

UNIVERSIDADE DE PASSO FUNDO
INSTITUTO DE CIÊNCIAS EXATAS E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM
COMPUTAÇÃO APLICADA

**SeedFlow: Sistema de Visão
Computacional para Classificação de
Grãos de Aveia**

Diego Inácio Patrício

Passo Fundo

2019

UNIVERSIDADE DE PASSO FUNDO
INSTITUTO DE CIÊNCIAS EXATAS E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

**SEEDFLOW: SISTEMA DE VISÃO
COMPUTACIONAL PARA
CLASSIFICAÇÃO DE GRÃOS DE
AVEIA**

Diego Inácio Patrício

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Computação Aplicada na Universidade de Passo Fundo.

Orientador: Prof. Dr. Rafael Rieder

Passo Fundo

2019

CIP – Catalogação na Publicação

P314s Patrício, Diego Inácio
SeedFlow : Sistema de Visão Computacional para
classificação de grãos de Aveia / Diego Inácio Patrício. – 2018.
81 f.: il. color. ; 30 cm.

Orientador: Prof. Dr. Rafael Rieder.

Dissertação (Mestrado em Computação Aplicada) –
Universidade de Passo Fundo, 2018.

1. Software. 2. Agricultura. 3. Informática na agricultura. 4.
Aveia. I. Rieder, Rafael, orientador. II. Título.

CDU: 631:004

**ATA DE DEFESA DO
TRABALHO DE CONCLUSÃO DE CURSO DO ACADÊMICO**

DIEGO INÁCIO PATRÍCIO

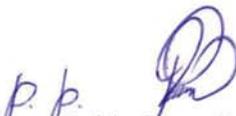
Aos dezesseis dias do mês de janeiro do ano de dois mil e dezenove, às 14 horas, realizou-se, no Instituto de Ciências Exatas e Geociências, prédio B5, da Universidade de Passo Fundo, a sessão pública de defesa do Trabalho de Conclusão de Curso "SeedFlow: Sistema de Visão Computacional para classificação de Grãos de Aveia", de autoria de Diego Inácio Patrício, acadêmico do Curso de Mestrado em Computação Aplicada do Programa de Pós-Graduação em Computação Aplicada – PPGCA/UPF. Segundo as informações prestadas pelo Conselho de Pós-Graduação e constantes nos arquivos da Secretaria do PPGCA, o aluno preencheu os requisitos necessários para submeter seu trabalho à avaliação. A banca examinadora foi composta pelos doutores Rafael Rieder, José Maurício Cunha Fernandes e Leonel Pablo Carvalho Tedesco. Concluídos os trabalhos de apresentação e arguição, a banca examinadora considerou o candidato APROVADO. Foi concedido o prazo de até quarenta e cinco (45) dias, conforme Regimento do PPGCA, para o acadêmico apresentar ao Conselho de Pós-Graduação o trabalho em sua redação definitiva, a fim de que sejam feitos os encaminhamentos necessários à emissão do Diploma de Mestre em Computação Aplicada. Para constar, foi lavrada a presente ata, que vai assinada pelos membros da banca examinadora e pela Coordenação do PPGCA.



Prof. Dr. Rafael Rieder - UPF
Presidente da Banca Examinadora
(Orientador)



Prof. Dr. José Maurício Cunha Fernandes - UPF
(Avaliador Interno)



Prof. Dr. Leonel Pablo Carvalho Tedesco - UNISC
(Avaliadora Externa)



Prof. Dr. Rafael Rieder
Coordenador do PPGCA

SEEDFLOW: SISTEMA DE VISÃO COMPUTACIONAL PARA CLASSIFICAÇÃO DE GRÃOS DE AVEIA

RESUMO

A aveia é um cereal de grande importância para a alimentação humana e animal devido aos benefícios nutricionais que oferece. Esses são oriundos principalmente das estruturas que compõem o grão. Em praticamente todas as etapas de seu processo de produção, a identificação correta da espécie e da cultivar que está sendo utilizada é uma informação essencial. O presente trabalho estabelece uma metodologia para aquisição, processamento e classificação de imagens digitais de grãos de aveia utilizando técnicas de visão computacional e de inteligência artificial. As técnicas de *aprendizado profundo*, aplicadas à imagens digitais, são caracterizadas pelo uso de redes neurais convolucionais capazes de reconhecer estruturas complexas presentes nessas imagens. Essas técnicas foram empregadas com duas finalidades: a primeira, identificar espécies de grãos de aveia, a exemplo de *Avena sativa* e *Avena strigosa*, e a segunda, classificar grãos quanto a cultivar para grãos da espécie *Avena sativa*. Entre as cultivares selecionadas estão a UPFA Ouro, UPFA Fuerza e UPFA Gaudéria. Diferentes arquiteturas de redes convolucionais estão disponíveis na literatura. Sendo assim, seis arquiteturas distintas foram comparadas para identificar qual produziria o melhor desempenho considerando o contexto deste trabalho. Essa abordagem proporcionou o resultado de acurácia de 99,7% para a identificação de espécies e 89,7% para a classificação de cultivares de aveia. Uma solução computacional denominada "SeedFlow" foi desenvolvida para a utilização da metodologia proposta. Essa solução é composta por três módulos: uma biblioteca de software; uma aplicação para treinamento e manipulação de bancos de imagens; e uma aplicação de análise e classificação de grãos utilizando redes neurais pré-treinadas. A abordagem proposta visa melhorar a eficiência em relação aos métodos manuais atualmente empregados. Segundo o experimento realizado, o seu emprego apresentou-se como viável, e pode servir como ferramenta para uso em testes de pré-seleção, em análise laboratorial ou no suporte à tomada de decisão em programas de melhoramento vegetal e avaliação de propriedade intelectual.

Palavras-Chave: cereal, cultivar, identificação, inteligência artificial.

SEEDFLOW: COMPUTER VISION SYSTEM FOR OAT GRAINS CLASSIFICATION

ABSTRACT

Oat is a cereal of great importance for human and animal food because of the nutritional benefits it offers from the structures that make up the grain. In practically every step of the production process, the correct identification of the species and the cultivar being used is essential information. The present work establishes a methodology for the acquisition, processing, and classification of digital images of oat grains using computer vision and artificial intelligence techniques. The techniques of deep learning, applied to digital images, are characterized by the use of convolutional neural networks capable of recognizing complex structures present in the acquired images. These techniques were used for two purposes: first, to identify species of oat grains, such as *Avena sativa* and *Avena strigosa*, and the second to classify grains for the cultivars of the species *Avena sativa*. Among the cultivars selected are UPFA Ouro, UPFA Fuerza, and UPFA Gaudéria. Different convolutional network architectures are available in the literature. Thus, six different architectures were compared to identify which would produce the best performance considering the context of this work. This approach provided the 99.7% accuracy result for species identification and the 89.7% accuracy result for the classification of oat cultivars. A computational solution called “SeedFlow” was developed for the use of the proposed methodology. This solution consists of three modules: a software library; an application for training and manipulation of image banks; and an application of grain analysis and classification using pre-trained neural networks. Our approach aims to provide better efficiency compared to the manual methods currently employed. According to the experiment carried out, its use proved to be viable, and it can be a useful tool in pre-selection tests, in laboratory analysis or in making-decision support for plant breeding programs and intellectual property evaluation.

Keywords: artificial intelligence, cereal, cultivar, identification.

LISTA DE FIGURAS

Figura 1.	Esquema de produção da aveia [16].	18
Figura 2.	Exemplos de transformações lineares [19].	22
Figura 3.	Exemplos de transformações geométricas [19].	23
Figura 4.	Modelos de cores [20].	24
Figura 5.	Procedimento para detecção de arestas [19].	26
Figura 6.	Estrutura de uma rede neural convolucional.	32
Figura 7.	LeNet5 [32].	32
Figura 8.	AlexNet [33].	33
Figura 9.	ResNet [37].	33
Figura 10.	DenseNet [37].	34
Figura 11.	Fluxograma do processo de revisão sistemática.	42
Figura 12.	Reconhecimento de colônias de fungos em arroz utilizando visão computacional e técnicas de aprendizado de máquina [50].	45
Figura 13.	Estrutura de grão cheios e “chochos” e suas sombras: (a) “chochos” (b) cheios [51].	46
Figura 14.	Imagens obtidas de grãos de trigo sadios (a) e danificados por germinação após 1 (b), 2(c), 3(d), 4(e), 5(f), 6(g) e 7(h) semanas [52].	47
Figura 15.	Classificação de grãos de trigo utilizando DSIFT e classificador SVM [54].	49
Figura 16.	Método de classificação de grãos de arroz utilizando as técnicas de SRC e MIF: Grãos com o centro (a) e a borda (b) em foco; (c) fusão das imagens (a) e (b); (d) imagem com o fundo removido; (e) imagem com a pilosidade removida; e (f) imagem com a região de interesse e o lema delimitada [55].	50
Figura 17.	Processo de separação de três grãos conectados [57].	52
Figura 18.	Características geométricas do grão do arroz [57].	52
Figura 19.	Visão geral da arquitetura do sistema SeedFlow.	55
Figura 20.	Material utilizado para aquisição das imagens.	57
Figura 21.	Decomposição no espaço de cor HSV.	58
Figura 22.	Visão geral do processo de preenchimento de buracos na imagem.	59
Figura 23.	Remoção do fundo de uma imagem de grãos de aveia.	60
Figura 24.	Identificação de contornos e caixas delimitadoras de uma imagem de grãos de aveia.	60
Figura 25.	Visão geral das etapas de pré-processamento e segmentação.	61

Figura 26.	Amostras de imagens de grãos de aveia após remoção do fundo e segmentação.	62
Figura 27.	Código-fonte em Python para especificação da rede neural convolucional LeNet5 utilizando Keras.	64
Figura 28.	Apresentação da tela principal do usuário do application seedflow.ui . .	65
Figura 29.	Imagem anotada contendo os pontos utilizados para obter as medidas morfológicas dos grãos identificados em uma imagem.	66
Figura 30.	Gráficos da acurácia (esq.) e da função de perda (dir.) obtidos no treinamento de redes neurais convolucionais para identificação de espécies de aveia: <i>Avena sativa</i> × <i>Avena strigosa</i>	69
Figura 31.	Gráficos da acurácia (esq.) e da função de perda (dir.) obtidos no treinamento de redes neurais convolucionais para classificação de cultivares de <i>Avena sativa</i> : UPFA Ouro × UPFA Fuerza × UPFA Gaudéria.	70
Figura 32.	Prancha para aferição da acurácia humana na identificação de espécies de aveia.	71
Figura 33.	Comparação da acurácia para identificação de espécies de aveia entre humanos (R1 a R6) e o método proposto utilizando redes neurais convolucionais (Máquina).	72

LISTA DE TABELAS

Tabela 1.	Classificação da aveia de acordo com o peso do hectolitro [13].	17
Tabela 2.	Classificação da aveia de acordo com a coloração [13].	17
Tabela 3.	Classificação da aveia quanto ao tipo [13].	17
Tabela 4.	Matriz de confusão [24].	28
Tabela 5.	Tabela de acurácia para caracterização de uma matriz de confusão 2x2 [24].	28
Tabela 6.	Resumo dos trabalhos relacionados	43
Tabela 7.	Ambiente utilizado para o desenvolvimento da solução SeedFlow.	56
Tabela 8.	Categorias e quantidades de imagens de grãos de aveia utilizadas para identificação de espécies.	61
Tabela 9.	Categorias e quantidades de imagens de grãos de aveia utilizadas para classificação de cultivares.	62
Tabela 10.	Visão parcial da tabela gerada pelo aplicativo seedflow.ui contendo parâmetros dos grãos identificados em uma imagem.	66
Tabela 11.	Comparação dos resultados para identificação de espécies de aveia utilizando redes neurais convolucionais.	68
Tabela 12.	Comparação dos resultados para classificação de cultivares de aveia utilizando redes neurais convolucionais.	69
Tabela 13.	Investimento necessário para implantação do sistema SeedFlow.	72

LISTA DE SIGLAS

API – Application Programming Interface

CPU – Central Processing Unit

DHE – Distinguibibilidade, Homogeneidade e Estabilidade

FAO – Food and Agriculture Organization of the United Nations

GPU – Graphics Processing Unit

OMC – Organização Mundial do Comércio

SDK – Software Development Kit

SIFT – Scale Invariant Feature Transform

SNPC – Serviço Nacional de Proteção de Cultivares

SVM – Support Vector Machine

UPF – Universidade de Passo Fundo

UPOV – International Union for the Protection of New Varieties of Plants

SUMÁRIO

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	AVEIA (<i>AVENA SPP.</i>)	16
2.1.1	Sistemas de Classificação	16
2.1.2	Sistema de Produção	17
2.1.2.1	Moagem	18
2.1.2.2	Produção de Derivados do Grão de Aveia	19
2.1.3	Desenvolvimento de Cultivares	19
2.2	VISÃO COMPUTACIONAL	21
2.2.1	Processamento de Imagens de Baixo Nível	21
2.2.2	Processamento de Imagens de Nível Intermediário	24
2.2.3	Processamento de Imagens de Alto Nível	27
2.3	APRENDIZADO DE MÁQUINA	27
2.3.1	Máquina de Vetores de Suporte (SVM)	29
2.3.2	Redes Neurais	30
2.3.3	Redes Neurais de Aprendizado Profundo (<i>Deep Learning</i>)	30
2.4	HARDWARE PARA VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA	34
2.4.1	Câmeras digitais	34
2.4.2	Scanners	35
2.4.3	Graphic Processing Units (GPU)	36
2.5	SOFTWARE PARA VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA	36
2.5.1	OpenCV	37
2.5.2	TensorFlow	37
2.5.3	Keras	38
2.5.4	Qt	38
3	TRABALHOS RELACIONADOS	40
3.1	PROTOCOLO DA REVISÃO SISTEMÁTICA	40
3.2	CLASSIFICAÇÃO DE GRÃOS DE TRIGO UTILIZANDO REDE NEURAL ARTIFICIAL	42

3.3	RECONHECIMENTO DE COLÔNIAS DE FUNGOS EM ARROZ UTILIZANDO VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA	43
3.4	UTILIZAÇÃO DE SOMBRAS PARA CLASSIFICAÇÃO DO PERCENTUAL DE GRÃOS CHEIOS DE ARROZ	44
3.5	ABORDAGEM DE VISÃO COMPUTACIONAL UTILIZANDO DUAS CÂMERAS PARA IDENTIFICAÇÃO DE GRÃOS GERMINADOS DE TRIGO	46
3.6	CLASSIFICAÇÃO DE GRÃOS DE ARROZ UTILIZANDO BPNN E DECOMPOSIÇÃO BASEADA EM <i>WAVELETS</i>	47
3.7	CLASSIFICAÇÃO DE GRÃOS DE TRIGO UTILIZANDO DSIFT E CLASSIFICADOR SVM	48
3.8	IDENTIFICAÇÃO DE CULTIVARES DE ARROZ UTILIZANDO ANÁLISE DE IMAGENS	49
3.9	IDENTIFICAÇÃO DE DOENÇAS EM SEMENTES DE SOJA UTILIZANDO BPNN	50
3.10	ANÁLISE DE GRÃOS GESSADOS DE ARROZ UTILIZANDO SVM	51
3.11	SISTEMA DE MEDIÇÃO AUTOMÁTICA DA PUREZA DO TRIGO	52
3.12	CONCLUSÕES DA REVISÃO SISTEMÁTICA	53
4	MATERIAIS E MÉTODOS	55
4.1	AQUISIÇÃO DE IMAGENS	56
4.2	PRÉ-PROCESSAMENTO	57
4.3	SEGMENTAÇÃO	58
4.4	TREINAMENTO DO CLASSIFICADOR	59
4.5	CLASSIFICAÇÃO	64
4.6	APRESENTAÇÃO DE RESULTADOS	65
5	RESULTADOS E DISCUSSÃO	67
5.1	ACURÁCIA HUMANA NA IDENTIFICAÇÃO DE ESPÉCIES DE AVEIA	70
5.2	ANÁLISE DA SOLUÇÃO SEEDFLOW	71
6	CONCLUSÃO	74
	REFERÊNCIAS	76

1. INTRODUÇÃO

A aveia tem significativa importância para a produção mundial de grãos. Segundo dados da FAO (*Food and Agriculture Organization of the United Nations*) [1], somente em 2016 foram produzidos 22,9 milhões de toneladas no mundo. É utilizada principalmente na alimentação humana e animal devido ao seu alto valor nutricional e elevado percentual de fibra alimentar. Recentemente, a demanda por produtos baseados em aveia tem aumentado devido aos seus benefícios nutricionais. A busca por parte do consumidor por dietas ricas em fibras encontra na aveia uma fonte alimentar ideal. Além disso, é utilizada de diversas outras formas na indústria de alimentos devido suas propriedades de retenção de umidade, de estabilizante e de incremento da resistência ao calor[2].

Sistemas de classificação auxiliam os processos de comercialização de grãos [3]. São concebidos para atender às necessidades dos compradores e, ao mesmo tempo, incentivar os produtores a buscarem a qualidade desejada na medida que proporcionam retornos equitativos em relação ao que está sendo comercializado. As operações de manuseio, transporte, processamento, tratamento e armazenamento são influenciados significativamente conforme as características do sistema de classificação de grãos adotado. Grãos produzidos e colhidos no campo são fisicamente heterogêneos devido a diversidade dos fatores agronômicos que os influenciam e dos demais elementos presentes como plantas, insetos, matéria inerte e solo. Portanto, o processo de classificação consiste em segregar o material heterogêneo em um conjunto de classes que refletem as características significativas para os consumidores [3].

Além do sistema de classificação, a seleção da cultivar a ser utilizada é outra importante decisão no processo de produção de cereais. Sua importância deve-se ao fato que é determinada por diferentes aspectos, como fatores econômicos, maturidade, resistência a doenças e pureza da cultivar. Considere, por exemplo, três principais cultivares de aveia desenvolvidas pela Universidade de Passo Fundo (UPF): UPFA 20 - Teixeira, UPFA 21 - Moreninha, UPFA 22 - Temprana. As cultivares Teixeira e Temprana são indicadas para produção de grãos para indústria com alto padrão de qualidade. Porém, a cultivar Temprana possui um ciclo super precoce, enquanto que a Teixeira possui um ciclo regular. Já a Moreninha é uma cultivar de aveia preta, indicada para o pastejo e ensilagem. Logo, essa última tem uma aplicação completamente diferente das duas primeiras [4].

A identificação correta da cultivar é uma informação essencial para o produtor. Quando uma cultivar é selecionada pelo produtor para o plantio, é importante também atentar para aspectos relativos a proteção da propriedade intelectual e a possibilidade de cobrança de *royalties*. A legislação brasileira [5] garante a pessoa física ou jurídica o direito de exploração comercial de uma cultivar registrada. Desse modo, é empregado um enorme esforço por parte das empresas obtentoras na fiscalização das lavouras visando garantir os seus direitos e evitar perdas econômicas. Porém, a identificação da cultivar a partir do grão não é uma tarefa simples. Al-

guns dos critérios utilizados para diferenciar as cultivares são influenciados pelo ambiente e, em alguns casos, estas características são fenotipicamente muito similares. As técnicas empregadas para a diferenciação de cultivares de aveia envolvem testes em laboratório por meio do uso de eletroforese e cromatografia. A extração dos descritores de DNA possibilitam a identificação do genótipo do indivíduo mesmo que esses sejam morfológicamente similares e geneticamente aparentados. Porém, exige profissionais e locais especializados para sua execução.

Na Agricultura, nota-se que o uso de aplicações de processamento de imagens e de visão computacional tem crescido devido à redução de custos dos equipamentos, o incremento do poder computacional e o interesse crescente por métodos não destrutivos de avaliação e classificação de alimentos [6]. A utilização dessas técnicas apresenta vantagens quando comparadas com métodos tradicionais de classificação baseados no trabalho manual e permite elevar a qualidade do produto final de forma automatizada, não destrutiva e eficiente do ponto de vista econômico.

Os métodos manuais para a avaliação de grãos são desafiadores mesmo para as pessoas capacitadas a realizar essas tarefas. Uma das dificuldades de tais métodos é a escassez de treinamento adequado aos avaliadores. São poucos os locais preparados para capacitar pessoas com a qualidade necessária. Outra dificuldade é o tempo necessário para realizar tais avaliações, o que inviabiliza tanto a tomada de decisão de forma rápida quanto a avaliação em larga escala [7]. Por outro lado, devido a variabilidade natural dos produtos, a tarefa de identificação e de classificação também é desafiadora do ponto de vista computacional, pois envolve um grande número de dados e parâmetros necessários para a identificação e a classificação de grãos [8]. Nesse âmbito, técnicas de visão computacional combinadas com algoritmos de reconhecimento de padrões e ferramentas de classificação automática têm sido desenvolvidas para suprir essas limitações e serem utilizadas no monitoramento de culturas e análise da qualidade de alimentos [9].

A utilização de algoritmos de aprendizado de máquina é comum em diversos segmentos como detecção de fraudes, análise de crédito, modelos de previsão de falhas, reconhecimento de padrões em imagens e filtros inteligentes de *spams*. Na Agricultura, são capazes de gerar efeitos disruptivos devido sua capacidade de análise de grandes quantidades de dados. Porém, diante da variedade de alternativas, é essencial que se conheça as características individuais de cada método e qual o melhor cenário para sua utilização.

A evolução das GPUs (*Graphic Processing Units*), desenvolvidas inicialmente para suprir as necessidades de *rendering* dos jogos digitais, foi um fator decisivo para a popularização da utilização de algoritmos de aprendizado de máquina. Devido sua alta capacidade de processamento paralelo, estas arquiteturas podem oferecer desempenho muito superior quando comparadas as CPUs (*Central Processing Unit*). Dessa forma, por meio da utilização desse tipo de hardware especializado, é possível obter soluções para problemas que antes eram inviáveis computacionalmente devido o tempo necessário de processamento em um processador convencional [10]. As GPUs são largamente empregadas na execução de algoritmos de inteligência

artificial devido seu alto poder de processamento. Embora algumas CPUs também apresentem grande capacidade de processamento, seu custo e a dificuldade do seu emprego quando comparada a utilização de GPUs tornam a sua utilização onerosa.

Logo, entende-se que há uma demanda por métodos rápidos para identificação de espécies e classificação de cultivares de aveia em diversos momentos ao longo da sua cadeia de produção. Essa demanda é presente desde a etapa de criação e registro de novas cultivares, passando pelo plantio e finalizando com o beneficiamento dos grãos pela indústria.

Diante desse contexto, é apresentado o seguinte problema de pesquisa: “É possível desenvolver um sistema de visão computacional capaz de automaticamente identificar espécies e de classificar cultivares de aveia por meio de imagens digitais de amostras de grãos?”. Os possíveis benefícios dessa solução seriam a redução do tempo de análise de uma amostra, redução da necessidade da presença de um especialista e incremento na confiabilidade dos resultados.

O presente trabalho apresenta uma metodologia para a identificação de espécies e a classificação de cultivares de aveia utilizando técnicas de visão computacional e inteligência artificial denominado de “SeedFlow”. A solução desenvolvida inclui também um sistema computacional para aplicação dessa metodologia que facilita sua utilização sem a necessidade de amplo treinamento prévio. O sistema foi projetado para ser utilizado em uma estação de trabalho utilizando hardware de fácil acesso e baixo custo. A disponibilidade de uso de diferentes modelos de scanners fotográficos e a possibilidade de uso de uma estação de trabalho genérica facilita a implementação da solução. A única restrição imposta é a necessidade de utilização de GPU. O custo de implantação da solução SeedFlow é inferior quando comparado a soluções similares como, por exemplo, a GroudEye da empresa TBit [11].

Sabe-se da diferença entre semente e grão. Em Agronomia, o termo “grão” é utilizado para denominar o produto obtido diretamente na colheita ou após sua transformação pela indústria. Já o termo “semente” é utilizado para denominar a estrutura que contém o embrião no estado de vida latente capaz de germinar para produzir outra planta. O nome da solução proposta, SeedFlow, remete a um fluxo de análise computacional (do inglês *flow*) de sementes (do inglês *seed*). Apesar disso, é possível analisar tanto sementes como grãos de aveia. Sendo assim, utilizaremos ao longo desse trabalho o termo “grão” de maneira genérica para referenciar tanto semente quanto grão.

Este documento está organizado nos seguintes capítulos:

- **Capítulo 2:** Apresenta a fundamentação teórica necessária para contextualização do tema abordado. Inicialmente, são apresentados os conceitos básicos relativos a produção da aveia no Brasil. Após, são apresentados os conceitos de visão computacional e técnicas de processamento de imagens. São abordados, também, os aspectos fundamentais do aprendizado de máquina bem como as principais arquiteturas de redes neurais convolucionais utilizadas no reconhecimento de padrões em imagens. Por último, são apresentados

os componentes de software livre disponíveis para construção de sistemas de visão computacional;

- **Capítulo 3:** Relata os trabalhos analisados para inspiração na construção da metodologia de análise de grãos de aveia elaborada. São descritos nove trabalhos disponíveis na literatura que fazem uso de técnicas de visão computacional e de inteligência artificial na análise de grãos das culturas do Trigo, Arroz e Soja;
- **Capítulo 4:** Descreve a metodologia de identificação de espécies e classificação de cultivares de aveia e o sistema computacional desenvolvido para sua utilização. Na descrição da metodologia, são detalhadas as técnicas de processamento de imagens empregadas e a forma adotada para o uso de redes neurais convolucionais na classificação das imagens. Posteriormente, são apresentadas as principais características presentes no sistema desenvolvido e os componentes de software empregados na sua concepção;
- **Capítulo 5:** Detalha e discute os resultados alcançados a partir do emprego do sistema SeedFlow. Nesse capítulo é descrito o experimento conduzido para comparar a eficácia do sistema automatizado com a forma manual de identificação de espécies de aveia;
- **Capítulo 6:** Relata as conclusões finais obtidas por meio da realização deste trabalho.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais das áreas de Agronomia e Ciência da Computação utilizadas para solucionar o problema de pesquisa. Inicialmente, são apresentadas as características da cultura da aveia no cenário brasileiro, os sistemas de classificação e o processo de produção de seus subprodutos. Depois, são abordados conceitos das áreas de Visão Computacional e Inteligência Artificial. Finalmente, é apresentado quais opções de hardware e software estão disponíveis para construção de sistemas de visão computacional e que podem ser utilizadas no desenvolvimento da solução SeedFlow.

2.1 AVEIA (*AVENA* SPP.)

No Brasil, as principais espécies de aveia cultivadas são: aveia branca (*Avena sativa*), aveia preta (*Avena strigosa*) e aveia amarela (*Avena byzantina*). O grão é formado por estruturas semelhantes a folhas que se encaixam firmemente, proporcionando, assim, eficiente proteção durante todo o período de desenvolvimento da semente. A casca, cerca de 30% do peso total do grão, não é comestível e, portanto, é um subproduto do processo de moagem da aveia. A morfologia do grão é semelhante a de outros cereais. Porém, o grão é mais longo e delgado quando comparado ao do trigo e a cevada e pode ou não apresentar pilosidade [2]. Atualmente estão registradas no Brasil 20 cultivares de aveia (*Avena sativa* L.), três de aveia (*Avena brevis* Roth), 55 de aveia branca (*Avena sativa* L. = *Avena byzantina* K. Koch) e 14 de aveia preta (*Avena strigosa*).

Segundo dados da CONAB [12], a área plantada em 2017 foi de 340 mil hectares. Desse, 248 mil hectares foram plantados no Estado do Rio Grande do Sul. A maior concentração dos municípios produtores encontra-se nos Estados do Rio Grande do Sul e Paraná. O sudoeste do Mato Grosso do Sul também é uma das regiões produtoras de aveia do país.

2.1.1 Sistemas de Classificação

De acordo com a Portaria do Ministério da Agricultura nº 191 publicada em 14 de abril de 1975 [13], a aveia é classificada de três formas: peso do hectolitro, cor e qualidade. A Tabela 1, a Tabela 2 e a Tabela 3 apresentam os critérios de classificação da aveia conforme a legislação brasileira. A classificação “abaixo do padrão” é atribuída a aveia caso não se enquadre em nenhum dos tipos de qualidade apresentados na Tabela 3. Nesse caso, apresentando bom estado de conservação, pode ser rebeneficiada para efeito de reenquadramento. Caso apresente mau estado de conservação, aspecto generalizado de mofo e fermentação, outras

sementes que sejam prejudiciais a utilização normal do produto ou odor estranho, a aveia é desclassificada e descartada.

Tabela 1. Classificação da aveia de acordo com o peso do hectolitro [13].

Grupo	Peso do hectolitro (kg)
1	igual ou superior a 50 kg
2	de 47 kg a 49 kg
3	de 41 kg a 46 kg
4	inferior a 41 kg

Tabela 2. Classificação da aveia de acordo com a coloração [13].

Classe	Descrição	Máximo aceitável de outras classes
Branca	Coloração que varia do branco ao amarelo	10%
Vermelha	Coloração avermelhada	10%
Cinzenta ou Moura	Coloração acinzentada	10%
Preta	Coloração preta	10%
Mista	Nenhuma das classes anteriores	Informar

Tabela 3. Classificação da aveia quanto ao tipo [13].

Tipo	Umidade (%)	Grãos carunchados e/ou danificados por insetos (%)	Grãos avariados (%)	Impurezas (%)
1	14 %	1 %	2 %	0,5 %
2	14 %	2 %	4 %	1 %
3	14 %	3 %	6 %	2 %
4	14 %	5 %	8 %	3 %

2.1.2 Sistema de Produção

A aveia é um cereal que possui diversas finalidades na alimentação animal e humana. Aproximadamente apresenta a proporção de 25% de casca, 9 a 12% de pericarpo, 63 a 84% de endosperma e 2,8% a 3,7% de germe entre das estruturas que formam o seu grão [14]. As dimensões do grão variam de 6 a 13 mm em comprimento e de 1 a 4,5mm em largura. O peso médio é de 32 mg e a densidade de 520 a 756 kg/m³ [15]. Ao longo do processo de produção, um importante fator que deve ser levado em consideração é o cuidado no armazenamento e manuseio do grão. Esse cuidado tem a finalidade de minimizar a rancidez resultante da oxidação lipídica e, portanto, é um importante fator na qualidade final do produto. Devido a sua casca altamente resistente e altamente indigesta, é necessário uma sequência de ações para transformar o grão em um produto consumível. Basicamente podemos dividir essas ações em dois grupos: moagem e produção de derivados do grão [16].

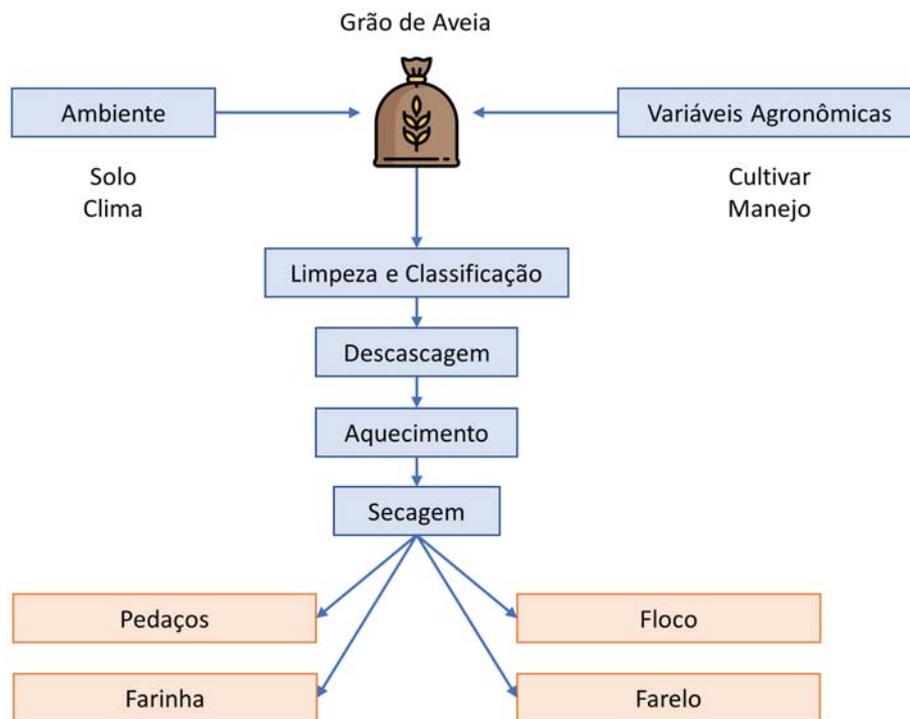


Figura 1. Esquema de produção da aveia [16].

2.1.2.1 Moagem

Basicamente, a moagem consiste em remover materiais estranhos, isolar e estabilizar os grumos e converter os grumos na forma que seja fácil cozinhá-los. Para isso, são necessárias as seguintes ações: limpeza, classificação, descascagem e secagem.

- **Limpeza:** Remoção de corpos estranhos para tornar a aveia adequada ao consumo. Consiste em utilizar separadores magnéticos e peneiras para reter os objetos indesejáveis;
- **Classificação:** A aveia é dividida em diferentes tamanhos utilizando suas características de densidade, peso e dimensões para garantir maior eficiência do moinho;
- **Descascagem:** Remoção da casca do grão por meio de discos rotativos. É ajustado para garantir a melhor relação entre o percentual de grãos descascados e o de grumos quebrados.
- **Secagem:** Aquecimento dos grumos para reduzir a umidade e o teor final de água. É vantajoso para garantir a qualidade, pois inativa as bactérias, leveduras e fungos que podem diminuir prazo de validade e representam riscos para a segurança alimentar.

2.1.2.2 Produção de Derivados do Grão de Aveia

Após o processo de moagem, os grumos estão prontos para serem convertidos em diferentes subprodutos. Os principais utilizados na indústria são: pedaços, flocos, farelo e farinha.

- **Pedaços:** Produzida simplesmente cortando os grumos inteiros em pedaços menores. Normalmente, cada grumo é cortado de dois a quatro vezes;
- **Farelo:** As farinha de aveia pode ser separadas em dois tipos: moagem grossa (rústica) e moagem fina. A moagem rústica é conhecida como farelo;
- **Flocos:** São obtidos simplesmente pelo achatamento de sêmolas inteiras;
- **Farinha:** Obtida por meio da moagem fina dos grumos de aveia.

2.1.3 Desenvolvimento de Cultivares

Cultivar é uma variedade de um gênero ou espécie vegetal que é claramente distinguível de outras variedades conhecidas por meio de um conjunto de descritores, sendo homogênea e estável quanto a estes descritores através de gerações sucessivas. O melhorista é o profissional que cria uma nova cultivar por meio de programas de melhoramento genético e o obtentor é o financiador da produção. Sendo assim, o melhorista detêm os direitos morais sobre a criação enquanto que o obtentor detêm os direitos patrimoniais. O descritor é uma característica morfológica, fisiológica, bioquímica ou molecular que é herdada geneticamente e pode ser utilizada na identificação da cultivar. Cabe aos órgãos competentes estabelecer estes descritores e aos obtentores avaliar a estabilidade das cultivares ao longo de gerações quando desejar protegê-las quanto à propriedade intelectual [17].

Países membros da OMC (Organização Mundial do Comércio) utilizam algum tipo de sistema de proteção intelectual de cultivares. A UPOV (*International Union for the Protection of New Varieties of Plants*) é uma convenção internacional que têm como missão propor e promover um sistema efetivo de proteção de cultivares. Neste âmbito, atua como um balizador no desenvolvimento das práticas adotadas pelos países em seus processos próprios de registros de cultivares. No Brasil este processo é regido pela Lei nº 9.456, de 25 de abril de 1997 por meio do SNPC (Serviço Nacional de Proteção de Cultivares). Esta lei define as condições para a proteção à propriedade intelectual e a forma com que uma nova cultivar deve ser registrada. Para o registro de uma nova cultivar é necessário o envio dos descritores mínimos definidos pelo SNPC. Esses devem ser suficientemente capazes de diferenciar uma nova cultivar ou uma cultivar essencialmente derivada das demais cultivares conhecidas.

O desenvolvimento de uma nova cultivar demanda um grande esforço no processo de melhoramento genético. Durante esse processo é essencial para o melhorista o controle preciso

de quais progênies estão sendo testadas e quais cruzamentos estão sendo realizados. Logo, é essencial a proteção da cultivar gerada para garantir ao obtentor o retorno comercial esperado relativo este intenso trabalho realizado. Uma das condições necessárias para a proteção de uma cultivar é a avaliação quanto ao atendimento aos critérios técnicos de distinguibilidade, homogeneidade e estabilidade. Esta validação é realizado por meio dos testes de DHE (Distinguibilidade, Homogeneidade e Estabilidade). Estes testes consistem em experimentos a campo ou laboratório que verificam se a cultivar atende aos requisitos técnicos esperados [5] [17].

O critério de distinguibilidade refere-se a uma cultivar ser claramente distinguível de quaisquer outras cultivares existentes. O critério de homogeneidade se refere à uniformidade da cultivar para suas características mais relevantes. No caso da aveia, a recomendação é que sejam realizados testes para pelo menos dois períodos de cultivo consecutivos. Por último, o critério de estabilidade refere-se a cultivar manter suas características relevantes inalteradas após sucessivos ciclos de cultivo. Para a aveia, cada ensaio deve incluir no mínimo 2000 plantas, em densidade normal de semeadura, as quais devem ser divididas em duas ou mais repetições em condições normais de cultivo. Solicita-se que pelo menos 20 plantas sejam analisadas quanto aos critérios de distinguibilidade e estabilidade. Quanto ao critério de análise visual de distinguibilidade, apenas 5 plantas atípicas são permitidas para cada 2000 plantas cultivadas. O processo é manual e oneroso. Logo, exige do melhorista um grande conhecimento técnico e senso crítico para avaliar as plantas cultivadas. A partir dos testes de DHE é possível determinar os descritores das cultivares de aveia.

Ainda, segundo a instrução de serviço do Ministério da Agricultura para a condução de ensaios de DHE para aveia (*Avena* spp.), são utilizados 28 descritores para a diferenciação de cultivares. Destacam-se alguns destes indicadores que possuem relação com a morfologia do grão: a) comprimento da gluma; b) forma da gluma; c) pilosidade da gluma; d) comprimento dos pelos basais; e) comprimento da ráquila; f) comprimento do lema; g) cor do lema; h) pilosidade externa do lema; i) tipo de arista. Fica claro, portanto, que o processo de descrição de uma cultivar é laborioso e envolve um série de características que devem ser mensuradas para garantir a estabilidade de uma nova cultivar.

Um aspecto importante que trata a Lei de Cultivares [5], mencionado anteriormente, é a proteção da propriedade intelectual. Segundo ela, é garantido a pessoa física ou jurídica, obtentora da nova cultivar, o direito à sua propriedade. Apenas o seu titular tem o direito à reprodução comercial da cultivar em todo o território brasileiro e, portanto, lhe é assegurado a possibilidade da cobrança de *royalties* para que terceiros utilizem a cultivar com fins comerciais. De fato, novamente fica claro que a eficaz identificação da cultivar facilita o processo de fiscalização e atribuição aos obtentores para posterior cobrança dos direitos garantidos por lei.

2.2 VISÃO COMPUTACIONAL

Visão computacional é o estudo de métodos e técnicas por meio das quais sistemas de visão artificial podem ser construídos e empregados de forma útil em aplicações práticas. Este estudo não inclui somente software, mas também hardware e técnicas de aquisição de imagens necessárias para aplicá-lo. No desenvolvimento de um sistema visão computacional, vários aspectos devem ser analisados levando em consideração o tipo do problema que será tratado. [18].

Basicamente, um sistema de visão computacional é composto por duas partes: **aquisição de imagens e processamento de imagens**:

- **Aquisição de imagens:** Compreende a transferência de sinais eletrônicos de um sensor para uma representação numérica por um dispositivo como uma câmera ou scanner [7]. A qualidade da imagem adquirida por um sistema de visão computacional é diretamente afetada pela iluminação usada durante a fase de aquisição. Dessa forma, todo o esforço investido no uso de iluminação adequada aumentará o desempenho e a confiabilidade do sistema, reduzindo a complexidade do software utilizado na fase de processamento. [19].
- **Processamento de imagens:** Compreende as tarefas executadas para manipular imagens digitais com o propósito de, por exemplo, melhorar sua qualidade, reduzir ruídos ou corrigir problemas de iluminação. Além disso, a análise de imagens refere-se ao processo de diferenciar regiões de interesse das outras regiões para extrair informações. Pode ser dividido em processamento de baixo nível, nível intermediário e alto nível [19].

2.2.1 Processamento de Imagens de Baixo Nível

Processamento de baixo nível, ou pré-processamento, inclui operações para ajuste de escala de cinza, correção de foco, aprimoramento de contraste ou nitidez e redução de ruído. Essas operações produzem uma nova imagem e são empregadas para melhorar a qualidade da imagem original ou para modificar a posição do objeto de interesse por transformações geométricas. [19].

No grupo de transformações de aprimoramento da imagem estão as transformações de intensidade. São consideradas operações pontuais na imagem pois o valor do pixel da imagem de saída dependerá apenas da função de transformação T aplicada ao ponto de mesma posição na imagem de entrada. O ajuste de brilho e contraste pode ser obtido por uma função linear $f(g) = ag + b$. A Figura 2 apresenta o efeito de diferentes valores de a (contraste) e b (brilho) para realçar as características de uma imagem de um circuito integrado. Idealmente estes valores podem ser determinados automaticamente em função das condições de iluminação da imagem. O método de normalização de contraste é uma alternativa para o cálculo automático desses valores [19].

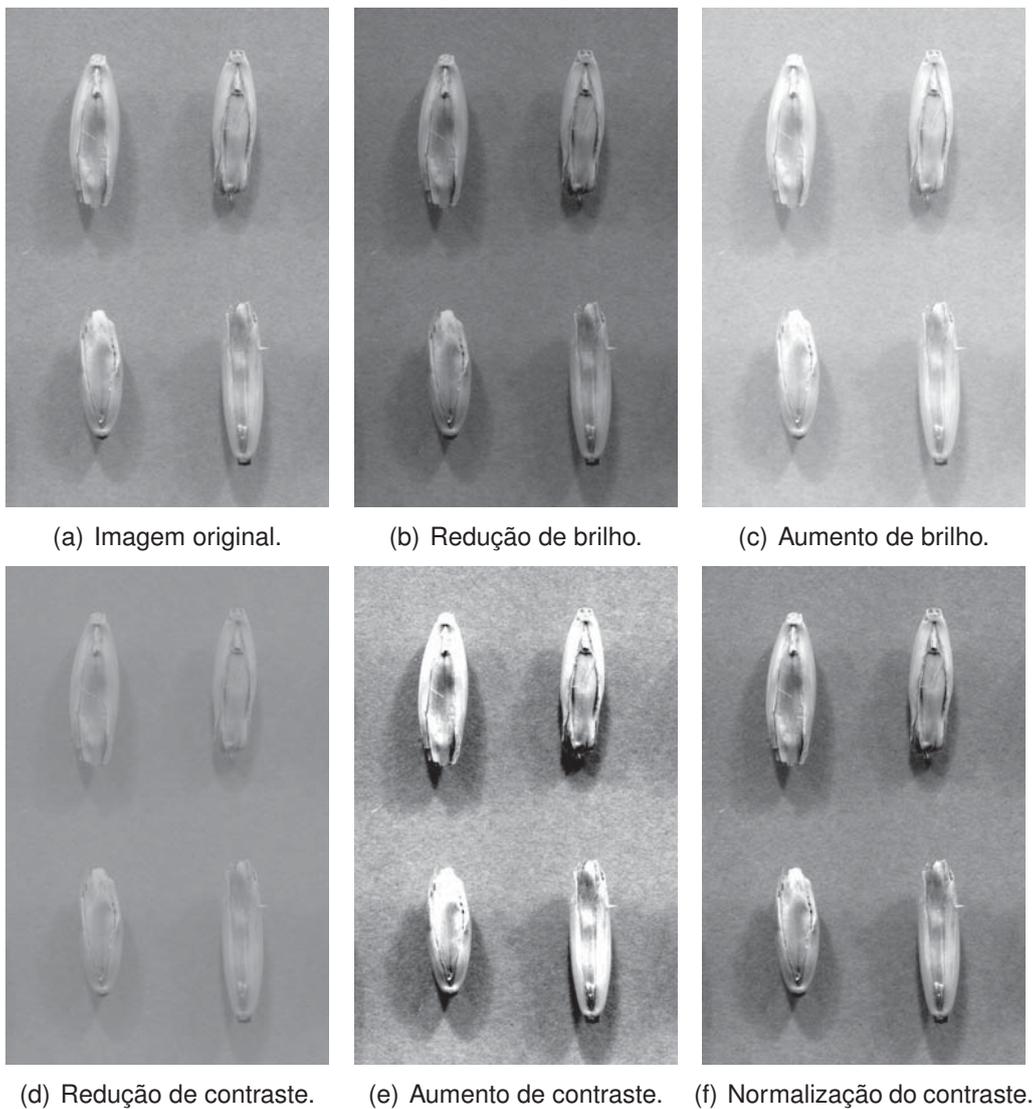


Figura 2. Exemplos de transformações lineares [19].

Em alguns casos, onde não é possível garantir que o objeto de interesse inspecionado estará sempre na mesma posição, é necessário que o algoritmo possa lidar com estas mudanças de posição. Para isto, transformações geométricas podem ser aplicadas na imagem com o objetivo de modificar a posição do objeto de interesse. Tais transformações podem ser de translação, rotação, dimensão e espelhamento. Ao utilizar este tipo de transformações é necessário que os valores dos pixels da imagem resultante sejam interpolados para evitar o serrilhamento da imagem de saída. Existem ainda as transformações de projeção utilizadas quando não é possível controlar a posição da câmera em relação ao objeto de interesse. Este tipo de projeção permite corrigir distorções de perspectiva na imagem [19].

Toda a imagem contém algum grau de ruído. Em alguns caso, é interessante reduzir o ruído de uma imagem por meio de um operador de suavização. Operações de filtragem espacial podem ser utilizadas para a redução de ruído em uma imagem. Dado um ponto (x,y) em uma imagem, e uma pequena região de vizinhança centrada em (x,y) , o processo de filtragem

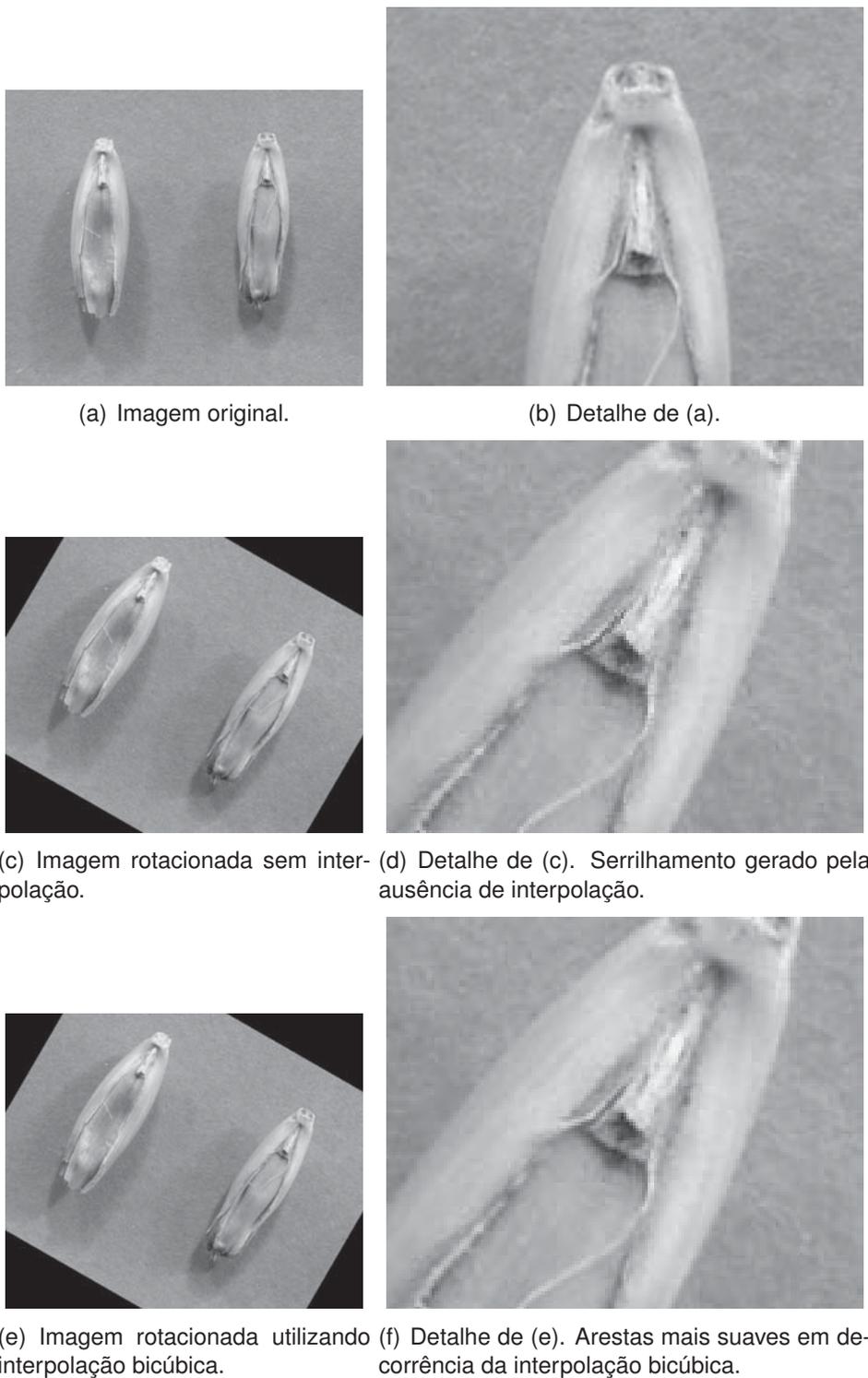


Figura 3. Exemplos de transformações geométricas [19].

consiste em mover, pixel por pixel, a origem da região de vizinhança e aplicar o operador de transformação T a estes pixels para obter o novo valor de (x,y) . Alguns exemplos são os filtros lineares que utilizam a média e a mediana da região de vizinhança e filtros adaptativos como o filtro gaussiano [18].

Modelo de cores, ou espaço de cores, consiste em um sistema de coordenadas no qual cada cor é representada por um único ponto. Para expressar efetivamente as informações de cor, é necessário estabelecer e selecionar modelos adequados de representação de cores. O modelo de cores RGB baseado em um sistema de coordenadas cartesiano, em que o espaço de cores é representado por um cubo, é mostrado na Figura 4. Por conveniência, o cubo é geralmente normalizado de modo que todos os valores R (vermelho), G (verde), B (azul) no modelo de cores RGB estejam no intervalo de $[0, 1]$. Utilizando esse modelo é possível decompor uma imagem em três planos distintos. No modelo de cores HSI, a componente intensidade (I) é desacoplada das componentes que transportam informações de matiz (H) e saturação (S). Como resultado, o modelo HSI é uma ferramenta ideal para o desenvolvimento de algoritmos de processamento de imagens com base nas descrições de cores que são naturais e intuitivas para os seres humanos. O modelo de cores HSV está mais próximo da percepção humana de cor do que o modelo HSI. O sistema de coordenadas do modelo HSV é também um sistema de coordenadas cilíndricas, mas é geralmente representado por um cone com base hexagonal [20].

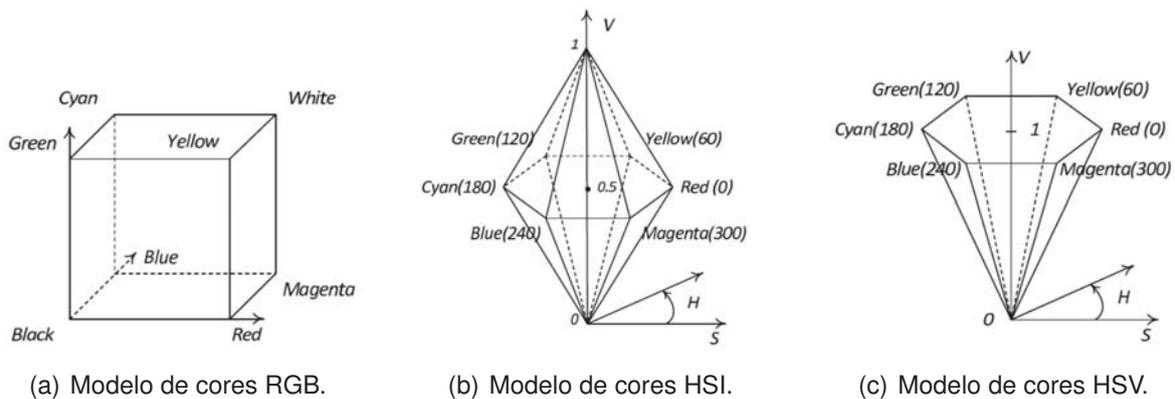


Figura 4. Modelos de cores [20].

2.2.2 Processamento de Imagens de Nível Intermediário

O processamento de nível intermediário envolve as operações de segmentação e extração de descritores dos objetos presentes na imagem.

Operações de segmentação são utilizadas para identificar regiões nas imagens. Ou seja, segmentação é operação que recebe como entrada uma imagem e produz como saída um conjunto de regiões ou contornos. A extração de atributos que caracterizam essas regiões permite a avaliação de uma série de características da região de interesse. Por exemplo, os parâmetros de elipse permitem determinar a orientação e o tamanho da região. A caixa delimitadora permite o cálculo da altura e largura da região de interesse. [19].

Filtros como *Sobel*, *Laplacian* e *Laplacian of Gaussian* podem ser usados para determinar regiões onde ocorrem mudanças abruptas de intensidade na imagem. O filtro *Gabor* é

seletivo para elementos na imagem dentro de um determinado intervalo de orientação e frequência. Os filtros do tipo *Haar* consistem em regiões retangulares adjacentes que são balanceadas onde o valor médio do filtro é zero e invariável às mudanças de iluminação.

O objetivo da detecção de bordas é produzir uma imagem binária em que valores diferentes de zero denotam a presença de uma borda na imagem. Os detectores podem, opcionalmente, também retornar outras informações, como escala e orientação, em relação a uma borda. Alguns exemplos de detectores de borda são *Canny*, *Harris* e o *SIFT*. O detector *SIFT* (*Scale Invariant Feature Transform*) é um método para identificar pontos de interesse. Ele associa informações de escala e orientação para cada ponto resultante do processo de detecção [21].

O *thresholding* é a operação que seleciona todos os pixels em um determinado intervalo de intensidade. Os valores constantes de *thresholding* funcionam bem quando não há variação no plano de fundo da imagem. Porém, mudanças de iluminação por exemplo, podem alterar os valores de intensidade da imagem e invalidar o método de *thresholding*. Desse modo, um método robusto para o cálculo do valor dos limiares é o que leva em consideração os valores mínimos e máximos do histograma da imagem.

Em algumas situações não é possível separar facilmente o plano de fundo da imagem apenas utilizando um valor de *thresholding*. Nestes casos, a técnica de *Watershed* (“linha divisória de águas”) é uma alternativa. Este algoritmo parte do gradiente de intensidade da imagem. Segundo seu funcionamento, inicialmente ele forma “vales” (pontos baixos) onde não há textura (áreas uniformes) e forma “montanhas” (pontos altos) nas regiões correspondentes as bordas, ou seja, onde existem linhas dominantes na imagem. Em seguida, seleciona pontos nos vales especificados pelo usuário ou um outro algoritmo e as preenche até que essas regiões se encontrem. Desta forma, os vales preenchidos a partir de um ponto passam a ser propriedade do mesmo. Assim sendo, é possível atribuir nomes a estas regiões [22].

Este método sempre produz contornos fechados para os objetos segmentados. Porém, como o algoritmo utiliza os valores mínimos de cinza, pode ocorrer o problema de sobre segmentação devido a natureza e a presença de ruído nas imagens. Uma alternativa é utilizar marcadores controlados na imagem para os objetos para o plano de fundo no sentido de auxiliar o algoritmo a identificar as regiões e obter melhores resultados da segmentação [22].

O resultado do método de *thresholding* é uma região a partir da imagem original. Porém, normalmente esta região contém múltiplos objetos. A identificação destes objetos é realizada utilizando um algoritmo de busca em profundidade e o operador de conectividade. A extração de regiões conectadas pode ser vista também como uma operação que nomeia as regiões conectadas.

O objetivo da detecção de arestas é produzir uma imagem binária onde os valores diferentes de zero denotem a presença de uma aresta na imagem. Detectores de aresta opcionalmente também podem retornar outras informações como escala e orientação relativos a aresta. Cantos são posições na imagem quem contém informações visuais importantes e podem ser utilizados para mapear pontos de um objeto de interesse em diferentes imagens obtidas

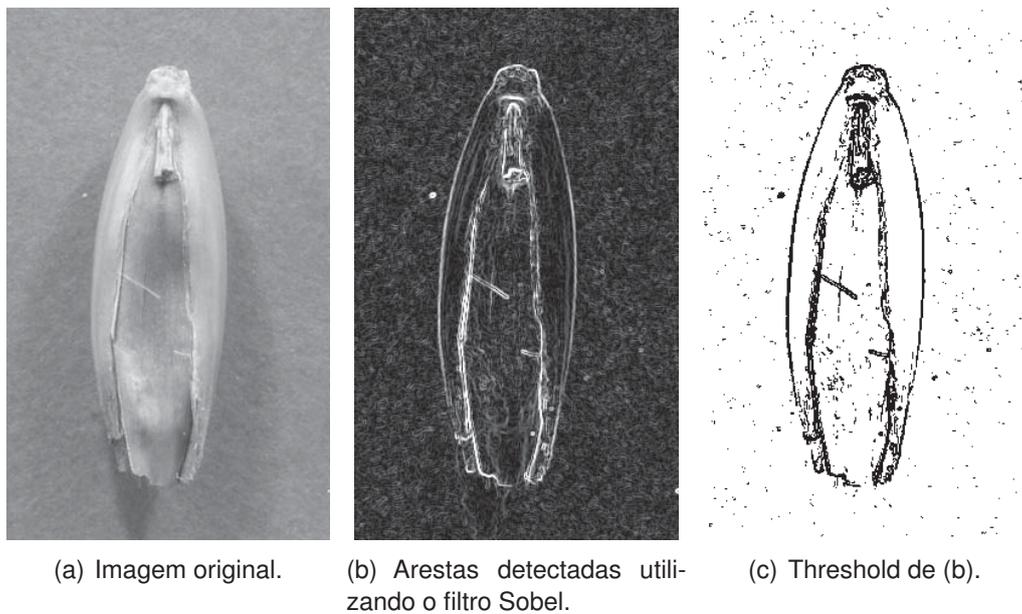


Figura 5. Procedimento para detecção de arestas [19].

de ângulos diferentes. Alguns exemplos de detectores de arestas são: *Canny*, *Harris* e *SIFT*. O detector *SIFT* (*Scale Invariant Feature Transform*) é um método para identificar pontos de interesse. Ele associa informações de escala e de orientação para cada ponto resultante do processo de detecção [21].

Descritores são uma forma de sumarizar o conteúdo de uma região da imagem. A abordagem mais simples é agregar a informação de uma imagem por meio de histogramas. Outra abordagem é utilizar descritores *SIFT* para caracterizar uma região. Inicialmente, gradientes são computados para cada pixel de uma região ao redor de um ponto de interesse. Depois, a região é subdividida em células. Estas informações são utilizadas para formar um histograma. O método *HOG* (*Histogram of Oriented Gradients*) trata de construir uma caracterização mais detalhada da estrutura espacial dos objetos de interesse. Este consiste em uma coleção de histogramas normalizados computados a partir de trechos da imagem distribuídos espacialmente. O resultado é um descritor que captura a estrutura espacial da imagem de forma aproximada porém, é invariante a pequenas deformações nos objetos. O descritor *BOV* (*Bag of Words*) é utilizado para caracterizar grandes regiões da imagem ou a imagem como um todo por meio da sumarização de dados estatísticos de descritores associados a todos os pontos de interesse em um determinada região. Cada descritor observado é considerado como sendo uma palavra de um vocabulário finito de possíveis descritores (palavras visuais). O resultado do descritor é um histograma simples que descreve a frequência com que estas palavras são observadas na imagem. Estes descritores são agrupados em *clusters* utilizando o método *k-Means* [21].

2.2.3 Processamento de Imagens de Alto Nível

Processos de alto nível, ou classificação de imagens, envolvem as técnicas utilizadas para identificação e compreensão de padrões presentes na imagem. Alguns tipos de comportamentos associados a esse tipo de inteligência são: a) habilidade de extrair informações a partir de imagens que possuem partes irrelevantes; b) capacidade de aprender a partir de exemplos e de generalizar o conhecimento de modo que possa ser utilizada em outros contextos; c) habilidade de fazer inferências a partir de informação incompleta [23].

Nesse contexto, técnicas de aprendizado de máquina são largamente empregadas para atender a esses comportamentos esperados. O reconhecimento e a classificação de regiões de interesse são geralmente realizados por classificadores estatísticos ou redes neurais.

2.3 APRENDIZADO DE MÁQUINA

Aprendizado de máquina compreende os métodos e técnicas para aplicações computacionais capazes de modificar ou adaptar suas ações, a fim de torná-las mais precisas. É possível classificar algoritmos de aprendizado nas seguintes categorias: aprendizado supervisionado, aprendizado não supervisionado, aprendizado por reforço e aprendizado evolutivo.

Algoritmos de aprendizado supervisionado são aqueles que partem de um conjunto de respostas corretas (conjunto de treinamento) para generalizar e responder corretamente a todas as entradas possíveis. Em contraste, os algoritmos de aprendizado não supervisionados comparam as entradas entre si para identificar semelhanças, a fim de classificá-las em categorias. Algoritmos de aprendizagem de reforço podem ser classificados como um estágio intermediário entre supervisionado e não supervisionado. O algoritmo é informado quando responde incorretamente a uma pergunta, mas isso não acontece quando responde corretamente. Você tem que explorar as diferentes respostas até que as respostas corretas sejam encontradas. Por fim, a aprendizagem evolutiva modela o processo de aprendizagem dos microrganismos, que consiste em se adaptar para aumentar as chances de sobrevivência e reprodução no ambiente em que vivem. Ambos os problemas de regressão e classificação podem ser tratados com o aprendizado de máquina [24].

É necessário garantir que haja treinamento suficiente para que um algoritmo possa generalizar bem um determinado problema. Porém, pouco treinamento é tão prejudicial quanto muito treinamento. Caso treinamento seja excessivo, os dados serão sobreajustados, ou seja, tanto a função real quanto o ruído e as imprecisões dos dados serão aprendidas. O modelo aprendido será muito complexo e não será capaz de generalizar o problema. É necessário que o processo de aprendizado seja encerrado antes deste ponto. Assim, serão necessários três conjuntos de dados: o conjunto de treinamento, o conjunto de validação para acompanhar o aprendizado e o conjunto de teste para validar os resultados [24].

A matriz de confusão é utilizada para analisar a qualidade dos resultados produzidos em problemas de classificação. Conforme apresentado na Tabela 4, o método consiste em uma matriz quadrada que contém todas as possíveis classes em ambas as direções (horizontal e vertical). No topo ficam as classes correspondentes as saídas previstas e ao lado as classes de entrada. Logo, o elemento na matriz (i, j) corresponderá a quantos padrões de entrada classificados na classe i foram designados pelo algoritmo na classe j . Qualquer valor na diagonal da matriz são as respostas esperadas [24].

Tabela 4. Matriz de confusão [24].

		Saídas		
		C_1	C_2	C_3
Entradas	C_1	5	1	0
	C_2	1	4	1
	C_3	2	0	4

No caso de utilização de métodos baseados em classificação binária, pode-se organizar os resultados na forma da Tabela 5 (uma tabela de confusão 2x2) com o objetivo de caracterizar a acurácia dos resultados. Dado duas classes de classificação 1 e 2, positivos verdadeiros são aqueles corretamente classificados na classe 1, enquanto que falso positivos são aqueles incorretamente classificados nesta classe. Negativos verdadeiros e falsos são aqueles corretamente e incorretamente classificados na classe 2, respectivamente [24].

Tabela 5. Tabela de acurácia para caracterização de uma matriz de confusão 2x2 [24].

1	2
Positivo Verdadeiro (TP)	Negativo Falso (FN)
Positivo Falso (FP)	Negativo Verdadeiro (TN)

Acurácia (ACC - Equação 1) pode ser definido como a soma do número de positivos e negativos verdadeiros dividido pelo número total de resultados [24]:

$$ACC = \frac{\#TP + \#TN}{\#TP + \#FP + \#TN + \#FN} \quad (1)$$

Outras medidas podem ser utilizadas para interpretar o desempenho do classificador [24]: sensibilidade (TPR - Equação 2), especificidade (TNR - Equação 3), precisão (PPV - Equação 4). Sensibilidade mede a proporção de positivos corretamente identificados enquanto que especificidade mede a proporção de negativos corretamente identificados. Neste sentido, a medida de precisão corresponde a proporção entre os positivos corretamente identificados em relação aos resultados positivos reais.

$$TPR = \frac{\#TP}{\#TP + \#FN} \quad (2)$$

$$TNR = \frac{\#TN}{\#TN + \#FP} \quad (3)$$

$$PPV = \frac{\#TP}{\#TP + \#FP} \quad (4)$$

Por ultimo, podemos ainda utilizar a medida *F1 Score* (Equação 5) que considera tanto a precisão *PPV* quanto a sensibilidade *TPR*. O *F1 Score* pode ser interpretado como um valor no intervalo de 0 (pior) a 1 (melhor) que corresponde a média harmônica da precisão e da sensibilidade.

$$F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} \quad (5)$$

O processo de aprendizado de máquina consiste nas seguintes etapas [24]:

- **Preparação e Aquisição de Dados:** Etapa inicial que consiste em obter os dados apropriados que contêm características a serem consideradas na aprendizagem. Em geral, os algoritmos de aprendizado exigem uma grande quantidade de dados e que sua quantidade é um fator a ser considerado. Essa demanda, de preferência sem muito ruído, com o custo computacional que aumenta proporcionalmente conforme a quantidade aumenta é um ponto relevante a ser considerado.
- **Seleção de Características de Interesse:** Consiste em identificar as características mais significativas para o problema que será abordado.
- **Seleção de Algoritmo:** Consiste em selecionar o algoritmo mais apropriado para tratar o problema em estudo.
- **Seleção de Parâmetros:** Alguns algoritmos precisam ser ajustados por parâmetros selecionados por meio de experimentação.
- **Treinamento:** Dado um conjunto de entradas, o algoritmo e os parâmetros, o treinamento consiste na construção do modelo computacional que será utilizado na predição das respostas aos novos dados.
- **Avaliação:** O sistema precisa ser avaliado quanto à precisão dos dados para os quais foi treinado.

Três tipos de classificadores são os mais utilizados em sistemas de visão computacional: máquina de vetores de suporte ou SVM, redes neurais e redes neurais profundas.

2.3.1 Máquina de Vetores de Suporte (SVM)

Máquina de vetores de suporte ou *support vector machine* (SVM) é um classificador linear não probabilístico que constrói um limiar de decisão com a maior distância possível entre

os pontos de exemplo. Um classificador SVM clássico é capaz de classificar os pontos do conjunto de entrada em duas classes possíveis. Os limiares de separação, embora lineares, têm a capacidade de incorporar dados em um espaço maior. Desta forma, é possível criar separadores lineares mesmo que no espaço original das entradas seja impossível defini-lo linearmente. Finalmente, o SVM é um método não paramétrico. Portanto, ele tem a flexibilidade de representar funções complexas e, ao mesmo tempo, é resiliente a *overfitting* [25].

2.3.2 Redes Neurais

Uma rede neural é composta de nós conectados por links diretos com um peso associado que determina a força e o sinal da conexão. Os pesos são a principal forma de armazenamento de longo prazo em redes neurais e os algoritmos de aprendizado tem como objetivo atualizar esses valores. Nós de entrada e de saída são aqueles que possuem conexões com o mundo externo. Cada nó é independente dos outros e, portanto, pode executar seus cálculos partindo apenas de seus valores de entrada e respectivos pesos. Cada unidade recebe os valores dos links de entrada, calcula o nível de ativação e, quando ativada, envia o sinal para todos os seus links de saída. Devido a sua característica de independência podem ser paralelizados [25].

Uma rede neural direta tem conexões em apenas uma direção e pode ser representada como um gráfico acíclico direcionado. Cada nó recebe entradas dos nós ascendentes e entrega a saída aos nós descendentes sem ciclos. Esse tipo de rede representa uma função e não possui estado interno, exceto os pesos das conexões. Em contraste, uma rede neural recorrente alimenta os nós de entrada com suas saídas. Desta forma, a rede forma um sistema dinâmico que pode ou não atingir um estado estável [25].

As redes neurais são organizadas em camadas, onde cada unidade recebe entrada apenas dos nós da próxima camada superior. Em redes que possuem uma ou mais camadas, são chamadas unidades ocultas para aquelas que não estão diretamente conectadas à saída da rede. Com apenas uma camada oculta suficientemente grande, é possível representar qualquer função contínua com precisão arbitrária. Com duas camadas, até funções não contínuas podem ser representadas [25].

2.3.3 Redes Neurais de Aprendizado Profundo (*Deep Learning*)

Deep Learning difere-se das redes neurais convencionais no sentido de que elas possuem mais nós, meios mais complexos de interconexão de camadas, demandam maior poder computacional para seu treinamento e têm extração automática dos parâmetros.

As quatro principais arquiteturas usadas são: redes pré-treinadas não-supervisionadas, redes neurais convolucionais, redes neurais recorrentes e redes neurais recursivas. Esses tipos

de redes já são usados na solução de vários problemas, como conversão de texto em fala [26], identificação de idioma [27], tradução [28], processamento de sinal [29], reconhecimento automático de caracteres [30], geração automática de legendas [31], entre outros. As redes neurais convolucionais ou *convolutional neural networks* (CNN) partem da premissa que a entrada será uma imagem. Dessa forma, são as mais utilizadas na identificação de padrões em imagens digitais. O objetivo principal de uma rede neural convolucional é aprender as características dessas imagens por meio de operações de convolução.

O LeNet5 foi uma das primeiras arquiteturas propostas [32]. Essa arquitetura de rede consiste em três tipos de camadas: camadas convolucionais, camadas de subamostragem e camadas totalmente conectadas. Inicialmente, o LeNet-5 era usado para reconhecer dígitos escritos à mão representados por imagens bidimensionais de 32×32 pixels. A última camada contém os nós que correspondem às probabilidades da imagem original pertencentes a uma das dez classes de dígitos [10]. Outras arquiteturas de rede são descritas na literatura: AlexNet (2012) [33], ZF Net (2013) [34], VGG Net (2014) [35], GoogLeNet (2015) [36], Microsoft ResNet (2015) [37] são alguns exemplos. Algumas redes R-CNN (CNNs baseadas na região) destinam-se a detectar objetos identificando as regiões com maior probabilidade de conter o objeto a ser identificado. Neste grupo estão R-CNN (2013) [38], Fast R-CNN (2015) [39] e Faster R-CNN (2015) [40]. A partir dessas arquiteturas, ainda é possível usá-las em combinação para obter resultados ainda mais expressivos.

A estrutura básica de uma rede convolucional inclui a camada de entrada, camadas convolucionais, camada de *pooling*, camada totalmente conectada, classificador SoftMax e a camada de saída (Figura 6). Durante um processo de convolução de uma determinada camada, um filtro desliza sobre essa camada. Após, é realizado o produto de Hadamard entre a matriz de peso e os valores dos pixels resultantes do processo de filtragem. A função de ativação mais comumente utilizada é a função sigmoide ou a função de tangente hiperbólica com um intervalo de $[-1, 1]$. No entanto, com a crescente profundidade da rede, os gradientes tendem a desaparecer ou explodir. A função de ativação de ReLu [33] foi proposta com a finalidade de reduzir esse efeito. Essa camada aplica uma função de ativação elementar, como $\max(0, x)$. A camada de pooling tem a finalidade de reduzir o tamanho das dimensões espaciais de entrada. O classificador SoftMax é utilizado para determinar a probabilidade de uma entrada x pertencer à classe t [41].

A camada convolucional é principal componente de uma rede convolucional. Essa camada consiste em um conjunto de filtros. À medida que deslizamos o filtro sobre a largura e a altura do volume de entrada, produzimos um mapa de ativação bidimensional que fornece as respostas desse filtro em todas as posições espaciais. Intuitivamente, a rede aprenderá os filtros que são ativados quando algum tipo estrutura visual é identificada. Essas estruturas podem ser uma borda ou uma mancha de alguma cor nas camadas iniciais ou, eventualmente, padrões inteiros nas camadas mais profundas da rede.

A inserção de uma camada de *pooling* entre camadas convolucionais sucessivas tem como função reduzir progressivamente o tamanho espacial da representação para reduzir a quantidade de parâmetros e computação na rede. É, portanto, um meio de controlar o *overfitting*. A camada de pooling é independente das camadas convolucionais.

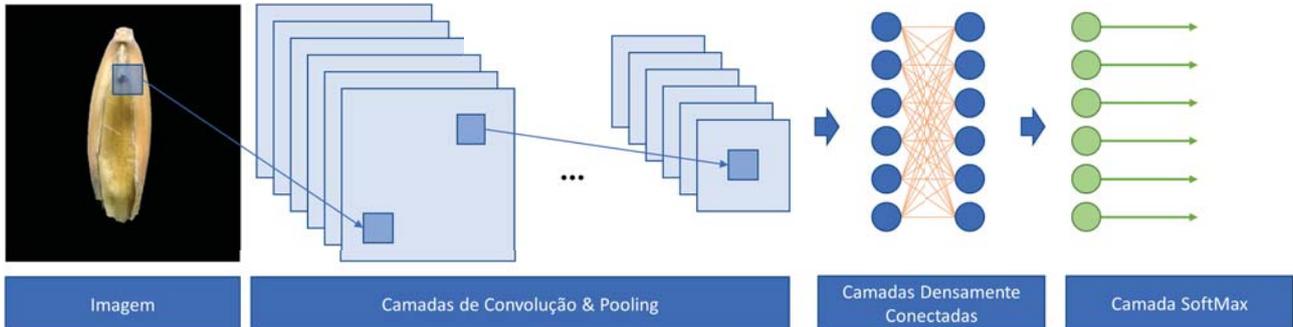


Figura 6. Estrutura de uma rede neural convolucional.

Dentre os mais utilizadas arquiteturas de redes neurais convolucionais, podemos destacar seis: **LeNet5**, **AlexNet**, **VGG 16**, **Inception V3**, **Resnet** e **DenseNet 121**.

- **LeNet5:** É composta por sete camadas: camada convolucional, camada de subamostragem, camada convolucional, camada de subamostragem, camada convolucional, camada totalmente conectada com 84 unidades e, finalmente, uma camada de saída (Figura 7) [32].

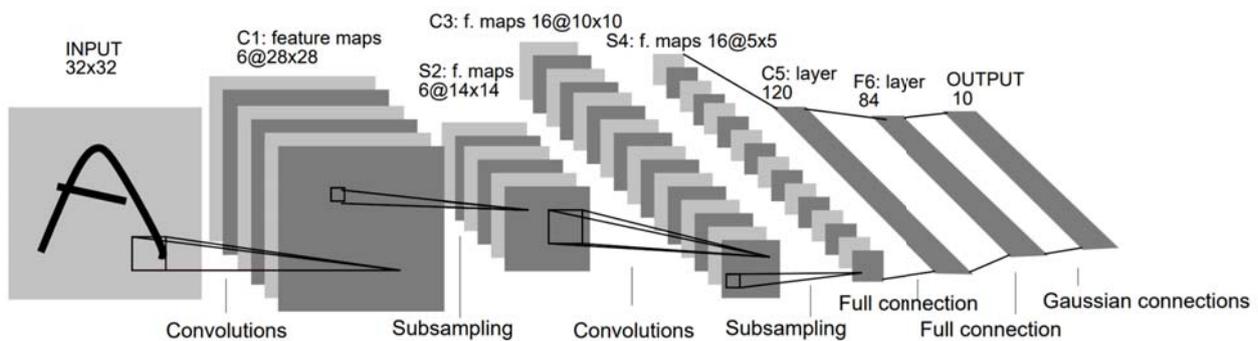


Figura 7. LeNet5 [32].

- **AlexNet:** É composta por cinco camadas convolucionais e três camadas totalmente conectadas. A segunda e terceira camadas convolucionais recebem a entrada da camada predecessora após as operações de *max-pooling* e normalização. A arquitetura AlexNet usa neurônios do tipo ReLu para reduzir o tempo necessário para treiná-la. Duas camadas de *Dropout* com uma probabilidade de 0,5 estão posicionadas entre a primeira e a segunda camada totalmente conectada (Figura 8) [33].
- **VGG:** Arquitetura composta por camadas convolucionais e camadas *max-pooling* organizadas na sequência: entrada, duas camadas convolucionais de tamanho 64, *max-pooling*,

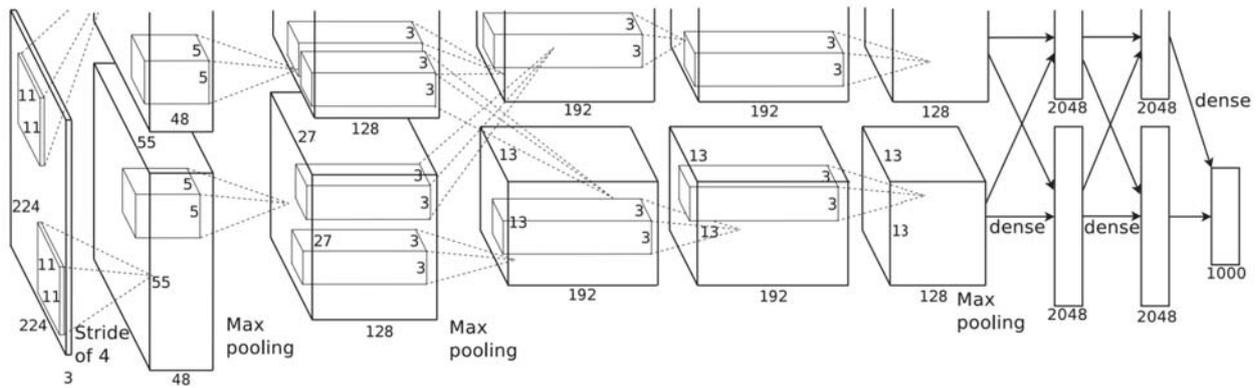


Figura 8. AlexNet [33].

duas camadas convolucionais de tamanho 128, *max-pooling*, três camadas convolucionais de tamanho 256, *max-pooling*, três camadas convolucionais de tamanho 512, *max-pooling*, três camadas convolucionais de tamanho 512, *max-pooling* e três camadas totalmente conectadas. Ela usa ReLu como modelo de neurônio e reforça a noção de que a profundidade da rede está relacionada à representação hierárquica de dados visuais [35].

- **Inception:** Arquiteturas mais profundas envolvem um número mais significativo de parâmetros e, portanto, são mais propensos ao *overfitting*. Adicionalmente, camadas convolucionais encadeadas também aumentam substancialmente o cálculo. A abordagem adotada foi a de utilizar vários tamanhos de filtro no mesmo nível em vez de encadear camadas convolucionais com diferentes tamanhos de filtro. Desse modo, a rede ficaria “mais larga” ao invés de “mais profunda”. A arquitetura possui 22 camadas de profundidade [36].
- **ResNet:** Apresenta o conceito de um bloco residual para abordar a degradação da precisão no treinamento utilizando arquiteturas mais profundas. Usando conexões de atalhos é possível transferir dados de ativação muito mais profundamente na rede e reduzindo o efeito do problema de desaparecimento do gradiente. Quando o gradiente é retropropagado para as camadas anteriores, esse pode ficar muito pequeno e o desempenho pode degradar severamente (Figura 9) [37].

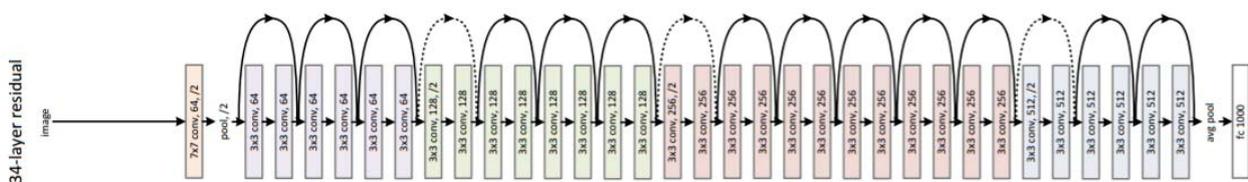


Figura 9. ResNet [37].

- **DenseNet:** É uma arquitetura densamente conectada baseada na ideia da ResNet de que redes mais profundas podem ser mais eficientes treinadas se contiverem conexões mais

curtas entre as camadas. Nesse sentido, em uma DenseNet, cada camada é conectada a todas as outras camadas em um modo de *feed-forward*. Como os criadores apontaram, há muitas vantagens em comparação com as arquiteturas simples, como a mitigação do problema de desaparecimento do gradiente, intensificação da propagação dos atributos e a redução do número de parâmetros (Figura 10) [42].

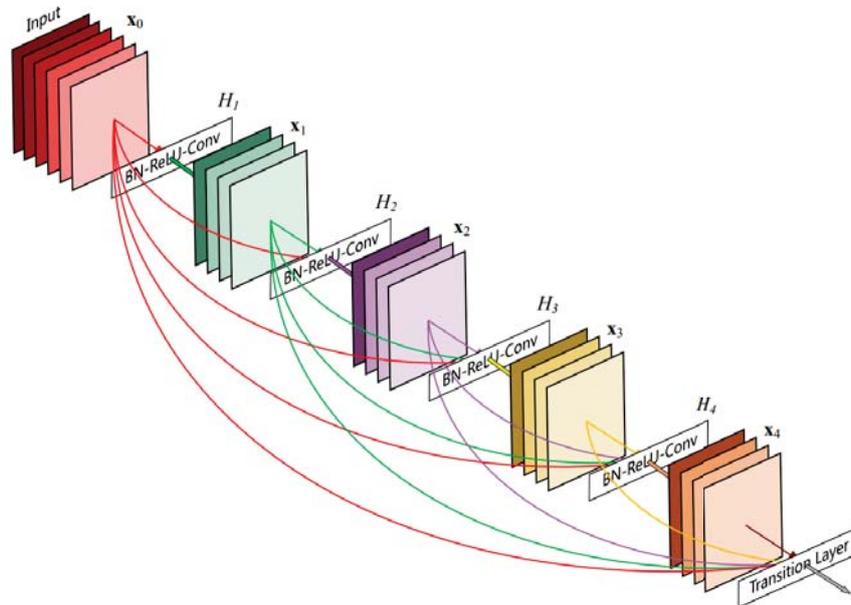


Figura 10. DenseNet [37].

2.4 HARDWARE PARA VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA

Diferentes tipos de hardware podem ser empregados na construção de um sistema de Visão Computacional. Os equipamentos utilizados para aquisição de imagem como, por exemplo, câmeras e scanners destacam-se por serem essenciais na conversão de luz em sinais eletrônicos. Esses possibilitam que técnicas de processamento de imagens sejam utilizadas para obtenção de informações a partir de tais sinais. Além disso, sabe-se que a complexidade utilizada na etapa de classificação exige um grande potencial de computação disponível. Nesse sentido é interessante que alternativas para o processamento de grandes quantidades de dados sejam encontradas. Uma dessas alternativas é a utilização de hardwares especializados em computação paralela como as GPUs.

2.4.1 Câmeras digitais

As câmeras possuem sensores que convertem a intensidade de luz recebida em um sinal eletrônico. Basicamente, dois tipos de câmeras que se distinguem pela forma de varredura

podem ser usadas: área ou linha. Câmeras convencionais ou de varredura de área geram uma imagem em cada ciclo de exposição. Em contraste, as câmeras de varredura de linha capturam apenas uma linha de pixels por vez. Para a aquisição de imagens bidimensionais é necessário mover o objeto a ser capturado usando um transportador ou mover a câmera ao longo de um objeto estacionário [43].

Há muita variação no tamanho do sensor, no tamanho do pixel e no número de pixels, levando a uma gama diversificada de designs de câmeras. Para câmeras compactas, pequenas dimensões de sensores não são realmente um problema. A resolução dessas câmeras é mais que suficiente. Porém, uma desvantagem do uso de pixels menores é mais ruído, portanto, uma menor relação sinal-ruído e faixa dinâmica reduzida em comparação com sensores que empregam pixels maiores. Além disso, os níveis de ruído mais altos exigem mais processamento de redução de ruído no processador da câmera, o que geralmente reduz a nitidez na imagem final. No entanto, o pequeno sensor permite o uso de lentes com menores distâncias focais e, portanto, facilita projetos de câmera mais compactas [43].

Quase todas as câmeras digitais disponíveis comercialmente usam uma matriz de filtro de cor posicionada diretamente na frente de um único sensor, capturando faixas de comprimento de onda separadas para pixels individuais. Cada pixel, portanto, contribui para apenas um dos canais de cor e os valores para os outros dois canais nesse pixel devem ser interpolados. O processo de interpolação de valores de cores é conhecido como demosaicagem [43].

2.4.2 Scanners

Existem vários tipos diferentes de scanner, sendo os mais comuns: scanners de tambor, usados para escanear filmes e transparências; scanners de mesa, usados para digitalizar fotografias, documentos impressos e, com alguns modelos, filmes; e scanners de filmes dedicados. Quando um material é escaneado, ele é iluminado por uma fonte de luz adequada e a luz transmitida ou refletida do material escaneado é capturada por um sensor digital. Os scanners coloridos geralmente têm três fileiras de elementos sensores CCD (ou seja, três CCDs lineares) cobertos com filtros vermelhos, verdes e azuis (RGB), para separar a imagem em três canais de cores [43].

O método de captura de imagem, tipo de sensor e fonte de luz depende do tipo de scanner. As lâmpadas de LED utilizadas em scanners mais modernos garantem a estabilidade temporal da fontes de luz durante a digitalização de uma imagem é essencial. Além disso, minimizam o tempo de aquecimento necessário quando comparado com scanners mais antigos equipados com lâmpadas fluorescentes. A resolução de um scanner está relacionada à sua taxa de amostragem e é medida em pixels por polegada (ppi) [43].

Um scanner de mesa com um sensor de imagem de contato (CIS) é mais fino e leve em comparação com um scanner com sensor CCD. Um CIS consiste em diodos emissores de luz (LEDs) vermelhos, verdes e azuis que iluminam a imagem em um ângulo de 45° e uma fileira de

sensores CCD ou CMOS que capturam a luz refletida através de um arranjo de lentes localizado acima dos sensores. Com o uso da tecnologia CIS, não há necessidade de um sistema óptico, lâmpadas ou filtros. As vantagens são o menor consumo de energia e custos de fabricação reduzidos. Porém, os objetos a serem escaneados devem estar em contato direto com o vidro do scanner. Caso contrário, o sensor CIS poderá escanear o objeto introduzindo distorções na imagem de saída [43].

2.4.3 *Graphic Processing Units (GPU)*

A diferença fundamental entre a CPU e a GPU está na finalidade básica para a qual cada um foi projetado. A CPU é otimizada para executar código sequencial. Possui uma lógica de controle sofisticada que permite as instruções de uma única thread serem executadas em paralelo ou mesmo fora de sua ordem sequencial, mantendo a aparência de execução sequencial. Além disso, são utilizadas grandes memórias de cache para reduzir as latências de acesso a instruções e dados. Em contrapartida, a GPU executa códigos especialmente construídos com uma grande quantidade de operações em paralelo e otimiza sua execução por meio de unidades de execução aritméticas mais simples de maior latência e uma grande largura de banda para acesso aos dados em memória. Portanto, seu design é orientado a taxa de transferência, uma vez que o esforço é concentrado em maximizar o throughput da execução total de um grande número de threads, mesmo que threads individuais levem um tempo potencialmente maior para serem executados. Ao reduzir a latência das operações em uma mesma thread, o hardware da CPU reduz a latência de execução de cada thread individual. No entanto, a grande memória cache, unidades aritméticas de baixa latência e lógica de controle sofisticada consomem área de chip e energia que poderiam ser usadas para fornecer mais unidades de execução aritméticas e canais de acesso à memória. Esse estilo de design é orientado à latência [10].

As GPUs são um tipo de hardware especializado para computação paralela e orientados para taxa de transferência muito maior que as CPUs. Esses não terão um bom desempenho em algumas tarefas nas quais as CPUs são projetadas para um bom desempenho. Para programas que possuem poucas threads, CPUs com latências de operação mais baixas podem obter um desempenho muito maior do que as GPUs. Quando um programa tem um grande número de threads, as GPUs com maior throughput de execução podem obter um desempenho muito superior ao das CPUs [10].

2.5 SOFTWARE PARA VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA

A grande demanda e necessidade do desenvolvimento de sistemas de Visão Computacional e Inteligência Artificial em um curto espaço de tempo fez com que grandes grupos desenvolvessem bibliotecas de software que pudessem ser utilizadas nas mais diversas situações. Seria impraticável o desenvolvimento de sistemas para a solução de problemas complexos

como os que existem hoje sem a utilização dos componentes de software disponíveis. Nessa seção iremos abordar algumas dessas bibliotecas.

2.5.1 OpenCV

A OpenCV (*Open Source Computer Vision Library*) [44] é uma biblioteca de código aberto distribuída sob a licença BSD e, portanto, pode ser utilizada tanto em aplicações comerciais quanto acadêmicas. Com a popularização das aplicações de visão computacional tem ganhado bastante atenção atualmente devido a grande quantidade de algoritmos de processamento de imagens que ela oferece. Suporta diferentes sistemas operacionais como Windows, Linux, Mac OS e pode ser utilizada inclusive em dispositivos móveis com sistema operacional iOS e Android. Os algoritmos por ela implementados são otimizados para explorar as características de paralelismo da CPU e utilizar inclusive, quando disponível, a GPU de forma transparente para o usuário. Dessa forma, foi desenvolvida com foco na eficiência computacional e possibilita o rápido desenvolvimento de aplicações de visão computacional. Sua grande comunidade garantem a essa biblioteca grande adoção por parte dos desenvolvedores.

A OpenCV foi projetada de forma modular. Ao longo do tempo, diversos módulos foram agregados a biblioteca expandindo o seu potencial. Atualmente, podemos destacar os seguintes módulos:

- *core*: estruturas de dados básicas incluindo arrays multidimensionais e funções básicas utilizadas pelos demais módulos;
- *imgproc*: algoritmos de filtragem linear e não-linear, transformações geométricas, conversões de espaço de cor, histogramas, entre outros;
- *video*: análise de vídeo em tempo real incluindo detecção de movimento, remoção de fundo e rastreamento de objetos;
- *calib3d*: algoritmos para calibração de câmera, estimativa de posição de objetos, visão *stereo* e reconstrução 3D;
- *ml*: algoritmos de aprendizado de máquina como PCA e SVM;
- *dnn*: detecção de objetos utilizando redes neurais convolucionais.

2.5.2 TensorFlow

De forma similar a OpenCV, a biblioteca TensorFlow [45] foi criada pelo Google e projetada para facilitar a criação e implantação de aplicações que fazem uso de técnicas de aprendizado de máquina. Essa biblioteca de código aberto utiliza grafos do fluxo de dados para representar a computação, o estado compartilhado e as operações que alteram esse estado. Explora

de maneira transparente a capacidade de computação em CPUs de múltiplos núcleos, *clusters*, GPUs e hardware especializado denominados de TPUs. Desse modo, facilita ao desenvolvedor a exploração de diferentes algoritmos.

Sua concepção é baseada no conceito de múltiplas camadas o que permite ao desenvolvedor optar pelo nível de especificidade que deseja utilizar. Na camada de baixo nível são disponibilizadas as função para manipular diretamente os tensores e o grafo de computação. Na camada de nível intermediário são oferecidas as funções para especificação de camadas de redes neurais, manipulação de conjuntos de dados e métricas para avaliação do treinamento e desempenho das redes. Na camada de alto nível é possível utilizar modelos completos pré-definidos e funções para o treinamento e geração de predições.

O *TensorBoard* compreende um conjunto de ferramentas que auxilia no entendimento do treinamento de redes neurais. Por meio dele é possível gerar métricas quantitativas sobre a computação definida pelo grafo de tensores.

2.5.3 Keras

Keras [46] é uma API de alto nível para especificação e manipulação de redes neurais criada para a linguagem de programação Python. Seu foco principal é o de possibilitar a experimentação rápida independente da biblioteca que será utilizada para computação numérica. Ao contrário da biblioteca Tensorflow, que realiza a computação, a Keras a delega para bibliotecas como CNTK, Theano e TensorFlow. Desse modo, a criação de protótipos é facilitada. Sua abrangência é ampla suportando redes convolucionais e redes recorrentes. Pode utilizar de maneira transparente tanto a CPU quanto a GPU.

A Keras facilita a construção de sistemas envolvendo o aprendizado de máquina na medida em que simplifica a utilização desses algoritmos. Oferece APIs consistentes e simples minimizando quantidade de código necessário para o uso comum. Sua característica modular permite a combinação de diferentes componentes de redes neurais para criar novos modelos específicos para o problema que se pretende tratar. Alguns exemplos desses componentes são: camadas neurais, funções de custo, otimizadores, esquemas de inicialização, funções de ativação, esquemas de regularização, entre outros. Caso os módulos desejados não estejam disponíveis, é possível facilmente adicionar módulos customizados devido sua extensibilidade.

2.5.4 Qt

Qt é um SDK multi-plataforma para o desenvolvimento e distribuição de aplicações. É utilizada em diferentes setores como o automotivo, médico, infotainment, wearables entre outros. A biblioteca oferece um grande conjunto de componentes que possibilitam ao desenvolvedor realizar seu trabalho sem se preocupar muito com as especificidades de um determinada pla-

taforma. A Qt lida com a integração e a compatibilidade dos sistemas. Dessa forma, a mesma aplicação pode ser distribuída para várias plataformas minimizando o custo de desenvolvimento da aplicação e ampliando o universo de utilização da solução.

O desenvolvimento do Qt foi iniciado em 1990 pelos programadores noruegueses Eirik Chambe-Eng e Haavard Nord. A Qt Company é o principal responsável pelo projeto do SDK porém muitas empresas e indivíduos em todo o mundo podem contribuir com ele por meio de um modelo de governança meritocrático. É distribuído sob a licença LGPL podendo ser adquiridas licenças caso deseje modificar o código fonte da biblioteca.

Oferece diversas opções para construção de interfaces para o usuários por meio das tecnologias Qt Quick (QML/JS), Qt Widgets (C++ ou Python) e Qt Web Engine (HTML/CSS/JS). Possui uma IDE própria, a Qt Creator, que auxilia na implementação utilizando essas tecnologias. Oferece componentes para a utilização de código independente de plataforma, onde a mesma base de código pode ser compilada e implantada em diferentes plataformas sem nenhuma alteração. Possui módulos estrutura de dados a exemplo de matrizes dinâmicas, filas, listas vinculadas e matrizes associativas. Auxilia na internacionalização de aplicações por meio do suporte ao Unicode. Oferece meios para suportar a comunicação entre processos e a funcionalidade para gerenciar threads e códigos paralelizados de maneira segura e independente de plataforma.

3. TRABALHOS RELACIONADOS

Revisões sistemáticas buscam coletar evidências para tratar questões de pesquisa específicas, desde que essas evidências se encaixem em critérios de elegibilidade pré-definidos. Deste modo, são um meio eficaz para identificar, avaliar e interpretar resultados relevantes para uma questão particular de pesquisa, uma área do conhecimento ou um fenômeno de interesse. Outro aspecto importante é que possuem a finalidade de reduzir o viés de pesquisa por meio da utilização de um método explícito e sistemático. Uma das razões para se realizar uma revisão sistemática é a de resumir as evidências relacionadas a um tratamento ou tecnologia, evidenciando seus benefícios e limitações [47].

A ideia de elaborar uma revisão sistemática para avaliar a aplicabilidade de visão computacional e inteligência artificial em culturas de grãos surgiu da necessidade de construir uma base de evidências que pudessem contribuir na solução do problema de pesquisa. O conteúdo integral da revisão sistemática foi publicado em 2018 [48].

3.1 PROTOCOLO DA REVISÃO SISTEMÁTICA

O protocolo especifica os métodos para condução de uma revisão sistemática específica. Sendo assim, é possível estabelecer a forma com que os estudos serão selecionados e garantir que a análise realizada não possa ser influenciada pelas expectativas do pesquisador [47]. Nesse sentido, o protocolo a seguir foi definido para a realização da revisão sistemática apresentada neste trabalho.

Background: Identificar na literatura trabalhos relacionados ao uso de técnicas de visão computacional e de inteligência artificial para as cinco maiores culturas de grãos, considerando o total de toneladas produzidas para estas culturas em 2014 [1]. Desta forma, este trabalho tem o intuito de prover um ponto de partida para trabalhos futuros nesta área.

Questões de Pesquisa: A presente revisão sistemática visa responder as seguintes questões de pesquisa:

1. Qual a aplicabilidade da visão computacional nas cinco maiores culturas de grãos?
2. Quais os principais resultados obtidos por meio do uso de técnicas de Inteligência Artificial / Deep Learning em conjunto com visão computacional na agricultura de precisão?
3. De que modo o uso de GPU (*Graphic Processing Unit*) pode aumentar a eficiência dos métodos de visão computacional estudados?

Estratégia de Busca: As bases de dados escolhidas para esta revisão são: ScienceDirect, Scopus, Springer, Web of Science, ACM (Association for Computing Machinery) e IEEE (Institute of Electrical and Eletronic Engineers).

Para o levantamento de trabalhos, definiu-se a seguinte expressão de busca: ("**computer vision**"OR "**machine vision**"OR "**artificial intelligence**"OR "**machine learning**"OR "**deep learning**"OR "**GPU**"OR "**Graphics Processing Unit**") AND (maize OR rice OR wheat OR soybean OR barley).

Devido a existência de outras revisões sistemáticas similares, o intervalo temporal de pesquisa utilizado foi definido de 2013 a 2017, ou seja, últimos 5 anos. Os estudos considerados nesta revisão devem ter sido publicados na língua inglesa em periódicos, desconsiderando capítulos de livros, anais e resumos de eventos e seminários.

Critério de Seleção dos Trabalhos: Os critérios de elegibilidade utilizados na avaliação dos trabalhos são:

1. O estudo deve fazer uso de alguma técnica de visão computacional para utilização na detecção, quantificação ou classificação de doenças em plantas, avaliação da qualidade de grãos, identificação de cultivares e avaliação de propriedades fisiológicas de plantas nas culturas de interesse;
2. O estudo deve utilizar alguma técnica de inteligência artificial para a identificação e padrões;
3. O estudo pode ou não fazer uso de GPU para otimizar o processamento do método proposto.

Procedimentos de Seleção dos Trabalhos: A avaliação dos trabalhos é realizada através das seguintes etapas: identificação, triagem e elegibilidade. Na etapa de identificação, as expressões de busca são utilizadas para obter os trabalhos de interesse. Os resultados das diferentes bases de dados são consolidados para a remoção de itens duplicados. Na etapa de triagem é realizada a avaliação de título e do resumo quanto ao enquadramento na área de interesse do presente estudo. Na etapa de elegibilidade é avaliado se o estudo atende aos critérios de elegibilidade previstos e se os resultados e a conclusão obtidos foram satisfatórios. Considera-se também a qualidade da descrição da técnica em análise e se há comparação com outros métodos similares.

Extração dos Dados: Para cada estudo elegível serão destacadas as seguintes informações: autores, ano de publicação, objetivos gerais, descrição do método utilizado para processamento de imagens, descrição do método utilizado para a classificação de imagens utilizando inteligência artificial, utilização ou não de GPU e os principais resultados obtidos.

Na fase de "Identificação", 1578 trabalhos não duplicados foram obtidos usando a expressão de busca proposta. Na fase de "Triagem" foram selecionados 183 itens para leitura e verificação integral dos resultados. Na etapa final da revisão sistemática, onde são verificados os critérios de elegibilidade, 25 trabalhos foram selecionados para inclusão na versão final da revisão sistemática. A Figura 11 apresenta o fluxograma da revisão sistemática.

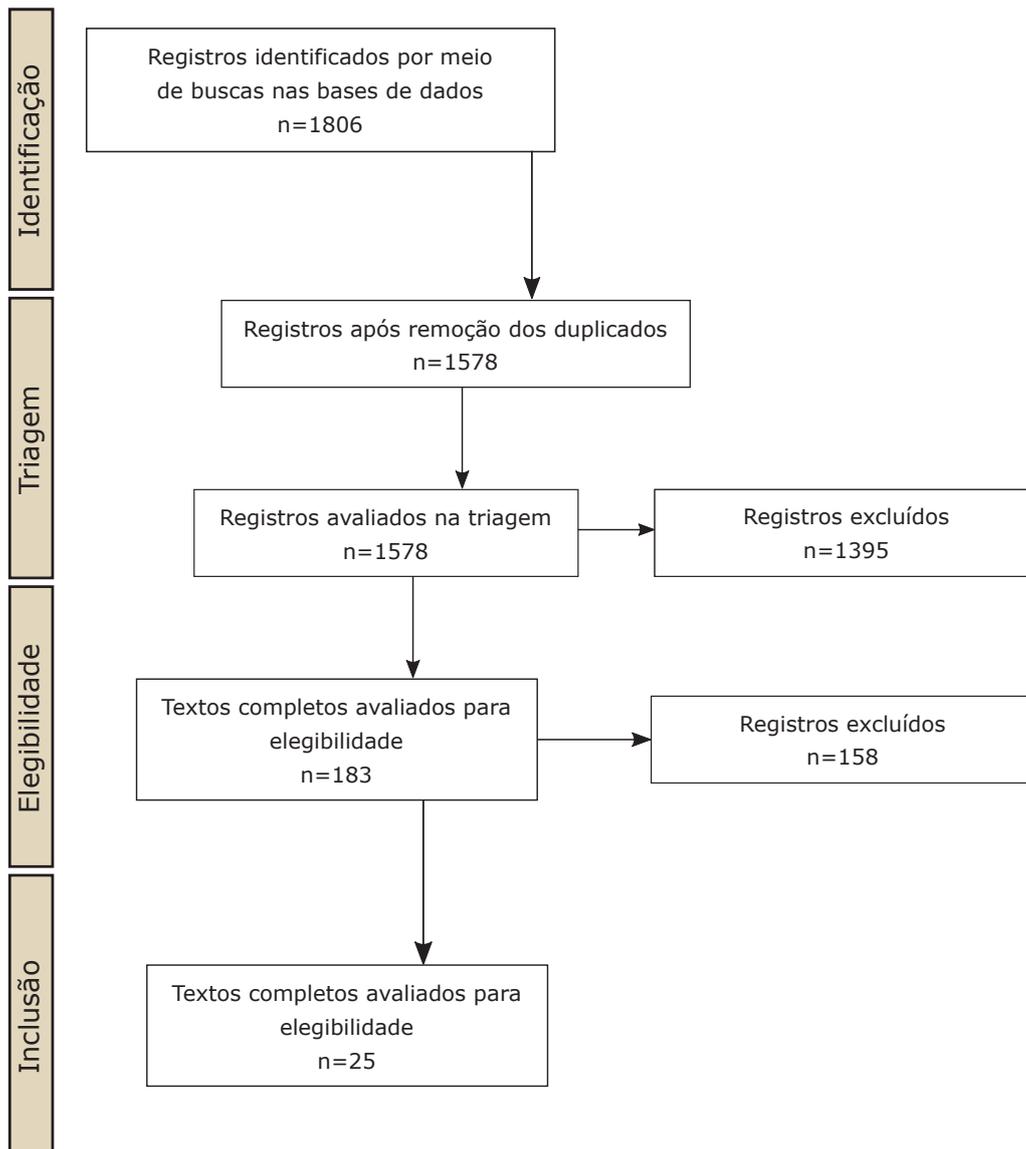


Figura 11. Fluxograma do processo de revisão sistemática.

Dos 25 trabalhos que compõe a revisão sistemática elaborada, selecionamos 10 que mais contribuíram na construção da metodologia para classificação de grãos de aveia, foco principal do presente trabalho. A Tabela 6 apresenta um sumário dos métodos empregados nos artigos selecionados.

O resumo dos trabalhos listados na Tabela 6 são apresentados nas seções a seguir.

3.2 CLASSIFICAÇÃO DE GRÃOS DE TRIGO UTILIZANDO REDE NEURAL ARTIFICIAL

Sabanci, Kayabasi e Toktas [49] apresentam um sistema de visão computacional que utiliza uma abordagem simplificada de classificação com um alto índice de acurácia. O objetivo do sistema é classificar grãos de trigo das espécies *Triticum aestivum* (comum) e *Triticum durum*

Tabela 6. Resumo dos trabalhos relacionados

#	Referência	Ano	Cultura	Dispositivo	Classificador
1	[49]	2017	Trigo	Câmera	ANN
2	[50]	2016	Arroz	Câmera	DBN
3	[51]	2016	Arroz	Câmera	SVM
4	[52]	2016	Trigo	Câmera	ANN
5	[53]	2016	Arroz	Câmera	BPNN
6	[54]	2016	Trigo	Câmera	SVM
7	[55]	2016	Arroz	Microscópio	SRC
8	[56]	2014	Soja	Câmera	BPNN
9	[57]	2014	Arroz	Câmera	SVM
10	[58]	2014	Trigo	Scanner	ANN

(duro) de acordo com suas características visuais baseado em ANN (*Artificial Neural Network*) do tipo MLP.

A imagens são obtidas por uma câmera em um ângulo perpendicular ao grão. Originalmente as imagens são obtidas em RGB. Na etapa de processamento das imagens, estas são convertidas em escala de cinza, binarizadas utilizando o método de Otsu e segmentadas utilizando o método de *thresholding*. São capturadas as características relativas a dimensão, a cor e a textura para cada grão, com o objetivo de servirem de entrada para o método de classificação. No estudo são selecionadas sete características visuais do grão: comprimento, proporção do comprimento e largura, G, B, G/TRGB, homogeneidade e entropia. As duas últimas, relativas a textura, são obtidas utilizando o método GLCM.

A ANN utilizada é composta por três camadas. O sistema de classificação foi treinado com 180 grãos e testado com 20 grãos. Destes, 19 foram classificados sem erro absoluto e 1 com erro absoluto de 195×10^{-6} . Desta forma, o autor propõe uso da rede de forma integrada com demais sistemas agrícolas de classificação automática.

3.3 RECONHECIMENTO DE COLÔNIAS DE FUNGOS EM ARROZ UTILIZANDO VISÃO COMPUTACIONAL E APRENDIZADO DE MÁQUINA

Sun *et al.*, 2016 [50] investigaram o potencial de utilizar visão computacional em conjunto com técnicas convencionais de aprendizagem de máquina e de *deep learning* para detecção de colônias de fungos em arroz causado por microrganismos como o *Aspergillus* e o *Penicillium*. A Figura 12 compara a imagem original de um amostra de grãos infectados e o resultado de detecção do método proposto.

O sistema para a obtenção das imagens é composto por uma câmera, duas fitas de led, um suporte para câmera e um suporte para a amostra em uma base preta. Na etapa de pré-processamento, é removido o fundo preto das imagens, a resolução é ajustada para 2490x2490 pixels e as coordenadas do pixel central calculadas para eliminar a interferência da placa utilizada para a deposição das amostras. A identificação das regiões infectadas é feita

por meio das informações de cor da imagem. Para isto, duas imagens de 512x512 pixels são geradas: uma em escala de cinza e outra em RGB. O vetor de 64 atributos de cor é obtido a partir do histograma de 16 níveis para a imagem em escala de cinza e para cada um dos canais de cor da imagem RGB. As áreas mofadas nas imagens das amostras não têm uma cor, textura ou formas uniformes. Devido a este fato, o método de segmentação utilizando *thresholding* é inútil para a segmentação das áreas de interesse.

Com o intuito de lidar com estas limitações, adotou-se para segmentação o método *pitch segmentation recognition*. Também é avaliado a utilização do algoritmo SPA (*Successive Projection Algorithm*) para reduzir o número dos atributos de cor das imagens de 64 para 14. Este número é validado utilizando o método RMSECV (*Root Mean Square Errors of Cross Validation*). Ao final, para a classificação são comparados os modelos SVM, BPNN (*Back-Propagation Neural Network*), CNN (*Convolution Neural Network*) e DBN (*Deep Belief Network*).

Para as taxas de acurácia dos conjuntos de testes, o modelo DBN apresentou o valor mais alto quando comparado com os demais modelos (99.4% sem SPA, 100% com SPA), seguido do modelo SVM (98.9% sem SPA, 99.4% com SPA) e do modelo BPNN (83.9% sem SPA, 95% com SPA). Outro ponto identificado é que a taxa de acurácia de classificação dos três modelos construídos utilizando SPA foi mais alta que a dos mesmos modelos construídos com o conjunto de parâmetros originais.

Os resultados indicam que a utilização do SPA é adequada para reduzir os parâmetros de entrada. Outro ponto apontado pelos autores é que o modelo SVM pode obter bons resultados para diferentes tarefas de classificação na agricultura, pois os valores obtidos são ligeiramente inferiores aos do modelo DBN. Entretanto, o desempenho do modelo SVM depende dos vetores de suporte que são extraídos a partir dos dados de entrada. Conforme o número de amostras de treinamento aumenta, o número de vetores de suporte do método SVM também aumenta, elevando a complexidade do modelo. Por último, o modelo BPNN não se mostrou estável quando o número de variáveis de entrada é grande.

3.4 UTILIZAÇÃO DE SOMBRAS PARA CLASSIFICAÇÃO DO PERCENTUAL DE GRÃOS CHEIOS DE ARROZ

O enchimento dos grãos de arroz é um fator crítico para determinar a produtividade de grãos. Liu *et al.* [51] apresentam um método para calcular o percentual de enchimento dos grãos de arroz utilizando sombras. O método utiliza quatro fontes luminosas para gerar a sombra dos grãos em quatro direções. A diferença das sombras dos grãos cheios e dos “chochos” é avaliada por meio da análise de imagens e de um classificador SVM. O sistema foi elaborado para ser utilizado como um método de avaliação on-line. Para isto, é utilizado uma correia de transporte de grãos que permite a análise a uma velocidade de 40 a 50 grãos/s.

Na etapa de pré-processamento, as imagens RGB das sementes com as sombras na quatro direções são segmentadas utilizando binarização. A partir da imagem binarizada é pos-

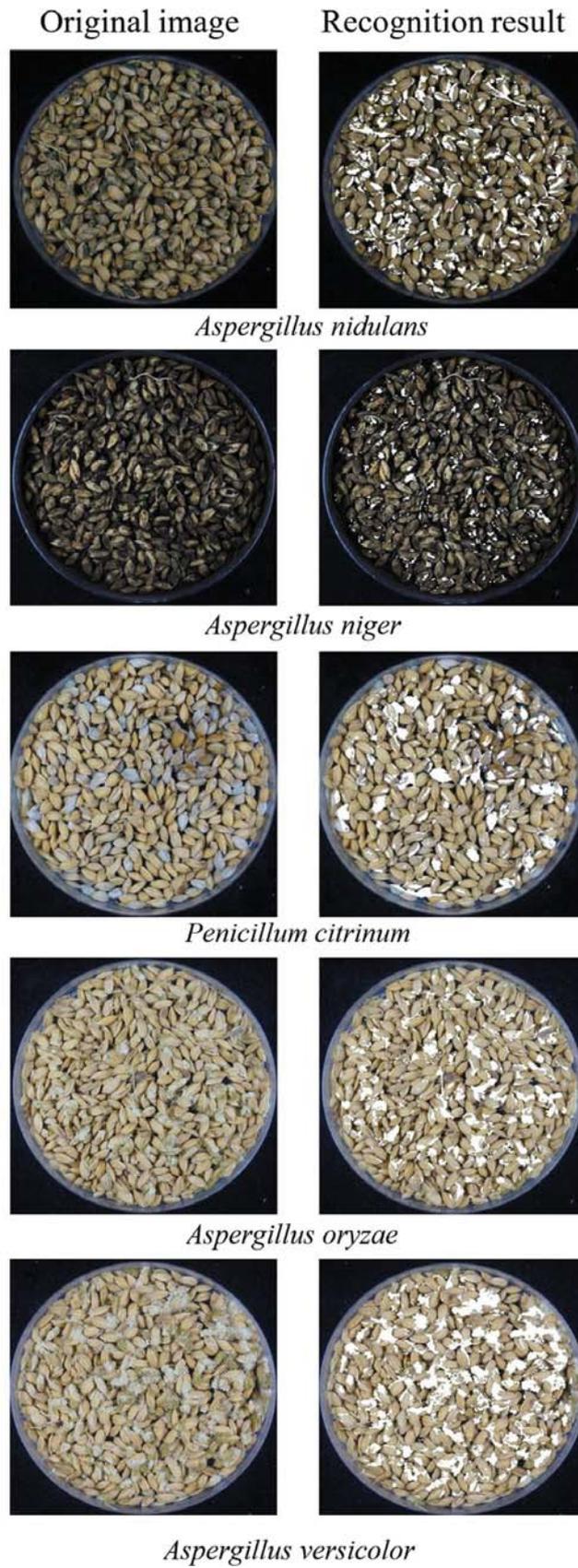


Figura 12. Reconhecimento de colônias de fungos em arroz utilizando visão computacional e técnicas de aprendizado de máquina [50].

sível extrair informações como a área do grão e a área da sombra. AR é a proporção da área retangular circunscrita da sombra para a do grão, enquanto que DR é a proporção da distância entre o centróide e o retângulo circunscrito da sombra a do grão. Ambas as medidas podem ser utilizadas para distinguir os grãos cheios dos “chochos”. Na Figura 13 é possível perceber a diferença das sombras geradas por estes dois tipos de grãos. O modelo de classificação adotado é o SVM.

A média de falso negativo e de falso positivo para a variedade Indica foi de 3,85% e 5,44%, respectivamente. Os resultados indicam que o método é confiável e pode ser utilizado como uma forma de avaliação rápida dos grãos de arroz. Como os grãos cheios são mais espessos que os “chochos”, a sombra destes será mais larga. Deste modo é possível utilizar um classificador para separar os grãos.

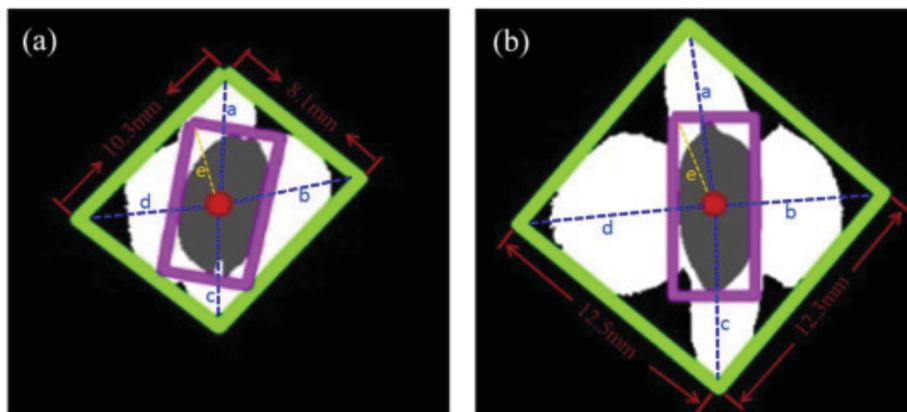


Figura 13. Estrutura de grão cheios e “chochos” e suas sombras: (a) “chochos” (b) cheios [51].

3.5 ABORDAGEM DE VISÃO COMPUTACIONAL UTILIZANDO DUAS CÂMERAS PARA IDENTIFICAÇÃO DE GRÃOS GERMINADOS DE TRIGO

A germinação precoce do grão de trigo ocorre quando as condições do campo apresentam um índice elevado de umidade antes ou durante a colheita. Esta condição acarreta na redução da qualidade do grão. Shrestha *et al.* [52] propõem um sistema de visão computacional que utiliza duas câmeras para separar e identificar grãos de trigo germinados em laboratório. O sistema possibilita a classificação em três categorias: grãos saudáveis, danificados por germinação e severamente danificados por germinação. A Figura 14 mostra a diferença de grãos saudáveis e grãos danificados pela germinação precoce.

O sistema de aquisição das imagens é composto por duas câmeras: uma posicionada no topo e outra posicionada na base. A técnica de segmentação adotada é a *Marker Controlled Watershed*. O algoritmo de *Watershed* sempre produz contornos fechados dos objetos além de ser computacionalmente viável. A abordagem do uso de marcadores visa lidar com o problema da segmentação exagerada dos objetos devido a natureza e ao ruído presente nas imagens.

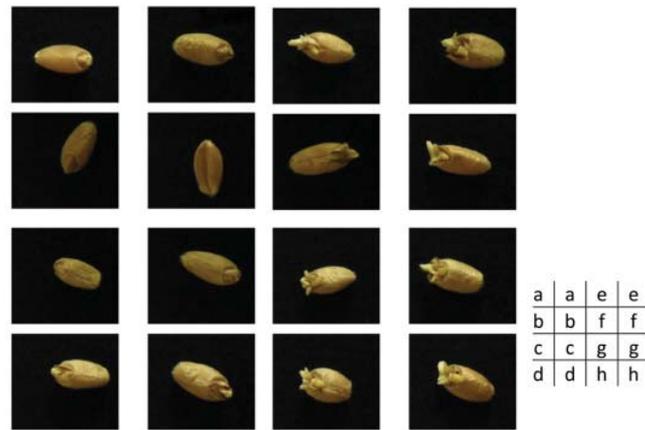


Figura 14. Imagens obtidas de grãos de trigo sadios (a) e danificados por germinação após 1 (b), 2(c), 3(d), 4(e), 5(f), 6(g) e 7(h) semanas [52].

Também é importante evitar que os grãos não sejam sobrepostos para melhorar a acurácia do método de classificação.

As imagens produzidas pelas duas câmeras são utilizadas para extrair as características morfológicas e não morfológicas dos grãos. As características morfológicas utilizadas são: perímetro, excentricidade elíptica, comprimento do eixo principal e relação do comprimento do casco convexo do grão com seu perímetro. As características não morfológicas utilizadas são: coloração RGB, desvio padrão da escala de cinza e textura. Grãos danificados por germinação usualmente apresentam a coloração amarela, verde ou marrom escuro podendo, ainda, apresentar coloração esbranquiçada. Tais alterações de cor não estão presentes em grãos sadios. Para melhorar a acurácia do método de classificação é importante que sejam evitados os grão sobrepostos na imagem.

Para classificação, o método utilizado é uma ANN que recebe 16 parâmetros de entrada. Do total de 555 grãos, 185 de cada classe, 389 (70%) foram selecionados aleatoriamente para o treinamento, 83 (15%) para a validação e 83 (15%) para o teste. O método apresenta acurácia de 72% e o tempo de processamento foi de 45s para o conjunto de teste utilizado.

3.6 CLASSIFICAÇÃO DE GRÃOS DE ARROZ UTILIZANDO BPNN E DECOMPOSIÇÃO BASEADA EM WAVELETS

Singh e Chaudhury [53] propõem a utilização de redes neurais com propagação reversa (BPNN) e decomposição utilizando wavelets para a classificação de grãos de arroz.

A abordagem utiliza dois tipos de câmeras para a aquisição das imagens: a primeira, presente em celulares, com resolução de 0.3 megapixels e a segunda, convencional do tipo *point-and-shoot* de 12 megapixels. As imagens são obtidas sob condição de luz natural no período da manhã para reduzir a interferência da sombra na imagem.

A extração de atributos das imagens é realizada utilizando três abordagens. A primeira consiste em extrair 18 atributos de cor (RGB e HSI). A segunda e a terceira abordagem consistem em utilizar, respectivamente, o método GLCM e a decomposição baseada em wavelets para cada um dos canais de cor da imagem (R,G e B). O método GLCM visa tratar os atributos de textura. Estes se referem a um conjunto conectado de pixels que ocorrem repetidamente em uma imagem e estão relacionados com a variação de intensidade em uma superfície. Para a descrição dos atributos de textura, os modelos mais aceitos são aqueles baseados no uso de co-ocorrências. Na transformada discreta por wavelets, um sinal de imagem pode ser analisado utilizando um conjunto de filtros seguido de uma operação de decimação.

O modelo de classificação utiliza uma rede neural BPNN de quatro camadas. O estudo compara ainda o método proposto com outros classificadores: SVM, k-nearest neighbour e naive Bayes. O classificador proposto apresentou 96% de acurácia. Dentre os demais resultados obtidos, podemos destacar três. O primeiro mostra que foi necessário a utilização de uma função de *thresholding* adaptativo para garantir a convergência da rede. O segundo aponta que a abordagem que utiliza tabela de GLCM demanda maior processamento que a baseada em wavelets. E, por último, a utilização de BPNN apresentou melhores resultados quando comparado com os demais métodos.

3.7 CLASSIFICAÇÃO DE GRÃOS DE TRIGO UTILIZANDO DSIFT E CLASSIFICADOR SVM

Olgun *et al.* [54] avaliam o desempenho da utilização da técnica DSIFT em conjunto com um classificador SVM para classificar grãos de trigo em 40 espécies diferentes.

SIFT é um algoritmo de visão computacional para representar e identificar objetos com algum atributo local diferenciado. O algoritmo DSIFT é derivado do algoritmo SIFT. A principal diferença entre eles é que o DSIFT utiliza um procedimento de amostragem para reduzir o custo do SIFT.

Inicialmente, a abordagem dos autores extrai os atributos DSIFT da imagem. Na sequência, o método de *clustering* k-means é utilizado sobre este conjunto de atributos. Por último, é construído o modelo BOW a partir do histograma dos atributos de clustering obtido no passo anterior. Os objetos de interesse são classificados, utilizando o modelo BOW, por meio de um classificador SVM. Antes da construção do modelo BOW, a dimensão do conjunto de atributos obtidos por DSIFT deve ser reduzida para otimizar o processamento e melhorar a acurácia. Neste sentido é aplicado o método de *k-means clustering*. A Figura 15 apresenta um modelo esquemático do método.

O desempenho do método foi extensamente avaliado com um conjunto de 6400 imagens. O método apresenta uma acurácia satisfatória de 88.33%.

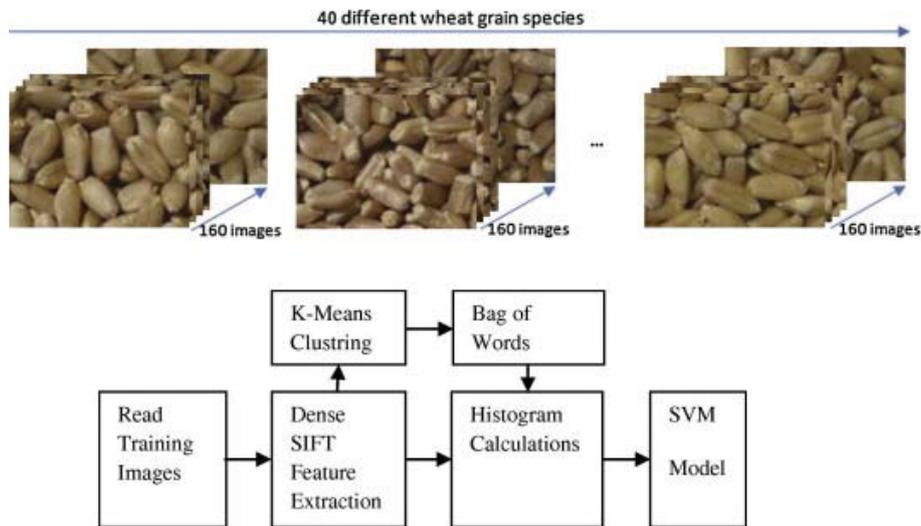


Figura 15. Classificação de grãos de trigo utilizando DSIFT e classificador SVM [54].

3.8 IDENTIFICAÇÃO DE CULTIVARES DE ARROZ UTILIZANDO ANÁLISE DE IMAGENS

Kuo *et al.* [55] desenvolveram um sistema de aquisição de imagens de alta resolução de grãos para classificação de 30 variedades de arroz (Figura 16).

As características morfológicas, textura e cor dos grãos de arroz são quantificadas utilizando processamento das imagens obtidas por meio de um microscópio. O corpo do grão de arroz é tipicamente composto pela casca e pelo lema. As características destes órgãos são utilizadas para distinguir as variedades. A técnica MIF (*Multifocus Image Fusion*) é utilizada para melhorar a qualidade das imagens obtidas. Esta técnica é empregada para minimizar as limitações do campo de visão do microscópio. As imagens dos grãos são segmentadas e tem seu fundo removido. A operação é realizada utilizando-se a técnica *k-Means* no espaço de cor HSV. A pilosidade do grão é removida por meio de operações morfológicas para que esta não interfira na determinação do perímetro do grão. O lema, que possui coloração diferente da casca, é identificado utilizando um limiar de cor. As características morfológicas calculadas para o grão são: perímetro, área da superfície, comprimento do eixo maior, comprimento do eixo menor, proporção entre os eixos, proporção entre os arcos, desvio padrão dos raios, proporção de Haralick e proporção entre área e perímetro. As características de cor são obtidas da casca. São elas: valores de RGB, valores de HSV e valores CIE $L^*a^*b^*$. Estes valores são calculados apenas para duas regiões de interesse: a primeira, centrada no grão e 50% menor que os comprimentos dos eixos, e a segunda, correspondente a região segmentada do lema. As características de textura são: pilosidade e seis níveis de cinza utilizando a técnica GLCM. São obtidos também os descritores de Fourier para o grão.

O classificador, que utiliza a técnica SRC (*Sparse Representation Based Classification*), foi construído para identificar as cultivares de arroz. O técnica SRC é um algoritmo de apren-

dizado de máquina adequado para resolver problemas com grande número de dimensões. Ele codifica as características representativas das amostras de treinamento no formato de átomos em um dicionário. Quando uma amostra é consultada, inicialmente ela é codificada da mesma forma. Sendo assim, será atribuída a ela a classe que apresentar o menor número de erros de codificação.

A acurácia do método é de 89.1% e o desvio padrão de 7.0%. Uma das limitações identificadas quando utiliza-se a abordagem de classificação baseada em imagens é que esta requer características morfológicas e de cor distintas. O classificador SRC não foi capaz de classificar corretamente variedades que possuem características muito semelhantes. Porém, de modo geral a técnica é capaz de distinguir de maneira não destrutiva as 30 variedades de arroz propostas no estudo. Um aspecto que contribuiu para isto foi o fato da utilização de imagens obtidas utilizando-se um microscópio. Deste modo é possível observar os detalhes finos dos grãos de arroz.

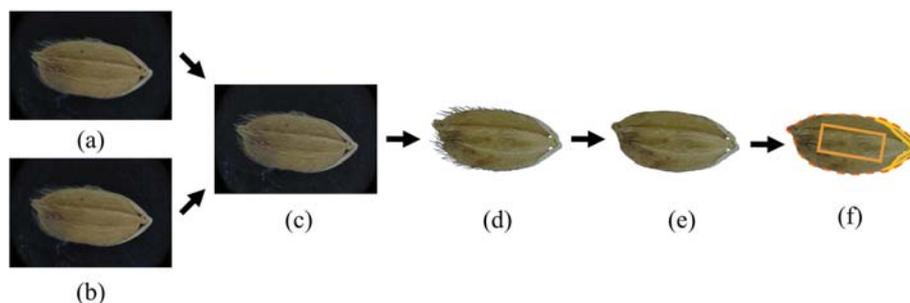


Figura 16. Método de classificação de grãos de arroz utilizando as técnicas de SRC e MIF: Grãos com o centro (a) e a borda (b) em foco; (c) fusão das imagens (a) e (b); (d) imagem com o fundo removido; (e) imagem com a pilosidade removida; e (f) imagem com a região de interesse e o lema delimitada [55].

3.9 IDENTIFICAÇÃO DE DOENÇAS EM SEMENTES DE SOJA UTILIZANDO BPNN

Kezhu *et al.* [56] propuseram um método de visão computacional combinado com BPNN para identificar doenças em sementes de soja como *soybean frogeye*, *mildewed soybean*, *worm-eaten soybean* e grãos danificados.

A aquisição das imagens é realizada a partir de um dispositivo específico para o experimento. Este dispositivo é composto por uma caixa de tamanho 100cm x 30cm x 30cm que não permite a penetração de luz exterior. Em sua base é colocado um tecido preto que impede a reflexão de luz. O sistema possui acoplado uma fonte de luz em formato de anel capaz de distribuir luz igualmente em todas as direções. A câmera é posicionada 60cm acima dos grãos, permitindo com que as imagens adquiridas sejam mais nítidas, acuradas e fáceis de se analisar. O sistema conta ainda com um iluminômetro para medir a intensidade de luz. O método METS (*Minimum Error Threshold Selecting*) é utilizado para separação do plano de fundo da imagem.

Os buracos e ruídos são tratados por meio de operações morfológicas. Ao final, o algoritmo *Watershed* é utilizado para tratar casos em que os grãos estão muito próximos. A etapa de extração de características consiste em obter informações morfológicas, de cor e parâmetros de textura dos grãos. O algoritmo de PCA é utilizado para reduzir a dimensionalidade do vetor dos atributos da imagem de 39 para 12 dimensões mantendo 99% dos dados primários do experimento.

A rede neural proposta foi modelada com três camadas. O método proposto apresenta taxas de êxito entre 96% e 90% na detecção das doenças tratadas no experimento.

3.10 ANÁLISE DE GRÃOS GESSADOS DE ARROZ UTILIZANDO SVM

Gessamento é um importante fator para determinar a qualidade do arroz e é causado por fatores adversos tais como a colheita de grãos imaturos e com alto teor de umidade. Normalmente a avaliação é realizada por inspeção manual. Entretanto, a determinação do percentual de gessamento dos grãos deve ser idealmente medida a partir de grãos distribuídos aleatoriamente de forma acurada e rápida, especialmente para melhorar a eficiência do processo de qualidade. Um desafio para a detecção automática são os grãos conectados que invariavelmente estão presentes na amostra e separá-los manualmente reduziria significativamente a eficiência do processo.

Sun *et al.* [57] propuseram um sistema automatizado de análise do percentual de grãos gessados no arroz. O algoritmo trata da separação dos grãos automaticamente, avaliando as áreas convexas em dois grãos conectados e segmentando-os com uma linha entre os cantos. O primeiro passo do método é identificar as regiões de grãos conectados em uma imagem binarizada. Grãos de arroz não conectados são ignorados neste passo. Para cada região identificada é subtraída a área dos grãos do polígono convexo mínimo e determinada a localização dos cantos dos grãos (Figura 17). Assim é possível separá-los conectando estes pontos.

Após a segmentação, é possível obter o número de grãos presentes na imagem e as informações de gessamento do grão. Sabe-se que as áreas gessadas são facilmente detectadas a a partir de sua tonalidade de cinza. Porém, quando o germe não é extraído totalmente a acurácia do método pode ser afetada. Para lidar com esta situação, o sistema detecta a localização da zona de provável de gessamento e considera esta informação no processo de classificação (Figura 18).

A classificação dos grãos é feita utilizando o método SVM. A acurácia do método é de aproximadamente 98%.

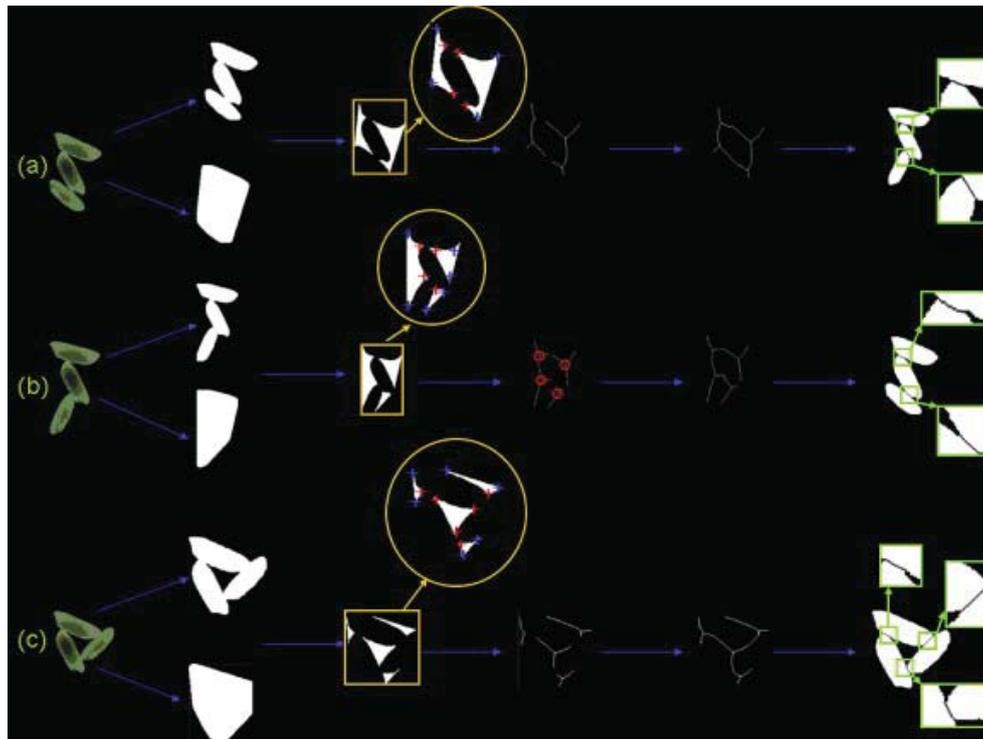


Figura 17. Processo de separação de três grãos conectados [57].

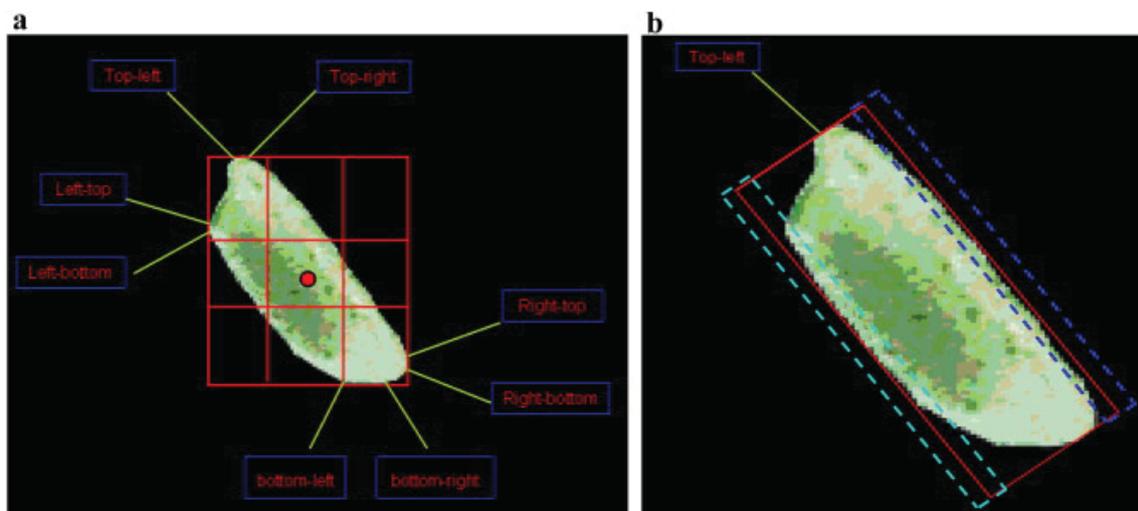


Figura 18. Características geométricas do grão do arroz [57].

3.11 SISTEMA DE MEDIÇÃO AUTOMÁTICA DA PUREZA DO TRIGO

Ebrahimi, Mollazade e Babaei [58] apresentam um sistema de visão computacional que utiliza redes neurais e algoritmo imperialista para detectar automaticamente a pureza do trigo. Pureza é um importante fator que precisa ser considerado no preço do trigo.

Os dados experimentais utilizados consistem em 52 parâmetros de cor, morfologia e textura extraídos de imagens. O algoritmo combina ICA (*Imperialist Competitive Algorithm*) e

ANN para atingir dois propósitos: encontrar o conjunto dos melhores parâmetros e criar um sistema robusto de classificação. O sistema é treinado para classificar em grãos de trigo e impurezas, classes de grãos de trigo e classes de impurezas.

A aquisição de imagens é realizada por meio de um scanner de mesa. A segmentação é composta por duas etapas. Na primeira é realizada a eliminação de sombras utilizando os espaços de cor RGB e HSI e operações matemáticas de mascaramento. A segunda etapa, a remoção de ruído e de pequenos objetos é realizada estabelecendo um limiar. Áreas menores que 15 pixels são removidas. A seleção de atributos que caracterizem os objetos de interesse é um dos mais significativos desafios no reconhecimento de padrões. O propósito do processo de seleção de características é identificar o número ideal de atributos que gerem os resultados esperados nos algoritmos de classificação. Este processo pode ser não intuitivo, porque os parâmetros que tem um desempenho baixo quando utilizados separadamente podem ter resultados melhores quando utilizados em conjunto. ICA é utilizado como uma forma de otimizar o número de características extraídas.

A acurácia do método obtida no testes foi de 96,25% para a classificação de grãos de trigo vs impurezas, de 87,50% para classes de grãos de trigo e 77,22% para classes de impurezas.

3.12 CONCLUSÕES DA REVISÃO SISTEMÁTICA

De forma geral, os trabalhos analisados utilizam a mesma sequência de operações: aquisição, processamento e classificação de imagens. No entanto, para cada situação específica, é necessário a seleção dos algoritmos mais adequados considerando as características individuais do problema que será sendo tratado. Não há um método genérico aplicável às culturas sem que ajustes em sua estratégia de detecção e processamento de imagens sejam necessários.

A segmentação de imagens também desempenha um papel importante na obtenção do resultado desejado. Embora essa seja uma tarefa mais simples nas imagens obtidas em ambientes controlados, essa tarefa não é trivial quando as imagens são obtidas diretamente do campo. Em tal situação, a variabilidade da iluminação, a complexidade do plano de fundo e a posição da captura da imagem são alguns dos fatores que tornam essa tarefa complexa. Nesse cenário, o uso da inteligência artificial já no estágio de segmentação pode trazer resultados positivos. Outro grande desafio a ser abordado pelos algoritmos é a variabilidade das características fisiológicas das plantas, pois, quando combinadas, aumentam ainda mais a diversidade de fatores que precisam ser analisados no processamento das imagens.

Nenhum dos trabalhos revisados teve como objeto de estudo a classificação de grãos aveia. Embora tenham sido abordados grãos de outras espécies, parece que existe uma lacuna no estudo de métodos de classificação para essa cultura. A importância da aveia na alimentação humana e animal tem motivado o trabalho de programas de melhoramento genético em diferen-

tes instituições de pesquisa. Desse modo, avaliar se a tecnologia de classificação utilizando inteligência artificial é capaz de classificar grão de aveia certamente é um dos diferenciais desse trabalho e uma grande ferramenta para esses programas. Outro ponto que cabe destacar é que a utilização de *deep learning* para identificação de material vegetal tem evoluído bastante recentemente. Sendo assim, é importante avaliar se os recentes avanços conquistados trouxeram ganhos no contexto em que se pretende empregar essa tecnologia.

4. MATERIAIS E MÉTODOS

A metodologia para a identificação de espécies e classificação de cultivares de aveia consiste na utilização de técnicas de visão computacional e inteligência artificial agrupadas em uma sequência lógica. A essa solução damos o nome de SeedFlow. Como visto anteriormente, um sistema de visão computacional, na etapa de processamento de alto nível, faz uso de um classificador para extrair informações relevantes da imagem que está sendo processada. No caso da solução proposta, a ideia é de utilizar um classificador baseado em redes neurais convolucionais pré-treinado para extrair as informações de espécie e de cultivar das imagens. Redes convolucionais são inspiradas no funcionamento do córtex visual. Sua arquitetura é composta por diversas camadas convolucionais, *pooling* e totalmente conectadas. Em linhas gerais, seu processo consiste em decompor a imagem em diferentes níveis de complexidade permitindo, assim, que estruturas possam ser identificadas pelas camadas totalmente conectadas. É um tipo de arquitetura de rede neural adequado para o processamento de imagens digitais.

A Figura 19 apresenta a visão geral da arquitetura do sistema SeedFlow. O sistema é composto por dois aplicativos, **seedflow.ui** e o **seedflow.ai** e uma biblioteca de software **seedflow.core**. Foi adotada como linguagem de programação o Python. Os componentes de software externos utilizados foram as bibliotecas OpenCV, TensorFlow e Keras. A finalidade de cada bloco da arquitetura da solução é apresentada abaixo:

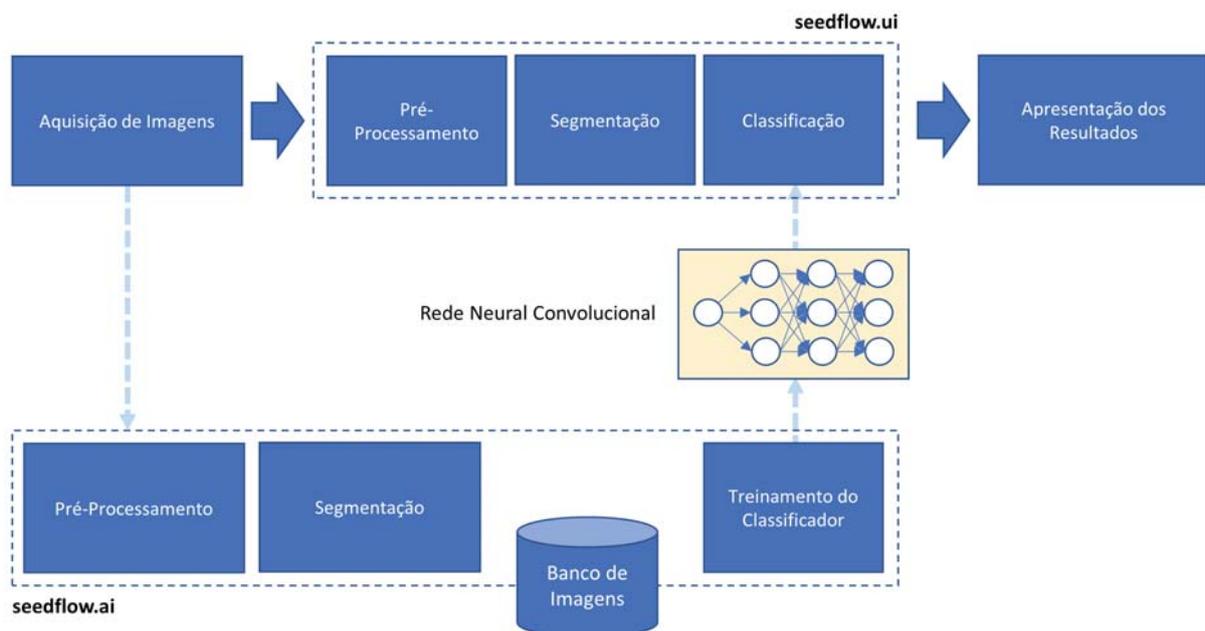


Figura 19. Visão geral da arquitetura do sistema SeedFlow.

- **seedflow.core**: Biblioteca de software que agrega os algoritmos comuns as duas aplicações. Nessa biblioteca são implementados os algoritmos de pré-processamento, processamento e de classificação de imagens.
- **seedflow.ui**: Aplicativo desktop que permite ao usuário classificar os grão presentes em uma imagem utilizando uma rede neural pré-treinada. Além da classificação, é possível visualizar também as características morfológicas de cada grão como por exemplo: comprimento do eixo maior e menor, perímetro e área do grão na imagem.
- **seedflow.ai**: Scripts utilizados a partir de linha de comando para treinamento das redes neurais convolucionais e manipulação de bibliotecas de imagens de grãos.

O ambiente de desenvolvimento utilizado para a construção da solução SeedFlow é apresentado na Tabela 7. Como pode ser verificado, o sistema não necessita de nenhum hardware especial exceto uma GPU. O custo de implantação da solução é inferior do que as soluções similares disponíveis no mercado. Hoje, com a popularização das GPUs, é comum sua utilização em estações de trabalho. Além disso, a escolha da linguagem de programação Python tem a vantagem de permitir que solução possa ser utilizada em diferentes sistemas operacionais como Windows, Linux e Mac OS.

Tabela 7. Ambiente utilizado para o desenvolvimento da solução SeedFlow.

Hardware	Software
CPU: Intel Core i7 6700	Windows 10
Memory: 16 GB DDR4	CUDA 9.0 + CuDNN 6.1
GPU: NVIDIA GTX1070 (8GB GDDR5)	Python 3.6
Scanner: Epson V370	Keras
	Tensorflow

Os componentes do sistema são detalhados nas próximas subseções.

4.1 AQUISIÇÃO DE IMAGENS

A aquisição das imagens é feita utilizando um scanner fotográfico equipado com sensor CCD (Charged Coupled Device). Um exemplo desse tipo de equipamento é o Epson Perfection V370 utilizado nos testes da solução. Os scanners que utilizam sensores do tipo CIS (Contact Image Sensor) não são adequados para serem utilizados em conjunto com a solução SeedFlow devido à sua limitação de digitalizar objetos tridimensionais. Os grãos, devido à sua forma cilíndrica, não terá toda a sua superfície em contato direto com o vidro do scanner. Os scanners do tipo CIS possuem uma limitação quanto a profundidade que é possível ser digitalizada. Sendo assim, a aquisição da imagem para o processo de classificação dos grãos será comprometida caso seja adotado esse tipo de equipamento.

O fundo branco padrão do scanner foi substituído para melhorar a eficácia da etapa de remoção de fundo. Foi utilizado um papel cartão Colorplus 180g na cor azul. Essa escolha foi adotada após comparar os resultados entre o grão escaneado com o fundo branco e o fundo azul. A utilização do fundo azul possibilitou maior acurácia na remoção do fundo da imagem. Mais detalhes serão apresentados na Seção 4.2. Outra decisão tomada na etapa de aquisição foi a de não tratar a ocorrência de sobreposição de grãos. A intenção inicial foi de obter as características morfológicas do grão além de sua classificação. Logo, para que as medidas fossem realizadas de modo mais acurado optou-se por restringir o processamento de grãos sobrepostos. Para facilitar a disposição dos grãos no scanner, foi elaborado em uma impressora 3D um espaçador para disposição dos grãos. O resultado é apresentado na Figura 20



(a) Espaçador de Grãos.

(b) Scanner.

Figura 20. Material utilizado para aquisição das imagens.

4.2 PRÉ-PROCESSAMENTO

A etapa de pré-processamento tem como objetivo principal remover o segundo plano da imagem. Inicialmente, as imagens RGB adquiridas por meio de um scanner são convertidas para o espaço de cores HSV. Usando esse espaço de cores, o canal de matiz (H) é selecionado para separar os grãos e o plano de fundo. O vetor de tonalidade azul é o que possui a maior distância angular quando comparado com as tonalidades predominantes nos grãos. Isso confirma a maior acurácia na remoção do fundo quando adotado o fundo azul. Na Figura 21 é apresentada uma imagem dos grãos de aveia no espaço de cor HSV. Note que a componente de matiz H (Hue) exibe tonalidade distintas para o fundo e para grão. Desta forma, é possível obter resultados mais acurados ao utilizar o algoritmo de limiarização automática.

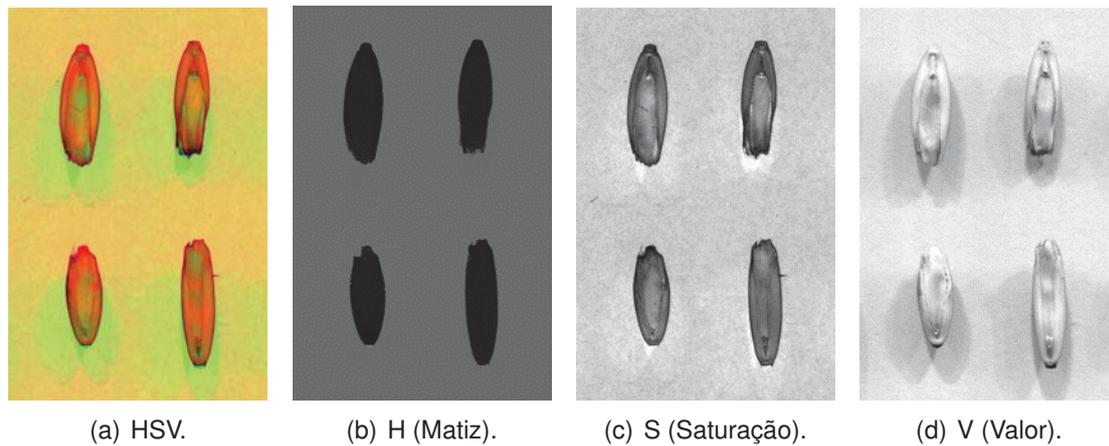


Figura 21. Decomposição no espaço de cor HSV.

O método de limiarização automática adotado é o Otsu [59]. O método parte do princípio de que a imagem consiste em apenas duas classes: primeiro plano e fundo. A heterogeneidade do fundo é ignorada. Desse modo, o Otsu estabelece qual limiar minimiza a sobreposição das duas classes. O resultado será o valor limite que separa o conjunto dos níveis de intensidade presentes na imagem. Ao primeiro grupo é atribuído valor 0 e ao segundo o valor máximo de intensidade na imagem (por exemplo, 255).

Em seguida, é utilizado um método de preenchimento para eliminar os buracos na máscara binária obtida no passo anterior. Esse passo não é determinante para o sucesso da metodologia, porém, possibilita um resultado visualmente melhor para a imagem final além de possibilitar que o valor obtido para o perímetro do grão seja mais preciso. As bordas dos grãos podem conter pequenas reentrâncias que causam esses artefatos. Sendo assim, sua remoção é desejada. O método consiste na aplicação de uma sequência de operações morfológicas e do algoritmo Watershed. Dada uma imagem de entrada X , é inicialmente aplicado o algoritmo de Watershed na posição $(0,0)$ da imagem com o valor de intensidade 255 resultando em Y . Depois, a imagem é invertida, $!Y$. Por último, a imagem final F é obtida pela operação binária OR entre a imagem de entrada X e $!Y$, ou seja, $F = X | !Y$. A sequência de operações aplicadas na imagem de entrada é apresentada na Figura 22

A remoção do fundo é obtida utilizando uma operação de subtração binária entre a imagem de entrada e a máscara gerada. O resultado da etapa de pré-processamento é mostrado na Figura 23.

4.3 SEGMENTAÇÃO

A etapa de segmentação inicia-se com a detecção das bordas. As bordas e as caixas delimitadoras são importantes para definir as coordenadas que serão utilizadas para segmentação da imagem. Sendo assim, a partir da imagem binarizada, é aplicado o método de definição

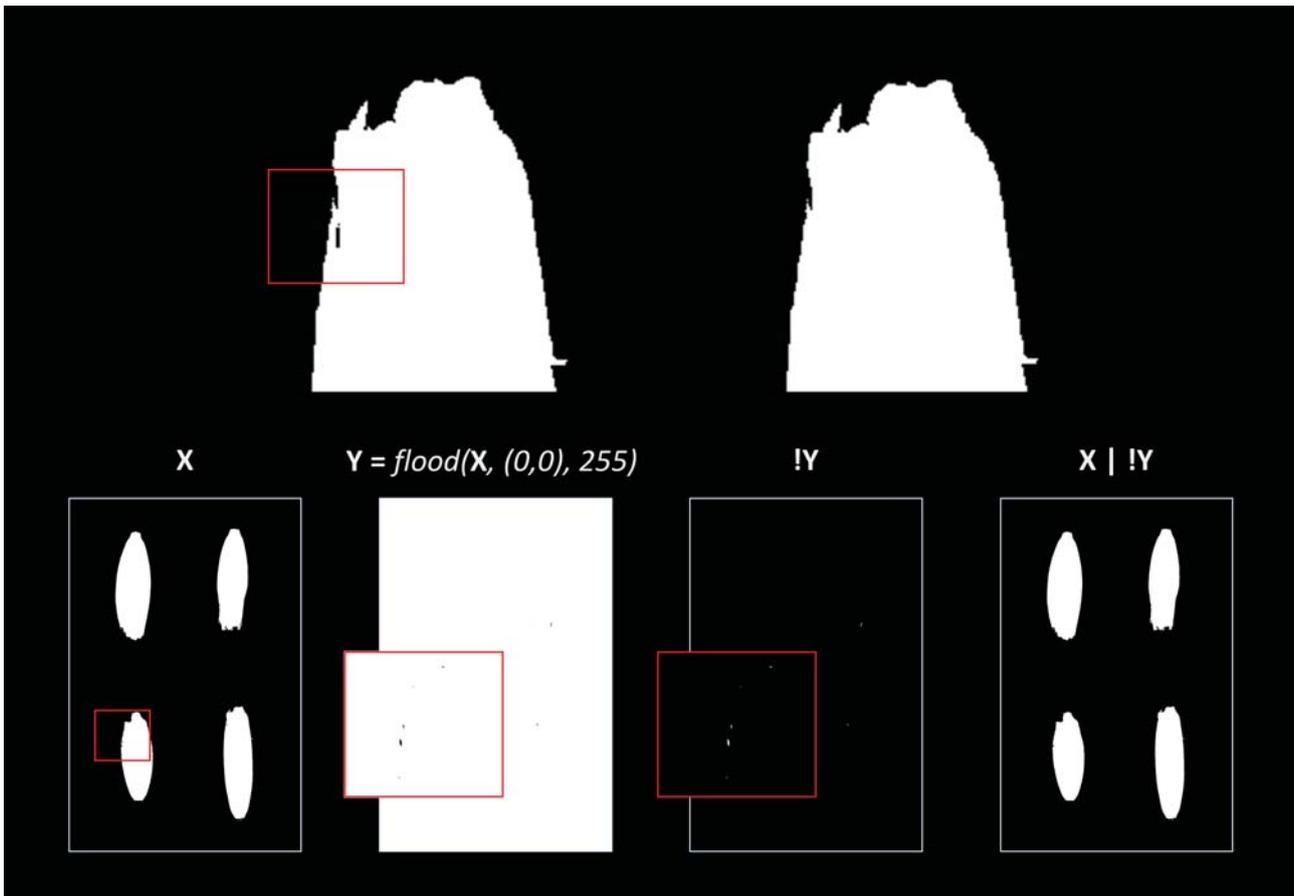


Figura 22. Visão geral do processo de preenchimento de buracos na imagem.

de contornos [60] (Figura 24). Em alguns casos, pequenos objetos podem ter sido detectados devido a ruídos ou pequenos artefatos capturados na imagem. A esse respeito, objetos com área menores do que 10% do tamanho da área mais significativa dos contornos encontrados na imagem são automaticamente ignorados. Finalmente, a imagem é segmentada usando as informações dos contornos resultantes. Para cada contorno é extraída a região de interesse mantendo as dimensões originais do grão. Cada segmento é transformado em uma nova imagem quadrada com fundo preto. Essa abordagem permite que cada segmento seja redimensionado para o tamanho mais apropriado de cada algoritmo de classificação sem que as proporções da imagem original seja modificada.

A sequência de operações das etapas de pré-processamento e segmentação é resumida na Figura 25.

4.4 TREINAMENTO DO CLASSIFICADOR

Dois casos foram considerados neste trabalho: qualidade de grãos e avaliação de propriedade intelectual. Para o primeiro caso, a identificação de diferentes espécies de aveia foi escolhida pois esse é um parâmetro comumente avaliado na análise de pureza de amostras de



Figura 23. Remoção do fundo de uma imagem de grãos de aveia.

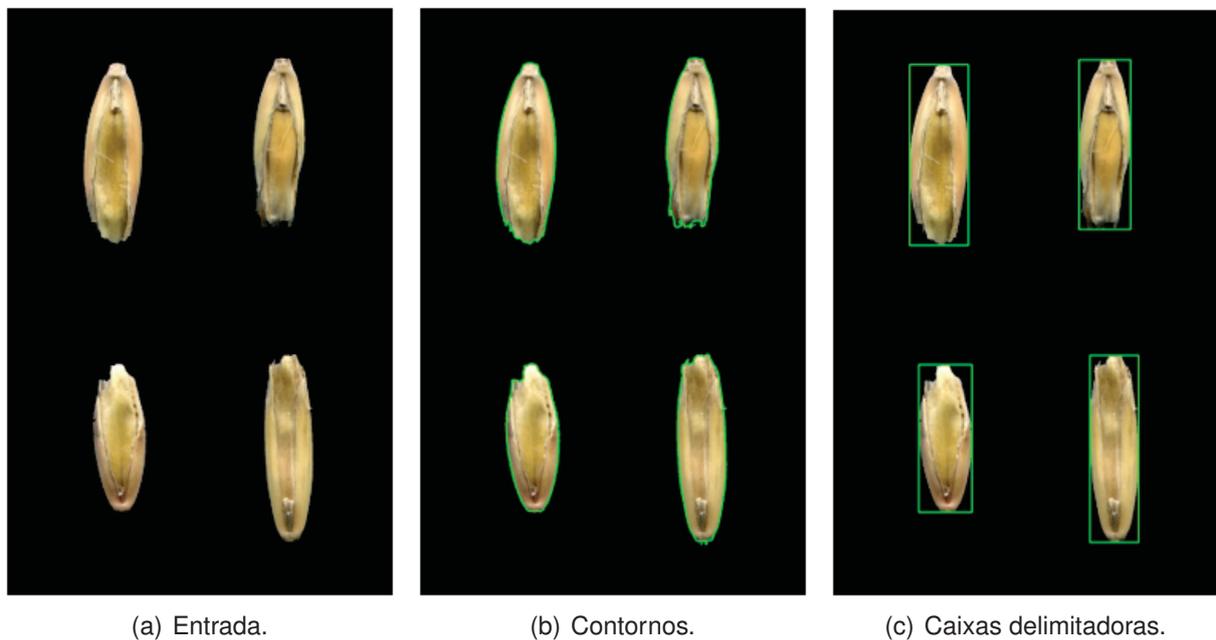


Figura 24. Identificação de contornos e caixas delimitadoras de uma imagem de grãos de aveia.

grãos. Para o segundo caso, a classificação de diferentes cultivares de aveia foi escolhida uma vez que essa informação é benéfica para programas de melhoramento e para os processos de verificação de propriedade intelectual.

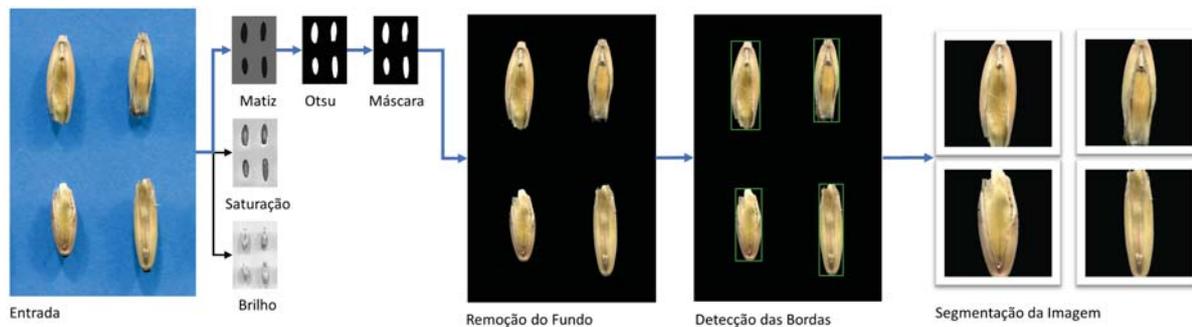


Figura 25. Visão geral das etapas de pré-processamento e segmentação.

Para que um classificador baseado em redes neurais convolucionais possa ser treinado é necessário inicialmente uma grande quantidade de imagens previamente anotada com suas respectivas classes as quais pretende-se classificá-las. Nesse sentido, um banco de imagens coloridas de grãos foi criado utilizando os processos de aquisição, pré-processamento e segmentação previamente apresentados. As imagens, inicialmente de tamanhos variados, foram convertidas para o tamanho 224×224 . Esse tamanho foi escolhido pois é possível utilizá-lo em todas as redes neurais convolucionais analisadas neste trabalho.

O banco de imagens é composto por dois conjuntos diferentes, um para a identificação de espécies de aveia e outro para a classificação de cultivares de aveia. A Tabela 8 e a Tabela 9 apresenta a subdivisão de categorias e a quantidade de imagens relacionadas a cada conjunto. A Figura 26 mostra algumas das imagens utilizadas de grãos de aveia no treinamento das redes neurais. Cada conjunto de imagens usa uma taxa de divisão de 75/25 para os subconjuntos de imagens de treinamento e de avaliação respectivamente. As imagens são selecionadas aleatoriamente para cada subconjunto. As de treinamento são usadas para obter os parâmetros ideais para o modelo enquanto que as de avaliação são usadas para aferir o desempenho final do classificador.

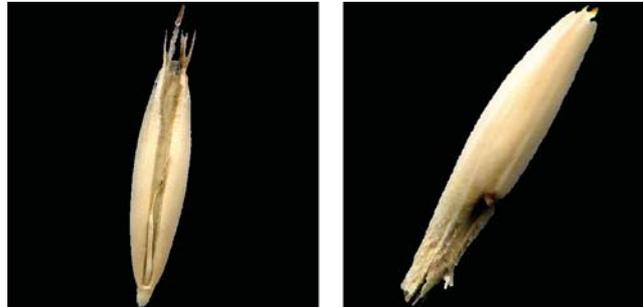
Tabela 8. Categorias e quantidades de imagens de grãos de aveia utilizadas para identificação de espécies.

Classe	Quantidade
<i>Avena strigosa</i>	5000
<i>Avena sativa</i>	5000
Total	10000

Para cada conjunto de imagens foram treinadas seis redes neurais convolucionais: Le-Net5, AlexNet, VGG 16, Inception V3, Resnet e DenseNet 121. Cada modelo foi treinado usando o otimizador Batch Stochastic Gradient Descent com Nesterov Momentum de valor 0.9. Ao treinar redes neurais, muitas vezes é útil reduzir a taxa de aprendizado para melhorar sua acurácia. Para isso, foi adotada uma taxa variável de aprendizado com valor inicial de 0,1. A cada cinco épocas sem melhora na acurácia, a taxa de aprendizado é dividida pela metade até o valor

Tabela 9. Categorias e quantidades de imagens de grãos de aveia utilizadas para classificação de cultivares.

Classe	Quantidade
UPFA Ouro	2500
UPFA Fuerza	2500
UPFA Gaudéria	2500
Total	7500



(a) *Avena strigosa* - UPFA Moreninha.



(b) *Avena sativa* - UPFA Fuerza.



(c) *Avena sativa* - UPFA Gaudéria.



(d) *Avena sativa* - UPFA Ouro.

Figura 26. Amostras de imagens de grãos de aveia após remoção do fundo e segmentação.

mínimo de 0,0001. Todos os modelos foram treinados por 100 épocas. Cada época consiste quando todo o dataset de entrada é processado uma única vez e os pesos são retroalimentados para todos os nodos da rede neural.

Os parâmetros de entrada e a sequência de passos adotada pelo script de treinamento é detalhada a seguir:

- **Parâmetros de entrada:** Conjunto de imagens de treinamento, nome do modelo da rede neural convolucional, número de épocas de treinamento, diretório de saída para a rede treinada.
- **Passo 1:** Instanciar a modelo selecionado. Normalmente é permitido pelos modelos de redes neurais a customização da camada SoftMax com o número de classes que será utilizado. Um exemplo de definição do modelo LeNet5 utilizando Keras é apresentado em 27.
- **Passo 2:** Compilação do modelo. Basicamente configura a rede e a prepara para ser treinada. Nesse passo é configurado o otimizador que será utilizado, a função de perda e qual métrica de aprendizado da rede será avaliada a cada época de treinamento.
- **Passo 3:** Inicialização dos `ImageDataGenerator`. Essa classe utilitária do Keras permite que as imagens sejam carregadas apenas no momento que forem necessárias. Dessa maneira, o treinamento não necessita que todo o conjunto de imagens esteja disponível em memória. Nesse momento os valores dos pixels são convertidos do intervalo de [0..255] para [0..1] com a finalidade de acelerar a convergência do modelo.
- **Passo 4:** Inicialização dos *callbacks*. A Keras disponibiliza uma série de *callbacks* (funções que são chamadas ao final de cada época) para customizar o processo de treinamento. Basicamente, são utilizados quatro: `ModelCheckpoint`, `EarlyStopping`, `CSVLogger` e `ReduceLRonPlateau`. O *callback* `ModelCheckpoint` monitora o processo de treinamento e apenas armazena o modelo treinado quando há algum incremento na métrica selecionada para avaliar o treinamento. Desse modo, é possível economizar espaço em disco e apenas armazenar os melhores resultados. O *callback* `EarlyStopping` permite que o treinamento seja interrompido caso um número x de épocas tenham se passado sem nenhum incremento na métrica de avaliação. O *callback* `CSVLogger` possibilita que os indicadores do treinamento sejam armazenados em um arquivo CSV (planilha). Finalmente, o *callback* `ReduceLRonPlateau` reduz a taxa de aprendizado após um número x de épocas sem incremento na métrica de avaliação.
- **Passo 5:** Execução do treinamento da rede neural. Nesse passo o treinamento é propriamente realizado. O algoritmo utilizado para treinar a rede neural, SGD, é um algoritmo de otimização que visa minimizar a função de perda parametrizada pelas variáveis do modelo atualizando seus valores na direção oposta do gradiente da função objeto. A taxa

de aprendizagem determina o tamanho dessa atualização a cada iteração para atingir um mínimo da função. Esse mínimo não necessariamente é o valor mínimo da função. Em alguns casos, pode ser um mínimo local. Devido a essa característica, é interessante o ajuste dinâmico da taxa de aprendizagem. A cada época, o conjunto de imagens é processado em *batches*, a função de perda é atualizada e os valores dos pesos dos nodos são retroalimentados para todos nodos que compõem a rede neural.

- **Passo 6:** Armazenar a rede treinada ao final do processamento.

```

1. def lenet5_model(num_classes):
2.     # Inicializar o modelo
3.     model = Sequential()
4.
5.     # Primeiro conjunto CONV / POOL
6.     model.add(Conv2D(20, (5, 5), input_shape=(224, 224, 3),
6.         activation='relu'))
7.     model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))
8.
9.     # Segundo conjunto CONV / POOL
10.    model.add(Conv2D(50, (5, 5), activation='relu'))
11.    model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))
12.
13.    # Camada densamente conectada FC
14.    model.add(Flatten())
15.    model.add(Dense(500, activation='relu'))
16.
17.    # Camada SoftMax
18.    model.add(Dense(num_classes, activation='softmax'))
19.    return model

```

Figura 27. Código-fonte em Python para especificação da rede neural convolucional LeNet5 utilizando Keras.

4.5 CLASSIFICAÇÃO

Compreende a operação de utilizar uma rede neural convolucional pré-treinada para classificar um imagem de um grão resultante do processo de segmentação. A biblioteca Keras possibilita que arquitetura, pesos e o estado do otimizador possam ser salvos em um arquivo HDF5. O formato HDF5 (*Hierarchical Data Format Version 5*) consiste em um formato de arquivo para armazenar e gerenciar grandes e complexas coleções de dados. Essa tecnologia permite que dados heterogêneos possam ser representados utilizando uma variedade de tipos de dados. Seu formato favorece a performance em operações de leitura e de escrita além de ser independente de plataforma, arquitetura ou linguagem de programação. Dessa forma, é

possível facilmente distribuir uma rede pré-treinada para ser utilizada pelo usuário no aplicativo `seedflow.ui`.

Basicamente, são dois tipos de redes pré-treinadas: as que identificam espécies de aveia e as que classificam cultivares de aveia. O sistema não depende da arquitetura de rede utilizada. Logo, a definição da arquitetura pode ser modificada sem que a aplicação do usuário necessite de modificações.

4.6 APRESENTAÇÃO DE RESULTADOS

O módulo de apresentação de resultados compreende uma interface de usuário desenvolvida para exibir as medidas morfológicas do grão, o resultado de classificação da rede bem como manipular imagens provenientes da etapa de aquisição de imagens para possibilitar a classificação dos grãos (Figura 28).



Figura 28. Apresentação da tela principal do usuário do aplicação `seedflow.ui`

Além da interface de usuário, duas saídas principais são geradas pela aplicação. A primeira é um arquivo CSV contendo todas os parâmetros dos grãos identificados em uma imagem. Esses parâmetros são formados pelas medidas morfológicas como área, perímetro, eixo maior e eixo menor e a classe proveniente da classificação do grão. Também é informado qual classificador foi selecionado pelo usuário. A Tabela 10) apresenta uma visão parcial dos dados obtidos em uma imagem.

A segunda saída gerada pela aplicação é uma imagem anotada com todos os pontos utilizados para obter as medidas morfológicas e as caixas delimitadoras dos grãos. A ideia é

Tabela 10. Visão parcial da tabela gerada pelo aplicativo seedflow.ui contendo parâmetros dos grãos identificados em uma imagem.

id	area (image) mm2	perimeter mm	x axis (minor) mm	y axis (major) mm	class	classifier
1	624.385	19.727	2.223	8.594	moreninha	species
3	923.692	22.598	2.815	9.885	moreninha	species
6	753.830	20.415	2.572	8.737	moreninha	species
11	593.196	20.904	2.280	9.139	moreninha	species
15	789.908	23.640	2.159	10.456	moreninha	species

permitir ao usuário a conferência dos resultados obtidos pela aplicação e possibilitar a realização posterior de análises estatísticas dos parâmetros das amostras (Figura 29).

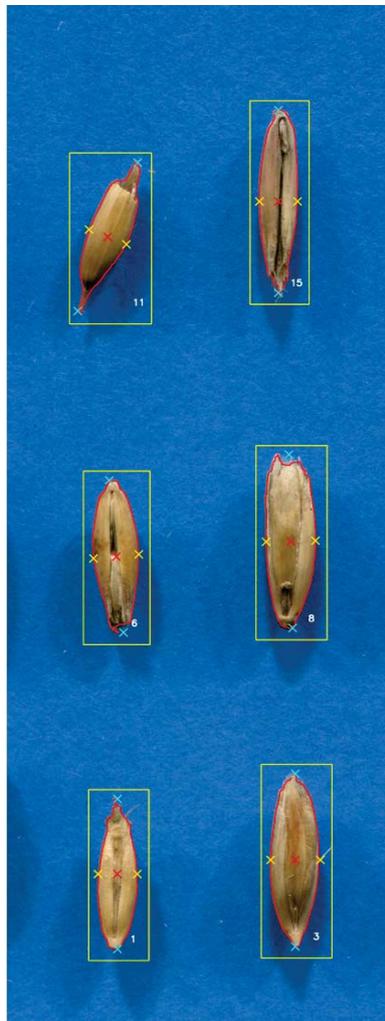


Figura 29. Imagem anotada contendo os pontos utilizados para obter as medidas morfológicas dos grãos identificados em uma imagem.

5. RESULTADOS E DISCUSSÃO

Um dos principais objetivos deste trabalho é mostrar que redes neurais convolucionais podem ser utilizadas na classificação de grãos de aveia. Dentro desse contexto, é importante também avaliar qual modelo possui a melhor acurácia para o conjunto de dados utilizado. De acordo com Barbedo [61], vários fatores podem afetar o desempenho dos modelos de classificação. No presente trabalho, alguns desses aspectos foram tratados com o intuito de reduzir seus impactos nos resultados finais.

Segundo o que foi apresentado no Capítulo 4, para a classificação são utilizadas imagens digitalizadas dos grãos de aveia. Ao utilizar uma rede neural convolucional faz-se necessário a disponibilidade de um banco de imagens corretamente anotado e grande o suficiente para permitir o treinamento da rede neural. Nesse sentido, foi dada uma atenção especial na construção de um banco de imagens de grãos de aveia. Isso só foi possível a partir da garantia de origem das amostras. As amostras de grãos utilizadas foram disponibilizadas pelo Laboratório de Sementes da UPF (LAS) contendo a identificação da origem e da cultivar. Dessa forma, cada cultivar foi escaneada em separado garantindo a correta atribuição da categoria correspondente. Esse processo é lento e custoso, não pelo custo dos equipamentos necessários, mas devido ao número elevado de horas de trabalho demandado. Grãos são difíceis de determinar sua origem caso sejam misturados. Os testes realizados mostraram que é uma tarefa difícil para uma pessoa determinar qual espécie corresponde a cada grão e o método manual normalmente aplicado é baseado na semelhança dos grãos entre si. Sendo assim, raramente um grão é analisado de forma isolada.

Outro aspecto importante considerado esteve relacionado as condições de captura de imagem. A iluminação é um aspecto relevante que deve ser observado para garantir a homogeneidade das imagens capturadas. Para esse fim, o scanner foi escolhido por fornecer as mesmas condições de iluminação e de plano de fundo para as imagens adquiridas. Dessa forma, o plano de fundo pode ser tratado facilmente visto que não contém elementos que interferem nesse processo. A não remoção do plano de fundo da imagem poderia reduzir a eficácia do processo de treinamento e classificação.

As redes neurais convolucionais são capazes de identificar estruturas em imagens e classificá-las de acordo com as classes pré-estabelecidas durante o processo de treinamento. A utilização desse tipo de redes possibilitou a obtenção de valores elevados para a acurácia na classificação dos grãos utilizando imagens digitais. Especificamente, a DenseNet obteve os maiores valores de acurácia: 99,7% para identificação da espécies de aveia e 89,7% para classificação da cultivar de aveia. A DenseNet mostrou ser também a mais eficiente do ponto de vista de uso de memória e a que necessitou do menor número de épocas de treinamento para convergir o modelo. A Tabela 11 e a Tabela 12 apresentam os valores de acurácia, tempo de treinamento para 100 épocas e tempo de classificação para 1600 grãos utilizando as redes neurais

convolucionais analisadas. Por outro lado, a rede AlexNet obteve excelentes resultados muito próximos aos obtidos pela DenseNet. A AlexNet (2012), por ser uma arquitetura mais antiga possui uma estrutura mais simples quando comparada a ResNet (2015) e a DenseNet (2016). As duas últimas, redes mais recentes, aplicam conceitos de redes neurais residuais para melhorar a precisão da classificação. As redes neurais residuais utilizam atalhos para pular algumas camadas. Uma motivação para essa abordagem é evitar o problema do desaparecimento de gradientes reutilizando ativações de uma camada anterior até que a camada próxima da atual aprenda. Ao resultado semelhante entre redes neurais convolucionais residuais e não-residuais atribuímos o fato de que as imagens de grãos possuem uma forma simples e com uma estrutura muito semelhante entre si, ou seja, forma oval e um pequeno espectro de cores. Portanto, uma rede mais simples, como o AlexNet, já seria capaz de abstrair e identificar as estruturas presentes nas imagens dos grãos. Ainda nessa tabela é interessante observar os tempos de treinamento e de classificação de 1600 grãos. Para o caso de identificação de espécies, o tempo de treinamento da rede AlexNet corresponde a aproximadamente 70% do tempo de treinamento da rede DenseNet. O tempo de classificação de 1600 grãos sem o uso de GPU utilizado pela AlexNet corresponde a 60% do tempo utilizado pela DenseNet. Desse modo, na ausência de hardware específico, o uso de uma rede mais simples como a AlexNet poderia ser justificado visto que a diferença de acurácia para o caso de identificação de espécies é muito pequena. Porém, com o aumento do número de classes, como no caso de identificação de cultivares, as diferenças de acurácia entre essas duas redes tendem a aumentar. Isso pode ser percebido na Tabela 12 onde a acurácia obtida pela rede AlexNet foi de 85,9% comparada a rede DenseNet 89,7%, uma diferença de aproximadamente 5%. O efeito do uso de GPU para o treinamento e classificação também foi mensurado. Segundo as Tabelas 11 e 12, esse tipo de hardware especializado reduz em média 25 vezes o tempo necessário para classificação de 1600 grãos. Já o tempo de treinamento não foi mensurado visto que seria impraticável o treinamento das redes neurais convolucionais sem o uso desse tipo de hardware.

Modelo	Treinamento				Classificação	
	Acurácia	Perda	Tempo	Época	Tempo - S/GPU	Tempo - C/GPU
LeNet5	0,978	0,143	3368s	72	41s	6s
AlexNet	0,992	0,019	9198s	79	269s	9s
VGG16	0,992	0,045	13598s	73	386s	14s
InceptionV3	0,989	0,050	8832s	61	148s	9s
ResNet	0,996	0,030	12207s	74	355s	10s
DenseNet121	0,997	0,012	12892s	22	440s	10s

Nota: Tempo de treinamento para 100 épocas. Tempo de classificação para 1600 grãos.

Tabela 11. Comparação dos resultados para identificação de espécies de aveia utilizando redes neurais convolucionais.

A Figura 30 e a Figura 31 apresentam os gráficos referentes a acurácia e a função de perda ao longo das épocas de treinamento para os dois casos analisados, o de identificação de espécies e o de classificação de cultivares de aveia. Ambos os gráficos permitem analisar a

Modelo	Treinamento				Classificação	
	Acurácia	Perda	Tempo	Época	Tempo - S/GPU	Tempo - C/GPU
LeNet5	0,657	2,244	2553s	97	39s	5s
AlexNet	0,859	0,439	6954s	75	264s	16s
VGG16	0,715	2,733	13489s	43	369s	9s
InceptionV3	0,852	0,749	6649s	45	160s	8s
ResNet	0,838	0,744	8921s	58	353s	9s
DenseNet121	0,897	0,414	9811s	53	465s	11s

Nota: Tempo de treinamento para 100 épocas. Tempo de classificação para 1600 grãos.

Tabela 12. Comparação dos resultados para classificação de cultivares de aveia utilizando redes neurais convolucionais.

quantidade de épocas necessárias para o treinamento e a qualidade da convergência do modelo. Note que no caso da identificação de espécies apenas duas classes foram consideradas: *Avena sativa* e *Avena strigosa*. Apesar do melhor resultado ter sido obtido na época 75 para AlexNet conforme a Tabela 11, por volta da época 25 o modelo já teria obtido uma convergência razoável. No caso de classificação de cultivares, onde três classes foram utilizadas: UPFA Ouro, UPFA Fuerza, UPFA Gaudéria, o resultado foi diferente. O modelo AlexNet atingiu uma convergência razoável aproximadamente na época 40. Foram necessárias mais épocas para que o modelo apresentasse uma convergência razoável. Também foram apresentados, para ambos os casos estudados, os gráficos da função de perda. Esse gráfico é útil para análise da ocorrência de *overfitting*. Quando essa situação ocorre, o valor de perda tende a aumentar ao longo das épocas fazendo com que o gráfico tenha a forma de “U”. Ambos os gráficos não apresentam tal situação.

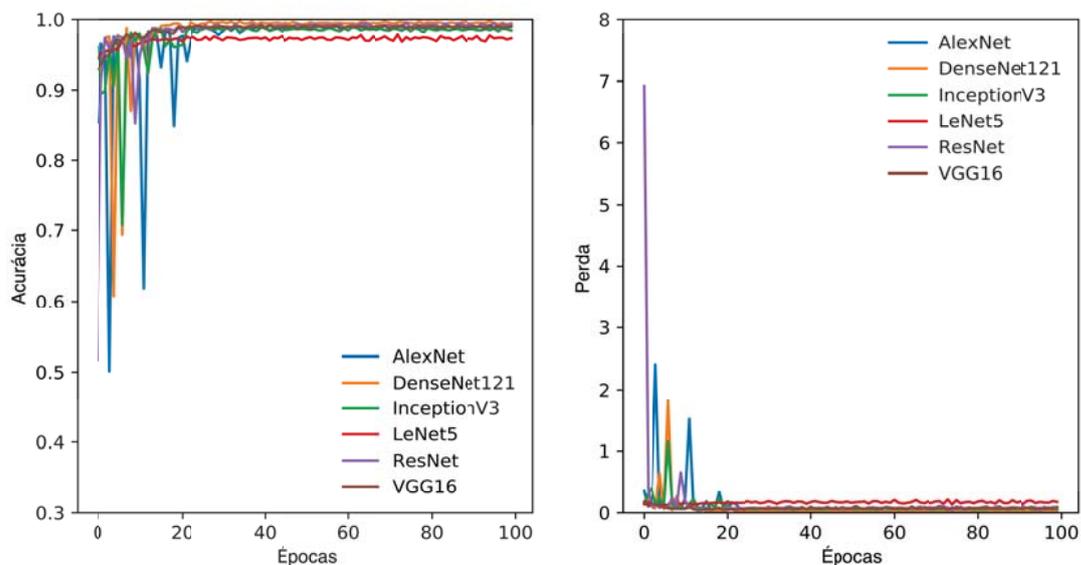


Figura 30. Gráficos da acurácia (esq.) e da função de perda (dir.) obtidos no treinamento de redes neurais convolucionais para identificação de espécies de aveia: *Avena sativa* × *Avena strigosa*.

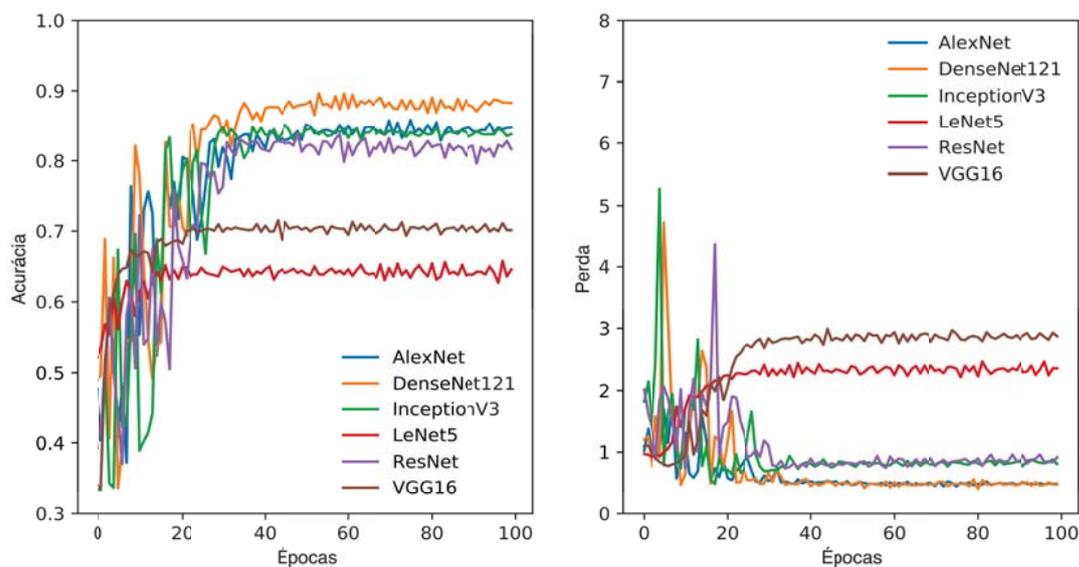


Figura 31. Gráficos da acurácia (esq.) e da função de perda (dir.) obtidos no treinamento de redes neurais convolucionais para classificação de cultivares de *Avena sativa*: UPFA Ouro × UPFA Fuerza × UPFA Gaudéria.

5.1 ACURÁCIA HUMANA NA IDENTIFICAÇÃO DE ESPÉCIES DE AVEIA

Um pequeno experimento foi realizado com seis participantes que possuem experiência prévia na identificação de espécies de aveia. Esse experimento foi realizado com a finalidade de comparar a acurácia humana e a obtida pelo uso de uma rede neural convolucional. Duas espécies foram utilizadas: *Avena sativa* (UPFA Ouro) e *Avena strigosa* (UPFA Moreninha). O kit foi composto por 20 pranchas de 100 grãos cada, totalizando 2000 grãos. Cada prancha (Figura 32) possui uma quantidade aleatória de cada espécie e, portanto, não necessariamente na proporção 50/50. Uma vez iniciado o experimento, os participantes foram instruídos a avaliar todas as pranchas até o final. Essa solicitação foi feita para permitir a avaliação da interferência da fadiga com a acurácia individual de cada participante. As pranchas foram marcadas com as coordenadas das células para que o usuário possa indicar a que espécie cada célula corresponde.

A acurácia humana média encontrada no experimento foi de 93,77% (5,72% do desvio padrão). Cada pessoa necessitou cerca de uma a duas horas para avaliar as 20 pranchas. Em comparação com o computador equipado com GPU, a mesma tarefa necessita de cerca de 10 segundos com uma acurácia média de 99,7%, desconsiderando o tempo gasto para capturar as imagens. Não foram identificadas razões para confirmar que a fadiga possa interferir na acurácia. Entretanto, a maioria dos participantes relatou em seus comentários de que a tarefa foi exaustiva e, portanto, esse resultado não é conclusivo. Outro ponto que merece destaque é que mesmo os participantes que foram treinados da mesma forma e que realizam essa tarefa rotineiramente apresentaram uma alta variabilidade na acurácia individual. A Figura 33 apresenta o gráfico da

acurácia obtido a partir das respostas dos seis participantes (R1 a R6) e a acurácia obtida pela máquina.



Figura 32. Prancha para aferição da acurácia humana na identificação de espécies de aveia.

O experimento sugere que a solução proposta é melhor que a classificação manual no sentido que demanda menos tempo e que apresenta menor variabilidade no resultado. Entretanto, é importante salientar que algumas desvantagens apresentadas na Seção 5.2 devem ser consideradas na hipótese da substituição da classificação manual pela automática.

5.2 ANÁLISE DA SOLUÇÃO SEEDFLOW

A solução proposta mostrou-se eficaz no auxílio da identificação de espécies e na classificação de cultivares de aveia. De modo geral, a aplicação é de uso intuitivo e de fácil acesso mesmo para usuários com pouco treinamento. Essa característica permite que ele possa ser utilizado mesmo por estagiários e profissionais recém-contratados. Sendo assim, é possível reduzir a necessidade da presença de um especialista durante a rotina do laboratório.

Alguns pontos mostraram-se inconvenientes e, portanto, demandariam aperfeiçoamentos. O primeiro ponto é o tempo necessário para a aquisição das imagens. O processo ainda é oneroso e poderia ser agilizado se não houvesse a necessidade de dispersão das sementes sobre o vidro do scanner. A eletricidade estática produzida pelo movimento das mãos pode fazer com que os grãos se movimentem sobre o vidro acarretando na necessidade de reposicionamento dos grãos. Uma das possíveis soluções seria a utilização de uma câmera fotográfica

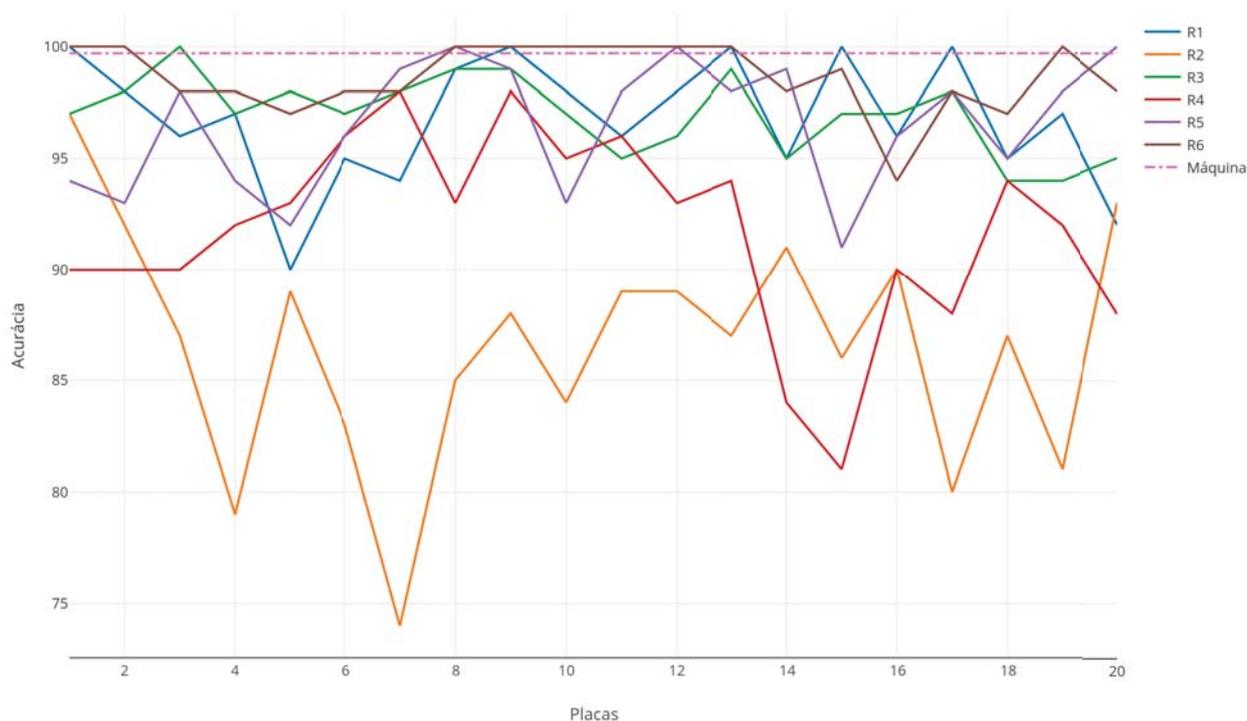


Figura 33. Comparação da acurácia para identificação de espécies de aveia entre humanos (R1 a R6) e o método proposto utilizando redes neurais convolucionais (Máquina).

equipada com flash circular posicionada perpendicularmente ao plano de dispersão dos grãos. Outro ponto que poderia ser reavaliado é a necessidade de uma estação de trabalho. Talvez com a utilização de computadores de placa única como o Raspberry Pi o sistema possa ser transportado com mais facilidade e, desse modo, utilizado nos diversos locais onde seja necessário.

Porém, apesar dos pontos destacados anteriormente, a solução cumpre bem com seu papel a um baixo custo de implantação. Abaixo, na Tabela 13, é listado o custo dos equipamentos necessários para a implantação do sistema:

#	Item	Valor (R\$)
1	Computador Desktop Core i7 8GB RAM DDR4 GeForce 1070	5.000,00
2	Monitor IPS FullHD 24"	700,00
3	Scanner Epson V370	1.000,00
	Total	6.700,00

Tabela 13. Investimento necessário para implantação do sistema SeedFlow.

Além disso, outros fatores positivos são apresentados pela solução SeedFlow. Como toda a solução é baseada em hardware comercializado por diversos fabricantes, essa não seria uma limitação para a disponibilização da solução em larga escala. As tecnologias adotadas

para o desenvolvimento da solução possibilitam a sua distribuição em um pacote instalável em qualquer máquina que possua a plataforma Python disponível. Além disso, a utilização de um scanner como equipamento de captura demonstrou-se adequada para o uso em ambiente laboratorial.

O ganho em termos de eficiência no processo de avaliação de grãos também é positivo. O desequilíbrio entre a capacidade de classificação de grão identificado entre diferentes indivíduos pode ser mitigado por meio do uso de sistemas de visão computacional. Isso garante maior confiança no processo como um todo e permite que o laboratório disponibilize um serviço de melhor qualidade.

6. CONCLUSÃO

Os sistemas de visão computacional tem sido utilizados amplamente na agricultura com o intuito de facilitar a tomada de decisão por parte do produtor. Sistemas automatizados de análise de material vegetal já são uma realidade e sua contribuição positiva é evidenciada pelos diversos relatos presentes na literatura. Ao longo da última década, algoritmos de aprendizado de máquina vivenciaram um grande desenvolvimento devido avanços em sua concepção e no hardware necessário para sua computação. Desde então, seu emprego na agricultura tem crescido ano após ano.

O presente trabalho apresentou uma metodologia para identificação de espécies e classificação de cultivares de aveia atingindo valores de acurácia superiores ao do ser humano quando comparado em situações similares de aplicação. Neste estudo, seis modelos CNN foram comparados com a finalidade de determinar o qual seria o melhor classificador de grãos de aveia em ambiente controlado. A acurácia obtida dos métodos avaliados mostrou que é viável utilizar inteligência artificial e visão computacional para classificar grãos de aveia. Para o caso das espécies de aveia, *Avena sativa* e *Avena strigosa*, a precisão do método proposto atingiu acurácia de 99,7% usando o modelo de rede neural convolucional DenseNet. Para o caso das cultivares de aveia, a acurácia, utilizando essa mesma abordagem foi de 89,7%.

O experimento que comparou a acurácia da metodologia proposta com a acurácia humana identificou que o resultado obtido pelo método é superior a de um humano no caso de identificação de espécies de aveia. A aplicação produzida para a utilização do método mostrou-se eficaz em facilitar a utilização do método proposto. Entendemos, desse modo, que a utilização de visão computacional e inteligência artificial pode trazer diversos benefícios na área de análise de grãos.

Rede neurais convolucionais correspondem a uma vasta área do conhecimento inserida na disciplina de inteligência artificial. O grande interesse por diversos grupos de pesquisa nos últimos anos tem gerado grandes avanços. Os recursos computacionais e as bases de dados disponíveis hoje facilitam em grande parte o desenvolvimento de novas aplicações utilizando esta tecnologia. Seu efeito disruptivo na agricultura é evidente e deve ser aproveitado.

Além de uma alternativa viável para o problema de pesquisa estudado, esse trabalho possibilitou a produção de artigos científicos submetidos a revistas de referência na área de visão computacional aplicada à agricultura. O primeiro, já publicado [48], apresenta a revisão sistemática realizada para identificar as técnicas de visão computacional e inteligência artificial mais utilizadas nas culturas de grãos. O segundo, ainda em avaliação, descreve a metodologia proposta e a sua aplicação em ambiente de laboratório. Além da produção científica, está em processo de tramitação o registro do software produzido no contexto desse trabalho.

Considerando a pergunta de pesquisa: “É possível desenvolver um sistema de visão computacional capaz de automaticamente identificar espécies e de classificar cultivares de aveia

por meio de imagens digitais de amostras de grãos?” entendemos que a resposta é sim. Porém, apesar do resultado satisfatório entendemos que diversos trabalhos futuros poderiam ser elaborados a partir da técnica proposta:

- **Aquisição de imagens:** Diferentes formas de aquisição de imagem poderiam ser avaliadas com o intuito de tornar mais eficiente o processo de aquisição dos grãos. Uma abordagem seria a utilização de câmeras fotográficas e outra seria a utilização de câmeras para análise em tempo real dos grãos. Além das câmeras, outro dispositivo que poderia ser utilizado para aquisição das imagens é o *smartphone*. Esse tipo de dispositivo já conta com câmeras de alta qualidade e, portanto, é possível que esse seja utilizado de maneira satisfatória para a aquisição das imagens dos grãos.
- **Hardware para o treinamento da rede neural:** O treinamento da rede neural poderia ser realizado em uma estação de trabalho diferente da que seria utilizada para análise de grãos. Dessa forma, cluster de computadores ou de GPUs poderiam ser explorados para aumentar o número de classes e de imagens utilizadas no treinamento da rede sem que o tempo necessário para o processamento inviabilize o processo.
- **Hardware para a utilização da rede neural:** Uma vez treinada a rede, o uso de hardware avançado não é mais necessário. Sendo assim, computadores de placa única poderiam ser explorados para tornar a solução mais compacta e possível de ser utilizada nas mais diversas situações.
- **Modelos de redes neurais convolucionais:** Além dos seis modelos explorados nesse trabalho, outros modelos apresentados na literatura poderiam ser avaliados. Embora tenhamos selecionado os mais representativos, um modelo diferente poderia trazer algum resultado relevante quando comparado com os que foram utilizados. Outra abordagem seria o desenvolvimento de um modelo específico para grãos considerando as suas características específicas.
- **Diferentes espécies e culturas:** A mesma metodologia poderia ser utilizada para as demais cultivares de aveia ou para as demais culturas de grãos.

Finalmente, cabe ressaltar a importância da Agência de Inovação Tecnológica da UPF - UPFTec. Seu objetivo principal é promover a inovação através da transferência de tecnologia aproximando a Universidade de Passo Fundo de empresas e demais organizações interessadas desenvolver inovações em conjunto com a universidade. Sua atuação foi essencial na aproximação e identificação de demandas entre o LAS (laboratório de Análise de Sementes) e o PPGCA (Programa de Pós-Graduação em Computação Aplicada) da UPF.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] FOOD; NATIONS, A. O. of the U. *FAOSTAT Database*. 2018. Disponível em: <<http://www.fao.org/faostat/>>.
- [2] RAMASWAMY, H.; RIAHI, E. Structure and composition of cereal grains and legumes. In: _____. *Handbook of Postharvest Technology*. CRC Press, 2003. p. 1–16. ISBN 978-0-8247-0514-5. Disponível em: <<https://doi.org/10.1201/9780203911310.pt1>>.
- [3] HULASARE, R.; DRONZEK, B.; JAYAS, D. Grain-grading systems. In: _____. *Handbook of Postharvest Technology*. CRC Press, 2003. p. 41–55. ISBN 978-0-8247-0514-5. Disponível em: <<https://doi.org/10.1201/9780203911310.ch3>>.
- [4] FUNDO, U. de P. *Cultivares de Aveia*. 2018. Disponível em: <<http://aveia.upf.br/>>. Acesso em: 2018-12-08.
- [5] LEI nº 9.456 de 25 de abril de 1997. 1997.
- [6] MAHAJAN, S.; DAS, A.; SARDANA, H. K. Image acquisition techniques for assessment of legume quality. *Trends in Food Science & Technology*, v. 42, n. 2, p. 116–133, 2015. ISSN 0924-2244. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0924224415000023>>.
- [7] ZAREIFOROUGH, H. et al. Potential Applications of Computer Vision in Quality Inspection of Rice: A Review. *Food Engineering Reviews*, v. 7, n. 3, p. 321–345, set. 2015. ISSN 1866-7910, 1866-7929. Disponível em: <<https://link.springer.com/article/10.1007/s12393-014-9101-z>>.
- [8] VISEN, N. et al. Ae-automation and emerging technologies: Specialist neural networks for cereal grain classification. *Biosystems Engineering*, v. 82, n. 2, p. 151–159, 2002. ISSN 1537-5110. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1537511002900644>>.
- [9] VITHU, P.; MOSES, J. A. Machine vision system for food grain quality evaluation: A review. *Trends in Food Science & Technology*, v. 56, p. 13–20, out. 2016. ISSN 0924-2244. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S092422441630084X>>.
- [10] KIRK, D. B.; HWU, W. mei W. Programming massively parallel processors (third edition). In: KIRK, D. B.; HWU, W.-m. W. (Ed.). Third edition. [S.l.]: Morgan Kaufmann, 2017. ISBN 978-0-12-811986-0.
- [11] TBIT. *TBit*. 2018. Disponível em: <<http://www.tbit.com.br>>.

- [12] CONAB. *Acompanamento da Safra Brasileira de Grãos - 2016/2017*. 2017. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>>. Acesso em: 2018-12-08.
- [13] PORTARIA Ministerial nº 191 de 14 de Abril de 1975. 1975.
- [14] CLARKE, B. Technology of cereals. *Experimental Agriculture*, Cambridge University Press, v. 25, n. 2, p. 273–273, 1989.
- [15] SHUKLA, T. P.; WELLS, G. H. Chemistry of oats: Protein foods and other industrial products. *CRC Critical Reviews in Food Science and Nutrition*, Taylor & Francis, v. 6, n. 4, p. 383–431, 1975. Disponível em: <<https://doi.org/10.1080/10408397509527196>>.
- [16] DECKER, E. A.; ROSE, D. J.; STEWART, D. Processing of oats and the impact of processing operations on nutrition and health benefits. *British Journal of Nutrition*, Cambridge University Press, v. 112, n. S2, p. S58–S64, 2014.
- [17] ACT 1991 - International Convention For The Protection of New Varieties of Plants. 1991. Disponível em: <<http://www.upov.int/upovlex/en/conventions/1991/act1991.html>>.
- [18] DAVIES, E. (Ed.). *Machine Vision*. Third edition. Burlington: Morgan Kaufmann, 2005. (Signal Processing and its Applications). ISBN 978-0-12-206093-9. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780122060939500393>>.
- [19] HORNBERG, A. *Handbook of Machine and Computer Vision*. Wiley-VCH Verlag GmbH & Co. KGaA, 2017. ISBN 9783527413409. Disponível em: <<http://dx.doi.org/10.1002/9783527413409.fmatter>>.
- [20] ZHANG, Y. *Image Processing (De Gruyter Textbook)*. [S.l.]: De Gruyter, 2017.
- [21] PRINCE, S. J. D. *Computer Vision: Models, Learning, and Inference*. 1st. ed. New York, NY, USA: Cambridge University Press, 2012. ISBN 1107011795, 9781107011793.
- [22] KAEHLER, A.; BRADSKI, G. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*. 1st. ed. [S.l.]: O'Reilly Media, Inc., 2016. ISBN 1491937998, 9781491937990.
- [23] GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.
- [24] MARSLAND, S. *Machine Learning: An Algorithmic Perspective, Second Edition*. 2nd. ed. [S.l.]: Chapman & Hall/CRC, 2014. ISBN 1466583282, 9781466583283.
- [25] RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach (3rd Edition)*. 3. ed. [S.l.]: Pearson, 2009. Hardcover. ISBN 0136042597.

- [26] XIE, F. et al. Sequence error (SE) minimization training of neural network for voice conversion. In: *INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, 2014*. [s.n.], 2014. p. 2283–2287. Disponível em: <http://www.isca-speech.org/archive/interspeech_2014/i14_2283.html>.
- [27] SAK, H.; SENIOR, A. W.; BEAUFAYS, F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *CoRR*, abs/1402.1128, 2014. Disponível em: <<http://arxiv.org/abs/1402.1128>>.
- [28] SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. Cambridge, MA, USA: MIT Press, 2014. (NIPS'14), p. 3104–3112. Disponível em: <<http://dl.acm.org/citation.cfm?id=2969033.2969173>>.
- [29] MARCHI, E. et al. Multi-resolution linear prediction based features for audio onset detection with bidirectional lstm neural networks. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2014. p. 2164–2168. ISSN 1520-6149.
- [30] BREUEL, T. M. et al. High-performance ocr for printed english and fraktur using lstm networks. In: *2013 12th International Conference on Document Analysis and Recognition*. [S.l.: s.n.], 2013. p. 683–687. ISSN 1520-5363.
- [31] ZAREMBA, W.; SUTSKEVER, I.; VINYALS, O. *Recurrent Neural Network Regularization*. 2014. Disponível em: <<https://arxiv.org/abs/1409.2329>>.
- [32] LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, Nov 1998. ISSN 0018-9219.
- [33] KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>.
- [34] ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013. Disponível em: <<http://arxiv.org/abs/1311.2901>>.
- [35] SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. Disponível em: <<http://arxiv.org/abs/1409.1556>>.
- [36] SZEGEDY, C. et al. Going deeper with convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*. [s.n.], 2015. p. 1–9. Disponível em: <<https://doi.org/10.1109/CVPR.2015.7298594>>.

- [37] HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. [s.n.], 2016. p. 770–778. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>.
- [38] GIRSHICK, R. B. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013. Disponível em: <<http://arxiv.org/abs/1311.2524>>.
- [39] GIRSHICK, R. B. Fast R-CNN. In: *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*. [s.n.], 2015. p. 1440–1448. Disponível em: <<https://doi.org/10.1109/ICCV.2015.169>>.
- [40] REN, S. et al. Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. [s.n.], 2015. p. 91–99. Disponível em: <<http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks>>.
- [41] CHENG, X. et al. Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, Elsevier, v. 141, p. 351–356, sep 2017. ISSN 01681699. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0168169917304854#f0030>>.
- [42] HUANG, G.; LIU, Z.; WEINBERGER, K. Q. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016. Disponível em: <<http://arxiv.org/abs/1608.06993>>.
- [43] ALLEN, E.; TRIANTAPHILLIDOU, S. *The Manual of Photography and Digital Imaging*. Taylor & Francis, 2012. ISBN 9781136091094. Disponível em: <<https://books.google.com.br/books?id=Sje8D392ufYC>>.
- [44] BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [45] ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>.
- [46] CHOLLET, F. et al. *Keras*. 2015. <https://keras.io>.
- [47] KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2007.
- [48] PATRÍCIO, D. I.; RIEDER, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, v. 153, p. 69–81, 10 2018.
- [49] SABANCI, K.; KAYABASI, A.; TOKTAS, A. Computer vision-based method for classification of wheat grains using artificial neural network. *Journal of the Science of Food and Agriculture*,

John Wiley & Sons, Ltd, v. 97, n. 8, p. 2588–2593, 2017. ISSN 1097-0010. Disponível em: <<http://dx.doi.org/10.1002/jsfa.8080>>.

- [50] SUN, K. et al. Recognition of Mould Colony on Unhulled Paddy Based on Computer Vision using Conventional Machine-learning and Deep Learning Techniques. *Scientific Reports*, v. 6, p. 37994, nov. 2016. ISSN 2045-2322. Disponível em: <<http://www-nature-com.ez103.periodicos.capes.gov.br/srep/2016/161129/srep37994/full/srep37994.html>>.
- [51] LIU, T. et al. A shadow-based method to calculate the percentage of filled rice grains. *Biosystems Engineering*, v. 150, p. 79–88, out. 2016. ISSN 1537-5110. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1537511016300174>>.
- [52] SHRESTHA, B. L. et al. A two-camera machine vision approach to separating and identifying laboratory sprouted wheat kernels. *Biosystems Engineering*, v. 147, p. 265–273, jul. 2016. ISSN 1537-5110. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1537511015303433>>.
- [53] SINGH, K. R.; CHAUDHURY, S. Efficient technique for rice grain classification using back-propagation neural network and wavelet decomposition. *IET Computer Vision*, v. 10, n. 8, p. 780–787, 2016. ISSN 1751-9632.
- [54] OLGUN, M. et al. Wheat grain classification by using dense sift features with svm classifier. *Computers and Electronics in Agriculture*, v. 122, p. 185 – 190, 2016. ISSN 0168-1699. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168169916300060>>.
- [55] KUO, T.-Y. et al. Identifying rice grains using image analysis and sparse-representation-based classification. *Computers and Electronics in Agriculture*, v. 127, p. 716–725, set. 2016. ISSN 0168-1699. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168169916305166>>.
- [56] KEZHU, T. et al. Identification of diseases for soybean seeds by computer vision applying BP neural network. *International Journal of Agricultural and Biological Engineering*, v. 7, n. 3, p. 43–50, jun. 2014. ISSN 1934-6352. Disponível em: <<https://ijabe.org/index.php/ijabe/article/view/1028>>.
- [57] SUN, C. et al. Evaluation and analysis the chalkiness of connected rice kernels based on image processing technology and support vector machine. *Journal of Cereal Science*, v. 60, n. 2, p. 426–432, 2014. ISSN 0733-5210. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0733521014000861>>.
- [58] EBRAHIMI, E.; MOLLAZADE, K.; BABAEI, S. Toward an automatic wheat purity measuring device: A machine vision-based neural networks-assisted imperialist competitive algorithm approach. *Measurement*, v. 55, p. 196–205, 2014. ISSN 0263-2241. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0263224114002097>>.

- [59] OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 9, n. 1, p. 62–66, Jan 1979. ISSN 0018-9472.
- [60] SUZUKI, S.; BE, K. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, v. 30, n. 1, p. 32 – 46, 1985. ISSN 0734-189X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0734189X85900167>>.
- [61] BARBEDO, J. G. Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, v. 172, p. 84 – 91, 2018. ISSN 1537-5110. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1537511018303027>>.