



unesp

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



ESTIMATIVA DE NITROGÊNIO EM *Annona emarginata* (Schltdl.) H. Rainer
UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIRS) -
UMA ABORDAGEM ESTATÍSTICA E COMPUTACIONAL

RAFAELA LANÇAS GOMES

**Dissertação de Mestrado apresentada
ao Instituto de Biociências, campus
de Botucatu, UNESP, para obtenção
do título de Mestre em Ciências
Biológicas (Botânica), Área de
concentração: Fisiologia Vegetal.**

**BOTUCATU - SP
- 2019 -**

UNESP - UNIVERSIDADE ESTADUAL PAULISTA

CÂMPUS DE BOTUCATU

INSTITUTO DE BIOCÊNCIAS

**ESTIMATIVA DE NITROGÊNIO EM *Annona emarginata* (Schltdl.) H. Rainer
UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIRS) -
UMA ABORDAGEM ESTATÍSTICA E COMPUTACIONAL**

RAFAELA LANÇAS GOMES

PROF^a DR^a GISELA FERREIRA

ORIENTADORA

PROF^o DR^o JOSÉ RAIMUNDO DE SOUZA PASSOS

CO-ORIENTADOR

**Dissertação de Mestrado apresentada
ao Instituto de Biociências, campus
de Botucatu, UNESP, para obtenção
do título de Mestre em Ciências
Biológicas (Botânica), Área de
concentração: Fisiologia Vegetal.**

**BOTUCATU - SP
- 2019 -**

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Gomes, Rafaela Lanças.

Estimativa de nitrogênio em *Annona emarginata* (Schltdl.)
H. Rainer utilizando espectroscopia no infravermelho próximo
(NIRS) : uma abordagem estatística e computacional / Rafaela
Lanças Gomes. - Botucatu, 2020

Dissertação (mestrado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Instituto de Biociências de
Botucatu

Orientador: Gisela Ferreira

Coorientador: José Raimundo de Souza Passos

Capes: 20303009

1. Plantas dos cerrados. 2. Bootstrap (Estatística).
3. Plantas - Nutrição. 4. Espectroscopia de infravermelho.
5. Análise espectral. 6. Redes neurais (Computação).

Palavras-chave: Aprendizado de máquinas; Bootstrap; Nutrição
mineral de plantas; Redes neurais artificiais; Índice
espectral de vegetação.

Agradecimentos

Á Deus, por ter colocado pessoas maravilhosas no meu caminho e me dado muitas oportunidades para seguir os meus sonhos.

Aos meus pais, avós, tia e irmão, que sempre me apoiaram em minhas decisões e acreditaram em mim, mesmo quando eu não acreditava.

Aos meus amigos Vitor Borges Pimenta, Carolina Aparecida Duarte Pereira Randovanovic, Hiran Tsai Sun e todos que me ajudaram a fazer este experimento acontecer e me apoiaram durante o mestrado.

Á minha família de laboratório, por toda a ajuda, apoio e companheirismo.

Ao Felipe Giroto Campos e a Marília Caixieta, por seus ensinamentos, e por me acompanhar durante todo o experimento e escrita.

Ao João Ricardo Favan e André Stefanini Jim pela ajuda e apoio com o equipamento e a programação.

Á todos os meus professores, que me guiaram a construir este caminho, em especial a professora Beatriz que me alfabetizou e aos professores Fábio e Isabela, cujas aulas de redação e as correções me proporcionaram base para conseguir escrever esta dissertação.

Ao professor José Raimundo de Souza Passos, meu mentor desde o meu primeiro ano, que me ensina até hoje a como ser um cientista.

Á professora Gisela Ferreira por ter me orientado e colaborado com a solução de todos os problemas que apareceram no caminho.

Ao programa de pós-graduação e á Capes, pelo apoio a pesquisa e a bolsa concedida.

Lista de abreviações:

ACP	Análise de Componentes Principais
CFS	<i>Correlation Based Feature Selection</i> (Seleção de Atributos Baseada em Correlação)
CSE	<i>Classifier Subset Evaluator</i> (Avaliador de Subconjunto Classificador)
E	Eficiência (índice criado)
FIR	<i>Far Infrared</i> (Infravermelho Distante)
FLD	Função Linear Discriminante
FQD	Função Quadrática Discriminante
IEV	Índice Espectral de Vegetação
LWIR	<i>Long-Wavelength Infrared</i> (Infravermelho de Ondas Longas)
MCC	<i>Multi Class Classifier</i> (Classificador Multi-Classe)
MIR	<i>Middle-Infrared</i> (Infravermelho Médio)
mSR	<i>Modified Simple Ratio</i> (Relação Simples Modificada)
MWIR	<i>Mid-Wavelength Infrared</i> (Infravermelho de Ondas Médias)
N	Nitrogênio
ND	<i>Normalized difference index</i> (Índice de Diferença Normalizada)
NDVI	<i>Normalized Difference Vegetation Index</i> (Índice de Diferença de Vegetação Normalizado)
NIR	<i>Near-Infrared</i> (Infravermelho Próximo)
NIRS	<i>Near-Infrared Spectroscopy</i> (Espectroscopia no Infravermelho Próximo)
PMC	<i>Perceptron</i> Multicamadas
PRI	<i>Photochemical Reflection Index</i> (Índice de Reflexão Fotoquímica)
PSND	<i>Pigment Specific Normalized Difference</i> (Diferença Normalizada Específica do Pigmento)
PSSR	<i>Pigment Specific Simple Ratio</i> (Relação Simples Específica de Pigmento)
R Log	Regressão Logística
RNA	Redes Neurais Artificiais
SR	<i>Simple Ratio</i> (Relação Simples)
SVM	<i>Suport Vector Machine</i> (Máquinas de Vetor Suporte)
SWIR	<i>Short-Wavelenght Infrared</i> (Infravermelho de Ondas Curtas)

Sumário:

1. Resumo	1
1.1. Abstract	2
2. Introdução Geral	3
1.1. Objetivos Gerais	6
1.2. Objetivos Específicos	6
3. Revisão Bibliográfica	7
3.1. Nutrição Mineral de Plantas	7
3.1.1. Tipos de Nutrientes	8
3.1.2. Sinais de Deficiência	8
3.1.3. Nitrogênio	9
3.2. Espectroscopia no Infravermelho Próximo (NIRS)	10
3.2.1. Funcionamento	11
3.2.2. A Reflectância Foliar	11
3.2.3. Vantagens e Aplicações em Ciências de Plantas	13
3.3. Aprendizado de Máquinas e Redes Neurais Artificiais	15
3.3.1. Como Máquinas Aprendem ?	16
3.3.2. O Neurônio Biológico	17
3.3.3. O Neurônio Artificial	18
3.3.4. WEKA	19
3.3.4.1. Alguns Algoritmos de Seleção de Atributos	19
3.3.4.2. Alguns Algoritmos de Classificação	21
3.3.5. Aplicações em Ciências de Plantas	22
3.4. análise Estatística	24
3.4. 1. Análise Multivariada	24
3.4.2. Simulação <i>Bootstrap</i>	25
3.4.3. Modelos Lineares Generalizados	25
3.5. Índices Espectrais de Vegetação	26
3.6. Família Annonaceae	27
3.6.1. Estudos Nutricionais com a Família	28
3.6.2. <i>Annona emarginata</i> (Schltdl.) H. Rainer	29
4. Capítulo 1 - Identificação da falta de nitrogênio em mudas de <i>Annona emarginata</i> (Schltdl.) H. Rainer antes do aparecimento de sinais visuais utilizando NIRS	31
4.1. Resumo	31
4.2. Introdução	33
4.3. Material e Métodos	34
4.4. Resultados	38
4.5. Discussão	44
4.6. Conclusões	49
4.7. Referências	49
5. Capítulo 2 - Aplicação de algoritmos de aprendizado de máquinas para identificação da falta de nitrogênio em mudas de <i>Annona emarginata</i> (Schltdl.) H. Rainer antes do aparecimento de sinais visuais utilizando NIRS e fotografias digitais	53

4.1. Resumo	53
4.2. Introdução	55
4.3. Material e Métodos	56
4.4. Resultados	59
4.5. Discussão	70
4.6. Conclusões	73
4.7. Referências	73
6. Considerações Finais	76
7. Referências Bibliográficas	77
8. Apêndices	
8.1. Apêndice A	95
8.2. Apêndice B	96

1. Resumo

O nitrogênio é um elemento mineral essencial para as plantas. Sua deficiência em fases iniciais do desenvolvimento pode gerar alterações fisiológicas e morfológicas que reduzem o crescimento e conflui na não expressão total do potencial genético vegetal. As técnicas mais difundidas para a quantificação do N nas plantas demandam tempo, são destrutivas e liberam compostos tóxicos para o ambiente. A NIRS (*Near-Infrared Spectroscopy* - Espectroscopia no Infravermelho Próximo), se apresenta como uma técnica alternativa, sendo indireta, mas instantânea, não destrutiva e que não utiliza reagentes químicos, mas necessita de calibração, que pode ser feita por métodos estatísticos e computacionais. Para mudas que são produzidas em viveiros, como as de *Annona emarginata* (Schltdl.) H. Rainer, é essencial manter o monitoramento de N, de forma rápida e não danosa, para garantir a qualidade e vigor das mudas. Desta forma, este trabalho visou detectar alterações na caracterização espectral foliar de *A. emarginata* em função do fornecimento de concentrações de nitrogênio e classificar as mudas em função dos níveis de nitrogênio, com base na caracterização espectral, utilizando algoritmos de aprendizado de máquinas e análise estatística multivariada. As mudas de *A. emarginata* (240) foram mantidas em sistema de hidroponia, com alterações na concentração de nitrogênio: 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4), com 60 repetições (mudas) para cada tratamento. Após 45 dias nestas soluções, foram coletadas três folhas de cada planta, fotografadas e aferidas suas caracterizações espectrais. Foram testadas diversas combinações de técnicas de redução e de classificação estatísticas e de aprendizado de máquinas e propostos índices espectrais com os comprimentos de onda mais selecionados. Quanto às variáveis fisiológicas, foram analisadas as trocas gasosas, fluorescência e pigmentos foliares. Pela caracterização espectral, percebe-se um afastamento total das curvas de mudas com falta e sem falta de N. Conclui-se que a técnica NIRS pode ser aplicada para a detecção de variações na concentração de N nas folhas de plantas jovens (mudas), antes do aparecimento de sinais visuais e que técnicas estatísticas são melhores para se fazer a calibração do método do que técnicas de aprendizado de máquinas.

Palavras chave: redes neurais artificiais, aprendizado de máquinas, *perceptron* multicamadas, nutrição mineral de plantas, *bootstrap*, índice espectral de vegetação.

1.1. Abstract

Nitrogen is an essential mineral element for plants. Its deficiency in early stages of development can lead to physiological and morphological changes that reduce growth and result in the total non-expression of plant genetic potential. The most widespread techniques for quantifying N in plants are time consuming, destructive and release toxic compounds into the environment. Near-Infrared Spectroscopy (NIRS) is an alternative technique, being indirect, but instantaneous, non-destructive and does not use chemical reagents, but needs calibration, which can be done by statistical and computational methods. For seedlings that are produced in nurseries, such as those of *Annona emarginata* (Schltdl.) H. Rainer, it is essential to keep N monitoring fast and harmless to ensure seedling quality. Thus, this work aimed to detect changes in the leaf spectral characterization of *A. emarginata* as a function of nitrogen concentration supply and to classify seedlings as a function of nitrogen levels, based on spectral characterization, using machine learning algorithms and multivariate statistical analysis. 240 *A. emarginata* seedlings were maintained in a hydroponic system, with modifications in nitrogen concentration: 0 mg.L⁻¹ of N (T1); 52.5 mg.L⁻¹ of N (T2); 105 mg.L⁻¹ N (T3) and 210 mg.L⁻¹ N (T4), with 60 replications (seedlings) for each treatment. After 45 days in these solutions, three leaves of each plant were collected, photographed and their spectral characterizations was measured. Several combinations of reduction and classification techniques and machine learning techniques were tested and proposed spectral indexes. Regarding the physiological variables, gas exchange, fluorescence and leaf pigment were analyzed. Due to the spectral characterization, it is possible to observe a total deviation of the curves with the N lacking and non-lacking seedlings. It can be concluded that the NIRS technique can be applied to detect variations in N concentration in the leaves of seedlings before visual symptoms appear, and which statistical techniques are better for method calibration than machine learning techniques.

Key words: artificial neural networks, machine learning, multilayer perceptron, plant mineral nutrition, *bootstrap*, vegetation spectral indexes.

2. Introdução Geral

A família *Annonaceae* é uma das maiores famílias dos trópicos e sub-trópicos, tendo mais de 2500 espécies distribuídas pantropicalmente (ESCRIBANO; VIRUEL; HORMAZA, 2007), sendo uma das mais diversificadas presentes nas florestas tropicais (COUVREUR *et al.*, 2011). Ela possui grande importância ecológica e tem gerado interesse econômico em decorrência de espécies cujos metabólitos especializados tem grande ação anti inflamatória (ATTIQ, JALIL, HUSAIN, 2011) e potencial antiproliferativo em células cancerígenas (PIEME *et al.*, 2014) e de espécies com frutos de sabor agradável e elevado valor nutritivo.

Dentre seus gêneros, o *Annona* é o mais cultivado no Brasil, graças a algumas espécies de importância econômica, como a pinha, ou fruta-do-conde (*Annona squamosa* L.), a graviola (*Annona muricata* L.), a cherimóia (*Annona cherimola*, Mill.) e o híbrido atemoia (*Annona x atemoya* Mabb.), cujos frutos possuem elevado valor comercial e geram interesse tanto no mercado interno como no externo, principalmente o norte americano e europeu (SOBRINHO, 2010). Para a produção de mudas de *Annona*, todas as copas comerciais são enxertadas em espécies nativas (SOBRINHO, 2010).

Para o cultivo de atemoia, que nos últimos anos vem tendo a preferência dos consumidores graças a suas boas características (SOBRINHO, 2010), um dos porta-enxerto mais utilizados para a sua propagação é a *Annona emarginata* (Schltdl.) H. Rainer (TOKUNAGA, 2005), uma espécie nativa do Brasil que fornece resistência para diversas podridões de colo e raiz e coleobrocas (TOKUNAGA, 2005) e garante copa vigorosa (SCALOPPI JUNIOR, 2007).

O restabelecimento do tecido na região enxertada é afetado por diversos fatores, dentre eles, se destacam as características nutricionais do porta-enxerto (BARON *et al.*, 2016), o que torna o monitoramento nutricional em viveiros de mudas de *A. emarginata* de extrema importância, a fim de permitir a detecção de deficiências antes, ou pelo menos na fase inicial da manifestação de sinais visuais (sintomas).

Dentre os nutrientes essenciais exigidos pelas plantas, o nitrogênio (N) é um dos elementos-chave cuja demanda é relativamente grande em comparação com a de outros elementos (WANG *et al.*, 2017), e portanto seu monitoramento precisa ser efetivo, principalmente em culturas perenes, como as frutíferas. Ele é essencial em todas as fases do desenvolvimento das plantas (FERNANDES; ROSSIELO, 1995), pois faz parte da composição de proteínas e compostos orgânicos essenciais, dentre eles

clorofilas, ácidos nucléicos, aminoácidos, auxina, alcaloides, citocininas, e algumas enzimas como a *Rubisco* (ribulose-1,5-bisfosfato carboxilase oxigenase) (MAATHUIS, 2009; OSAKI *et al.*, 1993). A falta de N nas fases iniciais do desenvolvimento gera consequências para a planta, como alterações fisiológicas e morfológicas que reduzem seu crescimento e faz com que não expressem totalmente seu potencial genético (EVANS, 1989), o que justifica constante monitoramento deste elemento em viveiros.

A concentração foliar de N pode ser aferida de diversas formas. Os métodos mais comuns para se fazer esta análise são os diretos, como o de Kjeldahl (JONES, 1991), mas tais análises demandam tempo, utilizam compostos tóxicos ao meio, como ácido bórico e requerem coletas com avaliações destrutivas do vegetal (MALAVOLTA *et al.*, 1997).

Uma alternativa para a avaliação nutricional não destrutiva de plantas é a Espectroscopia no Infravermelho Próximo (*Near-Infrared Spectroscopy* - NIRS), uma técnica analítica que permite inferir de modo indireto os compostos de um material com base em sua reflectância (FERRARESI, 2012). A NIRS possibilita a detecção mais rápida das características nutricionais das plantas em relação a determinado elemento, evitando os inconvenientes das análises convencionais, como a destruição da massa de matéria seca da planta, o elevado tempo e custo necessários para a quantificação em laboratório especializado.

Para que ocorra bom funcionamento desta técnica, é necessário primeiro calibrar, já que é um método indireto (OTTO, 1999). Isto pode ser feito com diversos métodos. Como a resposta adquirida com o NIRS é multivariada, geralmente são utilizadas técnicas estatísticas de análise multivariada para a calibração, mas, atualmente, existem metodologias alternativas computacionais que podem ser empregadas, como as técnicas de aprendizado de máquinas, que consistem na modelagem do comportamento de um ser inteligente (FACELI *et al.* 2011).

Apesar da larga utilização na área biológica e botânica, existem poucos estudos que mostram a aplicação de aprendizado de máquinas e particularmente redes neurais artificiais para nutrição mineral de plantas, sendo esta, uma técnica potencialmente promissora para a área.

Quando o sinal de deficiência se torna visível, o dano, já estará estabelecido, o que muitas vezes pode ser irreversível, resultando em redução da produtividade. Uma vantagem da técnica NIRS é que como a resposta no infravermelho está associada a alterações nos compostos foliares (MAHLEIN, 2016), em alguns casos é possível

detectar alterações em sua caracterização espectral antes do aparecimento de sinais visíveis, possibilitando efetuar correções prevenindo reduções significativas na produtividade primária.

Dentre os avanços na área de nutrição mineral com NIRS, não existem relatos para a família *Annonaceae*, desta forma, fazer a avaliação do estado nutricional de mudas de anonáceas com esta técnica será de grande importância no manejo das áreas de produção, que estão em constante expansão, podendo ser uma ferramenta de grande interesse, uma vez que os próprios estudos nutricionais para esta família são escassos (ROZANE; NATALE, 2014).

O estudo da *A. emarginata* durante a fase de desenvolvimento inicial representa também um avanço no fornecimento de mudas melhor nutridas e adaptadas para serem utilizadas em áreas de recuperação ambiental, além de redução no tempo de viveiro para a produção de porta-enxertos, o que pode resultar em redução de custo da muda.

Deste modo, o estabelecimento de métodos rápidos, simples e não destrutivos para quantificar o nitrogênio em anonáceas, como o NIRS e aprendizado de máquinas, são importantes para avaliar precocemente deficiências nutricionais que possam resultar em danos no metabolismo e que, conseqüentemente, comprometam a sobrevivência e o desenvolvimento das plantas desde os estádios iniciais de desenvolvimento em viveiro.

Neste contexto e considerando a importância da espécie *A.emarginata*, esta dissertação teve o propósito de avaliar se a NIRS é efetiva para detectar variações de N em folhas de mudas de *A.emarginata*, empregando-se de técnicas de aprendizado de máquinas e análises estatísticas multivariadas.

2. Objetivos

2.1. Objetivo Geral

a) Avaliar se a espectroscopia no infravermelho próximo (NIRS) é efetiva para detectar variações de nitrogênio em folhas de plantas jovens de *Annona emarginata* (Schltdl.) H. Rainer.

2.2. Objetivos Específicos

a) Avaliar se há aplicabilidade de algoritmos de aprendizado de máquinas para classificar mudas em função dos níveis de N com base na caracterização espectral.

b) Avaliar se a NIRS é mais efetiva que fotografias digitais para identificar variações no nitrogênio foliar em mudas de *Annona emarginata* (Schltdl.) H. Rainer.

c) Selecionar um conjunto de índices que, com base nas reflectâncias foliares prevejam com acuracidade o teor de nitrogênio em *Annona emarginata* (Schltdl.) H. Rainer.

3. Revisão Bibliográfica

3.1. Nutrição Mineral de Plantas

Plantas são seres autótrofos, mas, para sobreviverem precisam retirar do ambiente água, carbono (C), luz solar e nutrientes minerais. Todos estes fatores estão intrinsecamente ligados à sua produtividade, de forma que, a ausência ou insuficiência destes levam a problemas fisiológicos e, em casos extremos, a morte da planta.

Os nutrientes minerais, ou minerais essenciais, são elementos absorvidos pelas plantas na forma, principalmente, de íons inorgânicos (TAIZ; ZEIGER, 2017), que são necessários para completar o seu ciclo de vida, e são componentes de um metabólito ou constituinte essencial da planta (BUCHANAN; GRUISSEM; JONES, 2015). Em geral, os elementos são classificados de acordo com suas concentrações relativas nos tecidos vegetais em: macronutrientes, aqueles requeridos pelas plantas em grandes quantidades, como o nitrogênio (N), potássio (K), cálcio (Ca), magnésio (Mg), fósforo (P) e enxofre (S), e micronutrientes, aqueles requeridos em pequenas quantidades, como o cloro (Cl), ferro (Fe), boro (B), manganês (Mn), zinco (Zn), cobre (Cu), níquel (Ni) e molibdênio (Mo) (TAIZ; ZEIGER, 2017). Mas esta classificação é apenas quantitativa, e não relacionada com a importância dos nutrientes para as plantas, assim, os macronutrientes não são mais importantes que os micronutrientes, são apenas requeridos em maiores quantidades, um demonstrativo disto está na lei de Liebig.

Liebig (1841) estudou o efeito dos nutrientes minerais nas plantas e formulou uma lei, que ficou conhecida como lei de Liebig, ou lei dos mínimos; "A produtividade das culturas é limitada pelo nutriente que estiver em menor disponibilidade no solo, mesmo que os demais estejam em níveis adequados". Desta forma, podemos notar que o equilíbrio dos nutrientes na planta é essencial para a manutenção desta e que todos os nutrientes são igualmente essenciais.

O suprimento inadequado de nutrientes, tanto para o excesso quanto para a falta, vão resultar em desordem nutricional, percebida em forma de sintomas, que estão correlacionados com a função e a mobilidade do nutriente na planta (BARKER; PIELBEAM, 2015). A planta somente alcança sua produtividade máxima quando a concentração dos nutrientes minerais está adequada, ou seja, quando todos os nutrientes estão em quantidade suficiente, quando há falta de algum nutriente, ou o excesso a produtividade, ou seu crescimento, é afetado, sendo menor do que sua capacidade genética (TAIZ; ZEIGER, 2017).

3.1.1. Tipos de Nutrientes

Os nutrientes minerais podem ser distribuídos em quatro grupos de acordo com a sua função fisiológica e papel bioquímico (TAIZ; ZEIGER, 2017): nutrientes que são parte de compostos de carbono (N e S), nutrientes que são importantes no transporte de energia e integridade estrutural (P e B), nutrientes que permanecem na forma iônica (K, Ca, Mg, Cl e Zn) e nutrientes que agem em reações redox (Fe, Mn, Cu, Ni e Mo).

Também podem ser agrupados em duas classes de acordo com a sua mobilidade: elementos móveis, que são aqueles tem alta tendência de translocação durante deficiências, como N, P, K, Mg, Cl, Zn, Mo, e elementos imóveis, que são aqueles que não se translocam durante deficiências, como Ca, S, Fe, B e Cu (TAIZ; ZEIGER, 2017).

3.1.2. Sinais de deficiência

Os sinais de deficiência são a expressão final da desordem metabólica de uma planta, mas, quando aparecem, já ocorreu o comprometimento do metabolismo e do crescimento vegetal (MAUSETH, 2009). Quando a planta exibe sinais, já ocorreram alterações como a redução da velocidade dos processos metabólicos, a paralisação e desarranjo dos processos biológicos, a alteração de membranas, parede celular e organelas e a alteração e deformação das células (MAUSETH, 2009).

A maior parte dos sinais característicos de deficiência ou excesso de nutrientes ocorrem nas partes aéreas da planta e são facilmente observáveis. Em geral é fácil de se associar um sinal com um nutriente, graças a relação entre sinal e mobilidade e função do nutriente (SALISBURY; ROSS, 1992). No entanto, quase nunca é eficiente diagnosticar uma deficiência nutricional apenas pelos sinais exibidos pela planta, já que, como muitos nutrientes possuem funções parecidas, a falta pode resultar em sinais parecidos, como é o caso do magnésio e do nitrogênio, que são constituintes da molécula de clorofila e são móveis, de modo, que sua falta causará a clorose de folhas maduras, mas, no início o Mg causa clorose internerval e o N clorose total (MAUSETH, 2009). Outro problema que pode dificultar o diagnóstico do nutriente deficiente é que, em alguns casos, pode ocorrer a falta de dois ou mais elementos, somando os sinais.

3.1.3. Nitrogênio

Dentre os nutrientes minerais requeridos pelas plantas, o nitrogênio se destaca por ser vital na síntese de muitas moléculas de grande importância, como aminoácidos, ácidos nucleicos (SOUZA; FERNANDES, 2006), proteínas, dentre elas a ribulose 1,5 bisfosfato carboxilase/oxigenase (*Rubisco*) (LAWLOR, 1989) e metabólitos especializados, como os alcaloides (MARSCHNER, 2012).

Por ser essencial para a formação e desenvolvimento, a falta de N nas fases iniciais do desenvolvimento pode gerar alterações fisiológicas e morfológicas que reduzem o crescimento fazendo com que não ocorra a expressão total de seu potencial genético (EVANS, 1989).

O nitrogênio é absorvido pelas plantas em forma de íons nitrato (NO_3^-) ou íons amônio (NH_4^+). Como o nitrato entra em forma muito oxidada, é necessário que este seja reduzido até amônio para então sofrer o processo de aminação e ser incorporado (VIECELLI, 2017). Este processo de redução consome elevada quantidade de energia, o que limita sua ocorrência nas raízes (MARENCO; LOPES, 2005), desta forma, o principal local de redução, acaba sendo os órgãos fotossintéticos, já que este processo pode ser um dreno alternativo de elétrons, usando o excesso de energia decorrente de fotoinibição ou fotoxidação do fotossistema para a redução do nitrato presente nas folhas (BLOOM *et al.*, 2010).

Quase todo o nitrogênio presente na planta é absorvido pelas raízes, e depois distribuído em forma de aminoácidos, mas também pode ser translocado na forma de nitrato, dependendo do potencial redutor de suas raízes (VIECELLI, 2017).

O primeiro sinal de deficiência deste nutriente é o amarelecimento das folhas maduras (clorose), que ocorre devido à degradação e não reposição da clorofila nestas folhas, em decorrência da redução ou ausência do elemento no solo. Deste modo, folhas maduras translocam o elemento para as demais partes da planta, e, como não recebem nitrogênio das raízes, se tornam deficientes, apresentando o sinal (TAIZ; ZEIGER, 2017).

A concentração de nitrogênio de um órgão da planta pode ser aferido tanto de formas diretas, como pelo método de Kjeldahl e de Dumas, quanto de formas indiretas, como pelo índice SPAD e NIRS.

O método de Kjeldahl, considerado o método clássico e tradicional para determinação de N, precisa de quantidade relativamente grande de amostra (por volta de 1 grama) e que necessita ser destruída para poder isolar o conteúdo de N. Esta técnica é

composta basicamente de 3 fases: digestão, destilação e titulação (KJELDAHL, 1883). Sendo um processo delicado, pois necessita manipular reagentes perigosos em alta temperatura e se estes compostos não forem descartados de forma correta, podem poluir o meio.

Desta forma, a técnica NIRS se mostra vantajosa perante a técnica tradicional para quantificar N em folhas por não necessitar de compostos químicos, não destruir a pequena amostra que requer, ser muito mais rápido, menos trabalhoso e não oferecer riscos ao ambiente.

3.2. Espectroscopia no Infravermelho Próximo (*Near-Infrared Spectroscopy* - NIRS)

A espectroscopia no infravermelho próximo (*Near-Infrared Spectroscopy* - NIRS) é uma técnica analítica de detecção de substâncias, já consagrada no meio científico, que permite inferir indiretamente alguns compostos químicos de forma rápida, não destrutiva e com pouca amostra (FERRARESI, 2012).

O infravermelho é uma radiação não ionizante que abrange os comprimentos de onda de 700 nm à 1 mm e como seu espectro é muito largo, geralmente é subdividido em seções menores. Existem diferentes sub-divisões do espectro infravermelho, as mais comuns separam seu espectro em cinco (a) e em três (b) partes:

a) NIR (infravermelho próximo - *Near-infrared*), que vai de 700 à 1400 nm, SWIR (infravermelho de ondas curtas - *Short-wavelength infrared*), de 1400 à 3000 nm, MWIR (infravermelho de ondas médias - *Mid-wavelength infrared*), de 3000 à 8000 nm, LWIR (infravermelho de ondas longas - *Long-wavelength infrared*), de 8000 à 15000 nm e FIR (infravermelho distante - *Far infrared*), de 15000 à 100000 nm (BYRNES, 2009).

b) NIR (infravermelho próximo - *Near-infrared*), de 700 à 2500 nm, MIR (infravermelho médio - *Middle-infrared*), de 2500 à 5000 e FIR (infravermelho distante - *Far infrared*), de 5000 à 100000 nm (SKOOG; HOLLER; NIEMAN, 2002).

Considerando a divisão em três partes, que segue o ponto de vista da aplicação e instrumentação, na faixa do espectro NIR, de 780 à 2500 nm (WILLIAMS; NORRIS, 1990), ocorre a absorção, principalmente, de ligações do tipo C-H, O-H e N-H (BURNS; CIURCZACK, 2001), e este tipo de ligações estão muito presentes nos materiais orgânicos, como os de origem vegetal.

3.2.1. Funcionamento

A base teórica desta técnica está na física de cores, que se relaciona com a natureza de ligação das moléculas de uma amostra. Estas ligações vibram constantemente, causando estiramento e compressão dos átomos, de modo que estes tenham um movimento de onda. A frequência destas ondas é específica e dependente dos elementos envolvidos (VAN KEMPEN; JACKSON, 1996). Quando se incide luz em um alvo, parte desta será absorvida e parte será refletida, os comprimentos de onda que vibram na mesma frequência que as ligações serão absorvidos por estas e a energia é usada para intensificar as vibrações, gerando calor, por outro lado, os comprimentos de onda que não vibram na mesma frequência serão refletidos (NASSAU, 2001).

Desta forma, a técnica NIRS consiste em iluminar uma amostra com luz padrão conhecida no infravermelho próximo, e com um sensor captar a luz que foi refletida. A absorção de luz é então medida pela diferença entre a quantidade de luz incidida e a refletida (VAN KEMPEN; JACKSON, 1996). Com isto se realiza uma representação gráfica da refletância em faixas de comprimento de ondas estreitas e adjacentes, chamada de caracterização espectral de um alvo (FONSECA *et al.*, 2002). Ao se avaliar esta caracterização espectral, pode-se estimar, com base nas faixas absorvidas, a composição molecular do alvo.

Para um bom funcionamento da técnica, é necessário estabelecer uma relação matemática entre a resposta obtida pelo equipamento em uma amostra qualquer, com a resposta medida por um material de referência, ou seja, é necessário que ocorra uma calibração (OTTO, 1999). Após a calibração, a técnica pode ser utilizada de diversas formas, mas cuidadosamente, pois, podem ocorrer mudanças na caracterização espectral de um alvo em decorrência de outras variáveis não correlacionadas com a de desejo, que devem ser controladas no experimento de calibração, como a umidade e temperatura do material, características da amostra, morfologia e equipamento utilizado, gerando assim, erro na calibração (SIEMENS; DAUN, 2005; PANDORF; MAN, 1990).

3.2.2. A Refletância Foliar

Quando ocorre a incidência de radiação em uma folha, diversas interações vão ocorrer, uma parte dessa radiação vai ser refletida diretamente pela superfície foliar, e o resto vai penetrar na folha, onde vão ocorrer outras 3 interações: parte será transmitida pela folha, parte será absorvida pelos compostos químicos da folha e parte será refletida (Figura 1), estas são chamadas de propriedades ópticas das folhas (MAHLEIN, 2016).

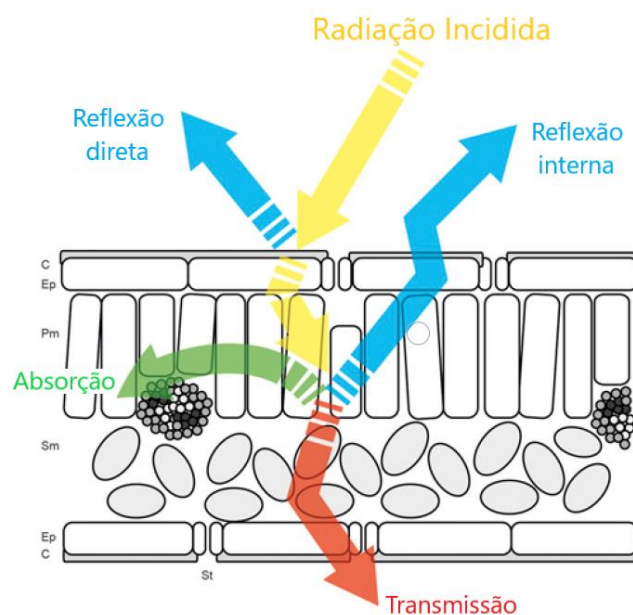


Figura 1: Interações de radiação luminosa com a folha.

Fonte: Adaptado de Mahlein (2016).

A reflectância é o resultado da reação da iluminação incidente com a matéria, que retorna a superfície convertida em energia radiante (BARANOSKI; ROKNE, 2001). A reflectância foliar é a soma da reflexão direta com a reflexão interna. Este é um fenômeno complexo pois depende de múltiplas interações biofísicas e bioquímicas (MAHLEIN, 2016). O padrão de reflectância de uma folha no infravermelho próximo se dá principalmente graças a estrutura anatômica da folha, tanto interna quanto externa (JENSEN, 2002), mas também é influenciada por sua composição química (MAHLEIN, 2016).

Em geral, pode-se associar algumas faixas de comprimento de onda com a absorção de determinados componentes bioquímicos das folhas. A faixa do visível (400 à 700 nm) é influenciada principalmente pelo conteúdo de pigmentos foliares, a faixa do NIR que compreende de 700 à 1100 nm varia muito de acordo com a estrutura da folha e com sua absorção de água, já a faixa do NIR que abrange de 1100 à 2500 nm é influenciada majoritariamente pela composição química e água (CARTER; KNAPP, 2001; JACQUEMOUD; USTIN, 2001).

Também existem algumas faixas curtas e comprimentos de onda específicos que podem ser associados a características e compostos da folha. Curran (1989) mostra que o comprimento de onda de 930 nm está associado a óleos, 1400 nm está associado a

água, 1510 nm a proteínas e nitrogênio e a faixa de 1530 à 1580 nm pode ser associada a amido.

3.2.3. Vantagens e Aplicações em Ciências de Plantas

Esta técnica possui ampla aplicação em diversas áreas, como na botânica, onde destaca-se por poder analisar qualquer material de origem vegetal, como sementes, frutos, flores e folhas tanto frescas como secas e preparadas (FERNANDEZ *et al.* 1994).

Lang *et al.* (2015), utilizando NIRS conseguiram diferenciar espécies amazônicas do gênero *Protium*, jovens e maduras, antes de seu florescimento. Foram construídos modelos considerando a caracterização espectral das folhas na faixa de 1000 à 2500 nm das plantas jovens e maduras conjuntamente, depois foi feita uma redução da dimensão dos dados com o uso de análise de componentes principais (PCA). Eles mostraram a possibilidade de uso de espectrorradiômetros para fins taxonômicos e como instrumento de controle de tráfico de madeira de lei.

A aplicação do NIRS pode substituir algumas técnicas caras e demoradas que utilizam reagentes altamente prejudiciais ao ambiente, um exemplo é para avaliar o conteúdo de óleos em sementes, cuja técnica usualmente utilizada é a de extração química com Soxhlet (ZENEBO; PASCUET; TIGLEA, 2008), que necessita de solventes como hexano, diclorometano e o acetato de etila. Rossato *et al.* (2013), utilizando NIRS, conseguiram prever o teor de óleo em sementes de canola (*Brassica napus* L.) com diferentes genótipos, provando que, com um bom método de calibração, esta técnica pode substituir a extração química com Soxhlet.

Jasinski *et al.* (2016) fizeram curvas de calibração preditivas de NIRS para quantificar o conteúdo de óleo, proteína, carbono e nitrogênio de sementes de *Arabidopsis*, e depois utilizaram essas equações para avaliar a variação natural destes principais componentes nas sementes. Com estas equações, provaram que o NIRS é uma ferramenta adequada para fenotipar a produtividade de sementes de *Arabidopsis*.

Xu *et al.* (2015) utilizaram NIRS para ver a degradação de qualidade de sementes de lótus chinesa (*Nelumbo nucifera* Gaertn.) por umidade de forma rápida e não destrutiva, podendo assim aferir a qualidade de um lote de sementes sem perda nas amostragens e de forma instantânea.

Pérez-Marín *et al.* (2019), com o intuito de estabelecer a adequação de espinafre (*Spinacia oleracea* L.) para diferentes usos após a colheita, utilizaram os espectros NIR

das folhas, associadas com modelos de regressão de mínimos quadrados modificados para prever os valores de nitrato, ácido ascórbico e sólidos solúveis em folhas frescas. Conseguiram resultados promissores ao classificar as folhas em "espinafre apta para ser usada fresca" e "espinafre conservada, ultracongelada ou congelada", provando a capacidade da NIRS para estabelecer o destino final da produção de espinafre, fornecendo triagem que leva em consideração parâmetros importantes para a segurança de qualidade.

Smis *et al.* (2014) utilizaram NIRS para determinar o conteúdo de sílica em diversos tecidos de plantas de diversas espécies. Eles observaram que a calibração era mais eficaz quando se usavam apenas espécies pertencentes ao mesmo grupo, sugerindo que, quanto mais próximas as características das amostras, mais eficaz será o método. Deste modo, podemos inferir que, ao se utilizar uma calibração já pronta, as amostras devem ser preparadas e ter características em comum com as utilizadas na calibração.

Wang *et al.* (2018) fizeram uma rápida predição de clorofila e carotenóides em folhas de chá com diferentes adubações nitrogenadas usando o imageamento hiperespectral (captação de uma maior número de bandas contíguas). Eles mostraram que a associação do imageamento a algoritmos de seleção de atributos, levou a um melhor desempenho da predição, ao encontrar faixas do espectro que correspondem as informações desejadas, provando que reduzir a dimensionalidade do conjunto de dados é essencial para um bom desempenho da técnica.

Ji-yonga *et al.* (2012) utilizaram NIRS para determinar o conteúdo total de flavonóides em folhas de *Ginkgo biloba*. Duas regiões de comprimentos de onda característicos foram os principais indicadores para esta estimativa, 1712,33 à 1642,04 nm e 1510,57 à 1453,49 nm, relacionados com o ponto de absorção dos anéis aromáticos dos flavonóides, sendo, desta forma, possível a criação de índices para detecção de flavonóides utilizando estas faixas do espectro.

Os órgãos da planta mais utilizados para aferir nutrição mineral são as folhas, por serem bons indicadores de deficiência e por facilidade de manejo, além de estarem presentes em todas as fases da vida de uma planta. As folhas utilizadas podem estar em estado fresco (WANG *et al.*, 2017), secas (PAILLAT *et al.* 2016) ou secas e maceradas até formar um pó (GALVEZ-SOLA *et al.*, 2015).

Com o uso da técnica NIRS, Galvez-Sola *et al.* (2015) estimaram nitrogênio (N), potássio (K), cálcio (Ca) e magnésio (Mg) em um *pool* de folhas de sete espécies de citrus. Wang *et al.* (2017) utilizaram NIRS em pomares de pêra para estimar a

quantidade de N em plantas adultas sem a necessidade de remoção da folha e de modo instantâneo. Lepine *et al.* (2016) mapearam nitrogênio em florestas com NIRS e estimaram a porcentagem de N em folhas. Os resultados sugerem que grande parte da variação da porcentagem de nitrogênio da copa das árvores está relacionada com a refletância na região do infravermelho próximo, e que a utilização de sensores NIRS é uma alternativa promissora para estimar este nutriente.

Tallada; Ramos (2018) utilizaram espectroscopia no infravermelho próximo e no visível (VIS-NIRS) para fazer estimativa rápida de nitrogênio nas folhas de cultivares de arroz filipinos. Paillat *et al.* (2016) utilizaram NIRS para determinar os conteúdos de nitrogênio e carbono em folhas de cana-de-açúcar (*Saccharum* spp.), comparando diferentes equipamentos e diferentes tipos de preparação das amostras. Este trabalho mostrou que o NIRS também oferece possibilidades analíticas em áreas de cultivo de cana-de-açúcar onde não há laboratórios disponíveis.

Após feita a calibração, correlacionando, com o uso de técnicas estatísticas ou computacionais, a caracterização espectral do alvo com as características desejadas, o método já pode ser utilizado para outras amostras, então, atualmente já existem trabalhos cuja análise nutricional foi feita com NIRS, utilizando calibrações prévias, como no caso de Evans *et al.* (2018), que determinaram o conteúdo total de nitrogênio e fenóis nas folhas de *Posidonia australis* usando a calibração feita previamente por Bain *et al.* (2013).

3.3. Aprendizado de Máquinas e Redes Neurais Artificiais

Atualmente, com o desenvolvimento de novas formas de coleta de dados, estão surgindo problemas não mais correlacionados com a falta de informações e sim com se excesso, uma vez que, a capacidade de processamento destes dados não cresce na mesma velocidade (NASSER; TARIQ, 2015).

Os métodos tradicionais de análise de dados são limitados por diversos fatores, como a falta de especialistas, a subjetividade presente em alguns dados e grandes volumes de dados, conhecidos como *Big Data* (FACELI *et al.* 2011). Estas análises clássicas estão limitadas a extrações de informações diretas, mas não conseguem extrapolar além do conjunto de dados (FACELI *et al.* 2011).

Com a evolução dos métodos de coletas de dados e dos estudos, estão surgindo problemas cada vez mais complexos e sem "solução algorítmica", deste modo, acabou

sendo necessário a modelagem do comportamento de um ser inteligente para possibilitar uma solução (FACELI *et al.* 2011).

Assim, surgiu o aprendizado de máquinas (*Machine Learning*), uma área da inteligência artificial que consiste em técnicas capazes de melhorar seu desempenho, em uma dada tarefa, utilizando experiências prévias (MITCHELL, 1997).

Essa área abrange diferentes algoritmos de processamento de informações, dentre esses estão as redes neurais artificiais (MITCHELL, 1997). Nahar (2012), definiu estas redes como um paradigma de processamento da informação baseado na forma como os sistemas nervosos biológicos processam as informações. Elas são formadas por um conjunto de neurônios artificiais organizados e altamente conectados entre si (NAHAR, 2012). Cada neurônio tem uma função individual, mas trabalham em conjunto para resolver um problema específico (NAHAR, 2012). Deste modo, as redes neurais artificiais, assim como as pessoas, aprendem com base no reconhecimento de padrões que se dá ao ajustar as conexões existentes entre os neurônios (NAHAR, 2012).

Estas possibilitam diversas aplicações na área de ciências de plantas, uma vez que, problemas cada vez mais complexos estão sendo levantados. Na agricultura de precisão, estas técnicas tem sido utilizadas de diversas maneiras, como recomendação de adubação (MODENESE-GORLA DA SILVA *et al.*, 2004) e classificação de estado nutricional (BACKHAUS; BOLLENBECK; SEIFFERT, 2003). No entanto, a aplicação em nutrição mineral de plantas ainda está pouco explorada. Nos poucos trabalhos disponíveis, pode-se perceber resultados promissores desses usos para facilitar e automatizar a vida no campo, além de aumentar a precisão das técnicas atualmente aplicadas ao se associarem com as redes.

3.3.1. Como Máquinas Aprendem?

Assim como humanos, existem diversas formas das máquinas aprenderem, em geral, o tipo de aprendizado está ligado com a função que a máquina exercerá (ADHIKARI; DENERO, 2017). Existem basicamente quatro aplicações do aprendizado de máquinas, regressão, classificação, associação e agrupamento, mas as mais utilizadas são regressão e classificação.

A regressão consiste em prever os resultados em uma saída contínua (HAYKIN, 2001), como por exemplo, prever o preço do dólar amanhã, baseado em um conjunto de dados. A classificação, que consiste em prever os resultados em uma saída discreta

(HAYKIN, 2001), como por exemplo, classificar o preço do dólar em baixo, médio e alto.

Existem quatro métodos de treinamento: o supervisionado, não supervisionado, semi-supervisionado e por reforço, mas os mais utilizados são o supervisionado e o não supervisionado (ADHIKARI; DENERO, 2017).

No treinamento supervisionado, os dados tem uma saída conhecida, se assemelhando a um aprendizado acompanhado de uma criança, mas acaba tendo a dificuldade de extrapolar para situações futuras não presentes no conjunto de dados, desta forma, é necessário um aprendizado contínuo para o aumento constante da máquina (ADHIKARI; DENERO, 2017). É utilizado principalmente para resolver problemas de regressão.

No treinamento não supervisionado, o conjunto de dados tem uma saída desconhecida, então o algoritmo irá explorar a base de dados para assim encontrar uma estrutura neles, se assemelhando ao aprendizado autodidata (ADHIKARI; DENERO, 2017).

3.3.2. O Neurônio Biológico

A unidade funcional do sistema nervoso biológico é o neurônio (KATZ, 1966). Ele foi inicialmente descrito detalhadamente pelo neurologista espanhol Santiago Ramón y Cajal em 1894 (KOVÁCS, 2006). Esta célula é composta com um corpo, ou soma, que é o centro metabólico da célula, e por extensões filamentosas que se projetam do soma, os dendritos e o axônio (KOVÁCS, 2006) (Figura 2). Ela se diferencia das demais por sua membrana, que possui propriedades essenciais para o funcionamento elétrico da célula nervosa e assim transmitir informações de uma célula a outra, a sinapse (CAJAL, 1894).

Os dendritos recebem os sinais elétricos de outros neurônios, estes sinais são processados pelo soma, e se forem fortes o suficiente para atingir o limiar de ativação, irão produzir um potencial de ativação, que levará o sinal elétrico pelo axônio até suas ramificações terminais (KATZ, 1966).

No potencial de ação está a informação que será passada adiante para sair como um impulso nervoso no axônio (KOVÁCS, 2006). Este potencial é constituído de 3 fases, inicialmente a membrana é mantida em um estado de repouso (potencial de repouso), ao receber o estímulo forte o suficiente, ela irá se despolarizar rapidamente (fase ascendente) e posteriormente se repolarizar (fase descendente), nesta repolarização

ocorre uma hiperpolarização, fazendo com que o neurônio entre em um período refratário, de modo que não consegue ser imediatamente reestimulado (HODGKIN; HUXLEY, 1952).

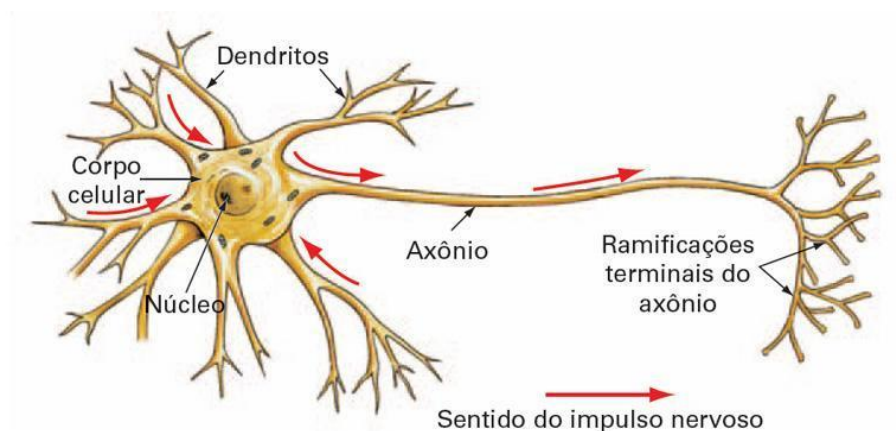


Figura 2: Representação simplificada de um neurônio biológico.
Fonte: Lopez (2000)

3.3.3. O neurônio Artificial

Assim como no sistema nervoso biológico, em que a unidade funcional é o neurônio, nas redes neurais artificiais a unidade funcional é o neurônio artificial. McCulloch e Pitts introduziram o conceito de neurônio artificial em 1943 (MCCULLOCH; PITTS, 1943). Este neurônio é uma modelagem matemática simplificada do neurônio biológico e acabou se tornando o principal componente de processamento para as redes neurais artificiais (HAYKIN, 2001).

Este neurônio artificial possui basicamente 7 componentes principais: Os sinais de entrada ($x_1, x_2, \dots, x_n$), os pesos sinápticos ($w_1, w_2, \dots, w_n$), a função agregadora (Σ), o limiar de ativação (θ), o potencial de ativação (u), a função de ativação ($g(\cdot)$) e o sinal de saída (y) (HAYKIN, 2001) (Figura 3).

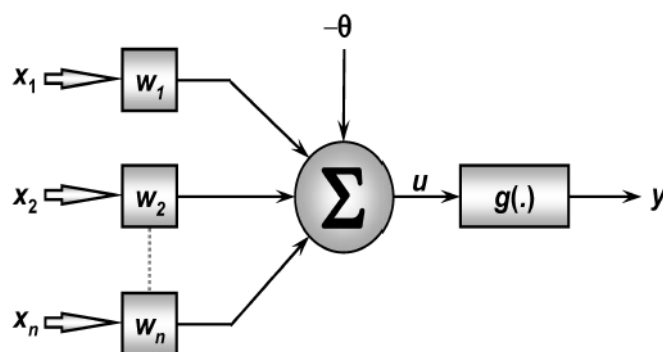


Figura 3: Modelo de neurônio artificial proposto por McCulloch e Pitts com sinais de entrada (x), pesos sinápticos (w), limiar de ativação (θ), função agregadora (Σ), potencial de ativação (u), função de ativação ($g(\cdot)$) e sinal de saída (y).
Fonte: Silva; Spatti; Flauzino (2010).

Nessa figura, os sinais de entrada representam os valores assumidos pelas variáveis do fenômeno a ser modelado (SILVA; SPATTI; FLAUZINO, 2010). Os pesos sinápticos ponderam os sinais de entrada, multiplicando o valor destas por um peso associado antes de ser processado pela função agregadora, que soma estes sinais para produzir um potencial de ativação para o neurônio (HAYKIN, 2001). O limiar de ativação define o valor mínimo que o potencial de ativação, que é a diferença entre o resultado da função de agregação e o limiar de ativação (HAYKIN, 2001), deve atingir para que haja o disparo em direção a saída do neurônio (SILVA; SPATTI; FLAUZINO, 2010). A função de ativação processa o valor do potencial de ativação para gerar o sinal de saída, que será transmitido às suas conexões (KOVÁCS, 2006).

3.3.4. WEKA

Atualmente existem diversos *softwares* nos quais podemos programar e treinar algoritmos de aprendizado de máquinas. Dentre estes, podemos destacar o WEKA, (*Waikato Environment for Knowledge Analysis*), um *software* livre desenvolvido em Java pela Universidade de Waikato.

É uma coleção de algoritmos de aprendizado de máquina voltado para tarefas de mineração de dados e contém ferramentas para preparação e seleção de atributos, classificação, regressão, agrupamento, mineração de regras de associação e visualização de dados.

3.3.4.1. Alguns Algoritmos de Seleção de Atributos

Conjuntos de dados com muitas variáveis (atributos), são difíceis de ser utilizados, pois, algumas destas variáveis oferecem informações irrelevantes e redundantes, gerando um ruído na resposta e tornando esta pouco confiável. Desta forma, os dados de entrada para um algoritmo de aprendizado de máquinas com função de classificação devem passar por um outro tipo de algoritmo para fazer a redução da dimensionalidade, também chamada de seleção de atributos, que identifica e remove o máximo possível de informações redundantes e não importantes, permitindo que os algoritmos de classificação operem de forma mais rápida e efetiva (HALL; SMITH, 1998).

No WEKA, existem diferentes métodos para se fazer seleção, dentre eles, podemos destacar o *Correlation based Feature Selection* (CFS), o *Wrapper* e o *Classifier Subset Evaluator* (CSE).

A hipótese central da seleção de atributos é que um bom conjunto de variáveis é aquele que possui alta correlação com as classes, mas que não são correlacionados entre si (HALL, 2000). O CFS (*Correlation based Feature Selection* - Seleção de atributos com base na correlação), é um algoritmo que acopla essa fórmula de avaliação com uma medida de correlação apropriada e uma estratégia de pesquisa heurística (HALL, 2000). Ele avalia o valor de um subconjunto de atributos considerando a capacidade preditiva individual de cada variável, juntamente com o grau de redundância entre elas, assim, subconjuntos cujas variáveis são altamente correlacionadas com a classe e possuem baixa inter-correlação são preferidos (HALL, 2000). Primeiramente o método calcula uma matriz de correlação atributo-classe e atributo-atributo, depois particiona os subconjuntos e associa um peso (*score*) a cada um, de acordo com a seguinte fórmula:

$$S = \frac{k \times r_{ac}}{\sqrt{k + k(k - 1)r_{aa}}}$$

Onde S é o mérito de um subconjunto de atributos contendo k atributos; r_{ac} é a média da correlação entre atributo-classe e r_{aa} é a média da correlação entre atributo-atributo (HALL, 1998).

O CFS começa com um conjunto vazio de atributos e usa a heurística *Best First* com um critério de parada quando cinco subconjuntos consecutivos não melhoram o mérito, e depois, o subconjunto com o maior mérito é selecionado (HALL, 1998). O método de busca *Best First* permite o retrocesso ao longo do caminho de pesquisa (HALL, 1998). Ele pesquisa o espaço de subconjuntos de atributos por *greedy hill climbing* aumentado com um facilitador de retrocesso (HALL, 1998). Pode começar com o conjunto vazio de atributos e pesquisar para frente, pode começar com o conjunto completo de atributos e procurar para trás, ou começar em qualquer ponto e pesquisar em ambas as direções (considerando todas as possíveis adições e exclusões de atributos únicos em um determinado ponto) (HALL, 1998).

O *wrapper* avalia subconjunto de atributos utilizando um algoritmo de aprendizado de máquinas que funciona como uma caixa preta para encontrar os melhores subconjuntos de atributos (KOHAVI; GEORGE, 1997). O propósito do método é encontrar um conjunto de atributos que melhor se adequa ao algoritmo de

aprendizado, isso torna esta abordagem totalmente dependente do algoritmo de aprendizado utilizado (KOHAVI; GEORGE, 1997).

O CSE funciona de forma parecida com o CFS, pois utiliza de um classificador para avaliar um mérito de um subconjunto de atributos, mas ao invés de calcular o mérito com base na correlação, ele utiliza um classificador (CHIDAMBARAM; SRINIVASAGAN, 2016). Ele pode fazer a sua avaliação utilizando de divisão percentual (*percentage split*) nos dados de treino, ou seja, separando o conjunto de treino em dois e assim, fazer o processo de treinamento duas vezes.

3.3.4.2. Alguns Algoritmos de Classificação

A biblioteca de algoritmos de classificação do WEKA é muito ampla, possuindo desde algoritmos mais simples, como árvores de decisão, até algoritmos mais complexos, como redes *perceptron* multicamadas. Aqui vamos elucidar um pouco mais sobre o ZeroR, Regressão Logística e o *Multi Class Classifier*.

O ZeroR é o algoritmo de aprendizado mais primitivo presente no WEKA (WITTEN *et al.*, 1999). Este algoritmo, se as saídas são nominais, sempre prevê a classe de saída como a classe majoritária presente nos dados de treinamento e, se a saída é numérica, prevê seu valor médio (WITTEN; FRANK, 2005). Ele é utilizado como um "controle", quando o algoritmo testado tem uma performance inferior a do ZeroR, isso é um indicador substancial de *overfitting* (WITTEN *et al.*, 1999).

A análise de regressão é uma técnica de modelagem preditiva utilizada para investigar uma relação entre a variável dependente (Y_i) e uma ou mais variáveis independentes (X_i) (KUMAR; SAHOO, 2012). A regressão logística é uma generalização da regressão linear (MANTARAS; ARMENGOL, 1998) utilizada para estimar variáveis dependentes ordinais, binárias ou multi-classe (KUMAR; SAHOO, 2012). O classificador *Logistic* do WEKA constrói um modelo de regressão logística multinomial com um estimador de crista (*ridge*) (LE-CESSIE; VAN-HOUWELINGEN, 1992). Como a regressão logística original não lida com pesos de instâncias, o algoritmo tem algumas alterações no método tradicional para lidar com esses pesos.

As redes neurais artificiais do tipo *Perceptron* de Multicamadas (PMC - *Multilayer Perceptron*) utilizadas tem como algoritmo de treinamento o *backpropagation*, que é a configuração mais utilizada e apropriada para classificação de padrões (SILVA; SPATTI; FLAUZINO, 2010).

O MCC é um classificador "meta" (*meta-classifier*), os classificadores meta são *proxy's* para o classificador principal, usados para fornecer um pré-processamento de dados adicionais, ou seja, eles funcionam como um filtro que altera os dados de entrada para o algoritmo *core* (algoritmo base), e assim melhoram o desempenho ou ampliam as capacidades de um algoritmo básico (WITTEN *et al.*, 1999). O MCC transforma um problema multi-classe em vários problemas de duas classes e combina seus resultados, além de ser capaz de aplicar códigos de saída com correção de erros para maior precisão (WITTEN *et al.*, 1999).

3.3.5. Aplicações em Ciências de Plantas

Uma das características das redes neurais artificiais que permitem a sua ampla utilização é a sua versatilidade, podendo ser aplicadas com diferentes intuitos, como aproximação de funções, previsão de séries temporais, reconhecimento de padrões e classificação (KOVÁCS, 2006) além de aceitar dados de entrada de diversas fontes, como calorímetros (AQUINO *et al.*, 2016), imagens digitais (PATIL; KUMAR, 2011), constituintes fitoquímicos (SUÁREZ *et al.*, 2015), refletância espectral (ZHANG *et al.* 2011) dentre outros.

Deste modo, Aquino *et al.* (2016) utilizaram dados calorimétricos como entrada de uma RNA do tipo *perceptron* multicamadas para fenotipar de modo indireto o teor de vitamina A em banana. Suárez *et al.* (2015) utilizaram GLM-AID (*General Linear Model Linked to Automatic Interaction Detection* - Modelo linear generalizado vinculado a detecção de interação automática) para identificar os constituintes fitoquímicos relevantes para a caracterização e autenticação de tomates e em seguida uma RNA, para identificar o cultivar, a data de colheita e o tipo de produção. Lu *et al.* (2016) utilizaram RNAs com algoritmos de treinamento do tipo *backpropagation* para modelar a concentração de clorofila *a* em lagos, usando como variáveis de entrada o fósforo dissolvido, fósforo total, nitrogênio total, cloreto, oxigênio dissolvido, pH, profundidade e temperatura da água. Ramos (2003) programou uma rede *Perceptron* de múltiplas camadas, com algoritmo de treinamento tipo *backpropagation* para classificar frutos com base em seus padrões, como peso e diâmetro.

Binoti; Binoti; Leite (2014) testaram diversas arquiteturas de redes neurais artificiais para estimar o volume de árvores tendo como entrada variáveis numéricas e categóricas. Bucene; Rogrigues (2004) classificaram terras para a irrigação, com relação a sua fertilidade aparente (produtividade) em cinco classes de acordo com variáveis do

solo utilizando uma rede *perceptron* multicamadas com o algoritmo de treinamento do tipo *backpropagation*. Eles obtiveram taxas de acerto de 78%, mostrando resultados promissores das redes para esta aplicação.

Apesar de encontrarmos uma ampla gama de aplicações possíveis, ao se comparar as produções voltadas para a área botânica com o total de artigos produzidos em outras áreas, podemos perceber que ainda são escassas, principalmente para nutrição mineral de plantas.

Qiu-Xiang *et al.* (2007) utilizaram sensoriamento remoto associado a redes neurais artificiais para monitorar a concentração foliar de nitrogênio em lavouras de arroz. Eles mostraram que existe um potencial promissor para a aplicação destas técnicas associadas, além de que, ao reduzir a dimensão dos dados de entrada com a análise de componentes principais (ACP) o acerto das redes aumenta, ou seja, quando se reduz os dados de entrada, a qualidade do ajuste melhora.

Backhaus; Bollenbeck; Seiffert (2003) associaram diversas arquiteturas de redes neurais artificiais a 3 câmeras hiperespectrais para fazer o monitoramento nutricional de mudas de tabaco em condição de viveiro. Eles obtiveram bons desempenhos, principalmente com as redes rRBF, que são adaptações das rede de base radial (RBF - *radial basis network*), e também mostraram que ao decompor as imagens e criar um subespaço, se aumenta a taxa de acerto das redes do tipo máquinas de vetor suporte (*Support Vector Machine* - SVM).

Modense-Gorla da Silva *et al.* (2004) utilizaram redes *perceptron* multicamadas com diferentes topologias para recomendar adubação de cálcio para goiabeira com base em uma série temporal de concentração do nutriente no solo. Eles obtiveram classificações bem precisas, a topologia que apresentou menor erro quadrático médio foi a de 20 neurônios nas camadas intermediárias.

Desta forma pode-se perceber que, apesar das redes neurais artificiais serem pouco utilizadas, apresentam resultados promissores para a área de nutrição mineral, podendo ser aplicadas para diferentes funções. Elas tem uma grande flexibilidade de dados de entrada, podendo processar dados coletados de diversas fontes com diferentes equipamentos. Também possuem alto nível de precisão, e podem fazer parte de sistemas autônomos, integrando diferentes conjuntos de dados e auxiliando a tomada de decisões em campo e em viveiros.

3.4. Análise Estatística

3.4.1. Análise Multivariada

As análises multivariadas levam em consideração mais de uma medida ou variável do mesmo indivíduo concomitantemente, desta forma, consegue se obter mais informações a respeito do problema tratado (HAIR *et al.*, 1995).

Em geral, as análises multivariadas são utilizadas com propósitos distintos: redução da dimensionalidade (simplificação), ordenação, agrupamento, classificação ou teste de hipóteses.

As técnicas de redução da dimensionalidade são geralmente aplicadas em conjuntos com muitas variáveis. Estes são difíceis de se analisar pois algumas destas variáveis dão informações redundantes, irrelevantes e insignificantes e acabam gerando ruído. Dentre as técnicas de redução da dimensionalidade, podemos destacar a análise de componentes principais - ACP (JOHNSON; WICHERN, 2014) e a metodologia *stepwise*.

A ACP se fundamenta na construção de combinações lineares das variáveis originais. Essas combinações lineares, denominadas de componentes principais, possuem propriedades ótimas em termos de variância, buscando-se novas variáveis (coordenadas) que maximizem a variância e não sejam correlacionadas entre si (MINGOTI, 2005).

O procedimento *proc stepdisc* do programa SAS - *Free Statistical Statistical Software*, SAS University Edition seleciona um subconjunto de variáveis quantitativas por análise discriminante linear, pressupondo que as variáveis possuem distribuições normais multidimensionais e matriz de variância comum. Permite a seleção *forward* (o modelo começa vazio e aos poucos as variáveis vão sendo incluídas e retiradas segundo um critério definido) e *backward* (o modelo se inicia com todas as variáveis inclusas e então vão sendo retiradas e colocadas segundo um critério definido) (MINGOTI, 2005). Os critérios escolhidos podem ser o nível de significância do teste F da análise de covariância ou o coeficiente de correlação parcial do ajuste do modelo considerando as variáveis já selecionadas anteriormente.

Já as técnicas de classificação, permitem classificar os conjuntos de dados em classes pré-definidas. Podemos destacar a técnica da função linear discriminante de Fisher (JOHNSON; WICHERN, 2014). Esta se baseia na classificação de k grupos através de funções lineares no espaço multidimensional, utilizando-se de métricas como

distâncias euclidianas e distância de Mahalanobis (que leva em consideração a matriz de variância e covariância das observações). A função linear discriminante (FLD) considera variâncias comuns (“*pool*”), com matriz de covariância comum nas distribuições normais multivariadas, já a função quadrática discriminante (FQD) considera variâncias distintas nas distribuições normais multivariadas.

3.4.2. Simulação *Bootstrap*

A técnica de simulação *bootstrap* (EFRON, 1979) para reamostragem é amplamente aplicada, em diferentes situações. Em suma, consiste em obter novos conjuntos (subconjuntos) de dados por reamostragem do conjunto de dados original (EFRON; TIBISHIRANI, 1993). Esta reamostragem pode ser feita diretamente ou via um modelo ajustado.

Ao ser empregada, permite assim, avaliar a variabilidade, pois, não teremos mais uma medida única de acerto das técnicas, ou um único valor para os parâmetros e sim uma média com suas medidas de variabilidade, sem usar cálculos analíticos (LEPAGE; BILLARD, 1992).

3.4.3. Modelos Lineares Generalizados

Os modelos lineares generalizados (MLG) foram criados por Nelder e Wedderburn (1972) para reunir vários modelos estatísticos que eram tratados separadamente em uma mesma família. São uma extensão dos modelos de regressão linear clássicos (MCCULLAGH; NELDER, 1989). Eles relacionam uma função de ligação da média da distribuição da variável aleatória dependente (função de distribuição) com o preditor linear (parte não aleatória) através da função de ligação (MCCULLAGH; NELDER, 1989).

Os MLG mais conhecidos são a regressão gama, regressão poisson, e a regressão logística. A regressão gama são utilizados para modelar dados contínuos, positivos e assimétricos e geralmente a função de ligação utilizada é a logarítmica. A regressão poisson é utilizada para modelar dados discretos, de contagem, e a principal função de ligação utilizada também é a logarítmica. A regressão logística é utilizada na modelagem de dados discretos binomiais, e a principal função de ligação utilizada é a logito.

Quando se ajustam modelos lineares generalizados com distribuição binomial e função de ligação *logito*, é chamada de regressão logística binomial, já que se consideram as variáveis resposta binárias. Quando se ajustam modelos lineares generalizados com distribuição multinomial e função de ligação *glogito*, é chamada de regressão logística multinomial, já que se consideram as variáveis resposta multinomiais.

3.5. Índices Espectrais de Vegetação

Dados espectrais são multivariados e contém muitas informações, que abrangem diferentes aspectos do alvo, dificultando a interpretação. Quando se deseja aferir apenas uma característica da planta ou floresta, é interessante a utilizar índices espectrais de vegetação (IEV), pois simplificam a análise e a interpretação destes conjuntos complexos, indo diretamente atrás da informação desejada (parâmetros aos quais estão correlacionados).

Existem atualmente diversos IEV's. Estes são calculados a partir de dados espectrais e estão correlacionados a parâmetros específicos da planta, como água, pigmentos, biomassa, etc. Como são combinações de comprimentos de onda específicos, ao se utilizar índices, em geral, ocorre uma redução na quantidade de variáveis de entrada, já que cada uma dessas novas variáveis carrega uma maior quantidade de informações.

Alguns dos IEV mais conhecidos são o NDVI (*normalized difference vegetation index* - Índice de diferença de vegetação normalizado), que é um indicador de biomassa e área foliar (ROUSE *et al.*, 1974), o SR (*simple ratio* - relação simples), indicador de biomassa (BIRTH; McVey, 1968), o mSR (*Modified simple ratio* - relação simples modificada), indicador de biomassa (SIMS; GAMON, 2002), o ND (*normalized difference index* - índice de diferença normalizada) que é indicador do conteúdo de clorofila (GITELSON; MERZLYAK, 1994), o PRI (*photochemical reflection index* - índice de reflexão fotoquímica), que é um indicador de pigmentos e eficiência de uso de radiação fotossintética (GAMON; PEÑEULAS; FIELD, 1992) e os PSSR (*pigment specific simple ratio* - relação simples específica de pigmento) e PSND (*pigment specific normalized difference* - diferença normalizada específica do pigmento), indicadores de clorofila a, clorofila b e carotenóides (BLACKBURN, 1998). Todos estes índices são formados pela divisão de um comprimento de onda pelo outro ou pela divisão da diferença de dois comprimentos de onda pela sua soma.

Dentre os índices mais renomados na literatura, são escassos os relacionados exclusivamente com o conteúdo de N nas folhas, quando se deseja aferir sobre este elemento, geralmente se utilizam índices correlacionados com a cor das folhas (concentração de clorofilas) e com a área foliar e biomassa, que são sinais da falta ou excesso deste elemento na planta, desta forma, são aplicáveis apenas em casos onde a já se estabeleceram danos as plantas. Também são escassos os índices que utilizem comprimentos de onda presentes na faixa do infravermelho próximo, especialmente para aferir N.

3.6. A família Annonaceae

A família Annonaceae foi, em geral, originada na América do Sul e Antilhas (PINTO *et al.*, 2005) e é uma das maiores famílias dos trópicos, com cerca de 135 gêneros e 2500 espécies (ESCRIBANO; VIRUEL; HORMAZA, 2007). Destas espécies, aproximadamente 386 são nativas do Brasil (LOPES; MELLO-SILVA, 2014), fazendo parte da composição de diversas fitofisiomias, como o cerrado (*Annona crassiflora* Mart.) e áreas alagadas (*Annona glabra* L.). Apresenta grande importância evolutiva, ecológica e econômica (CHATROU *et al.*, 2012).

A família é rica em óleos e outros metabólitos especializados, como alcaloides e acetogeninas de grande importância medicinal, por sua ação anti-inflamatória, analgésica e anti-hipertensiva, potencial antiproliferativo em células cancerígenas, além de importância na agricultura como inseticida e fungicida (LEBOEUF *et al.*, 1982; ATTIQ *et al.*, 2017; PIEME *et al.*, 2014).

Algumas de suas espécies possuem frutos explorados comercialmente, como o biribá (*Rollinia deliciosa* Saff), a anona-lisa ou condessa (*Annona reticulata* L.), a llama ou papuse (*Annona diversifolia* Saff), o araticum-de-casca-lisa (*Annona coriacea* Mart.) e o marolo (*Annona crassiflora* Mart.), mas no Brasil, as principais produções comerciais são as de graviola (*Annona muricata* L.), fruta-do-conde, ata ou pinha (*Annona squamosa* L.), chirimoia (*Annona cherimola* Mill.) e do híbrido atemoia (*Annona* X *atemoya* Mabb.) (SOBRINHO, 2010).

O Brasil é um dos principais produtores mundiais, junto com os Estados Unidos da América, Índia, Espanha, Chile, México e Venezuela (DONADIO, 1997). Aqui a demanda pela polpa das frutas de Anonáceas está em constante expansão, o que estimula pequenos e grandes produtores a investir nas culturas (ROZANE; NATALE, 2014).

Em geral, as espécies de anonáceas exploradas comercialmente são propagadas por enxertia, o que garante pomares homogêneos, com alta produção frutos com elevada qualidade (TOKUNAGA, 2005; SAVAZAKI, 2013). Dentre os porta-enxerto mais utilizados nas principais culturas brasileiras, está a *Annona emarginata* (Schltdl.) H. Rainer, que garante grande duração da muda no campo, e apresenta resistência para diversas podridões de colo e raiz e coleobrocas (TOKUNAGA, 2005), garantindo copa vigorosa (SCALOPPI JUNIOR, 2007).

3.6.1. Estudos Nutricionais com a Família

Os estudos nutricionais para esta família são escassos (ROZANE; NATALE, 2014). Dentre estudos convencionais com nutrição mineral em plantas jovens de Annonaceae, relatam-se os resultados com mudas de gravioleira (*Annona muricata* L.) no qual, para o crescimento, N, P, Ca e Mg são limitantes e K, S, B e Zn, tem uma demanda nutricional pequena (SILVA; FARNESSI, 2009), e que, após a repicagem, as mudas apresentam inicialmente uma taxa baixa de crescimento, mas, a partir de 105 dias a taxa de absorção dos nutrientes aumenta e consequentemente o crescimento das plantas (BARBOSA *et al.*, 2003).

Em relação a outros trabalhos com nutrição em mudas de Annonaceae, relata-se Baron *et al.* (2013), que verificaram que o uso de 50% da força iônica da solução nutritiva Hoagland & Arnon nº2 para o cultivo de mudas de *Annona emarginata* (Schltdl.) H. Rainer, em condições de viveiro, promoveram maiores valores de assimilação de carbono, eficiência do uso da água, eficiência da enzima Rubisco, taxa de crescimento relativo, taxa de assimilação líquida de carbono e acúmulo iônico das folhas, provando adaptação à redução da disponibilidade nutricional quando comparada a força iônica de 100%, o que permite reduzir a quantidade de elementos minerais na produção de mudas. Para esta espécie, o maior acúmulo de nitrogênio ocorreu em plantas mantidas em solução com 50% da força iônica.

Batista *et al.* (2003) estudaram o efeito da omissão de macronutrientes no crescimento, nos sintomas de deficiências nutricionais e na composição mineral em mudas de gravioleiras (*Annona muricata*). Eles utilizaram hidroponia para o estudo e provaram que a omissão de N, P, K, Ca, Mg e S na solução nutritiva resulta no aparecimento de sinais visuais (alterações morfológicas) nas mudas e que geram redução no crescimento e produção de matéria seca em mudas de gravioleira.

Costa *et al.* (2019) avaliaram o efeito de diferentes doses de N e K no crescimento de mudas de *Annona crassiflora*. Mostraram que a adubação nitrogenada teve efeito positivo no crescimento das mudas em doses entre 100 and 200mg dm⁻³, e que a adubação potássica teve melhores efeitos com doses de 150 mg dm⁻³. Souza *et al.* (2016) avaliaram o efeito de N e K no desenvolvimento dos botões florais de atemoia (*Annona X atemoya*) em duas épocas distintas, primavera-verão e outono-inverno. Não houve interação entre N e K para o vigor dos botões florais, e o K teve efeito no mesmo. Com relação no N, relataram que vigor dos botões florais tendeu a aumentar com a elevação das doses no período mais frio do ano, já no período mais quente, ocorreu a redução do vigor conforme aumentaram as doses de N.

Baron *et al.* (2017) observaram, em *Annona emarginata* (Schltdl.) H. Rainer, que o acúmulo de matéria seca e as trocas gasosas se alteram em função da força iônica, e isto causou um acúmulo diferencial de Fe, Mn e Zn, sendo que as plantas cultivadas com 75% da força iônica tiveram efeitos mais positivos. Também registraram que existe a relação inversa entre o aumento da força iônica e o teor de Fe e Mn e a relação direta entre a força iônica e a concentração de Zn.

Trabalhando com adubação nitrogenada e boro em plantas adultas de *Annona*, Costa *et al.* (2002) observaram que o peso médio dos frutos de pinha (*Annona squamosa* L.) variou em função do nível da adubação nitrogenada e também aumentou o número de flores e frutos.

Lima *et al.* (2007) avaliaram o efeito da amostragem para fazer a diagnose foliar em *Annona muricata*. Coletaram folhas de três posições diferentes da copa e dos ramos e então determinaram os teores de P, K, Ca, Mg, S, Fe, Cu, Zn e Mn, chegando a conclusão que as folhas da parte mediana, tanto dos ramos quanto da copa, representam melhor o estado nutricional das plantas.

3.6.2. *Annona emarginata* (Schltdl.) H. Rainer

Annona emarginata (Schltdl.) H. Rainer, é uma espécie nativa da América do Sul, que ocorre em diversas partes do Brasil (MAAS *et al.*, 2001), em áreas de cerrado e de Mata Atlântica . Segundo a Flora do Brasil (2020), já teve sua ocorrência confirmada nas regiões Nordeste (Bahia), Centro-Oeste (Distrito Federal, Mato Grosso do Sul, Mato Grosso), Sudeste (Minas Gerais, São Paulo) e Sul (Paraná, Rio Grande do Sul, Santa Catarina).

Possui habito arbóreo de pequeno a médio porte, é caducifólea, suas folhas são lisas e lanceoladas, suas flores brancas em formato de hélice (LORENZI, 2009) e seus frutos são cordiformes e lisos, sincárpicos amarelos com grande variação de tamanho (MAAS *et al.*, 2001).

Segundo a Lista vermelha da flora do Rio Grande do Sul, está considerada em estado vulnerável (CONSEMA-RS, 2002), então atualmente é cultivada com o intuito de reflorestamento.

Tem importância econômica devido ao seu uso para fazer caixotes de madeira e a larga utilização como porta-enxerto para a produção de mudas de atemoia. Devido a sua rusticidade, consegue adaptar-se a diferentes condições ambientais de solos "encharcados" e secos (MANTOAN *et al.*, 2015), além uma maior resistência a pragas como broca do caule (TOKUNAGA, 2005; KAVATI, 2013). Ela também apresenta, no óleo essencial de suas folhas, substâncias com ação anti-inflamatória (CAMPOS *et al.*, 2014).

4. Capítulo 1 - Identificação da falta de nitrogênio em mudas de *Annona emarginata* (Schltdl.) H. Rainer antes do aparecimento de sinais visuais utilizando NIRS.

4.1. Resumo

O monitoramento de nitrogênio é essencial em viveiros para garantir a produção de mudas de alta qualidade, e pode ser feito por métodos diretos e indiretos. Dentre os indiretos podemos destacar a Espectroscopia no Infravermelho Próximo (*Near-Infrared Spectroscopy* - NIRS), que é instantânea e não destrutiva, mas que precisa de calibração para adequado funcionamento. Desta forma o objetivo deste trabalho foi avaliar se a NIRS é efetiva para detectar variações da concentração de nitrogênio em folhas de plantas jovens de *Annona emarginata* (Schltdl.) H. Rainer, assim como sugerir índices a partir de reflectâncias foliares no infravermelho próximo que prevejam com precisão o teor de nitrogênio. O experimento foi instalado em sistema de hidroponia, com alterações na concentração de nitrogênio: 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4), com 60 repetições para cada tratamento. Ao longo do experimento, foram analisadas as variáveis das trocas gasosas e fluorescência da clorofila *a*. Após 45 dias de exposição aos tratamentos, foram analisados os pigmentos foliares, a quantificação de nitrogênio pelo método convencional e realizadas as caracterizações. Para os dados espectrais foi feita a redução da dimensionalidade e classificação com técnicas multivariadas. Os índices espectrais de vegetação foram sugeridos com os quatro últimos comprimentos de onda aferidos. A combinação de redução por *stepwise* e classificação com função linear discriminante chegou aos melhores resultados, com 93,55% de acerto para classificação nos quatro tratamentos e 100% de acerto para classificação da falta, ou não de N. Concluimos que a NIRS é efetiva para detectar variações de nitrogênio em folhas de plantas jovens de *Annona emarginata* (Schltdl.) H. Rainer e que há uma combinação de índices que permite maior acuracidade para prever as variações N.

Palavras chave: nutrição mineral de plantas, *bootstrap*, índice espectral de vegetação.

(*) Artigo escrito seguindo as normas ABNT, pois ainda não foi definida a revista para a qual será submetido.

4. Chapter 1 - Identification of nitrogen deficiency in *Annona emarginata* (Schltdl.)

H. Rainer seedlings before visual signs appearance using NIRS.

4.1. Abstract

Nitrogen monitoring is essential in nurseries to ensure the production of high quality seedlings, and can be done by direct and indirect methods. Indirect ones include Near-Infrared Spectroscopy (NIRS), which is instantaneous and non-destructive, but needs calibration for proper operation. Thus, the objective of this work was to evaluate if NIRS is effective to detect variations of nitrogen concentration in leaves of seedlings of *Annona emarginata* (Schltdl.) H. Rainer, as well as to suggest indexes from near infrared leaf reflectances that predict with accuracy the nitrogen content. The experiment was implemented in a hydroponics system, with modifications in nitrogen concentration: 0 mg.L⁻¹ of N (T1); 52.5 mg.L⁻¹ of N (T2); 105 mg.L⁻¹ N (T3) and 210 mg.L⁻¹ N (T4), with 60 replications for each treatment. Throughout the experiment, the variables of gas exchange and chlorophyll *a* fluorescence were analyzed. After 45 days of exposure to treatments, leaf pigments, nitrogen quantification by conventional method and leaf spectral characterization were analyzed. Were performed for the spectral data dimensionality reduction and classification with multivariate techniques. Vegetation spectral indexes were suggested with the last four measured wavelengths. The combination of stepwise reduction and classification with discriminant linear function reached the best results, with 93.55% of correct answers for classification in the four treatments and 100% of correct answers for classification, whether or not N. We conclude that NIRS is effective to detect nitrogen variations in leaves of *Annona emarginata* seedlings and that there is a combination of indexes that allows greater accuracy to predict N variations.

Key words: plants mineral nutrition, *bootstrap*, vegetation spectral index.

4.2. Introdução

A família Annonaceae tem distribuição pantropical e apresenta grande importância evolutiva, ecológica e econômica (CHATROU *et al.*, 2012). *Annona emarginata* (Schltdl.) H. Rainer é uma espécie arbórea considerada em estado vulnerável segundo a Lista vermelha da flora do Rio Grande do Sul (CONSEMA-RS, 2002). É produzida em viveiros com o intuito de reflorestamento, recuperação de áreas degradadas e para serem utilizadas como porta-enxerto de anonáceas frutíferas, como a atemoia (TOKUNAGA, 2005).

No atual contexto global de mudanças climáticas, os trópicos estão muito propensos a alterações no regime de chuvas e temperatura (WALSH *et al.*, 2015), além do aumento da concentração de CO₂ na atmosfera. Estas mudanças afetam o metabolismo das plantas, alterando sua respiração (RYAN, 1991) e o balanço de alguns nutrientes, como o nitrogênio (RENNENBERG *et al.*, 2009).

O monitoramento da concentração de nutrientes de uma planta é muito importante, tanto em situações de campo como em laboratório. Dentre os elementos minerais, o nitrogênio (N) destaca-se dos demais por ser vital na síntese de muitas moléculas de grande importância, como aminoácidos, ácidos nucleicos (SOUZA; FERNANDES, 2006), proteínas, como a *Rubisco* (LAWLOR, 1989) e metabólitos especializados (MARSCHNER, 2012), sendo portanto um importante elemento associado ao crescimento das plantas. Por ser formador da *Rubisco*, é essencial para a fotossíntese da planta, desta forma, sua falta ou excesso afetará as trocas gasosas e a fluorescência da clorofila *a* (BARON *et al.*, 2013, KALAJI *et al.*, 2018).

Novos métodos indiretos estão surgindo para se aferir a composição mineral de plantas, como a Espectroscopia no Infravermelho Próximo (*Near-Infrared Spectroscopy* - NIRS), que é instantânea e não destrutiva. Ela se relaciona com a natureza de ligação das moléculas de uma amostra, assim, com base na caracterização espectral de uma amostra, é possível aferir a sua composição molecular (FONSECA *et al.*, 2002). Mas, é necessário comparar os padrões de reflectância adquiridos em um alvo com um padrão, fazendo assim uma calibração do método (OTTO, 1999).

Para contornar esta necessidade de calibração, muitas vezes são utilizados Índices espectrais de vegetação (IEV). Estes são combinações de comprimentos de onda específicos que estão relacionados a parâmetros da planta que desejamos aferir. Em geral, os IEV's mais conhecidos que são utilizados para aferir N são formados por comprimentos de onda localizados no espectro visível e correlacionados aos sinais

visuais (sintomas) da falta deste mineral nas plantas. Neste trabalho são propostos índices formados por comprimentos de onda localizados no infravermelho próximo e que estão correlacionados com a concentração de nitrogênio na planta e não com os sinais de sua falta.

Este trabalho visa avaliar se a espectroscopia no infravermelho próximo (NIRS) é efetiva para detectar variações da concentração de nitrogênio em folhas de plantas jovens de *Annona emarginata* (Schltdl.) H. Rainer, assim como sugerir índices a partir de reflectâncias foliares no infravermelho próximo que prevejam com acurácia o teor de nitrogênio, para assim poder inferir sobre uma possível deficiência ou falta deste mineral.

4.3. Material e Métodos

Experimento e Coleta de Dados

O experimento foi conduzido no Departamento de Botânica, do Instituto de Biociências, UNESP, Campus de Botucatu/SP, em casa de vegetação tipo “*Paddy Fan*” sem controle de temperatura. As mudas de *Annona emarginata* (Schltdl.) H. Rainer ‘araticum-de-terra-fria’ foram obtidas no Núcleo de Produção de mudas de São Bento do Sapucaí, da CATI (Coordenadoria de Assistência Técnica e Integrada) cidade de São Bento do Sapucaí.

As plantas foram mantidas em sistema de hidroponia, em vasos com aproximadamente 1,5 litros de solução nutritiva Hoagland & Arnon nº 2 (1950), com força iônica de 50% (BARON, *et al.*; 2013). Após um período de adaptação, foram aplicados os tratamentos, constituídos por variações na concentração de nitrogênio na solução, durante 45 dias. Os quatro tratamentos foram: 0 mg.L⁻¹ de nitrogênio (T1); 52,5 mg.L⁻¹ de N (T2), que equivale a 50% da concentração recomendada para espécie (BARON, *et al.*; 2013); 105 mg.L⁻¹ de N (T3), a dose recomendada para a espécie (BARON, *et al.*; 2013) e 210 mg.L⁻¹ de N (T4), que é o dobro da dose recomendada para a espécie (BARON, *et al.*; 2013) (Tabela 1). As soluções nutritivas foram arejadas constantemente com sopradores e trocadas a cada 2 semanas. Também tiveram o pH e condutibilidade elétrica medidos e controlados diariamente.

O delineamento experimental foi inteiramente casualizado. Foram utilizadas 240 mudas, separadas em 60 por tratamento. A partir dessas mudas foram realizadas as

análises fisiológicas (trocas gasosas, fluorescência, pigmentos foliares e análise foliar completa) e a caracterização espectral das folhas.

Tabela 1: Tratamentos (T1, T2, T3, e T4) com solução nutritiva de Hoagland e Arnon n°2, diluída a 50% da força iônica, com diferentes concentrações de nitrogênio, aos quais foram submetidas as plantas de *A. emarginata* cultivadas.

	Solução estoque (g.L ⁻¹)	0 mg.L ⁻¹ de N (T1)	52,5 mg.L ⁻¹ de N (T2)	105 mg.L ⁻¹ de N (T3)	210 mg.L ⁻¹ de N (T4)
Macronutrientes		ml.L ⁻¹	ml.L ⁻¹	ml.L ⁻¹	ml.L ⁻¹
M NH ₄ H ₂ PO ₄	115,03	-	0,25	0,5	0,5
M KNO ₃	101,1	-	1,5	3	3
M Ca(NO ₃) ₂	236,16	-	1	2	2
M MgSO ₄	246,5	1	1	1	1
0,05M Ca(H ₂ PO ₄) ₂	11,70	5	5	-	-
0,01M CaSO ₄	1,36	100	75	-	-
0,5M K ₂ SO ₄	87,13	2,5	1,5	-	-
M NH ₄ NO ₃	80,04	-	-	-	3,75
Micronutrientes		ml.L ⁻¹	ml.L ⁻¹	ml.L ⁻¹	ml.L ⁻¹
H ₃ BO ₃	2,86	0,5	0,5	0,5	0,5
MnCl ₂ .4H ₂ O	1,81				
ZnSO ₄ .7H ₂ O	0,22				
CuSO ₄ .5H ₂ O	0,08				
H ₂ MoO ₄ .H ₂ O	0,02				
Solução de Fe		ml.L ⁻¹	ml.L ⁻¹	ml.L ⁻¹	ml.L ⁻¹
EDTA Bissódico (C ₁₀ H ₁₄ N ₂ O ₈ Na ₂ .2H ₂ O)	26,10	0,5	0,5	0,5	0,5
FeSO ₄ .7H ₂ O	24,90				

*M = Molar

Ao longo do experimento, a partir da aplicação dos tratamentos, a cada 15 dias, em 5 plantas de cada tratamento, avaliou-se fluorescência da clorofila *a* (energia quântica efetiva - Φ PSII; energia quântica potencial - F_v'/F_m' ; quentin fotoquímico - qP; taxa de transporte de elétrons - ETR; dissipação de energia na forma de calor pelas antenas - D e a energia fotoquímica não dissipada e não utilizada na fase fotoquímica - Ex (DEMMING-ADAMS *et al.*, 1996)) e de trocas gasosas (taxa de assimilação de CO₂ - A; taxa de transpiração - E; condutância estomática - Gs e concentração interna de CO₂ na folha - ci) com o sistema fechado portátil de fotossíntese (IRGA - *Infra Red Gas Analyser*), e a partir destas medidas foram calculadas a eficiência do uso da água (EUA), pela relação entre assimilação de CO₂ e taxa de transpiração e, a eficiência instantânea de carboxilação da enzima ribulose 1,5-difosfato carboxilase (Rubisco)

(A/ci) pela relação da taxa de assimilação de CO₂ e concentração interna de CO₂ na folha (ZHANG *et al.*, 2001).

Após 45 dias nas soluções com os respectivos tratamentos, de cada planta foram retiradas aleatoriamente 3 folhas maduras, localizadas na parte mediana do caule (LIMA *et al.*, 2007), para realizar a caracterização espectral. A caracterização espectral das folhas foi feita com espectroradiômetro da *Ocean Optics*, modelo *Flame* - NIR, com faixa de leitura de 966,03 a 1685,09 nm, e sensibilidade de 5,51 nm.

Para averiguar a efetividade dos tratamentos aplicados foi feita a quantificação de nitrogênio nas folhas pelo método de análise de tecido vegetal (folha) no laboratório de análise de solos do departamento de Ciência do Solo, UNESP, da fazenda experimental do Lageado, Botucatu, SP. Foram usadas para cada tratamento quatro repetições formadas por um *pool* aleatório de folhas de 15 mudas separadas por sorteio.

Também foi feita a quantificação de pigmentos foliares pelo método acetona-TRIS (SIMS; GAMON, 2002), utilizando 4 repetições, de 3 folhas de mudas selecionadas por sorteio, para cada tratamento.

Análise de dados

Variáveis fisiológicas

Para as análises fisiológicas, foram ajustados modelos lineares generalizados com distribuição gama e função de ligação logarítmica para cada variável considerando os tratamentos como fatores. A qualidade do ajuste foi feita inicialmente pela análise dos desvios por graus de liberdade e posteriormente pelos gráficos dos resíduos de Pearson padronizados (NELDER; WEDDERBURN, 1972). Para comparações entre tratamentos foi utilizado o teste de Tukey-Kramer (WESTFALL *et al.*, 1999) do procedimento *genmod* do programa estatístico SAS – *Free Statistical Software*, SAS University Edition.

Reflectância foliar

Para os dados de reflectância foliar (NIRS), primeiramente feita a média das 3 folhas, para se obter uma resposta por planta. Em seguida, foi feita uma amostragem casual simples com 1000 (mil) repetições de 70% do conjunto de dados para se fazer a modelagem e os 30% restantes para fazer a classificação, utilizando a técnica de simulação *bootstrap* (EFRON, 1979), com o comando *proc surveyselect* do programa SAS - *Free Statistical Software*, SAS University Edition. A classificação foi

feita de duas formas: considerando 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e 2 classes (Com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4).

Para facilitar a modelagem, foi feita a redução da dimensionalidade dos dados de refletância utilizando três técnicas distintas: a Análise de Componentes Principais (ACP) (JOHNSON; WICHERN, 2014), a metodologia *stepwise* (procedimento *proc stepdisc* do programa SAS - *Free Statistical Statistical Software*, SAS University Edition) e modelos lineares generalizados com distribuição binomial e função de ligação *logito* (regressão logística binomial) considerando as variáveis resposta binárias (com e sem falta de N) e modelos lineares generalizados com distribuição multinomial e função de ligação *glogito* (regressão logística multinomial) considerando as variáveis resposta multinomiais (Concentrações de N - Tratamentos) utilizando o procedimento *proc logistic* do programa SAS - *Free Statistical Statistical Software*, SAS University Edition.

Posteriormente foi feita a validação do modelo (classificação), com a técnica da função linear (FLD) e quadrática (FQD) discriminante de Fisher (JOHNSON; WICHERN, 2014).

Índices Espectrais de Vegetação - IEV

A partir da combinação dois a dois dos quatro últimos comprimentos de onda captados pelo equipamento (1685,089 nm, 1679,279 nm, 1673,471 nm e 1638,673 nm), foram sugeridos 24 índices espectrais de vegetação (IEV), se baseando na arquitetura dos principais IEV's presentes na literatura: divisão de um comprimento de onda pelo outro, divisão da soma de dois comprimentos de onda pela sua subtração e divisão da subtração de dois comprimentos de onda pela sua soma:

$$\begin{aligned}
 I1 &= \frac{1685.089}{1679.279}; I2 = \frac{1685.089}{1673.471}; I3 = \frac{1685.089}{1638.673}; I4 = \frac{1679.279}{1673.471}; I5 = \frac{1673.471}{1638.673}; \\
 I6 &= \frac{1679.279}{1638.673}; I7 = \frac{1679.279}{1685.089}; I8 = \frac{1673.471}{1685.089}; I9 = \frac{1638.673}{1685.089}; I10 = \frac{1673.471}{1679.279}; \\
 I11 &= \frac{1638.673}{1673.471}; I12 = \frac{1638.673}{1679.279}; I13 = \frac{1685.089-1679.279}{1679.279 + 1685.089}; I14 = \frac{1685.089-1673.471}{1673.471 + 1685.089}; \\
 I15 &= \frac{1685.089-1638.673}{1638.673 + 1685.089}; I16 = \frac{1679.279-1673.471}{1673.471 + 1679.279}; I17 = \frac{1679.279-1685.089}{1685.089 + 1679.279}; \\
 I18 &= \frac{1673.471-1638.673}{1638.673 + 1673.471}; I19 = \frac{1679.279 + 1685.089}{1685.089-1679.279}; I20 = \frac{1673.471 + 1685.089}{1685.089-1673.471}; \\
 I21 &= \frac{1685.089-1638.673}{1638.673 - 1685.089}; I22 = \frac{1673.471 + 1679.279}{1679.279 - 1673.471}; I23 = \frac{1638.673 + 1679.279}{1679.279 - 1638.673}; \\
 I24 &= \frac{1638.673 + 1673.471}{1673.471-1638.673}
 \end{aligned}$$

Após calculados, foram selecionados os índices mais representativos utilizando a função *proc stepdisc* e regressão logística multinomial e então testados na classificação com função linear discriminante e função quadrática discriminante. Desta forma, propomos a utilização de um pequeno conjunto de índices para se aferir sobre a concentração de N em folhas de *A. emarginata* antes do aparecimento de sinais visuais, ou seja, antes que estas concentrações fora do recomendado causem danos as mudas.

4.4. Resultados

Reflectância foliar

Com as curvas de reflectância foliar média para cada tratamento (figura 1), observa-se que as mudas sem N (T1 - vermelho) e as mudas com 50% da dose recomendada de N para a espécie (T2 - laranja) se agrupam em uma curva abaixo - com falta de N e que as mudas com a dose de N recomendada (T3 - azul) e as mudas com o dobro de N (T4 - verde) se agrupam em uma curva acima - sem falta de N.

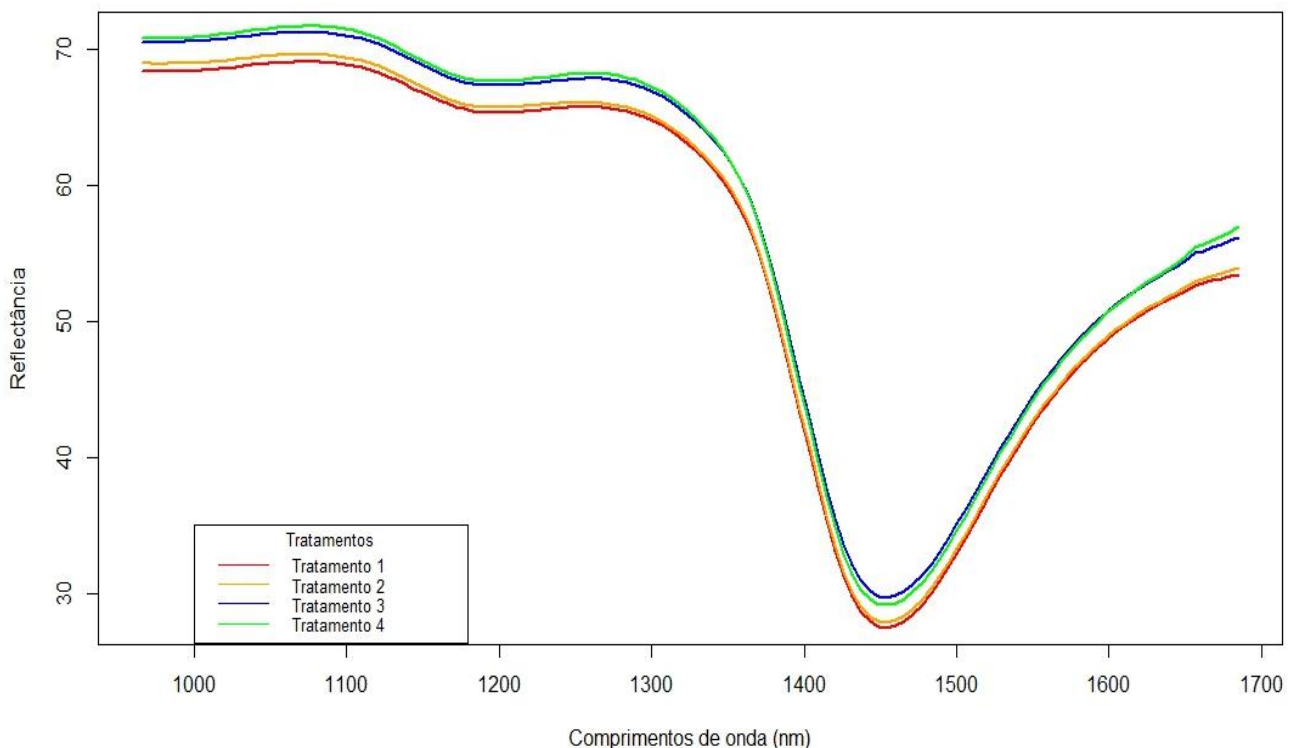


Figura 1: Curva de reflectância foliar média para cada tratamento. Os tratamentos são concentração de nitrogênio - 0 mg.L⁻¹ de N (Tratamento 1); 52,5 mg.L⁻¹ de N (Tratamento 2); 105 mg.L⁻¹ de N (Tratamento 3) e 210 mg.L⁻¹ de N (Tratamento 4).

Redução da Dimensionalidade

Diferentes conjuntos de comprimentos de onda foram selecionados para a classificação em 4 e em 2 classes. A técnica que selecionou maiores conjuntos de dados foi a *stepdisc* (Tabela 2). A técnica de componentes principais reduziu os dados para 2 autovetores compostos por uma combinação linear de todos os comprimentos de onda, acumulando 91,7% da variabilidade dos dados.

Tabela 2: Comprimentos de onda (em nm) advindos da caracterização espectral de folhas de plantas jovens de *Annona emarginata*, selecionados em função das diferentes seleções de atributos utilizadas e do número de classes.

Redução da dimensionalidade	Nº Classes	Comprimentos de onda selecionados (nm)
<i>Stepwise proc Stepdisc</i>	4	966,031; 971,544; 988,099; 1015,737; 1032,348; 1054,529; 1071,19; 1076,748; 1093,437; 1099,005; 1104,575; 1110,147; 1115,722; 1138,045; 1143,631; 1154,811; 1160,405; 1171,599; 1194,016; 1222,09; 1227,711; 1233,336; 1244,591; 1261,492; 1267,13; 1312,322; 1317,981; 1329,307; 1334,974; 1340,643; 1346,314; 1351,988; 1357,664; 1363,342; 1369,023; 1374,706; 1380,391; 1386,079; 1391,769; 1414,553; 1471,679; 1483,132; 1511,806; 1523,293; 1569,332; 1592,408; 1598,183; 1609,74; 1615,522; 1621,306; 1627,093; 1638,673; 1644,467; 1650,263; 1656,062; 1673,471; 1679,279; 1685,089
Regressão Logística Multinomial	4	1363,342; 1369,023; 1391,769; 1540,54; 1546,293; 1575,097; 1603,96; 1621,306; 1627,093; 1632,882; 1638,673; 1650,263; 1656,062; 1673,471; 1679,279; 1685,089
Análise de Componentes Principais	4	2 autovetores (Combinações lineares de todos os comprimentos de onda)
<i>Stepwise proc Stepdisc</i>	2	966,031; 971,544; 988,099; 1015,737; 1043,434; 1054,529; 1071,19; 1076,748; 1093,437; 1099,005; 1104,575; 1115,722; 1126,879; 1132,46; 1143,631; 1160,405; 1171,599; 1182,803; 1194,016; 1205,238; 1222,09; 1227,711; 1233,336; 1244,591; 1261,492; 1267,13; 1284,059; 1295,357; 1301,01; 1312,322; 1323,643; 1329,307; 1340,643; 1346,314; 1357,664; 1363,342; 1369,023; 1374,706; 1380,391; 1386,079; 1391,769; 1397,462; 1414,553; 1483,132; 1523,293; 1534,788; 1540,54; 1546,293; 1552,049; 1563,569; 1569,332; 1580,865; 1598,183; 1609,74; 1615,522; 1621,306; 1627,093; 1632,882; 1638,673; 1644,467; 1650,263; 1656,062; 1667,666; 1673,471; 1679,279; 1685,089
Regressão Logística Binomial	2	1374,706; 1650,263; 1661,862; 1679,279; 1685,089
Análise de Componentes Principais	2	2 autovetores (Combinações lineares de todos os comprimentos de onda)

Classificação

Observando as porcentagens de acerto médias e erro padrão para a classificação em duas e em quatro classes (tabela 3), percebemos que a redução por componentes principais, que é comumente utilizada para reduzir dados espectrais, apresentou as piores porcentagens de acerto, tanto quando associada a função linear discriminante quando para função quadrática discriminante.

A função linear discriminante foi mais efetiva para estes subconjuntos de variáveis selecionados, sugerindo igualdade de variâncias entre os grupos. A combinação de redução por *stepwise proc stepdisc* com a classificação por função linear discriminantes teve a maior porcentagem de acerto para a classificação em 4 classes e em 2 classes.

Tabela 3: Porcentagem média de classificação correta em 2 classes (com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4) e em 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e erro padrão (entre parênteses) dos dados espectrais de folhas de plantas jovens de *Annona emarginata* para os diferentes subconjuntos de comprimentos de onda selecionados pelas diferentes técnicas de redução de dimensionalidade e de seleção de covariáveis.

Redução da dimensionalidade	Classificação - 2 Classes		Classificação - 4 Classes	
	FLD	FQD	FLD	FQD
ACP	63,41 (0,16)	62,54 (0,14)	32,5 (0,15)	31,59 (0,15)
R Logística	98,62 (0,04)	98,80 (0,04)	91,57 (0,11)	88,61 (0,11)
Stepdisc	100,00 (0)	99,91 (0,01)	93,55 (0,09)	74,76 (0,16)

(*) ACP é análise de componentes principais; R Logística é a regressão logística; *Stepdisc* é o *stepwise* função *proc stepdisc*; FLD é função linear discriminante e FQD é função quadrática discriminante.

(**) Resultantes de 1000 reamostragens (*bootstrap*).

Índices Espectrais de Vegetação - IEV

A técnica *stepwise proc stepdisc* selecionou seis índices: I1, I5, I9, I13, I23 e I24 e a regressão logística selecionou quatro: I1, I13, I23 e I24. Na tabela 4 estão as porcentagens de classificação correta utilizando os conjuntos de índices selecionados como variáveis de entrada.

Tabela 4: Porcentagem média de classificação correta nas 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e erro padrão (entre parênteses) para os índices selecionados pelas diferentes técnicas de redução de dimensionalidade e de seleção de covariáveis.

Redução da dimensionalidade	Classificação	
	FLD	FQD
R Logística	80,76 (0,13)	80,74 (0,13)
Stepdisc	80,22 (0,14)	81,23 (0,13)

(*) Resultantes de 1000 reamostragens (*bootstrap*).

Quanto às classificações para os conjuntos de índices selecionados, a maior taxa de acerto foi a combinação de *stepwise proc stepdisc* com a função quadrática discriminante, de 81,23% de acerto, mas ela é apenas 1% maior do que a pior classificação, da combinação de *stepwise proc stepdisc* com a função linear discriminante, de 80,22% de acerto. Como esta diferença é pequena, é melhor se utilizar o conjunto de índices selecionados pela regressão logística, que é menor que o conjunto selecionado pela *stepdisc*.

Esta homogeneidade na porcentagem de acerto na classificação, nos leva a crer que a informação contida nos índices I5 e I9, que foram selecionados pela *stepdisc* além dos I1, I13, I23 e I24, são pouco relevantes para a classificação, não gerando incremento substancial nas porcentagens de acerto e sendo assim, dispensáveis.

O comprimento de onda de 1679,279 nm aparece em 3 dos 4 índices selecionados, sendo o que mais aparece nos índices escolhidos.

Análises Fisiológicas

Quanto às análises fisiológicas, o ajuste dos modelos foi bom, pois o valor dos desvios por graus de liberdade se manteve próximo da unidade, e a análise de resíduo de Pearson padronizado mostrou um resíduo aleatório em torno de zero e flutuando, na sua grande maioria, entre -2 e +2 (NELDER; WEDDERBURN, 1972).

Quanto as variáveis de trocas gasosas e fluorescência da clorofila *a*, para a transpiração (E), condutância estomática (Gs), assimilação de carbono (A), carbono interno (ci), eficiência do uso da água (EUA), eficiência instantânea da Rubisco (A/ci), energia quântica efetiva (øPSII), taxa de transporte de elétrons (ETR) e quentin fotoquímico (qP) não há diferença estatisticamente significativa detectadas pelos

modelos lineares generalizados ($\alpha=0,05$) entre os tratamentos em nenhum dos momentos analisados (Apêndice A e B).

Para a energia quântica potencial (F_v'/F_m'), dissipação de energia na forma de calor pelas antenas (D) e energia fotoquímica não dissipada e não utilizada na fase fotoquímica (Ex), não há diferença estatisticamente significativa ($\alpha=0,05$) entre os tratamentos nos 3 primeiros tempos analisados, mas, 45 dias após a implementação dos tratamentos (época da coleta das folhas - tempo 4) há diferenças significativas, detectadas pelo teste de Tukey-Kramer. A média do T1 é estatisticamente maior que a dos demais tratamentos para D e estatisticamente menor que a dos demais tratamentos para Ex e F_v'/F_m' (Tabela 5).

Tabela 5: Médias da energia quântica potencial (F_v'/F_m'), dissipação de energia na forma de calor pelas antenas (D) e energia fotoquímica não dissipada e não utilizada na fase fotoquímica (Ex) de mudas de *Annona emarginata* com os seus erros padrão (entre parênteses) 45 dias após a aplicação dos tratamentos (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)).

Tratamento	F_v'/F_m'	D	Ex
0 mg.L ⁻¹ de N (T1)	0,26122 B (0,04591)	0,73878 A (0,04591)	0,21891 B (0,03920)
52,5 mg.L ⁻¹ de N (T2)	0,53108 A (0,03129)	0,46892 B (0,03129)	0,47952 A (0,03728)
105 mg.L ⁻¹ de N (T3)	0,58807 A (0,03910)	0,41193 B (0,03910)	0,54468 A (0,03431)
210 mg.L ⁻¹ de N (T4)	0,50363 A (0,02731)	0,49637 B (0,02731)	0,47190 A (0,02916)

(*) Médias seguidas de mesma letra maiúscula não diferem estatisticamente pelo teste de Tukey-Kramer ($p<0,05$)

A análise foliar completa (Tabela 6) mostrou diferença nas concentrações de N nas folhas, sendo que plantas mantidas em solução sem N (T1) apresentaram concentração significativamente menor do que nos demais tratamentos segundo o teste de Tukey-Kramer.

O fato dos demais tratamentos (T2, T3 e T4) não separarem pode ser decorrência de dois fatores: o pouco tempo de exposição aos tratamentos (45 dias), que foi insuficiente para o início do aparecimento de sinais visíveis nas folhas; e do modo como a amostra é formada, já que utilizamos um *pool* de 15 plantas para cada amostra, e estas plantas possuem um alta variabilidade genética, podendo assim diluir os resultados. A amostragem foi feita deste modo em decorrência de necessitar de 2g de massa seca de folhas, que não é possível de conseguir em apenas uma muda da idade trabalhada.

Tabela 6: Quantidades médias de macronutrientes (nitrogênio - N, fósforo - P, potássio - K, cálcio - Ca, magnésio - Mg e enxofre - S) e desvio padrão (entre parênteses) em g.kg^{-1} e micronutrientes (Boro - B, cobre - Cu, manganês - Mn, ferro - Fe e zinco - Zn) em mg.kg^{-1} para os 4 tratamentos - Trat (Concentração de nitrogênio - 0 mg.L^{-1} de N (T1); 52,5 mg.L^{-1} de N (T2); 105 mg.L^{-1} de N (T3) e 210 mg.L^{-1} de N (T4)).

Trat	Macronutrientes						Micronutrientes				
	N	P	K	Ca	Mg	S	B	Cu	Fe	Mn	Zn
T1	20,17 B (1,67)	2,00 (0,04)	13,02 (0,3)	8,87 (0,51)B	3,72 (0,05)	1,84 (0,07)	73,81 (18,34)	2,75 (0,96)	179,5 (34,46)	646,25 (55,36)	29,50 (5,07)
T2	41,33 A (0,82)	2,79 (0,15)	18,52 (1,02)	7,95 (0,3)B	2,27 (0,09)	2,18 (0,23)	62,69 (3,19)	1,25 (0,5)	119,75 (19,19)	431,75 (20,37)	24,25 (3,77)
T3	39,76 A (2,51)	2,88 (0,31)	11,75 (1,45)	13,27 (1,27)A	3,17 (0,35)	2,28 (0,12)	105,34 (10,26)	3,00 (1,41)	266,5 (67,43)	1052,50 (149,04)	26,75 (1,71)
T4	43,47 A (0,61)	3,63 (0,27)	9,92 (0,59)	11,82 (0,54)A	3,35 (0,13)	2,31 (0,36)	102,17 (11,32)	2,00 (0,82)	311,75 (57,39)	944,00 (108,43)	23,25 (1,26)

(*) Médias seguidas de mesma letra não diferem estatisticamente pelo teste de Nemenyi ($p < 0,05$)

Para os pigmentos foliares não houve diferença estatística para nenhuma das variáveis testadas segundo os modelos lineares generalizados. Na figura 2 estão seus boxplots.

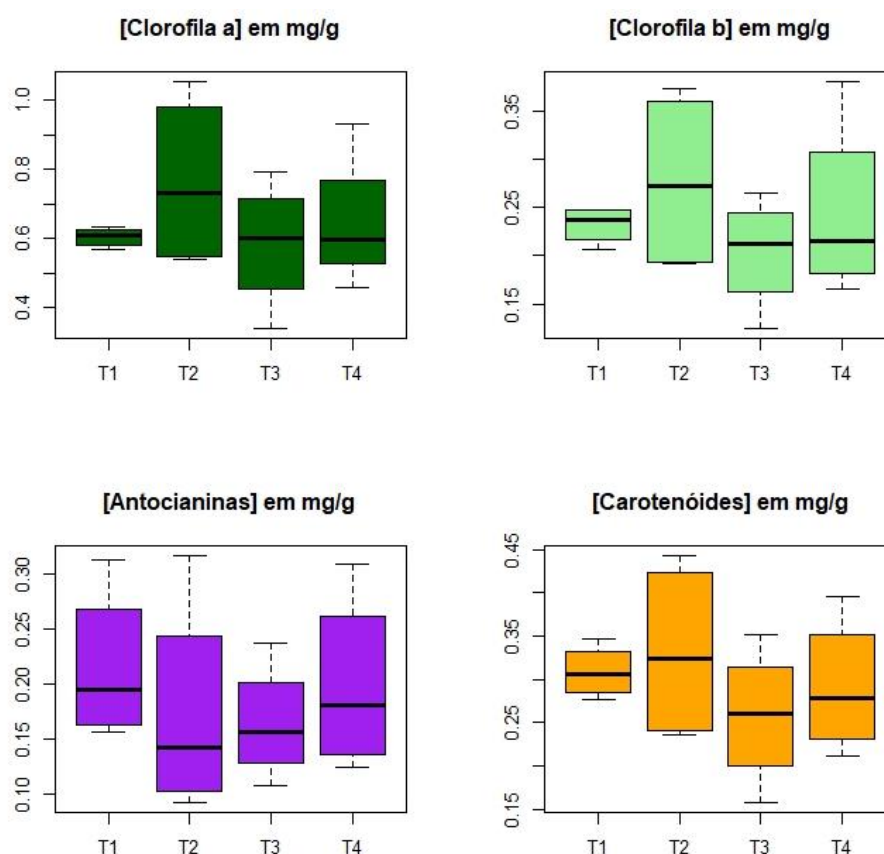


Figura 2: Boxplots das concentrações de pigmentos foliares para os quatro tratamentos (Concentração de nitrogênio - 0 mg.L^{-1} de N (T1); 52,5 mg.L^{-1} de N (T2); 105 mg.L^{-1} de N (T3) e 210 mg.L^{-1} de N (T4)).

4.5. Discussão

Neste trabalho a técnica NIRS se mostrou útil para a identificação da falta de N foliar antes do aparecimento de sinais visuais. Assim, medidas profiláticas podem ser tomadas antes que ocorra dano, evitando consequências como a redução do crescimento das mudas.

Annona emarginata (Schltdl.) H. Rainer é uma espécie produzida em viveiros com o intuito de reflorestamento e de se tornarem porta-enxerto para culturas de Annonaceae frutíferas. Desta forma, são mantidas em viveiros principalmente nas fases iniciais do desenvolvimento, até estarem aptas para a enxertia ou plantio em campo, o que ocorre após aproximadamente um ano desde a semeadura. Este trabalho estabelece um método de calibração que nos possibilita utilizar uma nova ferramenta no monitoramento destas mudas nas suas fases iniciais de desenvolvimento, sendo também uma base para construir no futuro um processo de automação no monitoramento de nitrogênio destas plantas em viveiro, ou plantas jovens em fase inicial de desenvolvimento no campo, em áreas de recuperação ou de reflorestamento, a fim de garantir o melhor estabelecimento das plantas no campo pela adequada manutenção dos níveis de nitrogênio.

Quanto a calibração, em geral, espera-se que quanto maior a porcentagem de redução que uma técnica atinge, mais eficiente se torna. Wang *et al.* (2019) mostraram que colocar como variáveis de entrada apenas faixas de espectro que correspondem as informações desejadas, selecionadas previamente, aumenta o poder preditivo das técnicas de classificação.

A ACP reduziu muito bem o conjunto de variáveis, saindo de 128 comprimentos de onda para apenas duas variáveis resposta, que são combinações lineares de todos os comprimentos de onda, mas, quando estas foram colocadas como variáveis de entrada para os classificadores, geraram porcentagens de classificação muito baixas em comparação com as demais técnicas (tabela 2). Este resultado foi oposto ao obtido por Yi *et al.* (2007), cuja redução da dimensionalidade dos dados hiperespectrais por ACP gerou aumento na classificação correta das redes neurais artificiais utilizadas para monitorar as concentrações foliares de nitrogênio em arroz.

A melhor combinação de técnica de redução de dimensionalidade do conjunto de dados com técnica de classificação, foi a redução por *stepwise* e a classificação por função linear discriminante, tanto para classificação em 4 quanto para classificação em 2 classes. A *stepwise proc stepdisc* foi técnica que menos reduziu o conjunto de dados

(tabela 2), mas as maiores porcentagens de acerto foram com seus subconjuntos, de forma que provavelmente, conjuntos grandes são melhores para este caso.

Geralmente, em viveiros, a falta de nitrogênio só é percebida quando começam a aparecer sinais visuais, como a clorose de folhas maduras, mas, neste estágio, já ocorreu um dano a planta. Antes do aparecimento dos sinais visuais, ocorrem alterações na planta, principalmente relacionadas ao seu metabolismo, como redução da velocidade metabólica, redirecionamento de esqueletos carbônicos e alteração na produção de metabólitos especializados (BARKER; PILBEAM, 2015). Estas modificações bioquímicas são captadas pela técnica NIRS, pois a caracterização espectral das folhas é muito sensível.

Assim, podemos supor que as baixas concentrações de N fornecidas a estas mudas, mesmo que por apenas 45 dias, já são suficientes para gerar pequenas alterações bioquímicas, nas folhas, o que culminou na separação em dois conjuntos de curvas (Figura 1). Campos *et al.* (2019) cultivaram *Annona emarginata* com diferentes níveis de nitrogênio e mostraram que as baixas concentrações de N contribuíram para que houvesse um redirecionamento de agentes redutores e esqueletos de carbono para a aumentar as substâncias voláteis foliares e carboidratos nas folhas, em especial o amido. Segundo Curran (1989), os comprimentos de onda na faixa de 990 nm, 1200 nm, 1450 nm, 1530 nm, 1540 nm, 1580 nm e 1690 nm estão correlacionados com a concentração de amido, assim, estas podem ser algumas das alterações químicas que influenciaram na separação de curvas na caracterização espectral.

A técnica NIRS aplicada na identificação da concentração de N foliar necessita de menos tempo de exposição aos tratamentos para indicar uma mudança metabólica em decorrência dos diferentes níveis de nitrogênio fornecidos as mudas do que as análises fisiológicas.

As análises de trocas gasosas e fluorescência da clorofila *a* são comumente usadas como indicadores de dano em decorrência da falta de um nutriente mineral (BARON *et al.*, 2013, KALAJI *et al.*, 2018), já que indicam dano a fotossíntese da planta. Neste trabalho, apenas 45 dias de exposição aos tratamentos não foi o suficiente para alterar as trocas gasosas, ou seja, as alterações bioquímicas que ocorreram ainda não são suficientes para afetar as trocas gasosas da planta. Já para a fluorescência da clorofila *a*, o tratamento sem N (T1) se diferenciou dos demais para as variáveis dissipação de energia na forma de calor pelas antenas (D), energia quântica potencial

(Fv'/Fm') e energia fotoquímica não dissipada e não utilizada na fase fotoquímica (Ex) (Tabela 5).

Como *Annona emarginata* é uma arbórea rústica, e seu metabolismo é lento, provavelmente apenas 45 dias de exposição a metade da concentração de N recomendada (T2) não foi o suficiente para causar alterações que afetem a fotossíntese das mudas, mas, após uma longa exposição a esta concentração de N, provavelmente haveriam alterações suficientes para serem detectadas. Por outro lado a NIRS conseguiu separar os 4 tratamentos com 93% de acerto (Tabela 2), possibilitando a recuperação destas mudas de T2 antes de qualquer alteração na sua fotossíntese.

A produção de energia fotoquímica das plantas sem N começou a ser afetada após 45 dias, então podemos teorizar que, como o N é um mineral é constituinte de todas as enzimas da planta, dentre elas a *rubisco*, a principal enzima da fase bioquímica da fotossíntese (LAWLOR, 1989), as antenas estavam dissipando mais energia solar na forma de calor, pois, provavelmente não havia *rubisco* suficiente para utilizar essa energia para assimilar carbono. Tendo como base a revisão de Murchie e Lawson (2013), acreditamos que plantas mantidas sem o fornecimento de N (T1) já tinham menor quantidade energia quântica potencial e, como dissipou grande quantidade dessa energia em forma de calor, tinha menor quantidade disponível para a fase bioquímica da fotossíntese, o que resultou em uma menor quantidade de energia fotoquímica não dissipada e não utilizada na fase fotoquímica .

A análise foliar completa demonstrou a diferença entre o não fornecimento de N e os demais tratamentos, mas foi pouco eficiente para detectar as diferentes concentrações de N fornecidas, separando apenas o tratamento sem N (T1) dos demais (Tabela 6). Identificar apenas a total ausência do mineral é pouco eficiente em condições de campo, já que raramente a muda estará totalmente sem N, apenas com uma concentração menor que a recomendada.

Um dos possíveis motivos desta separação apenas da ausência de N ocorrer é que o N é um elemento móvel, que será translocado das folhas mais velhas para as mais jovens (TAIZ; ZEIGER, 2017), mitigando o efeito de sua falta nas folhas da porção mediana da muda. Outro motivo possível é que com esta técnica, não se obtém uma resposta por planta, e sim por um grupo de plantas, já que necessita de uma grande quantidade de matéria seca de folha, impossível de se obter com apenas uma muda da idade trabalhada. Desta forma, como possuem uma alta variabilidade genética, pois são mudas seminais, há a possibilidade de que, ao se fazer esta amostragem conjunta, se

dilua os efeitos dos tratamentos. Além do que, o tempo pode não ter sido suficiente para o acúmulo de quantidades de N que pudessem ser detectadas pelo método tradicional de quantificação. Já a NIRS possibilitou uma resposta por cada muda individualmente, e não analisou apenas a concentração de N na folha, mas todos os efeitos bioquímicos desta concentração.

A análise foliar completa mostrou também que ocorreu uma redução do cálcio nos tratamentos com falta de nitrogênio. O cálcio, além de ser cofator para algumas enzimas envolvidas na hidrólise de ATP e de fosfolipídeos, também atua como mensageiro secundário na regulação metabólica (TAIZ; ZEIGER, 2017). Para funcionar como um sinalizador na planta, ele necessita se ligar a calmodulina, proteína encontrada no citosol de células vegetais e assim formar o complexo calmodulina- Ca^{2+} (TAIZ; ZEIGER, 2017), mas, como esta proteína precisa de N na sua constituição, a falta deste mineral pode causar a redução desta e, por consequência, isso pode refletir na concentração de cálcio nas folhas, o que parece estar evidenciado neste trabalho.

Para os pigmentos foliares, não houve diferença estatística entre os tratamentos (Figura 2), o que indica que apenas 45 dias de exposição aos tratamentos não foram suficientes para detecção de variações significativas e que resultassem no aparecimento de sinais visíveis. Assim, esta análise fisiológica não se mostra recomendada para estas mudas que estão expostas a tão pouco tempo as diferentes concentrações de N, pois não detecta variações de modo precoce, embora o dano já tenha sido iniciado.

Os comprimentos de onda utilizados para se fazer os índices foram 1685,089 nm, 1679,279 nm, 1673,471 nm e 1638,673 nm. Segundo Curran (1989) e Jensen (2002) a área próxima a 1690,00 nm está associada a nitrogênio e nitrato, desta forma, podemos dizer que ao menos um destes comprimentos de onda selecionados possui uma forte ligação com o problema apresentado. As folhas são as principais redutoras de nitrato em amônio, e, esta conversão é um dos drenos de energia fotoquímica não dissipada e não utilizada na fase fotoquímica da fotossíntese, impedindo a criação de formas reativas de oxigênio que podem causar lesões nas membranas (BLOOM *et al.*, 2010).

Geralmente, os trabalhos que visam utilizar índices espectrais de vegetação (IEV) para aferir N, utilizam índices que são formados por comprimentos de onda situados na banda do visível correlacionados a sinais visuais causados por uma longa exposição a falta ou excesso de N, como a degradação da clorofila e a alteração na biomassa. Zhou, Liu e Huang (2010), utilizaram o mSR, que é indicador de biomassa

(SIMS; GAMON, 2002), e Yang *et al.* (2018) utilizaram o NDVI, que é indicador de biomassa e área foliar (ROUSE *et al.*, 1974) para quantificar N em arroz. Nestes dois casos, são monoculturas de grande escala e metabolismo rápido, nas quais as plantas já apresentavam sintomas visuais das diferentes concentrações de N as quais as plantas estavam expostas.

Neste trabalho tentamos produzir índices espectrais de vegetação formados por comprimentos de onda situados no espectro NIR e que podem ser associados ao conteúdo de nitrogênio das folhas, e não aos sinais que este conteúdo induz. Segundo Xue e Su (2018), a maior parte dos IEV's são formados por comprimentos de onda de 400 à 900 nm. Eles previram que com o avanço do sensoriamento remoto hiperespectral e multiespectral, novos índices serão desenvolvidos, ampliando as áreas de pesquisas. Esta pode se tornar uma das principais áreas de pesquisa em sensoriamento remoto no futuro próximo.

Miphokasap *et al.* (2012) testaram diferentes IEVs com comprimentos de onda desde o visível até o final do NIRS (2500 nm) para inferir N em cana-de-açúcar, e obtiveram os melhores resultados com índices formados pelos comprimentos de onda 700 nm, 724 nm e 750 nm, que estão na faixa do visível. Como a cana-de-açúcar é uma planta C4 de crescimento rápido, os sinais visuais aparecem rapidamente, diferentemente de arboreas C3, assim, estes índices são pouco úteis para problemas como o apresentado neste artigo. Como a reflectância foliar é muito sensível, e a a concentração de nitrogênio na folha vai afetar todo o seu metabolismo, é interessante que para se aferir este mineral antes do aparecimento de sinais visuais, se use índices formados especificamente para o tipo de planta trabalhado.

É possível associar os índices criados neste trabalho com outros que abrangem esta faixa do visível, mas, provavelmente não vai ter um incremento significativo na porcentagem de acerto, já que não há diferença entre os tratamentos para pigmentos foliares (tabela 4), que são os compostos da folha de maior influência no espectro visível segundo Carter e Knapp (2001).

A regressão logística selecionou uma menor quantidade de índices que o *stepwise proc stepdisc*, mas, todos estes já haviam sido previamente selecionados pela *stepdisc*, provando que existem índices mais efetivos que outros. Nos índices selecionados pela regressão logística todos os comprimentos de onda aparecem, mas a combinação dos 1685,089 nm e 1679,279 nm aparecem em 50% do conjunto de

variáveis de entrada, nos levando a crer que a combinação destes dois comprimentos de onda tem uma grande importância para separação das 4 concentrações de N fornecidas.

Estes comprimentos de onda são os dois últimos do espectro mensurado, estando na área onde ocorre um reafastamento das quatro curvas (figura 1), também são os que estão mais perto de uma das regiões do espectro que se relaciona com o nitrogênio, sendo então, de alta importância para a classificação.

4.6. Conclusões

Concluimos que a espectroscopia no infravermelho próximo (NIRS) é efetiva para detectar variações de nitrogênio em folhas de plantas jovens de *Annona emarginata* (Schltdl.) H. Rainer e que há uma combinação de índices (I1, I13, I23 e I24) que permite maior acuracidade para prever as variações de nitrogênio em *Annona emarginata* (Schltdl.) H. Rainer.

4.7. Referências

- BARKER, A. V.; PILBEAM, D. J. **Handbook of plant nutrition**. 2 ed. Nova Iorque: CRC Press, 2015.
- BARON, D.; FERREIRA, G. ; RODRIGUES, J. D. ; BOARO, C.S.F.; MACEDO, A. C.. Gás exchange, physiological indexes and ionic accumulation in *Annona emarginata* (Schltdl.) H. Rainer seedlings in nutrients solution. **Revista Brasileira de Fruticultura**, Jaboticabal-SP, v. 36, p. 361-376, 2013.
- BLOOM, A. J. et al. Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. **Science**, v.328, p.899–903, 2010.
- CAMPOS, F.G.; VIEIRA, M.A.R.; AMARO, A.C.E.; DELA CRUZ-CHACÓN, I.; MARQUES, M.O.M.; FERREIRA, G., BOARO, C.S.F.. Nitrogen in the defense system of *Annona emarginata* (Schltdl.) H. Rainer. PLoS ONE, 14(6): e0217930, 2019. <https://doi.org/10.1371/journal.pone.0217930>
- CARTER, G. A.; KNAPP, A. K. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. **Am. J. Bot.** v.88, p.677-684, 2001. <https://doi.org/10.2307/2657068>
- CHATROU, L.W. et al. The natural history of Annonaceae. **Botanical Journal of the Linnean Society**, v.169, p.1–4, 2012.
- CONSELHO ESTADUAL DO MEIO AMBIENTE, RIO GRANDE DO SUL. Decreto estadual CONSEMA n. 42.099 de 31 de dezembro de 2002. Declara as espécies da

- flora nativa ameaçadas de extinção no estado do Rio Grande do Sul e da outras providências, Palácio Piratini, Porto Alegre, Rio Grande do Sul, 31 dez. 2002, 2002.
- CURRAN, P.J.. Remote sensing of foliar chemistry. **Remote Sensing of Environment**, v.30, p.271–278, 1989.
- DEMMIG-ADAMS, B.; ADAMS III, W.W.; BARKER, D. H.; LOGAN, B.A.; BOWLING, D. R.; VERHOEVEN, A.S. Using chlorophyll fluorescence to assess the fraction of absorbed light allocated to thermal dissipation of excess excitation. **Physiologia Plantarum**, v.98 ,p.253-264 , 1996.
- EFRON, B. **Bootstrap methods: Another look at the Jackknife**. Ann. Statistics, vol 7(1) p.1-26, 1979.
- FONSECA, E.L. da; ROSA, L.M.G.; FONTANA, D.C. Caracterização espectral de *Paspalum notatum* em diferentes níveis de adubação nitrogenada. **Pesquisa agropecuária brasileira**, Brasília, v.37, n.3, p. 365-371, mar. 2002.
- JENSEN, J.R.. **Remote sensing of the environment – An earth resource perspective**. [reprint.] edition. The MIT Press and MIT Press, Upper Saddle River, NJ, USA, 2002.
- JOHNSON, R. & WICHERN, D. **Applied Multivariate Statistical Analysis**. 6 Ed . Prentice Hall, New Jersey, 2014.
- KALAJI, H.M.; BABA, W., GEDIGA, K. et al. Chlorophyll fluorescence as a tool for nutrient status identification in rapeseed plants. **Photosynth Res**. v.136, p.329–343, 2018. doi:10.1007/s11120-017-0467-7
- LIMA, R.L.S.; FERREIRA, G.B.; WEBER, O.B.; CAZETTA, J.O.. Diagnose foliar da gravioleira (*Annona muricata* L.): efeito da posição de ramos e folhas. **Ciência e Agrotecnologia**. Editora da Universidade Federal de Lavras (UFLA), v. 31, n. 5, p. 1320-1325, 2007. Disponível em: <<http://hdl.handle.net/11449/30165>>.
- MIPHOKASAP, P.; HONDA, K.; VAIPHASA, C.; SOURIS, M.; NAGAI, M. Estimating Canopy Nitrogen Concentration in Sugarcane Using Field Imaging Spectroscopy. **Remote Sens**. v.4, p.1651-1670, 2012. doi:10.3390/rs4061651
- MURCHIE, E.H.; LAWSON, T. Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. Journal of Experimental Botany, V. 64:13, pp. 3983–3998, 2013. doi:10.1093/jxb/ert208
- NELDER, J.A.; WEDDERBURN, R.W. Generalized linear models. **Journal of the Royal Statistical Society Series A**, v.135, n.3, p.370–384, 1972.

- OTTO, M. Chemometrics: statistics and computer application in analitical chemistry. Chinchester: **wiley-VCH**. 1999.
- RENNENBERG, H.; DANNENMANN, M.; GESSLER, A.; KREUZWIESER, J.; SIMON, J.; PAPEN, H. Nitrogen balance in forest soils: nutritional limitation of plants under climate change stresses. **Plant Biology**, v.11, p.4-23, 2009. doi:10.1111/j.1438-8677.2009.00241.x
- ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In: **Proc. 3th Earth Resources Technology Satellite-1 Symposium**. Greenbelt, MD NASA, pp. 301– 317, 1974.
- RYAN, M.G. Effects of Climate Change on Plant Respiration. **Ecological Applications**, v.1, p.157-167, 1991. doi:10.2307/1941808
- SIMS, D. A.; GAMON, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. **Remote Sensing of Environment**, v. 81, n. 2, p. 337-354, 2002.
- TAIZ, L.; ZEIGER, E. Fisiologia vegetal. 6. ed., Artmed, 2017. 918 p.
- TOKUNAGA, T. **A cultura da atemoia**. 2.ed. Campinas: CATI, 2005. 80 p. (Boletim técnico)
- WALSH, K.J.E., et al. Tropical cyclones and climate change. *Wires Climate Change*. vol 7:1, pag. 65-89, 2015. <https://doi.org/10.1002/wcc.371>
- WANG, Y.; HU,X.; JIN,G.; HOU,Z.; NING, J.; ZHANG, Z. Rapid prediction of chlorophylls and carotenoids content in tea leaves under different levels of nitrogen application based on hyperspectral imaging. **J Sci Food Agric**, v.99, p.1997–2004, 2019.
- WESTFALL, P. H.; TOBIAS, R. D.; ROM, D.; WOLFINGER, R. D.; HOCHBERG, Y. **Multiple Comparisons and Multiple Tests Using the SAS System**, Cary, NC: SAS Institute Inc. 1999.
- XUE, J.; SU, B.. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, p.1–17, 2017. doi:10.1155/2017/1353691
- YANG, J.; DU, L.; GONG, W.; SHI, S.; SUN, J.; CHEN, B. Potential of vegetation indices combined with laser-induced fluorescence parameters for monitoring leaf nitrogen content in paddy rice. **PLoS ONE**, v.13(1): e0191068, 2018. <https://doi.org/10.1371>

- YI, Q. X.; HUANG, J. F.; WANG, F. M.; WANG, X. Z.; LIU, Z. Y. . Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network. **Environmental Science and Technology**, v.41, p.6770–6775, 2007.
- ZHANG, S. et al. Temperature-Dependent Gas Exchange and Stomatal/Non-Stomatal Limitation to CO² Assimilation of *Quercus Liaotungensis* under Midday High Irradiance. **Photosynthetica**, v.39, n.3, p.383-388, 2001.
- ZHOU, Qi-fa; LIU, Zhan-yu; HUANG, Jing-feng. Detection of nitrogen-overfertilized rice plants with leaf positional difference in hyperspectral vegetation index. **Zhejiang Univ-Sci B (Biomed & Biotechnol)**, v.11(6), p.465-470, 2010.

5. Capítulo 2 - Aplicação de algoritmos de aprendizado de máquinas para identificação da falta de nitrogênio em mudas de *Annona emarginata* (Schltdl.) H. Rainer antes do aparecimento de sinais visuais utilizando NIRS e fotografias digitais.

5.1. Resumo

A agricultura de precisão foca no aumento da produção e abrange diferentes técnicas de coleta e processamento de dados, como NIRS, fotografias digitais e aprendizado de máquinas. Neste estudo aplicamos estas técnicas para identificar a falta e o excesso de N antes do aparecimento de sinais visuais em mudas de *Annona emarginata* (Schltdl.) H. Rainer. Deste modo, este trabalho visa avaliar se algoritmos de aprendizado de máquinas são eficientes para classificar mudas em função dos níveis de N com base na caracterização espectral e em fotografias digitais e avaliar se a NIRS é mais efetiva que fotografias digitais para identificar variações no nitrogênio foliar em mudas de *Annona emarginata*. O experimento foi inteiramente casualizado, em sistema hidropônico com 4 tratamentos, que consistiram em variações da concentração de N na solução nutritiva: T1, sem N; T2, com 50% da concentração de N recomendada para a espécie; T3, com 100% de N recomendado para a espécie e T4, com o dobro da concentração de N recomendada para a espécie, por 45 dias. Foi feita a caracterização espectral e fotografias digitais de 3 folhas de cada muda. Com diferentes algoritmos de aprendizado de máquinas, foram feitas a seleção de atributos e classificação. As fotografias digitais são menos efetivas que a caracterização espectral para a detecção das variações de N. A regressão logística apresentou as maiores porcentagens de acerto na classificação. Assim, pode-se concluir que há aplicabilidade de algoritmos de aprendizado de máquinas para fazer a calibração da técnica NIRS e que a NIRS é mais efetiva que as fotografias digitais para identificar variações no nitrogênio foliar.

Palavras chave: redes neurais artificiais, aprendizado de máquinas, *perceptron* multicamadas, *bootstrap*.

(*) Artigo escrito seguindo as normas ABNT, pois ainda não foi definida a revista para a qual será submetido.

5. Chapter 2 - Application of machine learning algorithms to identify nitrogen deficiency in *Annona emarginata* (Schltdl.) H. Rainer seedlings prior to the appearance of visual signals using NIRS and digital photographs.

Abstract

Precision agriculture focuses on increasing production and encompasses different data collection and processing techniques such as NIRS, digital photography and machine learning. In this study we apply these techniques to identify the lack and excess of N before the appearance of visual signals in *Annona emarginata* (Schltdl.) H. Rainer seedlings. Thus, this work aims to evaluate if machine learning algorithms are efficient to classify seedlings as a function of N levels based on spectral characterization and digital photographs and to evaluate if NIRS is more effective than digital photographs to identify variations in leaf nitrogen in seedlings of *Annona emarginata*. The experiment was completely randomized in a hydroponic system with 4 treatments, which consisted of variations in the concentration of N in the nutrient solution: T1, without N; T2, with 50% of the recommended N concentration for the species; T3, with 100% N recommended for the species and T4, twice the recommended N concentration for the species, for 45 days. For each seedling, 3 leaves were used for spectral characterization and digital photographs. Different machine learning algorithms, attribute selection and classification were used to process the data. Digital photographs are less effective than spectral characterization for the detection of N variations. Logistic regression presented the highest percentages of accuracy in the classification. Thus, it can be concluded that there is applicability of machine learning algorithms to calibrate the NIRS technique and that NIRS is more effective than digital photographs to identify variations in leaf nitrogen.

Key words: artificial neural networks, machine learning, multilayer perceptron, bootstrap.

5.2. Introdução

A agricultura de precisão surgiu com a integração de avanços tecnológicos, como os de manejo, coleta e processamento de dados e sua aplicação na área agrária (WHELAN *et al.*, 1997). Ela tem apresentado importantes avanços nas últimas décadas (BERNARDI *et al.*, 2014) e está intrinsecamente associada a variabilidade (MULLA; SCHEPERS, 1997), tanto espacial quanto temporal (ZHANG; WANG; WANG, 2002). Assim, leva em consideração qualquer aspecto que varie de alguma forma, como os recursos do solo, clima, genética de plantas, diversidade de culturas e máquinas e operadores agrícolas (PIERCE; NOWAK, 1999).

Um dos seus principais objetivos é o aumento da produção sem precisar de mais terras e sem desperdiçar recursos, já que a agricultura de precisão deve ser lucrativa (LOWENBERG-DEBOER; SWINTON, 1997) e concomitantemente atender às necessidades e capacidades do agricultor (NOWAK, 1997).

Diversos recursos de obtenção de informação são aplicados nesta esta área de estudo, como o uso de imagens orbitais, fotografias digitais, imageamento hiperespectral e espectroscopia no infravermelho próximo (*Near-Infrared Spectroscopy* - NIRS). Estes permitem associação com algoritmos de aprendizado de máquinas para o processamento dos dados e obtenção de informações.

Essas técnicas podem ser aplicadas em diversas áreas da agricultura de precisão com diferentes objetivos, como na área de nutrição mineral em viveiros. O atual cenário de produção de mudas está cada vez mais acirrado, necessitando o surgimento de novas técnicas que possibilitam aumentar a qualidade das mudas. Dentre os principais fatores que influenciam nesta qualidade está o monitoramento nutricional. A associação do imageamento por NIRS e/ou fotografias digitais com técnicas de aprendizado de máquinas, possibilita a rápida identificação de desordens metabólicas decorrentes de problemas com a adubação, e conseqüentemente, possibilita a correção do problema, recuperando a muda antes de danos irreversíveis.

Muitas arbóreas frutíferas são mantidas em viveiros até a época de plantio. Para a produção de mudas comerciais de espécies do gênero *Annona* da família *Annonaceae*, as copas comerciais são enxertadas em espécies nativas (SOBRINHO, 2010). Dentre as espécies usadas como porta enxerto, a *Annona emarginata* (Schltdl.) H. Rainer se destaca para a produção do híbrido atemoia (*Annona x atemoya* Mabb.) por fornecer resistência para diversas podridões de colo e raiz e coleobrocas (TOKUNAGA, 2005) e

garantir copa vigorosa (SCALOPPI JUNIOR, 2007). As mudas de *A. emarginata* são produzidas a partir de sementes e mantidas em viveiros por aproximadamente 1 ano, até a época da enxertia ou de ser levada ao campo para o reflorestamento de áreas degradadas.

Assim, este estudo teve como objetivo avaliar se algoritmos de aprendizado de máquinas são eficientes para classificar mudas em função dos níveis de N com base na caracterização espectral e em fotografias digitais e avaliar se a NIRS é mais efetiva que fotografias digitais para identificar variações no nitrogênio foliar em mudas de *Annona emarginata*.

5.3. Material e Métodos

Experimento

Foi montado um experimento com delineamento inteiramente casualizado em casa de vegetação. 240 mudas de *Annona emarginata* (Schltdl.) H. Rainer foram colocadas em sistema de hidroponia com solução nutritiva Hoagland & Arnon nº 2 (1950), com força iônica de 50% (BARON, *et al.*; 2013). Foram feitas alterações nas concentrações de N nas soluções em decorrência dos tratamentos: 0 mg.L⁻¹ de nitrogênio (T1); 52,5 mg.L⁻¹ de N (T2), que equivale a 50% da concentração recomendada para espécie (BARON, *et al.*; 2013); 105 mg.L⁻¹ de N (T3), a dose recomendada para a espécie (BARON, *et al.*; 2013) e 210 mg.L⁻¹ de N (T4), que é o dobro da dose recomendada para a espécie (BARON, *et al.*; 2013) (Tabela 1). Foram separadas 60 mudas por tratamento. Mais detalhes do experimento estão descritos em Gomes et al. (2020).

Após 45 dias de exposição aos tratamentos, foram coletadas 3 folhas de cada muda, para fazer a caracterização espectral no infravermelho próximo e as fotografias digitais. A caracterização espectral das folhas foi feita com um espectrorradiômetro da Ocean Optics, modelo Flame - NIR, com faixa de leitura de 966,03 a 1685,09 nm, e sensibilidade de 5,51 nm. As fotografias digitais das folhas foram obtidas em ambiente padronizado (mesma condição de luminosidade do ambiente) utilizando uma câmera digital Canon EOS Rebel T5i Full HD 18 MP, lente 18-55mm, f/3,5-5,6 sem tripé com a seguinte configuração: distância focal 35mm, ISO 100, sem flash, temporizador de 2s, resolução máxima, gravação de arquivos nos formatos jpeg.

Foram tomadas as médias das 3 folhas, formando assim uma resposta por planta, tanto para as fotografias digitais, quanto para a caracterização espectral.

Além da caracterização espectral, também foram coletadas aleatoriamente folhas para fazer a análise de pigmentos foliares, como descrito em Gomes et al. (2020).

Aprendizado de Máquinas

Nesta fase foi utilizado o *software* livre WEKA (*Waikato Environment for Knowledge Analysis*) desenvolvido pela Universidade de Waikato. Primeiro foram separados subconjuntos de atributos com técnicas de redução de dimensionalidade para depois classificar em quatro classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e em duas classes (Com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4).

Foram feitas reduções na dimensionalidade dos conjuntos de dados utilizando as técnicas *Correlation based Feature Selection* (CFS), *Wrapper* e *Classifier Subset Evaluator* (CSE). Para o *Wrapper*, foram utilizados os algoritmos redes *Perceptron* Multicamadas (PMC) e a Regressão Logística, associadas aos algoritmos *Best First* e *Stepwise*. Para o CSE, foi utilizado o algoritmo redes *Perceptron* Multicamadas (PMC), associado aos algoritmos *Best First* e *Stepwise*, com e sem divisão do conjunto de treino.

Posteriormente foi feita a classificação utilizando os algoritmos ZeroR, Regressão Logística, *Perceptron* Multicamadas (PMC) e *Multi Class Classifier* (Classificador multi-classe - MCC) (Figura 1).

O conjunto de amostras aleatórias utilizadas para treinar as redes (*trainnig set*) foi composto de 66% dos dados amostrados, já o conjunto para testar as redes (*cross validation set*) foi os 33% restantes. Após programadas, foram rodadas 1000 vezes cada uma das redes, com diferentes conjuntos de treino e validação (*bootstrap*), obtendo assim a porcentagem de acerto médio e seu desvio padrão além das matrizes de confusão, essa medida de variabilidade, nos serve para saber como flutua a porcentagem de acerto da técnica com diversos conjuntos de treino e teste.

Após rodadas, as porcentagens de acerto médias das PMC, regressão logística e MCC foram compradas com um teste t pareado com as porcentagens de acerto médias do ZeroR ($\alpha=0,05$).

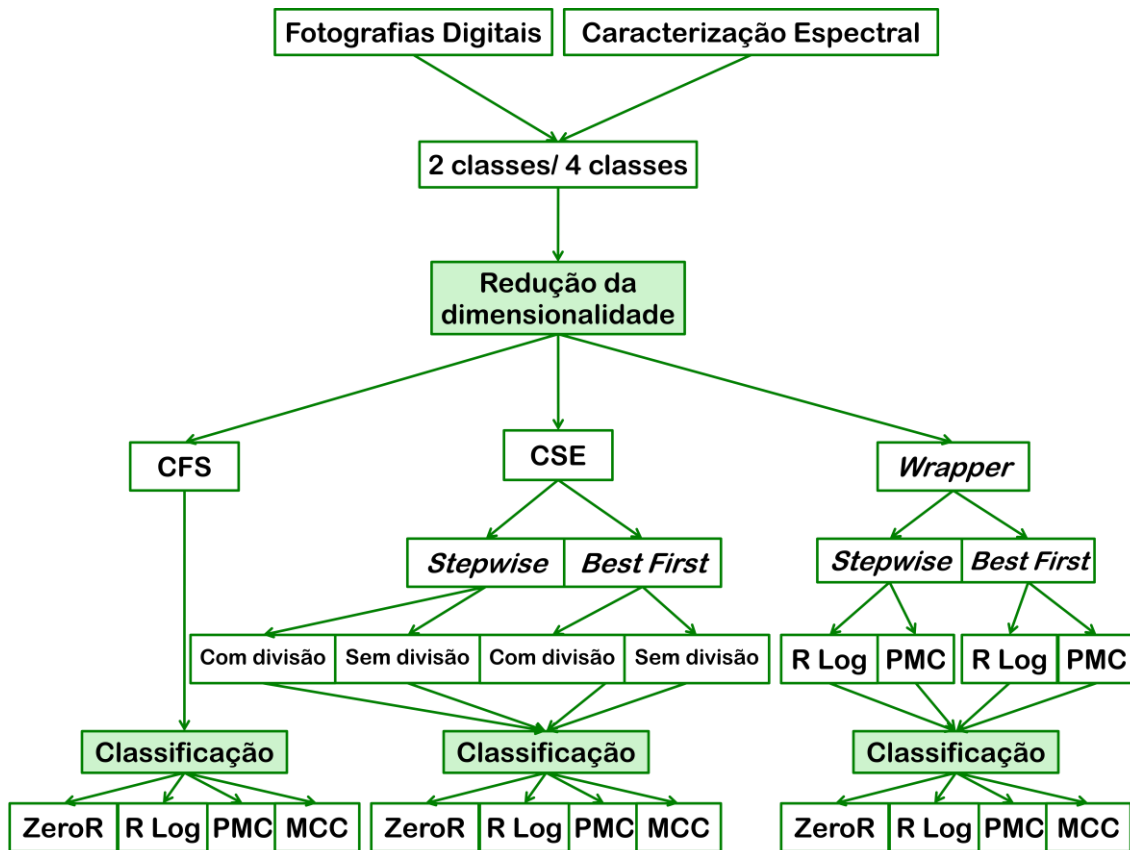


Figura 1: Metodologia de aprendizado de máquinas utilizada para análise das fotografias digitais e caracterização espectral, considerando 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e 2 classes (Com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4). Onde CFS é o *Correlation based Feature Selection*, CSE é *Classifier Subset Evaluator*, R Log é regressão logística, PMC é *perceptron multicamadas*, MCC é *Multi Class Classifier*, "Com divisão" é com divisão percentual do conjunto de treino e "Sem divisão" é sem divisão percentual do conjunto de treino.

Escolha da melhor combinação: redução-classificação

Para se avaliar a eficiência do binômio redução - classificação, criamos um índice chamado eficiência (E) para comparar as diferentes combinações de técnicas:

$$E = \frac{\% \text{ Acerto}}{\text{Dimensão}}$$

Onde E é a eficiência, uma medida adimensional, % Acerto é a porcentagem de acerto da técnica, e a Dimensão é o número de variáveis de entrada selecionadas pelas diferentes técnicas de redução de dimensionalidade.

Combinações de técnicas cuja E seja mais elevada, são as mais adequadas para o problema. Este índice só pode ser utilizado se o conjunto selecionado for multivariado

(mais de uma variável no conjunto), pois, como é uma divisão, se o denominador for muito baixo (< 2), resulta em um viés, obtendo valores altos mesmo quando a porcentagem de acerto da técnica de classificação é baixa.

Também foram feitos gráficos de dispersão, com a porcentagem de acerto no eixo X, a porcentagem de redução no eixo Y, e o tamanho dos pontos variando de acordo com o índice E para avaliar quais as melhores combinações de técnica de redução da dimensionalidade com técnica de classificação.

5.4. Resultados

Redução da Dimensionalidade dos Dados

Diferentes conjuntos de atributos foram selecionados para a classificação em 4 e em 2 classes. Em alguns casos, o *Best First* e *stepwise* chegaram aos mesmos subconjuntos de variáveis, não sendo separados nas tabelas, assim como a divisão ou não do conjunto de treino e a utilização de PMC ou regressão logística como algoritmo de seleção.

Na tabela 1 estão os tons de cinza selecionados por cada técnica de seleção de atributos para cada classe e na tabela 2 estão os comprimentos de onda. Observa-se nas tabelas uma tendência a ter uma menor quantidade de atributos selecionados para se fazer a classificação em um número menos de classes.

Tabela 1: Tons de cinza advindos matriz de pixels de fotografias digitais de folhas de plantas jovens de *Annona emarginata*, selecionados em função das diferentes técnicas de redução da dimensionalidade utilizadas e do número de classes.

Redução da dimensionalidade	Nº Classes	Tons de cinza
CFS	4	28; 30; 32; 33; 34; 60; 72; 78; 83; 151; 168; 174; 188; 199; 207; 214; 220; 222
<i>Wrapper</i> com PMC e <i>stepwise</i>	4	53; 80; 163; 190; 214; 248
<i>Wrapper</i> com PMC e <i>best first</i>	4	15; 53; 80; 163; 190; 213; 214; 219; 221; 248
<i>Wrapper</i> com Logística	4	14; 64; 83; 86; 164; 175; 184; 252
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	4	205; 213
CSE com <i>best first</i> e sem divisão do conjunto de treino	4	14; 29; 70; 84; 85; 87; 95; 97; 105; 117; 128; 129; 146; 170; 179; 187; 204; 213; 231; 238
CSE com <i>Stepwise</i> e com divisão do conjunto de treino	4	167
CSE com <i>best first</i> e com divisão do conjunto de treino	4	41; 134; 167
CFS	2	30; 31; 32; 33; 34; 35; 60; 73; 74; 82; 130; 184; 199; 207; 214; 218; 220; 222
<i>Wrapper</i> com PMC e <i>stepwise</i>	2	72; 73; 82; 84; 86; 183; 214
<i>Wrapper</i> com PMC e <i>best first</i>	2	22; 72; 73; 84; 86; 87; 183; 214; 231
<i>Wrapper</i> com Logística	2	25; 56; 72; 74; 164; 176; 246
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	2	48; 73; 87; 101; 119; 125; 131; 162; 176; 251
CSE com <i>best first</i> e sem divisão do conjunto de treino	2	13; 73; 81; 84; 87; 100; 101; 103; 114; 118; 119; 125; 131; 138; 162; 163; 176; 220; 251
CSE com <i>Stepwise</i> e com divisão do conjunto de treino	2	172; 231
CSE com <i>best first</i> e com divisão do conjunto de treino	2	23; 44; 117; 172; 232

(*) Onde PMC é *Perceptron Multicamadas*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*.

Tabela 2: Comprimentos de onda (em nm) advindos da caracterização espectral de folhas de plantas jovens de *Annona emarginata*, selecionados em função das diferentes seleções de atributos utilizadas e do número de classes.

Redução da dimensionalidade	Nº Classes	Comprimentos de onda selecionados (nm)
CFS	4	1414,553; 1667,666; 1685,089
CSE sem divisão do conjunto de treino	4	1171,599; 1408,854; 1420,255; 1431,666; 1477,404; 1511,806; 1523,293; 1632,882; 1638,673; 1673,471; 1685,089
CSE com divisão do conjunto de treino	4	1104,575; 1586,635; 1644,467; 1656,062; 1685,089
Wrapper com PMC e best first	4	1471,679; 1632,882; 1638,673; 1644,467; 1679,279; 1685,089
Wrapper com logística	4	1004,675; 1598,183; 1679,279; 1685,089
Wrapper com PMC e stepwise	4	966,031; 1471,679; 1632,882; 1638,673; 1685,089
CFS	2	1301,01; 1414,553; 1471,679; 1673,471; 1679,279; 1685,089
CSE com best first e sem divisão do conjunto de treino	2	1391,769; 1397,462; 1609,74; 1685,089
CSE com Stepwise e sem divisão do conjunto de treino	2	1471,679
CSE com divisão do conjunto de treino	2	1346,314
Wrapper com logística	2	1374,706; 1632,882; 1650,263; 1661,862; 1673,471; 1685,089
Wrapper com PMC e best first	2	1143,631; 1238,962; 1414,553; 1420,255; 1598,183; 1615,522; 1632,882; 1685,089
Wrapper com PMC e stepwise	2	1425,96

(*) Onde PMC é *Perceptron Multicamadas*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*.

Classificação em 4 classes

Na tabela 3 estão as porcentagens de acerto médias e erro padrão para a classificação em quatro classes utilizando como dados de entrada os comprimentos de onda selecionados previamente por técnicas de aprendizado de máquinas. Na tabela 4 estão as porcentagens de acerto médias e erro padrão para a classificação em quatro classes utilizando como dados de entrada os tons de cinza selecionados previamente por técnicas de aprendizado de máquinas.

Tabela 3: Porcentagem média de classificação correta nas 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e desvio padrão (entre parênteses) dos dados espectrais de folhas de plantas jovens de *Annona emarginata* para os diferentes subconjuntos de comprimentos de onda selecionados pelas diferentes técnicas de seleções de atributos utilizadas.

Redução da dimensionalidade	Classificação			
	ZeroR	PMC	R Log	MCC
CFS	24,68 (0,28)	74,54 (5,93)	79,71 (3,81)	76,23 (4,62)
CSE sem divisão do conjunto de treino	24,68 (0,28)	77,52 (4,48)	87,51 (3,75)	87,00 (3,69)
CSE com divisão do conjunto de treino	24,68 (0,28)	80,95 (3,77)	85,15 (3,36)	85,73 (3,20)
Wrapper com PMC e <i>best first</i>	24,68 (0,28)	78,45 (4,11)	79,69 (3,91)	76,88 (4,06)
Wrapper com logística	24,68 (0,28)	69,88 (4,50)	83,07 (3,71)	77,77 (4,01)
Wrapper com PMC e <i>stepwise</i>	24,68 (0,28)	79,58 (4,01)	81,66 (4,00)	78,31 (4,04)

(*) Onde PMC é *Perceptron Multicamadas*, R Log é regressão logística e MCC é *Multi Class Classifier*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*. (**) Resultantes de 1000 reamostragens (*bootstrap*).

Tabela 4: Porcentagem média de classificação correta nas 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) e desvio padrão (entre parênteses) dos dados das imagens (histogramas) de folhas de plantas jovens de *Annona emarginata* para os diferentes subconjuntos de tons de cinza selecionados pelas diferentes técnicas de seleções de atributos utilizadas.

Redução da dimensionalidade	Classificação			
	ZeroR	PMC	R Log	MCC
CFS	24,54 (0,28)	44,56 (5,02)	46,10 (5,15)	45,55 (4,97)
CSE com <i>best first</i> e sem divisão do conjunto de treino	24,54 (0,28)	47,73 (5,34)	47,49 (5,29)	47,91 (5,17)
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	24,54 (0,28)	39,62 (4,45)	44,01 (5,30)	39,17 (4,39)
CSE com <i>best first</i> e com divisão do conjunto de treino	24,54 (0,28)	45,64 (4,45)	43,85 (4,38)	46,24 (4,40)
CSE com <i>Stepwise</i> e com divisão do conjunto de treino	24,54 (0,28)	40,48 (4,23)	46,53 (4,28)	38,73 (4,16)
Wrapper com Logística	24,54 (0,28)	58,75 (4,86)	51,12 (5,16)	57,38 (4,80)
Wrapper com PMC e <i>best first</i>	24,54 (0,28)	51,27 (4,94)	49,21 (4,88)	51,29 (4,83)
Wrapper com PMC e <i>stepwise</i>	24,54 (0,28)	54,28 (4,74)	50,34 (4,96)	53,93 (4,92)

(*) Onde PMC é *Perceptron Multicamadas*, R Log é regressão logística e MCC é *Multi Class Classifier*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*. (**) Resultantes de 1000 reamostragens (*bootstrap*)

Todas as médias de classificação correta das redes PMC, regressão logística e MCC se diferem estatisticamente com 5% de significância pelo teste t pareado das médias de acerto do ZeroR, sendo assim, superiores estatisticamente. Isto é um indicativo de que não houve *overfitting*.

Ao se comparar as porcentagens de acerto das mesmas combinações de técnica de redução com técnica de classificação para os dados de fotografias digitais com os dados de NIRS, observa-se uma diferença até nas porcentagens de acerto da técnica ZeroR, sendo que as porcentagens de classificação correta para os dados de fotografias digitais são sempre inferiores às porcentagens para os dados de reflectância foliar no infravermelho próximo.

A maior porcentagem de acerto para os dados de NIRS foi da combinação de redução por CSE sem divisão do conjunto de treino com classificação por regressão logística (87, 51%). Já a maior porcentagem de acerto para os dados de fotografias digitais foi da combinação de redução por *Wrapper* com a regressão logística como algoritmo "caixa-preta" associado a classificação com PMC (58,75%). Os percentuais de acerto para dados de fotografias digitais é, em geral, 30% menor que o para dados espectrais.

Classificação em 2 classes

A redução do número de classes refletiu em um aumento considerável na porcentagem de acerto das classificações com técnicas de aprendizado de máquinas, tanto para os dados espectrais, quanto para os dados de fotografias digitais.

Com a simplificação do problema, reduzindo as classes, as porcentagens de acerto na classificação para os dados de fotografias digitais estão mais perto das porcentagens de acerto para os dados espectrais. A porcentagem de acerto da técnica ZeroR foi levemente superior para os dados de fotografias digitais.

Na tabela 5 estão as porcentagens de acerto médias e desvio padrão para a classificação em duas classes utilizando, como dados de entrada, os comprimentos de onda selecionados previamente por técnicas de aprendizado de máquinas. Na tabela 6 estão as porcentagens de acerto médias e erro padrão para a classificação em duas classes utilizando como dados de entrada os tons de cinza selecionados previamente por técnicas de aprendizado de máquinas.

Tabela 5: Porcentagem média de classificação correta nas 2 classes (Com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4) e desvio padrão (entre parênteses) dos dados espectrais de folhas de plantas jovens de *Annona emarginata* para os diferentes subconjuntos de comprimentos de onda selecionados pelas diferentes técnicas de seleções de atributos utilizadas.

Redução da dimensionalidade	Classificadores			
	ZeroR	PMC	Logística	MCC
CFS	49,68 (0,31)	88,50 (3,58)	91,45 (2,87)	82,93 (5,17)
CSE com <i>best first</i> e sem divisão do conjunto de treino	49,68 (0,31)	88,65 (3,30)	89,67 (2,93)	87,58 (3,68)
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	49,68 (0,31)	64,79 (4,14)	63,78 (4,30)	63,79 (4,33)
CSE com divisão do conjunto de treino	49,68 (0,31)	59,99 (4,23)	61,24 (4,45)	60,14 (4,34)
Wrapper com logística	49,68 (0,31)	87,71 (3,33)	97,05 (1,68)	87,17 (3,61)
Wrapper com PMC e <i>best first</i>	49,68 (0,31)	87,44 (3,66)	93,55 (2,65)	88,44 (3,38)
Wrapper com PMC e <i>stepwise</i>	49,68 (0,31)	65,39 (4,14)	64,36 (4,31)	64,76 (4,25)

(*) Onde PMC é *Perceptron* Multicamadas, R Log é regressão logística e MCC é *Multi Class Classifier*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*.

(**) Resultantes de 1000 reamostragens (*bootstrap*).

A maior porcentagem de acerto com dados espectrais é de 97,05% referente a redução por *Wrapper* com a regressão logística como algoritmo "caixa-preta" associado a classificação com regressão logística. Já a maior porcentagem de acerto com dados de fotografias digitais é de 79,05% da combinação de redução por *Wrapper* com o PMC como algoritmo "caixa-preta" e "rankeamento" por *best first* associado a classificação com PMC.

A classificação por regressão logística se destaca com as maiores porcentagens de acerto para os dados espectrais. Já para os dados de fotografias digitais, a classificação com PMC se destaca com as maiores porcentagens de acerto.

Tabela 6: Porcentagem média de classificação correta nas 2 classes (Com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4) e desvio padrão (entre parênteses) dos dados das imagens (histogramas) de folhas de plantas jovens de *Annona emarginata* para os diferentes subconjuntos de tons de cinza selecionados pelas diferentes técnicas de seleções de atributos utilizadas.

Redução da dimensionalidade	Classificadores			
	ZeroR	PMC	Logística	MCC
CFS	49,80 (0,29)	77,27 (4,44)	72,67 (4,25)	76,81 (4,40)
CSE com <i>best first</i> e sem divisão do conjunto de treino	49,80 (0,29)	73,43 (4,49)	71,71 (4,48)	73,55 (4,56)
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	49,80 (0,29)	72,47 (4,46)	73,91 (4,18)	74,17 (4,16)
CSE com <i>best first</i> e com divisão do conjunto de treino	49,80 (0,29)	67,08 (4,42)	67,24 (4,16)	66,84 (4,23)
CSE com <i>Stepwise</i> e com divisão do conjunto de treino	49,80 (0,29)	64,67 (7,61)	52,89 (4,26)	55,60 (6,15)
Wrapper com Logística	49,80 (0,29)	75,26 (4,56)	78,88 (4,04)	76,77 (4,25)
Wrapper com PMC e <i>best first</i>	49,80 (0,29)	79,05 (4,20)	75,64 (4,05)	78,82 (4,31)
Wrapper com PMC e <i>stepwise</i>	49,80 (0,29)	78,93 (4,10)	74,74 (4,08)	78,61 (4,59)

(*) Onde PMC é *Perceptron Multicamadas*, R Log é regressão logística e MCC é *Multi Class Classifier*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*.

(**) Resultantes de 1000 reamostragens (*bootstrap*).

Escolha da melhor combinação: redução-classificação

Nas tabelas 7 e 8 estão o percentual de redução da dimensionalidade dos dados e a eficiência da combinação das técnicas de redução e classificação para classificação em quatro e em duas classes respectivamente.

Nas tabelas 7 e 8 observa-se que a técnica ZeroR, o controle, tende a ter sempre os menores valores de E.

Tabela 7: Porcentagem de redução da dimensionalidade dos dados espectrais de folhas de plantas jovens de *Annona emarginata* e eficiência da combinação das técnicas (E) para os diferentes subconjuntos de comprimentos de onda selecionados pelas técnicas diferentes técnicas de aprendizado de máquinas, para classificação em 4 classes (Concentração de nitrogênio - 0 mg.L-1 de N (T1); 52,5 mg.L-1 de N (T2); 105 mg.L-1 de N (T3) e 210 mg.L-1 de N (T4)).

Redução	Classificação	Dimensão	% Redução	E
CFS	ZeroR	3	97,66	8,23
CFS	PMC	3	97,66	24,85
CFS	R Log	3	97,66	26,57
CFS	MCC	3	97,66	25,41
CSE sem divisão do conjunto de treino	ZeroR	11	91,41	2,24
CSE sem divisão do conjunto de treino	PMC	11	91,41	7,05
CSE sem divisão do conjunto de treino	R Log	11	91,41	7,95
CSE sem divisão do conjunto de treino	MCC	11	91,41	7,91
CSE com divisão do conjunto de treino	ZeroR	5	96,09	4,94
CSE com divisão do conjunto de treino	PMC	5	96,09	16,19
CSE com divisão do conjunto de treino	R Log	5	96,09	17,03
CSE com divisão do conjunto de treino	MCC	5	96,09	17,15
Wrapper com PMC e <i>best first</i>	ZeroR	6	95,31	4,11
Wrapper com PMC e <i>best first</i>	PMC	6	95,31	13,07
Wrapper com PMC e <i>best first</i>	R Log	6	95,31	13,28
Wrapper com PMC e <i>best first</i>	MCC	6	95,31	12,81
Wrapper com logística	ZeroR	4	96,87	6,17
Wrapper com logística	PMC	4	96,87	17,47
Wrapper com logística	R Log	4	96,87	20,77
Wrapper com logística	MCC	4	96,87	19,44
Wrapper com PMC e <i>stepwise</i>	ZeroR	5	96,09	4,94
Wrapper com PMC e <i>stepwise</i>	PMC	5	96,09	15,92
Wrapper com PMC e <i>stepwise</i>	R Log	5	96,09	16,33
Wrapper com PMC e <i>stepwise</i>	MCC	5	96,09	15,66

(*)Onde PMC é *Perceptron Multicamadas*, R Log é regressão logística e MCC é *Multi Class Classifier*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*.

Tabela 8: Porcentagem de redução da dimensionalidade dos dados espectrais de folhas de plantas jovens de *Annona emarginata* e eficiência da combinação das técnicas (E) para os diferentes subconjuntos de comprimentos de onda selecionados pelas técnicas diferentes técnicas de aprendizado de máquinas, para classificação em 2 classes (Com falta de nitrogênio - T1+T2 e sem falta de nitrogênio - T3+T4).

Redução	Classificação	Dimensão	% Redução	E
CFS	ZeroR	6	95,31	8,28
CFS	PMC	6	95,31	14,75
CFS	R Log	6	95,31	15,24
CFS	MCC	6	95,31	13,82
CSE com <i>best first</i> e sem divisão do conjunto de treino	ZeroR	4	96,87	12,42
CSE com <i>best first</i> e sem divisão do conjunto de treino	PMC	4	96,87	22,16
CSE com <i>best first</i> e sem divisão do conjunto de treino	R Log	4	96,87	22,42
CSE com <i>best first</i> e sem divisão do conjunto de treino	MCC	4	96,87	21,89
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	ZeroR	1	99,22	-
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	PMC	1	99,22	-
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	R Log	1	99,22	-
CSE com <i>Stepwise</i> e sem divisão do conjunto de treino	MCC	1	99,22	-
CSE com divisão do conjunto de treino	ZeroR	1	99,22	-
CSE com divisão do conjunto de treino	PMC	1	99,22	-
CSE com divisão do conjunto de treino	R Log	1	99,22	-
CSE com divisão do conjunto de treino	MCC	1	99,22	-
Wrapper com logística	ZeroR	6	95,31	8,28
Wrapper com logística	PMC	6	95,31	14,62
Wrapper com logística	R Log	6	95,31	16,17
Wrapper com logística	MCC	6	95,31	14,53
Wrapper com PMC e <i>best first</i>	ZeroR	5	96,09	9,93
Wrapper com PMC e <i>best first</i>	PMC	5	96,09	17,49
Wrapper com PMC e <i>best first</i>	R Log	5	96,09	18,71
Wrapper com PMC e <i>best first</i>	MCC	5	96,09	17,69
Wrapper com PMC e <i>stepwise</i>	ZeroR	1	99,22	-
Wrapper com PMC e <i>stepwise</i>	PMC	1	99,22	-
Wrapper com PMC e <i>stepwise</i>	R Log	1	99,22	-
Wrapper com PMC e <i>stepwise</i>	MCC	1	99,22	-

(*)Onde PMC é *Perceptron Multicamadas*, R Log é regressão logística e MCC é *Multi Class Classifier*, CFS é *Correlation based Feature Selection* e CSE é *Classifier Subset Evaluator*

Com os gráficos de dispersão (figura 2 e 3), podemos ver que a eficiência (tamanho dos pontos) varia conforme a posição no gráfico, de forma que tende a ser maior quando se localiza no canto superior direito do gráfico.

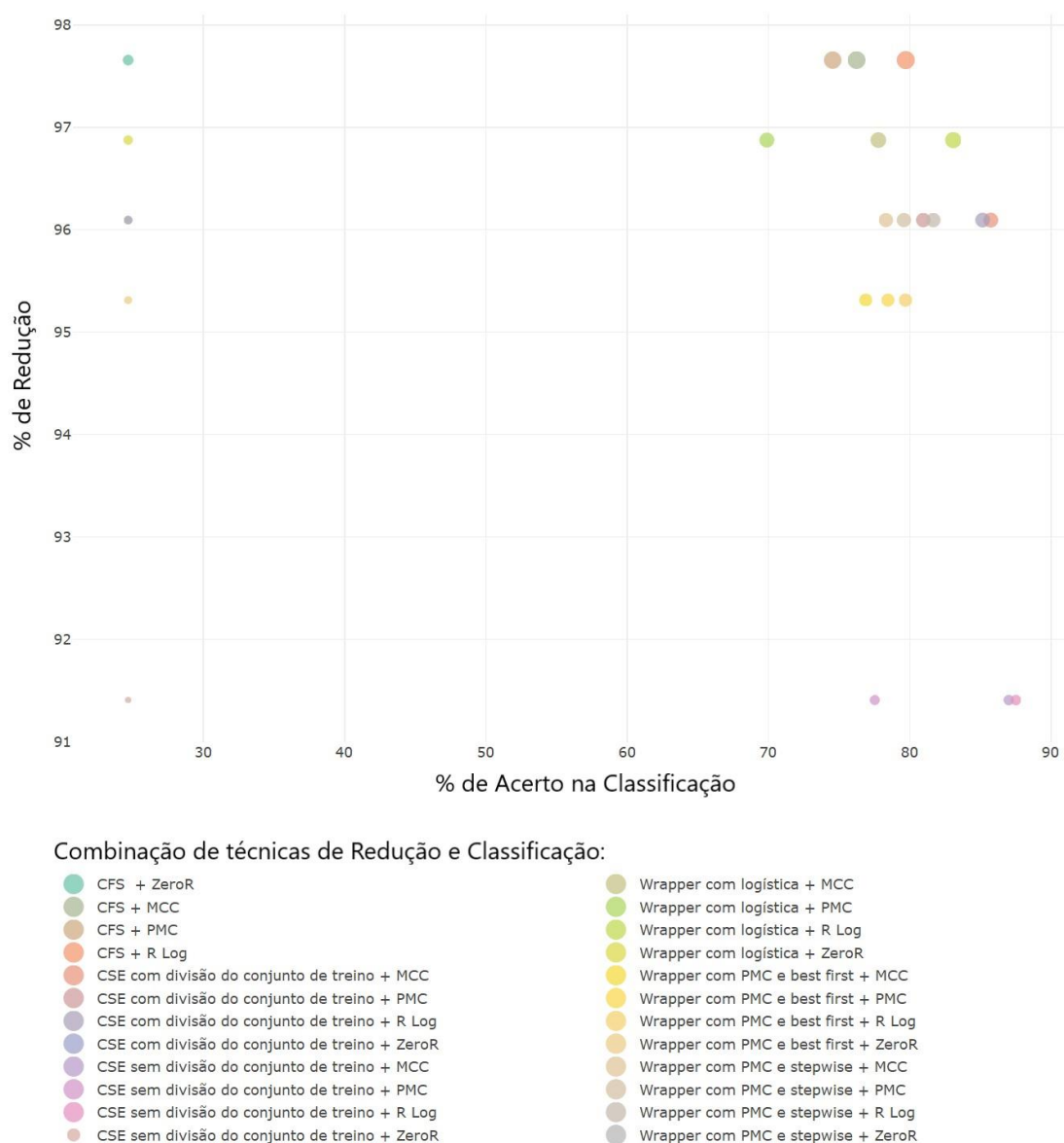
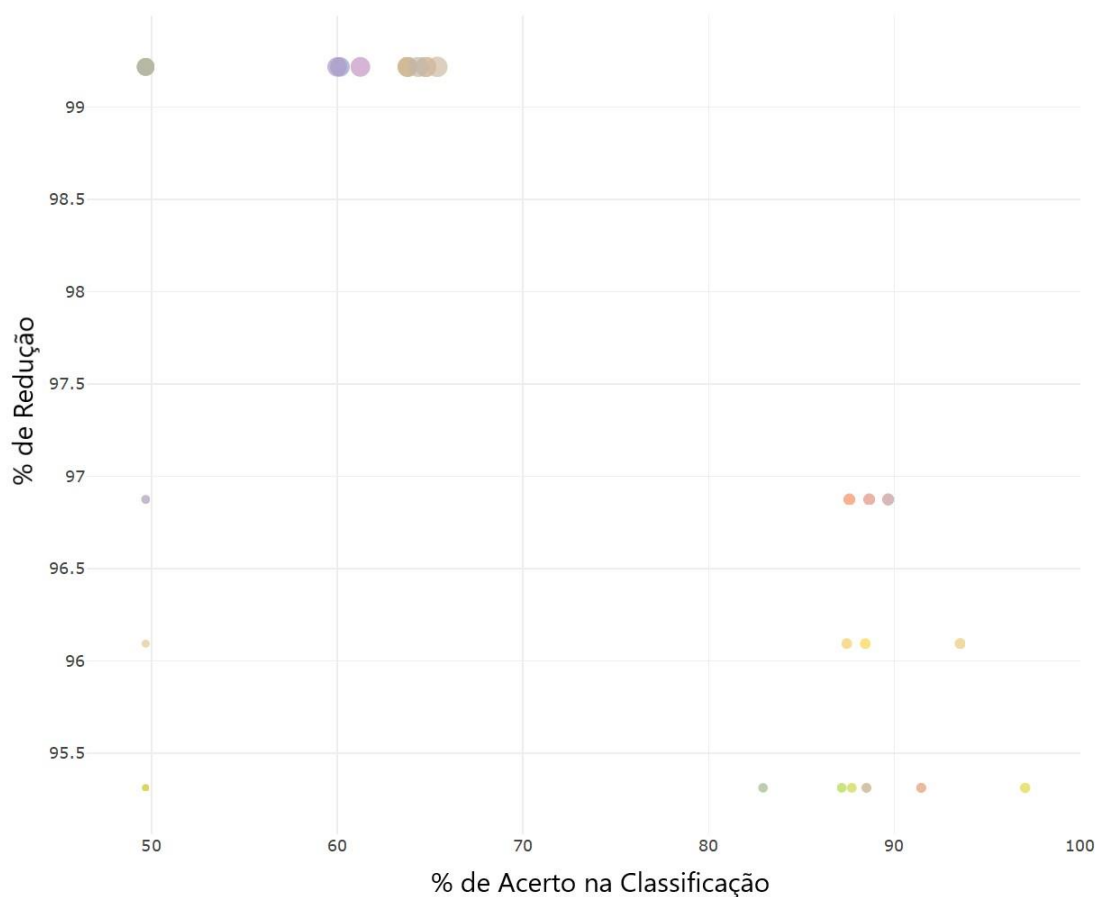


Figura 2: Gráfico de dispersão da porcentagem de redução de variáveis dos conjuntos de dados e as porcentagens de acerto na classificação das diferentes combinações de técnicas para classificação em 4 classes (Concentração de nitrogênio - 0 mg.L⁻¹ de N (Tratamento 1); 52,5 mg.L⁻¹ de N (Tratamento 2); 105 mg.L⁻¹ de N (Tratamento 3) e 210 mg.L⁻¹ de N (Tratamento 4)), com o tamanho do marcador variando de acordo com o valor do índice Eficiência.



Combinação de técnicas de Redução e de Classificação

- CFS + ZeroR
- CFS + MCC
- CFS + PMC
- CFS + R Log
- CSE com best first e sem divisão do conjunto de treino + MCC
- CSE com best first e sem divisão do conjunto de treino + PMC
- CSE com best first e sem divisão do conjunto de treino + R Log
- CSE com best first e sem divisão do conjunto de treino + ZeroR
- CSE com divisão do conjunto de treino + MCC
- CSE com divisão do conjunto de treino + PMC
- CSE com divisão do conjunto de treino + R Log
- CSE com divisão do conjunto de treino + ZeroR
- CSE com Stepwise e sem divisão do conjunto de treino + MCC
- CSE com Stepwise e sem divisão do conjunto de treino + PMC
- CSE com Stepwise e sem divisão do conjunto de treino + R Log
- CSE com Stepwise e sem divisão do conjunto de treino + ZeroR
- Wrapper com logística + MCC
- Wrapper com logística + PMC
- Wrapper com logística + R Log
- Wrapper com logística + ZeroR
- Wrapper com PMC e best first + MCC
- Wrapper com PMC e best first + PMC
- Wrapper com PMC e best first + R Log
- Wrapper com PMC e best first + ZeroR
- Wrapper com PMC e stepwise + MCC
- Wrapper com PMC e stepwise + PMC
- Wrapper com PMC e stepwise + R Log
- Wrapper com PMC e stepwise + ZeroR

Figura 3: Gráfico de dispersão da porcentagem de redução de variáveis dos conjuntos de dados e as porcentagens de acerto na classificação das diferentes combinações de técnicas para classificação em 2 classes (com falta de nitrogênio e sem falta de nitrogênio), com o tamanho do marcador variando de acordo com o valor do índice Eficiência.

Pigmentos Foliares

Para os pigmentos foliares não houve diferença estatística ($p>0,05$) para nenhuma das variáveis testadas segundo os modelos lineares generalizados. Na tabela 9 seguem as médias de clorofila *a*, clorofila *b*, carotenóides e antocianinas para cada tratamento em mg.g^{-1} .

Tabela 9: Quantidades médias de pigmentos foliares (clorofila *a*, clorofila *b*, antocianinas e carotenóides) em mg.g^{-1} e desvio padrão (entre parênteses) em mg.g^{-1} para os 4 tratamentos - Trat (Concentração de nitrogênio - 0 mg.L^{-1} de N (T1); 52,5 mg.L^{-1} de N (T2); 105 mg.L^{-1} de N (T3) e 210 mg.L^{-1} de N (T4)).

Tratamento	Clorofila a (mg/g)	Clorofila b (mg/g)	Antocianinas (mg/g)	Carotenóides (mg/g)
T1	0,6033	0,2318	0,2151	0,3088
	(0,0288)	(0,0194)	(0,0714)	(0,0303)
T2	0,7635	0,2769	0,1738	0,3319
	(0,2564)	(0,0968)	(0,1010)	(0,1061)
T3	0,5830	0,2032	0,1646	0,2575
	(0,1892)	(0,0588)	(0,0543)	(0,0807)
T4	0,6452	0,2439	0,1991	0,2909
	(0,2021)	(0,0949)	(0,0829)	(0,0798)

Discussão

São escassas as aplicações de aprendizado de máquinas para dados de nutrição mineral de plantas arbóreas e nativas em decorrência de seu pouco apelo comercial em larga escala. Os resultados deste trabalho se mostraram muito promissores e possibilitam no futuro a criação de um sistema autônomo de monitoramento de mudas Anonáceas em viveiros ou de plantas jovens no início de sua adaptação no campo.

Ao se comparar as porcentagens de acerto das técnicas utilizando os dados espectrais com as mesmas técnicas com os dados das imagens, pode-se observar que o infravermelho próximo obtém sempre melhores classificações que as imagens com as bandas do visível, isso nos leva a crer que as informações contidas na faixa que abrange o NIR são mais relevantes para discriminar nitrogênio em plantas arbóreas jovens do que as informações na faixa do visível.

Existem fatores intrínsecos e extrínsecos que dificultam a classificação automática de plantas utilizando imagens no visível (BARBEDO, 2016). Desta forma, o baixo desempenho em comparação com o NIRS pode se dar em decorrência de efeito de fundo da imagem e condição da luz na captura da imagem (BARBEDO, 2016), problemas que não afetam a NIRS.

Outro fator que pode ter influenciado uma maior efetividade da caracterização espectral em detrimento às fotografias digitais é que não houve diferença dos pigmentos foliares entre os tratamentos, e estes são os principais componentes químicos da planta detectados na faixa do visível (CARTER; KNAPP, 2001), desta forma, a classificação mais efetiva ocorre quando se utiliza a faixa do espectro mais relacionada com os demais compostos das folhas.

Tran *et al.* (2019) utilizaram redes neurais artificiais para classificar imagens no visível de tomateiros com deficiência de macronutrientes. Esta classificação foi possível pois, nesse caso, a cultura utilizada tem um metabolismo mais rápido e as plantas já apresentavam sinais visuais desta deficiência, o que não ocorreu no nosso caso.

De acordo com Chilingaryan, Sakkarieh e Whelan (2019), os algoritmos de aprendizado de máquinas têm sido cada vez mais utilizados pela agricultura de precisão para estimação de N em plantas, mas, quando aplicadas com dados de imagens como variáveis de entrada, são utilizados os comprimentos de onda de 400 à 900 nm ou índices espectrais de vegetação formados por comprimentos de ondas presentes nesta faixa do espectro. Desta forma, este trabalho torna possível identificar a falta de N antes do aparecimento de sinais visuais por explorar outra faixa do espectro, de 900 à 1700 nm, que não está ligada apenas aos sinais visuais da deficiência, mas a composição química da folha (MAHLEIN, 2016).

A associação de imageamento hiperespectral no NIR com aprendizado de máquinas já foi utilizado para detectar dano mecânico em manga, antes do aparecimento de manchas marrons nas frutas (RIVERA *et al.*, 2014), mostrando a capacidade de captar pequenas alterações antes de se tornarem visíveis. Esse poder preditivo, associado as altas porcentagens de acerto das técnicas de aprendizado de máquinas possibilita a criação de um sistema autônomo de monitoramento da qualidade de mudas.

Quando se simplifica o problema, reduzindo de quatro para duas classes, a quantidade de comprimentos de onda selecionados também tende a ser menor (tabela 3), isso prova que quanto mais simples um problema, uma menor quantidade de informação é necessária para se fazer a classificação. Em contrapartida, quando apenas um comprimento de onda foi selecionado, as porcentagens de acerto também reduziram consideravelmente, o que nos leva a crer que a informação contida individualmente não é suficiente para uma classificação apurada, mesmo para um problema simples de apenas duas classes. Em geral, a regressão logística teve as maiores porcentagens de acerto que as demais técnicas, se mostrando mais adequada para este tipo de problema.

Quanto a escolha da melhor combinação de técnica de seleção de atributos e técnica de classificação, o índice E (Eficiência) ajuda a tomar a decisão para a classificação em 4 classes, mas apresentou um problema para a classificação em duas classes, já que pode apenas ser aplicado quando o conjunto selecionado pela técnica de seleção de atributos tem mais de 1 atributo.

Deste forma, focando na classificação em 4 classes, observamos que na figura 2, as técnicas com maior porcentagem de acerto e de redução tem os maiores pontos, ou seja, nestes casos o índice eficiência é maior. A melhor combinação de técnica de seleção e classificação para classificar em 4 classes os dados de NIRS foi a redução por CFS e classificação por regressão logística, atingindo o E de 26,57 (Tabela 7). Esta combinação acertou em média 79,71% (tabela 3), utilizando apenas 3 comprimentos de onda como conjunto de entrada (tabela 7). A maior média de porcentagem de acerto, de 87,51% foi da combinação de redução por CSE sem divisão do conjunto de treino com classificação por regressão logística, que utilizou 11 comprimentos de onda no seu conjunto de entrada, o maior conjunto de entrada selecionado para classificação em 4 classes. Neste caso, apenas 11 comprimentos de onda ainda é um número de atributos baixo.

Quando a dimensão dos dados é muito grande, acaba gerando um ruído na resposta em decorrência de informações irrelevantes e redundantes (HALL; SMITH, 1998). Mas, quando a quantidade de atributos é muito pequena, também acaba sendo um problema, pois informações importantes são perdidas. Desta forma, é necessário selecionar um conjunto de atributos pequeno, não ínfimo, para se poder fazer uma boa classificação. No caso deste trabalho, para a classificação em 4 classes, talvez seja mais interessante manter 11 atributos e ter uma maior porcentagem de acerto na classificação do que reduzir os dados de entrada, facilitando o processamento, mas perdendo na porcentagem de acerto.

Diferentes técnicas de imageamento de plantas associadas á aprendizado de máquinas podem ser aplicadas com diversos objetivos. Neste trabalho mostramos uma aplicação desta combinação, mas seu potencial ainda deve ser muito explorado em plantas nativas no futuro.

Conclusões

Podemos concluir que há aplicabilidade de algoritmos de aprendizado de máquinas para fazer a calibração da técnica NIRS e que a NIRS é mais efetiva que as fotografias digitais para identificar variações no nitrogênio foliar.

Referências

- BARBEDO, J.C.A. A review on the main challenges in automatic plant disease identification based on visible range images. **Biosyst. Eng.**, v.144, p.52–60, 2016.
- BARON, D.; FERREIRA, G. ; RODRIGUES, J. D. ; BOARO, C.S.F.; MACEDO, A. C.. Gs exchange, phisiological indexes and ionic accumulation in *Annona emarginata* (Schltdl.) H. Rainer seedlings in nutrients solution. **Revista Brasileira de Fruticultura**, Jaboticabal-SP, v. 36, p. 361-376, 2013.
- BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: resultados de um novo olhar**. Brasília, DF: Embrapa, 2014. 596 p. (disponível no link: <http://www.alice.cnptia.embrapa.br/alice/handle/doc/1003690>)
- CARTER, G. A.; KNAPP, A. K. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. **Am. J. Bot.** v.88, p.677-684, 2001. <https://doi.org/10.2307/2657068>
- CHILINGARYAN, A.; SAKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. **Computers and Electronics in Agriculture**, v.151, p.61-69, 2018. <https://doi.org/10.1016/j.compag.2018.05.012>
- HALL, M. A.; SMITH, L. A. . **Practical feature subset selection for machine learning**. In C. McDonald (Ed.), Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth, 4-6 February, (pp. 181-191). Berlin: Springer, 1998.
- LOWENBERG-DEBOER. J.; SWINTON, S. M. **Economics of site-specific management in agronomic crops**. In 'The State of Site-Specific Management for Agriculture' (F. J. Pierce and E. J. Sadler, Eds.), ASA Miscellaneous Publication, pp. 369-396. ASA, CSSA, and SSSA, Madison, 1997.
- MAHLEIN, Anne-Katrin. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. **Plant Disease**, v.100, pp. 241-251, 2016. <https://doi.org/10.1094/PDIS-03-15-0340-FE>

- MULLA, D. J.; SCHEPERS, J. S. **Key processes and properties for site-specific management.** In “The State of Site-Specific Management for Agriculture” (F. J. Pierce and E. J. Sadler, Eds.), ASA Miscellaneous Publication, pp. 1-18. ASA, CSSA, and SSSA, Madison, WI, 1997.
- NOWAK, P. **A sociological analysis of site specific management.** In “The State of Site-Specific Management for Agriculture” (F. J. Pierce and E. J. Sadler, Eds.), ASA Miscellaneous Publication, pp. 397-422. ASA, CSSA, and SSSA, Madison, WI, 1997.
- PIERCE, F. J.; NOWAK, P. Aspects of Precision Agriculture. **Advances in Agronomy**, Vol. 67, Pp. 1-85, 1999. [https://doi.org/10.1016/S0065-2113\(08\)60513-1](https://doi.org/10.1016/S0065-2113(08)60513-1)
- RIVERA, N.V.; GÓMEZ-SANCHIS, J.; CHANONA-PÉREZ, J.; CARRASCO, J.J.; MILLÁN-GIRALDO, M.; LORENTE, D.; CUBERO, S.; BLASCO, J. Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning. **Biosystems Engineering**, v.122, p.91-98, 2014. <https://doi.org/10.1016/j.biosystemseng.2014.03.009>
- SCALOPPI JUNIOR, E.J. Propagação de espécies de Annonaceae com estacas caulinares. 2007. 92f. Tese (Doutorado) – Faculdade de Ciências Agronômicas e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2007.
- SOBRINHO, R.B. Potencial de exploração de anonáceas no Nordeste do Brasil. In: 17º Semana internacional da fruticultura, floricultura e agroindústria [**Anais...**] Fortaleza: Instituto Frutal, 2010.
- TOKUNAGA, T. **A cultura da atemoia.** 2.ed. Campinas: CATI, 2005. 80 p. (Boletim técnico)
- TRAN, T.; CHOI, J.; LE, T. H.; KIM, J. A Comparative Study of Deep CNN in Forecasting and Classifying the Macronutrient Deficiencies on Development of Tomato Plant. **Appl. Sci.** v.9(8), p.1601, 2019. <https://doi.org/10.3390/app9081601>
- WHELAN, B.M., MCBRATNEY, A.B., BOYDELL, B.C.. The Impact of Precision Agriculture. **Proceedings of the ABARE Outlook Conference**, ‘The Future of Cropping in NW NSW’, Moree, UK, July 1997, p. 5. 1997.
- ZHANG, N.; WANG, M.; WANG, N. Precision agriculture — a worldwide overview. **Computers and Electronics in Agriculture**, v.36, pp.113-132. 2002. [https://doi.org/10.1016/S0168-1699\(02\)00096-0](https://doi.org/10.1016/S0168-1699(02)00096-0)

6. Considerações Finais

Este estudo é inovador por aplicar NIRS em uma espécie nativa e arbórea com o intuito de averiguar o status nutricional de N, sendo o primeiro a associar NIRS,

aprendizado de máquinas e fotografias digitais a uma espécie do gênero *Annona*. Ele abre portas para novas aplicações de NIRS e aprendizado de máquinas em plantas da família Annonaceae.

As plantas desta família desempenharão um importante papel no futuro do controle biológico de pragas, sendo essenciais para a produção de extratos. Isso ocorre em decorrência de sua alta produção de alcaloides, que são moléculas nitrogenadas. Assim, o monitoramento de N nestas plantas é essencial para uma futura linha de produção de Annonáceas com o intuito da fabricação destes extratos. Desta forma, este trabalho é o começo da produção em silvicultura de precisão para esta família.

7. Referências Bibliográficas

ADHIKARI, A. DENERO, J. **Artificial Intelligence: Foundations of Computational Agents**, second edition, Cambridge University Press, 2017.

- Annona* in **Flora do Brasil 2020 em construção**. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB110243>>. Acesso em: 26 Nov. 2019
- AQUINO, C.F.; Salomão, L.C.C.; Azevedo, A.M. Fenotipagem de alta eficiência para vitamina A em banana utilizando redes neurais artificiais e dados colorimétricos. **Bragantia**, Campinas, v. 75, n. 3, p.268-274, 2016.
- ATTIQ, A.; JALIL, J.; HUSAIN, K. Annonaceae: Breaking the Wall of Inflammation. **Front. Pharmacol.** v.8 p.752-776. 2017. doi: 10.3389/fphar.2017.00752
- BACKHAUS, A.; BOLLENBECK, F.; SEIFFERT, U. Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks. Proceedings of the 3rd IEEE Workshop on Hyperspectral Imaging and Signal Processing: **Evolution in RemoteSensing WHISPERS** 2011, page 9. IEEE Press, 2011.
- BAIN, K.F.; VERGÉS A.; POORE, A.G.B. Using near infra red reflectance spectroscopy (NIRS) to quantify tissue composition in the seagrass *Posidonia australis*. **Aquatic Botany**, v.111, p.66–70, 2013.
- BARANOSKI, G.V.G.; ROKNE, J.G. . Efficiently simulating scattering of light by leaves. **The Visual Computer**, v.17, p.491–505, 2001.
- BARBOSA, Z.; SOARES, I.; CRISÓSTOMO, L.A. Crescimento e absorção de nutrientes por mudas de graviola. **Rev. Bras. Frutic.**, Jaboticabal - SP, v. 25, n. 3, p. 519-522, 2003.
- BARKER, A.V.; PILBEAM, D.J. **Handbook of plant nutrition**. CRS Press, v.2., 2015.
- BARON, D.; AMARO, A. C. E.; MACEDO, A. C.; BOARO, C. S. F.; FERREIRA, G.. Is *annona emarginata* capable of accumulate essential heavy metals without damages in the metabolism?. **Rev. Bras. Frutic.**, Jaboticabal , v. 39, n. 4, e-646, 2017 .
- BARON, D.; BRAVO, J.P.; MAIA, I.G.; PINA, A.; FERREIRA, G.. UGP gene expression and UDP-glucose pyrophosphorylase enzymatic activity in grafting annonaceous plants. **Acta Physiol Plant**, v.38, p.79, 2016.
- BARON, D.; FERREIRA, G. ; RODRIGUES, J. D. ; BOARO, C.S.F.; MACEDO, A. C.. Gs exchange, physiological indexes and ionic accumulation in *Annona emarginata* (Schltdl.) H. Rainer seedlings in nutrients solution. **Revista Brasileira de Fruticultura**, Jaboticabal-SP, v. 36, p. 361-376, 2013.

- BATISTA, M.M.F.; VIÉGAS, I.J.M.; FRAZÃO, D.A.C.; THOMAZ, M.A.A.; SILVA, R.C.L. EFEITO DA OMISSÃO DE MACRONUTRIENTES NO CRESCIMENTO, NOS SINTOMAS DE DEFICIÊNCIAS NUTRICIONAIS E NA COMPOSIÇÃO MINERAL EM GRAVIOLEIRAS (*Annona muricata*). **Rev. Bras. Frutic.**, Jaboticabal - SP, v. 25, n. 2, p. 315-318, Agosto 2003.
- BINOTI, D.H.B.; BINOTI, M.L.M.S.; LEITE, H.G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Ciência da Madeira (Braz. J. Wood Sci.)**, Pelotas, v. 05, n. 01, p. 58-67, Maio de 2014.
- BIRTH, S.G.; MCVEY, R.G. Measuring the color of growing turf with a reflectance spectrophotometer. **Agronomy Journal**, v.60, p.640–643, 1968.
- BLACKBURN, G.A. Quantifying chlorophylls and carotenoids at leaf and canopy scale: an evaluation of some hyperspectral approaches. **Remote Sensing of Environment**, v.66, p.273–285, 1998.
- BLOOM, A. J. et al. Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. **Science**, v.328, p.899–903, 2010.
- BUCENE, L.C.; RODRIGUES, L.H.A. Utilização de redes neurais artificiais para avaliação de produtividade do solo, visando classificação de terras para irrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.8, n.2/3, p.326-329, 2004.
- BUCHANAN, GRUISSEM, JONES. **Biochemistry & Molecular Biology of Plants**. Rockville: American Society of Plant Physiologists (ASPP0), 2nd ed., 2015.
- BURNS, D.A., CIURCZACK, E.W. Handbook of Near Infrared Analysis. New York: Marcel Dekker, 2nd ed. 2001.
- CAJAL, S. R. Y. The Croonian Lecture: La Fine Structure des Centres Nerveux. **Proceedings of the Royal Society of London**, v. 55, n. 331-335, p. 444–468, 1 jan. 1894.
- CAMPOS, F. G.; BARON, D.; MARQUES, R.M. de ; FERREIRA, G. ; BOARO, C.S.F . Characterization of the chemical composition of the essential oils from *Annona emarginata* (Schltdl.) H. Rainer 'terra-fria' and *Annona squamosa* L.. **Revista Brasileira de Fruticultura** (Impresso), v. 36, p. 202-208, 2014.
- CARTER, G. A.; KNAPP, A. K. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. **Am. J. Bot.** v.88, p.677-684, 2001. <https://doi.org/10.2307/2657068>

- CHATROU, L.W. et al. The natural history of Annonaceae. **Botanical Journal of the Linnean Society**, v.169, p.1–4, 2012.
- CHIDAMBARAM, S.; SRINIVASAGAN, K.G. **Optimization Approach for Feature Selection and Classification with Support Vector Machine**. In: Behera H., Mohapatra D. (eds) Computational Intelligence in Data Mining—Volume 1. Advances in Intelligent Systems and Computing, New Delhi:Springer, vol 410., 2016. https://doi.org/10.1007/978-81-322-2734-2_11
- CONSELHO ESTADUAL DO MEIO AMBIENTE, RIO GRANDE DO SUL. Decreto estadual CONSEMA n. 42.099 de 31 de dezembro de 2002. Declara as espécies da flora nativa ameaçadas de extinção no estado do Rio Grande do Sul e da outras providências, Palácio Piratini, Porto Alegre, Rio Grande do Sul, 31 dez. 2002, 2002.
- COSTA, A.M.; CARLOS, L.; SILVA, P.O.; BARBOSA, K.P.; RODRIGUES, C.R. . Nitrogen and Potassium Fertilization in the Initial Growth of *Annona crassiflora* Mart. **Floresta Ambient.**, Seropédica, v.26, n.2, e20170341, 2019.
- COSTA, S.L. da; CARVALHO, A.J.C. de; PESSANHA, P.G.O.; MONNERAT, P.H.; MARINHO, C.S. Produtividade da cultura da pinha (*Annona squamosa* L.) em função de níveis de adubação nitrogenada e formas de aplicação de boro. **Rev. Bras. Frutic.**, Jaboticabal , v. 24, n. 2, p. 543-546, Aug. 2002.
- COUVREUR, T.L.; PIRIE, M.D.; CHATROU, L.W.; SAUNDERS, R.M.; SU, Y.C.; RICHARDSON, J.E.; et al. . Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. **J. Biogeogr.**, v.38, p.664–680, 2011. doi: 10.1111/j.1365-2699.2010.02434.x
- CURRAN, P.J.. Remote sensing of foliar chemistry. **Remote Sensing of Environment**, v.30, p.271–278, 1989.
- DONADIO, L. C. Situação atual e perspectivas das anonáceas. In: SÃO JOSÉ, A. R.; VILAS BOAS, I.; MARAIS, O. M.; REBOUÇAS, T. N. H. (Ed.). **Anonáceas: produção e mercado** (pinha, graviola, atemóia e cherimólia). Vitória da Conquista: Universidade Estadual do Sudoeste da Bahia, 1997. p.1-4.
- EFRON, B. **Bootstrap methods: Another look at the Jackknife**. Ann. Statistics, vol 7(1) p.1-26, 1979.
- EFRON, B.; TIBISHIRANI, R. J. **An introduction to the bootstrap**. New York: John Wiley & Sons, p. 642, 1993.

- ESCRIBANO, P.; VIRUEL M.A.; HORMAZA, J.I. Molecular analysis of genetic diversity and geographic origin within an ex situ germplasm collection of cherimoya by using SSRs. **J Am Soc Hortic Sci**, v.132, p.357–367, 2007.
- EVANS, J.R. Photosynthesis and nitrogen relationship in leaves of C3 plants. **Oecologia**, Berlim, v. 78, p.9-19, 1989.
- EVANS, S.M.; SINCLAIR, E.A.; POORE, A.G.B.; BAIN, K.F.; VERGÉS, A. Assessing the effect of genetic diversity on the early establishment of the threatened seagrass *Posidonia australis* using a reciprocal-transplant experiment. **Restoration Ecology**, v. 26, n. 3, pp. 570–580, May 2018.
- FACELI, K.; LORENA, A.; GAMA, J.; CARVALHO, A. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina, **LTC**, 2011.
- FERNANDES, M.S.; ROSSIELLO, R.O.P. Mineral Nitrogen in Plant Physiology and Plant Nutrition. **Critical Reviews in Plant Sciences**, Boca Raton, v.14, n.2, p.111-148, 1995.
- FERNÁNDEZ, S.; VIDAL, D.; SIMÓN, E.; SUGRAÑES, L. Radiometric characteristics of *Triticum aestivum* cv. Astral under water and nitrogen stress. **International Journal of Remote Sensing**, London, v. 15, n. 9, p. 1867-1884, 1994.
- FERRARESI, Tatiana Maris et al . Espectroscopia de infravermelho na determinação da textura do solo. **Rev. Bras. Ciênc. Solo**, Viçosa , v. 36, n. 6, p. 1769-1777, 2012.
- FONSECA, E.L. da; ROSA, L.M.G.; FONTANA, D.C. Caracterização espectral de *Paspalum notatum* em diferentes níveis de adubação nitrogenada. **Pesquisa agropecuária brasileira**, Brasília, v.37, n.3, p. 365-371, mar. 2002.
- GALVEZ-SOLA L, GARCÍA-SÁNCHEZ F, PÉREZ-PÉREZ JG, et al. Rapid estimation of nutritional elements on citrus leaves by near infrared reflectance spectroscopy. **Frontiers in Plant Science**. v.6, a.571, 2015. doi:10.3389/fpls.2015.00571.
- GAMON, J.A.; PEÑEULAS, J.; FIELD, C.B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. **Remote Sensing of Environment**, v.41, p.35–44, 1992.
- GITELSON, A.A.; MERZLYAK, M.N. Spectral reflectance changes associate with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. **Journal of Plant Physiology**, v.143, p.286–292, 1994.

- HAIR, Joseph F., Jr.. ANDERSON, Rolph E., TATHAM, Ronald L., BLACK William C. **Multivariate Data Analysis: with readings** (fourth edition). Upper Saddle River, New Jersey: Prentice Hall, 1995. 745 pp.
- HALL, M.A.. **Correlation-based feature selection for machine learning**. PhD thesis, Department of Computer Science, University of Waikato, Hamilo, New Zealand, 1998.
- HALL, M.A. Correlation-Based Feature Selection for Machine Learning. **Department of Computer Science**, v.19, 2000.
- HALL, M. A.; SMITH, L. A. . **Practical feature subset selection for machine learning**. In C. McDonald (Ed.), Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth, 4-6 February, (pp. 181-191). Berlin: Springer, 1998.
- HAYKIN, S. **Redes Neurais: Principio e Prática**. 2. ed. Porto Alegre: Bookman, 2001.
- HODGKIN, A. L.; HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. **The Journal of Physiology**. v.117, n.4, p.500–544, 1952.
- JACQUEMOUD, S.; USTIN, S. L. Leaf optical properties: a state of the art. in: **Proceedings of the 8th International Symposium Physical Measurements & Signatures in Remote Sensing**, 8–12 January 2001, CNES, Aussois, France. p.223-232, 2001.
- JASINSKI, S.; LÉCUREUIL, A.; DURANDET, M.; BERNARD-MOULIN, P.; GUERCHE, P. .Arabidopsis Seed Content QTL Mapping Using High-Throughput Phenotyping: The Assets of Near Infrared Spectroscopy. **Front. Plant Sci**. v.7, 2016. doi: 10.3389/fpls.2016.01682
- JENSEN, J.R.. **Remote sensing of the environment – An earth resource perspective**. [reprint.] edition. The MIT Press and MIT Press, Upper Saddle River, NJ, USA, 2002.
- JI-YONGA, S.; XIAO-BOA, Z.; JIE-WENA, Z.; HOLMES, M.; KAI-LIANGA, W.; XUEA, W.; HONGA, C. Determination of total flavonoids content in fresh Ginkgo biloba leaf with different colors using near infrared spectroscopy. **Spectrochimica Acta Part A**, v. 94, p. 271-276, Aug. 2012.
- JONES, J. B. Jr. **Kjeldahl method for nitrogen determination**. 1 ed. Athens: Micro-Macro Publishing, Inc., 1991.

- JOHNSON, R. ; WICHERN, D. **Applied Multivariate Statistical Analysis**. 6 Ed . Prentice Hall, New Jersey, 2014.
- KATZ, B. Nerve, muscle, and synapse. 5. ed. Nova York: McGraw-Hill, 1966.
- KAVATI, R. **Porta-enxertos em anonáceas**. In: FERREIRA, G. et al. (Eds.) Anonaceas: Propagação e produção de mudas. Botucatu: FEPAF, 2013. p. 111-123.
- KJELDAHL, J. Neue Methode zur Bestimmung des Stickstoffs in organischen Körpern (New method for the determination of nitrogen in organic substances). **Zeitschrift für analytische Chemie**, v.22, p.366-383, 1883.
- KOHAVI, R.; GEORGE, J. H. Wrappers for feature subset selection. **Artificial Intelligence**. v. 97, p.273-324, 1997.
- KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 4. ed. São Paulo: Livraria da Física, 2006.
- KUMAR, Y.; SAHOO, G. Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA. **I.J. Information Technology and Computer Science**, v.7, p.43-49, 2012.
- LANG, C.; COSTA, F.R.C.; CAMARGO, J.L.C.; DURGANTE, F.M.; VICENTINI, A. Near Infrared Spectroscopy Facilitates Rapid Identification of Both Young and Mature Amazonian Tree Species. **PLoS ONE** 10(8):e0134521, 2015. doi:10.1371/journal.
- LAWLOR, D. W.; KONTTURI, M.; YOUNG A.T.. Photosynthesis by flag leaves of wheat in relation to protein, ribulose biphosphate carboxylase activity and nitrogen supply. **Journal of Experimental Botany**, v. 40, p. 43–52, 1989.
- LEBOEUF, M. et al. The phytochemistry of the Annonaceae. **Phytochemistry**, New York, v. 21, n. 12, p. 2783-2813, 1982.
- LE-CESSIE, S.; VAN-HOUWELINGEN, J.C. Ridge Estimators in Logistic Regression. **Applied Statistics**, v.41, p.191-201, 1992.
- LEPAGE, R.; BILLARD, L. **Exploring the limits of bootstrap**. New York: John Wiley & Sons, p.632, 1992.
- LEPINE, L.C.; OLLINGER, S.V.; OUIMETTE, A.P.; MARTIN M.E. Examining spectral reflectance features related to foliar nitrogen in forests: Implications for broad-scale nitrogen mapping. **Remote Sensing of Environment**. V. 173 p.174–186.2016.
- LIEBIG, J.. **Organic chemistry in its applications to agriculture and physiology**. Cambridge: J. Owen. 1841.

- LIMA, R.L.S.; FERREIRA, G.B.; WEBER, O.B.; CAZETTA, J.O.. Diagnose foliar da gravioleira (*Annona muricata* L.): efeito da posição de ramos e folhas. **Ciência e Agrotecnologia**. Editora da Universidade Federal de Lavras (UFLA), v. 31, n. 5, p. 1320-1325, 2007. Disponível em: <<http://hdl.handle.net/11449/30165>>.
- LOPES, S. **Bio**. 11. ed. São Paulo: Editora Saraiva, 2000. v. 1
- LOPES, J. C.; MELLO-SILVA, R. Diversity and characterization of Annonaceae from Brazil. **Rev. Bras. Frutic.**, Jaboticabal, v. 36, n. spe1, p. 125-131, 2014.
- LORENZI H. Árvores Brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Vol. 3, 1. Ed. Nova Odessa, SP: Instituto Plantarum, 384 p. 2009.
- LU, F.; CHEN, Z.; LIU, W.; SHAO, H. Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecological Engineering*, v. 95, p. 422-429, Oct. 2016.
- MAAS, P.J.; KAMER, H.M.D.; JUNIKKA, L.; MELLO-SILVA, R.D.; RAINER, H. Annonaceae from Central-eastern Brazil. **Rodriguésia**, v.52, n.80, p.65-98, 2001.
- MAATHUIS, F.J. Physiological functions of mineral macronutrients. **Curr. Opin. Plant Biol.** v.12, p.250-258, 2009.
- MAHLEIN, Anne-Katrin. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. **Plant Disease**, v.100, pp. 241-251, 2016. <https://doi.org/10.1094/PDIS-03-15-0340-FE>
- MALAVOLTA, E.; VITTI, G.C. & OLIVEIRA, S.A. **Avaliação do estado nutricional de plantas: Princípios e aplicações**. Piracicaba, Potafos, 1997. 308p.
- MANTARAS, D.; ARMENGOL, E. Machine learning from example: Inductive and Lazy methods. **Data & Knowledge Engineering**, v.25, p.99- 123, 1998.
- MANTOAN L.P.B.; FERREIRA, G.; BOARO, C.S.F. Chlorophyll a fluorescence in *Annona emarginata* (Schltdl.) H. Rainer plants subjected to water stress and after rehydration. **Scientia Horticulturae**, v.184, p.23–30, 2015.
- MARENCO, R.A.; LOPES, N.F. **Fisiologia Vegetal: fotossíntese, respiração, relações hídricas e nutrição mineral**. Viçosa: UFV, 2005. 439p.
- MARSCHNER, P. **Mineral nutrition of higher plants**. Australia: School of Agriculture, Food and Wine The University of Adelaide Australia, 3.ed., 2012, 651p.
- MAUSETH, J.D.. **Botany: an introduction to plant biology**. London: Jones and Bartlett Publishers International, 4th Edition, 2009.

- MCCULLOCH, P.; NELDER, J.A.. **Generalized Linear Models**. 2nd Edition, London: Chapman and Hall, 1989.
- MCCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of mathematical biophysics**, v. 5, n. 4, p. 115–133, dez. 1943.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005. 297 p
- MITCHELL, T.M. Machine Learning. **McGraw-Hill**, 1997.
- MODENESE-GORLA DA SILVA, S.H. et al. Redes neurais artificiais e agricultura de precisão para recomendação de adubação da cultura da goiabeira. **Revista Brasileira de Agrocomputação**, v.2, n.1, p.37-42, Jun.2004.
- NAHAR, K. **Artificial Neural Network**. COMPUSOFT, v. 1, n. 2, dez. 2012.
- NASSAU, K. **The Physics and Chemistry of Color: The Fifteen Causes of Color**. Wiley & Sons, New York, 2001.
- NASSER, T; TARIQ; R.S. Big Data Challenges. **J Comput Eng Inf Technol**, v. 4, n.3, 2015.
- NELDER, J.A.; WEDDERBURN, R.W. Generalized linear models. **Journal of the Royal Statistical Society Series A**, v.135, n.3, p.370–384, 1972.
- OSAKI, M., SHINANO, T., TADANO, T. Effect of nitrogen, phosphorus, or potassium deficiency on the accumulation of ribulose-1,5-bisphosphate carboxylase/oxygenase and chlorophyll in several field crops. **Soil Sci. Plant. Nutr.** v.39, p.417-425, 1993.
- OTTO, M. Chemometrics: statistics and computer application in analitical chemistry. Chinchester: **wiley-VCH**. 1999.
- PAILLAT, J.; DAVRIEUX, F.; THURIÈS, L.; MOUSSARD, G.; AZAMI, A.; MARION, D. Sugarcane foliar diagnosis with near-infrared spectroscopy: determination of nitrogen and carbon contents. **Proceedings of the International Society of Sugar Cane Technologists**, volume 29, pp. 38-42, 2016.
- PANFORD, J.A.; MAN, J.M. Determination of oil content of seeds by NIR: influence of fatty acid composition on wavelength selection. **Journal of the American Oil Chemists' Society**, v.67, p.473-482, 1990.
- PATIL, J. K.; KUMAR, R. Advances in image processing for detection of plant diseases. **Journal of Advanced Bioinformatics Applications and Research**, 2. v. 2, n. 2, p. 135–141, 2011.

- PÉREZ-MARÍN, D.; TORRES, I.; ENTRENAS, J.A.; VEGA,M.; SÁNCHEZ, M.T. Pre-harvest screening on-vine of spinach quality and safety using NIRS technology. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.207, p.242–250, 2019.
- PIEME, C.A.; KUMAR, S.G.; DONGMO, M.S.; MOUKETTE, B.M.; BOYOUUM, F.F.; NGOGANG, J.Y.; SAXENA, A.K. Antiproliferative activity and induction of apoptosis by *Annona muricata* (Annonaceae) extract on human cancer cells. **BMC Complementary and Alternative Medicine**, v.14, p.516-526, 2014.
- PINTO, A.C.Q.; CORDEIRO, M.C.R.; DE ANDRADE, S.R.M.; FERREIRA, F.R.; FILGUEIRAS, H.A.C.; ALVES, R.E.; KINPARA, D.J. *Annona* species. International Centre for Underutilized Crops, **University of Southampton**, Southampton, p 263, 2005.
- QIU-XIANG, Y. et al. Monitoring Rice Nitrogen Status Using Hyperspectral Reflectance and Artificial Neural Network. **Environ. Sci. Technol.**, v.41, p.6770-6775, 2007.
- RAMOS, J.P.S. Redes neurais artificiais na classificação de frutos: cenário bidimensional. **Ciênc. agrotec.**, Lavras. V.27, n.2, p.356-362, mar./abr., 2003.
- ROZANE, D.E.; NATALE, W. Calagem, adubação e nutrição mineral de Anonáceas. **Rev. Bras. Frutic.**, Jaboticabal, v. 36, n.spe1, p.166-175, 2014.
- ROSSATO, R.; PRETE, C. E.C.; CASTRO, C.; TOMM, G.O.; LEITE, R.S.; MANDARINO, J.M.G.; ARAÚJO, P.M.; CARVALHO, C. G. P. Predicting rapeseed oil content with near-infrared spectroscopy. **Pesq. agropec. bras.**, Brasília, v.48, n.12, p.1601-1604, dez. 2013. DOI: 10.1590/S0100-204X2013001200010
- ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In: **Proc. 3th Earth Resources Technology Satellite-1 Symposium**. Greenbelt, MD NASA, pp. 301– 317, 1974.
- SALISBURY, F.B.; ROSS, C.W.. **Plant physiology**. California: Wadsworth Publishing Company, 4 th Edition, 1992.
- SAVAZAKI, E.T. Efeito do Ácido Indolbutírico no enraizamento de estacas de três cultivares de atemóia (*Annona squamosa* L. x *Annona cherimola* Mill.) e formação da muda. In: FERREIRA, G. et al. (Eds.). **Anonáceas: Propagação e Produção de Mudas**. Botucatu:FEPAF, 2013, v.1, p.75-109.

- SCALOPPI JUNIOR, E.J. Propagação de espécies de Annonaceae com estacas caulinares. 2007. 92f. Tese (Doutorado) – Faculdade de Ciências Agronômicas e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2007.
- SIEMENS, B.J.; DAUN, J.K. Determination of the fatty acid composition of canola, flax, and solin by near-infrared spectroscopy. **Journal of the American Oil Chemists' Society**, v.82, p.153-157, 2005.
- SILVA, E.B.; FARNEZI, M.M.M. Limitações nutricionais para o crescimento de mudas de graviola em casa de vegetação em latossolo vermelho distrófico do norte de minas gerais. **Biosci. J.**, Uberlândia, v. 25, n. 6, p. 52-58, 2009.
- SILVA, I. N.; SPATTI, H. D.; FLAUZINO, R. A. Redes Neurais Artificiais: para engenharia e ciências aplicadas. São Paulo: **Artliber**, 2010. 399p.
- SIMS, D. A.; GAMON, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. **Remote Sensing of Environment**, v. 81, n. 2, p. 337-354, 2002.
- SKOOG, D.A.; HOLLER, F.J.; NIEMAN, T.A. **Princípios de análise instrumental**. São Paulo: Bookman, 2002.
- SMIS, A.; MURGUZUR, F. J. A.; STRUYF, E.; SOININEN, E. M.; JUSDADO, J. G. H.; MEIRE, P.; BRÅTHEN, K.A.. Determination of plant silicon content with near infrared reflectance spectroscopy. **Front. Plant Sci.** v.5, sep. 2014.
- SOBRINHO, R.B. Potencial de exploração de anonáceas no Nordeste do Brasil. In: 17º Semana internacional da fruticultura, floricultura e agroindústria [**Anais...**] Fortaleza: Instituto Frutal, 2010.
- SOUZA, I.V.B.; SÃO-JOSÉ, A.R.; SILVA, J.C.G.; BOMFIM, M.P.; JESUS, J.S. . The development of custard apple flower bud under nitrogen and potassium nutrition. **Rev.Bras.Frutic.**, Jaboticabal, v.38, n.2, e-507, 2016.
- SOUZA, S. R.; FERNANDES, M. S. Nitrogênio. In: FERNANDES, M. S. (Ed.) **Nutrição Mineral de Plantas**. Viçosa: Sociedade Brasileira de Ciência do Solo, p.215-252, 2006.
- SUÁREZ, H. M.; DOPAZO, A.G.; LÓPEZ, L.D.; ESPINOSA, F. (2015). Identification of Relevant Phytochemical Constituents for Characterization and Authentication of Tomatoes by General Linear Model Linked to Automatic Interaction Detection (GLM-AID) and Artificial Neural Network Models (ANNs). **PLoS ONE**, 10(6), e0128566. <http://doi.org/10.1371/journal.pone.0128566>

- TALLADA, J.G.; RAMOS, M.A. Visible-near-infrared absorbance spectroscopy for rapid estimation of leaf nitrogen contents of Philippine rice cultivars. **Cogent Food & Agriculture**, v.4, 2018. <https://doi.org/10.1080/23311932.2018.1487254>
- TAIZ, L.; ZEIGER, E. Fisiologia vegetal. 6. ed., Artmed, 2017. 918 p.
- TOKUNAGA, T. **A cultura da atemoia**. 2.ed. Campinas: CATI, 2005. 80 p. (Boletim técnico)
- VAN KEMPEN, T. A. T. G; JACKSON, D. NIRS may provide rapid evaluation of amino acids. **Feedstuffs**, v. 68, p.12-15, 1996.
- VIECELLI, C.A. **Guia de deficiências nutricionais em plantas**. Toledo, PR: PUCPR Câmpus Toledo / Grupo Marista, 2017.
- WANG J, SHEN C, LIU N, et al. Non-Destructive Evaluation of the Leaf Nitrogen Concentration by In-Field Visible/Near-Infrared Spectroscopy in Pear Orchards. **Sensors**, Basel, , v.17, n.3, p 538. 2017. doi:10.3390/s17030538.
- WANG, Y.; HU,X.; JIN,G.; HOU,Z.; NING, J.; ZHANG, Z. Rapid prediction of chlorophylls and carotenoids content in tea leaves under different levels of nitrogen application based on hyperspectral imaging. **J Sci Food Agric**, v.99, p.1997–2004, 2018.
- WILLIAMS, P., NORRIS, K.. Near Infrared Technology in the agricultural and Food Industries. Minnesota: **AACC**. 1990.
- WITTEN, IH.; FRANK, E. **Data mining: pratical machine learning tools and techniques**. 2nd ed. Amsterdam; Boston, MA: Morgan Kaufman, 2005.
- WITTEN, I.H.; FRANK, E.; TRIGG, L.; HALL, M.; HOLMES, G.; CUNNINGHAM, S.J. . **Weka: Practical machine learning tools and techniques with Java implementations**. (Working paper 99/11). Hamilton, New Zealand: University of Waikato, Department of Computer Science, 1999.
- XU, L.; FU, H.; CAI,C.; SHE, Y. Quality Degradation of Chinese White Lotus Seeds Caused by Dampening during Processing and Storage: Rapid and Nondestructive Discrimination Using Near-Infrared Spectroscopy. **Journal of Analytical Methods in Chemistry**. Article ID 345352, v.2015, 7 p. 2015. <http://dx.doi.org/10.1155/2015/345352>
- ZENEBON, O.; PASCUET, N.S.; TIGLEA, P. (Coord.). **Métodos físico-químicos para análise de alimentos**. 4. ed. São Paulo: Instituto Adolfo Lutz, Secretaria de Estado da Saúde, 2008. 1020 p.

ZHANG, H. *et al.* Estimation of rice neck blasts severity using spectral reflectance based on BP-neural network. **Acta Physiologiae Plantarum**, v. 33, n. 6, p. 2461–2466, 18 maio, 2011.

8. Apêndice A

Médias das variáveis de trocas gasosas (transpiração - E, condutância estomática - Gs, assimilação de carbono - A, carbono interno - ci, eficiência do uso da água - EUA e

eficiência instantânea da Rubisco - A/ci) de mudas de *Annona emarginata* com os seus erros padrão (entre parênteses) para os quatro tratamentos (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) ao longo do tempo (dia de aplicação dos tratamentos - Tempo 1, 15 dias após a aplicação dos tratamentos - Tempo 2, 30 dias após a aplicação dos tratamentos - Tempo 3 e 45 dias após a aplicação dos tratamentos - Tempo 4).

Tempo	Tratamento	E	Gs	A	ci	EUA	A/ci
1	T1	0,45400	22,2048	1,89980	300,851	0,28528	0,008358
		(0,13206)	(6,9225)	(1,03937)	(30,7447)	(0,04714)	(0,005490)
	T2	0,32185	14,3041	1,61786	231,787	0,24489	0,008580
		(0,08926)	(3,9691)	(0,62489)	(40,1233)	(0,06387)	(0,003830)
	T3	0,32078	15,3631	0,85236	317,597	0,33096	0,002827
		(0,06192)	(3,6600)	(0,23565)	(20,7497)	(0,04020)	(0,000864)
	T4	0,31290	13,9800	1,22853	251,596	0,25120	0,005251
		(0,06819)	(2,5893)	(0,12208)	(29,7053)	(0,04823)	(0,000908)
2	T1	0,39125	19,7329	2,33885	214,022	0,32102	0,015838
		(0,14706)	(7,9851)	(0,90060)	(50,2686)	(0,15988)	(0,005958)
	T2	0,43066	22,0703	1,82389	250,989	0,38669	0,010798
		(0,15771)	(8,2188)	(0,89775)	(52,4531)	(0,13706)	(0,005158)
	T3	0,27187	14,2378	1,31229	268,811	0,82075	0,006443
		(0,11159)	(5,8212)	(0,72164)	(49,8942)	(0,50411)	(0,003737)
	T4	0,23895	12,4504	0,97786	251,332	0,25386	0,003941
		(0,07529)	(4,2099)	(0,27746)	(28,6049)	(0,07363)	(0,000941)
3	T1	0,75928	46,3705	4,82145	229,759	0,15848	0,021735
		(0,11482)	(7,3957)	(0,60375)	(22,3120)	(0,01602)	(0,002990)
	T2	0,70512	43,2615	4,50288	222,254	0,17715	0,022136
		(0,17156)	(14,7980)	(1,13181)	(27,9764)	(0,03490)	(0,006107)
	T3	0,50539	30,3697	3,89628	180,150	0,12917	0,024525
		(0,10540)	(6,7544)	(0,77585)	(33,9423)	(0,01980)	(0,005209)
	T4	0,60611	34,3169	4,38494	211,952	0,23781	0,036228
		(0,18205)	(10,5262)	(2,12125)	(59,2405)	(0,08638)	(0,016033)
4	T1	0,69960	31,0499	2,91520	257,933	0,34199	0,012217
		(0,09906)	(4,7722)	(0,67951)	(22,5703)	(0,13216)	(0,003059)
	T2	1,68962	64,9574	6,45961	266,788	0,28703	0,024806
		(0,58977)	(27,2831)	(1,95196)	(12,6187)	(0,07497)	(0,007917)
	T3	1,34868	63,1510	5,52517	262,362	0,26060	0,021984
		(0,38473)	(17,5632)	(1,53935)	(12,8983)	(0,02794)	(0,006342)
	T4	0,73948	36,7108	3,03986	248,733	0,28703	0,016095
		(0,19834)	(6,6171)	(0,78977)	(34,7332)	(0,07497)	(0,007358)

9. Apêndice B

Médias das variáveis de fluorescência (energia quântica efetiva - ϕ PSII; energia quântica potencial - F_v'/F_m' ; quentin fotoquímico - qP; taxa de transporte de elétrons - ETR; dissipação de energia na forma de calor pelas antenas - D e a energia fotoquímica não dissipada e não utilizada na fase fotoquímica - Ex) de mudas de *Annona emarginata* com os seus erros padrão (entre parênteses) para os quatro tratamentos (Concentração de nitrogênio - 0 mg.L⁻¹ de N (T1); 52,5 mg.L⁻¹ de N (T2); 105 mg.L⁻¹ de N (T3) e 210 mg.L⁻¹ de N (T4)) ao longo do tempo (dia de aplicação dos tratamentos - Tempo 1, 15 dias após a aplicação dos tratamentos - Tempo 2, 30 dias após a aplicação dos tratamentos - Tempo 3 e 45 dias após a aplicação dos tratamentos - Tempo 4).

Tempo	Tratamento	ϕ PSII	F_v'/F_m'	ETR	qP	D	Ex
1	T1	0,02732 A (0,010962)	0,41579 A (0,05546)	13,760 A (5,52019)	0,05824 A (0,020641)	0,58421 A (0,05546)	0,38847 A (0,04804)
	T2	0,02894 A (0,010600)	0,37778 A (0,05625)	14,588 A (5,33936)	0,05085 A (0,025581)	0,62222 A (0,05625)	0,35835 A (0,04780)
	T3	0,01452 A (0,009511)	0,42508 A (0,05438)	7,322 A (4,79272)	0,02632 A (0,017455)	0,57492 A (0,05438)	0,41056 A (0,04649)
	T4	0,01700 A (0,003424)	0,51281 A (0,02152)	8,574 A (1,72677)	0,03302 (0,006467)	0,48719 A (0,02152)	0,49579 A (0,02044)
2	T1	0,02772 A (0,011455)	0,39868 A (0,03316)	13,976 A (5,78548)	0,06370 A (0,021937)	0,60132 A (0,03316)	0,37097 A (0,02424)
	T2	0,03766 A (0,009856)	0,44323 A (0,06548)	18,972 A (4,96466)	0,07988 A (0,017874)	0,55677 A (0,06548)	0,40557 A (0,05776)
	T3	0,01938 A (0,007785)	0,36783 A (0,10999)	9,768 A (3,91696)	0,05240 A (0,014339)	0,63217 A (0,10999)	0,34843 A (0,10356)
	T4	0,01000 A (0,004867)	0,38142 A (0,06432)	5,042 A (2,45022)	0,02056 A (0,009372)	0,61858 A (0,06432)	0,37142 A (0,06014)
3	T1	0,03950 A (0,010100)	0,42281 A (0,03775)	19,896 A (5,08864)	0,10466 A (0,034975)	0,57719 A (0,03775)	0,38333 A (0,04645)
	T2	0,03478 A (0,013692)	0,53078 A (0,02523)	17,526 A (6,89598)	0,06248 A (0,024247)	0,46922 A (0,02523)	0,49599 A (0,01906)
	T3	0,02060 A (0,006359)	0,54481 A (0,03271)	10,386 A (3,20244)	0,03714 A (0,011783)	0,45519 A (0,03271)	0,52419 A (0,03057)
	T4	0,03368 A (0,012806)	0,45987 A (0,07024)	16,980 A (6,45374)	0,06442 A (0,017934)	0,54013 A (0,07024)	0,42620 A (0,05990)
4	T1	0,04232 A (0,012597)	0,26122 A (0,04591)	21,322 A (6,34824)	0,14906 A (0,042315)	0,73878 A (0,04591)	0,21891 A (0,03920)
	T2	0,05156 A (0,019219)	0,53108 B (0,03129)	25,976 A (9,68127)	0,09932 A (0,038574)	0,46892 B (0,03129)	0,47952 B (0,03728)
	T3	0,04338 A (0,011685)	0,58807 B (0,03910)	21,862 A (5,88109)	0,07206 A (0,018301)	0,41193 B (0,03910)	0,54468 B (0,03431)
	T4	0,03174 A (0,005795)	0,50363 B (0,02731)	15,978 A (2,91366)	0,06452 A (0,013068)	0,49637 B (0,02731)	0,47190 B (0,02916)

(*) Médias seguidas de mesma letra maiúscula não diferem estatisticamente pelo teste de Tukey-Kramer (p<0,05)