematic review. *Revista da Sociedade Brasileira de Medicina Tropical*, 2008; 41(Suppl 2): 1-8.
18. NIITSUMA ENA, et al. Fatores associados ao adoecimento por hanseníase em contatos: revisão sistemática e metanálise. *Revista Brasileira de Epidemiologia*, 2021; 24: e210029.
19. OSKAM L, et al. Serology: recent developments, strengths, limitations and prospects: a state of the art overview. *Leprosy Review*, 2003; 74(3): 196-205.
20. SALES AM, et al. Leprosy among patient contacts: A multilevel study of risk factors. *PLoS Neglected Tropical Diseases*, 2011; 5(3): 1-6.
21. SILVA GAMA R, et al. Prospects for new leprosy diagnostic tools, a narrative review considering ELISA and PCR assays. *Revista da Sociedade Brasileira de Medicina Tropical*. 2020;53:e20190535.
22. SPENCER JS, et al. Identification of serological biomarkers of infection, disease progression and treatment efficacy for leprosy. *Memórias do Instituto Oswaldo Cruz*. 2012;107(Suppl 1):79-89.
23. VAN HOOIJ A, et al. Field-friendly serological tests for determination of *M. leprae*-specific antibodies. *Sci Rep.*, 2017; 7(1): 8868.
24. WORLD HEALTH ORGANIZATION. *Global leprosy strategy: accelerating towards a leprosy-free world*. New Delhi: World Health Organization; 2016.
25. WORLD HEALTH ORGANIZATION. *Global leprosy update, 2018: moving towards a leprosy-free world*. *Weekly Epidemiological Record*. 2019;94(35/36):389-412.
26. WORLD HEALTH ORGANIZATION. *Leprosy*. World Health Organization [Internet]. 2022 [cited 2022 Apr 4]. Available from: <https://www.who.int/news-room/fact-sheets/detail/leprosy>.