

ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR) JUNTAMENTE COM CLASSIFICAÇÃO MULTIVARIADA PARA DETERMINAÇÃO DE SEXO DE MOSQUITOS *Aedes Aegypti*

Jorge L. S. Viana^{1*}; Marfran C. D. Santos¹.

¹Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, Floresta 56400-000, Brazil.

*Jorge.soares@aluno.ifsertao-pe.edu.br

Palavras-Chave: Vigilância epidemiológica, Quimiometria, Análise espectroscópica.

Introdução

Arbovírus são vírus que podem ser transmitidos por espécies artrópodes. Aqui será destacada a espécie de mosquitos *Aedes aegypti*, que são artrópodes transmissores de uma série de vírus como os vírus da Dengue, Zika e a Chikungunya. A vigilância entomológica é um dos passos mais importantes na prevenção de surtos destes arbovírus sendo a principal ação de vigilância entomológica a detecção de focos de vetores nos estágios mais iniciais possíveis. (SANTOS et al., 2022). No trabalho da vigilância, é importante salientar a relevância dos mosquitos fêmeas, pois além de serem as únicas que possuem natureza hematófaga, sendo assim as responsáveis por infecções a humanos, também são as responsáveis diretas pelo aumento da população local da espécie.

A espectroscopia no infravermelho é uma técnica que tem sido amplamente utilizada na detecção de pequenas alterações na composição química de amostras biológicas. Esta técnica funciona a partir da incidência de um feixe de radiação infravermelha que penetrará por alguns micrômetros na amostra, produzindo uma excitação vibracional que pode ser detectada e traduzida em espectros que posteriormente poderão ser interpretados à luz das ligações químicas que produzem determinados sinais de absorvância ou transmitância e seus respectivos números ou comprimentos de onda. Estas técnicas possuem as vantagens de fornecerem resultados rápidos, serem de fácil manuseio, sensíveis às variações químicas e não serem destrutivas para a amostra (Paraskevaidi, et al., 2021)

Em regra, espectros de amostras semelhantes são muito semelhantes. Sendo assim, geralmente faz-se necessário a utilização de algoritmos computacionais capazes de detectar mínimas alterações nos espectros causadas pela presença de algum componente bioquímico diferente no meio. Deu-se o nome de quimiometria a esta área de estudos que envolvem a análise de informações químicas (espectros, por exemplo) a partir de modelos estatístico-matemáticos. Alguns destes algoritmos amplamente utilizados na quimiometria são o Algoritmo de Projeções Sucessivas (SPA), e Algoritmo Genético (GA) que podem ser utilizados na seleção de variáveis com maior potencial de discriminar espectros e classificá-los em suas respectivas classes. Estes algoritmos de seleção de variáveis podem ser combinados com um algoritmo de discriminação como Análise Discriminante Linear (LDA). O LDA utilizará as variáveis selecionadas por SPA e GA e buscará encontrar uma função linear que melhor discrimine as amostras analisadas em suas respectivas classes. Sendo assim, é possível construir os modelos SPA-LDA e GA-LDA no trabalho de classificação entre espectros obtidos de diversas amostras analisadas. (SANTOS et al., 2017).

Tendo em vista que os mosquitos fêmeas são os vetores que se alimentam de sangue e os transmissores de doenças, a presença elevada destes mosquitos representa um perigo para a população humana uma vez que representa uma elevada probabilidade de surto de arbovírus. (ARAÚJO, et al., 2018) Sendo assim, faz-se necessários meios capazes de detectar a presença destes mosquitos fêmeas acima dos níveis normais a fim de nortear uma melhor tomada de decisões proativa do ponto de vista do controle de surtos. Nesse sentido, a busca por ferramentas que auxiliem no monitoramento se torna relevante e as técnicas de espectroscopia baseada no

infravermelho têm se mostrado efetivas em análises envolvendo mosquitos. (LAMBERT et al., 2018; SANTOS et al., 2022; SIKULU-LORD et al., 2016)

Assim sendo, esse trabalho visa a construção de modelos quimiométricos baseados na espectroscopia no infravermelho próximo (NIR) junto com técnicas de análise multivariada SPA-LDA e GA-LDA para classificar, de forma efetiva, os mosquitos *Aedes aegypti* com base no seu sexo.

Material e Métodos

Foram coletados espectros de 66 mosquitos *Aedes aegypti*, sendo 33 machos e 33 fêmeas. A coleta de dados deu-se por meio da espectroscopia no infravermelho próximo (NIR) em um espectrofotômetro ARCoOptix FT-NIR Rocker (Arcoptix S.A., Suíça). As leituras foram feitas no modo de transmitância com resolução de 5 nm. Utilizou-se uma folha A4 branca de fundo e os espectros foram coletados na região abdominal dos mosquitos, sendo realizadas triplicatas na faixa de 900-2600nm. A cada nova leitura realizou-se a limpeza da sonda com álcool etílico (70% v/v), sendo feita a coleta de um novo branco e um novo papel para que não houvesse resquícios da análise anterior.

Para a importação, pré-tratamento e criação dos modelos quimiométricos foi utilizado o software MATLAB R2014b. O pré-processamento se deu com um corte entre 1200-2300 nm, EMSC (*Extended Scatter Correction*), Baseline (*Automatic Whittaker Filter*) e Savitzky-Golay *Smoothing*. Após o pré-processamento das amostras as mesmas foram divididas em dois subconjuntos: um de treinamento ($\approx 70\%$) e um de teste ($\approx 30\%$), aplicando o algoritmo de amostragem Kennard-Stone (KS). (KENNARD; STONE, 1969).

Para a criação dos modelos optou-se pela utilização de dois algoritmos de seleção de variáveis: para um dos modelos foi utilizado o Algoritmo de Projeção Sucessiva (SPA) cujo critério de seleção é eliminação da multicolinearidade dos dados; e para o outro modelo a seleção das variáveis se deu pelo Algoritmo Genético (GA), que seleciona variáveis mais relevantes imitando a teoria da evolução de Darwin, criando uma população com cromossomos (variáveis) diferentes e selecionando os mais promissores na previsão e na precisão. (SANTOS et al., 2018,2022;).

Por fim aplicou-se a Análise Discriminante Linear no conjunto de variáveis selecionadas por SPA e GA para reconhecimento dos padrões e assim classificar as amostras. O LDA é um dos algoritmos de classificação supervisionada mais utilizados, pois o mesmo calcula uma função capaz de discriminar classes com base em seus dados de treinamento. Neste processo de pontuação para classificar as amostras, o LDA considera a aproximação de que todas as classes possuem a mesma matriz de covariâncias, e que as probabilidades de distribuição de todas as classes são semelhantes. (WU et al., 1996)

Por último os resultados dos modelos foram aplicados nas equações (1) e (2) para calcular o desempenho e validação dos mesmos, sendo a sensibilidade (Sens) a confiança de um mosquito da classe estudada ser classificado corretamente, e a especificidade (Spec) a confiança que um mosquito da outra classe ser classificado corretamente.

$$\text{Sens (\%)} = \frac{VP}{VP+FN} \times 100 \quad (1)$$

$$\text{Spec (\%)} = \frac{VN}{VN+FP} \times 100 \quad (2)$$

Onde: VP = Verdadeiro positivo; FN = Falso negativo; VN = Verdadeiro negativo; FP = Falso positivo.

Resultados e Discussão

Após a realização do pré-tratamento dos espectros obteve-se os espectros médios em NIR para fêmeas e machos (fig.1), onde é possível perceber uma similaridade enorme, sendo extremamente difícil a classificação visual, sem o auxílio dos modelos quimiométricos.

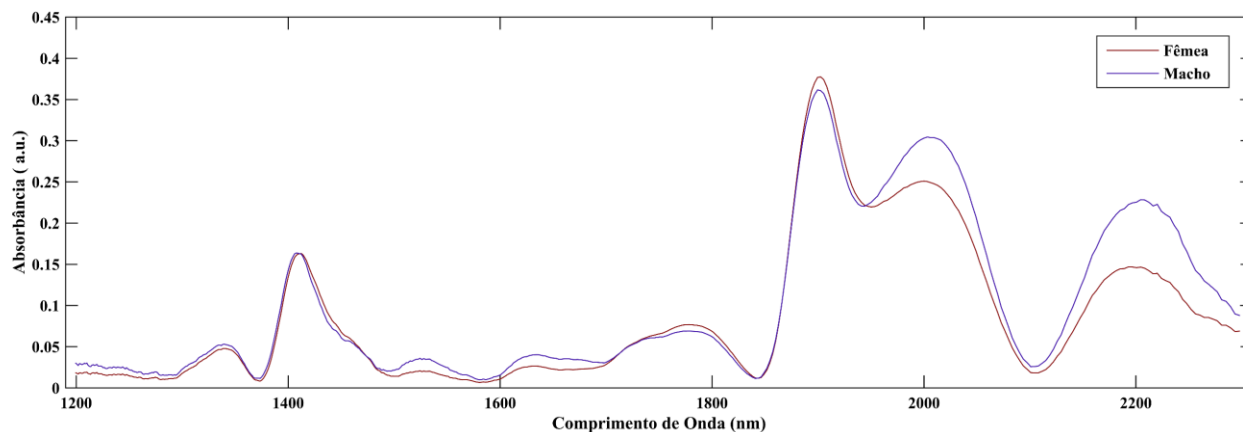


Fig.1: Espectros médios NIR dos mosquitos *Aedes aegypti* fêmea e macho após o pré-tratamento.

Modelo NIR-SPA-LDA

Após o treinamento do modelo o SPA selecionou um total de 5 variáveis significativas, apresentadas na tabela 1. Na figura 2 pode ser visto um espectro médio com as variáveis selecionadas por SPA identificadas.

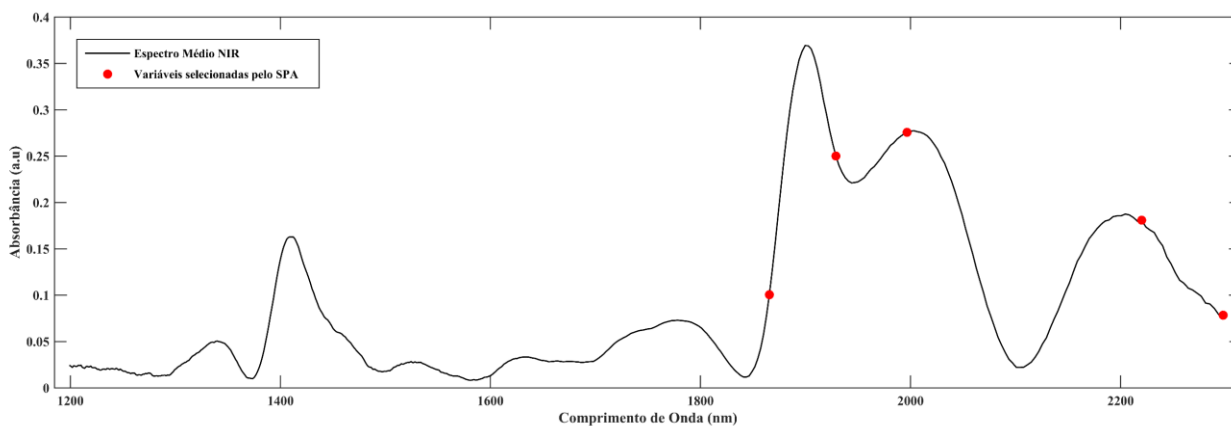


Fig.2: Variáveis selecionadas pelo SPA plotadas no espectro médio NIR.

Tabela 1: Variáveis selecionadas pelo SPA para a diferenciação de mosquitos *Aedes aegypt* fêmeas e machos.

| Modelo Quimiométrico | Variáveis Selecionadas | | | | |
|----------------------|------------------------|--------|--------|--------|--------|
| NIR-SPA-LDA | 1866nm | 1929nm | 1997nm | 2220nm | 2298nm |

Entre as variáveis selecionadas por SPA podemos destacar 2220 nm, que está associado a alcenos (alongamento e curvatura de CH_2), combinação de alongamentos em ligações N-H C-H; e 2300 nm, associado ao metileno (alongamento e curvatura de CH) (Türker-Kaya e Huck, 2017; Workman 1996).

Em seguida o LDA foi aplicado para a construção de um modelo supervisionado para a classificação dos mosquitos. A função discriminante gerada para o modelo NIR-SPA-LDA pode ser visualizada na figura 3, após a classificação calculou-se os valores de sensibilidade e especificidade para o mesmo obtendo assim os resultados apresentados na tabela 2.

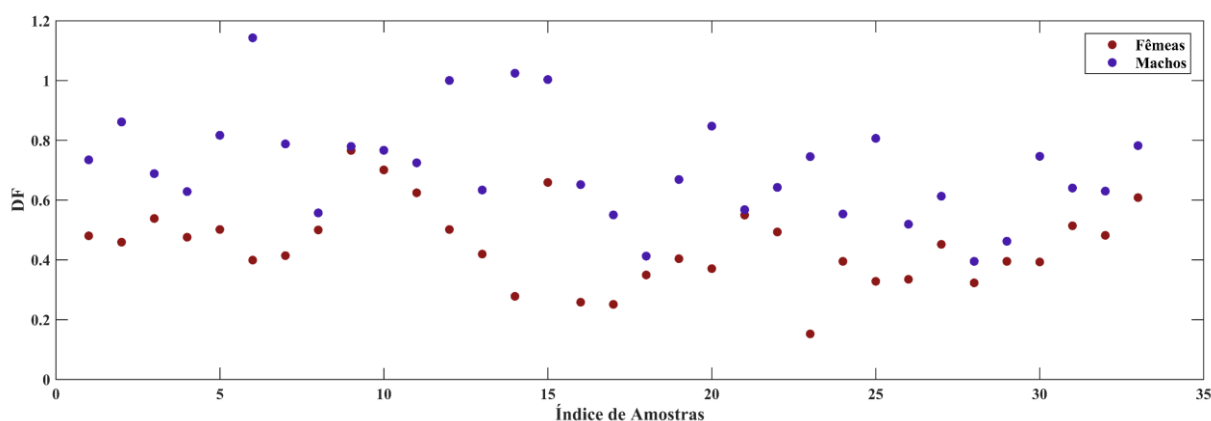


Fig.3: Modelo supervisionado gerado pelo LDA para o NIR-SPA-LDA.

Tabela 2: Valores obtidos para sensibilidade e especificidade do modelo NIR-SPA-LDA

| Modelo Quimiométrico | Valores para ambas as classes |
|----------------------|------------------------------------|
| NIR-SPA-LDA | Sens (%): 80,30 Spec (%): 80,30 |

Modelo NIR-GA-LDA

O GA após o treinamento selecionou um total de 16 variáveis significativas para a classificação das amostras, sendo apresentadas na tabela 3. Na figura 4 podem ser vistas as variáveis selecionadas identificadas no espectro médio (fig.4).

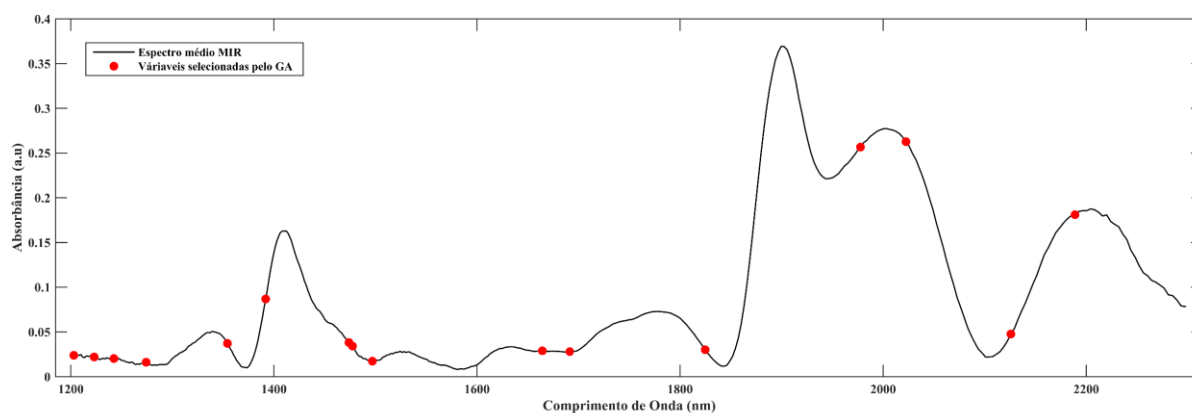


Fig.4: Variáveis selecionadas pelo GA plotadas no espectro médio NIR.

Tabela 3: Variáveis selecionadas pelo GA para a diferenciação de mosquitos *Aedes aegypti* fêmeas e machos.

| Modelo Quimiométrico | Variáveis Selecionadas | | | | | |
|----------------------|------------------------|--------|--------|--------|--------|--------|
| NIR-GA-LDA | 1203nm | 1223nm | 1243nm | 1274nm | 1354nm | 1392nm |
| | 1474nm | 1478nm | 1497nm | 1664nm | 1691nm | 1825nm |
| | 1978nm | 2023nm | 2126nm | 2189nm | | |

Entre as variáveis selecionadas por GA pode-se destacar 1223 nm que está associado com o 2º overtone de alongamento C-H; 1478 nm associado ao 1º overtone de alongamento O-H; e 1691 nm associado com o 1º overtone de alongamento C-H em grupos CH₃.

Em seguida, o LDA foi aplicado para a construção do modelo supervisionado para classificar os mosquitos. A função discriminante gerada para o modelo NIR-GA-LDA é apresentada na figura 5. Após a classificação, foram calculados os valores de sensibilidade e especificidade para o modelo, obtendo-se, assim, os resultados apresentados na tabela 4.

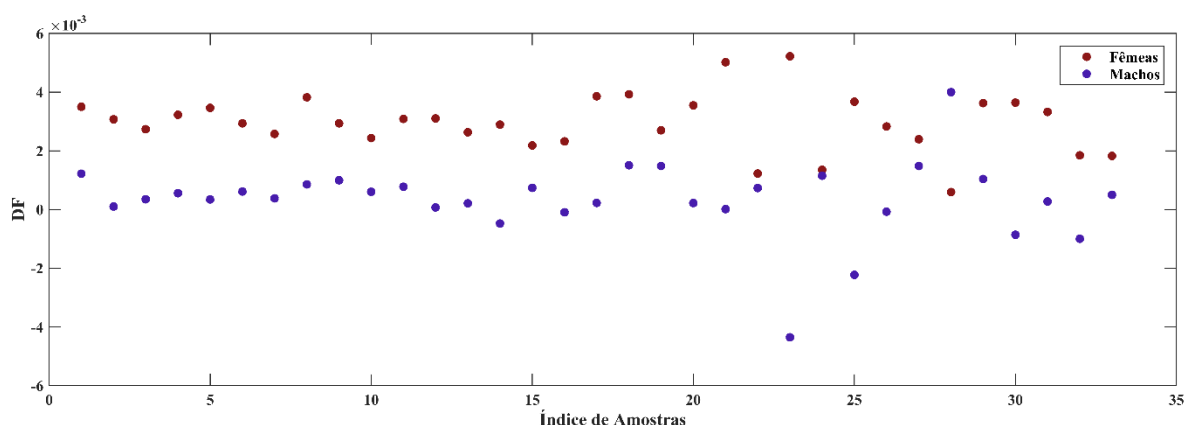


Fig.3: Modelo supervisionado gerado pelo LDA para o NIR-GA-LDA.

Tabela 4: Valores obtidos para sensibilidade e especificidade do modelo NIR-GA-LDA

| Modelo Quimiométrico | Valores para ambas as classes |
|----------------------|------------------------------------|
| NIR-GA-LDA | Sens (%): 93,94 Spec (%): 93,94 |

Um estudo recente descobriu o gene responsável por iniciar o desenvolvimento masculino nas espécies *Aedes Aegypti*. A este gene foi dado o nome de Nix. O cDNA de Nix é uma região de 985 pares de bases que codifica um polipeptídeo de 288 aminoácidos. Este cDNA está presente no DNA genômico masculino, mas ausente no feminino. Sua expressão está associada ao período anterior à determinação do sexo e, por fim, o Nix está localizado dentro do locus M (McDonald PT, Rai KS, 1970; Hall, et al., 2015). Adicionalmente ao Nix, na via de diferenciação sexual, dois outros genes foram caracterizados: *Fruitless* (*fru*) e *Doublesex* (*dsx*). Na presença de Nix, os genes *dsx* e *fru* produzem a transcrição específica do sexo masculino através de emendas alternativas e, na sua ausência, produz a transcrição feminina (HOANG, et al., 2016).

Tendo em vista os excelentes resultados observados nas figuras de mérito principalmente para a combinação NIR-GA-LDA, espera-se, em futuros estudos, associar algumas das variáveis espectrais selecionadas com a presença/ausência do Nix. Isto representaria um grande avanço no campo da vigilância entomológica destas espécies, uma vez que poderiam ser construídas armadilhas portáteis equipados com NIR capazes de detectar o sexo dos mosquitos presentes em determinada região, possibilitando a tomada de ações preventivas para eliminar a possibilidade de eventuais surtos de arboviroses em áreas de risco antes que estes surtos venham a acontecer.

Conclusões

Observando os valores obtidos de sensibilidade e especificidade, é possível identificar que o modelo NIR-GA-LDA possui potencial para ser utilizado em sistemas de previsão do crescimento de colônias de *Aedes aegypti* em conjunto com outras análises estatísticas para monitorar possíveis focos e locais de endemias. Mesmo com a possibilidade da criação de outros modelos em conjunto com outros algoritmos que possam vir a melhorar esse desempenho, acredita-se que o objetivo foi atingido ao desenvolver um modelo com uma confiabilidade aceitável capaz de classificar os mosquitos de acordo com seu sexo.

Os valores de figuras de mérito em 93,94% demonstra a efetividade do uso da espectroscopia no infravermelho próximo (NIR) como ferramenta para esse monitoramento, podendo ser implementado através de armadilhas equipadas com espectrofotômetros portáteis, ou pela implementação de um centro de análises para esse controle. Futuros estudos devem ser desenvolvidos a fim de relacionar o gene Nix com as variáveis selecionadas por SPA ou GA.

Agradecimentos

Ao IFSertãoPE pelo apoio financeiro e aos pesquisadores do Laboratório de Insetos e Vetores e do Laboratório de Química Analítica da UFRN pela aquisição das amostras e dos espectros.

Referências

- ARAÚJO, H. R. C.; KOJIN, B. B.; CAPURRO, M. L. Sex determination and Aedes population control. **Parasites & Vectors**, v. 11, n. 2, p. 644, 24 dez. 2018.
- HALL, A. B. et al. A male-determining factor in the mosquito Aedes aegypti. **Science (New York, N.Y.)**, v. 348, n. 6240, p. 1268–1270, 12 jun. 2015.
- HOANG, K. P. et al. Mechanisms of sex determination and transmission ratio distortion in Aedes aegypti. **Parasites & Vectors**, v. 9, n. 1, p. 49, 28 jan. 2016.
- KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137–148, fev. 1969.
- LAMBERT, B. et al. Monitoring the Age of Mosquito Populations Using Near-Infrared Spectroscopy. **Scientific Reports**, v. 8, n. 1, p. 5274, dez. 2018.
- MCDONALD, P. T.; RAI, K. S. Correlation of linkage groups with chromosomes in the mosquito, Aedes aegypti. **Genetics**, v. 66, n. 3, p. 475–485, nov. 1970.
- PARASKEVAIDI, M. et al. Clinical applications of infrared and Raman spectroscopy in the fields of cancer and infectious diseases. **Applied Spectroscopy Reviews**, v. 56, n. 8–10, p. 804–868, 26 nov. 2021.
- SANTOS, M. C. D. et al. ATR-FTIR spectroscopy with chemometric algorithms of multivariate classification in the discrimination between healthy vs. dengue vs. chikungunya vs. zika clinical samples. **Analytical Methods**, v. 10, n. 10, p. 1280–1285, 8 mar. 2018.
- SANTOS, M. C. D. et al. Infrared spectroscopy (NIRS and ATR-FTIR) together with multivariate classification for non-destructive differentiation between female mosquitoes of Aedes aegypti recently infected with dengue vs. uninfected females. **Acta Tropica**, v. 235, p. 106633, nov. 2022.
- SANTOS, M. C. D. et al. Spectroscopy with computational analysis in virological studies: A decade (2006-2016). **Trends in analytical chemistry: TRAC**, v. 97, p. 244–256, dez. 2017.
- SIKULU-LORD, M. T. et al. Near-Infrared Spectroscopy, a Rapid Method for Predicting the Age of Male and Female Wild-Type and Wolbachia Infected Aedes aegypti. **PLOS Neglected Tropical Diseases**, v. 10, n. 10, p. e0005040, 21 out. 2016.
- TÜRKER-KAYA, S.; HUCK, C. A Review of Mid-Infrared and Near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. **Molecules**, v. 22, n. 1, p. 168, 20 jan. 2017.
- WORKMAN, J. J. Interpretive Spectroscopy for Near Infrared. **Applied Spectroscopy Reviews**, v. 31, n. 3, p. 251–320, ago. 1996.
- WU, W. et al. Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. **Analytica Chimica Acta**, v. 329, n. 3, p. 257–265, ago. 1996.