



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



Curso de ENGENHARIA FÍSICA
Trabalho de Conclusão de Curso

Fotodiagnóstico de Tuberculose Bovina por Espectroscopia no Infravermelho

Paloma Dias de Oliveira

***Orientador:** Prof. Cícero Rafael Cena da Silva*

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia Física do Instituto de Física (INFI), da Universidade Federal de Mato Grosso do Sul (UFMS).

Campo Grande -MS
Novembro/2025



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



*“Eu só quero é ser feliz
Andar tranquilamente na favela onde eu nasci, é,
E poder me orgulhar
E ter a consciência que o pobre tem seu lugar”*

(Cidinho & Doca – Rap da Felicidade)



AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, por me fortalecer e permitir que eu chegasse até aqui, com honra e saúde.

A minha família e meus avós, pela educação, respeito e valores transmitidos ao longo de toda a minha vida. Um agradecimento especial à minha mãe e à minha irmã Paola, mulheres fortes e inspiradoras, pelas lições de força, coragem e generosidade.

Ao meu namorado Bruno, expresso minha profunda gratidão pelo companheirismo, apoio e incentivo constante, principalmente nos momentos de dificuldade. Agradeço por sua paciência, por nunca ter soltado minha mão, e por acreditar nas minhas conquistas mesmo diante dos desafios. Você foi essencial e sempre esteve ao meu lado.

Ao Professor Cícero, agradeço sinceramente pela orientação, paciência e conselhos valiosos durante todos estes anos. Carregarei seus ensinamentos por toda a minha trajetória acadêmica e profissional.

Às minhas amigas-irmãs Elaine e Milena, pela amizade verdadeira, pelo apoio durante as décadas de convivência, e aos amigos Danilo Dias, Alex Augusto, Rhuan Barbosa, Caíque Porsch, Rafael Castro, Rodrigo Fioravante, Luiz Eduardo, Lucas Yuji, Alessandra Lima e Rafael de Nadai, pelo auxílio, colaboração e compreensão, especialmente nesta reta final, suportando minhas ansiedades.

Aos colegas do FotoDiag-X, Camila, Thiago e Miller, expresso minha gratidão pela prontidão em ajudar, pelas explicações e pela paciência para tornar possível o desenvolvimento deste trabalho.

A todos que, de alguma forma, contribuíram para minha caminhada, minha sincera gratidão.



APRESENTAÇÃO

Este Trabalho de Conclusão de Curso foi desenvolvido no formato de artigo científico, com o objetivo de proporcionar ao estudante uma experiência prática de produção e comunicação do conhecimento acadêmico, em consonância com a estrutura usualmente exigida por periódicos científicos. Em vez de seguir integralmente o modelo tradicional de TCC, adotou-se a lógica de manuscrito, privilegiando a objetividade, a síntese e a organização textual típica de artigos submetidos à publicação.

Dessa forma, o trabalho não apresenta, de forma isolada e extensiva, capítulos convencionais de Introdução e Revisão Bibliográfica, como ocorre em monografias clássicas, uma vez que esses conteúdos são integrados de maneira sintética nas seções textuais próprias do artigo. A opção por essa organização visa aproximar o discente das práticas reais de escrita científica, permitindo o desenvolvimento de competências relacionadas à elaboração de título, resumo, metodologia, apresentação de resultados, discussão, conclusão e referências, conforme orientações correntes sobre estrutura de artigos científicos.



Fotodiagnóstico de Tuberculose Bovina por Espectroscopia no Infravermelho

Paloma Oliveira, Caíque Porsch, Camila Calvani, Thiago Franca e
Cícero Cena

paloma.oliveira@ufms.br

FotoDiag-X, Universidade Federal de Mato Grosso do Sul

RESUMO

A tuberculose bovina é uma zoonose de grande impacto causada por *Mycobacterium bovis*, que traz prejuízos econômicos e riscos à saúde pública, agravados pelas limitações dos métodos diagnósticos tradicionais, frequentemente invasivos e de baixa acessibilidade. Este estudo avaliou o potencial da espectroscopia no infravermelho por transformada de Fourier (FTIR) aliada ao pré-processamento SNV e, em seguida, submetidos à redução de dimensionalidade por PCA e à classificação por SVM para distinguir entre soros bovinos infectados e controles. As 80 amostras foram analisadas, e a abordagem revelou assinaturas espectrais capazes de discriminar os grupos. O modelo SVM linear apresentou validação cruzada (LOOCV) com acurácia de 70%, sensibilidade de 80% para os infectados e especificidade de 60% para os controles, destacando a viabilidade da técnica em contextos com infraestrutura limitada. Conclui-se que a integração entre FTIR e aprendizado de máquina permite diagnósticos mais eficientes e acessíveis, auxiliando o controle da tuberculose bovina.

Palavras-chave: tuberculose bovina, FTIR, aprendizado de máquina, análise multivariada.



ABSTRACT

Bovine tuberculosis is a significant zoonosis caused by *Mycobacterium bovis*, resulting in economic losses and public health risks, accentuated by the limitations of traditional diagnostic methods, which are often invasive and difficult to access. This study assessed the potential of Fourier Transform Infrared (FTIR) spectroscopy, combined with SNV preprocessing, PCA, and SVM, to distinguish between sera from infected and control cattle. Eighty samples were analyzed, and the approach revealed spectral signatures capable of differentiating the groups. The linear SVM model achieved leave-one-out cross-validation accuracy of 70%, with 80% sensitivity for infected animals and 60% specificity for controls, underscoring the method's feasibility in resource-limited settings. The integration of FTIR and machine learning enables more efficient and accessible diagnostics, aiding efforts to control bovine tuberculosis.

Keywords: bovine tuberculosis, FTIR, machine learning, multivariate analysis.

INTRODUÇÃO

A tuberculose bovina é uma zoonose crônica, causada pela bactéria *Mycobacterium bovis*, responsável por impactos significativos no rebanho bovino e risco direto à economia e à saúde humana, especialmente pelo consumo de produtos lácteos não pasteurizados e seus derivados, ou pelo contato com animais infectados [1-2]. O principal desafio para sua erradicação reside na presença contínua de animais infectados não detectados o que dificulta o controle da doença e evidencia a necessidade de métodos diagnósticos mais eficazes e rápidos [2].

No Brasil e em diversos outros países, o controle da tuberculose bovina enfrenta obstáculos significativos devido às limitações dos métodos diagnósticos convencionais. Entre eles, destacam-se os testes intradérmicos - simples (STI), da prega caudal (CFT) e o comparativo cervical (SICCT), que, embora dotados de alta sensibilidade, são invasivos, possuem especificidade restrita e podem gerar falso-negativos [2]. Ademais, dependem de infraestrutura laboratorial especializada, resultando em diagnósticos tardios e prejuízos produtivos, especialmente em regiões



extensas com recursos limitados. Neste contexto, surge uma lacuna por técnicas diagnósticas rápidas, precisas e de baixo custo [3].

Pesquisas indicam que a espectroscopia no Infravermelho por Transformada de Fourier (FTIR) destaca-se como técnica analítica capaz de identificar compostos químicos e investigar os modos vibracionais das ligações moleculares, por meio da absorção da radiação infravermelha [4,6,7,8]. Por meio desta abordagem, é possível realizar uma análise eficiente, rápida e minimamente invasiva, fornecendo informações diagnósticas sob a forma de “assinaturas espectrais” de amostras biológicas [5,6,9], dispensando o uso de reagentes e procedimentos potencialmente destrutivos, a depender da amostra. Dada a sobreposição de bandas e a alta dimensionalidade dos espectros de FTIR de soro bovino, a combinação com técnicas multivariadas, como PCA e SVM, permite extrair padrões relevantes que se correlacionam com a presença ou ausência de infecção, favorecendo um diagnóstico mais rápido e preciso [4,5,7].

Este estudo, demonstra o potencial dessa abordagem como uma ferramenta complementar ao diagnóstico tradicional, contribuindo para ampliar a precisão, agilidade e eficiência no monitoramento da doença. Para isso, foram obtidos espectros por meio de FTIR nas amostras de soro bovino, empregando-se pré-processamento por SNV, análise exploratória via PCA e classificação supervisionada com SVM [4-8]. A expectativa é que essa estratégia revele assinaturas espectrais distintivas, capazes de fundamentar o desenvolvimento de uma ferramenta complementar, eficiente e acessível para o diagnóstico da doença em rebanhos [4-7].

METODOLOGIA

4.1: Coleta das amostras

O estudo foi conduzido utilizando um total de 80 amostras de soro bovino provenientes da Faculdade de Medicina Veterinária e Zootecnia (FAMEZ), sendo de animais não infectados classificados como controle e de animais infectados classificados como tuberculose, com a aprovação do Comitê de Ética no Uso Animal (CEUA 765/2016) da Universidade Federal de Mato Grosso do Sul (UFMS).



4.2: Aquisição de dados

Os espectros foram obtidos por espectroscopia de Infravermelho por Transformada de Fourier (FTIR), que utiliza uma fonte de luz infravermelha que incide sobre o acessório de Reflectância Total Atenuada (ATR) e interage com a superfície da amostra, enquanto um detector mede a absorção da radiação, gerando o espectro característico para análise [6-10]. Utilizou-se o espectrômetro Agilent Cary 630, com resolução de 4 cm^{-1} e 12 varreduras por amostra, na faixa espectral de 1800 a 900 cm^{-1} [11]. Antes da leitura de cada amostra, realizou-se a aquisição de um espectro de background com água ionizada. Em seguida, 20 μL de soro foram aplicados diretamente sobre o ATR do equipamento para a coleta dos espectros. [12].

4.3: Pré Processamento

Os dados espectrais foram submetidos a um pré-processamento, utilizando a Normalização da Variável Padrão (SNV). Este método corrige variações indesejadas de espalhamento e diferenças de intensidade decorrentes da variabilidade das amostras ou de particularidades do equipamento [13-14], tornando os espectros mais comparáveis e adequados para análises subsequentes [11,12,17].

Matematicamente, o SNV atua centralizando e escalonando cada espectro individualmente. Para isso, subtraindo a média de todos os pontos do espectro ($\bar{x}$) de cada valor espectral (x_i) e, em seguida dividindo pelo desvio padrão (s) do mesmo espectro. Este procedimento resulta em espectros com média zero e variância unitária, o que minimiza tendências de fundo e efeitos multiplicativos. A transformação é representada pela seguinte equação [9-15]:

$$x_i^{SNV} = \frac{x_i - \bar{x}}{s} \quad (1)$$

Em que x_i representa o valor de transmitância no ponto espectral i , $\bar{x}$ é a média de todos os pontos do espectro e o s é o desvio-padrão do espectro. Com a aplicação do SNV, variações físicas relacionadas ao modo de preparo ou à espessura das amostras são atenuadas, realçando assim as diferenças químicas intrinsecamente relevantes para o estudo.



4.4: Análise de Componentes Principais (PCA)

A Análise de Componentes Principais (PCA) é um método não supervisionado que transforma múltiplas variáveis em alguns componentes principais por redução de dimensionalidade dos dados espectrais pré-processados [16], sem perder informações essenciais sobre a variação do conjunto analisado [12,17]. Essa técnica projeta os dados para um novo espaço, facilitando a visualização dos padrões e possíveis agrupamentos referentes ao perfil espectral das amostras de soro bovino [9,11,16]. Em seguida, gráficos de escores foram gerados para analisar tendências e identificar a separação entre amostras de animais infectados por tuberculose e controle, e a remoção de amostras irregulares (outliers) que podem distorcer a interpretação dos resultados e comprometer a clusterização das amostras, a técnica estatística, T^2 de Hotelling, que avalia a distância de cada amostra ao centro dos dados no novo espaço das componentes principais, esse procedimento é fundamental para garantir a qualidade da análise multivariada [15,18].

4.5: Aprendizado de Máquinas

A classificação das amostras de soro foi realizada utilizando Máquina de Vetores de Suporte (SVM), um algoritmo de aprendizado supervisionado usado para classificação, que separa diferentes classes ao encontrar o melhor hiperplano que maximiza a margem entre os grupos de dados no espaço de atributos [19,20], pode utilizar diferentes funções de kernel para permitir separações não lineares, tornando eficaz em conjuntos complexos ou com alta dimensionalidade [11,19,21]. Para avaliar o desempenho, adotou-se o método Leave-One-Out Cross-Validation (LOOCV) consiste em utilizar cada amostra do conjunto de dados, uma de cada vez, como conjunto de teste, enquanto todas as demais servem para treinar o modelo [20,21,23].

Para otimizar o SVM, foram testados diferentes kernels (linear, quadrático e cúbico) considerando o melhor resultado e o menor número de componentes principais. A matriz de confusão foi utilizada para avaliar o desempenho e sensibilidade do classificador entre os grupos analisados [20-21]. Devido ao número

reduzido de amostras, todo o conjunto foi utilizado para o treinamento do modelo. O fluxograma metodológico adotado neste estudo encontra-se ilustrado na Figura 1.

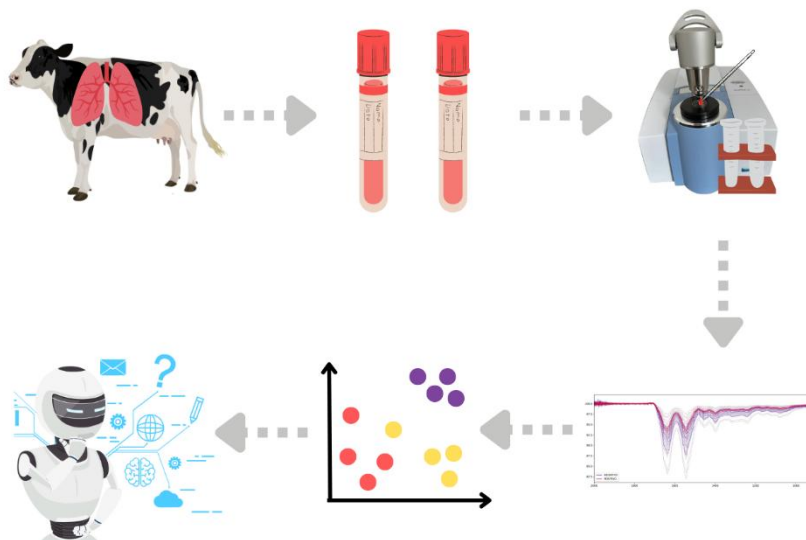


Figura 1: Fluxo metodológico do estudo: coleta de amostras de soro de bovinos (infectados e controle), processamento das amostras por espectroscopia FTIR, análise exploratória dos dados usando método não supervisionado e, por fim, aplicação de algoritmos de aprendizado de máquina para discriminação entre os grupos analisados.

RESULTADOS E DISCUSSÃO

Para avaliar as diferenças espectrais entre os grupos analisados, inicialmente foram obtidos os espectros FTIR brutos de todas as amostras de soro bovino, como demonstrado na Figura 2, nomeadas como “controle” e “tuberculose” em relação ao diagnóstico de tuberculose. A visualização desses dados é fundamental para identificar padrões gerais de absorção e possíveis variações que possam ser utilizadas nas etapas subsequentes de análise.

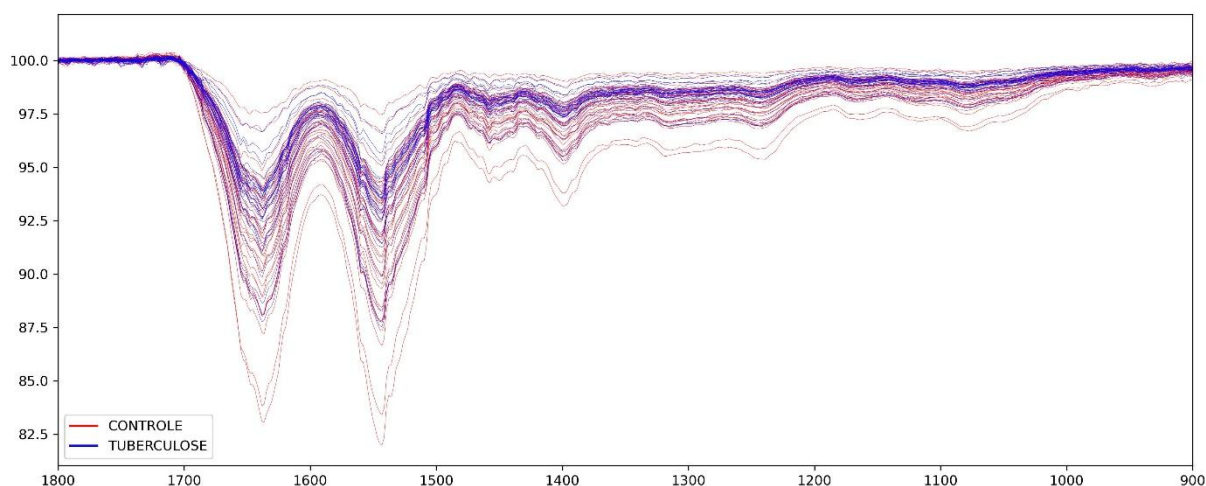


Figura 2: Perfis espectrais brutos obtidos por espectroscopia de infravermelho por Transformada de Fourier (FTIR) para as amostras de soro de bovinos controle e infectados por tuberculose, registrados na região de impressão digital ($1800 - 900 \text{ cm}^{-1}$), evidenciando padrões de absorção e variações entre os grupos de dados não tratados.

A Figura 2 mostra os perfis espectrais brutos, registrados por FTIR na região dos espectros na faixa de 1800 a 900 cm^{-1} conhecida como “impressão digital”, nessa faixa ocorrem absorções resultantes de diversas vibrações moleculares combinadas e específicas de cada molécula, tornando o espectro característico para identificação de compostos [21-22]. Por conter sinais únicos, essa região é fundamental para diferenciar moléculas, e por isso é largamente explorada em estudos de soro bovino, especialmente para diagnóstico e discriminação de grupos. As curvas apresentam padrões semelhantes, com banda de absorção intensa entre 1700 e 1500 cm^{-1} , região das proteínas de Amida I e Amida II, típica de vibrações de ligações e componentes estruturais do soro [22]. Observa-se uma diferença entre os perfis espectrais dos grupos, que pode indicar a infecção por tuberculose ou ruídos experimentais detectáveis nos espectros infravermelhos das amostras de soro. Essa diferença são fundamentais para etapas posteriores de análise multivariada e classificação.

Conforme demonstrado na Figura 3 apresenta-se o resultado do pré-processamento dos espectros por meio do método SNV, a média e o desvio padrão, facilitando a distinção entre as condições analisada.

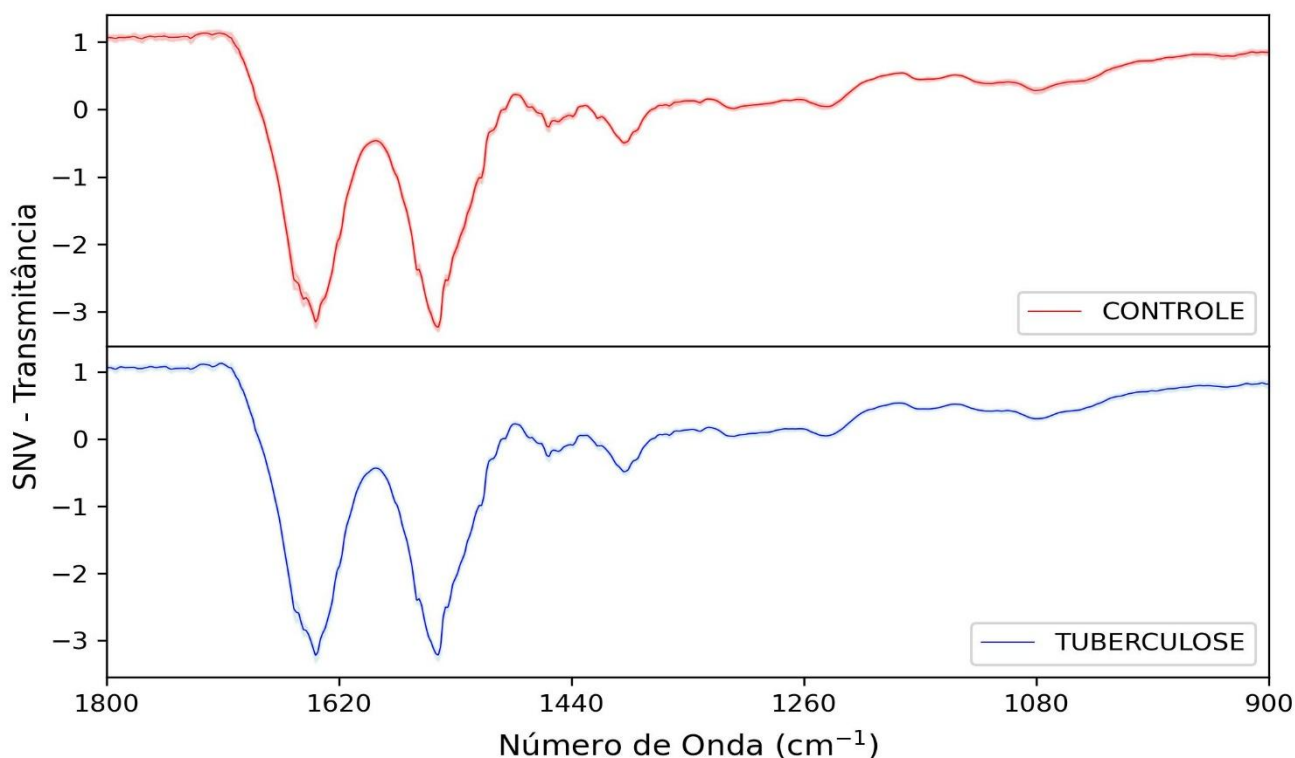


Figura 3: Média e desvio padrão para cada espectro de FTIR pré processado por SNV

O pré-processamento foi utilizado devido ao comportamento não linear dos espectros FTIR obtidos e à sua eficácia para o início das análises, facilitando a correção de espalhamento e diferenças de baseline. Na Figura 3, é possível identificar bandas associadas a vibrações específicas dos grupos funcionais presentes no soro, a banda em torno de 1650 cm^{-1} refere-se à Amida I (alongamento C=O das proteínas) enquanto próximo de 1550 cm^{-1} corresponde à Amida II. As bandas por volta de 1450 cm^{-1} indicam a flexão do grupo metileno (CH_2), a banda em torno de 1400 cm^{-1} é atribuída à vibração de estiramento simétrico (ν_s) do grupo carboxilato (COO^-). A banda, próxima de 1240 cm^{-1} , corresponde à Amida III e em torno de 1100 cm^{-1} é atribuído ao estiramento da ligação C-O-C [20-22].

A observação dessas bandas evidencia os principais componentes do soro que podem estar alterados pela condição infecciosa, servindo de base para as etapas subsequentes de análise multivariada e classificação.

Na Figura 4, são apresentados os gráficos de scores plot e *Loading* referentes à Análise de Componentes Principais dos espectros FTIR das amostras. Estes resultados destacam a distribuição das amostras projetadas nos componentes



principais e as regiões espectrais de maior contribuição para a variância explicada, evidenciando tendências na diferenciação entre grupos controle e tuberculose.

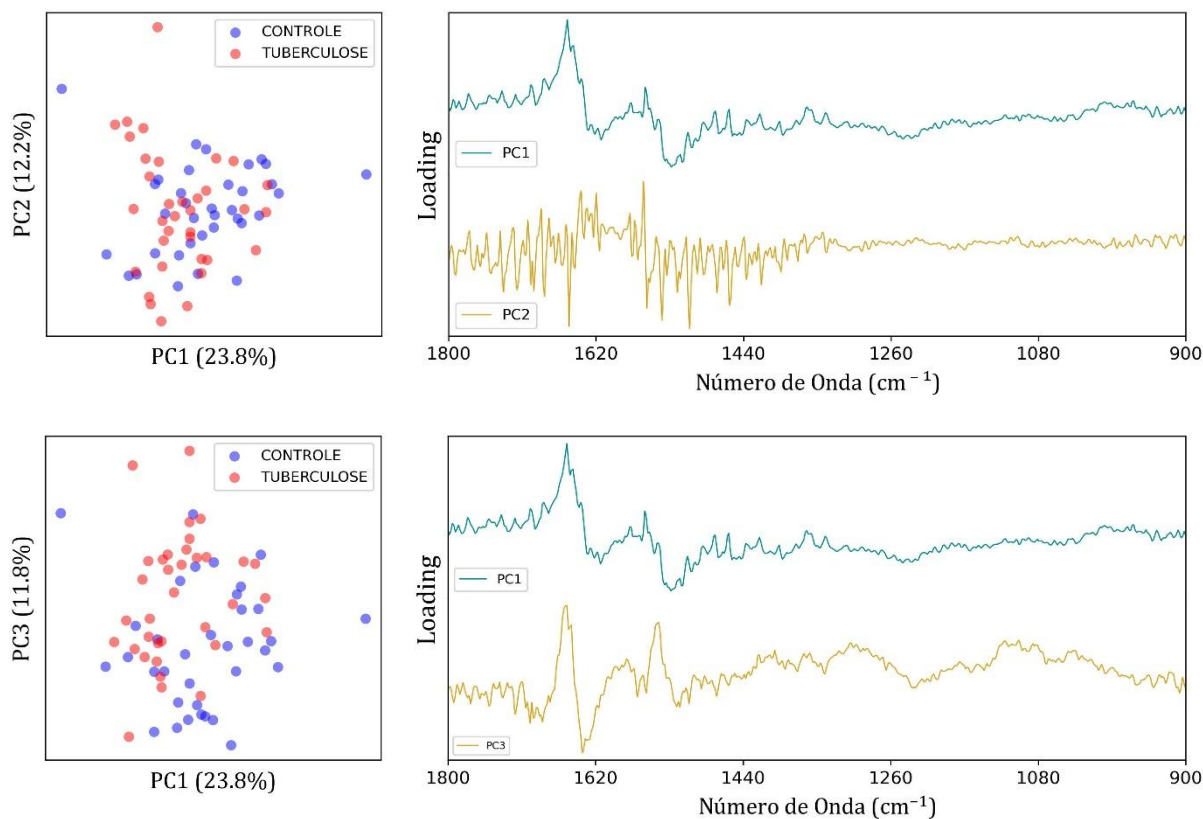


Figura 4: À esquerda, gráfico de score plot mostrando a projeção das amostras nos componentes principais das amostras indicando as variâncias por PC1, PC2 ou PC3 e à direita, gráficos de loadings das componentes principais no intervalo de 1800 - 900 cm^{-1} .

A visualização das amostras nos planos dos componentes principais permite avaliar a distribuição dos dados reduzidos em função das variáveis espectrais originais, com destaque para a tendência de agrupamento dos grupos controle e tuberculose e para sobreposição significativa dos escores no espaço bidimensional [16]. A variância explicada por cada componente, expressa nos eixos dos gráficos, reflete a capacidade do modelo em sintetizar as informações do conjunto espectral. O gráfico de *Loading*, por sua vez, evidencia as regiões espectrais de maior relevância para a construção dos componentes principais, indicando faixas associadas, apesar da influência de ruídos inerentes à matriz biológica e ao processo analítico.



Observa-se uma tendência de maior separação diagonal entre os grupos controle e tuberculose, embora não haja uma fronteira nítida entre eles, com muitos sobreposição dos escores (amostras) no espaço de componentes principais. No plano PC1 e PC2, a variância explicada é de 23,8% para PC1 e 12,2% para PC2, já no plano PC1 e PC3, PC3 capta 11,8% da variância total, evidenciando que PC1 e PC2 concentram maior variabilidade dos dados, embora PC2 carregue mais ruído do que informação, dificultando a separação efetiva dos grupos. A inclusão do PC3 se justifica pelo fato de que esse componente carrega parte relevante da variância dos dados, permitindo observar uma maior tendência de agrupamento [22].

Os gráficos de loading, mostram quais regiões do espectro mais contribuem para a formação dos componentes principais [13]. O PC1 evidencia bandas intensas em torno de 1700 e 1600 cm^{-1} , relacionadas, principalmente, às vibrações dos grupos amida I e II de proteínas. O PC2 destaca regiões próximas de 1700, 1600 e também de 1100 cm^{-1} , conectadas não apenas às proteínas, mas também a estruturas C-O-C. O PC3, por sua vez, contribui com informação adicional ao evidenciar padrões espectrais menos impactados por ruído, com menor sobreposição entre os grupos. Entretanto, a partir do PC3, os demais componentes passam a carregar predominantemente informações associadas a ruídos experimentais, não contribuindo de forma relevante para a interpretação dos dados. Essas regiões com ruídos, podem ser decorrentes de fatores como a natureza das amostras, manipulação ou limitações do equipamento.

Para a etapa de aprendizado de máquina, foi adotado o método de validação Leave-One-Out Cross-Validation (LOOCV) [12-20], que consistiu em utilizar todas as amostras como treinamento, sem separação entre treino e teste, proporcionando uma avaliação mais rígida do desempenho dos modelos. Para representar melhor as variações dos dados espectrais e aumentar a acurácia da classificação, foram selecionados os três primeiros componentes principais (PC1, PC2 e PC3).



Na Figura 5 expressa a acurácia do modelo de classificação SVM em função da quantidade de componentes principais utilizados, comparando os resultados obtidos com os núcleos linear, quadrático e cúbico. Observa-se que o kernel linear apresenta maior estabilidade e desempenho ao longo das diferentes quantidades de PCs, se destacando em relação aos demais. O parâmetro C foi definido como 1 por equilibrar adequadamente o grau de regularização e o ajuste aos dados, evitando overfitting, especialmente relevante em conjuntos com amostras limitadas [20].

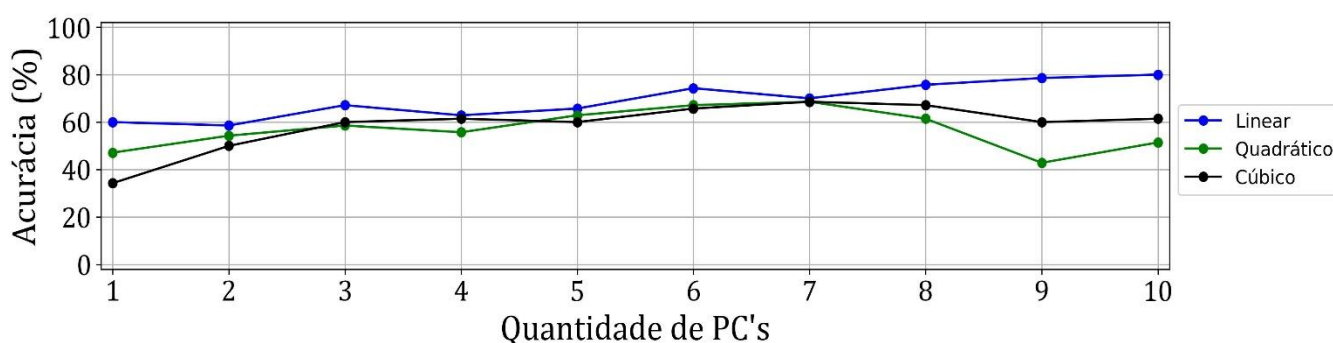


Figura 5: Acurácia (%) do modelo SVM nas diferentes configurações de núcleos (linear, quadrático e cúbico) em função da quantidade de componentes principais utilizadas. O kernel linear apresenta melhor resultado e foi selecionado para a etapa de classificação final.

A matriz de confusão, na Figura 6, subsequente foi empregada para a avaliação detalhada de desempenho do modelo SVM linear, indicando a taxa de acerto na classificação dos grupos controle e tuberculose, confirmando a eficácia da abordagem adotada e sua capacidade discriminativa com base nos dados espectrais pré-processados e reduzidos.

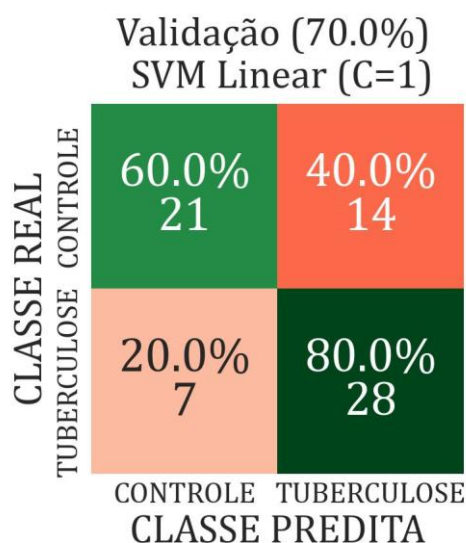


Figura 6: Matriz de confusão para o modelo SVM linear (C=1) com validação LOOCV, indicando o percentual de acerto na classificação das amostras de soro bovino entre os grupos controle e tuberculose.

A matriz de confusão obtida para o modelo SVM linear (C=1), com acurácia de 70%, apresenta os índices de acerto e erro na classificação das amostras de soro bovino entre os grupos controle e tuberculose. Observa-se que o classificador atingiu uma taxa de acerto de 60% para o grupo controle e 80% para o grupo tuberculose, indicando sensibilidade elevada na identificação dos casos positivos. Por outro lado, a especificidade para o grupo controle foi de 60%, enquanto 40% das amostras controle foram erroneamente classificadas como tuberculose. Entre as amostras tuberculose, apenas 20% foram distribuídas incorretamente como controle, reforçando a maior capacidade do modelo em reconhecer os casos doente [24]. Esses resultados validam a eficácia discriminativa do método linear adotado, evidenciando que a combinação das três primeiras componentes principais favoreceu o rendimento do classificador na distinção dos grupos estudados.



CONCLUSÃO

Os resultados obtidos neste trabalho evidenciam que a espectroscopia de infravermelho por transformada de Fourier (FTIR), aliada com algoritmo de aprendizado de máquinas, representa uma abordagem promissora e eficaz para a diferenciação entre soros de bovinos infectados e não infectados por tuberculose. A capacidade do método em extrair “impressão digital” permitiu alcançar índices adequados de acurácia e sensibilidade na classificação das amostras, mesmo diante da complexidade inerente dos dados biológicos.

Embora não substitua, no atual estágio, os métodos convencionais já estabelecidos, a combinação de FTIR com técnicas avançadas de análise multivariada desponta como ferramenta complementar de potencial uso em triagens rápidas, especialmente em cenários de baixa infraestrutura laboratorial. Ressalta-se a importância da ampliação de estudos com maior número de amostras, validação em diferentes populações e integração com outros biomarcadores, fortalecendo o caminho para tornar o diagnóstico da tuberculose bovina cada vez mais preciso, ágil e acessível.

Dessa forma, o presente estudo contribui para o avanço do monitoramento e controle da tuberculose bovina, abrindo perspectivas para a aplicação de tecnologias inovadoras na sanidade animal e proteção da saúde pública.

AGRADECIMENTOS

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT), pelo apoio financeiro e institucional fundamental para a realização deste estudo, incluindo os acordos de cooperação firmados entre a UFMS e demais parceiros.



REFERÊNCIAS BIBLIOGRÁFICAS

- [1] COLLINS, Áine B. *et al.* Prevalence of *Mycobacterium bovis* in milk on dairy cattle farms: An international systematic literature review and meta-analysis. *Tuberculosis* (Edinburgh, Scotland), v. 132, n. 102166, p. 102166, 2022.
- [2] WISEMAN, J.; CASSIDY, J. P.; GORMLEY, E. The problem that residual *Mycobacterium bovis* infection poses for the eradication of bovine tuberculosis. *Veterinary journal* (London, England: 1997), v. 308, n. 106266, p. 106266, 2024.
- [3] BARBOSA, Matheus E. P. *et al.* Easy and fast discrimination of female sand flies from *Lutzomyia* species with infrared spectroscopy and multivariate analysis. *Analytical chemistry*, v. 97, n. 22, p. 11581–11588, 2025.
- [4] DOU, Jingrui *et al.* Rapid discrimination of Brucellosis in sheep using serum Fourier transform infrared spectroscopy combined with PCA-LDA algorithm. *Photodiagnosis and photodynamic therapy*, v. 42, n. 103567, p. 103567, 2023.
- [5] CAMERON, James M. *et al.* Exploring pre-analytical factors for the optimisation of serum diagnostics: Progressing the clinical utility of ATR-FTIR spectroscopy. *Vibrational spectroscopy*, v. 109, n. 103092, p. 103092, 2020.
- [6] AGUIAR, Josafá C. *et al.* Differentiation of *Leishmania* species by FT-IR spectroscopy. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, v. 142, p. 80–85, 2015.
- [7] BEATTIE, J. Renwick; ESMONDE-WHITE, Francis W. L. Exploration of principal component analysis: Deriving principal component analysis visually using spectra. *Applied spectroscopy*, v. 75, n. 4, p. 361–375, 2021.
- [8] OLIVEIRA, Izabela C. *et al.* Fast and accurate discrimination of *Brachiaria brizantha* (A. rich.) stapf seeds by molecular spectroscopy and machine learning. *ACS agricultural science & technology*, v. 1, n. 5, p. 443–448, 2021.
- [9] FRANCA, Thiago *et al.* FTIR spectroscopy of biofluids for disease diagnosis: Data quality influence on group separation. *In: IEEE*, 2024.



[10] BATES, J. B. (ORG.). Fourier Transform Infrared Spectroscopy The basic principles and current applications of a rapidly expanding technique are reviewed. J. B. Bates, [S.d.].

[11] DE BRITO, Eliana C. A. *et al.* Paracoccidioidomycosis screening diagnosis by FTIR spectroscopy and multivariate analysis. Photodiagnosis and photodynamic therapy, v. 39, n. 102921, p. 102921, 2022.

[12] FRANCA, Thiago *et al.* Improving bovine brucellosis diagnostics: Rapid, accurate detection via blood serum infrared spectroscopy and machine learning. ACS omega, v. 10, n. 22, p. 22952–22959, 2025.

[13] BI, Yiming *et al.* A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. Analytica chimica acta, v. 909, p. 30–40, 2016.

[14] RINNAN, Åsmund; VAN DEN BERG, Frans; ENGELSEN, Søren Balling. Review of the most common pre-processing techniques for near-infrared spectra. Trends in analytical chemistry: TRAC, v. 28, n. 10, p. 1201–1222, 2009.

[15] OLIVEIRA, Vilma A. S. *et al.* Easy identification of Leishmania (Leishmania) amazonensis and Leishmania (Viannia) braziliensis species by using Fourier-transform infrared spectroscopy and machine learning algorithms. ACS omega, v. 10, n. 44, p. 52483–52491, 2025.

[16] JOLLIFFE, Ian T.; CADIMA, Jorge. Principal component analysis: a review and recent developments. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, v. 374, n. 2065, p. 20150202, 2016.

[17] DE REZENDE, Bruno Silva *et al.* Turning chaotic sample group clusterization into organized ones by feature selection: Application on photodiagnosis of Brucella abortus serological test. Journal of photochemistry and photobiology. B, Biology, v. 247, n. 112781, p. 112781, 2023.

[18] HOTELLING, H. Analysis of a complex of statistical variables into principal components. Journal of educational psychology, v. 24, n. 6, p. 417–441, 1933.

[19] JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. Science (New York, N.Y.), v. 349, n. 6245, p. 255–260, 2015.



[20] LARIOS, Gustavo *et al.* A new strategy for canine visceral leishmaniasis diagnosis based on FTIR spectroscopy and machine learning. *Journal of biophotonics*, v. 14, n. 11, p. e202100141, 2021.

[21] COELHO, Mateus Lotério *et al.* Canine visceral leishmaniasis diagnosis by UV spectroscopy of blood serum and machine learning algorithms. *Photodiagnosis and photodynamic therapy*, v. 42, n. 103575, p. 103575, 2023.

[22] PACHER, Gabriela *et al.* Diagnosis of cutaneous leishmaniasis using FTIR spectroscopy and machine learning: An animal model study. *ACS infectious diseases*, v. 10, n. 2, p. 467–474, 2024.

[23] FRANCA, Thiago; GONCALVES, Daniel; CENA, Cicero. ATR-FTIR spectroscopy combined with machine learning for classification of PVA/PVP blends in low concentration. *Vibrational spectroscopy*, v. 120, n. 103378, p. 103378, 2022.

[24] LÓPEZ-SANTIAGO, Rubén *et al.* Immune response to mucosal *Brucella* infection. *Frontiers in immunology*, v. 10, p. 1759, 2019.