Universidade
Estadual de Londrina

WILLIAM HITOSHI SUMIDA

DETECÇÃO DA FERRUGEM ASIÁTICA DA SOJA POR MEIO DE IMAGENS CAPTURADAS POR DRONES

LONDRINA–PR

2018

WILLIAM HITOSHI SUMIDA

DETECÇÃO DA FERRUGEM ASIÁTICA DA SOJA POR MEIO DE IMAGENS CAPTURADAS POR DRONES

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Alan Salvany Felinto
Coorientador: Msc. André Luis da Silva

LONDRINA–PR

2018

William Hitoshi Sumida
Detecção da ferrugem asiática da soja por meio de imagens capturadas por drones/ William Hitoshi Sumida. – Londrina–PR, 2018-
41 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Alan Salvany Felinto

– Universidade Estadual de Londrina, 2018.

1. Produtividade. 2. Produtos Fitossanitários. 3. Severidade 4. Análise de Imagens I. Alan Salvany Felinto. II. Universidade Estadual de Londrina. III. Detecção da Ferrugem Asiática da Soja Por Meio de Imagens Capturadas Por Drones

CDU 02:141:005.7

WILLIAM HITOSHI SUMIDA

DETECÇÃO DA FERRUGEM ASIÁTICA DA SOJA POR MEIO DE IMAGENS CAPTURADAS POR DRONES

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA

Prof. Dr. Alan Salvany Felinto
Universidade Estadual de Londrina
Orientador

Prof. Dr. Marcelo Giovanetti Canteri
Universidade Estadual de Londrina

Prof. Dr. Fabio Sakuray
Universidade Estadual de Londrina

Londrina–PR, 21 de Dezembro de 2018

Este trabalho é dedicado aos meus pais,
Vilma Yukie Hirata Sumida e Raul Shigueru Sumida.
Pois sem eles eu não poderia afirmar que ganhei na loteria do Universo

AGRADECIMENTOS

Agradeço aos meus pais e minha família literalmente por tudo. Agradeço ao Prof. Dr. Alan Salvany Felinto pela oportunidade de trabalhar juntos e contribuir com a comunidade por meio deste trabalho. Agradeço também aos companheiros da equipe Fitovision, André Luis, Marcelo Canteri e Maycon Gabriel por contribuirem tanto para meu desenvolvimento profissional e pessoal. E não menos importante, agradeço aos meus amigos por sempre estarem por perto me ajudando a criar histórias para serem contadas nos próximos churrascos! Não citarei nomes aqui, porque além da lista ser enorme, estou nos acréscimos do segundo tempo e ainda preciso imprimir este trabalho.

“Legacy is greater than currency.“
(Gary Vaynerchuck)

SUMIDA, W. H.. **Detecção da ferrugem asiática da soja por meio de imagens capturadas por drones**. 41 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina–PR, 2018.

RESUMO

Atualmente existem muitos problemas econômicos na cultura da soja causados por doenças. Já foram relatados danos na produtividade de 30% a 80% causados pela ferrugem asiática. Além da perda de produtividade, há o dano econômico em função do desperdício de produtos fitossanitários, pois sem as informações corretas o produtor aplica fungicida de forma imprecisa e em uma frequência desnecessária. Em 2015 a média do custo de defensivos agrícolas por hectare no sudeste foi de R$ 1.097,76. A avaliação da severidade da ferrugem asiática é um trabalho manual e lento para ter como resultado a porcentagem de área foliar afetada pelo fungo *Phakopsora phachyhizi*. Este trabalho de conclusão de curso tem como objetivo encontrar métodos computacionais precisos para prever a severidade da ferrugem asiática a partir de imagens capturadas por veículos aéreos não tripulados. O algoritmo mais eficaz na avaliação da severidade da ferrugem asiática foi o Support Vector Regression, com um coeficiente de correlação de 87,09 % e média do erro absoluto de 7,4871.

Palavras-chave: Produtividade. Produtos Fitossanitários. Severidade. Análise de Imagens

SUMIDA, W. H.. **Detection of soybean Asian rust by means of images captured by drones**. 41 p. Final Project (Bachelor of Science in Computer Science) – State University of Londrina, Londrina–PR, 2018.

ABSTRACT

Nowadays there are many problems in the soy culture caused by diseases. There were reported damage in the productivity from 30% to 80% caused by the Asian soybean rust. There are not only productivity problems caused by the soybean rust, there are economic damage based on waste of phytosanitary products caused by lack of information. In 2015 the average cost of agricultural pesticides per hectare in the southeast was R$ 1,097.76. The soybean rust evaluation is a manual and laborous task to have as a result, the percentage of foliar area affected by a fungus called *Phakopsora phachyhizi*. This paperwork has as an objective, research the most precise computational methods to correlate the asian soybean rust with technical features of images taken from unmanned aerial vehicles. The most effective algorithm in assessing the severity of Asian soybean rust was Support Vector Regression, with a correlation coefficient of 87.09 % and mean absolute error of 7.4871.

Keywords: Yield. Phytosanitary Products. Severity. Image Analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Trifólio da soja afetado pela ferrugem. [Fonte: Embrapa Agropecuária Oeste] 26
Figura 2 – Coeficiente de correlação de Pearsion. [Fonte: statistics.laerd.com] . . . 27
Figura 3 – Fluxograma. [Fonte: Autor] 29
Figura 4 – Ensaio de ferrugem asiática na fazenda escola UEL. [Fonte: André Luis da Silva 30
Figura 5 – Ensaio de fungicidas para soja e seus blocos. [Fonte: Autor] 30
Figura 6 – Histograma de severidade das 76 amostras coletadas. [Fonte: Autor] . . 37

LISTA DE TABELAS

Tabela 1 – Tratamentos (Trat.) e seus respectivos índices de severidade de ferrugem asiática da soja (Sev. %). 31
Tabela 2 – Ranking dos atributos 32
Tabela 3 – Matriz de Covariância. 33
Tabela 4 – Resultados com atributos selecionados a partir do Rank. 35
Tabela 5 – Resultados com conjunto de atributos selecionados a partir do CfsSubsetEval. 35
Tabela 6 – Resultados com conjunto de atributos selecionados a partir da matriz de covariância. 36
Tabela 7 – Severidades do experimento, previsões do Support Vector Regression (SVR), Regressão Linear (RL) e seus respectivos erros absolutos. . . . 36

LISTA DE ABREVIATURAS E SIGLAS

GLI	Green Leaf Index
NGRDI	Green-Red Vegetation Index
SVR	Support Vector Regression
RL	Regressão Linear

SUMÁRIO

1 INTRODUÇÃO 23

2 FUNDAMENTAÇÃO TEÓRICA 25
2.1 Cultura da Soja 25
2.2 Ferrugem Asiática da Soja 25
2.3 Fitopatometria 26
2.3.1 Escala diagramática 26
2.3.2 Radiômetro 26
2.4 Métricas 27
2.4.1 Coeficiente de Correlação de Pearson 27
2.5 Análise da Imagem 27
2.5.1 Imagem Digital 27
2.5.2 Descritores da imagem 28
2.5.3 SVR - Support Vector Regression 28
2.5.4 Weka 3: Data Mining Software in Java 28

3 MATERIAIS E MÉTODOS 29
3.1 Aquisição dos dados e imagens 29
3.2 Descritores das Imagens 31
3.3 Algoritmos de Regressão 32
3.3.1 Seleção de Atributos 32

4 CORRELAÇÃO DOS DADOS E RESULTADOS 35

5 CONCLUSÃO 39

REFERÊNCIAS 41

1 INTRODUÇÃO

Atualmente o Brasil é o maior produtor de soja do mundo, contando com 34,74% da participação mundial. Na safra 2017/2018 foram colhidos 116,996 milhões de toneladas do grão (CONAB 2018 [1]).

Apesar do número impressionar, a produtividade da soja pode aumentar com uma melhor aplicação de fungicidas nas plantações. Isso pode ser obtido com uma análise rápida da lavoura e ter um resultado preciso para a aplicação de doses sob medida nos focos de ferrugem.

No ramo de fitopatologia (estudo das doenças das plantas) é muito comum serem realizadas análises de campo para ter como resultado a severidade da ferrugem asiática, um fungo chamado *Phakopsora phachyhizi*, o qual ataca a soja.

O processo de análise consiste em recolher amostras de trifólios da soja em vários pontos da plantação e estimar uma porcentagem da área afetada da folha. É obtido um valor por meio de cálculos estatísticos que varia entre 0 e 100. A análise também pode ser realizada por meio de um radiômetro, o qual emite radiação ativa em dois comprimentos de onda, vermelho (660 nm) e infravermelho (780 nm). O equipamento realiza alguns cálculos e retorna o índice de vegetação por diferença normalizada (NDVI).

Tais métodos de análise são eficientes, pois é possível obter um resultado preciso da severidade da ferrugem asiática. Porém ambos são processos demorados, cansativos e muitas vezes caros, pois para acelerar o processo são contratadas equipes de 3 ou mais pessoas para realizar o serviço. Um outro ponto negativo também é o preço de um radiômetro, muitas vezes podendo custar até R$ 40.000,00.

Diante disso, este trabalho de conclusão de curso tem como objetivo estudar e apresentar os melhores soluções para assimilar os dados da severidade da soja obtidos no campo por meio da escala diagramática e radiômetro a partir de características das imagens RGB capturadas por veículos aéreos não tripulados.

O objetivo deste trabalho foi calcular a severidade da ferrugem asiática da soja a partir de imagens RGB obtidas por veículos aéreos não tripulados com dados de campo coletados por meio da escala diagramática.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordadas as fundamentações teóricas necessárias para o entendimento do trabalho. Inicialmente serão apresentados alguns conceitos sobre a cultura da soja (*Glycine max*) para um melhor entendimento sobre o contexto do problema. E em seguida serão descritos métodos computacionais.

2.1 Cultura da Soja

A soja (*Glycine max*) atualmente é a principal oleaginosa produzida e consumida a nível mundial. Na safra de 2017/2018 sua produção mundial foi de 336,699 milhões de toneladas. O Brasil é o segundo maior produtor do mundo, ficando atrás apenas do Estados Unidos, contando com 34,74% da produção mundial (116,996 milhões de toneladas)[1].

Apesar dos números impressionarem, o cultivo da soja possui obstáculos que atrapalham a produtividade da cultura. O obstáculo abordado neste trabalho será uma das doenças conhecida como ferrugem asiática da soja.

2.2 Ferrugem Asiática da Soja

A ferrugem asiática da soja é uma das doenças mais severas que incide na cultura da soja, ocorre no país inteiro durante o ano todo, pode causar perdas de produtividade de até 90% [2]. Causada pelo fungo *Phakopsora pachyrhizi*, foi constatada pela primeira vez no Brasil em 2001 [3].

O controle do fungo é de grande importância econômica levando em conta as perdas de produtividade causadas pela ferrugem. Na safra de 2003/2004 foram perdidos 4,6 milhões de toneladas de soja (US$ 1,22 bilhões), com custo de controle de US$ 2,08 bilhões. Durante os anos, notou-se uma melhora nas práticas de controle, na safra de 2011/2012, as perdas foram de 365,5 mil toneladas (US$ 191.6 milhões), com custo de controle de US$1,73 bilhões [4].

Em um período de 8 anos a perda de produtividade no Brasil foi otimizada em 92,05% com um gasto de 16,82% a menos em controle. Apesar do método apresentar melhora significativa, a produtividade e custo com fungicidas podem ser aperfeiçoadas com uma aplicação mais eficiente de produto nas plantações. Isso pode ser obtido com uma análise rápida da lavoura e ter um resultado preciso para a aplicação de doses sob medida nos focos de ferrugem.

2.3 Fitopatometria

O estudo da quantificação de doenças de plantas é chamado de fitopatometria e tem como objetivo identificar as reações que a planta expressa visivelmente ao ser atacada por um patógeno.

2.3.1 Escala diagramática

A escala diagramática [5] é uma forma de avaliar a severidade da ferrugem asiática da soja. O processo é realizado manualmente, no qual o agrônomo recolhe um trifólio da soja (Figura 1) e visualmente verifica a porcentagem de área afetada pelo fungo nas folhas. Na qual a severidade da ferrugem asiática é dada diretamente proporcional pela porcentagem de área, afetada pelo fungo.

Figura 1 – Trifólio da soja afetado pela ferrugem. [Fonte: Embrapa Agropecuária Oeste]

2.3.2 Radiômetro

O radiômetro é um equipamento o qual emite uma radiação ativa em dois comprimentos de onda, vermelho (660 nm) e infravermelho (780 nm) e coleta a refletância das folhas independente das condições de luminosidade, gerando o índice de vegetação por diferença normalizada (NDVI) [6], dado pela equação 2.1.

$$NDVI = \frac{(Infravermelho - Vermelho)}{(Infravermelho + Vermelho)} \tag{2.1}$$

2.4 Métricas

2.4.1 Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson retorna um valor r entre -1 e 1, logo indicando se a correlação foi positiva, negativa ou se não teve correlação (Figura 2).

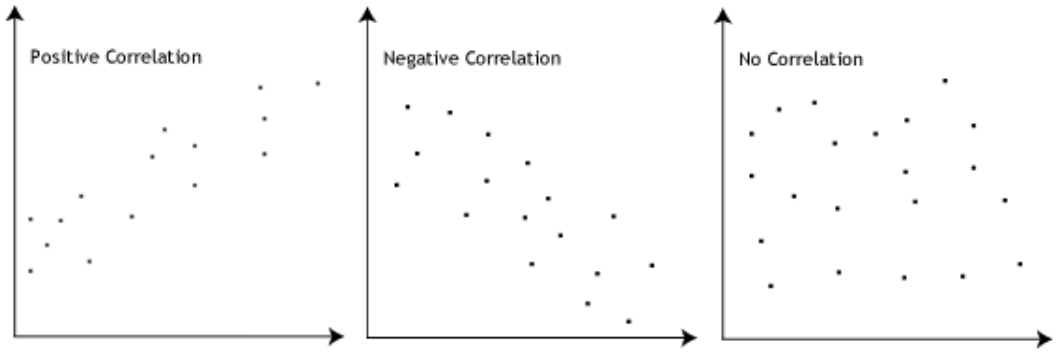

Figura 2 – Coeficiente de correlação de Pearsion. [Fonte: statistics.laerd.com]

Considerando que o valor de uma das variáveis aumenta e $r > 0$, significa que a outra variável possui tendência de aumentar, ou seja, a correlação foi positiva. Se $r = 0$, não houve correlação e caso o valor de uma das variáveis diminui e $r < 0$, significa que a outra variável possui tendência de diminuir, ou seja, a correlação foi negativa.

De acordo com Evans [7], podemos interpretar o valor do módulo de r da seguinte forma:

- Muito fraco: 0 - 0,19
- Fraco: 0,20 - 0,39
- Moderado: 0,40 - 0,59
- Forte: 0,60 - 0,79
- Muito forte: 0,80 - 1

2.5 Análise da Imagem

2.5.1 Imagem Digital

Uma imagem digital pode ser definida por uma função f(x,y), na qual x e y são variáveis espaciais finitas e discretas. Já a função f é um valor que representa uma cor.

Para representar as tonalidades das imagens coloridas, geralmente são utilizados 3 valores associados a um espaço de cor específico. Neste trabalho serão utilizados 2 espaços

de cor, o RGB e HSV. O primeiro é um modelo de cor aditivo no qual as cores vermelho, verde e azul são combinadas para produzir um espectro cromático vasto. Já o segundo representa as cores de uma imagem por meio da matiz, saturação e valor (brilho) [8].

2.5.2 Descritores da imagem

Descritores são informações essenciais extraídas da imagem as quais serão utilizadas para criar um modelo de regressão.

2.5.3 SVR - Support Vector Regression

A SVR (SVR - Support Vector Regression) [9] é uma técnica de regressão na mineração de dados a qual utiliza machine learning. É um algoritmo de análise de dados para escavar grandes volumes de dados e encontrar padrões que poderiam, por outros métodos, permanecer desconhecidos.

O SVM apresenta robustez e precisão mesmo com poucos exemplos e em uma base de dados com alta dimensão (Quantidade elevada de variáveis/atributos).

2.5.4 Weka 3: Data Mining Software in Java

Weka [10] é um software abrangendo um conjunto de algoritmos referentes à machine learning para data mining. Os algoritmos podem ser aplicados diretamente em um conjunto de dados ou utilizados diretamente no código Java. O Weka possui ferramentas para pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização.

É frequentemente utilizado para testar algoritmos e verificar quais são eficazes para assimilar dados.

3 MATERIAIS E MÉTODOS

Neste capítulo serão abordadas as etapas (Figura 3) que foram necessárias para atingir o objetivo proposto neste trabalho.

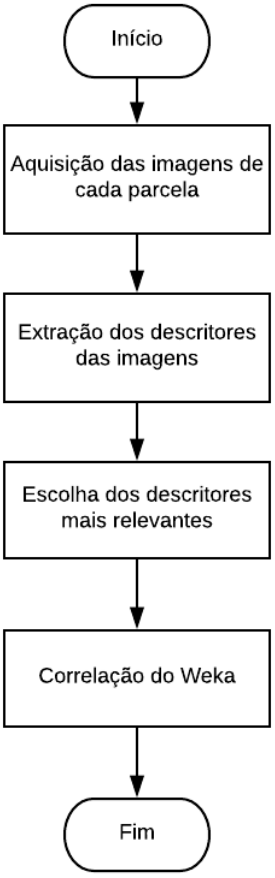

Figura 3 – Fluxograma. [Fonte: Autor]

3.1 Aquisição dos dados e imagens

As imagens foram obtidas a partir de um ensaio de fungicidas para a cultura da soja na Fazenda Escola UEL, no município de Londrina, Paraná. O delineamento experimental utilizado foi o de blocos casualizados, possuindo 19 tratamentos e 4 repetições, totalizando 76 amostras. Cada bloco media 3 metros de largura e 5,5 metros de comprimento.

A coleta das imagens foi realizada junto ao Departamento de Ciências Agrárias da UEL por meio do Phantom 4 PRO, um veículo aéreo não tripulado com a câmera padrão utilizada pelo mesmo. A imagem original pode ser visualizada na Figura 4.

Figura 4 – Ensaio de ferrugem asiática na fazenda escola UEL. [Fonte: André Luis da Silva

Para realizar a análise do experimento, foi necessário marcar os blocos (Figura 5), utilizando um software de edição de imagens, de acordo com a configuração do ensaio. Cada bloco media 3 m de largura e 5,5 m de comprimento. Após a identificação dos blocos (são as regiões quadriláteras de tamanhos semelhantes delimitadas pelas linhas brancas), foi realizado o recorte das mesmas e separadas em arquivos de imagem no formato PNG.

Figura 5 – Ensaio de fungicidas para soja e seus blocos. [Fonte: Autor]

Para realizar a avaliação, consideraram-se as linhas centrais de cada bloco, evitando

o efeito negativo das bordaduras. A severidade de ferrugem asiática da soja de cada amostra foi observada por meio da escala diagramática [5]. A avaliação de severidade da ferrugem foi dada em 4 pontos de cada bloco, cujo em cada ponto foram analisadas a parte superior, médio e inferior da planta. Foram constatadas porcentagens de severidade entre 40% e 100%, como pode ser observada na Tabela 1.

Tabela 1 – Tratamentos (Trat.) e seus respectivos índices de severidade de ferrugem asiática da soja (Sev. %).

Trat.	Sev. 1 (%)	Sev. 2 (%)	Sev. 3 (%)	Sev. 4 (%)
01	100,0	100,0	100,0	100,0
02	85,8	80,8	88,3	82,9
03	79,2	86,7	84,2	85,0
04	89,6	86,7	61,3	91,3
05	72,1	80,8	82,9	80,4
06	76,3	54,4	52,1	55,8
07	87,1	82,5	52,7	54,6
08	78,3	56,7	40,0	53,8
09	69,3	52,1	58,3	52,9
10	73,1	80,0	48,8	40,0
11	75,0	68,8	51,7	66,3
12	72,5	48,3	51,3	41,8
13	80,8	52,4	52,9	52,9
14	81,7	54,2	59,6	84,2
15	76,3	55,8	50,6	44,6
16	89,2	86,7	56,7	56,3
17	75,8	80,0	57,1	55,8
18	79,6	56,7	56,3	66,7
19	69,6	47,3	55,2	54,2

3.2 Descritores das Imagens

Para a extração dos dados, foi utilizado o OpenCV, uma biblioteca open source para visão computacional e a linguagem Python na versão 3.6 para o processamento das imagens e implementação do software.

Utilizamos como descritores das imagens a média dos canais RGB e HSV de cada bloco, imagem na escala de cinza, |R-B|, |B-G|, |G-R|, NGRDI e GLI.

Os descritores NGRDI [11] e GLI [12], dados pelas equações 3.1 e 3.2 respectivamente, são modelos matemáticos baseados em duas ou mais bandas que foram desenvolvidos para realçar a cobertura vegetal de uma área, assim como o NDVI [6]. Dentre os índices de vegetação foram as escolhidas, pois seguem a mesma linha deste trabalho, ou seja, utilizam apenas os canais RGB em suas respectivas equações.

$$NGRDI = \frac{(Verde - Vermelho)}{(Verde + Vermelho)} \tag{3.1}$$

$$GLI = \frac{(2 * Verde - Vermelho - Azul)}{(2 * Verde + Vermelho + Azul)} \tag{3.2}$$

Para cada bloco do experimento os descritores citados acima foram extraídos e salvos em um arquivo ARFF associados às suas respectivas taxas de severidade. O mesmo será utilizado como entrada de dados para o software Weka.

3.3 Algoritmos de Regressão

Após extrairmos os dados do experimento (Seção 3.2), o software Weka proporciona um ambiente para testar e comparar diversos algoritmos, assim como analisar a relevância dos descritores.

3.3.1 Seleção de Atributos

Antes de iniciar os testes dos algoritmos, é interessante descobrir quais são as melhores relações entre os descritores. Este processo é importante para identificar atributos irrelevantes aos métodos de regressão que serão utilizados posteriormente, e como resultado conseguimos aumentar o desempenho computacional dos mesmos [13].

Neste trabalho foram utilizados dois métodos disponíveis no Weka e uma matriz de covariância, para identificar quais os melhores conjuntos de atributos. O CorrelationAttributeEval cria um ranking de correlação dos atributos (Tabela 2). Já o CfsSubsetEval avalia o valor de um subconjunto de atributos considerando a capacidade preditiva individual de cada um, juntamente com o grau de redundância entre eles e seleciona um subconjunto de atributos que podem apresentar correlação forte [14].

Tabela 2 – Ranking dos atributos

Rank	Atributo	Descrição
1	B	Média do canal Blue
2	|G-R|	Módulo dos canais Green e Red
3	V	Média do canal Value
4	R	Média do canal Red
5	RGB	Média dos canais RGB
6	|R-B|	Módulo dos canais Red e Blue
7	G	Média do canal Green
8	|B-G|	Módulo dos canais Blue e Green
9	S	Média do canal Saturation
10	GLI	Green Leaf Index
11	H	Média do canal Hue
12	NGRDI	Green-Red Vegetation Index

Os atributos selecionados pelo algoritmo CfsSubsetEval foram NGRDI, GLI, média do canal Blue e módulo dos canais Green e Red.

Para realizarmos a última seleção de atributos, utilizamos a matriz de covariância e foi possível observar que alguns atributos possuem correlação alta entre si (Tabela 3) . Muitas vezes a correlação alta entre atributos indica que sua combinação pode prejudicar o modelo. Ao interpretar a matriz por tentativa e erro, chegamos no seguinte subconjunto de atributos: GLI, |R-B|, |G-R| e média dos canais R, B, RGB, S e V.

Tabela 3 – Matriz de Covariância.

	NGRDI	GLI	|R-B|	|G-R|	|B-G|	R	G	B	RGB	S	V	H
NGRDI	1	0.98	-0.76	-1	-0.21	-0.84	-0.56	-0.88	-0.78	-0.11	-0.82	1
GLI	0.98	1	-0.86	-0.97	-0.39	-0.91	-0.69	-0.86	-0.86	-0.3	-0.89	0.98
|R-B|	-0.76	-0.86	1	0.75	0.8	0.98	0.95	0.76	0.98	0.64	0.97	-0.77
|G-R|	-1	-0.97	0.75	1	0.2	0.83	0.56	0.9	0.78	0.09	0.82	-0.99
|B-G|	-0.21	-0.39	0.8	0.2	1	0.69	0.9	0.31	0.73	0.87	0.69	-0.23
R	-0.84	-0.91	0.98	0.83	0.69	1	0.92	0.87	0.99	0.49	1	-0.84
G	-0.56	-0.69	0.95	0.56	0.9	0.92	1	0.69	0.95	0.67	0.93	-0.58
B	-0.88	-0.86	0.76	0.9	0.31	0.87	0.69	1	0.87	0.03	0.88	-0.88
RGB	-0.78	-0.86	0.98	0.78	0.73	0.99	0.95	0.87	1	0.5	0.99	-0.79
S	-0.11	-0.3	0.64	0.09	0.87	0.49	0.67	0.03	0.5	1	0.48	-0.13
V	-0.82	-0.89	0.97	0.82	0.69	1	0.93	0.88	0.99	0.48	1	-0.83
H	1	0.98	-0.77	-0.99	-0.23	-0.84	-0.58	-0.88	-0.79	-0.13	-0.83	1

4 CORRELAÇÃO DOS DADOS E RESULTADOS

Para correlacionar os descritores obtidos com os valores de severidade da ferrugem asiática da soja, foi utilizado o software Weka. Com todos os descritores salvos em um arquivo com extensão arff, a tarefa de executar e comparar diversos algoritmos de regressão foi simplificada em selecionar e configurá-los por meio da funcionalidade experimenter.

Dentre os 9 algoritmos testados, foram selecionados 3 que tiveram as melhores correlações. São eles o Support Vector Regression [9], Linear Regression e Random Forest. O espaço amostral é formado por 76 blocos, sendo 80% (60 amostras) utlizadas para o treinamento dos algoritmos e 20% (16 amostras) para teste.

Utilizando o ranking gerado pelo CorrelationAttributeEval, foi sugerido pelo Weka que fossem removidos os 4 piores atributos (GLI e NGRDI, e a média dos canais Saturation e Hue). Utilizando os 8 melhores atributos, conforme Tabela 1, foi constatado que os 3 algoritmos utilizados neste trabalho atingiram um coeficiente de correlação alto. O coeficiente de correlação e a média do erro absoluto podem ser observados na Tabela 4.

Tabela 4 – Resultados com atributos selecionados a partir do Rank.

	Coeficiente de correlação(%)	Média do erro absoluto
Linear Regression	85,16	10,1547
Support Vector Regression	83,3	9,0318
Random Forest	81,22	10,6269

Utilizando o subconjunto de descritores gerado pelo CfsSubsetEval (NGRDI, GLI, média do canal Blue e módulo dos canais Green e Red), podemos observar que a Regressão Linear e o SVR atingiram coeficientes de correlação e média de erro absoluto semelhantes aos resultados anteriores (Tabela 4). Já o Random Forest se demonstrou inferior ao trabalhar com um conjunto de atributos reduzido, pois apresentou correlação baixa em função dos resultados obtidos anteriormente. O coeficiente de correlação e a média do erro absoluto podem ser observados na Tabela 5.

Tabela 5 – Resultados com conjunto de atributos selecionados a partir do CfsSubsetEval.

	Coeficiente de correlação(%)	Média do erro absoluto
Linear Regression	86,29	11,072
Support Vector Regression	83,06	9,1012
Random Forest	67,86	11,2452

Já utilizando os atributos selecionados pela análise da matriz de covariância, podemos observar que foi possível atingir resultados superiores aos encontrados anteriormente (Tabela 6).

Tabela 6 – Resultados com conjunto de atributos selecionados a partir da matriz de covariância.

	Coeficiente de correlação(%)	Média do erro absoluto
Linear Regression	88,96	10,5684
Support Vector Regression	87,09	7,4871
Random Forest	69,42	11,3152

Avaliando os resultados demonstrados nas tabelas 4, 5 e 6 podemos constatar que a Regressão Linear e o Support Vector Regression conseguiram estabelecer modelos matemáticos com correlações muito fortes [7] entre o valor previsto e a severidade real das blocos.

Com finalidade de verificar a precisão dos modelos, avaliamos os valores previstos e a média do erro absoluto do Support Vector Regression e a Regressão Linear com os valores reais de severidade. Podendo ser observados na Tabela 7.

Tabela 7 – Severidades do experimento, previsões do Support Vector Regression (SVR), Regressão Linear (RL) e seus respectivos erros absolutos.

Severidade	Previsão (SVR)	Erro Ebsoluto (SVR)	Previsão (RL)	Erro Absoluto (RL)
55,2	60,19	4,99	64,76	9,56
82,9	71,69	11,21	63,88	19,02
85	75,82	9,18	74,22	10,78
91,3	73,28	18,02	67,75	23,55
55,8	54,69	1,11	67,43	11,63
54,6	54,59	0,01	50,52	4,08
52,9	63,55	10,65	74,76	21,86
66,3	64,35	1,95	57,81	8,49
41,8	55,50	13,70	56,32	14,52
52,9	68,60	15,70	70,74	17,84
44,6	53,90	9,30	60,07	15,47
56,3	56,62	0,32	67,44	11,14
55,8	51,31	4,49	64,33	8,53
66,7	64,35	2,35	57,81	8,89
54,2	63,95	9,75	60,64	6,44

Apesar da Regressão Linear ter um coeficiente de correlação superior ao Support Vector Regression, podemos observar na Tabela 7, que o SVR alcançou resultados mais próximos dos valores reais. O que indica que é o modelo mais promissor a ser utilizado na prática.

Foi constatado que as blocos com maiores erros absolutos correspondem aos valores de severidades menos frequentes no conjunto de dados utilizados para treinamento. Como pode ser observado na Figura 6, o conjunto de dados está desbalanceado, o que indica a possibilidade do modelo gerar previsões precisas apenas em certos intervalos de severidade.

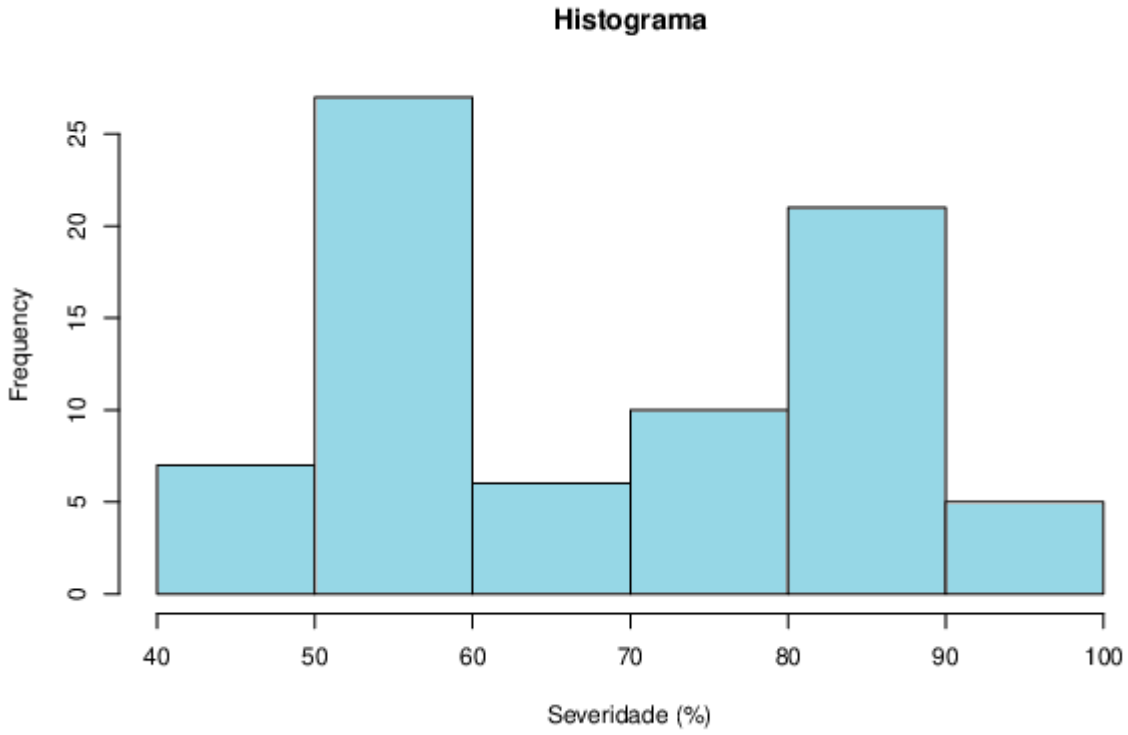

Figura 6 – Histograma de severidade das 76 amostras coletadas. [Fonte: Autor]

Foi possível observar neste trabalho que existe correlação muito forte entre descritores baseados nos canais RGB e HSV da imagem e a severidade da ferrugem asiática da soja. Tal correlação pode ser confirmada ao observar a média do erro absoluto do SVR utilizando os atributos selecionados a partir da matriz de covariância.

Apesar do algoritmo ter uma boa precisão, é necessário diminuir o erro absoluto para atingirmos níveis de severidade mais condizentes com os valores reais, sendo necessário balancear o conjunto de dados coletando um volume maior de amostras e consequentemente treinar os algoritmos de forma mais homogênea.

5 CONCLUSÃO

Neste trabalho, foram utilizados algoritmos de aprendizado de máquina para correlacionar as características de imagens aéreas RGB com a severidade da ferrugem asiática da soja.

Até o momento observou-se que o Support Vector Regression obteve o melhor desempenho, com um coeficiente de correlação de 87,09 % e média do erro absoluto de 7,4871, utilizando os atributos GLI, |R-B|, |G-R| e média dos canais R, B, RGB, S e V. Tal resultado indica uma correlação muito forte entre os dados brutos coletados em campo e a estimativa gerada pelo modelo.

Para trabalhos futuros buscaremos obter resultados mais precisos para que a análise de imagens aéreas RGB para a ferrugem asiática da soja se torne mais precisa. Com isto em mente, iremos explorar mais descritores que representem a cobertura vegetal da área experimental e aumentar o número de amostras usando outros ensaios de ferrugem para que o conjunto de dados atual se torne mais homogêneo. Também iremos utilizar métodos de classificação.

REFERÊNCIAS

[1] CONAB. 2018. Disponível em: <https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>.

[2] JT PAIVA WM, C. V. G. N. J. J. Y. *Epidemics of soybean rust (Phakopsora pachyrhizi) in Brazil and Paraguay from 2001 to 2003.* [S.l.: s.n.].

[3] SOARES SéRGIO DE ASSIS LIBRELOTTO RUBIN, A. P. W. J. G. O. R. M. *Fungicidas no controle da ferrugem asiática (Phakopsora pachyrhizi) e produtividade da soja.* [S.l.: s.n.], 2004.

[4] YORINORI, J. J. L. J. T. *Situação da Ferrugem Asiática da Soja no Brasil e na América do Sul.* [S.l.: s.n.], 2004.

[5] CANTERI CLAUDIA VIEIRA GODOY, L. K. M. G. *Diagramatic scale for assessment of soybean rust severity.* [S.l.: s.n.], 2006.

[6] CANTERI CLAUDIA VIEIRA GODOY, L. K. M. G. *Relação entre medidas de refletância e área foliar sadia, severidade da ferrugem asiática e produtividade da cultura da soja.* [S.l.: s.n.], 2007.

[7] D., E. J. *Straightforward statistics for the behavioral sciences.* [S.l.: s.n.], 1996.

[8] GONZALEZ, R.; WOODS, R. *Processamento Digital De Imagens.* ADDISON WESLEY BRA. ISBN 9788576054016. Disponível em: <https://books.google.com.br/books?id=r5f0RgAACAAJ>.

[9] BASAK SRIMANTA PAL, D. C. P. D. *Support Vector Regression.* 2007. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.6802&rep=rep1&type=pdf>.

[10] FRANK, M. A. H. E.; WITTEN, I. H. *The WEKA Workbench.* 2016. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf>.

[11] T. TIAN L., Z. Y. T. K. C. A. *A review of remote sensing methods for biomass feedstock production.* [S.l.: s.n.], 2011.

[12] RAYMOND DAUGHTRY C. S. T., E. J. U. H. L. D. S. H. J. E. *Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index.* [S.l.: s.n.], 2011.

[13] HALL, L. A. S. M. A. *Feature Subset Selection: A Correlation Based Filter Approach.* [S.l.: s.n.].

[14] HALL, M. A. *Correlation-based Feature Subset Selection for Machine Learning.* Tese (Doutorado) — University of Waikato, Hamilton, New Zealand, 1998.