

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CÂMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

JULIANO LUCAS CARDOSO JESUS

**APLICAÇÃO DA ESPECTROSCOPIA DO INFRAVERMELHO
PRÓXIMO E SENSOR HIPERESPECTRAL COMBINADOS COM
APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DA
QUALIDADE DO ARROZ BRANCO, PARBOILIZADO, PRETO E
VERMELHO**

CHAPADÃO DO SUL – MS

2025

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CÂMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

JULIANO LUCAS CARDOSO JESUS

**APLICAÇÃO DA ESPECTROSCOPIA DO INFRAVERMELHO
PRÓXIMO E SENSOR HIPERESPECTRAL COMBINADOS COM
APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DA
QUALIDADE DO ARROZ BRANCO, PARBOILIZADO, PRETO E
VERMELHO**

Orientador: Prof. Dr. Paulo Carteri Coradi

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul, como parte dos requisitos para obtenção do título de Mestre em Agronomia, área de concentração: Produção Vegetal.

CHAPADÃO DO SUL – MS

2025



Serviço Público Federal
Ministério da Educação

Fundação Universidade Federal de Mato Grosso do Sul
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA



CERTIFICADO DE APROVAÇÃO

DISCENTE: Juliano Lucas Cardoso Jesus

ORIENTADOR: Dr. Paulo Carteri Coradi

TÍTULO: Aplicação da espectroscopia do infravermelho próximo e sensor hiperespectral combinados com aprendizado de máquina para classificação da qualidade do arroz branco, parboilizado, preto e vermelho.

AVALIADORES:

Prof. Dr. Paulo Carteri Coradi

Prof. Dra. Dthenifer Cordeiro Santana

Prof. Dr. Jorge Gonzalez Aguilera

Chapadão do Sul, 19 de fevereiro de 2025

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Jorge González Aguilera, Usuário Externo**, em 19/02/2025, às 06:51, 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 **Paulo Carteri Coradi, Usuário Externo**, em 19/02/2025, às 07:06, 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 **Dthenifer Cordeiro Santana, Usuário Externo**, em 19/02/2025, às 10:04, 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 5420621 e o código CRC C5102EA5.

COLEGIADO DE PÓS-GRADUAÇÃO EM AGRONOMIA

Rod MS 306, Km 105, Caixa Postal 112

Fone: (67)3562-6351

CEP 79560-000 - Chapadão do Sul - MS

Referência: Processo nº 23455.000685/2024-27 SEI nº 5420621

AGRADECIMENTOS

Primeiramente agradeço a Deus por me guiar e fortalecer em cada etapa desta jornada. Aos meus amigos e familiares que estiveram ao meu lado durante todo o mestrado, oferecendo apoio e incentivo incondicionais. Em especial a minha mãe Lellis Cardoso Movais de Jesus e ao meu pai Nelson Almeida de Jesus por todo amor, dedicação e ensinamentos que me trouxeram até aqui.

Sou imensamente grato aos órgãos públicos FUNDECT, CAPES, CNPq e FAPEC pelo apoio financeiro - de bolsas de estudo e auxílios na pesquisa - que tornaram este trabalho possível.

Agradeço, também, à Universidade Federal de Mato Grosso do Sul (UFMS) por proporcionar um ambiente acadêmico de excelência, permitindo-me crescer no ensino, na pesquisa e na formação profissional.

Expresso minha gratidão aos grupos de pesquisa em *Postharvest Innovation: Technology, Quality & Sustainability* (POSITECHS-UFMS), em Melhoramento de Plantas (GEMP-UFMS) e ao Laboratório de Pós-Colheita (LAPOS) da UFMS pelo suporte e pela troca de conhecimento ao longo deste processo.

Um agradecimento especial ao meu orientador Prof. Dr. Paulo Carteri Coradi por sua orientação, paciência e incentivo, fundamentais para o desenvolvimento deste trabalho.

Agradeço também a minha amiga Nairiane dos Santos Bilhalva que, mesmo à distância, tornou-se uma grande companheira, sempre presente e disposta a me auxiliar em todos os processos dessa jornada. Sua amizade e apoio foram essenciais e sou imensamente grato por tê-la encontrado neste caminho.

A todos que, direta ou indiretamente, contribuíram para esta conquista, meu sincero muito obrigado!

LISTA DE FIGURAS

Figura 1. Amostras de arroz beneficiadas de arroz polido (A), arroz parboilizado (B), arroz vermelho (C) e arroz preto (D).....	12
Figura 2. Rede de correlação entre as variáveis da composição do arroz, especificamente entre todos os parâmetros de umidade (UM), proteína (PB), lipídios (LP), amido (AM), fibras (FB) e cinzas (CZ).....	18
Figura 3. Análise de componentes principais para o arroz branco, parboilizado, preto e vermelho em relação a composição centesimal.....	19
Figura 4. Assinatura hiperespectral NIR de diferentes tipos de arroz polido branco, parboilizado, vermelho e preto em relação a composição centesimal.....	21
Figura 5. Influência da umidade na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.....	22
Figura 6. Influência dos teores de proteína na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.....	24
Figura 7. Influência dos teores de lipídeos na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.....	26
Figura 8. Influência dos teores de fibra bruta na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.....	27
Figura 9. Influência dos teores de cinzas na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.....	28
Figura 10. Influência dos teores de amido na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.....	29
Figura 11. Boxplot do agrupamento de médias para a porcentagem de classificação correta (CC), coeficiente Kappa (Kappa) e F-score (F-score) para a classificação de arroz usando diferentes algoritmos de aprendizado de máquina (REPTree – DT, árvores de decisão – J48, floresta aleatória - RF, redes neurais artificiais – RNA, regressão logística - SL e máquina de vetor suporte – SVM).....	30

LISTA DE TABELAS

Tabela 1. Limites máximos de tolerância expressos em %/peso.....	12
Tabela 2. Relação dos modelos de aprendizagem de máquinas, utilizados na classificação da qualidade do arroz.....	14
Tabela 3. P-valores e resumo da análise de variância para os caracteres de qualidade do arroz.....	17

RESUMO: O arroz é um dos essenciais cereais mais consumido no mundo, a qualidade do cereal é avaliada no recebimento e na expedição dos lotes nas unidades armazenadoras e beneficiadoras, pela classificação física e tipificação. Entretanto, o método de classificação física tradicional é realizado de forma visual, enquanto a avaliação é subjetiva e imprecisa, sendo este um processo demorado. O objetivo deste estudo foi avaliar a utilização da espectroscopia do infravermelho próximo e sensor hiperespectral combinados com o aprendizado de máquina para determinar as propriedades físico-químicas e a classificação da qualidade do arroz branco, parboilizado, preto e vermelho. No Laboratório de Pós-Colheita (LAPOS) da Universidade Federal de Santa Marias, as amostras de arroz branco, parboilizado, preto e vermelho foram classificadas conforme a Instrução Normativa para determinar os defeitos físicos. Foram produzidas amostras com os percentuais de defeitos de 2kg cada que, posteriormente, foram divididas em 100 subamostras de 20g cada. Foram avaliadas a espectroscopia de infravermelho próximo, variáveis espectrais, análises multivariáveis e aprendizado de máquina. As análises de correlações de Pearson demonstram interações significativas entre as propriedades nutricionais do arroz. A análise de componentes principais discriminou as características físico-químicas do arroz, determinando os padrões nutricionais distintos com destaque ao alto valor nutricional do arroz preto em relação ao perfil amiláceo do arroz branco. As assinaturas hiperespectrais revelaram diferenças para os tipos de arroz e para cada arroz processado em função da composição físico-química. A combinação entre espectroscopia do infravermelho próximo e sensor hiperespectral e os algoritmos de aprendizado de máquina alcançaram maior precisão para todas as métricas de acurácia, sendo os modelos J48, SL, RF e SVM que obtiveram maior eficiência para classificação da qualidade do arroz, podendo constituir-se como um novo método para avaliação de lotes de grãos de arroz em unidades armazenadoras e beneficiadoras.

Palavras-Chave: Análise não destrutiva; Métodos de classificação de arroz; Processamento do arroz; Processamento de dados; Qualidade de grãos de arroz.

ABSTRACT: Rice is one of the most consumed essential cereals in the world. The quality of the cereal is assessed upon receipt and shipment of batches at storage and processing units, through physical classification and typing. However, the traditional physical classification method is performed visually, while the assessment is subjective and imprecise, and this is a time-consuming process. The objective of this study was to evaluate the use of near-infrared spectroscopy and hyperspectral sensor combined with machine learning to determine the physicochemical properties and quality classification of white, parboiled, black and red rice. At the Post-Harvest Laboratory (LAPOS) of the Federal University of Santa Marias, samples of white, parboiled, black and red rice were classified according to the Normative Instruction to determine physical defects. Samples with the percentage of defects of 2 kg each were produced and subsequently divided into 100 subsamples of 20 g each. Near-infrared spectroscopy, spectral variables, multivariate analysis and machine learning were evaluated. Pearson correlation analyses demonstrated significant interactions between the nutritional properties of rice. Principal component analysis discriminated the physicochemical characteristics of rice, determining distinct nutritional patterns, highlighting the high nutritional value of black rice in relation to the starchy profile of white rice. Hyperspectral signatures revealed differences for the types of rice and for each processed rice according to the physicochemical composition. The combination of near-infrared spectroscopy and hyperspectral sensor and machine learning algorithms achieved greater precision for all accuracy metrics, with the J48, SL, RF and SVM models achieving greater efficiency for classifying rice quality, and may constitute a new method for evaluating batches of rice grains in storage and processing units.

Keywords: Non-destructive Analysis; Rice grading methods; Rice processing; Data processing; Rice grain quality.

SUMÁRIO

1. INTRODUÇÃO	10
2. MATERIAL E MÉTODOS	11
2.1. Obtenção e preparação das amostras	11
2.2. Espectroscopia de Infravermelho Próximo (NIR)	12
2.3. Variáveis Espectrais	13
2.4. Análises multivariáveis	14
2.5. Aprendizado de Máquina	14
3. RESULTADOS E DISCUSSÕES	15
3.1. Análise da composição centesimal do arroz branco, parboilizado, preto e vermelho	15
3.2. Análise da correlação de Pearson e Componentes Principais das propriedades físico-químicas do arroz	17
3.3. Assinatura hiperespectral de diferentes tipos de arroz branco, parboilizado, vermelho e preto em relação a composição centesimal	20
3.4. Aplicação do Aprendizado de Máquina na classificação do arroz	30
4. CONCLUSÃO	31
5. REFERÊNCIAS BIBLIOGRÁFICAS	32

1. INTRODUÇÃO

O arroz (*Oryza sativa* L.) é uma cultura alimentar com grande importância mundialmente (Mukherjee *et al.*, 2025). A produção de arroz no Brasil para a safra 2024/25, foi de 11,7 milhões de toneladas com produtividade média de 6.895 kg ha⁻¹ (CONAB, 2025). É um alimento fundamental para a maioria da população, tornando a classificação dos tipos de arroz e a qualidade da semente essenciais para a agricultura e a proteção dos clientes (Islam *et al.*, 2025).

Para se obter sementes de arroz com qualidade é preciso ter elementos essenciais, como umidade, teor de proteína e teor de amido (Díaz *et al.*, 2022), dentre esses constituintes, as sementes de arroz são constituídas de amido, proteína de armazenamento, aminoácido, lipídio (Saleem *et al.*, 2023). Essas características têm o potencial de aumentar o rendimento das culturas em 5–20% (Sarkar *et al.*, 2024).

As sementes são submetidas ao processo de secagem, e posteriormente esses grãos de arroz são processados para se produzir arroz polido, parboilizado ou integrado (Coradi *et al.*, 2021). O arroz polido, consiste da remoção da casca externa e das camadas de aleurona do arroz, é um meio de remover fisicamente as proteínas do arroz (Choi *et al.*, 2025; Adnan *et al.*, 2025). Enquanto que o arroz parboilizado é uma nova variedade de arroz, em que seu processamento envolve: imersão, parboilização e secagem (Zhuang *et al.*, 2025).

Além desses tipos de arroz, tem-se os com colorações, como o arroz vermelho e preto decorrem de seus potenciais benefícios à saúde atribuídos às suas composições únicas de amido e fitoquímicos, notavelmente antocianinas, manifestam um espectro de atividades farmacológicas, como efeitos antidiabéticos (Bhuyan *et al.*, 2022; Feng *et al.*, 2022) e antioxidantes (Thilavech *et al.*, 2025; Santana *et al.*, 2025).

Normalmente para realizar a avaliação de qualidade dessas sementes é através do método de inspeção visual, sendo um método destrutivo que leva muito para processar e classificar (Wang *et al.*, 2024), o que o torna mais complexo para testes com grandes quantidades de amostras. Diante disso, existe uma necessidade de propor um método não destrutivo e rápido para diferenciar sementes de arroz danificadas. A espectroscopia no infravermelho próximo (NIR) conjuntamente com o sensor hiperspectral (HSI) oferecem muitos benefícios para classificação da qualidade de sementes, sendo análises analítica, rápida, não destrutivas, permitindo a avaliação de lotes grandes de produtos alimentícios sem alteração física ou desperdício (Jimoh *et al.*, 2025). Oferece também alta sensibilidade a características químicas e físicas de amostras de alimentos, permitindo análises detalhadas de

minerais essenciais, composição nutricional e até contaminação microbiana (Chen *et al.*, 2021; Song *et al.*, 2023).

Modelos de aprendizado de máquina (ML), em que consiste em algoritmos que tem a capacidade de aprenderem com dados e encontrarem padrões nas amostras de dados (Teodoro *et al.*, 2024), estão sendo utilizados para estabelecer relações quantitativas entre espectroscopia no infravermelho próximo e sensores hiperspectral na classificação da qualidade das sementes de arroz (Sampaio e Brites, 2022; Yang *et al.*, 2025). Os modelos de aprendizados de máquina agilizam e imediatamente avaliam as imagens, aprimorando o tratamento de informações provenientes de sensores (Silva *et al.*, 2025). Isso se deve ao fato de que a ML possibilita a criação de algoritmos para lidar com uma grande diversidade de dados e informações complexas, tais como dados de imagens espectrais, que necessitam de interação entre si (Della-Silva *et al.*, 2025). Diante dos aspectos discutidos o objetivo do estudo foi avaliar a utilização da espectroscopia do infravermelho próximo e sensor hiperspectral combinados com o aprendizado de máquina para determinar as propriedades físico-químicas e a classificação da qualidade do arroz branco, parboilizado, preto e vermelho.

2. MATERIAL E MÉTODOS

2.1. Obtenção e preparação das amostras

As amostras compostas de arroz beneficiadas foram obtidas logo após o beneficiamento dos grãos (sem armazenamento) em uma unidade de beneficiamento localizado no município de Cachoeira do Sul, no centro do Rio Grande do Sul, na região fisiográfica Depressão Central, ao lado do Rio Jacuí, latitude: 30° 0' 45" S, longitude: 52° 55' 11" W e altitude de 73 metros. Na Figura 1 estão representadas as amostras de arroz polido, parboilizado, vermelho e preto.



Figura 1. Amostras de arroz beneficiadas de arroz polido (A), arroz parboilizado (B), arroz vermelho (C) e arroz preto (D).

As amostras foram separadas conforme a Instrução Normativa nº 6 de 16 de fevereiro de 2009 com base no regulamento técnico do arroz (Brasil, 2009). As amostras do arroz foram constituídas pelos limites máximos de defeitos tolerados para os Tipo 1, baseando-se nas Tabelas 1 (Brasil, 2012). Foram constituídas amostras de 2kg cada que, posteriormente, foram divididas em 100 subamostras de 20g.

Tabela 1. Limites máximos de tolerância expressos em %/peso.

Arroz	Tipo	Matérias	Mofados	Picados ou Manchados	Gessados Verdes	Rajados	Amarelos	Total de Quebrados e Quirera
		Estranhas e Impurezas	e Ardidos					
Polido	1	1	0,2	0,5	1,50	1,75	0,3	14
Parboilizado	1	1	0,1	0,15	1,5	1,75	0,3	14
Vermelho	1	1	0,2	0,5	1,5	1,75	0,3	14
Preto	1	1	0,2	0,5	1,5	1,75	0,3	14

O limite máximo de tolerância admitido para grão não parboilizado é de 0,30% (zero vírgula trinta por cento) para todos os tipos. Acima desse limite o produto será considerado como Fora de Tipo.

2.2. Espectroscopia de Infravermelho Próximo (NIR)

A determinação, em percentual, da composição centesimal, constituída proteína bruta (PB), teor de água (UM), extrato etéreo (LP), fibra bruta (FB), cinzas (CZ) e amido (AM) das amostras foi realizada por espectroscopia de infravermelho próximo (Metrohm, espectrômetro DS2500, Herisau, Suíça), denominado NIR com alta precisão óptica, em triplicata. As amostras foram homogeneizadas e colocadas na cápsula de amostragem.

A análise tende a iluminar uma amostra com radiação de um comprimento de onda específico na região do infravermelho próximo e, em seguida, medir a diferença entre as quantidades de energia emitida pelo espectroscópio e refletida pela amostra para o detector. Essa diferença é medida em várias bandas, criando um espectro para cada amostra. O registro dos dados espectrais foi feito no modo de refletância, na faixa espectral de 400 a 2500 nm (Barnes *et al.*, 1989). O resultado obtido foi comparado a um conjunto de calibração (Horwitz *et al.*, 1970).

2.3. Variáveis Espectrais

A análise espectral foi realizada com espectrorradiômetro (FieldSpec 4 Jr, Analytical Spectral Devices, Boulder, EUA), utilizando um muglight. O equipamento tem a capacidade de medir o espectro na faixa de 350 a 2500 nm. O intervalo de leitura é de 1,4 nm na faixa de 50 a 1.050 nm e 2 nm na faixa de 1.000 a 2.500 nm. Nota-se que o equipamento leva uma resolução espectral de 3 nm na faixa de 350 a 700 nm e 30 nm na faixa de 1400 a 2100 nm. Uma das vantagens do referido método é a preservação das amostras, sem a interferência de fatores ambientais como a luminosidade, por exemplo, reduzindo erros derivados da utilização de iluminação difusa.

As leituras espectrais foram repetidas 3 vezes na mesma amostra e os dados utilizados no trabalho basearam-se na média das 3 leituras. Uma placa branca que reflete 100% da luz feita com sulfato de bário foi utilizada como padrão de referência. Os dados espectrais dessa placa foram armazenados pelo sistema para posterior determinação do fator de reflectância das amostras que foi multiplicado pelas leituras obtidas em cada uma delas. Para a análise dos dados o sensor foi conectado a um computador, com a instalação de software RS³, que busca o registro das leituras para que, posteriormente, os arquivos sejam importados pelo ViewSpectroPro visando a extração das informações em arquivo.txt e contribuindo com posteriores análises estatísticas.

2.4. Análises multivariáveis

Os dados foram submetidos à análise de rede de correlação de Pearson, utilizando o software Rbio (Bhering, 2017). Para realizar a separação dos clusters entre os diferentes tipos de arroz com base em suas características como amido, fibras, proteínas, lipídios e cinzas os dados foram submetidos a análise de componentes principais (PCA), contribuindo para a formação de um biplot. Essas análises foram executadas no software R utilizando o pacote “ggfortify” (Team, 2013). Posteriormente, foram construídos boxplots para cada cluster de acordo com os dados de características físico-químicas dos tipos de arroz.

2.5. Aprendizado de Máquina

Os dados foram submetidos a algoritmos de aprendizagem de máquina para a classificação da qualidade do arroz com a utilização de seis modelos distintos (Tabela 2) em que os tipos de arroz branco, parboilizado, vermelho e preto foram utilizados como variável de saída (output) dos modelos enquanto as variáveis espectrais foram utilizadas como entrada (input). A regressão logística (SL) foi utilizada como controle por se tratar de um algoritmo tradicional de regressão.

A classificação dos tipos de arroz foi feita utilizando a validação cruzada e estratificada com k-fold=10 realizada ao longo de 10 vezes, repetidas de forma constante, e foram configuradas a partir de 10 neurônios cada do em 100 execuções por modelo utilizado. Foi utilizado o padrão do software Weka 3.9.6, com exceção das Redes Neurais Artificiais, em que foram configuradas a partir de 10 neurônios cada. Ao final, o desempenho foi medido por algoritmo com a utilização de três métricas-padrões, considerando as porcentagens de classificações corretas (CC), o Kappa e o F-score.

Tabela 2. Relação dos modelos de aprendizagem de máquinas, utilizados na classificação da qualidade do arroz.

Sigla	Modelo de aprendizagem de máquinas	Referência
RNA	Redes neurais artificiais	Egmont-Petersen <i>et al.</i> (2002)
J48	Árvores de decisão J48	Quinlan (1993)
SL	Regressão logística	Štepanovský <i>et al.</i> (2017)
DT	REPTree	Snousy <i>et al.</i> (2011)

RF	Floresta aleatória	Belgiu & Drăguț (2016)
SVM	Máquina de vetor suporte	Nalepa & Kawulok (2019)

Os resultados foram representados por meio de boxplots, exibindo a acurácia dos modelos e das entradas para cada métrica. As médias dos resultados foram comparadas através do teste de Scott-Knott com nível de significância de 5% de probabilidade, no caso de efeitos significativos do teste ($F(p \leq 0,05)$), utilizando o software estatístico Sisvar versão 5.6 (Ferreira *et al.*, 2019). Esses dados são aprofundados no próximo capítulo que apresentam nossos resultados e a discussão deles.

3. RESULTADOS E DISCUSSÕES

3.1. Análise da composição centesimal do arroz branco, parboilizado, preto e vermelho

Na Tabela 3 encontram-se os resultados da análise de variância e do teste de média para as variáveis da composição centesimal em função das amostras de arroz branco, parboilizado, vermelho e preto. A partir dos resultados obtidos verificou-se a diferença estatística ($p \leq 0,05$) entre os tipos de arroz e suas respectivas composições. Os maiores desempenhos dos teores de umidade foram para o arroz vermelho (13,16%) e para o arroz preto (13,35%), enquanto o branco (11,12%) e o parboilizado (11,19%) exibiram valores inferiores. Isso pode ter ocorrido possivelmente, devido à estrutura física do grão, o arroz vermelho e o preto, retêm suas camadas de farelo, o que facilita a retenção de umidade devido à maior proteção natural do grão (Coradi *et al.*, 2016).

Tabela 3. P-valores e resumo da análise de variância para os caracteres de qualidade do arroz.

Processo	Umidade (%)	Amido (%)	Proteína (%)	Lipídios (%)	Fibras (%)	Cinzas (%)
Arroz Branco	11,12b	73,39a	8,04b	1,54c	2,10b	1,12c
Arroz Parboilizado	11,19b	70,11b	4,19d	1,44d	2,80a	1,23b
Arroz Vermelho	13,16a	66,40c	6,88c	1,94b	0,69d	1,02d
Arroz Preto	13,35a	57,06d	9,08a	2,04a	1,18c	1,43a
Pr>Fc	0,0000*	0,0000*	0,0000*	0,0000*	0,0000*	0,0000*
CV (%) =	8,37	7,26	10,38	13,39	10,63	8,67
Média Geral	12,21	66,74	7,05	1,74	1,69	1,2

*Significativo ($P < 0,05$), CV: Coeficiente de Variação; letras iguais na coluna não diferem estatisticamente pelo agrupamento de médias Scott-Knott a 5% de probabilidade.

O arroz branco apresentou maior teor de amido (73,39%) seguido pelo arroz parboilizado (70,11%). Enquanto o arroz vermelho (66,40%) e o preto (57,06%) exibiram menores valores de amido. O arroz branco possui maior concentração de amido devido o processo de refinamento que remove as camadas externas do grão, concentrando então, o amido na porção central (Martens *et al.*, 2023). Enquanto, o arroz parboilizado sofre uma alteração na estrutura do amido no grão, tornando-o mais denso e com menos amido solúvel (Pinta *et al.*, 2024), o que justifica sua menor quantidade em comparação ao arroz branco (Tabela 3).

Para o teor proteico, o arroz preto apresentou o maior teor sendo de 9,08%, seguido pelo arroz branco com teor de 8,04%. E o arroz parboilizado exibiu o menor teor de proteína 4,19%. A maior concentração de proteínas no arroz preto está relacionada a sua composição genética, podendo ter influência também dos fatores ambientais que favorecem o acúmulo de nutrientes, característica comum em grãos menos processados, como o arroz preto (Gogoi *et al.*, 2025). Em relação ao arroz branco, no momento que ocorre o polimento, isso contribui para a redução do conteúdo proteico, pois remove as camadas mais externas que contêm a maioria das proteínas, que geralmente é em torno de 7%, dependendo da variedade do grão, condições climáticas, ambientais, de manejo, processamento e armazenamento (Gürbüz *et al.*, 2025).

O arroz preto também apresentou o melhor desempenho para teor de lipídios (2,04%), enquanto o parboilizado apresentou o menor valor (1,44%). Os teores de lipídios encontram-se concentrados em embriões e farelos de arroz, é relacionado ao triglicerídeos, fosfolipídios, ácidos graxos e outros componentes, com teores de aproximadamente de 2 a 3% (Zhou *et al.*, 2002; Zhou *et al.*, 2024), como apresentado nesse estudo (Tabela 3). Enquanto que para o arroz parboilizado, no processo de parboilização envolve altas temperaturas que contribui para a redução dos lipídios no grão, uma vez que parte dos compostos oleosos é perdida durante o aquecimento (Pinta *et al.*, 2024), sendo um dos motivos pelo baixo teor apresentado no estudo (Tabela 3). No estudo desenvolvido por Becker-Algeri *et al.* (2017) há evidências de que os lipídios são encontrados em sua maior parte (1/3 do total) no gérmen do arroz e na camada de aleurona.

Para o teor de fibras, o arroz parboilizado apresentou o maior teor de 2,80%, enquanto o arroz vermelho obteve o menor teor 0,69%. A parboilização preserva parte da camada fibrosa do grão, resultando em um teor mais elevado desse componente (Sun *et al.*, 2025). Em contrapartida, o arroz vermelho, embora seja um grão integral, apresentou um teor de fibras

mais baixo, possivelmente devido às características específicas do grão e ao nível de processamento aplicado que podem variar dependendo do método de industrialização (Rahman *et al.*, 2025).

Os teores de cinzas, que indicam a quantidade de minerais presentes nos grãos, variaram, significativamente, entre as amostras. O arroz preto apresentou o maior valor de 1,43% enquanto o arroz vermelho exibiu o menor valor de 1,02% (Tabela 3). O arroz preto, conservar suas camadas externas o que faz com que retenha mais minerais, principalmente na casca e no farelo da semente de arroz (Obadi *et al.*, 2023), o que possivelmente ocorreu nesse trabalho. Além de possuir um elevado conteúdo de cinzas, o arroz preto é constituído também em grandes partes por metabólitos secundários, incluindo antocianinas e flavonas. Essas antocianinas possuem muitos benéficos na digestão de amido e proteína, regulação da flora intestinal, usos em alimentos e também impactam a digestão de proteínas induzindo mudanças estruturais nas proteínas, retardando sua degradação (Reddy *et al.*, 2025). Outros trabalhos mostraram que as antocianinas reduziram a digestibilidade do amido formando complexos amido-antocianina (Zhang *et al.*, 2023).

O arroz vermelho, caracterizado pela coloração avermelhada devido ao acúmulo de antocianinas e proantocianidinas no pericarpo (Martins *et al.*, 2021), oferece uma rica composição nutricional e um perfil fenólico distinto, incluindo altos níveis de antocianinas e flavonoides (Ramos *et al.*, 2022), como apresentado no presente estudo (Tabela 3).

3.2. Análise da correlação de Pearson e Componentes Principais das propriedades físico-químicas do arroz

É possível observar a rede de correlação que foi elaborada a partir da matriz de Pearson entre as variáveis de composição centesimal do arroz branco, parboilizado, vermelho e preto (Figura 2), em que as correlações positivas são expressas em linhas verdes e correlações negativas em linhas vermelhas, sendo a magnitude de correlação proporcional à espessura das linhas.

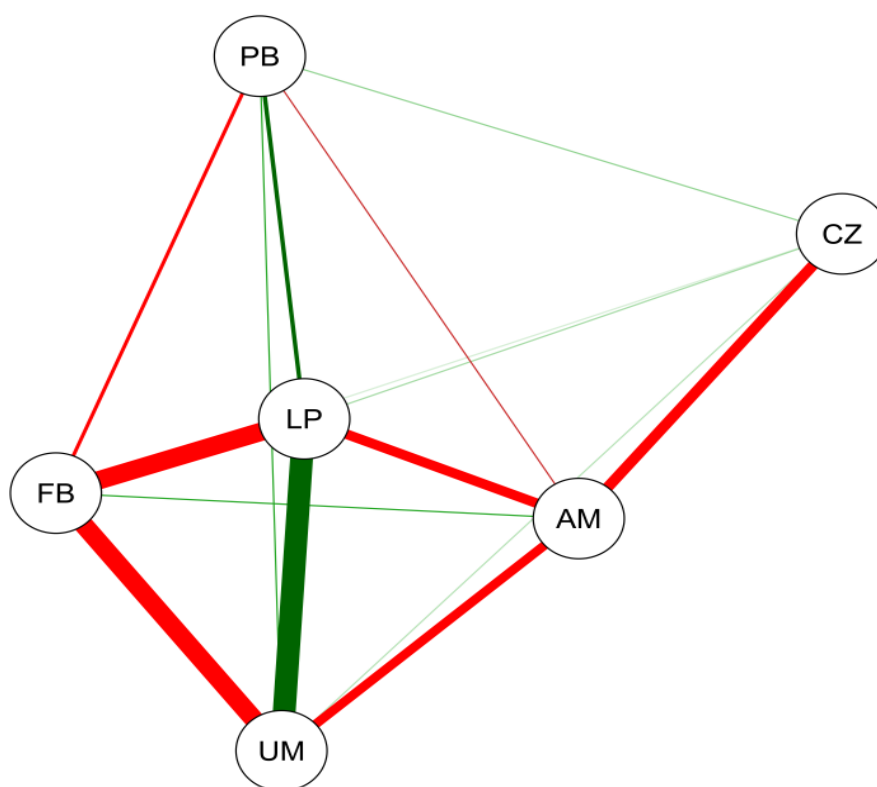


Figura 2. Rede de correlação entre as variáveis da composição do arroz, especificamente entre todos os parâmetros de umidade (UM), proteína (PB), lipídios (LP), amido (AM), fibras (FB) e cinzas (CZ).

Observa-se uma correlação positiva expressa entre lipídios e umidade, indicando que o aumento na umidade tende a estar associado a um aumento na quantidade de lipídios. Entre fibras e amidos pode ser observado uma correlação moderada, sugerindo que o aumento de teor de fibras está associado a redução no teor de amido. Já para a proteína apresentou relações positivas com lipídeos contribuindo com maiores valores para a umidade.

Os lipídios demonstram um papel central na rede de correlação, indicando ser significativo na interação entre os componentes de arroz, possivelmente influenciada por fatores genéticos ou ambientais durante o cultivo do arroz. A correlação fraca ou inexistente entre outros componentes sugere que esses componentes podem não interagir de maneira significativa.

A rede de correlação de Pearson apresenta ligações fortes e negativas entre amidos e fibras, cinzas e lipídeos, ou seja, quanto maior os teores dos compostos menor o teor de amido (Yanjie *et al.*, 2018), o que modifica as qualidades nutricionais do grão, uma vez que o amido

é o principal constituinte e suas propriedades físico-químicas determinam aspectos da qualidade como a capacidade de cozimento e processamento. O que ocorreu nesse trabalho, em que as ligações fortes entre lipídeo e umidade contribuindo negativamente para o amido (Figura 2).

Foi possível observar na análise de componentes principais (PCA) a variabilidade e as correlações entre os diferentes tipos de arroz (branco, parboilizado, preto e vermelho) com base em suas características físico-químicas, incluindo amido, fibras, proteínas, lipídios e cinzas (Figura 3). A variância acumulada foi de 90%, dando credibilidade à representação em um gráfico bidimensional em ambos os experimentos, uma vez que as porcentagens de variância acumulada nas duas primeiras variáveis canônicas devem ser maiores que 80% (Mingoti, 2005).

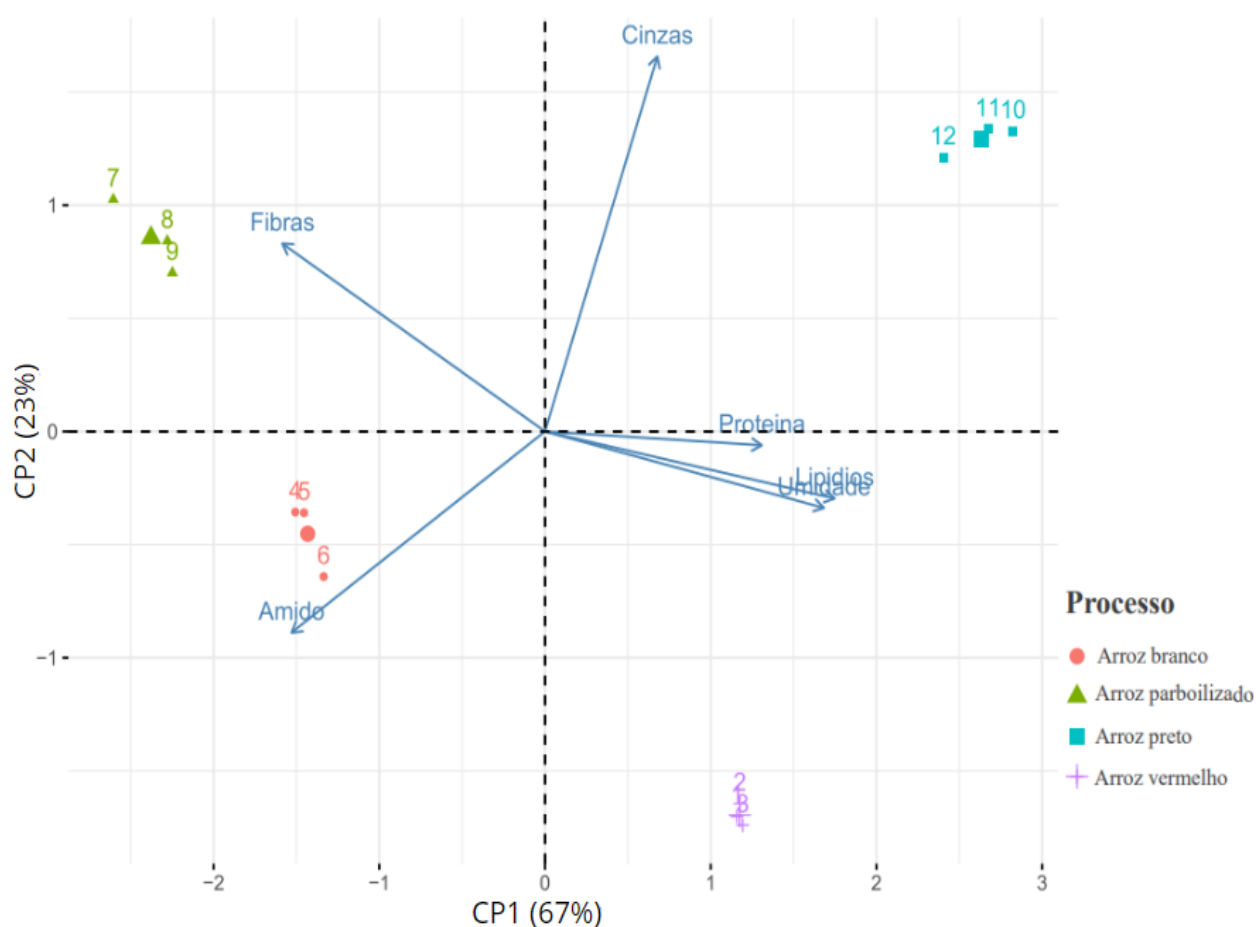


Figura 3. Análise de componentes principais para o arroz branco, parboilizado, preto e vermelho em relação a composição centesimal.

O arroz branco mostrou-se fortemente correlacionado ao teor de amido, evidenciando seu perfil de alto carboidrato e baixo valor nutricional em comparação a outras diferentes variedades de arroz, padrão também observado por SubbuThavamurugan *et al.* (2023) que destacam a predominância de carboidratos de rápida digestão. O arroz parboilizado, por outro lado, associa-se a um elevado teor de fibras, demonstrando que isso ocorre devido ao efeito do processo de parboilização que retém mais nutrientes essenciais na camada externa do grão (Muchilisyiah *et al.*, 2023).

Durante a fase de imersão, certos nutrientes solúveis em água são transferidos da camada de aleurona do arroz integral para o endosperma (Balbinoti *et al.*, 2018). Pesquisas indicaram que o arroz parboilizado é mais nutritivo do que o arroz moído (Wu *et al.*, 2021).

O arroz preto apresentou altos teores de cinzas (minerais), proteínas e lipídios, caracterizando-se como uma variedade de alto valor nutricional, devido possivelmente, possuir em sua composição grandes quantidades de minerais, proteínas e passar por menos processos de beneficiamento em relação aos outros tipos, o que preserva suas características básicas (Panda *et al.*, 2022).

Em relação ao arroz vermelho, apresentou-se incremento para proteínas, lipídios e umidade, o que lhe confere um perfil nutricional equilibrado e relativamente denso, embora menos intenso que o do arroz preto. Em outro estudo, também foi observado que o arroz vermelho possui concentrações consideráveis de proteínas e lipídios, mas em menor proporção, quando comparado ao arroz preto, o que o posiciona como uma alternativa nutricional (Muchilisyiah *et al.*, 2020).

Esses resultados relatam que o arroz preto fornece maiores valores nutricionais, seguido pelo arroz vermelho e parboilizado. O arroz branco, comumente utilizado na culinária brasileira foi o tipo de arroz que apresentou menor densidade de nutrientes, uma vez que a remoção de sua casca e gérmen influencia neste ponto em específico (Olaoni *et al.*, 2025).

3.3. Assinatura hiperespectral de diferentes tipos de arroz branco, parboilizado, vermelho e preto em relação a composição centesimal

A curva espectral de cada tipo de arroz foi construída de acordo com as suas respectivas médias e suas respectivas reflectâncias dos comprimentos de onda de 350-2450

nm (Figura 4). Nos comprimentos de onda entre 350nm e 750 nm é observável uma forte elevação do arroz branco quando comparado ao arroz preto, provavelmente devido ao sobretom das ligações CH e ao sobretom das vibrações de alongamento de OH e NH, que geralmente estão ligados a compostos orgânicos como proteínas e lipídios (Song *et al.*, 2023).

Entre 900 nm e 1170 nm as duas curvas apresentaram comportamento muito semelhante, diferenciando-se novamente entre 1811 nm e 1901 nm em que houve uma elevação da curva do arroz preto. Essa semelhança pode estar atribuída ao sobretom do alongamento de OH, que se refere a uma composição química específica conectada à presença de umidade e macronutrientes (Qiao *et al.*, 2022; Zheng *et al.*, 2024). Assim, as propriedades espectrais realçam a capacidade do HSI de classificar a qualidade do arroz.

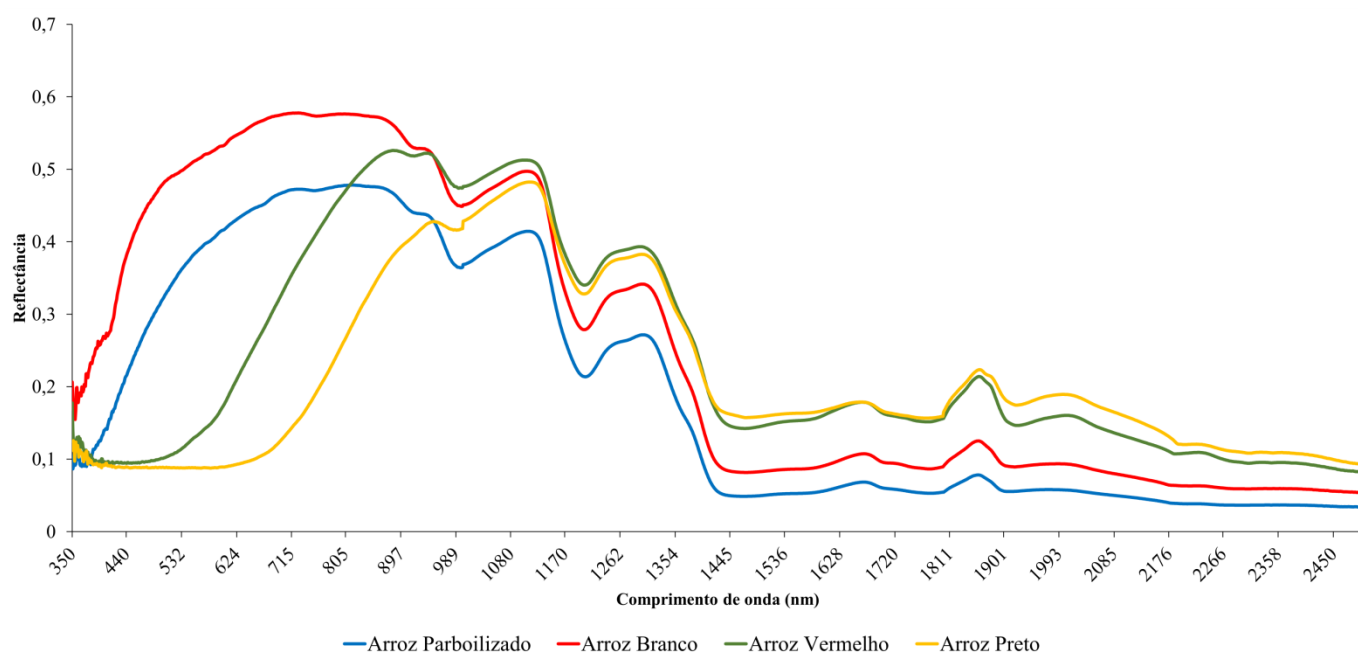


Figura 4. Assinatura hiperespectral NIR de diferentes tipos de arroz polido branco, parboilizado, vermelho e preto em relação a composição centesimal.

A curva espectral formada pelo arroz branco apresentou maior fator de reflectância do que as demais curvas formadas pelos outros tipos de arroz, entre 350 nm e 897 nm (Figura 4), cuja diferença diminuiu a partir de 989 nm. Ambos os grupos funcionais $-CH_2$ e $-OH$, estão relacionados a picos de aproximadamente 915 e 990 nm, respectivamente, indicam constituintes do arroz, como amido, umidade e proteína (Díaz *et al.*, 2022).

Para o arroz branco (Figura 5), os picos mais intensos são seguidos por uma queda abrupta refletindo menor concentração de compostos bioativos após o polimento. A curva do

arroz vermelho revela o pico mais alto com intensidade próxima de 0,6, diminuindo a alta concentração de compostos fenólicos e antocianinas que aumentam para a coloração e o potencial antioxidante como observado em Priyanthi *et al.* (2021).

Essas variações espectrais refletem como os componentes químicos, como amido, lipídios, proteínas e compostos fenólicos influenciam as respostas espectrais e sugerem a relevância de métodos não destrutivos para avaliar a qualidade físico-química do arroz (Shi *et al.*, 2023). Grãos de arroz com altos teores de umidade estão sujeitos ao aumento da atividade biológica e da respiração, e são suscetíveis à deterioração, devido a isso, torna-se necessário secar o arroz antes do processamento e armazenamento (Müller *et al.*, 2022).

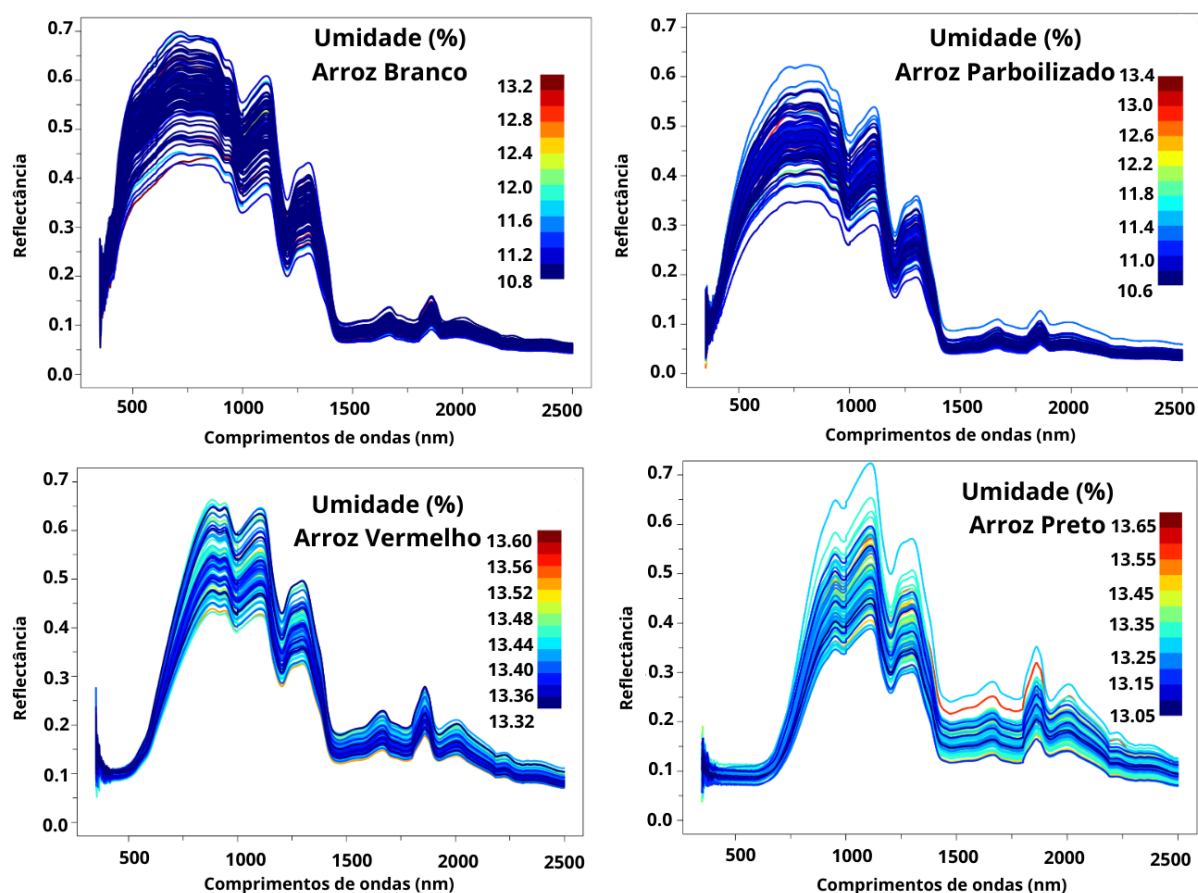


Figura 5. Influência da umidade na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.

A umidade exerce um impacto significativo na reflectância dos grãos de arroz com uma tendência observada de que níveis mais altos de umidade estão associados a menores valores de reflectância, isso ocorre possivelmente, devido a maior absorção da luz pela água presente nos grãos (Uivarasan *et al.*, 2024). Essa relação é especialmente evidente no arroz preto e vermelho onde as variações de reflectância são mais acentuadas, indicando uma interação

mais pronunciada entre a umidade e os componentes estruturais desses tipos de arroz (Chen *et al.*, 2025).

Enquanto que o arroz branco e o parboilizado, as variações de reflectância são menos expressivas com espectros mais homogêneos ao longo do intervalo espectral analisado. Essa diferença de comportamento sugere que a composição química do arroz, incluindo teor de amido, proteínas e pigmentos influencia a resposta espectral de maneira distinta em cada tipo de arroz (Kaur *et al.*, 2018).

Em relação à distribuição espectral, o arroz preto apresenta picos de reflectância mais elevados nas faixas de 1000 a 1500 nm, o que indica uma maior influência da umidade na absorção dessa região do espectro (Zhong *et al.*, 2023). Tanto o arroz branco quanto o arroz parboilizado exibem uma queda mais acentuada de reflectância após os 1000 nm com espectros que permanecem mais homogêneos em comparação ao arroz preto (Zahir *et al.*, 2025). Esse comportamento pode ser explicado pelas diferenças nas composições químicas e pelos diferentes processos de processamento aplicados a cada tipo de arroz que afetam a interação entre os grãos e a luz (Rhowell *et al.*, 2024).

Esses resultados confirmam a influência dos fatores físicos e químicos nos espectros de reflectância com o arroz preto apresentando uma maior variação devido as suas características estruturais e composição, enquanto o arroz branco e parboilizado exibem um padrão mais estável ao longo do espectro (Zhuang *et al.*, 2021). O estudo sugere que tanto o teor de umidade quanto o tipo de processamento desempenham papéis cruciais na determinação da resposta espectral dos grãos de arroz.

Na Figura 6 observa-se a variação da reflectância dos diversos tipos de arroz conforme a influência da proteína. O pico de absorção observado em 985 nm se relaciona com a vibração de alongamento da ligação NH da proteína (Wei *et al.*, 2021). Assim, neste estudo, a análise espectral foi conduzida utilizando amostras no estado de arroz com casca, com subsequente remoção da casca de arroz após a dessecação completa para gerar amostras de arroz integral. Consequentemente, foi observada uma redução no teor de umidade durante esse processo.

O pico de absorção evidente no espectro do arroz em casca supera o do arroz integral, atribuível à presença de celulose, um constituinte da casca do arroz (Weng *et al.*, 2023). Além disso, o pico de absorção na faixa de 2000–2250 nm se correlaciona com a lignina, o principal constituinte do xilema da planta, levando assim a uma maior taxa de absorção no espectro do

arroz em casca com casca em comparação com o arroz integral sem casca (Kästner *et al.*, 2022).

Pesquisas anteriores identificaram uma banda espectral indicativa do conteúdo de proteína na vizinhança de 1570, 985, 1200 e 1925 nm (Weng *et al.*, 2023). O estudo atual revela picos positivos significativos correlacionados com o conteúdo de proteína em torno dos comprimentos de onda de aproximadamente 985 e 1950 nm, enquanto picos negativos foram observados nas faixas de aproximadamente 985–1000 nm e 1150–1200 nm. Essas descobertas estão alinhadas com as de um estudo anterior, que confirmou 985, 1200 e 1925 nm como comprimentos de onda principais para identificação do conteúdo de proteínas (Weng *et al.*, 2023).

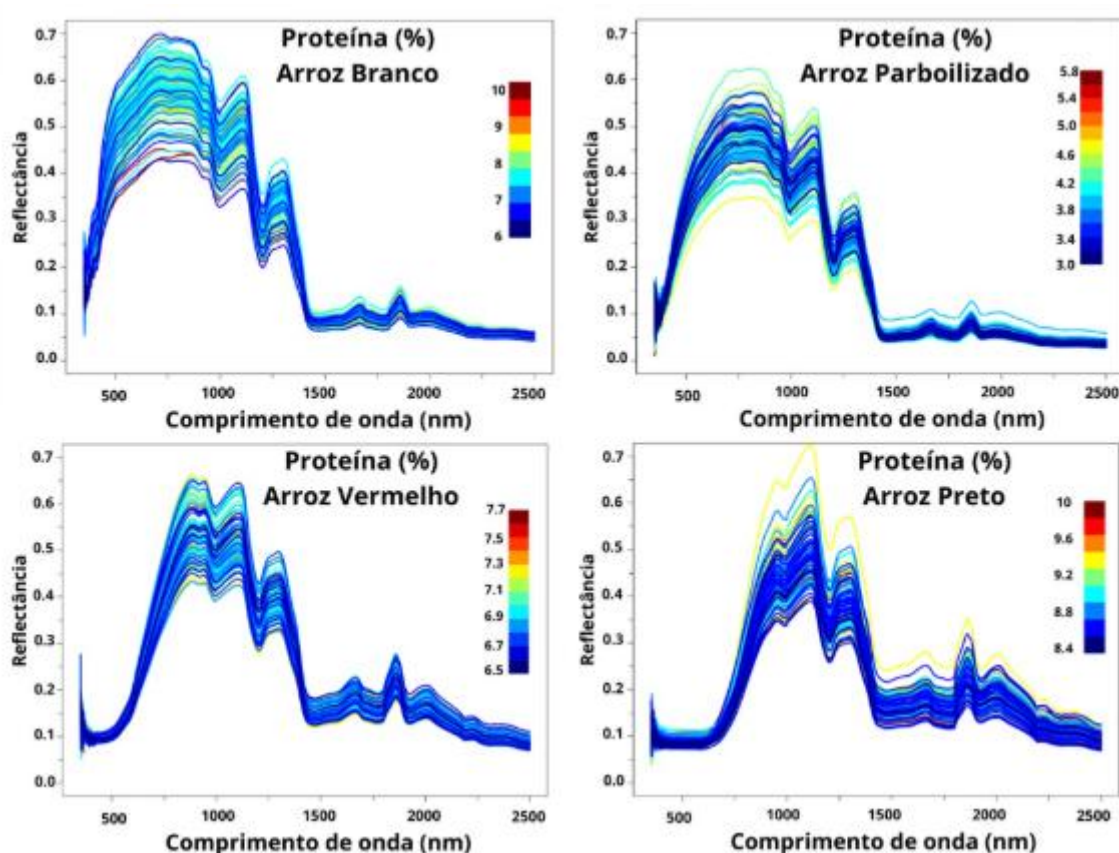


Figura 6. Influência dos teores de proteína na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.

A influência da proteína na reflectância dos grãos de arroz é evidente nas variações observadas nos picos ao longo do espectro. O arroz preto, que apresenta os maiores teores proteicos, demonstra um aumento significativo na reflectância na faixa de 1000 a 1500 nm, essa elevação sugere uma maior absorção de luz devido à presença das proteínas e outros compostos estruturais presentes nos grãos (Zhong *et al.*, 2023). De maneira semelhante, o

arroz vermelho também apresenta uma variação na reflectância, mas de forma menos pronunciada, resultando em espectros mais agrupados em comparação ao arroz preto (Kaur *et al.*, 2018).

Por outro lado, o arroz branco e parboilizado, com teores proteicos menores, mostram valores de reflectância mais baixos, especialmente após os 1000 nm, o que confirma a menor presença de proteínas nesses tipos de arroz (Zahir *et al.*, 2025).

Essas diferenças nos picos de reflectância revelam o impacto direto da composição proteica na resposta espectral dos grãos. O arroz preto e vermelho, com maior teor de proteína, exibe maiores valores de reflectância em faixas específicas enquanto o arroz branco e parboilizado apresentam espectros mais homogêneos e com menor variação, refletindo uma menor quantidade de proteínas (Uivarasan *et al.*, 2024). Esses resultados podem estar associados à estrutura física dos grãos e ao processo de parboilização que altera a composição e as propriedades estruturais, influenciando a forma como a luz interage com os grãos durante a análise espectral (Zhuang *et al.*, 2025).

A figura 6 ilustra como a reflectância dos diferentes tipos de arroz é afetada pelos lipídios. O impacto dos lipídios na reflectância dos grãos de arroz é claramente observado nas variações dos picos espectrais. Da mesma forma, o pico de absorção detectado em 1200 nm é atribuído à vibração de alongamento CH de lipídios (Wei *et al.*, 2021).

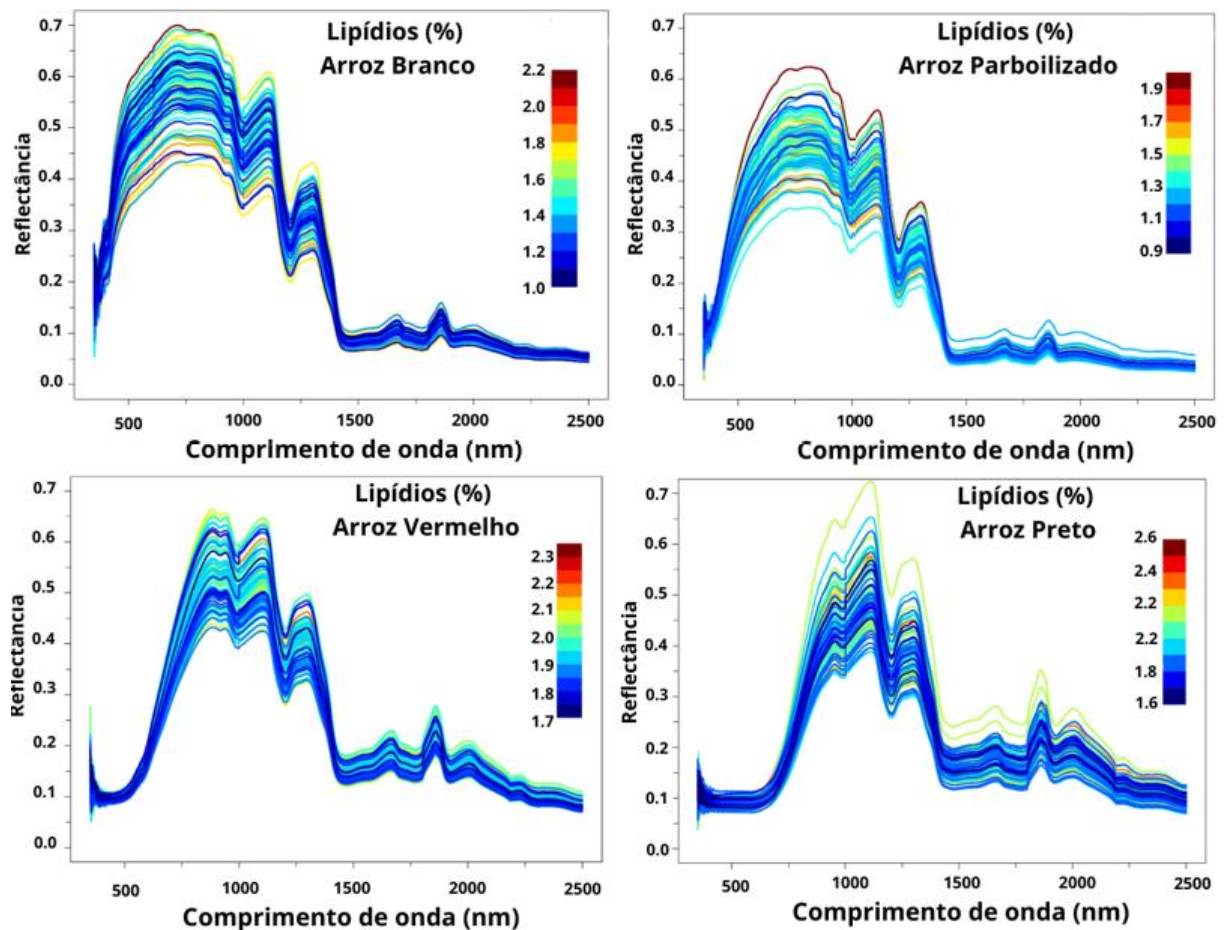


Figura 7. Influência dos teores de lipídeos na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.

Essas diferenças nos picos de reflectância entre os tipos de arroz estão diretamente relacionadas às variações na composição lipídica de cada variedade. O arroz preto apresenta uma maior intensidade de reflectância nas faixas acima de 1000 nm, sugerindo uma interação mais pronunciada dos lipídios com a radiação incidente. O arroz vermelho segue um padrão similar, mas com menor intensidade, enquanto o arroz branco e parboilizado demonstram espectros mais estáveis com menor dispersão de reflectância (Zhong *et al.*, 2023). Esse comportamento pode ser atribuído às diferenças nas quantidades de lipídios e às variações estruturais nos grãos, influenciadas tanto pela composição quanto pelo processamento (Zhu *et al.*, 2024).

A Figura 8 demonstra a resposta espectral dos variados tipos de arroz sob a ação de fibras.

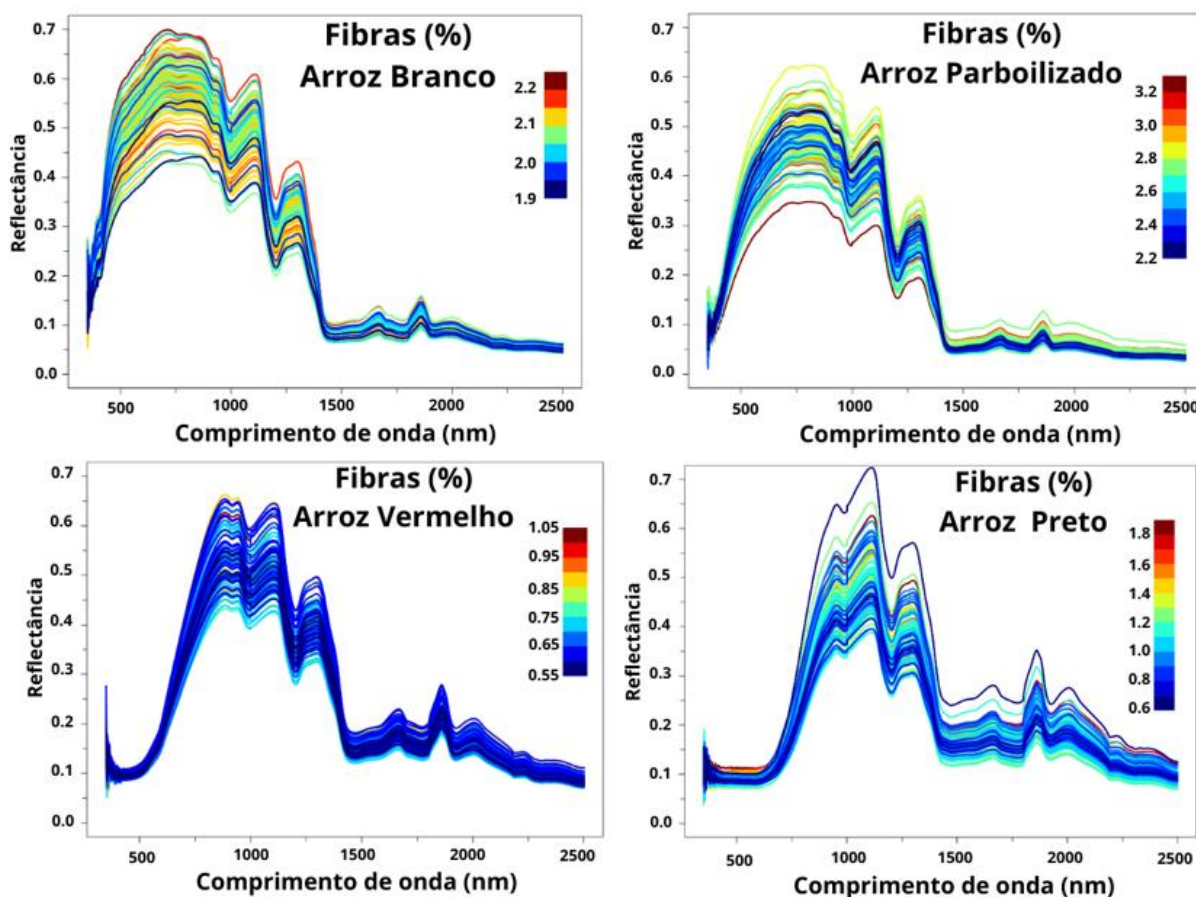


Figura 8. Influência dos teores de fibra bruta na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.

A relação entre fibras e reflectância é evidenciada pela intensidade e pela variação nos picos espectrais dos diferentes tipos de arroz. O arroz parboilizado, que possui um maior teor de fibras, apresenta uma elevação mais expressiva na faixa de 1000 a 1500 nm, indicando que as fibras têm uma maior influência na absorção da radiação nessa região do espectro (Kaur *et al.*, 2018).

As diferenças observadas nos picos de reflectância indicam que a presença de fibras impacta, diretamente, a resposta espectral dos grãos. O arroz parboilizado, devido ao seu maior teor de fibras, exibe uma variação mais ampla na reflectância, especialmente após os 1000 nm, enquanto os outros tipos de arroz apresentam espectros mais estáveis e com menor dispersão (Zahir *et al.*, 2025). Esse comportamento pode ser atribuído à estrutura física do grão e ao processo de parboilização, que altera a composição e a forma como a luz interage com os compostos presentes no arroz (Haruni *et al.*, 2024).

A Figura 9 mostra a refletividade dos diferentes tipos de arroz em função da influência de Cinzas. A relação entre os teores de cinzas e a reflectância dos grãos de arroz é observada

na intensidade e dispersão das curvas espectrais. O arroz preto, que apresenta os maiores teores de cinzas, exibe um espectro mais disperso com uma variação maior na faixa de 1000 a 1500 nm. Essa dispersão pode ser explicada pela maior absorção e dispersão da luz, influenciada pela presença de minerais, como as cinzas que alteram a interação com a radiação (Gogoi *et al.*, 2024). O arroz parboilizado, por sua vez, também apresenta um teor de cinzas superior ao do arroz branco e vermelho, mostrando diferenças notáveis na dispersão das curvas.

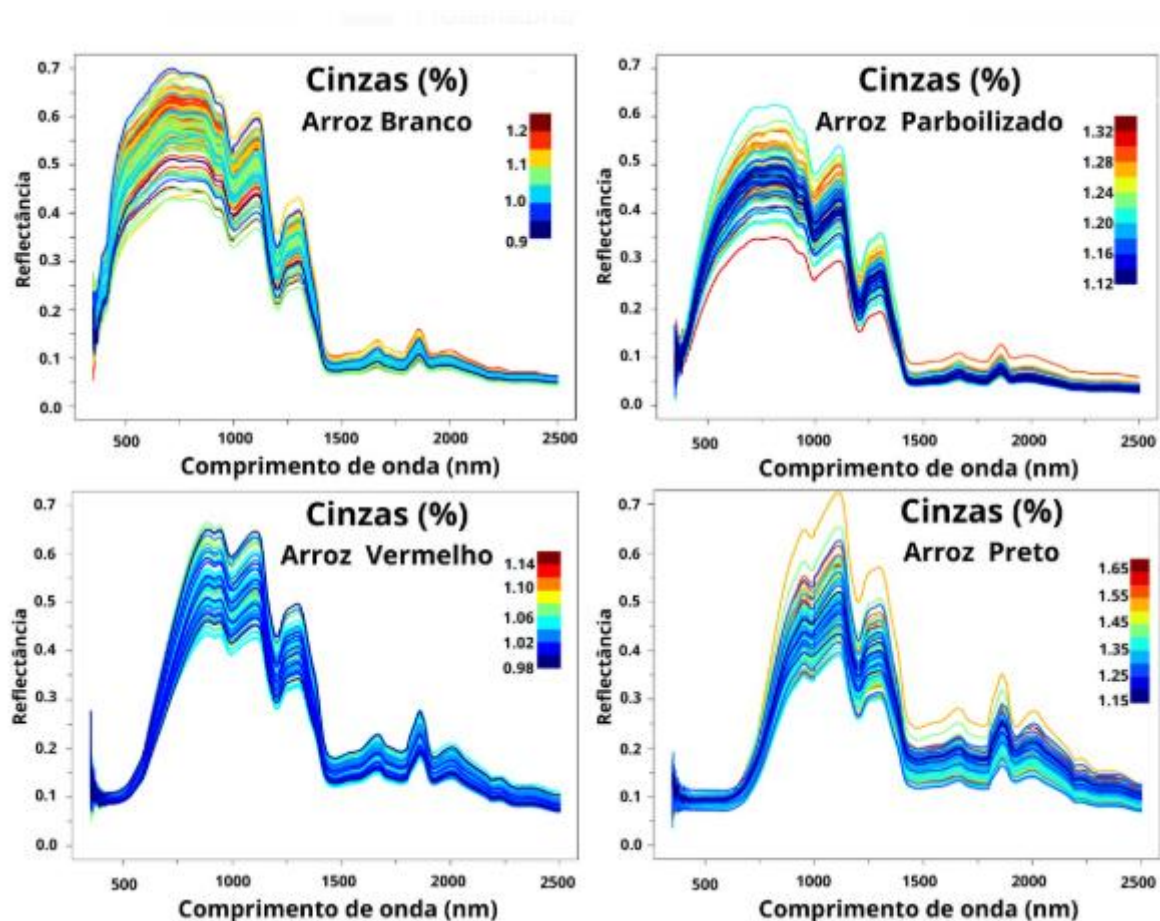


Figura 9. Influência dos teores de cinzas na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.

Essas diferenças nos espectros sugerem que o teor de cinzas impacta, diretamente, a absorção e a dispersão da luz, afetando a resposta espectral dos grãos. O arroz preto, com seu maior teor de cinzas, exibe uma variação espectral mais ampla enquanto os outros tipos de arroz apresentam espectros mais compactos e com menor dispersão (Zhu *et al.*, 2024). Esse comportamento pode ser relacionado à composição mineral de cada tipo de arroz, bem como

ao processo de parboilização que pode modificar a estrutura e a quantidade de minerais presentes nos grãos (Kumari *et al.*, 2025).

Na Figura 10 apresenta-se a resposta óptica dos distintos tipos de arroz sob a interferência de amido. A composição química do arroz depende muito da quantidade de amido da variedade, que varia de 60% a 80% de amido (Idrishi *et al.*, 2024).

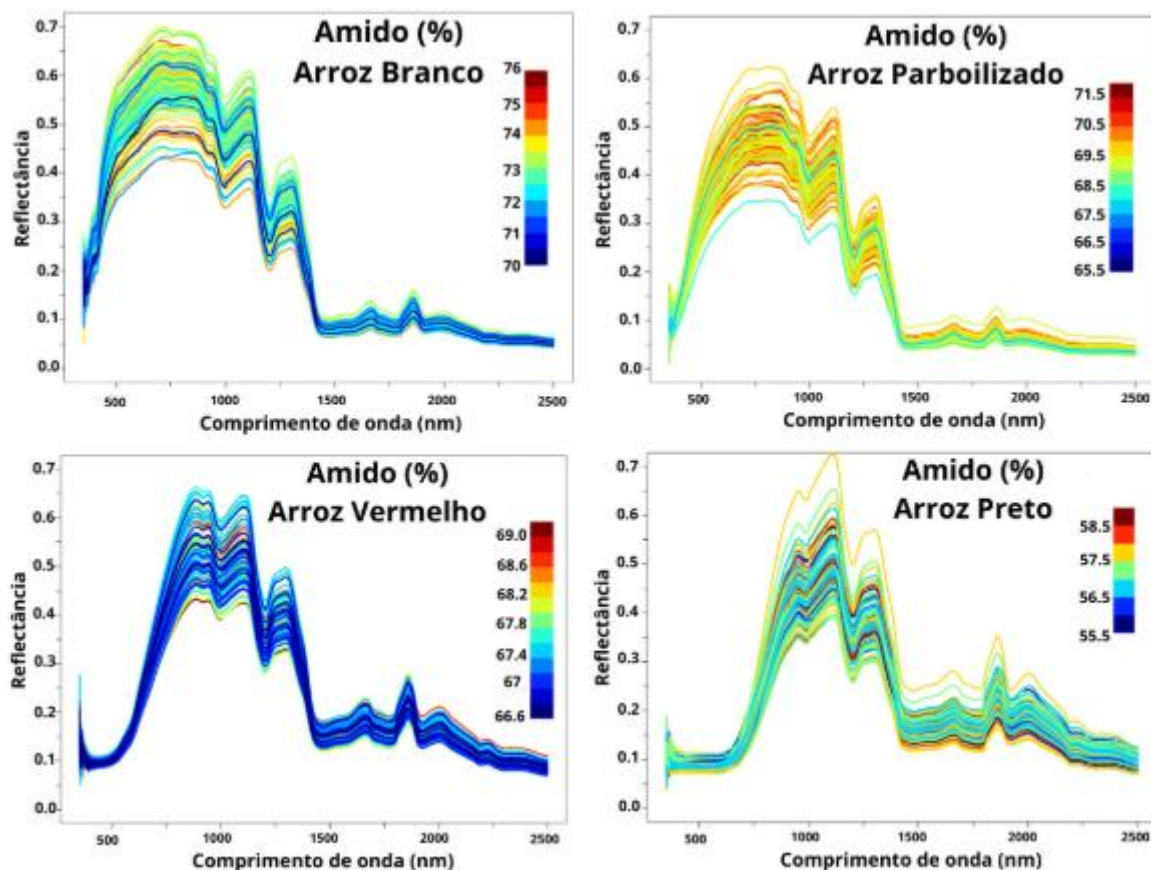


Figura 10. Influência dos teores de amido na reflectância dos grãos de arroz branco, parboilizado, vermelho e preto.

O amido é o principal componente do endosperma da semente do arroz, acumula-se durante o desenvolvimento do grão, é responsável por cerca de 80% do peso seco do arroz, o que está correlacionado com o envelhecimento do arroz e sua variação de qualidade (Shoukat *et al.*, 2025). Quanto maior o teor de amido de cadeia linear e maior teor de amido ramificado de cadeia longa geralmente apresenta maior quantidade de amido de digestão lenta e amido resistente, além de possuir mais propriedades funcionais (Gani *et al.*, 2017; Shen *et al.*, 2021). Dessa forma, as diferenças observadas nas curvas espectrais podem ter ocorrido, possivelmente devido a composição química do arroz, sendo o amido um dos fatores predominantes na variação da reflectância, especialmente nas faixas espectrais mais próximas do infravermelho (Uivarasan *et al.*, 2024).

3.4. Aplicação do Aprendizado de Máquina na classificação do arroz

Os gráficos contêm as três métricas de acurácia dos algoritmos para a classificação da qualidade das sementes dos diferentes tipos de arroz. Para a classificação correta, os modelos J48, SL, RF e SVM promoveram maior acurácia para classificar os tipos de arroz, obtendo resultados com 100% de porcentagem de classificação correta.

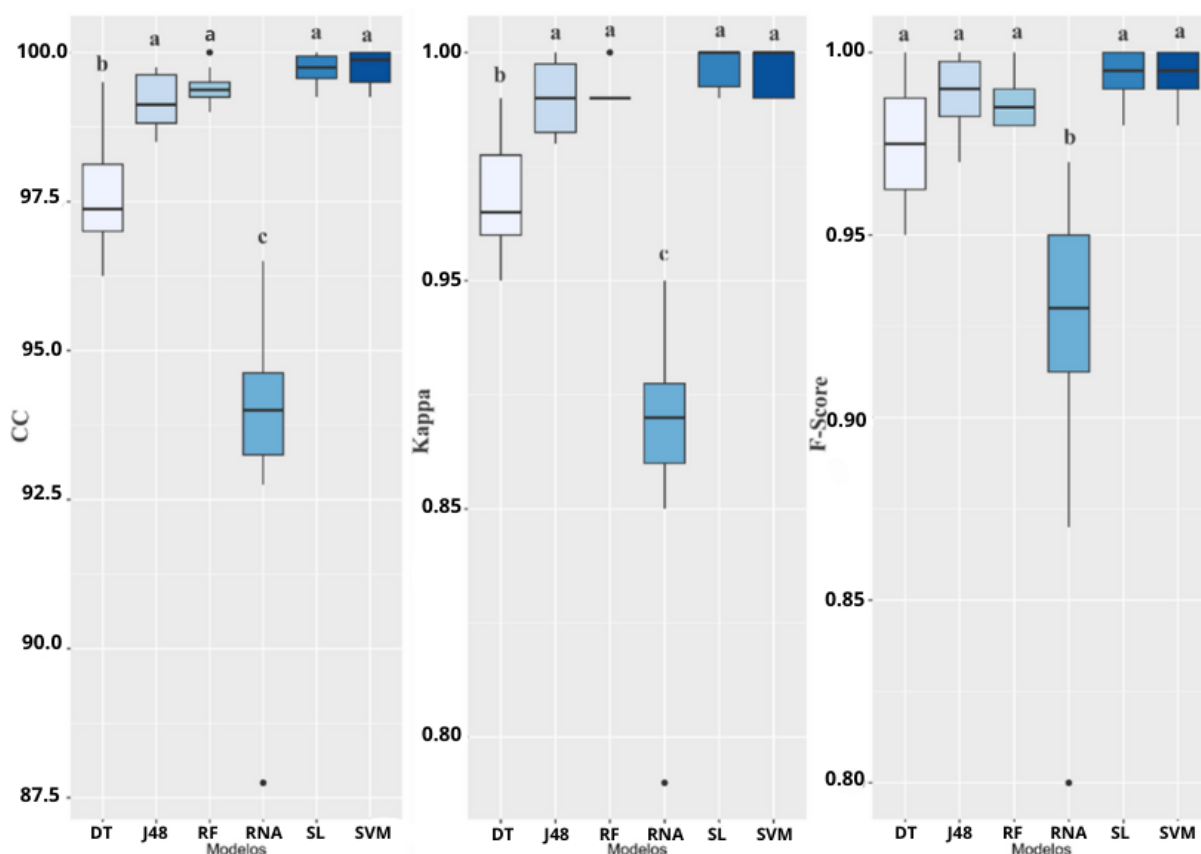


Figura 11. Boxplot do agrupamento de médias para a porcentagem de classificação correta (CC), coeficiente Kappa (Kappa) e F-score (F-score) para a classificação de arroz usando diferentes algoritmos de aprendizado de máquina (REPTree – DT, árvores de decisão – J48, floresta aleatória - RF, redes neurais artificiais – RNA, regressão logística - SL e máquina de vetor suporte – SVM). Médias seguidas pelas mesmas letras maiúsculas para as diferentes entradas e pelas mesmas letras minúsculas para os diferentes algoritmos de aprendizado de máquina não diferem pelo teste de Scott–Knott a 5% de probabilidade.

O desempenho do SVM reforça o que foi observado por Mishra *et al.* (2021) que destacam a capacidade deste modelo para lidar com dados agrícolas, oferecendo classificações precisas para diferentes variáveis agrônomicas. Os algoritmos SL e SVM também obtiveram bom desempenho refletido nos altos valores de índice Kappa e F-Score. Esses modelos mantiveram uma boa consistência na previsão, mesmo com as variações naturais dos dados. O SVM pode ser utilizado em aplicações agrícolas, especialmente por sua capacidade de generalizar situações com alta variabilidade nos dados como ocorre com imagens hiperespectrais (Morellos *et al.*, 2025). O objetivo é prever valores contínuos através da identificação de uma função que se afaste o mínimo possível dos valores-alvo reais para todos os dados de treino, mantendo um equilíbrio entre a complexidade do modelo e a acurácia da previsão (Vahidi *et al.*, 2025).

O modelo florestal aleatória (RF) é um algoritmo que difere do tradicional algoritmo que é a árvore de decisão, que possui a capacidade de prevenir o overfitting de dados sem poda, além de uma velocidade de treinamento mais rápida e ajustes de parâmetros simples, e apresenta um desempenho superior na modelagem de classificação com os parâmetros padrão (Hu *et al.*, 2023).

Utilizar dados da espectroscopia do infravermelho próximo e sensor hiperspectral para classificar a qualidade do arroz com base na composição centesimal e dados espectrais das sementes de arroz pode otimizar as avaliações da qualidade de diversas sementes, como a da cultura do arroz em trabalhos de pós-colheita. Agrupar os algoritmos de aprendizagem de máquinas a utilização de avaliações espectrais torna todo o processo mais ágil e acurado. Para estudos futuros, outras sementes de demais culturas e algoritmos de aprendizagem de máquina devem ser avaliadas visando maior precisão de classificação. É um método muito preciso, eficaz e fácil p manuseio. Em breve, essa tecnologia poderá ser disponibilizada a agricultores, universidades e empresas para fornecer informações para classificar a qualidade da semente de arroz. Dessa forma, perdas e desperdícios de sementes podem ser minimizadas.

4. CONCLUSÃO

A combinação entre espectroscopia do infravermelho próximo e sensor hiperspectral e os algoritmos de aprendizado de máquina alcançaram maior precisão para todas as métricas de acurácia, sendo os modelos J48, SL, RF e SVM que obtiveram maiores eficiência para classificação da qualidade do arroz, podendo constituir-se como um novo método para avaliação de lotes de grãos de arroz em unidades armazenadoras e beneficiadoras.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Adnan, A. S. M., Karruppan, J., Sulaiman, S. A., Sanusi, N., Sanusi, S. A., Razali, N., Abdullah, M. I. S. (2025). Study on the Potential of MARDI Specialty Rice for the Production of Local Parboiled Rice. **Advances in Agricultural and Food Research Journal**, v. 6, n. 1.

Balbinoti, T. C. V., de Matos Jorge, L. M., Jorge, R. M. M. (2018). Modeling the hydration step of the rice (*Oryza sativa*) parboiling process. **Journal of Food Engineering**, v. 216, p. 81-89.

Barnes, R. J., Dhanoa, M. S., Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied spectroscopy**, v. 43, n. 5, p. 772-777.

Becker-Algeri, T. A., Mendes, G. L., Colli, R., Grilo, B. C., Furlong, E. B. (2017). Efeito do beneficiamento e aplicação de tratamento térmico na composição química e digestibilidade proteica do arroz. **Revista CSBEA-v**, v. 3, n. 1, p. 1.

Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, [s. l.], v. 114, p. 24–31, 2016.

Bhering, LL Rbio: Uma ferramenta para análise biométrica e estatística usando a plataforma R. **Crop Breed. Appl. Biotechnol.** 2017, v. 17, p. 187–190.

Bhuyan, P., Ganguly, M., Baruah, I., Borgohain, G., Hazarika, J., Sarma, S. (2022). Propriedades inibitórias da alfa glucosidase de alguns compostos bioativos isolados do farelo de arroz preto: evidências combinadas in vitro e in silico que apoiam o efeito antidiabético do arroz preto. **RSC Advances**, v. 12, n. 35, p. 22650–22661.

Brasil. **Instrução Normativa 2/2012**. Ministério da Agricultura, Pecuária e Abastecimento. 2012.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 6, de 16 de fevereiro de 2009**. Aprova o regulamento técnico do arroz. Diário Oficial da União, Brasília, 2009.

Chen, X., Siesler, H. W., Yan, H. (2021). Rapid analysis of wheat flour by different handheld near-infrared spectrometers: A discussion of calibration model maintenance and performance

comparison. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 252, p. 119504.

Choi, Y. C., Ye, S. J., Shin, J. S., Kim, H. Y., Bae, J. E., Oh, S. M., Baik, M. Y. (2025). Synergistic effect between polishing and high-pressure enzyme treatment for low protein rice production. **Food Science and Biotechnology**, v. 34, n. 2, p. 383-390.

CONAB - Companhia Nacional De Abastecimento (2025) Acompanhamento da safra brasileira de grãos, v. 12 - Safra 2024/25, n. 5: 5º levantamento. Brasília, CONAB. 108p.

Coradi, P. C., Fernandes, C. H., Helmich, J. C., Goneli, A. L. (2016). Effects of drying air temperature and grain initial moisture content on soybean quality (*Glycine max* (L.) Merrill). **Engenharia Agrícola**, v. 36, p. 866-876.

Coradi, P. C., Müller, A., André, G. D. S., Teodoro, P. E., Campos, C. N. S., Anschau, K. F., Flores, E. M. M. (2021). Effects of cultivars and fertilization levels on the quality of brown and polished rice. **Cereal Chemistry**, v. 98, n. 6, p. 1238-1249.

Della-Silva, J. L., de Oliveira Faleiro, V., Pelissari, T. D., Ferreira, A., Dias, N. F., dos Santos, D. H., da Silva Junior, C. A. (2025). Evaluation of soybean plants affected by *Aphelenchoides besseyi* using remote sensing and machine learning techniques. **Remote Sensing Applications: Society and Environment**, p. 101461.

Díaz, E. O., Kawamura, S., Ishizu, H., Nagata, T., Koseki, S. (2022). Non-destructive assessment of amylose content in rice using a quality inspection system at grain elevators. **Food Chemistry**, v. 379, p. 132144.

Egmont-Petersen, M., De Ridder, D., Handels, H. Image processing with neural networks—a review. **Pattern Recognition**, [s. l.], v. 35, n. 10, p. 2279–2301, 2002.

Feng, SY, Wu, SJ, Chang, YC, Ng, LT, Chang, SJ (2022). Estimulação da captação de glicose GLUT4 por extrato rico em antocianina de arroz preto (*Oryza sativa* L.) via sinalização PI3K/Akt e AMPK/p38 MAPK em células C2C12. **Metabolites**, v. 12, n. 9, p. 856.

Ferreira, D. F. Sisvar: A computer analysis system to fixed effects split plot type designs. **Brazilian Journal of Biometrics**, v. 37, n. 4, p. 529-535, 2019.

Gani, A.; Ashwar, B.A.; Akhter, G.; Shah, A.; Wani, I.A.; Masoodi, F.A. Physico-Chemical, Structural, Pasting and Thermal Properties of Starches of Fourteen Himalayan Rice Cultivars. **Int. J. Biol. Macromol.** 2017, 95, 1101–1107.

Gogoi, P., Sharma, P., Goudar, G., Mahajan, A., Dutta, H., Sasikumar, R., Longvah, T. (2025). Nutritional profile and mineral bioaccessibility of pigmented rice landraces. **Journal of Food Measurement and Characterization**, v. 19, n. 2, p. 1513-1530.

Gürbüz, B. N., Pastrana, L. M., Pereira, R. N., & Cerqueira, M. A. (2025). Alternative protein-based meat and fish analogs by conventional and novel processing technologies: a systematic review and bibliometric analysis. **Foods**, v. 14, n. 3, p. 498.

Haruni, S. A., Padjung, R., Musa, Y., Farid, M., Anshori, M. F., Fadhilah, A. N. (2024). Functional food biofortification in increasing red and black rice production through the use of nano silica organic fertilizer. **Chilean journal of agricultural research**, v. 84, n. 3, p. 362-371.

Horwitz W, Chichilo P, Reynolds H (1970) Official methods of analysis of the Association of Official Analytical Chemists, Washington, DC, USA: **Association of Official Analytical Chemists**.

Hu, Jianchang; Szymczak, Silke. A review on longitudinal data analysis with random forest. **Briefings in bioinformatics**, v. 24, n. 2, p. bbad002, 2023.

Idrishi, R., Shelke, R. G., Singha, S., Rangan, L. (2024). Black rice starch: physicochemical characterization and nutritional properties. In: Non-Conventional Starch Sources. **Academic Press**, 2024. p. 197-223.

Islam, M. M., Himel, G. M. S., Moazzam, M. G., & Uddin, M. S. (2025). Artificial Intelligence-based Rice Variety Classification: A State-of-the-Art Review and Future Directions. **Smart Agricultural Technology**, p. 100788.

Jimoh, K. A., Hashim, N., Shamsudin, R., Man, H. C., Jahari, M. (2025). Development of near-infrared hyperspectral-based smart interface for glutinous rice quality detection. **Food Control**, p. 111252.

Kästner, F., Sut-Lohmann, M., Ramezany, S., Raab, T., Feilhauer, H., Chabrillat, S. (2022). Estimating heavy metal concentrations in Technosols with reflectance spectroscopy. **Geoderma**, v. 406, p. 115512.

Kata, L., Kuna, A., Naresh, P., Sohan, Z. (2023). Impact of parboiling conditions on milling quality and the nutritional value of popular rice varieties of Telangana. **In Biol. Forum Int. J.** Vol. 15, p. 773-778.

Kaur, P., Singh, N., Pal, P., Kaur, A. (2018). Variation in composition, protein and pasting characteristics of different pigmented and non pigmented rice (*Oryza sativa* L.) grown in Indian Himalayan region. **Journal of food science and technology**, v. 55, p. 3809-3820.

Kumari, A., Sadh, P. K., Kamboj, A., Yadav, B., Kumar, A., Sivakumar, S., Duhan, J. S. (2025). Exploring the Benefits of Nutrition of Little Millet: Unveiling the Effect of Processing Methods on Bioactive Properties. **Journal of Food Biochemistry**, v. 2025, n. 1, p. 2488816.

Martens, S., Coradi, P. C., Maldaner, V., de Oliveira Carneiro, L., Teodoro, P. E., Rodrigues, D. M., Flores, É. M. M. (2023). Drying and intermittence processes on the polished and brown rice physicochemical and morphological quality by near-infrared spectroscopy, X-ray diffraction, and scanning electron microscopy. **Food Chemistry: X**, v. 19, p. 100753.

Martins, G. M. V., de Sousa, S., Duarte, M. E. M., Cavalcanti-Mata, M. E. R. M., Oliveira, H. M. L. (2021). Modeling the combinatory effects of parboiling and cooking on red paddy rice (*Oryza sativa* L.) properties. **LWT**, v. 147, p. 111607.

Mingoti, S. A. **Análise de Dados Através de Métodos de Astatística Multivariada: uma Abordagem Aplicada**. 1th. Editora UFMG: Belo Horizonte, 2005.

Mishra, P., Angileri, M., Woltering, E. (2021). Identifying the best rice physical form for non-destructive prediction of protein content utilising near-infrared spectroscopy to support digital phenotyping. **Infrared Physics & Technology**, v. 116, p. 103757.

Morellos, A., Pantazi, X. E., Almoujahed, M. B., Kriaučiūnienė, Z., Kazlauskas, M., Šarauskis, E., Mouazen, A. M. (2025). Forecasting of Fusarium head blight spatial distribution in winter wheat using machine learning. **Computers and Electronics in Agriculture**, v. 231, p. 109967.

Muchlisyyah, J., Shamsudin, R., Kadir Basha, R., Shukri, R., How, S., Niranjan, K., Onwude, D. (2023). Parboiled rice processing method, rice quality, health benefits, environment, and future perspectives: A review. **Agriculture**, v. 13, n. 7, p. 1390.

Mukherjee, R., Ghosh, A., Chakraborty, C., De, J. N., Mishra, D. P. (2025). Rice leaf disease identification and classification using machine learning techniques: A comprehensive review. **Engineering Applications of Artificial Intelligence**, v.139, p. 109639.

Müller, A., Nunes, M. T., Maldaner, V., Coradi, P. C., de Moraes, R. S., Martens, S., Marin, C. K. (2022). Rice drying, storage and processing: effects of post-harvest operations on grain quality. **Rice Science**, v. 29, n. 1, p. 16-30.

Nalepa, J.; Kawulok, M. Selecting training sets for support vector machines: a review. **Artificial Intelligence Review**, [s. l.], v. 52, n. 2, p. 857–900, 2019.

Obadi, M., Xu, B. (2023). Effect of processing methods and storage on the bioactive compounds of black rice (*Oryza sativa* L.): a review. **Food & Function**, v. 14, n. 20, p. 9100-9122.

Olaoni, S. O., Regonda, B., Luthra, K., Atungulu, G. G. (2025). Optimizing lab methods for consistent rice milling analysis. **Cereal Chemistry**, v. 102, n. 1, p. 199-210.

Panda, D. K., Jyotirmayee, B., Mahalik, G. (2022). Black rice: A review from its history to chemical makeup to health advantages, nutritional properties and dietary uses. **Plant Sci. Today**, v. 9, p. 1-15.

Pinta, W., Aninbon, C., Kaewtaphan, P., Kunyane, K. (2024). Effects of Parboiling on Chemical Properties, Phenolic Content and Antioxidant Capacity in Colored Landrace Rice. **Foods**, v. 13(3), 393.

Priyanthi, C., Sivakanesan, R. (2021). The Total Antioxidant Capacity and the Total Phenolic Content of Rice Using Water as a Solvent, **International Journal of Food Science**, 5268584, p. 6.

Qiao, M., Xu, Y., Xia, G., Su, Y., Lu, B., Gao, X., Fan, H. (2022). Determination of hardness for maize kernels based on hyperspectral imaging. **Food Chemistry**, v. 366, p. 130559.

Quinlan, J. R. C4. 5: Programming for machine learning. Morgan Kaufmann, [s. l.], v. 38, n. 48, p. 49, 1993.

Rahman, A. N. F., Tahir, M., Haddada, J., Bakri, R., Hambali, A., Husnah, A. (2025). Assessing the Effects of Diverse Germination Methods on Red Rice Quality: A Physicochemical and Sensory Approach. **Current Research in Nutrition and Food Science Journal**, v. 13, n. 1.

Ramos, A. H., da Silva Timm, N., Rockenbach, B. A., Ferreira, C. D., Hoffmann, J. F., de Oliveira, M. (2022). Red rice drying and storage: Effects on technological properties and phenolic compounds of the raw and cooked grains. **Journal of Cereal Science**, v. 103, p. 103405.

Reddy, M. P., Sen, A., Swamy, C. T., Chelladurai, P., Sivadurga, K., Pandey, A., Purewal, S. S. (2025) Chemistry of Macro-and Micronutrients. **Colored Cereals: Properties, Processing, Health Benefits, and Industrial Uses**, p. 96.

Rhowell Jr, N. T., Alseekh, S., Fernie, A. R., Bonto, A. P., Yu, J., Buenafe, R. J. Q., Sreenivasulu, N. (2024). Comprehensive lipidomic insights of differentially accumulating lipids in large pigmented rice sprout collection and the changes in the starch composition upon germination. **Food Chemistry**, v. 460, p. 140677.

Saleem, S., Mushtaq, N. U., Shah, W. H., Rasool, A., Hakeem, K. R., Seth, C. S., Rehman, R. U. (2023). Millets as smart future food with essential phytonutrients for promoting health. **Journal of Food Composition and Analysis**, v. 124, p. 105669.

Sampaio, P. S., Brites, C. M. (2022). Near-Infrared Spectroscopy and Machine Learning: Analysis and. **Integrative Advances in Rice Research**, p. 257.

Santana, J. C. O., Pereira de Gusmão, R. P., Tejo Cavalcanti, M., de Luna Freire, K. R., Moreira de Carvalho, L., Sousa Galvão, M., Madruga, M. S., Abrantes da Silva Souza, T., Lisboa, H. M., & Nascimento, A. P. S. (2025). The role of red rice in craft beer: A sensory and nutritional evaluation. **Cereal Chemistry**, v. 102, p. 211–225.

Sarkar, M. A. R., Habib, M. A., Sarker, M. R., Rahman, M. M., Alam, S., Manik, M. N. I., Bhandari, H. (2024). Rice farmers' preferences for seed quality, packaging, and source: A study from northern Bangladesh. **Plos one**, v. 19, n. 6, p. e0306059.

Shen, S.; Chi, C.; Zhang, Y.; Li, L.; Chen, L.; Li, X. (2021) New Insights into How Starch Structure Synergistically Affects the Starch Digestibility, Texture, and Flavor Quality of Rice Noodles. **Int. J. Biol. Macromol.** v.184, p. 731–738.

Shi, S., Tang, Z., Ma, Y., Cao, C., Jiang, Y. (2023). Application of spectroscopic techniques combined with chemometrics to the authenticity and quality attributes of rice. **Critical Reviews in Food Science and Nutrition**, v. 65, n. 5, p. 913–935.

Shoukat, R., Cappai, M., Pilia, L., Pia, G. (2025). Rice Starch Chemistry, Functional Properties, and Industrial Applications: A Review. **Polymers**, v. 17, n. 1, p. 110.

Silva, C. S., Santana, D. C., Baio, F. H. R., Seron, A. C. D. S. C., Alvarez, R. D. C. F., Teodoro, L. P. R., Teodoro, P. E. (2025). High-Precision Phenotyping in Soybeans: Applying Multispectral Variables Acquired at Different Phenological Stages. **AgriEngineering**, v.7, n. 2, p. 47.

Snousy, M. B. Al; El-Deeb, H. M.; Badran, K.; Khilil, I. A. Al. (2011). Suite of decision tree-based classification algorithms on cancer gene expression data. **Egyptian Informatics Journal**, [s. l.], v. 12, n. 2, p. 73–82.

Song, Y., Cao, S., Chu, X., Zhou, Y., Xu, Y., Sun, T., Liu, X. (2023). Non-destructive detection of moisture and fatty acid content in rice using hyperspectral imaging and chemometrics. **Journal of Food Composition and Analysis**, v. 121, p. 105397.

Štepanovský, M., Ibrová, A., Buk, Z., Velemínská, J. Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. **Forensic Science International**, [s. l.], v. 279, p. 72–82, 2017.

SubbuThavamurugan, Dhivyadharchini, M., Suresh, P., Manikandan, T., Vasuki, A., Nandhagopalan, V., Prabha, A. M. L. (2023). Investigation on nutritional, phytochemical, and

antioxidant abilities of various traditional rice varieties. **Applied Biochemistry and Biotechnology**, v. 195, n. 4, p. 2719-2742.

Sun, Z., Li, A., Ji, S., Li, H., Li, Z., Gao, H., Zhao, D. (2025). Numerical Study of Rice Grain Milling Uniformity in the Abrasive Milling Process. **Foods**, v. 14, n. 4, p. 630.

Team, R. C. R: A language and environment for statistical computing. [s. l.], 2013.

Teodoro, L. P. R., Silva, M. O., dos Santos, R. G., de Alcântara, J. F., Coradi, P. C., Biduski, B., Teodoro, P. E. (2024). Machine learning for classification of soybean populations for industrial technological variables based on agronomic traits. **Euphytica**, v. 220, n. 3, p. 40.

Thilavech, T., Suantawee, T., Chusak, C., Suklaew, P. O., Adisakwattana, S. (2025). Black rice (*Oryza sativa* L.) and its anthocyanins: mechanisms, food applications, and clinical insights for postprandial glycemic and lipid regulation. **Food Production, Processing and Nutrition**, v. 7, n. 1, p. 15.

Uivarasan, A., Lukinac, J., Jukić, M., Šelo, G., Peter, A., Nicula, C., Mihaly Cozmata, L. (2024). Characterization of Polyphenol Composition and Starch and Protein Structure in Brown Rice Flour, Black Rice Flour and Their Mixtures. **Foods**, v. 13, n. 11, p. 1592.

Vahidi, Milad; Shafian, Sanaz; Frame, William Hunter (2025). Precision Soil Moisture Monitoring Through Drone-Based Hyperspectral Imaging and PCA-Driven Machine Learning. **Sensors**, v. 25, n. 3, p. 782.

Wang, L., Wang, W., Huang, Z., Zhen, S., Wang, R. (2024). Discrimination of internal crack for rice seeds using near infrared spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 319, p. 124578.

Wei, L., Ma, F., Du, C. (2021). Application of FTIR-PAS in rapid assessment of rice quality under climate change conditions. **Foods**, v. 10, n. 1, p. 159.

Weng, S., Tang, L., Wang, J., Zhu, R., Wang, C., Sha, W., Chu, Z. (2023). Detection of amylase activity and moisture content in rice by reflectance spectroscopy combined with spectral data transformation. **Spectrochimica acta part A: Molecular and biomolecular spectroscopy**, v. 290, p. 122311.

Wu, H., Nakamura, T., Nakamura, Y. (2021). Predominant rice phytochemicals and their disease-preventive effects. **Journal of Environmental Science for Sustainable Society**, v. 10, n. Supplement, p.01.

Yang, Y., Zhao, Y., Qin, K., Zhao, N., Zhang, D., Cui, X., Li, X. (2025). Application of hyperspectral remote sensing technology in the quality evaluation of black soil. **In: Advanced Fiber Laser Conference (AFL 2024)**. SPIE, 2025. p. 446-465.

Yanjie, X. U., Yining, Y., Shuhong, O., Xiaoliang, D., Hui, S., Shukun, J., Jinsong, B. (2018). Factors affecting sensory quality of cooked japonica rice. **Rice Science**, v. 25, n. 6, p. 330-339.

Zahir, S. A. D. M., Jamlos, M. F., Omar, A. F., Nordin, M. A. H., Raypah, M. E. A., Mamat, R., Muncan, J. (2025). Quantifying the Impact of Varied NPK Fertilizer Levels on Oil Palm Plants during the Nursery Stage: A Vis-NIR Spectral Reflectance Analysis. **Smart Agricultural Technology**, p. 100864.

Zhang, W., Zhu, H., Rong, L., Chen, Y., Yu, Q., Shen, M., Xie, J. (2023). Purple red rice bran anthocyanins reduce the digestibility of rice starch by forming V-type inclusion complexes. **Food Research International**, v. 166, p. 112578.

Zheng, R., Jia, Y., Ullagaddi, C., Allen, C., Rausch, K., Singh, V., Kamruzzaman, M. (2024). Optimizing feature selection with gradient boosting machines in PLS regression for predicting moisture and protein in multi-country corn kernels via NIR spectroscopy. **Food chemistry**, v. 456, p. 140062.

Zhong, Y., Zhang, Y., Liu, X., Liu, C., Wu, J., Huang, H., Zeng, Z. (2023). Structural characteristics of cooked black rice influenced by different stabilization treatments and their effect mechanism on the in vitro digestibility. **Food and Bioprocess Technology**, v. 16, n. 4, p. 897-908.

Zhou, K., Luo, Z., Huang, W., Liu, Z., Miao, X., Tao, S., Zeng, X. (2024). Biological roles of lipids in rice. **International Journal of Molecular Sciences**, v. 25, n. 16, p. 9046.

Zhou, Z., Robards, K., Helliwell, S., Blanchard, C. (2002). Composition and functional properties of rice. **International Journal of Food Science and Technology**, v. 37, n. 8, p. 849-868.

Zhu, Z., Li, X., Liu, R., Wang, Z., Zhang, H., Zhao, D., Li, W. (2024). Workability and Environmental Evaluation of Using Sunflower Stalk Ash as an Alkali Activator with Blast Furnace Slag and Black Rice Hull Ash to Prepare Geopolymer Grouts. **Waste and Biomass Valorization**, p. 1-18.

Zhuang, K., Zhang, Z., Shang, S., Zheng, K., Zhou, X., Huang, W., Ding, W. (2025). Comparing parboiling and milling for selenium-enriched rice (*Oryza sativa* L.): Differences in selenium speciation, texture, microstructure, and sensory. **Food Chemistry: X**, v. 25, p. 10216.