

UNIVERSIDADE FEDERAL DE VIÇOSA

DIANA SOARES MAGALHÃES

**DESENVOLVIMENTO E AVALIAÇÃO DE UM SISTEMA AUTOMATIZADO
PARA CLASSIFICAÇÃO DE GRÃOS DE SOJA A PARTIR DE IMAGENS DIGITAIS**

**VIÇOSA – MINAS GERAIS
2022**

DIANA SOARES MAGALHÃES

**DESENVOLVIMENTO E AVALIAÇÃO DE UM SISTEMA AUTOMATIZADO
PARA CLASSIFICAÇÃO DE GRÃOS DE SOJA A PARTIR DE IMAGENS DIGITAIS**

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

Orientador: Paulo Cesar Corrêa

Coorientadores: Luís César da Silva
Domingos Sárvio Magalhães
Valente

**VIÇOSA – MINAS GERAIS
2022**

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

T

M188d
2022

Magalhães, Diana Soares, 1996-
Desenvolvimento e avaliação de um sistema automatizado
para classificação de grãos de soja a partir de imagens digitais /
Diana Soares Magalhães. – Viçosa, MG, 2022.
1 dissertação eletrônica (72 f.): il. (algumas color.).

Orientador: Paulo Cesar Corrêa.
Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Engenharia Agrícola, 2022.

Referências bibliográficas: f. 63-72.

DOI: <https://doi.org/10.47328/ufvbbt.2022.180>

Modo de acesso: World Wide Web.

1. Soja - Grãos. 2. Raspberry Pi (Computador).
3. Automação. 4. Python (Linguagem de programação de
computador). 5. Soja - Fisiologia pós-colheita. 6. Propriedades
físicas. I. Corrêa, Paulo Cesar, 1951-. II. Universidade Federal
de Viçosa. Departamento de Engenharia Agrícola. Programa de
Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 633.34

DIANA SOARES MAGALHÃES

**DESENVOLVIMENTO E AVALIAÇÃO DE UM SISTEMA AUTOMATIZADO
PARA CLASSIFICAÇÃO DE GRÃOS DE SOJA A PARTIR DE IMAGENS DIGITAIS**

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

Aprovada: 26 de janeiro de 2022.

Assentimento:



Diana Soares Magalhães

Autora



Paulo Cesar Corrêa

Orientador

“A única forma de chegar ao impossível é acreditar que é possível.”

(Alice in wonderland)

AGRADECIMENTOS

A Deus, pelo dom da vida e pelas oportunidades que me trouxeram até aqui.

À minha família por todo carinho e amor depositados e que, mesmo com a distância, não deixaram de me apoiar e sempre vibrar com minhas conquistas.

Ao Danley que mesmo longe esteve sempre ao meu lado, me apoiando e me incentivando com muito amor, carinho e paciência, por me ajudar a superar os momentos difíceis e sempre acreditar em mim.

Ao professor Paulo Cesar Corrêa pela orientação e transmissão de conhecimentos que foram essenciais para a conclusão dessa etapa e que serão levados para o resto da vida.

Ao professor e coorientador Luís César pela contribuição tão importante no desenvolvimento desse trabalho, por acreditar em mim e em meu trabalho, pelas conversas, conselhos e ensinamentos tão importantes.

Aos colegas de laboratório de Propriedades Físicas, Juliana, Camilla e Marcos por todo companheirismo, momentos de descontração e trocas de conhecimento. Agradeço em especial a Juliana por acreditar nesse projeto e firmar parcerias tão importantes para a conclusão desse trabalho.

Ao Augusto pela ajuda imprescindível, pela paciência em sanar minhas dúvidas e toda a assistência durante esse período.

Agradeço aos demais colegas que conheci durante essa jornada, pelo convívio e troca de experiências.

Ao professor Paulo Roberto Cecon pelo auxílio nas análises estatísticas.

A Procer Automação e Sistemas pela confiança em meu trabalho e por acreditar no desenvolvimento do setor de pós-colheita de grãos.

Ao professor Domingos Savio pela coorientação e aos membros da banca avaliadora que despuseram de tempo e atenção em ler o material e por contribuir para o aprimoramento desse trabalho

À Universidade Federal de Viçosa por meio do Departamento de Engenharia Agrícola pela oportunidade.

Aos colegas e funcionários do Centro Nacional de Treinamento em Armazenagem (CENTREINAR) pela colaboração.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de estudos.

Agradeço a todos que estiveram presentes durante esse processo e que de alguma forma contribuíram para essa conquista.

BIOGRAFIA

DIANA SOARES MAGALHÃES, filha de Leondina Magalhães Castilho e Divino Soares Pereira, nasceu em Januária, Minas Gerais, em 8 de março de 1996.

Em março de 2014, iniciou o curso de Engenharia Agrícola e Ambiental no Instituto Federal do Norte de Minas Gerais – Campus Januária, graduando-se em maio de 2019. Durante a graduação atuou em diversas atividades dentre elas como bolsista em treinamento no laboratório de Biologia, monitora de Física Fundamental I, foi cofundadora de empresa júnior e bolsista de iniciação científica pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Em agosto de 2019, ingressou no Programa de Pós-Graduação em Engenharia Agrícola, em nível de Mestrado, na área de Armazenamento e Processamento de Produtos Agrícolas da Universidade Federal de Viçosa, submetendo-se à defesa de dissertação, requisito indispensável para a obtenção de título de *Magister Scientiae*, em janeiro de 2022.

RESUMO

MAGALHÃES, Diana Soares, M.Sc., Universidade Federal de Viçosa, janeiro de 2022. **Desenvolvimento e avaliação de um sistema automatizado para classificação de grãos de soja a partir de imagens digitais.** Orientador: Paulo Cesar Corrêa. Coorientadores: Luís César da Silva e Domingos Sárvio Magalhães Valente.

A soja (*Glycine max*) é uma das culturas mais difundidas pelo mundo e de grande importância econômica. Os procedimentos de classificação e determinação das propriedades físicas de grãos e sementes normalmente são realizados de forma manual e empregam diversos equipamentos, tornando-se assim um trabalho laborioso e sujeito a subjetividades. Sendo assim, o presente trabalho foi realizado com o objetivo de desenvolver um sistema automatizado para classificação de amostras de grãos de soja por meio de imagens digitais. O protótipo para obtenção das imagens digitais foi confeccionado em MDF preto e possuía embarcado o microprocessador Raspberry Pi integrada a uma Pi Camera para captura das imagens. A câmera era acionada por meio de um código em Python, em que foram padronizadas configurações como resolução, frame rate, brilho, contraste e saturação a fim de obter imagens uniformes. Para determinação das propriedades físicas foram capturadas imagens de grãos inteiros e através desses foram determinados largura, comprimento, área projetada e perímetro de 50 grãos analisados por meio de um aplicativo computacional utilizando linguagem de programação em Python. As medidas para os grãos obtidas pelo aplicativo foram comparadas com as determinadas por meio de um paquímetro. Aplicou-se teste de Bland-Altman para as variáveis de largura, comprimento e circularidade. Obteve-se resultados satisfatórios para a variável comprimento, em que o método de obtenção das medidas por imagens digitais não se diferiu estatisticamente do método do paquímetro, tornando-os equivalentes. Para as variáveis de largura e circularidade o método digital não se demonstrou equivalente, entretanto grande parte dos dados se encontram dentro do intervalo de confiança de 95%. Para as variáveis de área projetada e perímetro, a concordância entre os dados foi analisada por meio de uma ANOVA seguido do teste Tukey a 5% de probabilidade. A variável área projetada não diferiu significativamente segundo os métodos utilizados, enquanto o perímetro se diferiu apenas para o método em que se utilizou funções da biblioteca OpenCV. Para a classificação de grãos de soja, foram adquiridas 750 imagens digitais de amostras contendo combinações aleatórias dos atributos dos grãos sadios (A1), ardidos (A2), mofados (A3), queimados (A4), imaturos (A5),

danificados (A6), fermentados (A7), partidos e quebrados (A8), amassados (A9), chochos (A10), esverdeados (A11) e germinados (A12). Desenvolveu-se um aplicativo computacional para processamento de imagens e extração de características por meio das bibliotecas OpenCV e NumPy. Os dados obtidos foram utilizados para simulação de Redes Neurais Artificiais (RNA) em sete simulações de desempenho com o intuito de selecionar uma rede neural eficiente e capaz de classificar os grãos de soja. Os testes foram realizados com diferentes configurações de rede por meio do programa Matlab e as validações cruzadas foram obtidas pelo Orange Canvas. A maior acurácia foi de 80,01% encontrada para o segundo caso de simulação, utilizando 60% dos dados para treinamento em uma RNA com 30 neurônios. Para a validação da RNA foram utilizadas imagens de 10 amostras contendo os atributos e quantidade conhecidos de grãos de soja, as imagens foram submetidas ao melhor treinamento de rede encontrado e obteve taxa média de acerto de 80,8%. Dentre os 12 atributos empregados na classificação de grãos no modelo desenvolvido, esses apresentaram uma taxa de acerto de: A1 – 71,2%; A2 – 87,6%; A3 – 62,1%; A4 – 99,3%; A5 – 75,8%; A6 – 67,0%; A7 – 73,5%; A8 – 76,4%; A9 – 95,3%; A10 – 96,1%; A11 – 92,4% e A12 – 54,3.

Palavras-chave: Raspberry Pi. Automação. Python. Pós-colheita. Propriedades físicas.

ABSTRACT

MAGALHÃES, Diana Soares, M.Sc., Universidade Federal de Viçosa, January, 2022. **Development and evaluation of an automated system for classification of soybeans from digital images.** Adviser: Paulo Cesar Corrêa. Co-advisers: Luís César da Silva and Domingos Sárvio Magalhães Valente.

Soybean (*Glycine max*) is one of the most widespread crops in the world and of great economic importance. The procedures for classifying and determining the physical properties of grains and seeds are usually carried out manually and employ different equipment thus becoming laborious work and subjective. Therefore, the present work was carried out with the objective of developing an automated system for classifying samples of soybeans through digital images. The prototype for obtaining the digital images was made of black MDF and had an integrated Raspberry Pi microprocessor and a Pi Camera to capture the images. The camera was activated through a Python code, in which settings such as resolution, frame rate, brightness, contrast, and saturation were standardized in order to obtain uniform images. To determine the physical properties, images of whole grains were captured and through these, the width, length, projected area, and perimeter of 50 grains were analyzed through a computer application using Python programming language. The measurements for the grains obtained by the application were compared with those determined by a caliper. The Bland-Altman test was applied for the variables of width, length, and circularity. Satisfactory results were obtained for the length variable seeing that the method of obtaining measurements by digital images did not differ statistically from the caliper method making them equivalent. For the width and circularity variables, the digital method was not equivalent. However, most of the data are within the 95% confidence interval. For the variables of the projected area and perimeter, the agreement between the data was analyzed by means of an ANOVA followed by the Tukey test at 5% probability. The projected area variable did not differ significantly according to the methods used while the perimeter differed only for the method in which OpenCV library functions were used. For the classification of soybeans, 750 digital images of samples were acquired containing random combinations of the attributes of healthy (A1), burnt (A2), moldy(A3), burnt (A4), immature (A5), damaged (A), fermented (A7), broken and fragmented (A8), crushed (A9), hollow (A10), greenish (A12), and germinated grains(A12). A computational application was developed for image processing and feature extraction using OpenCV and NumPy libraries.

The data obtained were used to simulate artificial neural networks in seven performance simulations in order to select an efficient neural network capable of classifying soybeans. Tests were performed with different network configurations using the software Matlab and cross-validations were obtained using Orange Canvas. The highest accuracy was 80.01% found for the second simulation case using 60% of the data for training in an ANN with 30 neurons. For the validation of the ANN images of 10 samples containing the known attributes and quantity of soybeans were used. The images were submitted to the best network training found and obtained an average accuracy rate of 80.8%. Among the 12 attributes used in the classification of grains in the model developed these presented an accuracy rate of: A1 – 71.2%; A2 - 87.6%; A3 - 62.1%; A4 - 99.3%; A5 - 75.8%; A6 - 67.0%; A7 - 73.5%; A8 - 76.4%; A9 - 95.3%; A10 - 96.1%; A11 - 92.4% and A12 - 54.3%.

Keywords: Raspberry Pi. Automation. Python. Post-harvest. Physical properties.

LISTA DE ILUSTRAÇÕES

Figura 1. Raspberry Pi 4 Model B.....	24
Figura 2. Arquitetura MLP com duas camadas intermediárias.....	29
Figura 3. Fluxograma da metodologia utilizada no presente estudo.....	31
Figura 4. Protótipo para obtenção de imagens digitais de amostras de grãos de soja.....	32
Figura 5. Esferoide triaxial com suas respectivas dimensões características: (a) comprimento, (b) largura e (c) espessura.....	37
Figura 6. Parte 1 - fluxograma do algoritmo para o processamento da imagem.....	40
Figura 7. Parte II - fluxograma do algoritmo para segmentação e extração de características dos grãos de soja.....	41
Figura 8. Gráfico de Bland-Altman para as diferenças entre métodos de medição das variáveis de comprimento (A), largura (B) e circularidade (C).....	44
Figura 9. Validação cruzada com funções de ativação ReLU (a) e Tanh (b) para as Simulações S01 (A), S02 (B), S03 (C) e S04 (D).....	52
Figura 10. Validação cruzada para CS05 com função de ativação ReLU (a) e Tanh (b).	55
Figura 11. Validação cruzada para S06 com função de ativação ReLU (a) e Tanh (b).	56
Figura 12. Validação cruzada para S07 com função de ativação ReLU (a) e Tanh (b).	57

LISTA DE TABELAS

Tabela 1. Grupo I - destinada ao consumo humano in natura.	19
Tabela 2. Grupo II – destinada a outros usos (indústria e exportação).	20
Tabela 3. Resolução espacial para conversão de pixel em milímetros.	34
Tabela 4. Resumo da ANOVA para as diferenças entre métodos de medição.	45
Tabela 5. Resumo da ANOVA para as variáveis de área projetada e perímetro dos grãos de soja.	46
Tabela 6. Médias das áreas e perímetros para os respectivos métodos.	46
Tabela 7. Acurácia da RNA simulação 01.	47
Tabela 8. Acurácia da rede neural - Simulação 02.	48
Tabela 9. Dados do modelo ajustados para as variáveis de entrada e saída da RNA.	48
Tabela 10. Acurácia da rede neural para caso de simulação 03.	49
Tabela 11. Acurácia da rede neural para caso de simulação 04.	49
Tabela 12. Resultados das métricas para as funções de ativação ReLU (a) e Tanh (b) para as Simulações 01 a 04.	50
Tabela 13. Matriz de confusão - Simulação 05.	53
Tabela 14. Acurácia da rede neural para caso de simulação 05.	54
Tabela 15. Matriz de confusão desconsiderando o atributo G12 - Germinados.	54
Tabela 16. Matriz de confusão - Simulação 06.	56
Tabela 17. Matriz de confusão a - Simulação 07.	58
Tabela 18. Validação da RNA utilizando 10 amostras teste.	59
Tabela 19. Porcentagem de acerto das amostras de validação da rede neural artificial.	60

LISTAS DE SIGLAS E ABREVIATURAS

ANOVA	Análise de variância
CONAB	Companhia Nacional de Abastecimento
CPU	Central Processing Units
CV	Coeficiente de Variação
E.V.A.	Espuma vinílica acetinada
GB	Gigabytes
gl	Grau de Liberdade
GLP	General Public License
GPIO	General Purpose Input/Output
GPU	Graphics Processing Units
HDMI	High Definition Multimedia Interface
HSV	Hue, Saturation and Value
JPG	Joint Photographic Experts Group
LED	Light Emitting Diode
MAPA	Ministério da Agricultura Pecuária e Abastecimento
MDF	Medium Density Fiberboard
ML	Machine Learning
NASA	National Aeronautics and Space Administration
OpenCV	Open Source Computer Vision
OS	Operating System
PFS	Python Software Foundation
RAM	Random Access Memory
ReLU	Rectified Linear Unit
ROC	Receiver Operator Characteristic Curve
RGB	Red, Blue and Green
RNA	Redes Neurais Artificiais
SoC	System on Chip
SD	Secure Digital
Tanh	Hyperbolic Tangent Activation
USB	Universal Serial Bus

SUMÁRIO

1. INTRODUÇÃO.....	16
2. REVISÃO DE LITERATURA	18
2.1. Soja.....	18
2.1.1. Histórico	18
2.1.2. Classificação de grãos de soja	18
2.2. Propriedades físicas de produtos agrícolas.....	21
2.3. Linguagem de programação Python.....	21
2.4. Raspberry Pi	23
2.5. Processamento Digital de Imagens.....	24
2.6. <i>Machine Learning</i>	26
2.6.1. Aprendizado Supervisionado.....	27
2.6.2. Aprendizado Não Supervisionado.....	27
2.6.3. Aprendizado por reforço.....	28
2.6.4. Redes Neurais Artificiais.....	28
3. MATERIAIS E MÉTODOS.....	31
3.1. Matéria-prima	32
3.2. Protótipo para obtenção de imagens.....	32
3.3. Aquisição de imagens.....	33
3.4. Desenvolvimento dos algoritmos de processamento de imagens.....	34
3.4.1. Determinação e Cálculo Propriedades Físicas.....	34
3.4.2. Validação	36
3.4.2. Classificação dos grãos de soja	38
3.5.2. Validação do modelo de classificação.....	42
4. RESULTADOS E DISCUSSÃO	43
4.1. Propriedades Físicas	43

4.1.1. Largura, comprimento e circularidade.....	43
4.1.2. Área projetada e perímetro	45
4.2. Classificação de amostras de soja.....	46
4.2.1. Simulação 01 (S01)	47
4.2.2. Simulação 02 (S02)	47
4.2.3. Simulação 03 (S03)	48
4.2.4. Simulação 04 (S04)	49
4.2.6. Validação cruzada para as Simulações 01 a 04	50
4.2.5. Simulação 05 (S05)	52
4.2.7. Simulação 06 (S06)	55
4.2.8. Simulação 07 (S07)	57
4.2.9. Validação do modelo de classificação.....	58
5. CONCLUSÕES	62
5.1. Trabalhos Futuros	62
6. REFERÊNCIAS BIBLIOGRÁFICAS	63

1. INTRODUÇÃO

A cada ano, recordes de produção e produtividade de soja são batidos para atender a demanda crescente do mercado cada vez mais exigente quanto aos padrões de qualidade. Esses padrões são estabelecidos segundo legislações nacionais e internacionais e são adotados em transações comerciais. No Brasil, a Lei nº 9.972 de 25 de maio de 2000, regulamentada pelo Decreto nº 6.268 de 22 de novembro de 2007 normatizam a classificação dos produtos vegetais e de seus produtos e resíduos de valor econômico (BRASIL, 2000; BRASIL, 2007a).

Os procedimentos de classificação e determinação das propriedades físicas de grãos e sementes normalmente são manuais e empregam equipamentos como balanças, peneiras, balanças de peso hectolitrico, homogeneizadores, quarteadores e determinadores de umidade. Esses procedimentos são laboriosos e morosos, estão sujeitos a interpretações subjetivas por parte do classificador e impactam as tomadas de decisões quanto às rotinas de processamento e às práticas de comercialização.

Para agilizar a classificação de grãos com confiabilidade, pode-se empregar o Processamento Digital de Imagens (PDI) que consiste na obtenção, processamento e análise de imagens digitais empregando modelos matemáticos devidamente verificados e validados (JITANANI; CHIMLEK, 2019, YANG et al., 2015).

O uso de ferramentas computacionais e do processamento de imagens vem aumentando devido à redução de custos dos equipamentos, o crescente poder computacional dos *hardwares* e o interesse por soluções rápidas e não destrutivas (MAHAJAN et al., 2015). Complementando, Jitanani e Chimlek (2019) afirmam que o uso dessas tecnologias na caracterização e classificação de grãos e sementes se deve ao aprimoramento das técnicas de processamento de imagem e ao emprego da modelagem de redes neurais empregando *machine learning*.

Em conjunto com os métodos de processamento de imagens digitais, aplicação de técnicas da inteligência artificial e tecnologias como Redes Neurais Artificiais possibilita a automatização de processos como a classificação de grãos. Isso é possível devido a tarefas de reconhecimento de padrões que podem ser executadas implementando algoritmos de aprendizado supervisionado para o treinamento da RNA, em que é necessário um conjunto de exemplos previamente classificados (OSÓRIO; BITTENCOURT, 2000).

Apesar dos avanços quanto a classificação de grãos e sementes, o mercado ainda demanda dispositivos automatizados que agilizem com confiabilidade a classificação de amostras de grãos de soja conforme os regulamentos do Ministério da Agricultura, Pecuária e

Abastecimento. Desse modo, foi estabelecido como objetivo na condução desse trabalho o desenvolvimento e avaliação de um dispositivo computadorizado aplicado à classificação de amostras de soja utilizando imagens digitais e recursos de inteligência artificial.

2. REVISÃO DE LITERATURA

2.1. Soja

2.1.1. Histórico

A soja (*Glycine max*) é uma das culturas mais difundidas pelo mundo e de grande importância econômica. A palavra soja vem do japonês ‘shoyu’ e refere-se a uma leguminosa, pertencente à família fabaceae, que compreende o feijão, lentilha e ervilha (MELLO, BRUM; 2020). Originária da região de Manchúria, nordeste da China, o cultivo da soja é datado por volta de 2500 a.C., quando então a principal finalidade era alimentação da população local (ALVES, 2003).

Através dos séculos a soja foi difundida pelo mundo, chegando na Europa por volta do século XIII e apenas no século XIX foi introduzida nos Estados Unidos. O primeiro relato de implantação da cultura em território brasileiro foi em 1882 na Bahia, introduzida por Gustavo D’utra, mas devido às condições climáticas locais não houve boa adaptação (BONATO; BONATO, 1987). Em seguida foi implantado no estado de São Paulo, e em 1914 chegou ao Rio Grande do Sul onde teve melhor ambientação e começou a ser cultivada com finalidade de autoconsumo e alimentação animal (BLACK, 2000; BONETTI, 1981).

Além de ser um alimento versátil, o interesse econômico pela soja se deve também a sua composição química e valor nutritivo, em que se tem 40% de proteínas, 21% de óleo e 34% de carboidratos (TURZI, 2017; ALVES, 2003). Estudos indicam que a soja e derivados são excelentes fontes de compostos bioativos como ácidos fenólicos, flavonoides, isoflavonas, carotenóides e tocoferóis (DUEÑAS et al., 2012; KUMAR et al., 2010).

A soja é normalmente consumida *in natura* em países orientais e na forma desidratada em outros países, quando processada pode ser utilizada como extrato solúvel (leite de soja), molho shoyu, tofu, entre outros. Na industrialização da soja são obtidos derivados como farelo, óleo, tintas, vernizes, sabonetes e biodiesel. Desses derivados o farelo é utilizado principalmente na elaboração de ração animal, enquanto do processamento do óleo se obtém margarinas, óleos de cozinha e gorduras hidrogenadas (MELLO; BRUM, 2020).

2.1.2. Classificação de grãos de soja

A identificação e quantificação de defeitos e avarias em amostras de soja permite qualificar uma carga e estabelecer a finalidade de uso. Portanto, a classificação de grãos é

essencial para comercialização e industrialização, pois ao se determinar as características intrínsecas e extrínsecas do produto segundo padrões preestabelecidos, se tem maior controle da qualidade (PARIZZI, 2009). Durante a classificação é determinado o grupo, classe e tipo dos grãos avaliados. Esse processo deve ser realizado por um profissional qualificado e habilitado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Especificamente, no Brasil, a Lei nº 9.972 de 25 de maio de 2000, regulamentada pelo Decreto nº 6.268 de 22 de novembro de 2007 normatiza a classificação dos produtos vegetais e de seus produtos e resíduos de valor econômico (BRASIL, 2000; BRASIL, 2007a).

Conforme as Instruções Normativas, IN nº 11 de 15 de maio de 2007 e IN nº 37 de 27 de julho de 2007 (BRASIL, 2007b; BRASIL, 2007c), a classificação de soja se dá pela execução das seguintes etapas: amostragem, homogeneização, quarteamento, determinação de matérias estranhas e impurezas, determinação do teor de água, identificação de grupo/classe e tipo. Cada uma dessas etapas deve ser conduzida com rigor técnico e empregar equipamentos devidamente calibrados.

Os grãos de soja são classificados em dois grupos. O Grupo I destina-se ao consumo *in natura*, enquanto o Grupo II tem por finalidade a industrialização e/ou exportação. Segundo a quantificação dos percentuais de grãos avariados, esverdeados e o total de partidos, quebrados e amassados, são definidos no Grupo I os Tipos I ou II (Tabela 01) e para o Grupo II se os quantitativos percentuais forem iguais ou inferiores aos constantes na Tabela 02, o tipo corresponderá ao “padrão básico”.

Tabela 1. Grupo I - destinada ao consumo humano *in natura*.

Tipo	Avariados					Partidos, quebrados e amassados
	Total de ardidos e queimados	Máximo de queimados	Mofados	Total Avariados*	Esverdeados	
I	1,0	0,3	0,5	4,0	2,0	8,0
II	2,0	1,0	1,5	6,0	4,0	15,0

Tabela 2. Grupo II – destinada a outros usos (indústria e exportação).

Tipo	Avariados				Esverdeados	Partidos, quebrados e amassados
	Total de ardidos e queimados	Máximo de queimados	Mofados	Total Avariados*		
Padrão básico	4,0	1,0	6,0	8,0	8,0	30,0

* Soma de queimados, ardidos, mofados, fermentados, danificados, imaturos e cochos.

Fonte: Brasil (2007b)

Segundo a IN nº 11 de 15 de maio de 2007 entende-se por:

“

IV - avariados: grãos ou pedaços de grãos que se apresentam queimados, ardidos, mofados, fermentados, germinados, danificados, imaturos e chochos:

a) queimados: grãos ou pedaços de grãos carbonizados;

b) ardidos: grãos ou pedaços de grãos que se apresentam visivelmente fermentados em sua totalidade e com coloração marrom escura acentuada, afetando o cotilédone;

c) mofados: grãos ou pedaços de grãos que se apresentam com fungos (mofo ou bolor) visíveis a olho nu;

d) fermentados: grãos ou pedaços de grãos que, em razão do processo de fermentação, tenham sofrido alteração visível na cor do cotilédone que não aquela definida para os ardidos;

e) germinados: grãos ou pedaços de grãos que apresentam visivelmente a emissão da radícula;

f) danificados: grãos ou pedaços de grãos que se apresentam com manchas na polpa alterados e deformados, perfurados ou atacados por doenças ou insetos, em qualquer de suas fases evolutivas;

g) imaturos: grãos de formato oblongo, que se apresentam intensamente verdes, por não terem atingido seu desenvolvimento fisiológico completo e que podem se apresentar enrugados;

h) chochos: grãos com formato irregular que se apresentam enrugados, atrofiados e desprovidos de massa interna.

V - amassados: grãos que se apresentam esmagados, com os cotilédones e tegumento rompidos por danos mecânicos, estando excluídos deste defeito os grãos que se apresentam trincados em seu tegumento;

VI - partidos e quebrados: pedaços de grãos, inclusive cotilédones, que ficam retidos na peneira de crivos circulares de 3,0 mm (três milímetros) de diâmetro;

VII - *esverdeados: grãos ou pedaços de grãos com desenvolvimento fisiológico completo que apresentam coloração totalmente esverdeada no cotilédono (BRASIL, 2007b, Art. 2º).*

”

Os grãos de soja podem ainda ser classificados como amarelos, quando o tegumento possui coloração amarela, verde ou pérola, e misturada, quando não se enquadra na classe amarela.

2.2. Propriedades físicas de produtos agrícolas

O conhecimento das propriedades físicas de produtos agrícolas é essencial para condução de atividades como caracterização das dimensões e formatos dos grãos, descrição da cinética de secagem, definição de parâmetros para monitoração da qualidade, e dimensionamento de secadores, transportadores e equipamentos destinados a limpeza e, ou, classificação dos grãos (GONELI et al., 2011; RIBEIRO et al., 2002; BROOKER et al., 1992).

Na caracterização das dimensões e formatos dos grãos são conduzidos procedimentos que visam a mensuração das dimensões de comprimento, largura e espessura e o cálculo da circularidade, perímetro e área projetada.

2.3. Linguagem de programação Python

Python é uma linguagem de programação de alto nível de uso geral, orientada a objetos e de tipagem dinâmica, que dispõe de coleções de bibliotecas prontas para uso, além de ser possível a implantação de *frameworks* de terceiros. É uma linguagem de código aberto, com licença compatível com a *General Public License* (GPL), entretanto com menores restrições, o que permite que seja incorporada em produtos proprietários (BORGES, 2014). Desenvolvido no final de 1980 por Guido van Rossum no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda (CWI), o Python inicialmente tinha como objetivo auxiliar físicos e engenheiros.

Além de ser usada como linguagem principal, também é possível utilizar o Python como linguagem *script* em automatização de tarefas e adição de funcionalidades, também sendo possível integrar a outras linguagens como C e Fortran (BORGES, 2014). Os programas em Python são compatíveis a diferentes sistemas operacionais, o que torna possível mover os

programas de uma plataforma a outra. Tais características fizeram com que se tornasse uma linguagem de programação popular ao longo dos anos.

Em 2001 foi fundada a PSF – *Python Software Foundation*, organização sem fins lucrativos que coordena o uso da linguagem Python. Hoje, a PSF é apoiada por empresas com o Google, Microsoft, YouTube, NASA e a brasileira Globo Play. Isto tem contribuído para que o Python se tornasse uma das linguagens mais utilizada no mercado e já vir instalada de fábrica em sistemas operacionais como AmigaOS, MacOS e todas as distribuições do Linux (SILVA; SILVA, 2019).

A linguagem Python é comumente referida por desenvolvedores como uma linguagem que já vem com baterias inclusas (*batteries included*) devido ao vasto repertório de bibliotecas. Dentre a bibliotecas NumPy, SciPy, Pandas, OpenCV e Matplotlib são muito utilizadas no meio acadêmico por possuírem uma diversidade de funções (OLIPHANT, 2007). Visando realizar tarefas de processamento de dados e imagens, o presente trabalho teve Numpy e OpenCV como principais bibliotecas aplicadas.

A biblioteca NumPy é fundamental para a computação científica em Python. Abreviação de *Numerical Python*, é uma biblioteca de código aberto que suporta *arrays* e possibilita a realização de cálculos científicos, incluindo matemática lógica, manipulação de forma, classificação e seleção, operações estatísticas e álgebra linear básica, transformações discretas de Fourier, dentre outras funções (BRESSERT, 2012). Hoje, o Numpy é responsável por sustentar quase todas as bibliotecas utilizadas para computação científica ou numérica, como o caso do SciPy, Matplotlib, Pandas, Scikit-learn e scikit-image (HARRIS et al., 2020).

Open Source Computer Vision, ou simplesmente OpenCV, é uma biblioteca de código aberto desenvolvida para aplicações em visão computacional e *machine learning* (KAEHLER; BRADSKI, 2016). Um dos principais objetivos do OpenCV é fornecer uma infraestrutura de visão computacional simples e fácil de usar, possibilitando a criação de aplicativos sofisticados de maneira simplificada (KAEHLER; BRADSKI, 2016). A biblioteca OpenCV é bastante vasta e possui diversas funcionalidades aplicadas ao processamento de imagens digitais, como *cv2.imread*, ferramentas para leitura de imagem, *cv2.imshow* que é responsável por exibir a imagem em uma janela, redimensionamento de imagem através da ferramenta *cv2.resize*, através da *cv2.cvtColor* é possível transformar o espaço de cores da imagem, *cv2.medianBlur* aplica um filtro de desfoque médio, *cv2.threshold*: emprega um limite de nível fixo a cada elemento da matriz, *cv2.findContours* é responsável por encontrar o contornos de um objeto, por meio da ferramenta *cv2.drawContours* é possível desenhar os contornos de um objeto.

Muitas vezes, durante o processamento de imagens é preciso determinar medidas de um objeto. Sendo assim, a função `cv2.minAreaRect` encontra um retângulo da área mínima que circunda um objeto, enquanto `cv2.contourArea` determina a área do contorno, já `cv2.arcLength` determina o perímetro do contorno e `cv2.cv.BoxPoints(rect)`: encontra os quatro vértices de uma reta rotacionada (OPENCV, 2021).

2.4. Raspberry Pi

A Raspberry Pi é um Computador de Placa Única (SBC em inglês para *Single Board Computer*), de baixo custo desenvolvido no Reino Unido com intuito de aprimorar o ensino de informática, principalmente entre jovens.

O design da Raspberry é denominado *System on Chip* (SoC), em que uma única placa do tamanho de um cartão de crédito carrega todos os circuitos essenciais como Unidade Central de Processamento (CPU), Unidade de Processamento Gráfico (GPU) e demais entradas, saídas e circuitos de processamento, dentre outros, além da disponibilidade de recursos como os pinos *General Purpose Input/Output* (GPIO) que possibilita programação de *hardware*, condução de circuitos eletrônicos e coleta de dados (YAMANOOR; YAMANOOR, 2017; UPTON et al., 2016). A Raspberry Pi 4 (Figura 01) é a versão mais atual do dispositivo e vem com três opções de memória RAM: 2GB, 4GB e 8GB, conta com entradas USB 2, USB 3, USB-C e micro HDMI, entrada para ethernet, conexão *wireless*, *bluetooth*, além de entrada SD para cartão de memória (RASPBERRY PI, 2021). Por ser um pequeno computador, a Raspberry Pi suporta a instalação de um sistema operacional. O sistema operacional oficial com suporte é o Raspberry Pi OS, anteriormente chamado Raspbian, uma variante do Debian (RASPBERRY PI, 2021a).

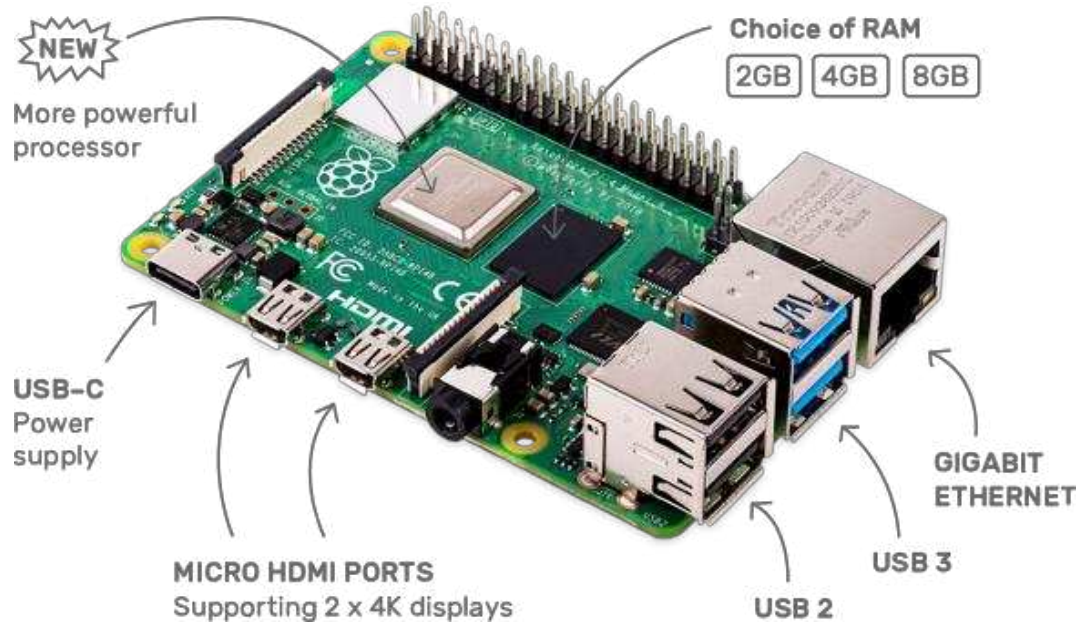


Figura 1. Raspberry Pi 4 Model B.

Fonte: Raspberry Pi Foundation, 2021.

2.5. Processamento Digital de Imagens

Uma imagem digital constitui na representação de uma imagem real empregando uma matriz bidimensional, em que cada elemento é definido como pixel “*picture elements*”, (KALAIVANI et al., 2013; GONZALES; WOODS, 2000). O pixel é o elemento básico de uma imagem e é caracterizado pelas cores primárias vermelho, verde e azul, e a proporção entre essas cores produz a intensidade do pixel (KALAIVANI et al., 2013).

Em geral, o processamento de imagens digitais envolve passos fundamentais para se atingir o objetivo como aquisição de imagens, pré-processamento, segmentação, reconhecimento e interpretação. O objetivo geral é produzir um resultado a partir do domínio do problema através do processamento de imagens. Inicia-se pela aquisição da imagem, ou seja, adquirir a imagem a ser analisada por meio de sensores de imageamento (câmeras, *scanners*, etc.). As imagens devem ser obtidas apropriadamente a fim evitar influências de fatores externos. No pré-processamento podem ser utilizadas técnicas de realce de contrastes e remoção de ruídos. Outro passo é a segmentação, uma das etapas mais importantes e laboriosas do processamento de imagens, consiste na divisão de uma imagem de entrada em partes ou objetos constituintes, simplificando ou alterando a representação da imagem com o intuito de facilitar a análise (COELHO et al., 2020; GONZALES; WOODS, 2000). Pode-se empregar diversos

métodos na segmentação de imagem, sendo os principais as técnicas com base no limiar, contorno da imagem ou regiões, agrupamentos de dados, métodos baseados em regiões e Redes Neurais Artificiais (KAUR; KAUR, 2014). As etapas de representação e descrição são parte da transformação dos dados iniciais em uma forma adequada para o processamento, visando a extração de características de uma imagem a fim de obter as informações, os dados iniciais são denominados descritores e contém propriedades que diferenciam as classes dos objetos. A etapa final do processamento de imagens é a de reconhecimento e interpretação, responsável por rotular os conjuntos de objetos conforme os descritores (COELHO et al., 2020; GONZALES; WOODS, 2000).

Segundo Gonzalez (2016) o Processamento Digital de Imagens (PDI) se constitui na manipulação de imagens empregando recursos computacionais de forma que a entrada e saída do processo são imagens ou características associadas a ela. Jitanani e Chimlek (2019) definem que o PDI consiste na obtenção, processamento e análise de imagens digitais empregando recursos computacionais que devem trazer embarcados modelos matemáticos devidamente verificados e avaliados. O PDI visa melhorar a informação visual para a interpretação humana (GONZALEZ, WOODS; 2000). Hoje, o PDI é considerado uma das principais áreas de pesquisa da engenharia e ciências da computação, e pode ser aplicado em diversas outras áreas como medicina, automação industrial, agricultura, dentre outros (COELHO et al., 2020).

O uso de imagens digitais com intuito de caracterizar e classificar grãos e sementes segundo padrões de qualidade tem sido crescente, o que segundo Jitanani e Chimlek (2019) se deve ao aprimoramento das técnicas de processamento de imagem e ao emprego da tecnologia *machine learning*. Esses autores desenvolveram um modelo para classificação de sementes segundo cinco padrões de qualidade, em que imagens obtidas no padrão RGB são analisadas segundo a configuração de histogramas quanto matiz, saturação e intensidade. O modelo desenvolvido apresentou níveis de acerto acima de 98,1%. Baek et al. (2020), empregando a linguagem de programação Python, desenvolveram um modelo para caracterização fenotípica de variedades de soja avaliando aspectos morfológicos e padrões de cores a partir de imagens digitais. O modelo apresentou uma correlação de 94,0%.

Effendi et al. (2019) criaram um sistema capaz de determinar a qualidade e variedades de grãos de milho utilizando o processamento digital de imagem, com base nas características de cores e textura. O sistema foi capaz de classificar os grãos de milho em dez categorias de qualidade com uma precisão de 90%, ao usar o valor de k (vizinho mais próximo) como cinco. Kuo et al. (2016) propuseram uma solução a partir do processamento de imagens para

caracterização fenotípica de trinta variedades de arroz ao considerar características morfológicas, cores e textura. O classificador alcançou uma precisão geral de 89,1%.

Yang et al. (2015) ao analisarem duas faces (lado germinativo e lado do endosperma) de três cultivares de milho ceroso através da obtenção de imagens em infravermelho visível e infravermelho próximo, obtiveram precisão de classificação de 98,2% e 96,3% para os lados germinativos e do endosperma, respectivamente, para as análises de cinco características morfológicas (área, circularidade, proporção, esfericidade e solidez) e oito recursos de textura (energia, contraste, correlação, entropia e seus desvios padrão).

2.6. *Machine Learning*

Machine learning ou aprendizado de máquina é um segmento da inteligência artificial que permite que máquinas e computadores sejam capazes de aprender e melhorar previsões ou comportamentos a partir de amostras de banco de dados fornecidas previamente e assim realizar tarefas sem que precisem ser explicitamente programados (MOLNAR, 2020; GOODFELLOW et al., 2016; SAMUEL, 1959).

Os sistemas de *machine learning* utilizam mecanismos de classificação como base de funcionamento, esses recebem informações de entrada no sistema e atribuem um nome para designar a classe pertencente (CARVALHO et al., 2014). Os mecanismos de classificação consistem na separação de indivíduos, objetos e fenômenos através de reconhecimento de padrões, ou seja, características em comum. Existem diversas estratégias que possibilitam a criação de máquinas capazes de aprender e melhorar por meio da observação, como aprendizado por hábito, instrução, dedução, analogia e indução (MITCHELL et al., 1997).

O aprendizado por indução é um dos mais utilizados em *machine learning*. A partir da indução é possível realizar uma inferência lógica e obter conclusões genéricas sobre um conjunto específico de dados. Mitchell et al. (1997) dizem que um programa computacional aprende por meio da experiência E , relacionada a uma classe de tarefas T e uma medida de performance P , se a performance nas tarefas em T , avaliadas através de P , aumentar com a experiência E . O aprendizado de máquina pode ser categorizado em: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço (MONARD; BARANAUSKAS, 2003).

2.6.1. Aprendizado Supervisionado

O aprendizado de máquina supervisionado é o tipo mais comum de aprendizado. No aprendizado de máquina supervisionado existe a interferência do usuário e cada exemplo no conjunto de dados é rotulado com o valor de saída (ou destino) esperado (KELLEHER, 2019).

O método de aprendizado supervisionado depende da existência de um conjunto de dados rotulados (conjunto de treinamento) que serão aplicados como referência do objetivo a ser alcançado, ou seja, como um treinamento com o conhecimento do ambiente (LORENA; CARVALHO, 2007). Nesse método é fornecido ao algoritmo de aprendizado (indutor) um conjunto de exemplos de treinamento para quais o rótulo da classe é conhecido e a partir dos dados rotulados o indutor realiza uma predição para os demais dados (não rotulados). Esse método parte da premissa que existe uma relação entre os dados e seus rótulos. O aprendizado supervisionado pode ser particionado em classificadores e regressores, sendo os classificadores aqueles que utilizam um conjunto finito de rótulos e geram rótulos de classe como saída e os regressores que geram valores contínuos, reais (MONARD; BARANAUSKAS, 2003).

O aprendizado supervisionado tem como principais vantagens a possibilidade de treinamento para tarefas específicas ao fornecer um conjunto de dados bem definidos, e rápido tempo de resposta para a inferência de um novo dado. Apesar de ser um dos métodos mais utilizados, nem sempre é indicado por ser muito sensível aos dados utilizados no treinamento. Se não houver um volume de dados adequados, ou se esses forem limitados, pode não ser possível realizar um treinamento que ofereça um nível satisfatório de generalização. Outro problema que pode ser encontrado ao utilizar o método supervisionado é de sobre ajuste (*overfitting*), que ocorre quando o modelo se ajusta demasiadamente ao conjunto de treinamento o que reduz a eficiência de classificação de novos dados (KUNCHEVA, 2014).

2.6.2. Aprendizado Não Supervisionado

No método de aprendizado não supervisionado, também conhecido como aprendizado por observação e descoberta, os dados utilizados não são rotulados e não necessitam de nenhuma interferência do usuário ou treinamento. Esse método tem como principal objetivo explorar os dados e encontrar estruturas ocultas neles. Esse tipo de método é amplamente aplicado para processos de redução de dimensionalidade, extração de informações, visualização de dados, detecção de valores discrepantes (*outliers*) (GE et al., 2017). Um exemplo típico de aprendizado não supervisionado é o agrupamento de dados, em que esses são atribuídos a

grupos de forma que cada grupo tenha propriedades em comum (CARLEO, 2019), esse processo de agrupamento de objetos em classes similares é chamado clusterização.

2.6.3. Aprendizado por reforço

Aprendizado por reforço (*Reinforcement Learning*) é uma técnica de aprendizado onde o agente ou sistema deve aprender a selecionar as ações disponíveis, que alteram o estado do ambiente e utilizam uma recompensa (reforço) para definir a qualidade da sequência de ações (MITCHELL, 1997). Nesse tipo de aprendizado, o agente interage com o meio que o cerca e aprende de maneira autônoma uma política ótima de atuação e passa a agir de acordo com o que é observado, quando a resposta obtida for a desejada ocorre a recompensa, caso contrário o algoritmo será punido, fazendo com que o mesmo busque sempre se ajustar melhor as situações de recompensa (SUTTON; BARTO, 2018; BRAGA et al., 2007). Para a Inteligência Artificial, o aprendizado por reforço representa técnicas computacionais para que agentes melhorem seu desempenho. O objetivo do agente é aprender uma política que maximize uma soma esperada de reforços a longo prazo. O aprendizado por reforço ocorre quando o sistema tem como objetivo não apenas processar as entradas disponíveis, mas também aprender as ‘regras do jogo’, baseado não na interação humana direta, mas nos dados acumulados das respostas ambientais fornecidas ao sistema (GESING et al., 2018).

2.6.4. Redes Neurais Artificiais

Redes Neurais Artificiais – RNA (*Artificial Neural Network*) são sistemas de aprendizado de máquina que tendem a simular o mecanismo de aprendizado do sistema nervoso de organismos biológicos. Assim como nos organismos vivos, o processamento de dados de uma RNA é realizado por unidades denominadas neurônios, conectados entre si por ligações sinápticas (AGGARWAL, 2018; LIN, 2017).

Uma RNA é composta, basicamente, pelas camadas de entrada, onde são fornecidos os dados iniciais do modelo; camada intermediária ou oculta, responsável pelo processamento de dados; e camada de saída, onde são obtidos os resultados processados (FURTADO, 2019). Cada entrada para um neurônio possui um peso, que modifica a função calculada nessa unidade. A rede neural mais simples é conhecida como *perceptron*, e contém apenas uma única camada de entrada e um nó de saída (AGGARWAL, 2018). O modelo *perceptron* de uma única camada só classifica padrões lineares separáveis, sendo necessário utilizar um Perceptron de

multicamadas (MLP). Uma rede MLP (Figura 02) é subdividida em camadas: camada de entrada, camada(s) intermediária(s) ou oculta(s) e camada de saída (FLECK, 2016).

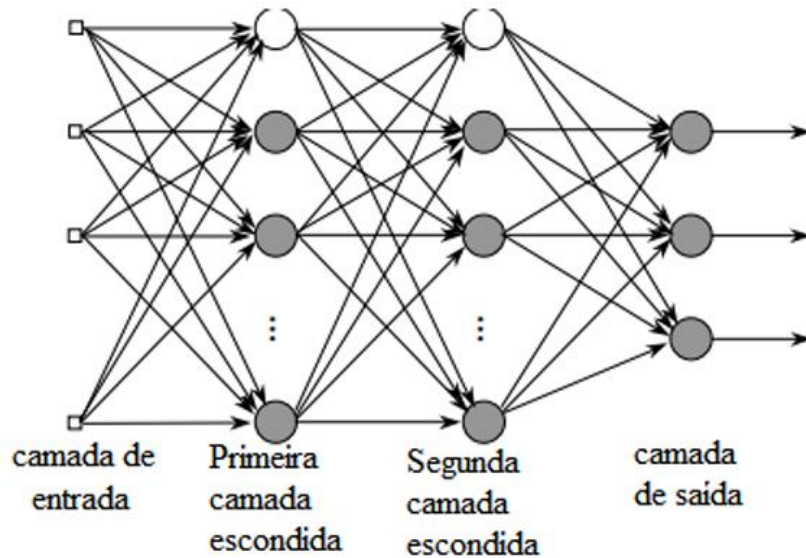


Figura 2. Arquitetura MLP com duas camadas intermediárias.

Fonte: Fleck, 2016.

O uso das RNAs para reconhecimento de padrões tem se mostrado cada vez mais atraente, eficazes e eficientes, capazes de modelar tarefas complexas. Destaca-se pelo processamento em alta velocidade, proporciona uma implementação paralela massiva, possuem excelentes propriedades de autoaprendizagem, adaptabilidade, tolerância a falhas, dentre outras vantagens (WANG et al., 2017; IZEMBOUDJEN et al., 2014). Tais características fazem com que o uso de RNAs sejam tão difundidos, sendo capazes de lidar com problemas do mundo real em áreas como agricultura, ciência, medicina, educação, finanças, gestão de negócios, segurança, comércio de commodities e arte (ABIODUN, 2018).

As RNAs aplicam equações matemáticas nas operações para o processamento de dados, entretanto essas operações são lineares. Para tornar uma rede capaz de modelar relações não lineares, aplica-se as funções de ativação. São ferramentas de aprendizado utilizadas pelas RNAs para calcular as informações de entrada recebida por cada nó, o processamento de cada neurônio de uma RNA é realizado por meio de funções de ativação. A função de ativação ReLu (Unidade Linear Retificada) é atualmente a função mais aplicada em todo o mundo. Inspirada em neurônios cerebrais que retornam como valor positivos ao zero (intervalo de 0 a $+\infty$), essa função tem se mostrado a mais eficiente na aplicação em redes neurais convolucionais ou

aprendizado profundo e tem como principal vantagem a não ativação de todos neurônios ao mesmo tempo, fazendo com que a rede se torne mais eficiente. A função Tanh (Tangente Hiperbólica), se assemelha a um sigmoide, em formato de ‘S’, com intervalo variando de -1 a 1. A função Tanh se destaca em problemas de separação com valores neutros, fortemente positivos e fortemente negativos (RIZZO, 2020).

Para avaliar o desempenho dos modelos de aprendizado de máquina, são utilizadas métricas, que variam de acordo com o modelo utilizado, regressão ou classificação (DATAIKU, 2021). Para os modelos de regressão, observa-se o erro quadrado médio e o R-quadrado (R^2). No modelo de classificação, como utilizado para o presente estudo, as métricas analisadas são: (i) Área sob a curva - AUC: corresponde a área abaixo da curva ROC (*Receiver Operator Characteristic Curve* – refere-se a taxa de verdadeiros positivos dada pela sensibilidade e a taxa de falsos positivos equivalente à especificidade) e indica o desempenho dos modelos principalmente para casos de classes balanceadas; (ii) Acurácia – CA: taxa geral certo, ou seja, proporção de previsões corretas em relação ao conjunto de dados; (iii) F-score – F1: média harmônica entre precisão e sensibilidade; (iv) Precisão – P: porcentagem de padrões classificados corretamente como pertencentes à classe positiva; (v) Sensibilidade - Recall: proporção de padrões de classes positivas identificados corretamente e indica o quão bom o classificador é na identificação de classes positivas.

3. MATERIAIS E MÉTODOS

O presente trabalho foi desenvolvido no Laboratório de Propriedades Físicas e Avaliação de Qualidade de Produtos Agrícolas, do Centro Nacional de Treinamento e Armazenagem (CENTREINAR), localizado no Campus Viçosa da Universidade Federal de Viçosa (UFV), Viçosa – MG.

Desenvolveu-se um algoritmo capaz de classificar e determinar propriedades físicas de grãos genéricos. Para o presente trabalho, utilizou-se grãos de soja como estudo de caso com o objetivo de validar a metodologia aplicada. A Figura 03 representa a metodologia utilizada para a realização deste trabalho.

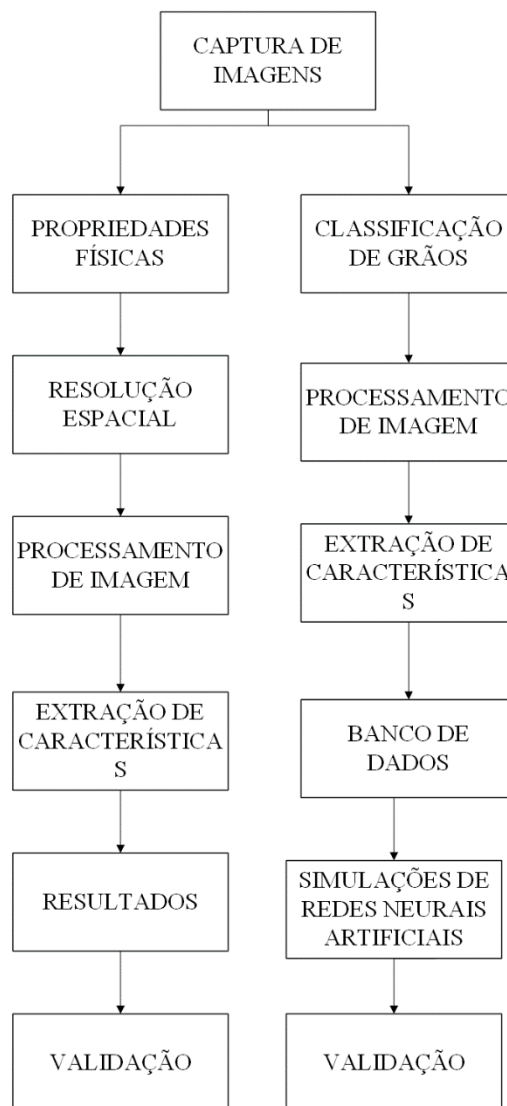


Figura 3. Fluxograma da metodologia utilizada no presente estudo.

3.1. Matéria-prima

Foram utilizadas amostras de soja “*in natura*” oriundas de unidades armazenadoras localizadas na Região Centro-Oeste do Brasil. Para o envio, as amostras, classificadas por profissionais habilitados segundo as Instruções Normativas IN n.º 11 de 15 de maio de 2007 (BRASIL, 2007b) e IN n.º 37 de 27 de julho de 2007 (BRASIL, 2007c), foram embaladas em sacos plásticos. Na embalagem de cada amostra os grãos e defeitos foram acondicionados separadamente.

3.2. Protótipo para obtenção de imagens

Para obtenção das imagens digitais das amostras foi confeccionado um protótipo em MDF (*Medium Density Fiberboard*) na cor preta com as dimensões comprimento, altura e largura iguais a 0,20 m (Figura 04). Na face superior foi inserido um orifício com seção quadrática de 2,5 x 2,5 cm para fixar uma câmera module V2 NoIR com resolução de 8 megapixel e junto a ela instalou-se o microprocessador Raspberry Pi 4 Model B com 4 GB de memória RAM. Internamente, para padronizar a intensidade de iluminação, foram instaladas fita de luzes de LED brancas nas laterais e parte superior do compartimento.



Figura 4. Protótipo para obtenção de imagens digitais de amostras de grãos de soja.

Fonte: A autora (2021).

Na parte inferior do protótipo se tem uma gaveta, onde as amostras de grãos de soja são acondicionadas para o registro das imagens, formando assim o campo de análise, com comprimento e largura de 15,0 e 10,0 cm, respectivamente. O campo de análise foi revestido com espuma vinílica acetinada (E.V.A.) na cor azul, com os propósitos de evitar a movimentação dos grãos durante a abertura e fechamento da gaveta, reduzir o reflexo causado pelo sistema de iluminação interno e facilitar a segmentação das imagens (MEDEIROS, 2020).

3.3. Aquisição de imagens

Para o acionamento da câmera, foi estruturado um código em Python com a inclusão dos módulos “*picamera*” e “*time*”. Nesse código, foram definidos os seguintes parâmetros de configuração que melhor se adaptaram ao ambiente de captura, sendo: rotação (180), resolução (3280 x 2464 pixels), framerate (30 frames por segundo), brilho (45), nitidez (100), saturação (10), contraste (20) e iso = 10. A definição desses parâmetros visou a obtenção imagens uniformes, para facilitar o processo de segmentação. O código abaixo descreve os parâmetros adotados para a captura de imagens:

```
#CAPTURA DE IMAGENS
from picamera import PiCamera
from time import sleep

camera = PiCamera()

#CONFIGURAÇÃO DE CÂMERA
camera.rotation = 180
camera.resolution = (3280,2464)
camera.framerate = 30
camera.brightness = 45
camera.sharpness = 100
camera.saturation = 10
camera.contrast = 20
camera.iso = 10

#Acionamento e captura
camera.start_preview()
```

```

sleep(15)
camera.capture('/home/pi/Documents/imagens/img.jpg')
camera.stop_preview()

```

Para a determinação da resolução espacial das imagens foi utilizada uma moeda com dimensões conhecidas, Tabela 03, que teve imagens obtidas sob as mesmas condições empregadas para capturar as imagens dos grãos de soja. Assim, pela comparação dos valores de comprimento, largura, área e perímetro obtidos no dispositivo com aqueles encontrados por meio de paquímetro e os cálculos apropriados foram determinados os fatores de conversão.

Tabela 3. Resolução espacial para conversão de pixel em milímetros.

Resolução espacial				
	Comprimento	Largura	Área	Perímetro
Objeto Real	24,91 mm	24,89 mm	486,82 mm ²	78,22 mm
Imagem	111,55 pixels	113,14 pixels	10.171,00 pixels ²	377,50 pixels
Fator de conversão	0,22	0,22	0,0484	0,22

3.4. Desenvolvimento dos algoritmos de processamento de imagens

Para implementação do algoritmo foi empregada a linguagem de programação Python versão 3.8 utilizando o ambiente de desenvolvimento Jupyter Notebook™, como também foram empregadas as bibliotecas OpenCV e Numpy.

3.4.1. Determinação e Cálculo Propriedades Físicas

Imagens dos grãos inteiros, posicionados no campo de análise, foram capturadas no padrão RGB (*Red, Green and Blue*) e em sequência transformadas para o padrão HSV (*Hue, Saturation, Value*) para facilitar a segmentação (KOSCHAN e ABIDI, 2008). Procedendo assim, é reduzida relação cor e luminância, o que leva a menores influências de áreas

sombreadas ou com maior incidência de luz. Consequentemente, se tem maior precisão na determinação das dimensões e contornos dos grãos.

Durante o processo de segmentação as bandas H e V foram somadas para facilitar a binarização, para tanto, foi empregado método Otsu Threshold disponível na biblioteca OpenCV. A etapa de binarização consiste na conversão das imagens em duas cores, preto (digito 0) e branco (digito 1). O fundo da imagem (campo de análise) foi classificado como 1 (branco) e os grãos como 0 (preto). Desse modo, foram obtidas as dimensões e formato dos grãos (GRANITO et al., 2002). Anterior ao processo de binarização foi aplicado um filtro de desfoque médio (*medianblur*) com os propósitos de suavizar os ruídos nas imagens e minimizar problemas decorrentes de sombras, sujidades e, ou demais falhas.

A função *findContours* disponível no módulo OpenCV foi aplicada em *loop* com objetivo de localizar, delimitar e contabilizar o contorno das imagens capturadas de cada grão. Foram empregadas as funções *cv2.contourArea*, *cv2.arcLength* e *cv2.boxPoints(rect)* para determinar a área projetada, perímetro e as dimensões largura e comprimento de cada grão, respectivamente. E em função da área e raio encontrados pelas funções do OpenCV, foram calculados a circularidade (C) por meio da Equação 1.

$$C = \frac{4 \cdot \text{área}}{\pi \cdot (2 \cdot \text{raio})^2} \quad (1)$$

Em que,

C – circularidade, %

área – área do contorno, mm²

raio – raio do contorno, mm

O algoritmo computacional determina ainda o maior eixo de comprimento e largura do contorno de cada grão. Para fins comparativos, calculou-se a área e perímetro por meio dos eixos de comprimento (a) e largura (b) encontrados pelo algoritmo computacional por meio das equações 2 e 3, respectivamente (WU et al., 2018; MOSHENIN, 1986; SAKAI; YONCKAWN, 1991).

$$A_p = \frac{ab\pi}{4} \quad (2)$$

$$P = \frac{2\pi(a+b)}{4} \quad (3)$$

Em que,

a – dimensão do maior eixo (comprimento), mm;

b – dimensão do eixo médio (largura), mm;

A_p – área projetada, mm²;

P – perímetro, mm.

3.4.2. Validação

Para validação dos procedimentos para determinação das dimensões e cálculos de circularidade, área projetada e perímetro, estes valores foram comparados aos obtidos por meio de rotinas laboratoriais (WU et al., 2018; MOSHENIN, 1986; SAKAI; YONCKAWN, 1991).

Foram obtidas as dimensões de comprimento (a) e largura (b) para 50 grãos de soja através de um paquímetro Mitutoyo, com precisão de 0,01 mm, por medições dos eixos ortogonais de comprimento (a) e largura (b) conforme indicado na Figura 05. Cada eixo foi medido em triplicata e se calculou a média com o intuito de reduzir o erro do operador. Por meio dos eixos mensurados foi possível calcular a circularidade (C), área projetada (A_p) e perímetro (P) por meio das equações citadas no item 3.4.1.

Os mesmos grãos foram submetidos ao aplicativo computacional, onde foi possível obter suas dimensões características pelo processamento digital de imagens. A validação se deu por meio da comparação dos resultados encontrados por meio do processamento das imagens digitais e aqueles obtidos a partir de medidas realizado com o uso do paquímetro.

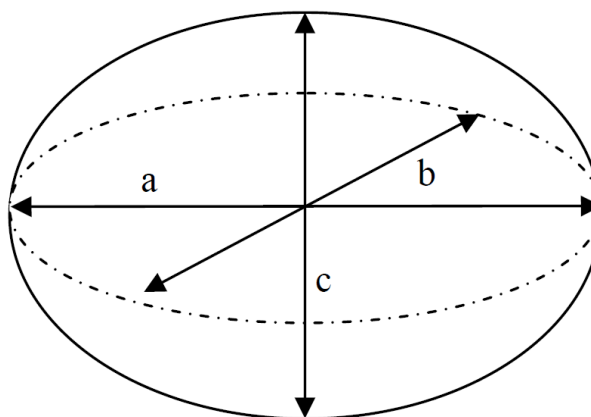


Figura 5. Esferoide triaxial com suas respectivas dimensões características: (a) comprimento, (b) largura e (c) espessura.

Fonte: Corrêa e Goneli (2008).

Utilizando o programa IBM SPSS Statistic (IBM, 2021), empregou-se o teste de Bland-Altman (BLAND; ALTMAN, 1995) com o objetivo de analisar a relação entre os métodos de obtenção de dimensões características obtidas por (i) paquímetro e (ii) dimensões por aplicativo para as variáveis de largura (a), comprimento (b) e circularidade (C). Através do teste de Bland-Altman é possível verificar a dispersão entre as diferenças de duas variáveis e as médias obtidas entre os métodos utilizados. Atualmente é o teste mais indicado para a análise de concordância entre métodos (DOGAN, 2018; HIRAKATA, 2009). Primeiro realizou-se o teste T para avaliar se as diferenças entre os métodos são significativamente diferentes de zero, em seguida foi plotado o gráfico de dispersão em que foi possível analisar a variabilidade dos dados em relação à média das diferenças e entre os limites de confiança de 95%, calculado empregando a Equação 4. Por fim, realizou-se uma regressão linear simples para avaliar a existência de viés de proporção, o que indica a ocorrência de tendência de variação dos dados abaixo ou acima da média.

$$\begin{aligned} LS &= \bar{X} + 1,96 SD \\ LI &= \bar{X} - 1,96 SD \end{aligned} \quad (4)$$

Em que:

LS – limite de superior do intervalo de confiança;

LI – limite de inferior do intervalo de confiança;

\bar{X} – média;

SD – desvio padrão da média das diferenças

As variáveis de área projetada e perímetro foram calculadas segundo três métodos em que as dimensões dos grãos foram obtidas por meio do (i) Paquímetro, (ii) Aplicativo computacional e (iii) Funções da biblioteca OpenCV (*cv2.contourArea* para área e *cv2.arcLength* para perímetro). Na avaliação dos métodos foram realizadas análise de variância (ANOVA) e a comparação das médias segundo teste de Tukey a 5% de significância.

3.4.2. Classificação dos grãos de soja

O modelo computacional desenvolvido compreende o processamento e análise das imagens digitais de amostras de grãos de soja, com o intuito de classificar amostras de grãos de soja segundo os preceitos estabelecidos nas Instruções Normativas IN n.º 11 de 15 de maio de 2007 (BRASIL, 2007b) e IN n.º 37 de 27 de julho de 2007 (BRASIL, 2007c).

Para o desenvolvimento do modelo foram obtidas 750 imagens digitais de amostras contendo combinações aleatórias de grãos sadios, ardidos, mofados, queimados, imaturos, danificados, fermentados, partidos e quebrados, amassados, chochos, esverdeados e germinados. Nas Figura 06 e 07 são apresentados os fluxogramas dos algoritmos empregados para o desenvolvimento do modelo computacional.

Na Figura 06, tem-se o fluxograma relacionado ao processamento da imagem, que tem por início leitura da imagem. Em sequência se procede recorte das regiões de interesse, para a retirada das bordas do campo de análise. Para adequar a visualização das imagens recortadas, procedeu à redução das dimensões das imagens em 30%.

Para aprimorar a segmentação das imagens, o processo de binarização foi conduzido segundo a quatro categorias de amostra: 1 - Grãos claros, 2 – Grãos escuros, 3 – Prevalência de grãos claros e 4 – Prevalência de grãos escuros. Para categoria 1 (Grãos claros) a binarização se deu pela condução das seguintes etapas: o OpenCV lê imagens no modelo BGR (*blue, green, red*) sendo assim, realizou-se a conversão das imagens em BGR para HSV, seleção da banda S e aplicação do filtro mediana. Quanto a categoria 2 (Grãos escuros) as etapas foram: seleção da banda B das imagens em RGB e aplicação do filtro mediana. Ao que se refere à categoria 3

(Prevalência de grãos claros) se procedeu: conversão das imagens em BGR para HSV, soma das bandas S e V e aplicação do filtro mediana. Por fim, na binarização aplicada a categoria 4 (Prevalência de grãos escuros) se tem as seguintes etapas: seleciona-se separadamente as bandas B e G, aplica-se o filtro mediana e assim é obtidas duas imagens IMG-1 e IMG-2, que são comparadas pixel a pixel, para o caso de os pixels serem diferente na IMG-2 esse pixel receberá o valor 0 e caso diferente de zero permanecerá com o padrão de cor original.

Na Figura 07 é apresentada a segunda parte do fluxograma do algoritmo para segmentação e extração de características das amostras. Como representado, a partir de uma imagem processada procede-se as varreduras vertical e horizontal para localizar os grãos. Para isso, era necessário informar o número de linhas e colunas de grãos presentes na amostra a ser analisada.

Dos procedimentos de segmentação, são disponibilizados os valores de quatro parâmetros para cada grão ou defeito na amostra: largura, comprimento, área de projeção e perímetro. E quanto a extração de caracterização, para cada imagem em padrão RBG, foram compilados nove valores referentes aos valores mínimo, médio e máximo nas bandas R, G e B. Ao final os treze parâmetros para cada grão ou defeito foram transcritos em um banco de dados configurado como um arquivo com extensão .xlsx contendo uma planilha. Esses dados foram normalizados em intervalo de 0 a 1 para evitar discrepâncias e diminuir a redundância do banco de dados, por se tratar de atributos com diferentes escalas numéricas.

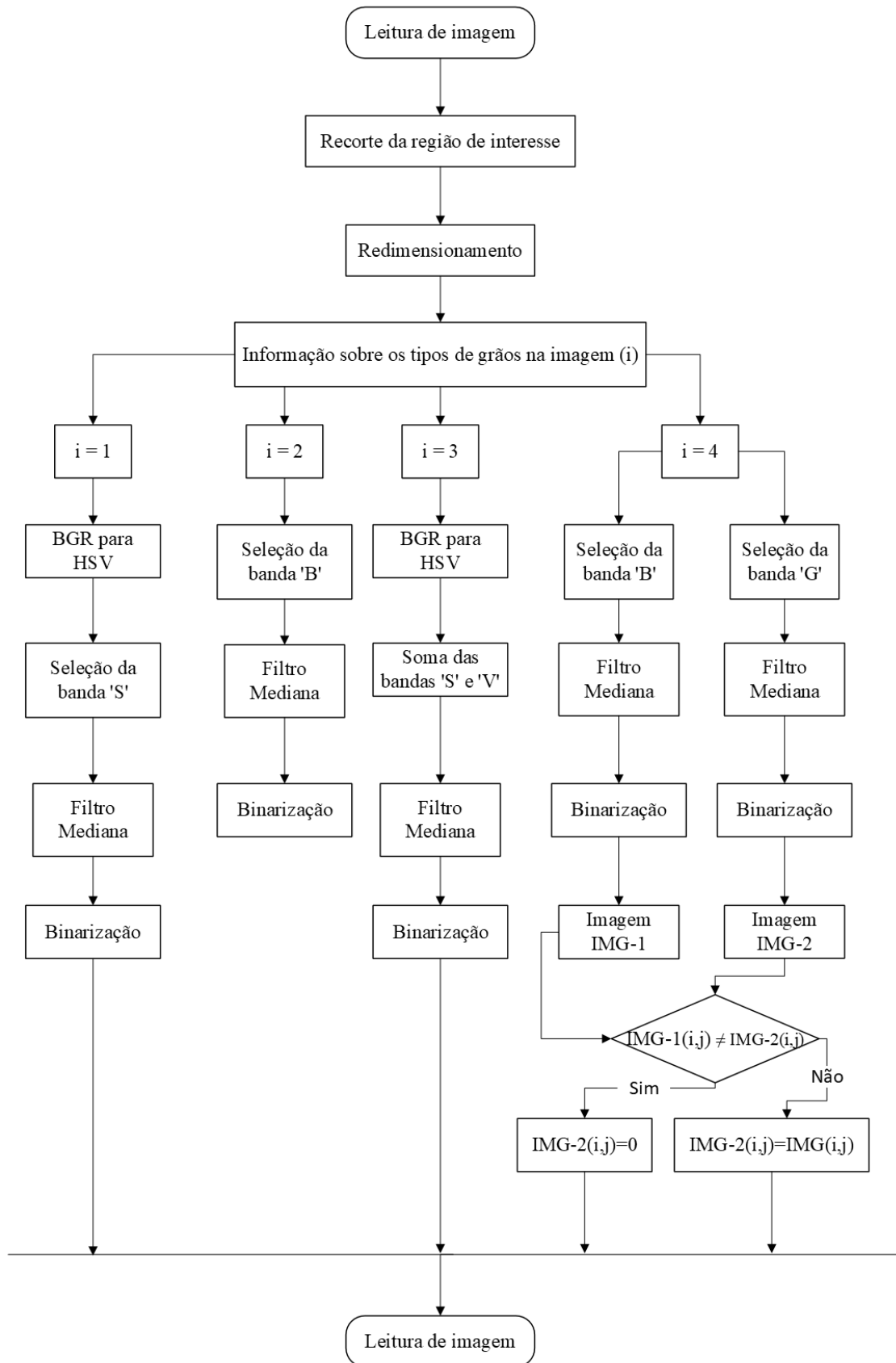


Figura 6. Parte 1 - fluxograma do algoritmo para o processamento da imagem.

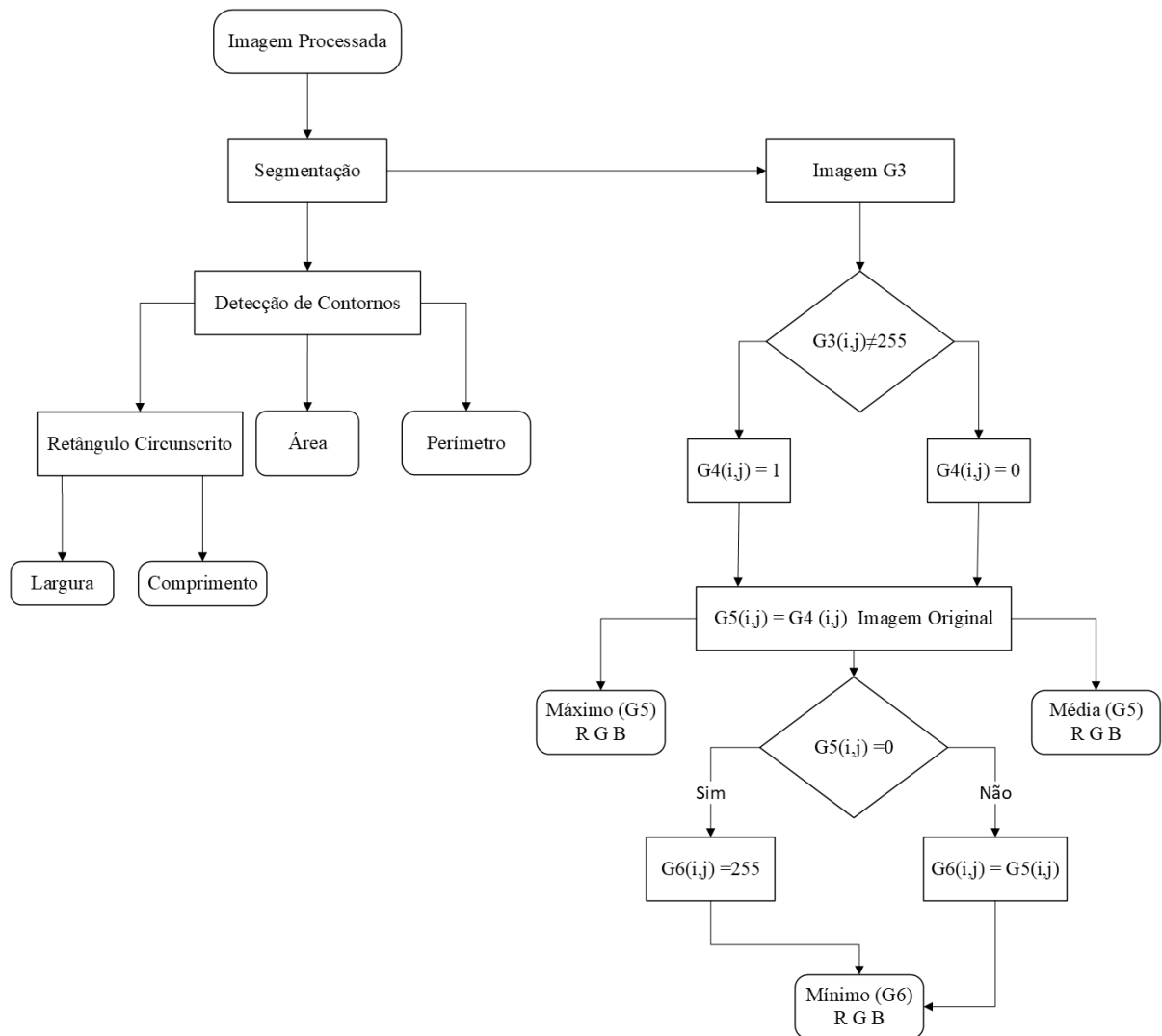


Figura 7. Parte II - fluxograma do algoritmo para segmentação e extração de características dos grãos de soja.

Utilizando o banco de dados e recursos dos programas MatlabTM para treinamento de redes neurais e o Orange CanvasTM na realização de matrizes de confusões, procedeu-se simulações das redes neurais artificiais – RNA, em que foram ponderadas as percentagens de dados empregados no treinamento e conjuntos de números de neurônios. O uso dessas ferramentas se justifica devido a facilidade oferecida por este para a resolução de problemas computacionais, sendo uma ferramenta bastante intuitiva e com uma diversidade em *toolboxes* capazes de simplificar a fase de simulações de redes neurais.

A ferramenta *nprtool* do Matlab, utilizada para reconhecimento de padrões, foi aplicada na realização dos testes da RNA em diferentes configurações. O método de treinamento adotado foi baseado no algoritmo de gradiente conjugado (CHARALAMBOUS, 1992). Foram realizadas validações cruzadas a fim de verificar o comportamento e nível de assertividade dos atributos para cada simulação de redes neurais. Para elaboração da avaliação cruzada foi utilizado o programa Orange Canvas™, empregando as funções de ativação linear retificada (ReLU) e tangente hiperbólica (Tanh). O método de treinamento adotado foi o Adam, que se trata de um algoritmo de otimização que calcula taxa de aprendizagem adaptativa para cada parâmetro (KINGMA; BA, 2014).

Como os procedimentos na classificação de amostras de grãos envolvem procedimentos de reconhecimento e especificações de padrões em *machine learning* foi aplicado o método de aprendizagem supervisionada, em que se adotou os seguintes códigos para descrição dos atributos dos grãos: 1 - Sadios, 2 - Ardidos, 3 - Mofados, 4 - Queimados, 5 - Imaturos, 6 - Danificados, 7 - Fermentados, 8 - Partidos e quebrados, 9 - Amassados, 10 - Chochos, 11 - Esverdeados e 12 - Germinados.

As métricas utilizadas nas avaliações das simulações das RNAs foram (i) Área sob a curva – AUC; (ii) Acurácia – CA; (iii) F-score – F1; (iv) Precisão – P e (v) Sensibilidade - Recall. A melhor configuração da RNA foi obtida a partir da variação do número de neurônios e porcentagem de treinamento por meio da análise estatísticas de uma regressão linear múltipla.

3.5.2. Validação do modelo de classificação

A validação do modelo computacional se deu pela comparação da classificação de grãos de soja realizada por meio do sistema desenvolvido e a realizada pelos classificadores devidamente habilitados.

Foram analisadas dez amostras contendo grãos com atributos e quantidades conhecidas de grãos de soja. A imagem de cada amostra foi obtida e segmentada, gerando um banco de dados salvos em arquivo no formato .xlsx. Esse banco de dados foi analisado por uma RNA previamente treinada para o caso de simulação 02 (S02). Com isso, foi gerada uma matriz de confusão para cada amostra, reportando as quantidades de grãos por atributo e as porcentagens de erro e acerto.

4. RESULTADOS E DISCUSSÃO

4.1. Propriedades Físicas

4.1.1. Largura, comprimento e circularidade

Ao empregar o teste T ao nível de significância de 5% ($p = 0,05$), foram avaliados métodos aplicados a determinação da largura, comprimento e circularidade, que se referem ao uso do (i) Paquímetro e (ii) Aplicativo computacional. Conforme as análises estatísticas, entre métodos empregados para determinação do comprimento (a) dos grãos de soja, obteve-se p-valor de 0,230, ou seja, maior que 0,05; portanto não há diferença significativa entre os métodos empregados. Quanto a determinação da largura (b) o p-valor $< 0,001$, indicando que há diferença estatística entre os métodos empregados. Como a circularidade é a razão entre largura e comprimento (b/a) foi constatado o mesmo comportamento ao obtido para b, com teste T indicando a não correlação entre os métodos para essa variável (p-valor $< 0,001$).

Na Figura 08 estão apresentados os gráficos de Bland-Altman, quanto a diferenças entre métodos de medição de comprimento (a), largura (b) e circularidade (C). Observa-se para o comprimento (a) a ocorrência de maior acurácia para os valores acima da média, enquanto para os abaixo da média se tem menor dispersão. Dos 50 valores calculados, observou-se um *outlier* ao se considerar o intervalo de confiança de 95% (Figura 08A). O gráfico de Bland-Altman para a variável de largura (Figura 08B) apresenta maior dispersão dos dados e presença de três *outliers*. Observa-se que para essa variável o aplicativo computacional superestimou os valores dos dados, resultando em mais diferenças negativas entre os métodos. A diferença média foi de $-0,10 \text{ mm} \pm 0,11$ com diferença máxima de $-0,31 \text{ mm}$, sendo assim é importante ressaltar que as diferenças encontradas entre as metodologias podem ser referentes a erros aleatórios ou não aleatórios ocasionados na aquisição das dimensões por paquímetro tanto quanto durante o processamento de imagens digitais (ARAUJO et al., 2020). A Figura 08C traz a representação gráfica de Bland-Altman para a variável de circularidade (C), em que é possível observar o mesmo comportamento observado para a variável de largura e assim como a presença de *outliers*.

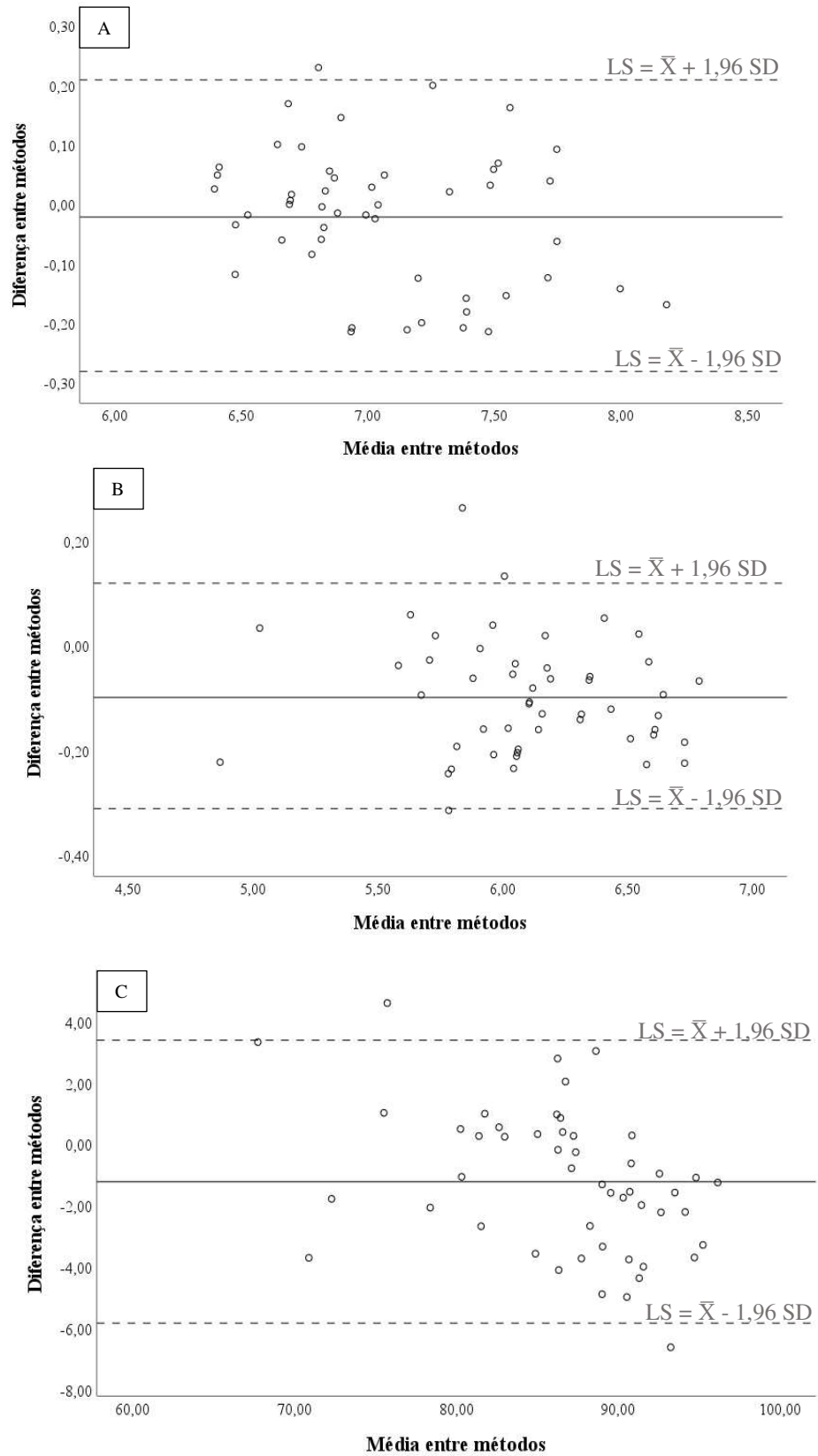


Figura 8. Gráfico de Bland-Altman para as diferenças entre métodos de medição das variáveis de comprimento (A), largura (B) e circularidade (C).

Os resultados da ANOVA para as diferenças entre os métodos de determinação estão indicados na Tabela 04. Os métodos para determinação do comprimento não apresentam diferença significativa entre os dados ($F \geq 0,05$), sendo assim pode-se dizer que não há tendências em subestimar ou superestimar valores, desse modo os métodos são concordantes e equivalentes. Para a largura, apesar dos métodos serem estatisticamente diferentes, a ANOVA indica que não há um viés, sendo assim não há tendências em subestimar ou superestimar valores em relação à média. A análise de variância para circularidade indicou um viés de proporção demonstrando uma tendência em superestimar os dados.

Tabela 4. Resumo da ANOVA para as diferenças entre métodos de medição.

FV	gl	Soma dos quadrados	Quadrado médio	F
Comprimento (a)				
Regressão	1	0,048	0,048	0,057
Resíduo	48	0,612	0,013	
Total	49	0,660		
Largura (b)				
Regressão	1	0,008	0,008	0,425
Resíduo	48	0,586	0,012	
Total	49	0,594		
Circularidade (C)				
Regressão	1	45,72	45,72	0,003
Resíduo	48	226,12	4,71	
Total	49	271,84		

4.1.2. Área projetada e perímetro

Na Tabela 05 é apresentado o resumo das Análise de Variância (ANOVA) e o Coeficiente de Variação (CV) para as variáveis de área projetada e perímetro segundo parâmetros determinados empregando: (i) medições por paquímetro, (ii) dimensões características por aplicativo e (iii) funções do OpenCV. Observa-se que apenas para a variável

de perímetro obteve-se F significativo ao nível de 1% de probabilidade, indicando que pelo menos uma das médias é estatisticamente diferente.

As médias obtidas para cada método estão indicadas na Tabela 06. Observa-se que para ambas as variáveis, os métodos que utilizam as funções do OpenCV obtiveram médias superiores aos dos demais métodos. Hariadi et al., (2018) ao estudarem o uso do OpenCV para determinação da área foliar em campo por imagens digitais observaram que a resolução da imagem afeta diretamente a precisão das medidas observadas. Sendo assim, um dos possíveis motivos para as diferenças encontradas entre as médias pode estar relacionado a baixa resolução do sensor de captura da Pi Câmera utilizada nesse estudo.

Tabela 5. Resumo da ANOVA para as variáveis de área projetada e perímetro dos grãos de soja.

FV	gl	Quadrado médio	
		Área	Perímetro
Métodos	2	22,61	32,79**
Resíduo	147	11,89	1,17
CV (%)	-	10,34	5,11

** F significativo a 1% de probabilidade

Tabela 6. Médias das áreas e perímetros para os respectivos métodos.

Métodos	Área	Perímetro
Paquímetro	32,51 a	20,60 a
Dimensões por aplicativo	33,61 a	20,79 a
Função do OpenCV	33,73 a	22,09 b

*Médias seguidas de pela mesma letra na coluna não diferem estatisticamente ao nível de 5% de probabilidade pelo teste Tukey.

4.2. Classificação de amostras de soja

Empregados banco de dados gerados pela extração das características dos grãos de soja por meio do processamento de imagens, foram avaliadas sete simulações com intuito de

selecionar a configuração da rede neural que com maior eficácia diferencia os atributos para classificação dos grãos de soja. Os testes foram realizados com diferentes configurações de rede por meio do Matlab e as validações cruzadas foram obtidas pelo Orange Canvas.

4.2.1. Simulação 01 (S01)

Para simulação 01, utilizou-se como entrada o banco de dados obtidos pelo processamento de imagens. Nas simulações, Tabela 07, foram procedidas 45 combinações ao se considerar cinco percentagens de treinamento e nove conjuntos de neurônios. Conforme os resultados a máxima acurácia foi 77,04% o que se deu para porcentagem de treinamento igual a 70% e número de neurônios 50. Nesta simulação apurou-se a média geral de acurácia igual a 75,26% e desvio padrão de 0,99%.

Tabela 7. Acurácia da RNA simulação 01.

Treinamento (%)	Conjuntos de número de neurônios								
	10	20	30	50	100	150	120	300	500
30	72,99	74,49	75,31	75,53	75,35	75,34	75,05	73,88	74,11
40	73,20	75,32	75,62	75,43	75,80	75,84	74,84	74,95	75,05
50	73,60	75,52	75,76	76,15	76,20	76,88	75,69	74,91	76,32
60	73,29	75,07	76,60	76,94	75,94	75,33	75,85	75,20	75,95
70	73,41	75,58	75,85	77,04	76,32	75,70	74,99	74,66	73,96

4.2.2. Simulação 02 (S02)

Para a simulação 02, optou-se pela remoção de discrepâncias (*outliers*) do banco de dados. Removeu-se 10% dos dados que apresentavam *outliers* superiores à média e 10% dos dados inferiores à média.

Para condução da simulação 02, a modelo da simulação 01, foram consideradas 45 combinações envolvendo percentagens de treinamento e o número de neurônios, Tabela 08. Nesta simulação a máxima acurácia foi igual a 80,26% para porcentagem de treinamento igual a 60% e número de neurônios 100. No entanto, ao se combinar porcentagem de treinamento igual a 60% e número de neurônios 30 a acurácia foi 80,01%. Devido à proximidade dos valores

das acurácias para as duas combinações, a segunda deve ser a selecionada por demandar menor esforço computacional.

Tabela 8. Acurácia da rede neural - Simulação 02.

Treinamento (%)	Conjuntos de número de neurônios								
	10	20	30	50	100	150	200	300	500
30	77,00	78,35	78,39	78,70	78,50	78,72	79,12	79,04	78,31
40	77,05	78,00	78,55	78,76	79,30	79,34	79,49	78,61	78,35
50	77,11	78,32	78,73	79,40	79,24	79,92	79,57	79,35	79,64
60	77,03	79,37	80,01	79,78	<u>80,26</u>	79,62	78,92	78,97	80,15
70	76,87	79,04	78,78	79,51	79,68	79,73	79,95	78,64	79,06

4.2.3. Simulação 03 (S03)

Através da aplicação de uma regressão linear múltipla na seleção de dados da simulação anterior (S02), foi possível avaliar os atributos de entrada e verificar se esses eram significativos para o presente estudo. Foi feito o ajuste do modelo e empregou-se a estatística de teste T aplicada às constantes de ajuste, obtendo-se o p-valor. Conforme a Tabela 09 observa-se que quatro entradas (5, 6, 7 e 9) apresentam constantes de ajuste não significativas com p-valor superior ao nível de 5%.

Tabela 9. Dados do modelo ajustados para as variáveis de entrada e saída da RNA

Entrada	Const. de ajuste	p-valor	Entrada	Const. de ajuste	p-valor
1	β_1	0,00013	8	β_8	$1,5695 \times 10^{-18}$
2	β_2	0,00016	9	β_9	0,5891
3	β_3	$2,5434 \times 10^{-23}$	10	β_{10}	$1,003 \times 10^{-75}$
4	β_4	$8,9750 \times 10^{-7}$	11	β_{11}	$1,0831 \times 10^{-251}$
5	β_5	0,10374	12	β_{12}	$3,9426 \times 10^{-245}$
6	β_6	0,23565	13	β_{13}	$8,894 \times 10^{-211}$
7	β_7	0,11184			

Desse modo, considerando os resultados apresentados na Tabela 09, para simulação 03 foram removidas duas variáveis com menor ajuste (6 e 9), passando assim de 13 para 11 variáveis de entradas. Sendo assim, ao se proceder as simulações foi constatado que a acurácia média reduziu e a maior taxa de acerto foi de 75,28% para o percentual de treinamento igual a 70% e o número de neurônios igual 50 (Tabela 10).

Tabela 10. Acurácia da rede neural para caso de simulação 03.

Treinamento (%)	Conjuntos de número de neurônios								
	10	20	30	50	100	150	200	300	500
30	71,39	72,93	73,78	73,50	73,84	74,19	73,69	73,45	73,41
40	72,48	73,36	74,41	74,19	74,03	73,12	74,18	74,21	74,16
50	72,29	73,31	73,85	74,84	75,10	74,97	74,52	74,05	72,50
60	72,20	73,78	74,75	74,22	74,75	74,35	74,82	74,84	74,25
70	72,16	73,89	74,19	75,28	73,79	75,00	74,15	73,94	73,96

4.2.4. Simulação 04 (S04)

Para a quarta simulação, além da retirada das variáveis de entrada anteriores, foram removidas as entradas 5 e 7 com intuito de observar a influência dessas no sistema em treinamento. Entretanto, apesar do pequeno aumento da acurácia em relação ao caso anterior, a remoção dessas variáveis não apresentou os resultados esperados. Com isso, ainda que a regressão linear multivariada apontasse tais variáveis como não significativas, ao removê-las, a acurácia média máxima foi de apenas 75,82% para uma RNA com 150 neurônios e 60% de dados de treinamento (Tabela 11).

Tabela 11. Acurácia da rede neural para caso de simulação 04.

Treinamento (%)	Conjuntos de número de neurônios								
	10	20	30	50	100	150	200	300	500
30	72,64	73,88	73,93	74,47	74,55	74,19	74,74	74,30	74,32
40	72,68	72,76	74,70	74,95	74,98	75,23	75,10	74,22	74,41
50	72,80	74,95	75,10	75,49	75,39	74,89	75,35	75,45	75,36

60	73,44	75,07	75,11	75,62	75,00	75,82	75,16	74,99	75,02
70	72,78	75,02	75,22	74,85	74,68	75,76	75,57	74,59	74,67

4.2.6. Validação cruzada para as Simulações 01 a 04

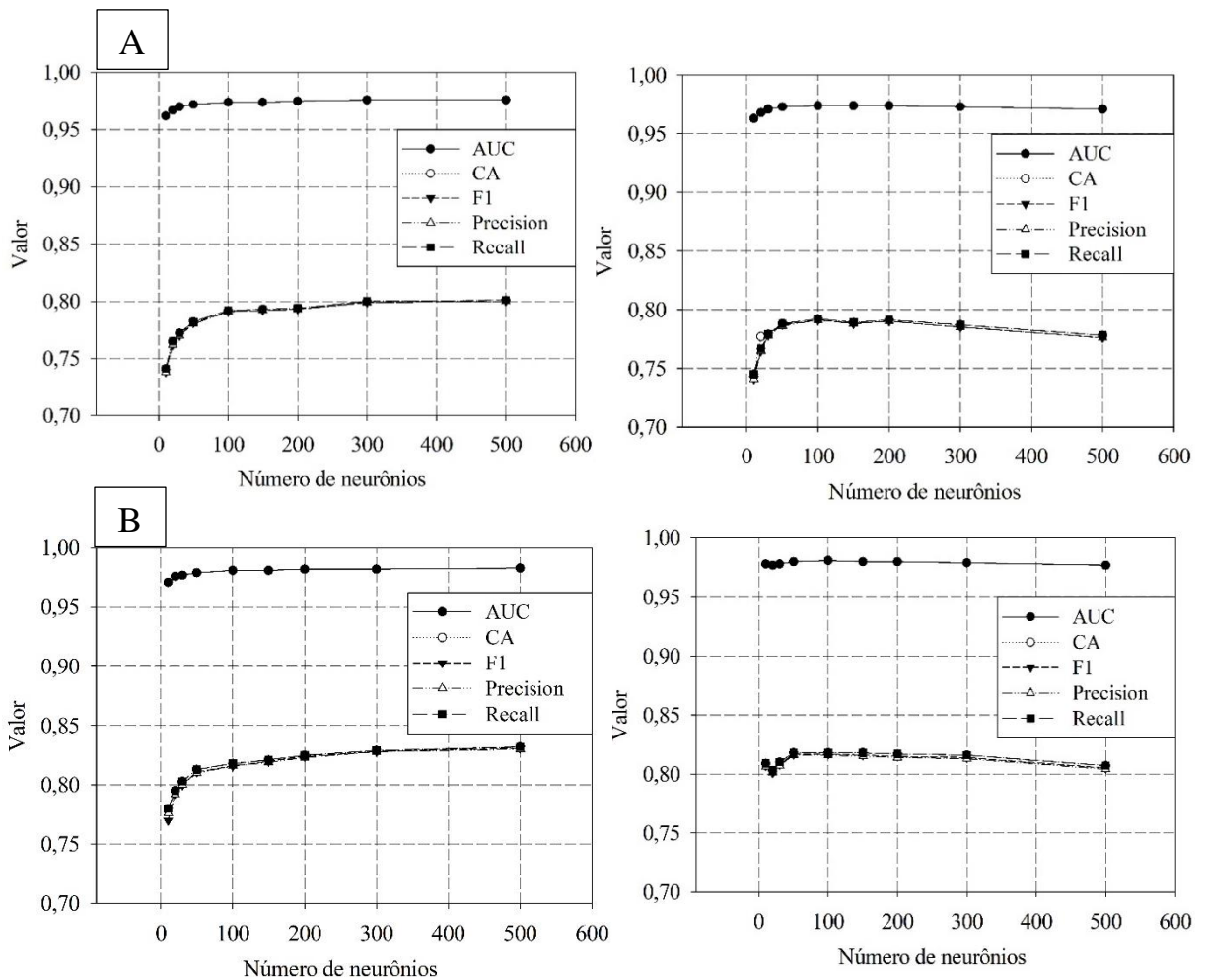
Na Tabela 12 estão representados os resultados das métricas de AUC, CA, F1, Precision e Recall para as simulações de 01 a 04. A Figura 09 ilustra as validações cruzada geradas para as simulações. Para Simulação 01 (Figura 09A) a função de ativação ReLU (a) foi superior à função Tanh (b) em todas as métricas observadas para este caso. Na Simulação 02 apenas a métrica F1 apresentou diferença entre as funções de ativação, sendo Tanh superior a ReLU (Figura 09B). A Figura 09C apresenta os valores médios das métricas analisadas para Simulação 03, em que as métricas médias para as funções de ativação não apresentaram diferenças significativas entre si. Na Simulação 04 os valores médios das métricas analisadas apresentaram maiores médias para a função ReLU (Figura 09D).

Desse modo, foi constatado que a Simulação 02 apresentou melhor desempenho em relação as demais simulações tanto para a função de ativação de unidade linear retificada (ReLU) quanto para a função de tangente hiperbólica (Tanh).

Tabela 12. Resultados das métricas para as funções de ativação ReLU (a) e Tanh (b) para as Simulações 01 a 04.

Simulação	Métricas	Função de ativação	
		ReLU (a)	Tanh (b)
Simulação 01	AUC	0,972 ± 0,005	0,971 ± 0,004
	CA	0,782 ± 0,020	0,781 ± 0,015
	F1	0,781 ± 0,021	0,778 ± 0,016
	Precision	0,781 ± 0,021	0,778 ± 0,016
	Recall	0,782 ± 0,020	0,780 ± 0,015
Simulação 02	AUC	0,979 ± 0,004	0,979 ± 0,001
	CA	0,813 ± 0,017	0,813 ± 0,006
	F1	0,810 ± 0,020	0,811 ± 0,006
	Precision	0,810 ± 0,018	0,810 ± 0,006

	Recall	$0,813 \pm 0,017$	$0,813 \pm 0,006$
Simulação 03	AUC	$0,970 \pm 0,004$	$0,970 \pm 0,002$;
	CA	$0,767 \pm 0,018$	$0,767 \pm 0,006$
	F1	$0,765 \pm 0,019$	$0,764 \pm 0,006$;
	Precision	$0,763 \pm 0,019$	$0,763 \pm 0,007$
	Recall	$0,767 \pm 0,018$	$0,767 \pm 0,006$
Simulação 04	AUC	$0,970 \pm 0,005$	$0,968 \pm 0,003$
	CA	$0,768 \pm 0,019$	$0,761 \pm 0,011$
	F1	$0,767 \pm 0,021$	$0,758 \pm 0,011$
	Precision	$0,766 \pm 0,021$	$0,757 \pm 0,011$
	Recall	$0,768 \pm 0,019$	$0,761 \pm 0,011$



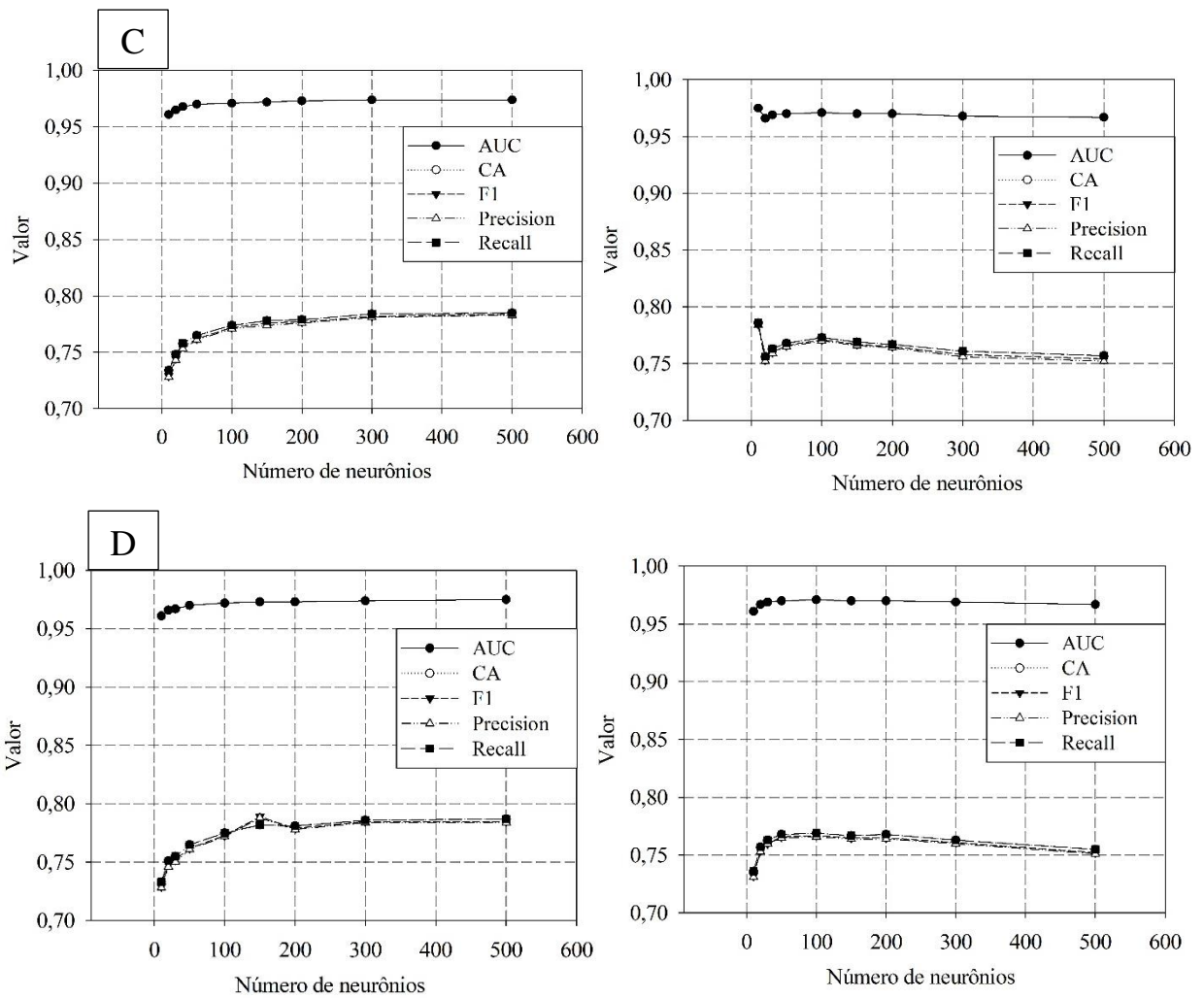


Figura 9. Validação cruzada com funções de ativação ReLU (a) e Tanh (b) para as Simulações S01 (A), S02 (B), S03 (C) e S04 (D).

4.2.5. Simulação 05 (S05)

Com o intuito de aprimorar a acurácia da simulação S02 (60% de dados para treinamento com rede de 30 neurônios) foi elaborada uma matriz de confusão (Tabela 13) em que se observa o atributo A12 - Germinados como o de menor acurácia, 54,3%, enquanto a acurácia média geral foi de $79,55 \pm 0,96\%$.

Foi observado que o atributo A12 - Germinados foi confundido constantemente com os atributos A3 - Mofados, A5 - Imaturos e A6 - Danificados. Uma das explicações plausíveis para baixa acurácia na previsão do atributo A12 seria presença da radícula que quando bem formada muitas das vezes é excluída durante o processo de segmentação e quando ainda em

desenvolvimento não difere significativamente das demais partes dos grãos nas imagens adquiridas. Genze et al. (2020), também obtiveram erros na predição da germinação de sementes de milho, centeio e milho utilizando processamento de imagens, principalmente no início do processo germinativo.

Tabela 13. Matriz de confusão - Simulação 05

		Classificação processada dos atributos											
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Classificação real dos atributos	A1	71.2	0.0	10.5	0.2	2,4	11,0	2,3	5,3	0,0	0,0	0,4	4,3
	A2	0.0	87.6	1.1	0.4	1,3	0,0	7,1	0,0	0,1	1,4	0,1	0,0
	A3	12.3	2.0	62.1	0.0	5,2	4,0	8,1	13,0	2,8	0,6	0,2	11,1
	A4	0.0	0.3	0.0	99.3	0,1	0,0	0,0	0,0	0,0	0,0	0,2	0,0
	A5	1.2	0.9	4.9	0.0	75,8	5,3	5,5	1,8	0,6	0,2	4,6	10,6
	A6	6.8	0.0	2.3	0.0	2,8	67,0	2,2	2,0	0,3	0,0	1,3	13,0
	A7	2.2	8.4	6.5	0.0	6,9	3,9	73,5	0,9	0,0	1,6	0,2	0,5
	A8	5.3	0.0	9.6	0.1	2,1	1,9	0,7	76,4	0,0	0,1	0,8	0,5
	A9	0.0	0.0	0.5	0.1	0,2	0,0	0,0	0,2	95,3	0,0	0,0	3,4
	A10	0.0	0.8	0.0	0.0	0,0	0,0	0,1	0,1	0,0	96,1	0,0	0,0
	A11	0.3	0.0	0.5	0.0	1,4	1,3	0,3	0,4	0,2	0,0	92,4	2,4
	A12	0.6	0.0	1.9	0.0	1,7	5,5	0,2	0,0	0,7	0,0	0,0	54,3

Código dos atributos - A1 - Sadios, A2 - Ardidos, A3 - Mofados, A4 - Queimados, A5 - Imaturos, A6 - Danificados, A7 - Fermentados, A8 - Partidos e quebrados, A9 - Amassados, A10 - Chochos, A11 - Esverdeados e A12 - Germinados.

Diante dos resultados apresentados para matriz de confusão, removeu-se o atributo A12 e procedeu-se novas simulações conforme a Tabela 14, em que foram consideradas as 45 combinações definidas segundo cinco percentagens de treinamento e nove conjuntos de neurônios. Observa-se que a máxima acurácia aumentou para 80,99% ao se tomar a percentagem de treinamento igual a 60% e número de neurônios igual a 100, reforçando assim um problema de classificação gerada pelo atributo A12 - Germinados.

Tabela 14. Acurácia da rede neural para caso de simulação 05.

Treinamento (%)	Conjuntos de número de neurônios								
	10	20	30	50	100	150	200	300	500
30	77,50	79,05	78,85	79,38	79,43	79,75	79,91	77,35	79,45
40	77,16	78,78	79,60	79,60	79,60	80,15	80,32	79,80	78,60
50	77,78	79,32	79,85	79,85	80,42	80,90	79,74	78,83	79,69
60	77,88	79,42	80,04	80,04	80,99	80,73	80,04	79,45	79,98
70	77,74	80,13	79,96	79,96	80,18	80,85	79,97	80,47	80,39

A Tabela 15 indica a nova matriz de confusão realizada para verificar a influência da retirada do atributo A12 - Germinados das demais classes. É possível observar que a taxa de acerto média nos demais atributos mostrou-se bastante estável mesmo com a remoção de A12, sendo em apenas A6 (Danificados) ocorrido um aumento expressivo de acurácia que migrou de 67% para 70,5%.

Tabela 15. Matriz de confusão desconsiderando o atributo G12 - Germinados.

	Classificação processada											
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	
Classificação real	A1	72.7	0.0	10.5	0.3	2.6	10.1	2.4	6.1	0.0	0.0	0.3
	A2	0.0	88.9	0.9	0.4	1.3	0.0	7.6	0.0	0.0	1.5	0.3
	A3	12.3	1.7	62.5	0.0	4.8	5.0	7.6	13.8	2.6	0.5	0.2
	A4	0.0	0.3	0.0	99.3	0.1	0.0	0.0	0.0	0.0	0.1	0.0
	A5	1.1	0.7	5.3	0.0	78.1	6.5	5.7	1.9	0.3	0.2	4.3
	A6	6.5	0.0	3.0	0.0	2.6	70.5	2.3	1.5	0.2	0.0	1.1
	A7	2.6	7.8	6.8	0.0	6.6	3.9	73.3	1.1	0.0	1.7	0.1
	A8	4.5	0.0	10.0	0.0	2.0	2.3	0.5	75.0	0.0	0.3	0.9
	A9	0.0	0.0	0.4	0.0	0.1	0.1	0.0	0.2	96.7	0.0	0.0
	A10	0.0	0.5	0.0	0.0	0.0	0.0	0.2	0.0	0.0	95.7	0.0
	A11	0.2	0.0	0.7	0.0	1.8	1.7	0.3	0.4	0.2	0.0	92.9

Código dos atributos - A1 - Sadios, A2 - Ardidos, A3 - Mofados, A4 - Queimados, A5 - Imaturos, A6 - Danificados, A7 - Fermentados, A8 - Partidos e quebrados, A9 - Amassados, A10 - Chochos, A11 - Esverdeados e A12 - Germinados.

Na Figura 10 pode-se observar as métricas obtidas para a Simulação 05, que foram AUC (a) $0,979 \pm 0,004$ e (b) $0,976 \pm 0,004$; CA (a) $0,820 \pm 0,017$ e (b) $0,807 \pm 0,017$; F1 (a) $0,819 \pm 0,017$ e (b) $0,805 \pm 0,017$; Precision (a) $0,819 \pm 0,017$ e (b) $0,805 \pm 0,018$ e Recall (a) $0,820 \pm 0,017$ e (b) $0,807 \pm 0,017$.

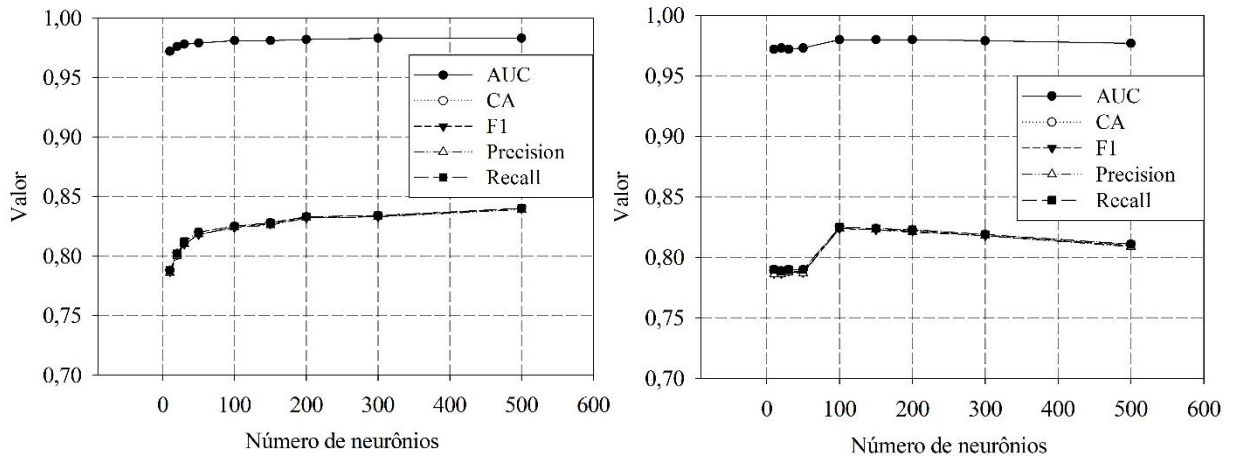


Figura 10. Validação cruzada para CS05 com função de ativação ReLU (a) e Tanh (b).

4.2.7. Simulação 06 (S06)

Para simulação 06, além da remoção do atributo A12 - Germinados, foram removidos os dados referentes ao atributo A3 - Mofados. Neste caso, primeiro foram obtidas as métricas para as funções de ativação ReLU (a) e Tanh (b). Os valores médios das métricas analisadas foram AUC (a) $0,988 \pm 0,002$ e (b) $0,987 \pm 0,001$; CA (a) $0,874 \pm 0,009$ e (b) $0,871 \pm 0,005$; F1 (a) $0,874 \pm 0,009$ e (b) $0,870 \pm 0,005$; Precision (a) $0,873 \pm 0,009$ e (b) $0,869 \pm 0,005$ e Recall (a) $0,874 \pm 0,009$ e (b) $0,871 \pm 0,005$. As métricas para a função de ativação ReLU foram maiores do que aquelas para a função Tanh (Figura 11).

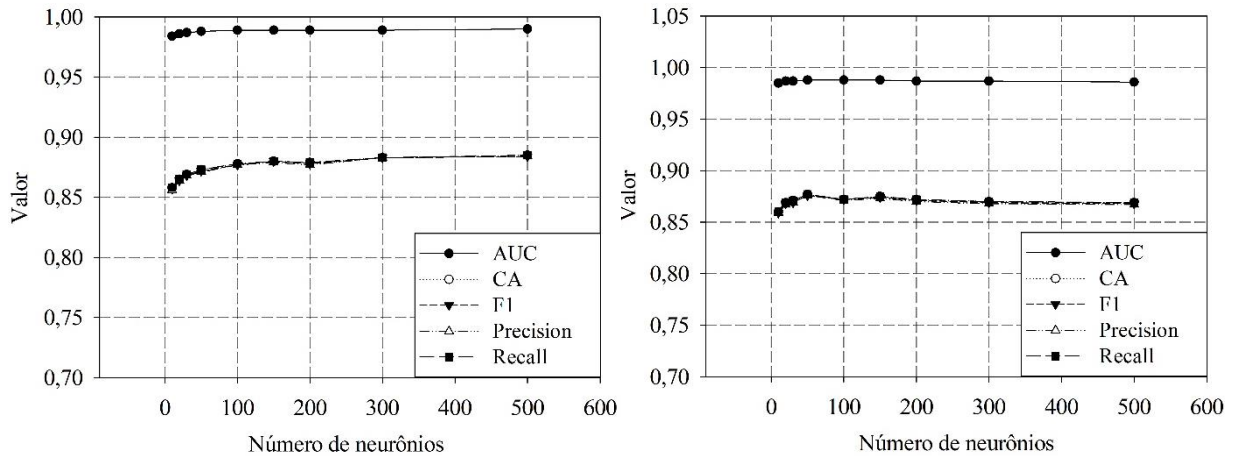


Figura 11. Validação cruzada para S06 com função de ativação ReLU (a) e Tanh (b).

Escolheu-se a melhor precisão entre as funções de ativação e através dessa criou-se uma matriz de confusão (Tabela 16). O melhor resultado encontrado foi para a função de ativação ReLU usando 70% dos dados para treinamento em uma RNA com 150 neurônios. É possível observar através da matriz de confusão que houve um aumento considerável na taxa de acerto para todas as classes de grãos.

Tabela 16. Matriz de confusão - Simulação 06.

		Classificação processada									
		A1	A2	A4	A5	A6	A7	A8	A9	A10	A11
Classificação real	A1	79,9	0,1	0,2	2,6	9,0	3,0	7,8	0,0	0,0	0,3
	A2	0,0	90,1	0,5	1,3	0,0	7,6	0,0	0,0	1,2	0,2
	A4	0,0	0,3	99,2	0,0	0,0	0,0	0,0	0,1	0,2	0,1
	A5	2,4	1,0	0,0	82,1	6,5	5,2	2,0	0,9	0,0	3,7
	A6	7,3	0,0	0,0	2,9	76,5	2,8	1,6	0,5	0,0	1,0
	A7	3,5	7,6	0,0	7,0	3,7	79,5	1,7	0,0	2,0	0,0
	A8	6,7	0,0	0,0	2,2	1,5	1,3	85,9	0,1	0,3	1,7
	A9	0,0	0,0	0,0	0,5	0,0	0,0	0,3	98,4	0,0	0,0
	A10	0,0	0,7	0,0	0,0	0,0	0,2	0,0	0,0	96,3	0,0
	A11	0,3	0,1	0,0	1,3	2,7	0,4	0,6	0,0	0,0	93,0

Código dos atributos - A1 - Sadios, A2 - Ardidos, A3 - Mofados, A4 - Queimados, A5 - Imaturos, A6 - Danificados, A7 - Fermentados, A8 - Partidos e quebrados, A9 - Amassados, A10 - Chochos, A11 - Esverdeados e A12 - Germinados.

4.2.8. Simulação 07 (S07)

No caso de simulação 07, além das dos Atributos 12 e 3 previamente retiradas, excluiu-se também os dados referentes ao Atributo 6 (Danificados). Assim como na Simulação 06, o procedimento adotado foi o uso da validação cruzada para a determinação das métricas para as duas funções de ativação estudadas e a escolha do melhor resultado para gerar a matriz de confusão.

Os resultados para as métricas analisadas estão representados pela Figura 12, as médias encontradas foram AUC: (a) $0,990 \pm 0,002$ e (b) $0,989 \pm 0,001$; CA: (a) $0,896 \pm 0,007$ e (b) $0,893 \pm 0,004$; F1: (a) $0,896 \pm 0,008$ e (b) $0,892 \pm 0,004$; Precision: (a) $0,896 \pm 0,008$ e (b) $0,892 \pm 0,004$ e Recall: (a) $0,896 \pm 0,007$ e (b) $0,893 \pm 0,004$. Para o CS07 a função ReLU apresentou valores médios maiores para todas as métricas avaliadas. O melhor resultado da função ReLU foi obtido para 50 neurônios e utilizou-se 70% dos dados para treinamento.

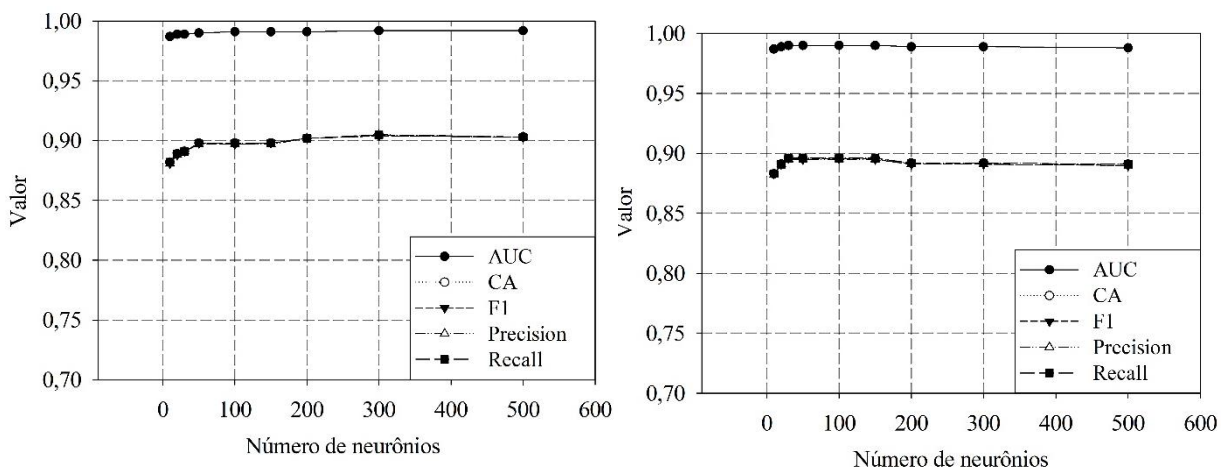


Figura 12. Validação cruzada para S07 com função de ativação ReLU (a) e Tanh (b).

A matriz de confusão apresentada na Tabela 17 demonstra um bom desempenho para todos os atributos após a remoção dos atributos A3, A6 e A12. Acredita-se que as propriedades definidas para a extração de características neste estudo não foram suficientes para gerar um padrão de reconhecimento eficiente para os atributos removidos. Desse modo, faz-se necessário buscar outros parâmetros de entradas para os atributos removidos. Oyama et al. (2012) ao desenvolverem de um modelo capaz de classificar automática de grãos de café por meio do processamento de imagem utilizaram características visuais de forma e coloração como atributos de classificação, entretanto não obtiveram um bom desempenho na taxa de acerto da

rede neural, pois assim como neste estudo, os descritores escolhidos não foram suficientes para distinguir algumas classes.

Tabela 17. Matriz de confusão a - Simulação 07.

		Classificação processada dos atributos								
		A1	A2	A4	A5	A7	A8	A9	A10	A11
Classificação real	A1	86.6	0.0	0.3	3.0	3.5	8.3	0.0	0.0	0.8
	A2	0.0	89.1	0.5	1.1	7.5	0.0	0.1	1.7	0.4
	A4	0.0	0.4	99.2	0.0	0.0	0.0	0.0	0.0	0.0
	A5	2.6	0.8	0.0	83.2	6.9	2.5	2.0	0.2	4.3
	A7	3.7	8.7	0.0	7.4	80.3	1.8	0.0	2.0	0.2
	A8	6.3	0.0	0.0	3.0	1.2	86.4	0.0	0.2	1.1
	A9	0.0	0.0	0.0	0.2	0.0	0.4	97.5	0.0	0.0
	A10	0.0	0.8	0.0	0.0	0.2	0.1	0.5	95.9	0.0
	A11	0.8	0.1	0.0	2.0	0.4	0.6	0.0	0.0	93.2

Código dos atributos - A1 - Sadios, A2 - Ardidos, A3 - Mofados, A4 - Queimados, A5 - Imaturos, A6 - Danificados, A7 - Fermentados, A8 - Partidos e quebrados, A9 - Amassados, A10 - Chochos, A11 - Esverdeados e A12 - Germinados.

4.2.9. Validação do modelo de classificação

Na Tabela 18 são apresentados os resultados para a validação da RNA ao se tomar 10 amostras aleatórias de grãos de soja.

Conforme os resultados das análises das amostras os quantitativos para os atributos A1 (Sadios) e A7 (Fermentados) foram subestimados em 26,08 e 23,3%, respectivamente. Enquanto para os atributos A2 (Ardidos), A5 (Imaturos) e A6 (Danificados) foram constatados menores quantitativos. Para os atributos A3 (Mofados), A4 (Queimados), A9 (Amassados), A10 (Chochos) e A11 (Esverdeados) os quantitativos foram estimados corretamente. Por outro lado, o atributo A12 (Germinados) apresentou a maior taxa de erro, não ocorrendo nenhum acerto nas previsões de classificação. Para o atributo A8 (Partidos e Quebrados) o erro de reconhecimento foi de 3,3%.

Tabela 18. Validação da RNA utilizando 10 amostras teste.

		Atributos											
Amostras /		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Valores													
01	Real	2	1	0	4	0	1	2	2	1	0	5	2
	Predito	2	0	0	4	2	2	2	2	1	0	5	0
	Erro	0	1	0	0	2	1	0	0	0	0	0	2
02	Real	4	3	4	0	1	0	3	3	1	3	2	1
	Predito	3	3	4	0	2	0	3	4	1	3	2	0
	Erro	1	0	0	0	1	0	0	1	0	0	0	1
03	Real	3	3	1	4	1	2	2	2	2	1	2	2
	Predito	2	3	0	4	4	3	3	1	2	1	2	0
	Erro	1	0	1	0	3	1	1	1	0	0	0	2
04	Real	4	5	1	6	2	3	3	4	2	1	3	2
	Predito	3	5	0	6	5	4	4	3	2	1	3	0
	Erro	1	0	1	0	3	1	1	1	0	0	0	2
05	Real	0	0	0	0	0	0	0	7	2	5	1	0
	Predito	1	0	0	0	0	0	0	6	2	5	1	0
	Erro	1	0	0	0	0	0	0	1	0	0	0	0
06	Real	3	7	2	0	0	0	0	0	0	0	0	0
	Predito	2	7	1	0	0	1	0	1	0	0	0	0
	Erro	1	0	1	0	0	1	0	1	0	0	0	0
07	Real	1	1	1	1	2	0	1	5	0	0	0	0
	Predito	1	1	2	1	2	1	0	4	0	0	0	0
	Erro	0	0	1	0	0	1	1	1	0	0	0	0
08	Real	4	4	0	4	4	0	2	0	0	0	0	0
	Predito	2	6	1	4	4	0	0	1	0	0	0	0
	Erro	2	2	1	0	0	0	2	1	0	0	0	0
09	Real	2	1	0	1	1	0	0	7	0	4	3	1
	Predito	1	1	0	1	1	2	0	7	0	4	3	0
	Erro	1	0	0	0	0	2	0	0	0	0	0	1
Real		0	11	1	0	3	0	17	0	0	0	0	0

10	Predito	0	17	2	0	1	1	11	0	0	0	0	0
	Erro	0	6	1	0	2	1	6	0	0	0	0	0
	Real	23	36	10	20	14	6	30	30	8	14	16	8
Total	Predito	17	43	10	20	21	14	23	29	8	14	16	0
	Erro	-6	+7	0	0	+7	+8	-7	-1	0	0	0	-8

Código dos atributos - A1 - Sadios, A2 - Ardidos, A3 - Mofados, A4 - Queimados, A5 - Imaturos, A6 - Danificados, A7 - Fermentados, A8 - Partidos e quebrados, A9 - Amassados, A10 - Chochos, A11 - Esverdeados e A12 - Germinados.

A Tabela 19 traz a porcentagem de acerto calculada por meio das matrizes de confusão para as amostras avaliadas. Observa-se que a menor porcentagem de acerto foi para a amostra 07. Essa amostra apresentou erros de classificação para os atributos A3, A6, A7 e A8, sendo os atributos A3 e A6 os que mais tem apresentado dificuldades de classificação durante todo esse estudo.

Alguns estudos foram realizados para a detecção de defeitos em amostras de grãos de soja por imagens com o intuito de otimizar os processos de classificação. Liu et al. (2015) desenvolveram um método capaz de identificar e eliminar grãos de sojas danificados por meio de visão computacional e processamento de imagens digitais aplicados a redes neurais artificiais e tecnologias de controle mecânico automatizado e obtiveram uma precisão média de 97%. Medeiros et al. (2020) encontraram taxas médias de precisão de identificação de 99,24, 98,2, 96,4, 85,6, 92,4 e 85,2 para grãos sadios, mofados, danificados por insetos, quebrados e parcialmente defeituosos, respectivamente, através do processamento de imagem e configuração de um modelo de classificação de rede neural com precisão média de 97,25%.

Tabela 19. Porcentagem de acerto das amostras de validação da rede neural artificial.

Taxa de acerto na classificação processada	
Amostra	Percentual médio de acerto (%)
01	80,0
02	84,0
03	80,0
04	86,1
05	93,3
06	83,3

07	58,3
08	77,8
09	90,0
10	75,0
Média	80,8

5. CONCLUSÕES

Considerando as avaliações realizadas neste trabalho, pode-se concluir que foi possível desenvolver e avaliar um programa computacional capaz de realizar a classificação de grãos de soja e determinar suas propriedades físicas.

Em geral o modelo se mostrou eficiente na determinação das propriedades físicas de grãos de soja. Entretanto, deve-se buscar formas de aprimorar o método de captura a fim de reduzir erros de dimensões para largura (b).

É possível concluir que a adoção das tecnologias de processamento de imagens e classificação por meio de redes neurais é um processo fundamental para automatizar processos de classificação de grãos de soja. Foi possível criar uma rede neural artificial capaz de classificar grãos de soja a uma confiabilidade de 80,04% empregando 60% dos dados para treinamento e 100 neurônios.

Dentre os 12 atributos empregados na classificação de grãos o modelo desenvolvido, esses apresentaram uma taxa de acerto de: A1 – 71,2%; A2 – 87,6%; A3 – 62,1%; A4 – 99,3%; A5 – 75,8%; A6 – 67,0%; A7 – 73,5%; A8 – 76,4%; A9 – 95,3%; A10 – 96,1%; A11 – 92,4% e A12 – 54,3.

A acurácia total pode atingir aproximadamente 81% caso seja removida o atributo A12 - Germinados. Porém, ainda é necessário aprimorar os métodos de captação de imagens e buscar variáveis alternativas para aprimorar a acurácia de previsão de classificação dos atributos que apresentaram baixa taxa de acerto.

5.1. Trabalhos Futuros

- Testar demais modelos de Redes Neurais Artificiais para reconhecimento de padrão e classificação de grãos;
- Implementar variáveis de entradas capazes de minimizar os erros de classificação dos atributos que apresentaram baixa taxa de acerto;
- Aprimorar protótipo;
- Buscar nova ferramenta de captura de imagem;
- Analisar impactos do sistema em controles de processos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABIODUN, Oludare Isaac et al. State-of-the-art in artificial neural network applications: A survey. **Heliyon**, v. 4, n. 11, p. e00938, 2018.

AGGARWAL, Charu C. An introduction to neural networks. In: **Neural Networks and Deep Learning**. Springer, Cham, 2018. p. 1-52.

ALVES, B. J. R.; BODDEY, R. M.; URQUIAGA, S. The success of BNF in soybean in Brazil. **Plant and Soil**, v.252, n.1, p. 1-9, 2003.

ARAÚJO, Marcos Eduardo Viana et al. Physical properties of yellow passion fruit seeds (*Passiflora edulis*) during the drying process. **Scientia Horticulturae**, v. 261, p. 109032, 2020.

BAEK, J. H.; LEE, E.; KIM, N.; KIM, S. L.; CHOI, I.; JI, H.; CHUNG, Y. S.; CHOI, M. S.; MOON, J. K.; KIM, K. H. High Throughput Phenotyping for Various Traits on Soybean Seeds Using Image Analysis. **Sensors**, v. 20, n.1, p. 248, 2020.

BEYELER, Michael. **OpenCV with Python blueprints**. Packt Publishing Ltd, 2015.

BONATO, E. R.; BONATO, A. L. V. A Soja no Brasil - História e Estatística. EMBRAPA – CNPSo, Londrina – PR, 1987.

BONETTI, L. P. Distribuição da soja no mundo: Origem, história e distribuição. In: MIYASAKA, S.; MEDINA, J.C. (Ed.). **A soja no Brasil**. Campina: ITAL, p. 1-6, 1981.

BORGES, Luiz Eduardo. **Python para desenvolvedores**. Novatec Editora Ltda. São Paulo - SP, ed. 1, out./ 2014.

BLACK, Robert Joseph. Complexo soja: fundamentos, situação atual e perspectiva. **Soja: tecnologia de produção II**. Piracicaba: ESALQ, p. 1-18, 2000.

BLAND, J. Martin; ALTMAN, Douglas G. Comparing methods of measurement: why plotting difference against standard method is misleading. **The Lancet**, v. 346, n. 8982, p. 1085-1087, 1995.

BRAGA, A. P.; FERREIRA, A. C. P. L.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro, Brazil: LTC editora, 2007.

BRAHMBHATT, Samarth. **Practical OpenCV**. Apress, 2013.

BRASIL. Lei nº 9.972 de 25 de maio de 2000. **Diário Oficial da União**, seção 1. Brasília – DF, maio 2000. Disponível em: < <https://www2.camara.leg.br/legin/fed/lei/2000/lei-9972-25-maio-2000-377281-publicacaooriginal-1-pl.html>>. Acesso em 18 de jun. 2020.

BRASIL. Decreto nº 6.268, de 22 de novembro de 2007. Regulamenta a Lei nº 9.972, de 25 de maio de 2000, que institui a classificação de produtos vegetais, seus subprodutos e resíduos de valor econômico, e dá outras providências. **Diário Oficial da União**, Brasília, 22 de novembro de 2007a, Seção I.

BRASIL. Instrução Normativa nº 11, de 15 de maio de 2007. Estabelece o Regulamento Técnico da Soja, definindo o seu padrão oficial de classificação, com os requisitos de identidade e qualidade intrínseca e extrínseca, a amostragem e a marcação ou rotulagem, na forma do anexo. **Diário Oficial da União**, Brasília, 16 de maio de 2007b, Seção I.

BRASIL. Instrução Normativa nº 37 de 27 de julho de 2007. Altera o inciso IV, do art. 2º, do Capítulo I, do Anexo da Instrução Normativa nº 11, de 15 de maio de 2007, que passa a vigorar com as seguintes alterações, dando-se nova redação às alíneas "b" e "g" e acrescentando-se a alínea "h". **Diário Oficial da União**, Brasília, 30 de julho de 2007c, Seção I.

BRESSERT, Eli. **SciPy and NumPy: an overview for developers**. O'Reilly Media, Inc, 2012.

BROOKER, D.B.; BAKKER-ARKEMA, F.W; HALL, C.W. Drying and storage of grains and oilseeds. **Springer**, New York, USA, 1992.

CARLEO, Giuseppe et al. Machine learning and the physical sciences. **Reviews of Modern Physics**, v. 91, n. 4, p. 045002, 2019.

CARVALHO, J. N.; GURJÃO, E. C.; MATA, M. E. R. M.; DUARTE, E. M. Classificação automática de frutas por análise de imagem—o caso da manga Tommy Atkins. **Memórias do WTA 2014**, p. 77-91, 2014.

COELHO, Juliana da Mota et al. CARACTERIZAÇÃO MORFOLÓGICA DE GRÃOS EM LÂMINAS DELGADAS POR MEIO DE TÉCNICAS DE PROCESSAMENTO DE IMAGEM. **Geociências (São Paulo)**, v. 38, n. 3, p. 735-750, 2020.

CONAB – Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos – Décimo segundo levantamento, safra 2019/2020, v. 7, Brasília – DF, set./2020.

CONAB – Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos – Décimo segundo levantamento, safra 2021/2022, v. 8, n. 12, Brasília – DF, set./2021a.

CONAB – Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos – Primeiro levantamento, safra 2021/2022, v. 9, n. 1, Brasília – DF, out./2021a.

CHANEL, Andrew A. et al. High-resolution linear polarimetric imaging for the event horizon telescope. **The Astrophysical Journal**, v. 829, n. 1, p. 11, 2016.

CHUNG, Gyuhwa; SINGH, Ram J. Broadening the genetic base of soybean: a multidisciplinary approach. **Critical Reviews in Plant Sciences**, v. 27, n. 5, p. 295-341, 2008.

DATAIKU. Guidebook – Machine Learning Basics: An Illustrated Guide for Non-Technical Readers. Disponível em: <<https://pages.dataiku.com/machine-learning-basics-illustrated-guidebook>>. Dezembro 2021.

DOĞAN, Nurettin Özgür. Bland-Altman analysis: A paradigm to understand correlation and agreement. **Turkish journal of emergency medicine**, v. 18, n. 4, p. 139-141, 2018.

DUEÑAS, M.; HERNÁNDEZ, T.; ROBREDO, S.; LAMPARSKI, G.; ESTRELLA, I.; MUÑOZ, R. Bioactive phenolic compounds of soybean (*Glycine max* cv. Merit): modifications by different microbiological fermentations. **Polish journal of food and nutrition sciences**, v. 62, n. 4, p. 241-250, 2012.

EFFENDI, M.; JANNAH, M.; EFFENDI, U. Corn quality identification using image processing with k-nearest neighbor classifier based on color and texture features. In: **IOP Conference Series: Earth and Environmental Science**, 2019. p. 012-066.

FELICETI, M. L.; SIEGA, T. C.; SILVA, M.; MASIERO, M. A.; SANTOS, I. N. T.; QUISINI, R.; BAHRY, C. A.; POSSENTI, J. C. Atributos fisiológicos das sementes de soja submetidas ao resfriamento artificial. **Research, Society and Development**, v. 9, n.11, 2020.

FURTADO, Maria Inês Vasconcellos. Redes neurais artificiais: uma abordagem para sala de aula. **Atena Editora**, 2019.

FLECK, Leandro et al. Redes neurais artificiais: Princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 1, n. 13, p. 47-57, 2016.

FREITAS, Márcio de Campos Martins. A cultura da soja no Brasil: O crescimento da produção brasileira e o surgimento de uma nova fronteira agrícola. **Enciclopédia Biosfera**, Centro Científico Conhecer - Goiânia, vol. 7, n. 12, p. 1, 2011.

GE, Z.; SONG, Z.; DING, S. X.; HUANG, B. Data mining and analytics in the process industry: The role of machine learning. **Ieee Access**, v. 5, p. 20590-20616, 2017.

GENZE, Nikita et al. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. **Plant methods**, v. 16, n. 1, p. 1-11, 2020.

GESING, Ben; PETERSON, Steve J.; MICHELSON, Dirk. **Artificial Intelligence in Logistics**. DHL Customer Solutions & Innovation, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. The MIT Press, Cambridge, MA, USA, 2016.

GONELI, A. L. D.; CORRÊA, P. C.; MAGALHÃES, F. E. A.; BAPTESTINI, F. M. Contração volumétrica e forma dos frutos de mamona durante a secagem. **Acta Scientiarum. Agronomy**, v. 33, n. 1, p. 1-8, 16 fev. 2011.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento de imagens digitais**. Editora Blucher, 2000.

HARIADI, Tony K. et al. Development of Leaf Area Meter Using Open CV for Smartphone **Application**. **Telkomnika**, v. 16, n. 4, p. 1857-1863, 2018.

HARRIS, Charles R. et al. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357-362, 2020.

HIRAKATA, Vânia Naomi; CAMEY, Suzi Alves. Análise de concordância entre métodos de Bland-Altman. **Clinical & Biomedical Research**, v. 29, n. 3, 2009.

HYMOWITZ, Theodore. On the domestication of the soybean. **Economic Botany**, v. 24, n. 4, p. 408-421, 1970.

IDRIS, Ivan. **NumPy: Beginner's Guide**. Packt Publishing Ltd, 2015.

IZEBOUDJEN, Nouma; LARBES, C.; FARAH, Ahcene. A new classification approach for neural networks hardware: from standards chips to embedded systems on chip. **Artificial Intelligence Review**, v. 41, n. 4, p. 491-534, 2014.

JITANAN, S.; CHIMLEK, P. Quality grading of soybean seeds using image analysis. **International Journal of Electrical and Computer Engineering**, vol. 9, n. 5, p. 3495-3503, 2019.

KAEHLER, Adrian; BRADSKI, Gary. **Learning OpenCV 3: computer vision in C++ with the OpenCV library**. O'Reilly Media, Inc., 2016.

KAUR, D.; KAUR, Y. Various image segmentation techniques: a review. **International Journal of Computer Science and Mobile Computing**, v. 3, n. 5, p. 809-814, 2014.

KALAIVANI, R.; MURUGANAND, S.; PERIASAMY, Azha. Identifying the quality of tomatoes in image processing using matlab. **International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering**, v. 2, n. 8, p. 3525-3531, 2013.

KELLEHER, John D. **Deep learning**. MIT press, 2019.

KOSCHAN, Andreas; ABIDI, Mongi. **Digital color image processing**. John Wiley & Sons, 2008.

KUO, T. Y.; CHUNG, C. L.; CHEN, S. Y.; LIN, H. A.; KUO, Y. F. Identifying rice grains using image analysis and sparse-representation-based classification. **Computers and Electronics in Agriculture**, v. 127, p. 716-725, 2016.

KUMAR, V.; RANI, A.; DIXIT, A. K.; PRATAP, D.; BHATNAGAR, D. A comparative assessment of total phenolic content, ferric reducing-anti-oxidative power, free radical-scavenging activity, vitamin C and isoflavones content in soybean with varying seed coat colour. **Food Research International**, v. 43, n. 1, p. 323-328, 2010.

KUNCHEVA, Ludmila I. **Combining pattern classifiers: methods and algorithms**. John Wiley & Sons, 2014.

LIU, Dejun et al. Discriminating and elimination of damaged soybean seeds based on image characteristics. **Journal of Stored Products Research**, v. 60, p. 67-74, 2015.

LIN, Jyh-Woei. Artificial neural network related to biological neuron network: a review. **Advanced Studies in Medical Sciences**, v. 5, n. 1, p. 55-62, 2017.

LORENA, Ana Carolina; DE CARVALHO, André CPLF. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43-67, 2007.

MAHAJAN, Shveta; DAS, Amitava; SARDANA, Harish Kumar. Image acquisition techniques for assessment of legume quality. **Trends in Food Science & Technology**, v. 42, n. 2, p. 116-133, 2015.

MEDEIROS, André Dantas et al. Interactive machine learning for soybean seed and seedling quality classification. **Scientific reports**, v. 10, n. 1, p. 1-10, 2020.

MELLO, E. S.; BRUM, A. L. A cadeia produtiva da soja e alguns reflexos no desenvolvimento regional do Rio Grande do Sul. **Brazilian Journal of Development**, Curitiba, v. 6, n. 10, p. 74734 - 74750, out. 2020.

MITCHELL, Tom M. et al. Machine learning. Burr Ridge, IL: McGraw Hill, 1997.

MOHSENIN, N. N. Physical properties of plant and animal materials. Gordon and Breach Sci. Pub, p. 734. New York, 1972.

MOLNAR, Christoph. **Interpretable machine learning**. Lulu. com, 2020.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

MCKINNEY, Wes. Python para Análise de Dados: Tratamento de dados com Pandas, NumPy e IPython. **Novatec Editora**, 2019.

OLIPHANT, Travis E. Python for scientific computing. **Computing in Science & Engineering**, v. 9, n. 3, p. 10-20, 2007.

OSÓRIO, Fernando S.; BITTENCOURT, João R. Sistemas inteligentes baseados em redes neurais artificiais aplicados ao processamento de imagens. In: **I WORKSHOP DE INTELIGÊNCIA ARTIFICIAL UNISC**–Universidade de Santa Cruz do Sul Departamento de Informática-Junho. 2000.

OPENCV, Open Source Computer Vision. About. OpenCV.org. Disponível em: < <https://opencv.org/about/>>. Acesso em 17 de março de 2021.

OYAMA, PI de C. et al. Sistema para classificação automática de café em grãos por cor e forma através de imagens digitais. **Embrapa Instrumentação-Capítulo em livro científico (ALICE)**, 2012.

PARIZZI, F.C. Curso de classificação e análise de grãos, 23. 2004, Viçosa, MG. **Apostila**. Viçosa: Centreinar, 2009. 131p.

RASPBERRY PI FOUADATION. Hardware. [Online]. Disponível em: < <https://www.raspberrypi.org/> >. Acesso em 24 de março de 2021.

RASPBERRY PI FOUADATION. Software. [Online]. Disponível em: < <https://www.raspberrypi.org/> >. Acesso em 24 de março de 2021a.

RIBEIRO, V. S.; SOBRAL, M. C.; ALMEIDA, M. M.; SILVA, G. F. Propriedades físicas de produtos agrícolas. **Revista Brasileira de Produtos Agroindustriais**, Campina Grande, v. 4, n.1, p. 1-6, 2002.

RIZZO, Igor Vilela; CANATO, Robson Leandro Carvalho. Inteligência artificial: funções de ativação. **Prospectus** (ISSN: 2674-8576), v. 2, n. 2, 2020.

SAMUEL, Arthur L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, v. 3, n. 3, p. 210-229, 1959.

SILVA, I. R. S.; SILVA, R. O. Linguagem de Programação Python. **Revista Tecnologias em Projeção**, v. 10, n.1, p.55, 2019.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement learning: An introduction**. MIT press, 2018.

TURZI, Mariano. **The political economy of agricultural booms: managing soybean production in Argentina, Brazil, and Paraguay**. Springer, 2016.

UPTON, Eben et al. **Learning computer architecture with Raspberry Pi**. John Wiley & Sons, 2016.

VAN DER WALT, S.; COLBERT, S. C.; VAROQUAUX, G. The NumPy array: a structure for efficient numerical computation. **Computing in Science and Engineering**, Institute of Electrical and Electronics Engineers, 13 (2), pp.22-30, 2011.

WANG, Ding; HE, Haibo; LIU, Derong. Intelligent optimal control with critic learning for a nonlinear overhead crane system. **IEEE Transactions on Industrial Informatics**, v. 14, n. 7, p. 2932-2940, 2017.

YAMANOOR, Narasimha Saii; YAMANOOR, Srihari. High quality, low cost education with the Raspberry Pi. In: **2017 IEEE Global Humanitarian Technology Conference (GHTC)**. IEEE, 2017. p. 1-5.

YANG, X.; HONG, H.; YOU, Z.; CHENG, F. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. **Sensors**, v. 15, n. 7, p. 15578-15594, 2015.