

**UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CAMPUS DE CHAPADÃO DO SUL
CURSO DE BACHARELADO EM AGRONOMIA**

EVELYN RODRIGUES DE LACERDA

**ESPECTROSCOPIA VIS-NIR-SWIR, INDICADORES
MICROBIANOS E CLASSIFICAÇÃO DO USO DO SOLO
UTILIZANDO MODELOS DE APRENDIZADO DE MÁQUINA**

CHAPADÃO DO SUL-MS

2026

**UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CAMPUS DE CHAPADÃO DO SUL
CURSO DE BACHARELADO EM AGRONOMIA**

**ESPECTROSCOPIA VIS-NIR-SWIR, INDICADORES
MICROBIANOS E CLASSIFICAÇÃO DO USO DO SOLO
UTILIZANDO MODELOS DE APRENDIZADO DE MÁQUINA**

Trabalho de Conclusão de Curso
apresentado à Universidade Federal de
Mato Grosso do Sul, como parte dos
requisitos para obtenção do título de
Agrônomo (a).

Orientadora: Prof. Dra. Meire
Aparecida Silvestrini Cordeiro.

Co-Orientador: Énio Antonio Manfroi
Filho.

**CHAPADÃO DO SUL-MS
2026**



CERTIFICADO DE APROVAÇÃO

AUTOR: Evelyn Rodrigues de Lacerda.

ORIENTADORA: Profa. Dra. Meire Aparecida Silvestrini Cordeiro.

Aprovada pela Banca Examinadora como parte das exigências do Componente Curricular Não Disciplinar TCC, para obtenção do grau de BACHARELA EM AGRONOMIA, pelo curso de Bacharelado em Agronomia da Universidade Federal de Mato Grosso do Sul, Campus de Chapadão do Sul.

Profa. Dra. Meire Aparecida Silvestrini Cordeiro

Presidente da Banca Examinadora e Orientador

Msc. Enio Antônio Manfroi Filho

Membro da Banca Examinadora e Coorientador

Profa. Dra. Larissa Pereira Ribeiro Teodoro

Membro da Banca Examinadora

Profa. Dra. Nívea Maria Mafra Rodrigues

Membro da Banca Examinadora

Chapadão do Sul, 12 de junho de 2026.

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Meire Aparecida Silvestrini Cordeiro, Professora do Magistério Superior**, em 30/06/2026, às 08:25, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Nívea Maria Mafra Rodrigues, Professora do Magistério Superior**, em 30/06/2026, às 08:51, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Larissa Pereira Ribeiro, Professora do Magistério Superior**, em 30/06/2026, às 13:42, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Enio Antonio Manfroi Filho, Assistente em Administração**, em 30/06/2026, às 17:53, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufms.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6499338** e o código CRC **C9D09C30**.

COORDENAÇÃO DE GESTÃO ACADÊMICA DO CÂMPUS DE CHAPADÃO DO SUL

Avenida Engenheiro Douglas Ribeiro Pantaleão, nº 5167

Fone:

CEP 79560-000 - Chapadão do Sul - MS

Referência: Processo nº 23455.000504/2026-24

SEI nº 6499338

AGRADECIMENTOS

A Deus, força maior que rege minha vida, e aos meus guias espirituais, que iluminaram meus caminhos, intuíram minhas decisões e me sustentaram nos momentos de maior cansaço e incerteza durante toda essa jornada acadêmica.

À Universidade Federal de Mato Grosso do Sul, campus de Chapadão do Sul e ao Programa de Educação Tutorial (PET) Agroflorestal, por toda a infraestrutura disponibilizada que tornou este trabalho possível. À minha orientadora Prof. Dra. Meire Aparecida Silvestrini Cordeiro e ao meu co-orientador Énio Antonio Manfroi Filho, por terem aceitado o desafio de guiar esta pesquisa. Agradeço imensamente pela paciência, pelas correções minuciosas, pelo conhecimento compartilhado e por sempre me impulsionar a buscar o rigor científico.

Aos meus pais, Claudio Francisco de Lacerda e Silvana Rodrigues, e familiares, Kauã Rodrigues de Lacerda, Naiara Rodrigues de Lacerda e John Lenon Rodrigues da Silva, pelo amor incondicional, pelo suporte financeiro e emocional, e por compreenderem as minhas ausências durante as fases mais intensas deste trabalho. Vocês são a minha base para tudo.

Aos meus amigos, Evelyn Gabrielly Boff, Gustavo Casadei Sabino e João Victor Xavier dos Santos, que me ouviram desabafar sobre algoritmos e solos, e que, mesmo sem entender a fundo o tema, sempre me deram forças para continuar e deixaram esta caminhada mais leve.

Aos meus colegas de laboratório e de campo, Necivero Santos de Jesus Júnior, Évellen de Alencar dos Santos e Marcelo José da Silva, que dividiram comigo a rotina de longas horas de análises microbiológicas. Sem o trabalho em equipe e ao apoio mútuo nos momentos de exaustão, esta pesquisa não teria se concretizado.

Às proprietárias da Fazenda Rancho Ideal, Maria Otilia Zardo, Livia Zardo Lovato e Eliane Cristina Rosin pela imensa generosidade ao abrirem as portas da propriedade e permitirem a coleta das amostras de solo necessárias para esta pesquisa. Agradeço pela confiança depositada e pelo apoio logístico concedido durante o trabalho de campo.

ESPECTROSCOPIA VIS-NIR-SWIR, INDICADORES MICROBIANOS E CLASSIFICAÇÃO DO USO DO SOLO UTILIZANDO MODELOS DE APRENDIZADO DE MÁQUINA

Resumo

A microbiologia do solo demanda métodos analíticos e técnicas não destrutivas em análises laboratoriais que sejam capazes de avaliar a qualidade do solo e seus marcadores microbiológicos. Diante disso, o sensoriamento proximal hiperespectral que é amplamente utilizado na agricultura como ferramenta eficaz e precisa no diagnóstico agrícola, associado às técnicas de aprendizado de máquina, pode apresentar potencial para realizar classificações por meio de modelos que utilizam algoritmos especializados. Portanto, o objetivo desta pesquisa foi realizar a classificação dos diferentes tipos de uso do solo por meio de indicadores microbianos e sensoriamento espectral utilizando modelos de aprendizado de máquina. O estudo foi conduzido na fazenda Rancho Ideal, localizada no município de Chapadão do Sul, Mato Grosso do Sul. As amostras de solo foram coletadas em quatro diferentes usos do solo presentes nos talhões da fazenda, sendo: Mata Nativa (Cerrado), Eucalipto, Pastagem e Lavoura. Em cada uso do solo foram coletadas 25 amostras, contabilizando 100 amostras ao total. A aquisição espectral foi realizada utilizando o equipamento Ocean Optics constituído de dois sensores: STS-VIS e Flame. Os indicadores microbianos de qualidade do solo analisados foram: Carbono da biomassa microbiana, respiração basal do solo e quociente metabólico, por meio de metodologias específicas. Para modelagem, seis modelos foram treinados globalmente com o conjunto de dados completo para cada indicador, a avaliação foi realizada por meio de validação cruzada com dez repetições. Os resultados evidenciam que, utilizando a técnica de pré-processamento Savitzky-Golay, os modelos testados foram capazes de identificar padrões complexos e correlações entre as bases de dados usadas, especialmente os baseados em gradient boosting, alcançando acurácia média de 94%. Dessa forma, conclui-se que, para a classificação, em um padrão espectral consistente, modelos de diferentes naturezas obtiveram desempenho significativo, sem depender de um algoritmo específico.

Palavras-chave: Sensor espectral, algoritmos, microrganismos do solo.

VIS-NIR-SWIR SPECTROSCOPY, MICROBIAL INDICATORS, AND LAND USE CLASSIFICATION USING MACHINE LEARNING MODELS

Abstract

Soil microbiology demands analytical methods and non-destructive techniques in laboratory analyses that are capable of evaluating soil quality and its microbiological markers. Therefore, spectral sensing, which is widely used in agriculture as an effective and precise tool for agricultural diagnosis, associated with machine learning techniques, may present potential to perform classifications through models that use specialized algorithms. Thus, the objective of this research was to classify different types of land use through microbial indicators and spectral sensing using machine learning models. The study was conducted at Rancho Ideal farm, located in the municipality of Chapadão do Sul, Mato Grosso do Sul. Soil samples were collected from four different land uses present in the farm plots, namely: Native Forest (Cerrado), Eucalyptus, Pasture and Cropland. In each land use, 25 samples were collected, totaling 100 samples. Spectral acquisition was performed using the Ocean Optics equipment consisting of two sensors: STS-VIS and Flame. The microbial indicators of soil quality analyzed were: microbial biomass carbon, basal soil respiration and metabolic quotient, through specific methodologies. For modeling, six models were trained globally with the complete dataset for each indicator, and the evaluation was performed through cross-validation with ten repetitions. The results show that, using the Savitzky-Golay preprocessing technique, the tested models were capable of identifying complex patterns and correlations between the databases used, especially those based on gradient boosting, reaching an average accuracy of 94%. Therefore, it is concluded that for classification, in a consistent spectral pattern, models of different natures obtained significant performance, without depending on a specific algorithm.

Keywords: Spectral sensor, algorithms, soil microorganisms.

LISTA DE FIGURAS

	Página
01. Caracterização da localização do estudo no espaço geográfico e pontos de coleta de solo nas áreas da fazenda Rancho Ideal.....	3
02. Esquemática visual da distribuição das 25 amostras de solo em cada tipo de uso do solo com áreas de 50m x 50m.....	4
03. Comparação entre a análise de componentes principais (PCA) dos dados espectrais do solo e dos indicadores microbiológicos laboratoriais em diferentes tipos de uso do solo.....	8
04. Comparação entre análise discriminante linear (LDA) dos dados espectrais do solo e dos atributos microbiológicos laboratoriais em diferentes tipos de uso do solo.....	11
05. Assinaturas espectrais médias dos diferentes usos do solo nas regiões VIS, NIR e SWIR, com variação expressa pelo desvio padrão.....	13
06. Ilustração visual das cores do solo em relação á reflectância.....	14
07. Boxplots dos indicadores microbianos de qualidade do solo em diferentes usos do solo.....	16
08. Desempenho do pré-processamento SG1 e dos modelos de machine learning com base nas métricas Accuracy, F1-score, Precision e Recall.	19

LISTA DE TABELAS

	Página
01. Hiperparâmetros usados para os modelos de aprendizado de máquina....	5
02. Relação dos modelos de aprendizado de máquina, utilizados na classificação.....	7
03. Desempenho médio dos modelos de aprendizado de máquina submetidos ao pré-processamento SG1, com base nas métricas Accuracy, F1-score, Precision e Recall.....	18

SUMÁRIO

Resumo	iii
Abstract	iv
LISTA DE FIGURAS	v
LISTA DE TABELAS	vi
1 INTRODUÇÃO	1
2 MATERIAL E MÉTODOS	2
2.1 Localização do estudo	2
2.2 Coleta de solo e amostragem	3
2.3 Análises laboratoriais	4
2.4 Leituras espectrais	5
2.5 Modelos de Classificação e Avaliação Estatística.....	5
3 RESULTADOS E DISCUSSÃO	7
3.1 Análise de Componentes Principais aplicada aos dados microbianos laboratoriais e aos dados espectrais.....	7
3.2 Análise discriminante linear dos dados espectrais e microbiológicos laboratoriais do solo	10
3.3 Assinaturas espectrais médias dos diferentes usos do solo nas regiões VIS, NIR e SWIR	13
3.4 Indicadores microbiológicos de qualidade do solo para avaliação da aderência à distribuição normal.....	15
3.5 Pré-processamentos e modelos de aprendizado de máquina na classificação multiclasse dos usos do solo.....	17
4 CONCLUSÕES	21
REFERÊNCIAS	22

1 INTRODUÇÃO

A existência de sistemas de uso do solo que sejam eficientes e estáveis para a produção agrícola é de grande importância para a subsistência da sociedade contemporânea (Souza *et al.*, 2020). A agricultura convencional com o tempo foi revelando diversas problemáticas que consequentemente enfraqueceram a vida no solo e a resiliência dos cultivos (Sudarshan *et al.*, 2024). Nesse contexto, a utilização de novas práticas sustentáveis de uso do solo, como a agricultura regenerativa, possibilitou aos produtores rurais restaurar áreas degradadas, fertilidade do solo e biodiversidade, garantindo produção de alimentos a longo prazo (Chen *et al.*, 2022; Khangura *et al.*, 2023).

Para garantir a eficiência de práticas sustentáveis, é muito importante medir a qualidade do solo por meio de indicadores específicos, uma vez que eles monitoram e medem os atributos do solo (Chaudhry *et al.*, 2024). Entre os indicadores, os microbiológicos são mais discriminativos a processos e perturbações, o que garante maior confiabilidade, entretanto a interpretação dos resultados é crucial (De Oliveira *et al.*, 2024). Alguns exemplos de métodos mais utilizados relacionados ao ciclo do carbono são as avaliações de carbono da biomassa microbiana (C-BM) e respiração basal do solo (RBS) (Semenov *et al.*, 2025). Para a execução desses métodos é utilizado reagentes químicos que desempenham um papel fundamental para o sucesso das análises (FAO, 2024).

Contudo, a toxicidade de reagentes químicos insalubres representa riscos significativos à saúde dos laboratoristas e seu descarte inadequado acarreta ameaças de contaminação do solo e lençol freático (Fatemi *et al.*, 2022; Hedman, 2026). De acordo com a norma regulamentadora NR-15, que visa atividades e operações insalubres, os seguintes reagentes (Ácido clorídrico e clorofórmio) são considerados com grau de insalubridade máximo, os quais são amplamente utilizados nas análises de RBS e C-BM. Pensando nisso, se faz necessário a criação de métodos não destrutivos e sua aplicação para a ciência do solo, visando segurança, rapidez e confiabilidade.

Entre os métodos não destrutivos, o sensoriamento proximal hiperespectral é amplamente utilizado na agricultura como ferramenta de rapidez e precisão no diagnóstico agrícola (Chen *et al.*, 2025). Ele funciona através da refletância da luz, detectando usualmente uma faixa de 400 – 2500 nm, abrangendo as bandas visível, infravermelho próximo (NIR) e de ondas curtas (SWIR), dessa forma, identifica atributos conforme a sua assinatura espectral única (Vairavan *et al.*, 2024). Estudos recentes da ciência do solo mostram que o sensoriamento espectral foi capaz de estimar textura, fertilidade, matéria orgânica e metais pesados do solo

(Shen *et al.*, 2020; Tiruneh *et al.*, 2022; Zhong *et al.*, 2023). No entanto, por possuir alta dimensão de dados e espectros complexos, dificulta a interpretação precisa. Dessa forma, é necessário a utilização de aprendizado de máquina (*Machine learning*) para extração de informações úteis (Liu *et al.*, 2026).

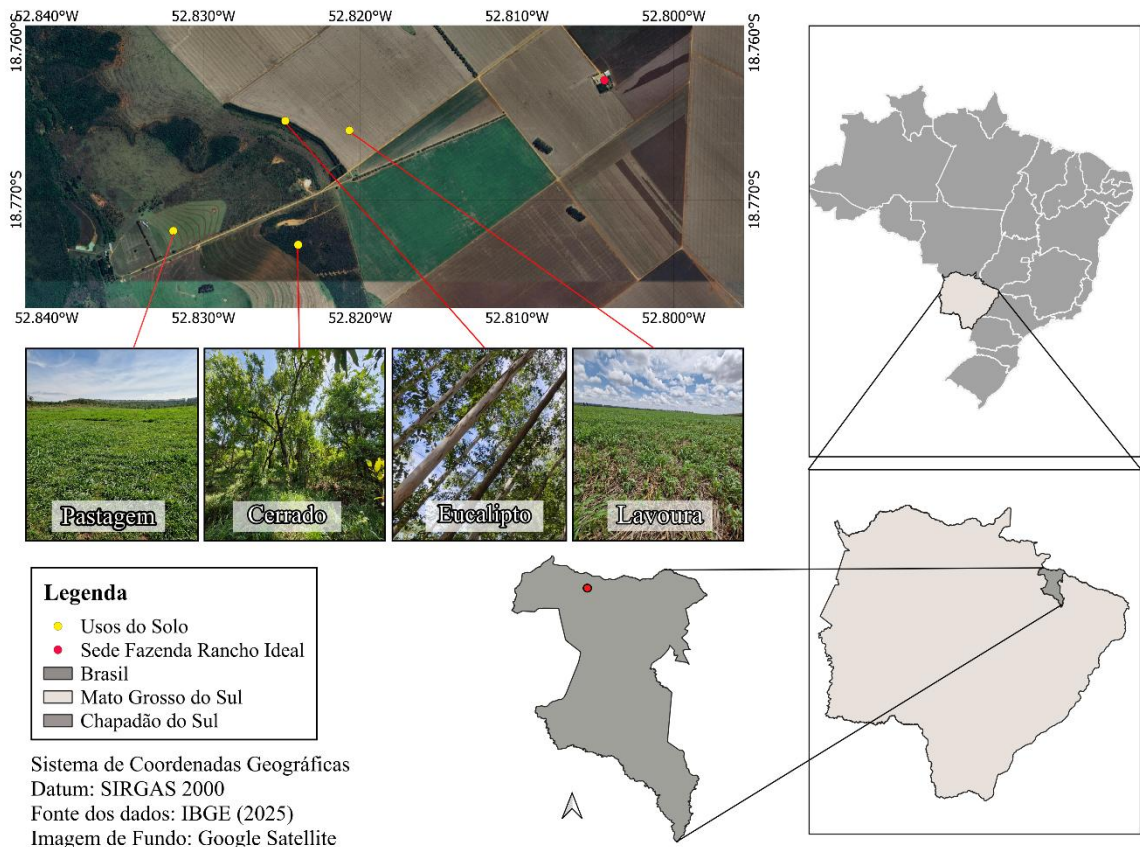
O aprendizado de máquina é uma ferramenta essencial na agricultura para análise de dados, sendo capaz de realizar previsões e classificações por meio de modelos que utilizam algoritmos especializados, aptos a processar um grande volume de medições. Em um estudo de (Dos Santos *et al.*, 2025) foi possível distinguir e classificar genótipos de soja por meio da análise de suas características espectrais utilizando diferentes algoritmos de aprendizado de máquina. Na ciência do solo (Camargo *et al.*, 2019) concluíram que imagens de satélites associadas ao aprendizado de máquina foram eficientes para classificar o uso e cobertura do solo do Cerrado, destacando os algoritmos *Random Forest*, *Multilayer Perceptron* e *Support Vector Machine* como os mais precisos.

Entretanto, estudos que utilizam base de dados de indicadores microbianos de qualidade do solo são incipientes. (Lima *et al.*, 2024) desenvolveram um modelo com alta precisão para distinguir entre áreas afetadas por rejeitos de mineração e aquelas em equilíbrio ecológico utilizando modelos de aprendizado de máquinas, e as características mais cruciais para essa discriminação foram os bioindicadores de qualidade do solo. Dessa forma, nota-se o grande potencial desses métodos de classificação aplicados à microbiologia do solo, bem como a importância de uma maior exploração em pesquisas futuras. Portanto o objetivo deste estudo é realizar a classificação dos diferentes tipos de usos do solo por meio de indicadores microbianos de qualidade do solo e sensoriamento espectral utilizando modelos de aprendizado de máquina.

2 MATERIAL E MÉTODOS

2.1 Localização do estudo

O estudo foi conduzido na fazenda Rancho Ideal, localizada no município de Chapadão do Sul, MS. O local da fazenda apresenta altitude de 837 metros com latitude de 18°45'44.26" Sul, longitude 52°48'15.82" Oeste (Figura 1), caracterizado com uma estação seca no inverno e outra chuvosa no verão, segundo a classificação de Köppen-Geiger. A região possui clima tropical (Aw), com uma temperatura média anual variando de 13 °C a 28 °C. A precipitação média anual é de 1.850 mm e o solo é classificado como Latossolo Vermelho Distrófico (Cunha *et al.*, 2013; Santos *et al.*, 2018).

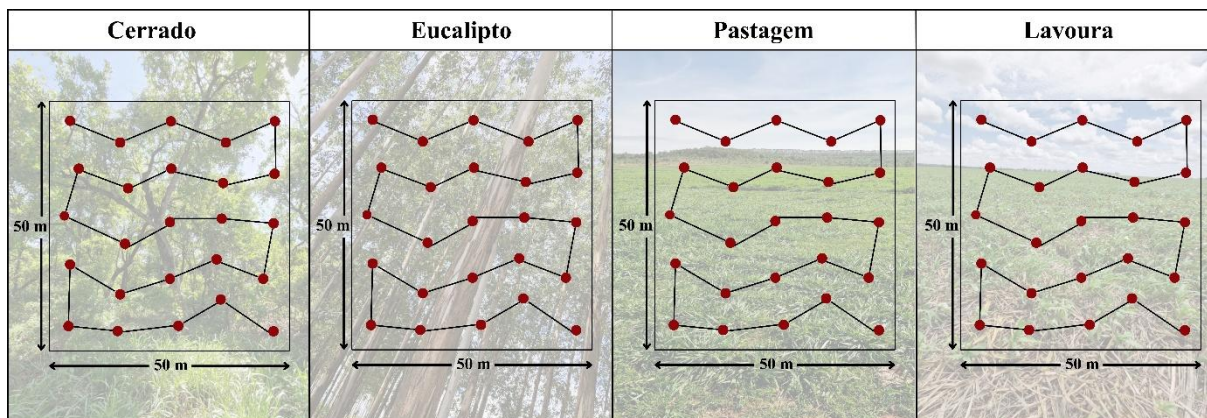


Fonte: Elaborado pela autora (2026).

Figura 1 - Caracterização da localização do estudo no espaço geográfico e pontos de coleta de solo nas áreas da fazenda Rancho Ideal.

2.2 Coleta de solo e amostragem

A coleta de solo ocorreu no mês de novembro de 2025, início do período chuvoso na região. As amostras de solo foram coletadas em em zigue – zague, com o auxílio do trado holandês, em 4 diferentes usos do solo presentes nos talhões da fazenda, sendo: Mata Nativa (Cerrado) (CER), Eucalipto (EUC), Pastagem (PAS) e Lavoura (LAV). Em cada uso do solo foram coletadas 25 amostras simples com profundidade de 0-10 cm, contabilizando 100 amostras ao total, distribuídas em áreas de 50 m x 50 m (Figura 2).



Fonte: Elaborado pela autora (2026).

Figura 2 - Esquematisação visual da distribuição das 25 amostras de solo em cada tipo de uso do solo com áreas de 50 m x 50 m.

No manejo da propriedade, práticas de agricultura regenerativa são adotadas há 20 anos. O histórico das áreas de amostragem define-se da seguinte forma: a área de Cerrado constitui a reserva nativa da fazenda, enquanto a área cultivada é composta pela espécie *Eucalyptus grandis*, com aproximadamente 14 anos de implantação.

A área de pastagem era composta pela espécie *Brachiaria Brizantha* BRS Piatã, como cultura de inverno e anteriormente foi semeada a soja no verão. Já a área de lavoura encontrava-se em sistema de rotação com mix de plantas de cobertura, contendo: Milheto, *Brachiaria Brizantha* BRS Piatã, níger e trigo mourisco, semeadas a lanço e incorporadas com niveladora na safrinha. Posteriormente na maturação dessas plantas, foi passado o rolo-faca, logo após ocorreu o rebrote, e foram coletadas as amostras de solo nesse talhão uma semana antes da semeadura da soja. Após coletadas, foram reservadas em incubadora de demanda bioquímica de oxigênio (BOD) e em seguida foi realizada as avaliações dos indicadores microbiológicos de qualidade do solo.

2.3 Análises laboratoriais

Os indicadores microbianos de qualidade do solo foram: Carbono da biomassa microbiana (C-BM), respiração basal do solo e quociente metabólico. O carbono da biomassa microbiana do solo foi avaliado por fumigação-extração (incubação por 24 h), com extração em K_2SO_4 ($0,5 \text{ mol L}^{-1}$), oxidação via $K_2Cr_2O_7$ ($0,0667 \text{ mol L}^{-1}$) e titulação com sulfato ferroso amoniacal $0,0333 \text{ mol L}^{-1}$, valores obtidos por (mg de C kg^{-1}) (Vance, Brookes e Jenkinson, 1987). A respiração basal do solo (RBS) foi determinada pelo CO_2 liberado a partir de 10 g de solo durante sete dias, utilizando $NaOH$ ($0,05 \text{ mol L}^{-1}$) e titulação com HCl ($0,05 \text{ mol L}^{-1}$),

dados mensurados via (mg de C-CO₂ Kg⁻¹) (Alef; Nannipieri, 1995). O quociente metabólico (*q*CO₂) foi obtido pela divisão da RBS e C-BM (Anderson; Domsh, 1993).

2.4 Leituras espectrais

A aquisição espectral foi realizada utilizando o equipamento *Ocean Optics* constituído de dois sensores: STS-VIS e Flame que abrangem um intervalo espectral de 350 a 2200 nm. Os dados foram coletados em triplicado para cada subamostra, e o espectro médio foi usado como entrada do modelo. A calibração do equipamento foi realizada uma placa branca que reflete 100% da luz, feita com sulfato de bário. Esse sensor estava conectado a um computador para os registros de cada leitura por meio de um software *OceanView* responsável pelas leituras. Foi realizado um pré tratamento dos dados, sendo retirados os ruídos entre 339.88 até 349.65 nm e entre 873.63 e 1017.41nm de forma manual. O banco de dados possui 2203 ondas analisadas. A média das leituras triplicadas foi usada como valor final do atributo.

2.5 Modelos de Classificação e Avaliação Estatística

Para modelagem, modelos foram treinados globalmente (Tabela 1), usando o conjunto de dados completo (todos os tipos de uso do solo combinados) para cada indicador, com o objetivo de desenvolver uma estrutura aplicável a lotes heterogêneos. A avaliação do modelo foi realizada usando validação cruzada 10 vezes. As análises foram realizadas em Python, utilizando as bibliotecas Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn e SciPy para manipulação dos dados, visualização gráfica, análise multivariada, pré-processamento espectral e modelagem supervisionada. Os dados foram organizados a partir de planilhas em formato Excel, contendo, para cada amostra, a identificação do uso do solo e os respectivos dados laboratoriais e espectrais.

Tabela 1 - Hiperparâmetros usados para os modelos de aprendizado de máquina.

Modelos	Técnica	Hiperparâmetros
LR	Regressão logística (sklearn.linear_model.LogisticRegression)	max_iter=5000; multi_class='multinomial'; solver='lbfgs'; class_weight='balanced'; random_state=42; demais parâmetros: padrão
RF	Floresta aleatória (sklearn.ensemble.RandomForestClassifier)	n_estimators=500; random_state=42; n_jobs=-1; class_weight='balanced_subsample'; demais parâmetros: padrão

SVM	Máquina de vetor de suporte (sklearn.svm.SVC)	kernel='rbf'; C=10; gamma='scale'; class_weight='balanced'; random_state=42; demais parâmetros: padrão
LGBM	Light Gradient Boosting Machine (lightgbm.LGBMClassifier)	n_estimators=800; learning_rate=0.05; num_leaves=31; subsample=0.8; colsample_bytree=0.8; random_state=42; class_weight='balanced'; verbose=-1; demais parâmetros: padrão
HGB	Classificador HistGradientBoosting (sklearn.ensemble.HistGradientBoostingClassifier)	max_iter=300; learning_rate=0.05; max_depth=6; random_state=42; demais parâmetros: padrão
PLS-DA	Análise discriminante por mínimos quadrados parciais (sklearn.cross_decomposition.PLSRegression)	scale=False; n_components definido adaptativamente em cada fold como $\min(10, n_{\text{train}} - 1, n_{\text{features}})$; a predição da classe foi obtida por argmax da matriz dummy predita.

Para os atributos microbiológicos do solo, foram considerados C-BM, RBS e qCO₂. Inicialmente, foi realizada análise exploratória para avaliar o comportamento individual e conjunto dessas variáveis.

Para os dados espectrais, foram construídas curvas médias por classe de uso do solo a partir da reflectância média em cada comprimento de onda, acompanhadas da faixa de variação expressa pelo desvio padrão. As assinaturas espectrais foram apresentadas ao longo das regiões VIS, NIR e SWIR, permitindo avaliar o comportamento global das curvas e a separação visual entre os usos do solo.

A estrutura multivariada dos dados foi investigada por meio da Análise de Componentes Principais (PCA), aplicada separadamente aos dados laboratoriais e aos dados espectrais. Antes da PCA, os dados foram padronizados por meio do StandardScaler, de forma a garantir comparabilidade entre variáveis de diferentes magnitudes. A PCA foi utilizada para reduzir a dimensionalidade dos dados e projetar as amostras em um espaço bidimensional, permitindo avaliar padrões de agrupamento e sobreposição entre os usos do solo. Além da PCA, foi empregada a Análise Discriminante Linear (LDA) para avaliar a separação supervisionada entre os usos do solo com base nos dados laboratoriais e espectrais. Diferentemente da PCA, a LDA incorporou a informação prévia das classes, permitindo identificar o potencial discriminante das variáveis quando o objetivo é maximizar a separação entre os grupos. Os resultados foram apresentados em gráficos bidimensionais com as amostras coloridas conforme o uso do solo.

Para a etapa de classificação supervisionada dos usos do solo foi utilizado primeira derivada de Savitzky-Golay (SG1). Os modelos avaliados ao longo do fluxo analítico incluíram

diferentes algoritmos de aprendizado de máquina (Tabela 2). A avaliação foi realizada por validação cruzada estratificada em 10 folds, utilizando como métricas Accuracy, F1-score, Precision e Recall.

Tabela 2 - Relação dos modelos de aprendizado de máquina, utilizados na classificação.

Sigla	Modelo de aprendizado de máquina	Referência
RF	Floresta aleatória	(Breiman, 2001)
SVM	Máquina de vetor suporte	(Cortes e Vapnik, 1995)
LGBM	Light Gradient Boosting Machine	(Ke et al., 2017)
LR	Regressão logística	(Cox, 1958)
HGB	Gradient boosting baseado em histogramas	(Friedman, 2001)
PLS-DA	Análise discriminante por mínimos quadrados parciais	(Barker e Rayens, 2003)

Os dados de Accuracy, F1-score, Precision e Recall foram examinados por meio de análise de variância (ANOVA), usando o teste de Scott-Knott em um nível de significância de 5%. Para a comparação gráfica do desempenho dos modelos, foram gerados boxplots com base nos resultados obtidos em cada fold. As métricas foram organizadas em dois conjuntos gráficos, sendo um composto por Accuracy e F1-score, e outro por Precision e Recall, de forma a melhorar a visualização e a comparação entre os modelos dentro de cada pré-processamento. As tabelas de resultados foram salvas separadamente por pré-processamento e por modelo, contendo os valores de cada fold e a média final, permitindo posterior construção dos gráficos comparativos dos modelos de melhor desempenho.

3 RESULTADOS E DISCUSSÃO

3.1 Análise de Componentes Principais aplicada aos dados microbianos laboratoriais e aos dados espectrais

A Figura 3 demonstra a comparação entre o PCA dos dados espectrais e o PCA dos indicadores microbianos laboratoriais. No PCA dos dados espectrais, Figura 3A, verificou-se que o PC1 explicou 83,40% da variância total e o PC2 explicou 9,96%, totalizando cerca de 93,36% da variância explicada nos dois primeiros componentes. Apesar dessa alta variância acumulada, observa-se forte sobreposição entre as classes de uso do solo, na qual os tipos de uso do solo não formaram agrupamentos compactos e nitidamente separados no espaço multivariado.

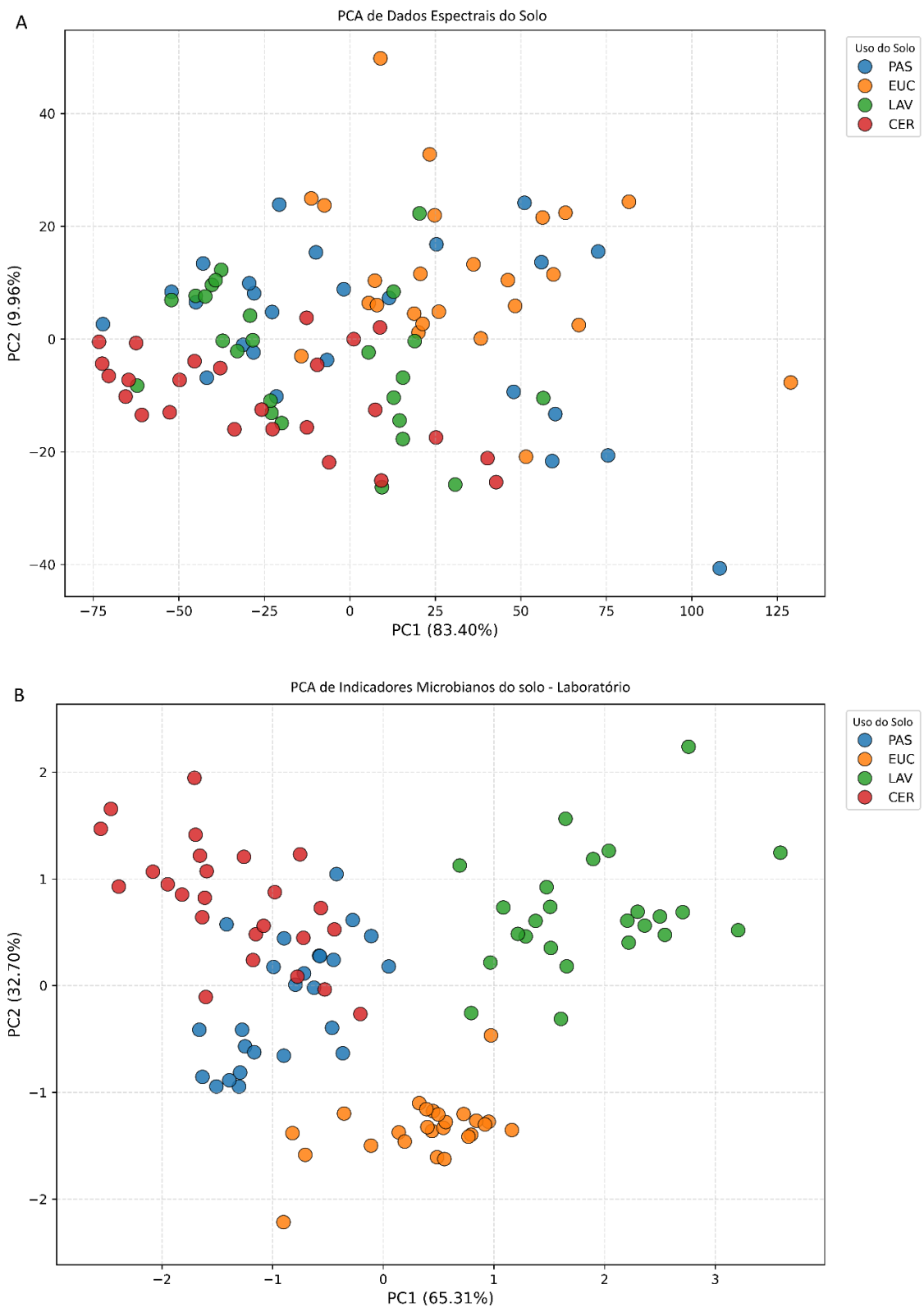


Figura 3 - Comparação entre a análise de componentes principais (PCA) dos dados espectrais do solo (A) e dos indicadores microbiológicos laboratoriais em diferentes tipos de uso do solo (B).

O comprimento de onda detectado pelo sensoriamento espectral é advindo da luz refletida por sólidos, líquidos ou gases, o qual, em solos, fornece informações físicas, químicas e biológicas, incluindo materiais orgânicos e inorgânicos presentes, sendo assim, a espectroscopia de reflectância não mede uma propriedade específica, fazendo-se necessário a utilização de modelos para melhorar a precisão da estimativa (Shin *et al.*, 2025). Sinalizando, portanto, que no espectro bruto, a discriminação entre os usos do solo ainda é limitada quando analisada por uma técnica não supervisionada como o PCA.

Visualmente, nota-se apenas uma tendência de agrupamento parcial ao longo do PC1. As amostras de cerrado (CER) concentram-se predominantemente em valores negativos, enquanto os sistemas manejados eucalipto (EUC) e pastagem (PAS), possuem parte das amostras deslocadas para valores positivos. Deste modo, nota-se uma ampla interseção entre os grupos. A análise de PCA é um método não supervisionado e linear, seu objetivo é diminuir a dimensionalidade dos dados e maximizar a variância global, assim, não capturando relações não lineares entre bandas (Uddin *et al.*, 2020).

Por refletir as propriedades intrínsecas do solo (textura, mineralogia, matéria orgânica e umidade), os dados espectrais não conseguem separar usos do solo pelo comportamento natural dos dados, pois eles ocupam a mesma nuvem no espaço de componentes (Lacerda *et al.*, 2016; Wang *et al.*, 2022), existindo assim a possibilidade de semelhança entre essas propriedades presentes no solo dos sistemas de uso da propriedade. Esse comportamento indica que, embora exista estrutura espectral no conjunto de dados, ela não é suficientemente forte, no espectro original, para promover uma separação clara dos sistemas de uso por meio dos dois primeiros componentes principais. Ainda assim, segundo (Shokr *et al.*, 2024) mesmo apresentando espectros com formas semelhantes, existem diferenças que aparecem em faixas específicas de 1150–2500 nm. Esses resultados reforçam a necessidade de empregar métodos supervisionados e estratégias de pré-processamento espectral para explorar de forma mais eficiente o potencial discriminatório das assinaturas espectrais.

No PCA dos indicadores microbianos laboratoriais, Figura 3B, o PC1 explicou 65,31% e o PC2 32,70% da variância, somando aproximadamente 98,01% da variância total. Nesse caso, a separação entre os grupos é visualmente mais evidente do que no PCA espectral. As amostras de EUC formam um agrupamento relativamente compacto em valores mais baixos de PC2, enquanto as de LAV se concentram em valores positivos de PC1, mostrando distinção clara em relação aos demais grupos. As amostras de CER tendem a ocupar a porção de PC1 negativo e PC2 positivo, enquanto PAS aparece em posição intermediária, com alguma

sobreposição principalmente com CER. Ainda assim, a estrutura geral dos dados laboratoriais mostra-se mais organizada e com maior capacidade de discriminação entre os usos do solo.

Os dados laboratoriais, por possuir metodologia específica, captam indicadores característicos e muito sensíveis e discriminantes a qualquer perturbação no manejo do solo (Mondal *et al.*, 2024; Abanto-Rodríguez *et al.*, 2026). Em um estudo de Liu *et al.* (2018), esses indicadores foram os que melhor distinguiram os tipos de uso do solo, superando atributos físicos e químicos. Em uma análise de PCA, Meena *et al.* (2021) mostraram que componentes principais associados a biomassa microbiana e respiração basal do solo explicam grande parte da variância entre os usos do solo.

Em síntese, a comparação entre os dois gráficos indica que os atributos microbiológicos laboratoriais apresentaram maior poder de separação entre os usos do solo do que os dados espectrais brutos.

3.2 Análise discriminante linear dos dados espectrais e microbiológicos laboratoriais do solo

Ao utilizar métodos supervisionados, na análise discriminante linear (LDA) nota-se que a separação no painel espectral ficou muito concentrada em LD1 (81,59%), enquanto LD2 explicou 14,63% (Figura 4A). Observa-se então, que grande parte da discriminação espectral ocorreu em um eixo dominante, ou seja, a diferença entre as classes foi capturada fortemente por uma direção linear, e isso revela que esses espectros das classes diferem, em termos discriminantes, “quase unidimensional” (Li *et al.*, 2019). Antes da classificação por algoritmos, o LDA é utilizado para redução supervisionada da dimensionalidade de um grande volume de dados, determinando uma projeção linear que otimiza o poder discriminatório dos dados, reduzindo a variância dentro de cada classe ao mesmo tempo em que aumenta a distância entre os diferentes grupos (Li *et al.*, 2019; Shambulinga *et al.*, 2020). Revelando assim, maior capacidade de separação entre os tipos de usos do solo do que a observada na PCA dos dados espectrais. Em suma, esses dados contêm informações discriminantes relevantes que podem ser eficientemente explorados por métodos supervisionados, reforçando seu potencial para a classificação.

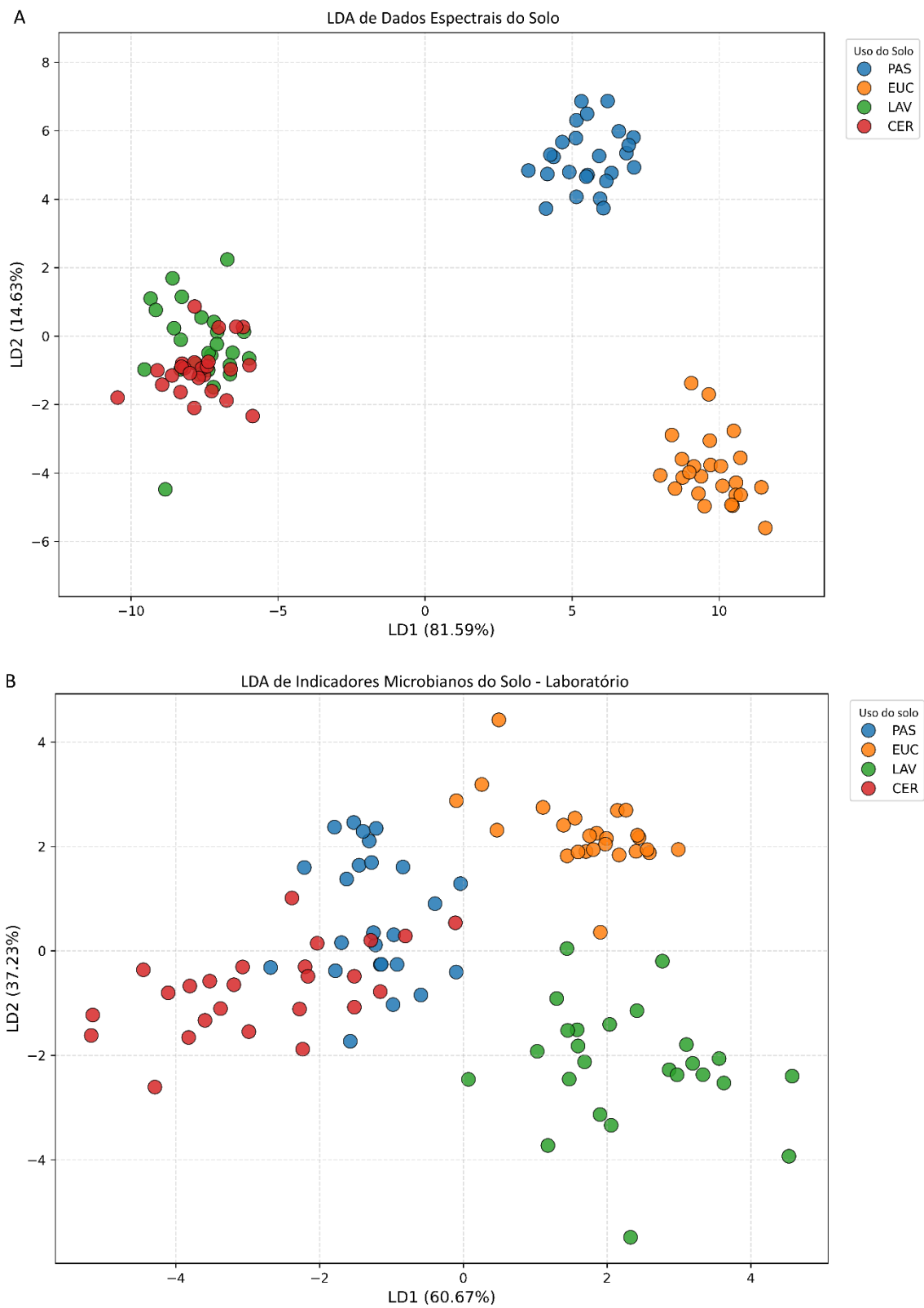


Figura 4 - Comparação entre análise discriminante linear (LDA) dos dados espectrais do solo (A) e dos atributos microbiológicos laboratoriais em diferentes tipos de uso do solo (B).

Os manejos PAS e EUC apresentaram agrupamentos bem definidos e espacialmente separados, enquanto CER e LAV mostraram maior proximidade e sobreposição parcial (Figura 4A). Áreas de CER e LAV são distintas do ponto de vista de qualidade microbiana do solo (Zago *et al.*, 2018), entretanto, os resultados apresentados demonstram o contrário.

Alguns trabalhos evidenciam que, solos submetidos a usos diferentes podem ter curvas espectrais semelhantes, sendo diferenciados apenas em faixas específicas do NIR-SWIR. Dessa maneira, ao avaliar o comportamento parecido das assinaturas espectrais médias, métodos multivariados tendem a aproximar grupos próximos e às vezes sobrepô-los, principalmente em solos com material de origem similar (Starks *et al.*, 2021; Shokr *et al.*, 2022; Greschuk *et al.*, 2022). Isso pode ser observado e explicado detalhadamente no item 3.3 deste estudo, na qual a reflectância dos usos do solo CER - LAV são próximas. Outro ponto importante é entender que a análise discriminante linear transforma dados em um subespaço que maximiza a separação entre médias de grupos a partir de matrizes de variância e covariância, ou seja, apenas reproduz correlações numéricas, não interpreta qualidade ecológica (Zhao *et al.*, 2024). No item 3.4 deste trabalho, há evidências que os usos do solo com maiores taxas de respiração são LAV e em seguida CER, dessa forma possivelmente existiu uma aproximação numérica estatística. Gelsomino e Azzellino (2011) utilizaram a análise discriminante para agrupar comunidades bacterianas, determinadas por variáveis biológicas, e perceberam que o agrupamento foi guiado por gradientes numéricos. Não por rótulos de “boa” ou “má” qualidade. Meena *et al.* (2021), mostraram que em diferentes usos do solo, índices microbianos são fortes discriminantes, organizando-os em gradientes de atividade e eficiência. A proximidade dos dados das área CER e LAV deste estudo, também pode ser resultado de duas décadas de manejos regenerativos que vêm sendo utilizados na mesma propriedade, que pode ter contribuído em partes para esses resultados.

Para os indicadores microbianos laboratoriais (Figura 4B), diferentemente do painel espectral, no laboratorial a separação ficou distribuída de forma mais equilibrada entre os dois eixos: com LD1 explicando 60,67% e LD2 37,23%. Neste caso, embora a LDA também tenha promovido discriminação entre os grupos, a distribuição das amostras foi visualmente menos segregada, com maior interseção entre algumas classes.

Essa distribuição equilibrada e a sutil sobreposição entre classes ocorre devido à natureza dinâmica e multifatorial das comunidades microbianas. Sendo assim, respondem simultaneamente à variabilidade local dos recursos, a chamada heterogeneidade espacial (Mondal *et al.*, 2024), aos gradientes ambientais amplos e ao manejo do solo. Isso resulta em

padrões metabólicos parcialmente sobrepostos entre diferentes usos do solo, como observado entre CER e PAS (Zhu *et al.*, 2020; Patoine *et al.*, 2022).

Tanto os dados espectrais quanto os atributos microbiológicos laboratoriais apresentaram capacidade de discriminar os usos do solo, porém com padrões distintos de separação. Esses achados sugerem que diferentes abordagens analíticas revelam dimensões complementares da condição do solo. A comparação entre as análises discriminantes indica que a similaridade entre os usos do solo depende do conjunto de variáveis analisado (Gama-Rodrigues *et al.*, 2018).

3.3 Assinaturas espectrais médias dos diferentes usos do solo nas regiões VIS, NIR e SWIR

Ao observar a Figura 5 que trata das assinaturas espectrais médias dos diferentes usos do solo nas regiões VIS, NIR e SWIR, percebe-se uma variação expressa pelo desvio padrão.

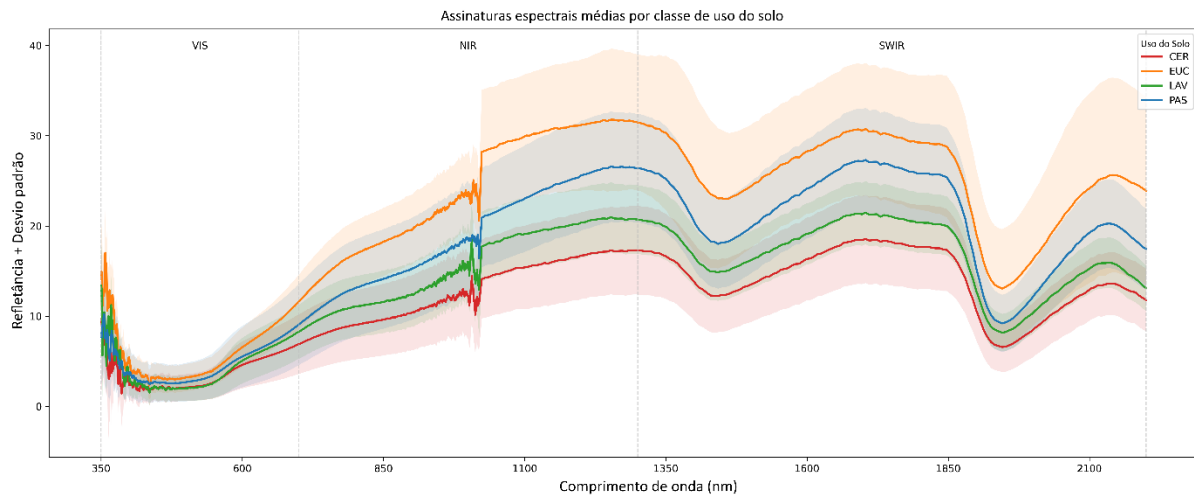


Figura 5. Assinaturas espectrais médias dos diferentes usos do solo nas regiões VIS, NIR e SWIR, com variação expressa pelo desvio padrão.

As quatro classes apresentaram padrões espectrais semelhantes na forma geral da curva, mas com diferenças consistentes na intensidade da reflectância ao longo das regiões VIS, NIR e SWIR. De modo geral, EUC apresentou os maiores valores de reflectância em praticamente todo o espectro, seguido por PAS, depois LAV, enquanto CER exibiu os menores valores médios. Isso indica que os usos do solo não diferiram apenas em feições pontuais, mas também no comportamento global da resposta espectral.

A reflectância está diretamente associada aos atributos presentes no solo. Maiores teores de matéria orgânica, óxidos de ferro e argila levam à diminuição da reflectância, deixando o solo mais escuro e absorvendo mais luz do que reflete (Figura 6). Solos mais claros, com menos matéria orgânica, possuem a tendência de refletir mais (Lopez, 2009; Demattê *et al.*, 2017;

Pearlshtien *et al.*, 2020; Çullu *et al.*, 2024). Outro fator importante é a umidade de solo, que pode diminuir a reflectância e alterar as feições espectrais, principalmente no NIR e SWIR (Dhawale *et al.*, 2021). Segundo Silva *et al.* (2026), solos com mata nativa (Cerrado) apresentam maiores teores de matéria orgânica e maior retenção e disponibilidade de água do que sistemas mais perturbados.

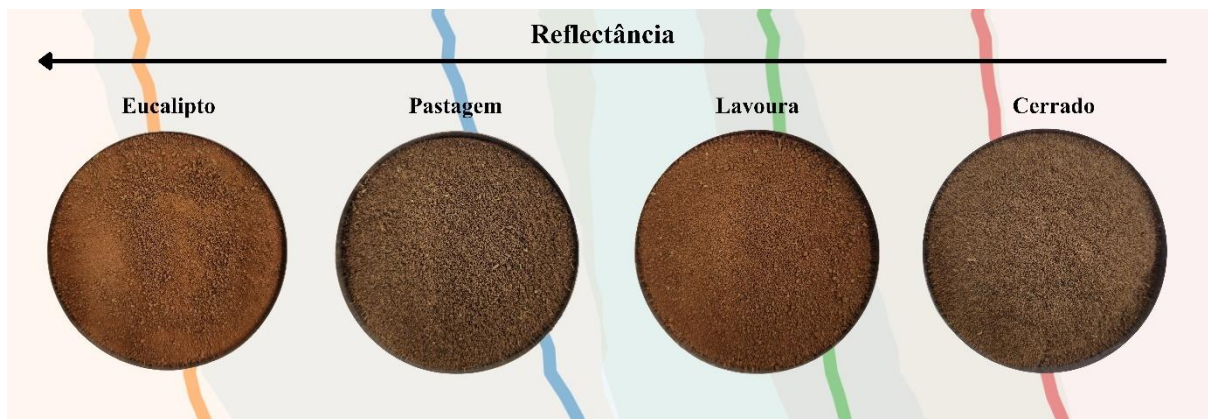


Figura 6. Ilustração visual das cores do solo em relação a reflectância.

A presença das faixas sombreadas mostra a variação interna de cada classe, permitindo observar que, apesar da dispersão entre amostras, a ordenação geral das curvas foi relativamente estável. Em vários trechos do espectro, especialmente do NIR ao SWIR, as curvas médias mantiveram separação visual consistente, sugerindo que essas regiões concentram informação relevante para discriminar os usos do solo.

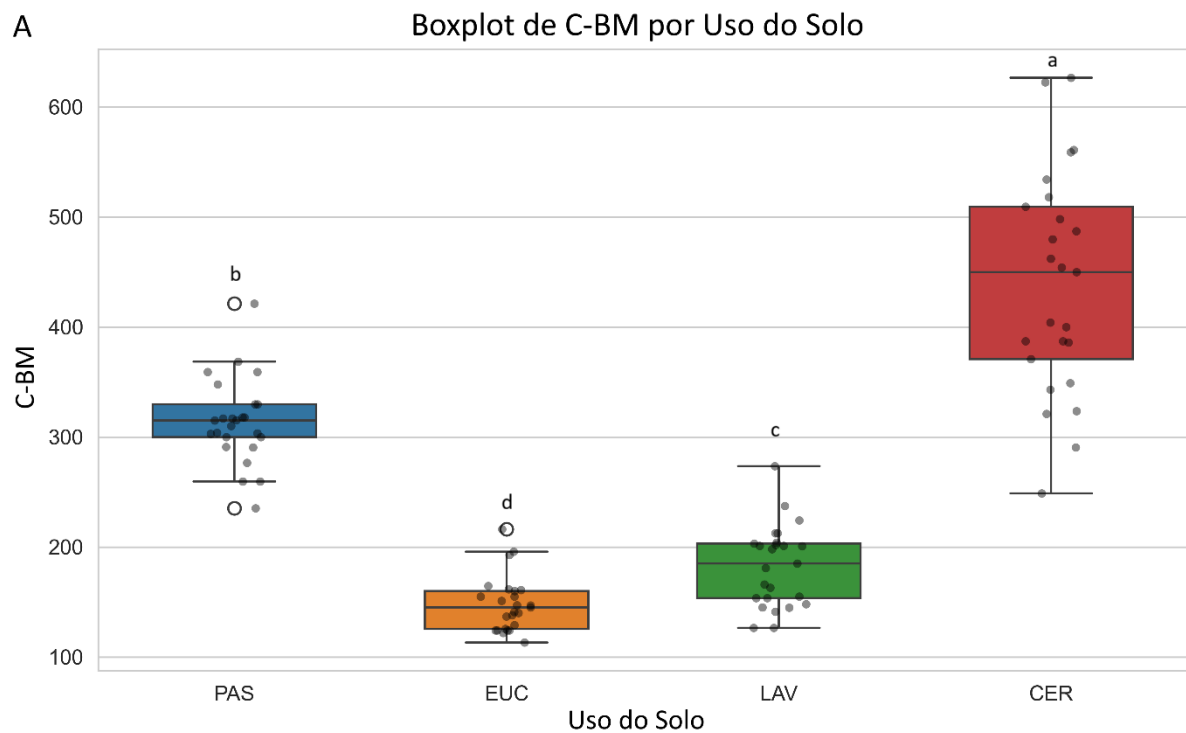
Na região do visível (VIS), as curvas apresentaram maior oscilação nas menores bandas (Figura 5). Shin *et al.*, 2025 ressaltam que o espectro VIS necessita de pré-processamento adequado, pois é fortemente afetado por ruído, deriva de baseline e espalhamento. Contudo, após essa região inicial, se percebe uma tendência de separação entre as classes. Isso sugere que mesmo no VIS existem diferenças no comportamento óptico dos solos entre os usos avaliados.

No NIR, a separação entre os grupos ficou mais evidente (Figura 5). As curvas tornam-se mais suaves e a distância entre as classes aumenta, especialmente entre EUC e CER. Como essa região é sensível a propriedades estruturais e ao comportamento geral da reflectância do material (Koirala *et al.*, 2023), é possível que os usos do solo tenham contribuído para essa diferenciação. A distinção entre PAS e LAV também se torna mais visível nessa faixa. No SWIR, as curvas mantêm o padrão de separação e exibem feições de absorção mais pronunciadas (Figura 5). Essa região é particularmente importante porque costuma concentrar informação relacionada a componentes minerais, matéria orgânica, água e grupos funcionais do

solo (Bannari *et al.*, 2018; Angelopoulou *et al.*, 2020; Rodríguez-Pérez *et al.*, 2021). A permanência da diferenciação entre as classes no SWIR reforça que os usos do solo influenciam a resposta espectral de maneira consistente ao longo do espectro, e não apenas em bandas isoladas.

3.4 Indicadores microbiológicos de qualidade do solo para avaliação da aderência à distribuição normal

De maneira geral, o solo sob a área CER apresentou os maiores valores de CBM (Figura 7A), enquanto da área LAV apresentou os maiores valores de qCO_2 e RBS (Figuras 7 B e 7C). O EUC demonstrou os menores valores de CBM e RBS. Enquanto na área PAS foi verificado comportamento intermediário em alguns indicadores, valores próximos aos observados no CER (Figura 7A).



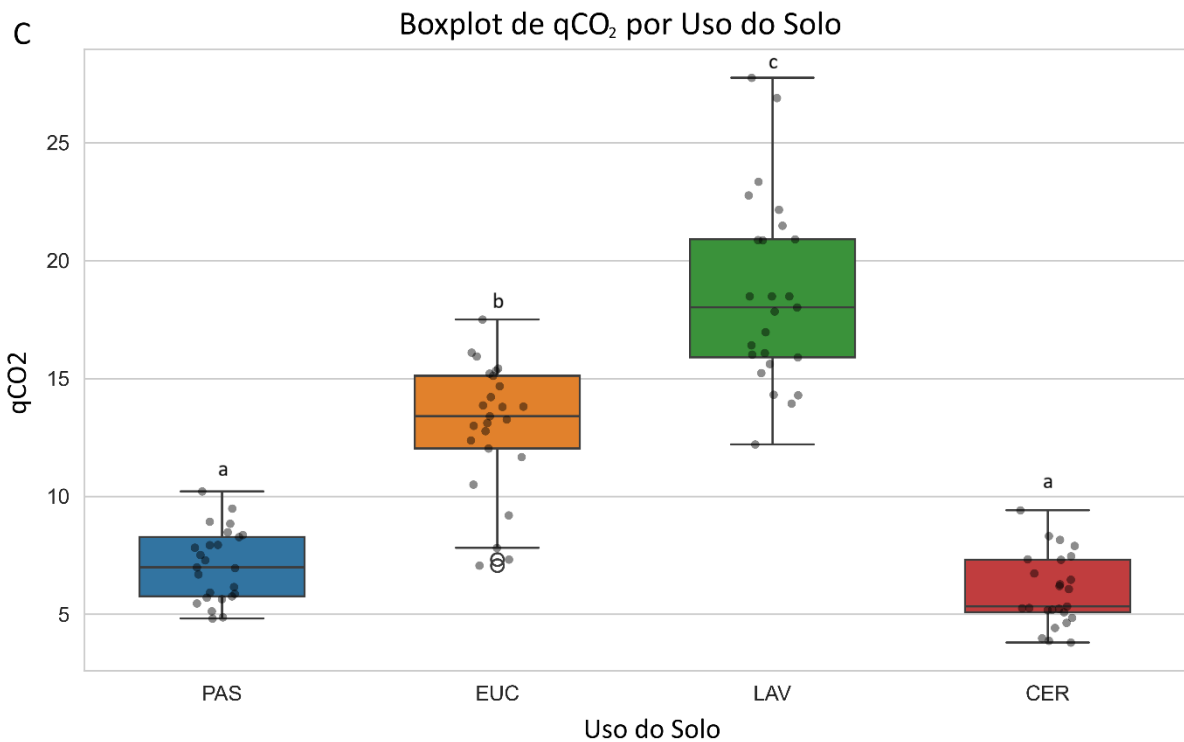
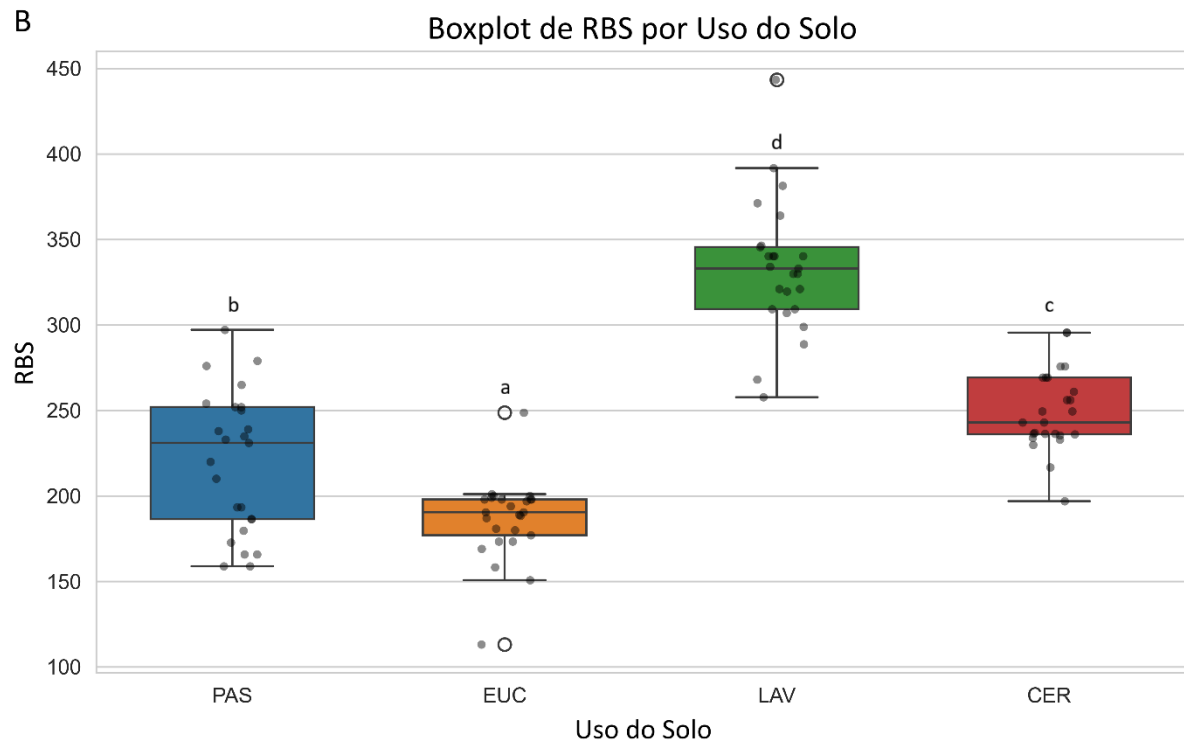


Figura 7. Boxplots dos indicadores microbianos de qualidade do solo em diferentes usos do solo.

Sistemas de uso do solo com pastagens bem manejadas e bom aporte de resíduos, podem manter valores de C-BM e RBS próximos a mata nativa (CER), sugerindo preservação parcial

do funcionamento microbiológico do solo, embora geralmente com características de atributos estruturais abaixo do ecossistema nativo (Lopes *et al.*, 2010). Em sentido oposto, as áreas EUC e LAV influenciaram na redução relativa da biomassa microbiana do solo (Figura 7A). Mesmo utilizando-se de práticas de agricultura regenerativa, estudos indicam que valores de CBM são melhores do que sistemas convencionais, porém não superam necessariamente sistemas nativos, não perturbados e em equilíbrio ecológico. O sistema pode apresentar limitações em diversidade de rotações, quantidade de resíduos e grau de distúrbio físico do solo (Gonçalves *et al.*, 2019; Stieven *et al.*, 2020).

Para RBS (Figura 7B), os maiores valores foram observados no solo da área LAV, seguida pela área CER, que também apresentou valores relativamente elevados. Cabe ressaltar que maiores valores de respiração não indicam, necessariamente, melhor qualidade do solo, mas podem refletir maior mineralização da matéria orgânica, liberação de nutrientes e melhor saúde do solo, especialmente quando acompanhados de maior biomassa microbiana (Semenov *et al.*, 2025). Dessa forma, a interpretação da RBS deve ser realizada em conjunto com o qCO_2 (Figura 7C). Nesse contexto, o fato de a LAV apresentar simultaneamente valores elevados de RBS e qCO_2 sugere uma microbiota metabolicamente mais ativa, porém menos eficiente em uso do carbono, possivelmente submetida a um distúrbio recente ou estresse ambiental (Wardle; Ghani, 2018).

A avaliação do qCO_2 normalmente é interpretada como indicador de estresse microbiano e de eficiência. Valores mais altos podem indicar maior gasto energético para manutenção por unidade de biomassa, menor eficiência metabólica ou ambiente mais perturbado (Feng *et al.*, 2021). Neste contexto, a área LAV influenciou nos maiores valores para este indicador, e a área sob CER proporcionou os menores valores (Figura 7C). Podendo-se inferir que o sistema agrícola promoveu maior estresse à microbiota (Uchôa *et al.*, 2025), enquanto na área sob CER verificou-se maior estabilidade funcional.

3.5 Pré-processamentos e modelos de aprendizado de máquina na classificação multiclasse dos usos do solo

Observa-se que não houve diferença estatística nas métricas de accuracy, F1-score, precision e recall entre (Floresta aleatória, Máquina de vetor de suporte, Gradient Boosting, Regressão logística, Hist Gradient Boosting e Análise discriminante por mínimos quadrados parciais) (Tabela 3). O desempenho dos modelos de classificação sob o pré-processamento SGI

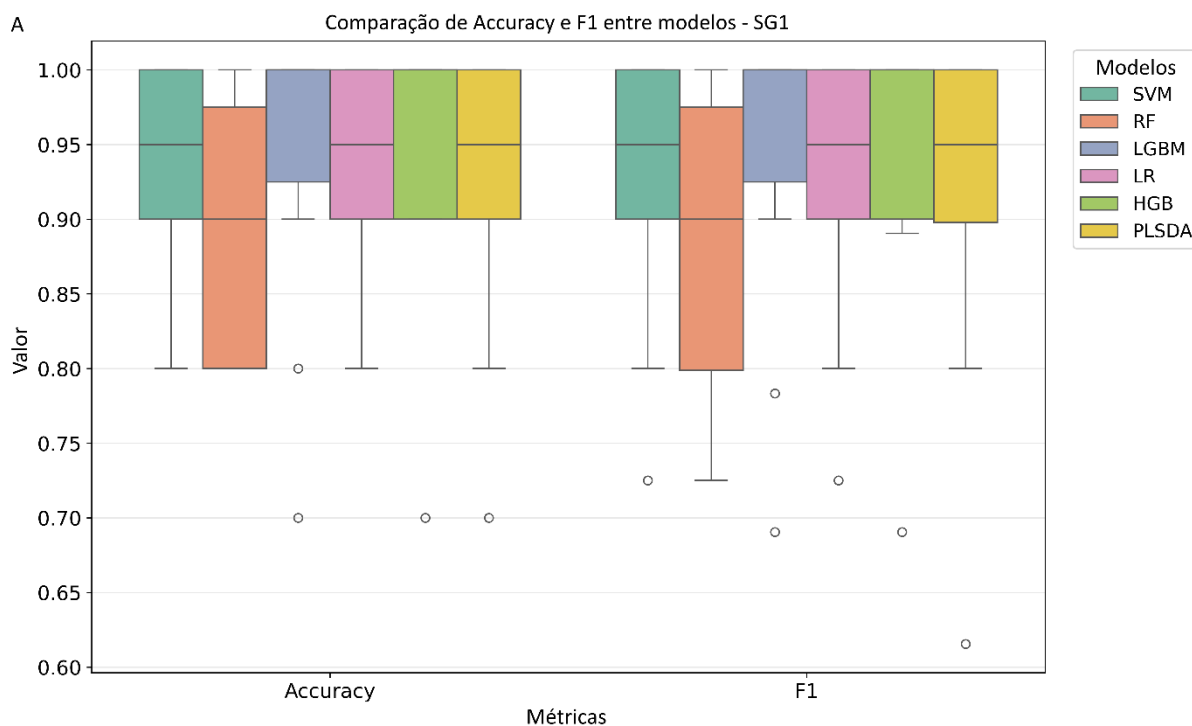
com base nas métricas Accuracy, F1-score, Precision e Recall estão representados na (Figura 8).

Tabela 3 - Desempenho médio dos modelos de aprendizado de máquina submetidos ao pré-processamento SG1, com base nas métricas Accuracy, F1-score, Precision e Recall.

Modelo	Accuracy	F1-score	Precision	Recall
RF	0.890000 a	0.880600 a	0.912600 a	0.890000 a
PLS-DA	0.920000 a	0.911200 a	0.926300 a	0.920000 a
SVM	0.930000 a	0.922500 a	0.934600 a	0.930000 a
LR	0.930000 a	0.922500 a	0.934600 a	0.930000 a
HGB	0.940000 a	0.937300 a	0.951600 a	0.940000 a
LGBM	0.940000 a	0.938000 a	0.955800 a	0.940000 a
Prob>F	0.5925 ^{ns}	0.5703 ^{ns}	0.7518 ^{ns}	0.5925 ^{ns}
CV (%) =	7.40	8.28	7.42	7.40
Média Geral:	0.9250000	0.9186833	0.9359167	0.9250000

Médias seguidas pela mesma letra na coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

^{ns} não significativo pelo teste F a 5% de probabilidade.



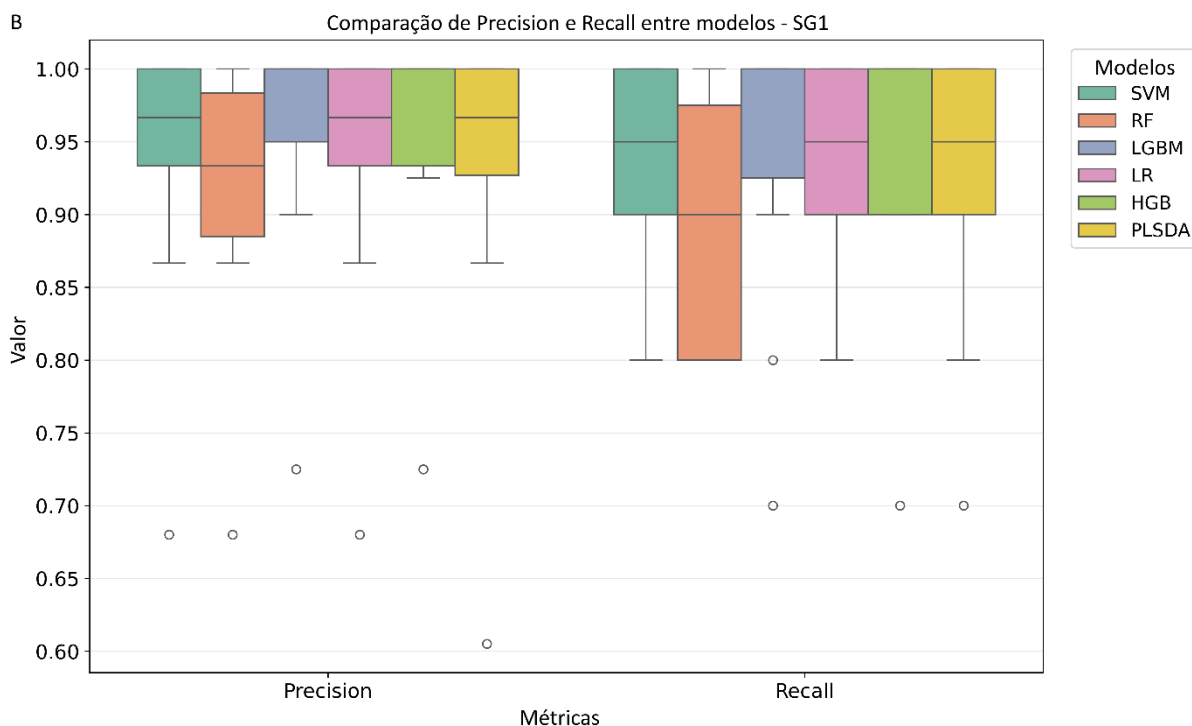


Figura 8. Desempenho do pré-processamento SG1 e dos modelos de machine learning com base nas métricas Accuracy, F1-score, Precision e Recall.

Percebe-se que dentro do pré-processamento SG1, os modelos apresentaram desempenho elevado e com claras diferenças hierárquicas (Figura 8A). De forma geral, HGB e LGBM foram os modelos mais precisos, ambos com Accuracy média de 0,94, além de F1-score médio de 0,938 para HGB e 0,9374 para LGBM. Em seguida os modelos LR e SVM ficaram ambos com Accuracy média de 0,93 e F1-score médio de 0,9225. O modelo PLS-DA também apresentou desempenho alto com Accuracy = 0,92 e F1 = 0,9113, confirmando que métodos clássicos da quimiometria continuam competitivos nesse tipo de dado. O RF foi o que apresentou o menor desempenho entre os seis modelos selecionados, com Accuracy média de 0,89 e F1 médio de 0,8807.

Outro ponto importante é que, além da média elevada, HGB e LGBM também mostraram boa consistência geral entre os folds, ainda que com alguma oscilação natural (Figura 8A). Para LR e SVM, mesmo ligeiramente abaixo na média, mostraram comportamento bastante competitivo e estável. Isso é relevante porque indica que, após o pré-processamento por primeira derivada de Savitzky-Golay ajusta polinômios locais, obtendo derivadas suavizadas, o que melhora resoluções de bandas, assim modelos mais simples ou lineares conseguem extrair informação discriminante importante das assinaturas espectrais (Agustika *et al.*, 2022).

Esses resultados reforçam que o SG1 foi decisivo para ampliar a separação entre as classes. A derivada provavelmente destacou melhor as mudanças locais da curva espectral e reduziu parte do efeito de baseline e de variações menos informativas do espectro bruto (Roger *et al.*, 2022). Com isso, os modelos passaram a trabalhar com um sinal mais discriminante. O fato de HGB e LGBM liderarem sugere que os padrões que diferenciam os usos do solo não são apenas lineares, mas envolvem relações mais complexas entre as bandas processadas. Ao mesmo tempo, o bom desempenho de LR e SVM (Figura 8A) evidencia que, após o pré-processamento adequado, a estrutura entre as classes se tornou suficientemente organizada para também ser capturada por modelos mais simples.

Do ponto de vista aplicado, isso indica que o tratamento espectral por SG1 tornou os dados suficientemente informativos para discriminar os usos do solo com alta precisão, o que fortalece o potencial da espectroscopia como ferramenta não destrutiva de classificação. Além disso, o bom desempenho simultâneo de modelos de naturezas distintas sugere que a diferenciação entre os usos do solo não dependeu de um único algoritmo, mas de um padrão espectral consistente que pôde ser explorado por diferentes abordagens supervisionadas.

Precision e Recall também confirmam o alto desempenho geral dos modelos com LGBM e HGB apresentaram os melhores resultados, com medianas mais altas e distribuição mais concentrada nas faixas superiores, indicando maior capacidade de acerto e melhor estabilidade entre os folds (Figura 8B).

Por fim, este estudo abre caminhos para o desenvolvimento de metodologias em larga escala e de forma não destrutiva. Sugere-se, para trabalhos futuros, a validação desses modelos de machine learning, visando diminuir a dependência de análises laboratoriais que se utilizam de reagentes químicos insalubres e acelerar a tomada de decisão no campo. Também abre caminho para outras classificações e até mesmo predição da atividade microbiana do solo, garantindo assim um futuro sustentável para a agricultura.

4 CONCLUSÕES

Foi possível classificar os tipos de usos do solo com alta eficiência da integração entre a espectroscopia VIS-NIR-SWIR e os indicadores microbiológicos por meio de modelos de machine learning.

Com técnica de pré-processamento SGI com os modelos testados-Floresta aleatória, Máquina de vetor suporte, Gradient Boosting, Regressão logística, Hist Gradient Boosting e Análise discriminante, por mínimos quadrados parciais, foram capazes de identificar padrões complexos e correlações entre as bases de dados usadas, em especial HGB e LGBM, alcançando Accuracy média de 0,94.

Para a PCA se vê um melhor desempenho dos dados laboratoriais em comparação aos dados espectrais, dessa forma a discriminação entre os usos do solo por meio de dados espectrais ainda é limitada quando analisada por uma técnica não supervisionada. Em contraste a isso, com uma técnica supervisionada, como LDA, os dados espectrais possuíram maior capacidade de separação entre as classes, apresentando melhor desempenho.

As assinaturas espectrais médias dos usos do solo, mostraram que o solo sob a área de Eucalipto (EUC) obteve mais reflectância em relação ao solo da área de Cerrado (CER), com mostrou-se com menor reflectância. A permanência da diferenciação entre as classes nas bandas do NIR e SWIR, reforça que os usos do solo influenciam a resposta espectral de maneira consistente ao longo do espectro. Para os indicadores microbianos de qualidade do solo, a área de Cerrado (CER) influenciou no maior valor de carbono da biomassa microbiana (C-BM) e menor quociente metabólico (qCO_2). O solo sob pastagem (PAS) proporcionou resultados inferiores para esses mesmos indicadores de qualidade, em relação aos encontrados na área de Cerrado (CER), porém superiores aos encontrados no solo das áreas de lavoura (LAV) e eucalipto (EUC). A área de lavoura influenciou nos maiores valores de respiração basal do solo (RBS) e quociente metabólico (qCO_2).

REFERÊNCIAS

- ABANTO-RODRÍGUEZ, C. et al. Soil functioning indicators decline with land-use intensification in the Peruvian Amazon: evidence from Ucayali. **Frontiers in Soil Science**, v. 6, p. 1741629, 2026. DOI: <https://doi.org/10.3389/fsoil.2026.1741629>.
- AGUSTIKA, D. K. et al. Fourier transform infrared spectrum pre-processing technique selection for detecting PYLCV-infected chilli plants. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 278, p. 121339, 2022. DOI: <https://doi.org/10.1016/j.saa.2022.121339>.
- ALEF, K.; NANNIPIERI, P. *Methods in Applied Soil Microbiology and Biochemistry*. Elsevier Science, 1995. 576 p.
- ANDERSON, J.P.E.; DOMSH, K.H. The metabolic quotient (qCO₂) as a specific activity parameter to assess the effects of environment conditions, such as pH, on te microbial biomass of forest soils. **Soil Biololy and Biochemisty**, v.25, n.3, p.393-395, 1993.
- ANGELOPOULOU, T. et al. From laboratory to proximal sensing spectroscopy for soil organic carbon estimation—A review. **Sustainability**, v. 12, n. 2, p. 443, 2020. DOI: <https://doi.org/10.3390/su12020443>.
- BANNARI, A. et al. Sentinel-MSI VNIR and SWIR bands sensitivity analysis for soil salinity discrimination in an arid landscape. **Remote Sensing**, v. 10, n. 6, p. 855, 2018. DOI: <https://doi.org/10.3390/rs10060855>.
- BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of Chemometrics: A Journal of the Chemometrics Society**, v. 17, n. 3, p. 166-173, 2003. DOI: <https://doi.org/10.1002/cem.785>.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001. DOI: <https://doi.org/10.1023/A:1010933404324>.
- CAMARGO, F. F. et al. A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. **Remote Sensing**, v. 11, n. 13, p. 1600, 2019. DOI: <https://doi.org/10.3390/rs11131600>.
- CHAUDHRY, H. et al. Evaluating the soil quality index using three methods to assess soil fertility. **Sensors**, v. 24, n. 3, p. 864, 2024. DOI: <https://doi.org/10.3390/s24030864>.

CHEN, L. et al. Advances in hyperspectral and diffraction imaging for agricultural applications. **Agriculture**, v. 15, n. 16, p. 1775, 2025. DOI: <https://doi.org/10.3390/agriculture15161775>.

CHEN, Y. et al. Evolutions and managements of soil microbial community structure drove by continuous cropping. **Frontiers in Microbiology**, v. 13, p. 839494, 2022. DOI: <https://doi.org/10.3389/fmicb.2022.839494>.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, v. 20, n. 3, p. 273-297, 1995. DOI: <https://doi.org/10.1007/BF00994018>.

COX, D.R. The Regression Analysis of Binary Sequences. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 20, p. 215-232, 1958. <https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>.

ÇULLU, M. A. et al. Rapid characterization of soil horizons for different soil series utilizing Vis-NIR spectral information. **Geoderma Regional**, v. 38, p. e00853, 2024. DOI: <https://doi.org/10.1016/j.geodrs.2024.e00853>.

CUNHA, F. F.; MAGALHÃES, F. F.; CASTRO, M. A. Métodos para estimativa da evapotranspiração de referência para Chapadão do Sul - MS. **Engenharia na Agricultura**, v. 21, p. 159-172, 2013. DOI: <https://doi.org/10.13083/reveng.v21i2.346>.

DE OLIVEIRA, E. P. et al. Microbiological attributes as indicators of soil quality in coffee growing systems in Southwest Bahia, Brazil. **Brazilian Journal of Microbiology**, v. 55, n. 2, p. 1841-1852, 2024. DOI: <https://doi.org/10.1007/s42770-024-01279-0>.

DEMATTE, J. A. M. et al. Genesis and properties of wetland soils by VIS-NIR-SWIR as a technique for environmental monitoring. **Journal of Environmental Management**, v. 197, p. 50-62, 2017. DOI: <https://doi.org/10.1016/j.jenvman.2017.03.014>.

DHAWALE, N. M. et al. Evaluating the precision and accuracy of proximal soil vis-NIR sensors for estimating soil organic matter and texture. **Soil Systems**, v. 5, n. 3, p. 48, 2021. DOI: <https://doi.org/10.3390/soilsystems5030048>.

DOS SANTOS, R. G. et al. Classification of soybean genotypes based on physiological clustering (PCA+ k-means) integrated with VIS-NIR hyperspectral data and machine learning models. **Smart Agricultural Technology**, v. 13, p. 101702, 2026. DOI: <https://doi.org/10.1016/j.atech.2025.101702>.

FAO. Standard operating procedure for soil microbial biomass (carbon): chloroform fumigation-extraction method. Roma: **FAO**, 2024. Disponível em: <https://doi.org/10.4060/cc9423en>. Acesso em: 8 jun. 2026.

FATEMI, F.; DEHDASHTI, A.; JANNATI, M. Implementation of chemical health, safety, and environmental risk assessment in laboratories: a case-series study. **Frontiers in public health**, v. 10, p. 898826, 2022. DOI: <https://doi.org/10.3389/fpubh.2022.898826>.

FENG, J. et al. Soil microbial trait-based strategies drive metabolic efficiency along an altitude gradient. **ISME communications**, v. 1, n. 1, p. 71, 2021. DOI: <https://doi.org/10.1038/s43705-021-00076-2>.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, p. 1189-1232, 2001. DOI: <https://doi.org/10.1214/aos/1013203451>.

GAMA-RODRIGUES, E. F. et al. Atributos físicos, químicos e microbiológicos dos solos sob diferentes usos em topossequências no Noroeste Fluminense. **Agrária-Revista Brasileira de Ciências Agrárias**, v. 13, n. 3, p. 1-9, 2018. DOI: <https://doi.org/10.5039/agraria.v13i3a5543>.

GELSOMINO, A.; AZZELLINO, A. Multivariate analysis of soils: microbial biomass, metabolic activity, and bacterial-community structure and their relationships with soil depth and type. **Journal of Plant Nutrition and Soil Science**, v. 174, n. 3, p. 381-394, 2011. DOI: <https://doi.org/10.1002/jpln.200900267>.

GONÇALVES, V. A. et al. Biomassa e atividade microbiana de solo sob diferentes sistemas de plantio e sucessões de culturas. **Revista de Ciências Agrárias**, 2018. DOI: <https://doi.org/10.22491/rca.2019.2611>.

GRESCHUK, L. T. et al. Combining spectral ranges for soil discrimination: A case study in the State of Maranhão-Brazil. **Geoderma Regional**, v. 29, p. e00507, 2022. DOI: <https://doi.org/10.1016/j.geodrs.2022.e00507>.

HEDMAN, H. D. Development of an Exploratory Simulation Tool: Using Predictive Decision Trees to Model Chemical Exposure Risks and Asthma-like Symptoms in Professional Cleaning Staff in Laboratory Environments. **Laboratories**, v. 3, n. 1, p. 2, 2026. DOI: <https://doi.org/10.3390/laboratories3010002>.

KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. **Advances in neural information processing systems**, v. 30, 2017. Disponível em: [LightGBM: A Highly Efficient Gradient Boosting Decision Tree](#). Acesso em: 10 Jun. de 2026.

KHANGURA, R. et al. Regenerative agriculture—A literature review on the practices and mechanisms used to improve soil health. **Sustainability**, v. 15, n. 3, p. 2338, 2023. DOI: <https://doi.org/10.3390/su15032338>.

KOIRALA, B. et al. A multisensor hyperspectral benchmark dataset for unmixing of intimate mixtures. **IEEE Sensors Journal**, v. 24, n. 4, p. 4694-4710, 2023. DOI: <https://doi.org/10.1109/jsen.2023.3343552>.

LACERDA, M. P. et al. Tropical texture determination by proximal sensing using a regional spectral library and its relationship with soil classification. **Remote Sensing**, v. 8, n. 9, p. 701, 2016. DOI: <https://doi.org/10.3390/rs8090701>.

LI, X.; ZHANG, L.; YOU, J. Locally weighted discriminant analysis for hyperspectral image classification. **Remote Sensing**, v. 11, n. 2, p. 109, 2019. DOI: <https://doi.org/10.3390/rs11020109>.

LIMA, H. S. et al. Machine learning-based soil quality assessment for enhancing environmental monitoring in iron ore mining-impacted ecosystems. **Journal of environmental management**, v. 356, p. 120559, 2024. DOI: <https://doi.org/10.1016/j.jenvman.2024.120559>.

LIU, D. et al. Soil physicochemical and microbial characteristics of contrasting land-use types along soil depth gradients. **Catena**, v. 162, p. 345-353, 2018. DOI: <https://doi.org/10.1016/j.catena.2017.10.028>.

LIU, Y. et al. Recent Advances in Raman Spectral Classification with Machine Learning. **Sensors**, v. 26, n. 1, p. 341, 2026. DOI: <https://doi.org/10.3390/s26010341>.

LÓPEZ, L. R. et al. Pedologia quantitativa: espectrometria VIS-NIR-SWIR e mapeamento digital de solos. Univ. de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, Brasil, 2009. DOI: <https://doi.org/10.11606/d.11.2009.tde-23062009-140151>.

MEENA, A; RAO, K. S. Assessment of soil microbial and enzyme activity in the rhizosphere zone under different land use/cover of a semiarid region, India. **Ecological Processes**, v. 10, n. 1, p. 16, 2021. DOI: <https://doi.org/10.1186/s13717-021-00288-3>.

MONDAL, B. P. et al. Spatial variability of soil microbiological properties under different land use systems. **African Journal of Agricultural Research**, v. 20, n. 9, p. 825-839, 2024. DOI: <https://doi.org/10.5897/ajar2024.16720>.

PATOINE, Guillaume et al. Drivers and trends of global soil microbial carbon over two decades. **Nature Communications**, v. 13, n. 1, p. 4195, 2022. DOI: <https://doi.org/10.1038/s41467-022-31833-z>.

PEARLSHTIEN, D. H.; BEN-DOR, E. Effect of organic matter content on the spectral signature of iron oxides across the VIS–NIR spectral region in artificial mixtures: An example from a red soil from Israel. **Remote Sensing**, v. 12, n. 12, p. 1960, 2020. DOI: <https://doi.org/10.3390/rs12121960>.

RODRÍGUEZ-PÉREZ, J. R. et al. Estimating soil properties and nutrients by visible and infrared diffuse reflectance spectroscopy to characterize vineyards. **Agronomy**, v. 11, n. 10, p. 1895, 2021. DOI: <https://doi.org/10.3390/agronomy11101895>.

ROGER, J. M.; MALLET, A.; MARINI, F. Preprocessing NIR spectra for aquaphotomics. **Molecules**, v. 27, n. 20, p. 6795, 2022. DOI: <https://doi.org/10.3390/molecules27206795>.

SANTOS, H. G. et al. Sistema Brasileiro de Classificação de Solos. Brasília: **Embrapa**. 2018.

SEMENOV, M. V. et al. Microbiological indicators for assessing the effects of agricultural practices on soil health: a review. **Agronomy**, v. 15, n. 2, p. 335, 2025. DOI: <https://doi.org/10.3390/agronomy15020335>.

SHAMBULINGA, M.; SADASHIVAPPA, G. Supervised hyperspectral image classification using SVM and linear discriminant analysis. **International Journal of Advanced Computer Science and Applications**, v. 11, n. 10, 2020. DOI:10.14569/IJACSA.2020.0111050.

SHEN, L. et al. Hyperspectral estimation of soil organic matter content using different spectral preprocessing techniques and PLSR method. **Remote Sensing**, v. 12, n. 7, p. 1206, 2020. DOI: <https://doi.org/10.3390/rs12071206>.

SHIN, S. K.; LEE, S. J.; PARK, J. H. Prediction of soil properties using vis-nir spectroscopy combined with machine learning: A review. **Sensors**, v. 25, n. 16, p. 5045, 2025. DOI: <https://doi.org/10.3390/s25165045>.

SHOKR, M. S. et al. Integration of VIS–NIR spectroscopy and multivariate technique for soils discrimination under different land management. **Land**, v. 13, n. 12, p. 2056, 2024. DOI: <https://doi.org/10.3390/land13122056>.

SILVA, M. L. N. et al. Water Retention and Availability in an Ultisol Under an Integrated Crop–Livestock–Forest System in the Matopiba Region, Brazil. **Land**, v. 15, n. 3, p. 360, 2026. DOI: <https://doi.org/10.3390/land15030360>.

SOUZA, C. M. et al. Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat archive and Earth Engine. *Remote Sens* 12 (17): 2735. **Landsc Urban Plan**, v. 86, n. 3-4, p. 219-225, 2020. DOI: <https://doi.org/10.3390/rs12172735>.

STARKS, P. J.; FORTUNA, A. M. Comparable Discrimination of Soil Constituents Using Spectral Reflectance Data (400–1000 nm) Acquired with Hyperspectral Radiometry. **Soil Systems**, v. 5, n. 3, p. 45, 2021. DOI: <https://doi.org/10.3390/soilsystems5030045>.

STIEVEN, A. C. et al. Atributos do solo em sistemas diferenciados de uso e manejo do solo. **Colloquium Agrariae**, Presidente Prudente, v. 16, n. 2, p. 1-15, 2020. DOI: <https://doi.org/10.5747/ca.2020.v16.n2.a354>.

SUDARSHAN, S. et al. Effects of intensive agricultural management practices on soil biodiversity and implications for ecosystem functioning: A review. **Int. J. Res. Agron**, v. 7, p. 166-169, 2024. DOI: <https://doi.org/10.33545/2618060x.2024.v7.i9sc.1446>.

TIRUNEH, G. A. et al. Use of soil spectral reflectance to estimate texture and fertility affected by land management practices in Ethiopian tropical highland. **Plos One**, v. 17, n. 7, p. e0270629, 2022. DOI: <https://doi.org/10.1371/journal.pone.0270629>.

UDDIN, M. P. et al. Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification. **International Journal of Remote Sensing**, v. 42, n. 1, p. 286-321, 2021. DOI: <https://doi.org/10.1080/01431161.2020.1807650>.

VAIRAVAN, C. et al. Hyperspectral imaging of soil and crop: A review. **Journal of Experimental Agriculture International**, v. 46, n. 1, p. 48-61, 2024. DOI: <https://hal.science/hal-05145978>.

VANCE, E.D.; BROOKES, P.C.; JENKINSON, D.S. An extraction method for measuring soil microbial biomass C. **Soil Biology and Biochemistry**, v.9, n.6, p.703-707, 1987. DOI: [http://doi:10.1016/0038-0717\(87\)90052-6](http://doi:10.1016/0038-0717(87)90052-6).

WANG, Z.; DING, J.; ZHANG, Z. Estimation of soil organic matter in arid zones with coupled environmental variables and spectral features. **Sensors**, v. 22, n. 3, p. 1194, 2022. DOI: <https://doi.org/10.3390/s22031194>.

WARDLE, D. A.; GHANI, A. A tale of two theories, a chronosequence and a bioindicator of soil quality. **Soil biology and biochemistry**, v. 121, p. A3-A7, 2018. DOI: <https://doi.org/10.1016/j.soilbio.2018.01.005>.

Zago, L. M. S. et al. Biochemical activity in Brazilian Cerrado soils is differentially affected by perennial and annual crops. **Australian Journal of Crop Science**, v. 12, n. 2, p. 235-242, 2018. <https://doi.org/10.21475/ajcs.18.12.02.pne716>.

ZHAO, S. et al. Linear discriminant analysis. **Nature Reviews Methods Primers**, v. 4, n. 1, p. 70, 2024. DOI: <https://doi.org/10.1038/s43586-024-00346-y>.

ZHONG, L. et al. Multi-scale stereoscopic hyperspectral remote sensing estimation of heavy metal contamination in wheat soil over a large area of farmland. **Agronomy**, v. 13, n. 9, p. 2396, 2023. DOI: <https://doi.org/10.3390/agronomy13092396>.

ZHU, X. et al. The soil microbial carbon pump: From conceptual insights to empirical assessments. **Global Change Biology**, v. 26, n. 11, p. 6032-6039, 2020. DOI: <https://doi.org/10.1111/gcb.15319>.