A New Strategy for Canine Visceral Leishmaniasis Diagnosis Based on Molecular Spectroscopy and Machine Learning

Gustavo Larios

UFMS – Universidade Federal de Mato Grosso do Sul

Matheus Ribeiro

UFMS – Universidade Federal de Mato Grosso do Sul

Carla Arruda

UFMS – Universidade Federal de Mato Grosso do Sul

Samuel L. Oliveira

UFMS – Universidade Federal de Mato Grosso do Sul

Thalita Canassa

UFMS – Universidade Federal de Mato Grosso do Sul

Matthew J. Baker

University of Strathclyde, Technology and Innovation Centre

Bruno Marangoni

UFMS – Universidade Federal de Mato Grosso do Sul

Carlos Ramos

UFMS – Universidade Federal de Mato Grosso do Sul

Cícero Cena (✉ cicero.cena@ufms.br)

UFMS – Universidade Federal de Mato Grosso do Sul

Research Article

Keywords: Visceral leishmaniasis, biofluids, diagnosis, FTIR spectroscopy, machine learning.

Posted Date: April 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-382827/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Visceral leishmaniasis is a neglected disease caused by protozoan parasites of the genus *Leishmania*. The successful control of the disease depends on its accurate and early diagnosis, which is usually made by combining clinical symptoms with laboratory tests such as serological, parasitological, and molecular tests. However, early diagnosis based on serological tests may exhibit low accuracy due to lack of specificity caused by cross-reactivities with other pathogens, and sensitivity issues related, among other reasons, to disease stage, leading to misdiagnosis. In this work was investigated the use of mid-infrared spectroscopy and multivariate analysis to perform a fast, accurate and easy canine visceral leishmaniasis diagnosis. Canine blood sera of non-infected, *Leishmania infantum*, and *Trypanosoma evansi* infected groups were studied. The data demonstrate that principal component analysis with machine learning algorithms achieved an overall accuracy above 85% in the diagnosis.

Introduction

Leishmaniasis is a group of neglected infectious diseases, affecting over 100 countries worldwide (1), caused by a protozoan parasite from over 20 *Leishmania* species. Three leading clinical forms of the disease affect humans: visceral leishmaniasis (VL), cutaneous leishmaniasis, and mucocutaneous leishmaniasis. In Brazil, VL is caused by *Leishmania infantum*, and its fatal if untreated in over 95% of cases. The successful treatment depends on its accurate, and early diagnosis (2).

There is a high occurrence of canine visceral leishmaniasis (CVL) in Brazil, and dogs become the main reservoir of the protozoan. The human infection is mainly caused by the bites of infected female phlebotomine sandflies, responsible for parasite transmission from dogs to humans (3). In this scenario, the euthanasia of infected dogs has been adopted as one of the main strategies for disease control. Therefore, more accurate VL testing in dogs became essential to prevent and control the disease and avoid the controversial culling of dogs with false-positive diagnosis delivered by the available diagnostic tests.

Tests recommended by the World Health Organization exhibit some drawbacks. Based on the correct identification of the parasite under a microscope, the parasitological test shows high accuracy. However, it depends on the capture of the parasite by aspiration of skin lesions, bone marrow, and lymph nodes, which is invasive and presents low diagnostic sensitivity (4). Enzyme-Linked Immune Sorbent Assay (ELISA), indirect fluorescent antibody (IFA), and immunochromatographic assay are the most common serological tests. They show high sensitivity but low specificity, often exhibiting cross-reactivity with other pathogens, such as the *Trypanosoma* genus (5–8). Molecular diagnosis by polymerase chain reaction (PCR), which replicates specific DNA fragments, is highly accurate, but it is a time-consuming and expensive test, hindering its use on a large scale (4, 9, 10).

Most alternative strategies to leishmaniasis diagnosis have been based on the development of biosensors. Impedance-based biosensors depend on the immobilization of a DNA fragment on an

electrode; among them, NiO nanostructured film was used for *L. donovani* diagnosis (11), and Au nanoleaves were employed in a highly sensitive sensor for *L. major* (12). In turn, *L. infantum* diagnosis was carried out using either Au electrodes (13) or Au nanoparticles dispersed in a polyaniline matrix (14). Gold nanoparticle-based colorimetric biosensors from kDNA were developed to diagnose and differentiate *L. major*, *L. tropica*, and *L. infantum* (15). Electrical capacitance measurements of interdigitated electrodes with immobilized proteoliposomes detected specific anti-*L. amazonensis* antibodies and distinguished them from *Trypanosoma cruzi* (16). Biosensors can provide accurate results to different species of *Leishmania*. Nevertheless, the need for DNA manipulation, electrical probes, chemicals, and, in some cases, controlled storage conditions difficult in its implementation as standard and widely deployed diagnostic tests (17). New alternatives to overcome these limitations have emerged from multivariate analysis of molecular spectra of biofluids, skin tissue, and cell samples to disease diagnosis.

Fourier transform infrared (FTIR) and Raman spectroscopy have shown great ability to enable the detection of infections diseases. They have successfully discriminated *Candida albicans* isolates, allowing candidiasis diagnosis (18). Discriminated strains of *Staphylococcus aureus* from other pathogens (19), and also identified bacteria related to dairy cow diseases (*Escherichia coli*, *Trueperella pyogenes* and *Streptococcus* spp (20, 21)) was possible by using FTIR. Serological screening of echinococcosis, a zoonotic parasitic disease, was carried out using Raman spectroscopy (22). Cerebrospinal fluid investigated by micro-Raman spectroscopy allowed the meningitis diagnosis (23). Malaria diagnosis by FTIR spectroscopy was developed from infected blood spectra (24). Other virus related diseases could be diagnosed from blood serum and spectroscopic analysis, such as dengue by Raman spectroscopy (25), and hepatitis B and C by FTIR (26). Papilloma virus diagnosis was performed from Raman spectra of cervical secretion (27), and FTIR spectra of cervical exfoliated cells (28).

Studies based on molecular spectroscopy diagnosis related to Trypanosomatidae family are limited to an investigation of *L. chagasi* (sin.=*L. infantum*), *T. cruzi* and *T. rangeli* strain discrimination (29), and Phlebotomine sandflies specimen's classification based on FTIR spectroscopy (30). This is the first report about canine blood serum evaluated by FTIR spectroscopy and machine learning algorithms for the *Leishmania* infection diagnosis. Three groups of canine blood serum were investigated, non-infected, *L. infantum*, and *T. evansi* infected animals. Different spectral range, and group combination were evaluated to provide the higher overall accuracy to the correct diagnosis. These results show the ability of molecular spectroscopy to provide a sensitive and specific analysis that is rapid and cost effective making it suitable for wide scale translation to enable effective canine disease control.

Methods

Blood serum collection

The samples were obtained from the "Laboratório de Biologia Molecular, FAMEZ (Faculdade Veterinária e Zootecnia)" biobank from "Universidade Federal do Mato Grosso do Sul (UFMS)", Brazil. The samples

were collected with approval of the UFMS Ethical Committee of Research in Animals (CEUA 765/2016). All methods were carried out in accordance with relevant guidelines and regulations.

Canine blood serum samples were collected from 48 dogs, and distributed into three groups: (i) 20 animals naturally infected by *L. infantum*; (ii) 8 animals naturally infected by *T. evansi*; and (iii) 20 non-infected animals (control group). The *L. infantum* infected animals exhibited clinical symptoms, such as onychogryphosis, lymphadenopathy, and progressive weight loss, etc. ELISA and immunochromatographic assays were employed as validation tests. The control group showed no clinical symptoms and negative results in the validation tests. The *T. evansi* infected group was confirmed by PCR and DNA sequencing, as described elsewhere (31). Serological tests excluded *Leishmania* infection in this group.

Sample preparation

The canine blood serum samples were removed from storage at -20 °C, and allowed to thaw at room temperature (27 °C) for 15 min. After homogenization, three deposition steps were performed by depositing 30 µL of serum on silicon oxide (SiO₂) substrate, followed by drying at 40 °C for 30 min. Each sample was analysed in triplicate, and a total of 144 samples were measured (120 from the *L. infantum* and non-infected animals, and 24 from the *T. evansi* infected animals).

FTIR spectra acquisition and data analysis

The mid-infrared spectra were acquired in a Fourier transform infrared spectrophotometer (Perkin Elmer, Spectrum 100) with an attenuated total reflectance accessory. The spectra were collected from 4000 to 600 cm⁻¹ with a resolution of 4 cm⁻¹ and 5 scans. The FTIR spectrum of each sample was obtained by averaging the spectra acquired in three different spots, from the center to the substrate's border, to avoid any inhomogeneity in the sample composition due to the drying process (32).

The average FTIR spectra of canine blood sera were investigated using MATLAB R2015b software after standard normal variate (SNV) pretreatment. Initially, the data were dimensionally reduced by principal component analysis (PCA) to evaluate the potential separation of the samples. The PCA analysis reduces the original amount of data into a new set of uncorrelated variables (principal components, PCs) by a linear orthogonal transformation of this data, revealing the data variance. The component scores (PCs plot) reveal the data separation enabling a predictive model for sample classification. Additionally, it is possible to identify the main spectral data variance by analyzing the weight associated with the linear transformation used to obtain the component score (loadings)(33).

Machine learning (ML) algorithms based on k-nearest neighbors (KNN), support vector machine (SVM), and discriminant analysis (DA) methods were performed by using the PCA data with the aid of MATLAB R2018b software. The number of PCs used in ML training was optimized to avoid overfitting. Briefly, the DA algorithm uses the training data set to determine the contour between the classes distributions, classifying the test sample according to this criterion. The SVM organizes the data in a spatial

distribution and groups it in categories through the optimization of a hyperplane between classes. SVM did not not assume a normal distribution, differing from DA in this point. The hyperplane can be linear or non linear, being adjusted to reach the best performance. The KNN is also based on the spatial distribution, the classification is based on the label of the k nearest neighbors. The number of closest neighbors for classification can be changed: k = 1 (Fine KNN) and k = 10 (Medium KNN). In this case, Euclidean distance is considered. In order to weight the distance between samples, the weighed KNN classifier can be applied using some specific function(34).

These methods were compared, and the accuracy results on the validation set were summarized accordingly. The validation process employed was the leave one out cross-validation (LOO-CV). In this case, a sample was withdrawn from the training set and used to test the predictive modeling power. The process is repeated, alternating the withdrawn sample to test all sample sets. The sample classification was performed adopting the proper ML algorithms from the validation process so that each sample was tested and assigned to its corresponding group. The samples were correctly classified when the predicted class, from ML classification, corresponded to the true class identified from the standard method.

Results And Discussion

The average FTIR spectrum for non-infected, *L. infantum*, and *T. evansi* infected canine blood serum are shown in Fig. 1, exhibiting remarkable similarity for all investigated groups. The vibrational bands can be assigned to proteins, carbohydrates, lipids, and fatty acid molecules. The vibrational assignments and molecular groups in the canine blood serum can be exhibited in Table 1.

The broad band centered around 3276 cm^{-1} is assigned to the overlap of molecular vibrational modes of protein and water molecules. The variation of water content in biological samples is common, and significant variation between the samples can contribute to a wrong discrimination between the groups. The drying step, and the same sample preparation conditions for all samples are important to not cause interference in our data discrimination due to water content.

The absorption in the $1700 - 1450\text{ cm}^{-1}$ spectral range is attributed to proteins molecules (Table 1). The main antibodies produced in response to Leishmania antigens belong to this group of molecules. Although, very similar FTIR spectra were obtained, including for the non-infected group, probably due to the overlap, for example, of amide I and II absorption bands in proteins. This spectral range is very information rich and useful for sample discrimination as previously shown by other researchers(26, 35, 36). In addition, other molecules that are assayed by FTIR (e.g. carbohydrates, lipids, and fatty acids) can be equally important for sample classification; the immune response of the organism during infection may impact these molecules, and consequently, cause subtle spectral alterations.

Table 1

Vibrational assignments and related molecular groups for canine blood serum FTIR spectra (37–46). The symbols meaning: ν = symetric and/or assimetric stretching; σ = scissoring bending; and δ = bending vibrational modes.

band (cm ⁻¹)	Vibrational Assignments	Organic Group	Biomolecule
929	$\nu(\text{C-O})$ $\nu(\text{C-O-C})$	Saccharides; carbohydrates	
1033		Saccharides; DNA; carbohydrates	Glucose; α_2 -macroglobulin
1073	$\nu_s(\text{PO}_2^-)$	Saccharides; nucleic acids; RNA/DNA; carbohydrates	Lactate; α_2 -macroglobulin
1164		Saccharides; carbohydrates	
1236	$\nu_{as}(\text{PO}_2^-)$	Saccharides; Nucleic acids; phosphate diester; RNA/DNA; carbohydrates	Serine; tyrosine; threonine
1314 1343	$\nu(\text{CH}_2)$	Collagen or asymmetric phosphate (Amide III)	Transferrin; α_1 -Acid Glycoprotein
1397	$\nu_s(\text{COO}^-)$ $\delta_s(\text{CH}_3)$ $\nu_s(\text{CH}_3)$	Amino acids; proteins	Fibrinogen; IgG ₁ ; IgM; IgA; Haptoglobin
1451	$\nu(\text{C-O-O})$ $\nu_{as}(\text{CH}_3)$ $\sigma(\text{CH}_2)$	Amino acids; lipids	Apolipoprotein-A ₁ ;
1515	$\delta(\text{N-H})$	α -helix of proteins (Amide II)	Albumin; IgG ₄
1536	$\nu_s(\text{C-N})$ $\sigma(\text{N-H})$		
1633	$\nu(\text{C=O})$ $\delta(\text{N-H})$ $\nu_s(\text{C-N})$	β -sheet and a helix of proteins (Amide I)	IgG ₂ ; IgG ₃
1738	$\nu(\text{C=O})$ Amide I $\nu(\text{C-N})$ $\delta(\text{H-N})$ Amide II	Lipids; phospholipids; cholesterol; esters; Fatty acids	
2854	$\nu_s(\text{CH}_2)$	Lipids	Apolipoprotein-B
2872	$\nu(\text{CH}_2)$	Fatty acids; esters; glycerol Phospholipids; triglycerides	
2928	$\nu_{as}(\text{CH}_2)$		
2960	$\nu_{as}(\text{CH}_3)$	Cholesterol esters; lipids; fatty acids	

band (cm^{-1})	Vibrational Assignments	Organic Group	Biomolecule
3067	$\nu(\text{CH}_2)$	Lipids; unsaturated lipids; protein (Amide II)	
3276	$\nu(\text{N-H})$ $\nu(\text{O-H})$	Water; and Proteins (Amide)	
3412			

Figure 2 shows the score plot and loadings for the $1800 - 800 \text{ cm}^{-1}$ range, which excludes the O-H/N-H overlapped bands, and lipid/amide II contributions for sample classification. Data selection is a rational alternative to improve the clustering in PCA analysis (30), when highly correlated data are used. Detailed results related to other spectral ranges can be found in the Supplementary Material (Figures S1-S4).

The score plot exhibited in Fig. 2(a), shows a broad data distribution with no cluster separation for non-infected, *L. infantum*, and *T. evansi* infected canine blood serum. The loadings, Fig. 2(b), indicate that the first three PCs are responsible for over 91% of the data variance, mainly in the spectral range assigned to proteins around 1600 cm^{-1} , including vibrational modes of immunoglobulins. However, the group separation was not effective by PCA because the infection did not induce measurable spectral differences between the groups with basis on the observed vibrational bands .

The PCA analysis was applied to 1x1 group classification to find a better strategy to sample identification. Figure 2(c) shows the score plot for non-infected versus *L. infantum* infected canine blood serum with no apparent clustering. The three first PCs represent 92.34% of data variation. The loading data exhibited in Fig. 2(d) is similar to the one shown in Fig. 2(b), with prominent peaks around 1600 cm^{-1} . Analogous behavior was observed for the score plot and loadings for non-infected versus *T. evansi* (Figs. 2(e) and 2(f)) and *L. infantum* versus *T. evansi* infected canine blood (Figs. 2(g) and 2(h)). Therefore, 1x1 group classification did not enable cluster formation and separation, probably due to the high data correlation.

Machine learning algorithms that rely on the PCA data were then employed to maximize sample classification accuracy, testing different spectral ranges and group comparisons. The results are described in the Supplementary Material (Figures S5 and S6).

Figure 3 summarizes the overall accuracy obtained from ML results in the LOO-CV, for different group combination at 1800 until 800 cm^{-1} range. The classification of non-infected, *L. infantum*, and *T. evansi* infected samples used only the first 5 PCs, responsible for 96.88% of data variance, to achieved an overall accuracy of 85.42% by using the linear KNN. Similar result was obtained for non-infected versus *L. infantum* infected group, with overall accuracy of 85% by using the first 4 PCs, responsible for 95.67% of data variance, in the quadractic SVM. However, the best accuracy achieved for non-infected versus *L.*

infantum group classification, was 87.5% in the 1700 until 1450 cm^{-1} range, by using the first 10 PCs with Quadratic Discriminant Analysis (Figure S6).

In turn, the classification of *L. infantum* and *T. evansi* infected samples yielded an overall accuracy of 100% using either linear or cubic SVM with 10 PCs (Fig. 3). Identical accuracy was achieved in the classification of non-infected and *T. evansi* infected sera by applying linear SVM with 6 PCs. The weighted KNN algorithm successfully classified only the non-infected and *T. evansi* infected samples, while linear SVM classified both groups. Cubic and quadratic SVM with 10 PCs also provided a 100% overall accuracy to classify *L. infantum* and *T. evansi* infected sera.

Figure 4 shows the confusion matrix arising from the leave-one-out cross-validation tests (LOO-CV) for the group classification strategies with higher overall accuracy achieved in the 1800 – 800 and 1700 – 1450 cm^{-1} ranges. The higher overall accuracy (85.42%) in the classification of non-infected, *L. infantum*, and *T. evansi* infected samples was achieved adopting fine KNN with 5 PCs for the 1800 – 800 cm^{-1} range, correctly classifying all samples infected by *T. evansi* (100%), 17 samples infected by *L. infantum* (85%), and 16 non-infected samples (80%) (Fig. 4(a)). Applying quadratic DA with 10 PCs for the data set from 1700 to 1450 cm^{-1} , the overall accuracy classification of 87.42% was determined for the non-infected and *L. infantum* classification. The number of samples correctly classified was the same for *L. infantum*, and only 2 samples (10%) were incorrectly classified in the non-infected group (Fig. 4(b)). The 100% overall accuracy was reached in the classification of both *L. infantum* and *T. evansi* infected sera employing linear or cubic SVM with 10 PCs (Fig. 4(c)), and non-infected and *T. evansi* utilizing linear SVM or weighted KNN with 6 PCs in the 1800 – 800 cm^{-1} range (Fig. 4(d)).

ML algorithms allied to PCA were able to distinguish *T. evansi* infection successfully in the group comparisons and provide good classification accuracy of *L. infantum* infection when compared with standard methods (4, 9, 11). Additionally, methods based on these strategies could present advantages such as easy sample manipulation, low cost, and fast diagnosis. There is no need for sample preprocessing and any input, and the collection and analysis of the spectra are performed in a few minutes. For comparison, serological tests currently available for CVL diagnosis require prior processing of biological samples and inputs such as antibodies, enzymes, and buffers. ELISA or IFA tests take around 60 to 180 min to deliver the diagnosis.

The false-positive classification for non-infected versus *L. infantum* versus *T. evansi* infected samples (Fig. 4(a)) and non-infected versus *L. infantum* infected samples (Fig. 4(b)) is not a significant issue since the present proposal may be used as a trial test. Positive tested animals would be tested again by high accurate methods before the specialist takes any decision. On the other hand, the false-negative classification was kept constant for three or two group tests, requiring, for example, new methods of sample preparation or measurement to improve the overall accuracy.

Additionally, plasma protein kinetics is used as a biomarker for diagnosis and clinical follow-up of many infectious diseases (47, 48). Acute-phase proteins (APPs) such as haptoglobin (Hp), serum amyloid A

(SAA), and C-reactive protein (CRP) have been investigated in dogs infected by *L. infantum* (49). Compared with the increase in specific antibodies (IgG), plasma proteins may change earlier due to infections, thus favoring an earlier disease diagnosis (49). In the present study, of the 20 animals in the control group (not infected), 4 (in the LI x N x TE comparison) and 2 (LI x N) were classified as LI (infected by *L. infantum*), which could be possible cases of recent infection, with an increase in APPs, but with no detectable IgG production. Thus, the animals could be infected but without clinical symptoms and with negative results in the serological tests. It is important to highlight that the animals used in this study did not undergo any selection, providing samples with high heterogeneity from animals with different sexes, ages, stages of infection, nutritional condition, and so on. Thus, the approach reported in this work presents a remarkable ability to classify animals infected by either *L. infantum* or *T. evansi*, taking into account the high inhomogeneity of the data closely related to the clinical reality.

Conclusions

FTIR measurements and machine learning algorithms together with PCA provided an alternative for canine visceral leishmaniasis diagnosis, discriminating non-infected and *L. infantum*, *T. evansi* infected animals with about 85% of classification accuracy. *L. infantum* from *T. evansi* infected animals with 100% accuracy. The high inhomogeneity of the samples that constituted each group proved that this approach is a robust test for canine visceral leishmaniasis and trypanosomiasis diagnosis.

Declarations

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. The authors are grateful to Prof. Alem-Mar Gonçalves (UFMS), PhD Ingrid Barcellos (LNLS), and MSc. Silvano Naves (UFMS) for their assistance throughout this work. Samuel Oliveira thanks the financial support provided by CNPq (309636/2017-5). The author Cicero Cena thanks Perovskita, his pet, for motivating this study and a new line of research in his career.

References

1. WHO. No Title [Internet]. Leishmaniasis - World Health Organization. 2020 [cited 2020 Jan 12]. p. 6. Available from: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>
2. Burza S, Croft SL, Boelaert M. Leishmaniasis. *Lancet*. 2018;392(10151):951–70.
3. Gramiccia M, Gradoni L. The current status of zoonotic leishmaniases and approaches to disease control. *Int J Parasitol*. 2005;35(11–12):1169–80.
4. Vega-López F. Diagnosis of cutaneous leishmaniasis. *Curr Opin Infect Dis*. 2003;16(2):97–101.
5. Sundar S, Reed SG, Singh VP, Kumar PCK, Murray HW. Rapid accurate field diagnosis of Indian visceral leishmaniasis. *Lancet*. 1998;351(9102):563–5.

6. Gomes-Silva A, Souza MA, Afonso-Cardoso SR, Andrade LR, Dietze R, Lemos E, et al. Serological reactivity of different antigenic preparations of *Leishmania (Leishmania) amazonensis* and the *Leishmania braziliensis* complex. *Rev Soc Bras Med Trop*. 2008;41(2):135–41.
7. Fargeas C, Hommel M, Maingon R, Dourado C, Monsigny M, Mayer R. Synthetic peptide-based enzyme-linked immunosorbent assay for serodiagnosis of visceral leishmaniasis. *J Clin Microbiol*. 1996;34(2):241–8.
8. Reithinger R, Dujardin JC. Molecular diagnosis of leishmaniasis: Current status and future applications. *J Clin Microbiol*. 2007;45(1):21–5.
9. Volpini ÂC, Passos VMA, Oliveira GC, Romanha AJ. PCR-RFLP to identify *Leishmania (Viannia) braziliensis* and *L. (Leishmania) amazonensis* causing American cutaneous leishmaniasis. *Acta Trop*. 2004;90(1):31–7.
10. Nasereddin A, Bensoussan-Hermano E, Schönián G, Baneth G, Jaffe CL. Molecular diagnosis of old world cutaneous leishmaniasis and species identification by use of a reverse line blot hybridization assay. *J Clin Microbiol*. 2008;46(9):2848–55.
11. Mohan S, Srivastava P, Maheshwari SN, Sundar S, Prakash R. Nano-structured nickel oxide based DNA biosensor for detection of visceral leishmaniasis (Kala-azar). *Analyst*. 2011;136(13):2845–51.
12. Moradi M, Sattarahmady N, Rahi A, Hatam GR, Sorkhabadi SMR, Heli H. A label-free, PCR-free and signal-on electrochemical DNA biosensor for *Leishmania major* based on gold nanoleaves. *Talanta* [Internet]. 2016;161:48–53. Available from: <http://dx.doi.org/10.1016/j.talanta.2016.08.030>
13. Cordeiro TAR, Gonçalves MVC, Franco DL, Reis AB, Martins HR, Ferreira LF. Label-free electrochemical impedance immunosensor based on modified screen-printed gold electrodes for the diagnosis of canine visceral leishmaniasis. *Talanta* [Internet]. 2019;195(November 2018):327–32. Available from: <https://doi.org/10.1016/j.talanta.2018.11.087>
14. Garcia MFK do. S, Andrade CAS, de Melo CP, Gomes DS, Silva LG, Dias R V., et al. Impedimetric sensor for *Leishmania infantum* genome based on gold nanoparticles dispersed in polyaniline matrix. *J Chem Technol Biotechnol*. 2016;91(11):2810–6.
15. Sattarahmady N, Movahedpour A, Heli H, Hatam GR. Gold nanoparticles-based biosensing of *Leishmania major* kDNA genome: Visual and spectrophotometric detections. *Sensors Actuators, B Chem* [Internet]. 2016;235:723–31. Available from: <http://dx.doi.org/10.1016/j.snb.2016.05.023>
16. Perinoto ÂC, Maki RM, Colhone MC, Santos FR, Migliaccio V, Daghashtanli KR, et al. Biosensors for efficient diagnosis of leishmaniasis: Innovations in bioanalytics for a neglected disease. *Anal Chem*. 2010;82(23):9763–8.
17. Souto DEP, Volpe J, Gonçalves C de C, Ramos CHI, Kubota LT. A brief review on the strategy of developing SPR-based biosensors for application to the diagnosis of neglected tropical diseases. *Talanta* [Internet]. 2019;205(July):120122. Available from: <https://doi.org/10.1016/j.talanta.2019.120122>
18. Potocki L, Depciuch J, Kuna E, Worek M, Lewinska A, Wnuk M. FTIR and raman spectroscopy-based biochemical profiling reflects genomic diversity of clinical candida isolates that may be useful for

- diagnosis and targeted therapy of candidiasis. *Int J Mol Sci*. 2019;20(4).
19. Guliev RR, Suntsova AY, Vostrikova TY, Shchegolikhin AN, Popov DA, Guseva MA, et al. Discrimination of *Staphylococcus aureus* Strains from Coagulase-Negative Staphylococci and Other Pathogens by Fourier Transform Infrared Spectroscopy. *Anal Chem*. 2020;92(7):4943–8.
 20. Jaureguiberry M, Madoz LV, Giuliadori MJ, Wagener K, Prunner I, Grunert T, et al. Identification of *Escherichia coli* and *Trueperella pyogenes* isolated from the uterus of dairy cows using routine bacteriological testing and Fourier transform infrared spectroscopy. *Acta Vet Scand*. 2016;58(1):1–6.
 21. Schabauer L, Wenning M, Huber I, Ehling-Schulz M. Novel physico-chemical diagnostic tools for high throughput identification of bovine mastitis associated gram-positive, catalase-negative cocci. *BMC Vet Res*. 2014;10(1):1–11.
 22. Yue X, Li H, Tang J, Liu J, Jiao J. Rapid and label-free screening of echinococcosis serum profiles through surface-enhanced Raman spectroscopy. *Anal Bioanal Chem*. 2020;412(2):279–88.
 23. Harz M, Kiehntopf M, Stöckel S, Rösch P, Straube E, Deuffer T, et al. Direct analysis of clinical relevant single bacterial cells from cerebrospinal fluid during bacterial meningitis by means of micro-Raman spectroscopy. *J Biophotonics*. 2009;2(1–2):70–80.
 24. Heraud P, Chatchawal P, Wongwattanakul M, Tippayawat P, Doerig C, Jearanaikoon P, et al. Infrared spectroscopy coupled to cloud-based data management as a tool to diagnose malaria: A pilot study in a malaria-endemic country. *Malar J* [Internet]. 2019;18(1):1–11. Available from: <https://doi.org/10.1186/s12936-019-2945-1>
 25. Rehman A, Anwar S, Firdous S, Ahmed M, Rasheed R, Nawaz M. Dengue blood analysis by Raman spectroscopy. *Laser Phys*. 2012;22(6):1085–9.
 26. Roy S, Perez-Guaita D, Bowden S, Heraud P, Wood BR. Spectroscopy goes viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy. *Clin Spectrosc* [Internet]. 2019;1(January):100001. Available from: <https://doi.org/10.1016/j.clispe.2020.100001>
 27. Chen C, Wang J, Chen C, Tang J, Lv X, Ma C. Rapid and efficient screening of human papillomavirus by Raman spectroscopy based on GA-SVM. *Optik (Stuttg)* [Internet]. 2020;210(January):164514. Available from: <https://doi.org/10.1016/j.ijleo.2020.164514>
 28. Mo H, Yang L, Wu G, Zheng X, Wang J, Yin L, et al. Rapid and non-invasive screening of high-risk human papillomavirus using Fourier transform infrared spectroscopy and multivariate analysis. *Optik (Stuttg)* [Internet]. 2020;206(January):164292. Available from: <https://doi.org/10.1016/j.ijleo.2020.164292>
 29. Aguiar JC, Mittmann J, Caetano PC, Raniero L. Using FT-IR spectroscopy for the identification of the *T. cruzi*, *T. rangeli*, and the *L. chagasi* species. *Exp Parasitol* [Internet]. 2018;192(July 2017):46–51. Available from: <https://doi.org/10.1016/j.exppara.2018.07.008>
 30. Casaril AE, Santos CG, Marangoni BS, Lima SM, Andrade LHC, Fernandes WS, et al. Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate analysis. *J Biophotonics*. 2021;(October 2020):1–9.

31. Echeverria JT, Soares RL, Crepaldi BA, de Oliveira GG, da Silva PMP, Pupin RC, et al. Clinical and therapeutic aspects of an outbreak of canine trypanosomiasis. *Rev Bras Parasitol Vet.* 2019;28(2):320–4.
32. Cameron JM, Butler HJ, Palmer DS, Baker MJ. Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying. *J Biophotonics.* 2018;11(4):1–12.
33. Larios GS, Nicolodelli G, Senesi GS, Ribeiro MCS, Xavier AAP, Milori DMBP, et al. Laser-Induced Breakdown Spectroscopy as a Powerful Tool for Distinguishing High- and Low-Vigor Soybean Seed Lots. *Food Anal Methods.* 2020;13(9):1691–8.
34. Larios G, Nicolodelli G, Ribeiro M, Canassa T, Reis AR, Oliveira SL, et al. Soybean seed vigor discrimination by using infrared spectroscopy and machine learning algorithms. *Anal Methods [Internet].* 2020;12(35):4303–9. Available from: <http://dx.doi.org/10.1039/D0AY01238F>
35. Scaglia E, Sockalingum GD, Schmitt J, Gobinet C, Schneider N, Manfait M, et al. Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis C using serum Fourier transform infrared spectroscopy. *Anal Bioanal Chem.* 2011;401(9):2919–25.
36. Pickering CL, Hands JR, Fullwood LM, Smith JA, Baker MJ. Rapid discrimination of maggots utilising ATR-FTIR spectroscopy. *Forensic Sci Int [Internet].* 2015;249:189–96. Available from: <http://dx.doi.org/10.1016/j.forsciint.2015.01.036>
37. Bruet V, Dumon H, Bourdeau P, Desfontis JC, Martin L. Prospective pilot study to detect dogs with non food-induced canine atopic dermatitis using fourier transform infrared spectroscopy. *Vet Dermatol.* 2016;27(5):356–61.
38. Petibois C, Cazorla G, Cassaigne A, Délérís G. Plasma protein contents determined by fourier-transform infrared spectrometry. *Clin Chem.* 2001;47(4):730–8.
39. Baker MJ, Hussain SR, Lovergne L, Untereiner V, Hughes C, Lukaszewski RA, et al. Developing and understanding biofluid vibrational spectroscopy: A critical review. *Chem Soc Rev.* 2016;45(7):1803–18.
40. Mordechai S, Shufan E, Porat Katz BS, Salman A. Early diagnosis of Alzheimer's disease using infrared spectroscopy of isolated blood samples followed by multivariate analyses. *Analyst.* 2017;142(8):1276–84.
41. Wang X, Shen X, Sheng D, Chen X, Liu X. FTIR spectroscopic comparison of serum from lung cancer patients and healthy persons. *Spectrochim Acta - Part A Mol Biomol Spectrosc.* 2014;122:193–7.
42. Elmi F, Movaghar AF, Elmi MM, Alinezhad H, Nikbakhsh N. Application of FT-IR spectroscopy on breast cancer serum analysis. *Spectrochim Acta - Part A Mol Biomol Spectrosc.* 2017;187:87–91.
43. Sheng D, Liu X, Li W, Wang Y, Chen X, Wang X. Distinction of leukemia patients' and healthy persons' serum using FTIR spectroscopy. *Spectrochim Acta - Part A Mol Biomol Spectrosc.* 2013;101:228–32.
44. Santos MCD, Nascimento YM, Araújo JMG, Lima KMG. ATR-FTIR spectroscopy coupled with multivariate analysis techniques for the identification of DENV-3 in different concentrations in blood and serum: A new approach. *RSC Adv.* 2017;7(41):25640–9.

- ## Figures

Average FTIR spectra for non-infected (black line), *L. infantum* (red line), and *T. evansi* (blue line) infected canine blood serum. The main vibrational assignments are indicated, below each spectral band, and their related molecular groups (proteins, carbohydrates, lipids, and fatty acids) are identified by different colors

(detailed description can be found at Table 1). The symbols ν , σ , and δ stand for the stretching (symmetric or asymmetric), scissoring, and bending modes, respectively.

Figure 2

Score plot (left) and loadings (right) from principal component analysis (PCA) for different sample groups in the 1800 until 800 cm^{-1} range, related to lipids, proteins, carbohydrates, and fatty acids molecules. (a, b) non-infected (black square), *L. infantum* (red circle), and *T. evansi* (blue diamond)

infected, canine blood serum. (c, d) non-infected, and *L. infantum* infected, canine blood serum. (e, f) *L. infantum*, and *T. evansi* infected, canine blood serum. (g, h) non-infected, and *T. evansi* infected, canine blood serum.

Figure 3

Accuracy of the different learning machine methods, in the LOO-CV, to discriminate and classify the canine blood serum groups from PCA data in the 1800 until 800 cm⁻¹ range. Gray bars (inclined dash pattern) for non-infected vs. *L. infantum* vs *T. evansi* (N x LI x TC); Red bars for non-infected vs. *L. infantum* (N x LI); Blue bars (mesh pattern) for *L. infantum* vs. *T. evansi* (LI x TC); and Green bars (horizontal dash pattern) for non-infected vs. *T. evansi* (N x TC), classification. ML algorithm testes: Discriminant Analysis (DA), k-nearest neighbor (KNN), and support vector machine (SVM) with different functions. The letter G stands for "gaussian".

Figure 4

Confusion matrix for the performance of the best supervised methods. (A) Non-infected (N) vs. L. infantum (LI) vs. T. evansi (TE) classification by using k-nearest neighbors (KNN) method with fine function and 5 PCs, in the 1800-800 cm⁻¹ range. (B) Non-infected vs. L. infantum classification by using quadratic discriminant analysis (DA) with 10 PCs, in the 1700-1450 cm⁻¹ range. (C) L. infantum vs. T. evansi classification by using cubic or linear support vector machine (SVM) with 10 PCs, in the 1800-800 cm⁻¹ range. (D) Non-infected vs. T. evansi classification by using linear support vector machine, or weighted k-nearest neighbors (KNN) method, with 6 PCs, in the 1800-800 cm⁻¹ range.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.pdf](#)
- [GraphicalAbstract.jpg](#)