

UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM
CIÊNCIAS AMBIENTAIS E SUSTENTABILIDADE AGROPECUÁRIA

Gercina Gonçalves da Silva

**Superpixel e Aprendizagem Supervisionada para a Identificação de Doenças da
Soja em Imagens obtidas por Veículos Aéreos Não Tripulados**

CAMPO GRANDE – MS

2017

UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM
CIÊNCIAS AMBIENTAIS E SUSTENTABILIDADE AGROPECUÁRIA

**Superpixel e Aprendizagem Supervisionada para a Identificação de Doenças da
Soja em Imagens obtidas por Veículos Aéreos Não Tripulados**

Autora: Gercina Gonçalves da Silva

Orientador: Hemerson Pistori

Coorientador: Denilson de Oliveira Guilherme

**"Tese apresentada como parte das exigências para
obtenção do título de DOUTOR EM CIÊNCIAS AMBIENTAIS
E SUSTENTABILIDADE AGROPECUÁRIA, no Programa de
Pós-Graduação Stricto Sensu em Ciências Ambientais e
Sustentabilidade Agropecuária da Universidade Católica
Dom Bosco - Área de concentração: Sustentabilidade
Ambiental e Produtiva Aplicada ao Agronegócio e
Produção Sustentável".**

CAMPO GRANDE – MS

2017

Silva, Gercina Gonçalves da. Superpixel e Aprendizagem Supervisionada para a Identificação de Doenças da Soja em Imagens obtidas por Veículos Aéreos Não Tripulados / Gercina Gonçalves da Silva; Hemerson Pistori. – Campo Grande - MS, 2017.

113 f.

Tese (Programa de Pós-Graduação Stricto Sensu em Ciências Ambientais e Sustentabilidade Agropecuária – Universidade Católica Dom Bosco – UCDB).

1: Reconhecimento de doenças da soja, 2: aprendizagem de máquina, 3: Superpixel, 4: VANTs

**UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM
CIÊNCIAS AMBIENTAIS E SUSTENTABILIDADE AGROPECUÁRIA**

FOLHA DE APROVAÇÃO

Gercina Gonçalves da Silva

**Superpixel e Aprendizagem Supervisionada para a Identificação de Doenças da
Soja em Imagens obtidas por Veículos Aéreos Não Tripulados**

Tese submetida ao Programa de Pós-Graduação *Stricto Sensu* em Ciências Ambientais e Sustentabilidade Agropecuária da Universidade Católica Dom Bosco, como parte dos requisitos para obtenção do título de doutor.

(Prof. Dr. Hemerson Pistori, UCDB).

(Prof. Dr. Denilson de Oliveira Guilherme, UCDB).

(Prof. Dr. Marcelo Theophilo Folhes, UCDB).

(Prof. Dr. Naji Rajai Nasri Ama, UCDB).

(Prof. Dr. Wesley Nunes Gonçalves, UFMS).

(Prof. Dr. Oswaldo Barbosa Loureda, Acrux Aerospace Technologies).

Campo Grande, 10 de fevereiro de 2017.

Ensina-nos a contar os nossos dias
para que o nosso coração alcance sabedoria.

(Salmos 90:12)

Sumário

LISTA DE TABELAS.....	8
LISTA DE FIGURAS	9
LISTA DE ABREVIATURAS E SIGLAS.....	10
RESUMO	11
ABSTRACT	12
1 INTRODUÇÃO.....	13
1.1 OBJETIVOS.....	16
1.2 JUSTIFICATIVA	16
1.3 ORGANIZAÇÃO DA TESE	19
2 REVISÃO DA LITERATURA	20
2.1 CULTURA DA SOJA	20
2.2 VISÃO COMPUTACIONAL	28
2.2.1 Segmentação	31
2.2.2 Segmentação por superpixel.....	32
2.2.3 Extração de atributos	35
2.3 APRENDIZAGEM AUTOMÁTICA.....	39
2.3.1 Algoritmos de classificação ou Classificadores	40
2.3.2 Aprendizagem automática e classificação de doenças.....	41
2.4 MÉTRICAS DE AVALIAÇÃO	43
2.5 VEÍCULOS AÉREOS NÃO TRIPULADOS - VANTS	48
3 METODOLOGIA.....	54
 ARTIGO 1 - SEGMENTAÇÃO DE IMAGENS DE VANTS PARA DETECÇÃO DE FERRUGEM DA SOJA	 55
1 Introdução.....	55
2 Material e Métodos	58
2.1 Instalação do Campo Experimental	58
2.2 Aquisição e Processamento das Imagens.....	59
2.2.1 Proposta para reconhecimento de doenças na soja.....	60
2.2.2 Segmentação por Superpixel	61
2.3 Procedimento de validação	63
3 Resultados e Discussões.....	64
4 Conclusão.....	68

Referências Bibliográficas.....	69
ARTIGO 2 - RECONHECIMENTO DE DOENÇAS DE SOJA COM BASE NA SEGMENTAÇÃO DE IMAGENS DE VANTS.....	73
1 Introdução.....	73
2 Material e Métodos	75
2.1 Instalação da lavoura de soja	76
2.2 Delineamento de Vãos e Captura de Imagens	77
2.3 Seleção de imagens para composição do Banco de Imagens de Doenças na Soja....	78
2.4 Segmentação das imagens.....	79
2.5 Extração de atributos	83
2.6 Classificação de imagens	84
2.7 Avaliações agronômicas	85
3 Resultados e Discussão.....	87
3.1 Avaliação agronômica com escalas diagramáticas.....	87
3.2 Avaliação por aprendizagem de máquina.....	89
4 Conclusão.....	95
Referências Bibliográficas.....	96
CONSIDERAÇÕES FINAIS:.....	100
REFERÊNCIAS BIBLIOGRÁFICAS	102

LISTA DE TABELAS

Tabela 1 - Produção de soja dos principais países produtores (Em mil toneladas) .20	20
Tabela 2 - Principais países importadores de soja (dados em mil toneladas).....	211
Tabela 3 - Principais países exportadores de soja (dados em mil toneladas).....	21
Tabela 4 - Produção de Soja na Região Centro-Oeste do Brasil (Em mil toneladas)	22
Tabela 5 - Estádios da soja e suas principais características.	23
Tabela 6 - Principais doenças do estágio vegetativo da soja.....	25
Tabela 7 - Principais doenças do estágio reprodutivo da soja	26
Tabela 8 - Desempenho da classificação para duas classes	44

Artigo 1 - SEGMENTAÇÃO DE IMAGENS DE VANTS PARA DETECÇÃO DE FERRUGEM DA SOJA

Tabela 1 - Avaliação dos classificadores considerando os índices de similaridade. .65	65
Tabela 2 - Classe melhor classificada considerando os coeficientes de Jaccard e Yule.	66
Tabela 3 - Desempenho dos classificadores com o DIAGSOJA	67

ARTIGO 2 - RECONHECIMENTO DE DOENÇAS DE SOJA COM BASE NA SEGMENTAÇÃO DE IMAGENS DE VANTS'

Tabela 1 - Produtos utilizados para o manejo de doenças.....	76
Tabela 2 - Avaliação dos classificadores através de métricas	89
Tabela 3 - Classes similares de acordo com os coeficientes de similaridade	90
Tabela 4 - Teste de Tukey para os classificadores	91

LISTA DE FIGURAS

Figura 1 - Reboleira de Ferrugem Asiática na plantação de soja	27
Figura 2 - Etapas de um sistema de visão computacional	30
Figura 3 - Exemplo de segmentação de imagem da soja demonstrando a imagem original e a imagem segmentada em superpixel.....	32
Figura 4 - Segmentação por superpixel com o SLIC	34
Figura 5 - Exemplo da Área sob a Curva ROC obtida junto ao software Weka.....	47
Figura 6 - Ilustração das etapas utilizadas para o desenvolvimento dos artigos.	54

ARTIGO 1 - SEGMENTAÇÃO DE IMAGENS DE VANTS PARA DETECÇÃO DE FERRUGEM DA SOJA

Figura 1 - Exemplo de imagens originais segmentadas com o extrator de superpixels Slic.	58
Figura 2 - Fluxograma: procedimentos metodológicos.....	59
Figura 3 - VANTs utilizados na captura de imagens	60
Figura 4 - Evolução dos sintomas da Ferrugem asiática no ensaio experimental	65
Figura 5 - Matriz de confusão - Banco de imagens A e B, Classificadores SMO e J48.....	66

ARTIGO 2 - RECONHECIMENTO DE DOENÇAS DE SOJA COM BASE NA SEGMENTAÇÃO DE IMAGENS DE VANTS

Figura 1 - Temperatura e umidade relativa do ar durante a Safra 2015/2016 de soja na Fazenda São José.....	77
Figura 2 - Precipitação durante a Safra 2015/2016 de soja na Fazenda São José	77
Figura 3 - Segmentador de imagens para geração do banco de imagens.	81
Figura 4 - Fragmento de uma imagem segmentada demonstrando visualmente a presença do míldio no folíolo da soja.	82
Figura 5 - Exemplo de imagens de classes que compõem o banco de imagens.....	82
Figura 6 - Escalas diagramáticas usadas na avaliação da severidade de doenças da soja .	86
Figura 7 - Severidade do míldio na cultura da soja	88
Figura 8 - Severidade de mancha-alvo na cultura da soja em função dos dias de plantio sob diferentes formas de controle.	88
Figura 9 - Matriz de confusão gerada no ambiente Weka com o classificador SVM	90
Figura 10 – Diagramas de caixa representando as diferenças entre os desempenhos dos classificadores.....	92
Figura 11 - Classificação visual das doenças da soja nas classes: míldio, macha alvo, ferrugem asiática, folha saudável, solo, e palha realizada pelo programa de computador	93

LISTA DE ABREVIATURAS E SIGLAS

ANAC - Agência Nacional de Aviação Civil

CEPEA - Centro de Estudos Avançados em Economia Aplicada

CONAB – Companhia Nacional de Abastecimento

DECEA - Departamento de Controle do Espaço Aéreo

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária

EUA – Estados Unidos da América

GLCM - Gray-Level Co-occurrence Matrix

HOG – Histogram of Oriented Gradients

INOVISÃO - Grupo de Pesquisa de Desenvolvimento e Inovação em Visão Computacional

IPEA – Instituto de Pesquisa Econômica Aplicada -

LBP - Local Binary Patterns

MAPA - Ministério da Agricultura, Pecuária e Abastecimento

KNN - K-Nearest Neighbor

SLIC - Simple Linear Iterative Clustering

SMO - Sequential Minimal Optimization

SVM - Support Vector Machines

USDA – United States Department of Agriculture

VANT – Veículo Aéreo não Tripulado

RESUMO

O agronegócio brasileiro tem um papel importante na economia do Brasil, mostrando bom desempenho nas exportações de produtos do segmento e construindo novos mercados em diferentes partes do mundo, tornando-se o segundo maior produtor de soja do mundo. A soja tem alguns fatores que podem limitar a produção, entre os quais o papel das doenças é destacado como causando perda de produção. O monitoramento do plantio e o diagnóstico preciso nas fases iniciais da doença podem contribuir para uma redução das perdas. Atualmente, o monitoramento é realizado em amostras distribuídas por toda a plantação, realizadas por técnicos que usam lupas para diagnosticar as patologias ou, simplesmente, a olho nu do especialista. Esta pesquisa utilizou o Veículo Aéreo Não Tripulado para realizar o monitoramento da plantação de soja através da coleta de imagens. O objetivo foi reconhecer as doenças que afetam a plantação experimental de soja através de um programa de computador que utiliza a visão computacional e a aprendizagem supervisionada. O algoritmo *SLIC* (*Simple Linear Iterative Clustering*) foi utilizado na segmentação das imagens e geração de superpixels. Com os segmentos foram gerados em três bancos de imagens. Dois dos bancos de imagens foram gerados com imagens de plantio experimental na safra 2014/2015 e deram origem ao primeiro artigo que compõe esta tese. O terceiro banco de imagens foi construído a partir de imagens coletadas na plantação experimental durante um ano agrícola de 2015/2016 e deu origem ao segundo artigo que compõe esta tese. Destaca-se que os dois primeiros bancos de imagens contam com apenas quatro classes (ferrugem asiática, palha, solo e folha saudável) enquanto o terceiro banco de imagens é composto por seis classes (ferrugem asiática, mancha alvo, míldio, solo, palha e folha saudável). As etapas para desenvolvimento dos artigos foram: coleta de imagens, segmentação, criação do banco de imagens, extração de atributos, utilização das seguintes técnicas SVM, J48, Floresta Aleatória e KNN para classificação das imagens que compõe o banco e testes com métricas para avaliação. O segundo artigo diferencia-se do primeiro pelo delineamento experimental, pela robustez do banco de imagens, pela realização da avaliação de severidade com uso de escala diagramática e, por comparar os resultados da classificação realizada pelo programa proposto a partir do banco de imagens com os resultados da classificação realizada pelo programa a partir das anotações na imagem segmentada nas classes devidas por um especialista. Verificou que o programa foi eficiente para distinguir entre uma classe e outra, onde as técnicas que obtiveram melhor desempenho foi o SVM e Florestas Aleatórias, considerando os resultados obtidos com as métricas de avaliação empregadas na tese, com desempenho superior a 95%.

Palavras-chave: Reconhecimento de doenças da soja, aprendizagem de máquina, Superpixel, Veículos aéreos não tripulados.

ABSTRACT

Brazilian agribusiness plays an important role in the country's economy, showing good performance in segment exports and building new markets in different parts of the world, making it the second largest soybean producer in the world. Soy has some factors that can limit production, among which the role of diseases is highlighted as causing loss of production. Planting monitoring and accurate diagnosis in the early stages of the disease can contribute to a reduction in losses. Currently, monitoring is performed on samples distributed throughout the plantation, performed by technicians who use magnifying glasses to diagnose the pathologies or, simply, to the naked eye of the specialist. This research uses the Unmanned Aerial Vehicle to carry out the monitoring of the soybean plantation through the collection of images. The objective was to recognize the diseases that affect experimental soybean plantation through a computer program that uses computer vision and supervised learning. The SLIC (Simple Linear Iterative Clustering) algorithm was used in the segmentation of images and generation of superpixels. With the segments were generated in three banks of images. Two of the image banks were generated with images of experimental planting in the 2014/2015 crop and gave rise to the first article that composes this thesis. The third image bank was constructed from images collected in the experimental plantation during a crop year of 2015/2016 and gave rise to the second article that composes this thesis. It should be noted that the first two image banks have only four classes (Asian rust, straw, soil and healthy leaf) while the third image bank consists of six classes (Asian rust, target spot, mildew, soil, straw and Healthy leaf). The steps for developing the articles were: image collection, segmentation, image bank creation, extraction of attributes, use of the following techniques SVM, J48, Random Forest and KNN to classify the images that compose the bank and tests with metrics for evaluation. The second article differs from the first one by the experimental design, by the robustness of the image bank, by the performance of the severity evaluation using a diagrammatic scale, and by comparing the results of the classification performed by the proposed program from the image bank with the Results of the classification performed by the program from the annotations in the segmented image in the classes due by a specialist. It verified that the program was efficient to distinguish between one class and another, where the techniques that obtained the best performance were SVM and Random Forests, considering the results obtained with the evaluation metrics used in the thesis, with a performance of over 95%.

Keywords: Recognition of soybean diseases; Machine learning; Superpixel; unmanned aerial vehicles.

1 INTRODUÇÃO

A economia brasileira vive um cenário caracterizado pela redução da atividade econômica, inflação e redução no crescimento econômico. Nessa conjuntura, o agronegócio, que representa 23% do PIB nacional, pode ser o diferencial no desempenho econômico do país (IPEA, 2016; CEPEA, 2014).

A produção agropecuária brasileira tem apresentado importante crescimento produtivo. Especificamente a produção de grãos, por exemplo, em 1975 teve uma colheita de 45 milhões de toneladas, expandindo-se em 1990 para 58 milhões (BUAINAIN et al., 2014), atingindo, na safra 2015/2016, o número de 186,4 milhões de toneladas (CONAB, 2016). A produção de soja no Brasil, na Safra de 2015/2016 foi de 95,4 milhões de toneladas. A região Centro-Oeste do Brasil foi responsável por 45,8% da produção nacional de soja, e, nessa região, o estado de Mato Grosso do Sul produziu 7,2 milhões de toneladas (CONAB, 2016) destacando, dessa forma a importância do produto para esse estado.

Embora a soja seja uma leguminosa amplamente cultivada e difundida com produtividade superior a 4.000 kg.ha^{-1} , a exploração econômica do seu potencial de rendimento dificilmente é alcançada devido, dentre outros fatores, ao manejo inadequado. Os principais fatores que limitam o aumento do rendimento são as doenças e as pragas que, de forma geral, são de difícil controle (YORINORI, 1993).

A expansão da cultura para novas fronteiras agrícolas leva ao aumento de doenças causadas por patógenos a cada safra, variando de ano para ano a importância do controle de cada doença que ataca a soja principalmente em função das variedades de plantas e condições climáticas (YORINORI, 1993). Estima-se que, no Brasil, as perdas anuais de produção por doenças da soja variem entre 15 a 20%, sendo que algumas doenças podem ocasionar 100 % de perda da safra (ZARBIN, 2009; EMBRAPA, 2013).

No Brasil, as primeiras e mais comuns doenças da soja foram o míldio, a mancha púrpura, o crestamento foliar, a mancha parda e a antracnose. Atualmente, as principais doenças da cultura da soja são: necrose da haste da soja, mancha olho-de-rã, oídio, ferrugem asiática da soja, cancro-da-haste, antracnose e nanismo amarelo da soja (ZARBIN, 2009; EMBRAPA, 2011). Para Grigolli (2016), considerando, especificamente, a região Centro Oeste do Brasil, as principais doenças podem ser classificadas em função do sintoma causado por cada patógeno formando os seguintes grupos:

- Doenças foliares: ferrugem asiática da soja, mancha alvo, crestamento foliar de cercospora, míldio, oídio;
- Doenças da haste, vagem e sementes: antracnose e mofo branco; e
- Doenças radiculares: podridão de carvão, murcha de sclerotium, mela da soja.

Os sintomas iniciais das doenças podem ser identificados através de realizações de monitoramentos freqüentes da plantação, devendo ocorrer durante o crescimento vegetativo e principalmente antes do início do florescimento da soja, independente da ocorrência de doenças (YORINORI et al., 2003). Usualmente, o monitoramento da soja é feito por amostragem e se dá de forma sistemática ou ao acaso, em que o tipo e tamanho da amostra variam conforme características das doenças ou objetivos do levantamento (MORAES, 2007).

A observação a olho nu é utilizada pelos agricultores para identificação e detecção de doenças foliares tendo como base experiências acumuladas. Essas técnicas de monitoramento podem ocasionar erros no diagnóstico da severidade das doenças (ROTHER; KSHIRSAGAR, 2015). Além disso, de acordo com Afridi *et al.* (2014), essa técnica de observação possui inconvenientes como:

1. As avaliações são subjetivas, pois dependem de quem avalia, de modo que outro observador pode chegar a conclusões diferentes a partir da mesma amostra;
2. A técnica é trabalhosa, ou seja, exige tempo e dedicação dos especialistas;
3. Alguns sintomas de doenças foliares, bem como suas evoluções podem não ser perceptíveis aos olhos humanos.

Entre os vários instrumentos que podem ser utilizados para monitoramento de culturas, os VANTs - Veículos Aéreos não Tripulados – têm ganhado destaque nos últimos anos, sendo comumente utilizados para aplicações de sensoriamento remoto de baixa altitude e alta resolução em agricultura de precisão. Essa tecnologia permite observações de campo freqüentes para capturar a variação nas plantações ao longo do tempo (HEE SUP *et al.*, 2016).

Adicionalmente, a detecção precoce de doenças e, por conseguinte, as medidas preventivas correspondentes podem minimizar a perda de rendimento significativamente. A busca de objetividade na detecção de doenças foliares beneficia a supervisão de grandes culturas, como é o caso dos grãos, podendo atuar no reconhecimento de doenças ainda em seus sintomas iniciais nas folhas da planta (ROTHER; KSHIRSAGAR, 2015). A utilização de técnicas computacionais de aprendizagem automática e reconhecimento de padrões que podem diagnosticar a presença de doenças em estágios iniciais (GUI *et. al.*, 2015).

Nesse contexto, propõe-se a utilização de um programa de computador para reconhecimento de doenças foliares em plantações de soja a partir das imagens capturadas por VANTs. O programa utiliza conceitos de visão computacional, além do algoritmo SLIC - *Simple Linear Iterative Clustering* - para gerar superpixels, ou seja, para agrupar pixels em regiões coesas, com características comuns de cor, textura e brilho. As imagens utilizadas na validação do programa foram obtidas em um experimento de plantio de soja.

Assim, a pesquisa parte dos seguintes problemas:

- 1- O programa proposto consegue distinguir entre as seguintes doenças foliares da soja: Ferrugem Asiática, Míldio e Mancha Alvo?
- 2- Entre as técnicas de aprendizagem automática utilizadas para classificação qual se mostrou mais adequada ao reconhecimento das doenças da soja?

Essa pesquisa faz parte dos projetos do INOVISÃO - Grupo de Pesquisa de Desenvolvimento e Inovação em Visão Computacional – que tem como principal objetivo a integração entre pesquisa, desenvolvimento e inovação para contribuir com o desenvolvimento do estado do Mato Grosso do Sul. A pesquisa também está vinculada a rede VANTAGRO que reúne universidades, empresas e demais

instituições, e pessoas interessadas no desenvolvimento de novas tecnologias baseadas em VANTs para aplicações na agricultura e na pecuária no estado do Mato Grosso do Sul.

1.1 OBJETIVOS

Esta tese tem como objetivo geral a detecção de doenças foliares na lavoura de soja, empregando um programa de computador por meio de um sistema de visão computacional utilizando imagens capturadas por VANTs.

Especificamente pretendeu-se realizar coletas de imagens em plantios de soja, utilizando um VANT. As imagens foram utilizadas para construção de 3 bancos de imagens anotadas sendo dois deles com apenas uma classe de doença (ferrugem asiática) e o terceiro contendo três classes de doenças (ferrugem asiática, mancha alvo e míldio). Esses bancos de imagens serão disponibilizados ao público para novos testes e implementações de programas que envolvam reconhecimento de doenças da soja tendo como base a visão computacional e aprendizagem supervisionada.

Além disso, com a implementação do programa para segmentação de imagens, por aprendizagem supervisionada, e para inspeção de imagens da plantação de soja, objetivou-se sua validação através da realização de testes e quantificação do resultado por medidas de desempenho.

1.2 JUSTIFICATIVA

Os avanços tecnológicos são considerados como motor do crescimento econômico global em longo prazo (THEIS; TOMKIN, 2012). Para o caso específico das atividades agrícolas onde o objetivo é o aumento da produção, há um processo de inovação constante, como o ocorrido com a Revolução Verde, na década de 1950, que trouxe ao sistema produtivo agrícola sementes melhoradas, insumos industriais (fertilizantes e agrotóxicos), mecanização e tecnologia (plantio, irrigação e colheita) (MATOS, 2010).

Informações da USDA – United States Department of Agriculture (2016) no que se refere a Safra 2015/2016, dão conta que a produção de soja no mundo foi de 313,313 milhões de toneladas, tendo como principal produtor os EUA - United States of America - com uma produção de 106,857 milhões de toneladas. O Brasil ocupa a segunda posição enquanto produtor no ranking mundial, produzindo na última Safra um total de 95.434,6 milhões de toneladas. O MAPA - Ministério da Agricultura, Pecuária e Abastecimento (2016) projeta, para a Safra de 2025/26, uma produção de 129,2 milhões de toneladas de soja, um acréscimo de 35% em relação à produção de 2015/16.

Para o MAPA (2016) há projeções de estagnação na produtividade de soja para um período de 10 anos, já que a projeção é de que a média nacional fique em torno de 3,0 toneladas por hectare. A tendência é que a expansão de soja ocorra por combinação das seguintes variáveis: expansão de fronteira em regiões onde ainda há terras disponíveis, ocupação de terras de pastagens e pela substituição de lavouras onde não há terras disponíveis para serem incorporadas. No Brasil, a expansão da soja ocorre principalmente sobre terras de pastagens. Além disso, outros fatores devem ser considerados como limitantes a expansão da produtividade da soja como as doenças (EMBRAPA, 2013).

Farias (2011) afirma que, quando se conhece o rendimento máximo de uma cultura, pode-se utilizar tecnologias e práticas de manejo para alcançá-lo. Nesse aspecto a utilização dos VANTs no monitoramento de culturas, via coleta de imagens, pode somar e/ou substituir a atual estratégia de monitoramento que se dá através de busca por sintomas, em algumas amostras dos talhões a olho nu ou com o uso da lupa (ROTHER; KSHIRSAGAR, 2015; BONALDO; RIEDO; LIMA, 2009).

Dado que o monitoramento constante da lavoura de soja é imprescindível para localizar possíveis sintomas de doenças na soja, ainda que a aplicação preventiva de fungicida já tenha sido efetuada (RICHETTI; ROESE, 2011) e que o custo dessa prevenção utilizando o conceito hora-homem seja de R\$1200,00 a cada 2.500 hectares (equivalente a 40 horas de trabalho), o sensoriamento remoto através de VANTs pode ser usado na ampliação da eficiência da produção, coletando imagens que podem definir de forma mais ágil a tomada de decisão (MACHOVINA; FEELEY; MACHOVINA, 2016).

Os programas de computadores compõem outra classe de tecnologias que podem ser utilizadas para atingir o rendimento máximo de uma cultura, com o emprego de técnicas de aprendizagem automática e visão computacional que tem sido amplamente difundida nos últimos anos (GUI *et al.*, 2015). As técnicas empregadas em programas de computadores tendem a proporcionar resultados mais precisos que os obtidos através de avaliações humanas, a exemplo dos resultados obtidos com a utilização de escalas diagramáticas que foram propostas para avaliarem a severidade de doenças na soja (GODOY *et al.*, 2006; SOARES *et al.*, 2009, KOWATA *et al.*, 2008).

Embora a soja possua doenças com características complexas e outras similares entre si (BONALDO; RIEDO; LIMA, 2009) o uso de tecnologias que empregam visão computacional e reconhecimento de padrões podem diagnosticar a presença das doenças na cultura e identificar os tipos de doenças (GUI *et al.*, 2015).

Os sistemas de visão computacional envolvem reconhecimentos de objetos em imagens e transformações dos objetos em informações que são processadas e posteriormente utilizadas. Um sistema de visão computacional tem como etapas principais a aquisição de imagens, pré-processamento, segmentação, extração de atributos, e processamento de alto nível (MILANO; HONORATO, 2010).

O desenvolvimento de um programa computacional para detecção automática de fitopatologias da parte aérea das plantas de soja passa pela necessidade de desenvolvimento de novas metodologias: novos processos ou adaptação de métodos existentes. Nessa pesquisa o algoritmo SLIC, proposto por Achanta *et al.* (2010), é utilizado para segmentações e geração de superpixels nas imagens da soja. A segmentação de imagens é considerada um dos passos-chave e a qualidade da segmentação deve afetar a confiabilidade no momento de extrair os atributos bem como na precisão no reconhecimento de padrões (GUI *et al.*, 2015).

Essa pesquisa traz contribuições acadêmicas por apontar a utilização de uma nova ferramenta para reconhecimento de doenças na soja, além de disponibilizar bancos de imagens com as classes investigadas para testes com novas implementações e parâmetros. Supõe-se que o reconhecimento das doenças foliares através do programa proposto pode contribuir com a redução dos danos provocados por essas nos talhões com a cultura da soja, bem como ampliar o lucro econômico do produtor.

Outro aspecto relevante desse estudo é a possibilidade de colaborar com a redução dos impactos que a produção de soja proporciona, bem como as atividades agrícolas de forma geral, como o agravamento de problemas ambientais, erosão dos solos e a contaminação dos recursos hídricos.

1.3 ORGANIZAÇÃO DA TESE

Esta tese está organizada no seguinte formato: além dessa introdução onde são apresentados o problema, objetivos e justificativas da pesquisa, a Seção 2 apresenta a revisão de literatura onde são abordadas informações acerca da cultura da soja, visão computacional, aprendizagem automática, métricas de avaliação e informações acerca dos veículos aéreos não tripulados. A Seção 3 apresenta a metodologia utilizada para o alcance dos resultados visualizados no texto.

Posteriormente a Tese é composta por dois artigos, sendo o primeiro decorrente do monitoramento e coleta de imagens em um ensaio experimental de plantio de soja durante a Safra 2014/2015. O segundo artigo foi elaborado a partir da coleta de imagens obtidas durante o monitoramento no experimento de plantio de soja durante a Safra 2015/2016.

2 REVISÃO DA LITERATURA

2.1 CULTURA DA SOJA

A soja é uma oleaginosa com origem na China. No Brasil o início de seu cultivo deu-se na década de 70, nos estados do sul, passando a ser cultivada posteriormente no Centro-oeste e Sudeste. Entre os principais subprodutos da soja, destacam-se, por meio do grão, o farelo e o óleo de soja. Além disso, o produto atende também a indústria farmacêutica e química dentre outros segmentos do mercado (DOMINGUES et al., 2014). A safra brasileira de soja 2015/2016 foi de 95,4 milhões de toneladas, mantendo-se em segundo lugar na produção de soja mundial, atrás apenas dos Estados Unidos (USDA, 2016), conforme Tabela 1.

Tabela 1 - Produção de soja dos principais países produtores (Em mil toneladas)

Países	2011/12	2012/13	2013/14	2014/15	2015/16
Estados Unidos	84.291	82.791	91.389	106.878	106.857
Brasil	66.500	82.000	86.700	97.200	95.434
Argentina	40.100	49.300	53.400	61.400	56.800
China	14.485	13.050	11.950	12.150	11.785
Índia	11.700	12.186	9.477	8.711	7.125
Paraguai	4.043	8.202	8.190	8.100	9.000
Canadá	4.467	5.086	5.359	6.049	6.371
Outros	14.841	15.956	15.997	19.234	18.875
Total	240.427	268.571	282.462	319.722	313.313

Fonte: USDA (dez/2016)

As exportações totais do Brasil para a safra 2016/2017 foram estimadas em aproximadamente 54 milhões de toneladas. A demanda pelo produto brasileiro tem se ampliado principalmente devido ao aumento de renda per capita e crescimento econômico da China, principal país importador do produto (conforme Tabela 2), bem como pela abertura de mercados que anteriormente eram considerados fechados para o comércio (CONAB, 2015; USDA, 2016).

Tabela 2 - Principais países importadores de soja (dados em mil toneladas)

Países Importadores	2011/12	2012/13	2013/14	2014/15	2015/16
China	59.865	70.364	78.350	83.230	86.000
União Européia	12.538	13.293	13.917	14.730	13.000
México	3.409	3.842	3.819	4.126	4.300
Japão	2.830	2.894	3.004	3.186	3.100
Tailândia	1.867	1.798	2.411	2.798	2.650
Taiwan	2.286	2.335	2.520	2.476	2.600
Egito	1.730	1.694	1.947	1.300	2.400
Indonésia	1.795	2.241	2.006	2.250	2.400
Turquia	1.249	1.608	2.197	2.283	2.250
Iran	126	301	1.311	1.864	2.120
Outros	9.497	12.699	12.882	14.747	15.393
Total	97.192	113.069	124.364	132.988	136.213

Fonte: USDA (dez/2016)

Esses fatores põem o Brasil em posição de principal país exportador de soja, conforme se observa na Tabela 3. De acordo com a CONAB – Companhia Nacional de Abastecimento (2015), dessa produção, 43% dos grãos exportados seguem in natura. Os 57% restantes são moídos, de onde 72% são transformados em farelos (50% são exportados), 18% são transformados em óleo (20% são exportados), e 10% são contabilizados como perdas.

Tabela 3 - Principais países exportadores de soja (dados em mil toneladas)

Países Exportadores	2011/12	2012/13	2013/14	2014/15	2015/16
Brasil	41.904	46.829	50.612	54.383	58.400
Estados Unidos	36.129	44.594	50.143	52.688	55.792
Argentina	7.738	7.842	10.573	9.920	9.250
Paraguai	5.518	4.800	4.488	5.300	5.300
Canadá	3.470	3.469	3.854	4.258	4.000
Outros	6.043	5.143	6.548	5.404	6.417
Total	100.802	112.677	126.218	131.953	139.159

Fonte: USDA (dez/2016)

A expansão da soja no território brasileiro deu-se a partir da demanda interna e externa pelo produto, dado que o grão encontrou no sul do Brasil um ecossistema semelhante ao verificado de forma predominante no sul dos Estados Unidos facilitando a importação e adoção de tecnologias (EMBRAPA, 2013). A soja se

expandiu para o cerrado na década de 80 alcançando os estados de Minas Gerais, Mato Grosso e Mato Grosso do Sul (DOMINGUES et al., 2014).

Participações institucionais foram importantes para a expansão produtiva, a exemplo de políticas cambiais e de comércio exterior, que impactaram positivamente o aumento das exportações e importações, bem como a participação do Banco do Brasil, enquanto subsidiadora de crédito rural. Outro fato relevante, foi a criação da EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária, em 1973 e, em 1975 a EMBRAPA/CNPSo. Efetivamente, após os anos de 1980 a soja tornou-se mais notória no Centro-Oeste brasileiro, onde se destacou a expansão da fronteira agrícola que contou com a participação, principalmente, de gaúchos, catarinenses e paranaenses (CAMPOS, 2011).

A produção brasileira de soja conta com diversos pontos favoráveis, entre os quais se destacam a disponibilidade de fatores de produção e acesso à tecnologia. Entre os desafios destacam-se questões de logística e políticas públicas, bem como questões relacionadas a fertilidade das terras disponíveis para agricultura e o uso de fertilizantes e defensivos. Outro aspecto que pode ser considerado desafiador refere-se à necessidade de desenvolvimento de cultivos adaptados às condições de determinadas microrregiões ou ainda, resistentes a doenças, como é o caso da ferrugem asiática (PINAZZA, 2008).

A região Centro Oeste produziu na última safra 43.752,6 milhões de toneladas de grãos de soja. O estado de Mato Grosso do Sul produziu pouco mais de 7 milhões de toneladas (Tabela 4).

Tabela 4 - Produção de Soja na Região Centro-Oeste do Brasil (Em mil toneladas)

	2011/12	2012/13	2013/14	2014/15	2015/16*
Mato Grosso	21.849	23.532,8	26.441,6	28.018,6	26.030,7
Mato Grosso do Sul	4.628,3	5.809,0	6.148,0	7.177,6	7.241,4
Goiás	8.251,5	8.562,9	8.994,9	8.625,1	10.249,5
Distrito Federal	176	186,7	216	147,3	231
Centro-Oeste	34.904,8	38.091,4	41.800,5	43.968,6	43.752,6

Fonte: CONAB (2016).

A soja possui dois estágios de desenvolvimento: estádios vegetativos e estádios reprodutivos, respectivamente denominados pelas letras V e R (Tabela 5). O estágio vegetativo vai até o 50º dia após a semeadura, e estágio reprodutivo se

concentra entre o 51° ao 125° dia após a semeadura (GAVA, 2014). Tão logo a maturação seja atingida, deve-se iniciar a colheita objetivando evitar perdas na qualidade do produto.

Tabela 5 - Estádios da soja e suas principais características.

Estádio	Denominação	Características
VE	Emergência (1° ao 7° dia)	Cotilédones acima da superfície do solo formando um ângulo de 90° com seus respectivos hipocótilos.
VC	Cotilédone (8° ao 14° dia)	Cotilédones completamente abertos e expandidos. As bordas de suas folhas unifolioladas não mais se tocam.
V1	Primeiro nó (15° ao 21° dia)	Folhas unifoliadas completamente abertas
V2	Segundo nó (a partir do 22° dia)	Primeira folha trifoliada aberta.
V3	Terceiro nó (a partir do 29°)	Segunda folha trifoliada aberta.
V(n)	Enésimo nó (Até o 49° dia)	“Enésimo” nó ao longo da haste principal com trifólio aberto.
R1	Florescimento (50° ao 58° dia)	Início da floração: até 50% das plantas com flor.
R2	Pleno Florescimento (59° ao 65° dia)	Floração plena: maioria dos racemos com flores abertas.
R3	Início da Formação de Vagens (66° ao 75° dia)	Final da floração: flores e vagens com até 1,5cm.
R4	Plena Formação das vagens (76° ao 87° dia)	Maioria das vagens no terço superior com 2-4cm.
R5	Início do enchimento das sementes (88° ao 100° dia)	R5.1. Grãos perceptíveis ao tato a 10% da granação; R5.2. Maioria das vagens com granação de 10%-25%; R5.3. Maioria das vagens entre 25-50% de granação; R5.4. Maioria das vagens entre 50-75% de granação, e R5.5. Maioria das vagens entre 75-100% de granação.
R6	Pleno enchimento das vagens (101° ao 111° dia)	Vagens com granação de 100% e folhas verdes.
R7	Início da maturação (112° ao 118° dia)	R7.1. Início: 50% de amarelecimento de folhas e vagens R7.2. Entre 51-75% de folhas e vagens amarelas, e R7.3. Mais de 76% de folhas e vagens amarelas.
R8	Maturação plena (119° ao 125° dia)	R8.1. Início a 50% de desfolha, e R8.2. Mais de 50% de desfolha à pré-colheita.
R9	126° dia.	R9. Ponto de maturação de colheita.

Fonte: Embrapa (2011)

O custo de produção, por hectare, da soja convencional no estado de Mato Grosso do Sul para a Safra 2015/2016 foi estimado em R\$ 2.475,05. O manejo da área corresponde a 17,7%, totalizando R\$ 438,09; os tratos culturais (22,6%) somam R\$ 559,36, por hectare. Com a soja transgênica, o custo por hectare foi de R\$ 2.451,52. O manejo da área corresponde a 17,9%, totalizando R\$ 438,82 e os

tratos culturais (21,1%) somam R\$ 517,27, por hectare. Ressalta-se que numa circunstância em que doenças fúngicas já ocorrem na plantação, o controle químico com fungicida é a principal medida de controle. Esse fato é justificado pela severidade da doença cujo alastramento dentro da lavoura pode chegar a 3 metros ao dia, sendo que o clima influencia na taxa de progresso e na severidade das epidemias (GRIGOLLI, 2015; RICHETTI, 2015).

A soja é uma planta herbácea, incluída na classe *Dicotyledoneae*, ordem *Rosales*, família *Leguminosae*, subfamília das *Papilionoideae*, gênero *Glycine* L.. Durante todo o ciclo da planta são distinguidos quatro tipos de folha: cotiledonares, folhas primárias ou simples, folhas trifolioladas ou compostas e prófilos simples. Sua cor, na maioria dos cultivares, é verde pálida e, em outras, verde escura (GAVA, 2014).

Relevante para a economia brasileira, a soja (*Glycine max*) é uma das plantas mais fáceis de serem cultivadas, porém, a exploração econômica do seu potencial de rendimento (mais de 4.000 em kg.ha⁻¹) dificilmente é alcançada (YORINORI, 1997). As doenças da soja encontram-se inseridas entre os principais fatores limitantes do rendimento, e em geral são de difícil controle, e as perdas anuais na produção de soja no Brasil causadas por doenças, são estimadas entre 15 a 20%, sendo que algumas podem ocasionar 100% de perda da safra (ZARBIN, 2009; EMBRAPA, 2013).

O número de doenças causadas por bactérias, fungos, nematóides e vírus, continua aumentando a cada safra com a expansão da cultura para novos ambientes. No Brasil, já foram identificadas aproximadamente 40 doenças. A importância de cada doença varia de ano para ano e de região para região, dependendo das variedades predominantes e da condição climática de cada safra (EMBRAPA, 2011). A Tabela 6 ilustra as principais doenças que acometem a soja nos estádios vegetativos (entre o primeiro e o 50° dia após a semeadura) no Brasil:

Tabela 6 - Principais doenças do estágio vegetativo da soja

Doenças/Estádios	VC	V1	V2	V3	VN	V9
Tombamento						
Antracnose						
Míldio da Soja						
Ferrugem da Soja						
Mancha parda da Folha						
Mancha Alvo						
Oídio da Soja						
Mofo Branco						
Mancha do mirotécio						
Seca da haste e da vagem						
Crestamento						
Viroses						
Crestamento foliar da cercóspera						
Mela da folha						

Fonte: EMBRAPA 2011

De acordo com a EMBRAPA (2011) entre as doenças causadas por nematóides destacam-se: nematóide de galhas, nematóide de cisto da soja, nematóide reniforme e nematóide das lesões radiculares. Destacam-se entre as doenças causadas por vírus: mosaico comum da soja, queima do broto, mosaico cálico e necrose da haste. Entre as doenças bacterianas destacam-se: crestamento bacteriano, pústula bacteriana, fogo selvagem.

As doenças da soja causadas por fungos foram separadas em dois grupos: doenças da parte aérea e doenças do sistema radicular. As seguintes doenças compõem o grupo de doenças da parte aérea: crestamento foliar de cercóspera, ferrugem americana, ferrugem asiática, mancha foliar de alternária, mancha foliar de ascoquita, antracnose, cancro da haste, mancha púrpura da semente, seca da haste e da vagem, seca da vagem, mancha de levedura, mancha foliar de mirotécio, mancha parda, mancha “olho-de-rã”, míldio, mancha foliar de filosticta, mancha alvo, mela ou requeima da soja, podridão branca da haste ou mofo branco, oídio.

Já o grupo das doenças do sistema radicular é composto pelas seguintes doenças: podridão de carvão, podridão parda da haste, podridão de fitóftora, podridão radicular de cilindrocládio, tombamento de escleródio, murcha de escleródio, tombamento de rizoctonia, morte em reboleira, podridão da raiz e da

base da haste, podridão vermelha da raiz (síndrome da morte súbita), podridão radicular de roselínia, podridão radicular de corinéspora.

Durante os estádios reprodutivos da soja (51° ao 125° dia após a semeadura), a produção pode ser acometida por diversas doenças, conforme pode-se visualizar na Tabela 7. Entre as doenças verificadas em ambos os estádios, a ferrugem asiática teve presença tanto no início quanto no final do ciclo. A ferrugem asiática é causada pelo fungo *Phakopsora pachyrhizi* Syd. & P.Syd. e ocorreu no Brasil pela primeira vez em 2001 (GOULART; ROESE; MELO, 2015).

Tabela 7 - Principais doenças do estágio reprodutivo da soja

Doenças/Estádios	R 1	R 2	R 3	R 4	R5. 1	R5. 2	R5. 3	R5. 4	R5. 5	R 6	R 7	R 8
Antracnose												
Míldio da Soja												
Ferrugem da Soja												
Mancha parda da Folha												
Mancha Alvo												
Oídio da Soja												
Mofo Branco												
Mancha do mirotécio												
Seca da haste e da vagem												
Crestamento												
Viroses												
Crestamento foliar da cercóspera												
Mela da folha												
Podridão por Phytophthora												
Podridão parda da haste												
Mancha olho de rã												
Mancha parda da Folha												
DFCs*												

DFCs – Doença de final de ciclo.

Na Safra 2001/2002 a ferrugem asiática atingiu cerca de 60% da área brasileira plantada com a soja, incluindo o estado de Mato Grosso do Sul (YORINORI; LAZZAROTTO, 2004). A rápida disseminação do fungo *Phakopsora pachyrhizi* eleva a progressão da doença, pois com apenas 6 ou 8 dias seu ciclo primário está completo, dando início a novos ciclos de reinfecção. O alto potencial para danos à

produção pode ser ampliado diante de condições favoráveis como temperatura entre 19 e 30°C e umidade relativa maior que 90% (BARRETO, 2011).

O fungo, ao provocar a doença na soja, leva a planta a um rápido amarelecimento e a queda prematura das folhas, interferindo na plena formação dos grãos, como demonstrado pela Figura 1. A depender da fase em que a ferrugem asiática acomete a planta, esta pode levar ao aborto e queda das vagens, podendo provocar perda total na produtividade (PERIN, 2013).



Figura 1 - Reboleira de Ferrugem Asiática na plantação de soja

De acordo com as informações do Consórcio Anti Ferrugem, no Brasil foram verificadas 399 ocorrências de ferrugem asiática, na Safra 2015/2016. Nos últimos 5 anos o número de ocorrências da doença no estado de Mato Grosso do Sul tem se reduzido, indicando a preocupação em manter o controle. Na safra 2015/2016 foram verificadas 67 ocorrências de ferrugem asiática nas sojas. Monitorar a produção da soja em todos os seus estádios é um processo valioso para a identificação da ferrugem da soja em estágio inicial, visando a aplicação de fungicidas após detecções dos sintomas iniciais da doença na lavoura. Além disso as seguintes estratégias de controle são recomendadas pela EMBRAPA (2011):

“a utilização de cultivares de ciclo precoce e semeaduras no início da época recomendada, a eliminação de plantas de soja voluntárias e a ausência de cultivo de soja na entressafra por meio do vazio sanitário, o monitoramento da lavoura desde o início do desenvolvimento da cultura, a utilização de fungicidas no aparecimento dos sintomas ou preventivamente e a utilização de cultivares resistentes, quando disponíveis”.

Além da ferrugem da soja ser uma limitação ao potencial máximo produtivo da soja no estado de Mato Grosso do sul, de acordo com Grigolli (2015) outras doenças marcaram a produção no estado durante a Safra 2014/2015, sendo as principais a antracnose e a mancha-alvo.

De acordo com a EMBRAPA (2011) a maior preocupação com a ferrugem se deve a severidade da doença, que após infectar a produção, causa amarelecimento e queda prematura das folhas, influenciando no tamanho dos grãos, ampliando perdas de rendimento e de qualidade. Quando a plantação é atingida pela doença na fase de formação das vagens ou início da granação, pode causar o aborto e a queda das vagens, o que pode levar a perda total do rendimento. No Brasil foram observadas reduções de produtividade de até 80% ao comparar áreas tratadas e não tratadas com fungicidas.

2.2 VISÃO COMPUTACIONAL

A visão tem sido objeto de pesquisa por se tratar de um notável sistema de percepção do ser humano, capaz de processar e interpretar grandes quantidades de dados de natureza visual. A visão computacional é a ciência responsável pela forma como um computador enxerga o meio à sua volta, extraindo informações significativas a partir de imagens capturadas por dispositivos imageadores como o scanner, o digitalizador de sinal de vídeo (frame/vídeo grabber), a câmera digital, entre outros (MOLZ, 2001; MILANO; HONORATO, 2010).

A forma como o ser humano identifica uma cena e como uma imagem digital é interpretada diferem-se entre si, pois o primeiro é capaz de perceber a estrutura tridimensional do mundo que o rodeia com aparente facilidade enquanto que o segundo identifica uma imagem a partir de atributos que devem ser extraídos da imagem, e que estão relacionados entre si (SZELISKI, 2010; RUDEK *et al.*, 2008).

Na visão computacional, as soluções e pesquisas desenvolvidas para aplicações específicas foram obtidas por métodos de tentativas e erros, o que faz da mesma uma ciência experimental (MOLZ, 2001). No início da década de 1970, a visão de computador era tida como um componente de percepção visual, participando de uma agenda ambiciosa que tinha por objetivo imitar a inteligência humana, dotando robôs de comportamentos inteligentes. Já na década de 1980, o foco e atenção centravam-se em técnicas matemáticas mais sofisticadas para a análise de imagens e cenários (SZELISKI, 2010).

Na década de 1990, diversos tópicos de pesquisas da década anterior continuaram a ser explorados, e alguns deles tornaram-se significativamente mais ativos. Um dos trabalhos iniciados na década de 1980 que tratava do uso de medições detalhadas de cor e intensidade, combinados com modelos físicos precisos de cenários e formação de imagens coloridas veio a instituir seu próprio subcampo conhecido como visão com base na física. Ainda nessa década, desenvolveu-se o notável estudo no campo da visão de computador: o aumento da interação com computação gráfica especialmente na área interdisciplinar de modelagem e renderização baseada em imagem (SZELISKI, 2010).

A década dos anos 2000 continuou com um aprofundamento da interação entre os campos de visão e gráficos. A tendência final é a aplicação de técnicas de aprendizado de máquina sofisticadas para problemas de visão de computador. Atualmente o uso da visão computacional se dá em diversas áreas, como por exemplo: reconhecimento óptico de caracteres, fotogrametria, imagens médicas, segurança automotiva, jogo de vídeo e estabilização, captura de movimento, vigilância, reconhecimento de impressão digital e biometria, detecção de face, autenticação visual entre outros (SZELISKI, 2010).

A visão computacional, de forma geral, resolve problemas demandados por outras áreas de pesquisa. Os sistemas de visão computacional basicamente envolvem reconhecimento de objetos em imagens e transformações dos objetos em

informações que são processadas e posteriormente utilizadas em algum sistema. Logo, a visão computacional propicia ao computador informações precisas extraídas de imagens e vídeos, levando o computador a executar tarefas inteligentes, imitando e até mesmo aproximando-se da inteligência humana. As principais etapas de um sistema de visão computacional (Figura 2) são: aquisição de imagens, pré-processamento, detecção e segmentação, extração de atributos ou características e, processamento de alto nível (MILANO; HONORATO, 2010).

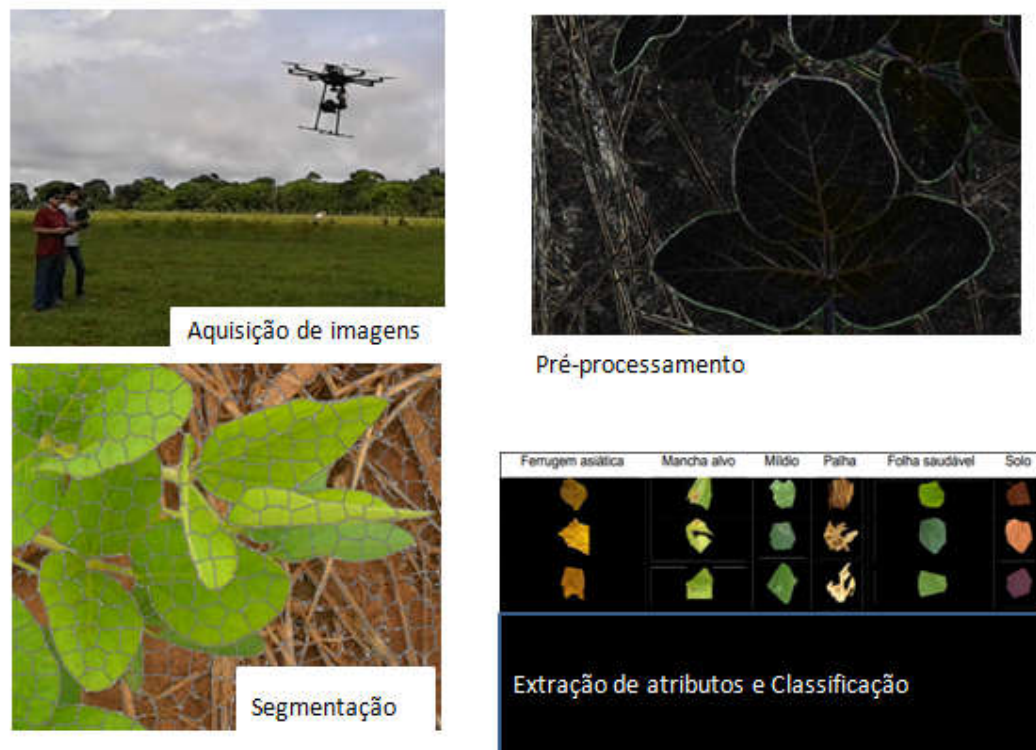


Figura 2 - Etapas de um sistema de visão computacional

Aquisição de Imagem é o primeiro passo no sistema de visão computacional, dando-se a partir de sensores de câmeras. A imagem pode ser bidimensional, tridimensional ou uma sequência de imagens. O próximo passo é o pré-processamento que ocorre antes da obtenção das informações de uma imagem, Nesse passo aplicam-se métodos que facilitam a identificação de um objeto para posteriormente realizar a extração e verificar as características que compõem uma imagem. Nessa etapa pode-se optar pela redução de ruídos e o realce das imagens, com o objetivo de eliminar imperfeições e proporcionar melhor qualidade da imagem para etapas posteriores (MILANO; HONORATO, 2010; QUINTA, 2013).

A segmentação é o processo que é realizado com o objetivo de destacar regiões relevantes da imagem, segmentando-as para um processamento posterior. Na etapa de segmentação, busca-se um particionamento da imagem em regiões de forma a separar elementos de interesse, para o problema a ser resolvido, de elementos que são irrelevantes para o problema. Em alguns casos, em problemas que envolvem contagem ou reconhecimento de múltiplos objetos, a segmentação, além de separar os elementos irrelevantes, separa os objetos de interesse em regiões distintas (MILANO; HONORATO, 2010).

Após a segmentação, realiza-se a extração de atributos que tem como objetivo caracterizar os objetos permitindo a identificação de pontos de interesse que são relevantes para realizar o reconhecimento de uma imagem ou classe (OLIVEIRA et al., 2009; BORTH et al., 2013).

O último passo refere-se ao processamento de alto nível que é um processo que inclui validação dos dados obtidos, estimativa de parâmetros sobre a imagem e classificação dos objetos obtidos em diferentes categorias (MILANO; HONORATO, 2010). A classificação pode ser obtida com a utilização de algoritmos de reconhecimento de padrões. Embora a visão computacional se utilize dessa sequência de etapas para o desenvolvimento de algumas tarefas, em algumas circunstâncias não é necessário o emprego de todas as etapas para a solução de um problema (QUINTA, 2013).

2.2.1 Segmentação

A segmentação de imagens é uma etapa muito importante para a visão computacional e tem como objetivo a simplificação da imagem, possibilitando realizações de operações computacionais e matemáticas no domínio de um problema para solucioná-lo. É um passo que pode determinar o sucesso ou o fracasso de uma análise (DLUGOSZ et al., 2009; PINTO, 2014).

A segmentação consiste na subdivisão da imagem em suas partes ou objetos que a constitui, considerando algumas de suas características intrínsecas, como, por exemplo, a cor ou a proximidade dos pixels (FIGUEIREDO et al., 2009; GONZALEZ; WOODS, 2010; NEVES; PELAES, 2001). O resultado final esperado de um processo de segmentação é um número finito de regiões homogêneas, que

individualizam as diferentes regiões contidas numa imagem. Dessa forma, a imagem seria formada por um conjunto de regiões, conectadas pelas suas bordas e que não se sobrepõe, de tal forma que cada pixel da imagem pertence unicamente a uma determinada região (SALDANHA; FREITAS, 2008).

A segmentação de uma imagem é o processo pelo qual se extraem, do domínio da imagem, uma ou mais regiões conectadas que satisfaçam o critério de uniformidade. Para essa segmentação, é importante escolher um algoritmo adequado para imagens de cenas naturais em cores e com texturas, pois a qualidade da segmentação, de forma geral, é determinada pelo algoritmo de segmentação utilizado e pelos valores de seus parâmetros (CECHINEL, 2000; CAVANI et al., 2006; MARTIN, 2002; FEITOSA et al., 2007).

2.2.2 Segmentação por superpixel

Em aplicações de Visão Computacional, a idéia da utilização de superpixels tem se tornado bastante comum. O uso do termo diz respeito a um agrupamento de pixels em regiões coesas, com características comuns de cor, textura e brilho dentro da região, conforme Figura 3. Para tanto foi desenvolvido um classificador que combina características como contorno, textura, brilho e continuidade, agrupando-as. A esse agrupamento deu-se o nome de superpixels (REN; MALIK, 2003).



Figura 3 - Exemplo de segmentação de imagem da soja demonstrando a imagem original e a imagem segmentada em superpixel

O trabalho de Ren e Malik teve como base os princípios da Gestalt (SILVA, 2010). Trata-se de uma teoria psico-visual proposta em 1935 por Max Wertheimer, Wolfgang Kohler e Kurt Koffka, objetivando explicar a percepção visual humana através de mecanismos conhecidos como Leis da Gestalt (KLAVDIANOS et al., 2013). Gestalt é um substantivo alemão que possui dois significados. O primeiro refere-se a forma, e o segundo refere-se a uma entidade concreta que possui entre seus vários atributos a forma, sendo esse o significado que os 'gestaltistas' utilizam (ENGELMANN, 2002).

Princípios de agrupamento como, por exemplo, a proximidade, similaridade e boa continuação foram desenvolvidas pela teoria clássica da Gestalt. O princípio da boa continuação afirma que uma boa segmentação deve ter limites suaves. O princípio da similaridade parte de dois pressupostos (REN; MALIK, 2003):

- Semelhança intra-região: os elementos em uma região são semelhantes, o que inclui brilho, textura, e baixo consumo de energia de contorno dentro da região;
- (Dis)similaridade inter-região: os elementos em diferentes regiões são diferentes, incluindo brilho, textura, e de alta energia de contorno na região de fronteiras.

Embora estes princípios clássicos de agrupamento tenham inspirado abordagens diversas no que se relaciona a segmentação, os princípios da Gestalt são *ceteris paribus* regras, indicando que distinguem segmentações concorrentes somente quando o resto é, em seu todo, igual (REN; MALIK, 2003).

Ao longo de sua utilização vários algoritmos de superpixels foram propostos, no entanto, a escolha de um algoritmo de superpixel depende da aplicação a qual se destina, embora seja de senso comum que os seguintes requisitos básicos estejam presentes (ACHANTA et al., 2010; STUTZ; HERMANS; LEIBE, 2014):

- Superpixels devem respeitar os limites do objeto;
- Superpixels devem ser gerados de forma tão eficiente quanto possível;
- Superpixels quando usados para reduzir a complexidade computacional como uma etapa de pré-processamento, devem ser rápidos para computar, eficientes no uso da memória e de fácil utilização.

Os algoritmos de superpixels capturam redundâncias da imagem e reduzem a complexidade das tarefas subseqüentes de processamento da imagem. Todavia sua utilização exige atenção na escolha do algoritmo e parâmetros a serem empregados devido aos riscos quanto a perda de dados relativos às bordas da imagem quando estas são inseridas dentro do superpixel (NEUBERT; PROTZEL, 2012).

Nesta pesquisa foi utilizado o método SLIC (*Simple Linear Iterative Clustering*) que foi introduzido por Achanta *et al.* (2010) e é uma adaptação do método de agrupamento *k-means* para geração de superpixels. O parâmetro 'k' do algoritmo refere-se ao número de superpixels que se pretende ajustar na imagem. Opcionalmente é possível ajustar o parâmetro 'compacidade' que permite controlar a forma do superpixel tornando-a mais quadrada/cúbica. Neste trabalho também foi utilizada a configuração do parâmetro 'sigma', que permite aplicar uma suavização na imagem, utilizando filtros gaussianos, antes da segmentação, através da utilização da biblioteca *scikit-image*. O algoritmo SLIC proporciona redução do número de cálculos de distância, pois o espaço de busca passa a limitar-se ao tamanho de um superpixel (LV, 2015). A Figura 4 exemplifica o emprego do SLIC na segmentação.



Figura 4 - Segmentação por superpixel com o SLIC

De acordo com Achanta *et al.* (2010), o algoritmo superpixel SLIC agrupa locais de pixels no espaço 5-D definido por L, a, b (valores do espaço de cor CIELAB) e as coordenadas x e y dos pixels. Para tanto, tendo a imagem de entrada, realiza-se o particionamento da imagem em regiões retangulares, definindo-se o número k correspondente a quantidade de unidades, levando cada superpixel a ter aproximadamente $\frac{N}{k}$ pixels, onde N é o número de pixels da imagem. Cada região compõe um superpixel inicial de dimensões $S \times S$, onde $S = \sqrt{\frac{N}{k}}$. Os centros dos agrupamentos de superpixel $C_k = [l_k, a_k, b_k, x_k, y_k]$ com $k = [1, k]$ são escolhidos, espaçados em um grid regular para formar os agrupamentos de tamanho aproximado S^2 . Os centros são movidos para o valor de menor gradiente sobre uma vizinhança de 3×3 pixels, evitando a alocação do centróide em regiões de bordas ou que tenham pixels ruidosos. Em vez de utilizar uma norma euclidiana simples no espaço 5D, usa-se uma medida de distância D definida como se segue:

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \quad (1)$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \quad (2)$$

$$D_s = d_{lab} + \frac{m}{S} * d_{xy} \quad (3)$$

onde D_s é a soma da distância d_{lab} (1) e a distância d_{xy} (2) normalizada pelo intervalo S. A variável m corresponde ao controle de compactação do superpixel, quanto maior o seu valor mais a proximidade espacial é enfatizada e mais compactado é o agrupamento. Cada pixel da imagem é associado a um centro mais próximo de um grupo cuja área de busca sobrepõe este pixel, depois de todos os pixels estarem associados a um centro, um novo centro é calculado com o vetor Labxy de todos os superpixels que pertencem ao grupo. No final do processo alguns pixels podem ficar conectados a um grupo incorretamente, e por isso o algoritmo reforça a conectividade na última etapa atribuindo os pixels sozinhos aos maiores grupos vizinhos (ACHANTA *et al.*, 2010; SARATH, 2014).

2.2.3 Extração de atributos

O processo de extrair e selecionar atributos de uma imagem pode determinar os bons resultados na classificação. Atributos são os aspectos, qualidades ou

características de um objeto, e podem ser simbólicos, numéricos ou ambos. Considera-se como padrão, um conjunto de atributos que definem um objeto. Assim, a extração de características mais relevantes da imagem evidencia diferenças e similaridades entre os objetos (OLIVEIRA, 2005).

Em um sistema eficiente de reconhecimento de padrões, o vetor de atributos criado pela extração deve conter atributos pertinentes sobre as classes a serem reconhecidas. Um vetor de atributos é composto de vários atributos que representa um objeto (OLIVEIRA *et al.*, 2009). Diante do exposto, ressalta-se que a proposta implementada nesse trabalho inclui os seguintes extratores de atributos:

1. Atributos de cor RGB, HSV, Cielab (mínimo, máximo, média e desvio padrão)
2. Descritor de forma, invariante a escala, translação e rotação: 7 momentos de Hu
3. Atributos de Textura – GLCM (contrastes, dissimilaridades, homogeneidades, asm (segundo momento angular), energias, correlações)
4. Forma e orientação: HOG
5. Atributos de textura: LPB .

2.2.3.1 Atributos de cor

A cor pode ser considerada uma das características mais importantes por possibilitar ao ser humano o reconhecimento das imagens. Além disso, a cor depende da reflexão da luz para os olhos, bem como do processamento dessa informação pelo cérebro (LEÃO; ARAÚJO; SOUZA, 2005). Portanto, faz-se importante definir qual espaço de cor será utilizado para extração desses atributos.

Para a visão computacional, a imagem a ser analisada é a fonte de informação. As imagens são formadas por conjuntos de pixels e os pixels são as unidades mínimas que compõem a imagem (GONZALEZ; WOODS, 2010). As imagens digitais podem assumir diversos espaços de cores, onde os pixels assumem valores diferentes, podendo citar os espaços de cores em tons de cinza, RGB, HSV e CIELAB.

Os espaços baseados em tons de cinza possuem uma variação de 256 (2^8) tonalidades, iniciando com a cor preta (0) ao branco (255) (QUINTA, 2013). O

espaço de cor RGB combina três canais de cor: *Red*, *Green* e *Blue* (respectivamente: vermelho, verde e azul), e atribui um valor de intensidade a cada pixel. Representa a cor natural onde as cores são criadas por adição e mistura das cores primárias. Esse espaço de cor foi proposto em 1801, por Thomas Young (1773-1829), para o qual “os cones dos olhos seriam subdivididos em três categorias, cujos valores máximos de sensibilidade situam-se próximos do vermelho, verde e azul” (PEDRINI, 2014).

A base do espaço de cor é um sistema de coordenadas, que tem como subespaço de interesse o cubo no qual os valores RGB (vermelho, verde e azul) primários estão em três vértices, as cores secundárias (ciano, magenta e amarelo) estão em outros três vértices, o preto está na origem e o branco está no vértice mais distante da origem (GONZALEZ; WOODS, 2010).

Em uma imagem colorida, utilizando imagens de 8 bits por canal (as imagens RGB usam três cores ou canais para reproduzir as cores), os valores de intensidade variam de 0 (preto) a 255 (branco) para cada um dos componentes de cores RGB (STOPPA, 2009).

Na observação de um objeto colorido, o ser humano procura descrevê-lo por sua nuance, saturação e brilho (GONZALEZ; WOODS, 2010). No espaço de cor HSB a cor é definida por esses 3 valores distintos: *Hue* (nuance ou matiz), *Saturation* (saturação) e *Brightness* (brilho). O espaço de cor baseia-se na percepção humana da cor, sendo representada por um sistema 3D de coordenadas polares. O matiz determina a cor ou tonalidade (amarelo, laranja, vermelho, etc). O brilho determina a intensidade percebida (cor mais clara, ou mais escura) e a saturação determina a profundidade da cor (esmaecida, intensa, etc) (STOPPA, 2009).

O espaço de cores CIELAB foi adotado em 1976 utilizando-se de três coordenadas denominadas com as letras L^* , a^* e b^* . O modelo também tem como base a percepção humana da cor. Assim, o L^* é o eixo referente a claridade. Sobre essa coordenada encontram-se todas as tonalidades de cinzas neutras, entre o preto e o branco. O eixo a^* define a tonalidade da cor, tendo como limites o verde e o vermelho. O eixo b^* define a intensidade ou pureza da cor, tendo como limites a cor azul e o amarelo (TAKATSUI, 2011).

2.2.3.2 Atributos de textura

A textura refere-se a estrutura que define as superfícies formadas pela repetição de um determinado elemento ou mesmo de vários elementos em diferentes posições espaciais relativas. Geralmente, a repetição envolve variações locais de escala, orientação ou outras características geométricas e ópticas dos elementos. Ela contém informações importantes sobre o arranjo estrutural da superfície. Além de descrever a relação da superfície com o ambiente circundante também descreve a composição física distinta de uma superfície (SARATH, 2014).

Na extração de atributos de texturas a Matriz de Co-ocorrência GLCM (*Gray-Level Co-occurrence Matrix*) é uma técnica muito utilizada, tendo sido desenvolvida na década de 70 e refere-se a ocorrência de combinações diferentes de valores de intensidade dos pixels (níveis de cinza) em uma imagem. A matriz de co-ocorrência descreve a textura através de um conjunto de características para as ocorrências de cada nível de cinza nos pixels da imagem considerando múltiplas direções (HARALICK; SHANMUGAM; DINSTEN, 1973; SCHWARTZ; HÉLIO PEDRINI, 2003).

Após o cálculo da matriz de co-ocorrência, outra matriz é calculada: a matriz de probabilidade de ocorrência das combinações entre os níveis de cinza. Com essa matriz são realizados os seguintes cálculos dos atributos de textura: a energia, entropia, variância (contraste), homogeneidade, dissimilaridade e medidas de correlação (OLIVEIRA et. al., 2012).

Outro extrator de atributos de textura é o *Local Binary Patterns* – LBP que foi proposto inicialmente em (OJALA; PIETIKÄINEN; HARWOOD; 1996) e mais tarde generalizado em (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2002). São considerados um dos melhores extratores de textura e tem como vantagens sua invariância a mudanças em tons de cinza e eficiência computacional. Sua estratégia para detecção de textura é observar para um ponto central a variação da sua cor em relação aos seus vizinhos.

2.2.3.3 Atributos de forma

Na extração dos atributos de forma foram utilizados como descritores de forma a técnica de momentos invariantes, também chamada de momentos Hu e Histogramas

de Gradientes Orientados – HOG. Os momentos invariantes de uma imagem permitem calcular a área de um objeto, ou mesmo identificá-lo ainda que sofra mudança de tamanho ou que seja rotacionado (HU, 1962). O algoritmo HOG é um descritor que calcula o histograma da orientação dos gradientes na imagem. É uma técnica de extração de atributos que inicialmente teve por objetivo auxiliar na detecção de pessoas em imagens (DALAL; TRIGGS, 2005).

2.3 APRENDIZAGEM AUTOMÁTICA

Na área de Inteligência Artificial os métodos de Aprendizagem Automática (*Machine Learning*) desempenham um importante papel, dado que resolução de problemas, prova de teoremas, processamento da linguagem natural, robótica e sistemas especialistas são sub áreas que utilizam tais métodos (ANNES, 1999).

O objetivo da Aprendizagem Automática é o desenvolvimento de técnicas que permitam aos computadores aprender com a experiência, ou seja, criar programas capazes de induzir conhecimentos a partir de informação estruturada e não estruturada em forma de exemplos. A Aprendizagem Automática trabalha com os métodos computacionais que torna possível que computadores não somente aprendam, mas que também melhorem o seu desempenho com a experiência (CAPELA; GEORGIEVA, 2012).

Na aprendizagem automática, quatro tipos de algoritmos de aprendizagem são utilizados (OSÓRIO, 1999; LORENA; CARVALHO, 2007):

- Aprendizagem supervisionada: as informações das classes são previamente conhecidas.
- Aprendizagem não supervisionada: as informações das classes não são conhecidas.
- Aprendizagem semi-supervisionada: quando algumas das informações referentes às classes são obtidas.
- Aprendizagem por reforço: permite o aprendizado a partir da interação com o ambiente no qual se está inserido.

Destaca-se que todo exemplo possui um atributo especial, denominado classe ou rótulo, que descreve o fenômeno de interesse que se deseja aprender para fazer previsões a respeito (REZENDE, 2003). Alguns métodos típicos de classificação têm sido usados de forma bem sucedidas. Assim, nessa pesquisa são utilizados os seguintes algoritmos de classificação: máquinas de vetores de suporte, k vizinhos mais próximos e os algoritmos de árvores de decisão: j48 e floresta aleatória.

2.3.1 Algoritmos de classificação ou Classificadores

Os estudos teóricos que embasaram as SVMs (*Support Vector Machines*) foram iniciados por Vladimir Vapnik e estabeleceram princípios a serem seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu. O emprego da SVM possibilita resoluções de problemas de classificação de dados, gerando classificadores que apresentam bons resultados (LORENA; CARVALHO, 2007). O SMO (Sequential Minimal Optimization) é um algoritmo considerado eficiente para a implementação da técnica SVM (Support Vector Machine), para resoluções de problemas de programação quadrática, utilizado para acelerar o treinamento de SVM. Esse algoritmo particiona grandes problemas em séries de pequenos possíveis problemas, a serem resolvidos analiticamente (GIRARDELLO, 2010).

O *K-Nearest Neighbor* (KNN) é um método baseado em instâncias, ou seja, agrupamento dos dados que faz a classificação entre os k-vizinhos mais próximos. Na determinação de um elemento não pertencente ao conjunto de treinamento, esse classificador procura k elementos do conjunto de treinamento que possua a menor distância em relação a esse elemento desconhecido (GALVÃO; HRUSCHKA JÚNIOR, 2004; SILVA, 2005). De forma geral, a medida de proximidade mais utilizada é a distância euclidiana.

De acordo com Vasconcellos et al. (2011) o J48 é um algoritmo de indução de árvore de decisão derivado da proposta de Quinlan (1993) como C4.5 que representa uma significativa evolução do ID3 (QUINLAN, 1986). O algoritmo evoluiu do ID3 até a versão C4.8, que é a versão implementada no Weka em linguagem

Java. Tem como base a árvore de decisão, que é formada a partir de nós de decisão que tomam uma escolha no valor de um atributo. Utilizando um conjunto de dados, o algoritmo constrói árvores de decisão, compostas das folhas que contém as classes que devem ser classificadas e os nós são atributos da imagem que apresentam melhor eficiência (VASCONCELOS, 2011 et al.; BHARGAVA et al., 2013).

Já o classificador Random Forest (Floresta Aleatória) é um método proposto por Breiman (2001) e consiste em um conjunto de árvores de decisão combinadas. Cada árvore de decisão é construída utilizando uma amostra aleatória inicial dos dados e, a cada divisão desses dados, um subconjunto aleatório de m atributos é utilizado para a escolha dos atributos mais informativos. Ao fim, a floresta aleatória gera uma lista dos atributos mais importantes no desenvolvimento da floresta, que são determinados pela importância acumulada do atributo nas divisões dos nós de cada árvore da floresta (JAMES; HASTIE; TIBSHIRANI, 2013).

2.3.2 Aprendizagem automática e classificação de doenças

As características de doenças que atuam nas culturas, a depender das técnicas de monitoramento escolhidas pelo produtor, podem ser de difícil identificação. Além disso, monitoramentos superficiais podem dificultar a correta identificação da patologia deixando passar alguns sintomas das patologias despercebidas principalmente porque alguns sintomas são comuns a várias doenças em sua fase inicial, e a evolução de alguns sintomas não são perceptíveis ao olho humano (AFRIDI et al, 2014; BONALDO; RIEDO; LIMA, 2009).

Nesse contexto os programas de computadores podem ser utilizados para o reconhecimento de doenças através da análise de imagens, através do emprego de técnicas de aprendizagem automática e visão computacional que tem sido amplamente difundida nos últimos anos (GUI et al, 2015).

As técnicas empregadas em programas de computadores tendem a ser mais precisas que os resultados obtidos através de avaliações humanas, a exemplo dos resultados obtidos com a utilização de escalas diagramáticas que foram propostas para avaliarem a severidade de doenças na soja (GODOY, 2006; SOARES et al., 2009, KOWATA et. al, 2008). Assim, o uso de tecnologias que empreguem visão

computacional e reconhecimento de padrões pode diagnosticar a presença das doenças na cultura e identificar com precisão os tipos de doenças (GUI et al, 2015).

Rumpf et al (2010) propuseram um sistema computacional para a detecção e diferenciação de doenças na cultura da beterraba sacarina. As plantas foram cultivadas na estufa em vasos de plástico sendo regadas conforme necessário e fertilizadas semanalmente, e quando as plantas estavam com quatro folhas completamente desenvolvidas, 15 plantas foram inoculadas com os agentes patogênicos de cada tratamento (cercosporiose, ferrugem de beterraba sacarina e oídio) e como controle 15 plantas saudáveis foram mantidas não inoculadas. A reflectância espectral foi medida usando um espectrofotômetro portátil, e os dados hiperespectricos foram registrados, por um período de 21 dias após inoculação. Nove índices de vegetação espectral, relacionados aos parâmetros fisiológicos foram utilizados como características para a classificação automática.

O programa utiliza Maquinas de Vetores de Suporte (SVM) e índices espectrais de vegetação gerados a partir do processamento de imagens aéreas para classificar, com 97% de precisão, folhas saudáveis e doentes de beterraba. Na classificação múltipla de folhas de beterraba saudáveis e doentes (cercosporiose, ferrugem de beterraba sacarina e oídio) o programa alcançou uma precisão superior a 86% (RUMPF et al, 2010).

Dandawate e Kokare (2015) criaram um sistema que classifica as imagens das folhas de soja como saudáveis e doentes usando Maquinas de Vetores de Suporte (SVM). A aquisição de imagens foi realizada com o uso de uma câmera móvel com resolução de 5 mega pixels. O banco de imagens foi composto de um total de 120 imagens, incluindo imagens de folhas saudáveis e doentes, no formato JPG no tamanho de 1920 x 2560 pixels. O programa utiliza o algoritmo SIFT (Scale Invariant Feature Transform) que reconhece as espécies de plantas com base na forma de folha. Os resultados demonstram uma precisão média de 93,79% na classificação das folhas.

Pires et al. (2016) propuseram uma abordagem para o reconhecimento da doença de soja com base em técnicas que utilizam cinco descritores locais (SURF, HOG, DSIFT, SIFT, e PHOW) e Histograma de Palavras Visuais (BoVW - do inglês Bag-of-Visual-Words) para definir uma metodologia capaz de representar

computacionalmente imagens de folhas de soja, mantendo a informação visual sobre possíveis doenças.

Na formação do banco de imagens verificou-se na plantação de soja a presença de folhas saudáveis, folhas com míldio e folhas com ferrugem, todavia essa última classe foi subdividida em duas conforme as suas características de cor das lesões em: em castanho-avermelhada (RB -reddish-brown) e, castanho-claras (TAN). As folhas foram escaneadas com o scanner HP Scanjet 1200, com resolução de 1200 dpi, com imagens geradas em formato tiff. Foram selecionadas 300 imagens de cada classe (Folhas saudáveis, Míldio, Ferrugem RB e Ferrugem Tan) totalizando 1.200 imagens secundárias. Como resultado obteve-se mais de 98% de precisão, detectando as doenças da soja em tempo médio de 0,1s por imagem. Além disso, concluíram através dos resultados, que o descritor PHOW funciona melhor para espaços coloridos se comparado com a escala de cinza, atingindo taxas corretas de classificação variando de 96,25 a 99,83.

2.4 MÉTRICAS DE AVALIAÇÃO

Métricas de avaliação precisam ser definidas para possibilitar a medição do desempenho dos segmentadores e/ou classificadores utilizando-se o método proposto neste trabalho. Para Machado et al. (2012) a comparação quantitativa do resultado de um algoritmo em relação as imagens de referência (*ground-truth*) é a maneira mais conhecida de se medir o desempenho cuja avaliação pode ser obtida pelo acerto em termos de pixels ou em regiões da imagem. No que se refere a avaliação baseada em pixel, os resultados tem como base a comparação entre os pixels das imagens de referência e das imagens de saída. Para melhor exemplificar, será utilizada a matriz de confusão para mostrar o número de classificações corretas em oposição às classificações preditas para cada classe (Tabela 8).

Destaca-se que todo exemplo possui um atributo especial, denominado classe, que descreve o fenômeno de interesse que se deseja aprender para fazer previsões a respeito. Assim, considerando problemas para classificação com apenas duas classes (geralmente rotuladas como “+” e “-”), as escolhas são estruturadas para predizer a ocorrência ou não ocorrência de um evento ou hipótese. Assim, os dois

erros possíveis são denominados de falsos positivos e falsos negativos (REZENDE, 2003).

Tabela 8 - Desempenho da classificação para duas classes

Classe	Predito como C+	Predito como C-
C+	Verdadeiros Positivos	Falsos Negativos
C-	Falsos Positivos	Verdadeiros Negativos

Considerando a Tabela 8, quatro situações podem ocorrer (REZENDE, 2003):

- O fenômeno pertence à classe C+ e é predita corretamente pelo classificador como pertencente à classe C+. Neste caso, o exemplo é um verdadeiro positivo (VP);
- O fenômeno pertence à classe C- e é predita corretamente pelo classificador como pertencente à classe C-. Neste caso, o exemplo é um verdadeiro negativo (VN);
- O fenômeno pertence à classe C- e é predita pelo classificador como pertencente à classe C+. Neste caso, o exemplo é um falso positivo (FP);
- O fenômeno pertence à classe C+ e é predita pelo classificador como pertencente à classe C-. Neste caso, o exemplo é um falso negativo (FN).

A predição diz respeito ao aprendizado de critérios de decisão que serão utilizados para a classificação de novas imagens, ou seja, dado um conjunto de imagens conhecidas (conjunto de treino) pode-se prever ou diagnosticar o que irá acontecer a outras imagens (BUCENE, 2002). Por meio da classificação dos pixels pode-se calcular medidas quantitativas.

Em problemas que envolvam classificação com mais de duas classes as medidas de desempenho mais comuns consideram a capacidade do modelo de discernir uma classe versus todas as outras. A classe de interesse é conhecida como a classe positiva, enquanto todas as outras são conhecidas como negativas (LANTZ, 2015).

Coeficiente de Jaccard - CJ, também conhecido como o coeficiente de similaridade de Jaccard é uma estatística utilizada para comparar a semelhança e

diversidade entre as imagens (MANNING; SCHUTZE, 1999). O coeficiente de Jaccard é definido pela Equação (4):

$$CJ = \frac{VP}{VP+FP+FN} \quad (4)$$

Onde: VP = verdadeiro positivo, FP= falso positivo e FN= falso negativo.

Para medir o grau de associação entre duas imagens pode-se utilizar os coeficientes de associação, como o **Coeficiente de Yule (CY)** que mede o grau de associação entre duas imagens. Esse coeficiente é definido pela Equação (5):

$$CY = \left| \frac{VP}{VP+FP} + \frac{VN}{VN+FN} - 1 \right| \quad (5)$$

Onde: VP = verdadeiro positivo, FP= falso positivo, FN= falso negativo e VN=verdadeiro negativo.

Destaca-se que o Coeficiente de Yule varia entre -1 e +1, inclusive, isto é, $-1 \leq CY \leq +1$. Logo a interpretação do valor do coeficiente de associação se dá da seguinte forma:

CY=1: associação perfeita positiva;

CY=-1: associação perfeita negativa;

CY=0: associação nula;

$0 < CY < 1$: associação positiva;

$-1 < CY < 0$: associação negativa.

Porcentagem de Classificação Correta (PCC) - é a taxa de classificação correta obtida em um conjunto de imagens e pode ser obtida a partir da Equação (6).

$$PCC = \frac{VP+FP}{VP+FP+VN+FN} \quad (6)$$

Onde: VP = verdadeiro positivo, FP= falso positivo, FN= falso negativo e VN=verdadeiro negativo.

Medida-F - é uma medida que avalia a capacidade de predição de um modelo. Ela é a média harmônica entre precisão e sensibilidade. Precisão é a fração correta do total previsto como pertencente a uma classe pelo total que foi previsto como pertencente a tal classe. A **precisão** é calculada para cada classe conforme a Equação 7:

$$p = \frac{VP}{VP+FP} \quad (7)$$

Onde: VP = verdadeiro positivo e FP= falso positivo.

Sensibilidade é a fração correta do total previsto como pertencente a uma classe pelo total de exemplos cuja classe verdadeira é positiva, ou seja, é a proporção das amostras corretamente classificadas como positivas por todas as amostras realmente positivas. A precisão é calculada para cada classe conforme a Equação 8:

$$r = \frac{VP}{VP+FN} \quad (8)$$

Onde: VP = verdadeiro positivo, e FN= falso negativo.

Dessa forma quanto melhor o modelo avaliado mais próximo de 1 será o valor da medida, e caso o modelo tenha uma capacidade de predição reduzida ele irá possuir uma Medida-F próxima a 0. Para calcular a Medida-F do modelo como um todo, pode-se utilizar dois diferentes métodos de média, o primeiro método é a Micro média da Medida F, o segundo é a Macro média da Medida -F. Para esse estudo será utilizada a macro média da Medida-F. Nesse método, a Medida-F é primeiramente calculada localmente em cada classe e então faz-se o cálculo da média de todas as classes (ASCH, 2013). Essa medida pode ser obtida a partir da Equação 9:

$$F(\text{Macro média}) = \frac{\sum_{i=1}^M F_i}{M} \quad (9)$$

Onde M é o número total de classes. A macro média dá um peso igual para cada classe, independente de sua frequência, porém esta média acaba sendo mais influenciada pelo desempenho do classificador por classes que não possuem tantas amostras (ASCH, 2013).

Kappa – Para Perroca e Gaidzinski (2003) é um método estatístico que tem por objetivo avaliar o nível de concordância entre dois conjuntos de dados gerando confiabilidade e precisão dos dados classificados. Essa medida pode ser obtida a partir da Equação 10:

$$k = \frac{P_0 - P_e}{1 - P_e} = 1 - \frac{1 - P_0}{1 - P_e} \quad (10)$$

Onde P_0 é a taxa de aceitação relativa e P_e é a taxa hipotética de aceitação.

Área sob a Curva ROC - A curva ROC (*Receiver Operating Characteristic - ROC*) permite a visualização, organização e seleção da técnica com base no seu desempenho. A curva ROC é construída em um plano cartesiano em que os dados falsos positivos (eixo X) e verdadeiros positivos (eixo Y) são distribuídos em um gráfico formando uma curva ascendente. É possível encontrar um único valor da curva ROC, através do cálculo da área abaixo da curva, que é chamada de área sob a curva. Quanto maior a área sob a curva, melhor é o desempenho (PRATI et al., 2008). A Figura 5 ilustra um exemplo da curva ROC.

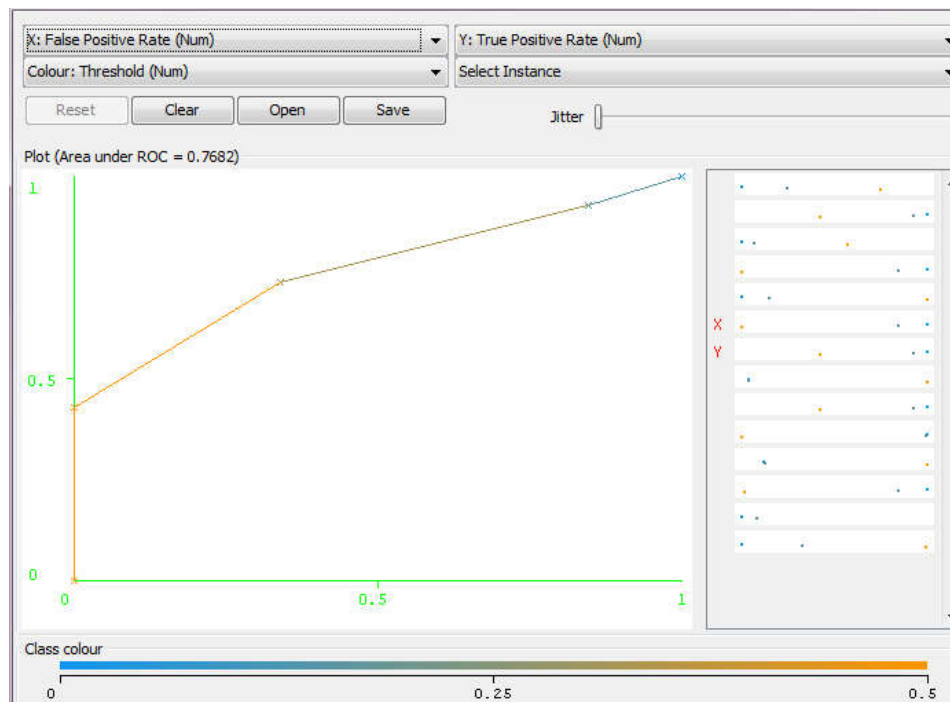


Figura 5 - Exemplo da Área sob a Curva ROC obtida junto ao software Weka.

2.5 VEÍCULOS AÉREOS NÃO TRIPULADOS - VANTS

A utilização dos VANTS - Veículos aéreos não tripulados alcançaram crescimento ao redor do mundo, impulsionados, principalmente, por avanços na tecnologia computacional, desenvolvimento de software, materiais mais leves, sistemas globais de navegação, avançados links de dados, sofisticados sensores e a miniaturização (JORGE; INAMASU; CARMO, 2011).

Para o DECEA - DEPARTAMENTO DE CONTROLE DO ESPAÇO AÉREO (2015) os VANTS são aeronaves que são projetadas para operar sem pilotos a bordo, não tem objetivos recreativos e possuem carga útil embarcada, conforme (Circular de Informações Aéreas AIC N 21/10). Destaca-se que embora o termo utilizado nessa pesquisa seja 'VANT', o termo adotado pela Organização da Aviação Civil Internacional (AOCI) é RPAS (Remotely Piloted Aircraft System).

Os VANTS podem ser classificados em dois tipos: VANTS com asas fixas e VANTS com asas rotativas. Os que possuem asas fixas são aviões convencionais pilotados remotamente e possuem uma autonomia tipicamente superior de vôo, bem como maior capacidade de carregamento de peso em virtude de sua estrutura, podendo assim cobrir uma área extensa. Os VANTS com asas rotativas geralmente assumem a forma de um disco com 4 hélices (quadricóptero), 6 (hexacóptero) ou ainda 8 (octacóptero). Esses possuem melhores capacidade de vôos em baixa altitude e em ambientes fechados devido as suas características de manobra, decolagem, pouso vertical e vôo pairado (KERROW, 2004).

Blom (2010) afirma que os VANTS surgiram em pesquisas militares da necessidade de reconhecimento do inimigo durante o período de guerra. Uma vez que o homem havia descoberto a arte do vôo, era inevitável que essa capacidade fosse utilizada na batalha. Durante a Revolução Francesa foram utilizados balões para reconhecimento do inimigo, porém devido a limitações técnicas da época, a utilização de balões logrou pouco sucesso.

De balões, passou-se hoje para uma vasta gama de máquinas tecnologicamente avançadas. Embora balões mais leves que o ar pareçam distante dos VANTS hoje utilizados, as missões realizadas ao longo da Frente Ocidental, em

1917 e 1918, são quase idênticas às realizados por VANTs no Iraque e no Afeganistão. O Corpo Aéreo do Exército dos EUA, e mais tarde a Força Aérea independente, desenvolveu a aeronave que opera sob o controle dos comandantes em solo.

O segmento evoluiu até os anos 1980. Com o surgimento da aviação, a capacidade da utilização dos VANTS foi utilizada pelo exército para uma variedade de missões. Muitos dos conceitos básicos dos VANTS surgiram na década de 1950 e 1960. Durante a Guerra do Vietnã, a Força Aérea utilizou VANTs para a coleta de dados, informações de sinais de guerra e foto/vídeo para reconhecimento (BLOM, 2010).

No uso civil há uma vasta área de aplicações possíveis para a utilização de VANTs, a exemplo de pesquisas ambientais remotas, monitoramento e certificação de poluição ambiental, gerenciamento de queimadas, segurança, monitoração de fronteira, oceanografia, agricultura e aplicações de pesca entre outras. No Brasil, iniciativas de utilização de VANTs têm sido procuradas por setores governamentais e privados. O primeiro registro de VANT em solo brasileiro refere-se ao BQM1BR, fabricado pela extinta CBT (Companhia Brasileira de Tratores), de propulsão a jato (IESA, 2013).

A partir do ano 2000, os VANTs para uso civil começaram a ganhar força no mercado. Nessa década, surgiu o Projeto Arara desenvolvido numa parceria do Instituto de Ciências Matemáticas e Computação da Universidade de São Paulo (ICMC-USP) e a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), especialmente para utilização em agricultura de precisão. O projeto deu origem, em abril de 2005, ao primeiro VANT de asa fixa desenvolvido com tecnologia 100% brasileira, cujo desenho industrial foi patenteado pela EMBRAPA (IESA, 2013).

Medeiros (2007) afirma que na Agricultura de Precisão as imagens obtidas são utilizadas principalmente para monitoramento de lavouras e estimativas de volume de produção. As fotografias aéreas obtidas a partir do VANT auxiliam no mapeamento das culturas, na avaliação de áreas cultivadas, na detecção de áreas afetadas, em cadastros rurais e no mapeamento do solo.

Herwitz et al. (2004) afirmam que a agricultura está se tornando uma indústria cada vez mais baseada no conhecimento em resposta a considerações econômicas

e ambientais. Os autores realizaram um estudo junto a Kauai Coffee Company no Hawaí, utilizando VANT para a coleta de imagens objetivando vigilância e apoio às decisões na plantação de café. Para os autores, existem vários aspectos de manejo da cultura que podem se beneficiar de observação aérea. O estudo demonstrou a capacidade de um VANT de sobrevoar a plantação, equipado com sistemas de imagem para monitorar uma região agrícola por um período de tempo prolongado, estando os VANTs a oferecer uma valiosa contribuição para o futuro monitoramento de recursos agrícola. O estudo demonstrou que a alta resolução de imagens foi útil para mapear os focos de capim-colonião, bem como para mostrar diferenças na cobertura do solo. Assim, para os autores, os VANTs desempenham um papel mais amplo, complementar ao de satélites e aviões convencionalmente testados em apoio à agricultura.

No ano de 2008, Apan et al. (2010) realizaram um estudo que investigou o uso de um veículo aéreo não tripulado (VANT) em aplicações agrícolas. A área de estudo foi localizada em Watts Bridge Memorial Airfield no sudeste de Queensland, Austrália. O objetivo do estudo foi avaliar um sistema totalmente autônomo de aquisição de imagem. Para tanto, testou-se a capacidade do piloto automático para desencadear um sistema de câmara de detecção remota, bem como avaliou-se a precisão tridimensional do piloto automático. A capacidade de adquirir imagens com precisão sobre pontos pré-determinados foi essencial para garantir a cobertura e agilizar o mosaico das imagens. A conclusão do estudo fez referência a necessidade de mais desenvolvimento para superar os problemas de precisão.

Em 2011, Primicerio et al. (2012) projetaram um VANT para realização de uma pesquisa com base em Índice de Vegetação por Diferença Normalizada (NDVI), mapeando vinhedos na Itália, justificando que os mapas podem ser um instrumento adequado para gestão viticultura, com a possibilidade de fornecimento de informações úteis aos viticultores para aumentar o potencial enológico da vinha. O fato de o VANT pesar menos de 7 kg faz com que esse seja submetido a limitações mínimas do espaço aéreo, tornando-se um sistema de sensoriamento remoto muito flexível quando se trata de pequenas culturas. Para os autores, os VANTs poderão ser considerados ferramentas comuns para uso de produtores rurais se forem de fácil utilização, automáticos e economicamente acessíveis; objetivos atendidos pelo equipamento. Todavia são necessárias melhorias no que se refere ao tempo de voo,

redução de vibrações da câmera na aquisição de imagens e faz-se necessário investigar a possibilidade de autonomia para decolagem e aterrissagem.

Peña et al. (2013) utilizaram um VANT para gerar um mapa de plantas daninhas em uma plantação de milho experimental num campo produtivo na Espanha. Para os autores, o equipamento possui potencial que não se obteve previamente com imagens aéreas convencionais ou de satélite para a concepção de tratamentos detalhados no controle de plantas daninhas específicas do local da plantação no início de pós-emergência. A utilização de VANTs associada à Análise de Imagem Baseada em Objeto, do inglês Object-Based Image Analysis (OBIA), pode contribuir com a implementação da legislação europeia que se refere a utilização sustentável dos pesticidas, promovendo reduções em aplicações de herbicidas e a utilização de doses adequadas para os níveis de infestação de plantas daninhas presentes.

Gómez-Candon et al. (2014) realizaram um estudo em dois campos de trigo na província de Sevilha, na Andaluzia ao sul da Espanha. Os campos estavam naturalmente infestados por plantas daninhas de folhas largas e gramíneas. As imagens foram coletadas por um VANTs, e posteriormente foram mosaicadas para originarem uma orto-imagem da lavoura de trigo. Os autores concluíram que um VANT voando a uma distância de 30 a 100 m de altura e com um número moderado de pontos de controle é capaz de gerar alta resolução espacial nas orto-imagens com a precisão de georeferenciamento necessária para mapear joios do trigo em um estágio fenológico inicial.

Torres-Sánchez et al. (2015) utilizaram VANTs para cálculo de características geométricas em 3 dimensões de árvores individuais e fileiras de árvores. Foram gerados modelos de superfície e utilizadas técnicas de Análise de Imagem Baseada em Objeto. As métricas foram usadas para monitorar o status de colheita, analisar a capacidade de produção da árvore e otimizar tarefas agronômicas, como o uso da água, aplicação de nutrientes, as operações de poda e controle de pragas. De acordo com os autores, essa associação no uso dessas ferramentas proporcionou confiabilidade e precisão no monitoramento agrícola das árvores sob dois padrões: a altura e o volume da copa de todas as árvores em toda a plantação. Esses dados podem ser utilizados para melhor dosagem de insumos agrícolas e otimização de operações de manejo da cultura, com implicações agro-ambientais relevantes.

Atualmente, no Brasil já se pode falar em VANTs gerando imagens importantes para a Agricultura de Precisão, bem como em softwares para análises das imagens. Medeiros (2007) propôs a criação de um VANT que pudesse ser utilizado para monitoramento e sensoriamento das atividades agrícolas. O trabalho se concentrou na construção e desempenho do protótipo, finalizando com o monitoramento e tratamento das imagens aéreas capturadas, que foi realizado através do software GIMP. O autor concluiu que as imagens geradas na pesquisa são de boa qualidade e poderiam servir de apoio a tomada de decisão do produtor.

Jesus et al. (2015) propõe a construção e uso de VANT para aquisição de imagens de plantações de soja que, em conjunto com um software de processamento de imagens, possibilite detecção de possíveis pragas e doenças, todavia o estudo está em andamento. Também nesse sentido foram verificados na literatura, aplicativos para celulares e tablets que tem por proposta o reconhecimento de doenças e pragas em áreas produtivas.

O monitoramento de safras a partir de imagens possibilita a aquisição de dados da área das lavouras, desde a fase do plantio até a fase da colheita. Tais informações são úteis para o manejo e monitoramento de safras, bem como na gestão e logística da produção, entre outros. As imagens tomadas por VANT aliadas a uma boa técnica de geoprocessamento trás resultados satisfatórios acarretando uma melhor ocupação e tratamento do solo, plantio e colheita especializados (SILVA NETO, 2013).

Honkavaara et al. (2013) afirmam que os métodos que se utilizam de VANTs possibilitam a coleta de dados eficiente, em termos de custo, com o espaço desejado e resoluções temporais. Para os autores, uma importante vantagem dessa tecnologia se refere ao fato que os dados de sensoriamento remoto podem ser obtidos ainda que sob condições de iluminação restritiva ou desfavoráveis, ou seja, sob a cobertura de nuvens, fato que torna o método operacional em uma ampla gama de aplicações.

Os VANTs na agricultura são utilizados em situações diversas como, por exemplo, para detecção e controle de invasores na plantação de milho. Uma grande vantagem desse tipo de uso, é que os VANTs podem operar a altitudes mais baixas e, portanto, a captura de imagens com uma resolução espacial muito elevada (de alguns centímetros ou milímetros), o que não seria viável a partir de vôos

convencionais ou satélites. Isso é fundamental para discriminar entre as pequenas mudas de plantas daninhas e de culturas em estágios iniciais, na maioria dos campos (PEÑA-BARRAGÁN et al., 2012).

A expectativa é que os VANTs possam fornecer ferramentas de sensoriamento remoto que sejam eficientes para a Agricultura de Precisão, atuando nos objetivos de permitir o uso eficiente de recursos, proteger o ambiente e fornecer informações relacionadas a tratamentos de gestão (utilização de máquinas para aplicações orientadas, semeadura, fertilização e proteção fitossanitária) (HONKAVAARA et al., 2013).

O fato da utilização dos VANTs ter alcançado grande crescimento, somado aos riscos envolvidos como, por exemplo, uma possível queda destes veículos sobre pessoas, prédios ou mesmo outros aviões, mostram a necessidade de regulamentação e fiscalização quando ao uso indiscriminado. No Brasil a insegurança jurídica quanto a esse aspecto tem representado uma importante barreira para a criação de um mercado consistente em torno dos VANTs. A ANAC (Agência Nacional de Aviação Civil) apresentou à sociedade, através de uma consulta pública, uma proposta de regulamentação que poderá destravar algumas barreiras, principalmente para a utilização na agricultura, que diferentemente das aplicações em áreas urbanas, apresenta riscos relativamente baixos e administráveis. Todavia a existência de um arcabouço jurídico que esclareça as regras de segurança e as responsabilidades de cada ator no caso de acidentes é muito importante.

3 METODOLOGIA

Esta tese é composta por dois artigos. O primeiro, Artigo 1, está intitulado como 'Segmentação de imagens de VANTS para detecção de ferrugem da soja', e o segundo artigo, Artigo 2, é intitulado por 'Reconhecimento de doenças de soja com base na segmentação de imagens de VANTS'. O método utilizado está detalhado na seção 'Material e Métodos' de cada um dos artigos, todavia, de forma simplificada seguem o ciclo expresso na Figura 6.



Figura 6 - Ilustração das etapas utilizadas para o desenvolvimento dos artigos

ARTIGO 1 - SEGMENTAÇÃO DE IMAGENS DE VANTS PARA DETECÇÃO DE FERRUGEM DA SOJA

Resumo: Este trabalho apresenta os resultados da avaliação de um programa de computador desenvolvido para reconhecimento da ferrugem asiática na soja. A ferramenta utiliza conceitos de visão computacional, empregando o algoritmo *Simple Linear Iterative Clustering –SLIC* para segmentação e geração de superpixels que compõem os bancos de imagens. Os VANTs multirrotor octacoptero e Phantom foram utilizados para a coleta de imagens numa plantação de soja dando origem, respectivamente, ao banco de imagens A com 1125 sub-imagens, e banco de imagens B com 702 sub-imagens, distribuídas nas classes: Palha, Solo, Folha Sã e Ferrugem. Foram extraídos dessas imagens atributos de texturas, cor e forma, que foram utilizados para medir o desempenho dos classificadores SVM, KNN, J48 e Florestas Aleatórias no ambiente Weka através das métricas: coeficientes de similaridade de Jaccard e de Yule, PCC, Medida-F, área sob a Curva ROC e Kappa. Os resultados obtidos junto aos coeficientes de similaridade demonstraram que o banco de imagens A obteve melhores resultados com o classificador SVM enquanto que o banco de Imagens B com o classificador J48. Esses classificadores foram utilizados para a geração da matriz de confusão e posterior cálculo de métricas para cada classe. A classe ‘ferrugem’ obteve melhores resultados quando avaliada através do coeficiente de Jaccard, observada no banco de imagens B, com o classificador SVM (0,953) e J48 (0,932). Entre os classificadores o melhor desempenho obtido foi verificado com o algoritmo Florestas Aleatórias através da métrica Área sob a Curva ROC em ambos os bancos de imagens. Os resultados indicam que a atuação do programa desenvolvido para reconhecimento da patologia que foi implementado e avaliado permite um bom desempenho no reconhecimento da ferrugem asiática.

Palavras-chave: doenças da soja, ferrugem asiática, superpixel, VANT

1 Introdução

A soja é uma das principais *commodities* produzidas mundialmente, faz parte do conjunto de atividades agrícolas com maior destaque e é cada vez mais importante no contexto econômico do Brasil, integrando o segmento do agronegócio que possui uma representatividade de 23% no Produto Interno Bruto – PIB nacional. É uma planta facilmente cultivada e pode proporcionar um rendimento superior a 4.000 kg por hectare. Esse potencial de rendimento é limitado por fatores diversos, entre os quais se destacam as doenças, que de forma geral, são de difícil controle (Yorinori et al., 1993; EMBRAPA, 2011). Estima-se que no Brasil as perdas anuais na produção causadas por doenças da soja variam entre 15 a 20%, sendo que algumas doenças podem ocasionar 100% de perda da safra. Cada doença pode agir de forma mais ou

menos severa, dependendo dos mecanismos de resistência ou tolerância das cultivares de soja, das estratégias de controle e manejo adotadas e da influência das condições climáticas favoráveis a ocorrências das doenças (EMBRAPA, 2013).

Em um sistema integrado de manejo da cultura da soja, o monitoramento freqüente da plantação favorece a identificação dos sintomas iniciais das principais doenças. Assim, a vistoria da lavoura deve ocorrer durante o crescimento vegetativo e principalmente antes do início do florescimento da soja, independente da ocorrência ou não de doenças (Yorinori et al., 2003). Usualmente, o monitoramento da soja é feito por amostragem e se dá de forma sistemática ou ao acaso, em que o tipo e tamanho da amostra variam conforme características das doenças ou objetivos do levantamento (Moraes, 2007). Destaca-se que, comumente, as vistorias são realizadas pelos próprios produtores e/ou técnicos, sendo uma tarefa que demanda tempo e recursos financeiros.

Os veículos aéreos não tripulados (VANTs) surgem como ferramenta de auxílio às atividades agrícolas entre as quais se destaca o monitoramento de lavouras de soja (Gómez-Candón et al., 2014; Torres-Sánchez et al., 2015), possibilitando a coleta de dados com resoluções espaciais e temporais desejadas (Honkavaara et al., 2013). Estudos como o de Apan et al. (2010) que investigou o uso de VANTs em aplicações agrícolas de café no sudeste de Queensland, na Austrália, de mapeamento de vinhedos na Itália (Primicerio et al., 2012) e de trigo em estágio fenológico inicial na província de Sevilha, ao sul da Espanha (Gómez-Candón et al., 2014) apontaram para a viabilidade do uso dessa ferramenta. Esses autores concluíram que o VANT proporcionou respostas positivas, demonstrando o grande potencial de sua utilização como auxílio a tomada de decisões na produção agrícola.

A análise de imagens aéreas obtidas por VANTs requer a utilização de programas especialistas. Para suprir esta demanda, pesquisadores vêm trabalhando no desenvolvimento de programas aplicados à resolução de problemas agrícolas, a exemplo da pesquisa de Rumpf et al. (2010) que propôs um sistema para a detecção e diferenciação de doenças na cultura da beterraba sacarina. O programa utiliza Maquinas de Vetores de Suporte (SVM) e índices espectrais de vegetação gerados a partir do processamento de imagens aéreas para classificar, com 97% de precisão, folhas sadias e doentes de beterraba. Na classificação múltipla de folhas de beterraba sadias e doentes (cercosporiose, ferrugem de beterraba sacarina e oídio) o programa alcançou uma precisão superior a 86%.

Pires et al (2016) propuseram uma abordagem para o reconhecimento da doença de soja com base em técnicas que utilizam descritores locais e histograma de palavras visuais (*Bag of Visual Words* - BOVW). Os autores trabalharam com cinco descritores locais (SURF, HOG, DSIFT, SIFT, e PHOW) e histogramas de palavras visuais (BOVW) para definir uma metodologia capaz de representar computacionalmente imagens de folhas de soja, mantendo a informação visual sobre possíveis doenças. Utilizaram, para tanto, um conjunto de imagens digitais (escaneadas) composto por 1.200 folhas (escala de cinza e colorido) obtidas em uma plantação de soja comercial. Os resultados mostraram que a abordagem proposta obteve mais de 98% de precisão, sendo eficiente na tarefa de detectar computacionalmente as doenças da soja em tempo médio de 0,1s por imagem. Os resultados demonstraram, também, que o descritor PHOW funciona melhor para espaços coloridos se comparado com a escala de cinza, atingindo taxas corretas de classificação variando de 96,25 a 99,83.

Considerando o potencial produtivo da soja no Brasil, bem como a importância da utilização de mecanismos que inovem o processo de monitoramento da fitossanidade da produção agrícola, justifica-se o desenvolvimento de um software que possibilite a detecção rápida e precoce de doenças através do processamento de imagens aéreas obtidas por VANTs, podendo incidir em economia e/ou redução de prejuízo através da automatização da detecção de doenças em plantações agrícolas (Resende et al., 2010; Medeiros, 2007; Jesus et al., 2015).

Propõe-se nesse artigo a validação do programa DIAGSOJA¹ que utiliza conceitos de visão computacional e reconhecimento de padrões. Tem-se por objetivo reconhecer doenças em plantações de soja a partir das imagens capturadas por VANTs. As imagens utilizadas na validação do programa foram obtidas em um ensaio experimental de plantio de soja, utilizando um multirrotor octacóptero e um Phantom 2.

Com o DIAGSOJA realiza-se a segmentação das imagens, utilizando o algoritmo Simple Linear Iterative Clustering – SLIC - para gerar superpixels, ou seja, para agrupar pixels em regiões coesas, com características comuns de cor, textura e

¹ O DIAGSOJA é um programa desenvolvido pelo INOVISÃO, que é um grupo de pesquisa de desenvolvimento e inovação em visão computacional e tem como principal objetivo a integração entre pesquisa, desenvolvimento e inovação para contribuir com o desenvolvimento do estado do Mato Grosso do Sul.

brilho, a exemplo da Figura 1, com os quais criou-se o banco de imagens para testes de classificadores e testes estatísticos.



Figura 1 - Exemplo de uma imagem original (à esquerda) segmentada com o extrator de superpixels Slic (à direita).

O restante deste trabalho está organizado da seguinte forma: a Seção 2 descreve os procedimentos metodológicos que regeram o artigo, como a condução do plantio experimental e captura das imagens, a proposta para reconhecimento de doenças na soja, segmentação por superpixel e construção dos bancos de imagens. A Seção 3 descreve e discute os resultados da abordagem proposta. Finalmente, a Seção 4 conclui o artigo e sugere trabalhos futuros.

2 Material e Métodos

2.1 Instalação do Campo Experimental

O ensaio foi conduzido na fazenda São José, pertencente a Universidade Católica Dom Bosco, localizada em Campo Grande-MS, Brasil, sob as coordenadas geográficas Latitude 20°24'9.88"S, Longitude 54°36'31.49"O. O campo com o plantio da soja foi instalado em uma área de 1 ha. A variedade utilizada foi a BMX Potencia RR, plantada no dia 25 de novembro de 2014, quando foram semeadas 16 sementes por metro linear de plantio. A adubação de plantio consistiu na utilização

do formulado N-P-K (02-23-23), foram aplicados 320 kg ha⁻¹. O tratamento das sementes foi feito com Standak Top® (Fipronil 25%), Comofix® (Co 1% e Mo 10%) na concentração de 2 mL kg⁻¹ de semente. O inoculante utilizado foi o Masterfix L (inoculante líquido para soja) na concentração de 1,5 mL kg⁻¹ de semente. Durante o período de observação da soja, não foi realizada aplicação de fungicidas, mesmo após detecção de focos de ferrugem asiática, única patologia verificada na plantação, ainda no estágio vegetativo da soja. A Figura 2 demonstra os procedimentos utilizados no decorrer no artigo.

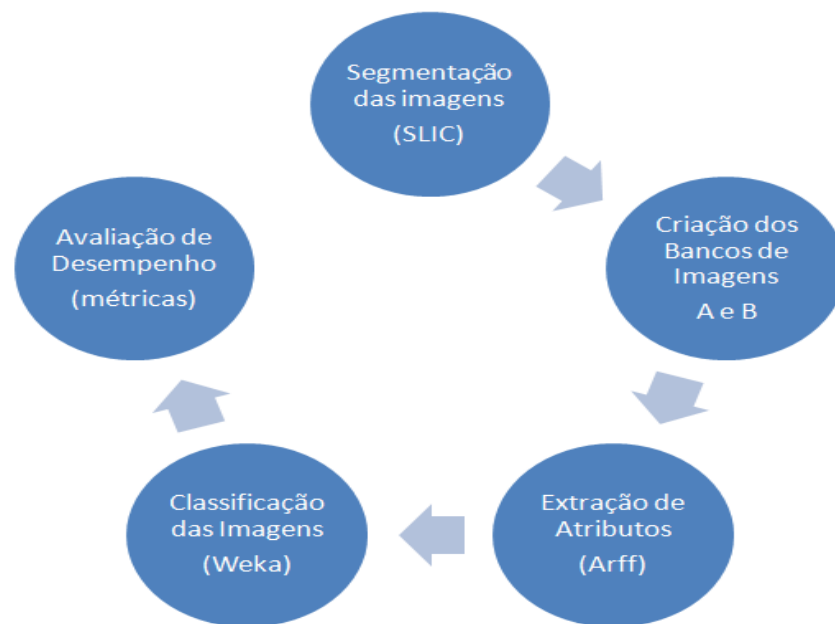


Figura 2 – Fluxograma: procedimentos metodológicos

2.2 Aquisição e Processamento das Imagens

As coletas de imagens aéreas foram realizadas ao longo do ciclo produtivo da cultura, entre os dias 01/12/2014 e 27/03/2015, através de vôos com altura média de 5 metros. Criou-se então um banco de imagens digitais de plantas de soja saudáveis e doentes para teste do software. As imagens foram capturadas por dois VANTs, conforme visualizado na Figura 3:



Figura 3 – VANTs utilizados na captura de imagens

- VANT multirotor, octacoptero: Tendo em torno 12kg, com duração de 5 a 8 minutos de voo por bateria, câmera Canon 70d, lente Canon EF-S 10-18mm IS STM, com gimbal (suporte de 3 eixos que permitem a rotação e estabilização da imagem) e controladora de voo DJI A2. A lente utilizada nas coletas das imagens possui 18 mm, com uma resolução de 72dpi, a uma altura de 10 e 5 metros.
- VANT DJI Phantom com uma câmera GoPro HERO3 Black Edition acoplada, com uma resolução de vídeo de 1440p e resolução de foto de 12 megapixels.

Entre as imagens capturadas com o primeiro equipamento, foi gerado o banco de imagens intitulado aboveR3C4, e a partir do segundo equipamento gerou-se o banco de imagens phantomR6C4.

2.2.1 Proposta para reconhecimento de doenças na soja

O desenvolvimento do software teve como base conceitual a visão computacional a partir da utilização de técnicas computacionais (Szeliski, 2010). Num sistema de visão computacional a aquisição de imagem é o primeiro passo, dando-se a partir de sensores de câmeras, a exemplo das utilizadas nos VANTs. O próximo passo é o pré-processamento onde aplica-se métodos que facilitam a identificação de um objeto para posteriormente extrair atributos. Na etapa da segmentação o objetivo é destacar regiões relevantes da imagem para um processamento posterior, e o último passo refere-se ao processamento de alto nível

que é um processo que inclui estimativa de parâmetros sobre a imagem e classificação dos objetos obtidos em diferentes categorias (Milano; Honorato, 2010). A classificação é obtida com a utilização de algoritmos de reconhecimento de padrões.

Para o reconhecimento automático dos sintomas de doenças na soja, optou-se por segmentar as imagens em regiões que compartilham certas propriedades comuns, tais como brilho, textura e forma. Para tanto, utilizou-se o programa DIAGSOJA, criado pelo grupo de pesquisa INOVISÃO, da UCDB, e desenvolvido com o auxílio do Play Framework 2.3.7, em linguagem de programação Java, e concebido para a criação de partições com o algoritmo SLIC (*Simple Linear Iterative Clustering*) que ao agrupar vários pixels da imagem em uma única região (superpixel) reduz o custo computacional do agrupamento (Achanta et. al., 2010).

Para extrair atributos de texturas foram realizados cálculos estatísticos (correlação e entropia) com a matriz de co-ocorrência (Haralick; Shanmugam; Dinstein, 1973; Schwartz; Pedrini, 2003). Ainda na descrição da textura, utilizou-se também de Padrões Binários Locais (Local BinaryPattern - LBP) proposto por Ojala; Pietikäinen e Mäenpää (2002). A extração de atributos de cor ocorre através de cálculos de média, desvio padrão, valor máximo e mínimo dos canais dos espaços de cores: RGB, HSV e CieLab (Pedrini, 2014; Gonzalez; Woods, 2010). Também foram extraídos histograma de gradiente nas imagens em níveis de cinza, bem como atributos de forma (formfactor, roundness), que extraem a relação da área do objeto dividida por uma circunferência em torno (Dalal; Triggs, 2005). A máquina utilizada para o desenvolvimento do trabalho foi um Notebook da marca Samsung NP-RV411-AD3 c/ Intel Core i3, 3GB.

2.2.2 Segmentação por Superpixel

O superpixel tem sido cada vez mais utilizado em estudos relacionados à visão computacional. O superpixel foi desenvolvido por Ren e Malik (2003), com base nos princípios da Gestalt, e refere-se ao desenvolvimento de um classificador que combina características com contorno, textura, brilho e continuidade. Ao longo de sua utilização vários algoritmos de superpixels foram propostos e assim, a escolha de um algoritmo de superpixel depende da aplicação a qual se destina, todavia os

seguintes requisitos são desejáveis: os superpixels devem respeitar os limites do objeto; devem ser gerados de forma tão eficiente quanto possível, e, quando usado para reduzir a complexidade computacional como uma etapa de pré-processamento, deve ser rápido para computar, eficiente no uso da memória do computador e simples de usar (Achanta et al, 2010).

Nesse artigo o algoritmo utilizado para geração de superpixel foi o SLIC (*Simple Linear Iterative Clustering*) que foi introduzido por Achanta et al. (2010) e é uma adaptação do método de agrupamento *k-means* para geração de superpixels. O SLIC agrupa pixels no espaço 5-D definido por L, a, b (valores do espaço de cor CIELAB) e as coordenadas x e y dos pixels (Achanta et. al., 2010; Lv, 2015). Com a definição do parâmetro *k*, correspondente a quantidade de superpixels, realiza-se o particionamento da imagem em regiões retangulares, levando cada superpixel a ter aproximadamente $\frac{N}{k}$ pixels, onde N é o número de pixels da imagem. O processo do agrupamento começa com a etapa de inicialização, na qual os k centros dos agrupamentos $C = [l_i a_i b_i x_i y_i]$, são em cada pixel i, expressos por uma matriz regularmente espaçada em S, conforme Equação (1). Para que os superpixels tenham aproximadamente o mesmo tamanho, o intervalo da matriz é determinado pela seguinte equação:

$$S = \sqrt{\frac{Largura * Altura}{k}} \quad (1)$$

Assim, os centros do superpixel são movidos para locais com baixa magnitude de gradiente, numa vizinhança 3x3, para evitar que um superpixel tenha seu centróide colocado sobre regiões de borda e para reduzir as chances dele conter pixels ruidosos. No passo de atribuição, cada pixel i é associado com o centro mais próximo do agrupamento, cuja região de busca se sobrepõe à sua localização, esta é a chave para acelerar esse algoritmo, pois ao limitar o tamanho da região da busca, reduz-se significativamente o número de cálculos de distância, o que resulta em uma vantagem de velocidade significativa sobre o agrupamento K-means convencional (Achanta et al., 2010). O processo anteriormente descrito só é possível através da introdução da medida de distância D, a qual determina o centro mais próximo para cada pixel:

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \quad (2)$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \quad (3)$$

$$D_s = d_{lab} + \frac{m}{S} * d_{xy} \quad (4)$$

onde D é a soma da distância d_{lab} e a distância d_{xy} normalizada pelo intervalo S . A variável m corresponde ao controle de compactação do superpixel, quanto maior o seu valor, mais a proximidade espacial é enfatizada e mais compactado é o agrupamento. Este procedimento é repetido até a convergência ou até um número máximo de iterações T . Na etapa de pós-processamento, os superpixels não representam necessariamente componentes ligados, de tal forma que o algoritmo precisa reforçar a conectividade através da re-atribuição de pixels disjuntos para superpixels próximos (Achanta et al., 2010).

2.3 Procedimento de validação

No programa DIAGSOJA procedeu-se a segmentação das imagens, utilizando o algoritmo SLIC para geração de superpixels, utilizando para tanto 5 iterações e 40 segmentos. A partir da segmentação de imagens, foram gerados superpixels que compuseram dois bancos de imagens, pois as duas aeronaves geraram imagens com padrões diferentes entre si. Destaca-se que nesse experimento, apenas uma doença afetou a plantação utilizada para o estudo: ferrugem asiática. Assim, em ambos os bancos, as imagens foram distribuídas nas classes: Ferrugem Asiática, Folha Sã, Palha e Solo. Para o banco de imagens aboveR3C4 foram gerados 1125 superpixels, e para as imagens obtidas com o phantomR6C4 foram obtidos 702 superpixels. Com esses superpixels obtiveram-se informações fornecidas pelos atributos das imagens de cada classe, em ambos os bancos de imagens.

A partir de então, utilizou-se de conhecimentos acerca da aprendizagem automática, especificamente da aprendizagem supervisionada. A aprendizagem automática trabalha com os métodos computacionais que torna possível que computadores não somente aprendam, mas que também melhorem o seu desempenho com a experiência (Capela; Georgieva, 2012), e a utilização da aprendizagem supervisionada ocorre quando as informações das classes são previamente conhecidas (Lorena; Carvalho, 2007). Empregou-se então, alguns métodos típicos de classificação (SVM, KNN, J48 e Florestas Aleatórias) com os quais foram realizados os testes no ambiente Weka versão 3.7. Foi utilizada a

validação cruzada (Santos et al., 2009) que distribuiu o conjunto de dados em 10 partições, onde 9 partes foram utilizadas para treinamento dos classificadores e uma parte foi utilizada para validação.

Para verificar qual classificador obteve melhor desempenho utilizou-se dos coeficientes de similaridade: coeficientes de Jaccard (Lara, 2007) e de Yule (Meyer, 2002), onde os resultados com valores próximos a 0 são considerados de pior resultado e valores próximos a 1 são considerados de melhor resultado. Com os resultados obtidos junto aos coeficientes de similaridade, foram geradas as matrizes de confusão para os classificadores com melhor desempenho para ambos os bancos de imagens (aboveR3C4 e phantomR6C4). Os valores exibidos nas matrizes permitiram calcular os coeficientes de similaridade para cada classe abordada, verificando assim em qual classe o programa obteve melhor desempenho. Para avaliar os classificadores foram utilizadas as seguintes métricas: PCC (Lara, 2007), Medida-F (Asch, 2013), área sob a Curva ROC (Prati et al., 2008) e Análise de concordância Kappa (Perroca; Gaidzinski, 2003). Para verificar se houve diferença significativa entre os classificadores foi aplicada a Análise de Variância (Anjos, 2009).

3 Resultados e Discussões

O fungo *Phakopsora pachyrhizi*, responsável por provocar a ferrugem asiática na soja, foi verificado no ensaio experimental ainda no estágio vegetativo da planta, em período que foi caracterizado por incidência de muita chuva na região. A doença iniciou-se na parte baixa da planta, todavia não foi realizada aplicação de defensivos, mesmo após detecção de focos de ferrugem asiática, pois o objetivo foi obter imagens da patologia para testes com o programa de computador.

As primeiras características da doença percebidas durante o monitoramento do ensaio experimental e coleta de imagens foi a visualização de lesões na cor verde claro. Posteriormente a coloração alterou-se, evoluindo para a cor amarelo ferrugem, havendo também alterações no tamanho das lesões na face inferior dos folíolos, conforme progresso da doença demonstrado na Figura 4.



Estádio R1

Estádio R5

Figura 4 – Evolução dos sintomas da Ferrugem asiática no ensaio experimental

Após a segmentação das imagens através do DIAGSOJA, procedeu-se as classificações dos dois bancos de imagens (aboveR3C4 – banco de imagens A, e phantomR6C4 – banco de imagens B), utilizando os algoritmos de classificação SVM, KNN, J48 e Florestas Aleatórias (FA), e a técnica de amostragem validação cruzada com 10 dobras. Os resultados das classificações foram avaliados através dos coeficientes de similaridade de Jaccard e de Yule. Verificou-se que, tanto o coeficiente de Jaccard como o coeficiente de Yule, indicaram que as imagens do banco de imagens A (aboveR3C4) apresentaram melhores resultados com o classificador SVM, e as imagens do banco de imagens B (phantomR6C4), com o classificador J48, conforme Tabela 1.

Tabela 1 - Avaliação dos classificadores considerando os índices de similaridade.

Métricas	SVM		KNN		J48		FA	
	A	B	A	B	A	B	A	B
Coeficiente de Jaccard	0,990	0,842	0,935	0,624	0,971	0,861	0,971	0,780
Coeficiente de Yule	0,990	0,857	0,935	0,712	0,970	0,873	0,971	0,820

A Figura 5 apresenta as matrizes de confusão para os bancos de imagens A e B, para as quais foram utilizados, respectivamente, os classificadores SVM e J48, pois obtiveram melhores desempenhos com ambos os coeficientes, objetivando avaliar qual classe - ferrugem asiática, folha sã, palha ou solo- foi melhor classificada (Prina; Trentin, 2015), onde as imagens em tom vermelho representam os valores mais altos. Destaca-se que quanto mais próximos os valores das diagonais principais da matriz de confusão sejam do valor total de imagens contidas na classe, menor foi a confusão entre as classes.

A-SVM	Folha Sã	Ferrugem	Solo	Palha
Folha Sã	699	0	0	1
Ferrugem	4	37	0	1
Solo	1	0	104	11
Palha	1	0	16	250

A-J48	Folha Sã	Ferrugem	Solo	Palha
Folha Sã	691	7	0	2
Ferrugem	7	27	4	4
Solo	0	2	95	19
Palha	1	2	20	244

B-SVM	Folha Sã	Ferrugem	Solo	Palha
Folha Sã	36	3	0	2
Ferrugem	5	97	0	13
Solo	0	2	46	1
Palha	0	6	2	489

B-J48	Folha Sã	Ferrugem	Solo	Palha
Folha Sã	36	3	1	1
Ferrugem	6	92	4	13
Solo	1	0	47	1
Palha	2	17	1	477

Figura 5 - Matriz de confusão - Banco de imagens A e B, Classificadores SVM e J48

Utilizando uma matriz de confusão, Rezende (2003) demonstra como calcular medidas quantitativas através da classificação de pixels, considerando problemas em que a classificação envolve apenas duas classes. Conforme observado na Figura 5, este artigo trabalha com 4 classes e, assim, para calcular as medidas quantitativas, a classe de interesse foi considerada como a classe positiva, enquanto todas as outras foram consideradas como negativas (LANTZ, 2015).

A partir das matrizes, foram calculados os coeficientes de Jaccard e Yule para cada classe - ferrugem asiática, folha sã, palha e solo, em cada banco de imagens, conforme resultados apresentados na Tabela 2. Os resultados fornecidos através do coeficiente de Jaccard indicaram que tanto o classificador SVM quanto o J48 obtiveram melhores resultados com a classe Palha para o banco de imagens A, e com a classe Ferrugem, para o banco de imagens B. Com os resultados fornecidos através do coeficiente de Yule verificou-se que para o banco de imagens A, os melhores resultados foram com a classe Solo (classificador SVM), e com a classe Palha (classificador J48), enquanto que com o banco de imagens B, o melhor resultado foi obtido na classe Folha Sã, tanto com o classificador SVM quanto com o J48.

Tabela 2 - Classe melhor classificada considerando os coeficientes de Jaccard e Yule.

Classes	Coeficiente de Jaccard				Coeficiente de Yule			
	SVM		J48		SVM		J48	
	A	B	A	B	A	B	A	B
Palha	0,990	0,783	0,976	0,720	0,989	0,870	0,967	0,792
Solo	0,881	0,770	0,509	0,681	0,995	0,868	0,697	0,782
Folha Sã	0,788	0,902	0,679	0,855	0,855	0,954	0,777	0,884
Ferrugem	0,893	0,953	0,836	0,932	0,931	0,928	0,880	0,874

O desempenho de cada classificador foi testado através das métricas PCC, Medida-F, área sob a Curva ROC e Análise de concordância Kappa, e o resultado está demonstrado na Tabela 3. Com a métrica PCC, verificou-se que a maior porcentagem de imagens corretamente classificadas como parte das classes ocorreu com o classificador SVM no que se refere ao banco de imagens A e, com o classificador Florestas Aleatórias (FA) no banco de imagens B.

Na utilização da métrica Medida-F, optou-se pelo uso da macro Medida-F, por proporcionar um peso igual a cada classe (Asch, 2013). Assim, verificou-se que com o banco de imagens A, o melhor desempenho foi obtido com o classificador SVM, enquanto que para o banco de imagens B, os classificadores SVM e Florestas Aleatórias obtiveram igual desempenho.

Tabela 9 - Desempenho dos classificadores com o DIAGSOJA

Métricas	Banco de Imagens	Classificadores			
		SVM	KNN	J48	FA
PCC	A	97,06±1,29	91,17±2,30	94,12±1,90	95,87±1,90
	B	95,07±2,43	91,51±2,93	94,53±2,94	95,38±2,06
Medida-F	A	97,00±0,01	91,00±0,02	94,00±0,02	95,00±0,02
	B	95,00±0,02	91,00±0,03	94,00±0,03	95,00±0,02
Área sob a Curva ROC	A	99,00±0,01	97,00±0,02	99,00±0,01	100±0,00
	B	97,00±0,05	81,00±0,11	94,00±0,07	100±0,01
Kappa	A	95,00±0,02	84,00±0,04	89,00±0,03	92,00±0,03
	B	89,00±0,05	82,00±0,06	88,00±0,06	89,00±0,05

As Curvas de Características de Operação do Receptor (Curvas ROC Receiver Operating Characteristic) é uma representação gráfica da taxa de verdadeiros positivos versus taxa de falsos positivos. O desempenho entre os dois bancos de imagens, avaliado através Área sob a Curva ROC foi o mesmo utilizando o classificador Florestas Aleatórias. Na avaliação com a métrica Kappa verificou-se que tanto no banco de imagens A quanto no B a concordância geral é estatisticamente diferente de zero e superior a 80, ou seja, a concordância é quase perfeita.

Para verificar se houve diferença significativa entre os classificadores foi aplicada a Análise de Variância para os dois bancos de imagens. Considerando os resultados valor-p=0,31542 e 0,33745, a um nível de significância de 0,05%, pode-

se afirmar que não existem evidências estatísticas de que os desempenhos sejam diferentes entre os classificadores.

Automatizar tanto quanto possível o processo de reconhecimento de doenças é uma necessidade para os produtores já que, a depender da patologia que atinge a plantação, uma demora no diagnóstico pode comprometer o ganho do produtor (AFRIDI *et al.*, 2014).

O emprego de aprendizagem automática e visão computacional como base para programas de reconhecimento tem se mostrado de grande valia em tarefas de mapeamento de plantação, bem como na identificação de focos de pragas ou doenças (UGALE; GUPTA, 2016). Nesse sentido o experimento demonstrou que é possível detectar sintomas de doenças foliares em plantações de soja utilizando imagens obtidas por VANTs e, os testes realizados utilizando o programa de computador para reconhecimento de doenças demonstraram realizar distinção entre as classes conforme os resultados estatísticos já apresentados.

4 Conclusão

Considerando a importância da soja, bem como o potencial produtivo que não é alcançado devido, entre outros fatores, ao impacto de doenças, verificou-se com essa pesquisa que a utilização de um sensor para aquisição de imagens, acoplado aos VANTs, permite agilidade na aquisição de imagens e monitoramento de lavouras produtivas, mais a atuação do programa desenvolvido para reconhecimento das patologias que foi implementado e avaliados por meio de várias métricas, permite um bom desempenho no reconhecimento da ferrugem asiática. Os resultados obtidos junto aos classificadores demonstraram que a proposta, bem como sua utilização, pode atuar como nova alternativa para o reconhecimento dos sintomas da doença nas folhas de soja.

O software encontra-se em fase de teste e deve contar com ajustes. Outros fatores poderão proporcionar melhores resultados, como testes com os parâmetros a serem estabelecidos pelo usuário na geração de arquivo para análise. Além disso, é possível que a utilização de um banco de imagens mais robusto possa proporcionar melhores desempenhos.

Outro fator limitante para a análise foi o surgimento de apenas uma doença no experimento, quando doenças com características diferentes são relevantes para avaliar o desempenho do programa. Além disso, o fato da doença ter como uma de suas características seu início na parte baixa da planta pode dificultar o reconhecimento da doença no seu estágio inicial. Nesse sentido, a busca por uma forma de obtenção de imagens da parte baixa da planta pode favorecer o reconhecimento dos sintomas da doenças no estágio inicial, onde é possível ao produtor entrar com o uso de defensivos e reduzir o prejuízo.

Referências Bibliográficas

ACHANTA, R.; SMITH, K.; LUCCHI, A.; FUA, P.; and SUSSTRUNK, S.. Slic superpixels. Technical report, EPFL, Tech.Rep. 149300, 2010.

AFRIDI, M. J.; LIU, X.; MITCHELL MCGRATH, J.. "An Automated System for Plant-level Disease Rating in Real Fields," in Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014), Stockholm, Sweden, August 24-28, 2014.

ANJOS, A. dos.. Análise de Variância.. Capítulo 7, 2009.
<http://www.est.ufpr.br/ce003/material/cap7.pdf>

APAN, A.; TROY, J.; LES, C. Z..The use of an unmanned aerial vehicle as a remote sensing platform in agriculture. Australian. Journal of Multi-disciplinary Engineering. Nov. 2010.

ASCH, V. V.. "Macro- and micro-averaged evaluation measures," Tech. Rep., University of Antwerp, 2013.

CAPELA NF, GEORGIEVA P. Aplicação de técnicas de aprendizagem automática para classificação de emoções humanas com sinais de EEG. Electrónica e Telecomunicações, vol. 5, nº 4, dezembro 2012

DALAL, N.; TRIGGS, B.. Histograms of Oriented Gradients for Human Detection. In CVPR, pages 886-893, 2005

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA SOJA. Tecnologias de produção de soja – região central do Brasil 2012 e 2013. - Londrina: Embrapa Soja, 2011.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA SOJA. Tecnologias de produção de soja – região central do Brasil 2014. - Londrina: Embrapa Soja, 2013.

GÓMEZ-CANDÓN, D.; CASTRO, A.I.D.; LÓPEZ-GRANADOS, F. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precis. Agric.* 2014, 15, 44–56.

GONZALEZ, R. C.; WOODS, R. E. (2010). *Processamento Digital de Imagens*. 3.ed. São Paulo: Pearson.

HARALICK, R.M.; SHANMUGAM, K.; DINSTEN, I. "Textural Features for Image Classification", *IEEE Trans. on Systems, Man and Cybernetics*, Vol. SMC-3, pp. 610-621, 1973.

HONKAVAARA, E.; SAARI, H.; KAIVOSOJA, J.; PÖLÖNEN, I.; HAKALA, T.; LITKEY, P.; MÄKYNEN, J.; PESONEN, L. Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture. *Remote Sens.* 2013, 5, 5006-5039.

JESUS, L.F.; SILVA, V.B.; ROCHA, F.G.. Uso de software para detecção de doenças na cultura da soja com o auxílio de um drone autônomo. *Computer on the Beach 2015 - Resumo Estendido*. Página 429. Disponível em <<https://danielapezzini.files.wordpress.com/2015/04/projeto.pdf>>. Acesso em 05/08/2015.

LANTZ, B.. *Machine Learning with R*. Packt, 2º Edição, 2015.

LARA, A. C.. Segmentação de movimento usando morfologia matemática. *Dissertação*. Universidade de São Paulo, 2007

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. *RITA – Revista de Informática Teórica*, v. 14, n. 2, p. 43-67, 2007

LV, J.. An Improved SLIC Superpixels using Reciprocal Nearest Neighbor Clustering. *International Journal of Signal Processing, Image Processing and Pattern Recognition* Vol. 8, No. 5 (2015), pp. 239-248. Disponível em <<http://dx.doi.org/10.14257/ijsp.2015.8.5.25>> Acesso em 20 de jul. de 2015.

MEDEIROS, F.A.. Desenvolvimento de um veículo aéreo não tripulado para aplicação em Agricultura de Precisão. *Dissertação*. Universidade Federal de Santa Maria. Centro de Ciências Rurais. Programa de Pós Graduação em Engenharia Agrícola. Santa Maria-RS. 2007. Disponível em http://cascavel.cpd.ufsm.br/tede/tde_busca/arquivo.php?codArquivo=878. Acesso em 17/07/2015

MEYER A. S.. Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes. Dissertação de Mestrado. ESALQ, Universidade de São Paulo. Piracicaba SP. 2002.

MILANO, D. D.; HONORATO, L. B. Visão computacional. – Universidade Estadual de Campinas– Faculdade de Tecnologia. 2010.

MORAES, S.A. de Quantificação de doenças de plantas. 2007. Artigo em Hypertexto. Disponível em: <http://www.infobibos.com/Artigos/2007_1/doencas/index.htm>. Acesso em: 22/9/2016.

OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(7), p. 971-987, 2002.

PEDRINI, H.. Processamento e Análise de Imagens. Instituto de Computação UNICAMP. 2014. Disponível em <http://www.ic.unicamp.br/~helio/disciplinas/MO445/aula_cores.pdf> Acesso em Ago/2015.

PERROCA, M.G.; GAIDZINKI, R. R.. Assessing the interrater reliability of an instrument for classifying patients - kappa quotient. *Rev Esc Enferm USP*. [Internet]. 2003. [acesso 23 abr 2013];37(1):72-80. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342003000100009. Acesso em Ago/2015.

PIRES, R. D. L. et al.. Local descriptors for soybean disease recognition. *Computers and Electronics in Agriculture*. Volume 125, July 2016, Pages 48–55.

PRATI, R.C.; BATISTA, G.E.A.P.A.; MONARD, M.C. (2008). Curvas ROC para avaliação de classificadores. *IEEE América Latina*, Vol. 6, N. 2.

PRIMICERIO, J., Di GENNARO, S. F., FIORILLO, E., GENESIO, L., LUGATO, E., MATESE, A. and VACCARI, F. P. 2012. A flexible unmanned aerial vehicle for precision agriculture. *Precision Agriculture* 13 517–523.

PRINA, B. Z.; TRENTIN, R.. GMC: Geração de Matriz de Confusão a partir de uma classificação digital de imagem do ArcGIS®. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, João Pessoa-PB, Brasil, 25 a 29 de abril de 2015, INPE.

REN, X.; MALIK, J.. Learning a classification model for segmentation. *IEEE ICCV*, pp. 10–17, 2003.

RESENDE et al., Agricultura de Precisão no Brasil: Avanços, Dificuldades e Impactos no Manejo e Conservação do Solo, Segurança Alimentar e Sustentabilidade. XVIII Reunião Brasileira de Manejo e Conservação do Solo e da

Água, Agosto de 2010, Teresina, Piauí. Disponível em: <http://www.alice.cnptia.embrapa.br/bitstream/doc/870646/1/Agriculturaprecisao.pdf>. Acesso em 15 setembro de 2016.

REZENDE, S. O.. Sistemas inteligentes: fundamentos e aplicações. Barueri, SP: Manole, 2003.

RUMPF, T.; MAHLEIN, A.K.; STEINER, U.; OERKE, E.C.; DEHNE, H.W.; PLÜMER, L.. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Comput. Electron. Agric.*, 74 (1) (2010), pp. 91–99.

SANTOS, L. D. M. dos; MIKAMI, R.; VENDRAMIN, A. C. B. K.; KAESTNER, C. A. A.. Procedimentos de Validação Cruzada em Mineração de Dados para ambiente de Computação Paralela. ERAD 2009 — Caxias do Sul, 17 a 20 de março de 2009.

SCHWARTZ, W.R.; PEDRINI, H.. Método para classificação de imagens baseada em matrizes de coocorrência utilizando características de textura. III Colóquio Brasileiro de Ciências Geodésicas, Curitiba-PR, Brasil, 2003.

SZELISKI, R.. *Computer Vision: Algorithms and Applications*. Springer, 2010.

TORRES-SÁNCHEZ, J., LÓPEZ-GRANADOS, F., PEÑA, J. M. (2015). An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. *Computers and Electronics in Agriculture*, 114, 43-52.

UGALE, V.; GUPTA, D.. A Comprehensive Survey on Agricultural Image Processing. *International Journal of Science and Research (IJSR)*. Volume 5 Issue 1, January 2016.

YORINORI, J. T.; CHARCHAR, M.J.D'A.; NASSER, L.C.B.; HENNING, A.A...Doenças da soja e seu controle. In: ARANTES, N.E.; SOUZA, P.I. de M. de., ed. *Cultura de soja nos cerrados*. Piracicaba: POTAFOS, 1993. p.333-397.

YORINORI, J. T.; GODOY, C. V.; PAIVA, W. M.; FREDERICK, R. D.; COSTAMILAN, L. N.; BERTAGNOLLI, P. F.; NUNES JUNIOR, J.. Evolução da ferrugem da soja (*Phakopsora pachyrhizi*) no Brasil, de 2001 a 2003. *Fitopatologia Brasileira*, Brasília, v. 28, p. S210, ago. 2003.

ARTIGO 2 - RECONHECIMENTO DE DOENÇAS DE SOJA COM BASE NA SEGMENTAÇÃO DE IMAGENS DE VANTS

Resumo: A soja é um produto importante para a economia brasileira, todavia possui fatores que podem limitar seu rendimento produtivo, a exemplo das doenças que são geralmente de difícil controle. Assim, este artigo objetiva a utilização de um programa de computador para o reconhecimento de doenças em imagens obtidas por um VANT em uma plantação de soja. O programa tem como base a visão computacional e aprendizagem de máquina, utilizando o algoritmo SLIC para a segmentação das imagens em superpixels. Para alcançar o objetivo, após a segmentação das imagens, foi criado um banco de imagens com as seguintes classes: míldio, mancha alvo, ferrugem asiática, solo, palha e folha saudável, totalizando 22.140 imagens. Foram utilizadas escalas diagramáticas para realização de avaliações da severidade das doenças. O programa de computador para reconhecimento de doenças explorou quatro técnicas de aprendizagem supervisionadas: SVM, J48, Floresta Aleatória e KNN. Os resultados da classificação realizada pelo programa proposto a partir do banco de imagens foram comparados com os resultados da classificação obtidos através da classificação Software x Especialista. As técnicas que obtiveram melhor desempenho foram o SVM e Florestas Aleatórias, levando em consideração os resultados obtidos com todas as métricas de avaliação utilizadas. Verificou-se que o programa é eficiente para diferenciar as classes de doenças tratadas nesse artigo.

Palavras – chave: Doenças da soja, Segmentação, VANTS.

1 Introdução

A agropecuária brasileira é uma das principais bases econômicas do país. O índice de produtividade brasileira ampliou-se com a realização de investimentos em ciência e tecnologia, colocando o Brasil entre os maiores produtores mundiais de alimentos, fibras e energias renováveis (EMBRAPA, 2014). Em se tratando especificamente da produção brasileira de grãos, a safra 2015/2016 foi de 186.3 milhões de toneladas e a estimativa para a Safra 2016/17 é que a produção de grãos situe-se entre 210,5 e 214,8 milhões de toneladas, demonstrando o potencial produtivo (CONAB, 2016).

A soja se destaca pela série de produtos e subprodutos que são derivados da sua cadeia produtiva, demonstrando sua relevância para o agronegócio brasileiro. Todavia, na última safra, aspectos climáticos comprometeram lavouras em diversos estados produtores no Brasil, ocasionando impactos na produtividade média, estabelecendo um patamar inferior à safra 2014/2015 (CONAB, 2016).

Para a obtenção de níveis de produtividade satisfatórios é necessário realizar um bom manejo de pragas e doenças, sendo essas últimas de fundamental importância devido ao seu grande impacto negativo na cultura da soja. Dentre as doenças de maior destaque nas regiões produtoras, como o estado de Mato Grosso do Sul (5º maior produtor brasileiro na safra 2015/2016), pode-se citar a ferrugem asiática (*Phakopsora pachyrhizi*), a antracnose (*Colletotrichum truncatum*) e a mancha alva (*Corynespora cassiicola*) (GRIGOLLI, 2015).

Para o manejo adequado das doenças na cultura da soja deve-se realizar inspeções diárias nas lavouras para identificar sinais do ataque de doenças. O monitoramento deve ser realizado por profissionais treinados, entretanto a avaliação, mesmo obedecendo a padrões (escalas diagramáticas), pode ser subjetiva já que pode haver variação entre a percepção de um profissional para outro profissional. Essa variação pode ser fundamental na tomada de decisão no controle preventivo das doenças (AFRIDI *et al.*, 2014).

Bonaldo, Riedo e Lima (2009) afirmam que alguns fatores podem dificultar a correta identificação das doenças foliares da soja, destacando alguns sintomas comuns a várias doenças em sua fase inicial. Além disso, a correta identificação é dificultada por monitoramentos superficiais, que deixam passar alguns sintomas das patologias despercebidas quando feitas a olho nu, já que a evolução de alguns sintomas não são perceptíveis ao olho humano, e a visualização com o auxílio de uma lupa de 10 a 20 aumentos pode significar um melhor auxílio a observação.

Afridi *et al.* (2014) destacam que o monitoramento realizado visualmente, por humano, pode não ser adequado, já que duas avaliações, realizadas por pessoas diferentes, podem ter resultados diferentes considerando a percepção de cada um quanto à patologia e sua severidade. Além disso, essa técnica de monitoramento pode causar enfado ao observador.

Assim, novas tecnologias são propostas objetivando facilitar a identificação e a tomada de decisão quanto ao controle de doenças em grandes culturas assim como a soja. O uso de veículos aéreos não tripulados - VANTs tem se destacado como uma nova tecnologia que pode ser empregada para o monitoramento agrícola a exemplo de Jia *et al.* (2016) que propuseram o uso do VANT para monitorar o crescimento do arroz através de imagens multiespectrais. Hunt *et al.* (2016) afirmam que o potencial uso de VANTs em aplicações agrícolas se deve principalmente a

possibilidade do voo em baixa altitude que permite obtenção de imagens de alta resolução espacial.

A visão computacional tem se mostrado como uma importante aliada para a análise das imagens obtidas através dos VANTs, sendo empregada em programas de computador para mapear e identificar, por exemplo, focos de pragas ou doenças em cultivos agrícolas (PEÑA-BARRAGÁN *et al.*, 2012 ; GÓMEZ-CANDON *et al.*, 2014; UGALE; GUPTA, 2016). Este trabalho teve como objetivo a utilização de um software que emprega um sistema de visão computacional para automatizar a detecção de doenças na cultura da soja.

Além disso, essa pesquisa contribui academicamente pela construção de banco de imagens anotadas contendo 22.140 imagens distribuídas entre as seguintes classes: míldio, macha alvo, ferrugem asiática, folhas saudáveis, solo e palhas. As imagens foram capturadas num experimento de plantio de soja utilizando um Phantom 3 Professional e foram segmentadas com a utilização do algoritmo SLIC para geração de superpixels. O Banco de imagens será disponibilizado para colaborar com o desenvolvimento de novos sistemas de visão computacional.

A classificação e detecção de doenças são realizadas por meio de avaliação visual humana, que pode ocorrer com o uso da escala diagramática para avaliação de severidade da patologia na soja, por amostragens. Assim, avaliações visuais humanas para detecção do grau de severidade das doenças da soja foram realizadas durante o experimento e essa informação foi utilizada para aferir o desempenho do software. Além disso, foram implementados extratores de cor, forma e textura e, quatro técnicas de aprendizado de máquina supervisionado foram exploradas, a fim de construir um sistema de visão computacional que pode classificar patologias em imagens de produção da soja.

2 Material e Métodos

A condução do experimento passou por procedimentos necessários para o alcance dos objetivos propostos. Cada uma das etapas do desenvolvimento da pesquisa, pode ser verificada nas seções subseqüentes.

2.1 Instalação da lavoura de soja

O ensaio foi conduzido na fazenda São José, pertencente à Universidade Católica Dom Bosco (UCDB), localizada em Campo Grande, MS, Brasil, sob as coordenadas geográficas: Latitude 20°24'9.88"S, Longitude 54°36'31.49"O. O campo com o plantio da soja foi instalado em uma área de um hectare. A cultivar utilizada foi BMX Potencia RR, que foi semeada no dia 10 de dezembro de 2015, na densidade de semeadura de 16 sementes por metro de linha. A adubação de semeadura consistiu na utilização do formulado N-P-K (02-23-23), utilizado na dosagem de 320 kg ha⁻¹. O tratamento das sementes foi feito com Standak Top® (Fipronil 25,00% + piraclostrobina 2,50% + tiofanato metílico 22,50%) e Comofix® (Co 1% e Mo 10%), ambos na concentração de 2 mL kg⁻¹ de semente. O inoculante utilizado foi o Masterfix L® (inoculante líquido para soja) na concentração de 1,5 mL kg⁻¹ de semente de soja.

O delineamento experimental adotado foi em blocos ao acaso, com quatro tratamentos (níveis de doença) e cinco repetições. As parcelas tiveram dimensões de 6x10m. Para a manutenção da área foi utilizado o inseticida Avatar a cada 10 dias para o controle de lagartas e Galil para controle do percevejo. No manejo das doenças as aplicações consistiram em momentos distintos com produtos de princípio ativo distintos, conforme demonstrado na Tabela 1.

Tabela 10 - Produtos utilizados para o manejo de doenças

Fungicida	Fox® (400 mL ha ⁻¹) + Aureo (200mL/ha)	Orkestra® (300 mL ha ⁻¹) + Assist (500 mL/ha)	Aproach Prima® (300 mL/ha) + Nimbus (600 mL/ha)	Priori Xtra® (300 mL/há) + Nimbus (600 mL/ha)
Tratamento	Estádio de Aplicação			
	V6	R1	R1+18	R1+35
100%	X	X	X	X
60%	X	X		
30%			X	
0%				

As condições meteorológicas durante a condução da lavoura foram monitoradas pela estação meteorológica modelo Davis Vantage Pro instalada em uma fazenda ao lado da área experimental na qual foram monitorados dados de temperatura e umidade relativa do ar, conforme demonstrado na Figura 1.

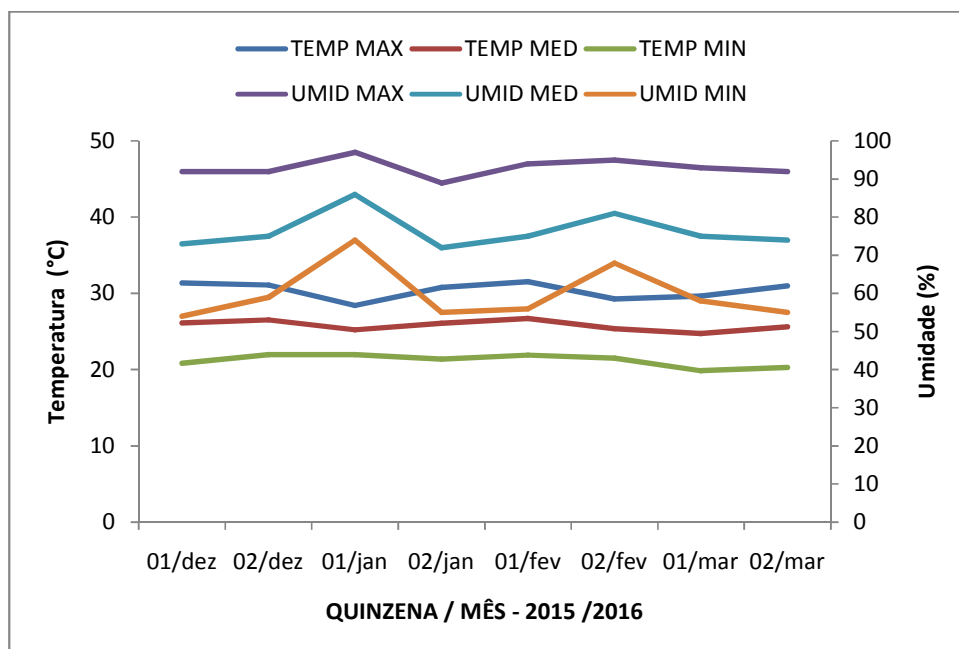


Figura 1 - Temperatura e umidade relativa do ar durante a Safra 2015/2016 de soja na Fazenda São José.

Também foram monitoradas a pluviometria com os dados expressos na Figura 2. Destaca-se que as chuvas, se constantes, podem ocasionar encharcamento do solo levando a problemas diversos nas plantas, como a vulnerabilidade das raízes ao ataque de patógenos.

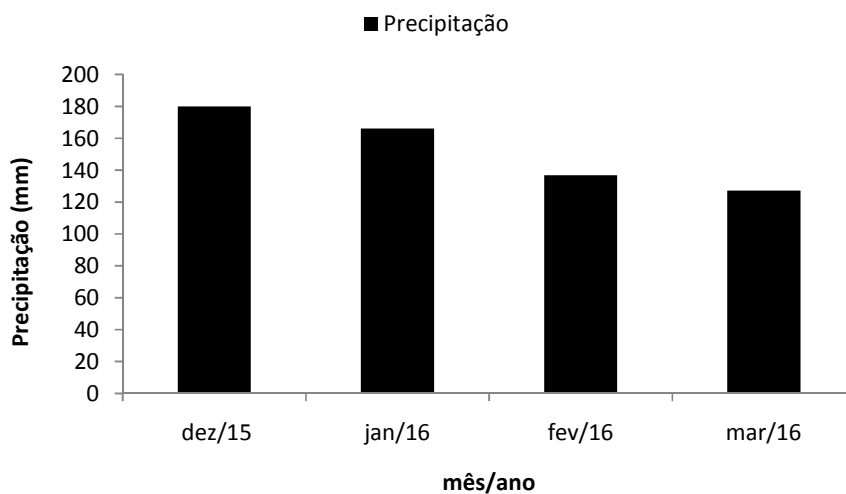


Figura 2 - Precipitação durante a Safra 2015/2016 de soja na Fazenda São José

2.2 Delineamento de Vãos e Captura de Imagens

Os registros de imagens foram realizados utilizando-se do equipamento VANT DJI Phantom 3 Professional, com peso de 1.280 gramas e velocidade máxima de 16

m/s. O Phantom utilizou a bateria LiPo 4x3,7V que possui uma autonomia de voo de aproximadamente 23 minutos. Além disso, ele é equipado com uma câmera Sony EXMOR 1/2.3", 12.4 M, lente FOV 94° 20 mm, suportando os formatos de arquivo FAT32/exFAT, JPEG, DNG e MP4, MOV (MPEG-4 AVC/H.264), e possui também um gimbal com estabilização nos 3 eixos e suporte a Micro SD com capacidade máxima de 64 GB.

O aplicativo DJI GO APP foi utilizado, objetivando o emprego, no Phantom3, de cinco categorias de vôos inteligentes: *Follow Me*, *Course Lock*, *Waypoints*, *Home Lock* e *Point of Interest*. Afim de utilizar a modalidade *Waypoints*, foram realizados testes, mas a cada vez que o Phantom 3 foi reiniciado, os *Waypoints* se perderam, inviabilizando assim o uso dos mesmos.

Considerando o tamanho da área experimental (24x50m) a demarcação de cada parcela de 60m² foi sinalizada com estacas de bambu (1,3 metros de altura). Essa sinalização não permaneceu em sua totalidade devido à entrada de máquinas para aplicação dos defensivos. Os vôos foram realizados a uma altura de 5 metros do solo e, a cada início de coleta de imagens, uma imagem única de todo o experimento foi obtida. Os vôos foram realizados entre os meses de dezembro de 2015 e março de 2016 pelo menos uma vez por semana, no período das oito às dez horas da manhã. Devido à grande quantidade de chuvas durante o mês de janeiro, algumas visitas foram canceladas. Foram realizadas coleta de imagens e filmagem da área experimental, sendo a proporção da imagem utilizada 4x3, com resolução de 4000x3000px e a filmagem realizada em Full HD com todos os parâmetros na configuração original de fábrica.

2.3 Seleção de imagens para composição do Banco de Imagens de Doenças na Soja

Com a realização dos vôos na área experimental de soja, as imagens foram coletadas e posteriormente foram armazenadas em 12 diretórios diferentes, conforme a data de captura. No total foram coletadas 711 imagens referentes às doenças da soja durante a Safra 2015/2016, correspondendo a 3,3GB. Como cada imagem possuía uma dimensão de 4000 x 3000 pixels e, em média, 4,7MB foi necessário o particionamento das imagens. Para tanto, foi implementado um programa que executou um *script* para automatizar a ação. Com isso, cada imagem

foi particionada em 12 novas imagens (1000 x 1000 = 1 MP) alcançando o total de 8532 imagens.

Posteriormente, efetuou-se o cálculo do tamanho da amostra aleatória simples para descrição da proporção populacional (BUSSAB; MORETTIN, 2011), utilizando um intervalo de confiança (IC) de 95% e erro padrão (EP) de 5%, chegando-se a uma amostra de 368 imagens. Como as imagens foram obtidas em 12 datas diferentes, a amostra calculada foi dividida pelo total de datas: $\frac{368}{12} = 30,7$ imagens de cada data. Devido a essa impossibilidade optou-se pelo sorteio de 31 imagens de cada data, totalizando 372 imagens. O sorteio das imagens foi realizado através da função =ALEATORIOENTRE (x,y) do software Excel. A máquina utilizada para o desenvolvimento do trabalho foi um Notebook Samsung NP-RV411-AD3 c/ Intel Core i3, 3GB, 320GB.

2.4 Segmentação das imagens

Na segmentação das imagens para posterior anotação e criação do banco de imagens anotadas, foi utilizado o algoritmo para geração de superpixel que tem sido cada vez mais utilizado em estudos relacionados à visão computacional. O superpixel refere-se ao desenvolvimento de um classificador que combina características como contorno, textura, brilho e continuidade (REN; MALIK, 2003). Para Achanta et al. (2010) os seguintes requisitos são desejáveis em relação aos superpixels:

1. superpixels devem respeitar os limites do objeto;
2. superpixels devem ser gerados de forma tão eficiente quanto possível;
3. superpixels quando usados para reduzir a complexidade computacional como uma etapa de pré-processamento, deve ser rápida para computar, memória eficiente e simples de usar.

O SLIC (Simple Linear Iterative Clustering) foi introduzido por Achanta et al. (2010) e é uma adaptação do método de agrupamento k-means para geração de superpixels. O algoritmo superpixel SLIC agrupa locais de pixels no espaço 5-D definido por L, a, b (valores da escala CIELAB de cor) e as coordenadas x e y dos pixels (Achanta et. al., 2010; Lv, 2015). Para tanto, tendo a imagem de entrada, realiza-se o particionamento da imagem em regiões retangulares, definindo-se o

número k correspondente a quantidade de unidades, levando cada superpixel a ter aproximadamente $\frac{N}{k}$ pixels, onde N é o número de pixels da imagem.

O processo do agrupamento começa com a etapa de inicialização, na qual os k centros dos agrupamentos $C = [l_i a_i b_i x_i y_i]$, são em cada pixel i , expressos por uma matriz regularmente espaçada em S , conforme equação (1). Para que os superpixels tenham aproximadamente o mesmo tamanho, o intervalo da matriz é determinado pela seguinte equação:

$$S = \sqrt{\frac{\text{Largura} \cdot \text{Altura}}{k}} \quad (1)$$

Assim, os centros do superpixel são movidos para locais com baixa magnitude de gradiente, numa vizinhança 3x3, para evitar que um superpixel tenha seu centróide colocado sobre regiões de borda e para reduzir as chances dele conter pixels ruidosos. Posteriormente, no passo de atribuição, cada pixel i é associado com o centro mais próximo do agrupamento, cuja região de busca se sobrepõe à sua localização, esta é a chave para acelerar esse algoritmo, pois ao limitar o tamanho da região da busca, reduz-se significativamente o número de cálculos de distância, o que resulta em uma vantagem de velocidade significativa sobre o agrupamento *k-means* convencional, no qual cada pixel deve ser comparado com todos os centros de agrupamento (Achanta et. al., 2010). O processo anteriormente descrito só é possível através da introdução da medida de distância D , a qual determina o centro mais próximo para cada pixel:

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \quad (2)$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \quad (3)$$

$$D_s = d_{lab} + \frac{m}{S} * d_{xy} \quad (4)$$

Onde D é a soma da distância d_{lab} e a distância d_{xy} normalizada pelo intervalo S . A variável m corresponde ao controle de compactação do superpixel, quanto maior o seu valor, mais a proximidade espacial é enfatizada e mais compactado é o agrupamento. Este procedimento é repetido até a convergência ou até um número máximo de iterações T . Na etapa de pós-processamento, os superpixels não representam necessariamente componentes ligados, de tal forma que o algoritmo precisa reforçar a conectividade através da re-atribuição de pixels disjuntos para superpixels próximos (ACHANTA et. al., 2010).

De acordo com Achanta et. al. (2010), por padrão, o único parâmetro de entrada do algoritmo SLIC Superpixel é o número de superpixels, de aproximadamente mesmo tamanho, k . Todavia, opcionalmente é possível ajustar o parâmetro compacidade, m , que permite controlar a forma do superpixel tornando-a mais quadrada/cúbica. Neste trabalho também foi utilizada a configuração do parâmetro sigma, que permite aplicar uma suavização na imagem, utilizando filtros gaussianos, antes da segmentação, através da utilização da biblioteca scikit-image (<http://migre.me/wutyR>), Figura 3.

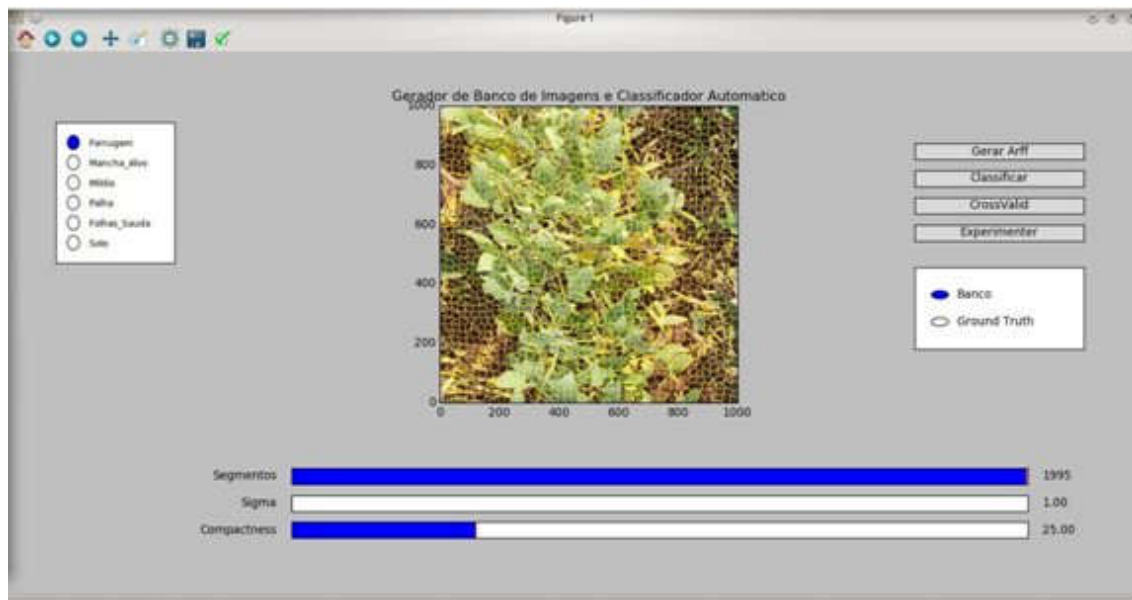


Figura 3 - Segmentador de imagens para geração do banco de imagens.

Nesta pesquisa, os parâmetros utilizados para a segmentação das imagens foram: Segmentos (k) = 1995, Sigma = 1 e Compacidade (m) 25. As doenças foliares da soja, principalmente no estágio inicial, possuem sintomas discretos, como pequenas manchas, conforme demonstrado na Figura 4, o que justifica o número de segmentos (1995) que foi adequado para separar os sintomas conforme patologia verificada nos folíolos da soja. Além disso, o parâmetro $m = 25$ para o contorno da patologia, e o valor 1 do sigma proporcionou a suavização necessária na imagem.



Figura 4 - Fragmento de uma imagem segmentada demonstrando visualmente a presença do míldio no folíolo da soja.

Após a segmentação, as imagens foram rotuladas nas classes: palha, solo, míldio, mancha alvo, ferrugem asiática e folhas saudáveis. A Figura 5 expõe algumas amostras que compõem o banco de imagens. O banco de imagens final conta com 22.140 imagens/segmentos, distribuídas em classes com as seguintes quantidades:

Ferrugem Asiática: 3894

Palha: 7170

Mancha Alvo: 75

Folha Saudável: 5049

Míldio: 1819

Solo: 4133



Figura 5 – Exemplo de imagens de classes que compõem o banco de imagens

2.5 Extração de atributos

Após a conclusão do banco de imagens passou-se a etapa de extração de atributos utilizados como entrada para os classificadores explorados neste trabalho. Essa extração foi realizada utilizando uma coleção de extratores de forma, cor, textura e orientação da imagem implementados nas bibliotecas OpenCV e scikit-image. Foram utilizados os seguintes extratores: atributos de cor RGB, HSV, Cielab (Mín., Máx., média e Desvio); descritor de forma, invariante a escala, translação e rotação: 7 momentos de Hu; atributos de textura – GLCM (contrastes, dissimilaridades, homogeneidades, asm, energias, correlações); forma e orientação: HOG; atributos de textura: LPB.

Para Hu (1962) os momentos invariantes de uma imagem permitem calcular a área de um objeto, ou mesmo identificá-lo mesmo que sofra mudança de tamanho ou que seja rotacionado. Assim o extrator de atributos implementado calcula os 7 momentos de Hu e os momentos raw e centralizados de ordem 1 e 2.

Uma Matriz de Co-ocorrência GLCM (Gray-Level Co-occurrence Matrix) é uma técnica utilizada dentro da área de análise de texturas que foi desenvolvida na década de 70 (HARALICK; SHANMUGAM; DINSTEN, 1973). Na extração dos atributos de textura baseados em matrizes de coocorrência (GLCM), utilizando matrizes 4x4 nas distâncias 1 e 2 e com ângulos 0, 45 e 90, emprega-se as seguintes propriedades de texturas definidas na GLCM: energia, contraste, correlação, homogeneidade e dissimilaridade.

Histogramas de Gradientes Orientados (HOG) é um descritor que calcula o histograma da orientação dos gradientes na imagem. É uma técnica de extração de atributos que inicialmente teve por objetivo auxiliar na detecção de pessoas em imagens (DALAL, TRIGGS, 2005).

Os Padrões Binários Locais (do inglês Local Binary Pattern - LBP) foi proposto inicialmente em (OJALA; PIETIKÄINEN; HARWOOD; 1996) e mais tarde generalizado em (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2002). São considerados um dos melhores extratores de textura e tem como vantagens sua invariância a mudanças em tons de cinza e eficiência computacional. Sua estratégia para detecção de textura é observar para um ponto central a variação da sua cor em relação aos seus vizinhos.

2.6 Classificação de imagens

Após segmentar e formar o banco de imagens anotadas passou-se a classificação utilizando o programa de computador proposto para reconhecimento de doenças na soja. O software Weka versão 3.8 executado no Windows 64 bits foi utilizado. O Weka é um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados (HALL et al, 2009) e tem como entrada arquivos no formato ARFF (Attribute-Relation File Format) que é um arquivo de texto ASCII que descreve uma lista de instâncias que compartilham um conjunto de atributos.

O ARFF foi desenvolvido pelo Projeto Machine Learning no Departamento de Ciência da Computação da Universidade de Waikato para uso com o software de aprendizado de máquina Weka. Os ARFFs utilizados como entrada para os testes no Weka neste trabalho foram gerados a partir dos extratores de atributos citados na seção 2.5. Os algoritmos utilizados para os testes comparativos foram Máquina de Vetores de Suporte (SVM), J48 (evolução do algoritmo C4.5), Florestas Aleatórias (no inglês Random Forest) e KNN (IBK no Weka). Todos os algoritmos foram executados com as configurações definidas por padrão no software Weka.

Os estudos teóricos que embasaram as SVMs (Support Vector Machines) foram iniciados por Vladimir Vapnik e estabeleceram princípios a serem seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu. O emprego da SVM possibilita resoluções de problemas de classificação de dados, gerando classificadores que apresentam bons resultados (LORENA; CARVALHO, 2007). O SMO (Sequential Minimal Optimization) é um algoritmo considerado eficiente para a implementação da técnica SVM (Support Vector Machine), para resoluções de problemas de programação quadrática, utilizado para acelerar o treinamento de SVM. Esse algoritmo particiona grandes problemas em séries de pequenos possíveis problemas a serem resolvidos (GIRARDELLO, 2010).

De acordo com Vasconcellos et al. (2011) o J48 é um algoritmo de indução de árvore de decisão derivado da proposta de Quinlan (1993) como C4.5 que representa uma significativa evolução do ID3 (QUINLAN, 1986). O algoritmo evoluiu do ID3 até a versão C4.8, que é a versão implementada no Weka em linguagem Java. Tem como base a árvore de decisão, que é formada a partir de nós de decisão

que tomam uma escolha no valor de um atributo. Utilizando um conjunto de dados, o algoritmo constrói árvores de decisão, compostas das folhas que contém as classes que devem ser classificadas e os nós são atributos da imagem que apresentam melhor eficiência (VASCONCELOS et al., 2011; BHARGAVA et al., 2013).

Já o classificador Floresta Aleatória (Random Forest) é um método proposto por Breiman (2001) e consiste em um conjunto de árvores de decisão combinadas. Cada árvore de decisão é construída utilizando uma amostra aleatória inicial dos dados e, a cada divisão desses dados, um subconjunto aleatório de m atributos é utilizado para a escolha dos atributos mais informativos. Ao fim, a floresta aleatória gera uma lista dos atributos mais importantes no desenvolvimento da floresta, que são determinados pela importância acumulada do atributo nas divisões dos nós de cada árvore da floresta (JAMES et al., 2013).

O IBK é uma implementação do K-Nearest Neighbor (KNN) e se baseia em instâncias, ou seja, agrupamento dos dados que faz a classificação entre os k -vizinhos mais próximos. Na determinação de um elemento não pertencente ao conjunto de treinamento, esse classificador procura k elementos do conjunto de treinamento que possua a menor distância em relação a esse elemento desconhecido (GALVÃO; HRUSCHKA JÚNIOR, 2004; SILVA, 2005). De forma geral, a medida de proximidade mais utilizada é a distância euclidiana.

2.7 Avaliações agronômicas

Foram realizadas avaliações para medir a severidade das doenças com o uso de escalas diagramáticas propostas para cada tipo de doença verificada nesse experimento, conforme Figura 6. A avaliação teve início 60 dias após o plantio da lavoura. Foram realizadas 5 amostragens espaçadas de 15 em 15 dias a partir do momento de avaliação da cultura.

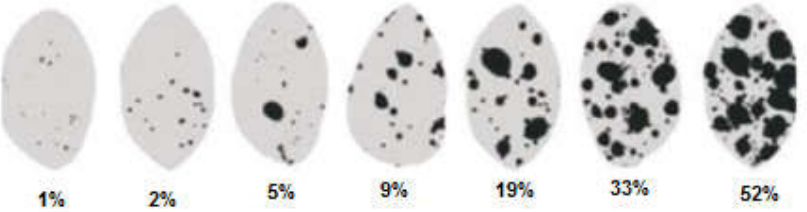
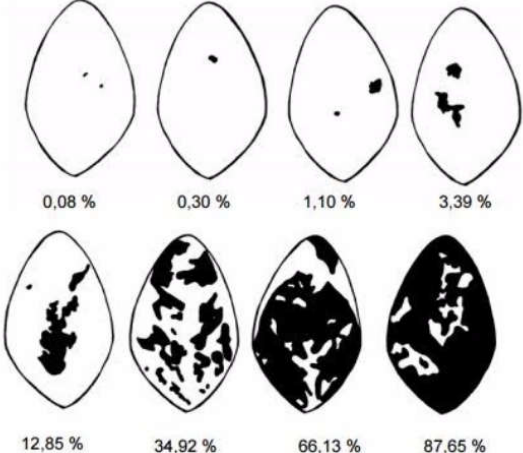
Escala diagramática - ferrugem da soja Godoy et al. (2006)	 0,6% 2,0% 7,0% 18,0% 42,0% 78,5%
Escala diagramática - mancha-alvo em soja (Soares et al. 2009)	 1% 2% 5% 9% 19% 33% 52%
Escala diagramática - míldio da soja Kowata et. al (2008)	 0,08 % 0,30 % 1,10 % 3,39 % 12,85 % 34,92 % 66,13 % 87,65 %

Figura 6 - Escalas diagramáticas usadas na avaliação da severidade de doenças da soja

A avaliação da severidade das doenças foliares medidas através da escala diagramáticas ocorre através de notas que informam a porcentagem do tecido foliar da planta que se encontra afetado pela patologia. Essas notas dependem da percepção do avaliador e exige treinamento prévio para reconhecimento da doença e precisão do grau de ataque. A avaliação da severidade tem como objetivo expressar a intensidade da doença e os danos causados, todavia é uma tarefa trabalhosa e subjetiva por depender da acuidade do avaliador.

3 Resultados e Discussão

Nesta seção são apresentados os resultados das avaliações realizadas durante o experimento. Os resultados obtidos através da avaliação agronômica com o uso da escala diagramática demonstra a intensidade da doença ao longo das avaliações e é apresentada na Seção 3.1. O reconhecimento das patologias testados por algoritmos de aprendizagem automática são apresentados na Seção 3.2.

3.1 Avaliação agronômica com escalas diagramáticas

As doenças da soja verificadas no experimento foram: míldio, mancha alvo e ferrugem asiática. A primeira doença visualizada foi o míldio, durante o início do estágio R1, seguida pela mancha alvo e ferrugem asiática, sendo que essa última foi verificada no final do ciclo da soja. As doenças foram verificadas em todos os tratamentos e parcelas.

O míldio foi identificado no experimento através de observação visual, onde foram percebidas pontuações amarelas na parte superior dos folíolos, e na parte inferior, na região correspondente as pontuações amarelas foram observadas estruturas com coloração rosada. Na identificação da mancha alvo no plantio experimental, verificou-se que os folíolos apresentavam pontuações pardas com halo amarelo, evoluindo posteriormente para manchas circulares, de coloração castanho-clara a castanho-escuro. Já a ferrugem asiática apresentou lesões que variaram na cor castanho-clara a castanho-avermelhada, acompanhado de lesões semelhantes a bolhas na parte inferior da folha, que são as estruturas de reprodução do fungo (EMBRAPA, 2003).

Com a avaliação agronômica realizada com o uso da escala diagramática, pode-se avaliar a severidade das doenças presentes no experimento (míldio, mancha-alvo e ferrugem asiática) e acompanhar a evolução da patologia (MORAES, 2007). Todavia, as avaliações com a ferrugem asiática não foi significativa e a doença só foi verificada no experimento a partir da terceira avaliação. Assim a Figura 7 demonstra a evolução da severidade do míldio em função dos tratamentos aplicados e dos dias de plantio.

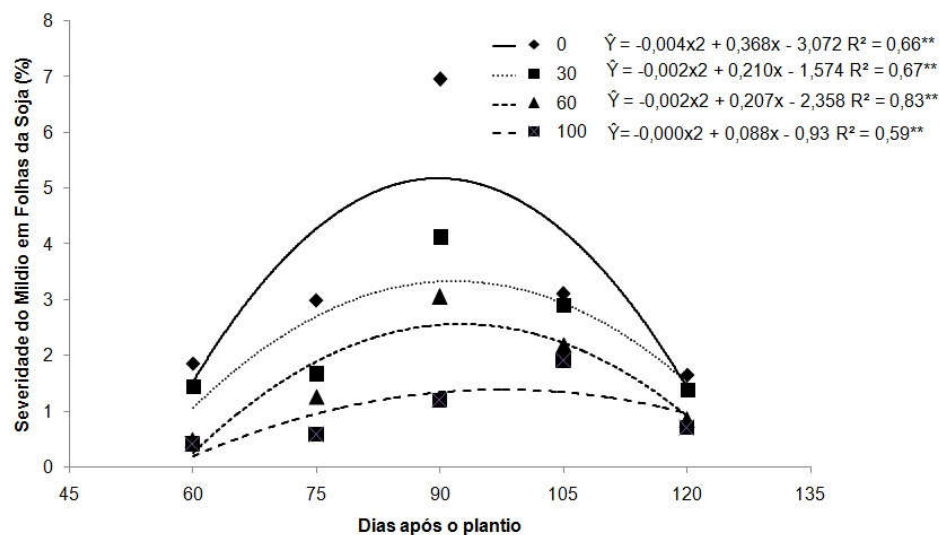


Figura 7 - Severidade do míldio na cultura da soja

Observa-se que para mancha alva da soja (Figura 8) a severidade no tratamento com aplicação de 100% do controle de doenças foi a menor, e como esperado, o tratamento onde não houve aplicação de defensivo na plantação de soja, a presença da doença foi mais evidente. O progresso da doença comportou-se de forma linear com sua maior incidência próximo aos 90 dias de cultivo (mês de março) coincidindo com o final de ciclo para a cultura da soja (GODOY et al, 2012).

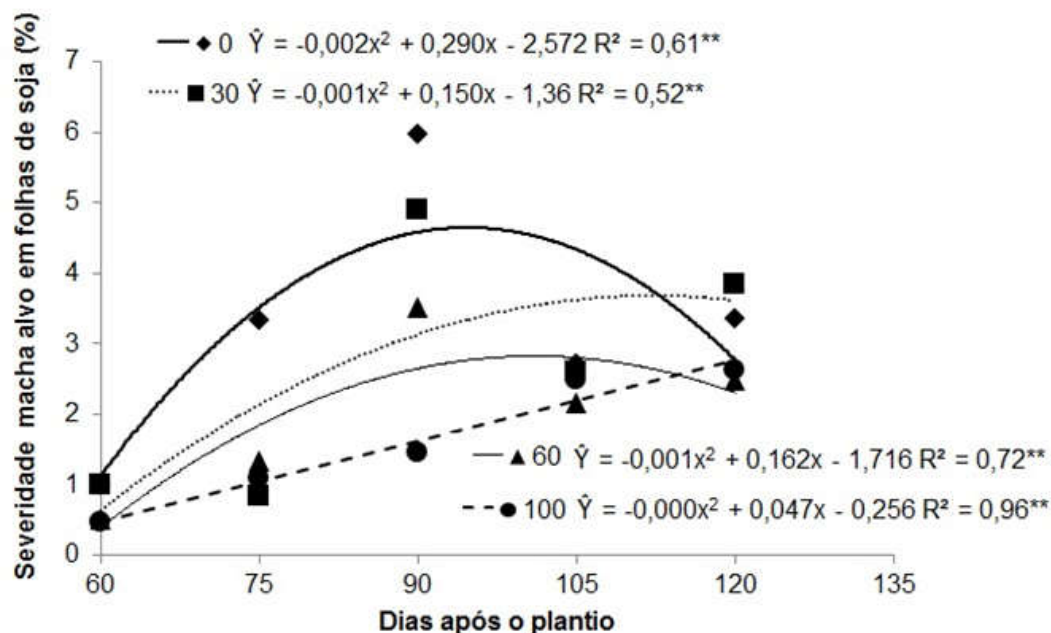


Figura 8- Severidade de mancha-alva na cultura da soja em função dos dias de plantio sob diferentes formas de controle.

Gonçalves (2010) ressalta a importância de conhecer a doença e quantificar os sintomas causados pelo patógeno, o que foi realizado com o uso das escalas diagramáticas. Destaca-se que a queda apresentada na severidade das doenças, nas avaliações, tanto na Figura 7 quanto na Figura 8, pode ter sido influenciada pelo aparecimento da ferrugem asiática no plantio experimental que ocorreu a partir da terceira avaliação, coincidindo com o período em que começou a ocorrer o processo de desfolha.

3.2 Avaliação por aprendizagem de máquina

Para a construção de um sistema de visão computacional capaz de reconhecer patologias em imagens de produção da soja quatro técnicas de aprendizagem supervisionada foram exploradas: SVM, J48, FA (Floresta Aleatória) e KNN, e a técnica de amostragem utilizada foi a validação cruzadas com 10 dobras. Todos os algoritmos foram executados com as configurações definidas por padrão no software Weka. Assim, após a segmentação das imagens através do SLIC e extração de atributos para aprendizagem supervisionada, procedeu-se a classificação do banco de imagens no ambiente Weka. Os resultados das classificações foram avaliados por meio das métricas Porcentagem de Classificação Correta – PCC, Medida-F, Área Sob a Curva ROC, Kappa, Revocação e Precisão. Os resultados estão demonstrados na Tabela 2.

Tabela 11 - Avaliação dos classificadores através de métricas

Métricas	SVM	J48	FA	KNN
PCC	94,75±0,42	89,76±0,57	92,90±0,45	83,38±0,68
Medida F	0,950±0,00	90,00±0,01	92,00±0,01	83,00±0,01
Curva ROC	0,950±0,00	78,00±0,03	97,00±0,00	68,00±0,02
Kappa	93,00±0,01	87,00±0,01	91,00±0,01	78,00±0,01
Revocação	95,00±0,03	90,00±0,04	93,00±0,04	83,00±0,03
Precisão	94,00±0,03	90,00±0,03	93,00±0,04	83,00±0,03

As métricas PCC, Medida F, Kappa, Revocação e Precisão indicaram que o classificador SVM obteve melhores resultados nas avaliações. Todavia, o melhor resultado obtido foi com a Curva ROC com o algoritmo Floresta Aleatória. A Figura 9 apresenta a matriz de confusão gerada com a utilização do classificador SVM por obter melhores desempenhos com ambas as métricas (exceto a área sob a curva ROC) onde as imagens em tom vermelho representam os valores mais altos.

Destaca-se que quanto mais próximos os valores das diagonais principais da matriz de confusão sejam do valor total de imagens contidas na classe, menor foi a confusão entre as classes.

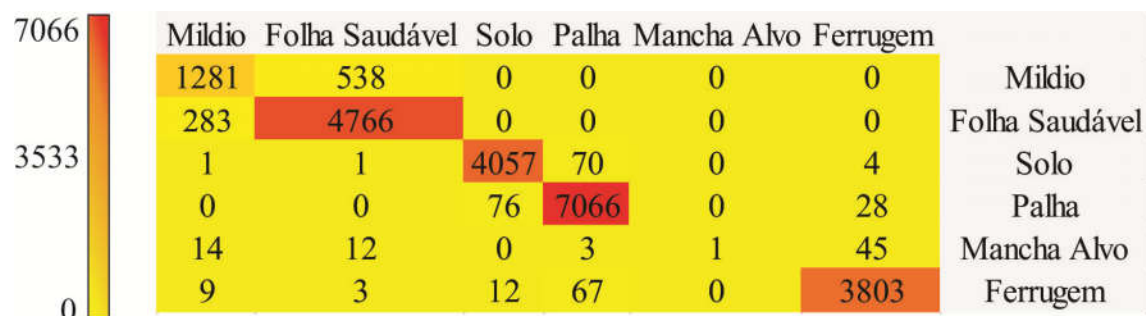


Figura 9 - Matriz de confusão gerada no ambiente Weka com o classificador SVM

Com as informações contidas na matriz, foram calculadas as métricas Coeficiente de Jaccard (CJ), Coeficiente de Yule (CY), precisão e revocação para cada classe - mildio, folhas saudáveis, solo, palha, mancha-alvo e ferrugem asiática-conforme resultados apresentados na Tabela 3. Considerando se tratar de um problema de classificação com mais de duas classes, o cálculo das medidas quantitativas, foi realizado, considerando a classe de interesse como a classe positiva, enquanto todas as outras foram consideradas como negativas (LANTZ, 2015).

Tabela 12 - Classes similares de acordo com os coeficientes de similaridade

Classes	CJ	CY	Precisão	Revocação
Mildio	0,6025	0,7805	0,8067	0,70423
Folhas Saudáveis	0,8506	0,8790	0,8959	0,94395
Solo	0,9611	0,9745	0,9788	0,98161
Palha	0,9666	0,9736	0,9806	0,98550
Mancha Alvo	0,0133	0,9967	1,0000	0,01333
Ferrugem	0,9577	0,9752	0,9802	0,97663

As métricas calculadas (Coeficiente de Jaccard (CJ), Coeficiente de Yule (CY), precisão e revocação) tem como base a comparação dos pixels das imagens segmentadas com as imagens de referencia (ANDRADE et al, 2012). Os resultados apresentados para o coeficiente de Jaccard e revocação indicaram que a melhor

classificação foi com a classe palha, todavia entre as doenças, a melhor classificação ocorreu na classe ferrugem asiática. A mancha-alvo obteve melhor classificação quando observada através do coeficiente de Yule e da métrica precisão.

Para verificar se houve diferença significativa entre os classificadores foi aplicada a Análise de Variância, através do software R. Considerando o valor- $p=0,0311$, a um nível de significância de 0,05%, pode-se afirmar que há evidências estatísticas de que os desempenhos sejam diferentes entre os classificadores. Dado que o teste ANOVA indica evidências estatísticas de que há diferença no desempenho entre os classificadores mas não indica quais classificadores se diferem foi aplicado o teste de Tukey para essa identificação (ABDI; WILLIAMS, 2010). A Tabela 4 indica que há evidências estatísticas de que o desempenho é diferente entre os algoritmos SVM e KNN.

Tabela 13 - Teste de Tukey para os classificadores

Classificadores	Valor-p
KNN-J48	0.1267587
FA-J48	0.9921373
SVM-J48	0.7903252
FA-KNN	0.2012547
SVM-KNN	0.0226619
SVM-FA	0.6320562

A diferença entre os classificadores pode ser melhor evidenciadas através da visualização da Figura 10 onde se pode analisar as medianas da taxa de acerto de cada classificador.

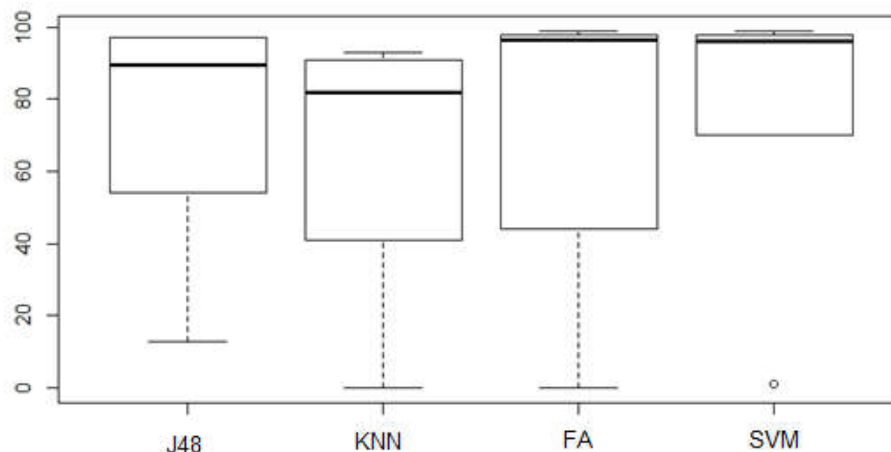


Figura 10 – Diagramas de caixa representando as diferenças entre os desempenhos dos classificadores

Após as técnicas de aprendizagem supervisionada (SVM, J48, Floresta Aleatória e KNN) serem avaliadas com o uso das métricas, o programa foi utilizado para realizar classificação de algumas imagens obtidas através do VANT na plantação de soja, conforme demonstrado na Figura 11.

Conforme se observa na Figura 11, a primeira imagem da esquerda para a direita refere-se a imagem original e a segunda refere-se a imagem já classificada pelo programa. As imagens foram obtidas pelo VANT na plantação de soja, durante o experimento, sendo capturadas, respectivamente nas seguintes datas: 25/02/2016, 04/03/2016 e 08/03/2016.

Para a realização dessa classificação visual a imagem foi segmentada pelo algoritmo SLIC (utilizando os mesmos parâmetros empregados na segmentação da criação do banco de imagens) e posteriormente os segmentos foram classificados, de maneira automática pelo programa, como pertencentes a uma das classes: míldio, mancha alvo, ferrugem asiática, folha saudável, solo e palha. O algoritmo utilizado na classificação foi o SVM.

A primeira imagem não apresentou a presença da mancha alvo da soja, mas teve evidências de ferrugem e míldio (6,71% e 0,32%, respectivamente). Na segunda imagem o ataque das doenças míldio, mancha alvo e ferrugem foi de, respectivamente 13,83%, 0,62% e 14,79%. Na última imagem as doenças míldio, mancha alvo e ferrugem alcançaram o total de 21,56%, 0,04% e 5,91%.

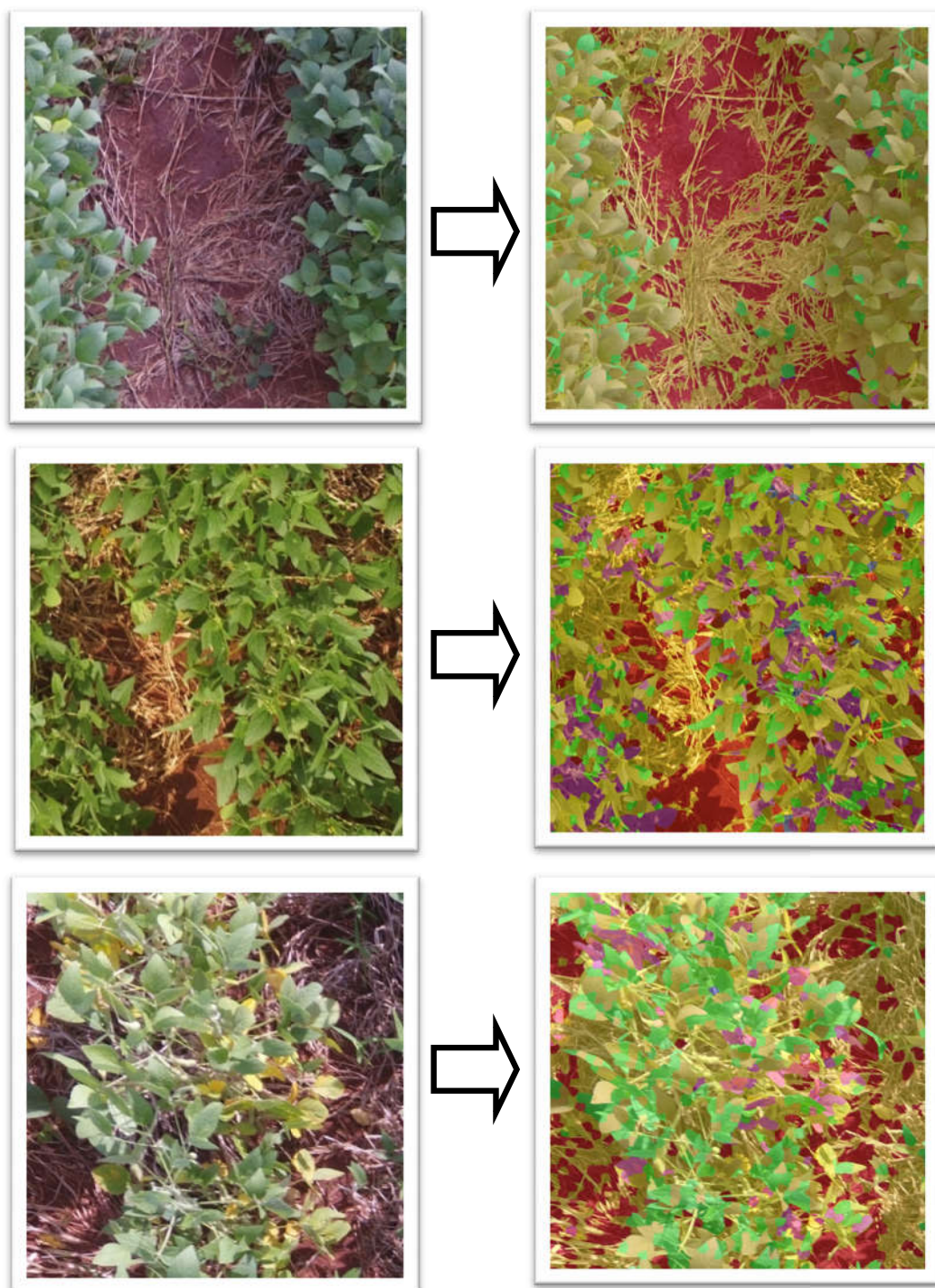


Figura 11 – Classificação visual das doenças da soja nas classes: míldio ■, macha alvo ■, ferrugem asiática ■, folha saudável ■, solo ■ e palha ■ realizada pelo programa de computador.

A observação visual das imagens demonstra uma variação de cor para as patologias. Todavia, no caso da ferrugem asiática, a coloração das regiões foliares afetadas são predominantemente castanho-claras ("TAN"). Todavia, cultivares onde o patógeno é resistente as lesões são predominantemente castanho-avermelhadas ("reddish-brown - RB") (EMBRAPA, 2003).

Além disso, no final do ciclo algumas doenças podem ser confundida (BONALDO; RIEDO; LIMA, 2009). A semelhança nos sintomas da ferrugem e de outras doenças pode levar a confusão e, conseqüentemente, a não identificação correta das patologias (EMBRAPA, 2003). Nesse sentido, a classificação dos segmentos que compõem cada imagem em suas respectivas classes, conforme aprendizado recebido pelo programa de computador proporciona maior confiabilidade acerca dos diagnósticos obtidos.

4 Conclusão

Esse artigo teve como objetivo reconhecer doenças no experimento de soja, e para tanto foi implementado um programa que teve como base aprendizagem automática e visão computacional. No experimento foram verificadas a presença das doenças: ferrugem asiática, mancha-alvo e míldio.

Nas avaliações utilizando as escalas diagramáticas pode-se verificar diferentes graus de severidade das doenças conforme se alterava o tratamento que se deu em função da porcentagem de defensivos utilizados no experimento. Assim, quanto menor a porcentagem de defensivo, maior a severidade do ataque das doenças no experimento.

Na classificação das imagens anotadas pertencentes ao banco de imagens realizadas com os algoritmos de classificação do Weka, por meio da avaliação com as métricas, os resultados indicaram que o melhor desempenho foi obtido pelo classificador SVM, exceto para a métrica Área Sob a Curva ROC que obteve o melhor desempenho com o classificador Florestas Aleatórias. Todavia o classificador utilizado para a construção das matrizes de confusão, para cálculo das demais métricas e avaliação entre as classificações foi o SVM por obter melhor desempenho na maioria das métricas utilizadas na avaliação. Entre as doenças, os resultados apresentados para o coeficiente de Jaccard e revocação na classe ferrugem asiática. A mancha-alvo obteve melhor classificação quando observada através do coeficiente de Yule e da métrica precisão.

Na classificação com imagens de referência anotadas por um especialista logrou-se bom desempenho somente nas classes solo, folha saudável e ferrugem, todavia com resultados inferiores a da classificação realizada com as imagens anotadas pertencentes ao banco de imagens. Todavia apenas uma imagem foi anotada pelo especialista, que mencionou cansaço na tarefa.

Assim, verifica-se que o programa é apto a diferenciar as classes, onde as técnicas que obtiveram melhor desempenho foi o SVM e Florestas Aleatórias, levando em consideração os resultados obtidos com todas as métricas de avaliação utilizadas.

Destaca-se a importância de um bom resultado na classificação principalmente porque um erro pode induzir um produtor de soja a prejuízos, seja prejuízo econômico, social ou ambiental, já que a demora no uso do defensivo pode levar a

um resultado, e o apressar em defender o plantio com o uso dos defensivos pode levar a outros prejuízos. Assim a proposta demonstra alcançar o objetivo proposto para esta pesquisa e contribui com outros pesquisadores pela disponibilização de um banco de imagens robusto com as classes trabalhadas nesse artigo. Além disso, o programa pode atuar enquanto um redutor da subjetividade existente nas avaliações realizadas por técnicos ou especialistas, no momento de reconhecer as doenças, melhorando o nível de acerto em relação a avaliação humana na classificação das imagens.

Referências Bibliográficas

ABDI, H., WILLIAMS, L. Newman-Keuls and Tukey test. In Salkind, N., Frey, B., & Dougherty, D. (Eds.), *Encyclopedia of Research Design*, pp. 897–904. Sage, Thousand Oaks, CA, (2010).

ACHANTA, R.; SMITH, K.; LUCCHI, A.; FUA, P.; and SUSSTRUNK, S.. Slic superpixels. Technical report, EPFL, Tech.Rep. 149300, 2010.

AFRIDI, M. J.; LIU, X.; MITCHELL MCGRATH, J.. "An Automated System for Plant-level Disease Rating in Real Fields," in *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014)*, Stockholm, Sweden, August 24-28, 2014.

ANDRADE, W. T.; QUINTA, L. N. B. ; GONCALVES, A. B. ; CEREDA, M. P. ; H. Pistori . Segmentação Baseada em Textura e Watershed aplicada a Imagens de Pólen. In: *SIBGRAPI 2012 - Conference on Graphics, Patterns and Images, Workshop of Undergraduate Work (WUW)*, 2012, Ouro Preto - MG. *Anais do SIBGRAPI*, 2012.

BHARGAVA, N.; SHARMA, G.; BHARGAVA, R.; MATHURIA, M. Decision Tree Analysis on J48 Algorithm for Data Mining. *Journal of Advanced Research in Computer Science and Software Engineering*. v. 3, 2013.

BUSSAB, W. O.; MORETTIN, P. A. (2011). *Estatística básica* (7a ed.). São Paulo: Saraiva.

BONALDO, M. S.; RIEDO, I. C.; LIMA, A. R. Monitoramento e diagnóstico de doenças foliares da cultura da soja na região COMCAM na safra 2007/2008. Campo Digital, Campo Mourão, v. 4, n. 1, p. 127–136, 2009.

BREIMAN, L.. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.

COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB (2016). Levantamentos de Safra 2015/2016. Disponível em <<http://www.conab.gov.br/conteudos.php?a=1252&t=>>>. Acesso em: 02 dezembro de 2016.

DALAL, N.; TRIGGS, B.. HISTOGRAMS OF ORIENTED GRADIENTS FOR HUMAN DETECTION. IN CVPR, PAGES 886-893, 2005.

EMBRAPA SOJA. Tecnologias de produção de soja: região Central do Brasil, 2003. Londrina, 2004. 239p.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA SOJA. Tecnologias de produção de soja – região central do Brasil 2014. - Londrina: Embrapa Soja, 2013.

GALVÃO, D. C. DE O; HRUSCHKA JUNIOR, E. R.. Ponderação do knn utilizando-se de métodos tradicionais de seleção de atributos. Da Vinci , Curitiba, v. 1 , n. 1, p. 115-125, 2004. Disponível em < <http://www.up.edu.br/davinci/pdf08.pdf>>. Acesso em 17/05/2015.

GIRARDELLO, A. D.. Um Estudo sobre o Uso de Máquinas de Vetores de Suporte em Problemas de Classificação. CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS. Colegiado de Ciência da Computação. CASCAVEL. 2010.

GODOY, C.V.; KOGA, L.J.; CANTERI, M.G.. Diagrammatic scale for assessment of soybean rust severity. Fitopatologia Brasileira 31:063-068. 2006.

GODOY, C.V. et al.. Eficiência de fungicidas para o controle da mancha-alvo, *Corynespora cassiicola*, na safra 2011/12: resultados sumarizados dos ensaios cooperativos. EMBRAPA. Circular Técnica 94. Londrina-PR, 2012.

GÓMEZ-CANDÓN, D.; CASTRO, A.I.D.; LÓPEZ-GRANADOS, F. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precis. Agric.* 2014-15, 44–56.

GONÇALVES, F. P.. Progresso poliético e quantificação de danos da Clorose Variegada dos Citros em laranjeiras 'Natal' submetidas a déficits hídricos. Tese (Doutorado) – Escola Superior de Agricultura 'Luiz de Queiroz'. Piracicaba. 2010.

GRIGOLLI, J. F. J.. Manejo de Pragas e Doenças na Cultura da Soja. Palestra: Fitossanidade Safra 2015. Fundação MS. Disponível em < <http://migre.me/rdAYC>> Acesso em Ago/2015.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update. SIGKDD Explorations, v. 11, n. 1., 2009.

HARALICK, R.M.; SHANMUGAM, K.; DINSTEN, I. "Textural Features for Image Classification", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-3, pp. 610-621, 1973.

HU, M. K.. "Visual Pattern Recognition by Moment Invariants", IRE Transactions on Information Theory, vol. IT-8, pp.179-187, 1962. Artigo disponível em: <http://www.sci.utah.edu/~gerig/CS7960-S2010/handouts/Hu.pdf>, Acesso em: 20/02/2015.

HUNT, E.R.; RONDON, S. I. ; HAMM , P. B. ; TURNER, R. W. ; BRUCE, A. E., et al. " Insect detection and nitrogen management for irrigated potatoes using remote sensing from small unmanned aircraft systems ", Proc. SPIE 9866, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, 98660N (May 17, 2016); doi:10.1117/12.2224139; <http://dx.doi.org/10.1117/12.2224139>

JAMES, G.; HASTIE, T.; TIBSHIRANI, R. An introduction to statistical learning: with applications in R. London: Ed. Springer, 2013. 429 p.

JIA, Y.; SU, Z.; SHEN, W.; YUAN, J.; XU, Z.. UAV Technology and Its Application in Agriculture. Advanced Science and Technology Letters. Vol.137 (SUComS 2016), pp.107-111 <http://dx.doi.org/10.14257/astl.2016.137.20>

KOWATA, L. S.; MAY-DE-MIO, L. L.; DALLA-PRIA, M.; SANTOS, H. A. A.. Escala diagramática para avaliar severidade de mildio na soja. Scientia Agraria, Curitiba, v.9, n.1, p.105-110, 2008.

LANTZ, B.. Machine Learning with R. Packt, 2º Edição, 2015.

LORENA, A. C; CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. RITA – Revista de Informática Teórica, v. 14, n. 2, p. 43-67, 2007

LV, J.. An Improved SLIC Superpixels using Reciprocal Nearest Neighbor Clustering. International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 5 (2015), pp. 239-248. Disponível em<<http://dx.doi.org/10.14257/ijsp.2015.8.5.25>> Acesso em 20 de jul. de 2015

MORAES, S.A. de Quantificação de doenças de plantas. 2007. Artigo em Hypertexto. Disponível em: <http://www.infobibos.com/Artigos/2007_1/doencas/index.htm>. Acesso em: 10/1/2017

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A Comparative study of Texture Measures with Classification Based on Feature Distributions, *Pattern Recognition* 29 (1), p. 51-59, 1996.

OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(7), p. 971-987, 2002.

PEÑA-BARRAGAN JM, KELLY M, DE CASTRO AI, LOPEZ-GRANADOS F.. Object-based approach for crop row characterization in UAV images for site-specific weed management. In Queiroz-Feitosa et al., editors. 4th International Conference on Geographic Object-Based Image Analysis (GEOBIA 2012), Rio de Janeiro, Brazil: 426-430.

QUINLAN, J. R. (1993). C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

QUINTA, L. N. B. Visão Computacional aplicada na classificação de grãos de pólen. Dissertação de Mestrado em Biotecnologia. Programa de Pós Graduação em Biotecnologia - Universidade Católica Dom Bosco. Janeiro/2013.

REN, X.; MALIK, J.. Learning a classification model for segmentation. *IEEE ICCV*, pp. 10–17, 2003.

SILVA, L. M. Uma aplicação de Árvores de Decisão, Redes Neurais e KNN para a Identificação de Modelos ARMA não Sazonais e Sazonais. Tese de Doutorado em Engenharia Elétrica – Pontifícia Universidade Católica do Rio de Janeiro. 2005.

SOARES, R.M.; GODOY, C.V.; OLIVEIRA, M.C.N. Escala diagramática para avaliação da severidade da mancha alva da soja. **Tropical Plant Pathology**, v.34, n.5, p.333-338, 2009.

UGALE, V.; GUPTA, D.. A Comprehensive Survey on Agricultural Image Processing. *International Journal of Science and Research (IJSR)*. Volume 5 Issue 1, January 2016

VASCONCELOS, B.R.; GONÇALVES, A.B.; VILPOUX, O.F.; CEREDA, M.P. Georeferencing Methodology For Honey Tracking. In: INTERNATIONAL APICULTURAL CONGRESS, XLII, 2011, Buenos Aires. Anais...Buenos Aires, APIMONDIA, 2011.

CONSIDERAÇÕES FINAIS:

A pesquisa documentada através dos dois artigos demonstra que o programa implementado é apto para distinguir entre as classes trabalhadas durante os testes com os classificadores e sua avaliação via métricas de desempenho. Ainda que no ensaio experimental (Safrá 2014/2015), cujas imagens foram trabalhadas no primeiro artigo, a única doença verificada tenha sido a ferrugem asiática, o programa distinguiu entre uma classe e outra. Todavia, como foram utilizados dois VANTS na coleta de imagens no ensaio, as imagens ficaram com padrões diferentes de imagens, levando a dois bancos de imagens anotadas, conforme o instrumento de coleta utilizado (banco de imagens A -aboveR3C4-, e banco de imagens B -phantomR6C4).

Os bancos de imagens foram classificados através dos algoritmos de classificação SVM, KNN, J48 e Florestas Aleatórias. A métrica Área Sob a Curva ROC, que é uma representação gráfica da taxa de verdadeiros positivos versus taxa de falsos positivos, demonstrou desempenho de 100% para ambos os bancos de imagens, com a técnica de aprendizagem automática Florestas Aleatórias.

Com o experimento delineado e exposto no Artigo 2, cujo texto compõe essa tese, verificou-se a presença de, além da ferrugem asiática, mais duas patologias: míldio e mancha-alvo. Esse experimento, além da avaliação do programa, contou também com duas avaliações de cunho agrônomo: avaliação do programa contra a avaliação do especialista e avaliação com o emprego da escala diagramática para medir a severidade da patologia na soja.

Com a avaliação realizada com o auxílio da escala diagramática, verificou-se que as avaliações para a ferrugem asiática não foi significativa, pois a doença foi verificada no experimento já no início do período da desfolha. A severidade no tratamento da mancha-alvo e míldio, com aplicação de 100% do controle de doenças foi menor. O progresso da doença comportou-se de forma linear com sua maior incidência próximo aos 90 dias de cultivo (mês de março) coincidindo com o final de ciclo para a cultura da soja.

Nas classificações através da aprendizagem supervisionada empregou-se os algoritmos: SVM, J48, Florestas Aleatórias e KNN, e para avaliar as classificações

foram utilizadas as métricas: PCC, Medida F, Kappa, Revocação, e Precisão. O melhor desempenho obtido foi com o classificador Florestas Aleatórias avaliado com o emprego da métrica Área Sob a Curva ROC. Todavia, todas as demais métricas indicaram melhor desempenho com o classificador SVM.

Por fim, tendo a classificação do especialista como referência, testou-se a classificação do programa de computador proposto, utilizando as métricas Porcentagem de Classificação Correta -PCC, Coeficiente de Jaccard e Coeficiente de Yule. Tendo com base os segmentos anotados pelo especialista, os melhores resultados das avaliações do programa foi na classificação das classes solo, folha saudável e ferrugem asiática. Além disso, a classificação realizada pelo programa proposto a partir do banco de imagens obteve resultados superiores aos resultados obtidos tendo como base a classificação do especialista.

O programa proposto consegue detectar doenças foliares na plantação de soja, conforme demonstrado pela distinção entre uma classe e outra no decorrer da pesquisa. Todavia aqui foram trabalhadas apenas três doenças da soja. Em caso de patologia diferente, há a necessidade de banco de imagens anotadas da nova doença para testar o reconhecimento. Além disso, a robustez do banco de imagens não permitiu a realização de testes com todos os classificadores disponíveis no ambiente WEKA por provocar travamento da máquina, de modo que apenas quatro classificadores foram explorados. Nesse sentido, o uso de uma máquina mais robusta pode colaborar na busca de classificadores que proporcionem melhores resultados, pois o programa proposto dispõe de ferramentas para essa busca.

Destaca-se também que embora os extratores implementados no programa demonstrem robustez, conforme resultados das avaliações com as métricas, a depender das características de novas doenças a serem investigados, outros extratores podem ser requisitados.

Na coleta de imagens a principal limitação ocorreu pela ferramenta utilizada para a obtenção das imagens tanto no ensaio experimental quanto no experimento em si, tem uma autonomia de voo média de 23 minutos, incorrendo em necessidade de substituição de bateria ou recarregamento, fator que pode restringir o uso do mesmo, a depender dos objetivos do usuário. Dessa forma, trabalhos futuros podem melhor explorar as limitações constantes nessa tese.

REFERÊNCIAS BIBLIOGRÁFICAS

ACHANTA, R.; SMITH, K.; LUCCHI, A.; FUA, P.; and SUSSTRUNK, S.. Slic superpixels. Technical report, EPFL, Tech.Rep. 149300, 2010.

AFRIDI, M. J.; LIU, X.; MITCHELL MCGRATH, J.. "An Automated System for Plant-level Disease Rating in Real Fields," in Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014), Stockholm, Sweden, August 24-28, 2014.

ANNES, R.. Abordagem Multiagente para Processamento de Múltiplos Corpora: uma arquitetura multiagente de rotuladores de categorias gramaticais – Dissertação (mestrado). Faculdade de Informática da Pontifícia Universidade Católica do Rio Grande do Sul, Faculdade de Informática, Porto Alegre, 1999. 79p.

APAN, A.; TROY, J.; LES, C. Z.. The use of an unmanned aerial vehicle as a remote sensing platform in agriculture. Australian. Journal of Multi-disciplinary Engineering. Nov. 2010.

ASCH, V. V.. "Macro- and micro-averaged evaluation measures," Tech. Rep., University of Antwerp, 2013. Disponível em <<http://www.cnts.ua.ac.be/~vincent/pdf/microaverage.pdf>>. Acesso em 17/03/2015

BARRETO, A. F.. Avaliação de parâmetros da tecnologia de aplicação para o controle da ferrugem asiática da soja. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias. Jaboticabal, 2011 v, 81f.

BHARGAVA, N.; SHARMA, G.; BHARGAVA, R.; MATHURIA, M. Decision Tree Analysis on J48 Algorithm for Data Mining. Journal of Advanced Research in Computer Science and Software Engineering. v. 3, 2013.

BLOM, J. D.. Unmanned Aerial Systems: a historical perspective. Occasional paper 37. Combat Studies Institute Press. US Army Combined Arms Center. Fort Leavenworth, Kansas. September 2010.

BONALDO, M. S.; RIEDO, I. C.; LIMA, A. R. Monitoramento e diagnóstico de doenças foliares da cultura da soja na região COMCAM na safra 2007/2008. Campo Digital, Campo Mourão, v. 4, n. 1, p. 127–136, 2009.

BORTH, M. R.; PISTORI, H.; GONÇALVES, A. B.; FREITAS, U.. Análise da Extração de Atributos do Algoritmo SURF em Espécies de Peixe 6º Encontro Científico de Administração, Economia e Contabilidade (ECAECO), Ponta Porã, MS, Brasil, 2 a 4 de Setembro de 2013.

BREIMAN, L.. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.

BUAINAIN, A. M.; ALVES, E.; SILVEIRA, J. M.; NAVARRO, Z.. O mundo rural no Brasil do século 21 : a formação de um novo padrão agrário e agrícola. Brasília, DF : Embrapa, 2014. 1182 p.

BUCENE, L. C.. Mineração de dados climáticos para previsão de geada e deficiência hídrica para as culturas do café e da cana-de-açúcar para o Estado de São Paulo / Luciana Corpas Bucene, Luiz Henrique Antunes Rodrigues, Carlos Alberto Alves Meira. – Campinas : Embrapa Informática Agropecuária, 2002.

CAMPOS, M. C.. Modernização da Agricultura, Expansão da Soja no Brasil e as Transformações Socioespaciais no Paraná. Revista Geografar. Curitiba, v.6, n.1, p.161-191, jun./2011.

CAPELA NF, GEORGIEVA P. Aplicação de técnicas de aprendizagem automática para classificação de emoções humanas com sinais de EEG. Electrónica e Telecomunicações, vol. 5, nº 4, dezembro 2012

CAVANI, F. A.; SOUSA, R. V.; PORTO, A. J. V.; TRONCO, M. L.. Segmentação e classificação de imagens de laranjeiras utilizando JSEG e perceptron multicamadas. Revista Minerva, v. 3, p. 189-197, 2006.

CECHINEL, C.. Técnicas de Segmentação de Imagens a Cores. Seminário Visão Computacional - CPGCC/UFSC – 2000. Disponível em <<http://www.inf.ufsc.br/~visao/2000/Cores/>>. Acesso em janeiro/2014.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA – CEPEA. Relatório PIBAgroBrasil. Dezembro 2014. Disponível em: <http://www.cepea.esalq.usp.br/comunicacao/Cepea_PIB_BR_dez14.pdf>. Acesso em: 08 abr. 2015.

COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB (2015). Levantamentos de Safra 2014/2015. Disponível em <<http://www.conab.gov.br/conteudos.php?a=1253&#this>>. Acesso em: 06 jun. 2015.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. Levantamentos de Safra 2015/2016. Disponível em <<http://www.conab.gov.br/conteudos.php?a=1252&t=>>>. Acesso em: 02 dezembro de 2016.

DALAL, N.; TRIGGS, B.. Histograms of Oriented Gradients for Human Detection. In CVPR, pages 886-893, 2005

DANDAWATE, Y.; KOKARE, R.. An automated approach for classification of plant diseases towards development of futuristic Decision Support System in Indian perspective. *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, 2015, pp. 794-799.

DECEA - DEPARTAMENTO DE CONTROLE DO ESPAÇO AÉREO. Vãos de VANT (drones). Entenda melhor! 2015. Disponível em: < http://www.decea.gov.br/?i=midia-e-informacao&p=pg_noticia&materia=autorizacoes-para-voos-de-vant-entenda-melhor>. Acesso em: 21 Dez. 2015.

DLUGOSZ, F. L.; ROSOT, N. C.; ROSOT, M. A. D.; OLIVEIRA, Y. M. M. Índice para a avaliação de segmentação de imagens. *Floresta*, Curitiba, v. 39, n. 1, p. 131-143, 2009.

DOMINGUES, M. S.; BERMANN, C.; MANFREDINI, S. A produção de soja no Brasil e sua relação com o desmatamento na Amazônia. *RPGeo* N. 01 (2014). Disponível em < <http://www.periodicos.unir.br/index.php/RPGeo/article/download/887/1125>>

EMBRAPA. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Tecnologias de Produção de Soja - Região Central do Brasil 2012 e 2013. Embrapa Soja Londrina, PR. 2011. Disponível em <http://migre.me/rdzAs> Acesso em Ago/2015.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA SOJA. Tecnologias de produção de soja – região central do Brasil 2014. - Londrina: Embrapa Soja, 2013.

ENGELMANN, A.. A psicologia da Gestalt e a ciência empírica contemporânea. *Psicologia: Teoria e Pesquisa*, 18, 1-16. 2002.

FARIAS, J. R. B.. LIMITAÇÕES CLIMÁTICAS À OBTENÇÃO DE RENDIMENTOS MÁXIMOS DE SOJA. Embrapa Soja, Londrina, PR, Brasil. MERCOSOJA 2011. Quinto congresso de La soja Del Mercosur. Setembro/2011. Argentina.

FEITOSA, R. Q.; COSTA, G. A.; OSTWAL, P. da.; CAZES, T.. Ajuste Automático de Parâmetros de Segmentação. In: Thomas Blaschke; Hermann Johann Heinrich Kux. (Org.). Sensoriamento Remoto e SIG avançados. 2 ed. São Paulo: Oficina de Textos, 2007, v. 1, p. 129-144.

FIGUEIREDO, E. M. P.; RODRIGUES, D. L. ; TAVARES, J. M. . Método de segmentação de objetos em imagens baseado em contornos activos e algoritmo genético. Congreso de Métodos Numéricos en Ingeniería 2009, ISBN: 978-84-96736-66-5, pag. 491, 16 pag., 29 Junho - 2 Julho 2009, Barcelona, Espanha

GALVÃO, D. C. DE O; HRUSCHKA JUNIOR, E. R.. Ponderação do knn utilizando-se de métodos tradicionais de seleção de atributos. Da Vinci , Curitiba, v. 1 , n. 1, p. 115-125, 2004. Disponível em < <http://www.up.edu.br/davinci/pdf08.pdf>>. Acesso em 17/05/2015.

GAVA, R.. O efeito do estresse hídrico na cultura da soja. Tese de Doutorado. ESALQ. Piracicaba, 2014.

GIRARDELLO, A. D.. Um Estudo sobre o Uso de Máquinas de Vetores de Suporte em Problemas de Classificação. CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS. Colegiado de Ciência da Computação. CASCABEL. 2010

GODOY, C.V.; KOGA, L.J.; CANTERI, M.G.. Diagrammatic scale for assessment of soybean rust severity. Fitopatologia Brasileira 31:063-068. 2006.

GÓMEZ-CANDÓN, D.; CASTRO, A.I.D.; LÓPEZ-GRANADOS, F. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. Precis. Agric. 2014-15, 44–56.

GONZALEZ, R. C.; WOODS, R. E. (2010). Processamento Digital de Imagens. 3.ed. São Paulo: Pearson.

GOULART, A. C. P. ; ROESE , A. D.; MELO , C. L. P.. Integração do tratamento de sementes com pulverização de fungicidas para controle da ferrugem Asiática da soja. Biosci. J., Uberlândia, v. 31, n. 3, p. 737-747, May/June. 2015.

GRIGOLLI, J. F. J.. Manejo de Pragas e Doenças na Cultura da Soja. Palestra: Fitossanidade Safra 2015. Fundação MS. Disponível em < <http://migre.me/rdAYC>> Acesso em Ago/2015.

GRIGOLLI, J. F. J.. Pragas da Soja e Seu Controle. Cap. 07 Manejo de Pragas. Pag. 134-155. Novembro/2016. Fundação MS. Disponível em <<http://migre.me/vKG8T>> Acesso em Dez/2016.

GUI, J., HAO, L., ZHANG, Q., BAO, X., 2015. A new method for soybean leaf disease detection based on modified salient regions. International Journal of Multimedia and Ubiquitous Engineering 10 (6), 45 – 52.

HARALICK, R.M.; SHANMUGAM, K.; DINSTEN, I. Textural Features for Image Classification, IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-3, pp. 610-621, 1973.

HEE SUP, Y et al. 2016, Use of Unmanned Aerial Vehicle for Multi-temporal Monitoring of Soybean Vegetation Fraction, Journal of Biosystems Engineering, vol. 2, no. 2, Retrieved from <http://dx.doi.org/10.5307/JBE.2016.41.2.126>

HERWITZ ,S.R; et al.. Imaging from an unmanned aerial vehicle: agricultural surveillance and decision support. Computers and Electronics in Agriculture. Volume 44, Issue 1, July 2004, Pages 49–61.

HONKAVAARA, E.; SAARI, H.; KAIVOSOJA, J.; PÖLÖNEN, I.; HAKALA, T.; LITKEY, P.; MÄKYNEN, J.; PESONEN, L. Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture. Remote Sens. 2013, 5, 5006-5039.

HU, M. K.. Visual Pattern Recognition by Moment Invariants, IRE Transactions on Information Theory, vol. IT-8, pp.179-187, 1962. Artigo disponível em: <http://www.sci.utah.edu/~gerig/CS7960-S2010/handouts/Hu.pdf>, Acesso em: 20/02/2015.

INSTITUTO DE ESTUDOS SÓCIO-AMBIENTAIS – IESA. Um pouco sobre VANT no mundo e no Brasil. Universidade Federal de Goiás. 2013. Disponível em http://www.iesa.ufg.br/uploads/51/original_Folder1_I_Workshop_Vant_UFG.pdf Acesso em 23 de fevereiro de 2014.

INSTITUTO DE PESQUISA ECONOMICA APLICADA – IPEA. Visão Geral da Conjuntura. 8 de julho de 2016. Disponível em <http://www.ipea.gov.br/cartadeconjuntura/index.php/tag/crescimento-economico/> acesso em 29/11/2016.

JAMES, G.; HASTIE, T.; TIBSHIRANI, R. An introduction to statistical learning: with applications in R. London: Ed. Springer, 2013. 429 p.

JESUS, L.F.; SILVA, V.B.; ROCHA, F.G.. Uso de software para detecção de doenças na cultura da soja com o auxílio de um drone autônomo. Computer on the Beach 2015 - Resumo Estendido. Página 429. Disponível em <<https://danielapezzini.files.wordpress.com/2015/04/projeto.pdf>>. Acesso em 05/08/2015.

JORGE, L. A. C.; INAMASU, R. Y. ; CARMO, R. B. . Desenvolvimento de um VANT totalmente configurado para aplicações em Agricultura de Precisão no Brasil. In: XV Simpósio Brasileiro de Sensoriamento Remoto-SBSR, 2011, Curitiba. Anais XV Simpósio Brasileiro de Sensoriamento Remoto-SBSR, 2011.

KERROW, P. M. (2004). Modeling the Draganflyer four-rotor helicopter. International Conference on Robotics & Automation, New Orleans, LA, USA

KLAVDIANOS, P. ; MANSOURI, A.; MERIAUDEAU, F.. Gestalt-inspired features extraction for object category recognition. IEEE International Conference on Image Processing, Sep 2013, Melbourne, Australia. pp.1-5. disponível em <https://hal.archives-ouvertes.fr/hal-00839640/document>> acesso em 23/03/2015.

KOWATA, L. S.; MAY-DE-MIO, L. L.; DALLA-PRIA, M.; SANTOS, H. A. A.. Escala diagramática para avaliar severidade de mildio na soja. Scientia Agraria, Curitiba, v.9, n.1, p.105-110, 2008.

LANTZ, B.. **Machine Learning with R**. Packt, 2º Edição, 2015.

LEÃO , A.C.; ARAÚJO, A.; SOUZA, L.A.C.. Implementação de sistema de gerenciamento de cores para imagens digitais. PUC-Minas, Poços de Caldas, MG, Brazil, cap.3: 61 – 96. 2005.

LORENA, A. C; CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. RITA – Revista de Informática Teórica, v. 14, n. 2, p. 43-67, 2007

LV, J.. An Improved SLIC Superpixels using Reciprocal Nearest Neighbor Clustering. International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 5 (2015), pp. 239-248. Disponível em <<http://dx.doi.org/10.14257/ijsp.2015.8.5.25>> Acesso em 20 de jul. de 2015

MACHADO, B. B.; GONÇALVES, W. N.; PISTORI, H; SILVA, J. A.; SAUEIA, V.; SOUZA, K. P.; TOLEDO, B.; TESSARO, W. Segmentação de Gestos e Camundongos por Subtração de Fundo, Aprendizagem Supervisionada e Watershed In: Avanços em Visão Computacional. Editores: L. A. Pereira, H. Vieira Neto e A. Gonzaga. Editora Omnipax, Curitiba, 2012.

MACHOVINA BRIAN L., FEELEY KENNETH J., MACHOVINA BRETT J. (2016) UAV remote sensing of spatial variation in banana production. *Crop and Pasture Science* , -<http://dx.doi.org/10.1071/CP16135>

MANNING, C. D.; SCHUTZE, H.. Foundations of statistical natural language processing. Massachusetts: MIT Press, 1999. 620p.

MARTIN, D. An empirical approach to grouping and segmentation. PhD thesis, University of California, Berkeley, 2002.

MATOS, A. K. V. de.. Revolução Verde, biotecnologia e tecnologias alternativas. Cadernos da FUCAMP, v.10, n.12, p.1-17/2010.

MEDEIROS, F.A.. Desenvolvimento de um veículo aéreo não tripulado para aplicação em Agricultura de Precisão. Dissertação. Universidade Federal de Santa Maria. Centro de Ciências Rurais. Programa de Pós Graduação em Engenharia Agrícola. Santa Maria-RS. 2007. Disponível em http://cascavel.cpd.ufsm.br/tede/tde_busca/arquivo.php?codArquivo=878. Acesso em 17/07/2015

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO –MAPA. Projeções do agronegócio: Brasil 2015/2016 a 2025/2026, 2016. Disponível em: <Disponível em: <http://www.agricultura.gov.br> >. Acesso em: Dez/2016.

MILANO, D. D.; HONORATO, L. B. Visão computacional. – Universidade Estadual de Campinas– Faculdade de Tecnologia. 2010.

MOLZ, R. F.. Uma Metodologia para o Desenvolvimento de Aplicações de Visão Computacional utilizando um projeto conjunto de Hardware e Software. Porto Alegre: PPGC da UFRGS. 80 f.: il. PhD Thesis – Universidade Federal do Rio Grande do Sul. PhD Program in Computer Science, Porto Alegre, BR-RS., pp. 18-19. 2001.

MORAES, S.A. de Quantificação de doenças de plantas. 2007. Artigo em Hypertexto. Disponível em: <http://www.infobibos.com/Artigos/2007_1/doencas/index.htm>. Acesso em: 22/9/2016.

NEUBERT, P.; PROTZEL, P.. Superpixel benchmark and comparison. Forum Bildverarbeitung, 2012.

NEVES, S. C. M.; PELAES, E. G. Estudo e implementação de técnicas de segmentação de imagens. Revista Virtual de Iniciação Acadêmica da UFPA. Vol 1, No 2, Julho 2001.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A Comparative study of Texture Measures with Classification Based on Feature Distributions, *Pattern Recognition* 29 (1), p. 51-59, 1996.

OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(7), p. 971-987, 2002.

OLIVEIRA, J. A. Classificação de regiões usando atributos de forma e seleção de atributos. – São José dos Campos: INPE, 2005.

OLIVEIRA, J. A.; DUTRA, L. V.; RENNÓ, C. D.; SANTOS, P. S.. Extração de Atributos de Forma para Classificação de Imagens de Alta Resolução do Satélite HRC/CBERS-2B. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil. Abril/2009. Disponível em <<http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.17.22.33/doc/7015-7022.pdf>> Acesso em Ago/2015.

OLIVEIRA, L. F. et al.. Extração de Características de Forma utilizando matriz de co-ocorrência e Atributos de Haralick. Workshop de Visão computacional 2012. Disponível em http://iris.sel.eesc.usp.br/wvc/Anais_WVC2012/pdf/97967.pdf. Acesso em Dez/2016.

OSÓRIO, F. S.. INSS: A hybrid system for constructive machine learning. *Neurocomputing* 28 (1999) 191-205. Elsevier Press 1999.

PADILHA, A. J. M. N.. Processamento e Análise de Imagem. Fichero de aula – Capítulo 4. DEEC - Departamento de Engenharia Electrotécnica e de Computadores. Faculdade de Engenharia da Universidade do Porto. Porto – Portugal. 2005. Disponível em <http://paginas.fe.up.pt/~padilha/PAI/ficheiros/> Acesso em janeiro/2014.

PEDRINI, H.. Processamento e Análise de Imagens. Instituto de Computação UNICAMP. 2014. Disponível em <http://www.ic.unicamp.br/~helio/disciplinas/MO445/aula_cores.pdf> Acesso em Ago/2015.

PEÑA-BARRAGAN JM, KELLY M, DE CASTRO AI, LOPEZ-GRANADOS F.. Object-based approach for crop row characterization in UAV images for site-specific weed management. In Queiroz-Feitosa et al., editors. 4th International Conference on Geographic Object-Based Image Analysis (GEOBIA 2012), Rio de Janeiro, Brazil: 426-430.

PEÑA, J. M.; TORRES-SÁNCHEZ, J.; CASTRO, A. I.; KELLY, M.; LÓPEZ-GRANADOS, F.. Weed Mapping in Early-Season Maize Fields Using Object-Based

Analysis of Unmanned Aerial Vehicle (UAV) Images. PLOS ONE. October 2013. Volume 8. Issue 10. 77151. Disponível em < <http://migre.me/rdgd3>> consulta em 10/08/2015.

PERIN, B. G.. Assistência técnica na produção de soja e milho em querência – mato grosso. Centro de Ciências Agrárias, da Universidade Federal de Santa Catarina. FLORIANÓPOLIS – SC. Julho/2013.

PERROCA, M.G.; GAIDZINKI, R. R.. Assessing the interrater reliability of an instrument for classifying patients - kappa quotient. Rev Esc Enferm USP. [Internet]. 2003. [acesso 23 abr 2013];37(1):72-80. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342003000100009

PINAZZA, G. G, M..Análise da competitividade da cadeia produtiva da soja no Brasil vis-à-vis os demais países exportadores sul-americanos. Dissertação de Mestrado. Universidade Federal de São Carlos. UFSCar, 2008.

PINTO, T. W.. Segmentação de imagens digitais combinando watershed e corte normalizado em grafos. Limeira, SP : [s.n.], 2014.

PIRES, R. D. L. et al.. Local descriptors for soybean disease recognition. Computers and Electronics in Agriculture. Volume 125, July 2016, Pages 48–55.

PRATI, R.C.; BATISTA, G.E.A.P.A.; MONARD, M.C. (2008). Curvas ROC para avaliação de classificadores. IEEE América Latina, Vol. 6, N. 2.

PRIMICERIO, J., Di GENNARO, S. F., FIORILLO, E., GENESIO, L., LUGATO, E., MATESE, A. and VACCARI, F. P. 2012. A flexible unmanned aerial vehicle for precision agriculture. Precision Agriculture 13 517–523.

QUINLAN, J. R. (1986). Induction of decision trees. Machine Learning, 1(1):81-106.

QUINLAN, J. R. (1993). C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

QUINTA, L. N. B. Visão Computacional aplicada na classificação de grãos de pólen. Dissertação de Mestrado em Biotecnologia. Programa de Pós Graduação em Biotecnologia - Universidade Católica Dom Bosco. Janeiro/2013.

REN, X.; MALIK, J.. Learning a classification model for segmentation. IEEE ICCV, pp. 10–17, 2003.

REZENDE, S. O.. Sistemas inteligentes: fundamentos e aplicações. Barueri, SP: Manole, 2003.

RICHETTI, A.; ROESE, A. D. Custo do controle químico da ferrugem asiática da soja para a Safra 2010/11. Comunicado técnico, Dourados: Embrapa Agropecuária Oeste, n. 161, p. 1-6, 2011.

RICHETTI, A. 2015. Viabilidade econômica da cultura da soja na safra 2015/2016, em Mato Grosso do Sul. Comunicado Técnico n o 202, Dourados, MS, Julho/2015. 13p.

ROTHER, P. R.; KSHIRSAGAR, R. V.. Cotton leaf disease identification using pattern recognition techniques, *2015 International Conference on Pervasive Computing (ICPC)*, Pune, 2015, pp. 1-6. doi: 10.1109/PERVASIVE.2015.7086983

RUDEK, M.; COELHO, L. dos S.; CANGIOLIERI, O.. Visão Computacional Aplicada a Sistemas Produtivos: Fundamentos e Estudo de Caso. Disponível em: <http://www.abepro.org.br/biblioteca/ENEGEP2001_TR10_0917.pdf>. Acesso em: 12 set. 2008.

RUMPF, T.; MAHLEIN, A.K.; STEINER, U.; OERKE, E.C.; DEHNE, H.W.; PLÜMER, L.. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Comput. Electron. Agric.*, 74 (1) (2010), pp. 91–99.

SALDANHA, M.; FREITAS, C.. Segmentação de Imagens Digitais: Uma Revisão. 2008. Disponível em: <<http://www.lac.inpe.br/cap/arquivos/pdf/P19.pdf>>. Acesso em janeiro de 2014.

SARATH, D. S.. Medição automática de efeito de herbivoria utilizando técnicas de segmentação e aprendizagem supervisionada. TCC. Engenharia de Computação, UCDB, 2013-2014. Campo Grande –MS.

SCHWARTZ, W.R.; PEDRINI, H.. Método para classificação de imagens baseada em matrizes de coocorrência utilizando características de textura. III Colóquio Brasileiro de Ciências Geodésicas, Curitiba-PR, Brasil, 2003.

SILVA, L. M. Uma aplicação de Árvores de Decisão, Redes Neurais e KNN para a Identificação de Modelos ARMA não Sazonais e Sazonais. Tese de Doutorado em Engenharia Elétrica – Pontifícia Universidade Católica do Rio de Janeiro. 2005.

SILVA, N. Identificação de trajetórias espaço-temporais de movimentos em vídeo. Dissertação de mestrado Universidade Federal de Uberlândia. 2010. Disponível em <http://repositorio.ufu.br/handle/123456789/670>. Acesso em 17/03/2015

SILVA NETO, M.. Como utilizar imagens aéreas na Agricultura de Precisão? 13 de Setembro de 2013. Disponível em <http://migre.me/rcp58> Acesso em 26/11/2013.

SOARES, R.M.; GODOY, C.V.; OLIVEIRA, M.C.N. Escala diagramática para avaliação da severidade da mancha alvo da soja. Tropical Plant Pathology, v.34, n.5, p.333-338, 2009.

STOPPA , M. H.. Matemática no ambiente digital: imagens e TV. Synergismus scyentifica UTFPR. XIII ERMAC. Mini Curso. Departamento de Matemática – Campus Catalão – UFG. Universidade Tecnológica Federal do Paraná. Pato Branco. Setembro de 2009.

STUTZ, D.; HERMANS, A.; LEIBE, B.. Superpixel Segmentation using Depth Information. Bachelor thesis, RWTH Aachen University, Aachen, Germany, 2014.v

SZELISKI, R..Computer Vision: Algorithms and Applications. Springer, 2010.

TAKATSUI, F.. Sistema CIE LAB: análise computacional de fotografias. Dissertação (Mestrado) – Universidade Estadual Paulista, Faculdade de Odontologia. Araraquara: 2011. 100 f.

THEIS, T.; TOMKIN, J. (Orgs.). Sustainability: A Comprehensive Foundation. Rice University, Houston, Texas, 2012.

TORRES-SÁNCHEZ, J.; LÓPEZ-GRANADOS, F.; SERRANO, N.; ARQUERO, O.; PEÑA, J. M.. High-Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. PLOS ONE | DOI:10.1371/journal.pone.0130479 June 24, 2015.

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. Oilseeds: World Markets and Trade. Foreign Agricultural Service. Dez, 2016. Disponível em <<https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf>>

VASCONCELOS, B.R.; GONÇALVES, A.B.; VILPOUX, O.F.; CEREDA, M.P. Georeferencing Methodology For Honey Tracking. In: INTERNATIONAL APICULTURAL CONGRESS, XLII, 2011, Buenos Aires. Anais...Buenos Aires, APIMONDIA, 2011.

YORINORI, J. T.; CHARCHAR, M.J.D'A.; NASSER; L.C.B.; HENNING, A.A.. Doenças da soja e seu controle. In: ARANTES, N.E.; SOUZA, P.I. de M. de., ed. Cultura de soja nos cerrados. Piracicaba: POTAFOS, 1993. p.333-397.

YORINORI, J.T. Soja (*Glycine max* (L.) Merrill) - controle de doenças. In: Ribeiro do Vale, F.X. & Zambolim, L. (Eds.) Controle de doenças de plantas. V. 2. Viçosa, UFV/MAA, 1997. Cap. 21, pp.953-1023.

YORINORI, J. T.; GODOY, C. V.; PAIVA, W. M.; FREDERICK, R. D.; COSTAMILAN, L. N.; BERTAGNOLLI, P. F.; NUNES JUNIOR, J.. Evolução da ferrugem da soja (*Phakopsora pachyrhizi*) no Brasil, de 2001 a 2003. Fitopatologia Brasileira, Brasília, v. 28, p. S210, ago. 2003.

YORINORI, J.T.; LAZZAROTTO, J.J. Situação da ferrugem asiática da soja no Brasil e na América do Sul. Londrina: Embrapa Soja, 2004. 27p. (Embrapa Soja. Documentos, 236. Disponível em: <<http://www.cnpso.embrapa.br/download/alerta/documentos236.pdf>> Acesso em: Jul. 2015.

ZARBIN, P. H. G.; RODRIGUES, M. A. C. M.; LIMA, E. R.. Feromônios de insetos: tecnologia e desafios para uma agricultura competitiva no Brasil. Quím. Nova [online]. 2009, vol.32, n.3, pp. 722-731.

WILLKOMM, M.; BOLTEN, A.; BAETH, G. (2016): Non-destructive monitoring of rice by hyperspectral in-field spectrometry and UAV-based remote sensing: Case study of field-grown rice in North Rhine-Westphalia, Germany. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B1, 1071-1077, doi:10.5194/isprs-archives-XLI-B1-1071-2016