

Universidade Federal de Mato Grosso do Sul  
Instituto de Física

Caíque Nery Rodrigues Porsch

FTIR Spectroscopy for Seed Specie  
(*Brachiaria Decumbens* vs *Brachiaria  
Ruziziensis*) Classification

Campo Grande  
2025

**Caíque Nery Rodrigues Porsch**

**FTIR Spectroscopy for Seed Specie (*Brachiaria Decumbens* vs *Brachiaria Ruziziensis*) Classification**

Trabalho de Conclusão de Curso apresentado como parte das atividades do Curso de Engenharia Física da Universidade Federal do Mato Grosso do Sul - UFMS, Campus Campo Grande, como requisito parcial para a obtenção do título de Bacharel em Engenharia Física.

Orientador(a) Prof. Dr. Cícero Rafael Cena da Silva

Campo Grande  
2025

## **AGRADECIMENTOS**

Agradeço ao meu orientador, Cícero Cena, pelos ensinamentos, pelas correções e, principalmente, pelo exemplo. Uma pessoa que fala tanto sobre a importância do equilíbrio na vida para que nunca nos esforcemos além do necessário e possamos preservar a saúde mental e que, mesmo assim, mostra com atitudes que é possível se dedicar de verdade aos nossos sonhos sem abrir mão do bem-estar.

À Camila e ao Thiago, Paloma, Miler minha profunda gratidão pela grande ajuda. Eles dedicaram seu tempo para me ensinar, analisar o que estava errado e me auxiliar, cada vez mais, na construção do meu futuro.

Aos meus amigos, que estiveram ao meu lado e compartilharam as dificuldades do curso e da iniciação científica. Minha gratidão pelos momentos para reclamar da vida, pela parceria e pelo apoio mútuo que tornaram toda a caminhada mais fácil.

Aos professores do INFI, minha sincera gratidão pela dedicação, pelo esforço contínuo em aprimorar o trabalho e pelo compromisso com um ensino cada vez melhor.

Agradeço também aos meus pais e minha avó, que sempre estiveram ao meu lado, me apoiando e incentivando a seguir meus sonhos. E, por fim, agradeço a todos que fizeram parte e aos que ainda farão da minha trajetória de vida.

*Precisamos pensar direitinho! Hmm... hmmhmm... Ah! Entendi  
(League of Leagens/Lulu)*

## RESUMO

A identificação de sementes de *Brachiaria* é normalmente realizada por profissionais treinados por meio de inspeção visual, o que pode levar a erros de classificação e impactar operações comerciais. Este estudo explora a espectroscopia no infravermelho com transformada de Fourier (FTIR) como um método automático para classificar sementes de *Brachiaria decumbens* e *Brachiaria ruziziensis*. As sementes comerciais foram limpas e moídas antes da análise. O pó das sementes foi caracterizado utilizando um espectrômetro FTIR Agilent Cary 630 com acessório ATR, cobrindo a faixa de 4000–650 cm<sup>-1</sup>. Os dados de FTIR foram pré-processados usando o método Modified Standard Normal Variate (MSNV), e a classificação das sementes foi realizada por meio da Análise de Componentes Principais (PCA), que revelou agrupamentos e separações claras. Esses resultados sugerem que essa metodologia oferece uma ferramenta robusta para análise agrícola.

**Palavras-chave:** FTIR espectroscopia; PCA; Sementes; cultivar; *Brachiaria*

## ABSTRACT

The identification of *Brachiaria* seeds is typically performed by trained personnel through visual inspection, which can lead to misclassification and impact commercial operations. This study explores Fourier-transform infrared (FTIR) spectroscopy as an automatic method for classifying *Brachiaria decumbens* and *Brachiaria ruziziensis* seeds. Commercial seeds were cleaned and ground before analysis. The seed powder was characterized using an Agilent Cary 630 FTIR spectrometer with an ATR accessory, covering the 4000–650 cm<sup>-1</sup> range. FTIR data was preprocessed using the Modified Standard Normal Variate (MSNV) method, and seed classification was performed through Principal Component Analysis (PCA), which revealed clear clustering and separation. These results suggest that this methodology provides a robust tool for agricultural analysis.

**Keywords:** FTIR spectroscopy; PCA; Seeds; cultivar; *Brachiaria*

## SUMÁRIO

1	INTRODUÇÃO . . . . .	7
2	METODOLOGIA . . . . .	9
3	RESULTADOS . . . . .	12
	REFERÊNCIAS . . . . .	18

## 1 INTRODUÇÃO

A pecuária brasileira, um dos pilares do agronegócio nacional, é sustentada por vastas áreas de pastagem, estimadas em mais de 99 milhões de hectares. Neste cenário, as graminéas do gênero *Brachiaria* são fundamentais, servindo como a principal base forrageira para rebanhos e como um componente essencial nos modernos sistemas de Integração Lavoura-Pecuária (ILP). Diferentes espécies e cultivares possuem papéis agronômicos distintos: a *Brachiaria decumbens*, por exemplo, é valorizada por sua rusticidade e adaptação a solos menos férteis, enquanto a *Brachiaria ruziziensis* é preferida em sistemas de ILP para a formação de palhada e restauração de nutrientes, exigindo solos de maior fertilidade[1].

O sucesso na formação da pastagem depende diretamente da qualidade e da correta identificação das sementes utilizadas. Contudo, a indústria de sementes de forrageiras enfrenta um significativo gargalo logístico e técnico. O método oficial para validação de lotes, o "padrão-ouro" estabelecido pelo Ministério da Agricultura e Pecuária (MAPA) através das Regras para Análise de Sementes (RAS), é a "Análise de Mistura de Sementes". Este processo, embora rigoroso, é notoriamente demorado, podendo levar em média 28 dias para a emissão de um laudo. A origem desta demora é dupla: reside tanto na complexa análise de pureza, que depende da inspeção visual de taxonomistas treinados, quanto, principalmente, no tempo biológico inerente aos testes de germinação e viabilidade[1].

A etapa de identificação visual, além de demorada, é um ponto crítico de falha. As sementes de espécies distintas, como *B. decumbens* e *B. ruziziensis*, são morfologicamente similares, o que torna a classificação manual propensa a erros. Um erro de classificação (misclassification) tem impactos comerciais e agronômicos severos, podendo levar ao fracasso total da lavoura caso uma espécie seja semeada em condições edafoclimáticas inadequadas para seu desenvolvimento. Este cenário demanda o desenvolvimento de métodos analíticos alternativos que ofereçam maior rapidez, objetividade e precisão.

Neste contexto, a Espectroscopia de Infravermelho com Transformada de Fourier (FTIR), acoplada a ferramentas quimiométricas, surge como uma alternativa tecnológica promissora. Esta técnica analítica fornece uma "impressão digital molecular" da amostra, baseada na vibração das suas ligações químicas. O FTIR é capaz de analisar a composição bioquímica global da semente (lipídios, proteínas, carboidratos) de forma rápida (análises em menos de 5 minutos) de baixo custo operacional. Estudos recentes demonstram que a metodologia é robusta, conseguindo diferenciar espécies morfologicamente semelhantes, como *B. decumbens* e *B. ruziziensis*, com "clara separação e agrupamento" através da Análise de Componentes Principais (PCA)[6].

Mais significativamente, o potencial do FTIR excede a simples classificação de espécies. Pesquisas indicam que a técnica pode medir diretamente as reservas de energia metabólica

(proteínas e lipídios) da semente, permitindo a classificação de lotes de alto e baixo vigor com índices de precisão próxima a 96%[3]. A espectroscopia vibracional, portanto, não ataca apenas o gargalo da análise de pureza, mas tem o potencial de predizer o resultado do demorado teste de germinação. Com a abertura regulatória das novas RAS para métodos alternativos, mediante "comprovação científica de sua equivalência", a validação desta tecnologia torna-se estratégica.

## 2 METODOLOGIA

Este capítulo detalha os procedimentos experimentais e as técnicas de análise de dados empregadas para a discriminação das sementes de *Brachiaria decumbens* e *Brachiaria ruziziensis*. A metodologia abrange desde o preparo das amostras até a análise estatística multivariada dos dados espectrais.

As amostras de sementes das duas cultivares (*B. decumbens* e *B. ruziziensis*), adquiridas diretamente do fornecedor, foram submetidas a um rigoroso processo de limpeza. O objetivo era garantir que a análise espectral refletisse apenas a composição química das sementes. Inicialmente, as sementes foram lavadas em água destilada, procedimento que visava remover impurezas macroscópicas, como poeira, solo aderido e pequenos detritos vegetais, a exemplo de fragmentos de galhos. Após a lavagem, as sementes foram secas e inspecionadas visualmente. Detritos remanescentes que não foram eliminados pela água, como cascas aderidas ou outras sementes, forammeticulosamente separados de forma manual, com o auxílio de pinças de laboratório. Para garantir a homogeneidade da amostra e aumentar a superfície de contato para a análise espectroscópica, as sementes limpas foram trituradas em um moinho de facas. Este processo de moagem transforma as sementes em um pó fino, o que é essencial para obter espectros de alta qualidade e reproduzibilidade na técnica de ATR.

A análise da composição química das amostras moídas foi realizada por espectroscopia no infravermelho com transformada de Fourier (FTIR). Foi utilizado um espectrofotômetro modelo Agilent Cary 630, dotado de um acessório de Reflectância Total Atenuada (ATR). O uso do ATR é vantajoso por permitir a análise direta de amostras sólidas em pó com mínimo preparo adicional, garantindo bom contato entre a amostra e o cristal de ATR. Para cada cultivar, foram analisadas 40 amostras (replicatas). Os espectros foram coletados na faixa de frequência de 4000 a 650 cm<sup>-1</sup>, com resolução de 4 cm<sup>-1</sup> e um total de 12 varreduras (scans) por amostra fazendo o background na atmosfera. A utilização de 12 scans visa melhorar a relação sinal-ruído (S/N) do espectro final, resultando em um espectro médio mais limpo e representativo, 12 varreduras por amostra.

Os espectros brutos obtidos pelo FTIR contêm, além da informação química relevante, ruídos e variações físicas (como espalhamento de luz) que podem prejudicar a análise. Para extrair informações fidedignas, foi aplicada uma rotina de pré-processamento.

Embora os espectros tenham sido coletados de 4000 a 650 cm<sup>-1</sup>, a análise multivariada concentrou-se na região de 1800 a 650 cm<sup>-1</sup>. Esta faixa, conhecida como "região da impressão digital"(fingerprint region), é de particular interesse. Ela contém vibrações de deformação angular e de ligação (C-C, C-O, C-N) que formam um padrão espectral complexo e único para cada composto, sendo altamente específica. A seleção desta região [4] justifica-se por conter as maiores diferenças espectrais entre as amostras, aumentando

o potencial de classificação.

A primeira etapa do tratamento de dados foi a correção da linha de base (baseline correction). Este procedimento é para mitigar distorções e desvios no espectro, frequentemente causados por espalhamento de luz (especialmente em amostras em pó), aquecimento do detector ou outras derivações instrumentais. A correção reposiciona a linha de base do espectro, garantindo que as intensidades das bandas de absorção sejam comparáveis[2].

Após a correção da linha de base, os dados foram normalizados. A normalização é uma etapa importante para corrigir efeitos físicos indesejados, como diferenças na densidade da amostra ou no contato com o cristal de ATR, que afetam a intensidade total do espectro.

Foi aplicada a técnica robusta de Modified Standard Normal Variate (MSNV). Diferente de outras técnicas de outros métodos de pré-processamento, o MSNV é menos sensível a outliers no conjunto de dados, pois utiliza medidas estatísticas robustas: a Mediana (em vez da média) e o Desvio Absoluto Mediano (MAD) (em vez do desvio padrão)[9].

O MSNV é o cálculo para cada ponto de dado ou seja para cada número de onda no espectro conforme a Equação 1

$$\text{MSNV}(x_i) = \frac{x_i - \text{Mediana}(x)}{\text{MAD}(x)} \quad (\text{Equação 1})$$

Este método centraliza os dados (subtraindo a mediana) e os escala (dividindo pelo MAD), tornando os espectros comparáveis entre si.

Com os dados pré-processados, foram aplicadas técnicas de análise multivariada para identificar padrões e classificar as amostras.

Para garantir a integridade e a robustez do modelo de classificação, foi necessário identificar e remover amostras atípicas (outliers). Foi utilizado o método estatístico  $T^2$  de Hotelling [8].

Este teste é uma medida multivariada da distância de uma observação (amostra) em relação ao centro (média) do conjunto de dados, levando em consideração a covariância entre as variáveis (números de onda). Amostras que apresentaram um valor de  $T^2$  acima de um limite de confiança estabelecido (ex: 95% ou 99%) foram consideradas anômalas e removidas da análise subsequente[8].

Após a remoção de outliers, o conjunto de dados foi submetido à Análise de Componentes Principais (PCA). A PCA é uma técnica de análise exploratória e não supervisionada, amplamente utilizada na quimiometria[6].

Seu objetivo principal é reduzir a dimensionalidade de um conjunto de dados complexo (onde cada número de onda é uma variável) e identificar os padrões de variação subjacentes. A PCA transforma as variáveis originais (milhares de pontos de dados do espectro) em um novo conjunto de variáveis ortogonais (não correlacionadas) chamadas Componentes Principais (PCs)[3].

As primeiras PCs (PC1, PC2, etc.) capturam a maior parte da variância total dos dados. A PCA permite a visualização dos dados em gráficos de escores bidimensionais (ex: PC1 vs PC2), nos quais é possível observar a tendência de formação ou dispersões entre as amostras, indicando suas semelhanças ou diferenças químicas.

A simples visualização dos scores da PCA (os gráficos de pontos) apenas indica se os grupos se separam, mas não explica o porquê. Para interpretar quais características químicas são responsáveis pela separação observada, é indispensável realizar a análise dos gráficos de loadings (cargas) [8, 9].

Os loadings revelam qual a contribuição (ou peso) de cada variável original (cada número de onda do espectro FTIR) para a construção de uma determinada Componente Principal. Ao analisar os picos e vales no gráfico de loadings da PC que separou os grupos, é possível identificar exatamente quais bandas espectrais (e, consequentemente, quais grupos funcionais químicos) tiveram a maior influência na discriminação entre as sementes de *B. decumbens* e *B. ruziziensis*.

Por fim, para a classificação supervisionada das amostras, foi empregado o algoritmo Support Vector Machine (SVM). O SVM é um método de aprendizado de máquina que busca determinar uma separação com a fronteira de decisão linear entre as classes no espaço multidimensional dos dados. Para verificar a eficácia e a capacidade de generalização do modelo SVM, utilizou-se a técnica de validação cruzada Leave-One-Out (LOOCV). Neste método, o modelo é treinado iterativamente utilizando  $N - 1$  amostras e testado com a amostra restante, repetindo-se o processo para todas as amostras. O desempenho final do classificador foi avaliado por meio da Matriz de Confusão, a qual quantifica os acertos e erros do modelo, permitindo o cálculo das taxas de sensibilidade, especificidade e acurácia global da classificação[3].

### 3 RESULTADOS

# FTIR spectroscopy for seed specie (*Brachiaria Decumbens* vs *Brachiaria Ruziziensis*) classification

line 1: Caique Porsch 1<sup>st</sup>  
line 2: Federal University of Matogrosso do Sul  
line 3: Grupo de Optica e Fotonica  
line 4: caique.porsch@ufms.br

line 1: Camila Calvani 4<sup>th</sup>  
line 2: Federal University of Matogrosso do Sul  
line 3: Grupo de Optica e Fotonica  
line 4: calvanicamila@gmail.com

line 1: Afonso Henrique Silva Leite 2<sup>nd</sup>  
line 2: Federal Institute of Matogrosso do Sul  
line 3: Grupo de Optica e Fotonica  
line 4: afonso.leite@ifms.edu.br

line 1: Cicero Cena 5<sup>th</sup>  
line 2: Federal University of Matogrosso do Sul  
line 3: Grupo de Optica e Fotonica  
line 4: cicero.cena@ufms.br

line 1: Thiago França 3<sup>rd</sup>  
line 2: Federal University of Matogrosso do Sul  
line 3: Grupo de Optica e Fotonica  
line 4: t.franca@ufms.br

**Abstract**— The identification of *Brachiaria* seeds is typically performed by trained personnel through visual inspection, which can lead to misclassification and impact commercial operations. This study explores Fourier-transform infrared (FTIR) spectroscopy as an automatic method for classifying *Brachiaria decumbens* and *Brachiaria ruziziensis* seeds. Commercial seeds were cleaned and ground before analysis. The seed powder was characterized using an Agilent Cary 630 FTIR spectrometer with an ATR accessory, covering the 4000–650 cm<sup>-1</sup> range. FTIR data was preprocessed using the Modified Standard Normal Variate (MSNV) method, and seed classification was performed through Principal Component Analysis (PCA), which revealed clear clustering and separation. These results suggest that this methodology provides a robust tool for agricultural analysis.

**Keywords**— FTIR spectroscopy, PCA, Seeds, cultivar, *Brachiaria*.

## I. INTRODUCTION

*Brachiaria decumbens*, also known as *Brachiaria* or basilisk, is a forage grass that reaches heights between  $5.0 \times 10^1$  and  $1.0 \times 10^2$  cm. It is characterized by its high vegetative vigor and moderate tolerance to waterlogged soils, making it widely used in pastures to feed cattle. This creeping plant tends to spread quickly throughout the environment. Although it is not the most nutritious of forages, it has a reasonable balance of proteins and other nutrients, in addition to being highly resistant to pests and diseases [1].

*Brachiaria ruziziensis*, on the other hand, is another forage grass, although its use is more restricted. It has been utilized in areas of crop-livestock integration and requires more fertile soils for growth. Unlike *B. decumbens*, *B. ruziziensis* is more resistant to drought and is distinguished by its medium-tall height and upright growth. It is widely used after harvest in crop-livestock integration to restore nutrients to the soil, consuming a greater amount of nitrogen [1].

The main difference between *B. decumbens* and *B. ruziziensis* is that *B. decumbens* is more suitable for planting in less fertile land, although it does not produce well during periods of drought. In contrast, *B. ruziziensis*, with its greater tolerance to drought, is often used to replenish soil nutrients, especially in straw production, but requires fertile soils.

Although the benefits of these plants are clear and their use in livestock, agronomy, and physiology is well established, distinguishing the two species when mature is straightforward. However, when it comes to their seeds, they are morphologically similar, which makes separation difficult. This becomes a problem when purchasing these seeds, especially when stored in large quantities .

A widely used method for classifying samples, especially due to its effectiveness in fast and low-cost analyses, is Fourier Transform Infrared Spectroscopy (FTIR). When combined with Principal Component Analysis (PCA), this method allows for the efficient separation of seeds, mitigating the negative impact that a wrong choice of grass could have on the soil, whether due to unsuitability for the environment or the search for specific results in mass production for cattle feeding or soil management [3-5].

## II. MATERIALS AND METHODS

Two packages of *Brachiaria* seeds (*decumbens* and *ruziziensis*) were purchased directly from the supplier. The seeds underwent a rigorous cleaning process, initially being washed in distilled water to remove impurities such as twigs and soil. Subsequently, the remaining debris was manually separated using tweezers. After cleaning, the seeds were crushed in a knife mill to increase the contact surface for subsequent analysis by Fourier Transform Infrared Spectroscopy (FTIR).

FTIR analysis was performed using an Agilent Cary 630 spectrophotometer, equipped with an attenuated total reflectance (ATR) accessory. Forty samples from each group were analyzed in the frequency range of 4000 to 650 cm<sup>-1</sup>, with a resolution of 4 cm<sup>-1</sup>, and 12 scan. Special attention was given to the region from 1800 to 650 cm<sup>-1</sup>, known as the fingerprint region, due to its unique spectral characteristics that allow for better distinction between samples [4]. This region was selected because it contains the largest spectral differences relevant to classification.

For data pre-processing, the baseline was first corrected to mitigate possible distortions introduced by the equipment. Next, normalization was applied using Modified Standard Normal Variate (MSNV), a robust technique for adjusting data dispersion. MSNV is calculated by subtracting the median of each data point and then dividing the result by the

Median Absolute Deviation (MAD), thus normalizing the data set. The formula for MSNV can be expressed as:

$$\text{MSNV}(x_i) = \frac{x_i - \text{Median}(x)}{\text{MAD}(x)} \quad (1)$$

where  $x_i$  represents the individual data value,  $\text{Median}(x)$  is the median of the data set  $x$ , and  $\text{MAD}(x)$  is the median of the absolute differences from the median, providing a robust measure of data dispersion.

After data normalization, Principal Component Analysis (PCA) was applied, an unsupervised technique that reduces the dimensionality of data using an orthogonal linear transformation, identifying underlying patterns in the original variables. PCA facilitates the visualization of data in a two-dimensional graph, highlighting the relationships between samples, whether grouped or dispersed and allowing a better understanding of the variables that most contribute to variation in the data. Loading analysis is essential to interpret PCA, as it reveals which bands of the spectrum have the greatest influence on sample separation. Therefore, PCA must always be analyzed together with the loadings to achieve high precision when applying the methods [8,9].

In addition to the pre-processing and initial analysis steps, Hotelling's  $T^2$  statistical method was applied to identify and remove outliers, an essential procedure to guarantee the integrity of the data used in Principal Component Analysis (PCA) [8]. Hotelling's  $T^2$  is a multivariate measure that evaluates the distance of each sample in relation to the mean of the data set, taking into account the variance and covariance of the variables.

### III. RESULTS AND DISCUSSION

The average FTIR spectra of each sample collected are represented in Figure 1. This figure illustrates the standard average spectra along with their respective dispersions, providing a comprehensive overview of the spectral characteristics across different samples. The spectra exhibit notable absorption bands in specific regions: (i) 3700-3100  $\text{cm}^{-1}$ , (ii) 3000-2800  $\text{cm}^{-1}$ , and (iii) 1800-650  $\text{cm}^{-1}$  [1]. These spectral regions are indicative of various biochemical components, such as lipids, fatty acids, carbohydrates, and proteins.

In particular, the absorption peaks observed at 1558 and 1543  $\text{cm}^{-1}$  are attributed to the N-H deformation coupled with the C=N stretching of amides II, which are essential for identifying protein structures. Additionally, the band at 3317  $\text{cm}^{-1}$  is characteristic of vibrational modes associated with water and amide II groups, including O-H and N-H bonds [4]. This peak reflects the presence of moisture and protein interactions within the samples. Furthermore, medium-sized bands around 1748  $\text{cm}^{-1}$ , corresponding to C-O stretching and C-H bending vibrations, play a significant role in the spectral analysis presented in Figure 2.

Each of these spectral regions was critically evaluated to determine the optimal range for applying Principal Component Analysis (PCA). The selected ranges enable a detailed differentiation of the samples based on their chemical composition, enhancing the accuracy of the PCA results.

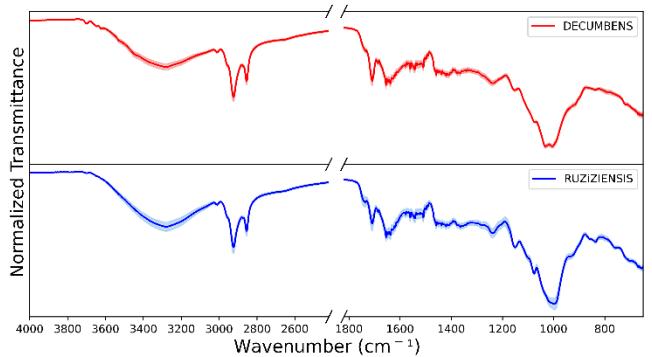


Fig. 1. Normalized spectra using MSNV for seed lots of the *B. decumbens* and *B. ruziziensis* classes.

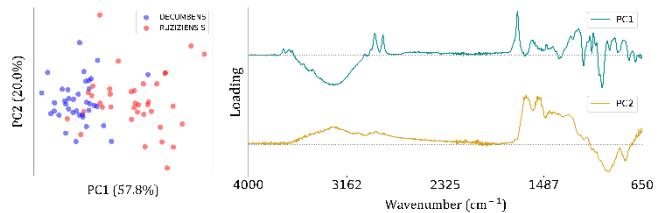


Fig. 2. Two-dimensional PCA containing PC1 (57.8%) and PC2 (20.0%) and the loadings of each PC in the range of 4000-650  $\text{cm}^{-1}$ .

When analyzing only PC1 and PC2, a clear separation between the two classes is observed. It is also noted that PC1 is the main separation factor. Upon examining the loadings for the entire region, it was found that the band at 3317  $\text{cm}^{-1}$ , related to water and amide II, has the highest load, suggesting that water, which is not inherent to the seeds, is influencing the result. To avoid this problem and improve the accuracy of seed analysis, it was decided to focus only on the region of 1800-650  $\text{cm}^{-1}$ , which is considered the fingerprint region, thus eliminating the influence of the water band.

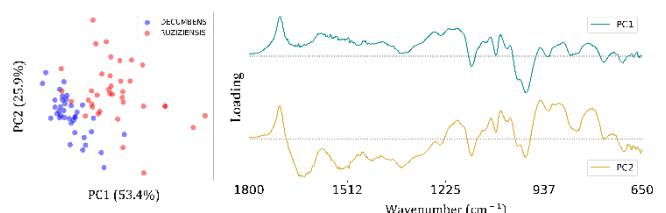


Fig. 3. Two-dimensional PCA plot showing PC1 (57.8%) and PC2 (20.0%), along with the loadings of each PC in the range of 1800-650  $\text{cm}^{-1}$ .

When comparing Figures 2 and 3, it is possible to observe a significant change in scores in relation to the total region, which was expected, given the sensitivity of PCA to changes in the analyzed variables. Now, loadings are no longer affected by the water band. It can be observed that the points at 1753 and 1558-1543  $\text{cm}^{-1}$  exert great influence on the samples, as well as the band at 1040  $\text{cm}^{-1}$ , attributed to C-N stretching, and the bands at 915-847  $\text{cm}^{-1}$ , related to C-C stretch, which are important molecules in seeds and contribute to the discrimination of PCs [4]. Furthermore, it is noted that the scores are now more defined, making it possible to draw a line of separation between the two classes. The *decumbens*

group presents a more compact grouping, while the *ruziziensis* group is more dispersed.

It is observed that PC1 captured only 53.4% of the total variation, and PC2 captured 25.9%, totaling 79.3%. The remaining 20.7% represents residual variability that can be exploited to further improve seed separation and achieve greater classification efficiency.

#### IV. CONCLUSION

The application of Fourier Transform Infrared Spectroscopy (FTIR), after normalization by Modified Standard Normal Variate (MSNV), followed by Principal Component Analysis (PCA), proved to be highly effective in discriminating methodology for *Brachiaria (decumbens* and *ruziziensis*) seed samples. The analysis, focused on the spectral region of 1800-650 cm<sup>-1</sup>, known as the “fingerprint region,” revealed that even though the seeds are from the same genus, they exhibit distinct vibrational differences that allow a clear separation between the cultivars.

Based on these findings, implementing supervised machine learning methods, such as K-Nearest Neighbors (KNN) or Support Vector Machines (SVM), on data obtained from PCA could potentially further improve seed classification accuracy.

#### ACKNOWLEDGMENT

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), code 001. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), code 403651/2020-5; 302525/2022-0; 440214/2021-1. Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT), code 91/2023; 360/2022.

#### REFERENCES

- [1] Lutts, S., Ndikumana, J., & Louant, B. P. (1991). “Fertility of *Brachiaria ruziziensis* in interspecific crosses with *Brachiaria decumbens* and *Brachiaria brizantha*: Meiotic behaviour, pollen viability and seed” set. *Euphytica*, 57(3), 197-203. <https://doi.org/10.1007/BF00039673>
- [2] L. P. Gantt; C. R. McCurdy; J. T. Bernhard; C.; Pratt, S. Wang; A. C. Brand; P. Broun “Expanding the Scope of High-Throughput Screening for Photochemical Therapy: A Case Study on Near-Infrared Dyes.” *ACS Appl. Sci. Technol.* 2021, 1, 123–130. doi: 10.1021/acsagscitech.1c00067.
- [3] Oliveira, I. C., Franca, T., Nicolodelli, G., Moraes, C. P., Marangoni, B., Bacchetta, G., Milori, D. M. B. P., & Cena, C. (2021). “Fast and accurate discrimination of *Brachiaria brizantha* (A. Rich.) Stapf seeds by molecular spectroscopy and machine learning.” *ACS Agricultural Science & Technology*, 1(2), 67-73. <https://doi.org/10.1021/acsagscitech.1c00067>
- [4] Ribeiro, M. C. S., Senesi, G. S., Cabral, J. S., Cena, C., Marangoni, B. S., Kiefer, C., & Nicolodelli, G. (2021). “Evaluation of rice varieties using LIBS and FTIR techniques associated with PCA and machine learning algorithms.” *Applied Optics*, 60(5), 902-908. <https://doi.org/10.1364/AO.409029>
- [5] Larios, G., Nicolodelli, G., Ribeiro, M., Canassa, T., Reis, A. R., Oliveira, S. L., Alves, C. Z., Marangoni, B. S., & Cena, C. (2020). *Infrared spectroscopy and machine learning algorithms. Analytical Methods*, 12(35), 4303-4309. <https://doi.org/10.1039/d0ay01238f>
- [6] I.T. Jolliffe, J. Cadima, *Principal Component Analysis: a review and recent developments*, *Phil. Trans. R. Soc. A* 374 (2016) 20150202 doi:<https://doi.org/10.1098/rsta.2015.0202>
- [7] Casaril, A. E., Santos, C. G., Marangoni, B. S., Lima, S. M., Andrade, L. H. C., Fernandes, W. S., Infrac, J. O. M., Alves, N. O., Borges, M. D. G. L., & Cena, C. (2021). “Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate analysis”. *Journal of Biophotonics*, 14(4), Article e202000412. <https://doi.org/10.1002/jbio.202000412>
- [8] Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417-441 (1933)
- [9] G. M. C. M. Zosim, V. P. Koska, S. A. Khusainov, and S. R. Tyurin, “Application of near infrared spectroscopy for classification of essential oils,” *Chemom. Intell. Lab. Syst.*, vol. 96, no. 2, pp. 194–199, Feb. 2009, doi: 10.1016/j.chemolab.2008.11.006. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism, vol. III*, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [10] M. F. M. Z. Silva, A. S. Figueira, J. A. Santos, and L. S. Costa, “Application of chemometric methods for the determination of essential oil composition in plant extracts,” *Trends Anal. Chem.*, vol. 46, pp. 109–119, Aug. 2013, doi: 10.1016/j.trac.2013.04.015.
- [11] S. L. R. Moreira, D. D. Silva, R. G. T. Santos, and V. P. Carvalho, “Recent Advances in Functionalized Polymers for Drug Delivery Applications,” *Molecules*, vol. 26, no. 15, pp. 4588, Jul. 2021, doi: 10.3390/molecules26154588.
- [12] S. L. R. Moreira, D. D. Silva, R. G. T. Santos, and V. P. Carvalho, “Recent Advances in Functionalized Polymers for Drug Delivery Applications,” *Chemom. Intell. Lab. Syst.*, vol. 96, no. 2, pp. 194–199, Feb. 2009, doi: 10.1016/j.chemolab.2008.11.006.

A Figura 4 apresenta a matriz de confusão obtida para o modelo de classificação. Utilizando as quatro primeiras Componentes Principais (PCs), alcançou-se uma acurácia global de 94%. A análise por classe demonstrou uma taxa de acerto de 96% para a *B. decumbens* e de 92% para a *B. ruziziensis*.

**Figura 4 – Validação Cruzada (LOOCV) para o Comprimento de Onda 4000 à 650 cm<sup>-1</sup>**

		Validação (94.0%) SVM Linear (C=0.11)	
		DECUMBENS	RUZZIENSIS
CLASSE REAL	DECUMBENS	96.0% 24	4.0% 1
	RUZZIENSIS	8.0% 2	92.0% 23
		DECUMBENS	RUZZIENSIS
		CLASSE PREDITA	

Fonte: Autoria Própria

A Figura 5 apresenta os resultados obtidos mantendo-se o mesmo número de Componentes Principais (4 PCs), porém restringindo a análise ao intervalo de 1800 a 650 cm<sup>-1</sup>. Esta região espectral foi selecionada para eliminar a influência das bandas de estiramento da ligação O-H (associadas à umidade), focando exclusivamente na região da impressão digital.

**Figura 5 – Validação Cruzada (LOOCV) para o Comprimento de Onda 1800 à 650 cm<sup>-1</sup>**

		Validação (94.0%) SVM Linear (C=0.11)	
		DECUMBENS	RUZZIENSIS
CLASSE REAL RUZZIENSIS/DECUMBENS	DECUMBENS	96.0% 24	4.0% 1
	RUZZIENSIS	8.0% 2	92.0% 23
CLASSE PREDITA			

Fonte: Autoria Própria

A comparação entre as validações cruzadas (Figuras 4 e 5) revela que a restrição da análise espectral à região da impressão digital otimizou o desempenho do modelo, atingindo uma acurácia global de 96%. Estes resultados atestam a eficácia da metodologia proposta, confirmando que o acoplamento da técnica FTIR com o algoritmo SVM é viável e robusto para a classificação precisa das sementes analisadas.