

Um Conjunto de Dados para Classificação da Qualidade de Sementes de Feijão a Partir de Imagens Hiperespectrais

Eneri M. D’Angelis¹, Tsang Ing Ren¹, Maria Conceição M. de Souza²

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife, PE – Brasil

²Instituto Agrônômico de Pernambuco (IPA)
Recife, PE – Brasil

{emd,tir}@cin.ufpe.br, maria.martiniano@ipa.br

Abstract. *This article presents the construction of a dataset containing reflectance values in near-infrared wavelengths, obtained through a hyperspectral sensor, on Phaseolus vulgaris L. bean seeds, with germination quality classification as purpose. The exploratory analysis conducted on the acquired data reveals that the data obtained for normal and abnormal seeds exhibit a certain degree of distinction, enabling the use of the dataset for classification purposes.*

Resumo. *Este artigo apresenta a construção de um conjunto de dados com valores de refletância em comprimentos de onda próximo à região do infravermelho, obtidos através de um sensor hiperespectral, em sementes de feijão Phaseolus vulgaris L., para ser utilizado em classificação quanto à qualidade de germinação. A análise exploratória feita nas informações adquiridas mostram que os dados obtidos para sementes normais e anormais possuem um certo grau de distinção, o que torna possível o uso do conjunto de dados para classificação.*

1. Introdução

Um problema que afeta principalmente as lavouras de feijão no Brasil é a utilização de sementes clandestinas ou sementes produzidas pelo próprio agricultor, que não possuem certificação alguma, ou seja, não possuem garantia de bom rendimento ou mesmo de que irão germinar [Prezepiorski 2023]. Para evitar problemas como este, o Ministério da Agricultura possui normativas a serem seguidas por produtores de sementes, seguindo estas normativas garante-se a certificação das sementes produzidas.

Sementes certificadas possuem garantia de qualidade de 80% do lote quanto à capacidade de germinação, assim como estarem livres de fungos e doenças [Prezepiorski 2023]. Alguns dos testes feitos para alcançar essa garantia são análise de pureza, teste de tetrazólio, determinação do grau de umidade e teste de germinação, para a realização dos testes citados é selecionada uma amostragem de sementes que irá ser responsável pela validação de todo o lote seguindo as normativas [Ministério da Agricultura 2009].

A categorização para o conjunto de dados que será descrito neste artigo teve como base o teste de germinação, que apesar de existir diferentes métodos, o mais convencional é feito de forma manual, no qual a amostragem obtida para validação de um lote de sementes é submetida a condições controladas e favoráveis a germinação. Após germinação,

a plântula, como é chamada as sementes após germinarem, mas que ainda não se desenvolveram por completo [Kellermann et al. 2016], precisa apresentar algumas estruturas essenciais, o que é avaliado por um técnico treinado após o período de germinação estabelecido para o teste [Ministério da Agricultura 2009].

Este artigo apresenta um conjunto de dados composto por imagens hiperespectrais de sementes de feijão *Phaseolus vulgaris* L., desenvolvido com o objetivo de ser utilizado na classificação das sementes quanto a sua germinação. As imagens hiperespectrais são capazes de capturar informações dos objetos, neste caso sementes, em uma ampla gama de comprimentos de onda, e como diferentes componentes químicos refletem o fluxo de radiação incidente, o que permite uma análise mais precisa e abrangente das características das sementes em diferentes estágios de germinação [Mayrinck et al. 2020].

Nas seções seguintes deste artigo, exploraremos a geração e a composição do conjunto de dados de imagens hiperespectrais de sementes de feijão preto da subespécie *Phaseolus vulgaris* L., discutiremos as técnicas de aquisição de dados e como foi feita a classificação de germinação. Além disso, serão abordadas aplicações potenciais desse conjunto de dados e como ele pode contribuir para avanços na análise de sementes. Este trabalho visa fornecer uma visão abrangente da relevância e do potencial impacto desse conjunto de dados na comunidade científica e na indústria agrícola.

2. Revisão de Literatura

Nesta seção, serão apresentados e discutidos trabalhos relevantes na área de imagens hiperespectrais com ênfase na análise de sementes.

2.1. Imagens Hiperespectrais

As imagens hiperespectrais são registros visuais adquiridos por meio de sensores que efetuam medições em uma extensa faixa de comprimentos de onda, abrangendo todo o espectro eletromagnético. Essas imagens proporcionam informações espectrais pormenorizadas para cada ponto, viabilizando a identificação e a caracterização de materiais e objetos com base nas suas distintas assinaturas espectrais [Plaza et al. 2009].

O processamento de imagens hiperespectrais refere-se à análise e manipulação de imagens obtidas por meio de sensores capazes de medir uma ampla faixa de comprimentos de onda em todo o espectro eletromagnético. Essas imagens hiperespectrais fornecem informações espectrais detalhadas para cada pixel, possibilitando a identificação e caracterização de materiais e objetos com base em suas distintas assinaturas espectrais. O processamento de imagens hiperespectrais engloba uma variedade de técnicas avançadas, incluindo classificação, segmentação, desmistificação espectral e integração espacial-espectral, todas destinadas a extrair informações significativas e úteis dos dados capturados.

2.2. Utilização de Imagens Hiperespectrais na Análise de Sementes

A imagem hiperespectral possui diversas vantagens na avaliação de qualidade e segurança de sementes:

- Ela permite a obtenção de informações espectrais e espaciais ao mesmo tempo, facilitando a análise em regiões específicas das imagens.

- É eficiente na rápida avaliação das composições e propriedades das sementes, especialmente em sementes individuais.
- Oferece a possibilidade de criar mapas de previsão visual úteis em aplicações industriais.
- Possibilita uma amostragem eficiente, permitindo a análise de mais sementes com maior variação.
- Pode ser combinada com aprendizado profundo para identificar variedades de sementes com resultados promissores.

Em resumo, a imagem hiperespectral é uma ferramenta valiosa para a avaliação de sementes devido à sua capacidade de adquirir informações espectrais e espaciais simultaneamente, sua eficiência na triagem de propriedades das sementes e seu potencial para aplicações industriais e aprendizado profundo [Feng et al. 2019].

2.3. Vantagens da Utilização de Imagens Hiperespectrais para Análise de Sementes

As técnicas de imagem óptica apresentam diversas vantagens em comparação aos métodos tradicionais de análise para a avaliação da qualidade das sementes:

- Velocidade e não destrutividade: As técnicas de imagem óptica oferecem uma forma rápida e que não danifica as sementes para avaliar sua qualidade. Elas permitem a visualização em tempo real de características físicas, químicas, fisiológicas e patológicas das sementes sem causar danos a elas.
- Medidas precisas e consistentes: Essas técnicas, como a fluorescência da clorofila e a imagem multiespectral, podem fornecer medidas precisas e consistentes da qualidade das sementes em comparação com os métodos convencionais. Elas são capazes de oferecer indicadores confiáveis do potencial fisiológico das sementes.
- Alta precisão e sensibilidade: A imagem multiespectral, em especial, possui uma precisão elevada ao distinguir sementes de maior vigor das sementes com menor vigor, mesmo antes de serem visíveis a olho nu. Isso ocorre porque ela consegue capturar informações tanto espaciais quanto espectrais ao mesmo tempo, fornecendo detalhes minuciosos sobre textura, cor, forma, tamanho e composição química das sementes.
- Potencial para automação e triagem em larga escala: As técnicas de imagem óptica podem ser incorporadas em sistemas automatizados, permitindo a triagem eficiente de grandes quantidades de sementes. Isso pode acelerar significativamente o processo de avaliação da qualidade das sementes na indústria agrícola.
- Utilidade em programas de melhoramento: A imagem de fluorescência da clorofila pode ser uma ferramenta valiosa no desenvolvimento de novas variedades em programas de melhoramento de sementes. Ela é capaz de distinguir com precisão entre diferentes cultivares e auxiliar na identificação de sementes com menor potencial fisiológico.

Em resumo, as técnicas de imagem óptica oferecem benefícios em termos de velocidade, precisão, alta precisão e capacidade de automação, tornando-as ferramentas valiosas na avaliação da qualidade das sementes nas indústrias de alimentos e agricultura [Galletti et al. 2020].

3. Metodologia

3.1. Pré Seleção das Sementes

Visando restringir a quantidade de sementes que precisariam ser analisadas para o conjunto de dados atingir a quantidade necessária para validação científica [Ministério da Agricultura 2009], foi realizada uma pré seleção nas sementes quanto à aspectos físicos da semente [Almeida et al. 2004].

Inicialmente foram selecionadas quatrocentas sementes pré classificadas como que não germinariam ou apresentariam anormalidades ou contaminação após germinação, estas foram selecionadas por possuírem danos mecânicos visíveis, manchas na superfície, alteração da coloração, enrugados na sua película (Figura 1).



Figure 1. Amostragem de sementes de feijão pré classificadas como anormais

Enquanto que foram selecionadas também quatrocentas e noventa e seis sementes pré classificadas como sementes que germinariam normalmente e as plântulas apresentariam todas as estruturas necessárias, assim como estariam também saudáveis e livres de contaminação para continuarem o crescimento [Ministério da Agricultura 2009]. Estas

foram escolhidas por não apresentarem dano algum, estarem com a superfície lisa e de aspecto brilhante e de cor uniforme (Figura 2).

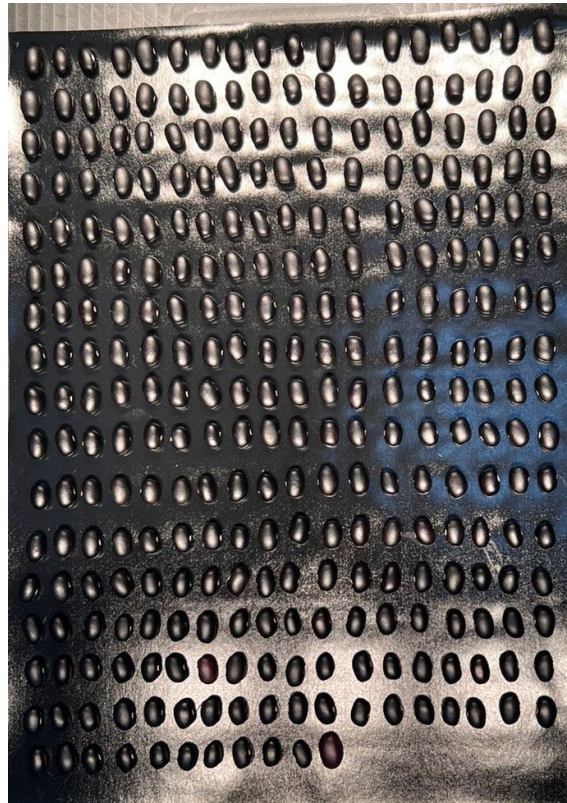


Figure 2. Amostragem de sementes de feijão pré classificadas como saudáveis

3.2. Obtenção das Imagens Hiperespectrais

Após seleção dos dois grupos de sementes foi realizada a captura das imagens hiperespectrais, assumindo que a pré classificação feita estava correta, as sementes foram coladas em fileiras em um papel adesivado preto, para evitar que a reflexão atrapalhasse na captura dos espectros das sementes. O uso do papel adesivado foi importante para não perder a ordem das sementes na hora de realizar o teste de germinação para classificação final do conjunto de dados.

As imagens foram obtidas com um sensor hiperespectral da SPECIM modelo FX17 (Figura 3), que captura espectros que formam uma imagem do alvo, de forma que cada pixel da imagem inclui um espectro completo [Specim 2020]. A assinatura espectral, que vai depender dos níveis de reflectância para a radiação que está sendo emitida em determinado espectro, de cada semente pode revelar a composição química ou até mesmo nível de umidade da mesma [Vicentini 2017].

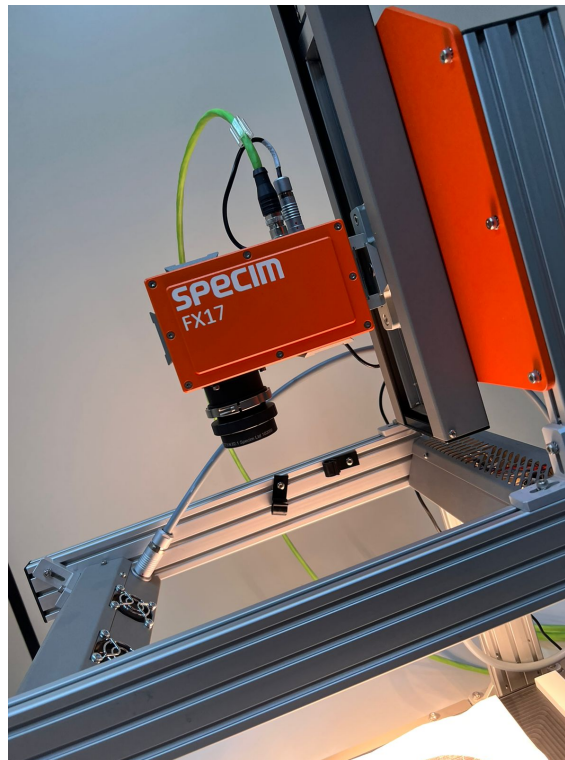


Figure 3. Sensor hiperespectral utilizado para captura das imagens

Como a germinação de uma semente depende da presença e quantidade de alguns componentes químicos armazenados em seu interior, como amido, proteínas e lipídios, que servem como fontes de energia e nutrientes para o embrião em crescimento, assim como a quantidade de água presente nas sementes também ser um dos fatores que influencia a germinação [Marcos Filho 2015]. A construção deste conjunto de dados parte do pressuposto de que esses componentes essenciais para o sucesso da germinação podem ser identificados de acordo com os valores obtidos pela reflectância da radiação.

3.3. Teste de Germinação para Classificação dos Dados

Apesar de ter sido feita uma pré classificação das sementes, esta foi feita baseada apenas no aspecto físico das sementes, então para ter-se um conjunto de dados realmente válido quanto a germinação foi realizado um teste de germinação com as sementes após a captura das imagens hiperespectrais. Para o teste de germinação as sementes foram colocadas na mesma ordem que estavam quando as imagens foram capturadas, para ter-se dados coerentes.

O teste de germinação foi feito seguindo as diretrizes do Ministério da Agricultura, o substrato escolhido para germinação foi papel toalha, as folhas foram previamente autoclavadas para eliminar toxicidade do material que pudesse vir a influenciar no teste [Ministério da Agricultura 2009]. Para realização do teste as folhas foram umedecidas com água na proporção recomendada, e foram utilizadas 3 folhas para cada 50 sementes.



Figure 4. Placa perfurada com 50 orifícios

As sementes foram posicionadas sobre 2 folhas do substrato utilizando uma placa perfurada (Figura 4) com 50 orifícios, para garantir o correto espaçamento entre as sementes. Após identificação da folha, para manter a mesma ordem das sementes referente a quando as imagens hiperespectrais foram obtidas, cobriu-se as sementes com uma terceira folha de papel toalha umedecido (Figura 5) e em seguida a folha foi enrolada para armazenamento durante o período de 7 dias.

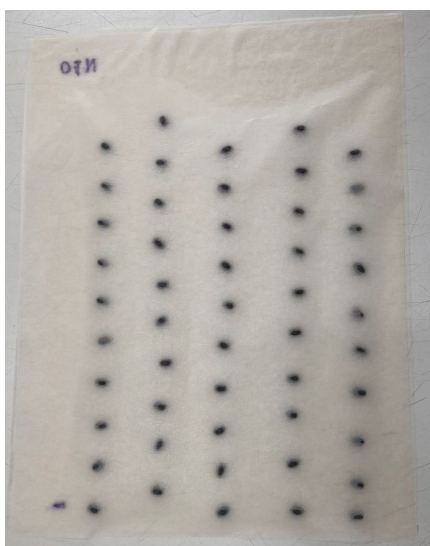


Figure 5. Sementes no substrato de papel toalha umedecido cobertas para teste de germinação

Após o período de germinação o resultado foi avaliado com a ajuda de uma profissional técnica certificada para análise de sementes, as sementes foram classificadas em dois grupos diferentes quanto a sua germinação, o grupo de sementes que germinaram normalmente foram colocadas no grupo de sementes normais (Figura 9), enquanto que as sementes que germinaram com alguma anormalidade ou infecção e as que não germinaram foram colocadas no grupo de sementes anormais (Figura 8). Nas Figuras 6 e 7, pode-se ver uma folha com 50 sementes anormais e normais, respectivamente, após teste de germinação.

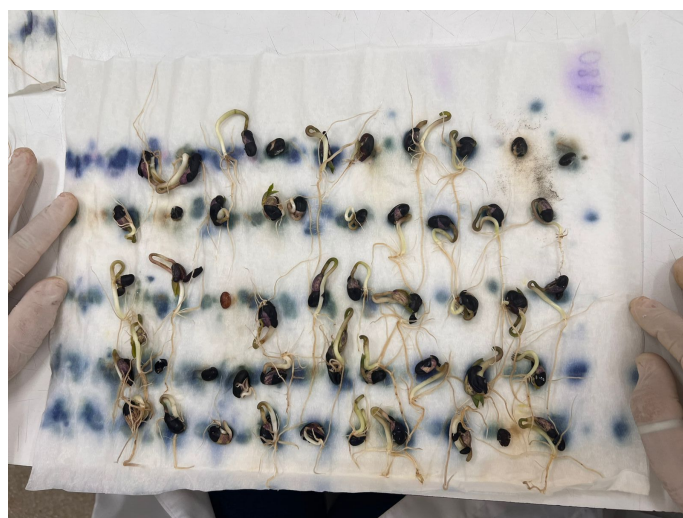


Figure 6. Resultado do teste de germinação de sementes pré classificadas como anormais



Figure 7. Resultado do teste de germinação de sementes pré classificadas como saudáveis



Figure 8. Exemplo de semente classificada como anormal após teste de germinação

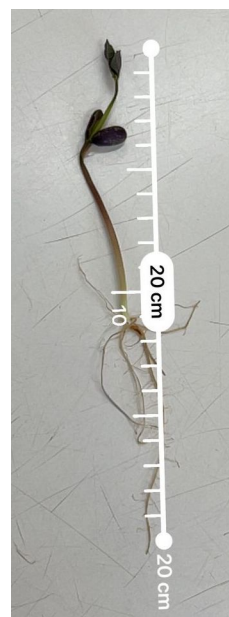


Figure 9. Exemplo de semente classificada como saudável após teste de germinação

3.4. Obtenção dos Dados

Após a captura das imagens com o sensor hiperespectral, como foi citado na subseção 3.2, foi gerado um arquivo de extensão *raw*. Para abrir o arquivo foi utilizado o software Evince, no qual além de apresentar os dados referentes a todos os espectros da imagem capturada, também fornece uma renderização RGB da imagem (Figura 10), na qual é possível selecionar apenas a área desejada para seleção dos dados referentes [Evince].

O próprio software, que foi utilizado para obtenção dos dados, proporciona a opção de normalização dos dados para substituição de valores de pico pela média. Para a construção do conjunto de dados descrito neste artigo, cada semente foi selecionada na imagem RGB, mapeando assim os dados na matriz *raw* referentes a semente específica, sendo estes dados copiados para um arquivo da semente em questão.

O conjunto de dados criado foi dividido em duas pastas, uma contendo os dados das sementes consideradas normais após teste de germinação e a outra com dados das sementes anormais. Cada semente possui um arquivo *csv* contendo uma matriz com 224 colunas e o número de linhas varia de acordo com a área da semente analisada.

4. Análise Exploratória dos Dados

Após validação feita através do teste de germinação, o conjunto de dados resultante possui 426 sementes anormais e 470 sementes normais. Os valores encontrados nas matrizes que representam cada semente se refere a refletância de determinado ponto da semente em um comprimento de onda específico, cada coluna da matriz refere-se a um comprimento de onda. Os comprimentos de onda do conjunto de dados variam de 927.87nm até 1714.72nm, se mantendo em um valor próximo do infravermelho.

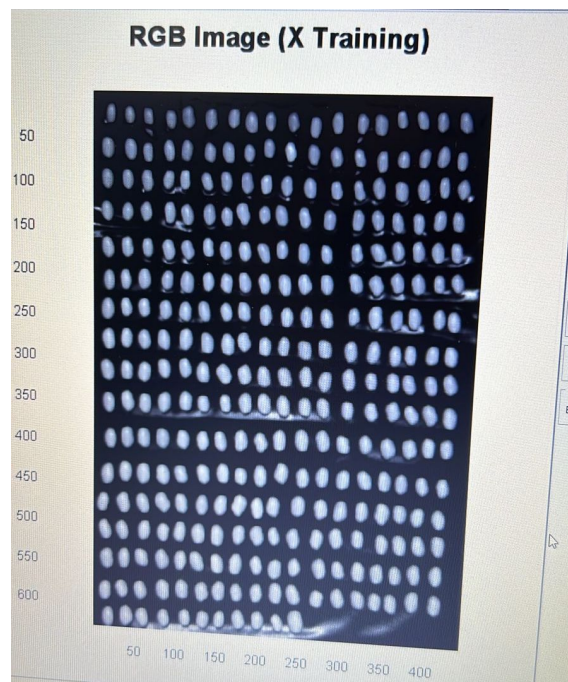


Figure 10. Imagem RGB extraída da câmera hiperespectral no *software* Evince

Algumas análises foram feitas com o conjunto de dados para compreendê-lo melhor e explorar a possibilidade de utilizá-lo para classificação de sementes quanto à sua qualidade. Nesta sessão tem-se algumas imagens das análises feitas, mas as mesmas podem ser encontradas no repositório do *Github* do conjunto de dados¹.

As imagens abaixo são gráficos *boxplot* comparativos entre os valores médios (Figura 11) para cada coluna das sementes consideradas anormais e as sementes normais, assim como valores máximos (Figura 12) e valores mínimos (Figura 13). Um ponto interessante para o propósito de classificação é que a média dos valores para sementes normais apresentam uma faixa de valores distintos aos apresentados pelas sementes anormais.

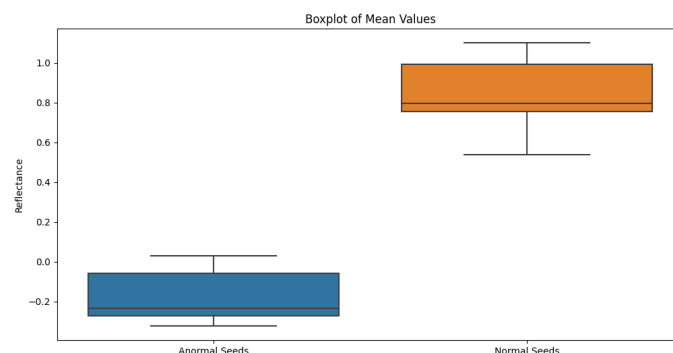


Figure 11. Comparativo de valores médios de cada coluna entre sementes anormais e normais

¹<https://github.com/eneridangelis/beansSeedAnalysis>

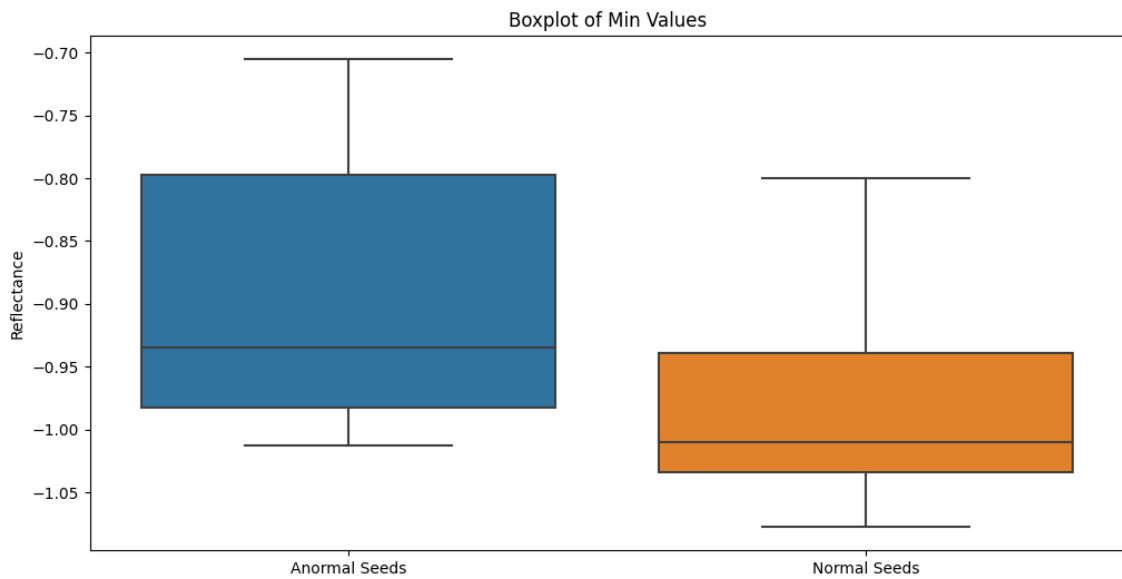


Figure 12. Comparativo de valores mínimos de cada coluna entre sementes anormais e normais

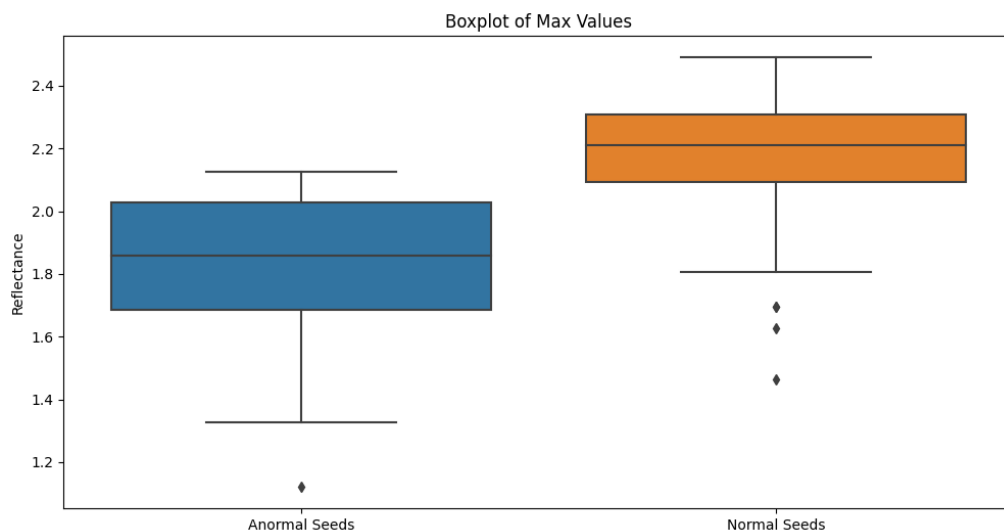


Figure 13. Comparativo de valores máximos de cada coluna entre sementes anormais e normais

Para comparação dos dados referentes a classificação das sementes, uma outra análise feita foi plotar as informações referentes à apenas uma semente normal (Figura 14) e uma anormal (Figura 15), para poder comparar os valores entre elas. Pode-se então atestar que a faixa de valores para cada comprimento de onda, representados pelas colunas, são bastante diferentes entre as sementes normais e anormais.



Figure 14. *Boxplot* de valores de refletância para uma semente anormal

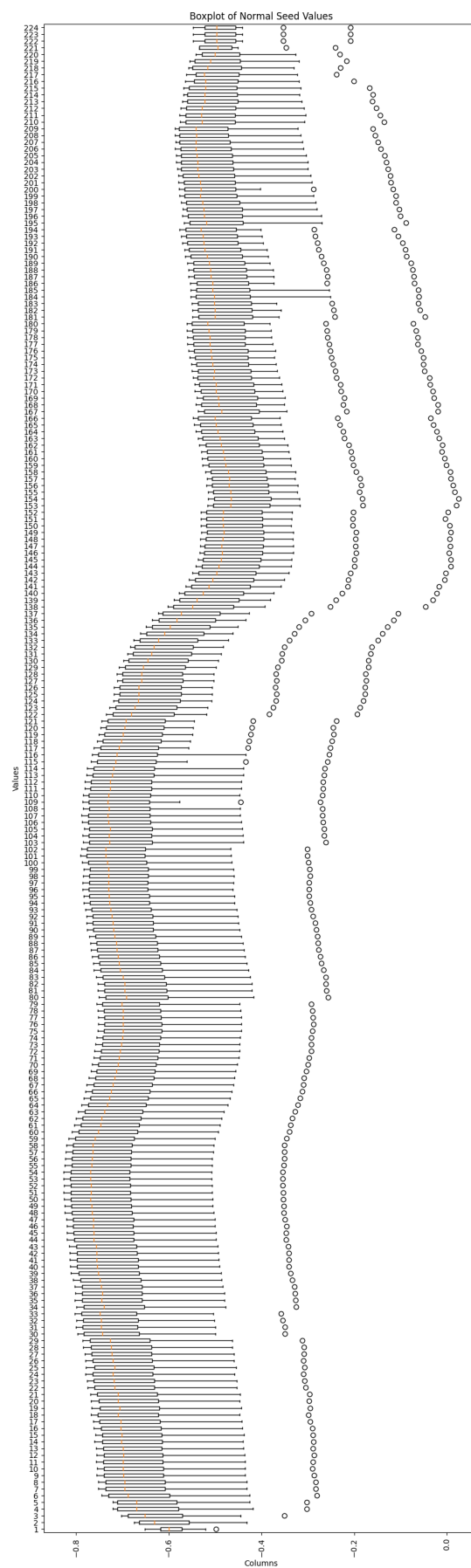


Figure 15. *Boxplot* de valores de refletância para uma semente normal

Por fim, foram plotados dois gráficos de linha para comparar os valores médios da reflectância para cada comprimento de onda por semente entre as sementes anormais (Figura 16) e normais (Figura 17). Nas imagens citadas, o eixo y representa cada semente, e o eixo x representa as colunas do conjunto de dados, referentes aos comprimentos de ondas. A diferença entre as sementes normais e anormais fica nítida observando a cor dominante nos gráficos, observando também a legenda das cores.

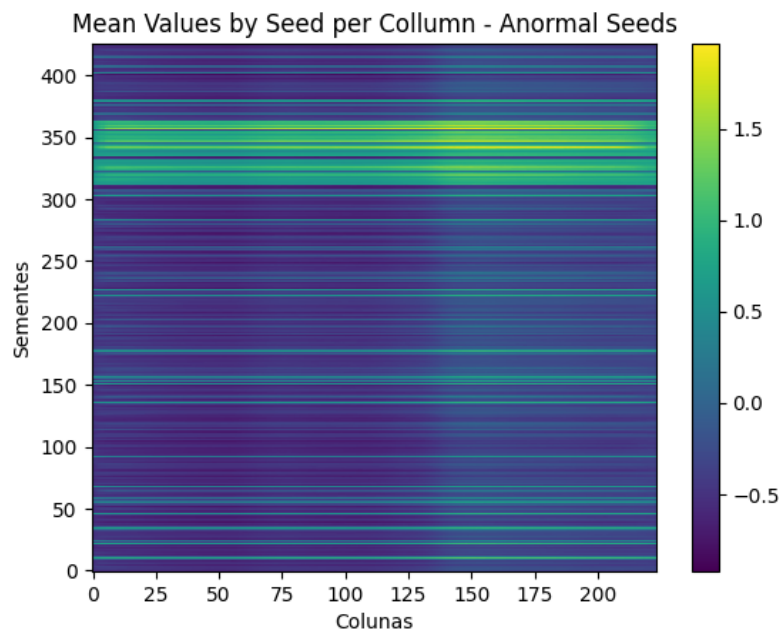


Figure 16. Gráfico de linha de valores médios por coluna para cada semente anormal

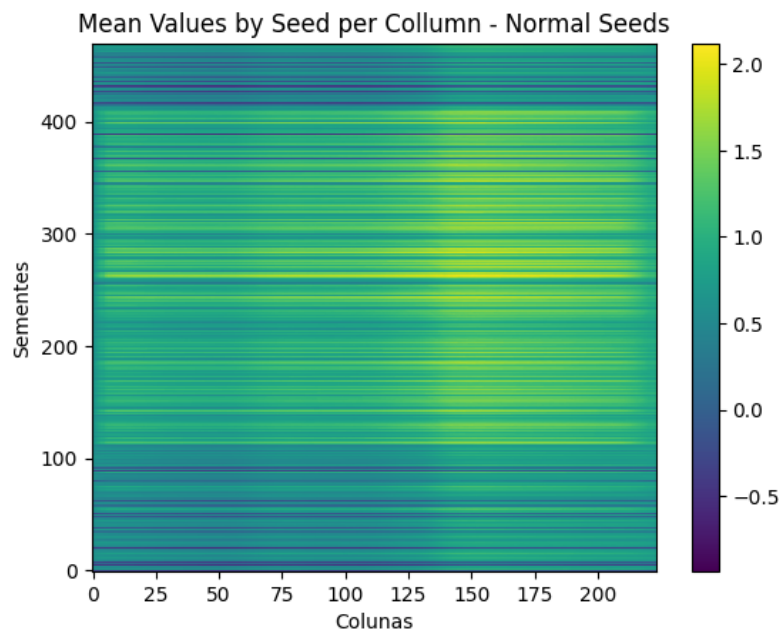


Figure 17. Gráfico de linha de valores médios por coluna para cada semente normal

5. Conclusão

Dado que a partir de informações quanto à refletância nos comprimentos de onda próximos ao infravermelho pode fornecer informações sobre componentes químicos essenciais para germinação das sementes [Mayrinck et al. 2020]. A construção do conjunto de dados apresentado neste artigo partiu do princípio de que os dados obtidos seriam ideais para aplicação de um modelo de classificação para identificação das sementes anormais e normais quanto à germinação.

A partir dos gráficos mostrados na seção anterior pode-se notar que são estabelecidos padrões distintos entre o conjunto de dados referentes às sementes consideradas normais e ao das sementes consideradas anormais. Essa distinção faz-se acreditar que é possível classificar as sementes através dos dados obtidos por imagens hiperespectrais, que é uma possível aplicação para o conjunto de dados fornecido neste artigo.

References

- Almeida, F. d. A. C., Figueiredo Neto, A., Costa, R. F. d., Gouveia, J. P. G. d., and Oliveira, M. E. C. d. (2004). Danos mecânicos em sementes de feijão vigna, causados pelas operações na unidade de beneficiamento. *Rev. Bras. Eng. Agric. Ambient./Braz. J. Agric. Environ. Eng.*, 8(2-3):254–259.
- Evince. Evince - tutorials and documentation. <https://prediktera.com/evince-samples-files-and-tutorials/>. Accessed: 2023-9-13.
- Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., and Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods*, 15(1).
- Galletti, P. A., Carvalho, M. E. A., Hirai, W. Y., Brancaglioni, V. A., Arthur, V., and Barboza da Silva, C. (2020). Integrating optical imaging tools for rapid and non-invasive characterization of seed quality: Tomato (*solanum lycopersicum* l.) and carrot (*daucus carota* l.) as study cases. *Front. Plant Sci.*, 11.
- Kellermann, B., Bona, C., and Souza, L. A. d. (2016). Morfoanatomia da plântula e comparação da folha nas fases juvenis e adulta de *piptocarpha angustifolia* (asteraceae). *Rodriguésia*, 67(3):627–638.
- Marcos Filho, J. (2015). *Fisiologia de sementes de plantas cultivadas*. Abrates.
- Mayrinck, L. G., Lima, J. M. E., Guimarães, G. C., Nunes, C. A., and Oliveira, J. A. (2020). Use of near infrared spectroscopy in cotton seeds physiological quality evaluation. *J. Seed Sci.*, 42.
- Ministério da Agricultura, P. e. A. (2009). https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise_sementes.pdf. Accessed: 2023-9-7.
- Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., and Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:S110–S122. Imaging Spectroscopy Special Issue.

Prezepiorski, E. C. F. (2023). Comércio de sementes clandestinas gera prejuízo de R\$ 2,5 bi por ano. <https://www.gazetadopovo.com.br/parana/comercio-sementes-clandestinas-gera-prejuizo/>. Accessed: 2023-9-12.

Specim (2020). Specim FX17. <https://www.specim.com/products/specim-fx17/>. Accessed: 2023-9-12.

Vicentini, T. (2017). O que é uma câmera hiperspectral? <http://astro34.com.br/o-que-e-uma-camera-hiperspectral/>. Accessed: 2023-9-13.