UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

CAIO BRUNO WETTERICH

Imagem de fluorescência aplicada em doenças de citros

São Carlos
2016

CAIO BRUNO WETTERICH

Imagem de fluorescência aplicada em doenças de citros

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

Área de concentração: Física Básica
Orientador: Prof. Dr. Luis Gustavo Marcassa

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2016

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica revisada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Wetterich, Caio Bruno

Imagem de fluorescência aplicada em doenças de citros / Caio Bruno Wetterich; orientador Luis Gustavo Marcassa - versão corrigida -- São Carlos, 2016.

139 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Básica) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2016.

1. Doenças cítricas. 2. Agricultura de precisão. 3. Espectroscopia de fluorescência. 4. SVM. 5. Redes neurais artificiais. I. Marcassa, Luis Gustavo, orient. II. Título.

Aos meus pais e à minha noiva pelo carinho,
motivação, paciência e apoio durante o
desenvolvimento deste trabalho.

AGRADECIMENTOS

A Deus por me guiar e ajudar nos momentos de angústia e dificuldades, me revelando caminhos e oportunidades durante essa jornada.

Ao meu orientador e amigo Prof. Dr. Luis Gustavo Marcassa, pelo seu suporte e apoio durante o desenvolvimento deste trabalho. Seus conhecimentos, sinceridade, opinião e nossas conversas me ajudaram a crescer profissionalmente e como pessoa.

Aos meus pais, Carmo e Isabel, que tanto me apoiaram e incentivaram a continuar este trabalho, não me deixando abater pelas dificuldades encontradas. Pela confiança em meu sucesso. O apoio de vocês foi fundamental durante esta jornada. Amo muito vocês.

À minha noiva, namorada e amiga, Denise, que nesta trajetória me apoiou de forma incondicional em minhas escolhas com sua paciência e amor. Nossas conversas me tornaram uma pessoa melhor. Sua confiança em meu trabalho me motivou em cada dia de angústia. Sem você ao meu lado as dificuldades seriam maiores. Amo muito você. Além de seu amor, você me concedeu o maior presente que recebemos em nossas vidas, nossa princesinha Laurinha. Com você todos os problemas são esquecidos, tornando alegre cada dia ao seu lado.

Aos meus demais familiares, em especial meu irmão Cássio, pela sua sinceridade e alegria, que me ajudaram em momentos difíceis. Às minhas avós, Antônia e Rosa, madrinha Inês, tios e tias, primos e primas, pelas orações realizadas durante o doutorado.

Aos meus amigos, Paulinho, Jamile, Paulo e Tamiris, o companheirismo e alegria de vocês é fundamental para superar todas as adversidades. Ao Jorge e a Marina que me deram uma segunda família em que posso contar de maneira incondicional.

Ao pesquisador José Belasque Jr., pelo suporte durante o desenvolvimento deste trabalho. Ao Fundecitrus, na pessoa do Júlio, pela paciência e suporte durante a coleta das amostras neste projeto.

À todos os meus amigos, em especial os que estavam presentes no dia a dia: Emerson, Otuka, Moser, Nó, Renatinho, Gustavão, Chico, Anderson, Ruan, Henry, Ruben, que fizeram do estudo da física um pouco mais fácil e alegre. Agradeço pelo apoio e amizade de vocês.

Aos eternos amigos do futebolzinho, Bruno, Éder, Diogo, Murilo, Tucci e Fernando. Ao meu amigo de longa data, Sérgio, sua amizade é fundamental.

Ao Laboratório de Interações Atômicas e ao Grupo de Fotônica pelo suporte para o desenvolvimento do projeto e pelos amigos que lá fiz.

Aos membros da banca examinadora pelas contribuições e ideias acerca deste projeto.

Ao Instituto de Física de São Carlos, pela oportunidade de realização deste projeto de doutorado.

À Fundação de Amparo à Pesquisa do Estado de São Paulo, pela concessão da bolsa de doutorado e pelo apoio financeiro para o desenvolvimento desta pesquisa.

“Talvez não tenha conseguido fazer o melhor, mas lutei
para que o melhor fosse feito. Não sou o que deveria ser,
mas Graças a Deus, não sou o que era antes.”

Martin Luther King

RESUMO

WETTERICH, C. B. **Imagem de fluorescência aplicada em doenças de citros**. 2016. 139 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016.

Nos últimos anos, tem havido um crescente interesse na detecção precoce das doenças que afetam as culturas agrícolas a fim de evitar grandes perdas econômicas devido à contaminação de novas plantas. As principais doenças cítricas, cancro cítrico e greening, são uma séria ameaça à produção de citros em todo o mundo, incluindo regiões do Brasil e dos Estados Unidos. A disseminação rápida das doenças leva à redução do número de pomares cultivados, resultando em danos econômicos aos produtores e às indústrias relacionadas. O desenvolvimento de métodos para o diagnóstico precoce pode resultar em uma importante ferramenta para o controle e gestão dos citros. Algumas deficiências nutricionais como a de ferro e zinco apresentam sintomas visuais semelhantes com o greening, enquanto que o cancro cítrico pode ser confundido com a verrugose ou leprose dos citros, podendo levar ao diagnóstico incorreto. Atualmente, somente testes bioquímicos são capazes de detectar especificamente o cancro cítrico e o greening, e consequentemente diferenciá-los das demais doenças e deficiências de nutrientes. No presente trabalho, a técnica de espectroscopia por imagens de fluorescência em conjunto com os métodos de aprendizado e classificação, SVM (do inglês, *Support Vector Machine*) e ANN (do inglês, *Artificial Neural Network*), foram utilizadas a fim de identificar e discriminar as principais doenças que afetam a citricultura nos estados de São Paulo/Brasil e da Flórida/EUA. As amostras em estudo são cancro cítrico, verrugose, greening e deficiência de zinco. O objetivo principal é discriminar as doenças com sintomas visuais semelhantes, no caso, cancro cítrico de verrugose e greening de deficiência de zinco para as amostras do Brasil, e greening de deficiência de zinco para as amostras dos Estados Unidos. Os resultados mostram que é possível utilizar a técnica de espectroscopia por imagens de fluorescência em conjunto com os métodos de classificação na discriminação das doenças que apresentam sintomas visuais semelhantes. Ambos os classificadores apresentaram uma elevada precisão na classificação tanto das amostras do Brasil como dos Estados Unidos, destacando assim eficácia da técnica sob condições diferentes.

Palavras-chave: Doenças cítricas. Agricultura de precisão. Espectroscopia de fluorescência. SVM. Redes neurais artificiais.

ABSTRACT

WETTERICH, C. B. **Fluorescence image applied on citrus diseases**. 2016. 139 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016.

In recent years, there has been an increasing interest in early detection of diseases that affect agricultural crops to avoid great economic losses due to contamination of new plants. The main citrus diseases, citrus canker and HLB, are a serious threat to citrus production worldwide, including regions in Brazil and the United States. The rapid spread of the diseases leads to the reduction of cultivated orchards resulting in economic losses to producers and industries. The development of methods for early diagnosis can result in an important tool for the control and management of citrus. Some nutritional deficiencies such as iron and zinc have similar visual symptoms to HLB, while the citrus canker can be confused with citrus scab and citrus leprosis, which may lead to incorrect diagnosis. Currently, only biochemical tests are able to detect, specifically, citrus canker and HLB, and thus distinguish them from other diseases and nutrient deficiencies. In this work, the fluorescence imaging spectroscopy technique with the learning and classification methods, SVM (Support Vector Machine) and ANN (Artificial Neural Network), were used to identify and discriminate the main diseases that affect citrus production in the states of São Paulo/Brazil and Florida/USA. The samples studied are citrus canker, citrus scab, HLB and zinc deficiency. The objective is to discriminate the diseases with similar visual symptoms, such as citrus canker from citrus scab and HLB from zinc deficiency for samples from Brazil and HLB from zinc deficiency for samples from the United States. The results show that it is possible to use the fluorescence imaging spectroscopy technique together with the classification methods for the discrimination of the diseases that have similar visual symptoms. Both classifiers showed high accuracy in the classification of the samples from Brazil and the United States, highlighting the efficiency of the technique under different conditions.

Keywords: Citrus diseases. Precision agriculture. Fluorescence spectroscopy. SVM. Artificial neural network.

LISTA DE FIGURAS

Figura 1.1 - Produção de laranja no estado da Flórida (EUA) nos últimos vinte anos.	25
Figura 2.1 - Espectros de absorção molar (linha fina) e emissão (linha grossa) da fluorescência induzida por radiação ultravioleta dos principais pigmentos presentes na folha. As clorofilas absorvem na região espectral do azul e do vermelho, e fluorescem na região do vermelho e infravermelho. As clorofilas são responsáveis pela coloração verde das folhas.	34
Figura 2.2 - Estrutura química das moléculas de clorofilas <i>a</i> e <i>b</i> destacando a posição e o grupo de radicais que compõe cada uma das moléculas.	35
Figura 2.3 - Espectros de absorção dos fotossistemas FSI e FSII, destacando os picos de máxima absorção	36
Figura 2.4 - Precisão dos diferentes classificadores em função do número de características extraídas das imagens hiperespectrais.	44
Figura 3.1 - Espectro de transmissão do filtro de cristal líquido variável em função do comprimento de onda.	48
Figura 3.2 - Componentes do filtro óptico sintonizável. (1) Filtro óptico variável; (2) Cabo de 25 pinos; (3) Módulo de controle; (4) Cabo USB ou RS232; (5) Fonte de alimentação; (6) Computador com o software do fabricante.	49
Figura 3.3 - (a) Esquema da montagem do SIF-1, destacando os principais componentes ópticos: (1) câmera CCD, (2) lentes objetivas, (3) filtro de cristal líquido variável, (4) sistema de iluminação à LEDs em diferentes comprimentos de onda: 365, 405, 470 e 530 nm; (b) Filtro óptico de cristal líquido.....	50
Figura 3.4 - (a) Esquema da montagem do SIF-2 , destacando os principais componentes ópticos: (1) câmera CCD, (2) lentes objetivas, (3) roda de filtros, (4) sistema de iluminação à LEDs em 405 e 470 nm. (b) Imagem digital do sistema completo com a excitação em 470 nm ligada.....	52
Figura 3.5 - Interface do software de seleção dos comprimentos de onda do filtro variável.	54
Figura 3.6 - Interface do software de aquisição das imagens de fluorescência.....	55
Figura 3.7 - Resposta da câmera CCD em função do tempo de exposição.....	56

Figura 3.8 -	(a) Gráfico da sensibilidade espectral da câmera CCD. (b) Gráfico da curva de transmissão do filtro variável.	56
Figura 3.9 -	Comportamento da função obtida a partir do produto da equação resposta da câmera CCD pela equação da curva de transmissão do filtro.	57
Figura 3.10 -	Exemplos de imagens de fluorescência segmentadas para amostras de (a) cancro cítrico, (b) verrugose, (c) greening e (d) deficiência de zinco.	59
Figura 3.11 -	Separação de um conjunto de dados pelo hiperplano em um espaço de maior dimensão implementada através da função kernel.	63
Figura 3.12 -	Arquitetura de uma rede MLP com duas camadas ocultas.	65
Figura 3.13 -	Diferentes tipos de funções de ativação: (a) limiar, (b) rampa, (c) sigmoideal e (d) Gaussiana. Fonte: Elaborada pelo autor.	65
Figura 4.1 -	Curvas ROC representativas de três graus de capacidade de discriminação.	70
Figura 4.2 -	Curvas PR para dois algoritmos diferentes.	71
Figura 4.3 -	Imagens de fluorescência em 690 nm para amostras (a) sadia e (b) contaminada com greening do Brasil.	72
Figura 4.4 -	Curvas PR e ROC para a classificação dos dados do Brasil (a) e (b), e dos Estados Unidos (c) e (d), respectivamente.	73
Figura 5.1 -	(a) Imagem de fluorescência em 690 nm utilizando a excitação em 470 nm para uma amostra de cancro cítrico destacando o ROI selecionado. (b) Imagem resultante após a aplicação da máscara binária gerada pelo ROI na imagem original.	76
Figura 5.2 -	Gráficos da intensidade em função do comprimento de onda comparando amostras de cancro cítrico e verrugose para excitações em (a) 405 nm e (b) 470 nm, além de amostras de greening e deficiência de zinco em (c) 405 nm e (d) 470 nm.	77
Figura 5.3 -	Mapa de contorno da otimização dos parâmetros C e γ obtidos através da validação cruzada. (a) Pesquisa abrangente das melhores regiões com as maiores acurácias. (b) Ajuste fino em torno da região selecionada. A cor vermelha representa as regiões com acurácia maior que 90%.	81
Figura 5.4 -	Acurácia do classificador SVM em função do número de amostras de treinamento para os dois grupos em estudo.	86

Figura 6.1 - Gráficos da intensidade em função do comprimento de onda comparando amostras de greening e deficiência de zinco para excitações em (a) 405 nm e (b) 470 nm.....	100
Figura 6.2 - Exemplos de imagens segmentadas obtidas a partir da excitação em 470 nm e fluorescência em 690 nm para amostras de (a) greening e (b) deficiência de zinco.	102
Figura 6.3 - Mapa de contorno da otimização dos parâmetros C e γ destacando a etapa inicial da seleção das melhores regiões (a) e o ajuste fino em torno da região selecionada (b).	103
Figura A. 1 - (a) <i>Diaphorina citri</i> adulto se alimentando em uma planta hospedeira; (b) Sintomas característicos do greening; (c) Desfolha e morte de ponteiros em estágio mais avançado da doença; (d) Fruto doente com coloração verde clara, deformado, pequeno e assimétrico; (e) Frutos de ramos saudáveis; (f) Folhas de citros com sintomas de deficiência de zinco.	131
Figura A. 2 - Ciclo epidemiológico do cancro cítrico: (a) disseminação, (b) infecção, (c) colonização e (d) sobrevivência.	133
Figura A. 3 - Sintomas característicos do cancro cítrico em (a) folhas, (b) frutos e (c) ramos. Sintomas visuais de cancro cítrico e verrugose (d) em folhas.	135
Figura B. 1 - Ângulos que podem ser utilizados para o cálculo da matriz de co-ocorrência.	137
Figura B. 2 - Processo de extração da matriz de co-ocorrência em níveis de cinza a partir de uma imagem considerando $d=1$ e $\theta = 0^\circ$	138

LISTA DE TABELAS

Tabela 1.1 - Quantidade percentual de plantas contaminadas com greening segundo as regiões citrícolas do estado de São Paulo.	26
Tabela 1.2 - Quantidade percentual de talhões contaminados com cancro cítrico segundo as regiões citrícolas do estado de São Paulo.	27
Tabela 4.1 - Resultados da classificação das imagens de fluorescência pelo classificador SVM.	72
Tabela 5.1 - Bandas espectrais selecionadas para as excitações em 405 e 470 nm a partir do experimento exploratório.	78
Tabela 5.2 - Resultados da classificação do Grupo 1 segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).	82
Tabela 5.3 - Desempenho da discriminação de amostras de cancro cítrico e verrugose pelo método SVM.	83
Tabela 5.4 - Resultados da classificação do Grupo 2 segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).	84
Tabela 5.5 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método SVM.	85
Tabela 5.6 - Desempenho do classificador SVM considerando as combinações para a detecção do cancro cítrico (Grupo 1).	87
Tabela 5.7 - Desempenho do classificador SVM considerando as combinações para a detecção do greening (Grupo 2).	88
Tabela 5.8 - Resultados da classificação do Grupo 1 pelo método ANN segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).	90
Tabela 5.9 - Desempenho da discriminação de amostras de cancro cítrico e verrugose pelo método ANN.	91

Tabela 5.10 - Resultados da classificação do Grupo 2 pelo método ANN segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).....	92
Tabela 5.11 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método ANN.....	93
Tabela 5.12 - Desempenho do classificador ANN considerando as combinações para a detecção do cancro cítrico (Grupo 1).	94
Tabela 5.13 - Desempenho do classificador ANN considerando as combinações para a detecção do greening (Grupo 2).	95
Tabela 6.1 - Bandas espectrais selecionadas para as excitações em 405 e 470 nm a partir do experimento exploratório para a discriminação de amostras de greening e deficiência de zinco.	101
Tabela 6.2 - Resultados da classificação pelo método SVM segundo as taxas VP, VN, FP e FN.	104
Tabela 6.3 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método SVM.....	105
Tabela 6.4 - Desempenho do classificador SVM considerando as combinações para a detecção do greening.	106
Tabela 6.5 - Resultados da classificação das amostras de greening e deficiência de zinco pelo método ANN segundo as taxas de VP, VN, FP e FN.....	107
Tabela 6.6 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método ANN.....	108
Tabela 6.7 - Desempenho do classificador ANN considerando as combinações para a detecção do greening.	109
Tabela 6.8 - Desempenho dos classificadores para as amostras da Flórida de acordo com a combinação das bandas espectrais do Brasil.	110

LISTA DE ABREVIATURAS E SIGLAS

ANN	Artificial Neural Network
ART	Adaptive Resonance Theory
BDT	Bagged Decision Tree
BF	Blue Fluorescence
BGF	Blue-Green Fluorescence
BP-ANN	Backpropagation ANN
CART	Classification and Regression Trees
CCD	Charge Coupled Device
CDA	Coordenadoria de Defesa Agropecuária
ChlF	Chlorophyll Fluorescence
CREC	Citrus Research and Education Center
CVC	Clorose Variegada dos Citros
DNA	Deoxyribonucleic Acid
EUA	Estados Unidos da América
FAD	Flavina Adenina Dinucleotídeo
FN	Falsos Negativos
FP	Falsos Positivos
FRF	Far Red Fluorescence
FSI	Fotossistema I
FSII	Fotossistema II
Fundecitrus	Fundo de Defesa da Citricultura
GF	Green Fluorescence
GLCM	Gray Level Co-occurrence Matrix
HLB	Huanglongbing
LDA	Linear Discriminant Analysis
LEDs	Light Emitting Diodes
LIA	Laboratório de Interações Atômicas
LIFS	Laser-Induced Fluorescence Spectroscopy
K-nn	K-nearest neighbor
MAPA	Ministério da Agricultura, Pecuária, e Abastecimento
MIR	Mid-infrared
MLP	Multilayer Perceptron

MRF	Markov Random Field
MSE	Mean Square Error
MTMF	Mixture Tuned Matched Filtering
NADPH	Nicotinamida Adenina Dinucleotídeo Fosfato, forma reduzida
NIR	Near-infrared
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PLS	Partial Least Square
PLSR	Partial Least Square Regression
PR	Precision Recall
QDA	Quadratic Discriminant Analysis
qPCR	quantitative Polymerase Chain Reaction
qrt-PCR	quantitative real-time PCR
RBF	Radial Basis Function
RF	Red Fluorescence
RFA	Radiação fotossinteticamente ativa
RGB	Red, Green, Blue
ROC	Receiver Operating Characteristics
ROI	Region of Interest
SAM	Spectral Angle Mapping
SOM	Self-Organizing Maps
SIF-1	Sistema de Imagens de Fluorescência 1
SIF-2	Sistema de Imagens de Fluorescência 2
SLP	Single-Layer Perceptron
SRM	Structural Risk Minimization
SVM	Support Vector Machine
TVM	Tobacco Mosaic Virus
USDA	United States Department of Agriculture
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos

SUMÁRIO

1	INTRODUÇÃO	23
2	REVISÃO DA LITERATURA	31
2.1	Espectroscopia de fluorescência.....	32
2.1.1	Técnicas e aplicações na agricultura	38
2.2	Principais aplicações dos classificadores SVM e ANN	43
3	INSTRUMENTAÇÃO E METODOLOGIA	47
3.1	Sistemas de espectroscopia por imagens de fluorescência.....	47
3.1.1	Sistema de imagens de fluorescência utilizando filtro óptico de cristal líquido variável (SIF-1)	47
3.1.2	Sistema de imagens de fluorescência utilizando filtros ópticos passa-banda associados a rodas de filtros (SIF-2)	51
3.2	Metodologia.....	52
3.2.1	Amostras de citros	52
3.2.2	Aquisição e pré-processamento das imagens	54
3.2.3	Análise das imagens de fluorescência	58
3.2.3.1	Segmentação das imagens	59
3.2.3.2	Extração das características.....	60
3.2.3.3	Classificação.....	61
4	ESTUDO PRELIMINAR ENVOLVENDO O USO DA TÉCNICA SVM	67
4.1	Contextualização da pesquisa.....	67
4.2	Procedimentos experimentais.....	68
4.3	Resultados e discussões	71
5	DISCRIMINAÇÃO DAS PRINCIPAIS DOENÇAS/ESTRESSE QUE AFETAM A CITRICULTURA NO BRASIL.....	75
5.1	Experimento exploratório utilizando o SIF-1	75
5.2	Experimento utilizando o SIF-2	78
5.2.1	Resultados da classificação das imagens pelo SVM	80
5.2.2	Resultados da classificação das imagens pelo ANN	89
5.2.3	Discussão comparativa dos resultados dos classificadores	96
6	DISCRIMINAÇÃO DAS PRINCIPAIS DOENÇAS/ESTRESSE QUE AFETAM A CITRICULTURA NA FLÓRIDA/EUA.....	99
6.1	Experimento exploratório utilizando o SIF-1	99
6.2	Experimento utilizando o SIF-2	101
6.2.1	Resultados da classificação das imagens pelo SVM	102
6.2.2	Resultados da classificação das imagens pelo ANN	107
6.2.3	Análise do desempenho dos classificadores baseada nas combinações das bandas espectrais do Brasil.....	110
6.2.4	Discussão comparativa dos resultados dos classificadores	111
7	CONCLUSÕES	115

REFERÊNCIAS	119
APÊNDICE A – Principais doenças que afetam a citricultura	129
APÊNDICE B – Exemplo da obtenção da matriz de co-ocorrência GLCM	137

1 INTRODUÇÃO

O agronegócio é atualmente o principal setor da economia brasileira. O Brasil se destaca como um dos grandes países exportadores de *commodities* agrícolas e insumos do agronegócio. Em 2013, o valor financeiro referente ao comércio agrícola mundial foi de US\$ 1,14 trilhões, no qual o Brasil obteve participação recorde de US\$ 99,97 bilhões na exportação de produtos no setor, cifra 4,3% superior à alcançada em 2012 (US\$ 95,81 bilhões), segundo dados divulgados pelo Ministério da Agricultura, Pecuária, e Abastecimento (MAPA) em 2014.(2) No ano de 2015, o valor financeiro exportado pelo agronegócio apresentou uma queda de 8,8% em relação a 2014, totalizando cerca de US\$ 88,22 bilhões, contra uma importação que caiu 21,3% atingindo aproximadamente US\$ 13,07 bilhões, produzindo assim um superávit comercial de US\$ 75,15 bilhões.(3) Apesar da queda apresentada, a participação do agronegócio na balança comercial brasileira é de fundamental importância para a estabilidade econômica do país, uma vez que sem esse setor a balança comercial passaria de um superávit de US\$19,69 bilhões para um déficit de US\$ 55,46 bilhões.(3)

As exportações agrícolas do Brasil desempenham um papel importante nos mercados internacionais. O Brasil é o segundo maior exportador agrícola mundial e o maior fornecedor de açúcar, suco de laranja e café.(4) Em 2013, tornou-se maior fornecedor de soja, assim como é um importante exportador de carnes bovinas e de frango.(4, 5) Além disso, o Brasil é considerado um dos maiores produtores mundiais de frutas.(4) Nesse contexto, destaca-se a citricultura, em especial, a produção de laranjas que fornece a fruta para o consumo *in natura* e para as indústrias produtoras de suco. A citricultura brasileira é responsável por cerca de 60% da produção mundial de suco de laranja com aproximadamente um milhão de toneladas, assim como detém participação de 85% na exportação do produto, seguido pelos Estados Unidos da América (EUA), com pouco mais de 100 mil toneladas de suco.(5-7) A maior parte das exportações é absorvida por apenas três mercados: Estados Unidos, União Europeia e Canadá.(6)

No cenário nacional, a região sudeste se destaca como o maior produtor de laranja, respondendo por aproximadamente 80% do total produzido no país.(8) O último levantamento da safra de laranja 2015/16 do cinturão citrícola de São Paulo e Triângulo/Sudoeste Mineiro realizado pelo Fundo de Defesa da Citricultura (Fundecitrus) indica uma safra de 286,14 milhões de caixas, de 40,8 kg, um aumento de 2,6% em relação à primeira estimativa realizada no início de 2015.(9) No entanto, houve uma redução na área de cultivo de citros no

Brasil de cerca de 1% em relação ao ano de 2014, segundo levantamento realizado pelo Departamento de Agricultura dos Estados Unidos (USDA, do inglês *United States Department of Agriculture*) em dezembro de 2015.(10) Esta diminuição é consequência, principalmente, da erradicação das árvores afetadas pelo greening, cujo inventário da USDA demonstrou também a diminuição de 1% na projeção do número de árvores em relação ao ano anterior.(10) Vale destacar o aumento da produção diante da redução da área plantada, conforme levantamento do Fundecitrus. Isto se deve aos constantes investimentos em pesquisa e tecnologia para o desenvolvimento de sistemas mais eficientes, associado a um melhor aproveitamento do espaço, consumo de água e utilização de insumos agrícolas no controle de pragas e doenças, assim como o ajuste do nível de nutrientes das plantas.(11) Além disto, as condições climáticas favoráveis também contribuíram para o aumento da produtividade, uma vez que o tamanho do fruto foi maior do que o previsto devido à precipitação constante durante o período de colheita.(9, 10)

Em relação à produção de suco de laranja, os estados da Flórida (EUA) e de São Paulo (Brasil) juntos possuem 81% da produção mundial do produto, sendo apenas o estado de São Paulo responsável pela produção de 53% deste total.(5) Entretanto, na última década, problemas com mão-de-obra, comercialização e, principalmente, doenças e pragas têm ameaçado a sustentabilidade deste mercado.(11) A USDA prevê uma queda de 25% na produção de suco de laranja nos Estados Unidos como resultado dos baixos rendimentos do estado da Flórida, o qual é responsável por 60% da produção de laranjas dos Estados Unidos.(12) A estimativa para a safra 2015/16 da Flórida, divulgada pela USDA, é de 74 milhões de caixas, representando a menor safra de laranjas em mais de 50 anos, caso a previsão se confirmar.(13) A disseminação do greening em larga escala pelos pomares da Flórida é o principal responsável pela queda na produção, doença que atinge mais 70% dos pomares do estado, ao mesmo tempo em que as áreas de cultivo permanecem em níveis reduzidos.(12-13). A Figura 1.1 apresenta a produção de laranjas no estado da Flórida nos últimos vinte anos, destacando a queda significativa no setor atualmente.

Figura 1.1 - Produção de laranja no estado da Flórida (EUA) nos últimos vinte anos.
Fonte: FLÓRIDA...(13)

Atualmente, os principais fatores que limitam os lucros e a expansão da citricultura são as restrições comerciais na forma de barreiras ao comércio e, principalmente, problemas fitossanitários.(14) As pragas e doenças foram responsáveis pela erradicação de 40 milhões de árvores nesta última década, com perdas de quase 80 milhões de caixas por ano. A taxa de mortalidade das árvores cítricas saltou de 4% para 7,5%, valor considerado alarmante pelos produtores.(5) Neste contexto, uma das preocupações mais sérias do setor são os prejuízos causados pelas doenças cancro cítrico e greening, conhecida também como *Huanglongbing* (HLB). O monitoramento constante destas é fundamental para evitar a redução da produção de frutos cítricos em pomares comerciais e não-comerciais, assim como a necessidade de erradicação das plantações.(14)

O greening é a mais destrutiva e recente doença que afeta os citros no Brasil. O agente causal do greening é uma bactéria conhecida como *Candidatus Liberibacter spp.*, que habita e se desenvolve no floema da planta hospedeira.(15) A doença surgiu na Ásia há mais de cem anos e foi identificada pela primeira vez nos dois maiores produtores mundiais de suco de laranja, São Paulo (Brasil) e Flórida (Estados Unidos), em 2004 e 2005, respectivamente.(16) No Brasil, as primeiras plantas com sintomas do greening foram encontradas na região de Araraquara, São Paulo.(15) Hoje, está presente em todas as regiões cítricas de São Paulo, assim como em pomares de Minas Gerais e Paraná.(17)

De 2004 a 2012, a disseminação do greening foi bastante rápida e atingiu todas as regiões produtoras de citros, passando de 3,4% para 64,1% dos talhões contaminados. A

incidência de plantas com sintomas saltou de 0,58%, em 2008, para 6,91%, em 2012.(18) A Tabela 1.1 apresenta a quantidade percentual de plantas com sintomas nas regiões citrícolas. A doença apresenta maiores incidências nas regiões Centro e Leste, 9,9% e 14,8%, respectivamente.(18) Embora com incidências menores, as regiões Norte (1,8%) e Oeste (1,4%) apresentam taxas de avanço da doença semelhantes as das regiões Centro e Leste.(18) Por outro lado, as regiões Sul e Noroeste têm incidências (respectivamente 0,9% e 0,3%) e taxas de crescimento menores que as demais regiões.(18) Um levantamento, realizado pelo Fundecitrus em 2015, apontou que 17,89% das plantas do parque citrícola de São Paulo e das regiões do Triângulo e Sudoeste de Minas Gerais apresentam a doença. Na comparação com 2012, quando foi realizado o último levantamento, a incidência de greening teve um aumento de 159%.(19)

Tabela 1.1 - Quantidade percentual de plantas contaminadas com greening segundo as regiões citrícolas do estado de São Paulo.

	CENTRO	NORTE	NOROESTE	OESTE	LESTE	SUL
2008	1,12	0,00	0,01	0,01	0,73	0,07
2009	1,36	0,04	0,00	0,06	1,70	0,01
2010	3,51	0,39	0,01	0,34	2,76	0,29
2011	6,08	0,81	0,17	0,68	7,15	0,84
2012	9,89	1,78	0,28	1,35	14,8	0,85

Fonte: LEVANTAMENTOS...(18)

Em relação ao cancro cítrico, é uma doença que está presente nas principais regiões produtoras de citros do mundo, incluindo os estados de São Paulo e Flórida, e assim como o greening, é causada por uma bactéria denominada *Xanthomonas citri* subsp. *citri*, de origem asiática, que pode afetar todas as espécies e variedades de citros de importância comercial.(1, 20) O cancro cítrico foi identificado pela primeira vez na região de Presidente Prudente, São Paulo, em 1957, atingindo, atualmente, grande parte do estado paulista e outros estados como Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná e Rio Grande do Sul.(20)

Os impactos diretos do cancro cítrico para a citricultura estão relacionados aos prejuízos à produção, devido à depreciação da qualidade dos frutos, restrição da comercialização e, principalmente, redução pela queda dos frutos afetados pela doença.(1) Os prejuízos vêm aumentando desde que a legislação do Programa de Erradicação do Cancro Cítrico foi amenizada, em meados de 2009, excluindo a exigência de erradicação de todo o talhão de plantas em situações que o índice de contaminação fosse igual ou maior que 0,5%,

mantendo-se somente o critério de eliminação das plantas sintomáticas e assintomáticas num raio de 30 metros em torno do foco.(20-21)

Entre 2001 e 2009, a doença manteve baixas incidências de talhões contaminados, variando entre 0,08% e 0,20%, devido às medidas definidas pelo Programa de Erradicação do Cancro Cítrico. A partir de 2010, com as mudanças da lei, o cancro começou a avançar e chegou a 0,44% de talhões contaminados, e em 2012 atingiu 1,39%.(21) A Tabela 1.2 apresenta a quantidade percentual de talhões contaminados com cancro cítrico nas regiões citrícolas do estado de São Paulo no período de 2008 a 2012. Pode-se notar que as maiores incidências estão nas regiões Noroeste e Oeste do estado, onde foi primeiro detectado e nas quais os citricultores foram mais resistentes à aplicação do Programa de Erradicação do Cancro Cítrico.(21) Nessas duas regiões, as altas incidências de talhões contaminados em 2012 são bastante preocupantes (11,5% na região Noroeste e 2,7% na Oeste). Além destas, outra região em que a doença cresceu bastante foi o Centro (0,8%), atingindo também a maior incidência registrada.(21)

Tabela 1.2 - Quantidade percentual de talhões contaminados com cancro cítrico segundo as regiões citrícolas do estado de São Paulo.

	CENTRO	NORTE	NOROESTE	OESTE	LESTE	SUL
2008	0,16	0,06	0,49	1,59	0,00	0,00
2009	0,11	0,00	0,87	0,30	0,00	0,39
2010	0,41	0,23	2,55	0,22	0,06	0,10
2011	0,06	0,51	7,07	2,21	0,00	0,81
2012	0,77	0,11	11,5	2,70	0,00	0,17

Fonte: LEVANTAMENTOS...(21)

O cenário já era alarmante para as regiões citrícolas em 2012, conforme os dados do levantamento divulgado pelo Fundecitrus, quando em 2013 houve uma nova alteração na lei e aboliu-se também a regra dos 30 metros, substituindo-a pela eliminação apenas da planta sintomática e pulverização periódica daquelas assintomáticas situadas dentro do raio de 30 metros, com calda cúprica na concentração de 0,1% de cobre metálico, repetindo a pulverização a cada brotação.(20,22) O resultado foi catastrófico: a incidência de cancro cítrico na região citrícola saltou de 1,39% em 2012 para cerca de 5,0% em 2015, um aumento de 260% em três anos.(19)

O processo de identificação das plantas contaminadas com cancro cítrico e greening envolve a inspeção visual constante dos pomares, seguido do diagnóstico laboratorial utilizando técnicas de detecção.(23-24) Um protocolo extenso de atividades é realizado antes

do diagnóstico final. Entretanto, trata-se de um processo demorado e com custo elevado, tornando-se inviável a aplicação em toda a plantação.

Isso posto, é possível observar a gravidade da situação da citricultura brasileira, bem como a ineficiência dos programas de controle e erradicação e a necessidade do desenvolvimento de técnicas de detecção rápida e eficiente de tais doenças. Neste contexto, os recentes avanços tecnológicos têm propiciado o aumento das aplicações envolvendo a interação radiação-matéria no estudo das alterações sofridas pelo metabolismo das plantas devido às doenças ou estresses bióticos.(24) Tais alterações podem ser detectadas através da aplicação de técnicas de espectroscopia adequadas.(24) Dentre essas técnicas, destaca-se a espectroscopia por imagens de fluorescência, que vêm sendo investigada como ferramenta no estudo das plantas (23,25-27), uma vez que fornece informação espectral e espacial das amostras. O aspecto mais importante da espectroscopia de fluorescência é que se trata de uma técnica não-destrutiva e não-invasiva para a fisiologia da planta.(28) Além disso, é uma técnica que pode ser utilizada tanto laboratório como fora deste, onde as condições são menos controladas.(29)

O objetivo desta pesquisa é aplicar a técnica de espectroscopia por imagens de fluorescência para diferenciar doenças da cultura cítrica utilizando os sistemas de imagens desenvolvidos pelo nosso grupo. Os sistemas desenvolvidos foram aplicados em dois estados, São Paulo (Brasil) e Flórida (Estados Unidos), a fim de discriminar as doenças com sintomas visuais semelhantes, no caso, cancro cítrico de verrugose e greening de deficiência de zinco para as amostras do Brasil, e greening de deficiência de zinco para as amostras dos Estados Unidos. A prévia discriminação das doenças utilizando estes sistemas poderia diminuir o tempo e custo do processo de diagnóstico, uma vez que somente amostras corretas estariam sendo submetidas aos testes laboratoriais. O outro objetivo da pesquisa é avaliar a viabilidade da aplicação dos sistemas desenvolvidos sob condições ambientais distintas sem a necessidade de considerá-las em nossas análises.

Assim, no capítulo 2 apresentamos uma revisão sobre a aplicação das técnicas espectroscópicas no diagnóstico de doenças em plantas e, em especial, para os citros. Destacaremos a teoria envolvida no processo de emissão da fluorescência pelas amostras, bem como alguns experimentos envolvendo a aplicação da técnica de espectroscopia por imagens de fluorescência. Também serão apresentados as principais aplicações e resultados dos dois métodos de classificação utilizados neste trabalho, SVM (do inglês, *Support Vector Machine*) e ANN (do inglês, *Artificial Neural Network*).

No capítulo 3 será tratada a instrumentação utilizada neste trabalho, destacando os dois sistemas de imagens de fluorescência desenvolvidos e suas principais características. Inicialmente será descrito o sistema de imagens que utiliza um filtro óptico de cristal líquido variável para a seleção das principais bandas de emissão das amostras em estudo. Em seguida, será descrito o sistema que utiliza filtros ópticos passa-banda para a coleta de imagens de fluorescência de uma grande quantidade de amostras. Por fim, serão descritos os procedimentos experimentais deste trabalho, destacando a coleta das amostras, a aquisição e o pré-processamento das imagens de fluorescência, assim como a análise destas imagens. Na seção, análise de imagens, apresentaremos os processos de segmentação, extração de características e classificação das imagens, discutindo, por fim, os principais conceitos das técnicas SVM e ANN.

No capítulo 4 será apresentado um estudo preliminar envolvendo a aplicação da técnica SVM pelo nosso grupo na diferenciação de amostras sadias e contaminadas com greening da Flórida/EUA. Iremos apresentar os resultados envolvidos, assim como a contribuição desse estudo no desenvolvimento deste trabalho.

No capítulo 5 serão apresentados os resultados obtidos na discriminação das principais doenças que afetam a citricultura no Brasil. Serão discutidos os resultados dos dois classificadores, destacando o potencial de cada um na classificação das imagens de fluorescência.

No capítulo 6 apresentaremos os resultados envolvidos na discriminação das amostras contaminadas com greening e com deficiência de zinco do estado da Flórida/EUA, destacando a viabilidade da aplicação da técnica de imagens por fluorescência em conjunto com esses dois classificadores em regiões com características distintas.

Por fim, no capítulo 7 serão apresentadas as conclusões obtidas neste trabalho.

2 REVISÃO DA LITERATURA

Este capítulo é dedicado à revisão dos principais métodos e experimentos que envolvem a aplicação de técnicas de espectroscopia e computacionais no diagnóstico de doenças em plantas e, em especial, em citros. Ao longo da última década, diversos sistemas de sensoriamento vêm sendo desenvolvidos desde o surgimento da agricultura de precisão, especialmente para culturas de citros no estado da Flórida.(30) Estes sistemas abrangem a identificação da fruta e da cultura a fim de monitorar a área cultivada, considerado como primeiro passo na implementação da agricultura de precisão (30), bem como a detecção de doenças através de diferentes técnicas óticas.(30)

É importante ressaltar que tais técnicas óticas não têm como objetivo substituir técnicas bioquímicas para confirmação das doenças que afetam os citros. A confirmação/diagnóstico das doenças usualmente é realizada através das seguintes técnicas: corrida do fluxo bacteriano, isolamento e reprodução do patógeno em meio de cultura apropriado, patogenicidade, sorologia e PCR (do inglês, *Polymerase Chain Reaction*) que consiste na amplificação do DNA (do inglês, *Deoxyribonucleic Acid*) da bactéria causadora da doença e comparação com um banco de dados a fim de determinar o agente causal.(14,20, 24,31-32)

No entanto, as técnicas bioquímicas não são ferramentas viáveis de triagem preliminar no processamento de uma grande quantidade de amostras devido ao tempo e alto custo envolvido para sua execução. Uma planta normal de citros tem em média cerca de 20.000 folhas ou mais, e somente algumas folhas (menos que 10) vão expressar algum sintoma visível. Apesar das técnicas bioquímicas apresentarem uma alta eficácia, o processo como um todo não é tão eficiente, já que depende de diversos parâmetros, tais como a capacidade humana em identificar corretamente as plantas contaminadas, o grau de severidade da doença, condições de campo, etc.(33-34)

Neste contexto, uma das preocupações mais sérias do setor são os prejuízos causados pelas doenças, cancro cítrico e greening. Atrasos na confirmação e diagnóstico favorecem a proliferação das doenças devido à espera para a erradicação das plantas sintomáticas. Essas técnicas bioquímicas podem ser utilizadas como uma ferramenta robusta a fim de garantir a presença das doenças, mas não com uma ferramenta de triagem preliminar.(24) Em vista disto, existe uma demanda pelo desenvolvimento de métodos que apresentem maior rapidez, sensibilidade e seletividade no diagnóstico das doenças.(29) Maiores detalhes sobre estas duas principais doenças que afetam a citricultura são apresentadas no Apêndice A.

Os recentes desenvolvimentos tecnológicos na área agrícola têm levado à busca por métodos não destrutivos automatizados para a detecção de doenças em plantas. É desejável que a ferramenta de detecção seja rápida, específica para uma determinada doença e sensível para a detecção nos estágios iniciais de desenvolvimento da doença.(24) O progresso na área de baixo custo através do uso de satélites de baixa altitude e sistemas aéreos não tripulados tem proporcionado à possibilidade de monitoramento contínuo das plantas.(30) Esses sistemas são constituídos por um sensor e uma plataforma para sua fixação e transporte, como por exemplo, veículos aéreos ou terrestres. Contudo, antes da aplicação dessas novas tecnologias, foi necessário desenvolver técnicas que detectassem os primeiros sinais de estresse nas produções agrícolas.(35) Dentre essas técnicas, destaca-se que a espectroscopia que vêm sendo amplamente utilizada pelos pesquisadores para o monitoramento das condições das plantas.(36) Alterações nas características espectrais das plantas nas regiões do visível, do infravermelho próximo (NIR, do inglês *near-infrared*) e do infravermelho médio (MIR, do inglês *mid-infrared*) do espectro eletromagnético são a base dos sensores mais comumente utilizados na detecção de doenças e estresses em plantas.(28,37-40)

Assim, as técnicas espectroscópicas são alternativas promissoras no diagnóstico de doenças em citros, uma vez que apresentam elevado potencial para a detecção rápida e eficiente de alterações no tecido biológico das plantas.(29) Há duas principais técnicas experimentais, espectroscopia à fibra que apresenta resolução espectral e espectroscopia por imagens, o qual apresenta tanto resolução espectral como espacial.(24,28,41-43) Em ambos os casos, as técnicas estudadas se baseiam em duas propriedades ópticas, refletância e fluorescência. A seguir, serão abordados os detalhes da espectroscopia por fluorescência, base do desenvolvimento deste trabalho, na citricultura.

2.1 Espectroscopia de fluorescência

A técnica de espectroscopia por fluorescência vem apresentando um grande número de aplicações no mais diversos ramos do conhecimento. Na citricultura, especificamente, permite analisar a fluorescência emitida pelos pigmentos constituintes das folhas e estudar, indiretamente, os estados fotossintéticos das plantas e consequentemente identificar alterações causadas por doenças, deficiências de minerais e estresses por água ou temperatura.(29, 42, 44)

Em geral, as plantas absorvem a radiação luminosa na região visível do espectro eletromagnético, mais precisamente na faixa entre 400 e 750 nm. Os principais pigmentos

responsáveis pela absorção da radiação são as moléculas de clorofilas *a* e *b* e os carotenóides.(45) Além da absorção da radiação incidente, os pigmentos refletem a luz não absorvida em cores diferentes. No caso das clorofilas, por exemplo, estas apresentam máxima absorção na região espectral do azul e do vermelho e refletem na região do verde.(44)

A fluorescência é um fenômeno óptico que ocorre após a absorção dos fótons incidentes pelas moléculas das amostras. Os fótons são absorvidos pela estrutura eletrônica da molécula, no estado fundamental, fazendo com que os elétrons da molécula transicionem para estados quânticos excitados e menos estáveis, permanecendo por um curto intervalo de tempo.(45) A molécula dissipa a energia acumulada por meio de processos de decaimento radiativo e/ou não radiativo para que ocorra a transição para um estado mais estável. Os processos de decaimento não radiativos envolvem a liberação da energia em excesso por meio de vibrações e rotações das moléculas ou perda de energia através de colisões com moléculas vizinhas.(45)

Em relação aos processos de decaimento radiativo, estes envolvem a emissão de radiação pela matéria, e são divididos em duas categorias: fluorescência e fosforescência, conforme a transição entre os estados quânticos. Nessa forma de decaimento, a emissão de fótons pode depender dos estados quânticos singletos ou dos estados quânticos singleto e tripleto da molécula. A transição ao estado fundamental a partir de um estado excitado singleto é denominada fluorescência. Trata-se de uma transição permitida por dipolo elétrico, uma vez que não ocorre variação do spin total ($\Delta S = 0$), satisfazendo assim uma das seguintes regras de seleção que determinam as transições permitidas por dipolo elétrico: $\Delta l = \pm 1$, $\Delta m = 0, \pm 1$ e $\Delta S = 0$. Além da transição para o estado fundamental, as moléculas no estado excitado singleto podem também fazer uma conversão para o primeiro estado tripleto, ocorrendo assim variação no spin total da molécula ($\Delta S = 1$). Essa conversão é denominada de cruzamento intersistema. A transição a partir deste estado para o estado fundamental é proibido por dipolo segundo as regras de seleção, mas pode ocorrer devido às interações do tipo quadrupolo elétrico. O tempo de vida dos dois processos é a característica fundamental para sua distinção, uma vez que a fluorescência evanesce em nanossegundos e a fosforescência pode durar até alguns segundos.(45)

A fluorescência também já foi estudada em outras moléculas além das moléculas de clorofilas *a* e *b*, e carotenóides. Cerovic e colaboradores (44) estudaram características como absorção e emissão de fluorescência de diversas moléculas importantes presentes na estrutura das folhas, com destaque para o ácido ferúlico, NADPH (Nicotinamida Adenina Dinucleotídeo Fosfato, forma reduzida), FAD (Flavina Adenina Dinucleotídeo) e o rubisco.

Estas biomoléculas também são responsáveis pela absorção da luz do ambiente, bem como exercem outras funções bioquímicas.(44) A Figura 2.1 apresenta um conjunto de gráficos adaptados da referência (44) que contém os espectros de absorção molar (linha fina) e fluorescência (linha grossa) induzida por radiação ultravioleta (UV-350 nm) destes principais pigmentos, incluindo as moléculas de clorofilas *a* e *b*. Os espectros de absorção e emissão de fluorescência foram obtidos para as moléculas de clorofilas e ácido ferúlico em solução de metanol, enquanto que NADPH, FAD e o rubisco estavam em solução aquosa com pH 7,9.

Figura 2.1 - Espectros de absorção molar (linha fina) e emissão (linha grossa) da fluorescência induzida por radiação ultravioleta dos principais pigmentos presentes na folha. As clorofilas absorvem na região espectral do azul e do vermelho, e fluorescem na região do vermelho e infravermelho. As clorofilas são responsáveis pela coloração verde das folhas.
Fonte: CEROVIC (44).

A partir da Figura 2.1 nota-se que a radiação UV é a principal fonte de radiação para induzir a fluorescência do ácido ferúlico, NADPH, FAD e do rubisco. Estes pigmentos fluorescem na região entre 400 e 550 nm. Vale destacar também a absorção das moléculas de clorofilas na região do azul e do vermelho, bem como se estende significativamente na região

do UV, principalmente, para a clorofila *a*. As moléculas de clorofilas fluorescem na região do vermelho e do infravermelho.(44)

As clorofilas são os pigmentos mais estudados através da técnica espectroscopia de fluorescência, pois fornecem a melhor informação bioquímica da planta uma vez que estão relacionadas diretamente com a fotossíntese.(45) A molécula de clorofila é formada por complexos derivados da porfirina contendo no seu centro um íon de Mg. Além dos quatros anéis de pirrol provenientes da porfirina, a molécula de clorofila possui um quinto anel adicional formado por cadeias hidrofóbicas. Na molécula de clorofila *a* ($C_{55}H_{72}N_4O_5Mg$), o anel de porfirina contém um grupo metil ($-CH_3$) no C-3, enquanto que a clorofila *b* ($C_{55}H_{70}N_4O_6Mg$) contém um grupo aldeído ($-CHO$) nessa posição.(46-47) A estrutura química das moléculas de clorofila está representada na Figura 2.2.

Figura 2.2 - Estrutura química das moléculas de clorofilas *a* e *b* destacando a posição e o grupo de radicais que compõe cada uma das moléculas.
Fonte: STREIT (46).

Os pigmentos presentes nas folhas estão localizados no interior dos cloroplastos, especificamente, nos tilacóides e lamelas (47), e são divididos em dois grupos segundo a sua funcionalidade, pigmentos acessórios e pigmento principal. Os pigmentos acessórios, sendo compostos principalmente pela clorofila *b* e os carotenóides, auxiliam na absorção da radiação luminosa e na transferência da energia radiante para o pigmento principal.(47) O pigmento principal é a clorofila *a*; ela também absorve luz do ambiente, mas sua principal

função é capturar a energia dos pigmentos acessórios e entregá-la aos centros de reação para ocorrência da fotossíntese.(45)

As plantas possuem dois centros de reação: o fotossistema I (FSI) e o fotossistema II (FSII), que interagem entre si através de transportadores de elétrons.(46) A radiação fotossinteticamente ativa (RFA) compreendida entre 390 e 760 nm (46) é captada pelos pigmentos que juntamente com proteínas estruturais constituem o complexo antena, direcionando a energia para os centros de reação em um processo conhecido como “esquema Z”.(48) O centro de reação do FSI apresenta pico máximo de absorção em 700 nm, enquanto que o FSII possui máxima absorção em 680 nm (48). A Figura 2.3 apresenta os espectros de absorção dos dois fotossistemas.(49)

Figura 2.3 - Espectros de absorção dos fotossistemas FSI e FSII, destacando os picos de máxima absorção. Fonte: DAVID (49).

Os fotossistemas oxidam a água a oxigênio molecular para formar ATP e reduzir o NADP^+ em NADPH. Desta forma, a energia absorvida é armazenada na forma de ATP e na forma de coenzimas orgânicas. Posteriormente, esses substratos serão utilizados no ciclo de Calvin e Benson para produção de açúcares e/ou cadeias de carbono para biossíntese.(48) Maiores detalhes sobre os processos da fotossíntese podem ser encontrados na literatura.(47,49)

A partir do estudo da interação da radiação com as folhas e as respostas emitidas por estas, alguns parâmetros importantes foram definidos. A excitação ultravioleta das folhas

verdes induz duas respostas distintas de fluorescência: a BGF (do inglês, *Blue-Green Fluorescence*) na região entre 400-630 nm e a fluorescência da clorofila *a*, denominada na literatura de ChlF (do inglês, *Chlorophyll Fluorescence*), na região entre 630-800 nm.(44) As intensidades relativas destes dois tipos de fluorescência são altamente sensíveis às propriedades intrínsecas das folhas e aos fatores ambientais. Os espectros de emissão de fluorescência podem ser considerados como uma assinatura revelando o estado fisiológico da planta.(44)

Segundo Cerovic e colaboradores (44), as duas regiões mencionadas acima podem ainda ser divididas em outras quatro bandas de fluorescência: fluorescência no azul (BF, do inglês *Blue Fluorescence*), fluorescência no verde (GF, do inglês *Green Fluorescence*), fluorescência no vermelho (RF, do inglês *Red Fluorescence*) e fluorescência no vermelho distante (FRF, do inglês *Far Red Fluorescence*). A fluorescência no vermelho e no vermelho distante juntas são chamadas de fluorescência da clorofila, ChlF, pois a clorofila *a* é o único fluoróforo nas folhas com emissão nesta parte do espectro; a clorofila *b* transfere toda sua energia absorvida para a clorofila *a* in vivo.(19) A análise da fluorescência da clorofila *a* é considerada uma medida precisa e não-destrutiva da eficiência fotossintética que direta ou indiretamente pode refletir os impactos dos fatores ambientais sob a planta.

A fluorescência na região do azul e do verde também pode ser agrupada e tratada como a fluorescência BGF. Diferentemente da fluorescência da clorofila que provém exclusivamente da molécula de clorofila *a*, diversos compostos podem contribuir para a emissão do BGF. Destaca-se entre estes compostos o ácido ferúlico, uma vez que se acredita que ele seja responsável pela maior parte da emissão nessa região.(44) Ainda assim, outros produtos do metabolismo secundário das plantas podem contribuir para o BFG, tais como os ácidos cinâmicos, ácidos fenólicos, nicotinamidas (NADPH), entre outros. Detalhes sobre os demais compostos que fluorescem na região do azul e do verde podem ser encontrados no trabalho de Cerovic.(44) Apesar dos estudos envolvendo a emissão destes compostos na região do BGF, a ChlF tem sido mais estudada uma vez que está relacionada diretamente com a eficiência fotossintética da planta.(19)

Na região do BGF têm-se picos de emissão da fluorescência em 440 e 520 nm, enquanto que na região da ChlF os picos são em 690 e 740 nm. Uma vez que o BGF e a ChlF têm origens distintas, eles podem se alterar de forma independente em resposta aos diferentes fatores fisiológicos e ambientais. Na literatura, essas mudanças são mais convenientemente descritas em termos das variações de intensidade dos picos BF, GF, RF e FRF ou ainda através das razões dos picos de fluorescência. As razões de fluorescência geralmente

utilizadas são: a) RF/FRF, b) BF/RF e BF/FRF e c) BF/GF.(44) Em geral, a RF/FRF é utilizada para estudos sobre a eficiência fotossintética da planta; já a BF/ChlF ou a GF/ChlF são sensíveis aos estresses que as plantas estão sofrendo, especialmente por déficit de minerais.(44-45) Doenças, estresses ou outras alterações podem ser detectadas através da espectroscopia de fluorescência, visto que estas podem levar à variações nas concentrações dos pigmentos que fluorescem.

2.1.1 Técnicas e aplicações na agricultura

Nesta seção, será abordada brevemente a técnica de espectroscopia de fluorescência à fibra, destacando os principais resultados e motivos que levaram nosso grupo a utilizar a técnica de espectroscopia por imagem de fluorescência para a análise das doenças que afetam a citricultura. Em seguida, serão apresentadas as principais aplicações da técnica por imagens na agricultura, bem como os resultados envolvendo a análise de doenças na citricultura.

A espectroscopia de fluorescência se refere a um tipo de método espectroscópico, no qual a fluorescência do objeto de interesse é medida após sua excitação por uma fonte de luz. Nos últimos vinte anos, a espectroscopia de fluorescência induzida por laser (LIFS, do inglês *Laser-Induced Fluorescence Spectroscopy*) tem sido utilizada em diversos estudos, tais como no monitoramento de níveis de estresse, deficiência de nutrientes, estado fisiológico e doenças em plantas.(24,28,42,50) Basicamente, um equipamento de espectroscopia de fluorescência é constituído por uma fonte de radiação para a excitação das amostras, um guia de ondas para conduzir a radiação de excitação da fonte à amostra e a fluorescência da amostra ao detector, um monocromador e uma base computacional para a análise e armazenamento dos dados. No caso do LIFS, a fonte de radiação é um laser e o guia de ondas uma fibra óptica responsável tanto pela excitação das amostras quanto pela coleta da fluorescência.

No que concerne à aplicação da técnica LIFS na citricultura, o Laboratório de Interações Atômicas (LIA) possui ampla experiência em sua utilização. Destacam-se os estudos envolvidos na diferenciação do cancro cítrico da clorose variegada dos citros (CVC) (35,37), detecção de estresse hídrico (37), detecção e discriminação do cancro cítrico de amostras com estresse mecânico (42) e diagnóstico do cancro cítrico em relação a outras doenças.(28) Neste último trabalho, Lins e colaboradores (28) utilizaram um sistema de espectroscopia de fluorescência portátil e coletaram espectros de folhas com sintomas de cancro cítrico, CVC, verrugose, e greening. O sistema é composto por um laser de estado sólido em 532 nm (segundo harmônico do laser de Nd:YAG) e potência de 10 mW para a

excitação das folhas, um espectrômetro que opera entre 350 e 850 nm, uma fibra óptica no formato de Y, a qual conduz a radiação por meio de uma fibra central e coleta a fluorescência das folhas através de seis fibras periféricas e, por fim, um filtro óptico para bloquear o sinal de retroespalhamento e um computador para o armazenamento dos espectros. A fim de reduzir ruídos nos dados, os autores coletaram espectros de amostras com sintomas visíveis das doenças e também de amostras sadias, com o objetivo de minimizar impactos ambientais como deficiência de minerais e água sobre os espectros. A análise dos espectros, a partir de figuras de mérito, mostrou que essa técnica apresenta uma alta sensibilidade (~90%), mas uma baixa especificidade (~70%) no diagnóstico do cancro cítrico. Segundo os autores a baixa especificidade está relacionada com a escolha do local da folha para a coleta dos espectros, bem como a pequena área de detecção, uma vez que fica delimitada pelo diâmetro da fibra.(28,45)

Neste sentido, a técnica de espectroscopia de fluorescência por imagens é um avanço em relação à técnica de espectroscopia de fluorescência por fibras, onde imagens de fluorescência, em vez de simples espectros, são obtidas utilizando uma câmera. Esta técnica permite obter tanto informação espectral quanto espacial ao mesmo tempo. Os sistemas de imagens requerem alguns componentes de instrumentação básicos, tais como: fontes de luz, componentes óticos, detectores, sistemas de aquisição de dados e computadores. Estes componentes podem variar de acordo com a técnica de espectroscopia e o tipo de aplicação. Por exemplo, imagens de refletância na região visível do espectro eletromagnético podem ser detectadas com câmeras digitais comerciais, enquanto que as de fluorescência necessitam de um sistema mais sensível e, portanto, mais caro. Imagens de fluorescência também são possíveis através de câmeras convencionais (51), mas para se obter análises e estatísticas mais precisas destas imagens, sistemas com câmeras CCD monocromáticas (do inglês, *Charge Coupled Device*) combinado com diferentes filtros de emissão são geralmente utilizados. O uso de CCDs coloridas não é muito empregado devido à sua baixa sensibilidade.(51) Em relação às aplicações, em alimentos, a maior parte da aquisição das imagens é realizada em condições de ambiente controlado; entretanto, na agricultura, a aquisição pode ser tanto em ambiente controlado como também diretamente no campo conforme a aplicação.(52)

As regiões do espectro eletromagnético frequentemente utilizadas na coleta de imagens de fluorescência são: azul (440 nm), verde (520-550 nm), vermelho (690 nm), vermelho distante (740 nm) e infravermelho próximo (800 nm).(24) A técnica de espectroscopia por imagens nas regiões do visível e do infravermelho próximo oferece um método rápido, não-destrutivo, e de baixo custo em diversas aplicações na agricultura. Esta

técnica vem sendo utilizada em estudos sobre a atividade fotossintética das plantas, conteúdo da clorofila nas plantas, monitoramento da qualidade das frutas, e detecção de plantas doentes.(44,51) As imagens capturadas podem ser analisadas tanto diretamente para cada faixa espectral em que foram obtidas como também através da composição das razões destas diferentes bandas espectrais.(51)

Há duas formas de compor imagens espectrais, mantendo o campo de visão fixo e varrendo a detecção espectral ou mantendo a detecção espectral fixa e varrendo o campo de visão. Na primeira técnica, é necessária a utilização de um elemento óptico que possibilite a seleção de uma banda espectral para a detecção da fluorescência. Exemplos desses elementos são os filtros ópticos; a seleção pode ser feita eletronicamente através do uso de filtros sintonizáveis, ou mecanicamente, utilizando filtros ópticos tradicionais associados a rodas de filtros. Na segunda técnica, a informação espectral é fornecida por elementos difrativos, espectrógrafos, elementos que dispersam o sinal óptico do campo de visão para a formação de imagens. Entretanto, para evitar a sobreposição da informação espacial com espectral, é necessário reduzir o campo de visão do sistema, o que implica a necessidade de varrer espacialmente a amostra.(29)

Desta forma, a evolução dos dispositivos ópticos com resolução espectral tem contribuído significativamente para o estudo das doenças e pragas que afetam as culturas agrícolas. Exemplos das principais aplicações da técnica de espectroscopia de fluorescência podem ser encontrados nos seguintes trabalhos (24,30,51). Lenk e colaboradores (51) discutem as diferentes instrumentações para a aquisição de imagens multiespectrais de fluorescência de folhas e frutas, bem como suas possíveis aplicações no monitoramento da qualidade das frutas, atividades fotossintéticas, estrutura do tecido foliar, e sintomas de doenças em plantas.(51) Nos trabalhos (30) e (24), os autores descrevem as tecnologias atuais que podem ser empregadas no desenvolvimento de técnicas de sensoriamento remoto a fim auxiliar no monitoramento de plantas saudáveis e contaminadas. Em especial, destacam-se os trabalhos para monitoramento do vírus mosaico em plantas de tabaco, bem como a aplicação de sistemas multiespectrais e hiperespectrais para a detecção do greening em plantações de citros.(26,53) Sistemas multiespectrais, geralmente, coletam os dados de 3 a 7 bandas espectrais em uma única observação na região do visível e do infravermelho próximo do espectro, ao passo que os sistemas hiperespectrais podem conter dezenas a centenas de bandas espectrais estreitas, fornecendo assim informação adicional em relação aos sistemas multiespectrais.(54)

Alterações na emissão da fluorescência na região do azul e do verde (BGF), assim como da fluorescência da clorofila (ChlF) em plantas expostas a radiação ultravioleta podem fornecer informações importantes sobre o estado fisiológico das plantas.(24) Chaerle e colaboradores exploraram esta característica em seu trabalho.(26) Utilizando imagens de fluorescência na região de emissão do BGF e da ChlF avaliaram a eficácia da técnica para observar a evolução da infecção pelo vírus mosaico em plantas de tabaco (TVM, do inglês *Tobacco Mosaic Virus*). Imagens de refletância em 550 nm e 800 nm foram coletadas e consideradas como imagens de referência. As imagens de fluorescência demonstraram uma diferença visível entre as folhas infectadas e não infectadas em um curto intervalo de tempo (50 horas) em comparação com as imagens de referência (14 dias para sintomas visíveis da infecção). As alterações na emissão da fluorescência foram detectadas antes que os sintomas fossem visíveis ao olho humano ou através em uma câmera RGB (do inglês, *Red, Green, Blue*). O trabalho demonstra o potencial da técnica de imagens de fluorescência para o monitoramento do vírus TVM em plantas de tabaco.(26)

Vale destacar o recente trabalho de Mo e colaboradores (55) que também exploraram a excitação ultravioleta para a detecção do estresse hídrico em culturas de soja. O objetivo desse trabalho foi desenvolver um método para a detecção rápida do estresse hídrico em folhas de soja utilizando a técnica de imagens de fluorescência hiperespectral. Utilizando um sistema hiperespectral e diodos emissores de luz (LEDs, do inglês *Light Emitting Diodes*) em 365 nm como fonte de excitação, imagens de fluorescência foram obtidas para duas culturas distintas de sojas entre 421 a 780 nm em intervalos de 4,8 nm durante 16 dias. O modelo de regressão PLSR (do inglês, *Partial Least Square Regression*) foi desenvolvido a fim de avaliar o estresse hídrico, assim como a técnica de imagens por PLS (do inglês, *Partial Least Square*) para verificar o modelo de predição desenvolvido. As imagens PLS foram obtidas aplicando o modelo PLSR nas imagens de fluorescência hiperespectrais. A análise das imagens mostrou que a precisão na discriminação do estresse hídrico nas duas culturas foi de 0,973 e 0,969 para o grupo de tratamento de 8 e 6 dias, respectivamente. Estes resultados validam o uso de imagens de fluorescência na detecção do estresse hídrico nas culturas de soja.(55)

Os sistemas de imagens multiespectrais e hiperespectrais também têm sido amplamente utilizados em conjunto com aeronaves e satélites em estudos de sensoriamento remoto do meio ambiente, detecção de deficiências minerais e doenças que afetam as culturas agrícolas.(53-54) A quantidade de informação obtida através das imagens de satélite tem permitido o monitoramento das características dinâmicas da paisagem, fornecendo assim um meio de detecção de importantes alterações na cobertura terrestre e quantificação das taxas de

variação.(54) Kumar e colaboradores propuseram em seu trabalho (53) um método para a detecção das áreas infectadas pelo greening em pomares de citros através da aquisição de imagens multiespectrais e hiperespectrais por meio de um drone. O sensor hiperespectral em conjunto com uma câmera CCD capturaram imagens de 128 bandas desde 457,2 a 921,7 nm em intervalos de 3,63 nm, enquanto que o sistema multiespectral coletou imagens para as seguintes bandas: azul (430 a 470 nm), verde (530 a 570 nm), vermelho (630 a 670 nm), e no infravermelho próximo (810 a 850 nm). As imagens foram analisadas através do software ENVI, sendo as áreas infectadas pelo greening identificadas por meio dos métodos SAM (do inglês, *Spectral Angle Mapping*) e MTMF (do inglês, *Mixture Tuned Matched Filtering*). A precisão do método SAM usando imagens multiespectrais (87%) foi comparável aos resultados obtidos pelo método MTMF, entretanto, a precisão foi maior quando comparada com a obtida para as imagens hiperespectrais. Segundo os autores, essa precisão pode ser melhorada aliando um levantamento terrestre mais abrangente da área em estudo com mais resultados de PCR, o que ajudaria na quantificação do número de amostras falsos positivos.(53)

Em relação à aplicação da espectroscopia por imagens de fluorescência na citricultura, destaca-se o trabalho de Pereira e colaboradores que aplicaram a técnica de microscopia por imagens de fluorescência para avaliar o greening (*Huanglongbing*) em culturas de citros.(27) O estudo investigou a fluorescência emitida por amostras sadias e contaminadas com greening, além de avaliar o potencial da técnica para o diagnóstico precoce da infecção causada pela bactéria. As plantas foram avaliadas mensalmente e a evolução da bactéria inoculada nas plantas foi monitorada através da técnica qrt-PCR (do inglês, *quantitative real-time PCR*). Na análise dos dados, utilizou-se um método estatístico para a seleção das variáveis dos histogramas de cores das imagens. O método analítico proposto é baseado no estresse causado pela bactéria nas amostras e foi capaz de diagnosticar folhas sadias e doentes entre quinto ao oitavo mês após a infecção.(27)

Dentre os estudos envolvendo imagens de fluorescência em doenças cítricas, vale destacar também a publicação do nosso grupo utilizando a técnica de espectroscopia por imagens em um estudo comparativo com amostras sadias e contaminadas com greening do estado de São Paulo/Brasil e da Flórida/EUA.(23) Os resultados indicaram que a aplicação desta técnica em conjunto com o classificador SVM apresenta uma precisão de até 90% na comparação das amostras contaminadas e sadias.(23) Embora as intensidades de fluorescência das folhas coletadas no Brasil e EUA fossem diferentes, este método apresenta um elevado

potencial para a detecção do greening. Maiores detalhes sobre esse trabalho serão abordados no capítulo Estudo Preliminar.

Por fim, outro trabalho que vale destacar é o de Pourreza e colaboradores (38) que desenvolveram um sistema de detecção por imagens em tempo real do greening no campo. Um dos sintomas característicos do greening é o acúmulo de amido nas folhas infectadas de citros criando manchas amareladas num padrão assimétrico. Esse sintoma pode ser confundido com a deficiência de determinados nutrientes, tais como magnésio e zinco. Em um estudo anterior (56), eles demonstraram que o acúmulo de amido resulta na rotação do plano de polarização em 90° da luz no comprimento de onda em 591 nm, podendo esta característica ser utilizada na identificação das folhas contaminadas com greening, assim como na diferenciação da doença de amostras com deficiência de zinco ou magnésio. O sistema desenvolvido é constituído basicamente por uma câmera monocromática altamente sensível, 10 LEDs de alta potência em 591 nm e filtros polarizadores. O sistema foi testado e calibrado primeiramente no laboratório simulando as condições do campo e, em seguida, foi aplicado diretamente nos pomares. A fim de confirmar a presença da doença foi realizado o teste qrt-PCR. A partir das imagens foram extraídos os descritores, média e desvio padrão, os quais foram utilizados no processo de classificação pelo SVM. Os resultados mostraram que o sistema claramente destaca o acúmulo de amido nas folhas contaminadas por greening e as diferencia de amostras com deficiência de zinco, com acurácias variando entre 95,5% a 98,5%, tanto no laboratório quanto no campo.(38)

2.2 Principais aplicações dos classificadores SVM e ANN

A técnica de espectroscopia por imagens em conjunto com métodos de classificação são uma alternativa para a detecção de doenças no campo. O caráter multivariado dos dados obtidos das imagens, tanto na forma de valores espectrais como características, pode ser estudado através de análises qualitativas ou quantitativas.(52) A análise qualitativa envolve a classificação, onde grupos ou categorias de características são identificados, por exemplo, classificação de plantas saudáveis de plantas doentes. Em relação à análise quantitativa, esta envolve a predição, onde propriedades específicas dos objetos são estimadas, tais como, concentração de nutrientes das folhas e sólidos solúveis de sucos.(52) Técnicas de aprendizado supervisionado e não supervisionado são utilizados para o tratamento dos dados em resultados interpretáveis. Dentre as técnicas de aprendizado supervisionado, destacam-se o

classificador Naïve Bayes, a análise discriminante, K-nn (do inglês, *K-nearest neighbor*), lógica fuzzy, árvores de decisão, SVM e ANN.(52)

Nos últimos anos, as técnicas SVM e ANN têm sido amplamente utilizadas em sistemas de visão para classificação de imagens e reconhecimento de padrões.(23,38,57-60) Dentre esses estudos, destaca-se o trabalho de Melgani e colaboradores (58) que investigaram a performance do SVM na classificação de imagens hiperespectrais coletadas na região do visível e do infravermelho por um sistema aerotransportado. Dois tipos de SVM foram avaliados, o linear (SVM-linear) e o de função de base radial (SVM-RBF, do inglês *Radial Basis Function*). Para sustentar a análise do SVM, seu desempenho foi comparado com outros dois classificadores não-paramétricos, as redes neurais utilizando a função de base radial e o K-nn. Os autores adotaram quatro estratégias diferentes para a classificação multiclases dos dados hiperespectrais: um contra todos, um contra um e duas estratégias baseadas em árvore hierárquicas. Os resultados obtidos em função do número de características extraídas para os diferentes classificadores são apresentados na Figura 2.4. A partir desses resultados, é possível concluir que o SVM, particularmente o que utiliza a função RBF, é uma alternativa válida para o reconhecimento de padrões na classificação de imagens hiperespectrais obtidas pelo sensoriamento remoto.(58) Também é possível observar que a rede neural, em determinados casos, apresenta precisão da ordem de 90%.

Figura 2.4 - Precisão dos diferentes classificadores em função do número de características extraídas das imagens hiperespectrais.
Fonte: MELGANI (58).

Vale destacar também o trabalho de Zhao e colaboradores (61) que aplicaram os dois métodos, SVM e redes neurais, para o monitoramento da qualidade do chá verde. Os autores desenvolveram um sistema de imagens hiperespectrais para a aquisição dos dados entre 408 e

1117 nm com resolução espectral de 2,8 nm. Inicialmente, os autores aplicaram a técnica de análise das componentes principais (PCA, do inglês *Principal Component Analysis*) sobre os dados hiperespectrais a fim de reduzir sua dimensionalidade e determinar as três melhores bandas espectrais. Os comprimentos de onda selecionados foram 762, 793 e 838 nm, e características como média, desvio padrão, energia e entropia foram extraídas de cada imagem para essas bandas específicas. Maiores detalhes sobre o cálculo e a definição destes quatros descritores estatísticos podem ser encontrados em (61).

Os autores compararam o desempenho do algoritmo SVM com os seguintes classificadores: LDA (do inglês *Linear Discriminant Analysis*), e dois tipos de redes neurais, a retropropagação ANN (BP-ANN, do inglês *backpropagation ANN*) e a RBF-ANN. O desempenho do BP-ANN se mostrou superior ao do RBF-ANN, sendo então utilizado na comparação com o SVM e o LDA. As taxas de precisão da classificação para o conjunto de dados de validação foram de 95%, 90% e 85% para o SVM, BP-ANN e LDA, respectivamente.(61) Portanto, o algoritmo SVM apresenta um excelente desempenho na classificação das imagens hiperespectrais de amostras de chá verde, seguido pelas redes neurais.

As diferenças observadas nas precisões da classificação acima se deve primeiramente ao fato que o SVM e o BP-ANN são abordagens de aprendizagem não-lineares, enquanto que o LDA é um método linear. Em geral, o método não-linear é mais eficaz do que o método linear no nível de autoaprendizagem e auto ajuste aos dados.(61) No caso do SVM e do BP-ANN, o primeiro método incorpora uma excelente generalização em sua teoria, o que resulta em melhores resultados preditivos do que o método BP-ANN.(61-62)

Classificadores estatísticos, redes neurais e SVMs vêm sendo utilizados na detecção e contagem de frutos não maduros a fim de ajudar os produtores na criação de mapas de produtividade.(57, 63) Estes mapas são ferramentas úteis nos ajustes das práticas de manejo durante os estágios de maturação das frutas. Destaca-se o trabalho de Kurtulmus e colaboradores (57), que testaram diferentes tipos de classificadores a fim de determinar o melhor classificador para a detecção e contagem de pêssegos não maduros a partir das características extraídas de imagens coloridas obtidas em condições de iluminação natural. As características extraídas das imagens representam informações referentes à forma e textura dos objetos. Os classificadores utilizados foram: LDA, Naïve Bayes, K-nn, árvores de regressão e classificação (CART, do inglês *Classification and Regression Trees*), SVM e ANN. A detecção dos pêssegos no campo está sujeita a obstáculos específicos, tais como iluminação desigual e oclusão dos objetos da copa das árvores. No entanto, os algoritmos

propostos conseguiram detectar com sucesso cerca de 85% dos frutos de pêssegos não maduros nas imagens. Os classificadores SVM e ANN, seguido do LDA, apresentaram os melhores desempenhos na classificação dos pêssegos a partir das características extraídas. O desenvolvimento de um sistema de predição precoce dos frutos não maduros a partir de câmeras instaladas em uma plataforma móvel e com aquisição contínua de imagens em conjuntos com estes classificadores é o objetivo final do estudo desses autores.(57)

Com relação aos estudos que envolvem a aplicação destes classificadores na citricultura, mencionou-se na seção de Técnicas e Aplicações na Agricultura os trabalhos de Wetterich e colaboradores (23), assim como o de Pourreza e colaboradores (38). Nestes trabalhos, os autores utilizaram o classificador SVM para a detecção e discriminação do greening nas folhas de citros obtendo altas taxas de precisão. Vale destacar também o trabalho de Sankaran e colaboradores (60), que demonstraram a aplicabilidade de imagens térmicas e na região do visível e do infravermelho próximo para a detecção do greening. Os classificadores utilizados foram: LDA, análise discriminante quadrático (QDA, do inglês *Quadratic Discriminant Analysis*), BDT (do inglês, *Bagged Decision Tree*) e SVM. Nesse estudo, o classificador SVM apresentou a maior precisão na classificação em relação aos demais classificadores, com uma precisão média de 87% e um número mínimo de falsos negativos.(60)

Isso posto demonstra o potencial da técnica de espectroscopia em conjunto com métodos de reconhecimento de padrões e classificação na detecção de doenças em citros. Dessa forma, uma vez que temos experiência no uso do método SVM na classificação de imagens de fluorescência de folhas infectadas com greening (23), assim como o método ANN também se destaca na classificação de imagens, resolvemos expandir nossas análises e utilizar, além do SVM, o classificador ANN para a discriminação de doenças com sintomas visuais semelhantes. No capítulo seguinte, será abordada a metodologia envolvida para o desenvolvimento deste trabalho, bem como a base teórica dos classificadores SVM e ANN.

3 INSTRUMENTAÇÃO E METODOLOGIA

Este capítulo é dedicado à instrumentação desenvolvida, bem como à metodologia empregada. Serão descritos os dois sistemas de imagens de fluorescência construídos para a aquisição das imagens, a metodologia envolvida na coleta das amostras e aquisição das imagens fluorescência, assim como seu pré-processamento. Também será descrito as duas técnicas empregadas para a análise e classificação das doenças em estudo.

3.1 Sistemas de espectroscopia por imagens de fluorescência

Como mencionado anteriormente no capítulo de Revisão da Literatura, os sistemas de espectroscopia por imagens de fluorescência evitam o problema da análise de uma pequena região da amostra e a necessidade de se realizar várias medidas na mesma, como no caso da técnica LIF, uma vez que permite a análise dos sinais ópticos com resolução espacial e espectral.(45) Neste trabalho foram desenvolvidos dois diferentes sistemas de imagens de fluorescência. A seguir, serão descritos os dois sistemas desenvolvidos, destacando a instrumentação envolvida e suas principais características.

3.1.1 Sistema de imagens de fluorescência utilizando filtro óptico de cristal líquido variável (SIF-1)

O primeiro sistema desenvolvido, baseado no uso de um filtro sintonizável, tem como objetivo varrer a região visível do espectro eletromagnético para as doenças/estresse em estudo que afetam os citros e, consequentemente, selecionar as principais bandas espectrais. Este primeiro sistema de imagens de fluorescência (SIF-1) construído é composto por uma câmera CCD científica, um sistema de lentes, um filtro óptico de cristal líquido variável, um conjunto de iluminação composto por LEDs, um laptop e uma estrutura metálica para fixar os componentes e as amostras.

O sistema é constituído por um filtro óptico de cristal líquido variável (*Liquid Crystal Tunable Filter*, Meadowlark Optics, EUA) que responde à tensão aplicada, variando assim sua transmissão. Este filtro é composto por múltiplos retardadores de cristal líquido variável, retardadores invariantes e polarizadores, todos protegidos e com temperatura controlada, o qual permite selecionar o comprimento de onda de interesse na faixa entre 420 e 750 nm. O filtro permite também a escolha de três diferentes larguras de banda para cada comprimento

de onda: estreito, médio e largo. No intervalo entre 420 e 750 nm a largura de banda varia de 5 a 19 nm, 13 a 35 nm e de 22 a 65 nm, para as configurações estreita, média e larga, respectivamente. Estas diferentes larguras permitem alterar a transmissão da luz através do filtro óptico sintonizável e as características de transmissão espectral. A Figura 3.1 apresenta o espectro de transmissão em função do comprimento de onda.

Figura 3.1 - Espectro de transmissão do filtro de cristal líquido variável em função do comprimento de onda.
Fonte: LIQUID...(64)

O filtro utiliza de uma configuração padrão 8-32 hardware, sendo conectado a um módulo de controle por um cabo de 25 pinos, responsável pelo controle e estabilização da temperatura do filtro. O controle da temperatura é importante uma vez que a birrefringência dos retardadores de cristal líquido variável é uma função da temperatura, assim como da voltagem. Por esta razão, todos os componentes do filtro são mantidos a 40°C, de modo que as condições do ambiente não afetem a calibração ou a velocidade de troca dos comprimentos de onda. O tempo para a estabilização é de aproximadamente 90 minutos. O módulo de controle se conecta a um computador através de um cabo USB ou cabo RS232. O software fornecido pelo fabricante permite a seleção do comprimento de onda, a largura de banda, bem como informa sobre o status da temperatura do filtro. A Figura 3.2 apresenta os componentes que constituem este conjunto do sistema do filtro óptico sintonizável.

Figura 3.2 - Componentes do filtro óptico sintonizável. (1) Filtro óptico variável; (2) Cabo de 25 pinos; (3) Módulo de controle; (4) Cabo USB ou RS232; (5) Fonte de alimentação; (6) Computador com o software do fabricante.

Fonte: LIQUID...(64)

O conjunto de iluminação é composto por LEDs de alta potência, aproximadamente 300 mW por diodo. O sistema é formado por LEDs com picos de emissão em 405 (ProLight Opto Technology Corporation, USA) e 470 nm (Cree®, USA) e largura de banda de 10 nm. Os diodos estão fixados a um suporte metálico com inclinação adequada para a distribuição homogênea da luz sobre as superfícies das folhas e são alimentados por uma fonte de tensão. O suporte metálico também tem como função dissipar o calor dos diodos. Os LEDs em 405 e 470 nm tem máxima eficiência quando são fornecidos 1.2 A e 0.7A, respectivamente.

Acoplados ao filtro variável temos um sistema de lentes e uma câmara CCD para a obtenção das imagens. A câmera CCD utilizada é uma mvBlueFox 120G (Matrix Vision, Alemanha). Ela possui resolução de 480 x 640 pixels na área ativa do sensor, controle eletrônico de exposição que varia entre 1 a $10^7 \mu s$ e ganho do sensor de 0 a 30 dB. As imagens formadas são monocromáticas com 8, 10 ou 12 bits por pixels. Possui comunicação com computadores via USB 2.0 e drivers de captura de imagens para o software LabVIEW® (National Instruments, USA). A lente objetiva utilizada é uma PENTAX TV LENS 6 mm que possui controle manual do foco da imagem e da íris para atender as condições de visualização e iluminação. O foco das imagens uma vez ajustado não é modificado mesmo com a troca das amostras. O campo de visão desta lente é um retângulo de dimensões 6,7 x 8,7 cm, aproximadamente. A resolução do pixel é de 0,014 cm, valor estimado a partir dos dados da resolução da câmera CCD e do campo de visão da lente.

Além disto, uma estrutura metálica no formato de uma caixa, anodizada na cor preta, foi confeccionada para alinhar e fixar os componentes e posicionar as amostras de citros para

a coleta das imagens espectrais. Todos os componentes acima descritos estão posicionados na parte interna da caixa com a intenção de evitar a influência de luz externa. O esquema do sistema de imagem de fluorescência é apresentado na Figura 3.3.

Inicialmente, o módulo de controle do filtro variável é ligado a fim de estabilizar sua temperatura. Após a estabilização, as amostras das folhas de citros são fixadas na base da caixa e os LEDs são ligados através de uma fonte de tensão que permite o controle da intensidade luminosa. Em seguida, seleciona-se o comprimento de onda de interesse através do software do filtro óptico sintonizável e captura-se a respectiva imagem de fluorescência através do software da câmera CCD, a qual é salva no formato *Bitmap*. Vale destacar que este sistema apresenta como principal característica seu fácil manuseio e portabilidade, entretanto é inviável de se aplicar no campo uma vez que o tempo para a estabilização da temperatura do filtro é significativa.

Figura 3.3 - (a) Esquema da montagem do SIF-1, destacando os principais componentes ópticos: (1) câmera CCD, (2) lentes objetivas, (3) filtro de cristal líquido variável, (4) sistema de iluminação à LEDs em diferentes comprimentos de onda: 365, 405, 470 e 530 nm; (b) Filtro óptico de cristal líquido. Fonte: Elaborada pelo autor.

3.1.2 Sistema de imagens de fluorescência utilizando filtros ópticos passa-banda associados a rodas de filtros (SIF-2)

A partir do sistema de imagens de fluorescência descrito no tópico anterior foi possível selecionar as principais bandas de emissão para as diferentes doenças analisadas neste trabalho. Após a seleção destas bandas, passamos a utilizar um segundo sistema (SIF-2), similar ao descrito, porém este faz uso de uma roda de filtros no qual a seleção é feita mecanicamente a partir de filtros ópticos passa-banda. Este sistema está descrito em maiores detalhes na dissertação de mestrado de WETTERICH, C. B.(12)

O sistema é constituído pelos seguintes filtros ópticos passa-banda: 530, 550, 560, 570, 580, e 690 nm (modelos FB530-10, FB550-10, FB560-10, FB570-10, FB580-10, e FB690-10, Thorlabs, EUA). Os filtros apresentam uma largura de banda de 10 nm, uma transmissão típica de 50% dentro da banda, e menor que 0,01% fora da banda.(30) Tais modelos de filtros foram selecionados conforme a análise dos espectros de emissão de cada doença, sendo o procedimento descrito no capítulo dos Resultados.

A roda de filtros utilizada foi o modelo FW102B (Thorlabs, USA) que comporta até seis filtros ópticos. Possui comunicação com computadores via USB 2.0, permitindo o controle dos filtros através do próprio software fornecido pelo fabricante. Além disto, o fabricante fornece drivers que podem ser utilizados para o controle da roda via o software LabVIEW[®] (National Instruments, USA). Em relação ao sistema de iluminação, este é composto por diodos emissores de luz (LEDs) de alta potência, totalizando 8 LEDs divididos em dois conjuntos. Os conjuntos são formados por LEDs em 405 e 470 nm. Os diodos estão fixados a um suporte metálico com inclinação de 45° para a distribuição homogênea da luz sobre as superfícies das folhas, e são alimentados por uma fonte de tensão.

Os demais componentes deste sistema para a aquisição das imagens de fluorescência são idênticos ao utilizado no primeiro sistema e estão posicionados na parte interna da caixa com a intenção de evitar a influência da luz externa na aquisição das imagens. O esquema do SIF-2 é apresentado na Figura 3.4.a, e na Figura 3.4.b uma foto digital do sistema completo inclusive com o laptop e a excitação em 470 nm ligada.

Este segundo sistema já havia sido utilizado pelo nosso grupo em um estudo preliminar para a detecção do greening em amostras do Brasil e dos Estados Unidos.(23) A principal vantagem deste sistema em relação ao primeiro é que não é necessário aguardar a estabilização da temperatura, além da troca dos comprimentos de onda ser mais rápido e pode ser utilizado diretamente no campo para a coleta das imagens.

Figura 3.4 - (a) Esquema da montagem do SIF-2 , destacando os principais componentes ópticos: (1) câmera CCD, (2) lentes objetivas, (3) roda de filtros, (4) sistema de iluminação à LEDs em 405 e 470 nm. (b) Imagem digital do sistema completo com a excitação em 470 nm ligada.
Fonte: Elaborada pelo autor.

3.2 Metodologia

Esta seção é dedicada à metodologia envolvida na coleta das amostras, na aquisição das imagens de fluorescência, assim como o pré-processamento destas. Também serão descritas as técnicas empregadas para a análise e classificação das doenças em estudo.

3.2.1 Amostras de citros

Os sistemas de imagens de fluorescência desenvolvidos foram aplicados em duas regiões, Brasil e Estados Unidos, a fim de verificar a eficácia da técnica em condições diferentes. Primeiramente no caso das amostras do Brasil, foi conduzido um experimento exploratório utilizando o SIF-1 para a seleção das principais bandas espectrais de fluorescência que distinguem cada par de doenças. As imagens foram coletadas para 5 amostras de cada uma das seguintes condições: i) cancro cítrico, ii) verrugose, iii) greening e iv) deficiência de zinco. Na sequência, foram coletadas imagens de fluorescência para 100 amostras de cada doença utilizando o SIF-2. Estas imagens foram obtidas para as excitações em 405 e 470 nm utilizando os filtros previamente selecionados no primeiro experimento. Os

experimentos foram conduzidos no LIA uma vez que a agência Fundecitrus se disponibilizou a coletar e encaminhar as amostras para o estudo.

As folhas foram coletadas das espécies de plantas cítricas *Citrus reticulata* (“Ponkan”), *Citrus sinensis* (“Hamlin”) e *Citrus latifolia* (“Limão-Tahiti”). O uso de diferentes espécies de plantas cítricas não influencia nos resultados, como já demonstrado pelo nosso grupo em estudos anteriores envolvendo a técnica de espectroscopia induzida por laser.(28) Além disso, as amostras possuíam diferentes níveis de severidade e idade, variando desde sintomas iniciais até situações em que estão gravemente infectadas. Uma vez que estamos interessados em desenvolver uma técnica que possa ser aplicada diretamente em amostras do campo independente de suas condições, os experimentos não levaram em conta essas variáveis. A única condição exigida foi que as amostras fossem coletadas para cada doença com base em seus sintomas no campo e somente se não apresentassem outros sintomas macroscópicos.

As amostras foram coletadas por técnicos qualificados da agência Fundecitrus entre os meses de Maio e Agosto de 2013 no município de Araraquara/SP e transportadas ao laboratório para a aquisição das imagens em caixa de isopor fechadas. Vale salientar que a fim de se evitar qualquer efeito do tempo de destacamento das folhas, as medidas de fluorescência foram realizadas imediatamente após sua coleta.(28) Após a coleta das imagens de fluorescência, as folhas foram transportadas para a agência Fundecitrus para o diagnóstico das doenças através do uso de técnicas tradicionais.(31,65-67) Amostras com cancro cítrico e verrugose foram diferenciadas através do isolamento da bactéria *Xanthomonas* em meio de cultura apropriado.(28,31) O greening e a deficiência de zinco foram diferenciadas através da técnica qPCR (do inglês, *quantitative Polymerase Chain Reaction*).(32) Estes testes são considerados como testes padrão-ouro e seus resultados foram utilizados para checar a validade da técnica proposta.

No caso do estudo realizado na Flórida/EUA, o procedimento adotado para a coleta das amostras é idêntico. As amostras foram coletadas no CREC (Citrus Research and Education Center), Universidade da Flórida, por técnicos qualificados, entre os meses de Janeiro e Fevereiro de 2014. Entretanto, no período da coleta não haviam amostras contaminadas com verrugose. Desta forma, o estudo foi conduzido de forma a diferenciar somente amostras sintomáticas com greening ou deficiência de zinco. Utilizando o SIF-1, imagens de fluorescência de 5 amostras de cada uma das doenças foram coletadas. A partir deste experimento exploratório, as principais bandas espectrais foram selecionadas e imagens de fluorescência foram capturadas para 100 amostras de cada condição utilizando o SIF-2. As

folhas foram coletadas da espécie de planta cítrica *Citrus sinensis* (“Hamlin”). Após a coleta das imagens, a técnica bioquímica qPCR foi utilizada para a confirmação da doença/estresse. Estes testes foram conduzidos pelo Prof. Dr. William Gurley da Universidade da Flórida, e os resultados utilizados na validação da técnica proposta.

3.2.2 Aquisição e pré-processamento das imagens

O software utilizado para a seleção dos comprimentos de onda a partir do filtro de cristal líquido variável foi desenvolvido pelo próprio fabricante. A Figura 3.5 apresenta a interface do software fornecido. Como mencionado anteriormente na seção 3.1.1, este software permite selecionar o comprimento de onda e a largura de banda, bem como informa sobre o status da temperatura do filtro.

Figura 3.5 - Interface do software de seleção dos comprimentos de onda do filtro variável.
Fonte: Elaborada pelo autor.

A aquisição das imagens de fluorescência é realizada através do software fornecido pela empresa fabricante da câmera CCD. Esta é controlada através de um computador via conexão USB 2.0. O software permite ajustar o ganho e o tempo de exposição da câmera CCD, salvar as imagens de fluorescência conforme os filtros são selecionados no filtro sintonizável, além de fornecer várias outras opções de controle. Os valores de “Ganho” e

“Tempo de Exposição” da CCD foram ajustados manualmente para que a imagem não apresentasse regiões de saturação. Alterações nestes valores podem ser visualizadas instantaneamente na tela do software para o seu devido controle. As imagens espectrais são armazenadas no formato *Bitmap*. A Figura 3.6 apresenta a interface do software fornecido pelo fabricante.

Figura 3.6 - Interface do software de aquisição das imagens de fluorescência.

Fonte: Elaborada pelo autor.

As imagens de fluorescência das amostras respondem de formas distintas a cada doença estudada e comprimento de onda selecionado, sendo necessário o ajuste do ganho e tempo de exposição da imagem a cada nova seleção no filtro variável. Durante a aquisição das imagens de fluorescência, o ganho da câmera CCD foi mantido fixo e variou-se somente o tempo de exposição para os diferentes comprimentos de onda. O motivo de mantermos fixo o ganho e variarmos o tempo de exposição está relacionado com o fato de que o sinal de resposta da CCD aumenta linearmente com o tempo, como apresentado na Figura 3.7.

Figura 3.7 - Resposta da câmera CCD em função do tempo de exposição.
Fonte: Elaborada pelo autor.

No entanto, o processo de análise das imagens de fluorescência das diferentes doenças se torna inviável, uma vez que não se tem um padrão para comparação. Primeiramente, devemos levar em consideração tanto a curva de resposta da câmera CCD quanto à curva de transmissão do filtro variável em função do comprimento de onda selecionado. A Figura 3.8 apresenta a sensibilidade espectral da câmera CCD, bem como a curva de transmissão do filtro variável. Vale observar que a câmera CCD é mais sensível na região da cor verde do espectro eletromagnético, e a transmissão do filtro variável aumenta conforme o aumento do comprimento de onda selecionado.

Figura 3.8 - (a) Gráfico da sensibilidade espectral da câmera CCD. (b) Gráfico da curva de transmissão do filtro variável.
Fonte: Elaborada pelo autor.

A partir destes dados foi realizado um ajuste polinomial de sexta ordem para cada curva obtendo assim duas funções que representassem tal comportamento. Uma nova equação foi gerada a partir do produto da função de transmissão da câmera CCD (T_{CCD}) pela função da curva de transmissão do filtro variável (T_{filtro}). Esta nova equação foi normalizada segundo seu valor máximo no intervalo entre 420 e 750 nm. Tal função é denominada de espectro normalizado e será representada pelo símbolo $En(\lambda)$. A Figura 3.9 apresenta o comportamento deste espectro.

Figura 3.9 - Comportamento da função obtida a partir do produto da equação resposta da câmera CCD pela equação da curva de transmissão do filtro.

Fonte: Elaborada pelo autor.

Sendo assim, foi realizada a eliminação da informação espectral para os diferentes comprimentos de acordo com a seguinte expressão:

$$I_s(\lambda) = \frac{I(\lambda)}{En(\lambda)} \quad (1)$$

onde $I(\lambda)$ representa a imagem selecionada para um comprimento de onda específico.

Em seguida novas imagens foram obtidas, $I_s(\lambda)$, para os diferentes comprimentos de onda, restando eliminar a informação dos diferentes tempos de exposição da câmera CCD (t_{exp}). O procedimento para a eliminação do tempo de exposição consiste em selecionar uma imagem para um determinado comprimento de onda, $I_s(\lambda)$, e multiplicar pela razão entre o menor tempo de exposição (t_{min}) e o tempo de exposição da imagem neste comprimento de onda ($t(\lambda)$). No caso, t_{min} representa o menor tempo de exposição dentre todas as amostras

que compõem os grupos de comparação, ou seja, o menor tempo de exposição utilizado na comparação do cancro cítrico e verrugose para uma excitação específica, por exemplo. O mesmo procedimento foi adotado no caso da comparação de amostras contaminadas com greening ou afetadas por deficiência de zinco. A eliminação da informação do tempo de exposição só é possível uma vez que a resposta da câmera CCD é linear com o tempo. A equação a seguir exemplifica o procedimento descrito para eliminação do tempo de exposição:

$$I_{final} = I_s(\lambda) * \frac{t_{min}}{t(\lambda)} \quad (2)$$

onde t_{min} representa o menor tempo de exposição das amostras para uma excitação específica comparando as doenças de cada grupo de estudo, e $t(\lambda)$ representa o tempo de exposição da imagem no comprimento de onda analisado. No caso das imagens obtidas pelo SIF-2, o procedimento adotado é semelhante ao descrito, onde substituímos a transmissão do filtro variável pela porcentagem de transmissão do filtro óptico passa-banda no comprimento de onda de interesse. O pré-processamento das imagens é realizado para as diferentes excitações e doenças através de uma rotina computacional desenvolvida no software Matlab® (The MathWorks, EUA). As imagens obtidas após este pré-processamento são salvas no formato *Bitmap*.

3.2.3 Análise das imagens de fluorescência

É importante salientar que a aquisição das imagens de fluorescência e seu pré-processamento é apenas o primeiro passo; o próximo passo consiste em discriminar as doenças através de técnicas de análise de imagens. Neste trabalho, dois métodos de aprendizado supervisionado e não-paramétricos foram testados a fim de determinar qual classificador é mais eficaz no diagnóstico de doenças em citros a partir de características extraídas das imagens de fluorescência. Os classificadores utilizados foram: SVM (do inglês, *Support Vector Machine*) e ANN (do inglês, *Artificial Neural Network*). Estes métodos têm sido utilizados com sucesso na classificação de imagens multiespectrais.(23,57,68) A seguir serão descritos os procedimentos envolvidos na análise das imagens, desde a segmentação e extração das características das imagens, até uma breve introdução sobre os dois métodos de classificação.

3.2.3.1 Segmentação das imagens

Os procedimentos de análise das imagens de fluorescência foram desenvolvidos no software Matlab[®]. A segmentação das imagens obtidas pelo SIF-2 é o primeiro passo da análise na identificação de possíveis áreas úteis para a extração das características. A seleção da região de interesse de uma imagem é um processo crítico, uma vez que pode refletir diretamente no desempenho do classificador. Neste trabalho, a segmentação foi realizada manualmente para cada amostra, semelhante ao nosso estudo preliminar (23), selecionando as regiões com sintomas visíveis das doenças/estresse. As novas imagens são salvas no formato *JPEG* e são utilizadas na etapa seguinte para a extração das características. A Figura 3.10 apresenta exemplos de imagens de fluorescência segmentadas em 690 nm e excitação 470 nm para amostras de cancro cítrico (Figura 3.10.a), verrugose (Figura 3.10.b), greening (Figura 3.10.c), e deficiência de zinco (Figura 3.10.d) do experimento conduzido no Brasil.

Figura 3.10 - Exemplos de imagens de fluorescência segmentadas para amostras de (a) cancro cítrico, (b) verrugose, (c) greening e (d) deficiência de zinco.

Fonte: Elaborada pelo autor.

3.2.3.2 Extração das características

A etapa seguinte consiste na extração das características de textura das imagens segmentadas. As características de textura foram extraídas com base na matriz de co-ocorrência de níveis de cinza (GLCM, do inglês *Gray Level Co-occurrence Matrix*).^(69,70) Trata-se de uma matriz que caracteriza as texturas das imagens através de um conjunto de estatísticas para as ocorrências de cada nível de cinza em pixels diferentes ao longo de direções distintas. A matriz GLCM representa a distribuição das intensidades e também, fornece informações relevantes sobre as posições relativas dos pixels vizinhos de uma imagem.⁽⁷¹⁾ Tal matriz considera a relação entre dois pixels por vez, um denominado de “pixel de referência” e o outro de “pixel vizinho”.⁽⁷²⁾ Para uma dada imagem I , de tamanho $N \times N$, a matriz GLCM pode ser definida como ⁽⁷¹⁾:

$$P(i, j) = \sum_{x=1}^N \sum_{y=1}^N \begin{cases} 1, & \text{se } I(x, y) = i \text{ e } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{caso contrário} \end{cases} \quad (3)$$

onde o deslocamento $(\Delta x, \Delta y)$ especifica a distância entre o pixel de interesse e seu vizinho. A parametrização deste “*offset*” torna a matriz de co-ocorrência sensível à rotação. O deslocamento $(\Delta x, \Delta y)$ pode assumir valor igual a 1 ou maior e é representado pelo símbolo Δ quando $\Delta x = \Delta y$ ⁽⁷¹⁾. Cada elemento da matriz P especifica o número de vezes que o pixel com valor i ocorreu de acordo com um determinado ângulo de rotação em relação a um pixel com valor j . O Apêndice B ilustra o processo de obtenção da matriz GLCM a partir de uma imagem qualquer a fim de facilitar sua compreensão.

Em 1973, Haralick e colaboradores ⁽⁶⁹⁾ mostraram que 14 características podem ser calculadas a partir da matriz GLCM. No caso, primeiramente determinam-se as matrizes de co-ocorrência correspondentes aos diferentes ângulos de rotação e, em seguida, calcula-se o valor da característica com relação aos respectivos ângulos e realiza-se a média destes valores, obtendo assim o valor final da métrica de interesse. Seis características de textura são consideradas as mais relevantes entre as 14 inicialmente propostas: energia, contraste, variância, correlação, entropia e homogeneidade.⁽⁷³⁾ Neste trabalho, o valor do parâmetro de deslocamento Δ foi definido como sendo igual a 1 e selecionamos três características de textura entre as seis mais relevantes. As características extraídas foram contraste, energia e homogeneidade, e foram calculadas segundo os ângulos definidos por Haralick e colaboradores.

Os três descritores estatísticos selecionados foram calculados de acordo com as Equações (4) - (6):

$$\text{Energia} = \sum_{i,j} (p(i,j))^2 \quad (4)$$

$$\text{Constraste} = \sum_{i,j} |i-j|^2 p(i,j) \quad (5)$$

$$\text{Homogeneidade} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (6)$$

onde $p(i,j)$ representa a matriz de co-ocorrência normalizada. Esta matriz normalizada foi obtida dividindo-se cada elemento da matriz GLCM pela somatória dos seus valores. As métricas estatísticas são calculadas automaticamente pelo software Matlab[®] através do “*toolbox*” de processamento de imagens. Mais detalhes a respeito da matriz GLCM e dos descritores estatísticos podem ser encontrados nas referências (69,71,73).

3.2.3.3 Classificação

As características extraídas das imagens segmentadas foram utilizadas como vetor de entrada nos classificadores SVM e ANN. Primeiramente será apresentada uma breve introdução a respeito do método SVM e, em seguida, serão discutidos os principais conceitos sobre redes neurais (ANN).

A técnica SVM, proposta por Vapnik (62), é baseada no princípio da minimização do risco estrutural (SRM, do inglês *Structural Risk Minimization*) proveniente da teoria de aprendizado estatístico a fim de evitar o problema do super ajuste (“*overfitting*”).(62,68) Essa teoria estabelece uma série de princípios que devem ser seguidos para a obtenção de classificadores com boa generalização, os quais são definidos conforme a sua capacidade de prever corretamente a classe dos novos dados no domínio em que o aprendizado ocorreu.(62) As SVMs foram originalmente propostas para a classificação de problemas de duas classes, mas algumas estratégias vêm sendo adotadas a fim de resolver problemas multiclases, tais como, um contra todos, um contra um, etc.(61,74) Neste trabalho, utilizamos o processo de classificação binário do SVM, onde dada uma entrada este a classifica em duas possíveis classes distintas.

No processo de classificação binária supomos que x_i é um elemento do conjunto de dados de treinamento composto por n amostras e $y_i \in \{-1, +1\}$ é o rótulo da classe associada

com x_i , onde cada elemento é atribuído à classe +1 ou -1. O objetivo do algoritmo SVM é criar um hiperplano ideal entre os dois conjuntos de dados, maximizando a distância entre a margem e o ponto dos dados mais próximos.(61,75) Os pontos mais próximos dos dados em relação ao hiperplano são chamados de vetores de suporte e contêm todas as informações necessárias ao classificador.(61,74) O hiperplano ideal pode ser determinado resolvendo-se o seguinte problema de otimização:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \quad (7)$$

devendo satisfazer a condição $y_i [(\omega^t x_i) + b] \geq 1 - \xi_i$ com $\xi_i \geq 0$; onde o parâmetro ω é ortogonal ao hiperplano, b é deslocamento a partir da origem do hiperplano, ξ_i representa o grau de classificação incorreta de x_i e C é o parâmetro de penalidade que controla o “*trade-off*” entre o número de amostras classificadas erroneamente e a complexidade do modelo.(74,76)

No caso em que os conjuntos de dados não possam ser linearmente classificados no espaço de entrada, este espaço é mapeado para um espaço de características de dimensão maior onde os dados sejam linearmente separáveis.(61) Esta transformação para um espaço de maior dimensão é implementada através de uma função kernel. A Figura 3.11 ilustra a ideia desta transformação para um espaço de maior dimensão evidenciando a separação dos dois diferentes conjuntos de dados pelo hiperplano. Existem várias funções kernel que podem ser empregadas nos casos em que os dados não são linearmente separáveis, mas as frequentemente utilizadas são: função polinomial, função de base radial Gaussiana (RBF) (do inglês, *Gaussian Radial Basis Kernel Function*), e função sigmóide.(23,61,74) O desempenho do classificador SVM está relacionado diretamente com a função kernel utilizada.(61) No presente estudo, optamos por utilizar a função RBF pois trata-se de uma função simples, rápida computacionalmente, e pode obter melhores resultados na classificação em relação às demais funções. A função RBF é apresentada na Equação (8):

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right) \quad (8)$$

onde x e x' representam os dados no espaço original e γ controla a largura da função RBF.

Figura 3.11 - Separação de um conjunto de dados pelo hiperplano em um espaço de maior dimensão implementada através da função kernel.
Fonte: Elaborada pelo autor.

O desempenho do SVM também pode ser fortemente afetado tanto pelo parâmetro C (77), uma vez que determina quão severamente os erros de classificação devem ser penalizados, quanto pelo parâmetro γ que determina a largura da Gaussiana.(75) Desta forma, é necessário algum método para a seleção do melhor par (C, γ) de modo que o SVM obtenha a máxima precisão na classificação dos dados desconhecidos. A técnica de validação cruzada é geralmente usada a fim de definir estes parâmetros.(61,77-78) Esta técnica consiste em dividir o conjunto de treinamento em n subconjuntos de tamanhos iguais e testar um subconjunto utilizando o classificador treinado com os $n-1$ subconjuntos restantes.(79) Vários pares de (C, γ) são testados e o com melhor precisão na validação cruzada é selecionado. O procedimento de validação cruzada pode evitar o problema do super ajuste. Sendo assim, utilizamos esta técnica para estimar os valores ideais destes parâmetros. Uma vez definido o modelo de classificação do SVM através do cálculo destes parâmetros no processo de treinamento, este é aplicado em um conjunto de testes e os dados são classificados com base em qual lado do plano cada ponto se localiza.

Com relação ao segundo classificador utilizado neste trabalho, a rede neural artificial é um dos métodos mais comuns de aprendizado e classificação.(57,80) Trata-se de uma técnica de aprendizado assistido que reproduz o funcionamento do cérebro humano (80) capaz de capturar características mais complexas dos dados.(81) A principal vantagem desta técnica está relacionada com fato de que as ANNs não necessitam de informações detalhadas acerca

dos processos físicos do sistema a ser modelado. A arquitetura das redes neurais é basicamente caracterizada por um conjunto de neurônios interconectados, também conhecidos como nós, e conexões entre o vetor de entrada e o de saída.(82) O número de conexões torna a ANN capaz de encontrar as relações mais importantes entre as variáveis dos dados de treinamento.(80) Tais relações são utilizadas para criar o modelo, o qual em seguida é aplicado no conjunto de dados de teste para sua validação.

As redes neurais podem ser agrupadas em duas categorias conforme suas arquiteturas: (1) redes *feed-forward*, que não possuem loops, e (2) redes *feedback*, onde os loops ocorrem devido às conexões de realimentação (81). Redes neurais típicas que compõem a categoria *feed-forward* são as SLP (do inglês, *Single-Layer Perceptron*), MLP (do inglês, *Multilayer Perceptron*) e as redes de função de base radial. A categoria *feedback* é composta principalmente pelas redes competitivas, rede SOM (do inglês, *Self-Organizing Maps*) do Kohonen, rede Hopfield e pelo modelo ART (do inglês, *Adaptive Resonance Theory*). Maiores detalhes sobre estas redes podem ser encontradas nas seguintes referências (81,82). Neste trabalho, optamos por utilizar a rede MLP, uma vez que Gardner e Dorling demonstraram ser uma alternativa eficaz comparada às técnicas tradicionais (82). Além disso, o gasto computacional é baixo, pois a resposta das redes para uma determinada entrada não depende do seu estado anterior.(81)

A rede MLP é constituída por um sistema de neurônios interconectados organizados em camadas que possuem conexões unidirecionais entre elas, assim como apresentado na Figura 3.12. Em geral, a arquitetura de uma rede MLP é constituída por três camadas divididas da seguinte forma: camada de entrada, camada oculta e camada de saída. As entradas e saídas podem ser representadas por simples vetores, e a rede pode ter uma ou mais camadas ocultas.

A informação é inserida no sistema a partir dos vetores $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$, onde cada vetor representa uma característica específica extraída das imagens. Essa informação é processada e transferida para a primeira camada oculta da rede por meio da seguinte função matemática:

$$S_i = \sum_1^N w_{ni} x_n + b_i \quad (9)$$

onde N é o número de características de entrada, w_{ni} é o peso entre os neurônios na camada de entrada n e os neurônios na camada oculta i , e b_i corresponde aos desvios associados aos neurônios na camada oculta.(80) O peso indica o quão importante são as conexões entre os neurônios.

Figura 3.12 - Arquitetura de uma rede MLP com duas camadas ocultas.
Fonte: Elaborada pelo autor.

Na etapa seguinte, os valores obtidos são utilizados por uma função de transferência, denominada função de ativação, a fim de fornecer um valor de saída (y_i) na camada de saída.(80) Há diversos tipos de funções de ativação que podem ser empregadas, tais como: (a) função limiar (ou degrau), (b) função rampa, (c) função sigmoidal e (d) função Gaussiana. A Figura 3.13 apresenta o comportamento desses quatro tipos de funções.

Figura 3.13 - Diferentes tipos de funções de ativação: (a) limiar, (b) rampa, (c) sigmoidal e (d) Gaussiana.
Fonte: Elaborada pelo autor.

Neste estudo, utilizamos apenas a função sigmoideal, representada pela Equação (10), uma vez que é a função mais utilizada em redes neurais e também possibilita a formação de limites de decisão mais suaves em relação às demais funções.(81)

$$y_i = \frac{1}{1 + e^{-s_i}} \quad (10)$$

As redes MLP utilizam a técnica de aprendizado supervisionado para realizar seu treinamento. O algoritmo utilizado foi o de treinamento retro propagado (do inglês, *backpropagation training*) pois é o mais simples computacionalmente para o treinamento de redes multicamadas.(82) Trata-se de um algoritmo que minimiza a função desvio quadrático (E) representada pela Equação (11):

$$E = \frac{1}{2} \sum_{i=1}^N (d_i - y_i)^2 \quad (11)$$

onde d_i representa o valor experimental.

De forma sucinta, os valores de saída gerados são comparados com os valores experimentais e o algoritmo de retropropagação calcula o gradiente local da superfície de erro. Em seguida, o erro é retro propagado através da rede e, a partir desta informação, o algoritmo ajusta o peso de cada conexão de modo a minimizar o erro total. Maiores detalhes e um aprofundamento matemático sobre rede neurais pode ser encontrado no livro de CHRISTOPHER, M. B.(83)

Neste trabalho, os classificadores SVM e ANN foram aplicados a dois grupos de imagens distintos no caso dos dados coletados no Brasil. O grupo 1 compreende imagens de amostras contaminadas com cancro cítrico e verrugose; o grupo 2 compreende imagens de amostras contaminadas com greening e deficiência de zinco. Esta separação se deve ao fato de que a diferenciação visual da doença tanto em amostras com cancro cítrico e verrugose, como também greening e deficiência de zinco é difícil. Além disso, aplicamos as duas técnicas na classificação das amostras do CREC/Flórida. No caso, as técnicas foram aplicadas somente na classificação de amostras de greening e deficiência de zinco, pois o CREC não possuía amostras contaminadas com verrugose na época da coleta das imagens de fluorescência.

Os processos de análise e classificação das amostras a partir do SVM foram realizados utilizando-se a biblioteca LIBSVM, compatível com o software Matlab®. A LIBSVM se encontra disponível para download na referência.(84) Com relação à técnica ANN, foi utilizado o “*toolbox*” de redes neurais do Matlab® para construir e treinar a rede e, em seguida, para a classificação das amostras.

4 ESTUDO PRELIMINAR ENVOLVENDO O USO DA TÉCNICA SVM

Este capítulo é dedicado à apresentação dos resultados obtidos a partir de um estudo preliminar utilizando a técnica de classificação SVM. Inicialmente será realizada uma breve introdução apresentando o contexto envolvido nesse estudo e, em seguida, os detalhes da coleta e processamento das imagens, assim como os respectivos resultados. O trabalho foi publicado na revista *Journal of Spectroscopy* no ano de 2013 e encontra-se disponível para consulta na referência (23).

4.1 Contextualização da pesquisa

O método SVM foi inicialmente empregado pelo nosso grupo em um estudo comparativo com o objetivo de avaliar o potencial da técnica de espectroscopia por imagens de fluorescência na diferenciação de amostras contaminadas com greening e sadias em duas regiões produtoras com características distintas, Brasil e Flórida. Esse trabalho foi nosso primeiro contato com esta técnica de classificação.

No entanto, a complexidade envolvida no desenvolvimento de algoritmos de reconhecimento de padrões e de classificação, assim como nosso pouco conhecimento nessa área de processamento na época deste trabalho, nos levou a estabelecer uma colaboração científica, em meados de 2011, com o grupo de pesquisa do Prof. Reza Ehsani do CREC (*Citrus Research and Education Center*) da Universidade da Flórida, Estados Unidos.

O CREC é o maior centro de pesquisa do mundo dedicado ao estudo de um único produto, o citros. As pesquisas no CREC envolvem desde o estudo das deficiências de nutrientes em plantas até o desenvolvimento de novas espécies resistentes a doenças e equipamentos para a área de agricultura de precisão.(85) Com relação ao grupo do Prof. Reza Ehsani, este se destaca pela pesquisa nas áreas de automação de máquinas de colheita e agricultura de precisão para a produção de citros, com foco principalmente no desenvolvimento de veículos aéreos não-tripulados para a detecção em tempo real da saúde das plantas e detecção de doenças.

A partir da excelência desse centro de pesquisa no estudo de citros e também do interesse do Prof. Reza Ehsani em avaliar a eficácia da técnica de espectroscopia por imagens, a colaboração foi estabelecida de forma que o nosso grupo ficou responsável por fornecer as imagens coletadas no Brasil, bem como empregar o sistema desenvolvido para a coleta de imagens de amostras do CREC. O grupo do Prof. Reza Ehsani ficou responsável pelo

processamento e análise destas imagens. A rotina de classificação das amostras foi desenvolvida pelo seu então aluno de doutorado Ratnesh Kumar. A seguir serão apresentados os procedimentos envolvidos na coleta e processamento das imagens.

4.2 Procedimentos experimentais

O sistema de imagens de fluorescência utilizado foi o SIF-2. Como já mencionado, este sistema utiliza filtros ópticos passa-banda para a seleção dos comprimentos de onda. Na época da coleta das imagens não havíamos ainda desenvolvido o SIF-1 para a varredura dos comprimentos de onda e seleção das melhores bandas espectrais para o uso no SIF-2. No entanto, os filtros ópticos foram selecionados com base no trabalho do aluno de doutorado Lins, E. C. C. (45) Os comprimentos de onda utilizados foram: 570, 610, 690 e 740 nm (modelos FB570-10, FB610-10, FB690-10 e FB740-10, Thorlabs, USA). Outra diferença nos experimentos foi o uso de LEDs em 365 e 530 nm, além dos LEDs em 405 e 470 nm, para a excitação das amostras.

As imagens de fluorescência foram coletadas para amostras de folhas de citros contaminadas com greening no Brasil e nos Estados Unidos. No Brasil, 20 árvores de citros da espécie *Citrus sinensis* (“Hamlin”) de quatro municípios do estado de São Paulo foram avaliadas. As imagens foram coletadas para folhas sintomáticas e assintomáticas para cada planta, totalizando 100 amostras sintomáticas para o greening e 55 amostras assintomáticas (confirmadas como sadias pelo método PCR). A coleta das imagens foi realizada entre Junho e Julho de 2009 diretamente nas fazendas, uma vez que nosso sistema é portátil. Após as medidas, as folhas foram transportadas em caixas de isopor fechadas para o Fundecitrus para a confirmação do greening através do método PCR. Nos Estados Unidos, as folhas foram coletadas de 8 árvores de citros da mesma espécie do Brasil no CREC entre 14 e 17 de Julho de 2011, totalizando 68 amostras contaminadas com greening e 17 sadias. A quantidade de amostras nos Estados Unidos é menor devido ao pouco tempo que tivemos disponível para a coleta das imagens. Este estudo foi conduzido a fim de avaliar a aplicabilidade da técnica de espectroscopia por imagens de fluorescência em diferentes condições.

A análise das imagens de fluorescência foi realizada através do software Matlab[®] e envolveu as três etapas já descritas na seção 3.2.3: segmentação, extração das características e classificação. A segmentação das imagens foi realizada manualmente para a seleção das regiões de interesse de cada amostra. A partir das imagens segmentadas, características de textura como energia, contraste, correlação e homogeneidade foram extraídas com base na

matriz de co-ocorrência. No trabalho de doutorado resolvemos não utilizar a característica de correlação, uma vez que nossos resultados se mostraram melhores sem esta. Por fim, na etapa de classificação as quatro características extraídas das imagens foram utilizadas como vetor de entrada no classificador SVM. Neste artigo também optamos por utilizar a função kernel RBF, mas nenhum procedimento foi adotado a fim de selecionar o melhor par (C, γ) de modo a obter a máxima precisão possível na classificação dos dados desconhecidos. No caso, foram utilizados os valores padrão da biblioteca LIBSVM.(84)

Inicialmente, o SVM foi aplicado em um conjunto de dados de treinamento constituído pelas características extraídas de 20 amostras, sendo 10 amostras sadias e 10 contaminadas com greening. A partir do modelo gerado, este foi aplicado no conjunto de teste composto pelas características extraídas das demais amostras. O conjunto de teste do Brasil foi composto pelas características extraídas de 135 amostras, enquanto que dos Estados Unidos por 65 amostras. Este procedimento foi realizado separadamente para as imagens de fluorescência de cada país.

O desempenho do classificador SVM foi avaliado segundo o parâmetro acurácia, e as curvas ROC (do inglês, *Receiver Operating Characteristic*) e PR (do inglês, *Precision Recall*). A acurácia está relacionada com a capacidade da técnica SVM em fornecer resultados corretos na diferenciação entre amostras sadias e contaminadas com greening quando comparada ao teste PCR, considerado como padrão-ouro. Em relação à curva ROC, esta ilustra o desempenho do classificador binário a partir de um gráfico da taxa de verdadeiros positivos (VP) no eixo y contra a taxa de falsos positivos (FP) no eixo x. A probabilidade de ocorrência de resultados verdadeiros positivos pode ser expressa como a sensibilidade do teste, enquanto que a probabilidade de resultados falsos positivos pode ser expressa como 1-especificidade.(86) No que diz respeito ao desempenho do classificador, a curva ROC mais próxima do canto superior esquerdo fornece o maior poder de discriminação dos dados.(87) As curvas consideradas boas estão entre a linha diagonal e a linha perfeita, onde quanto maior a distância da linha diagonal, melhor a classificação. A linha diagonal indica uma classificação aleatória, ou seja, um sistema que seleciona aleatoriamente saídas como positivas ou negativas.(86) Pontos acima da linha diagonal representam bons resultados na classificação, enquanto que pontos abaixo representam resultados ruins. A Figura 4.1 apresenta exemplos de curvas ROC para três graus diferentes segundo a capacidade de discriminação. Neste tipo de representação, a área sob a curva define a probabilidade do teste em detectar os verdadeiros positivos e os verdadeiros negativos, ou seja, expressa a acurácia da técnica.(88)

Figura 4.1 - Curvas ROC representativas de três graus de capacidade de discriminação.
Fonte: SOUZA (86).

Por fim, as curvas PR são geralmente utilizadas na classificação binária para avaliar o classificador e são representadas por um gráfico da precisão em função do “*recall*”. A precisão é uma medida da relevância dos resultados, enquanto o “*recall*” representa a medida de quantos resultados realmente relevantes são retornados.(89) Um alto valor na área da curva PR representa tanto uma precisão alta como também um “*recall*” alto, onde a alta precisão está relacionada com a baixa taxa de falsos positivos e o alto “*recall*” com uma baixa taxa de falsos negativos. Altos valores para ambos os parâmetros mostram que o classificador está retornando resultados precisos (alta precisão), bem como retornando a maioria de todos os resultados positivos (alto “*recall*”).(89) A Figura 4.2 apresenta exemplos de curvas PR para dois algoritmos quaisquer.(90) O algoritmo pode ser selecionado de acordo com um dos seguintes requisitos: alta precisão com baixo “*recall*” ou alto “*recall*” mas com baixa precisão.

Figura 4.2 - Curvas PR para dois algoritmos diferentes.
Fonte: DIFFERENCES...(90)

4.3 Resultados e discussões

O pré-processamento das imagens neste trabalho consistiu em avaliar visualmente as imagens de fluorescência e determinar a banda espectral com maior potencial de discriminação da doença. A partir dessa avaliação preliminar, o comprimento de onda selecionado foi o 690 nm para a excitação em 530 nm. A fluorescência medida nesse comprimento de onda representa a emissão das moléculas de clorofila *a*, e foi útil na discriminação do greening, uma vez que está relacionada diretamente com o estado fisiológico das plantas.(44, 91) A Figura 4.3 apresenta imagens de fluorescência de uma amostra sadia e contaminada com greening para a banda espectral selecionada. A partir da figura, pode-se observar que a textura das amostras sadias e com greening são diferentes; a amostra sadia tem uma textura mais uniforme em relação à amostra infectada com greening.

Figura 4.3 - Imagens de fluorescência em 690 nm para amostras (a) sadia e (b) contaminada com greening do Brasil.

Fonte: WETTERICH (23).

Após a seleção da banda espectral, as respectivas imagens de fluorescência foram segmentadas e suas características extraídas. Os descritores de textura foram utilizados como vetor de entrada no classificador SVM. Os resultados da classificação das amostras são apresentados na Tabela 4.1.(23) Os valores de acurácia obtidos pelo SVM foram de 90% e 61% na classificação dos dados do Brasil e EUA, respectivamente. Os resultados obtidos na classificação das amostras do Brasil foram significativamente superiores não só pelo alto valor da acurácia, mas também pela baixa taxa de falsos negativos (9%) comparado aos EUA (40%). Na seleção de um classificador é mais importante ter uma baixa taxa de falsos negativos do que taxas baixas de falsos positivos, pois não é interessante que amostras contaminadas com greening sejam classificadas erroneamente como sadias.

Tabela 4.1 - Resultados da classificação das imagens de fluorescência pelo classificador SVM.

Amostras	Falsos positivos	Falsos negativos	Acurácia
Sadias e com greening do Brasil	12%	9%	90%
Sadias e com greening dos EUA	35%	40%	61%

Fonte: WETTERICH (23).

Os resultados da classificação das amostras pelo método SVM também foram avaliados segundo as curvas ROC e PR. Ambas as curvas são apresentadas na Figura 4.4 para os dados do Brasil e EUA. A partir dos gráficos é possível comprovar o melhor desempenho do classificador SVM na diferenciação do greening nas amostras do Brasil. Comparando inicialmente as curvas PR obtidas, observa-se que a área sob a curva dos dados do Brasil (Figura 4.4.a) é significativamente maior que a dos EUA (Figura 4.4.c), o que evidencia a

baixa taxa de falsos positivos e falsos negativos da classificação. Os valores menores das taxas também podem ser observados na Tabela 4.1. Com relação às curvas ROC, é possível notar que a curva ROC dos dados do Brasil (Figura 4.4.b) está mais distante da linha diagonal que a curva dos Estados Unidos (Figura 4.4.d). Além disto, a área maior sob a curva dos dados do Brasil expressa que a acurácia da técnica é maior, comprovando assim os resultados apresentados na Tabela 4.1. Estes resultados demonstram o potencial da técnica em discriminar as amostras do Brasil.

Figura 4.4 - Curvas PR e ROC para a classificação dos dados do Brasil (a) e (b), e dos Estados Unidos (c) e (d), respectivamente.

Fonte: Elaborada pelo autor.

A diferença nas acurácias da classificação pelo SVM está provavelmente relacionada com as intensidades da fluorescência emitidas pelas amostras. Ao comparar as intensidades emitidas pelas amostras de cada país verificou-se que a intensidade emitida pelas folhas dos EUA era cerca de 20% menor em relação à intensidade emitida pelas amostras do Brasil, sob as mesmas condições de iluminação e detecção. A possível razão para essa observação pode ser a variação da fisiologia da planta resultado das diferenças do solo das duas regiões,

Flórida e São Paulo. Devido à baixa intensidade da fluorescência emitida pelas amostras dos EUA, a acurácia da classificação foi muito menor (61%) comparada aos resultados do Brasil. Outra razão para a baixa acurácia pode estar relacionada com o número reduzido de amostras da Flórida.

Os resultados indicaram que a aplicação da técnica de espectroscopia por imagens de fluorescência em conjunto com o classificador SVM apresenta uma precisão de até 90% na discriminação das amostras contaminadas e sadias do Brasil. Embora as intensidades de fluorescência das folhas coletadas no Brasil e EUA fossem diferentes, este método apresenta um elevado potencial para a detecção do greening.

Além dos resultados obtidos, o desenvolvimento deste trabalho em conjunto com o grupo do Prof. Reza Ehsani nos propiciou o primeiro contato com a técnica de classificação SVM e forneceu a base necessária para o uso desse método no decorrer deste projeto de doutorado. Vale ressaltar também nosso contato com os processos de segmentação e extração das características, etapas consideradas fundamentais no processamento de imagens.

No doutorado avançamos significativamente em relação ao primeiro contato com a técnica SVM, uma vez que fomos capazes de otimizar o algoritmo de classificação com o objetivo de determinar os melhores parâmetros do kernel utilizado e consequentemente aumentar a acurácia dos nossos dados. Entretanto, o conhecimento adquirido foi imprescindível para o delineamento dos processos envolvidos na classificação, assim como permitiu buscarmos outras técnicas de análise de imagens. No caso, a técnica escolhida foi as redes neurais artificiais devido ao seu potencial em capturar características mais complexas dos dados.

5 DISCRIMINAÇÃO DAS PRINCIPAIS DOENÇAS/ESTRESSE QUE AFETAM A CITRICULTURA NO BRASIL

Este capítulo é dedicado à apresentação dos experimentos e resultados envolvidos na discriminação das principais doenças que afetam a citricultura brasileira. Inicialmente, apresentaremos o experimento no qual utilizamos o SIF-1 para a seleção das principais bandas espectrais das doenças em estudo, assim como os resultados obtidos. Em seguida, será apresentado o experimento utilizando o SIF-2 e os resultados referentes à classificação das imagens através das técnicas SVM e ANN.

5.1 Experimento exploratório utilizando o SIF-1

O trabalho se concentrou primeiramente na coleta das imagens de fluorescência a partir do filtro óptico sintonizável a fim de selecionar as principais bandas espectrais para a discriminação das doenças. No caso, quando menciona-se banda espectral no decorrer da apresentação dos resultados estamos nos referindo à seleção e uso de filtros ópticos para comprimentos de onda específicos. Nesta etapa, imagens de fluorescência foram coletadas utilizando o SIF-1 para as duas excitações disponíveis (405 e 470 nm) e para 5 amostras das seguintes doenças/deficiências: cancro cítrico, verrugose, greening e deficiência de zinco. As amostras foram fornecidas por técnicos do Fundecitrus e submetidas a exames laboratoriais convencionais para o diagnóstico e confirmação das doenças conforme descrito no tópico 3.2.1.

As imagens foram obtidas para comprimentos de onda a partir de 480 nm para as duas excitações. Uma vez que o filtro variável permite a escolha de três diferentes larguras de banda, resolvemos trabalhar com a faixa estreita nesta região do espectro (480 a 750 nm) a fim de obter imagens mais precisas e menos ruidosas. Vale salientar que neste intervalo a largura de banda varia de 9 nm (em 480 nm) até 19 nm (em 750 nm). Após a coleta das imagens de fluorescência das folhas contaminadas pelo SIF-1, estas foram pré-processadas conforme descrito no tópico 3.2.2 para a eliminação da informação espectral tanto da câmera CCD como do filtro variável, assim como do tempo de exposição da câmera.

Com as imagens normalizadas após o pré-processamento, foi possível comparar os grupos de doenças e consequentemente determinar as melhores bandas espectrais. Como já mencionado, dividimos as doenças em dois grupos, grupo 1 e grupo 2, segundo o critério de doenças que apresentem sintomas visuais semelhantes. No caso, o grupo 1 é constituído por

imagens de amostras contaminadas com cancro cítrico e verrugose, enquanto que o grupo 2 é composto por imagens de greening e deficiência de zinco.

De modo a selecionar as principais bandas espectrais, analisamos a intensidade média da fluorescência emitida em função do comprimento de onda das imagens pré-processadas de cada doença. Foi desenvolvida uma rotina no Matlab[®] que utiliza o “*toolbox*” de Processamento de Imagens para a seleção da região de interesse (ROI, do inglês *Region of Interest*) das imagens de cada amostra e o respectivo cálculo da média da intensidade dos pixels destas regiões. Um ROI quadrado de dimensões 20 x 20 mm foi selecionado aleatoriamente em cada amostra segundo o critério de que a região definida deve conter sintomas da doença. Uma vez selecionado o ROI, uma nova imagem é gerada sendo esta denominada de máscara binária. A máscara binária possui o mesmo tamanho que a imagem original, mas com valores de pixels iguais a 1 na região interna do ROI e iguais a 0 na região externa. A partir da multiplicação da imagem original com a máscara, obtêm-se a imagem final da região selecionada. A Figura 5.1.a e 5.1.b apresenta a imagem de fluorescência de uma amostra de cancro cítrico destacando o ROI selecionado e a imagem resultante após a aplicação da máscara binária, respectivamente.

Figura 5.1 - (a) Imagem de fluorescência em 690 nm utilizando a excitação em 470 nm para uma amostra de cancro cítrico destacando o ROI selecionado. (b) Imagem resultante após a aplicação da máscara binária gerada pelo ROI na imagem original.

Fonte: Elaborada pelo autor.

A partir da máscara binária criada para uma determinada amostra, comprimento de onda e excitação, esta é aplicada nas imagens dos demais comprimentos de onda e excitação. Há a necessidade de se criar outra máscara somente quando se altera a amostra. Com as imagens resultantes de cada comprimento de onda para as excitações em 405 e 470 nm, a rotina realiza o cálculo da média das intensidades dos pixels dessas imagens. As Figuras 5.2.a

e 5.2.b apresentam os gráficos da intensidade em função do comprimento de onda para amostras de cancro cítrico e verrugose utilizando as excitações em 405 e 470 nm, respectivamente, enquanto que as Figuras 5.2.c e 5.2.d se referem às amostras contaminadas com greening e deficiência de zinco.

Figura 5.2 - Gráficos da intensidade em função do comprimento de onda comparando amostras de cancro cítrico e verrugose para excitações em (a) 405 nm e (b) 470 nm, além de amostras de greening e deficiência de zinco em (c) 405 nm e (d) 470 nm.

Fonte: Elaborada pelo autor.

A partir da Figura 5.2, foi possível determinar as principais bandas de fluorescência para o processo de discriminação das doenças. As bandas foram selecionadas com base nas regiões onde é possível observar uma distinção nos espectros das doenças/estresse. Comparando inicialmente o cancro cítrico e a verrugose, selecionamos as bandas de fluorescência em 550 e 580 nm para a excitação em 405 nm, bem como a banda em 550 nm para a excitação em 470 nm, como pode ser observado nas Figuras 5.2.a e 5.2.b. Também decidimos utilizar a banda em 570 nm para a excitação em 470 nm. Embora esta banda não demonstre uma clara distinção entre as doenças, ela já havia sido utilizada pelo nosso grupo

em um trabalho anterior para a detecção de cancro cítrico utilizando a técnica de espectroscopia de fluorescência induzida por laser.(35)

No caso da comparação entre o greening e a deficiência de zinco, selecionamos as bandas em 560 e 580 nm para a excitação em 405 nm, assim como as bandas em 530 e 550 nm para a excitação em 470 nm conforme as regiões de máxima intensidade observadas nas Figuras 5.2.c e 5.2.d. Além destes comprimentos de onda selecionados, também consideramos as imagens de fluorescência em 690 nm para todas as excitações e doenças. O motivo que nos levou a adicionar este comprimento de onda se deve ao fato que a fluorescência em 690 nm representa a emissão da molécula de clorofila *a*, e seu estudo pode ser útil para a discriminação das doenças uma vez que está diretamente relacionado com o estado fisiológico das plantas.(23,44,91) A tabela 5.1 apresenta as bandas espectrais selecionadas para os dois grupos de doenças em estudo.

Tabela 5.1 - Bandas espectrais selecionadas para as excitações em 405 e 470 nm a partir do experimento exploratório.

Doença/Deficiência	Excitação em 405 nm	Excitação em 470 nm
Cancro cítrico e Verrugose (Grupo 1)	Filtros em 550, 580 e 690 nm	Filtros em 550, 570 e 690 nm
Greening e Deficiência de zinco (Grupo 2)	Filtros em 560, 580 e 690 nm	Filtros em 530, 550 e 690 nm

Fonte: Elaborada pelo autor.

5.2 Experimento utilizando o SIF-2

Na fase exploratória, foi possível determinar as principais bandas espectrais a partir do SIF-1. Contudo, este sistema não é viável para a coleta de imagens de fluorescência para um grande número de amostras, uma vez que o tempo para a coleta de imagens para uma única amostra é relativamente longo e também é necessário aguardar a estabilização da temperatura do filtro óptico variável. Logo, adquirimos filtros ópticos passa-banda para os comprimentos de onda apresentados na Tabela 5.1, e utilizamos o sistema SIF-2 descrito no tópico 3.1.2 para a coleta de imagens de fluorescência.

Nesta etapa, imagens de fluorescência foram coletadas para 100 amostras de cada uma das doenças/estresse que compõem os dois grupos de análise. Como já mencionado anteriormente, esta separação em dois grupos está relacionado ao fato de que a diferenciação

visual da doença tanto em amostras com cancro cítrico e verrugose, como também greening e deficiência de zinco é difícil.

As imagens de fluorescência obtidas foram pré-processadas, conforme descrito no tópico 3.2.2, considerando a porcentagem de transmissão dos filtros ópticos passa-banda e a resposta espectral da CCD. Em seguida, foi realizado o processo de análise das imagens de fluorescência, como descrito no tópico 3.2.3. Este processo é constituído de três etapas: segmentação, extração das características e classificação. As imagens de fluorescência foram segmentadas manualmente, de forma semelhante ao realizado em nosso estudo preliminar.(23) A partir das imagens segmentadas, as seguintes características de textura foram extraídas: energia, contraste e homogeneidade. As características de textura extraídas foram utilizadas como vetor de entrada nos classificadores SVM e ANN.

A classificação das imagens foi realizada mediante a aplicação dos classificadores nos conjuntos de dados compostos pelas características extraídas das 200 amostras de cada grupo. Os dois grupos foram analisados separadamente, mas o procedimento envolvido é idêntico. Os dados de cada grupo foram divididos aleatoriamente em dois subconjuntos. O primeiro subconjunto é denominado de conjunto de treinamento e foi usado para treinar os classificadores, especificamente, construir o modelo do SVM e criar a rede do ANN. Este subconjunto é composto pelas características extraídas de 20 imagens de fluorescência, sendo 10 imagens para cada doença do grupo em análise. O outro subconjunto é denominado de conjunto de teste e foi utilizado para avaliar a capacidade de predição dos classificadores. O conjunto de teste é constituído pelas características das 180 amostras restantes, 90 para cada doença. Os classificadores foram aplicados nos mesmos subconjuntos para a comparação dos resultados. Visto que estamos trabalhando com processos de classificação binários tanto pelo SVM como pelo ANN, as amostras serão classificadas como sendo positivas ou negativas para cancro cítrico para o grupo 1, e positivas e negativas para greening no caso do grupo 2.

Os desempenhos dos classificadores foram avaliados segundo três parâmetros: acurácia, sensibilidade e especificidade. Uma vez que estamos diante de um diagnóstico dicotômico (positivo ou negativo para a doença), os resultados da classificação devem apresentar um equilíbrio entre esses parâmetros. A acurácia, como já explicado anteriormente, está relacionada com a capacidade dos classificadores em fornecer resultados corretos na diferenciação entre as amostras. O parâmetro sensibilidade reflete a capacidade que a técnica utilizada apresenta em detectar as amostras verdadeiramente positivas, ou seja, de identificar corretamente as amostras doentes. No caso da especificidade, esta reflete a capacidade que a técnica possui em detectar as amostras verdadeiramente negativas, isto é, diagnosticar as

amostras que são negativas para a doença em estudo. No entanto é necessário outro teste, considerado como padrão-ouro, na diferenciação entre as amostras.(88) Neste caso, os resultados dos testes bioquímicos foram utilizados como referência para checar a validade da técnica proposta. Vale mencionar que não utilizamos as curvas ROC e PR na avaliação dos desempenhos dos classificadores como no estudo preliminar (23), pois estes três parâmetros já nos fornecem diretamente as informações que podem ser extraídas das curvas. Os parâmetros sensibilidade, especificidade e acurácia foram calculados de acordo com as Equações (12) - (14), respectivamente:

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (12)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (13)$$

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (14)$$

onde VP representa a taxa de amostras classificadas como verdadeiros positivos, VN a taxa de verdadeiros negativos, FP a taxa de falsos positivos e FN a taxa de falsos negativos. Os resultados de cada classificador são apresentados nas seções seguintes.

5.2.1 Resultados da classificação das imagens pelo SVM

A fim de obter o melhor desempenho na classificação das imagens de fluorescência pela técnica SVM, os parâmetros C e γ foram ajustados através da técnica de validação cruzada. Nesta etapa, para cada par (C, γ) , o conjunto de dados de treinamento foi dividido em 5 subconjuntos de tamanhos iguais e um desses subconjuntos foi testado utilizando o classificador treinado a partir dos demais 4 subconjuntos. Este procedimento é repetido para todos os subconjuntos constituintes do conjunto de dados de treinamento e no final calcula-se a média das acurácias. Nós adotamos a metodologia empregada nos trabalhos (78,79).

Os parâmetros C e γ foram variados no intervalo entre 2^{-30} a 2^{30} com passo de $2^{0,25}$, e o par (C, γ) com a melhor acurácia na validação cruzada dos dados de treinamento foi selecionado. Em seguida, um ajuste fino foi conduzido em torno da região desse par utilizando um passo menor. Os valores ideais dos parâmetros foram determinados conforme o resultado dessa última etapa. A Figura 5.3 ilustra os resultados do procedimento adotado para a seleção dos melhores valores dos parâmetros para amostras de cancro cítrico e verrugose em 690 nm e excitação em 405 nm.

Figura 5.3 - Mapa de contorno da otimização dos parâmetros C e γ obtidos através da validação cruzada. (a) Pesquisa abrangente das melhores regiões com as maiores acurácias. (b) Ajuste fino em torno da região selecionada. A cor vermelha representa as regiões com acurácia maior que 90%.
Fonte: Elaborada pelo autor.

A partir do mapa de contorno apresentado na Figura 5.3.a, é possível observar que duas regiões apresentam as maiores acurácias, entre 90 e 100% (cor vermelha), além da

predominância de regiões com acurácias entre 60 e 70% (cor verde), 70 e 80% (cor amarela), e 80 a 90% (cor laranja). Nesta seleção preliminar, o parâmetro C e γ varia entre $2^{-30,0}$ e $2^{-0,75}$, e γ entre $2^{11,75}$ e $2^{30,0}$ para a primeira região. Na segunda região, a variação dos parâmetros é menor, com C variando aproximadamente entre $2^{28,25}$ e $2^{30,0}$, e γ entre $2^{-15,0}$ e $2^{-13,5}$. Com a seleção automática do melhor par (C, γ) e o ajuste fino em torno desse ponto, os resultados da Figura 5.3.b indicam que as maiores acurácias da validação cruzada são obtidas para valores de C entre $2^{-30,5}$ e $2^{-29,5}$, e γ entre $2^{-11,25}$ e $2^{-11,75}$. A acurácia nessa região varia entre 90 e 91%, conforme mostrado na Figura 5.3.b. Esses resultados são relativos à análise dos dados para a banda espectral em 690 nm e excitação em 405 nm na comparação de cancro cítrico e verrugose. O procedimento adotado é repetido para cada excitação e banda espectral para os dois grupos de doenças.

Após a seleção do melhor par (C, γ) , o conjunto de dados de treinamento foi treinado novamente no sentido de construir o modelo de classificação SVM. Nesta etapa de aprendizado, o SVM representa os exemplos como pontos no espaço, mapeados de forma que os exemplos das duas categorias distintas são separados o máximo possível. Em seguida, o modelo foi aplicado no conjunto de teste que se deseja classificar e estes dados foram atribuídos a uma classe ou outra. As primeiras tentativas envolveram a utilização de apenas uma única excitação e fluorescência. Ao compararmos os resultados da classificação do SVM com os resultados do teste padrão-ouro das doenças do Grupo 1 obtivemos a Tabela 5.2, conforme as taxas de amostras classificadas como VP, VN, FP e FN.

Tabela 5.2 - Resultados da classificação do Grupo 1 segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

			Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Resultado do SVM				
Excitação em 405 nm	F550	Positivo	VP = 86	FP = 22
		Negativo	FN = 4	VN = 68
	F580	Positivo	VP = 88	FP = 39
		Negativo	FN = 2	VN = 51
	F690	Positivo	VP = 71	FP = 71
		Negativo	FN = 19	VN = 19

(continua)

(continuação)

Tabela 5.2 - Resultados da classificação do Grupo 1 segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

		Resultado do SVM	Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Excitação em 470 nm	F550	Positivo	VP = 85	FP = 38
		Negativo	FN = 5	VN = 52
	F570	Positivo	VP = 90	FP = 85
		Negativo	FN = 0	VN = 5
	F690	Positivo	VP = 55	FP = 11
		Negativo	FN = 35	VN = 79

Fonte: Elaborada pelo autor.

Com base nos resultados apresentados na Tabela 5.2, foi possível calcular o desempenho do SVM na classificação das amostras segundo os parâmetros sensibilidade, especificidade e acurácia de acordo com as Equações (12) - (14). O desempenho do SVM é apresentado na Tabela 5.3 para cada excitação e banda espectral.

Tabela 5.3 - Desempenho da discriminação de amostras de cancro cítrico e verrugose pelo método SVM.

	Excitação em 405 nm			Excitação em 470 nm		
	F550	F580	F690	F550	F570	F690
Sensibilidade	95,6%	97,8%	78,9%	94,4%	100,0%	61,1%
Especificidade	75,6%	56,7%	21,1%	57,8%	5,6%	87,8%
Acurácia	85,6%	77,2%	50,0%	76,1%	52,8%	74,4%

Fonte: Elaborada pelo autor.

A análise do desempenho do SVM de acordo com a Tabela 5.3 demonstra que os parâmetros sensibilidade e especificidade variam significativamente, com valores muito baixos para a especificidade em determinadas bandas espectrais. Além disso, as acurácias obtidas ficaram aquém do esperado na classificação das amostras. O melhor resultado ficou por conta da classificação das imagens de fluorescência obtidas para o comprimento de onda em 550 nm e excitação em 405 nm, atingindo uma acurácia relativamente boa (~86,0%) e com uma alta sensibilidade (~96,0%). No entanto, a especificidade de 76,0% é baixa para um eventual uso comercial da técnica.

Adotando o mesmo procedimento na classificação das amostras do Grupo 2, obtivemos os resultados apresentados na Tabela 5.4 conforme as taxas de amostras classificadas como VP, VN, FP e FN.

Tabela 5.4 - Resultados da classificação do Grupo 2 segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

			Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Resultado do SVM				
Excitação em 405 nm	F560	Positivo	VP = 81	FP = 93
		Negativo	FN = 0	VN = 6
	F580	Positivo	VP = 81	FP = 99
		Negativo	FN = 0	VN = 0
	F690	Positivo	VP = 79	FP = 88
		Negativo	FN = 2	VN = 11
Excitação em 470 nm	F530	Positivo	VP = 49	FP = 57
		Negativo	FN = 32	VN = 41
	F550	Positivo	VP = 81	FP = 94
		Negativo	FN = 0	VN = 5
	F690	Positivo	VP = 53	FP = 27
		Negativo	FN = 28	VN = 72

Fonte: Elaborada pelo autor.

Como já mencionado, as amostras de greening e deficiência de zinco foram submetidas ao Fundecitrus para a realização da análise por qPCR a fim de confirmar a presença da doença e da deficiência de zinco. A partir dos testes realizados, 9 amostras que foram classificadas visualmente como greening pelos técnicos são negativas para a doença e foram contabilizadas em nossas estatísticas na elaboração da Tabela 5.4. Assim, podemos calcular os parâmetros sensibilidade, especificidade e acurácia. Os resultados do desempenho do SVM na classificação das amostras do Grupo 2 são apresentados na Tabela 5.5.

Tabela 5.5 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método SVM.

	Excitação em 405 nm			Excitação em 470 nm		
	F560	F580	F690	F530	F550	F690
Sensibilidade	100,0%	100%	97,5%	60,5%	100,0%	65,4%
Especificidade	6,1%	0%	11,1%	41,8%	5,1%	72,7%
Acurácia	48,3%	45,0%	50,0%	50,3%	47,8%	69,4%

Fonte: Elaborada pelo autor.

Os resultados mostram novamente que os parâmetros sensibilidade e especificidade variam significativamente apresentando baixos valores de acordo com a excitação e banda espectral selecionada. No caso das imagens de fluorescência em 580 nm para a excitação em 405 nm, observa-se que a sensibilidade alcançada foi de 100%, ou seja, a técnica consegue discriminar as amostras com greening. Porém, a especificidade de 0% obtida indica que a técnica não reconheceu as amostras com deficiência de zinco. Portanto, para essa determinada excitação e fluorescência a técnica não é útil uma vez que não consegue discriminar nada. Além disso, as acurácias também foram baixas variando desde 48% (fluorescência em 560 nm e excitação em 405 nm) até aproximadamente 70% (fluorescência em 690 e excitação em 470 nm).

De maneira geral, os resultados obtidos nessa primeira tentativa de classificação dos dois grupos de amostras pelo método SVM se mostraram relativamente baixos para um possível uso do teste na discriminação das doenças. Um bom teste de classificação deve resultar em valores altos para os três parâmetros analisados. No caso, estes resultados são relativos à classificação do SVM considerando 20 amostras de treinamento, sendo 10 amostras de cada doença, e 180 amostras de teste. No entanto, não sabemos como se comporta a acurácia da técnica em função do número de amostras de treinamento. Desta forma, resolvemos verificar se o aumento do número de amostras de treinamento implicaria consequentemente em uma melhora da precisão do classificador.

Inicialmente selecionamos o melhor resultado obtido para cada grupo de doenças e mantivemos fixos os parâmetros C e γ selecionados a partir da validação cruzada, variando somente o número de amostras de treinamento de 20 a 140 igualmente divididas para cada doença e, por fim, aplicamos o SVM nas amostras de teste restantes. Na comparação das amostras de cancro cítrico e verrugose foi selecionada a excitação em 405 nm e a fluorescência em 550 nm, enquanto que para a comparação das amostras de greening e deficiência de zinco foi utilizado a excitação em 470 nm e o comprimento de onda em 690

nm. A Figura 5.4 apresenta o gráfico da acurácia do SVM em função do número de amostras de treinamento segundo os critérios estabelecidos.

Figura 5.4 - Acurácia do classificador SVM em função do número de amostras de treinamento para os dois grupos em estudo.
Fonte: Elaborada pelo autor.

A partir da Figura 5.4, podemos afirmar que a acurácia do SVM não aumenta conforme o aumento do número de amostras de treinamento na etapa de aprendizagem do classificador para os dois grupos de doenças. No caso da classificação das amostras de cancro cítrico e verrugose, observa-se que o aumento do número de amostras implica diretamente no declínio da acurácia atingindo o menor valor para um total de 120 amostras de treinamento. Com relação à classificação das amostras de greening e deficiência de zinco, a acurácia também decai para todos os conjuntos de treinamento, exceto quando se utiliza um total de 60 amostras para treinar o classificador. No entanto, para esse número de amostras a acurácia é praticamente igual ao do conjunto com 20 amostras. Sendo assim, esses resultados mostram que as baixas acurácias não estão relacionadas com o número de amostras de treinamento utilizadas. Uma provável explicação para este resultado deve estar relacionado com o problema do “*overfitting*” onde o classificador SVM se especializa nos dados utilizados em seu treinamento mas apresenta uma baixa taxa de acerto quando aplicado no conjunto de testes.⁽⁹²⁾ Essa situação pode ocorrer quando os exemplos de treinamento disponíveis são pouco representativos e uma possível solução seria avaliar as demais bandas espectrais e verificar como se comporta a acurácia em função do aumento do número de amostras de

treinamento. No entanto, como já havíamos utilizado previamente este classificador no trabalho (23) com uma pequena quantidade de amostras de treinamento, resolvemos adotar como número ideal o conjunto de treinamento composto pelas características extraídas de 20 amostras, uma vez que as acurácias são máximas para esta quantia conforme apresentado na Figura 5.4.

Visto que o desempenho do classificador SVM ficou aquém do desejado na classificação das doenças dos dois grupos e também que não estava relacionado ao aumento do número de amostras de treinamento, resolvemos adotar uma nova estratégia. Essa estratégia consistiu em combinar as diferentes excitações e bandas de fluorescência. Para cada excitação, combinamos os dois comprimentos de onda com as acurácias mais elevadas, assim como as três bandas espectrais e, por fim, combinamos as duas excitações e todas as bandas de fluorescência.

Desta forma, na discriminação das doenças do Grupo 1 foram combinadas as bandas de fluorescência em 550 e 580 nm para a excitação em 405 nm, as bandas em 550 e 690 nm para a excitação em 470 nm, assim com as três bandas de cada excitação e por fim todas as bandas, conforme observado na Tabela 5.3. No caso das doenças do Grupo 2, combinamos as bandas de fluorescência em 560 e 690 nm para a excitação em 405 nm, as bandas em 530 e 690 para a excitação em 470 nm, bem como as três bandas de cada excitação e por fim todas as bandas espectrais, como observado na Tabela 5.5. As amostras foram consideradas como negativas para a doença em estudo nessas combinações quando para cada banda de fluorescência estas fossem classificadas como negativas pelo SVM; caso contrário, foram consideradas como positivas para a doença. A Tabela 5.6 apresenta os resultados da classificação para as amostras do Grupo 1 considerando essas combinações e procedimentos.

Tabela 5.6 - Desempenho do classificador SVM considerando as combinações para a detecção do cancro cítrico (Grupo 1).

	Excitação 405 nm		Excitação 470 nm		Combinação das duas excitações
	F550 + F580	F550 + F580 + F690	F550 + F690	F550 + F570 + F690	Todas as fluorescências
Sensibilidade	97,8%	97,8%	95,6%	100,0%	100,0%
Especificidade	75,6%	81,1%	93,3%	93,3%	95,6%
Acurácia	86,7%	89,4%	94,4%	96,7%	97,8%

Fonte: Elaborada pelo autor.

Segundo a Tabela 5.6, podemos verificar que a acurácia aumenta significativamente quando passamos a considerar a combinação das fluorescências, especialmente no caso da excitação em 470 nm. O melhor resultado (acurácia de 97,8%) foi obtido quando consideramos a combinação das duas excitações e todas as bandas de fluorescência. A partir dessa combinação, foi possível identificar corretamente todas as amostras positivas para cancro cítrico (sensibilidade de 100%). Além disso, as amostras negativas para cancro cítrico também foram identificadas com uma elevada precisão (especificidade de 95,6%). Vale destacar o resultado obtido utilizando a excitação em 470 nm e as duas bandas de fluorescência em 550 e 690 nm. O desempenho do classificador para essa combinação foi ligeiramente menor, mas apresenta um grande potencial uma vez que podemos utilizar somente essa excitação e os dois comprimentos de onda para identificar as doenças do Grupo 1. Isso pode economizar tempo em todo o processo. Por fim, é possível observar que a utilização da banda em 570 nm para a excitação em 470 nm não alterou significativamente a acurácia. Este resultado não foi uma surpresa visto que não havia muita diferença no espectro da Figura 5.2.b. Adotando o mesmo procedimento na classificação das amostras do Grupo 2, obtivemos os resultados apresentados na Tabela 5.7.

Tabela 5.7 - Desempenho do classificador SVM considerando as combinações para a detecção do greening (Grupo 2).

	Excitação 405 nm		Excitação 470 nm		Combinação das duas excitações
	F560 +	F560 +	F530 +	F530 +	Todas as fluorescências
	F690	F580 +	F690	F550 +	
		F690		F690	
Sensibilidade	100,0%	100,0%	87,7%	100,0%	100,0%
Especificidade	13,1%	13,1%	80,8%	77,8%	79,8%
Acurácia	52,2%	52,2%	83,9%	87,8%	88,9%

Fonte: Elaborada pelo autor.

Os resultados demonstram que a combinação das bandas espectrais para a excitação em 405 nm não alteram significativamente o desempenho do SVM na classificação dos dados, como observado ao compararmos com os resultados da Tabela 5.5. Além disso, a banda em 580 nm não contribui no resultado, pois o desempenho é igual para as duas combinações. No caso da excitação em 470 nm, observa-se uma boa acurácia na classificação dos dados na combinação das três bandas (87,8%) com uma sensibilidade de 100,0%, permitindo a

identificação de todas as amostras positivas para greening. No entanto, a especificidade é baixa para um eventual uso do classificador, pois somente 78,0% das amostras com deficiência de zinco são identificadas corretamente. A contribuição da combinação das duas excitações e todas as bandas espectrais não é significativa quando comparada ao desempenho do SVM considerando as três bandas para a excitação em 470 nm. Dessa forma, esses resultados indicam que o uso da excitação em 470 nm é melhor para a discriminação das amostras de greening e deficiência de zinco.

Os resultados discutidos para a técnica SVM são promissores com relação à discriminação das amostras de cancro cítrico e verrugose, especialmente, utilizando a excitação em 470 nm e as três bandas espectrais em 550, 570 e 690 nm. No caso da discriminação das amostras de greening e deficiência de zinco, os resultados são satisfatórios, mesmo com uma acurácia de aproximadamente 88,0%, uma vez que a especificidade de 78% não permite identificar a maior parte das amostras com deficiência de zinco. Na seção a seguir, serão apresentados os resultados da classificação dos dados através da técnica ANN.

5.2.2 Resultados da classificação das imagens pelo ANN

Na classificação das imagens de fluorescência através do método ANN utilizamos, especificamente, as redes neurais de multicamadas MLP estruturada com duas camadas ocultas, a função de ativação sigmoideal para a transferência de informação entre os neurônios, e a função de retro propagação para o treinamento da rede, como descrito na seção 3.2.3.3. O número máximo de iterações que o algoritmo realizava foi definido como 1000, e a função de desempenho do erro quadrático médio (MSE, do inglês *Mean Square Error*) definido como 0,01. O processo de treinamento da rede neural só cessa ao atingir o número de iterações máximas ou o MSE desejado. A camada de entrada contém um nó para cada característica extraída das imagens de fluorescência e a camada de saída um nó para cada parâmetro da classificação, como apresentado na Figura 3.11.

O número de neurônios nas camadas ocultas é parte fundamental da arquitetura das redes neurais.(93) Essas conexões tornam a técnica capaz de encontrar importantes relações entre as diferentes variáveis.(80) Sendo assim, variamos o número de neurônios da rede entre 5 a 4000 a cada 5 neurônios com o objetivo de determinar o número ideal de neurônios que fornece a máxima acurácia. O número de neurônios selecionados está relacionado com as configurações do computador utilizado na classificação das imagens. O processamento foi realizado em um notebook equipado com processador Intel(R) Core(TM) i7-3537U CPU @

2.00GH, 8.00 GB de memória RAM e Windows 8.1 *Home Edition*. Para cada conjunto de neurônios, o processo foi repetido 20 vezes a fim de minimizar o desvio padrão, uma vez que se trata de um processo aleatório.

A partir dessas configurações e para cada grupo de neurônios, as redes neurais foram treinadas usando o conjunto de dados de treinamento e, em seguida, com a rede gerada, esta foi aplicada sob o conjunto de teste, obtendo-se assim a acurácia da classificação. Tanto o conjunto de treinamento como de teste de cada grupo são iguais ao utilizado na classificação dos dados pela técnica SVM, visto que o melhor desempenho do SVM se obteve para um conjunto composto por 20 amostras de treinamento, 10 para cada doença em estudo. O algoritmo desenvolvido no software Matlab[®] salva todos os dados gerados da classificação para cada conjunto de neurônios e repetição no formato *xlsx*. Em seguida, é calculado a média e o desvio padrão dos resultados obtidos e selecionado o número de neurônios de acordo com o conjunto que apresentou a maior acurácia. Uma vez determinado o número de neurônios, o algoritmo verifica qual o melhor resultado entre as 20 repetições realizadas nesse conjunto de neurônios e seleciona o melhor resultado, o qual é salvo em outra planilha do Microsoft Excel. Todo este procedimento é realizado automaticamente e repetido para cada excitação e banda de fluorescência para cada grupo de doenças.

Assim como na classificação pela técnica SVM, a primeira tentativa consistiu em analisar os resultados da classificação das doenças para cada banda espectral e excitação. A Tabela 5.8 apresenta as taxas de amostras classificadas como VP, VN, FP e FN obtidas a partir da comparação dos resultados da classificação das amostras do Grupo 1 pelo ANN com os resultados do teste padrão-ouro.

Tabela 5.8 - Resultados da classificação do Grupo 1 pelo método ANN segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

		Condição determinada pelo teste padrão-ouro		
				Resultado do ANN
		Condição Positiva	Condição Negativa	
Excitação em 405 nm	F550	Positivo	VP = 86	FP = 19
		Negativo	FN = 4	VN = 71
	F580	Positivo	VP = 88	FP = 35
		Negativo	FN = 2	VN = 55
	F690	Positivo	VP = 67	FP = 60
		Negativo	FN = 23	VN = 30

(continua)

(continuação)

Tabela 5.8 - Resultados da classificação do Grupo 1 pelo método ANN segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

		Resultado do ANN	Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Excitação em 470 nm	F550	Positivo	VP = 87	FP = 39
		Negativo	FN = 3	VN = 51
	F570	Positivo	VP = 90	FP = 72
		Negativo	FN = 0	VN = 18
	F690	Positivo	VP = 72	FP = 17
		Negativo	FN = 18	VN = 73

Fonte: Elaborada pelo autor.

Uma vez determinado as taxas VP, VN, FP e FN, calculamos o desempenho do classificador ANN segundo os três parâmetros estabelecidos. O desempenho do ANN é apresentado na Tabela 5.9 para as duas excitações e cada banda espectral.

Tabela 5.9 - Desempenho da discriminação de amostras de cancro cítrico e verrugose pelo método ANN.

	Excitação em 405 nm			Excitação em 470 nm		
	F550	F580	F690	F550	F570	F690
Sensibilidade	95,6%	97,8%	74,4%	96,7%	100,0%	80,0%
Especificidade	78,9%	61,1%	33,3%	56,7%	20,0%	81,1%
Acurácia	87,2%	79,4%	53,9%	76,7%	60,0%	80,6%

Fonte: Elaborada pelo autor.

A análise do desempenho do ANN de acordo com a Tabela 5.9 demonstra que os parâmetros sensibilidade e especificidade variam significativamente conforme a banda espectral e excitação. A sensibilidade varia desde 74,4% para a banda em 690 nm e excitação em 405 nm até 100,0% no caso da fluorescência em 570 nm para a excitação em 470 nm. O intervalo de variação da especificidade é maior, desde 20,0% para a fluorescência em 570 nm até o máximo de 81,1% no caso da fluorescência em 690 nm, ambas para a excitação em 470 nm. A acurácia também não ultrapassa o valor de 90,0% para as bandas analisadas. Esse valor de acurácia combinado com valores altos na sensibilidade e especificidade possibilitaria a discriminação das doenças sem a necessidade da combinação das bandas. O melhor resultado

novamente ficou por conta da classificação das imagens de fluorescência em 550 nm com excitação em 405 nm. Comparando os resultados de ambos os classificadores através das Tabelas 5.3 e 5.9, demonstra-se que o desempenho do ANN nesta primeira tentativa é ligeiramente melhor em todas as bandas espectrais na discriminação das doenças do Grupo 1.

Adotando o mesmo procedimento, a Tabela 5.10 apresenta as taxas de amostras classificadas como VP, VN, FP e FN obtidas a partir da comparação dos resultados da classificação das amostras do Grupo 2 pelo ANN com os resultados do teste padrão-ouro.

Tabela 5.10 - Resultados da classificação do Grupo 2 pelo método ANN segundo as taxas de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN).

		Resultado do ANN	Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Excitação em 405 nm	F560	Positivo	VP = 81	FP = 86
		Negativo	FN = 0	VN = 13
	F580	Positivo	VP = 81	FP = 95
		Negativo	FN = 0	VN = 4
	F690	Positivo	VP = 78	FP = 79
		Negativo	FN = 3	VN = 20
Excitação em 470 nm	F530	Positivo	VP = 67	FP = 29
		Negativo	FN = 14	VN = 70
	F550	Positivo	VP = 81	FP = 93
		Negativo	FN = 0	VN = 6
	F690	Positivo	VP = 52	FP = 26
		Negativo	FN = 29	VN = 73

Fonte: Elaborada pelo autor.

A partir das taxas VP, VN, FP e FN, calculamos os parâmetros sensibilidade, especificidade e acurácia de cada banda espectral para as duas excitações. O desempenho do classificador ANN é apresentado na Tabela 5.11.

Tabela 5.11 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método ANN.

	Excitação em 405 nm			Excitação em 470 nm		
	F560	F580	F690	F530	F550	F690
Sensibilidade	100,0%	100%	96,3%	82,7%	100,0%	64,2%
Especificidade	13,1%	4,1%	20,2%	70,7%	6,1%	73,7%
Acurácia	52,2%	47,2%	54,4%	76,1%	48,3%	69,4%

Fonte: Elaborada pelo autor.

Os resultados demonstram novamente uma variação significativa nos parâmetros sensibilidade e especificidade, com valores baixos no caso da especificidade para todas as bandas espectrais, exceto na análise da fluorescência em 530 nm e excitação em 470 nm. O melhor resultado foi obtido justamente para esse comprimento de onda, atingindo uma acurácia de 76,1%, sensibilidade de 82,7% e especificidade de 70,7%. Entretanto, as precisões alcançadas são baixas, uma vez que um bom teste de classificação deve resultar em valores altos para os três parâmetros analisados. O classificador ANN se mostrou um pouco melhor em relação ao SVM, comparando os resultados das Tabelas 5.5 e 5.11. A maior diferença entre os resultados ocorreu para a fluorescência em 530 nm e excitação em 470 nm, onde verifica-se um aumento de mais de 50% no valor da acurácia do ANN em relação à acurácia obtida para o SVM, assim como um aumento também nos parâmetros sensibilidade e especificidade.

Visto que o desempenho do classificador ANN também ficou aquém do desejado na classificação das doenças dos dois grupos, resolvemos adotar a segunda estratégia que consiste em combinar, para cada excitação, os dois comprimentos de onda com as acurácias mais elevadas, assim como as três bandas espectrais e, por fim, as duas excitações e todas as bandas de fluorescência. A partir dos resultados da classificação das amostras pelo ANN, as combinações realizadas são iguais às adotadas no SVM para os dois grupos de doenças. Nas doenças do Grupo 1 foram combinadas as bandas de fluorescência em 550 e 580 nm para a excitação em 405 nm, as bandas em 550 e 690 nm para a excitação em 470 nm, assim com as três bandas de cada excitação e por fim todas as bandas. Nos casos das doenças do Grupo 2, combinamos as bandas de fluorescência em 560 e 690 nm para a excitação em 405 nm, as bandas em 530 e 690 para a excitação em 470 nm, bem como as três bandas de cada excitação e por fim todas as bandas espectrais. A Tabela 5.12 apresenta os resultados da classificação para as amostras do Grupo 1 considerando essa estratégia.

Tabela 5.12 - Desempenho do classificador ANN considerando as combinações para a detecção do cancro cítrico (Grupo 1).

	Excitação 405 nm		Excitação 470 nm		Combinação das duas excitações
	F550	F550	F550	F550	
	+	+	+	+	Todas as fluorescências
	F580	F580	F690	F570	
		+		+	
		F690		F690	
Sensibilidade	100,0%	100,0%	98,9%	100,0%	100,0%
Especificidade	81,1%	90,0%	88,9%	88,9%	94,4%
Acurácia	90,6%	95,0%	93,9%	94,4%	97,2%

Fonte: Elaborada pelo autor.

Os resultados apresentados na Tabela 5.12 mostram que o desempenho do classificador ANN aumenta significativamente quando considerado a combinação das bandas espectrais e excitações. Assim como observado na classificação pelo SVM, o melhor resultado foi obtido ao considerarmos a combinação das duas excitações e todas as bandas de fluorescência. A acurácia do ANN foi de 97,2% enquanto que do SVM foi de 97,8% para essa combinação. Também foi possível identificar corretamente todas as amostras positivas para cancro cítrico (sensibilidade de 100,0%), e as amostras negativas para cancro cítrico foram identificadas com uma precisão significativa (especificidade de 94,4%). Na comparação dos resultados dos dois classificadores, é possível notar que a diferença no desempenho entre eles não é significativa quando analisado os resultados da combinação das bandas para a excitação em 470 nm, assim como a combinação de todas as fluorescências. O SVM, nesse caso, é um pouco melhor. No entanto, os resultados melhoram consideravelmente para a combinação das três bandas com excitação em 405 nm comparados aos resultados do classificador SVM. A acurácia do ANN salta para 95,0% com sensibilidade de 100,0% e especificidade de 90,0%.

Seguindo os procedimentos adotados para a combinação das bandas espectrais das amostras do Grupo 1, obtivemos os resultados apresentados na Tabela 5.13 para a classificação das doenças do Grupo 2.

Tabela 5.13 - Desempenho do classificador ANN considerando as combinações para a detecção do greening (Grupo 2).

	Excitação 405 nm		Excitação 470 nm		Combinação das duas excitações
	F560	F560	F530	F530	
	+	+	+	+	
	F690	F580	F690	F550	Todas as fluorescências
		+		+	
		F690		F690	
Sensibilidade	100,0%	100,0%	96,3%	100,0%	100,0%
Especificidade	25,3%	26,3%	84,9%	84,9%	91,0%
Acurácia	58,9%	59,4%	90,0%	91,7%	93,2%

Fonte: Elaborada pelo autor.

A partir da Tabela 5.13, é possível observar que a combinação das bandas espectrais relativas à excitação em 405 nm não melhoram substancialmente o desempenho da classificação das amostras pelo ANN. Nota-se que os melhores resultados são obtidos para a combinação das bandas relacionadas à excitação em 470 nm, uma vez que a acurácia da técnica ultrapassa os 90,0% de acerto na diferenciação entre as doenças, assim como quando considera-se a combinação das duas excitações e fluorescências, atingindo 93,2% de acurácia. No entanto, no primeiro caso, a especificidade de praticamente 85,0% é considerada baixa. Em contrapartida, a combinação de todos os dados atingiu uma especificidade de 91,0% contribuindo significativamente para a identificação das amostras negativas para greening. Além disso, a sensibilidade de 100,0% permitiu identificar todas as amostras positivas para greening. Assim como observamos na classificação pelo SVM, a contribuição das bandas em 405 nm na combinação de todas as fluorescências não é tão significativa também no ANN, mas provavelmente colaborou para o aumento do valor da especificidade nesse caso. Por fim, ao compararmos os resultados do SVM e ANN através das Tabelas 5.7 e 5.13, respectivamente, é notório que o desempenho do classificador ANN é melhor em todas as combinações.

Os resultados discutidos para a técnica de classificação ANN são promissores, uma vez que a técnica pode ser aplicada na diferenciação das amostras com sintomas visuais semelhantes com uma precisão elevada, especialmente quando combinado os dados das duas excitações e de todas as bandas espectrais. Na seção seguinte, iremos realizar uma breve discussão comparando os resultados obtidos da classificação das doenças dos dois grupos através das técnicas SVM e ANN.

5.2.3 Discussão comparativa dos resultados dos classificadores

A classificação das imagens de fluorescência permitiu identificar e diferenciar as doenças com uma precisão elevada dependendo do método de classificação e estratégia adotado. A técnica de espectroscopia por imagens de fluorescência aplicada neste trabalho através dos sistemas desenvolvidos em conjunto com os métodos de classificação pode ser utilizada diretamente no campo para o controle das doenças sem a necessidade de procedimentos laboratoriais complexos. Mesmo que utilizado no laboratório, o diagnóstico seria muito mais rápido e menos caro do que os protocolos de detecção padrão. O atual sistema de aquisição das imagens de fluorescência permite a coleta de imagens de uma única amostra por vez. A partir do potencial apresentado por essa técnica, é possível, no futuro, o desenvolvimento de um único sistema, onde imagens de fluorescência possam ser coletadas de várias amostras simultaneamente e analisadas ao mesmo tempo, reduzindo consequentemente ainda mais o tempo de diagnóstico.

A discriminação das doenças pelos dois métodos de classificação produziu resultados interessantes. Primeiramente, a análise apenas das bandas de fluorescência dos respectivos grupos de doenças e excitações não permitiu distinguir as amostras com uma boa precisão por parte dos dois classificadores. Nessa primeira estratégia adotada, o melhor resultado ficou por conta da classificação das imagens de fluorescência de cancro cítrico e verrugose em 550 nm com excitação em 405 nm através da técnica ANN, superando o desempenho do classificador SVM. No entanto, a especificidade de 78,9% é baixa para diferenciarmos as amostras negativas para cancro cítrico.

Visto que os desempenhos dos classificadores estavam abaixo do esperado, verificamos a influência do número de amostras de treinamento sobre os resultados da classificação. A partir da análise da acurácia do SVM em função das amostras de treinamento para os dois grupos de doenças definimos o conjunto ideal para treinarmos os classificadores. O resultado mostrou que o conjunto constituído por 20 amostras de treinamento, 10 de cada doença em estudo, fornece a maior acurácia na classificação dos dados. Desta forma, definimos que esse número de amostras de treinamento seria utilizado na classificação das doenças através da técnica ANN.

Dado que já estávamos treinando os classificadores com o conjunto ideal de amostras, adotamos uma segunda estratégia a fim de melhorar o desempenho do SVM e ANN. Essa estratégia consistiu em combinar as diferentes excitações e bandas de fluorescência. Primeiramente, com relação à classificação das amostras de cancro cítrico e verrugose,

verificamos que o desempenho entre os classificadores não apresenta uma diferença significativa quando combinadas as bandas da excitação em 470 nm, mas os resultados do ANN são melhores quando combinadas as bandas da excitação em 405 nm. O melhor resultado, no geral, foi obtido para o classificador SVM com uma acurácia de 97,8% contra 97,2% do ANN ao considerarmos a combinação das duas excitações e todas as bandas de fluorescência. No caso, a especificidade para essa combinação também foi maior para o SVM com 95,6% contra 94,4% do ANN. Desta forma, podemos concluir que tanto o SVM como o ANN são interessantes na discriminação das amostras de cancro cítrico e verrugose, sendo cada um melhor para as bandas de uma determinada excitação.

No caso da análise das amostras de greening e deficiência de zinco, verificamos que os resultados obtidos para o classificador ANN são superiores aos do SVM em todas as combinações, mas em especial para a combinação das duas excitações e todas as bandas espectrais. Comparando essa combinação, observa-se que a acurácia salta de 88,9% (SVM) para 93,2% no caso do método ANN, com sensibilidade de 100,0% em ambos os casos. A mudança mais significativa se dá por conta do aumento da especificidade, que passa de 79,8% (SVM) para 91,0% (ANN). Dessa forma, um número maior de amostras negativas para greening são identificadas. O melhor desempenho do método ANN está relacionado provavelmente com a capacidade da técnica em capturar características mais complexas dos dados. Sendo assim, podemos concluir que o método ANN é mais indicado para a discriminação das amostras de greening e deficiência de zinco coletadas no Brasil.

É provável que o uso de outras excitações e bandas espectrais, bem como suas combinações, associadas a estas técnicas de classificação possam colaborar para o aumento dos níveis de precisão apresentados neste estudo para a detecção de cancro cítrico e greening, assim como para diferentes condições de estresses bióticos e abióticos. Estudos ainda devem ser conduzidos a fim de melhorar ainda mais os resultados, porém trata-se de um resultado muito promissor que pode ser utilizado na diferenciação das doenças que possuem sintomas similares e, eventualmente, o desenvolvimento de um sistema comercial.

Os resultados da discriminação das diferentes doenças que afetam a citricultura no Brasil através da técnica SVM foram publicados, recentemente, na revista *Applied Optics*. O trabalho encontra-se disponível para consulta na referência.(25)

6 DISCRIMINAÇÃO DAS PRINCIPAIS DOENÇAS/ESTRESSE QUE AFETAM A CITRICULTURA NA FLÓRIDA/EUA

Este capítulo é dedicado à apresentação dos experimentos e resultados envolvidos na discriminação da principal doença que afeta a citricultura na Flórida: o greening. Os experimentos foram conduzidos no CREC da Universidade da Flórida durante o estágio de pesquisa no exterior sob a orientação do Prof. Dr. Reza Ehsani. O objetivo do estágio foi avaliar a viabilidade da aplicação dos sistemas desenvolvidos para amostras em condições ambientais distintas sem a necessidade de considerar tais condições em nossas análises. Dentre as condições ambientais, podemos citar os níveis de nutrientes das plantas, características do solo, umidade, etc. Vale ressaltar que neste estágio realizamos somente a discriminação das amostras sintomáticas de greening e de deficiência de zinco. O procedimento adotado é igual ao descrito no Capítulo 5, onde inicialmente selecionamos os comprimentos de onda de interesse através do SIF-1, coletamos as imagens para as bandas selecionadas utilizando o SIF-2 e, por fim, classificamos estas imagens pelos métodos SVM e ANN. As seções a seguir apresentam os resultados obtidos em cada etapa.

6.1 Experimento exploratório utilizando o SIF-1

O trabalho se concentrou primeiramente na coleta das imagens de fluorescência a partir do SIF-1 a fim de selecionar as principais bandas espectrais para a discriminação das amostras de greening e deficiência de zinco. Imagens de fluorescência foram coletadas para 5 amostras de cada uma das condições para as duas excitações (405 e 470 nm). Após a coleta das imagens, testes laboratoriais convencionais foram conduzidos pelo Prof. Dr. William Gurley da Universidade da Flórida para o diagnóstico e confirmação da doença nas amostras.

Assim como no experimento exploratório realizado no Brasil, as imagens de fluorescência foram obtidas para comprimentos de onda a partir de 480 nm para as duas excitações e, em seguida, pré-processadas conforme descrito no tópico 3.2.2. A partir da normalização das imagens, foi possível comparar as amostras de greening e deficiência de zinco e selecionar as principais bandas espectrais através da análise da intensidade média da fluorescência emitida em função do comprimento de onda. Utilizando a mesma rotina computacional desenvolvida no Matlab[®], um ROI quadrado de dimensões 20 x 20 mm foi selecionado aleatoriamente em cada amostra considerando que a região selecionada deve conter sintomas visuais da doença e do estresse. Uma vez selecionado o ROI, a rotina calcula

a média das intensidades dos pixels das imagens para cada comprimento de onda, excitação e amostra. Esses procedimentos estão descritos em maiores detalhes na seção 5.1. Resolvemos repetir esse procedimento para as amostras de greening e deficiência de zinco da Flórida dado que não sabemos se as bandas espectrais definidas no Brasil são as melhores para a discriminação da doença em outro país, sob condições diferentes. As Figuras 6.1.a e 6.1.b apresentam os gráficos da intensidade em função do comprimento de onda para amostras de greening e deficiência de zinco utilizando as excitações em 405 e 470 nm, respectivamente.

Figura 6.1 - Gráficos da intensidade em função do comprimento de onda comparando amostras de greening e deficiência de zinco para excitações em (a) 405 nm e (b) 470 nm.
Fonte: Elaborada pelo autor.

A partir da Figura 6.1 foi possível determinar as principais bandas de fluorescência para o processo de discriminação das amostras contaminadas com greening. As bandas foram selecionadas de acordo com as regiões onde se observa uma distinção entre os espectros da doença e do estresse. Comparando os espectros de fluorescência obtidos através da excitação em 405 nm (Figura 6.1.a), selecionamos as bandas espectrais em 560 e 580 nm. No caso da excitação em 470 nm, foram selecionadas as bandas em 530 e 550 nm de acordo com as regiões de máxima intensidade observadas na Figura 6.1.b. Além destes comprimentos de onda selecionados, também consideramos a banda de fluorescência em 690 nm para as duas excitações, mesmo não havendo uma clara distinção entre os espectros, pois está relacionada com a emissão das moléculas de clorofila *a*.(44,91) A Tabela 6.1 apresenta as bandas espectrais selecionadas a partir deste experimento exploratório.

Tabela 6.1 - Bandas espectrais selecionadas para as excitações em 405 e 470 nm a partir do experimento exploratório para a discriminação de amostras de greening e deficiência de zinco.

Doença/Deficiência	Excitação em 405 nm	Excitação em 470 nm
Greening e Deficiência de zinco	Filtros em 560, 580 e 690 nm	Filtros em 530, 550 e 690 nm

Fonte: Elaborada pelo autor.

Comparando os resultados da seleção das bandas espectrais do Brasil e da Flórida é possível observar que as bandas selecionadas são idênticas na discriminação das amostras de greening e deficiência de zinco. A princípio, podemos concluir que a aplicação do sistema SIF-1 para a seleção das principais bandas não depende das diferentes características de cada região. Desta forma, uma vez definidas as melhores bandas, estas poderão ser utilizadas para a coleta das imagens de fluorescência pelo SIF-2 sem a necessidade de realizar um estudo prévio com o SIF-1 para as amostras com greening ou deficiência de zinco dos estados de São Paulo/Brasil e Flórida/EUA. Nas próximas seções, iremos avaliar o potencial desse resultado através da análise do desempenho dos métodos SVM e ANN na classificação das imagens de fluorescência obtidas pelo SIF-2 para as bandas espectrais definidas.

6.2 Experimento utilizando o SIF-2

Nesta segunda etapa, imagens de fluorescência foram coletadas para 100 amostras de greening e deficiência de zinco no CREC utilizando o SIF-2 para os comprimentos de onda apresentados na Tabela 6.1. Após a aquisição das imagens, estas foram pré-processadas conforme descrito no tópico 3.2.2 a fim de eliminar a informação espectral da câmera CCD e a porcentagem de transmissão dos filtros ópticos. Em seguida, foi realizado o processo de análise das imagens de fluorescência que consiste na segmentação, extração das características e classificação pelo SVM e ANN. Assim como nos dados do Brasil, as imagens foram segmentadas manualmente e os três descritores definidos na seção 3.2.3.2 foram extraídos dessas imagens, sendo utilizados como vetor de entrada nos classificadores. A Figura 6.2 apresenta exemplos de imagens segmentadas das amostras de greening e deficiência de zinco da Flórida, destacando as semelhanças visuais existentes entre estas.

Figura 6.2 - Exemplos de imagens segmentadas obtidas a partir da excitação em 470 nm e fluorescência em 690 nm para amostras de (a) greening e (b) deficiência de zinco.
Fonte: Elaborada pelo autor.

A classificação das imagens foi realizada mediante a aplicação dos classificadores nos conjuntos de dados compostos pelas características extraídas das 200 amostras. Visto que no Capítulo 5 definimos o número ideal de amostras que devem compor os conjuntos de treinamento e teste, os classificadores foram treinados com as características extraídas de 20 amostras, 10 para cada condição, e testados com as características das 180 amostras restantes. As amostras foram classificadas como positivas ou negativas para greening e, a partir da comparação com os resultados do teste padrão-ouro (qPCR), foi possível avaliar os desempenhos dos classificadores através dos parâmetros, já discutidos, acurácia, sensibilidade e especificidade. Os parâmetros foram calculados de acordo com as Equações (12) - (14) apresentadas no Capítulo 5, as quais são descritas em termos da taxa de amostras classificadas como verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN). Os resultados da classificação pelos métodos SVM e ANN são apresentados nas seções seguintes.

6.2.1 Resultados da classificação das imagens pelo SVM

Inicialmente, os parâmetros C e γ do kernel do SVM foram ajustados através da técnica de validação cruzada a fim de se obter o melhor desempenho na classificação das imagens de fluorescência. Analogamente à classificação dos dados do Brasil, os parâmetros C e γ foram variados no intervalo entre 2^{-30} a 2^{30} com passo de $2^{0,25}$ e, em seguida, um ajuste

fino foi conduzido em torno do par que apresentou a melhor acurácia na validação cruzada na etapa inicial. Os valores ideais dos parâmetros foram determinados conforme o resultado desse ajuste fino. A Figura 6.3 ilustra os resultados do procedimento adotado para a seleção do melhor par (C, γ) utilizando imagens de fluorescência em 560 nm obtidas a partir da excitação em 405 nm.

Figura 6.3 - Mapa de contorno da otimização dos parâmetros C e γ destacando a etapa inicial da seleção das melhores regiões (a) e o ajuste fino em torno da região selecionada (b).

Fonte: Elaborada pelo autor.

Os resultados mostram, especificamente para esses dados, que os melhores resultados na etapa de validação cruzada são obtidos para valores de C variando entre $2^{25,6}$ e $2^{25,7}$, e γ em torno da região de $2^{-4,0}$. A acurácia nessa região varia entre 85 e 90%, conforme apresentado na Figura 6.3.b. Esse procedimento adotado é repetido para cada excitação e banda espectral apresentada na Tabela 6.1. A seleção do melhor par (C, γ) é realizada automaticamente pela rotina desenvolvida no Matlab[®].

Após a seleção do melhor par (C, γ) , o conjunto de dados de treinamento foi treinado novamente no sentido de construir o modelo de classificação SVM e, em seguida, o modelo foi aplicado no conjunto de teste. Adotamos as mesmas duas estratégias utilizadas na análise das amostras do Brasil. A primeira estratégia consiste na aplicação do classificador SVM nos dados de apenas uma única excitação e fluorescência, enquanto que a segunda envolve a combinação das bandas espectrais com os melhores desempenhos na classificação. A Tabela 6.2 apresenta os resultados obtidos para a primeira estratégia, conforme as taxas de amostras classificadas como VP, VN, FP e FN.

Tabela 6.2 - Resultados da classificação pelo método SVM segundo as taxas VP, VN, FP e FN.

			Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Resultado do SVM				
Excitação em 405 nm	F560	Positivo	VP = 84	FP = 49
		Negativo	FN = 18	VN = 29
	F580	Positivo	VP = 100	FP = 72
		Negativo	FN = 2	VN = 6
	F690	Positivo	VP = 8	FP = 4
		Negativo	FN = 94	VN = 74
	F530	Positivo	VP = 83	FP = 35
		Negativo	FN = 19	VN = 43
Excitação em 470 nm	F550	Positivo	VP = 86	FP = 39
		Negativo	FN = 16	VN = 39
	F690	Positivo	VP = 33	FP = 32
		Negativo	FN = 69	VN = 46

Fonte: Elaborada pelo autor.

As taxas apresentadas na Tabela 6.2 foram obtidas comparando os resultados da classificação pelo SVM com os resultados da análise por qPCR realizadas pelo Prof. Dr. William Gurley. A partir destes dados, calculamos os parâmetros sensibilidade, especificidade e acurácia. Os resultados do desempenho do SVM na discriminação das amostras de greening e deficiência de zinco são apresentados na Tabela 6.3.

Tabela 6.3 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método SVM.

	Excitação em 405 nm			Excitação em 470 nm		
	F560	F580	F690	F530	F550	F690
Sensibilidade	82,4%	98,0%	7,8%	81,4%	84,3%	32,4%
Especificidade	37,3%	7,7%	94,9%	55,1%	50,0%	59,0%
Acurácia	62,8%	58,9%	45,6%	70,0%	69,4%	43,9%

Fonte: Elaborada pelo autor.

Os resultados mostram, assim como na classificação das amostras de greening e deficiência de zinco do Brasil, que o desempenho do classificador fica muito aquém do esperado, com os parâmetros sensibilidade e especificidade variando significativamente de acordo com a excitação e banda espectral selecionada. Além disso, as acurácias obtidas também foram baixas variando desde 43,9% (fluorescência em 690 nm e excitação em 470 nm) até 70,0% (fluorescência em 530 nm e excitação em 470 nm). Diferentemente do observado na classificação das amostras do Brasil, os piores desempenhos ficaram por conta das imagens de fluorescência em 690 nm para as duas excitações. No caso do Brasil, esse comprimento de onda para ambas as excitações forneceu os melhores resultados observados adotando a primeira estratégia.

Dado que o desempenho do classificador SVM foi relativamente baixo para as diferentes bandas espectrais e excitações, resolvemos adotar a segunda estratégia uma vez que os resultados melhoraram significativamente através da combinação das bandas espectrais para os dados do Brasil. Sendo assim, a partir da Tabela 6.3, combinamos as bandas de fluorescência em 560 e 580 nm para a excitação em 405 nm, as bandas em 530 e 550 nm para a excitação em 470 nm, bem como as três bandas de cada excitação e por fim todas as bandas espectrais. As combinações das bandas espectrais diferem das combinações realizadas no Brasil, dado que nos baseamos na seleção das combinações segundo o critério inicialmente definido. No entanto, ao final deste capítulo, iremos apresentar os resultados utilizando as mesmas combinações do Brasil no intuito de verificar se há diferenças significativas no

desempenho dos dois classificadores. A Tabela 6.4 apresenta os resultados da classificação das amostras de greening e deficiência de zinco considerando as combinações mencionadas.

Tabela 6.4 - Desempenho do classificador SVM considerando as combinações para a detecção do greening.

	Excitação 405 nm		Excitação 470 nm		Combinação das duas excitações
	F560 +	F560 +	F530 +	F530 +	Todas as fluorescências
	F580	F580 +	F550	F550 +	
		F690		F690	
Sensibilidade	92,2%	88,2%	88,2%	89,2%	90,2%
Especificidade	38,5%	96,2%	59,0%	87,2%	98,7%
Acurácia	68,9%	91,7%	75,6%	88,3%	92,8%

Fonte: Elaborada pelo autor.

A partir da Tabela 6.4 é possível observar que a combinação das bandas espectrais aumenta significativamente o desempenho do classificador SVM, especialmente, quando combinadas as três bandas de fluorescência de cada excitação. Para essa combinação, a acurácia do SVM foi de aproximadamente 93,0% com uma especificidade de 98,7%, permitindo identificar praticamente todas as amostras negativas para greening. A sensibilidade de 90,2% também é considerada alta na identificação das amostras positivas para greening. Vale destacar também o desempenho obtido na combinação das três bandas de fluorescência para a excitação em 405 nm, onde atingimos um acurácia de 91,7%, especificidade de 96,2% e sensibilidade de 88,2%. Através dessa combinação podemos diminuir o tempo de identificação das amostras contaminadas, uma vez que não seria necessário coletar imagens para a excitação em 470 nm. Os resultados demonstram um grande potencial da técnica de espectroscopia por imagens em conjunto com o método SVM na identificação tanto de amostras de greening e deficiência de zinco da Flórida como do Brasil. Na seção a seguir serão apresentados os resultados da classificação dos dados através do método ANN.

6.2.2 Resultados da classificação das imagens pelo ANN

A classificação das imagens de fluorescência das amostras de greening e deficiência de zinco através do método ANN foram realizadas utilizando as mesmas configurações definidas na análise das imagens do Brasil. No caso, utilizamos as redes neurais de multicamadas MLP com duas camadas ocultas, a função de ativação sigmoideal para a transferência da informação entre os neurônios, a função de retro propagação para o treinamento da rede, e definimos o número máximo de iterações e o desempenho MSE como 1000 e 0,01, respectivamente. O número de neurônios da rede foi variado entre 5 a 4000 em passo de 5 neurônios, e para cada conjunto de neurônios, o processo foi repetido 20 vezes, calculando as respectivas acurácias. O algoritmo desenvolvido calcula a média e o desvio padrão dos resultados obtidos e seleciona o número de neurônios de acordo com o conjunto que apresentou a maior acurácia, e em seguida, determina o melhor resultado entre as 20 repetições realizadas para esse conjunto de neurônios. As redes foram treinadas com o mesmo conjunto de dados que o SVM utilizou no processo de treinamento e testadas com as demais amostras. Esse procedimento foi realizado para as bandas espectrais definidas na Tabela 6.1.

Assim como na classificação pela técnica SVM, a primeira tentativa consistiu em analisar os resultados da classificação das doenças para cada banda espectral e excitação. A Tabela 6.5 apresenta as taxas de amostras classificadas como VP, VN, FP e FN obtidas a partir da comparação dos resultados da classificação das amostras pelo método ANN com os resultados do teste padrão-ouro.

Tabela 6.5 - Resultados da classificação das amostras de greening e deficiência de zinco pelo método ANN segundo as taxas de VP, VN, FP e FN.

		Condição determinada pelo teste padrão-ouro		
		Condição Positiva	Condição Negativa	
Resultado do ANN				
Excitação em 405 nm	F560	Positivo	VP = 74	FP = 34
		Negativo	FN = 28	VN = 44
	F580	Positivo	VP = 98	FP = 66
		Negativo	FN = 4	VN = 12
	F690	Positivo	VP = 44	FP = 24
		Negativo	FN = 58	VN = 54

(continua)

(continuação)

Tabela 6.5 - Resultados da classificação das amostras de greening e deficiência de zinco pelo método ANN segundo as taxas de VP, VN, FP e FN.

			Condição determinada pelo teste padrão-ouro	
			Condição Positiva	Condição Negativa
Resultado do ANN				
Excitação em 470 nm	F530	Positivo	VP = 69	FP = 18
		Negativo	FN = 33	VN = 60
	F550	Positivo	VP = 76	FP = 22
		Negativo	FN = 26	VN = 56
	F690	Positivo	VP = 75	FP = 38
		Negativo	FN = 27	VN = 40

Fonte: Elaborada pelo autor.

A partir dos resultados apresentados na Tabela 6.5, calculamos o desempenho do classificador ANN através dos parâmetros sensibilidade, especificidade e acurácia para cada banda espectral e excitação. O desempenho do classificador ANN é apresentado na Tabela 6.6.

Tabela 6.6 - Desempenho da discriminação de amostras de greening e deficiência de zinco pelo método ANN.

	Excitação em 405 nm			Excitação em 470 nm		
	F560	F580	F690	F530	F550	F690
Sensibilidade	72,6%	96,1%	43,1%	67,7%	74,5%	73,5%
Especificidade	56,4%	15,4%	69,2%	76,9%	71,8%	51,3%
Acurácia	65,6%	61,1%	54,4%	71,7%	73,3%	63,9%

Fonte: Elaborada pelo autor.

Com base nos resultados apresentados na Tabela 6.6, pode-se verificar que o desempenho do classificador ANN é relativamente superior ao desempenho do SVM apresentado na Tabela 6.3 para essa primeira tentativa adotada. No entanto, o desempenho ainda é baixo para um eventual uso da técnica na discriminação de amostras de greening e deficiência de zinco. As acurácias obtidas variaram desde 54,4% (fluorescência em 690 nm e excitação em 470 nm) até 73,3% (fluorescência em 550 nm e excitação em 470 nm), com valores baixos para os parâmetros sensibilidade e especificidade.

A segunda tentativa adotada foi a combinação das bandas de fluorescência com os melhores desempenhos. As combinações realizadas para a classificação pelo ANN são iguais às adotadas no SVM, uma vez que os melhores desempenhos foram obtidos para as mesmas bandas espectrais. Sendo assim, combinamos as bandas de fluorescência em 560 e 580 nm para a excitação em 405 nm, as bandas em 530 e 550 nm para a excitação em 470 nm, bem como as três bandas de cada excitação e por fim todas as bandas espectrais. A Tabela 6.7 apresenta os resultados da classificação considerando as combinações mencionadas.

Tabela 6.7 - Desempenho do classificador ANN considerando as combinações para a detecção do greening.

	Excitação 405 nm		Excitação 470 nm		Combinação das duas excitações
	F560 + F580	F560 + F580 + F690	F530 + F550	F530 + F550 + F690	Todas as fluorescências
Sensibilidade	88,2%	87,3%	82,4%	89,2%	90,2%
Especificidade	55,1%	89,7%	82,1%	89,7%	97,4%
Acurácia	73,9%	88,3%	82,2%	89,4%	92,2%

Fonte: Elaborada pelo autor.

Os resultados da classificação apresentados na Tabela 6.7 mostram que o melhor desempenho do classificador ANN é obtido quando combinado todas as bandas espectrais, alcançando assim um acurácia de 92,2% com uma alta sensibilidade (90,2%) e especificidade (97,4%). A partir dessa combinação, somos capazes de identificar as amostras positivas para greening com uma boa precisão, assim como praticamente todas as amostras negativas para greening. Comparando os resultados de ambos os classificadores, podemos notar que para determinadas combinações das bandas espectrais a acurácia das redes neurais é maior, enquanto que para outras combinações a acurácia do SVM se sobressai. Na combinação das duas excitações, os resultados são praticamente iguais para os dois classificadores.

Em relação aos resultados obtidos no Brasil na classificação das amostras de greening e deficiência de zinco pelo método ANN, observamos também que o melhor resultado ficou por conta da combinação de todas as bandas espectrais com uma acurácia de 93,2%. A principal diferença observada está relacionada com a detecção de amostras positivas e negativas para greening. No Brasil, verificamos que todas as amostras positivas para greening

foram identificadas, enquanto que para os dados da Flórida identificamos praticamente todas as amostras negativas para greening.

6.2.3 Análise do desempenho dos classificadores baseada nas combinações das bandas espectrais do Brasil

Nas duas seções anteriores, apresentamos os resultados obtidos da classificação das imagens de fluorescência das amostras de greening e deficiência de zinco da Flórida através dos métodos SVM e ANN, considerando apenas as combinações definidas com base no critério da seleção das bandas espectrais com os melhores desempenhos. No entanto, as bandas com maiores acurácias para a Flórida são diferentes em relação às bandas selecionadas no Brasil. Sendo assim, resolvemos utilizar a combinação das bandas do Brasil a fim de verificar se os resultados da classificação sofrem alguma alteração significativa. No caso, a única combinação a ser avaliada consiste na combinação dos dois comprimentos de onda com as acurácias mais elevadas do Brasil, uma vez que as demais já foram analisadas. Desta forma, combinamos as bandas espectrais em 560 e 690 nm para a excitação em 405 nm, assim como as bandas em 530 e 690 nm para a excitação em 470 nm. As combinações foram avaliadas para os dois classificadores. A Tabela 6.8 apresenta os resultados do desempenho de ambos os classificadores adotando essa combinação.

Tabela 6.8 - Desempenho dos classificadores para as amostras da Flórida de acordo com a combinação das bandas espectrais do Brasil.

		SVM		ANN	
		Excitação 405 nm	Excitação 470 nm	Excitação 405 nm	Excitação 470 nm
		F560 +	F530 +	F560 +	F530 +
		F690	F690	F690	F690
Flórida	Sensibilidade	80,4%	86,3%	80,4%	83,3%
	Especificidade	96,2%	87,2%	89,7%	88,5%
	Acurácia	87,2%	86,7%	84,4%	85,6%

Fonte: Elaborada pelo autor.

A partir da Tabela 6.8 podemos observar que o desempenho de ambos os classificadores na detecção das amostras com greening da Flórida são significativamente superiores considerando as combinações de bandas espectrais definidas para as amostras do Brasil, principalmente para o método SVM. As acurácias obtidas pelo SVM, segundo as

combinações das duas bandas com os melhores desempenhos da Flórida, foram 68,9% e 75,6% para as excitações em 405 e 470 nm, respectivamente, conforme apresentado na Tabela 6.4. No caso, para o SVM segundo as combinações das bandas do Brasil, a acurácia passou para 87,2% e 86,7% para as excitações em 405 e 470 nm, respectivamente, um aumento significativo em relação ao resultado anterior. Este aumento também foi observado na classificação dos dados pelo método ANN, passando de 73,9% para 84,4% utilizando a excitação em 405 nm, e de 82,2% para 84,4% segundo a excitação em 470 nm. Este aumento no desempenho dos classificadores está relacionado, provavelmente, com o fato de que a contribuição do comprimento de onda em 690 nm é mais significativa do que observamos quando o analisamos individualmente, conforme os resultados apresentados nas Tabelas 6.3 e 6.6. Uma possível explicação para essa diferença é que as bandas espectrais foram selecionadas a partir da análise da intensidade média dos espectros obtidos de apenas 5 amostras através do SIF-1 com um desvio padrão relativamente alto, o que pode ter contribuído para a diferença notada. Assim como para o Brasil, a combinação dessas bandas espectrais se mostraram interessantes, no entanto, a análise de um maior número de amostras no experimento exploratório pode permitir a seleção de melhores comprimentos de onda e, consequentemente, aumentar o desempenho na classificação dos dados.

6.2.4 Discussão comparativa dos resultados dos classificadores

Os resultados discutidos para as técnicas de classificação SVM e ANN são promissores, uma vez que demonstramos que podem ser aplicadas na diferenciação das amostras com sintomas visuais semelhantes com uma precisão elevada de acordo com a estratégia adotada. Assim como no Brasil, a técnica de espectroscopia por imagens em conjunto métodos de classificação pode ser utilizada diretamente no campo para o controle das doenças.

Os experimentos conduzidos na Flórida produziram resultados interessantes. Primeiramente, a discriminação das amostras de greening e deficiência de zinco através da análise dos espectros das imagens de fluorescência obtidas através do SIF-1 permitiu a seleção das principais bandas de fluorescência. Nesta primeira etapa, verificamos que as bandas selecionadas são idênticas às determinadas na análise das amostras do Brasil. Esse resultado nos permite afirmar que uma vez definidas as bandas espectrais, não há a necessidade de conduzir novamente um experimento exploratório com o SIF-1 para condições diferentes, podendo assim aplicar diretamente o SIF-2 para as bandas definidas para o estudo

do greening. Estudos ainda devem ser realizados a fim de verificar a dimensão deste resultado, porém trata-se de um resultado muito promissor.

Com relação à classificação das imagens de fluorescência pelos métodos SVM e ANN, verificamos que os desempenhos dos classificadores foram baixos quando analisadas separadamente as bandas de fluorescência de cada excitação. Nessa primeira estratégia adotada, o desempenho do classificador ANN foi relativamente superior ao do SVM, mas a acurácia máxima atingida de 73,3% não é alta para um eventual uso da técnica. Sendo assim, adotamos a segunda estratégia que consistiu em combinar as diferentes excitações e bandas de fluorescência com os melhores desempenhos. O melhor resultado, no geral, foi obtido para o classificador SVM com uma acurácia de 92,8% contra 92,2% do ANN ao considerarmos a combinação das duas excitações e todas as bandas de fluorescência. No caso, a especificidade para essa combinação também foi maior para o SVM com 98,7% contra 97,4% do ANN, enquanto que a sensibilidade se manteve igual. Ambos os classificadores se mostraram muito eficientes na discriminação do greening. Por fim, resolvemos combinar as duas bandas espectrais definidas no Brasil para a discriminação do greening e deficiência de zinco da Flórida. Os resultados se mostraram superiores para os métodos SVM e ANN na classificação das amostras da Flórida considerando essa combinação do que a combinação das duas bandas com melhores desempenhos. No entanto, mesmo com o aumento significativo, o desempenho não atingiu os valores desejados para utilizarmos somente a combinação de duas bandas espectrais na detecção do greening.

Os resultados demonstram o potencial dos dois classificadores na identificação tanto de amostras de greening e deficiência de zinco do Brasil como da Flórida. No caso da classificação das amostras do Brasil, verificamos que a acurácia do classificador ANN (93,2%) se mostra superior ao SVM (88,9%), enquanto que para os dados da Flórida o SVM (92,8%) se sobressai minimamente ao ANN (92,2%). Um artigo com os resultados apresentados está em fase de submissão.

Isso posto, podemos concluir que o resultado obtido no experimento exploratório é condizente com o que verificamos através do desempenho dos classificadores SVM e ANN, uma vez que, assim como no Brasil, a combinação de todas as bandas produziu o melhor resultado na diferenciação do greening, demonstrando dessa forma a viabilidade de se aplicar o sistema em regiões com características diferentes. O fato de não termos obtidos bons resultados em nosso estudo preliminar, envolvendo a diferenciação de amostras sadias de amostras contaminadas com greening na Flórida (23), está provavelmente relacionado à banda

espectral e excitação utilizada, assim como a não realização da etapa de pré-processamento das imagens, a qual nos permite realizar a comparação da doença em condições diferentes.

Estudos ainda devem ser conduzidos a fim de melhorar ainda mais os resultados, porém trata-se de um resultado muito promissor que pode ser utilizado na diferenciação das amostras com greening de amostras com deficiência de zinco através da técnica de espectroscopia por imagens em conjunto com esses dois métodos de classificação. O provável uso de outras bandas espectrais e combinações, assim como o aumento do número de características extraídas das imagens podem contribuir para melhorar o desempenho dos classificadores.

7 CONCLUSÕES

Neste trabalho, propusemos o uso da técnica de imagens de fluorescência como meio para uma detecção rápida e eficiente das principais doenças que afetam os pomares de citros. Dois sistemas de imagens de fluorescência foram desenvolvidos, sendo o primeiro voltado principalmente para a seleção das principais bandas de emissão, e o segundo para a coleta de imagens de fluorescência para um grande número de amostras, através do uso de filtros ópticos passa-banda. Os sistemas desenvolvidos foram utilizados para a coleta de imagens de amostras de dois estados com características completamente diferentes, São Paulo/Brasil e Flórida/EUA. As doenças analisadas no Brasil foram o cancro cítrico, verrugose, greening e deficiência de zinco, enquanto nos Estados Unidos analisamos somente greening e deficiência de zinco. As principais características das imagens de fluorescência foram extraídas através de um algoritmo implementado no software Matlab[®] e empregamos os classificadores SVM e ANN.

Os resultados indicaram que as técnicas de classificação empregadas apresentam uma elevada precisão quando aplicadas em amostras de doenças diferentes, mas com sintomas visuais semelhantes. No caso das amostras de cancro cítrico e verrugose, a acurácia máxima obtida foi de 97,8% para o SVM e 97,2% para o método ANN, com sensibilidade de 100,0% e especificidade maior que 95,0% para ambas as técnicas. Esse resultado é relativo à combinação de todas as bandas espectrais definidas no experimento exploratório utilizando o SIF-1. Na discriminação das amostras de greening e deficiência de zinco tanto do Brasil como dos EUA, a acurácia máxima também foi obtida para a combinação de todas as bandas espectrais. No caso das amostras do Brasil, a acurácia foi de 88,9% para o SVM e de 93,2% para o ANN, mas a especificidade da técnica ANN é significativamente maior que a obtida para o SVM. No caso das amostras dos EUA, a acurácia foi de 92,8% para o SVM e 92,2% para o ANN, com altos valores de sensibilidade e especificidade. Vale destacar que os melhores resultados foram obtidos combinando todas as bandas espectrais definidas para todas as análises realizadas.

Sendo assim, primeiramente, podemos concluir que a técnica de espectroscopia por imagens de fluorescência em conjunto com estes dois métodos de classificação pode ser utilizada para a detecção e discriminação das amostras de doenças que apresentem sintomas visuais semelhantes. Acreditamos também, devido aos resultados obtidos na discriminação do greening da deficiência de zinco, que os sistemas desenvolvidos podem ser aplicados em qualquer região dos estados de São Paulo e Flórida, sem a necessidade de considerar as

diferentes características de cada região. Especulamos que as mesmas bandas devam funcionar em outros países, mas por enquanto não podemos afirmar isso. Uma vez que definimos as melhores bandas espectrais, estas podem ser aplicadas diretamente na análise do greening e da deficiência de zinco na região de estudo. Este resultado é interessante, uma vez que o tempo do processo como um todo diminui pois estaríamos utilizando somente o SIF-2 para a coleta das imagens. Estudos devem ser conduzidos a fim de verificar se estes resultados se aplicam também na discriminação das amostras de cancro cítrico e verrugose em outros países.

A partir do potencial apresentado por essa técnica, é possível, no futuro, o desenvolvimento de um único sistema onde imagens de fluorescência possam ser coletadas de várias amostras e processadas simultaneamente, reduzindo, conseqüentemente, ainda mais o tempo de diagnóstico. Trata-se simplesmente de uma questão de engenharia o desenvolvimento de tal sistema, visto que os princípios básicos foram demonstrados em nosso trabalho. Vale destacar o caráter multidisciplinar desta pesquisa, a qual envolve áreas como física, engenharia e computação.

Além disso, acreditamos também que o desempenho de ambos os classificadores possa ser melhorado através da combinação de outras bandas espectrais, assim como pela extração de um número maior de características das imagens. No entanto, o principal desafio, para o futuro, é a implementação de um algoritmo de classificação multiclases, o qual permitiria discriminar todas as doenças em estudo de uma única vez e não de maneira binária como realizamos neste trabalho. A técnica SVM utilizada foi originalmente formulada para a solução de problemas de classificação binária, entretanto, pesquisas recentes indicam o potencial considerável do SVM na classificação de dados multiclases (94). A generalização do SVM para problemas multiclases pode ser realizada basicamente por meio da combinação dos preditores gerados dos subproblemas binários (95). Em relação à técnica ANN, esta pode ser também aplicada na classificação multiclases, assim como métodos de árvore de decisão e classificadores probabilísticos convencionais, como a classificação da máxima verossimilhança (94). Acreditamos que a partir das técnicas mencionadas e das bandas espectrais selecionadas seria possível a implementação de um algoritmo de classificação multiclases das doenças analisadas neste trabalho. Provavelmente, teríamos que investir em um computador com nível de processamento maior em relação ao utilizado, uma vez que para o SVM, segundo Hsu e colaboradores (79), a reformulação dessa técnica em versões multiclases leva a algoritmos computacionalmente caros.

Por fim, o mais importante resultado deste trabalho está relacionado com o fato de que os sistemas e técnicas aqui desenvolvidos podem ser aplicados em outras culturas para o estudo de pragas, deficiências, desidratação de uvas, etc., necessitando apenas de pequenos ajustes conforme a cultura de interesse. Nesse sentido, vale destacar o trabalho de doutorado do aluno Rodolpho Tinini sob a orientação da Prof. Dra. Bárbara Teruel, o qual desenvolveu em colaboração com o nosso grupo de pesquisa um sistema de imagens espectrais para a supervisão do processo de desidratação moderada de uvas e predição do teor de sólidos solúveis (96). Através do sistema desenvolvido, constatou-se que a desidratação moderada apresenta potencial para ser monitorada através do uso de imagens de fluorescência, garantindo assim a qualidade do produto com a manutenção de propriedades valorizadas das uvas (96). Dessa forma, o trabalho aqui desenvolvido representa o estado da arte no que concerne o estudo das diferentes culturas agrícolas.

REFERÊNCIAS

- 1 CANCRO CÍTRICO: medidas essenciais de controle. Disponível em: <http://www.fundecitrus.com.br/pdf/manuais/manualZcancro_web_final.pdf>. Acesso em: 26 jan. 2016.
- 2 BRASIL. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. **Intercâmbio comercial do agronegócio : principais mercados de destino**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2014. ISBN 978-85-7991-090-6.
- 3 BALANÇA comercial resumida. 2015. Disponível em: <<http://www.agricultura.gov.br/internacional/indicadores-e-estatisticas/balanca-comercial>>. Acesso em: 02 fev. 2016.
- 4 OECD/FAO Agricultural Outlook 2015-2024. Disponível em: <http://www.keepeek.com/Digital-Asset-Management/oecd/agriculture-and-food/oecd-fao-agricultural-outlook-2015_agr_outlook-2015-en>. Acesso em: 03 fev. 2016.
- 5 NEVES, M. F. T., V. G.; MILAN, P.; LOPES, F. F.; CRESSONI, F.; KALAKI, R. **O retrato da citricultura brasileira**. Ribeirão Preto: Markestrat, 2010.
- 6 CITRUS. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/citrus>>. Acesso em: 02 fev. 2016.
- 7 LARANJA. Disponível em: <<http://www.investe.sp.gov.br/setores-de-negocios/agronegocios/laranja/>>. Acesso em: 02 fev. 2016.
- 8 PRODUÇÃO brasileira de laranja em 2013. 2013. Disponível em: <https://http://www.embrapa.br/documents/1355135/1529009/Laranja_Brasil_2013.pdf/5c85ffa4-f792-4db8-b1e7-2940d1cf07e5>. Acesso em: 02 fev. 2016.
- 9 PRODUÇÃO de laranjas de São Paulo é reestimada em 286,14 milhões de caixas. 2015. Disponível em: <<http://www.fundecitrus.com.br/comunicacao/noticias/integra/producao-de-laranjas-de-sao-paulo-e-reestimada-em-28614-milhoes-de-caixas/345>>. Acesso em: 02 fev. 2016.
- 10 BRAZIL - Citrus Annual Report. 2015. Disponível em: <[http://gain.fas.usda.gov/Recent GAIN Publications/Citrus Annual_Sao Paulo ATO_Brazil_12-21-2015.pdf](http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Citrus%20Annual%20Sao%20Paulo%20ATO%20Brazil_12-21-2015.pdf)>. Acesso em: 03 fev. 2016.
- 11 NEVES, M. B. **Simulação do impacto do Huanglongbing na produção de pomares cítricos sob diferentes cenários de manejo, incidências da doença e idades dos pomares no início da epidemia**. 2013. 71f. Dissertação (Mestrado em Fitossanidade) - Fundo de Defesa da Citricultura, Fundecitrus, Araraquara, 2013.
- 12 CITRUS: World Market and Trade. 2016. Disponível em: <<http://apps.fas.usda.gov/psdonline/circulars/citrus.pdf>>. Acesso em: 03 fev. 2016.

13 FLÓRIDA deve ter a menor safra em 50 anos. 2015. Disponível em: <<http://www.fundecitrus.com.br/comunicacao/noticias/integra/florida-deve-ter-a-menor-safra-em-50-anos/342>>. Acesso em: 02 fev. 2016.

14 RANULFI, A. C. **Utilização de técnicas espectroscópicas no estudo e caracterização de doenças em citros: HLB (greening) e cancro cítrico**. 2014. 137f. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2014.

15 BELASQUE JUNIOR, J.; BASSANEZI, R. B.; YAMAMOTO, P. T.; AYRES, A. J.; TACHIBANA, A.; VIOLANTE, A. R.; TANK JR., A.; DI GIORGI, F.; TERSI, F. E. A.; MENEZES, G. M.; DRAGONE, J.; JANK JR., R. H.; BOVÉ, J. M. Lessons from huanglongbing management in São Paulo State, Brazil. **Journal Of Plant Pathology**, v. 92, n. 2, p. 18, 2010.

16 BELASQUE JUNIOR, J.; YAMAMOTO, P. T.; MIRANDA, M. P. D.; BASSANEZI, R. B.; AYRES, A. N. J.; BOVÉ, J. M. Controle do huanglongbing no estado de São Paulo, Brasil. **Citrus Research & Technology**, v. 31, n. 1, p. 53-64, 2010.

17 GREENING. Disponível em: <<http://www.fundecitrus.com.br/doencas/greening/10>>. Acesso em: 03 fev. 2016.

18 LEVANTAMENTOS: Greening. Disponível em: <<http://www.fundecitrus.com.br/levantamentos/greening/10>>. Acesso em: 26 jan. 2016.

19 CITRICULTOR. 2015. Disponível em: <http://www.fundecitrus.com.br/pdf/revistas/ed31Zcitricultor_web.pdf>. Acesso em: 03 fev. 2016.

20 BEHLAU, F.; BELASQUE JUNIOR, J. B. **Cancro cítrico: a doença e seu controle**. Araraquara: Fundecitrus, 2014.

21 LEVANTAMENTOS: Cancro cítrico. Disponível em: <<http://www.fundecitrus.com.br/levantamentos/cancro-citrico/>>. Acesso em: 26 jan. 2016.

22 SÃO PAULO(ESTado). Resolução SAA nº 147, de 31-10-2013. 2013. Estabelece procedimentos a serem adotados no Estado de São Paulo visando a supressão/erradicação da praga denominada cancro cítrico, causado pela bactéria *Xanthomonas citri* subsp. *citri*. Disponível em: <http://sistemas.cda.sp.gov.br/greening/lnk_resolsaa147_ctr/lnk_resolsaa147_ctr.php>. Acesso em: 29 jan. 2016.

23 WETTERICH, C. B.; KUMAR, R.; SANKARAN, S.; BELASQUE JUNIOR, J.; EHSANI, R.; MARCASSA, L. G. A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of Huanglongbing citrus disease in the USA and Brazil. **Journal of Spectroscopy**, v. 2013, p. 6, 2013. doi: 10.1155/2013/841738.

24 SANKARAN, S.; MISHRA, A.; EHSANI, R.; DAVIS, C. A review of advanced techniques for detecting plant diseases. **Computers and Electronics in Agriculture**, v. 72, n. 1, p. 1-13, 2010.

- 25 WETTERICH, C. B.; FELIPE DE OLIVEIRA NEVES, R.; BELASQUE JUNIOR, J.; MARCASSA, L. G. Detection of citrus canker and Huanglongbing using fluorescence imaging spectroscopy and support vector machine technique. **Applied Optics**, v. 55, n. 2, p. 400-407, 2016.
- 26 CHAERLE, L.; LENK, S.; HAGENBEEK, D.; BUSCHMANN, C.; VAN DER STRAETEN, D. Multicolor fluorescence imaging for early detection of the hypersensitive reaction to tobacco mosaic virus. **Journal of Plant Physiology**, v. 164, n. 3, p. 253-262, 2007.
- 27 PEREIRA, F. M. V.; MILORI, D. M. B.; PEREIRA-FILHO, E. R.; VENANCIO, A. L.; RUSSO, M. T.; MARTINS, P. K.; FREITAS-ASTUA, J. Fluorescence images combined to statistic test for fingerprinting of citrus plants after bacterial infection. **Analytical Methods**, v. 3, n. 3, p. 552-556, 2011.
- 28 LINS, E. C.; BELASQUE JUNIOR, J.; MARCASSA, L. G. Optical fiber laser induced fluorescence spectroscopy as a citrus canker diagnostic. **Applied Optics**, v. 49, n. 4, p. 663-667, 2010.
- 29 WETTERICH, C. B. **Uso de imagens de fluorescência para monitoramento da evolução do cancro cítrico**. 2012. 92f. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.
- 30 LEE, W. S.; EHSANI, R. Sensing systems for precision agriculture in Florida. **Computers and Electronics in Agriculture**, v. 112, p. 2-9, 2015. doi: 10.1016/j.compag.2014.11.005.
- 31 SCHAAD, N. W.; JONES, J. B.; CHUN, W. **Laboratory guide for identification of plant pathogenic bacteria**. 3th ed. Minnesota, USA: APS Press, 2001. 398.
- 32 TEIXEIRA, D. C.; SAILLARD, C.; COUTURE, C.; MARTINS, E. C.; WULFF, N. A.; EVEILLARD-JAGOUÉIX, S.; YAMAMOTO, P. T.; AYRES, A. J.; BOVÉ, J. M. Distribution and quantification of *Candidatus Liberibacter americanus*, agent of huanglongbing disease of citrus in São Paulo State, Brasil, in leaves of an affected sweet orange tree as determined by PCR. **Molecular and Cellular Probes**, v. 22, n. 3, p. 139-150, 2008.
- 33 FERNANDES, N. G.; BARBOSA, J. C.; AYRES, A. J.; MASSARI, C. A. Plantas doentes não detectadas nas inspeções dificultam a erradicação do cancro cítrico. **Summa Phytopathol**, v. 26, p. 6, 2000. doi:10.5935/2236-3122.20110021
- 34 BELASQUE JUNIOR, J.; BERGAMIN FILHO, A.; BASSANEZI, R. B.; BARBOSA, J. C.; FERNANDES, N. G.; YAMAMOTO, P. T.; LOPES, S. A.; MACHADO, M. A.; LEITE JUNIOR, R. P.; AYRES, A. J.; MASSARI, C. A. Base científica para a erradicação de plantas sintomáticas e assintomáticas de Huanglongbing (HLB, Greening) visando o controle efetivo da doença. **Tropical Plant Pathology**, v. 34, n.3, p. 137-145, 2009.
- 35 LINS, E.; BELASQUE JUNIOR, J.; MARCASSA, L. Detection of citrus canker in citrus plants using laser induced fluorescence spectroscopy. **Precision Agriculture**, v. 10, n. 4, p. 319-330, 2009.

- 36 POURREZA, A.; LEE, W. S.; ETXEBERRIA, E.; BANERJEE, A. An evaluation of a vision-based sensor performance in Huanglongbing disease identification. **Biosystems Engineering**, v. 130, p. 13-22, 2015.
- 37 MARCASSA, L. G.; GASPAROTO, M. C. G.; BELASQUE JUNIOR, J.; LINS, E. C.; DIAS NUNES, F.; BAGNATO, V. S. Fluorescence spectroscopy applied to orange trees. **Laser Physics**, v. 16, n. 5, p. 884-888, 2006.
- 38 POURREZA, A.; LEE, W. S.; EHSANI, R.; SCHUELLER, J. K.; RAVEH, E. An optimum method for real-time in-field detection of Huanglongbing disease using a vision sensor. **Computers and Electronics in Agriculture**, v. 110, n. 2015, p. 221-232, 2015.
- 39 NAIDU, R. A.; PERRY, E. M.; PIERCE, F. J.; MEKURIA, T. The potential of spectral reflectance technique for the detection of Grapevine leafroll-associated virus-3 in two red-berried wine grape cultivars. **Computers and Electronics in Agriculture**, v. 66, n. 1, p. 38-45, 2009.
- 40 MISHRA, A. R.; KARIMI, D.; EHSANI, R.; LEE, W. S. Identification of Citrus Greening (HLB) using a vis-nir spectroscopy technique. **Transactions of the Asabe**, v. 55, n. 2, p. 711-720, 2012.
- 41 WETTERICH, C. B.; KUMAR, R.; SANKARAN, S.; BELASQUE JUNIOR, J.; EHSANI, R.; MARCASSA, L. G. A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of Huanglongbing citrus disease in the USA and Brazil. **Journal of Spectroscopy**, v. 2013, p. 6, 2013. doi: 10.1155/2013/841738.
- 42 BELASQUE JUNIOR, J.; GASPAROTO, M. C. G.; MARCASSA, L. G. Detection of mechanical and disease stresses in citrus plants by fluorescence spectroscopy. **Applied Optics**, v. 47, n. 11, p. 1922-1926, 2008.
- 43 GARCIA-RUIZ, F.; SANKARAN, S.; MAJA, J. M.; LEE, W. S.; RASMUSSEN, J.; EHSANI, R. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. **Computers and Electronics in Agriculture**, v. 91, p. 106-115, 2013.
- 44 CEROVIC, Z. G.; SAMSON, G.; MORALES, F.; TREMBLAY, N.; MOYA, I. Ultraviolet- induced fluorescence for plant monitoring: present state and prospects. **Agronomie**, v. 19, n. 7, p. 36, 1999.
- 45 LINS, E. C. C. C. **Espectroscopia da fluorescência na citricultura**. 2009. 116p. Tese (Doutorado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2009.
- 46 STREIT, N. M.; CANTERLE, L. P.; CANTO, M. W. D.; HECKTHEUER, L. H. H. As clorofilas. **Ciência Rural**, v. 35, n.3, p. 748-755, 2005.
- 47 VIEIRA, E. L.; SOUZA, G. S.; SANTOS, A. R.; SILVA, J. S. **Manual de fisiologia vegetal**. São Luis: EDUFMA, 2010.

- 48 CASAROLI, D.; FAGAN, E. B.; SIMON, J.; MEDEIROS, S. P.; MANFRON, P. A.; DOURADO NETO, D.; LIER, Q. D. J. V.; MÜLLER, L.; MARTIN, T. N. Radiação solar e aspectos fisiológicos na cultura de soja - uma revisão. **Revista FZVA**, v. 14, n. 2, p. 102-120, 2007.
- 49 DAVID, W. L. **Photosynthesis**. 3th ed. New York: BIOS Scientific Publishers, 2001. ISBN 9781859961575.
- 50 CARDINALI, M. C. D. B. **Diagnóstico de Huanglongbing (HLB) em citros utilizando técnicas fotônicas**. 2012. 117f. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.
- 51 LENK, S.; CHAERLE, L.; PFÜNDEL, E. E.; LANGSDORF, G.; HAGENBEEK, D.; LICHTENTHALER, H. K.; VAN DER STRAETEN, D.; BUSCHMANN, C. Multispectral fluorescence and reflectance imaging at the leaf level and its possible applications. **Journal of Experimental Botany**, v. 58, n. 4, p. 807-814, 2007.
- 52 MANICKAVASAGAN, A.; JAYASURIYA, H. **Imaging with electromagnetic spectrum: applications in food and agriculture**. Berlin: Springer Verlag, 2014. 204p.
- 53 KUMAR, A.; LEE, W. S.; EHSANI, R. J.; ALBRIGO, L. G.; YANG, C. H.; MANGAN, R. L. Citrus greening disease detection using aerial hyperspectral and multispectral imaging techniques. **Journal of Applied Remote Sensing**, v. 6, n.1, 2012. doi: 10.1117/1.JRS.6.063542
- 54 USHA, K.; SINGH, B. Potential applications of remote sensing in horticulture - a review. **Scientia Horticulturae**, v. 153, n. 0, p. 71-83, 2013.
- 55 MO, C.; KIM, M. S.; KIM, G.; CHEONG, E. J.; YANG, J.; LIM, J. Detecting drought stress in soybean plants using hyperspectral fluorescence imaging. **Journal of Biosystems Engineering**, v. 40, n. 4, p. 335-344, 2015.
- 56 POURREZA, A.; LEE, W. S.; RAVEH, E.; EHSANI, R.; ETXEBERRIA, E. Citrus Huanglongbing detection using narrow-band imaging and polarized illumination. **Transactions of the Asabe**, v. 57, n. 1, p. 259-272, 2014.
- 57 KURTULMUS, F.; LEE, W. S.; VARDAR, A. Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network. **Precision Agriculture**, v. 15, n. 1, p. 57-79, 2014.
- 58 MELGANI, F.; BRUZZONE, L. Classification of hyperspectral remote sensing images with support vector machines. **IEEE Transactions on Geoscience and Remote Sensing**, v. 42, n. 8, p. 1778-1790, 2004.
- 59 CAMPS-VALLS, G.; BRUZZONE, L. Kernel-based methods for hyperspectral image classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 43, n. 6, p. 1351-1362, 2005.

- 60 SANKARAN, S.; MAJA, J. M.; BUCHANON, S.; EHSANI, R. Huanglongbing (Citrus Greening) detection using visible, near infrared and thermal imaging techniques. **Sensors**, v. 13, n. 2, p. 2117-2130, 2013.
- 61 ZHAO, J.; CHEN, Q.; CAI, J.; OUYANG, Q. Automated tea quality classification by hyperspectral imaging. **Applied Optics**, v. 48, n. 19, p. 3557-3564, 2009.
- 62 VAPNIK, V. N. **The nature of statistical learning theory**. New York, USA: Springer-Verlag, 1995.
- 63 SENGUPTA, S.; LEE, W. S. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. **Biosystems Engineering**, v. 117, p. 51-61, 2014. doi:10.1016/j.biosystemseng.2013.07.007.
- 64 LIQUID Crystal tunable filter: user manual. Disponível em: < http://www.meadowlark.com/store/D4020_Users_manual_1.03.pdf >. Acesso em: 29 dez. 2015.
- 65 GOLMOHAMMADI, M.; CUBERO, J.; PENALVER, J.; QUESADA, J. M.; LOPEZ, M. M.; LLOP, P. Diagnosis of *Xanthomonas axonopodis* pv. *citri*, causal agent of citrus canker, in commercial fruits by isolation and PCR-based methods. **Journal of Applied Microbiology**, v. 103, n. 6, p. 2309-2315, 2007.
- 66 CUBERO, J.; GRAHAM, J. H.; GOTTWALD, T. R. Quantitative PCR method for diagnosis of citrus bacterial canker. **Applied and Environmental Microbiology**, v. 67, n. 6, p. 2849-2852, 2001.
- 67 MAVRODIEVA, V.; LEVY, L.; GABRIEL, D. W. Improved sampling methods for real-time polymerase chain reaction diagnosis of citrus canker from field samples. **Phytopathology**, v. 94, n. 1, p. 61-68, 2004.
- 68 PYDIPATI, R.; BURKS, T. F.; LEE, W. S. Statistical and neural network classifiers for citrus disease detection using machine vision. **Transactions of the ASABE**, v. 48, n. 5, p. 2007-2014, 2005.
- 69 HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. **IEEE Transactions on Systems, Man and Cybernetics**, v. SMC-3, n. 6, p. 12, 1973.
- 70 PYDIPATI, R.; BURKS, T. F.; LEE, W. S. Identification of citrus disease using color texture features and discriminant analysis. **Computers and Electronics in Agriculture**, v. 52, n. 1-2, p. 49-59, 2006.
- 71 ELEYAN, A.; DEMIREL, H. Co-occurrence matrix and its statistical features as a new approach for face recognition. **Turkish Journal of Electrical Engineering and Computer Sciences**, v. 19, n. 1, p. 97-107, 2011.
- 72 MATRIZES de Co-Ocorrência. Disponível em: < <http://www2.ic.uff.br/~aconci/co-ocorrencia.pdf> >. Acesso em: 05 jan. 2016.

- 73 BARALDI, A.; PARMIGGIANI, F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. **IEEE Transactions on Geoscience and Remote Sensing**, v. 33, n. 2, p. 12, 1995.
- 74 LIANG, L.; ZHANG, T.; WANG, K.; TANG, H.; YANG, X.; ZHU, X.; DUAN, Y.; LI, H. Classification of steel materials by laser-induced breakdown spectroscopy coupled with support vector machines. **Applied Optics**, v. 53, n. 4, p. 544-552, 2014.
- 75 LI, L.; LIU, D.; CAO, P.; XIE, S.; LI, Y.; CHEN, Y.; YANG, Y. Automated discrimination between digs and dust particles on optical surfaces with dark-field scattering microscopy. **Applied Optics**, v. 53, n. 23, p. 5131-5140, 2014.
- 76 SONG, S.; ZHAN, Z.; LONG, Z.; ZHANG, J.; YAO, L. Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. **Plos One**, v. 6, n. 2, p. 11, 2011.
- 77 MUCHERINO, A.; PAPAJOGEI, P.; PARDALOS, P. Support vector machines. In: _____ (Ed.) **Data mining in agriculture**. New York: Springer Verlag, 2009. v.34, cap. 6, p.123-141. (Springer Optimization and Its Applications).
- 78 QI, B.; ZHAO, C.; YIN, G. Feature weighting algorithms for classification of hyperspectral images using a support vector machine. **Applied Optics**, v. 53, n. 13, p. 2839-2846, 2014.
- 79 HSU, C. W.; CHANG, C. C.; LIN, C. J. **A practical guide to support vector classification**. 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.224.4115>>. Acesso em: 11 fev. 2016.
- 80 HERNÁNDEZ SUÁREZ, M.; ASTRAY DOPAZO, G.; LARIOS LÓPEZ, D.; ESPINOSA, F. 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**, v. 10, n. 6, p. e0128566, 2015.
- 81 JAIN, A. K.; JIANCHANG, M.; MOHIUDDIN, K. M. Artificial neural networks: a tutorial. **Computer**, v. 29, n. 3, p. 31-44, 1996.
- 82 GARDNER, M. W.; DORLING, S. R. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. **Atmospheric Environment**, v. 32, n. 14-15, p. 2627-2636, 1998.
- 83 CHRISTOPHER, M. B. **Neural networks for pattern recognition**. Oxford: Oxford University Press, Inc., 1995. 482p.
- 84 CHANG, C. C.; LIN, C. J. **LIBSVM** - a library for support vector machines. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>>. Acesso em: 11 fev. 2016.

85 CITRUS RESEARCH AND EDUCATION CENTER. **History**. Disponível em: <<http://www.crec.ifas.ufl.edu/about/History/>>. Acesso em: 11 jan. 2016.

86 ANÁLISE do poder discriminativo através de curvas ROC. Disponível em: <<http://crsouza.com/2009/07/analise-de-poder-discriminativo-atraves-de-curvas-roc/>>. Acesso em: 12 jan. 2016.

87 BRAGA, A. C. D. S. **Curvas ROC: aspectos funcionais e aplicações**. 2000. 243f. Tese (Doutorado em Métodos numéricos e estatísticos) - Departamento de Engenharia de Produção e Sistemas, Universidade do Minho, Braga, 2000.

88 ARTIGOS sobre testes diagnósticos. Disponível em: <http://www.sboc.org.br/app/webroot/leitura-critica/LEITURA-CRITICA_C5.pdf>. Acesso em: 12 jan. 2016.

89 PRECISION-RECALL. Disponível em: <http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html>. Acesso em: 12 jan. 2016.

90 DIFFERENCES between Receiver Operating Characteristic AUC (ROC AUC) and Precision Recall AUC (PR AUC). Disponível em: <<http://www.chioka.in/differences-between-roc-auc-and-pr-auc>>. Acesso em: 12 jan. 2016.

91 PEREIRA, F. M. V.; MILORI, D. M. B. P.; PEREIRA, E. R.; VENANCIO, A. L.; RUSSO, M. D. T.; CARDINALI, M. C. D.; MARTINS, P. K.; FREITAS-ASTUA, J. Laser-induced fluorescence imaging method to monitor citrus greening disease. **Computers and Electronics in Agriculture**, v. 79, n. 1, p. 90-93, 2011.

92 LORENA, A. C.; CARVALHO, A. C. P. L. F. D. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 45-67, 2007.

93 SAURABH, K. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. **International Journal of Engineering Trends and Technology**, v.3, n.6, p.714-717, 2012.

94 PEREIRA, G. H. A. C., J. A. S. Utilização de support vector machine para classificação multiclases de imagens Landsat TM+. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16, 2013. Foz do Iguaçu. **Anais...Foz do Iguaçu: INPE**, 2013. p.2322-2328.

95 LORENA, A. C. **Investigação de estratégias para a geração de máquinas de vetores de suporte multiclases**. 2006. 203f. Tese (Doutor em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2006.

96 TININI, R. C. S. D. R. **Avaliação não destrutiva de sólidos solúveis por meio de imagens espectrais de uvas submetidas à desidratação parcial**. 2015. 135f. Tese (Doutor em Engenharia Agrícola) - Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2015.

- 97 GOTTWALD, T. A.; DA GRAÇA, J. V.; BASSANEZI, R. B. **Citrus huanglongbing**: the pathogen and its impact. 2007. Disponível em: <http://www.ars.usda.gov/SP2UserFiles/Place/66180000/Gottwaldetal._2007HLB-Thepathogenanditsimpact.PlantHealthProgress.pdf>. Acesso em: 11 fev. 2016.
- 98 ZAGHI, A. F. **Diagnóstico de doenças de citros utilizando espectroscopia de fluorescência**. 2008. 101f. Dissertação (Mestrado em Química Analítica) - Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 99 A CITRICULTURA em risco – o pesadelo da Greening. Disponível em: <<http://www.agrosoft.org.br/agropag/103543.htm>>. Acesso em: 25 jan. 2016.
- 100 MANUAL técnico: Greening. 2009. Disponível em: <http://www.citrusbr.com/manuaistecnicos/fundecitrus_greening.pdf>. Acesso em: 05 fev. 2016.
- 101 BOVÉ, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. **Journal of Plant Pathology**, v. 88, n. 1, p. 7-37, 2006.
- 102 FMC lança inseticida para citricultura. 2015. Disponível em: <<http://www.jornalentreposto.com.br/agricola/36-agroquimica/1184-fmc-lanca-inseticida-para-citricultura>>. Acesso em: 05 fev. 2016.
- 103 SOUTH AFRICAN citrus growers experience rollercoaster season. 2012. Disponível em: <<http://www.freshfruitportal.com/news/2012/07/23/south-african-citrus-growers-experience-rollercoaster-season/>>. Acesso em: 05 fev. 2016.
- 104 MANUAL técnico: cancro cítrico. 2008. Disponível em: <http://www.citrusbr.com/manuaistecnicos/fundecitrus_cancrocitrico.pdf>. Acesso em: 06 fev. 2016.
- 105 CANCRO cítrico. Disponível em: <<http://www.fundecitrus.com.br/doencas/cancro/7>>. Acesso em: 06 fev. 2016.
- 106 ALVES, W. A. L.; ARAÚJO, S. A. D.; LIBRANTZ, A. F. H. Reconhecimento de padrões de texturas em imagens digitais usando uma rede neural artificial híbrida. **Exacta**, v. 4, n. 2, p. 325-332, 2006.
- 107 CLASSIFICAÇÃO de texturas invariante a rotação usando matriz de co-ocorrência. Disponível em: <<http://www.lps.usp.br/hae/Ito-Kim-Salcedo-Textura.pdf>>. Acesso em: 08 jun. 2016.
- 108 GRAYCOMATRIX. Disponível em: <<http://www.mathworks.com/help/images/ref/graycomatrix.html?requestedDomain=www.mathworks.com>>. Acesso em: 08 jun. 2016.

APÊNDICE A – Principais doenças que afetam a citricultura

As pragas e doenças foram responsáveis pela erradicação de 40 milhões de árvores nesta última década, atingindo perdas de quase 80 milhões de caixas por ano. A taxa de mortalidade das árvores citrícolas saltou de 4% para 7,5%, valor considerado alarmante pelos produtores.(5) Neste contexto, uma das preocupações mais sérias do setor são os prejuízos causados pelas doenças greening e cancro cítrico. As principais características de cada doença, assim como suas formas de manejo e controle serão apresentadas neste Apêndice.

O greening é a mais importante e destrutiva doença que afeta a citricultura mundial, capaz de provocar perdas irreparáveis nos pomares dos produtores. Os primeiros indícios da doença foram relatados no sul da China no final do século XIX (97) e, desde então, atingiu as diferentes regiões produtoras de citros nos continentes Asiático, Africano, Americano e Oceania.(17,98) A doença foi identificada pela primeira vez nos dois maiores produtores mundiais de suco de laranja, São Paulo (Brasil) e Flórida (Estados Unidos), em 2004 e 2005, respectivamente.(16) No Brasil, as primeiras plantas com sintomas do greening foram encontradas na região de Araraquara, São Paulo.(15) Hoje, está presente em todas as regiões citrícolas de São Paulo, assim como em pomares de Minas Gerais e Paraná.(17) Até o momento não existe variedade comercial de copa ou porta-enxerto resistente à doença, assim como nenhum método de controle curativo.(16,17) Desta forma, prevenir a infecção das plantas é fundamental para o controle da doença.

O agente causal do greening é uma bactéria conhecida como *Candidatus Liberibacter* spp., que habita e se desenvolve no floema da planta hospedeira (15) e, possui três diferentes espécies patogênicas: *Candidatus Liberibacter asiaticus*, *Candidatus Liberibacter africanus* e *Candidatus Liberibacter americanus*.(99) Esta última espécie, mundialmente desconhecida até sua descoberta, foi encontrada nos pomares paulistas.(100) No entanto, no Brasil prevalece a forma asiática.(100)

No Brasil, as duas espécies da bactéria são transmitidas pelo psíldeo (inseto vetor) *Diaphorina citri*, um pequeno inseto de coloração cinza e com manchas escuras na asa, cujo tamanho na fase adulta pode variar de 2 a 3 mm de comprimento.(14,100) A Figura 1.a apresenta a imagem da *Diaphorina citri*. A bactéria não é disseminada pelo vento, água ou qualquer outro instrumento agrícola. Os psíldeos diferem dos demais insetos por permanecerem nas folhas e ramos em uma inclinação de 45 graus.(100) Tanto as bactérias quanto o psíldeo são encontrados na planta ornamental *Murraya* spp., conhecida como murta ou falsa murta.(100) O inseto vetor também pode se hospedar em todas as variedades cítricas,

assim como pode ser encontrado em todas as regiões citrícolas do país, pois não apresenta limitações climáticas nas nossas condições.(100) O psilídeo adquire a bactéria ao se alimentar em plantas contaminadas, geralmente brotos novos, onde também coloca seus ovos e ocorre o desenvolvimento das ninfas.(17,100) Insetos adultos, provenientes de ninfas criadas em plantas doentes, transmitem a bactéria com mais eficiência do que psilídeos que adquirem a bactéria somente na fase adulta (17). Este fato justifica, portanto, a necessidade de se eliminar as árvores assim que detectados os primeiros sintomas, evitando consequentemente o crescimento do inseto em plantas doentes. A bactéria obstrui o floema da planta hospedeira, impedindo a distribuição da seiva elaborada e prejudicando seu desenvolvimento.(100) Sendo assim, plantas novas contaminadas não chegam a produzir, enquanto que plantas adultas se tornam improdutivas no período de 2 a 5 anos.(100)

O período de incubação do greening, ou seja, período entre a infecção e o aparecimento dos primeiros sintomas varia, geralmente, de seis a dozes meses e, depende de diversos fatores como o ambiente (principalmente temperatura), idade e variedade da planta hospedeira, concentração bacteriana na transmissão, espécie de *Candidatus Liberibacter*, época do ano, entre outros.(34,101) O sintoma inicial do greening é mais característico que vem sendo observado independente do agente causal, do hospedeiro e da condição ambiental, é o surgimento de um ou mais ramos que se destacam pela presença de folhas de cor amarela em contraste com a coloração verde das folhas dos ramos não afetados, como apresentado na Figura 1.b.(34, 100) Com a evolução da doença, os sintomas começam a aparecer em outros ramos da planta até atingir toda a copa, podendo evoluir para a desfolha e morte de ponteiros, como ilustrado da Figura 1.c.(14,34,100) Os frutos de ramos sintomáticos não amadurecem normalmente e adquirem uma coloração verde clara manchada, apresentando-se deformados, pequenos e assimétricos, diferentemente de frutos de ramos saudáveis (17,34), assim como apresentado nas Figuras 1.d e 1.e. É comum ocorrer a queda precoce de frutos sintomáticos em plantas doentes. Além da queda, os frutos apresentam-se mais ácidos e com menores valores de *Brix*, reduzindo assim a qualidade do suco.(34)

Os sintomas do greening podem ser confundidos com sintomas de algumas deficiências nutricionais como de zinco, manganês, magnésio ou cobre, e até com a doença conhecida como gomose.(100) A Figura 1.f apresenta folhas com deficiência de zinco a fim de ilustrar a semelhança entre os sintomas visuais. A dificuldade em diferenciar o greening destas deficiências aliada à subjetividade humana na inspeção e ao longo período de incubação da doença pode levar ao diagnóstico incorreto e consequentemente favorecer a disseminação da doença.

Figura A. 1 - (a) *Diaphorina citri* adulto se alimentando em uma planta hospedeira; (b) Sintomas característicos do greening; (c) Desfolha e morte de ponteiros em estágio mais avançado da doença; (d) Fruto doente com coloração verde clara, deformado, pequeno e assimétrico; (e) Frutos de ramos saudáveis; (f) Folhas de citros com sintomas de deficiência de zinco.
 Fonte: (a) FMC...(102); (b) MANUAL...(99); (c) GOTTWALD (96); (d) GREENING (16); (e) SOUTH...(103); (f) CITRUS...(103)

Atualmente, considerando a fácil disseminação da doença e a inexistência de métodos curativos, o manejo integrado é a forma mais eficiente de combate ao greening e é baseado em cinco medidas fundamentais as quais devem ser adotadas por todos os citricultores de maneira coletiva.(100) As cinco medidas do manejo integrado são: inspeção frequente de

todas as plantas do pomar, erradicação das plantas com sintomas, monitoramento e controle do psíldeo, aquisição de mudas sadias produzidas em viveiros devidamente cadastrados na Coordenadoria de Defesa Agropecuária (CDA) e, por fim, eliminação das plantas de falsa murta.(17,100). O método analítico geralmente empregado no diagnóstico da doença é a técnica de Reação em Cadeia da Polimerase Quantitativo (qPCR, do inglês *quantitative Polymerase Chain Reaction*), que consiste na amplificação do DNA da bactéria causadora da doença.(14,24,31-32) No entanto, essas medidas requerem tempo, recursos humanos e financeiros, assim como uma ação conjunta entre governo e integrantes da cadeia citrícola para o controle dessa doença que coloca em risco esse importante setor do agronegócio brasileiro.(16,34)

Com relação ao cancro cítrico, esta doença afeta todas as espécies e variedades de citros de importância comercial, havendo apenas uma variação quanto à susceptibilidade.(102-103) A doença teve origem no continente Asiático e foi identificada no Brasil pela primeira vez na região de Presidente Prudente, São Paulo, em 1957.(20) Atualmente, a doença pode ser encontrada em diferentes estados do Brasil (Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná e Rio Grande do Sul) (20), assim como em mais de trinta países incluindo Ásia, ilhas dos oceanos Pacífico e Índico, América do Sul e Estados Unidos.(14) A principal forma de controle da doença é a erradicação das plantas contaminadas.(20)

O cancro cítrico é causado pela bactéria *Xanthomonas citri* subsp. *citri* e seus impactos para a citricultura estão relacionados aos prejuízos à produção devido à depreciação da qualidade; à restrição da comercialização impedindo o transporte dos frutos originários de áreas afetadas e, principalmente, à redução pela queda de frutos afetados pela doença.(1,14) Estes danos levam ao aumento considerável dos gastos de produção nas áreas onde são implementadas medidas de prevenção, quarentena, erradicação e controle da doença.

O ciclo epidemiológico do cancro cítrico é constituído de quatro etapas: disseminação, infecção, colonização e sobrevivência.(20) A disseminação da bactéria pode ocorrer por intermédio do homem no trânsito de mudas, material de colheita, implementos e veículos contaminados, e pela ação da natureza, por meio de chuvas acompanhadas de vento.(1,20,103) Eventos naturais como chuvas associadas a ventos são a principal forma de disseminação da bactéria nas áreas citrícolas, podendo deslocá-la por até centenas de metros a partir da planta afetada.(1,20) A disseminação da bactéria não garante a ocorrência do cancro cítrico, sendo necessária a penetração desta nos tecidos da planta para sua infecção. As formas mais comuns de penetração são por aberturas naturais, como os estômatos (tecido responsável

pelas trocas gasosas da planta); por ferimentos mecânicos, causados por equipamentos, atrito entre partes da própria planta, abrasão de partículas de poeiras, dentre outros; e ainda por ferimentos causados por insetos, como o minador dos citros, que ao se alimentar de folhas novas provoca aberturas que facilitam a penetração.(20,103) A presença de água na superfície da planta é o fator ambiental mais importante também nesta fase.(1) Ao final da fase de infecção inicia-se a última etapa do processo de desenvolvimento do cancro cítrico na planta, a colonização, caracterizada pela reprodução da bactéria dentro da planta e pelo início da doença propriamente dita.(20,103) É nesta fase que surgem os sintomas característicos do cancro cítrico e também o aumento da população bacteriana, que se multiplicam localmente, próximo ao ponto de penetração, de onde liberam enzimas para a degradação das células das plantas e obtenção de nutrientes.(20,103) A faixa mais favorável para a colonização e reprodução está situada entre 25°C e 35°C.(20) A sobrevivência e permanência da doença no pomar ao longo dos anos ocorre pelas lesões remanescentes de cancro cítrico em frutos, folhas e ramos da planta.(103) Na presença de água, ocorre a exsudação bacteriana das lesões e a disseminação da doença para tecidos saudáveis da mesma árvore ou pomares próximos, iniciando-se assim um novo ciclo.(20) A Figura 2 apresenta o ciclo epidemiológico do cancro cítrico destacando as quatro etapas envolvidas.

Figura A. 2 - Ciclo epidemiológico do cancro cítrico: (a) disseminação, (b) infecção, (c) colonização e (d) sobrevivência.

Fonte: CANCRO...(1)

Os primeiros sintomas da doença tornam-se visíveis de duas a cinco semanas após a infecção da planta, e podem ocorrer em folhas, frutos e, de maneira menos frequente, em ramos.(1,20) Inicialmente, os sintomas em folhas apresentam-se como pontos escurecidos e, geralmente, com amarelecimento ao redor, evoluindo para pústulas salientes de coloração marrom clara. Com o progresso da doença, essas lesões tornam-se maiores - podendo atingir mais de um centímetro de diâmetro -, circulares e, na maioria das vezes, apresenta um halo amarelado.(1) Nas folhas, as lesões são visíveis nos dois lados da folha, entretanto, podem ser observadas inicialmente na face inferior devido ao maior número de estômatos.(20) Os sintomas em frutos são semelhantes aos observados em folhas, sendo comum a ocorrência de sintomas na face voltada para o exterior da copa da planta, uma vez que há uma maior exposição desse lado a chuva e ventos, fatores que aumentam a predisposição à doença.(1) As lesões de cancro cítrico não afetam as características internas do fruto.(20) No caso dos ramos, as lesões apresentam as mesmas características observadas em folhas e frutos, como saliência e coloração marrom.(1) Embora menos frequente, dependendo da susceptibilidade da variedade da planta, as lesões em ramos podem resultar em rachaduras e, conseqüentemente, seca e morte destes, prejudicando assim o desenvolvimento da planta nos primeiros anos.(103)

O cancro cítrico apresenta sintomas característicos, no entanto, pode ser confundido com outras doenças da cultura dos citros. Dentre essas, destacam-se as doenças bacterianas como cancriose dos citros e a mancha bacteriana dos citros, e as doenças não bacterianas, como a leprose dos citros e verrugose.(20) Os sintomas semelhantes com estas doenças, principalmente com a verrugose, dificulta a inspeção visual do cancro cítrico. A Figura 3 apresenta os sintomas característicos do cancro cítrico, assim como ilustra sua semelhança com a doença verrugose.

Existem duas estratégias globais de controle para o cancro cítrico. A primeira baseia-se na exclusão e erradicação da bactéria, cujo objetivo consiste em manter os pomares de uma determinada região livres da doença ou com incidência muito baixa de plantas doentes.(1,20) De acordo com a resolução SAA 147, de 31 de outubro de 2013, deve-se realizar, como já mencionado, a eliminação apenas da planta com sintomas de cancro cítrico e pulverizar com cobre as demais plantas em um raio de 30 metros, sendo repetida a cada brotação.(1) Além disso, o citricultor deve realizar, no mínimo, uma inspeção trimestral em todas as plantas de citros e informar a Secretaria de Agricultura por meio de relatórios semestrais.(1) A segunda estratégia consiste no manejo integrado do cancro cítrico, cuja finalidade é reduzir danos e perdas em áreas com ampla disseminação da doença.(20) Independente da estratégia do

cancro cítrico ser a erradicação ou o manejo integrado, o produtor deverá adotar um conjunto de medidas de controle específicas para cada situação, sendo as principais: aquisição de mudas sadias produzidas em viveiros devidamente cadastrados na CDA, inspeções visuais constantes por pessoal treinando, desinfestação de equipamentos e pessoas que tenham acesso aos pomares, controle do minador dos citros, plantio de quebra-ventos para evitar a ação do vento e da poeira sobre o pomar, uso de bactericida cúprico e indutores de resistência e, por fim, a escolha da variedade comercial de citros, uma vez que existem diferenças significativas de susceptibilidade entre os genótipos.(1,20,103)

Figura A. 3 - Sintomas característicos do cancro cítrico em (a) folhas, (b) frutos e (c) ramos. Sintomas visuais de cancro cítrico e verrugose (d) em folhas.
Fonte: CANCRO...(1)

O diagnóstico baseado apenas nos sintomas atende a maioria das situações associadas ao manejo do cancro cítrico, no entanto, determinados casos como a liberação do comércio de frutos para áreas livres ou sob erradicação, exigem que a confirmação da doença esteja associada a técnicas específicas para a detecção do agente causal. Dentre essas técnicas, destacam-se a corrida do fluxo bacteriano, isolamento da bactéria em meio de cultura apropriado, patogenicidade, PCR e sorologia.(20) Essas técnicas são as mais comumente empregadas no diagnóstico da doença. Maiores detalhes sobre o cancro cítrico, as formas de

manejo e controle, assim como das técnicas bioquímicas, podem ser encontrados na seguinte referência.(20)

Apesar das técnicas bioquímicas apresentarem uma alta eficácia, o processo como um todo não é tão eficiente, já que depende de diversos parâmetros, tais como a capacidade humana em identificar corretamente as plantas contaminadas, o grau de severidade da doença, condições de campo, etc.(33,34) A incidência de cancro cítrico na região citrícola do país saltou de 1,39% em 2012 para cerca de 5,0% em 2015, um aumento de 260% em três anos, segundo o levantamento do Fundecitrus.(19) O Fundecitrus em colaboração com outras instituições têm focado no desenvolvimento de projetos de pesquisa com cancro cítrico em áreas endêmicas, para que no futuro os citricultores possam contar com ferramentas de controle e manejo mais eficazes.(20)

APÊNDICE B – Exemplo da obtenção da matriz de co-ocorrência GLCM

A extração de texturas em imagens é essencial na área de reconhecimento de padrões cuja ideia básica consiste em extrair da imagem um conjunto de características que permitam realizar posteriormente sua classificação. Para a descrição de texturas em imagens, geralmente são utilizadas as seguintes abordagens: estatística, estrutural ou espectral.(106) Neste trabalho, utilizou-se somente a abordagem estatística o qual propicia a descrição das texturas das imagens através de regras estatísticas que governam a distribuição e a relação entre os diferentes níveis de cinza.(106)

Em 1973, Haralick e colaboradores (69) propuseram uma metodologia para a descrição de texturas onde são definidas 14 características provenientes do cálculo das matrizes de co-ocorrência (GLCM). Trata-se de uma matriz bidimensional que consiste na contagem de quantas combinações diferentes de valores de intensidade dos pixels (níveis de cinza) ocorrem em uma imagem para uma determinada direção. Tal matriz considera a relação entre dois pixels por vez, um denominado de “pixel de referência” e o outro de “pixel vizinho”.(72) O pixel vizinho escolhido pode ser vizinho em qualquer direção e também a vizinhança não precisa estar exatamente à distância de 1 pixel em relação ao pixel de referência. Desta forma, a matriz de co-ocorrência é sensível à rotação. Cada pixel dentro da imagem torna-se o pixel de referência, iniciando do canto superior esquerdo e procedendo até o canto inferior direito. A Figura B.1 ilustra os possíveis ângulos que podem ser utilizados para o cálculo da matriz de co-ocorrência.(107)

Figura B. 1 - Ângulos que podem ser utilizados para o cálculo da matriz de co-ocorrência.
Fonte: CLASSIFICAÇÃO...(107)

A partir da definição de matriz de co-ocorrência e dos conceitos de pixels vizinhos e orientações, a construção desta matriz torna-se simples. A fim de ilustrar o processo de construção da matriz GLCM, a Figura B.2 apresenta uma imagem com 8 níveis de cinza e a respectiva matriz GLCM extraída considerando a distância entre os pixels sendo igual a 1 na direção 0° .(108)

Figura B. 2 - Processo de extração da matriz de co-ocorrência em níveis de cinza a partir de uma imagem considerando $d=1$ e $\theta = 0^\circ$.
Fonte: GRAYCOMATRIX...(108)

A matriz GLCM extraída é uma matriz quadrada de tamanho $N \times N$, onde N é número de níveis de cinza presentes na imagem original. No caso do exemplo da Figura B.2, a matriz GLCM possui dimensão 8×8 uma vez que a imagem original possui 8 níveis de cinza. Uma vez definidos os parâmetros distância e ângulo que passa pelos pixels vizinhos em relação à horizontal, cada posição da matriz conterá a variação de níveis de cinza segundo estes parâmetros considerando os índices da matriz. Sendo assim, por exemplo, a posição (1,1) da matriz GLCM irá conter a quantidade de vezes que houve a ocorrência do nível de cinza 1 com o vizinho no nível de cinza 1 na horizontal à direita. Neste exemplo, a combinação [1 1] ocorre apenas uma única vez em toda a imagem, ficando armazenado no elemento (1,1) da matriz GLCM o valor 1. Já para o elemento (1,2), o valor 2 simboliza a existência de duas ocorrências da combinação [1 2] na imagem original, como se pode observar na Figura B.2. Este processo se repete para toda a matriz da imagem original obtendo-se assim a matriz GLCM para uma distância e ângulo específico. A partir da matriz obtida, diversas características de textura pode ser extraídas como foi proposto por Haralick em 1973.(69)

Maiores detalhes a respeito dos procedimentos envolvidos na obtenção da matriz GLCM e da extração dos descritores de textura podem ser encontrados em (69,71,106).