



Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DOS MATERIAIS

Gustavo Sander Larios

Espectroscopia vibracional no infravermelho em soro sanguíneo para o diagnóstico da leishmaniose visceral e tripanossomíase canina

Campo Grande – MS

2020



Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DOS MATERIAIS

“Espectroscopia vibracional no infravermelho em soro sanguíneo para o diagnóstico da leishmaniose visceral e tripanossomíase canina”

Gustavo Sander Larios

Orientador: Prof. Dr. Cícero Rafael Cena

Coorientador: Prof. Dr. Bruno Spolon Marangoni

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul - UFMS – Instituto de Física - INFI, para obtenção do título de Mestre em Ciência dos Materiais.

Campo Grande – MS

2020

Agradecimento

O Presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço aos meus pais, Roberto e Mereide que me apoiaram em todo este trajeto me dando forças e sustentação para continuar. Agradeço também por todos os ensinamentos a mim oferecidos, pois foram a partir deles que me trouxeram até aqui.

Ao meu orientador, Prof. Dr. Cícero Rafael Cena, pela excelente orientação e confiança a mim atribuída. Sendo um grande amigo durante este percurso.

Agradeço aos meus amigos, companheiros e aos meus irmãos pelo apoio nas horas de cansaço e desânimo.

Agradeço a todos os professores por me proporcionar não somente conhecimento racional, mas auxiliarem no processo de manifestação do meu caráter e afetividade da educação e da formação profissional.

Agradeço ao grupo de pesquisa CABRA pela convivência e trocas de informações.

Agradeço a UFMS por proporcionar um ambiente criativo e amigável.

Agradeço ao grupo de Ótica e fotônica; e principalmente ao professor Samuel Leite, por confiarem a mim a utilização dos equipamentos e espaço laboratorial. Assim como, as críticas construtivas e interpretação de muitos resultados aqui apresentados.

Agradeço ao Professor Carlos Alberto do Nascimento, por confiar e dedicar seu tempo auxiliando este trabalho.

Agradeço ao professor Bruno Marangoni e ao estudante Matheus Cícero pelo auxílio, debates e ensinamentos proporcionados durante esta pesquisa.

"Há um ditado chinês que diz que, se dois homens vêm andando por uma estrada, cada um carregando um pão, ao se encontrarem, eles trocam os pães; cada um vai embora com um. Porém, se dois homens vêm andando por uma estrada, cada um carregando uma ideia, ao se encontrarem, trocam as ideias; cada um vai embora com duas.

Quem sabe, é esse mesmo o sentido do nosso fazer: repartir ideias, para todos terem pão..."
(Mario Sérgio Cortella)

Resumo

A Leishmaniose é um conjunto de doenças infectocontagiosas provocada por parasitas do gênero *Leishmania spp.*, transmitida por insetos da espécie flebotomíneos. Segundo a Organização Mundial de Saúde esta é uma das principais doenças negligenciadas de maior risco epidêmico. Sendo, incidente em cinco continentes e endêmica em 102 países; onde, estima-se que cerca de 350 milhões de pessoas correm riscos de infecção e surgem mais de 1,3 milhões de novos casos por ano. A forma mais grave e letal desta doença é a *L. visceral*, que em áreas urbanas prevalece e ocorre com maior frequência em populações caninas. Todavia, estes animais são em muitos casos assintomáticos, e servem como fonte silenciosa para a transmissão em humanos. De modo que, diagnosticar corretamente esta população impacta diretamente no controle e erradicação do número de casos da doença. No entanto, observa-se que os métodos tradicionais de diagnóstico, possuem uma série de impasses, que podem ser resumidas em: custo de operação elevado, baixa sensibilidade e especificidade, além, da alta taxa de confusão entre outras doenças parasitológicas, como a Tripanossomíase canina (TC), causada por parasitas do gênero *Trypanossoma* (espécies *T. evansi*, *T. equiperdum*, *T. vivax* e *T. cruzi*). Nesta vertente, pesquisas voltadas ao diagnóstico de doenças de difícil identificação, como cânceres, doenças virais, doenças cardiovasculares e doenças mentais, vêm associando as técnicas de Aprendizagem de Máquina (AM), juntamente a técnica de Espectroscopia de Infravermelho por Transformadas de Fourier (FTIR) para a elaboração de novas formas de diagnóstico, que resolvam as dificuldades dos métodos tradicionais e apresentem vantagens para implementação em ambiente clínico. Com isso, este trabalho busca associar a técnica de FTIR e AM para solucionar os problemas encontrado no diagnóstico da Leishmaniose visceral canina (LVC) e TC. Foi utilizado o espectro de absorção no infravermelho de soro sanguíneo de 3 grupos de animais. Sendo eles, soro negativo para LVC, soro positivo para LVC e soro positivo para TC. Nossos resultados demonstraram que a técnica de FTIR juntamente com as técnicas de AM, foram capazes de diferenciar indivíduos não infectados de indivíduos infectado por LVC e TC com sensibilidade e especificidade superior a 90%.

Palavras-chave: FTIR, Leishmaniose canina, *Trypanossoma evansi*, Análise Multivariada, Aprendizagem de máquina, Máquina de Suporte de Vetores, Análise de Componentes Principais, Validação Cruzada.

Abstract

Leishmaniasis is a set of infectious diseases, caused by parasites of the genus *Leishmania spp.*, Transmitted by insects of the phlebotomine species. According to the World Health Organization, this is one of the main neglected diseases of greatest epidemic risk, occurring on five continents and endemic in 102 countries; where, it is estimated that about 350 million people are at risk of infection and more than 1.3 million new cases appear each year. The most serious and lethal form of this disease is visceral L., which in urban areas prevails and occurs more frequently in canine populations. However, these animals are in many cases asymptomatic, and serve as a silent source for transmission in humans, so that correctly diagnosing this population directly impacts the control and eradication of the number of cases of the disease. However, it is observed that the traditional methods of diagnosis, have a series of impasses, which can be summarized in: high operating cost, low sensitivity and specificity, in addition to the high rate of confusion among other parasitological diseases, such as trypanosomiasis canine (TC), caused by parasites of the genus *Trypanosoma* (species *T. evansi*, *T. equiperdum*, *T. vivax* and *T. cruzi*). In this regard, research aimed at diagnosing diseases that are difficult to identify, such as cancers, viral diseases, cardiovascular diseases and mental illnesses, have been associating Machine Learning (AM) techniques, together with the Fourier Transform Infrared Spectroscopy (FTIR) technique for the development of new forms of diagnosis, which solve the difficulties of traditional methods and present advantages for implementation in a clinical environment. With this, this work seeks to associate the FTIR and AM technique to solve the problems found in the diagnosis of canine visceral Leishmaniasis (LVC) and CT. The infrared absorption spectrum of blood serum from 3 groups of animals was used, namely, negative serum for LVC, positive serum for LVC and positive serum for CT. Our results demonstrated that the FTIR technique, together with the AM techniques, were able to differentiate uninfected individuals from individuals infected by LVC and CT with sensitivity and specificity greater than 90%.

Keywords: FTIR, Canine Leishmaniasis, *Trypanosoma evansi*, Multivariate analysis, Machine learning, Support Vector Machine, Principal component analysis, Cross Validation.

Lista de Figuras

Figura 1: Trabalhos publicados nos últimos 20 anos segundo a plataforma Web Of Science.....	19
Figura 2: Metodologia para Análise de Componentes Principais.....	22
Figura 3: Metodologia para a técnica de Máquina de Suporte de Vetores com função Kernel.....	24
Figura 4: Metodologia para Validação Cruzada.	26
Figura 5: Espectro Infravermelho médio do soro sanguíneo canino.	29
Figura 6: Loading referente a projeção de maior variabilidade entre os espectros de FTIR.	32
Figura 7: Representação tridimensional das componentes principais para as três classes de animais.	34
Figura 8: Matriz de confusão para o sistema de aprendizagem de máquina pelo método SVM com função Kernel Quadrática.	36
Figura 9 - Loading referente a projeção de maior variabilidade entre os espectros de FTIR para a 3 ^a , 4 ^a e 5 ^a PC..	48
Figura 10 - Representação tridimensional da 4 ^a , 5 ^a e 6 ^a PC para as classes de animais.....	49
Figura 11 – Taxa de acurácia para os métodos de aprendizado supervisionado obtidos durante o treinamento com 10 dobras por VC.....	50
Figura 12 – Taxa de acurácia para os métodos de aprendizado supervisionado obtidos durante o treinamento com 15 dobras por VC.....	51

Lista de Tabelas

Tabela 1 - Estudos na área dos sensores e biossensores para o diagnóstico da Leishmaniose	17
Tabela 2: Estudos aplicando a técnica de FTIR para o diagnóstico de doenças. 19	
Tabela 3: Atribuição dos modos vibracionais do espectro IR para o soro canino..	
.....	30

Lista de abreviaturas e siglas

LV	Leishmaniose Visceral
LVC	Leishmaniose Visceral canina
TC	Tripanossomíase Canina
AM	Aprendizado de Máquina
FTIR	Espectroscopia de Infravermelho por Transformadas de Fourier
TR	Teste Imunocromatográfico Rápido
ELIZA	Ensaio Imuno Enzimático
IFI	Teste de imunofluorescência indireta
IR	Infravermelho
FP	Falsos Positivos
VN	Verdadeiros Negativos
PCR	Proteína C reativa
PCA	Análise de Componentes Principais
ATR	Refletância total atenuada
CDA	Análise de Discriminante Canônico
LDA	Análise de Discriminante de Linear
K-NN	K-Vizinhos mais próximos
PC	Componentes Principais
SiO ₂	Óxido de Silício
SVM	Máquina de Suporte de Vetores
TE	Tripanosoma evansi
ν	vibrações de alongamento
δ	vibrações de flexão
s	simétrica
as	assimétricas

Sumário

1. Introdução	11
2. Objetivos.....	14
2.1 Objetivo Geral	14
2.2 Objetivo Específico	14
3. Revisão bibliográfica	15
3.1 Métodos tradicionais para o diagnóstico LVC e TC	15
3.2 Métodos alternativos para o diagnóstico da LVC.....	16
3.3 Espectroscopia de IR para o diagnóstico de doenças	18
3.4 Métodos de Aprendizagem de Máquina	22
3.4.1 Análise de Componentes Principais	22
3.4.2 Máquina de Suporte de Vetores.....	23
3.4.3 Validação Cruzada.....	25
4. Metodologia.....	27
4.1 Amostras de Soro Canino	27
4.2 Diagnóstico das Amostras	27
4.3 Preparação das amostras para medições de FTIR.....	28
4.4 Espectroscopia de FTIR.....	28
4.5 Software para o desenvolvimento do Aprendizado de Máquina	28
5. Resultados e Discussões	29
5.1 Análise do espectro infravermelho	29
5.2 Análise Multivariada: Método não-supervisionado pela técnica de Análise de Componentes Principais.....	31
5.3 Análise Multivariada: Método supervisionado pela técnica de Máquina de Suporte de Vetores.....	35
6. Conclusão	38
7. Referencias	39
APÊNDICE A – Análise de PCA para a 3 ^a , 4 ^a , 5 ^a e 6 ^a PC.	48
APÊNDICA B – Métodos de análise supervisionada.....	50

1. Introdução

A Leishmaniose faz parte do conjunto de doenças negligenciadas e consideradas de alto risco epidêmico pela Organização Mundial de Saúde [1]. Sua incidência ocorre em 102 países, distribuídos principalmente na Ásia, África, América do Sul e no sul da Europa [2, 3]. Estima-se que cerca de 350 milhões de pessoas correm riscos de infecção, com a ocorrência de 1,3 milhões de novos casos por ano [1, 4].

A transmissão desta doença ocorre em animais mamíferos, através da picada de insetos vetores infectados de espécie flebotomíneo. Em áreas urbanas, a enzootia canina precede a ocorrência de casos de infecção em humanos, sendo mais frequente e difícil de identificar [5, 6].

O padrão ouro para o diagnóstico animal ou humano é o Exame Parasitológico, que tem como principal vantagem o custo de operação [7, 8] e como principal desvantagem a dificuldade de encontrar o parasita, sendo necessário dedicar diversas horas para analisar uma amostra e confirmar a infecção [9]. Devido a isso, alternativas mais rápidas são utilizadas como um processo de triagem para a aplicação do padrão ouro, dentre elas, destaca-se os exames sorológico/imunológico [8, 10–13] e o de Proteína C-Reativa (PCR) [9, 14, 15].

No caso da LVC os exames imunológicos são empregados devido ao baixo custo de operação, além de possuir alta eficiência devido a sensibilidade dos antígenos reagirem com os anticorpos específicos do organismo [7]. O Ensaio Imuno Enzimático (ELISA) e o teste Imunocromatográfico Rápido (TR) [8], são frequentemente aplicados para o diagnóstico da doença no animal. No entanto, no processo de triagem muitas outras doenças são identificadas erroneamente, graças à alta taxa de confusão com outras doenças causadas principalmente por outros parasitas, como por exemplo, os do gênero *Trypanosoma* [7, 16, 17].

Tal confusão ocorre principalmente pelo fato dos antígenos utilizados nos testes de triagem, serem derivados da mesma família da leishmaniose (*trypanosomatidae*) [18]. Assim, para diferenciar o agente causador da infecção, leva-se em consideração dois fatores, os sinais clínicos e se reside em áreas endêmicas, antes de desenvolver outros exames para confirmar a espécie do parasita [19]. De forma que, diferenciar se a infecção

é causada por parasitas do gênero *Leishmania* ou *Trypanosoma* é crucial para assegurar a vida do indivíduo infectado.

Como ambas as doenças possuem caráter de risco a saúde pública, e suas formas de diagnóstico apresentam diversos contrapontos, torna-se necessário desenvolver estudos para tentar sanar estas dificuldades. Nesta vertente, Mohan et. al (2011), desenvolveu um biossensor nano estruturados, a base de nanopartículas de Óxido de Níquel para identificação da LV em solução de DNA [20]. Moradi et. al. (2016), por meio de um dispositivo à base de nano folhas de ouro, identificou a leishmaniose tegumentar pelo análise do DNA [21]. Sattarahmady et. al (2016), criou um marcador visual para identificar a infecção por *Leishmania major*, através de nanopartículas de ouro dissolvidas junto a fitas de kDNA, [22]. Garcia et. al (2016), desenvolveu um dispositivo baseado em nanopartículas de ouro para identificação da *Leishmania infantum* em DNA por medições de impedância [23]. Perinoto et. al (2010), utilizando biossensor nanoestruturado com proteínas antigênicas ancoradas, foi capaz de detectar anticorpos da *Leishmania amazonensis* e *T. Cruzi* por medidas de capacitância no soro de camundongos [17]. E como trabalho mais recente, Cordeiro et. al (2019), desenvolveu um imunossensor capaz de analisar o sangue de animais infectados pela L. visceral, e por meio de medições com espectroscopia de impedância eletroquímica identificou a infecção[5]. Sobretudo, Souto et. al (2019) aponta que por mais promissor que os biossensores se mostrem para diagnosticar a leishmaniose, ou outras doenças tropicais, seu preço comercial os tornam inacessíveis no mercado atual [24].

Neste contexto, a técnicas de espectroscopia de absorção no infravermelho (IR), veem sendo aplicada para diagnosticar diversos tipos de doença, como: cânceres [25–34] doenças virais [35, 36] e doenças mentais [37], devido a sua simplicidade e versatilidade de uso. O que a torna uma nova forma de diagnóstico de baixo custo e de larga escala, que supera as limitações dos testes laboratoriais e de biópsia padrão, através de uma metodologia rápida, não invasiva e com mínima ou nenhuma preparação da amostra [38].

Em relação aos trabalhos publicados recentemente sobre a aplicação da técnica de espectroscopia de absorção no IR, destaca-se os resultados de Zlotogorski-Hurvitz et. al (2019), que diferenciou pacientes com câncer na boca de pacientes saudáveis pela análise de suas salivas [39]. Chaber et al. (2019) e Chaber et al. (2018), determinou a probabilidade de sucesso no tratamento de pacientes com sarcoma de Ewing pela análise do espectro FTIR obtido em tecido ósseo[40] e diferenciou este câncer de osteomielite

[41]. Xinhao et al. (2019), analisou pacientes saudáveis de pré-diabéticos, por meio do sangue periférico e os diagnosticou por esta técnica[42]. Já em relação a parasitas família *trypanosomatidae*, Aguiar et al. (2018) analisou cepas de *L. chagasi*, *T. cruzi* e *T. rangeli* e obteve uma diferenciação clara entre a estrutura morfológica destas espécies [43].

Todavia, não encontramos nenhum trabalho voltado ao diagnóstico da LVC e TC utilizando espectroscopia de FTIR. Portanto, devido ao alto risco epidemiológico e as dificuldades encontradas para o diagnóstico destas doenças em animais, este trabalho visa empregar esta técnica junto aos métodos de AM para solucionar os problemas no diagnóstico da LVC e TC.

2. Objetivos

2.1 Objetivo Geral

Propor uma nova de diagnosticar animais com leishmaniose visceral e tripanossomíase, por meio da espectroscopia de infravermelho com transformadas de Fourier junto aos métodos de aprendizado de máquina.

2.2 Objetivo Específico

- Identificar a composição molecular do soro sanguíneo canino através do espectro de FTIR, comparando indivíduos infectados com os não infectados;
- Determinar pela Análise de Componentes Principais, as regiões espectrais que melhor contribuem para diferenciação das classes de animais;
- Utilizar os espectros de FTIR para a criação de um banco de dados, que será empregado pelos métodos de aprendizagem supervisionada, servindo como fonte base para realização do diagnóstico;
- Determinar um método de supervisionado que melhor se adeque aos dados deste estudo;
- Com este método, desenvolver um sistema de aprendizado para diagnosticar a LVC e a TC através do espectro de FTIR;
- Comparar a sensibilidade, especificidade, custo e aplicabilidade do método proposto com os procedimentos tradicionais destas doenças, verificando o potencial para a implementação deste tipo de exame em ambiente clínico.

3. Revisão bibliográfica

3.1 Métodos tradicionais para o diagnóstico LVC e TC

A leishmaniose e a tripanossomíase são doenças parasitológicas provocadas por parasitas da família Trypanosomatidae, dos gêneros *Leishmania* e *Trypanosoma* [45, 46]. A transmissão da LVC é provocada por insetos vetores de espécie *flebotomíneo*, enquanto a TC é transmitida por insetos *triatomíneos*, através do contato entre suas fezes, ingestão direta do inseto, transfusão sanguínea ou pela ingestão de alimentos contaminados [45, 46].

Segundo o Ministério da Saúde e as Organizações mundiais de saúde, estas doenças são um grande problema para os serviços públicos devido a três fatores: a variedade de sintomas, alterações histopatológicas inespecíficas e a inexistência de um diagnóstico 100% específico e sensível [46–48]. Graças a grande variedade de sintomas, somente o diagnóstico clínico impossibilita a identificação destas doenças, de forma que sua confirmação deve ser feita por métodos laboratoriais específicos, que são baseados em exames parasitológico e sorológicos/imunológicos [49].

O método parasitológico é um procedimento altamente invasivo, onde as amostras são obtidas por meio de punções ou escarificação de pele e analisadas por microscopia ótica, sendo impraticáveis por programas de saúde públicas devido a sua dificuldade e demora para operação [46]. Todavia, pode-se afirmar que este um método seguro de confirmação da doença, visto que os resultados são baseados na observação direta do parasita, o que em casos avançados a especificidade deste exame pode chega 100%, no entanto, em casos de animais assintomáticos a sensibilidade fica abaixo dos 80% [46, 47].

Os exames sorológicos/imunológicos são praticados como alternativa de triagem para o exame parasitológico e são desenvolvidos em duas etapas, a do teste imunocromatográfico rápido (TR), que é a primeira alternativa para identificar os animais soro positivos e o Ensaio Imunoenzimático (ELISA), utilizado para confirmar os animais soro reagentes ao TR [46, 47].

O teste ELISA consiste na reação de anticorpos presentes no soro animal com antígenos específicos para a *Leishmania*, através de marcadores que geram um resultado observável por espectrofotometria ou por coloração [47, 48]. A sensibilidade e

especificidade deste exame depende restritamente da espécie do parasita, pois a mesma interfere na quantidade de anticorpos presentes no soro e dos antígenos utilizados para reação [8, 10–13]. Com isso, a principal vantagem do ELISA é a possibilidade de analisar diversas amostras ao mesmo tempo, e sua principal desvantagem é a alta taxa de reação cruzada com diversas outras doenças, como: tripanossomíase, pênfigo foliáceo sul-americano, paracoccidiodomicose, esporotricose, entre outras [16].

O teste TR consiste em um procedimento rápido e de fácil análise, onde é realizado uma reação direta no soro sanguíneo animal com proteínas recombinantes K39, sendo um procedimento altamente sensível e pouco específico, visto que pode apresentar interação com diversos anticorpos presentes no soro animal [50]. Desta forma, este procedimento é adotado como uma alternativa rápida para a triagem entre os exames ELISA e parasitológico [50].

3.2 Métodos alternativos para o diagnóstico da LVC

Em relação aos procedimentos tradicionais de diagnóstico da LVC, é observado variações nas metodologias para o desenvolvimento destes exames, como por exemplo, no caso do teste ELISA ao utilizar antígenos diferentes é possível aumentar a especificidade e a sensibilidade do teste [8, 10, 12], podendo ainda de identificar a espécie do parasita [11].

No casos dos exames de Reação em Cadeia de Polimerase (do inglês *Polymerase Chain Reaction*, PCR), é possível determinar a espécie e a subespécie do parasita com alta precisão [51], além de possibilitar diversas metodologias para melhorar sua sensibilidade e especificidade [9, 14]. No entanto, este procedimento não é empregado por órgãos de saúde pública, devido a necessidade de equipamentos robustos, estrutura laboratorial avançada e manuseio técnico especializado [52].

Por mais que os procedimentos tradicionais de diagnóstico da leishmaniose ainda são estudados e aperfeiçoados, os principais problemas em relação ao custo de operação, tempo de realização do exame, tempo para aquisição de resultados e metodologias que possibilitem aplicação em grandes quantidades de amostras ainda não foram definitivamente resolvidas.

Nesta vertente, pesquisas veem aplicando diversas metodologias alternativas para diagnósticas doenças, um dos principais ramos de diagnóstico são os sensores químicos e biossensores, que são dispositivos tecnológicos promissores para o diagnóstico rápido fácil aplicação campo [24].

Os sensores químicos identificam informações como a presença e concentração de moléculas específicas para a doença, e por sinal eletrônico faz uma mensuração, onde, por meio de padrões desenvolve seu diagnóstico[24]. Enquanto os biossensores identificam a doença pela interação entre os receptores e as moléculas ou antígenos específicos, gerando um sinal eletrônico ou marcador visual que possibilita determinar sua presença [24]. Na Tabela 1 é apresentado os trabalhos voltados ao emprego desta tecnologia no diagnóstico da leishmaniose.

Tabela 1 - Estudos na área dos sensores e biossensores para o diagnóstico da Leishmaniose

Sensor/Biossensor	Alvo	Indivíduos	Tempo	Ref.
Imunossensor de ressonância plasmônica de superfície (SPR)	Soro	Cães	-	[53]
Filme nanoestruturado de óxido de níquel (NiO)	DNA	Humanos	-	[20]
Imunossensor de impedância eletroquímica	Soro	Cães	10 a 60 min	[5]
Biossensor de compósito híbrido de polianilina com nanopartículas de ouro	DNA	Cães	-	[23]
Biossensor de nanopartículas de ouro	DNA	Humanos	-	[21]
Biossensor com antígenos de <i>Leishmania amazonensis</i> e <i>T. cruzi</i>	Soro	Camundongos	-	[17]
Biossensor com nanopartículas de ouro	DNA	Humanos	15 a 24 min	[22]

Segundo os trabalhos apresentados na Tabela 1, os sensores/biossensores desenvolvidos para o diagnóstico da leishmaniose apresentaram resultados comparativos aos exames ELISA, TR e PCR, com a vantagem de uma resposta diagnóstica muito rápida e simples, visto que basta aplicar uma pequena quantidade de amostra no dispositivo e por medidas elétricas o resultado é obtido em poucos minutos [5, 22].

Todavia, o custo de produção para um dispositivo deste porte é extremamente elevado, pois são altamente tecnológicos, sensíveis e compostos de nanomateriais [20–23], o que torna seu custo final inacessível pelos órgãos de saúde pública [24].

Apesar de diversos avanços para o diagnósticos da LVC e TC terem sido desenvolvidos, um exame de baixo custo, não invasivo, rápido, com percento específico e sensível ainda não foi alcançado[46–48]. Sobretudo para a população canina, que são os principais alvos de infecção e podem servir como fontes para a transmissão humana [54].

3.3 Espectroscopia de IR para o diagnóstico de doenças

Dentre as diversas formas de diagnóstico de doenças, as técnicas baseadas em espectroscopia de absorção no IR, como o FTIR, veem ganhando destaque por sua praticidade, rapidez e facilidade de manuseio. Sua metodologia se baseia na interação molecular de cunho vibracional, onde é possível identificar os grupos moleculares e quaisquer alterações que estes apresentam de forma rápida e altamente precisa, com pouca ou quase nenhuma preparação previa da amostra [55–57], permitindo a possibilidade de emprego em larga escala [38, 68] e com baixo custo operacional [57].

Segundo a plataforma de pesquisa *Web of Science*, nos últimos 20 anos, observa-se um crescente aumento no número de publicações desta técnica para o diagnóstico de diferentes tipos de doenças. A Figura 1 demonstra o número de publicações (registros) até o final de 2019, filtrando as palavras ‘FTIR’ e ‘diagnósticos’ no conteúdo dos trabalhos.

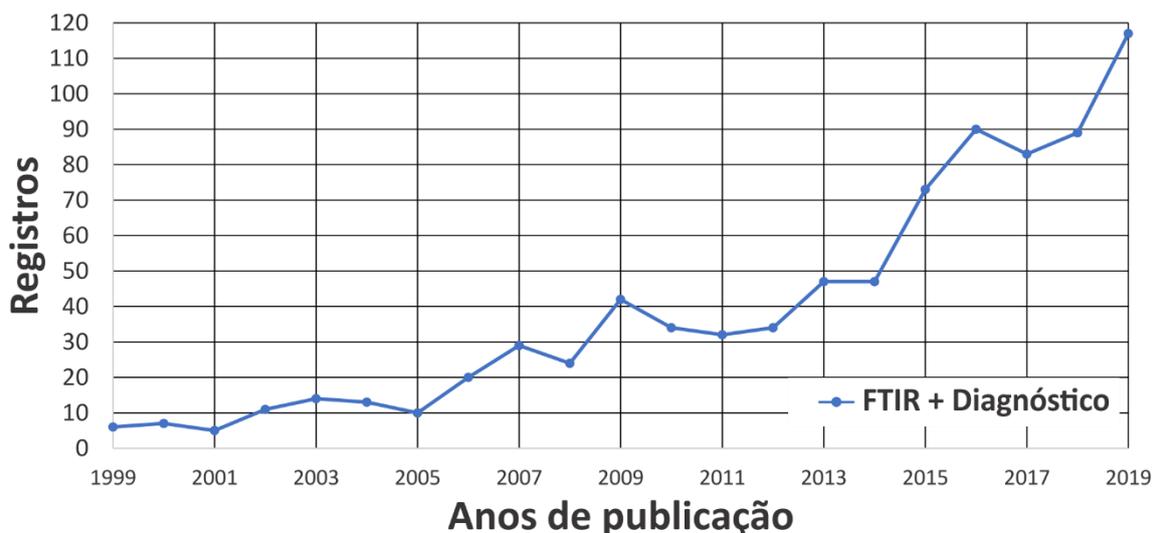


Figura 1: Trabalhos publicados nos últimos 20 anos segundo a plataforma Web Of Science. Fonte: Autoria própria.

Observa-se que desde 1999 a 2019 o número de publicações na área de diagnósticos com o uso do FTIR aumentou em 600 vezes, totalizando 827 publicações. Isso enfatiza que o FTIR está se mostrando como uma nova ferramenta promissora para auxiliar e facilitar o diagnóstico de diversas doenças [57, 59, 60]. Na Tabela 2, é apresentado algumas pesquisas utilizando o FTIR no diagnóstico de doenças de difícil identificação, dando destaque ao uso em diferentes alvos de análise, métodos de AM e nos resultados alcançados.

Tabela 2: Estudos aplicando a técnica de FTIR para o diagnóstico de doenças.

Método de análise	Doença	Alvo	Resultado	Ref.
PCA, XGBOOST e CART	Pré-diabétes	Sangue	Esp. - 100%; Sen. - 85%; Prec. - 93,33%	[42]
PCA, LDA, QDA, SVM, KNN e RF	Sarcoma de Ewing e Osteomielite	Tecido ósseo	Prec. - 88%	[41]
PCA, LDA, QDA, SVM, KNN e RF	Taxa de morte por Sarcoma de Ewing	Tecido ósseo	Espec. - 92%; Sens. - 75,3%	[40]
PCA, LDA E Regressão Parcial e PLSR	Dengue	Soro	Sens. - 89%; Espec. - 95%; $R^2 = 0,9980$	[61]
PCA e LDA	Câncer Pâncreas	Tecido		[62]
PCA, LDA e SVM	Câncer de Boca	Exossomos salivares	Sens. - 100%; Espec. - 89%; Prec. - 95%	[39]
PCA, PLS-D A	Bagomovírus	Folhas		[63]
PCA, SPA, GA, LDA	Vírus da Dengue	Soro	Prec. - 100%	[36]

PCA-LDA, SPA-LDA E GA-LDA	Chikungunha e Zika Virus	Sangue	Sens. - 100% para Chikungunya; Sens. - 92% para Zika vírus.	[35]
Teste-T	Câncer no ovário	Tecido	_____	[25]
Teste U	Câncer no ovário	Tecido	_____	[26]
LDC, QDC, ECLASS, K-NN e SVM	Câncer de ovário Câncer endométrio	Soro e plasma	Soro - 96,67%; Plasma - 95% (Cânc. Ovár.) Soro - 77,08%; Plasma - 81,67% (Cânc. Endom.)	[64]
Método de Yano's Teste-T	Leucemia	Soro	Espec. - 79%; Sens. - 83%	[65]
Teste -T e CDA	Câncer gástrico	Glóbulo vermelho	Sens. - 95%; Espec. - 70%; Prec. - 84,2%	[29]
PCA e GA	Câncer gástrico	Tecido	Sens. - 100%; Espec. - 68,2%; Prec. - 81,1%	[27]
PCA e LDA	Câncer Mama	Soro	Sens. - 84%; Prec. - 74%; Prec.: 83%	[28]
Análise de cluster e Redes Neurais	Câncer de mama	Soro	Espec. - 95%-100%; Sens. - 98%	[30]
Acompanhamento pelo Espectro	Câncer de mama	Tecido	_____	[31]
Teste-T	Câncer de Pulmão	Soro	Sens. -83,3%; Espec. - 100%	[66]
ANOVA, PCA, LDA e Teste-T	Câncer de colo	Tecido	_____	[34]
Teste-T, PCA e LDA	Câncer de colo	Tecido	Sens. - 96,6%; Espec. - 89,9%; Prec. - 93,8%	[32]
PCA-LDA; PCA-QDA; GA-LDA e GA-QDA	Câncer de Próstata	Tecido	Sens. - 75%; Espec. - 100%	[33]
PCA e LDA	Doença de Alzheimer	Sangue, soro e leucócito	Prec. geral: 90%; estágio grave: 85%; estágio moderado: 81,9%; diferenciação entre os estágios: 85%	[37]

Legenda: Prec. – Precisão; Sens. – Sensibilidade; Espec. – Especificidade; LDA – Análise de Discriminante Linear; QDA – Análise de discriminante Quadrático, SVM – Máquina de Suporte de Vetores, KNN – Vizinhos k-próximos; RF – Árvores aleatórias; PLSR - Mínimos Quadrados Parciais; GA – Análise de agrupamento; CDA - Análise de Discriminante Canônico.

Pode-se destacar que algumas doenças foram diagnosticadas tanto em tecido, quanto em biofluidos, como o soro, plasma e sangue total [25, 26, 28, 30, 31, 35, 36, 62]. O que demonstra a versatilidade que a espectroscopia de FTIR apresenta em desenvolver o mesmo diagnóstico em diferentes materiais, com alta sensibilidade e especificidade, visto que as medições no FTIR indicaram a composição molecular do material. Um

exemplo desta versatilidade foi o estudo desenvolvido por Depciuch, J. et al. (2017), que obteve biomarcadores visíveis no espectro de FTIR, que indicavam se o paciente estava curado ou não do câncer de mama [31]. No entanto, a maioria dos trabalhos apresentados precisaram utilizar as técnicas de AM para potencializar a análise dos espectros, no intuito de aproveitar 100% das informações para assim desenvolverem o diagnóstico por interpretação dos dados [66, 67].

Vale ressaltar que existem ainda alguns desafios científicos e tecnológicos que devem ser superados para implementação desta técnica em ambiente clínico, no caso dos biofluidos o processo de coleta e secagem da amostra deve ser bem esquematizado, pois estes fatores influenciam diretamente na quantidade de analitos e na heterogeneidade das amostras[38]. Visto que, durante o processo de secagem as moléculas dos biofluidos não se distribuem uniformemente, fazendo com que as bordas de uma gota de sangue, por exemplo, tenham mais material do que o centro [58].

As etapas de pré-analíticas e analíticas também são fundamentais para a padronização da coleta fora do laboratório, onde protocolos devem ser desenvolvidos para garantir um padrão nas medições[38]. Por exemplo, a forma, o local da coleta, proporção de diluição, armazenamento, cuidados, insumos, preparações e principalmente a triagem envolvendo as características biológicas dos indivíduos, como, sintomas, sexo, idade, entre outras, pode influenciar no padrão das amostras, causando alterações indesejadas [38].

Além destes fatores, os processos de AM utilizados são pontos importantes para um bom funcionamento da técnica em usos de diagnóstico, todavia, não existe um consenso sobre qual o melhor método a ser utilizado ou priorizado, visto que eles dependem restritamente da qualidade e da quantidade de informações. No entanto, é observado uma tendência entre estas pesquisas, empregando a técnica de Análise de Componentes Principais (ACP ou PCA, do inglês *Principal Component Analysis*), junto as técnicas de análise supervisionadas, como a Máquina de Suporte de Vetores (MSV ou SVM, do inglês *Support Vector Machine*).

3.4 Métodos de Aprendizagem de Máquina

3.4.1 Análise de Componentes Principais

A PCA é uma técnica multivariada utilizada para a redução do volume dos dados com a menor perda possível de qualidade, promovendo a observação do comportamento do sistema em coordenadas reduzidas [69].

Esta técnica se baseia na criação de componentes chamadas de Componentes Principais (CP ou PC do inglês, *Principal Component*), capazes de descrever a variância e a covariância dos dados originais $x_1, x_2, x_3, \dots, x_n$, em termos de variáveis não correlacionadas $y_1, y_2, y_3, \dots, y_n$. Onde cada y_i , consiste em uma combinação linear das variáveis originais $x_1, x_2, x_3, \dots, x_n$ [69, 70], conforme ilustrado na Figura 2.

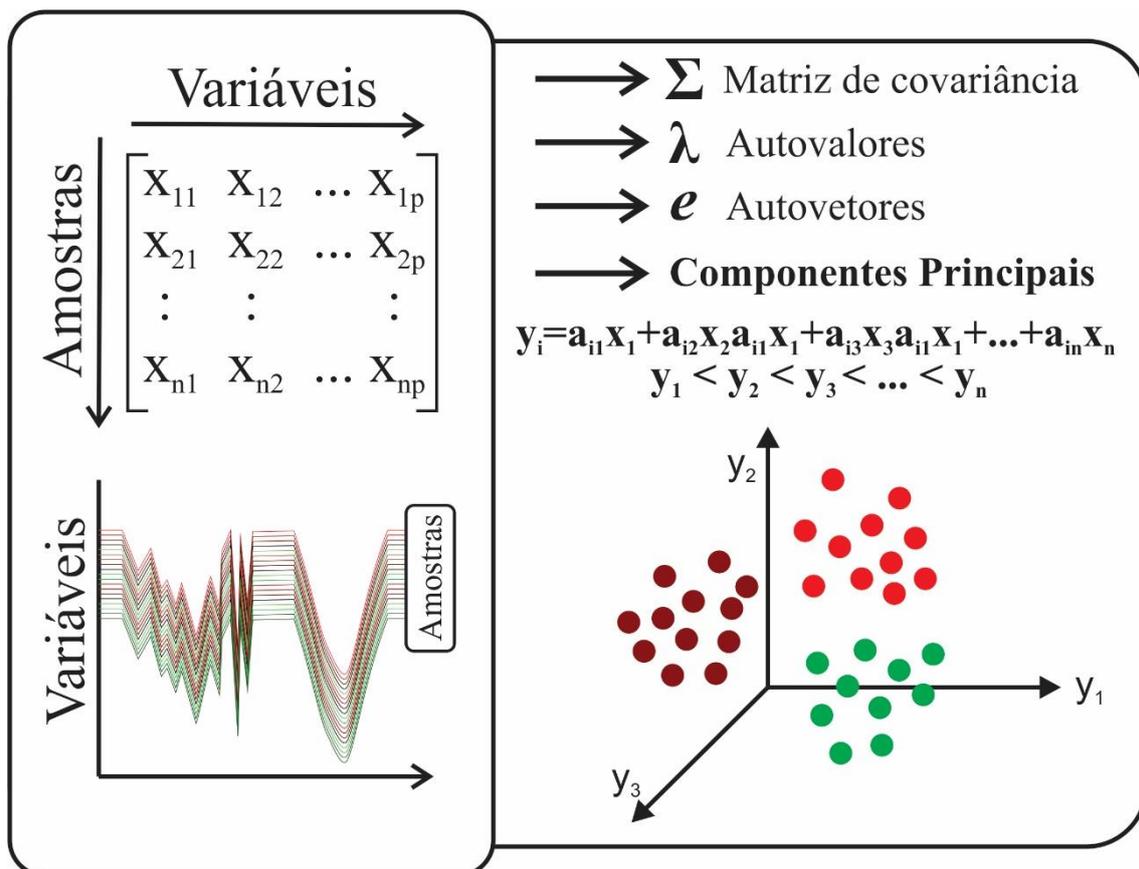


Figura 2: Metodologia para Análise de Componentes Principais.

Fonte: Autoria própria.

A PCA consiste na extração da matriz de covariância do sistema e a partir dela obtém-se os autovalores e autovetores, que definem a direção e a parcela de maior variabilidade dos dados, sendo estes referentes a direção das Componentes Principais (CP ou PC, do inglês *Principal Component*). Geometricamente pode-se representar o sistema de conversão realizada pela PCA imaginando a rotação e translação do sistema de dados $x_1, x_2, x_3, \dots, x_n$ original, com o eixo centrado na média; e através de combinações lineares entre estas informações, determina-se as projeções em direção as PC.

Nesta análise cada variável não correlacionada y_i possui uma ordem decrescente de importância, onde a y_1 expressa a maior variabilidade do sistema e é chamada de primeira componente principal (1ª PC); enquanto a segunda variável y_2 , chamada de segunda componente principal (2ª PC) não é correlacionada a y_1 e concentra a segunda maior parte da variação do sistema de coordenadas; da mesma forma que a terceira variável y_3 (3ª PC) não se correlaciona a nenhuma outra variável e representa a terceira maior parte da variação do sistema. Similarmente o sistema de coordenadas se repete até que todos os dados sejam completamente transformados em novas coordenadas não correlacionadas [70].

Desta forma, esta técnica se torna uma potente ferramenta para auxiliar na visualização com comportamento do sistema, quando o mesmo apresenta um grande volume de dados ou amostras, facilitando a diferenciação dos grupos por meio das coordenadas reduzidas, preservando grande parte da qualidade dos dados.

3.4.2 Máquina de Suporte de Vetores

O SVM é um algoritmo matemático utilizada para o reconhecimento de padrões, baseada em funções que divide os dados em regiões ou hiperplanos de separação. Para isso, a técnica SVM utiliza vetores de peso, juntamente com margens de risco e funções para diminuir o *overfitting* do sistema, estipular intervalos de confiança entre a separação das classes[71].

Seu fundamento se baseia no mapeamento dos dados de entrada por meio de funções e transformações, que podem ser auxiliadas pelas funções Kernel linear, quadrática ou polinomial. No SVM os dados que estão em um conjunto X , representados por x_i , são convertidos para um espaço de alta dimensão Z , onde cada conjunto x_i é

transformado em uma função $\phi(x)$ pelo algoritmo Kernel [76, 77]. A metodologia da técnica SVM associada as funções Kernel está representada em sua forma simples e resumida através da Figura 3.

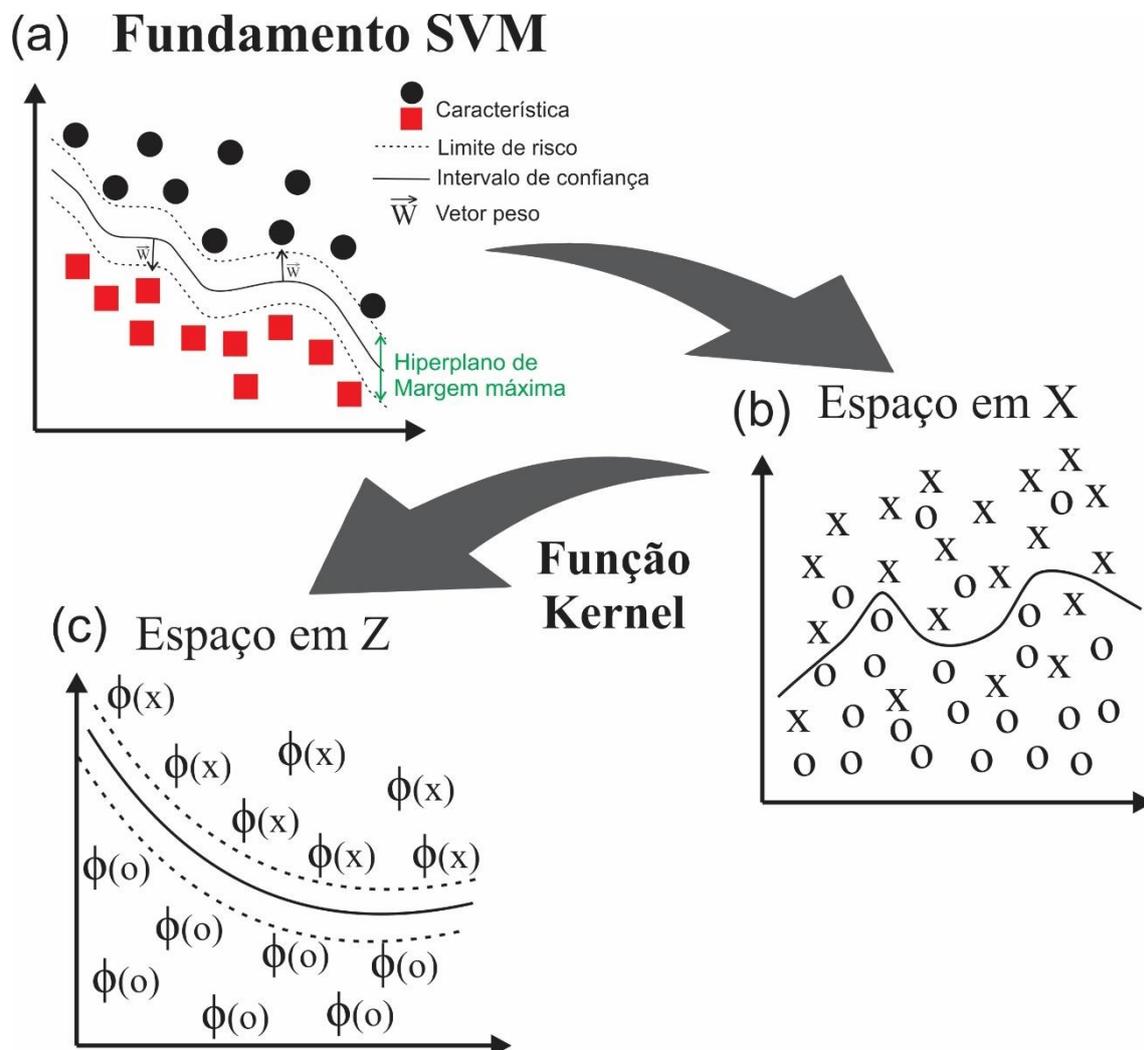


Figura 3: Metodologia para a técnica de Máquina de Suporte de Vetores com função Kernel. Fonte: Autoria própria.

Na Figura 3(a) a técnica SVM desenvolve uma análise em alta dimensão sobre o conjunto de dados, onde cada característica retém todas as informações e variáveis do sistema. A partir destas características, a técnica determina vetores capazes de diferenciar entre os grupos. E com isso, determina limites de riscos e intervalos de confiança para sua separação. No entanto, quando o sistema apresenta alta complexidade, como no caso da Figura 3(b), surge a necessidade de aplicar métodos para melhorar a interpretação do sistema.

Estes métodos são determinados pelas funções Kernel, que facilitam a interpretação destes sistemas convertendo os dados que estão em uma dimensão X , compostas por x_i informações, em um novo sistema Z . Onde, cada componente x , será agora interpretada por uma função Kernel característica, facilitando o desenvolvimento do método, como mostrado na Figura 3(c). Os cálculos para o desenvolvimento do sistema SVM e das funções Kernel, são apresentados com mais detalhes por Schölkopf, Burges, e Smola (1999)[71] e Muller et al. (2001)[72].

3.4.3 Validação Cruzada

A técnica de Validação Cruzada (VC ou CV, do inglês *Cross-Validation*) foi proposta inicialmente por Geisser (1975)[74], como um modelo analítico de previsão e validação de conjunto de dados. De modo que permita uma análise que não conflite ou interfira em um processo de aprendizado, evitando a tendenciosidade em sistemas computacionais de análise de dados [75].

Sua metodologia é baseada na divisão do conjunto de dados em k pastas ou dobras (do inglês *k-fold*) aleatoriamente, onde as informações de cada pasta possua a mesma ou quase a mesma quantidade de informações[74]. Com isso, a VC treina o sistema desenvolvido k vezes, utilizando $k-1$ em cada um dos treinamentos, e através de sucessíveis repetições treina todos os dados e os utiliza para validar o sistema, sem que estes interfiram no processo[74]–[76].

Por fim, a precisão final do processo de VC, é obtida através do cálculo de pontuação média de todos os resultados treinados [74]–[76]. A Figura 4 exemplifica o processo de validação cruzada proposto por Geisser (1975)[74], onde o conjunto de dados é treinado em quatro dobras e o resultado da validação é feito pela pontuação média da discrepância entre os treinamentos.

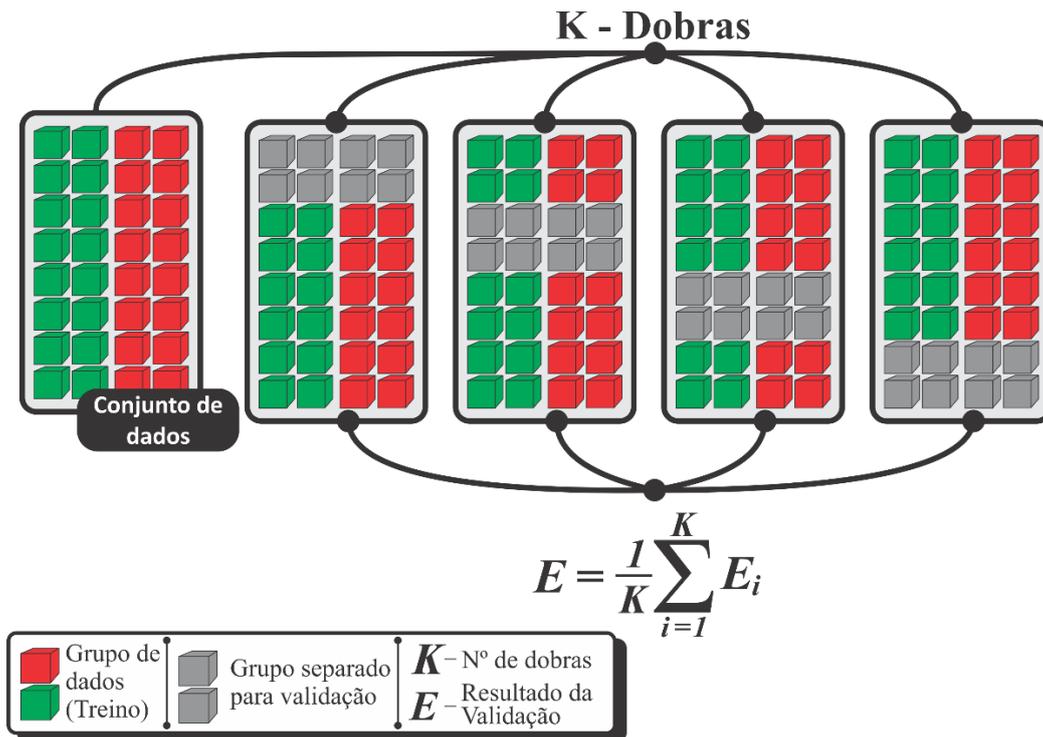


Figura 4: Metodologia para Validação Cruzada. Fonte: Autoria própria.

Em adicional, Rodriguez, Perez e Lozano (2010) observou que o processo de VC depende somente de dois fatores: a quantidade de dados para o treinamento e o número de dobras escolhidas[75]; de forma que para obter um resultado confiável neste tipo de validação, é necessário que o número k seja de 5, 10 ou mais dobras, dependendo da quantidade de dados [75], [77].

4. Metodologia

4.1 Amostras de Soro Canino

As amostras soro positivo e soro negativo para LV utilizadas nesse trabalho são oriundas do banco de amostras do laboratório de biologia molecular da FAMEZ (Faculdade Veterinária e Zootecnia) da Universidade Federal de Mato Grosso do SUL (UFMS), coletadas durante a execução do projeto "Caracterização genética de isolados de *Leishmania infantum* (sin. *L. chagasi*) de animais por MLSA (*Multilocus Sequence Analysis*), e sua associação com aspectos clínicos e epidemiológicos da leishmaniose visceral" (CEUA 765/2016). As amostras de soro positivos para TC, foram obtidas durante um surto epidemiológico na região do Pantanal-MS, Brasil [78].

O corpo amostral utilizado nesta pesquisa foi de 20 animais soro positivo para LVC, 8 animais soro positivo para TC e 20 animais soro negativo somente para LVC, totalizando 48 indivíduos.

4.2 Diagnóstico das Amostras

O diagnóstico clínico dos animais considerados positivos para LVC, ocorreu quando apresentavam sinais clínicos compatíveis, tais como: onicogribose, linfadenopatia, emagrecimento progressivo, etc. Além do, resultado sorológico positivo (imunocromatografia e ELISA) foi realizado a visualização de formas amastigotas de *Leishmania spp.* por microscopia óptica para confirmação da infecção.

Os animais considerados negativos para LVC, foram assim classificados quando não apresentavam sinais clínicos de doença, e obtiveram resultados sorológico negativo para a LVC (imunocromatografia e ELISA).

Os animais positivos para TC, foram diagnosticados por PCR, seguindo os protocolos propostos por Geysen et al. 2003, onde foi confirmando a infecção por *Trypanossoma evansi* através do sequenciamento genético deste parasita[79].

4.3 Preparação das amostras para medições de FTIR

O soro sanguíneo animal foi mantido congelado até o momento do uso a uma temperatura de -20°C . Para a preparação deste material, o soro canino foi descongelado a temperatura ambiente e depositado em substrato de óxido de silício (SiO_2), seguindo os seguintes passos:

1. Homogeneização do soro utilizando Vortex por aproximadamente 30 s;
2. Deposição de $30\ \mu\text{l}$ do material em substrato de SiO_2 ;
3. Secagem do material em estufa a 40°C por aproximadamente 30 min.

Estes passos foram repetidos três vezes no intuito de aumentar a concentração do material biológico no substrato. Além disso, para cada animal aqui analisado, seu soro sanguíneo foi distribuído em três substratos de SiO_2 . Sendo assim, esta pesquisa analisou 60 amostras com LVC, 24 amostras com TC e 60 amostras não infectadas (grupo controle), totalizando 144 amostras provenientes de 48 animais.

4.4 Espectroscopia de FTIR

As amostras foram analisadas uma única vez pela técnica de espectroscopia de infravermelho com Transformadas de Fourier (FTIR) por meio do equipamento PerkinElmer Spectrum 100N FT-NIR, com o acessório ATR (Refletância total atenuada) acoplado no modo transmitância, na região de $4000\text{-}600\ \text{cm}^{-1}$, utilizando 5 varreduras com uma resolução de $4\ \text{cm}^{-1}$.

4.5 Software para o desenvolvimento do Aprendizado de Máquina

Os métodos de AM foram desenvolvidos por meio do software MATLAB R2018b, sobre a região espectral de $4000\text{-}600\ \text{cm}^{-1}$ (total de 3401 variáveis). Nele foi desenvolvido as análises de PCA e as técnicas supervisionadas através da extensão *Classification Learner* versão 11.3. Os modelos de aprendizagem supervisionada foram validados por meio do sistema de Validação Cruzada (VC).

5. Resultados e Discussões

5.1 Análise do espectro infravermelho

Após a preparação do material, as amostras foram medidas uma única vez e como esperado os picos correspondentes as proteínas, carboidratos, lipídios e ácidos graxos são observáveis no espectro de FTIR (Figura 5). As posições dos picos junto aos modos vibracionais e as biomoléculas correspondentes são descritas na Tabela 3.

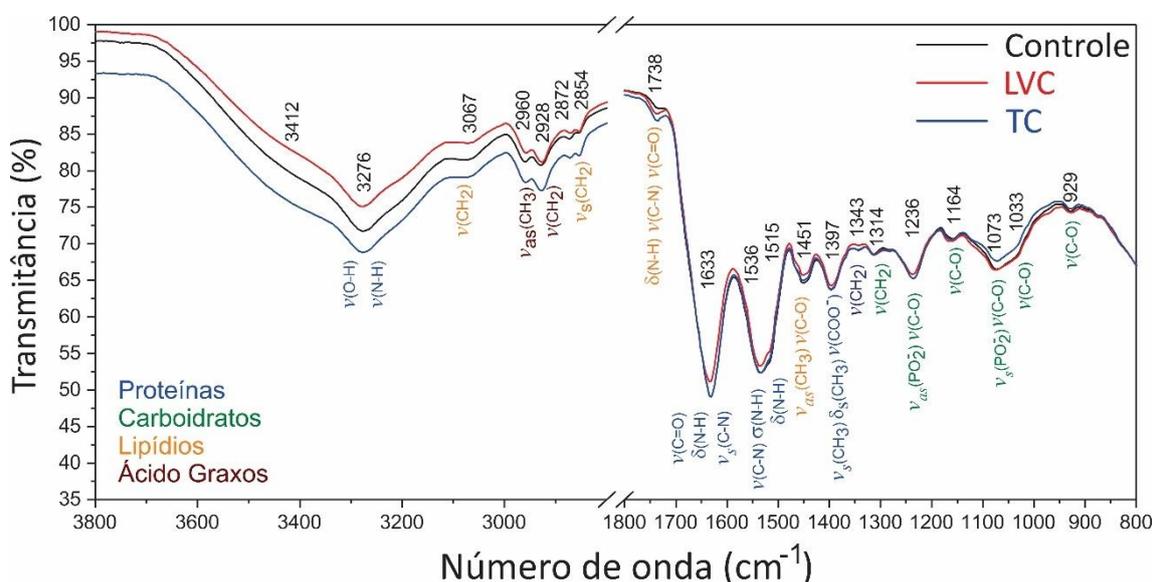


Figura 5: Espectro Infravermelho médio do soro sanguíneo canino.

Legenda: ν : vibrações de alongamento; δ : vibrações de flexão; s : simétrica; as : assimétrica[80]. Fonte: Autoria própria.

No processo de infecção, o organismo do hospedeiro desenvolve diversas alterações a fim de controlar o desenvolvimento da doença, neste sentido, era esperado que houvesse diferenças entre os picos representativos das espécies reativas de oxigênio e nitrogênio, além de variações nos anticorpos presentes no soro entre os indivíduos infectados para o grupo controle [12, 84–86]. No entanto, o espectro de FTIR para as três classes de animais apresentam grandes semelhanças entre si, o que torna necessário o uso de algoritmos (PCA, neste caso) que matematicamente encontram informações espectrais capazes de discriminar uma classe da outra.

Tabela 3: Atribuição dos modos vibracionais do espectro IR para o soro canino.

Modos vibracionais (cm ⁻¹)	Atribuições	Biomoléculas	Referências
≈929	$\nu(\text{C-O})$ $\nu(\text{C-O-C})$	Sacarídeos, carboidratos;	[80, 81]
≈1033	—	Sacarídeos, DNA, carboidratos;	[63, 80–82]
≈1073	— $\nu_s(\text{PO}_2^-)$	Sacarídeos, ácidos nucleicos, RNA/DNA, carboidratos;	[35, 36, 63, 64, 80–82]
≈1164	—	Sacarídeos, carboidratos; Sacarídeos, ácidos nucleicos, diéster de fosfato, RNA/DNA, carboidratos;	[35, 80, 81, 83]
≈1236	— $\nu_{as}(\text{PO}_2^-)$	Serina, Tirosina, Treonina;	[28, 35–37, 57, 63, 64, 80–82]
≈1314	$\nu(\text{CH}_2)$	Colágeno (Amida III);	[28, 36, 63, 64, 80, 81]
≈1343		Transferrina; Glicoproteína ácida $\alpha 1$	[82]
≈1397	$\nu_s(\text{COO}^-)$ $\delta_s(\text{CH}_3)$ $\nu_s(\text{CH}_3)$	Fibrinogênio; Aminoácidos, Proteínas; IgG1; IgM; IgA; Haptoglobina;	[37, 57, 64, 80, 81]
≈1451	$\nu(\text{C-O-O})$ $\nu_{as}(\text{CH}_3)$	Apolipoproteína-A1;	[37, 63, 64, 80–82]
≈1515	$\delta(\text{N-H})$	α -hélice da proteína	[28], [84], [85]
≈1536	$\nu_s(\text{C-N})$ $\sigma(\text{N-H})$	(Amida II); Albumina; IgG4	[28, 36, 57, 63, 80–82]
≈1633	$\nu(\text{C=O})$ $\delta(\text{N-H})$ $\nu_s(\text{C-N})$	Folha β e hélice de proteínas (Amida I);	[28, 36, 57, 63, 80–82]
≈1738	$\nu(\text{C=O})$ Amide I $\nu(\text{C-N})$ $\delta(\text{H-N})$ Amide II	Lipídios, fosfolipídios, colesterol, ésteres, ácidos graxos;	[35–37, 57, 63, 64, 80, 82]
≈2854	$\nu_s(\text{CH}_2)$	Lipídios;	Apolipoproteína-B [57, 80, 81]
≈2872	$\nu(\text{CH}_2)$		[28, 57, 80]

≈ 2928	$\nu_{as}(\text{CH}_2)$	Ácidos graxos, ésteres, glicerol, fosfolipídios, triglicerídeos;	[28, 57, 63, 64, 80]
≈ 2960	$\nu_{as}(\text{CH}_3)$	Ésteres de colesterol, lipídios, ácidos graxos;	[28, 57, 63, 64, 80]
≈ 3067	$\nu(\text{CH}_2)$	Lipídios, lipídios não saturados (Amida B);	[28, 36, 57, 63]
≈ 3276	$\nu(\text{N-H})$ $\nu(\text{O-H})$	Proteínas (banda Amida A);	[28, 57]
≈ 3412		Proteínas (Amida A e B);	[28]

5.2 Análise Multivariada: Método não-supervisionado pela técnica de Análise de Componentes Principais

O emprego da técnica de PCA permite observar o comportamento dos dados espectrais, através da redução de dimensionalidade do sistema, por meio de combinações lineares entre os dados do espectro de FTIR e as projeções das variáveis na direção de maior variabilidade dos dados.

A Figura 6 apresenta os coeficientes das informações espectrais das classes em relação a estas projeções, obtidos pela análise de PCA. Nela é exposto os coeficientes da 1ª, 2ª e 3ª PC, sendo estas, com índice de variância de 65,53%, 32,78% e 0,98%, respectivamente, totalizando 99,29% das informações dos espectros. Variância da qual, é suficiente para investigar o comportamento dos espectros de FTIR em uma observação tridimensional.

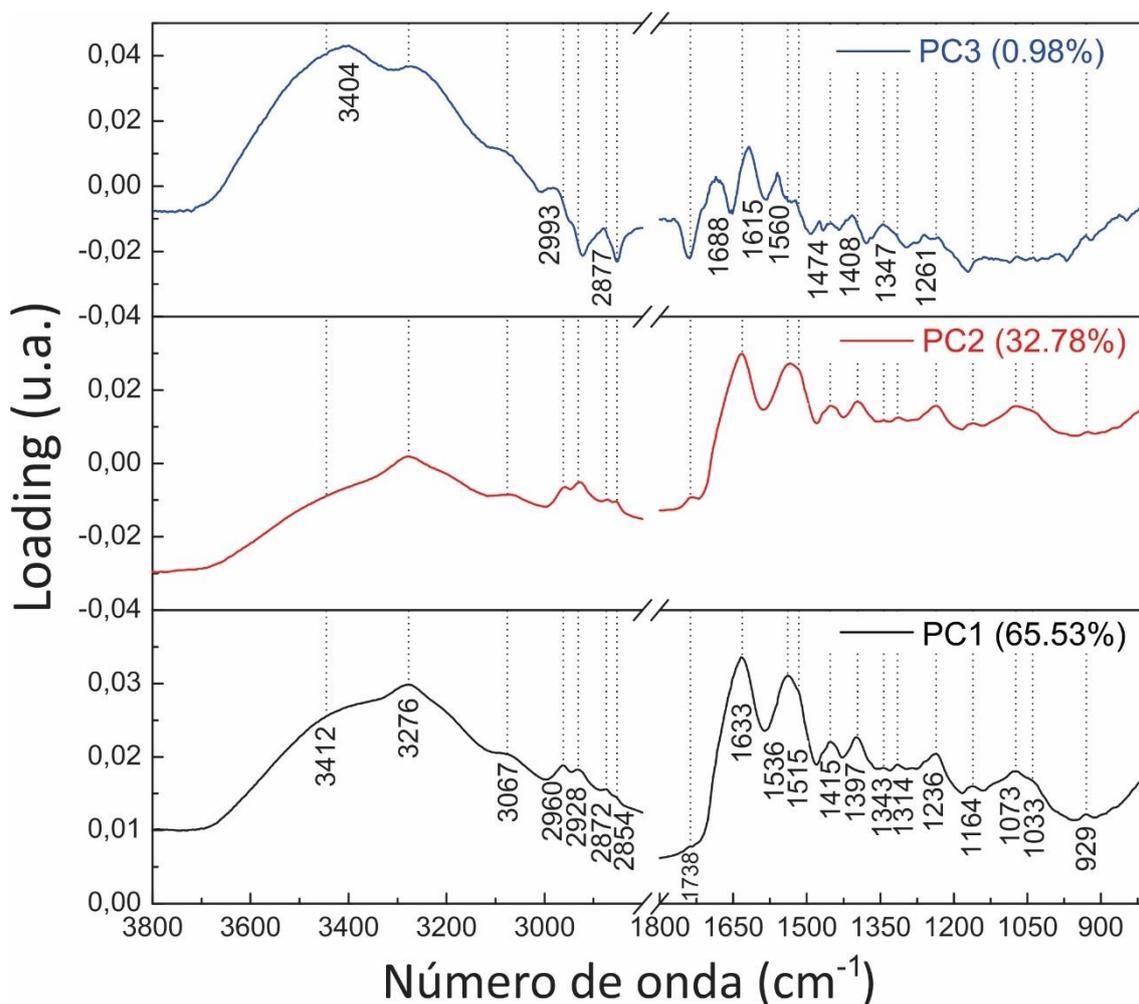


Figura 6: Loading referente a projeção de maior variabilidade entre os espectros de FTIR. Fonte: Autoria própria.

Pelas características entre os coeficientes formado através da variância entre a PC1 e a PC2 (98,31% do sistema), é possível confirmar que os espectros de FTIR entre as três classes de animais apresentam grande semelhança entre si. Todavia, os índices da PC3 indicam que os dados possuem variação entre os picos principais (observar as linhas pontilhadas da Figura 6) demonstrando que algumas informações sutis, como deslocamento de bandas ou atenuações, podem estar sendo obstruídas devido a intensidade das bandas principais do espectro de FTIR.

Assim, analisando os coeficientes da 3ª PC, pode-se destacar que diversos picos estão contribuindo para a diferenciação das classes, entre eles, o pico 3421 cm^{-1} das vibrações O-H e N-H referentes ao grupo amida das proteínas A e B [28, 57], que sofrem um deslocamento para a região de 3404 cm^{-1} ; o pico 2960 cm^{-1} e 2872 cm^{-1} referentes as gorduras do soro e provenientes das vibrações assimétrica e simétrica dos grupos CH_3 e

CH₂ [28, 57, 63, 64, 80] respectivamente, apresentou alterações para as regiões de 2993 cm⁻¹ e 2877 cm⁻¹; o pico 1397 cm⁻¹ referente aos fibrinogênios IgG1, IgM, IgA e haptoglobinas [37, 57, 64, 80, 81] tiveram sua banda deslocada para 1408 cm⁻¹; o pico 1343 cm⁻¹ referente a Transferrina e a Glicoproteína ácida α1 [28, 36, 63, 64, 80–82] apresentou alteração para 1347 cm⁻¹.

Na região dos principais compostos biológicos (região de 1800cm⁻¹ a 1000cm⁻¹) a 3ª PC demonstrou o aparecimento de novos picos, como: 1688 cm⁻¹, 1615 cm⁻¹ e 1560 cm⁻¹ entre as bandas de 1738 cm⁻¹ a 1536 cm⁻¹, composta por vibrações entre os grupos C=O e C-N da folha β-hélice da Amida I e a flexões e torção do grupo N-H da Amida II [35–37, 57, 63, 64, 80, 82]; o pico 1474 cm⁻¹ entre as bandas 1515 cm⁻¹ e 1451 cm⁻¹, referente as vibrações simétricas C-O-O, assimétricas CH₃, e flexão entre o grupo N-H da α-hélice da Amida II, proveniente das Apolipoproteína-A1, Albumina e ao IgG4 [37, 63, 64, 80–82]; e o pico 1261 cm⁻¹ entre 1314 cm⁻¹ a 1236 cm⁻¹, referente as vibrações simétricas C-O e CH₂, e as vibrações assimétricas entre o grupos fosfato, presentes na estrutura do DNA/RNA [28, 35–37, 57, 63, 64, 80–82].

Pela análise de PCA, é possível constatar que a hipótese inicial de que as espécies reativas de oxigênio e nitrogênio, além dos anticorpos estão realmente apresentando alterações[12, 84–86], e estas não são perceptíveis ao olho nu, mas são visíveis em uma análise matemática como no caso da PCA.

Analisando tridimensionalmente o comportamento dos espectros de FTIR para estas três componentes principais (Figura 7), é possível observar que o eixo que melhor diferencia as classes é dado pela 3ª PC, confirmando que há deslocamento nas bandas dos espectros ou atenuações que não são observáveis a olho nu, mas que a análise de PCA permite discriminar. Devido a esta observação, as contribuições das 4ª, 5ª e 6ª PC são apresentadas no APÊNDICE A, reforçando a hipótese de que as menores contribuições do sistema estão sendo fundamentais para melhor discriminação das classes de animais.

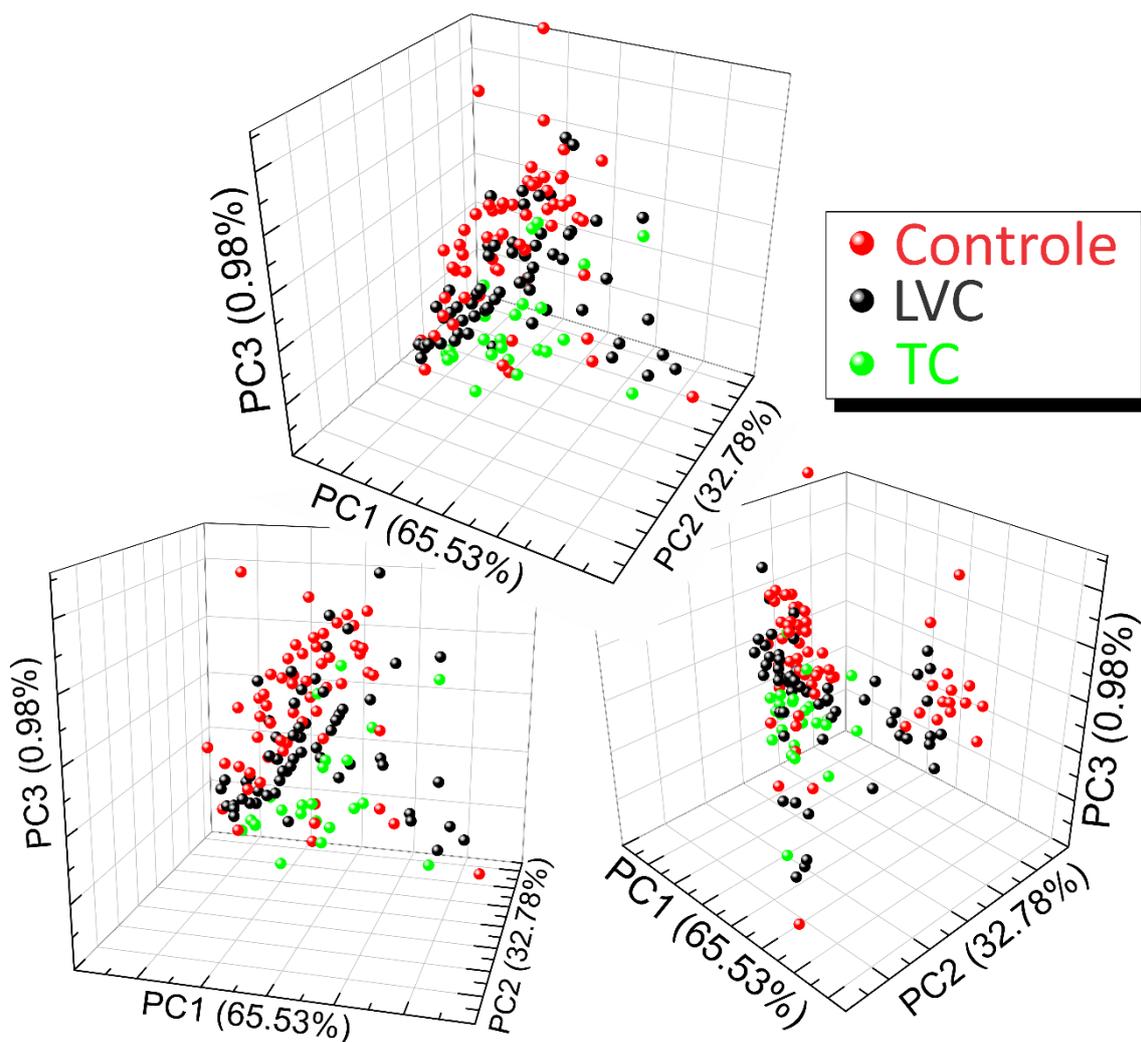


Figura 7: Representação tridimensional das componentes principais para as três classes de animais. Fonte: Autoria própria.

Na Figura 7 destaca o comportamento das informações contidas no espectro de FTIR para cada uma das amostras, em um intervalo de confiança de 99,29%, conforme identificado na análise da Figura 6, é possível confirmar que o plano formado pela 1ª e a 2ª PC apresentam grandes semelhanças, o que impossibilita a diferenciação dos grupos somente estas duas componentes. No entanto a 3ª PC representou uma variação de 0,98% dos dados, todavia os resultados observados por esta componente indicam que os deslocamentos dos picos principais estão contribuindo para a diferenciação dos grupos.

Isso demonstra que neste caso, as componentes de maior variação, como a PC1 e PC2 na verdade, estão atrapalhando a diferenciação, visto que elas estão indicando que os dados em sua maioria apresentam o mesmo comportamento. Com isso, a melhor diferenciação será dada pela variância das menores componentes principais.

Desta forma, o tratamento matemático proposto pelo PCA permitiu identificar as principais alterações em uma análise tridimensional dos espectros de FTIR, onde foram observados que os picos 3412 cm^{-1} , 2960 cm^{-1} , 2872 cm^{-1} , 1397 cm^{-1} , 1343 cm^{-1} e as regiões entre 1738 cm^{-1} - 1536 cm^{-1} , 1515 cm^{-1} - 1415 cm^{-1} e 1314 cm^{-1} - 1236 cm^{-1} , estão promovendo as diferenciações entre as classes. Sendo estas bandas representadas pelas proteínas, anticorpos (IgG, IgM e IgA) e o DNA/RNA presentes no soro. Todavia, mais estudos devem ser realizados a fim de confirmar as hipóteses aqui apresentadas.

5.3 Análise Multivariada: Método supervisionado pela técnica de Máquina de Suporte de Vetores

O aprendizado supervisionado parte em deduzir uma classe de dados a partir de um conjunto de treinamentos rotulados. De modo que o sistema se ajuste para prever a classe de dados desconhecidos ou conjunto de teste. Neste trabalho, este método foi empregado para a dedução de duas doenças, a LVC e a TC, criando assim uma nova proposta para o diagnóstico destas infecções.

Como o algoritmo mais eficiente para uma tarefa de classificação não pode ser discernido a priori, foram avaliados os seguintes classificadores: vizinhos mais próximos (KNN), SVM, análise de discriminante linear (LDA), análise de discriminante quadrática (QDA), florestas aleatórias (RF) e os classificador de conjunto (EC), com diversas funções Kernel associadas.

Dado a alta dimensionalidade dos dados espectrais provenientes do FTIR, foi gerado recursos a partir das representações de dimensão reduzida pela PCA, no intuito de diminuir a complexidade dos sistemas. Para isso, foi mapeado o número de PC que melhor se representava cada modelo supervisionado.

Todas as 144 amostras foram utilizadas no processo de treinamento e a validação destes modelos foi desenvolvida por meio do método de validação cruzada, dividindo o conjunto de amostras em 10 e 15 dobras. Dentre todos os sistemas de aprendizagem treinados neste estudo (Apêndice B), os que apresentaram maior acurácia foram os modelos SVM com função kernel quadrática, a matriz de confusão apresentada na Figura 8 demonstra o comportamento deste algoritmo perante as três classes de indivíduos (LVC, TC e Controle).

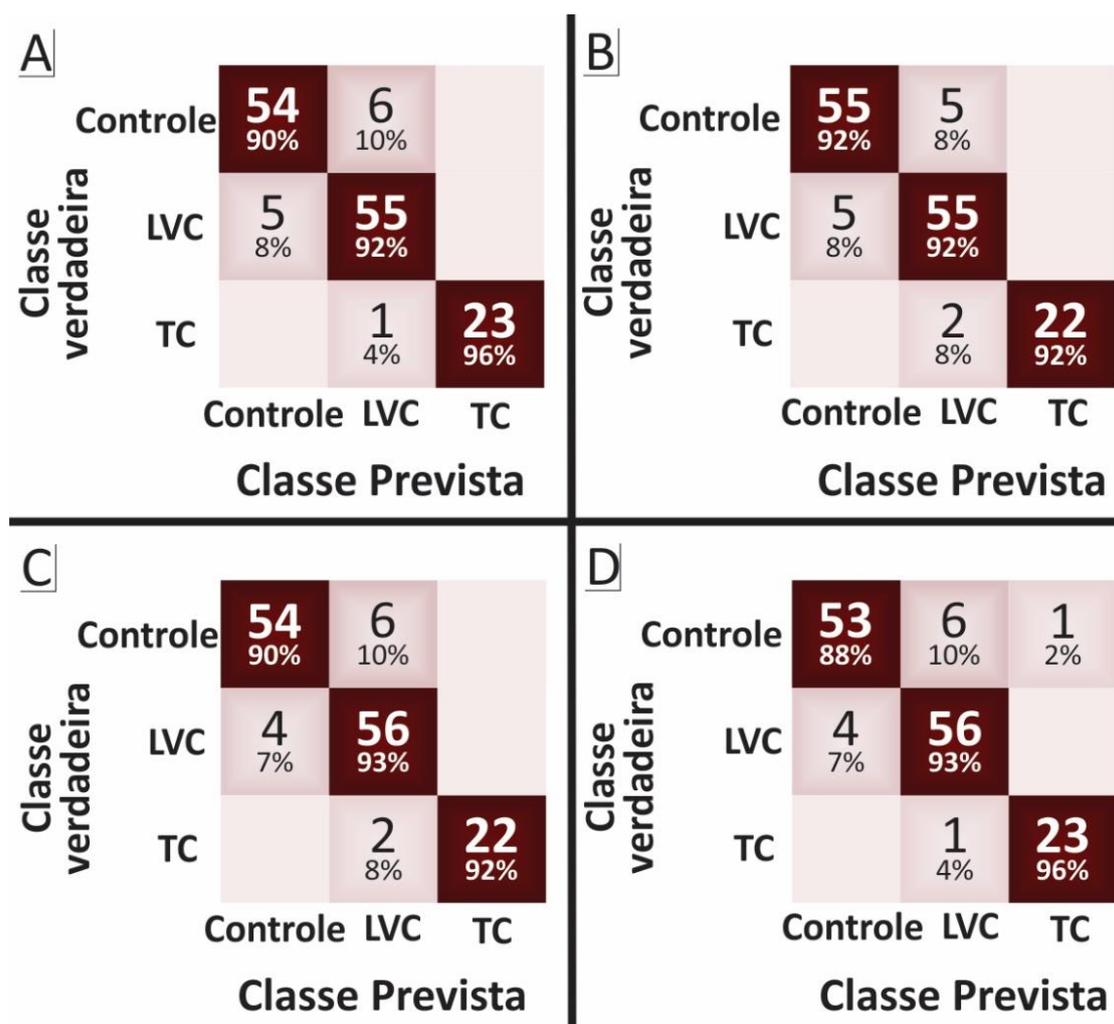


Figura 8: Matriz de confusão para o sistema de aprendizagem de máquina pelo método SVM com função Kernel Quadrática.

Legenda: A) Modelo treinado com 10 dobras +22PC; B) Modelo com Modelo treinado com 10 dobras +23PC; C) Modelo treinado com 15 dobras +20PC; D) Modelo treinado com 15 dobras +22PC. Fonte: Autoria própria.

O modelo exposto na Figura 8-A, obteve sensibilidade de 92% para a LVC e 96% para o TC, a especificidade para este modelo foi de 92%. Enquanto o modelo da Figura 8-B, obteve sensibilidade e especificidade de 92% para ambas as classes. O modelo da Figura 8-C obteve sensibilidade de 93% para LVC e 92% para TC, a especificidade alcançada neste modelo foi de 90%. E por fim, o modelo apresentado na Figura 8-D, obteve sensibilidade de 93% para LVC e 96% para TC, com especificidade de 88%.

Em relação aos falsos negativos a LVC teve uma taxa de 7 a 8% de confusão com os animais do grupo controle; para a TC, a taxa de confusão com a LVC foi de 4 a 8%, a

respeito dos falsos positivos a taxa foi de 8 a 10%. Isso indica que os animais do grupo controle poderiam estar infectados por LVC em ambos os modelos, enquanto que a taxa de confusão entre infectados por TC, foi de apenas 2% no modelo apresentado pela Figura 8-D.

Ao Comparar estes resultados com os procedimentos padrões de diagnóstico destas doenças, o método proposto neste trabalho se equipara as técnicas ELISA ou TR [9, 20, 88]. Em relação ao tempo de realização deste procedimento, é necessário 90 min para a preparação, deposição e secagem do soro no substrato de SiO₂, sem necessidade de nenhum produto para insumo, ou seja, o soro após seco, pode ser analisado diretamente pelo espectrômetro de FTIR.

O custo unitário para este procedimento diagnóstico pode ser estimado pelo valor do equipamento (\approx 150 mil reais) somado ao custo de manutenção da lâmpada de IR (\approx 10 mil reais), dividido pelo tempo de vida da lâmpada (2500 horas) e pela quantidade de medições possíveis dentro de 1 hora seguindo o protocolo proposto aqui proposto (60 medições), o que daria em média um valor sem lucros de R\$ 1,07 (um real e sete centavos).

Representando um valor inferior aos gastos cometidos por instituições públicas nos exames tradicionais, onde, para cada reação de testes de Ensaio Imunoenzimático são gastos R\$ 3,15 (três reais e quinze centavos) e para cada reação de Imunofluorescência Indireta é gasto R\$ 1,50 (um real e cinquenta centavos) [89]. Já em relação aos valores cobrados por entidades particulares, que visam lucros comerciais, a média de um TR unitário é de R\$ 31,59 (trinta e um reais e cinquenta e nove centavos) [90], enquanto um teste ELISA junto a Imunofluorescência Indireta é cobrado valores de 75,00 a 100,00 reais [91].

6. Conclusão

A Leishmaniose é uma das principais doenças negligenciadas de alto risco a saúde pública, seus métodos de diagnósticos tradicionais são de baixa sensibilidade e especificidade, além de apresentarem confusão com diversas doenças, principalmente as da mesma família, como a Tripanossomíase. Visando solucionar estes problemas, foi aplicado a técnica de espectroscopia de FTIR junto as técnicas de AM, no intuito de identificar e diferenciar animais com LVC de animais com TC e grupo controle, através da análise do soro sanguíneo.

A técnica de PCA indicou alterações nos picos do espectro de FTIR provenientes as proteínas, anticorpos e DNA/RNA entre as classes de animais. Enquanto a técnica SVM com função kernel quadrática, permitiu o desenvolvimento de modelos de aprendizagem capazes de identificar estas doenças com acurácia de 91,7%, especificidade de 92%, sensibilidade de 93% e 96% para LVC e TC, respectivamente.

Portanto, concluiu-se que o uso da espectroscopia de FTIR junto aos métodos de AM, apresentam vantagens em relação a sensibilidade, especificidade, custo de operação, facilidade de uso e uma rápida alternativa para diagnosticar a LVC e TC.

7. Referencias

- [1] WHO/PAHO, “Leishmaniasis: Epidemiological Report of the Americas N° 7 - Março, 2019,” *Inf. Leishmanioses N° 7 - Março, 2019*, vol. 1, pp. 1–27, 2019.
- [2] S. C. Ilgoutz and M. J. McConville, “Function and assembly of the *Leishmania* surface coat,” *Int. J. Parasitol.*, vol. 31, no. 9, pp. 899–908, Jul. 2001.
- [3] L. Firmino-Cruz *et al.*, “Immunomodulating role of IL-10-producing B cells in *Leishmania amazonensis* infection,” *Cell. Immunol.*, vol. 334, pp. 20–30, Dec. 2018.
- [4] S. Burza, S. L. Croft, and M. Boelaert, “Leishmaniasis,” *Lancet*, vol. 392, no. 10151, pp. 951–970, Sep. 2018.
- [5] T. A. R. Cordeiro, M. V. C. Gonçalves, D. L. Franco, A. B. Reis, H. R. Martins, and L. F. Ferreira, “Label-free electrochemical impedance immunosensor based on modified screen-printed gold electrodes for the diagnosis of canine visceral leishmaniasis,” *Talanta*, vol. 195, pp. 327–332, Apr. 2019.
- [6] M. Gramiccia and L. Gradoni, “The current status of zoonotic leishmaniasis and approaches to disease control,” *Int. J. Parasitol.*, vol. 35, no. 11–12, pp. 1169–1180, Oct. 2005.
- [7] R. Reithinger and J.-C. Dujardin, “Molecular Diagnosis of Leishmaniasis: Current Status and Future Applications,” *J. Clin. Microbiol.*, vol. 45, no. 1, pp. 21–25, Jan. 2007.
- [8] G. S. V. Tavares *et al.*, “A Pluronic® F127-based polymeric micelle system containing an antileishmanial molecule is immunotherapeutic and effective in the treatment against *Leishmania amazonensis* infection,” *Parasitol. Int.*, vol. 68, no. 1, pp. 63–72, Feb. 2019.
- [9] F. Vega-López, “Diagnosis of cutaneous leishmaniasis,” *Curr. Opin. Infect. Dis.*, vol. 16, no. 2, 2003.
- [10] S. Sundar, S. G. Reed, V. P. Singh, P. C. Kumar, and H. W. Murray, “Rapid accurate field diagnosis of Indian visceral leishmaniasis,” *Lancet*, vol. 351, no. 9102, pp. 563–565, Feb. 1998.
- [11] A. Gomes-Silva *et al.*, “Serological reactivity of different antigenic preparations of *Leishmania (Leishmania) amazonensis* and the *Leishmania braziliensis* complex,” *Rev. Soc. Bras. Med. Trop.*, vol. 41, no. 2, pp. 135–141, Apr. 2008.

- [12] H. W. Murray, J. D. Berman, C. R. Davies, and N. G. Saravia, "Advances in leishmaniasis," *Lancet*, vol. 366, no. 9496, pp. 1561–1577, Oct. 2005.
- [13] C. Fargeas, M. Hommel, R. Maingon, C. Dourado, M. Monsigny, and R. Mayer, "Synthetic peptide-based enzyme-linked immunosorbent assay for serodiagnosis of visceral leishmaniasis," *J. Clin. Microbiol.*, vol. 34, no. 2, pp. 241–248, 1996.
- [14] Â. C. Volpini, V. M. . Passos, G. C. Oliveira, and A. J. Romanha, "PCR-RFLP to identify *Leishmania* (*Viannia*) *braziliensis* and *L. (Leishmania) amazonensis* causing American cutaneous leishmaniasis," *Acta Trop.*, vol. 90, no. 1, pp. 31–37, Mar. 2004.
- [15] A. Nasereddin, E. Bensoussan-Hermano, G. Schonian, G. Baneth, and C. L. Jaffe, "Molecular Diagnosis of Old World Cutaneous Leishmaniasis and Species Identification by Use of a Reverse Line Blot Hybridization Assay," *J. Clin. Microbiol.*, vol. 46, no. 9, pp. 2848–2855, Sep. 2008.
- [16] Z. J. Attar *et al.*, "Latex agglutination test for the detection of urinary antigens in visceral leishmaniasis," *Acta Trop.*, vol. 78, no. 1, pp. 11–16, Jan. 2001.
- [17] A. C. Perinoto *et al.*, "Biosensors for Efficient Diagnosis of Leishmaniasis: Innovations in Bioanalytics for a Neglected Disease," *Anal. Chem.*, vol. 82, no. 23, pp. 9763–9768, Dec. 2010.
- [18] S. Sundar and M. Rai, "Laboratory Diagnosis of Visceral Leishmaniasis," *Clin. Vaccine Immunol.*, vol. 9, no. 5, pp. 951–958, Sep. 2002.
- [19] R. M. Luciano, S. B. Lucheis, M. Z. Troncarelli, D. M. Luciano, and H. Langoni, "Avaliação da reatividade cruzada entre antígenos de *Leishmania* spp e *Trypanosoma cruzi* na resposta sorológica de cães pela técnica de imunofluorescência indireta (RIFI)," *Brazilian J. Vet. Res. Anim. Sci.*, vol. 46, no. 3, p. 181, Jun. 2009.
- [20] S. Mohan, P. Srivastava, S. N. Maheshwari, S. Sundar, and R. Prakash, "Nano-structured nickel oxide based DNA biosensor for detection of visceral leishmaniasis (Kala-azar)," *Analyst*, vol. 136, no. 13, p. 2845, 2011.
- [21] M. Moradi, N. Sattarahmady, A. Rahi, G. R. Hatam, S. M. R. Sorkhabadi, and H. Heli, "A label-free, PCR-free and signal-on electrochemical DNA biosensor for *Leishmania major* based on gold nanoleaves," *Talanta*, vol. 161, pp. 48–53, Dec. 2016.
- [22] N. Sattarahmady, A. Movahedpour, H. Heli, and G. R. Hatam, "Gold

- nanoparticles-based biosensing of *Leishmania major* kDNA genome: Visual and spectrophotometric detections,” *Sensors Actuators B Chem.*, vol. 235, pp. 723–731, Nov. 2016.
- [23] M. F. K. do. S. Garcia *et al.*, “Impedimetric sensor for *Leishmania infantum* genome based on gold nanoparticles dispersed in polyaniline matrix,” *J. Chem. Technol. Biotechnol.*, vol. 91, no. 11, pp. 2810–2816, 2016.
- [24] D. E. P. Souto, J. Volpe, C. de C. Gonçalves, C. H. I. Ramos, and L. T. Kubota, “A brief review on the strategy of developing SPR-based biosensors for application to the diagnosis of neglected tropical diseases,” *Talanta*, vol. 205, p. 120122, Dec. 2019.
- [25] L. Li *et al.*, “Characterization of ovarian cancer cells and tissues by Fourier transform infrared spectroscopy,” *J. Ovarian Res.*, vol. 11, no. 1, p. 64, Dec. 2018.
- [26] M. M. Grzelak *et al.*, “Diagnosis of ovarian tumour tissues by SR-FTIR spectroscopy: A pilot study,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 203, pp. 48–55, Oct. 2018.
- [27] W. Yi, D. Cui, Z. Li, L. Wu, A. Shen, and J. Hu, “Gastric cancer differentiation using Fourier transform near-infrared spectroscopy with unsupervised pattern recognition,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 101, pp. 127–131, Jan. 2013.
- [28] F. Elmi, A. F. Movaghar, M. M. Elmi, H. Alinezhad, and N. Nikbakhsh, “Application of FT-IR spectroscopy on breast cancer serum analysis,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 187, pp. 87–91, Dec. 2017.
- [29] H. Liu, Q. Su, D. Sheng, W. Zheng, and X. Wang, “Comparison of red blood cells from gastric cancer patients and healthy persons using FTIR spectroscopy,” *J. Mol. Struct.*, vol. 1130, pp. 33–37, Feb. 2017.
- [30] J. Backhaus *et al.*, “Diagnosis of breast cancer with infrared spectroscopy from serum samples,” *Vib. Spectrosc.*, vol. 52, no. 2, pp. 173–177, Mar. 2010.
- [31] J. Depciuch *et al.*, “Monitoring breast cancer treatment using a Fourier transform infrared spectroscopy-based computational model,” *J. Pharm. Biomed. Anal.*, vol. 143, pp. 261–268, Sep. 2017.
- [32] L. Dong *et al.*, “Evaluation of FTIR spectroscopy as diagnostic tool for colorectal cancer using spectral analysis,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 122, pp. 288–294, Mar. 2014.

- [33] L. F. S. Siqueira, R. F. Araújo Júnior, A. A. de Araújo, C. L. M. Morais, and K. M. G. Lima, "LDA vs. QDA for FT-MIR prostate cancer tissue classification," *Chemom. Intell. Lab. Syst.*, vol. 162, pp. 123–129, Mar. 2017.
- [34] E. Kaznowska, J. Depciuch, K. Szmuc, and J. Cebulski, "Use of FTIR spectroscopy and PCA-LDC analysis to identify cancerous lesions within the human colon," *J. Pharm. Biomed. Anal.*, vol. 134, pp. 259–268, Feb. 2017.
- [35] M. C. D. Santos *et al.*, "ATR-FTIR spectroscopy with chemometric algorithms of multivariate classification in the discrimination between healthy vs. dengue vs. chikungunya vs. zika clinical samples," *Anal. Methods*, vol. 10, no. 10, pp. 1280–1285, 2018.
- [36] M. C. D. Santos, Y. M. Nascimento, J. M. G. Araújo, and K. M. G. Lima, "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.*, vol. 7, no. 41, pp. 25640–25649, 2017.
- [37] S. Mordechai, E. Shufan, B. S. Porat Katz, and A. Salman, "Early diagnosis of Alzheimer's disease using infrared spectroscopy of isolated blood samples followed by multivariate analyses," *Analyst*, vol. 142, no. 8, pp. 1276–1284, 2017.
- [38] L. B. Leal, M. S. Nogueira, R. A. Canevari, and L. F. C. S. Carvalho, "Vibration spectroscopy and body biofluids: Literature review for clinical applications," *Photodiagnosis Photodyn. Ther.*, vol. 24, no. February, pp. 237–244, Dec. 2018.
- [39] A. Zlotogorski-Hurvitz, B. Z. Dekel, D. Malonek, R. Yahalom, and M. Vered, "FTIR-based spectrum of salivary exosomes coupled with computational-aided discriminating analysis in the diagnosis of oral cancer," *J. Cancer Res. Clin. Oncol.*, vol. 145, no. 3, pp. 685–694, Mar. 2019.
- [40] R. Chaber *et al.*, "Predicting Ewing Sarcoma Treatment Outcome Using Infrared Spectroscopy and Machine Learning," *Molecules*, vol. 24, no. 6, p. 1075, Mar. 2019.
- [41] R. Chaber *et al.*, "Distinguishing Ewing sarcoma and osteomyelitis using FTIR spectroscopy," *Sci. Rep.*, vol. 8, no. 1, p. 15081, Dec. 2018.
- [42] X. Yang *et al.*, "Pre-diabetes diagnosis based on ATR-FTIR spectroscopy combined with CART and XGBoots," *Optik (Stuttg.)*, vol. 180, pp. 189–198, Feb. 2019.
- [43] J. C. Aguiar, J. Mittmann, P. C. Caetano, and L. Raniero, "Using FT-IR

- spectroscopy for the identification of the *T. cruzi*, *T. rangeli*, and the *L. chagasi* species,” *Exp. Parasitol.*, vol. 192, pp. 46–51, Sep. 2018.
- [44] A. G. D. S. PINTO, “CARACTERIZAÇÃO DE FLAGELADOS DO GÊNERO *Trypanosoma*, GRUBY 1843, EM CÃES DOMÉSTICOS (*Canis familiares*) NO MUNICÍPIO DO RIO DE JANEIRO,” Fundação Oswaldo Cruz, 2009.
- [45] M. I. L. D. M. MARQUES, “Leishmaniose Canina,” UNIVERSIDADE TÉCNICA DE LISBOA, 2008.
- [46] M. da S. (MS), *Guia de vigilância em saúde*, 1st ed., vol. 3. Brasília, DF, 2017.
- [47] M. da S. (MS), “Leishmaniose Visceral: o que é, causas, sintomas, tratamento, diagnóstico e prevenção,” 2019. [Online]. Available: <http://portalms.saude.gov.br/saude-de-a-z/leishmaniose-visceral>. [Accessed: 29-Mar-2020].
- [48] Organización Panamericana de la Salud, *Manual de procedimientos para la vigilancia y control de las leishmaniasis*. Washington, D.C, 2019.
- [49] L. Solano-Gallego *et al.*, “Directions for the diagnosis, clinical staging, treatment and prevention of canine leishmaniosis,” *Vet. Parasitol.*, vol. 165, no. 1–2, pp. 1–18, 2009.
- [50] C. O. S. Barbosa, “Desempenho do teste imunocromatográfico rápido DPP - Dual Path Platform para o diagnóstico da leishmaniose visceral canina e reação cruzada com hemoparasitos,” Universidade Federal de Uberlândia, 2015.
- [51] L. Galluzzi, M. Ceccarelli, A. Diotallevi, M. Menotta, and M. Magnani, “Real-time PCR applications for diagnosis of leishmaniasis,” *Parasit. Vectors*, vol. 11, no. 1, p. 273, Dec. 2018.
- [52] H. Goto and J. A. L. Lindoso, “Current diagnosis and treatment of cutaneous and mucocutaneous leishmaniasis,” *Expert Rev. Anti. Infect. Ther.*, vol. 8, no. 4, pp. 419–433, Apr. 2010.
- [53] D. E. P. Souto *et al.*, “Development of a label-free immunosensor based on surface plasmon resonance technique for the detection of anti-*Leishmania infantum* antibodies in canine serum,” *Biosens. Bioelectron.*, vol. 46, pp. 22–29, Aug. 2013.
- [54] M. S. Mehand, F. Al-Shorbaji, P. Millett, and B. Murgue, “The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts,” *Antiviral Res.*, vol. 159, no. September, pp. 63–67, Nov. 2018.

- [55] R. Etzioni *et al.*, “The case for early detection,” *Nat. Rev. Cancer*, vol. 3, no. 4, pp. 243–252, Apr. 2003.
- [56] M. J. Garnett *et al.*, “Systematic identification of genomic markers of drug sensitivity in cancer cells,” *Nature*, vol. 483, no. 7391, pp. 570–575, Mar. 2012.
- [57] M. S. Pepe *et al.*, “Phases of Biomarker Development for Early Detection of Cancer,” *JNCI J. Natl. Cancer Inst.*, vol. 93, no. 14, pp. 1054–1061, Jul. 2001.
- [58] J. M. Cameron, H. J. Butler, D. S. Palmer, and M. J. Baker, “Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying,” *J. Biophotonics*, vol. 11, no. 4, p. e201700299, Apr. 2018.
- [59] M. J. Baker *et al.*, “Using Fourier transform IR spectroscopy to analyze biological materials,” *Nat. Protoc.*, vol. 9, no. 8, pp. 1771–1791, Aug. 2014.
- [60] M. J. Baker *et al.*, “Clinical applications of infrared and Raman spectroscopy: state of play and future challenges,” *Analyst*, vol. 143, no. 8, pp. 1735–1757, 2018.
- [61] K. Naseer, S. Ali, S. Mubarik, I. Hussain, B. Mirza, and J. Qazi, “FTIR spectroscopy of freeze-dried human sera as a novel approach for dengue diagnosis,” *Infrared Phys. Technol.*, vol. 102, no. May, p. 102998, Nov. 2019.
- [62] V. Notarstefano *et al.*, “Investigation of human pancreatic cancer tissues by Fourier Transform Infrared Hyperspectral Imaging,” *J. Biophotonics*, no. September, pp. 1–10, Nov. 2019.
- [63] Q. M. I. Haq *et al.*, “Application of reflectance spectroscopies (FTIR-ATR & FT-NIR) coupled with multivariate methods for robust in vivo detection of begomovirus infection in papaya leaves,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 198, pp. 27–32, Jun. 2018.
- [64] K. Gajjar *et al.*, “Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer,” *Analyst*, vol. 138, no. 14, p. 3917, 2013.
- [65] D. Sheng, X. Liu, W. Li, Y. Wang, X. Chen, and X. Wang, “Distinction of leukemia patients’ and healthy persons’ serum using FTIR spectroscopy,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 101, pp. 228–232, Jan. 2013.
- [66] X. Wang, X. Shen, D. Sheng, X. Chen, and X. Liu, “FTIR spectroscopic comparison of serum from lung cancer patients and healthy persons,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 122, pp. 193–197, Mar. 2014.

- [67] S. Raghu and N. Sriraam, "Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms," *Expert Syst. Appl.*, vol. 113, pp. 18–32, Dec. 2018.
- [68] Y. B. Dibike, S. Velickov, D. Solomatine, and M. B. Abbott, "Model Induction with Support Vector Machines: Introduction and Applications," *J. Comput. Civ. Eng.*, vol. 15, no. 3, pp. 208–216, Jul. 2001.
- [69] K. Hongyu, V. L. M. Sandanielo, and G. J. Oliveira Junior, "Principal Component Analysis: theory, interpretations and applications," *Eng. Sci.*, vol. 5, no. 1, pp. 83–90, 2015.
- [70] D. W. Johnson, Richard A. WICHERN, *Applied multivariate statistical analysis*, 6^a edition. New Jersey: Prentice-Hall: Upper Saddle River, 2007.
- [71] A. Schölkopf, B., Burges, C., and Smola, *Advances in Kernel Methods - Support Vector Learning*. London, England: MIT Press, 1999.
- [72] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, no. 2, pp. 181–201, Mar. 2001.
- [73] J. Cristianini, N. and Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 1st ed. Cambridge, UK: Cambridge University Press, 2000.
- [74] S. Geisser, "The Predictive Sample Reuse Method with Applications," *J. Am. Stat. Assoc.*, vol. 70, no. 350, pp. 320–328, Jun. 1975.
- [75] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.
- [76] M. W. Browne, "Cross-Validation Methods," *J. Math. Psychol.*, vol. 44, no. 1, pp. 108–132, Mar. 2000.
- [77] Y. Jung, "Multiple predicting K -fold cross-validation for model selection," *J. Nonparametr. Stat.*, vol. 30, no. 1, pp. 197–215, Jan. 2018.
- [78] J. T. Echeverria *et al.*, "Clinical and therapeutic aspects of an outbreak of canine trypanosomiasis," *Rev. Bras. Parasitol. Veterinária*, vol. 28, no. 2, pp. 320–324, Apr. 2019.
- [79] D. Geysen, V. Delespaux, and S. Geerts, "PCR-RFLP using Ssu-rDNA amplification as an easy method for species-specific diagnosis of Trypanosoma

- species in cattle,” *Vet. Parasitol.*, vol. 110, no. 3–4, pp. 171–180, Jan. 2003.
- [80] C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum, and R. Garnotel, “Rapid screening of classic galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma,” *Analyst*, vol. 140, no. 7, pp. 2280–2286, 2015.
- [81] S. da C. Gonçalves-de-Albuquerque *et al.*, “The Equivocal Role of Th17 Cells and Neutrophils on Immunopathogenesis of Leishmaniasis,” *Front. Immunol.*, vol. 8, Oct. 2017.
- [82] I. Mesquita *et al.*, “The impact of IL-10 dynamic modulation on host immune response against visceral leishmaniasis,” *Cytokine*, vol. 112, pp. 16–20, Dec. 2018.
- [83] V. Rodrigues, A. Cordeiro-da-Silva, M. Laforge, R. Silvestre, and J. Estaquier, “Regulation of immunity during visceral Leishmania infection,” *Parasit. Vectors*, vol. 9, no. 1, p. 118, Dec. 2016.
- [84] V. Bruet, H. Dumon, P. Bourdeau, J. Desfontis, and L. Martin, “Prospective pilot study to detect dogs with non food-induced canine atopic dermatitis using Fourier transform infrared spectroscopy,” *Vet. Dermatol.*, vol. 27, no. 5, p. 356, Oct. 2016.
- [85] C. Petibois, G. Cazorla, A. Cassaigne, and G. Déléris, “Plasma Protein Contents Determined by Fourier-Transform Infrared Spectrometry,” *Clin. Chem.*, vol. 47, no. 4, pp. 730–738, Apr. 2001.
- [86] A. C. O. Neves, P. P. Silva, C. L. M. Morais, C. G. Miranda, J. C. O. Crispim, and K. M. G. Lima, “ATR-FTIR and multivariate analysis as a screening tool for cervical cancer in women from northeast Brazil: a biospectroscopic approach,” *RSC Adv.*, vol. 6, no. 102, pp. 99648–99655, 2016.
- [87] M. J. Baker *et al.*, “Developing and understanding biofluid vibrational spectroscopy: a critical review,” *Chem. Soc. Rev.*, vol. 45, no. 7, pp. 1803–1818, 2016.
- [88] J. G. S. Oliveira *et al.*, “Polymerase chain reaction (PCR) is highly sensitive for diagnosis of mucosal leishmaniasis,” *Acta Trop.*, vol. 94, no. 1, pp. 55–59, Apr. 2005.
- [89] R. A. de Lira, “Diagnóstico da Leishmaniose Visceral Canina: Avaliação do Desempenho dos Kits EIE-Leishmaniose-VisceralCanina-Bio-Manguinhos e IFI-Leishmaniose-VisceralCanina-Bio-Manguinhos,” Fundação Oswaldo Cruz, 2005.
- [90] Prime Cirurgica, “TESTE RÁPIDO LEISHMANIOSE KIT COM 20 TESTES

- ORANGELIFE.” [Online]. Available: <https://www.cirurgicaestilo.com.br/teste-rapido-leishmaniose-kit-com-20-testes-orangelife-p12162/>. [Accessed: 23-Mar-2020].
- [91] VETPAT, “Tabela de exames,” 2020. [Online]. Available: <https://vetpat.com.br/wp-content/uploads/2019/05/TABELA-EXAMES-VETPAT-2019.pdf>. [Accessed: 26-Mar-2020].

APÊNDICE A – Análise de PCA para a 3^a, 4^a, 5^a e 6^a PC.

Pela análise de PCA, foi observado que as menores parcelas do sistema de redução de dimensionalidade estão sendo responsáveis pela maior diferenciação do sistema, a Figuras 9 apresenta o coeficiente da 1^a PC, para comparação entre as 3^a, 4^a, 5^a e 6^a PC. Já a Figura 10 demonstra o comportamento tridimensional destas componentes perante a redução de dimensionalidade dos espectros de FTIR.

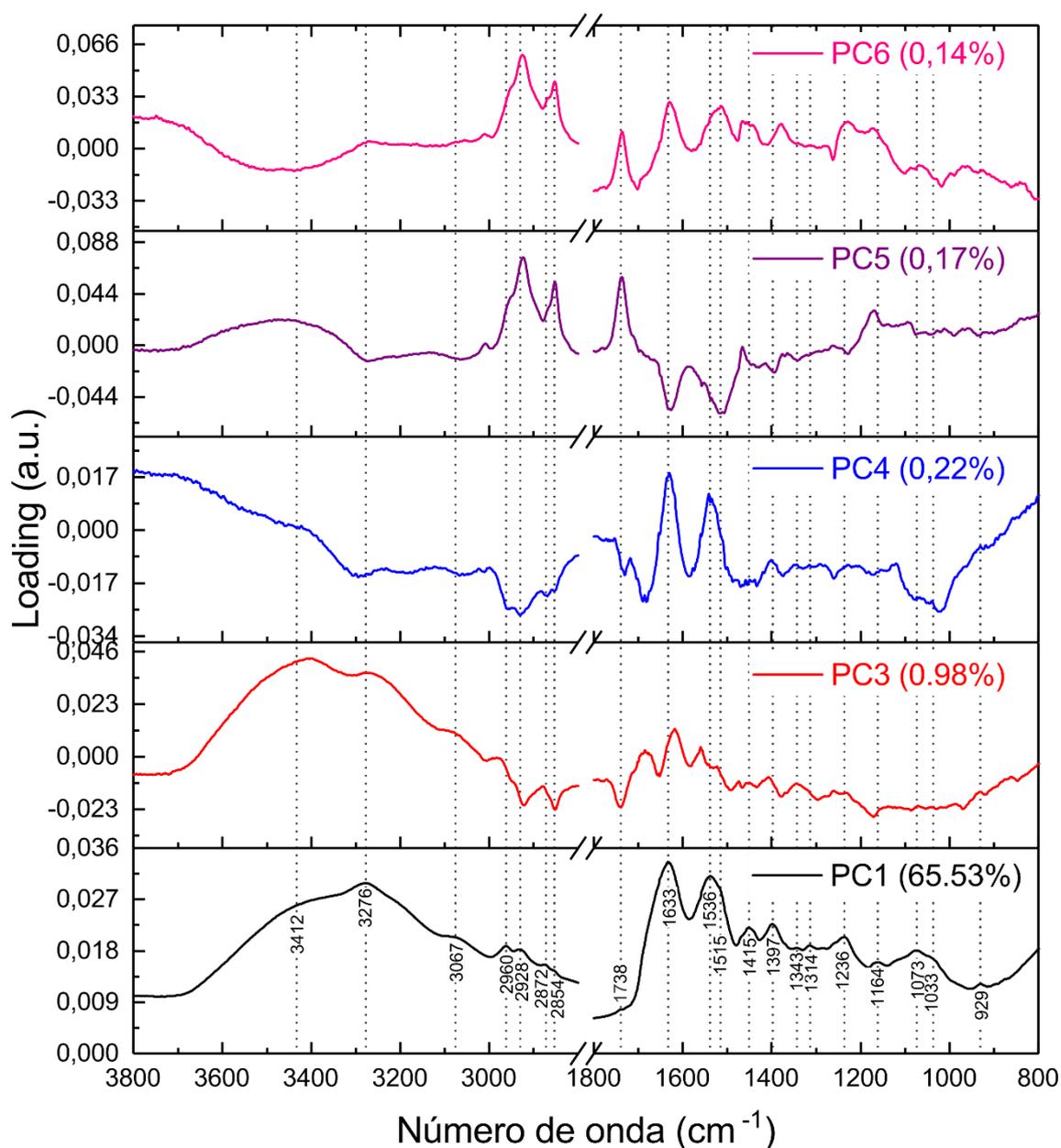


Figura 9 - Loading referente a projeção de maior variabilidade entre os espectros de FTIR para a 3^a, 4^a e 5^a PC. Fonte: Autoria própria.

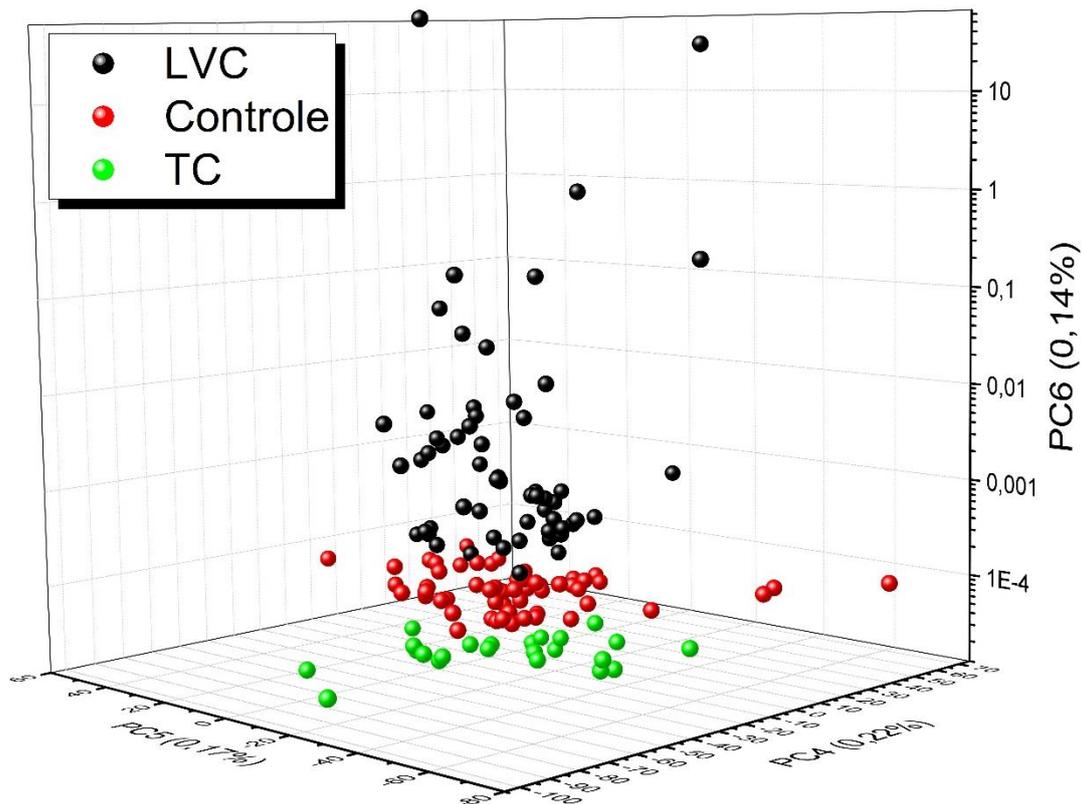


Figura 10 - Representação tridimensional da 4ª, 5ª e 6ª PC para as classes de animais. Fonte: Autoria própria.

Na sessão 5.2 é destacado que as principais contribuições dos espectros de FTIR para a separação das classes de animais, estariam contidas nas menores variações e deslocamento das bandas. Através da análise tridimensional da 4ª, 5ª e 6ª PC (Figura 10), é possível confirmar esta hipótese, de modo que as contribuições provenientes aos picos das proteínas, anticorpos e DNA estão de fato contribuindo para a discriminação entre as classes de animais.

Principalmente para a classe dos animais com LVC, onde é observado que a variação deste grupo na Figura 10 é muito maior do que os outros animais, com isso, é possível deduzir pelo espectro apresentado na PC6 que os animais com LVC apresentam uma variação extremamente superior quando comparado as outras classes. Além disso, as classes de animais foram perfeitamente discriminadas pelas PC4, PC5 e PC6, no entanto mais estudos devem ser realizados, já que estas três componentes somam apenas 0,53% da variação do sistema.

APÊNDICA B – Métodos de análise supervisionada

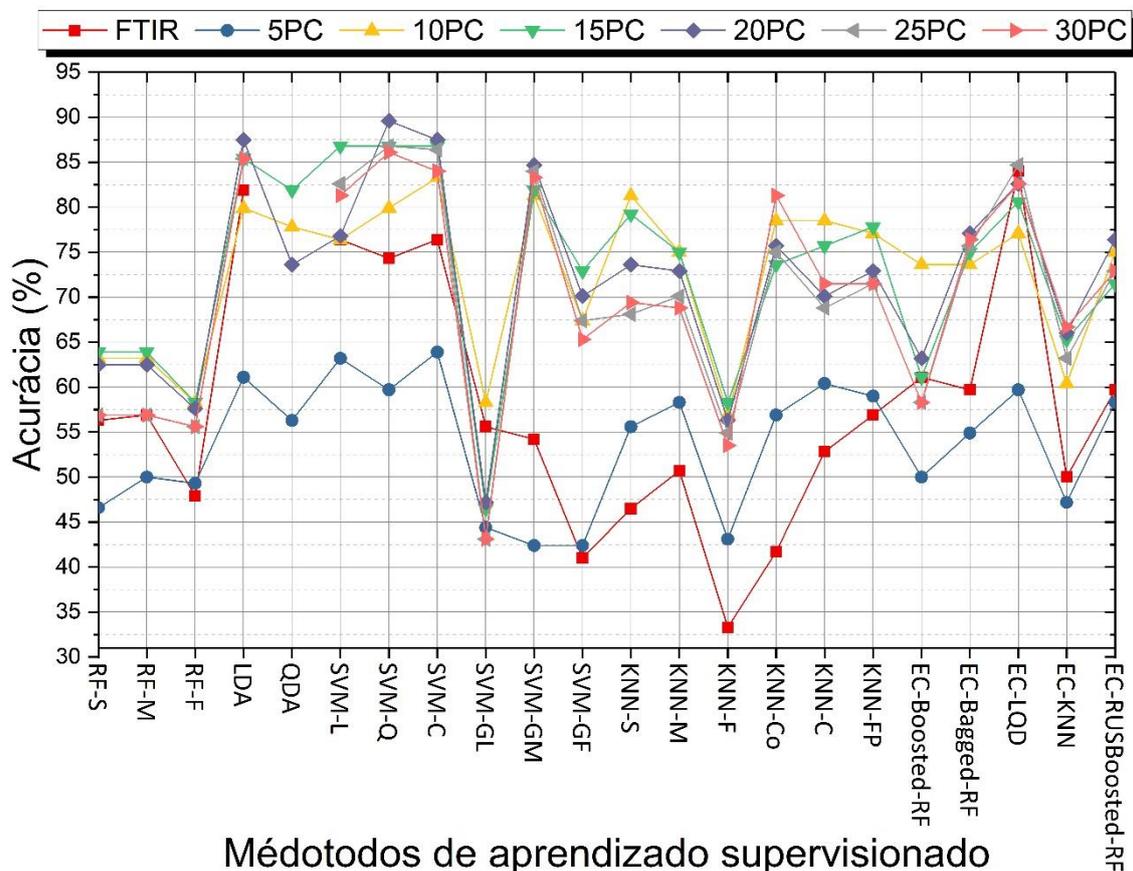


Figura 11 – Taxa de acurácia para os métodos de aprendizado supervisionado

obtidos durante o treinamento com 10 dobras por VC. Fonte: Autoria própria

Legenda: RF – Árvores aleatórias; LDA – Análise de discriminante linear; QDA – Análise de discriminante quadrática; SVM – Máquina de suporte de vetores; KNN – Vizinhos k mais próximos; EC – Aprendizado em conjunto; S – agrupamento suave; M – Agrupamento médio; F – agrupamento forçado; L – Linear; Q – Quadrático; C – Cúbico; GL – Gaussiana leve; GM – Gaussiana média; GF – Gaussiana forçada; Co – Cosseno; LQD – Discriminante linear e quadrático; - ROB – RUSBoosted.

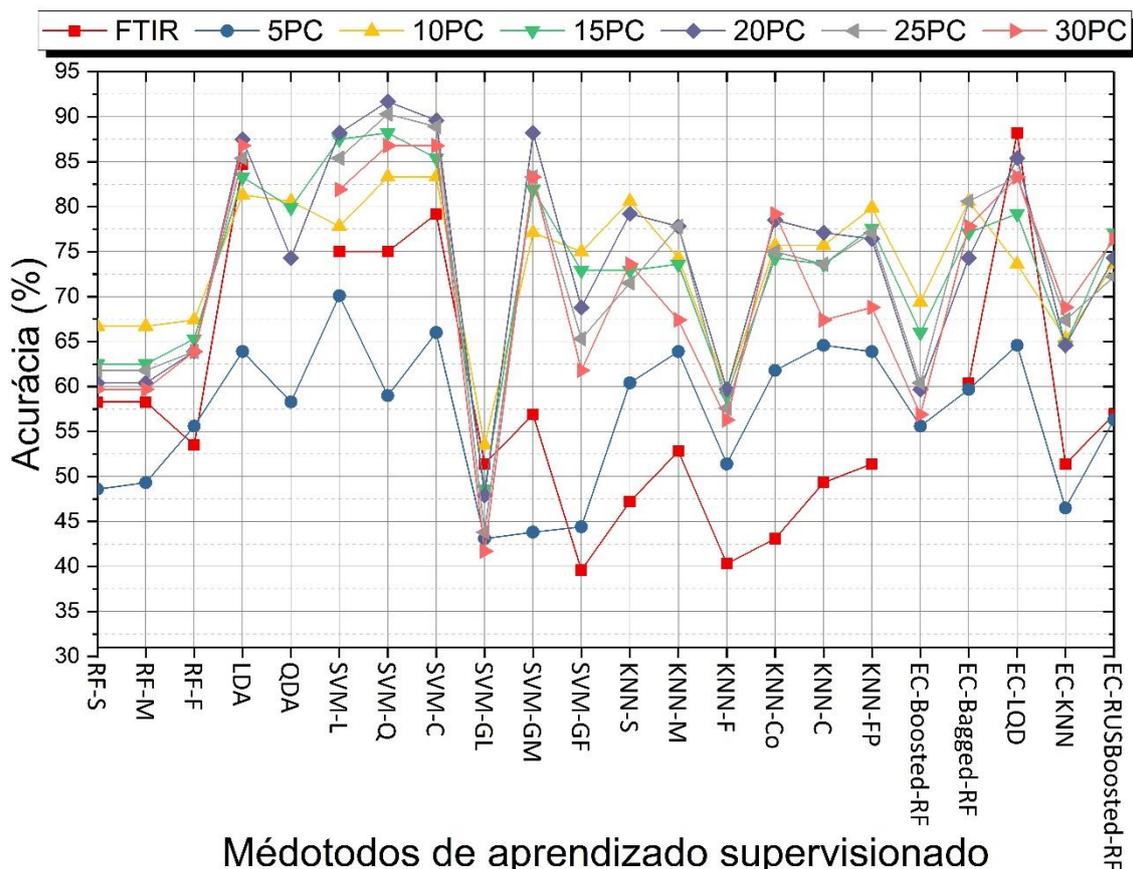


Figura 12 – Taxa de acurácia para os métodos de aprendizado supervisionado obtidos durante o treinamento com 15 dobras por VC. Fonte: Autoria própria

Legenda: RF – Árvores aleatórias; LDA – Análise de discriminante linear; QDA – Análise de discriminante quadrática; SVM – Máquina de suporte de vetores; KNN – Vizinhos k mais próximos; EC – Aprendizado em conjunto; S – agrupamento suave; M – Agrupamento médio; F – agrupamento forçado; L – Linear; Q – Quadrático; C – Cúbico; GL – Gaussiana leve; GM – Gaussiana média; GF – Gaussiana forçada; Co – Cosseno; LQD – Discriminante linear e quadrático; - ROB – RUSBoosted.