

Mineração de dados de LIFS para classificação de doença em citros

Stéfano de Ávila Souza Spindola¹
Thiago Massaiti Kuboyama Kubota²
Paulino Ribeiro Villas Boas³
Débora Marcondes B. P. Milori⁴

¹Aluno de Engenharia Física, Departamento de Física da Universidade Federal de São Carlos, São Carlos - SP. Estagiário-Embrapa, Embrapa Instrumentação, São Carlos, SP; stefano.engphysics@gmail.com;

²Aluno de mestrado em Física Aplicada, Instituto de Física de São Carlos, Universidade São Paulo, São Carlos, SP;

³Pesquisador da Embrapa Instrumentação, São Carlos, SP;

⁴Pesquisadora da Embrapa Instrumentação, São Carlos, SP.

A doença Huanglongbing (HLB), popularmente conhecida como “Greening”, tem causado grandes prejuízos à produção de citros mundialmente, com taxas crescentes de propagação. A principal causa do avanço da doença é a falta de método preciso para diagnóstico da doença, em especial em estágio precoce, antes mesmo da planta demonstrar sintomas. O Grupo de Ótica da Embrapa Instrumentação tem trabalhado ativamente nos últimos anos na busca de técnicas fotônicas que permitam o diagnóstico da doença ainda em fase assintomática. Em particular, a espectroscopia de fluorescência induzida por laser (LIFS) tem gerado índices de classificação superiores à 80% e tem grande potencial de ser levada a campo. Entretanto, os classificadores desenvolvidos com dados de um determinado mês não classificam adequadamente dados de outros meses. O objetivo deste trabalho foi comparar diferentes modelos de classificação com os dados de LIFS de 9 meses de coleta visando obter um classificador mais preciso e menos dependente do conjunto de calibração. Os modelos testados foram classificadores induzido via regressão por mínimos quadrados (PLSR), primeiro classificando as folhas entre sadia, assintomática ou sintomática, segundo classificando as folhas entre as estações de chuva ou seca, terceiro classificando as folhas utilizando níveis hierárquicos. Foram testados dois tipos de classificadores hierárquicos, o primeiro que classifica as folhas no 1º nível entre sintomática ou não sintomática e no 2º nível classifica as folhas não sintomática entre sadia ou assintomática e o segundo classificador hierárquico que classifica as folhas no 1º nível entre sintomática ou não sintomática, no 2º nível classifica as folhas não sintomática entre chuva ou seca, no 3º nível classificado as folhas de chuva entre sadia ou assintomática e no 4º nível é classificado as folhas de seca entre sadia ou assintomática, resultando em acertos para o primeiro classificador de $(68,4 \pm 0,3)\%$, para o segundo classificador $(79,3 \pm 0,3)\%$, para terceiro classificador $(67,3 \pm 0,3)\%$ e para o quarto classificador acertos de $(64,0 \pm 0,4)\%$, respectivamente. Também foram avaliadas diferentes regiões do espectro para normalizar os dados espectrais, e nenhuma melhora foi observada. Também foram avaliadas as regiões do espectro de fluorescência das folhas que mais contribuíram para a diferenciação da doença e da estação com a ferramenta *Moving Window* PLS. Com as regiões selecionadas, foram criados índices que foram usados como dados de entrada para o classificador PLSR, resultando em acertos de $(63,76 \pm 0,3)\%$ na classificação das folhas entre sadia, assintomática ou sintomática e $(69,30 \pm 0,4)\%$ na classificação das folhas entre chuva ou seca. Comparando os diversos classificadores propostos neste trabalho, os acertos ficaram entre 60% e 70% salvo pela classificação entre chuva e seca, acerto próxima a 80%. Os resultados deste trabalho motivam a busca de novos testes e outros modelos de classificação a fim de elevar o nível de acerto e de robustez para diagnóstico da doença HLB em citros.

Apoio financeiro: Estágio Embrapa

Área: Automação e Instrumentação agropecuária

Palavras chave: Classificação, mineração de dados, LIFS, diagnóstico de HLB.