

**JOHAN BERNARDO GARCIA CUADRADO**

**ANÁLISE DE IMAGEM E APRENDIZAGEM DE MÁQUINA NA AVALIAÇÃO  
DA QUALIDADE DE SEMENTES DE ARROZ**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Magister Scientiae*.

Orientadora: Denise Cunha Fernandes dos Santos Dias

**VIÇOSA – MINAS GERAIS  
2020**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

Garcia Cuadrado, Johan Bernardo, 1981-

G216a  
2020

Análise de imagem e aprendizagem de máquina na  
avaliação da qualidade de sementes de arroz / Johan Bernardo  
Garcia Cuadrado. – Viçosa, MG, 2020.

56 f. : il. (algumas color.) ; 29 cm.

Orientador: Denise Cunha Fernandes dos Santos Dias.  
Dissertação (mestrado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. *Oryza sativa* L. 2. Arroz - Qualidade. 3. Pureza da  
semente. I. Universidade Federal de Viçosa. Departamento de  
Fitotecnia. Programa de Pós-Graduação em Fitotecnia. II. Título.

CDD 22. ed. 633.18

**JOHAN BERNARDO GARCIA CUADRADO**

**ANÁLISE DE IMAGEM E APRENDIZAGEM DE MÁQUINA NA AVALIAÇÃO  
DA QUALIDADE DE SEMENTES DE ARROZ**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Magister Scientiae*.

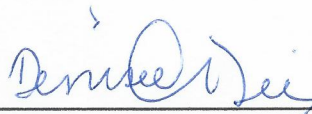
APROVADA: 21 de setembro de 2020.

Assentimento:



Johan Bernardo Garcia Cuadrado

Autor



Denise Cunha Fernandes dos Santos Dias

Orientadora

Ao meu pai Bernardo Julio (*in memoriam*)  
por todo amor, sabedoria e educação.

**DEDICO**

## AGRADECIMENTOS

Aos meus pais, Bernardo e Carmen, pela educação recebida através de bons exemplos e valores que levarei comigo por toda vida.

À minha esposa Monica e meus filhos Natalia e Johan David pelo amor e inspiração.

Às minhas irmãs, pelo apoio e por acreditarem em mim e a todos familiares que sempre estiveram ao meu lado.

À empresa FEDEARROZ, pela oportunidade, apoio e sempre acreditar em min.

À Universidade Federal de Viçosa ao Programa de Pós-Graduação em Fitotecnia por contribuir com minha formação profissional e pela oportunidade.

À professora, pesquisadora e orientadora Dra. Denise Cunha Fernandes dos Santos Dias, pelos ensinamentos, apoio e compreensão.

Ao professor Dr. Laércio Junio da Silva por sua amizade, sugestões, disponibilidade e suporte durante meus estudos.

A meus amigos André Medeiros e Manuel Zavala pela amizade, empenho e contribuição no desenvolvimento desta pesquisa.

Aos técnicos do laboratório de sementes e aos amigos do Grupo de Estudos em Sementes (GSEM): Wander, José Custódio, Daniel, Ariadne, Nayara, Rafaela, Ana Clara, Rubens, Fernanda, João Paulo, Ze Maria, Guilherme e Bruno pelo prazer da convivência, ensinamentos e por toda ajuda.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Eterna gratidão!

## RESUMO

CUADRADO, Johan Bernardo Garcia, M.Sc., Universidade Federal de Viçosa, setembro de 2020. **Análise de imagem e aprendizagem de máquina na avaliação da qualidade de sementes de arroz.** Orientadora: Denise Cunha Fernandes dos Santos Dias.

O Brasil é o maior produtor e consumidor de arroz fora do continente asiático, com crescente demanda anual na produção, o que aumenta também a necessidade de lotes de sementes de alta qualidade para a implantação da cultura. Para um eficiente controle da qualidade dos lotes a serem comercializados é importante disponibilizar para as empresas métodos de baixo custo e fácil execução que permitam obter informações seguras e rápidas de cada lote produzido. Neste contexto, destacam-se as análises automatizadas de imagens de sementes e de plântulas, que permitem obter informações sobre o potencial fisiológico das sementes, e métodos avançados capazes de auxiliar na identificação de misturas varietais. Diante disso, objetivou-se: i) avaliar a eficiência dos softwares SAPL<sup>®</sup> e ILASTIK na detecção de diferenças no potencial fisiológico de lotes de sementes de arroz; ii) propor um método rápido e não-destrutivo para auxiliar na identificação de cultivares arroz com uso de sistema de visão computacional combinado com algoritmos de aprendizagem de máquina. Foram conduzidos dois experimentos no Laboratório de Sementes do DAA/UFV, em delineamento inteiramente casualizado com quatro repetições. No experimento 1, foram utilizadas sementes de quatro lotes da cultivar F-2000, 6 lotes da cultivar Caçula, 6 lotes da cultivar Relâmpago, caracterizados inicialmente pelos testes de germinação, primeira contagem de germinação, envelhecimento acelerado, teste de frio, índice de velocidade de emergência e índice de velocidade de emissão de radícula. Com o software SAPL<sup>®</sup> foram obtidos os parâmetros comprimento total de plântula, raiz e parte aérea, índice de vigor e índice de uniformidade determinados aos 4 e 5 dias após o início do teste. Já com o software ILASTIK, realizou-se a classificação das plântulas quanto ao vigor obtendo-se os parâmetros número de plântulas vigorosas e número de plântulas fracas. Os dados foram submetidos à análise de variância realizada separadamente para cada cultivar e as médias obtidas foram comparadas pelo teste de Tukey a 5% de probabilidade. Os dados obtidos em cada teste foram também submetidos à análise de correlação de Pearson. No experimento 2, foram utilizadas sementes de oito cultivares arroz e sementes de arroz vermelho e arroz preto providas de amostras de lotes comerciais. As imagens das sementes foram adquiridas e processadas com auxílio do equipamento GroundEye<sup>®</sup>, obtendo depois da segmentação das regiões de interesse nas imagens, 312 variáveis de cor, textura e geometria de cada semente. Para a análise dos

dados, foi realizada estatística multivariada com a análise de componentes principais (PCA). Foram utilizados os algoritmos *Linear Discriminant Analysis* (LDA) e *Support Vector Machine - Radial* (SVM-R) para a criação dos modelos de classificação, avaliados com base na matriz de confusão para os cálculos das métricas acurácia, Kappa, sensibilidade, especificidade e acurácia balanceada. Concluiu-se que os softwares SAPL<sup>®</sup> e ILASTIK foram eficientes para a avaliação do potencial fisiológico de sementes de arroz, apresentando resultados correlacionados com os obtidos nos demais testes de vigor utilizados. Os parâmetros obtidos por meio da classificação das imagens geradas por ambos os sistemas computadorizados podem ser utilizados nos programas de controle de qualidade de sementes de arroz. Verificou-se também que os modelos desenvolvidos mostraram alto desempenho na identificação de cultivares arroz, principalmente com uso das características de cor e textura das imagens, obtendo-se acurácia de 78% e 80% com os algoritmos LDA e SVM-R, respectivamente. Esta alta precisão mostrou que os métodos de visão computacional com algoritmos de aprendizagem de máquina têm potencial para serem usados na identificação de sementes de cultivares de arroz.

Palavras-chave: *Oryza sativa* L. Qualidade fisiológica. Automatização.

## ABSTRACT

CUADRADO, Johan Bernardo Garcia, M.Sc., Universidade Federal de Viçosa, September, 2020. **Image analysis and machine learning in the evaluation of rice seed quality.** Advisor: Denise Cunha Fernandes dos Santos Dias.

Brazil is the largest producer and consumer of rice outside the Asian continent, with growing annual demand in production, which also increases the need for high quality seed lots for the implantation of the crop. For an efficient quality control of the lots to be sold it is important to make available to companies low cost and easy execution methods that allow to obtain safe and fast information of each lot produced. In this context, we highlight the automated analysis of images of seeds and seedlings that allow obtaining information about the physiological potential of seeds and also advanced methods capable of assisting in the identification of varietal mixtures. Therefore, the objective was to: i) evaluate the efficiency of the SAPL® and ILASTIK software in detecting differences in the physiological potential of rice seed lots; ii) to propose a fast and non-destructive method to assist in the identification of rice cultivars using a computer vision system combined with machine learning algorithms. Two experiments were conducted at the Seed Laboratory at DAA / UFV, in a completely randomized design with four replications. In experiment 1, seeds from four lots of cultivar F-2000 were used, 6 lots of cultivar Caçula 6 lots of cultivar Relâmpago, initially characterized by germination tests, first germination count, accelerated aging, cold test, speed index of emergency and radicle emission speed index. With the SAPL® software, the parameters of total length of seedling, root and shoot, vigor index and uniformity index determined at 4 and 5 days after sowing were obtained. With the ILASTIK software, seedlings were classified according to their vigor, obtaining the parameters number of vigorous plants and number of weak plants. The data were subjected to analysis of variance performed separately for each cultivar and the averages obtained compared by the Tukey test at 5%. The data obtained in each test were also submitted to Pearson's correlation analysis. In experiment 2, seeds from eight rice cultivars and seeds of red rice and black rice from commercial batch samples were used. The images of the seeds were acquired and processed with the aid of the Groundeye® equipment, obtaining after segmentation of the regions of interest in the images, 312 color, texture and geometry variables for each seed. For data analysis, multivariate statistics was performed with principal component analysis (PCA). Linear Discriminant Analysis (LDA) and Support Vector Machine - Radial (SVM-R) algorithms

were used to create the classification models, evaluated based on the confusion matrix for calculating the metrics accuracy, Kappa, sensitivity, specificity and balanced accuracy. It was concluded that the software SAPL® and ILASTIK were efficient for the evaluation of the physiological potential of rice seeds, corroborating the results obtained in the other vigor tests used. The parameters obtained by classifying the images generated by both computerized systems can be used in rice seed quality control programs. It was also found that the developed models showed high performance in the identification of rice cultivars, mainly using the color and texture characteristics of the images, obtaining accuracy of 78% and 80% with the LDA and SVM-R algorithms, respectively. This high precision showed that computer vision methods with machine learning algorithms have the potential to be used in the identification of rice cultivar seeds.

Keywords: *Oryza sativa* L. Physiological quality. Automation.

## LISTA DE ILUSTRAÇÕES

### Capítulo I

Figura 1 Padrão de plântula forte e fraca utilizado para o treinamento do software ILASTIK.....	26
Figura 2 Coeficientes de correlação simples de Pearson (r) estimados entre as variáveis germinação (G), primeira contagem de germinação (PC), envelhecimento acelerado (EA), teste de frio (F), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primária (IVR) e as variáveis obtidas na análise computadorizada.....	36

### Capítulo II

Figura 1 Análise de componentes principais considerando variáveis de cor, forma e textura de 5331 sementes de 10 genótipos de arroz.....	47
Figura 2 Vinte variáveis de maior importância para os modelos de classificação LDA e SVM-R de sementes de diferentes genótipos de arroz.....	51

## LISTA DE TABELAS

### Capítulo I

- Tabela 1 Caracterização da qualidade inicial de quatro lotes de arroz cv-F-2000 avaliados pelo grau de umidade (GU), primeira contagem de germinação (PCG), germinação (G), envelhecimento acelerado (EA), teste de frio (TF), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primária (IVR) .....27
- Tabela 2 Caracterização da qualidade inicial de seis lotes de arroz cv-Relâmpago, avaliados pelo grau de umidade (GU), primeira contagem de germinação (PCG), germinação (G), envelhecimento acelerado (EA), teste de frio (TF), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primaria (IVR) .....28
- Tabela 3 Caracterização da qualidade inicial de seis lotes de arroz cv-Caçula, avaliados pelo grau de umidade (GU), primeira contagem de germinação (PCG), germinação (G), envelhecimento acelerado (EA), teste de frio (TF), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primaria (IVR) .....29
- Tabela 4 Resultados médios dos parâmetros comprimento de parte aérea (CPA), comprimento radicular (CR), comprimento total (CT), índice de uniformidade (IU), índice de crescimento (IC) e índice de vigor corrigido (IVC) gerados pelo SAPL<sup>®</sup> e de número de plantas vigorosas (V), número de plantas fracas (F) obtidos pelo programa de aprendizagem de máquina ILASTIK, aos quatro e cinco dias após a semeadura de quatro lotes de sementes de arroz cv-F2000 .....31
- Tabela 5 Resultados médios dos parâmetros comprimento de parte aérea (CPA), comprimento radicular (CR), comprimento total (CT), índice de uniformidade (IU), índice de crescimento (IC), índice de vigor corrigido (IVC) gerados pelo SAPL<sup>®</sup> e do número de plantas vigorosas (V), número de plantas fracas (F) obtidos pelo programa de aprendizagem de máquina ILASTIK, aos quatro e cinco dias após a semeadura de seis lotes de sementes de arroz cv-Relâmpago .....33
- Tabela 6 Resultados médios dos parâmetros comprimento de parte aérea (CPA), comprimento radicular (CR), comprimento total (CT), índice de uniformidade (IU), índice de crescimento (IC), índice de vigor corrigido (IVC) gerados pelo SAPL<sup>®</sup> e de número de plantas vigorosas (V), número de plantas fracas (F) obtidos pelo programa de

aprendizagem de máquina ILASTIK, aos quatro e cinco dias após a semeadura de seis lotes de sementes de arroz cv-Caçula .....	35
--	----

## **Capítulo II**

Tabela 1 Matrizes de confusão para classificação de sementes de arroz com uso do algoritmo LDA .....	48
Tabela 2 Matrizes de confusão para classificação de sementes de arroz com uso do algoritmo SVMR .....	49

## SUMÁRIO

INTRODUÇÃO GERAL .....	13
REFERÊNCIAS BIBLIOGRÁFICAS .....	15
Capítulo I - análise de imagem e aprendizagem de máquina na avaliação do vigor de sementes de arroz	
RESUMO .....	16
ABSTRACT .....	17
INTRODUÇÃO.....	18
MATERIAL E MÉTODOS.....	22
RESULTADOS E DISCUSSÃO .....	26
CONCLUSÕES .....	37
REFERÊNCIAS BIBLIOGRÁFICAS.....	38
Capítulo II. Visão computacional e algoritmos de aprendizagem de máquina para classificação de sementes de cultivares de arroz	
RESUMO .....	40
ABSTRACT .....	42
INTRODUÇÃO.....	43
MATERIAL E MÉTODOS.....	44
RESULTADOS .....	47
DISCUSSÃO .....	51
CONCLUSÕES .....	54
REFERÊNCIAS BIBLIOGRÁFICAS .....	55

## INTRODUÇÃO GERAL

O arroz (*Oryza sativa* L.) é um dos principais cereais consumidos no mundo e reveste-se de grande importância social e econômica. O Brasil é o maior produtor de arroz das Américas, com um montante produzido de 15 milhões de toneladas em 2018 (FAO, 2019). A qualidade das sementes é um dos principais fatores responsáveis pelo sucesso da cultura, uma vez que a utilização de sementes de alta qualidade física, fisiológica, genética e sanitária é fundamental para a plantabilidade e uniformidade das plantas em campo, com reflexos na produtividade.

Os atributos de qualidade genética são importantes para se obter lavouras constituídas por plantas uniformes, com adaptabilidade às condições de ambiente onde são cultivadas, com características de resistência a pragas e doenças e que expressem todo o potencial produtivo da cultivar escolhida pelo produtor para o plantio.

Os atributos de qualidade fisiológica e física, igualmente importantes, devem atender a padrões estabelecidos por lei para a comercialização no país (BRASIL, 2009). Para a avaliação segura do potencial fisiológico dos lotes, faz-se necessária a utilização de testes de vigor para complementar as informações fornecidas pelo teste de germinação, principalmente quanto ao seu desempenho em campo e no armazenamento, e assim permitir tomadas de decisões quanto ao uso e destino dos lotes (MARCOS-FILHO, J. 2015). Na atualidade, há crescente demanda por métodos que permitam a obtenção de resultados confiáveis sobre o potencial fisiológico dos lotes de sementes, mas que sejam de baixo custo e fácil execução (MEDEIROS *et al.*, 2019).

A qualidade física do lote de sementes é fundamental para se evitar a contaminação da lavoura com plantas indesejáveis, que competirão em nutrientes, água e luz, reduzindo a produtividade e a eficiência da colheita (VLACHOS; ARVANITOYANNIS, 2008). O arroz vermelho tem sido um dos principais problemas para a produção de sementes de arroz, pois pertence a mesma espécie do arroz cultivado (*Oryza sativa* L.) e se distingue deste pela coloração vermelha do pericarpo das sementes, que sofrem degrana natural com alto grau de umidade, além de apresentarem dormência que lhes garante maior sobrevivência no solo. No entanto, devido ao cruzamento natural entre arroz vermelho e arroz cultivado, surgem plantas com características morfológicas semelhantes ao arroz cultivado, mas com pericarpo vermelho, tornando difícil sua diferenciação (MARCHEZAN, 1994). Portanto, a identificação eficiente da sua presença

nos lotes de sementes é essencial para o controle da sua disseminação e para a produção de lotes de elevado padrão de qualidade.

Neste contexto, as análises automatizadas de imagens podem se constituir em ferramenta importante para a avaliação da qualidade genética, física e fisiológica das sementes permitindo realizar um rigoroso controle de qualidade interno na empresa antes da comercialização dos lotes (HOFFMASTER *et al.*, 2003; MARCOS-FILHO *et al.*, 2009; CHAUGULE *et al.*, 2016; LIU *et al.*, 2016). Recentemente, novas e promissoras abordagens baseadas em técnicas de visão computacional e aprendizado de máquina têm sido aplicadas para a avaliação da qualidade das sementes (LUO *et al.*, 2019; HUANG *et al.*, 2017). Diversos sistemas foram também desenvolvidos para auxiliar na análise computadorizada de plântulas, destacando-se o Seed Vigor Imaging System (SVIS<sup>®</sup>) (SAKO *et al.*, 2001), o Sistema automatizado do vigor de sementes (Vigor-S<sup>®</sup>) (CASTAN *et al.*, 2018) e o Sistema de análise de plântulas (SAPL<sup>®</sup>) (MEDEIROS *et al.*, 2018). Assim, sistemas computacionais modernos podem auxiliar no controle de qualidade dos lotes de sementes de arroz, proporcionando a automatização das análises, garantindo menor subjetividade, rapidez e facilidade nos resultados de avaliação da qualidade das sementes (POURREZA *et al.*, 2012).

Portanto, o objetivo deste trabalho foi avaliar a eficiência dos softwares SAPL<sup>®</sup> e ILASTIK na detecção de diferenças no potencial fisiológico de lotes de sementes de arroz e também propor uma metodologia para auxiliar na identificação de cultivares arroz com uso de análise de simples imagens RGB de sementes e algoritmos de aprendizagem de máquina.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BRASIL. **Regras para Análise de Sementes**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2009.
- CASTAN, D. O. C.; GOMES-JUNIOR, F. G.; MARCOS-FILHO, J. Vigor-S, a new system for evaluating the physiological potential of maize seeds. **Scientia Agricola**, v. 75, n. 2, p. 167–172, 2018.
- CHAUGULE, A. A.; MALI, S. N. Identification of paddy varieties based on novel seed angle features. **Computers and Electronics in Agriculture**, v. 123, p. 415–422, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.compag.2016.03.012>>.
- FAO. **Food And Agricultural Organization**. [S.l: s.n.], abr. 2019
- HOFFMASTER, A.L.; FUJIMURA, K.; MCDONALD, M.B.; BENNETT, M. . An automated system for vigor testing three-day-old soybean seedlings. **Seed Science and Technology**, v. 31, n. 3, p. 701–713, 2003.
- HUANG, K.-Y. *et al.* A Novel Method of Identifying Paddy Seed Varieties. **Sensors**, v. 17, n. 4, p. 809, abr. 2017.
- LIU, W. *et al.* Online Variety Discrimination of Rice Seeds Using Multispectral Imaging and Chemometric Methods. **Journal of Applied Spectroscopy**, v. 82, n. 6, p. 993–999, 2016.
- LUO, Z. *et al.* Rapid identification of rice species by laser-induced breakdown spectroscopy combined with pattern recognition. **Applied Optics**, v. 58, n. 7, p. 1631, mar. 2019.
- MARCHEZAN, E. Arroz vermelho: Caracterization, damages e control. **Ciência Rural**, v. 24, n. 2, p. 415–421, 1994.
- MARCOS-FILHO, J.; KIKUTI, A. L. P.; LIMA, L. B. Métodos para avaliação do vigor de sementes de soja, incluindo a análise computadorizada de imagens. **Revista Brasileira de Sementes**, v. 31, n. 2, p. 102–112, 2009.
- MARCOS-FILHO, J. Seed vigor testing: an overview of the past, present and future perspective, **Scientia Agricola**, v.72, n.4, p.363-374, 2015.
- MEDEIROS, A. D. *et al.* Vigor of maize seeds determined by a free image analysis system. **Revista Ciencia Agronomica**, v. 50, n. 4, p. 616–624, 2019.
- MEDEIROS, A. D.; PEREIRA, M. D. SAPL®: A free software for determining the physiological potential in soybean seeds. **Pesquisa Agropecuaria Tropical**, v. 48, n. 3, p. 222–228, 2018.
- POURREZA, A. *et al.* Identification of nine Iranian wheat seed varieties by textural analysis with image processing. **Computers and Electronics in Agriculture**, v. 83, p. 102–108, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.compag.2012.02.005>>.
- SAKO, Y.; MCDONALD, M. B.; FUJIMURA, K. A System for Automated Seed Vigor Assessment. **Seed Science and Technology**, v. 29, n. 3, p. 625–636, 2001.
- VLACHOS, A.; ARVANITTOYANNIS, I. S. A review of rice authenticity/adulteration methods and results. **Critical Reviews in Food Science and Nutrition**, v. 48, n. 6, p. 553–598, 2008.

## RESUMO

CUADRADO, Johan Bernardo Garcia, M.Sc., Universidade Federal de Viçosa, setembro de 2020. **Capítulo I. Análise de imagem e aprendizagem de máquina na avaliação do vigor de sementes de arroz.** Orientadora: Denise Cunha Fernandes dos Santos Dias.

Na atualidade, novas ferramentas computacionais vêm se destacando e evidenciando resultados promissores para a avaliação rápida do potencial fisiológico de sementes. Tais ferramentas podem ser particularmente interessantes para sementes de arroz, cujo tempo necessário para obtenção dos resultados do teste de germinação é de 14 dias. Diante disso, objetivou-se avaliar a eficiência dos softwares SAPL<sup>®</sup> e ILASTIK na detecção de diferenças no potencial fisiológico de lotes de sementes de arroz, em comparação com testes de vigor recomendados para a espécie. O experimento foi conduzido no Laboratório de Sementes do DAA/UFV, em delineamento inteiramente casualizado com quatro repetições. Foram utilizadas sementes de quatro lotes da cultivar F-2000, seis lotes da cultivar Caçula e seis lotes da cultivar Relâmpago, caracterizados pelos testes de germinação, primeira contagem de germinação, envelhecimento acelerado, teste de frio, índice de velocidade de emergência e índice de velocidade de emissão de radícula. Com o software SAPL<sup>®</sup> foram obtidos os parâmetros comprimento total de plântula, raiz e parte aérea, índice de vigor e índice de uniformidade, determinados aos quatro e cinco dias após a semeadura. Já com o software ILASTIK, realizou-se a classificação das plântulas quanto ao vigor, obtendo-se os parâmetros número de plantas vigorosas e número de plantas fracas. Os dados foram submetidos à análise de variância, realizada separadamente para cada cultivar e as médias obtidas foram comparadas pelo teste de Tukey a 5% de probabilidade. Os dados obtidos em cada teste foram também submetidos à análise de correlação de Pearson. Os softwares SAPL<sup>®</sup> e ILASTIK foram eficientes para a avaliação do potencial fisiológico de sementes de arroz, corroborando os resultados obtidos nos demais testes de vigor utilizados. Ambos os softwares podem ser utilizados nos programas de controle de qualidade de sementes de arroz tendo como vantagens maior rapidez na obtenção de resultados, precisão e menor subjetividade.

**PALAVRAS-CHAVE:** *Oryza sativa* L. Qualidade fisiológica. Automatização.

## ABSTRACT

CUADRADO, Johan Bernardo Garcia, M.Sc., Universidade Federal de Viçosa, September, 2020. **Chapter I. Image analysis and machine learning to assess the vigor of rice seeds.** Advisor: Denise Cunha Fernandes dos Santos Dias.

Rice is one of the most produced cereals in the world, grown on five continents, and is an important part of the world's diet. Brazil is the largest producer and consumer of rice outside the Asian continent, with growing annual demand in production and the need for high quality physiological seed lots. The minimum time to obtain the results of the germination test of rice seeds is 14 days, thus, there is a need for methods that evaluate in a practical, efficient, fast and economical way the physiological potential of the seeds. Currently, new computational tools have been highlighted in the evaluation of the physiological potential of seeds. Therefore, the objective was to evaluate the efficiency of SAPL® and ILASTIK software in detecting differences in the physiological potential of rice seed lots, in comparison with the vigor tests recommended for the species. The experiment was carried out at the DAA / UFV Seed Laboratory, in a completely randomized design with four replications. Seeds from 4 lots of cultivar F-2000, 6 lots of cultivar Caçula and 6 lots of cultivar Relampago were used, initially characterized by germination tests, first germination count, accelerated aging, cold test, emergence speed index and radicle speed index. With the SAPL® software, the parameters of total length of seedling, root and shoot, vigor index and uniformity index determined at 4 and 5 days after sowing were obtained. With the ILASTIK software, seedlings were classified according to their vigor, obtaining the parameters number of vigorous plants and number of weak plants. The data were subjected to analysis of variance performed separately for each cultivar and the averages obtained compared by the Tukey test at 5%. The data obtained in each test were also submitted to Pearson's correlation analysis. The SAPL® and ILASTIK software were efficient for the evaluation of the physiological potential of rice seeds, corroborating the results obtained in the other vigor tests used. The parameters obtained through the classification of images generated by both computerized systems can be used in rice seed quality control programs, with the advantages of faster results, precision and less subjectivity.

**KEYWORDS:** *Oryza sativa* L. Physiological quality. Automation.

## INTRODUÇÃO

O arroz é um dos cereais mais produzidos no mundo, cultivado nos cinco continentes, faz parte importante na dieta alimentar mundial, com uma produção global estimada de 513,5 milhões de toneladas no ano 2019 (FAO, 2019), o que leva anualmente ao aumento das áreas cultivadas e à exigência por maiores rendimentos do cultivo.

O Brasil é o maior produtor e consumidor de arroz fora do continente asiático, com média de 15 milhões de toneladas ao ano em área total de 2,4 milhões de hectares (FAO, 2019). Com a demanda crescente da produção de arroz, aumenta também a necessidade de disponibilidade de lotes de sementes de alta qualidade.

O potencial fisiológico das sementes de arroz é avaliado pelos testes de germinação e de vigor. O teste de germinação, por ser realizado sob condições ideais, pode superestimar o real potencial fisiológico da semente não refletindo o seu desempenho em campo e no armazenamento. Contudo, é um teste importante para a comercialização dos lotes, uma vez que envolve procedimentos padronizados e permite a reprodutibilidade dos resultados. Os resultados do teste de germinação expressam o potencial máximo de germinação do lote (BRASIL, 2009).

Para a avaliação segura do potencial fisiológico dos lotes, faz-se necessária a utilização de testes de vigor para complementar as informações fornecidas pelo teste de germinação e permitir tomadas de decisões quanto à comercialização e destino dos lotes (MARCOS-FILHO, J. 2015). A utilização de sementes vigorosas é fundamental para se assegurar a emergência rápida e uniforme das plântulas em campo, assegurando a formação de estande de plantas adequado, evitando perdas na produtividade ou até a necessidade de ressemeadura.

Dentre as finalidades de se avaliar o vigor das sementes destacam-se: detectar diferenças entre lotes com germinação semelhante e obter estimativas sobre o seu desempenho em campo e no armazenamento (MARCOS-FILHO, J. 1999). Tais informações são importantes para se tomar as melhores decisões de uso e destino dos lotes, ou seja, para orientar o controle de qualidade interno nas empresas de semente.

Considerando que o potencial fisiológico das sementes de arroz pode ser afetado por fatores adversos em campo e no processamento, a avaliação criteriosa da qualidade das sementes torna-se ferramenta estratégica, sendo considerada essencial para assegurar a qualidade dos lotes que serão disponibilizados no mercado.

Diversos testes têm sido propostos para avaliar o vigor das sementes, porém a efetividade deles varia dependendo da espécie em estudo (KRZYZANOWSKI *et al.*, 1999; BAALBAKI *et al.*, 2009). Dentre estes, o teste de envelhecimento acelerado e o teste de frio têm demonstrado eficiência para sementes de arroz (BAEK, J. S. *et al.*, 2018). Contudo, para a obtenção dos resultados destes testes são necessários, no mínimo, 10 dias. Mesmo para o teste de germinação, o tempo mínimo para a obtenção dos resultados é de 14 dias (BRASIL, 2009).

Embora alguns testes de vigor produzam resultados confiáveis, sempre há espaço para a continuidade da pesquisa e aprimoramento de metodologias ou inclusão de alternativas inovadoras (MARCOS-FILHO, J. 2009). A agilidade na tomada de decisões quanto ao destino dos lotes em programas de controle de qualidade nas empresas de sementes tem sido uma necessidade cada vez mais frequente. Assim, tem aumentado a demanda por métodos que sejam de baixo custo, fácil execução e que permitam a obtenção de resultados confiáveis em curto espaço de tempo de maneira prática e eficiente sobre o potencial fisiológico das sementes (MEDEIROS *et al.*, 2019).

Neste contexto, destacam-se as análises automatizadas de imagens de sementes e de plântulas. A literatura tem documentado vários estudos envolvendo a análise de imagens em Tecnologia de Sementes, visando a estabelecer relação com o vigor de sementes (MARCOS-FILHO, J. 2009).

Os primeiros estudos utilizando análise computadorizada de plântulas foram realizados por McCormac *et al.* (1990), determinando o comprimento médio da raiz primária de tomate, alface e couve-flor. Outros estudos utilizando sistemas de análise de imagens correlacionaram crescimento de plântulas e emergência em couve-flor e tomate (GENEVE, R.; KESTER, R. 2001) e em rabanete (TOHIDLOO, G.; KRUSE, 2009).

A partir destes estudos diversos sistemas foram desenvolvidos para auxiliar na análise computadorizada de plântulas, destacando-se o Seed Vigor Imaging System (SVIS<sup>®</sup>) (SAKO *et al.*, 2001), o Sistema automatizado do vigor de sementes (Vigor-S<sup>®</sup>) (CASTAN *et al.*, 2018) e o Sistema de análise de plântulas (SAPL<sup>®</sup>) (MEDEIROS *et al.*, 2018). Utilizando a ferramenta matemática Matlab<sup>®</sup> para determinar o comprimento da parte aérea e da radícula de plântulas de trigo, Brunet *et al.* (2016) verificaram que o método foi adequado para separar as sementes de diferentes cultivares em níveis de vigor

fornecendo resultados similares aos obtidos por medição das plântulas descrito por Nakagawa *et al.* (1999).

O SVIS foi desenvolvido por SAKO *et al.* (2001) para avaliar o vigor de sementes de alfaca. Neste sistema, as plântulas são escaneadas, gerando índices de vigor, crescimento e uniformidade. Estudos conduzidos com diferentes espécies mostraram que este sistema automatizado foi eficiente para a avaliação do vigor de sementes de soja (HOFFMASTER *et al.*, 2003; MARCOS-FILHO *et al.*, 2009), milho (HOFFMASTER *et al.*, 2005), melão (MARCOS-FILHO *et al.*, 2006), amendoim (MARCHI *et al.*, 2011), milho doce (ALVARENGA *et al.*, 2012), trigo (SILVA, C. B. D *et al.*, 2012), pepino (CHIQUITO *et al.*, 2012), quiabo (KIKUTI, A.; MARCOS-FILHO, 2013), berinjela (SILVA, V. N.; CICERO, 2014) e cenoura (MARCHI, J. L.; CICERO, 2017). O procedimento para avaliação do vigor de sementes por meio do SVIS<sup>®</sup> é simples e os resultados da análise de imagens para uma amostra de 50 sementes podem ser obtidos dentro de dois a três minutos, com menor subjetividade em relação à análise realizada pelo analista, aumentando a confiabilidade dos dados para fins de comparação (GOMES-JUNIOR, 2014). Contudo, necessita de ajustes para a marcação das plântulas ou de suas partes após o processamento das imagens que requerem grande atenção do analista para que as partes marcadas sejam precisamente identificadas pelo sistema e os parâmetros gerados corretamente (CASTAN *et al.*, 2018).

Em virtude disto, um sistema automatizado de análise do vigor de sementes, denominado Vigor S, foi projetado em parceria entre USP/ESALQ e EMBRAPA Instrumentação, inicialmente para utilização em sementes de soja e milho (CASTAN *et al.*, 2018). Este sistema segue os mesmos princípios do SVIS<sup>®</sup>, mas com ajustes para aumentar a precisão das análises e reduzir os problemas detectados durante a execução do SVIS<sup>®</sup>. Os autores obtiveram resultados semelhantes aos do SVIS quanto aos índices obtidos, com vantagem de maior acurácia, velocidade, obtendo ainda informações compatíveis com as do teste de envelhecimento acelerado (MEDEIROS *et al.*, 2019).

O SAPL<sup>®</sup>, desenvolvido recentemente na Universidade Federal do Rio Grande do Norte, é um software que utiliza o processamento digital de imagens para a avaliação do vigor de sementes a partir do desempenho de plântulas. Este sistema permite obter informações sobre o comprimento total de plântulas, comprimento radicular e do hipocótilo, além de obter índices de vigor, de crescimento e de uniformidade de plântula (MEDEIROS *et al.*, 2018).

O índice de vigor é baseado na velocidade e uniformidade de desenvolvimento das plântulas em relação ao máximo valor possível para plântulas de arroz com 5 dias. A uniformidade é estabelecida em função do desvio em relação ao desenvolvimento padrão de plântulas com 5 dias de idade, também programado no software. Assim, após o processamento das imagens, o software gera automaticamente valores numéricos referentes ao vigor, crescimento e à uniformidade das plântulas, Medeiros e Pereira (2018) confirmaram a eficiência do SAPL<sup>®</sup> na avaliação do potencial fisiológico em sementes de soja. A análise automatizada de plântulas utilizando o SAPL<sup>®</sup> também foi eficaz na avaliação do vigor de sementes de moringa (PEREIRA *et al.*, 2020) e milho (MEDEIROS *et al.*, 2019). Em sementes de grão-de-bico, o comprimento total, o comprimento radicular e o índice de vigor corrigido determinados aos quatro dias de germinação pelo SAPL permitiu a separação dos lotes em níveis de vigor, o que só foi possível pelo comprimento de parte aérea aos oito dias do início do teste de germinação (ARAUJO *et al.*, 2020).

Outra ferramenta interessante é a aprendizagem de máquina, que é um subcampo da inteligência artificial dedicado ao desenvolvimento de algoritmos e técnicas que permitam ao computador aprender, isto é, que permitam ao computador aperfeiçoar seu desempenho em alguma tarefa (KAN, 2017). Algumas partes da aprendizagem de máquina estão intimamente ligadas à mineração de dados e estatística. Sua aplicação prática inclui o processamento de linguagem natural, motores de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional e locomoção de robôs e sistemas de previsão.

Uma abordagem promissora de aprendizado de máquina orientada ao usuário é o Interactive Machine Learning (IML), as abordagens IML são definidas como modelos interativos de aprendizagem automática que permitem ao usuário classificar e analisar imagens (BERG, 2019).

Neste contexto, o ILASTIK é uma ferramenta para classificação que utiliza a abordagem Interactive Machine Learning (IML) para segmentação e análise de imagens interativas. Ele é construído como uma estrutura de software modular, que atualmente possui fluxos de trabalho para classificação no nível de pixel e objeto, rastreamento de objetos automatizado e semi-automatizado, segmentação semi-automatizada e contagem de objetos sem detecção. A ferramenta pode ter diferentes aplicações na área de sementes, dentre elas buscar estabelecer classificação de plântulas obtidas no teste de germinação

em categorias de vigor, como por exemplo, em plântulas fortes e fracas. Estas informações poderiam ser promissoras para a avaliação do vigor das sementes, caso forneçam resultados compatíveis com outros testes de vigor quanto à classificação dos lotes em níveis de vigor.

Em síntese, o uso de tecnologias computacionais que permitam obter informações sobre o potencial fisiológico das sementes tem se mostrado uma ferramenta adicional de rápida resposta e de interpretação não subjetiva, que aumenta a capacidade de decisão sobre as condições fisiológicas das sementes. Para que tais tecnologias sejam eficientes para a avaliação do vigor de sementes é essencial a adequação de metodologias para cada espécie.

Não existem na literatura estudos que validem a análise de imagens com o Sistema de Análise de Plântulas (SAPL<sup>®</sup>) e a classificação do vigor das plântulas por meio da aprendizagem de máquina utilizando o software ILASTIK para a determinação do potencial fisiológico de sementes de arroz.

Diante do exposto, objetivou-se verificar a eficácia da análise computadorizada de imagens por meio do SAPL<sup>®</sup> e da classificação do vigor das plântulas utilizando o software ILASTIK para a avaliação do vigor de sementes de arroz.

## **MATERIAL E MÉTODOS**

A pesquisa foi conduzida no Departamento de Agronomia da Universidade Federal de Viçosa, no Laboratório de Pesquisa de Sementes, Viçosa, Minas Gerais. Foram utilizados seis lotes de sementes de arroz das cultivares Relâmpago e Caçula, fornecidos pela Empresa de Pesquisa Agropecuária de Minas Gerais – EPAMIG e quatro lotes da cultivar Fedearroz 2000, fornecido pela empresa Colombiana FEDEARROZ.

As sementes de cada lote foram submetidas aos seguintes testes para caracterizar a sua qualidade inicial.

### **Caracterização da qualidade inicial dos lotes**

#### **Determinação do grau de umidade**

Efetuada antes da instalação dos testes, pelo método da estufa a  $105\pm 3$  °C durante 24 horas (BRASIL, 2009), com duas repetições de 50 sementes para cada lote. Os resultados foram expressos em porcentagem de umidade (b.u.).

### **Teste de germinação**

Quatro repetições de 50 sementes foram distribuídas sobre papel toalha umedecido com quantidade de água equivalente a 2,5 vezes o peso do papel seco. Foram confeccionados rolos mantidos em germinador a 25 °C e as avaliações efetuadas no quinto e no decimo quarto dia (germinação) após a instalação do teste, de acordo com os critérios estabelecidos pelas Regras para Análises de Sementes (BRASIL, 2009). Os resultados foram expressos em porcentagem média de plântulas normais.

### **Primeira contagem de germinação**

Foi conduzido conjuntamente com o teste de germinação, consistindo no registro da porcentagem de plântulas normais obtidas no quinto dia após a semeadura (BRASIL, 2009).

### **Envelhecimento acelerado**

Conduzido com quatro repetições de 50 sementes para cada tratamento, que foram colocadas em caixas *gerbox* sobre uma tela, contendo 40 ml de água e mantidas em incubadora BOD a 42 °C durante 120 horas (MARCOS-FILHO, J. 1999). Após esse período, as sementes foram colocadas para germinar da mesma maneira descrita acima para o teste de germinação e a avaliação realizada no quinto dia após a semeadura. Os resultados foram expressos em porcentagem de plântulas normais.

### **Teste de frio sem solo**

Foi conduzido com quatro repetições de 50 sementes, em rolos de papel umedecido com uma quantidade de água destilada equivalente a 2,5 vezes o peso do papel. Após a semeadura, os rolos foram colocados no interior de sacos plásticos, vedados com fita adesiva e mantidos em B.O.D a 10 °C durante sete dias. Após este período, os rolos foram retirados dos sacos plásticos e transferidos para um germinador à temperatura de 25 °C, por mais cinco dias para então se proceder a avaliação da porcentagem de plântulas normais (BARROS *et al.*, 1999).

### **Velocidade de emissão da raiz primária**

Este teste foi conduzido com quatro repetições de 50 sementes de cada lote. As sementes foram colocadas para germinar da mesma maneira descrita para o teste de germinação. A cada 12 horas o material foi examinado de tal maneira que as sementes cujo desenvolvimento da raiz primária foi visível, foram contadas e removidas (TOLEDO

*et al.*, 1999). Após concluída esta etapa, foi calculado o índice de velocidade de emissão de raiz primária, adotando-se o critério sugerido por Maguire (1962).

### **Índice de velocidade de emergência (IVE)**

O teste foi conduzido com quatro repetições de 50 sementes, em bandejas de poliestireno expandido de múltiplas células, sendo semeada uma semente por célula preenchida com areia lavada, irrigadas diariamente com mesmo volume por célula e mantidas na sala de crescimento. Foram efetuadas contagens diárias de plântulas normais emergidas até a estabilização, avaliando a porcentagem de emergência, a velocidade de emergência (VE) e índice de velocidade de emergência (IVE) (NAKAGAWA, 1999).

### **Análise computadorizada de plântulas**

Para a aquisição das imagens, inicialmente foi realizado o teste de comprimento de plântulas com quatro repetições de 20 sementes por lote, sobre linha traçada no terço superior do papel toalha, no sentido longitudinal, seguindo a metodologia proposta por Nakagawa *et al.* (1999). O teste foi conduzido em germinador a 25 °C. Foram avaliadas as plântulas com cinco e seis dias, as quais foram transferidas do papel de germinação para uma folha de espuma vinílica acetinada (E.V.A) de coloração azul (40 x 60 cm), contendo dez células de cinco centímetros de largura, divididas por faixas brancas. A aquisição das imagens foi feita com uma câmera digital Nikon, modelo Coolpix P510, de 16 Megapixels, com 1/25 segundos de velocidade de disparo do obturador e F3.0 de abertura do diafragma. A câmera foi mantida a 60 cm das amostras e angulação de 90° em relação a base fotográfica, utilizando suporte do tipo *Copystand*.

### **Software SAPL®**

As imagens foram armazenadas e inseridas no arquivo *default* selecionado do software SAPL® instalado em notebook DELL Latitude 3470, Windows 7 Professional. O software forneceu medições do comprimento da parte aérea (CPA), raiz (CR) e plântula inteira (CT). Estes valores foram inseridos no pacote SeedCalc do software R (TEAM, 2019), para gerar índices de uniformidade (CASTAN; GOMES-JUNIOR; MARCOS-FILHO, 2018), crescimento (MEDEIROS *et al.*, 2019) e vigor corrigido (MEDEIROS; PEREIRA, 2018), conforme as fórmulas descritas abaixo:

$$IC = \left( \frac{\text{médiaCT}}{X_{\text{máx}}} \right) \times 1000$$

Em que:  $IC$ , o índice de crescimento;  $médiaCT$ , a média do comprimento total de plântulas da repetição;  $Xmáx$ , a média do comprimento total de 5% das maiores plântulas do experimento.

$$Uniformity\_2 = \left[ 1 - \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n \times X} \right] \times 1000 - \left[ n_{dead} \times \left( \frac{50}{n_{total}} \right) \right]$$

Em que:  $Uniformity\_2$  é o índice de uniformidade;  $X_i$ , o comprimento da plântula analisada;  $X$ , o comprimento médio de plântulas do lote de sementes analisado;  $n$ , a variável número de plântulas totais avaliadas;  $n_{dead}$ , o número de sementes não germinadas ou mortas presentes no lote avaliado; e  $n_{total}$ , o número total de plântulas.

$$Vigor\_Corr = \left[ (Growth \times wg) + (Uniformity \times wu) \right] \times \left( \frac{G}{100} \right)$$

Em que:  $Vigor\_Corr$  é o índice de vigor corrigido;  $Growth$  é o índice de crescimento e  $Uniformity$  é o índice de uniformidade escolhido pelo usuário;  $wg$  e  $wu$  são pesos ajustáveis na fórmula para o crescimento e uniformidade, porém com valores referência de 70 e 30, respectivamente;  $G$  é o percentual de germinação do lote.

### Software ILASTIK

As imagens obtidas foram carregadas no software e no pré-processamento e a segmentação foi utilizada a ferramenta de classificação de pixels, definindo duas classes de segmentação “semente ou plântula” e “fundo”, baseadas na intensidade de cores e pixels. Para criar os mapas de probabilidade, os pixels pertencentes às regiões de cada classe foram selecionados pintando pinceladas de diferentes cores diretamente nos dados de entrada. A probabilidade de o pixel pertencer às classes de segmentação foi estimada para cada pixel da imagem. O software foi treinado para o reconhecimento de duas classes de plântulas, as vigorosas (V) plântulas morfológicamente saudáveis mostrando crescimento vigoroso de hipocótilo e raiz com comprimento maior ou igual a seis centímetros; e Plântulas fracas (F), as quais mostraram ausência, subdesenvolvimento ou deformação de alguma estrutura essencial e comprimento de raiz menor a 6 centímetros (Figura 1). O classificador treinado foi aplicado a todas as imagens, e os mapas de probabilidade foram exportados para cada imagem, gerando dados de número de plântulas vigorosas (V) e número de plântulas fracas (F).

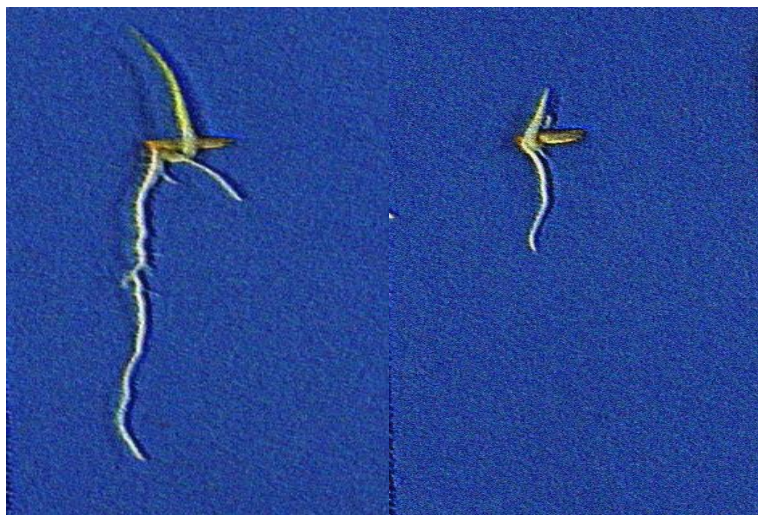


Figura 1 – Padrão de plântula forte e fraca utilizado para o treinamento do software ILASTIK.

### **Delineamento experimental e análise estatística**

O experimento foi conduzido em delineamento inteiramente casualizado, com quatro repetições. Os dados foram testados quanto à distribuição normal dos erros pelo teste de Shapiro-Wilk e à homogeneidade de variâncias pelo teste de Bartlett. Os dados referentes aos testes para a caracterização inicial da qualidade fisiológica dos lotes e análise computadorizada de plântulas foram submetidos à análise de variância realizada separadamente para cada cultivar e as médias obtidas para cada lote foram comparadas pelo teste de Tukey ( $p \leq 0,05$ ). Os dados obtidos em cada teste foram submetidos à análise de correlação de Pearson ( $r$ ,  $p \leq 0,05$ ). Todas as análises estatísticas foram processadas utilizando-se o software R (TEAM, 2019).

## **RESULTADOS E DISCUSSÃO**

### **Caracterização da qualidade fisiológica dos lotes**

Os resultados de grau de umidade das sementes variaram de 11,5 a 13,9% entre os lotes da cultivar F-2000 (Tabela 1), 13,2 a 15,9% no caso da cultivar Relâmpago (Tabela 2), e 13,9 a 16,0% para a cultivar Caçula (Tabela 3). Estas variações são aceitáveis, o que segundo Marcos-Filho (2015), não compromete a qualidade dos resultados e o adequado estabelecimento de comparações de qualidade fisiológica entre os lotes de cada cultivar.

Não houve diferença significativa na porcentagem de germinação e primeira contagem de germinação das sementes dos lotes da cultivar F-2000. Todos os lotes apresentaram valores de germinação superiores a 80% (Tabela 1), sendo este o valor mínimo estabelecido para comercialização de sementes de arroz no Brasil (BRASIL, 2009). Também não houve diferença entre os lotes quando ao índice de velocidade de emergência (IVE).

Foi possível observar, para esta cultivar, diferenças significativas de desempenho entre os lotes por meio do teste de envelhecimento acelerado, no qual os lotes 1 e 2 mostraram-se mais vigorosos que o lote 4 (Tabela 1). Comportamento similar foi obtido no teste de frio, destacando-se o lote 1 como o mais vigoroso, os lotes 2 e 3 em posição intermediária e o lote 4 menos vigoroso. Já pelo índice de velocidade de emissão da radícula (IVR), o lote 1 foi o superior aos demais, que não diferiram entre si.

Tabela 1- Caracterização da qualidade fisiológica de quatro lotes de sementes de arroz cv-F-2000 avaliados pelo grau de umidade (GU), primeira contagem de germinação (PCG), germinação (G), envelhecimento acelerado (EA), teste de frio (TF), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primária (IVR).

Lotes	GU (%)	PCG (%)	G (%)	EA (%)	TF (%)	IVE (índice)	IVR (índice)
1	11,5	68,0	94,5	76,0 a	77,5 a	40,5	1,01 a
2	13,0	54,0	95,5	64,5ab	64,5ab	42,4	0,95b
3	13,9	58,0	98,5	54,0bc	67,0ab	38,3	0,91b
4	13,8	68,0	96,0	46,0c	51,5b	40,2	0,90b
F <sub>C</sub>	-	4,25	2,43	19,04*	4,46*	0,86	13,42*
CV(%)	-	11,3	2,27	9,92	15,54	8,79	2,93

\* Significativo pelo teste F a 5% de probabilidade; F<sub>C</sub> - F calculado; CV - Coeficiente de variação. Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

Em geral, pelos testes utilizados para a caracterização da qualidade dos lotes (Tabela 1) foi constatado o lote 1 como o de melhor desempenho, lotes 2 e 3 como de médio vigor, e o lote 4 com pior desempenho.

Na cultivar Relâmpago, não houve diferença significativa entre os lotes quanto à germinação, IVE e envelhecimento acelerado. Nota-se que todos os lotes apresentaram alta germinação, acima de 90%. Todos os lotes apresentaram desempenho semelhante após terem sido submetidos às condições de estresse por calor e umidade no teste de

envelhecimento acelerado, o que indica que têm potencial de armazenamento semelhantes, já que esta é uma das finalidades deste teste (DANTAS *et al.*, 2019). Contudo, ao serem submetidos às condições de baixa temperatura proporcionadas pelo teste de frio, apresentaram diferenças significativas no seu desempenho, com maior vigor para as sementes do lote 6 em relação aos demais. Os lotes 1 e 4 ocuparam posição intermediária sendo superiores ao lote 5. Este lote também apresentou pior desempenho nos testes de primeira contagem de germinação e IVR, indicando menor velocidade de germinação. O arroz tem alta sensibilidade a baixas temperaturas, sementes com maior qualidade fisiológica podem apresentar desempenho nessa condição (GRZYBOWSKI *et al.*, 2015).

Tabela 2 - Caracterização da qualidade fisiológica de quatro lotes de sementes de arroz cv-Relâmpago avaliados pelo grau de umidade (GU), primeira contagem de germinação (PCG), germinação (G), envelhecimento acelerado (EA), teste de frio (TF), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primária (IVR).

Lotes	GU (%)	PCG (%)	G (%)	EA (%)	TF (%)	IVE (índice)	IVR (índice)
1	13,2	74,0 a	93,0	55,0	55,5b	45,6	1,09ab
2	15,9	74,5 a	91,5	51,5	55,0bc	43,7	1,12ab
3	15,0	75,5 a	92,0	57,5	57,5bc	42,7	1,17 a
4	14,0	63,0 a	90,5	44,7	63,0b	40,4	1,16 a
5	15,8	30,5b	93,5	49,2	46,7c	35,7	1,06b
6	15,9	62,5 a	94,0	49,0	88,5 a	41,1	1,15 a
Fc	-	30,67*	0,37	1,70	17,28*	1,93	5,5*
CV(%)	-	9,76	4,65	15,6	11,37	11,83	3,42

\* Significativo pelo teste F a 5% de probabilidade; Fc - F calculado; CV - Coeficiente de variação. Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

De modo geral, para a cultivar Relâmpago, pela caracterização da qualidade inicial dos lotes, observou-se que o lote 5 apresentou menor vigor, comparado aos demais, os quais apresentaram vigor semelhante, com destaque do lote 6 no teste de frio (Tabela 2).

Na caracterização da qualidade fisiológica dos lotes da cultivar Caçula, foi possível observar diferenças significativas entre eles pelos testes de vigor tradicionais, independentemente de não apresentarem diferenças na germinação (Tabela 3), que é uma

das últimas alterações decorrentes do processo de deterioração das sementes que ocorre com o decorrer do tempo após a maturidade fisiológica das sementes (EBONE; CAVERZAN; CHAVARRIA, 2019).

De modo geral, os resultados obtidos no teste de primeira contagem, envelhecimento acelerado e índice de velocidade de emergência foram semelhantes em termos de classificação de lotes em níveis de vigor para este cultivar (Tabela 3). Nestes testes, os lotes 1, 2, 3, 4 foram superiores, seguidos pelo lote 6, sendo o lote 5 o menos vigoroso. Já pelo teste de frio, os lotes 1 e 3 foram superiores ao lote 6, embora estes lotes não tenham diferido significativamente dos lotes 2, 4 e 5. Wrasse *et al.* (2009) relataram que o uso do teste de frio foi eficaz para a classificação do potencial fisiológico de lotes de sementes de arroz.

Tabela 3- Caracterização da qualidade fisiológica de quatro lotes de sementes de arroz cv- Caçula avaliados pelo grau de umidade (GU), primeira contagem de germinação (PCG), germinação (G), envelhecimento acelerado (EA), teste de frio (TF), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primária (IVR).

Lotes	GU (%)	PCG (%)	G (%)	EA (%)	TF (%)	IVE (índice)	IVR (índice)
1	13,9	70,0 a	89,0	32,5 a	55,5 a	40,0 a	0,90ab
2	13,8	59,5ab	86,0	28,7ab	50,5ab	36,6 a	0,88abc
3	15,7	56,0ab	88,0	27,0ab	55,5 a	37,5 a	0,92 a
4	15,4	57,5ab	84,0	23,2b	51,5ab	39,21 a	0,94 a
5	15,5	12,5c	85,0	10,7c	45,0ab	17,25c	0,81bc
6	16,0	50,0b	87,0	12,5c	38,5b	27,2b	0,79c
Fc	-	23,16*	0,24	25,32*	4,65*	27,58*	6,92*
CV(%)	-	16,25	7,84	15,7	12,39	10,24	5,4

\* Significativo pelo teste F a 5% de probabilidade; Fc - F calculado; CV - Coeficiente de variação. Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

Resultados semelhantes também foram encontrados ao se avaliar o índice de velocidade de emissão da radícula, onde os lotes 5 e 6 tiveram desempenho inferior aos demais, destacando-se como superiores os lotes 3 e 4. A rapidez na germinação é um importante componente do vigor de sementes com reflexos diretos na velocidade de emergência em campo. Lotes de sementes com germinação semelhante podem não atingir o valor máximo de germinação ao mesmo tempo, indicando diferenças na velocidade de

germinação. Assim, a rapidez na germinação é um indicativo de elevada atividade metabólica resultando em embebição mais rápida e alta taxa de crescimento do embrião e, conseqüentemente, da plântula (WRASSE *et al.*, 2009).

Em síntese, na caracterização da qualidade fisiológica dos lotes da cultivar Caçula, as sementes dos lotes 5 e 6 foram as que apresentaram menor vigor e o lote 1 maior vigor.

### **Caracterização dos lotes por processamento de imagens no software SAPL<sup>®</sup> e classificação de imagem com aprendizagem de máquina no software ILASTIK.**

A análise de imagens de plântulas da cultivar F-2000 pelo software SAPL<sup>®</sup> possibilitou a identificação de diferenças no potencial fisiológico por meio de alguns parâmetros avaliados aos quatro dias após o início do teste de germinação. Em geral, em todos os parâmetros de crescimento, o lote 1 apresentou melhor desempenho e os lotes 2, 3 e 4 não diferiram entre si (Tabela 4). Já para o índice de uniformidade não houve diferenças significativas, não permitindo a classificação dos lotes em níveis de qualidade fisiológica.

Com a classificação de imagens utilizando o software ILASTIK, o qual permitiu obter resultados referentes ao número de plântulas vigorosas e fracas de cada lote, foi possível a separação dos lotes quanto ao vigor aos quatro e cinco dias após o início do teste de germinação, sendo o lote 1 o que apresentou o maior número de plântulas vigorosas, e o lote 4 o menor. Os lotes 2 e 3 ficaram em posição intermediária e não diferiram entre si. No parâmetro plântulas fracas, maior valor foi obtido para o lote 4 e para o lote 1 o menor valor. Assim, observou-se que lotes com maior proporção de plântulas fortes tiveram menor o valor para plântulas fracas, havendo correspondência entre estas duas características. Segundo Medeiros *et al.* (2020), métodos de aprendizagem de máquina interativos, como o software ILASTIK, tiveram alta acurácia para classificar sementes de soja quanto à presença de danos e também classificar as plântulas quanto ao nível de vigor das plântulas.

Com o processamento e classificação das imagens, foi possível corroborar os resultados obtidos na caracterização inicial a partir dos testes tradicionais de vigor (Tabela 1).

Considerando que o teste de germinação com sementes de arroz dura 14 dias, a obtenção de resultados consistentes que permitam separar lotes em níveis de qualidade

fisiológica em quatro dias é de grande relevância para agilizar a tomada de decisões nas empresas quanto ao manejo racional de lotes para a comercialização. Além disto, pelo sistema automatizado de análise de plântulas o tempo gasto para as avaliações é consideravelmente reduzido. Diversos autores têm relatado redução no tempo das avaliações de lotes de sementes com o uso de sistemas computacionais como milho e soja (MEDEIROS *et al.*, 2019), feijão (GOMES *et al.*, 2014), trigo (BRUNES *et al.*, 2016), entre outros.

Tabela 4 - Comprimento de parte aérea (CPA), comprimento de raiz (CR), comprimento total (CT), índice de uniformidade (IU), índice de crescimento (IC) e índice de vigor corrigido (IVC) gerados pelo SAPL<sup>®</sup> e de número de plantas vigorosas (V), número de plântulas fracas (F) obtidos pelo programa de aprendizagem de máquina ILASTIK, aos quatro e cinco dias após a semeadura de quatro lotes de sementes de arroz cv-F2000.

Lote	CPA (mm. plântula <sup>-1</sup> )	CR	CT	IU	IC	IVC	Plântulas vigorosas V	Plântulas fracas F
				Índices				
4 dias								
1	8,2 a	41,6 a	49,8 a	617	816 a	714 a	15,7 a	4,5 b
2	3,8 ab	17,5 b	21,3 b	664	349 b	423 b	12,2 b	7,2 b
3	3,1 b	11,3 b	14,1 b	580	237 b	334 b	13,0 b	6,7 b
4	3,1 b	11,7 b	14,8 b	523	242 b	313 b	8,0 c	12,0 a
Fc	5,28*	13,86*	12,09*	1,83	12,1*	11,92*	14,7*	10,8
CV (%)	12,4	15,45	11,43	14,45	11,5	11,76	12,3	20,1
5 dias								
1	13,2	51,5	64,8	600	751	667	16,7 a	3,2 b
2	13,7	48,3	62,0	632	719	662	14,5 ab	5,5 ab
3	13,3	44,1	57,5	659	667	654	13,5 ab	6,2 ab
4	13,5	40,6	54,1	538	628	577	11,5 b	8,5 a
Fc	0,11	2,18	1,79	1,92	1,08	1,36	11,4	14,3
CV (%)	8,9	13,34	11,8	9,5	9,75	9,25	12,9	33,1

\* Significativo pelo teste F a 5% de probabilidade; Fc – F calculado; CV = Coeficiente de variação. Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

Para a cultivar Relâmpago, foram obtidas diferenças altamente significativas entre os lotes aos quatro e cinco dias após o início do teste de germinação, sendo que no quarto dia foi possível identificar, pelos parâmetros comprimento de parte aérea (CPA), comprimento de raiz (CR), comprimento total (CT) e índice de uniformidade (IU) que apenas os lotes 5 e 6 apresentaram comprimentos e índice maiores que os demais lotes, ou seja, obtendo-se dois grupos (Tabela 5).

No quarto dia, em geral, os lotes 5 e 6 não diferiram entre si e foram superiores aos demais lotes ao se avaliar o CPA, o CR, o CT e o IU. Já pelo IC e IVC, apenas o lote 6 foi superior aos demais. Ao se comparar estes resultados com os testes utilizados para a caracterização do potencial fisiológico dos lotes (Tabela 2), observa-se que o lote 5 foi inferior ao lote 6 em todos os testes que indicaram diferenças significativas entre os lotes, ou seja, nos testes de PC, frio e IVE. Desta forma, pode-se afirmar que aos quatro dias após a semeadura, os resultados obtidos com o SAPL não foram eficientes para a classificação dos lotes quanto ao vigor, pois apenas pelo IC e IVC foi possível identificar o lote 6 como superior aos demais. Já pela classificação de imagens com o programa ILASTIK, houve diferenças significativas entre os lotes quanto ao número de plântulas vigorosas e fracas, onde o lote 6 foi superior aos lotes 1, 2, 4 e 5. Estes resultados foram coerentes ao se comparar com os demais testes utilizados para a caracterização do potencial fisiológico dos lotes dessa cultivar (Tabela 2).

Aos cinco dias após início do teste de germinação, houve a separação dos lotes em níveis de vigor, exceto para a variável IU, que não detectou diferença significativa entre os lotes (Tabela 5). Em todas as demais variáveis analisadas, as sementes do lote 6 apresentaram desempenho superior aos demais lotes, ocorrendo variações quanto à indicação do lote de pior desempenho. Pelo CT, o lote 6 foi superior aos demais lotes, que não diferiram entre si. Os lotes foram classificados, portanto, em apenas dois níveis de qualidade fisiológica. Já pelo CPA, os lotes 1, 2 e 3 apresentaram os menores valores, não diferindo entre si, sendo inferiores ao lote 4, que por sua vez, foi inferior ao lote 5. Já pelos parâmetros CR, IC e IVC o lote 5 teve o pior desempenho, enquanto os lotes 1, 2, 3 e 4 ocuparam posição intermediária com maior vigor para o lote 6. Nas variáveis número de plantas vigorosas e fracas, os lotes 6 e 3 foram superiores, seguidos pelos lotes 2 e 4, e com o pior desempenho os lotes 1 e 5. Assim, ao se comparar estes resultados com os da Tabela 2, verifica-se que houve coerência entre as informações obtidas com as variáveis CR, IC, IVC, V e F, e a classificação dos lotes obtida nos testes de vigor, principalmente quanto ao maior vigor das sementes do lote 6 e o menor vigor das sementes do lote 5, ocorrendo pequenas variações com relação aos demais lotes que podem ser classificados como de vigor intermediário. Medeiros e Pereira (2018), relataram que tanto o índice de vigor corrigido como o de crescimento possibilitaram distinção dos lotes de soja pela sua qualidade fisiológica.

Nota-se que também nestas avaliações de processamento de imagens feitas com o SAPL®, conforme comentado para os resultados obtidos no quarto dia (Tabela 5), os

lotes 5 e 6 foram identificados como apresentando desempenho superior aos demais, o que não ocorreu nos demais testes de vigor empregados (Tabela 2), em que o lote 5 foi classificado junto aos de pior desempenho.

Tabela 5- Comprimento de parte aérea (CPA), comprimento de raiz (CR), comprimento total (CT), índice de uniformidade (IU), índice de crescimento (IC) e índice de vigor corrigido (IVC) gerados pelo SAPL<sup>®</sup> e de número de plantas vigorosas (V), número de plântulas fracas (F) obtidos pelo programa de aprendizagem de máquina ILASTIK, aos quatro e cinco dias após a semeadura de quatro lotes de sementes de arroz cv-Relâmpago.

Lote	CPA	CR	CT	IU	IC	IVC	Plântulas vigorosas	Plântulas fracas
	(mm. plântula <sup>-1</sup> )			Índices			V	F
4 dias								
1	3,4 b	19,6 b	23,0 b	488 b	378 b	382 b	14,0 bc	5,2 a
2	3,7 b	21,2 b	24,9 b	650 ab	409 b	440 b	14,7 bc	4,2 ab
3	3,9 b	19,9 b	23,9 b	627 ab	391 b	425 b	17,2 ab	1,7 bc
4	6,4 b	27,9 ab	34,3 ab	492 b	562 b	490 b	16,0 bc	3,0 abc
5	15,3 a	38,5 ab	52,3 a	792 a	299 b	418 b	13,7 bc	4,7 ab
6	14,7 a	43,4 a	58,1 a	830 a	952 a	861 a	19,0 a	1,0 c
Fc	36,08*	5,43*	8,25*	6,11*	8,33*	11,13*	9,45*	10,7*
CV (%)	11,28	13,06	13,70	14,87	12,54	11,31	7,65	38,0
5 dias								
1	10,7 d	52,3 b	63,0 b	610	731 b	646 b	14,5 c	4,7 a
2	10,9 d	51,9 b	62,9 b	667	730 b	650 b	15,7 bc	3,2 abc
3	10,8 d	48,5 bc	59,3 b	660	689 b	626 b	18,0 ab	0,5 d
4	13,9 c	51,6 b	65,6 b	702	761 b	673 b	17,0 abc	2,0 bcd
5	18,5 b	41,5 c	58,4 b	605	506 c	501 c	15,2 c	3,5 ab
6	23,5 a	64,4 a	88,0 a	806	1021 a	899 a	19,0 a	0,7 cd
Fc	118,48*	14,31*	20,03*	1,61	49,68*	26,77*	18,9*	19,6
CV (%)	6,54	7,61	7,42	17,22	6,35	7,51	7,32	47,21

\* Significativo pelo teste F a 5% de probabilidade; Fc – F calculado; CV = Coeficiente de variação. Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

Portanto, os resultados obtidos com o software ILASTIK em plântulas de arroz da cultivar Relâmpago aos 4 e 5 dias após o início do teste de germinação permitiram a estratificação dos lotes em níveis de vigor fornecendo informações coerentes à classificação obtida nos testes de vigor tradicionais utilizados. Já com o programa SAPL<sup>®</sup>, apenas nas avaliações realizadas aos cinco dias foram obtidas informações coerentes com os demais testes de vigor utilizados (Tabela 2).

Nos resultados obtidos com os softwares SAPL<sup>®</sup> para as sementes da cultivar Caçula no quarto dia de germinação (Tabela 6) houve diferença significativa entre os lotes apenas para as variáveis CPA e IU. Para o comprimento de parte aérea, maiores valores foram obtidos para os lotes 1 e 3 (11,2 e 13,3 mm plântula<sup>-1</sup>), que não diferiram entre si, e menor valor para o lote 6 (5,5 mm plântula<sup>-1</sup>), que não diferiu dos lotes 4 e 5. Pelo índice de uniformidade, os lotes foram classificados em dois grupos, lotes 1, 2 e 3 como os de maior vigor e os demais (4, 5 e 6) como os de pior desempenho. Com a classificação das imagens das plântulas obtidas para cada lote pelo software ILASTIK, verifica-se, de modo geral, que o número de plântulas vigorosas dos lotes 1 e 3 não diferiu entre si e foi superior ao observado para os demais lotes, indicando maior vigor. O menor valor para plântulas vigorosas foi observado no lote 6, que não diferiu dos lotes 2 e 5, o que permite indicar estes lotes como apresentando vigor inferior ao dos lotes 1 e 3. Quando se observa o número de plântulas fracas é possível classificar os lotes em dois grupos: lotes 2, 5 e 6 como menos vigorosos (com maior número de plântulas fracas) e lotes 1, 3 e 4 apresentando melhor desempenho.

Nas avaliações realizadas no 5º dia após o início do teste de germinação (Tabela 6), foram observadas diferenças significativas apenas no comprimento da parte aérea, com resultados similares aos 4 dias, e na classificação de imagens pelo número de plantas vigorosas e fortes. Em geral, os lotes 5 e 6 apresentaram menor vigor e os demais (1, 2, 3, 4) melhor desempenho, sendo significativamente iguais entre si.

Os dados obtidos na caracterização e classificação dos lotes da cultivar Caçula feita pelos softwares SAPL<sup>®</sup> e ILASTIK corroboram os resultados obtidos nos testes tradicionais de vigor (Tabela 3), nos quais o lote 1 foi classificado como superior e os lotes 5 e 6, que foram os menos vigorosos.

Embora diferenças visuais no vigor das plântulas, muitas vezes, possam ser óbvias, pode haver dificuldade para se obter reprodutibilidade na classificação das categorias de plântulas fortes e fracas. Plântulas consideradas morfologicamente normais diferem quanto à taxa de crescimento, robustez e tamanho (MEDEIROS E PEREIRA, 2018). Assim, os testes convencionais mesmo sendo realizados por analistas treinados, podem estar sujeitos a erros de interpretação além de consumirem tempo nesta tarefa de avaliação (KUO *et al.*, 2016). Neste contexto, as técnicas de análise de imagens são importantes ferramentas por terem maior eficácia no procedimento de avaliação. Estudos recentes utilizando o SAS<sup>®</sup> (PINTO *et al.*, 2015), Vigor-S<sup>®</sup> (CASTAN; GOMES-JUNIOR; MARCOSFILHO, 2018) e SAPL (MEDEIROS E PEREIRA, 2018);

MEDEIROS *et al.*, 2018; MEDEIROS *et al.*, 2019) têm demonstrado a eficiência destes sistemas na avaliação do vigor de sementes de diferentes espécies via análise automatizada de plântulas.

Tabela 6 - Comprimento de parte aérea (CPA), comprimento de raiz (CR), comprimento total (CT), índice de uniformidade (IU), índice de crescimento (IC) e índice de vigor corrigido (IVC) gerados pelo SAPL® e de número de plantas vigorosas (V), número de plântulas fracas (F) obtidos pelo programa de aprendizagem de máquina ILASTIK, aos quatro e cinco dias após a semeadura de quatro lotes de sementes de arroz cv-Caçula.

Lote	CPA (mm. plântula <sup>-1</sup> )	CR	CT	IU	IC	IVC	Plântulas vigorosas V	Plântulas fracas F
Índices								
4 dias								
1	11,2 a	30,2	41,4	794 a	679	635	15,7 a	4,0 b
2	8,3 b	39,1	47,5	667 a	779	641	10,5 bc	8,2 a
3	13,3 a	29,8	43,2	735 a	708	630	15,7 a	2,7, b
4	7,1 bc	28,0	35,1	321 b	576	446	13,5 ab	4,2 b
5	8,1 bc	40,7	48,8	358 b	800	636	9,7 bc	7,7, a
6	5,5 c	29,7	35,3	302 b	579	459	8,7 c	9,0 a
Fc	22,3*	1,4	1,35	23,68*	1,43	1,88	15,3*	13,4*
CV (%)	13,46	13,76	9,70	11,87	10,19	9,56	14,6	23,9
5 dias								
1	18,0 a	50,4	68,4	787	794	705	18,7 a	1,0 b
2	11,3 bc	53,1	64,4	677	748	625	15,5 a	3,5 b
3	19,1 a	47,9	67,0	707	777	665	16,2 a	1,5 b
4	12,1 b	49,1	61,2	533	710	555	15,5 a	2,5 b
5	9,6 c	55,2	66,7	663	774	633	11,0 b	6,5 a
6	11,4 bc	44,8	54,3	574	630	533	10,5 b	7,5 a
Fc	47,14*	0,65	1,07	2,33	1,07	3,89	10,21*	21,5*
CV (%)	8,48	14,05	15,89	14,33	15,89	13,74	12,5	35,0

\* Significativo pelo teste F a 5% de probabilidade; Fc – F calculado; CV = Coeficiente de variação. Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

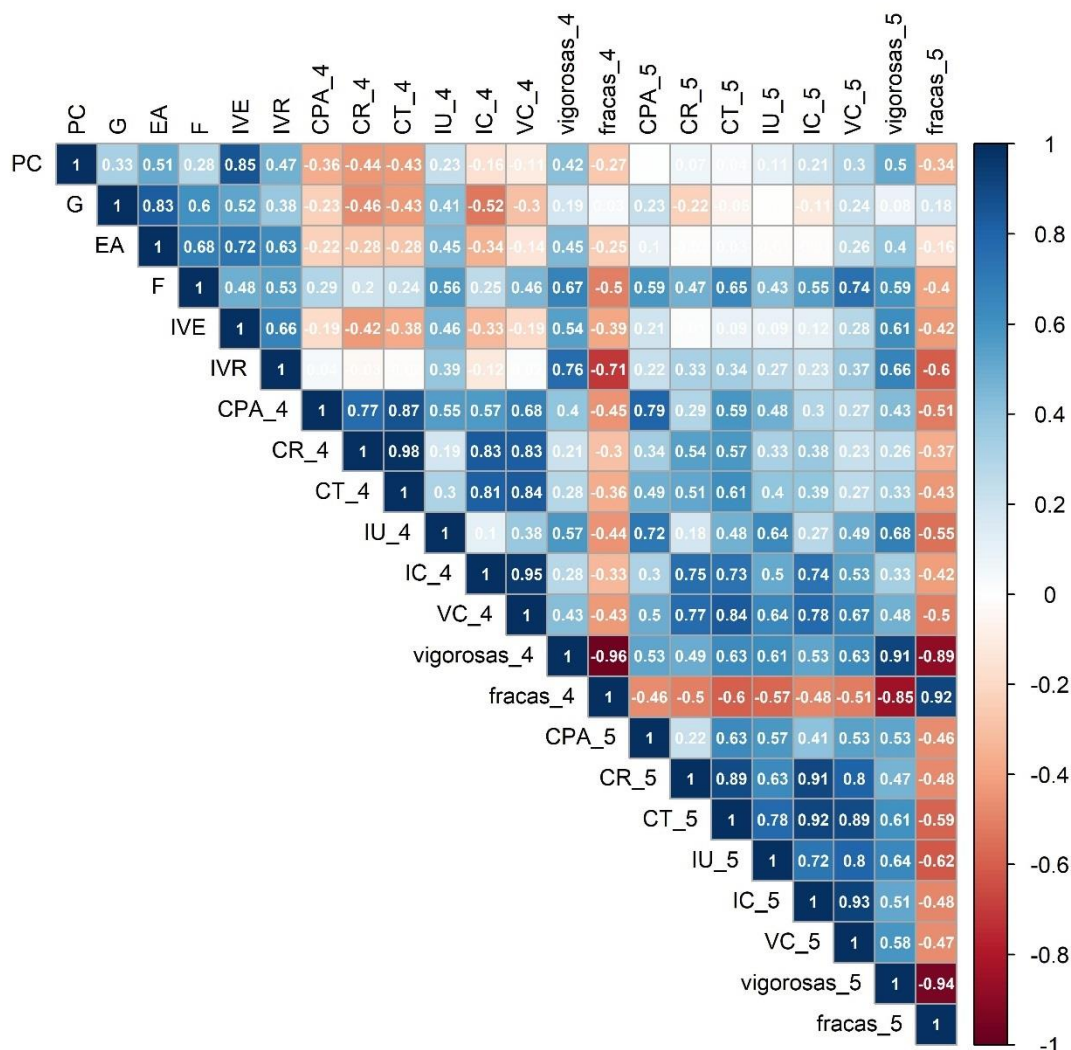


Figura 2 - Coeficientes de correlação de Pearson ( $r$ ) estimados entre as variáveis germinação (G), primeira contagem de germinação (PC), envelhecimento acelerado (EA), teste de frio (F), índice de velocidade de emergência (IVE), índice de velocidade de emissão raiz primária (IVR) e as variáveis obtidas na análise computadorizada.

Pela análise de correlação de Pearson (Figura 1), realizada entre os dados obtidos nos testes de avaliação da qualidade fisiológica e os das análises de plântulas pelo SAPL e ILASTIK, foi possível observar que não houve correlação entre os resultados destas avaliações e os dos testes de germinação e envelhecimento acelerado. Contudo, houve correlação significativa entre os resultados do SAPL e ILASTIK com os do teste de frio, primeira contagem de germinação, índice de velocidade de emergência e índice de velocidade da raiz primária. Observou-se correlação significativa entre a grande maioria das variáveis geradas pelo SAPL e o teste de frio. Entre as variáveis geradas pelo software ILASTIK, a variável número de plantas vigorosas apresentou correlação significativa

com o teste de frio, índice de velocidade de emergência, índice de velocidade da raiz primária e primeira contagem da germinação. Já com a variável número de plantas fracas, houve correlação significativa negativa com o teste de frio e índice de velocidade da raiz primária. Houve correlação significativa entre as variáveis obtidas com a análise de plântulas pelo SAPL e as variáveis do software ILASTIK número de plantas vigorosas e fracas.

Com os resultados obtidos foi demonstrada a relação de coerência entre as análises com testes tradicionais e os parâmetros de caracterização e classificação de imagens gerados pelos sistemas computadorizados SAPL<sup>®</sup> e ILASTIK, concluindo-se que estas ferramentas podem ser incluídas como metodologias para avaliação da qualidade de sementes de arroz, as quais têm vantagens como maior rapidez na obtenção de resultados, precisão e menor subjetividade.

## **CONCLUSÃO**

Em síntese, pode-se afirmar que ambos os softwares SAPL e ILASTIK constituem novas tecnologias em análise de sementes que permitem de forma direta identificar lotes de sementes de arroz com maior potencial fisiológico por meio da captura e armazenamento de imagens de plântulas. Ambos os softwares permitem obter, com acurácia, informações relativamente rápidas sobre o vigor de sementes de arroz, eliminando a subjetividade das interpretações dos testes convencionais.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALVARENGA, R. O.; MARCOS-FILHO, J.; JUNIOR, F. G. G. Avaliação do vigor de sementes de milho superdoce por meio da análise computadorizada de imagens de plântulas. **Revista Brasileira de Sementes**, v. 34, n. 3, p. 488–494, 2012.
- ARAÚJO, JOYCE DE OLIVEIRA *et al.* Chickpea seed vigor evaluated by computerized seedling analysis. **Semina: Ciências Agrárias**, v. 3, n. 2, p. 112–114, 2020.
- BAALBAKI, R.; ELIAS, S.; MARCOS-FILHO, J.; MCDONALD, M. B. Seed Vigor Testing Handbook. **AOSA**, v. 32, p. 343, 2009.
- BAEK, J. S. *et al.* Evaluation of Seed Vigor Tests for Predicting Seedling Establishment at Low Temperature in Rice (*Oryza sativa* L.). **Journal of Crop Science and Biotechnology**, v. 21, n. 2, p. 155–163, 2018.
- BARROS, A. S.; DIAS, M.C.; CICERO, S. M.; KRZYANANOWSKI, F. C. TESTES DE FRIO. IN: KRZYANANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. B. Vigor de sementes: conceitos e testes. **ABRATES**. [S.l: s.n.], p. 5. 1999.
- BERG, S. *et al.* Ilastik: Interactive Machine Learning for (Bio)Image Analysis. **Nature Methods**, v. 16, n. 12, p. 1226–1232, 2019.
- BRASIL. **Regras para Análise de Sementes**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2009.
- BRUNES, A. P. *et al.* Seedling length in wheat determined by image processing using mathematical tools. **Revista Ciencia Agronomica**, v. 47, n. 2, p. 374–379, 2016.
- CASTAN, D. O. C.; GOMES-JUNIOR, F. G.; MARCOS-FILHO, J. Vigor-S, a new system for evaluating the physiological potential of maize seeds. **Scientia Agricola**, v. 75, n. 2, p. 167–172, 2018.
- CHIQUITO, A. A.; GOMES JUNIOR, F. G.; MARCOS-FILHO, J. Assessment of physiological potential of cucumber seeds using the software Seedling Vigor Imaging System® (SVIS®). **Revista Brasileira de Sementes**, v. 34, n. 2, p. 255–263, 2012.
- DANTAS, A. F. *et al.* Loss of genetic integrity in artificially aged seed lots of rice (*Oryza sativa* L.) and common bean (*Phaseolus vulgaris* L.). **Mutation Research - Genetic Toxicology and Environmental Mutagenesis**, v. 846, n. April, p. 403080, 2019. Disponível em: <<https://doi.org/10.1016/j.mrgentox.2019.07.008>>.
- MARCHI, L.; CICERO, SILVIO MOURE; GOMES JUNIOR, F. Utilização Da Análise Computadorizada De Plântulas Na Avaliação Do Potencial Fisiológico De Sementes De Amendoim. **Revista Brasileira de Sementes**, v. 33, p. 652–662, 2011.
- EBONE, L. A.; CAVERZAN, A.; CHAVARRIA, G. Physiologic alterations in orthodox seeds due to deterioration processes. **Plant Physiology and Biochemistry**, v. 145, n. September, p. 34–42, 2019. Disponível em: <<https://doi.org/10.1016/j.plaphy.2019.10.028>>.
- FAO. **Food And Agricultural Organization**. [S.l: s.n.], abr. 2019
- GENEVE, R.; KESTER, S. Evaluation of Seedling Size Following Germination Using Computer-aided Analysis of Digital Images from a Flat-bed Scanner. **HortScience**, v. 36, n. 6, p. 117–120, 2001.
- GOMES-JUNIOR, F. G. Automated image analysis of seedlings for vigor evaluation of

- common bean seeds. **Acta Scientiarum. Agronomy**, v. 36, n. 2, p. 195–200, 2014.
- GRZYBOWSKI, C. R. DE S.; VIEIRA, R. D.; PANOBIANCO, M. Testes de estresse na avaliação do vigor de sementes de milho. **Revista Ciencia Agronomica**, v. 46, n. 3, p. 590–596, 2015.
- HOFFMASTER, A.F.; XU, L.; FUJIMURA, K.; MCDONALD, M. B. ; BENNETT, M.A.; EVANS, A. . The Ohio State University Seed Vigor Imaging System (SVIS) for soybean and corn seedlings. **Seed Technology**, v. 27, n. 1, p. 7–24, 2005.
- HOFFMASTER, A.L.; FUJIMURA, K.; MCDONALD, M.B.; BENNETT, M. . An automated system for vigor testing three-day-old soybean seedlings. **Seed Science and Technology**, v. 31, n. 3, p. 701–713, 2003.
- KAN, A. Machine learning applications in cell image analysis. **Immunology and Cell Biology**, v. 95, n. 6, p. 525–530, 2017.
- KIKUTI, A.; MARCOS-FILHO, J. Análise de imagens de plântulas e testes tradicionais para avaliação do vigor de sementes de quiabo. **Journal of Seed Science**, v. 35, n. 4, p. 443–448, 2013.
- KRZYZANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. B. Vigor de sementes: conceitos e testes. **ABRATES**. [S.l: s.n.], p. 2;10, 1999.
- KUO, T.-Y. *et al.* Identifying rice grains using image analysis and sparse-representation-based classification. **Computers and Electronics in Agriculture**, v. 127, p. 716–725, set. 2016.
- MAGUIRE, J. D. Seeds germination-aid selection and evaluation for seedling emergence and vigor. **Crop Science**, v. 2, n. 2, p. 176–177, 1962.
- MARCHI, J. L.; CICERO, S. M. Use of the software Seed Vigor Imaging System (SVIS®) for assessing vigor of carrot seeds. **Scientia Agricola**, v. 74, n. 6, p. 469–473, 2017.
- MARCOS-FILHO, J.; BENNETT, M. A.; EVANS, A. S.; GRASSBAUGH, E. M. Assessment of melon seed vigour by an automated computer imaging system compared to traditional procedures. . **Seed Science and Technology**, v. 34, n. 2, p. 507–519, 2006.
- MARCOS-FILHO, J. Seed vigor testing: an overview of the past, present and future perspective, **Scientia Agricola**, v.72, n.4, p.363-374, 2015.
- MARCOS-FILHO, J.; KIKUTI, A. L. P.; LIMA, L. B. Métodos para avaliação do vigor de sementes de soja, incluindo a análise computadorizada de imagens. **Revista Brasileira de Sementes**, v. 31, n. 2, p. 102–112, 2009.
- MARCOS-FILHO, J.; KRZYZANOWSKI, F.C.; VIEIRA, R.D.; FANÇA NETO, J. B. Vigor de sementes: conceitos e testes. **ABRATES**. [S.l: s.n.], p. 2–3, 1999.
- MCCORMAC, A. C. ; KEEFE, P. D. ; DRAPER, S. R. Automated vigour testing of field vegetables using image analysis. **Seed Science and Technology**, v. 18, n. 1, p. 103–112, 1990.
- MEDEIROS, A. D. *et al.* Vigor of maize seeds determined by a free image analysis system. **Revista Ciencia Agronomica**, v. 50, n. 4, p. 616–624, 2019.
- MEDEIROS, A. D.; PEREIRA, M. D. SAPL®: A free software for determining the physiological potential in soybean seeds. **Pesquisa Agropecuaria Tropical**, v. 48, n. 3, p. 222–228, 2018.

- MEDEIROS, A. D. DE *et al.* Interactive machine learning for soybean seed and seedling quality classification. **Scientia Agricola**, p. 1–13, 2020.
- NAKAGAWA, J.; KRZYZANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. B. Vigor de sementes: conceitos e testes. **ABRATES**. [S.l: s.n.], p. 2–10, 1999.
- PEREIRA, MÁRCIO DIAS; REIS, JOÃO ANTÔNIO DOS; FERRARI, C. DOS S. Processamento digital de imagens de plântulas na avaliação do vigor de sementes de Moringa oleifera Lam. **Revista Ciencia Agronomica**, v. 30, n. 2, p. 291–306, 2020.
- PINTO, C. A. G. *et al.* Image analysis in the evaluation of the physiological potential of maize seeds. **Revista Ciencia Agronomica**, v. 46, n. 2, p. 319–328, 2015.
- SAKO, Y.; MCDONALD, M. B.; FUJIMURA, K. A System for Automated Seed Vigor Assessment. **Seed Science and Technology**, v. 29, n. 3, p. 625–636, 2001.
- SILVA, C. B. DA *et al.* Automated system of seedling image analysis (SVIS) and electrical conductivity to assess sun hemp seed vigor. **Revista Brasileira de Sementes**, v. 34, n. 1, p. 55–60, 2012.
- SILVA, V. N.; CICERO, S. M. Avaliação do vigor de sementes de tomate durante o armazenamento por meio de análise computadorizada de imagens de plântulas. **Semina: Ciências Agrárias**, v. 35, n. 4, p. 2317–2326, 2014.
- TEAM, R. C. **R Development Core TeamR: A language and environment for statistical computing**, 2019.
- TOHIDLOO, G.; KRUSE, M. Development of an image analysis aided seedling growth test for winter oilseed rape and verification as a vigour test. **Seed Science and Technology**, v. 37, n. 1, p. 98–109, 2009.
- TOLEDO, F. F.; NOVENBRE, A. D. D. L. C.; CHAMMA, H. M. C. P.; WALDEMAR, R. Vigor de sementes de milho (*Zea mays* L.) avaliado pela precocidade de emissão da raiz primaria. **Scientia Agricola**, v. 56, n. 1, p. 112–115, 1999.
- WRASSE, C. F. *et al.* Testes de vigor para sementes de arroz e sua relação com o comportamento de hidratação de sementes e a emergência de plântulas. **Científica**, v. 37, n. 2, p. 107–114, 2009. Disponível em: <10.15361/1984-5529.2009v37n2p107-114>.

## RESUMO

CUADRADO, Johan Bernardo Garcia, M.Sc., Universidade Federal de Viçosa, setembro de 2020. **Capítulo II. Visão computacional e algoritmos de aprendizagem de máquina para classificação de sementes de cultivares de arroz.** Orientadora: Denise Cunha Fernandes dos Santos Dias.

Métodos avançados capazes de auxiliar na identificação de misturas varietais em lotes comerciais de sementes continuam sendo objeto de grande interesse para a indústria sementeira. O objetivo deste trabalho foi propor um método rápido e não-destrutivo para auxiliar na identificação de cultivares arroz com uso de sistema de visão computacional combinado com algoritmos de aprendizagem de máquina. Para tanto, foram utilizadas sementes de oito cultivares de arroz e sementes de arroz vermelho e arroz preto, providas de amostras de lotes comerciais. As imagens das sementes foram adquiridas e processadas com auxílio do equipamento GroundEye<sup>®</sup>, obtendo-se 312 variáveis de cor, textura e geometria de cada semente. Para a análise dos dados, foi realizada estatística multivariada com a análise de componentes principais (PCA). Foram utilizados os algoritmos *Linear Discriminant Analysis* (LDA) e *Support Vector Machine - Radial* (SVM-R) para a criação dos modelos de classificação, avaliados com base na matriz de confusão para os cálculos das métricas acurácia, Kappa, sensibilidade, especificidade e acurácia balanceada. Os modelos desenvolvidos mostraram alto desempenho na identificação de cultivares de arroz, principalmente com uso das características de cor e textura das imagens. A precisão atingida pelos modelos no conjunto de teste foi de 78% e 80% com os algoritmos LDA e SVM-R, respectivamente. A alta precisão dos classificadores mostrou que os métodos de visão computacional com algoritmos de aprendizagem de máquina podem ser considerados promissores para identificar sementes de cultivares de arroz e os resultados podem ser alcançados rápida e eficazmente, em comparação aos métodos convencionais.

Palavras-chave: *Oryza sativa* L. Pureza. Groundeye. LDA. SMV-R.

## ABSTRACT

CUADRADO, Johan Bernardo Garcia, M.Sc., Universidade Federal de Viçosa, September, 2020. **Chapter II. Computer vision and machine learning algorithms for seed classification of rice cultivars.** Advisor: Denise Cunha Fernandes dos Santos Dias.

ABSTRACT - Advanced methods capable of assisting in the identification of varietal mixtures in commercial seed lots remain an object of great interest for the seed industry. The objective of this work was to propose a fast and non-destructive method to assist in the identification of rice cultivars using a computer vision system combined with machine learning algorithms. For that, seeds of eight rice cultivars and seeds of red rice and black rice were used, coming from samples of commercial lots. The images of the seeds were acquired and processed with the aid of the Groundeye® equipment, obtaining, after segmenting the regions of interest in the images, 312 color, texture and geometry variables for each seed. For data analysis, multivariate statistics was performed with principal component analysis (PCA). The Linear Discriminant Analysis (LDA) and Support Vector Machine - Radial (SVM-R) algorithms were used to create the classification models, evaluated based on the confusion matrix for the calculations of the accuracy, Kappa, sensitivity, specificity and balanced metrics accuracy. The developed models showed high performance in the identification of rice cultivars, mainly using the color and texture characteristics of the images, the precision achieved by the models in the test set was 78% and 80% with the LDA and SVM-R algorithms, respectively. The high precision of the classifiers showed that computer vision methods with machine learning algorithms can be easily used to identify seeds of rice cultivars and the results can be achieved quickly and effectively, compared to conventional methods.

Keywords: *Oryza sativa* L. Purity. Groundeye. LDA. SMV-R.

## INTRODUÇÃO

O arroz (*Oryza sativa* L.) é um dos principais cereais consumidos no mundo e reveste-se de grande importância social e econômica. O Brasil é o maior produtor de arroz das Américas, com um montante produzido de 15 milhões de toneladas em 2018 (FAO, 2019). Sem dúvida, parte desses bons resultados são decorrentes do uso de sementes de boa qualidade que são levadas ao campo.

A qualidade das sementes é representada pelo somatório dos atributos de pureza física, genética, fisiológica e sanitária, todos eles igualmente importantes. A qualidade física que diz respeito à composição física do lote, está relacionada à porcentagem de sementes puras, material inerte e outras sementes presentes no lote. A consequência da utilização de um lote de sementes com baixa pureza física é a infestação da lavoura com plantas indesejáveis, que competirão em nutrientes, água e luz, reduzindo a produtividade e a eficiência da colheita (VLACHOS; ARVANITOYANNIS, 2008). A qualidade genética do lote está diretamente relacionada com a pureza varietal, ou seja, com as características da cultivar quando ao ciclo, resistência a doenças, produtividade dentre outras.

A presença de sementes nocivas, como o arroz do tipo vermelho ou preto, é considerada um dos principais problemas que pode comprometer a certificação para comercialização dos lotes de sementes de arroz (GEALY; BRYANT, 2009). Essas sementes causam problemas como difícil controle/erradicação no campo, contaminação de novas áreas, sobrevivência das sementes no solo devido à dormência. Portanto, a identificação eficiente da sua presença nos lotes de sementes é essencial para o controle da sua disseminação.

No Brasil, as outras sementes presentes no lote são identificadas visualmente com base em sua aparência. No entanto, esse processo é demorado e subjetivo (KUO *et al.*, 2016) e nem sempre as características externas das sementes de arroz permitem identificação adequada. A verificação da identidade varietal por meio de análises rápidas é crucial e altamente significativa para a produção agrícola em larga escala no cultivo de arroz (CHEN *et al.*, 2019). Além das inspeções visuais, diversos métodos baseados em marcadores genéticos têm sido aplicados para identificar variedades de sementes de arroz com alta precisão (BECERRA *et al.*, 2015; CHUANG *et al.*, 2011; CIRILLO *et al.*, 2009; STEELE *et al.*, 2008). Porém, esses métodos consomem muito tempo, são complexos e onerosos.

Recentemente, novas e promissoras abordagens baseadas em técnicas de visão computacional e aprendizado de máquina têm sido aplicadas para identificação e avaliação automática de cultivares de arroz. Luo et al. (2019) aplicaram espectroscopia de ruptura induzida por laser (LIBS) combinada com o reconhecimento de padrões utilizando análise discriminante, rede neural de função de base radial e rede neural *perceptron* multicamada, para discriminar espécies de arroz. Huang et al. (2017) estabeleceram um algoritmo para reconhecer sementes de três cultivares de arroz por meio de características geométricas e, alcançaram precisões superiores a 90% com uso de uma rede neural de propagação reversa como classificador. Kuo et al. (2016) distinguiram sementes de arroz de 30 variedades de forma não destrutiva usando processamento de imagens e classificação baseada em representação esparsa (SRC), com precisão de discriminação de 89,1%. Lee et al. (2011) usaram uma câmera CCD (dispositivo de carga acoplada) para capturar 17 características morfocolorimétricas de sementes de arroz, sendo estas processadas por análise discriminante linear (LDA) e utilizadas como entrada para a rede neural de retropropagação para classificar sementes de diferentes variedades de arroz. Embora essas técnicas sejam comprovadamente eficazes, a criação de novas metodologias, simples, equiparadamente eficientes e que atendam a novas demandas do mercado, tendem a agregar novas perspectivas de avanços para o setor.

Sistemas de visão computacional modernos podem atender a essa finalidade e proporcionar a automatização das análises, garantindo ainda menor subjetividade, rapidez e facilidade nos resultados de classificação e avaliação da qualidade das sementes (POURREZA *et al.*, 2012). Portanto, o objetivo deste trabalho foi propor uma metodologia para auxiliar na identificação de cultivares arroz com uso de análise de imagens RGB de sementes e algoritmos de aprendizagem de máquina.

## MATERIAL E MÉTODOS

### Local e material vegetal

A pesquisa foi conduzida no Laboratório de Pesquisa de Sementes do Departamento de Agronomia da Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil. Foram utilizadas sementes de arroz de dez genótipos: Alterosa, Caçula, Caravera, Fedearroz 2000, Predileta, Relâmpago, Rubelita, Seleta e também sementes de arroz vermelho e de arroz preto.

### Aquisição e processamento de imagens

Com auxílio do equipamento GroundEye<sup>®</sup>, foram adquiridas e processadas imagens de cinco amostras de cada genótipo, compostas de aproximadamente 105 sementes de cada genótipo. As amostras foram colocadas sobre bandeja de acrílico transparente e as imagens foram capturadas por câmera de alta resolução contida dentro do módulo de captura do equipamento. Ao todo foi feita a captura de 5.331 imagens de sementes de arroz.

Utilizando-se o sistema de análise de imagens do GroundEye<sup>®</sup>, foi realizada a segmentação das regiões de interesse (sementes) e remoção do fundo das imagens. Foram extraídas 267 variáveis de cor, textura e geometria de cada semente.

### **Análise de componentes principais**

Para a análise dos dados foi realizada estatística multivariada com a análise de componentes principais (PCA) usando o pacote *factoextra*. Uma matriz “n x p” foi obtida onde “n” é o número de categorias (n=10) e “p” o número de variáveis analisadas (p=312). Autovalores foram calculados a partir das matrizes de covariância e representados em gráficos bidimensionais.

### **Aprendizagem de máquina**

#### **Classes de sementes e particionamento de dados**

Cada cultivar foi tomado com uma classe independente. Ao todo foram utilizadas 10 classes: Classe 1-Alterosa; Classe 2-Caçula; Classe 3-Caravera; Classe 4-Fedearroz 2000; Classe 5-Predileta; Classe 6-Arroz preto; Classe 7-Relâmpago; Classe 8 -Rubelita; Classe 9-Seleta; Classe 10-Arroz vermelho.

O banco de dados total foi dividido em dois conjuntos de dados menores, um utilizado para o treinamento do modelo e outro para testá-lo. Para o treinamento, foram utilizados 70% dos dados e no teste os 30% dos dados restantes.

### **Algoritmos de aprendizagem de máquina**

Foram utilizados os algoritmos *Linear Discriminant Analysis* (LDA) e *Support Vector Machine - Radial* (SVM-R). O algoritmo LDA é um classificador que maximiza a distância entre classes e minimiza a distância dentro da classe para atingir uma máxima separabilidade entre elas. Já o SVM-R é um classificador que representa aos pontos de

amostra no espaço, separando as classes em múltiplos espaços, usando um hiperplano de separação definido como o vetor entre os pontos das classes chamado vetor suporte.

O desenvolvimento dos modelos foi implementado no software 3.6.3 (TEAM, 2019). Para o algoritmo LDA foi usado o pacote Mass (<https://cran.r-project.org/web/packages/MASS/index.html>), para o algoritmo SVM o pacote Caret (<https://cran.r-project.org/web/packages/caret/index.html>).

### Validação dos modelos

Os modelos desenvolvidos foram validados por meio de validação cruzada (10 folds) utilizando os dados de treinamento e pelo uso de um conjunto de teste (dados não utilizados previamente para o treinamento). Os dados foram avaliados com base na matriz de confusão com os valores de Verdadeiros Positivos (TP), Falsos Positivos (FP), Verdadeiros Negativos (TN) e Falsos Negativos (FN), para cálculo das métricas acurácia, sensibilidade, especificidade, acurácia balanceada e Kappa (COHEN, 1968). As variáveis mais importantes foram importadas e representadas graficamente usando a função varImp (TEAM, 2019).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Kappa} = \frac{2 * \text{TP} * \text{TN} - \text{FP} * \text{FN}}{\text{TP} * \text{FN} + \text{TP} * \text{FP} + 2 * \text{TP} * \text{TN} + \text{FN}^2 + \text{FN} * \text{TN} + \text{FP}^2 + \text{FP} * \text{TN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2}$$

## RESULTADOS

### Análise de componentes principais

A análise de componentes principais foi utilizada para reduzir a dimensionalidade dos dados, servindo como análise exploratória para visualização de padrões ocultos contidos no conjunto de dados. Os três primeiros componentes explicaram mais da metade da variabilidade total dos dados (Figura 1). Observa-se que as sementes do genótipo de arroz preto apresentaram diferença significativa dos demais genótipos no primeiro componente, localizando-se nos *scores* mais negativos. Os demais genótipos apresentaram diferenças menos acentuadas entre as sementes.

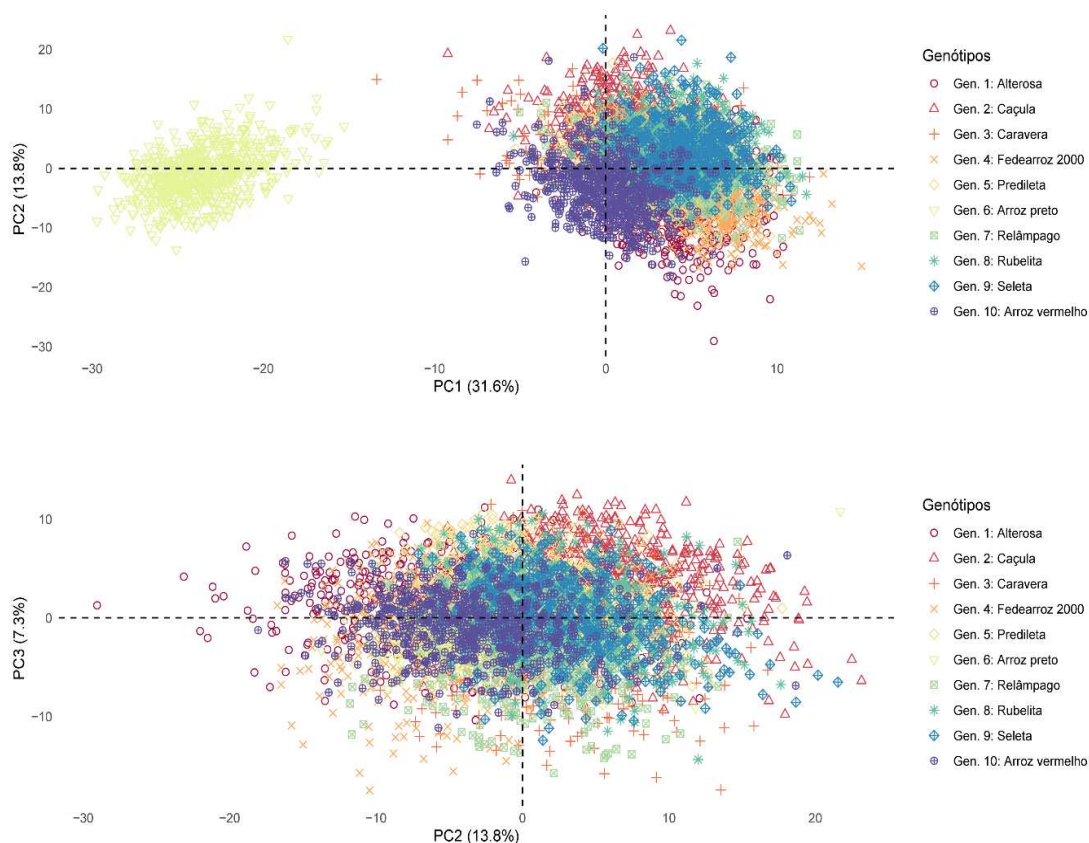


Figura 1 - Análise de componentes principais considerando variáveis de cor, forma e textura de 5331 sementes de 10 genótipos de arroz.

### Modelos de aprendizagem de máquina para classificação de cultivares de arroz

Os descritores de sementes gerados pelo software GroundEye<sup>®</sup> foram utilizados no desenvolvimento de modelos de aprendizagem de máquina com o uso de algoritmos LDA e SVM-R, para classificar sementes quanto aos genótipos que pertenciam. Em geral, foi observado alto desempenho nos dois modelos desenvolvidos.

Tabela 1- Matriz de confusão para classificação de sementes de genótipos de arroz com uso do algoritmo LDA.

Classes	Training set (n = 3737)									
	Alteros	Caçula	Caravera	F-2000	Predileta	Preto	Relâm	Rubelita	Seleta	Vermelho
Alterosa	293	0	0	0	0	0	0	3	0	1
Caçula	0	343	9	0	0	0	4	0	1	0
Caravera	0	13	282	0	1	0	69	3	0	0
F-2000	12	0	0	307	0	0	1	0	2	0
Predileta	13	1	0	2	315	0	0	48	22	7
Preto	0	0	0	0	0	445	0	0	0	0
Relâmpago	0	6	90	3	0	0	261	0	0	0
Rubelita	33	1	0	5	39	0	0	245	28	0
Seleta	41	5	0	3	16	0	1	44	296	0
Vermelho	2	0	0	2	0	0	0	2	1	416
Accuracy	0.86									
Kappa	0.84									
Sensitivity	0.74	0.93	0.74	0.95	0.85	1.00	0.78	0.71	0.85	0.98
Specificity	1.00	1.00	0.97	1.00	0.97	1.00	0.97	0.97	0.97	1.00
Balanced Accuracy	0.87	0.96	0.86	0.97	0.91	1.00	0.87	0.84	0.91	0.99
Cross-validation (folds = 10)										
Accuracy	0.80 ± 0.02									
Kappa	0.77 ± 0.02									
Sensitivity	0.79 ± 0.15									
Specificity	0.98 ± 0.02									
Balanced Accuracy	0.88 ± 0.08									
Testing set (n = 1.594)										
Alterosa	94	1	0	0	0	0	0	0	0	0
Caçula	0	148	6	0	0	0	3	0	1	0
Caravera	0	3	115	0	0	0	45	1	0	0
F-2000	11	0	0	121	0	0	0	0	0	1
Predileta	10	1	0	0	121	0	0	30	11	4
Preto	0	0	0	0	0	190	0	0	0	0
Relâmpago	0	3	39	3	0	0	92	0	0	1
Rubelita	20	1	0	3	22	0	0	77	21	4
Seleta	33	0	2	8	16	0	3	39	116	4
Vermelho	0	1	0	2	0	0	0	0	0	167
Accuracy	0.78									
Kappa	0.75									
Sensitivity	0.56	0.94	0.71	0.88	0.76	1.00	0.64	0.52	0.78	0.92
Specificity	1.00	0.99	0.97	0.99	0.96	1.00	0.97	0.95	0.93	1.00
Balanced Accuracy	0.78	0.96	0.84	0.94	0.86	1.00	0.81	0.74	0.85	0.96

Tabela 2 - Matriz de confusão para classificação de sementes de genótipos de arroz com uso do algoritmo SVM-R

Classes	Training set (n=3737)									
	Alterosa	Caçula	Caravera	F-2000	Predileta	Preto	Relâm	Rubelita	Seleta	Vermelho
Alterosa	352	0	0	2	1	0	0	5	3	1
Caçula	0	364	1	0	0	0	3	0	0	0
Caravera	0	3	351	0	0	0	27	2	0	0
F-2000	5	0	0	316	0	0	0	0	0	0
Predileta	9	0	0	2	343	0	0	27	10	1
Preto	0	0	0	0	0	445	0	0	0	0
Relâmp	0	2	29	0	0	0	306	0	0	0
Rubelita	12	0	0	0	20	0	0	286	9	0
Seleta	16	0	0	1	7	0	0	23	327	0
Vermelho	0	0	0	1	0	0	0	2	1	422
Accuracy	0.94									
Kappa	0.93									
Sensitivity	0.89	0.99	0.92	0.98	0.92	1.00	0.91	0.83	0.93	1.00
Specificity	1.00	1.00	0.99	1.00	0.99	1.00	0.99	0.99	0.99	1.00
Balanced Accuracy	0.94	0.99	0.96	0.99	0.96	1.00	0.95	0.91	0.96	1.00
Cross-validation (folds=10)										
Accuracy	0.80 ± 0.02									
Kappa	0.77 ± 0.02									
Sensitivity	0.79 ± 0.15									
Specificity	0.98 ± 0.01									
Balanced Accuracy	0.88 ± 0.08									
Testing set (n=1594)										
Alterosa	135	1	0	5	3	0	0	3	2	1
Caçula	0	146	5	0	2	0	4	1	0	0
Caravera	0	4	108	2	1	0	40	3	1	0
F-2000	4	1	2	125	0	0	2	2	2	0
Predileta	2	0	0	0	117	0	2	36	11	4
Preto	0	0	0	0	0	190	0	0	0	0
Relâmp	1	2	46	2	0	0	94	4	1	0
Rubelita	12	2	1	0	24	0	1	71	22	2
Seleta	14	2	0	0	12	0	0	26	110	1
Vermelho	0	0	0	3	0	0	0	1	0	173
Accuracy	0.80									
Kappa	0.77									
Sensitivity	0.80	0.92	0.67	0.91	0.74	1.00	0.66	0.48	0.74	0.96
Specificity	0.99	0.99	0.96	0.99	0.96	1.00	0.96	0.96	0.96	1.00
Balanced Accuracy	0.90	0.96	0.82	0.95	0.85	1.00	0.81	0.72	0.85	0.98

O modelo desenvolvido com o algoritmo LDA atingiu acurácia de 85% no treinamento, 80% na validação cruzada e 78% no conjunto de teste, com coeficiente Kappa  $\geq 75\%$  (Tabela 1). Com o algoritmo LDA, as classes correspondentes ao arroz preto e arroz vermelho apresentaram as maiores taxas de acerto, com valores  $\geq 96\%$ . As

taxas mais baixas de sensibilidade foram observadas para os genótipos Alterosa e Rubelita, apresentando também os menores valores de acurácia balanceada.

O modelo desenvolvido com o algoritmo SVM-R apresentou alto desempenho na classificação de genótipos de arroz (tabela 2), com acurácia de 80%, e coeficiente de Kappa de 77% no teste, sendo estes valores iguais aos calculados na validação cruzada. O modelo obteve 100% de acerto na classificação das sementes de arroz preto e 98% para as sementes de arroz vermelho. O modelo desenvolvido com o algoritmo SVM-R no teste gerou acurácia balanceada acima de 72%, apresentando a maior taxa de falsos negativos para as sementes do genótipo Rubelita.

### **Importância das variáveis**

As variáveis de cor e textura das sementes foram as mais importantes na classificação dos genótipos, apresentando valores acima de 80% de importância para a maioria das classes nos dois modelos desenvolvidos (Figura 2). Na classe 4, pertencente ao genótipo Fedearroz 2000, as variáveis de cor apresentaram valores altos e discriminantes em comparação às demais classes, obtendo 100% de importância na classificação. O arroz vermelho apresentou maior variação na importância das variáveis usadas para a classificação, em que as variáveis de cor se sobressairam em relação às demais. As características de geometria das sementes usadas não apresentaram grande relevância na classificação nos genótipos nos dois modelos. Em geral, as variáveis mais importantes para gerar classificação em todos os genótipos nos dois modelos foram: color 11 (Acima do otsu: CIELab- L), color 23 (Dominância amarela), color 56 (HSL: Matiz assimetria), color 60 (HSL: Matiz kurtosis), texture 40 (Run length: LGE), texture 43 (Length: LRLGE), texture 47 (Run length: SRHGE), texture 48 (Run length: SRLGE).

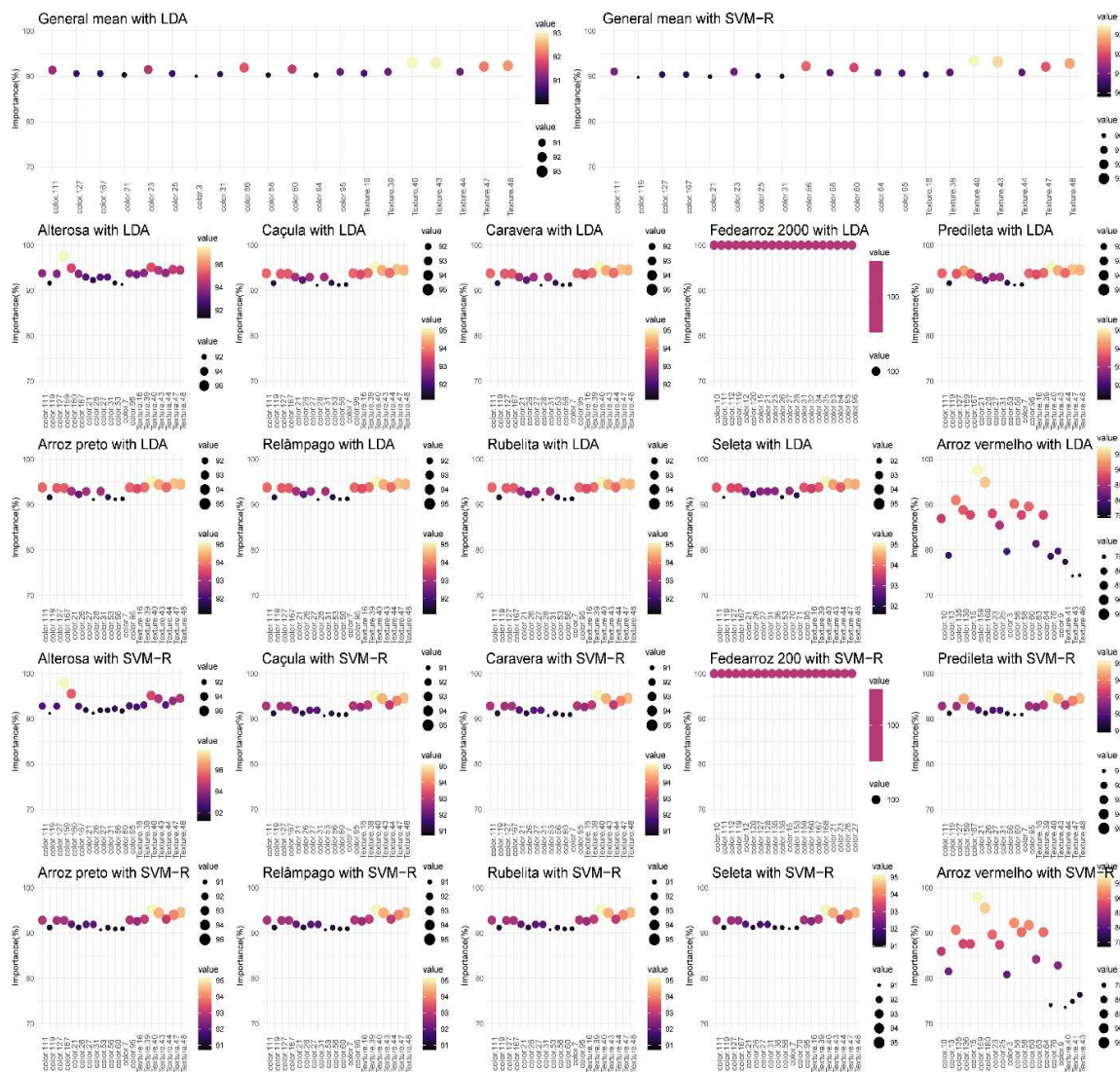


Figura 2 - Vinte variáveis de maior importância para os modelos de classificação LDA e SVM-R de sementes de diferentes genótipos de arroz.

## DISCUSSÃO

A pureza física e varietal são fatores primordiais no que diz respeito à qualidade de lotes de sementes de arroz, tendo impacto direto no preço de mercado e na uniformidade do cultivo. O padrão do MAPA para sementes de arroz no Brasil estabelece como aceitável no máximo 1 semente de arroz vermelho em uma amostra de 700g na classe certificada e 2 na não certificada na avaliação de outras sementes durante a análise de pureza. Já as sementes de arroz preto são consideradas proibidas, com tolerância 0.

Nos laboratórios de análise de sementes atualmente a classificação dos lotes quanto à pureza varietal é visual, sendo um processo difícil, demorado e subjetivo devido à quantidade de sementes a serem analisadas, além da semelhança nas características próprias de algumas sementes (ZHANG *et al.*, 2019). Nas sementes de arroz vermelho, a lema e pálea dificilmente diferem visualmente das sementes das variedades cultivadas, sendo necessário nos laboratórios descascar uma amostra de 1 Kg, para verificar a presença do pericarpo vermelho. A identificação de cultivares de arroz com base na avaliação visual de características externas das sementes é extremamente difícil, sendo importante a automação levando em consideração outras informações como cor, textura e geometria. Neste estudo propomos uma metodologia simples e eficiente para classificação automática de sementes de diferentes genótipos de arroz, utilizando informações de cor, textura e geometria de imagens RGB de sementes combinadas com o uso de algoritmos de aprendizagem de máquina.

Os modelos desenvolvidos usando LDA e SVM-R foram altamente eficientes na classificação dos genótipos de arroz. Foram obtidas altas acurácias e índice de Kappa no conjunto de teste e validação cruzada (Tabelas 1, 2). Os modelos propostos apresentaram sensibilidade acima de 96% na discriminação de sementes de arroz vermelho e preto, o que indica que são promissores para serem utilizados nos programas de controle de qualidade de empresas de sementes ou laboratórios de análises para determinar a contaminação dos lotes. Contudo, há ainda necessidade de mais estudos para tentar aumentar a sensibilidade do método. Pela alta eficácia dos modelos para a classificação dos genótipos usando características morfométricas e de cor nas sementes, estudos adicionais devem ser implementados buscando ajustes para uso tanto na análise de rotina em laboratório como para monitorar a qualidade dos lotes após a colheita, na recepção e beneficiamento e também para avaliação de homogeneidade de acessos dos bancos de germoplasma dos programas de melhoramento genético de arroz.

Em estudos recentes, pesquisadores tentaram classificar genótipos de arroz usando diferentes técnicas. Chaugule *et al.* (2016) usaram um método de aprendizagem de máquina para identificação de variedades de arroz com o uso das características angulares da semente, que foram comparadas com características como cor, forma e textura; algoritmo proposto atingiu uma precisão de 97.6% no conjunto de teste. Weng *et al.* (2020) desenvolveram uma técnica para classificação de genótipos de arroz usando imagem hiperespectral e deep learning network, obtendo uma precisa identificação das

variedades com acurácia média de 98.57% no conjunto de teste. Qiu *et al.* (2018) identificaram sementes de distintos genótipos de arroz usando imagens hiperespectrais combinadas com o método rede neural convolucional (CNN), mostrando que a CNN pode ser adotada na análise de dados espectrais com resultados promissores.

Os algoritmos de aprendizagem de máquina utilizados neste estudo têm sido utilizados com sucesso na ciência de plantas. O LDA é um algoritmo que aumenta a distância entre as classes e reduz a separabilidade dentro da classe, combinando linearmente recursos e criando estimativas de limites. Baek *et al.* (2019) obtiveram acurácia acima de 90% usando modelo LDA na seleção de bandas de ondas hiperespectrais ideais para detecção de sementes de arroz descoloridas e doentes. Segundo França-silva *et al.*, 2020, o algoritmo Análise Discriminante Linear (LDA) pode ser utilizado para criação de modelos de detecção de *Drechslera avenae* em sementes de aveia preta, usando os recursos de refletância, cor e textura das imagens, com precisão de 86% na validação independente.

O SVM-R é um algoritmo que projeta dados em um espaço de recursos dimensionais mais altos e identifica um hiperplano para separar classes com a margem mais significativa possível. O benefício de um hiperplano é encontrar uma linha de separação entre os dados de duas classes, reduzindo consideravelmente as classificações falsas (GUERRERO *et al.*, 2012). Kong *et al.* (2013) obtiveram 100% de accuracy com o algoritmo SVM na identificação de cultivares de sementes de arroz usando imagens hiperespectrais no infravermelho próximo e análise de dados multivariada. Diferenças claras entre variedades de sementes de arroz podem ser facilmente visualizadas usando a técnica de imagem multiespectral e uma excelente classificação pode ser alcançada combinando dados das características espectrais e morfológicas usando modelos com algoritmo SVM (LIU *et al.*, 2016). Zhou *et al.* (2020) relataram a eficiência do algoritmo SVM como modelo de classificação na discriminação de sementes de milho baseada em imagem hiperespectral. No nosso estudo, o modelo usando SVM apresentou maiores valores de acurácia e coeficiente de Kappa, quando comparado ao modelo com LDA. É importante enfatizar que o sistema proposto é baseado em simples imagens RGB, não sendo necessário procedimentos complexos para extração das características. Além disso, representa uma alternativa de baixo custo comparado a sistemas hiper e multiespectrais.

Os modelos desenvolvidos foram eficientes na classificação dos genótipos utilizando as variáveis de cor e texturas nas imagens, sendo que as variáveis de geometria não foram

significativamente importantes para os classificadores explicado pela alta homogeneidade na geometria das sementes dentro das classes analisadas. Sun *et al.* (2015), em estudos de identificação de origem de arroz usando imagem hiperespectral, obtiveram precisão do modelo SVM com base nas variáveis de textura de 73,33% e 54.17% usando unicamente variáveis de geometria.

Neste estudo, vale ressaltar a necessidade de utilização de sementes com alta pureza varietal no conjunto de treino e validação cruzada, para garantir altas métricas de sensibilidade e especificidade no teste. Além disso, fatores ambientais e sanitários regulam a aparência das sementes, o que pode interferir no desempenho dos modelos de classificação propostos. Uma limitação no uso de abordagem baseada em imagem para identificar genótipos de arroz é que se requerem sementes com características morfológicas e de cores distintas entre os genótipos a serem analisados. Verifica-se que os classificadores alcançaram menor precisão no genótipo Rubelita porque as aparências das sementes foram semelhantes às aparências dos genótipos Predileta e Seleta.

Na área de análise de sementes, o uso de ferramentas de visão computacional que possam garantir rápida e efetiva tomada de decisão nos resultados de qualidade, é de alta relevância para economizar tempo e recursos. O presente estudo demonstrou a possível classificação de diferentes genótipos de arroz, com base nas características de cor e textura das sementes.

## CONCLUSÃO

Os algoritmos LDA e SVM-R são eficazes na classificação de sementes de diferentes cultivares de arroz e também para a separação de sementes de arroz vermelho e preto com base nos descritores gerados a partir de imagens e têm potencial para serem utilizados em programas de controle de qualidade de empresas de sementes ou laboratórios de análises.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BAEK, I. *et al.* Selection of optimal hyperspectral wavebands for detection of discolored, diseased rice seeds. **Applied Sciences (Switzerland)**, v. 9, n. 5, p. 1–15, 2019.
- BECERRA, V. *et al.* Genetic diversity, identification, and certification of Chilean rice varieties using molecular markers. **Chilean journal of agricultural research**, v. 75, n. 3, p. 267–274, 2015.
- CHAUGULE, A. A.; MALI, S. N. Identification of paddy varieties based on novel seed angle features. **Computers and Electronics in Agriculture**, v. 123, p. 415–422, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.compag.2016.03.012>>.
- CHEN, J. *et al.* Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 219, p. 179–185, 2019. Disponível em: <<https://doi.org/10.1016/j.saa.2019.03.105>>.
- CIRILLO, A. *et al.* Molecular characterization of Italian rice cultivars. **European Food Research and Technology**, v. 228, n. 6, p. 875–881, abr. 2009.
- COHEN, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. **Psychological Bulletin**, v. 70, n. 4, p. 213–220, 1968.
- FAO. **Food And Agricultural Organization**. [S.l: s.n.], abr. 2019.
- FRANÇA-SILVA, F. *et al.* Detection of *drechslera avenae* (Eidam) sharif [*helminthosporium avenae* (eidam)] in black oat seeds (*avena strigosa* schreb) using multispectral imaging. **Sensors (Switzerland)**, v. 20, n. 12, p. 3343, 2020.
- GEALY, D. R.; BRYANT, R. J. Seed physicochemical characteristics of field-grown US weedy red rice (*Oryza sativa*) biotypes: Contrasts with commercial cultivars. **Journal of Cereal Science**, v. 49, n. 2, p. 239–245, 2009.
- GUERRERO, J. M. *et al.* Expert Systems with Applications Support Vector Machines for crop weeds identification in maize fields. **Expert Systems With Applications**, v. 39, n. 12, p. 11149–11155, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2012.03.040>>.
- HUANG, K.-Y. *et al.* A Novel Method of Identifying Paddy Seed Varieties. **Sensors**, v. 17, n. 4, p. 809, abr. 2017.
- KONG, W. *et al.* Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. **Sensors (Basel, Switzerland)**, v. 13, n. 7, p. 8916–8927, 2013.
- KUO, T.-Y. *et al.* Identifying rice grains using image analysis and sparse-representation-

based classification. **Computers and Electronics in Agriculture**, v. 127, p. 716–725, 2016.

LEE, C. Y. *et al.* Intelligent classification methods of grain kernels using computer vision analysis. **Measurement Science and Technology**, v. 22, n. 6, p. 064006, jun. 2011.

LIU, W. *et al.* Online Variety Discrimination of Rice Seeds Using Multispectral Imaging and Chemometric Methods. **Journal of Applied Spectroscopy**, v. 82, n. 6, p. 993–999, 2016.

LUO, Z. *et al.* Rapid identification of rice species by laser-induced breakdown spectroscopy combined with pattern recognition. **Applied Optics**, v. 58, n. 7, p. 1631, mar. 2019.

POURREZA, A. *et al.* Identification of nine Iranian wheat seed varieties by textural analysis with image processing. **Computers and Electronics in Agriculture**, v. 83, p. 102–108, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.compag.2012.02.005>>.

QIU, Z. *et al.* Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. **Applied Sciences (Switzerland)**, v. 8, n. 2, p. 1–12, 2018.

STEELE, K. A. *et al.* InDel markers distinguish Basmati from other fragrant rice varieties. **Field Crops Research**, v. 105, n. 1–2, p. 81–87, jan. 2008.

SUN, J. *et al.* a method for rapid identification of rice origin by hyperspectral imaging technology. **Journal of Food Process Engineering**, v. 40, n. 12297, 2015.

TEAM, R. C. **R Development Core TeamR: A language and environment for statistical computing**, 2019.

VLACHOS, A.; ARVANITTOYANNIS, I. S. A review of rice authenticity/adulteration methods and results. **Critical Reviews in Food Science and Nutrition**, v. 48, n. 6, p. 553–598, 2008.

WENG, S. *et al.* Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 234, p. 118237, 2020. Disponível em: <<https://doi.org/10.1016/j.saa.2020.118237>>.

ZHANG, J. *et al.* Purity analysis of multi-grain rice seeds with non-destructive visible and near-infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 164, n. April, p. 104882, 2019.

ZHOU, Q. *et al.* Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm. **Infrared Physics & Technology**, v. 109, n. 10, p. 103418, 2020.