



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



JOÃO RICARDO FAVAN

**DESENVOLVIMENTO DE SISTEMA INTELIGENTE PARA CLASSIFICAÇÃO DA
SEVERIDADE DA MANCHA FOLIAR BACTERIANA DO EUCALIPTO**

Botucatu

2019

JOÃO RICARDO FAVAN

**DESENVOLVIMENTO DE SISTEMA INTELIGENTE PARA CLASSIFICAÇÃO DA
SEVERIDADE DA MANCHA FOLIAR BACTERIANA DO EUCALIPTO**

Tese apresentada à Faculdade de Ciências Agronômicas da Unesp, Câmpus de Botucatu, para obtenção do título de Doutor em Ciência Florestal.

Orientador: Prof. Dr. José Raimundo de Souza Passos

Botucatu

2019

F272d Favon, João Ricardo
Desenvolvimento de sistema inteligente para classificação da
severidade da mancha foliar bacteriana do eucalipto / João
Ricardo Favon. -- Botucatu, 2019
76 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Agrônômicas, Botucatu
Orientador: José Raimundo de Souza Passos

1. Fitopatologia. 2. Inteligência artificial. 3. Reflectância. 4.
Eucalipto. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências Agrônômicas, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DESENVOLVIMENTO DE SISTEMA INTELIGENTE PARA CLASSIFICAÇÃO DA SEVERIDADE DA MANCHA FOLIAR BACTERIANA DO EUCALIPTO

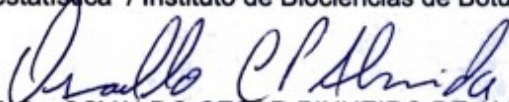
AUTOR: JOÃO RICARDO FAVAN

ORIENTADOR: JOSÉ RAIMUNDO DE SOUZA PASSOS

Aprovado como parte das exigências para obtenção do Título de Doutor em CIÊNCIA FLORESTAL, pela Comissão Examinadora:



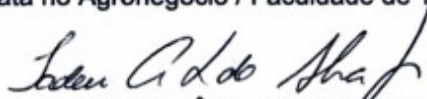
Prof. Dr. JOSÉ RAIMUNDO DE SOUZA PASSOS
Bioestatística / Instituto de Biociências de Botucatu - UNESP



Prof. Dr. OSVALDO CESAR PINHEIRO DE ALMEIDA
Análise e Desenvolvimento de Sistemas / Faculdade de Tecnologia de Botucatu



Prof. Dr. LUÍS HILARIO TOBLER GARCIA
Big Data no Agronegócio / Faculdade de Tecnologia Shunji Nishimura



Prof. Dr. TADEU ANTÔNIO FERNANDES DA SILVA JÚNIOR
Ciência e Tecnologia Ambiental / Universidade do Sagrado Coração



Prof. Dr. ROGERIO ANTÔNIO DE OLIVEIRA
Bioestatística / Instituto de Biociências de Botucatu - UNESP

Botucatu, 24 de maio de 2019

AGRADECIMENTOS

Agradeço primeiramente a Deus, fonte e princípio de tudo, por me conceder saúde, inteligência, foco, força, fé, e todos os dons necessários para chegar até aqui. Não porque mereço, mas por sua graça e bondade.

Agradeço a minha família, nas pessoas de Benedito Favan e Terezinha Favan, meus pais, por tudo que me ensinaram, educação, comprometimento, garra, vontade, respeito, valores, e todo o alicerce que me fez ser quem eu sou hoje. Obrigado por me dar tudo aquilo que nenhuma fortuna poderia pagar. Obrigado por todo o apoio desde o começo dessa jornada. Com certeza, se não por vocês, não chegaria até aqui.

Estendo o agradecimento às minhas irmãs, Fernanda e Paula, e aos meus cunhados Marcos e Fernando, por estarem presentes em minha vida, tornando a caminhada mais fácil e divertida.

Agradeço também a uma pessoa tão especial, com quem decidi compartilhar minha vida, Renata Coscolin (PhD), companheira, amiga, esposa. Obrigado por toda ajuda e apoio, por me animar nos momentos de tristeza, por me levantar nos momentos difíceis, por me fazer sorrir a cada dia, sem você ao meu lado, com certeza, a conclusão dessa etapa ainda seria um sonho. Obrigado pela paciência, amor, cumplicidade e pelas consultorias agrônômicas (rs). Te amo sempre e muito.

Ao meu orientador, Prof. Dr. José Raimundo de Souza Passos, profissional de excelência e pessoa de valor inestimável. Obrigado por acreditar no projeto, pelos conselhos durante o processo, pelo apoio a todo momento, pelas horas de conversas via Skype, levarei comigo mais que suas palavras, mas sim, seus exemplos. Será sempre uma referência de professor, pesquisador e pessoa para mim.

Estendo também os agradecimentos aos colegas de grupo de pesquisa, André Jin, Lara, Rafaela, por toda ajuda na condução do experimento e pelas conversas nos “meios tempos”. Ao Tadeu Fernandes, esta pesquisa não seria realizada sem a sua ajuda.

Agradeço aos novos colegas de trabalho, que fazem o nosso dia a dia mais leve. Obrigado a todos os professores e funcionários da Faculdade de Tecnologia “Shunji Nishimura” de Pompéia.

Ao M.E.C., pois o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -Brasil (CAPES) - Código de Financiamento 001.

Enfim, agradeço a todos que de forma direta ou indireta contribuíram para a conclusão desse projeto.

“Não há nada como um sonho para criar o futuro. O amanhã começa agora!”

WALKER, J. **A História de Johnnie Walker | Johnnie Walker.**
Disponível em: <<https://www.johnniewalker.com/pt-br/o-mundo-de-johnnie-walker/a-história-de-johnnie-walker/>>. Acesso em: 28 fev. 2019.

RESUMO

O setor florestal brasileiro e, principalmente, o setor de árvores plantadas tem suma importância para a economia do Brasil. Nessa cadeia produtiva, os viveiros florestais são os responsáveis por fornecer as mudas utilizadas em reflorestamentos, sendo que esses fornecem matéria prima para as indústrias de energia e papel e celulose. Dessa forma, o adequado manejo e a prevenção de doenças nas plantas comercializadas ocupam um lugar de destaque na produção dos viveiros. A Mancha Foliar Bacteriana (MFB) do eucalipto é uma doença preocupante para a produção de mudas de eucalipto, visto que sua incidência, sem o devido tratamento, pode levar a grandes perdas. A reflectância foliar de plantas pode ser um indicador para as respostas a diversos fenômenos biofísicos e bioquímicos em plantas. Este trabalho desenvolveu classificadores baseados em Inteligência Artificial para discriminar a ocorrência da mancha foliar bacteriana, assim como determinar sua severidade e seu período de latência, utilizando imagens digitais e assinaturas espectrais das folhas de eucalipto. Mudas de eucalipto foram inoculadas com uma suspensão de bactérias *Xanthomonas spp.* e suas assinaturas espectrais e imagens digitais foram coletadas durante oito dias consecutivos. Mudas de eucalipto não inoculados foram utilizados com controle negativo. Os dados coletados foram analisados utilizando técnicas estatísticas e de inteligência artificial, a fim de se obter a severidade da doença, seu período de latência, distinção entre mudas saudáveis e infectadas. Os modelos computacionais testados apresentaram bons resultados para a discriminação entre plantas doentes e saudáveis, permitindo a detecção pré sintomática da doença. A severidade da doença mensurada a partir da reflectância foliar espectral foi melhor ajustada pelos modelos estatísticos de regressão linear múltipla. Dessa forma, ferramentas computacionais e estatísticas devem ser usadas conjuntamente para obter os melhores resultados na classificação das doenças e na previsão da severidade da doença.

Palavras-chave: Aprendizado de máquinas. Regressão linear múltipla. Viveiros florestais. Mancha foliar.

ABSTRACT

Brazilian forestry sector has great importance to the Brazilian economy. Forest nurseries are responsible for supplying the seedlings used in reforestation, which provide raw material for the energy, pulp and paper industries. Adequate management and prevention of diseases in plants is important for nurseries. Bacterial leaf spot of eucalyptus is a worrying disease for the production of seedlings, since its incidence, without treatment, causes losses. Leaf reflectance can be an indicator for the responses to various biophysical and biochemical phenomena in plants. The objective of this work was to develop classifiers based on artificial intelligence to discriminate the existence of the bacterial disease of the leaf spot, as well as to determine its severity and period of latency using digital images and spectral signatures of eucalyptus leaves. *Eucalyptus* seedlings were infected with *Xanthomonas spp.* and their spectral signatures and digital images were collected during eight consecutive days. Collected data were analyzed using statistical and artificial intelligence techniques, obtain the severity of the infection, latency period distinction between healthy and infected seedlings. computational models showed good results for the discrimination between diseased and healthy plants, allowing the presymptomatic detection of the disease. Severity of the disease was better adjusted by the statistical models of multiple linear regression. Computational and statistical tools should be used together to obtain best results in the classification of diseases and the prediction of their severity.

Keywords: Machine learning. Multiple linear regression. Nurseries. Leaf spot

LISTA DE ILUSTRAÇÕES

- Figura 1 - Folha de *Eucalyptus* com sintomas de Mancha Foliar Bacteriana causada por *Xanthomonas spp.*.....21
- Figura 2 - Boxplot das porcentagens de acertos para as três melhores técnicas de seleção, utilizando algoritmo de Perceptron de multicamadas, para as mil rodadas de testes executadas.....39
- Figura 3 - Curva média da reflectância foliar espectral de mudas de *Eucalyptus grandis* x *E. urophylla* submetidos (B) e não submetidos (C) a infecção com *Xanthomonas spp.* destacado os comprimentos de onda (nm) elegidos pelas técnicas de seleção de atributos.....40
- Figura 4 - Boxplot das porcentagens de acertos entre plantas não inoculadas e inoculadas com *Xanthomonas spp.* em folhas de clone híbrido de *Eucalyptus grandis* x *E. urophylla* para cada dia após a inoculação utilizando algoritmo de Perceptron de multicamadas.....45
- Figura 5 - Boxplot dos coeficientes de correlação de Pearson entre a severidade real e a severidade predita pelo algoritmo *Support Vector Machine* para folhas de clone híbrido de *Eucalyptus grandis* x *E. urophylla*, inoculadas com *Xanthomonas spp.* para cada dia do experimento.....63

LISTA DE TABELAS

- Tabela 1 - Quantidade de atributos e comprimentos de onda (nm) selecionados para cada técnica de seleção de atributos testadas com dados da reflectância foliar espectral (%) de eucalipto submetidos a infecção com *Xanthomonas spp.*.....37
- Tabela 2 - Média (desvio padrão entre parênteses) do percentual de classificações corretas para curvas de reflectância espectral de mudas de eucaliptos, submetidas e não submetidas à inoculação de *Xanthomonas spp.*, para cada algoritmo de classificação (coluna) utilizando cada uma das técnicas de seleção de atributos (linha).....38
- Tabela 3 - Quantidade de atributos e comprimentos de onda (nm) selecionados para cada técnica de seleção de atributos testadas com dados da reflectância foliar espectral de folhas de eucaliptos submetidos a infecção com *Xanthomonas spp.* para classificação em dias após inoculação.....43
- Tabela 4 - Média (desvio padrão entre parênteses) do percentual de classificações corretas para curvas de reflectância espectral de mudas de eucaliptos, submetidas à inoculação de *Xanthomonas spp.*, para cada algoritmo de classificação (coluna) utilizando cada uma das técnicas de seleção de atributos (linha) para a classificação em dias após a inoculação.....44
- Tabela 5 - Quantidade de atributos e comprimentos de onda (nm) selecionados para cada técnica de seleção de atributos testadas com dados da reflectância foliar espectral de folhas de eucaliptos submetidos a infecção com *Xanthomonas spp.* para predição da severidade da doença.....60
- Tabela 6 - Média (desvio padrão entre parênteses) do coeficiente de correlação de Pearson para curvas de reflectância espectral de folhas de clone híbrido de *Eucalyptus grandis* x *E. urophylla*, submetidas à inoculação de *Xanthomonas spp.*, para cada algoritmo de predição (coluna) utilizando cada uma das técnicas de seleção de atributos (linha) para a predição da severidade da doença.....61
- Tabela 7 - Diagnósticos de ajustes dos modelos de regressão linear múltipla para a severidade (no. pústulas/cm²) em função dos comprimentos de ondas (nm).....64

SUMÁRIO

INTRODUÇÃO GERAL.....	19
CAPÍTULO 1: COMPARAÇÃO ENTRE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA DISCRIMINAÇÃO DE MUDAS DE EUCALIPTO INOCULADAS OU NÃO COM <i>XANTHOMONAS SPP.</i>.....	29
1.1 Introdução.....	31
1.1.1 Objetivo.....	33
1.2 Material e Métodos.....	34
1.3 Resultados e Discussão.....	36
1.3.1 Discriminação entre plantas sadias e plantas inoculadas com <i>Xanthomonas</i> spp.....	36
1.3.2 Classificação das mudas em dias após a inoculação.....	42
1.3.3 Detecção pré sintomática da doença.....	45
1.4 Conclusão.....	47
Referências.....	47
CAPÍTULO 2: MODELOS COMPUTACIONAIS E ESTATÍSTICOS PARA A PREDIÇÃO DA SEVERIDADE DA MANCHA FOLIAR CAUSADA POR <i>XANTHOMONAS SPP.</i> EM CLONE HÍBRIDO DE <i>EUCALYPTUS GRANDIS</i> X <i>EUCALYPTUS UROPHYLA</i>.....	52
2.1 Introdução.....	53
2.1.1 Objetivo.....	56
2.2 Material e métodos.....	56
2.3 Resultados e Discussão.....	60
2.3.1 Modelagem computacional com aprendizado de máquinas.....	60
2.3.2 Modelagem estatística com modelos de regressão linear múltipla.....	64
2.3.3 Considerações Adicionais.....	66
2.4 Conclusão.....	67
2.5 Referências.....	67
CONSIDERAÇÕES FINAIS.....	71
REFERÊNCIAS.....	73

INTRODUÇÃO GERAL

O *Eucalyptus* é um gênero de plantas da família das *Myrtaceae*, conhecido com o nome comum de eucalipto, composto majoritariamente por espécies arbóreas. O gênero corresponde a 730 espécies de plantas originárias da Austrália, Tanzânia, Nova Guiné, Indonésia e Filipinas. Suas principais características podem ser destacadas como seu rápido crescimento, capacidade de adaptação às diversas regiões ecológicas e pelo potencial econômico proporcionado pela utilização diversificada de sua madeira (PLANTSYSTEMATICS.ORG, 2019).

O gênero foi introduzido no Brasil por volta de 1855 no Jardim Botânico do Rio de Janeiro. No entanto, sua produção foi intensificada por Edmundo Navarro de Andrade, por volta de 1930, para suprir a necessidade de madeira para dormentes, moirões, batentes e outros produtos destinados a Companhia Paulista de Estradas de Ferro (FOELKEL, 2005).

Desde sua implantação até os dias de hoje a cultura do eucalipto tem sua importância no fornecimento de matéria-prima para outras diversas indústrias, dessa forma sua produção e comercialização se deu de maneira exponencial e tornou-se um marco na economia nacional.

O setor de florestas plantadas é muito importante para a economia brasileira, perfazendo uma receita bruta de 73,8 bilhões de Reais e participação no PIB de 1,1%. As florestas plantadas totalizam uma área de 7,84 milhões de hectares e seus produtos são destinados a diversos segmentos industriais como papel e celulose, siderurgia, carvão vegetal, painéis de madeira, pisos laminados, móveis, etc (IBA, 2018). As florestas plantadas de *Eucalyptus spp.* ocupam uma área de 5,67 milhões de hectares, sendo esta a principal cultura florestal no Brasil (IBA, 2017).

O manejo apropriado, a utilização de mudas que atenda a padrões de qualidade estabelecidos, livres de doenças, cultivadas em viveiros com manejo fitossanitário adequado são pontos fundamentais para o sucesso na condução de uma floresta plantada (AUER; SANTOS; NETO, 2011; KRATZ; WENDLING, 2013).

O tecido foliar é o principal responsável pelo processo fotossintético das plantas, o qual permite o crescimento vegetativo, no entanto, as manchas foliares causam necrose neste tecido, diminuindo a área fotossinteticamente ativa e podendo causar a morte da planta (BEDENDO, 2011).

As manchas foliares são causadas principalmente por fungos e bactérias, ocorrendo majoritariamente em regiões de clima quente e úmido, resultando na diminuição do desenvolvimento vegetativo da planta, seu rendimento no viveiro e na qualidade do produto final no campo (AUER; SANTOS; NETO, 2011). As manchas têm forma, coloração e propagação diferentes, sendo geralmente, empregadas como forma de controle, a utilização de variedades de plantas resistentes as doenças, aplicação de fungicidas, ou ainda, a erradicação das plantas hospedeiras (BEDENDO, 2011).

A Mancha Foliar Bacteriana (MFB) pode ser considerada como uma doença recente, se comparada com outras doenças do eucalipto, pois seu primeiro registro é datado de 1995 (AUER; SANTOS; NETO, 2011), no entanto, é uma das principais doenças da Eucaliptocultura no Brasil, principalmente quando se tratando de produção de mudas em viveiros. Estima-se que esta doença tenha causado perdas na ordem de 7,5 milhões de dólares entre os anos de 2003 e 2008 (ALFENAS *et al.*, 2009; GONÇALVES *et al.*, 2008).

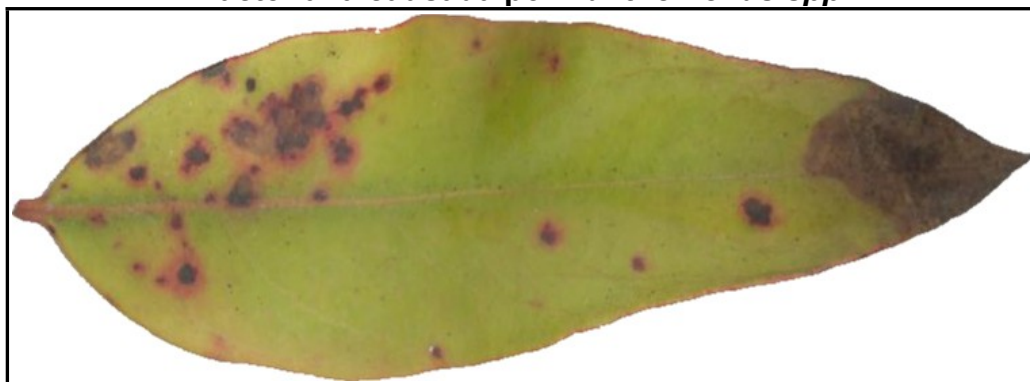
A MFB tem sua ocorrência registrada nos principais estados brasileiros produtores de eucalipto, sendo eles, Amapá, Bahia, Minas Gerais, São Paulo, Pará, Mato Grosso do Sul e Rio Grande do Sul. Há ocorrência da doença em países da América Latina como a Argentina, Paraguai e Uruguai (ALFENAS *et al.*, 2009; AUER; SANTOS; NETO, 2011; GONÇALVES *et al.*, 2008).

A MFB tem como principais agentes causadores as bactérias dos gêneros *Xanthomonas axonopodis*, *X. campestris*, *Pseudomonas syringae*, *P. cichorri* e *P. putida*. Por outro lado são consideradas as variedades de eucalipto mais suscetíveis a doença a *Eucalyptus cloeziana*, *E. grandis*, *E. globulus*, *E. maidenii*, *E. pellita*, *E. regnans*, *E. robusta*, *E. saligna*, *E. urophylla* e *E. viminalis* e *E. urograndis* (ALFENAS *et al.*, 2009; FERRAZ *et al.*, 2018; GONÇALVES *et al.*, 2008).

Os principais sintomas da MFB do eucalipto podem variar devido à idade da planta, do estado de desenvolvimento da lesão e da espécie do eucalipto. De forma mais generalizada, são caracterizados com pontuações nas folhas mais jovens e nos ponteiros, que evoluem na forma de manchas úmidas, do tipo anasarcas, com ocorrência angulares, internervurais, translúcidas, ocorrendo de ambos os lados da folha. As manchas podem estar concentradas ao longo da nervura principal, nas

margens das folhas ou distribuídas aleatoriamente sobre o limbo. Posteriormente, as manchas se tornam necróticas e deformam o limbo foliar (ALFENAS *et al.*, 2009; AUER; SANTOS; NETO, 2011; MAFIA; TEIXEIRA; FERREIRA, 2017). Como exemplificado pela Figura 1.

Figura 1 - Folha de *Eucalyptus* com sintomas de Mancha Foliar Bacteriana causada por *Xanthomonas spp.*



Fonte: Damasceno *et al.*(2014)

A evolução da doença é caracterizada pelo aumento do número de lesões e elas adquirem um aspecto ressecado, com coloração amarronzada, podendo conter orifícios no centro da lesão ou áreas recortadas do limbo nas folhas mais jovens, pode, também, haver necrose em pecíolo e ramos. O ápice da doença é caracterizado por uma intensa desfolha devido à senescência precoce das folhas infectadas.

O processo de diagnose da doença deve ser feito por meio da avaliação do material utilizando um microscópio estereoscópio, que não deve apresentar estruturas fúngicas, como hifas e esporos, que geralmente, são encontradas em folhas com lesões ocasionadas por fungos. Para confirmação do diagnóstico deve-se utilizar a metodologia de isolamento de fitobactérias com material doente em meio nutriente ágar, se realmente se tratar da Mancha Foliar Bacteriana do eucalipto, deve ser percebido o surgimento de colônias nesse meio (ALFENAS *et al.*, 2009; AUER; SANTOS; NETO, 2011; MAFIA; TEIXEIRA; FERREIRA, 2017).

O controle da MFB do eucalipto em viveiros deve ser feito por meio da multiplicação dos clones suscetíveis somente em épocas do ano que forem desfavoráveis à doença, no outono e inverno, onde são caracterizadas as baixas temperaturas e umidade relativa do ar, assim como menores precipitações

pluviométricas quando comparados com a primavera e o verão. Deve ser feita também a remoção das folhas ou das plantas doentes e evitar que haja o molhamento da parte aérea das mudas, no entanto, esta última medida se torna dificultada devido a utilização de irrigação por microaspersão em todo o processo produtivo dos viveiros comerciais (ALFENAS *et al.*, 2009; AUER; SANTOS; NETO, 2011; MAFIA; TEIXEIRA; FERREIRA, 2017).

A doença pode ser levada a campo por meio de mudas infectadas, nesse sentido, a recomendação é que os plantios sejam feitos com mudas saudáveis e certificadas, obtidas de viveiros idôneos. O plantio de material genético resistente pode ser considerado com alternativa viável dependendo do caso analisado (AUER; SANTOS; NETO, 2011).

A radiação solar que chega a superfície terrestre, ao atingir a planta, interage com a mesma em três frações, a primeira é absorvida pela folha e utilizado nos processos fotossintéticos, a segunda fração é refletida pelas folhas em um fenômeno chamado reflexão, a terceira fração sofre o processo de transmissão, adentrando as camadas da folha e suas estruturas (CUNHA, 2004; MOREIRA, 2011).

O comportamento espectral da vegetação se manifesta de forma distinta em três regiões desse espectro, sendo elas a região do visível, região do infravermelho próximo e região do infravermelho médio. A região do visível é situada na faixa espectral entre 400nm e 700nm, considerada a faixa de radiação fotossinteticamente ativa; A região denominada infravermelho próximo, compreende a faixa do espectro eletromagnético entre 700 m e 1300nm, onde a radiação incidente interage com a estrutura interna da folha; A região chamada de infravermelho médio é situada na faixa espectral entre 1300 nm e 3000 nm, onde é observado o conteúdo de água na folha (CUNHA, 2004; LAW; WARING, 1994; MOREIRA, 2011; ODUM, 1983).

A reflectância espectral foliar pode ser considerado como uma resposta da planta a diversos fenômenos biofísicos e bioquímicos que ocorrem em sua estrutura, dessa forma, pode ser associada a um padrão de resposta, denominado assinatura espectral, que pode variar para um dado binômio patógeno-hospedeiro, demonstrando que esta planta está saudável ou não, de acordo com suas respostas espectrais em determinadas faixas do espectro eletromagnético. Técnicas similares

veem sendo usadas em trabalho de sensoriamento remoto para monitorar plantações infectadas (COLWELL, 1974 ; FRANKE; MENZ, 2007; MAHLEIN *et al.*, 2012).

A utilização de reflectância espectral, e por conseguinte, a assinatura espectral das plantas, para detecção de doenças já foi utilizado com sucesso em diversas culturas para as mais diversas doenças, como por exemplo a ferrugem em trigo (HUANG *et al.*, 2007), o amarelamento em videiras (AL-SADDIK *et al.*, 2018), mancha foliar de *Cercospora* em beterraba (MAHLEIN *et al.*, 2012a) e a detecção de árvores danificadas em florestas (VALENZUELA; LÓPEZ-GARCÍA, 2017).

O aprendizado de máquina (AM) pode ser definido como o estudo e desenvolvimento de técnicas e algoritmos capazes de reconhecer padrões e “aprender” com base em seus erros e assim fazer previsões sobre os dados. Esses algoritmos constroem modelos a partir das amostras direcionadas como entradas a fim de fazer previsões ou decisões guiadas por esses dados, em vez de seguir instruções previamente programadas (SIMON, 2013).

O AM, diferentemente dos algoritmos tradicionais, não acontece por meio de uma sequência de comandos, mas sim, com o processamento de uma grande quantidade de dados que são fornecidos, deixando que a máquina identifique os resultados positivos e então selecione o melhor resultado entre os obtidos (ALPAYDIN, 2016).

O AM concentra-se especificamente em quatro tipos de problemas, a classificação, onde a resposta aprendida deve atribuir a uma amostra como pertencente a uma classe pré-definida; a regressão, onde a resposta aprendida deve ser um valor contínuo para um determinado problema; o agrupamento, onde a máquina deve encontrar semelhanças em diversas amostras e a associação, que busca associar determinado fenômeno com outros eventos (ALPAYDIN, 2016).

O processo de criação de modelos de aprendizado de máquinas consiste em duas fases bem definidas, a seleção de atributos e o processo de aprendizado. A primeira fase tem como objetivo selecionar dentre os atributos disponíveis aqueles que são mais relevantes para o problema proposto, e, a segunda fase utiliza-se dos atributos selecionados para realizar o aprendizado e resolver o problema adequadamente (MITCHELL, 1997; REZENDE; MARCACINI; MOURA, 2011).

Há, na literatura, diversas técnicas de seleção de atributos, sendo cada uma focada em uma estratégia diferente para a seleção dos atributos mais relevantes para um determinado problema, nesse trabalho, são trazidas algumas das técnicas mais utilizadas.

Avaliação de Atributo Classificador – *Classifier Attribute Evaluation* – avalia a importância de cada atributo utilizando um classificador específico, resultado em um ranqueamento dos atributos analisados (WITTEN; FRANK, 2005).

Avaliação de subconjunto por classificador – *Classifier Subset Evaluation* - Avalia subconjuntos de atributos nos dados de treinamento usando um classificador para estimar o mérito de cada conjunto de atributos, ao final, é selecionado o subconjunto com maior mérito de classificação (WITTEN; FRANK, 2005).

Avaliação de Atributos por ganho de informação – *Information Gain Attribute Evaluation* (InfoGainAttributeEval) - avalia a importância de um atributo medindo o ganho de informação que este obtém em relação ao atributo classe, resultando em um ranqueamento dos atributos com maior importância (KAREGOWDA; MANJUNATH; JAYARAM, 2010).

Avaliação de atributos por taxa de ganho - *Gain Ratio Attribute Evaluation* (GainRatioAttributeEval) - avalia a importância de um atributo medindo a taxa de ganho (Gain Ratio) em relação à classe, resultando em um ranqueamento dos atributos de acordo com sua importância. A taxa de ganho de um atributo é mensurado pela razão entre valor de ganho de informação desse atributo pelo maior ganho proporcionado no conjunto de dados (KAREGOWDA; MANJUNATH; JAYARAM, 2010).

A seleção de atributos com a técnica de Análise de componentes principais (PCA) foi feita executando a técnica de estatística multivariada conhecida pelo mesmo nome, conjuntamente com sua transformação dos dados pelos seus autovalores. A redução de dimensionalidade foi feita por meio da escolha dos autovetores suficientes para conjugar maior porcentagem da variação dos dados originais (WOLD; ESBENSEN; GELADI, 1987).

A seleção de atributos por meio da avaliação de Relief – *Relief Attribute Evaluation* (ReliefAttributeEval) – foi feita através de amostragens repetidas de uma instância (registro em um conjunto de dados) e considerando o valor de um

determinado atributo para a instância mais próxima da mesma e de outra classe, sendo a diferença dos valores do atributo para essas instâncias, consideradas no cálculo da importância para cada atributo, resultando em um ranqueamento dos atributos utilizados (KIRA; RENDELL, 1992; WITTEN; FRANK, 2005).

A técnica de seleção de atributos *Symmetrical Uncert* (SymmetricalUncertAttributeEval) avalia a importância de um atributo medindo sua incerteza em relação ao atributo classe (APPAVU *et al.*, 2011).

A técnica de seleção de atributos *Wrapper* avalia subconjuntos de atributos utilizando um algoritmo de aprendizagem determinado, sendo o procedimento de avaliação cruzada empregado para estimar a precisão desse algoritmo no subconjunto submetido a avaliação (KOHAVI; JOHN, 1997).

Com os atributos mais relevantes selecionados, a segunda fase se caracteriza pela escolha do algoritmo de classificação, ou de predição, a ser utilizado. Nesse sentido, há uma rica literatura sobre esses algoritmos, limitando este trabalho a apresentar uma breve descrição dos algoritmos mais utilizados.

Uma rede bayesiana é um modelo gráfico baseado em probabilidade. Uma rede bayesiana consiste em nós estocásticos, que podem ser variáveis observadas ou não observadas, caso primeiro, a distribuição do nó é dada pela distribuição do erro de observação ou distribuição de dados, nos demais casos, distribuições prévias são especificadas para os nós. Cada nó é conectado com um subconjunto dos outros nós, que influenciam a distribuição correspondente. Um nó depende apenas de seus nós “pais” e dado seus pais, cada nó é condicionalmente independente de todos os outros não-descendentes (MEYER-BAESE; SCHMID, 2014).

O classificador J48 é uma árvore de decisão C4.5 binária para problemas de classificação. Com essa técnica, uma árvore de decisão é construída para modelar o processo de classificação. Após a construção dessa árvore de decisão, as amostras são submetidas à classificação e seus resultados avaliados (QUINLAN, 1994).

O perceptron multicamadas consiste de um sistema de neurônios interconectados (nós), o que representa um mapeamento não-linear entre um vetor de entrada e um vetor de saída. Os nós são conectados por pesos e sinais de saída, que são uma função da soma das entradas para o nó, modificado por uma função de

ativação. A saída de um nó é dimensionada pelo peso de conexão e alimentada para ser uma entrada para os nós na próxima camada da rede, sendo assim, a arquitetura de um *perceptron* multicamadas consistirá em várias camadas de neurônios subsequente e interconectadas. A camada de entrada objetiva trazer o vetor de entrada para a rede. Um *perceptron* multicamadas pode ter uma ou mais camadas ocultas e uma camada de saída. Os *perceptrons* multicamadas são descritos como totalmente conectados, com cada nó conectado a cada nó na camada seguinte e anterior (GARDNER; DORLING, 1998). O Processo de aprendizagem por *backpropagation*, utilizada pelo *perceptron* multicamadas, consiste em ajustar os pesos das conexões entre os nós de forma a diminuir o erro na classificação da rede (RUMELHART, 1986).

O classificador OneR cria uma árvore de decisão de um nível para cada atributo, adicionando um ramo dessa árvore para valor desse atributo, para cada ramo, é atribuído o valor da classe mais frequente. O classificador procura por atributos com as menores taxas de erros, assim como, em caso de atributos numéricos, estes são discretizados (HOLTE, 1993). Embora seja um classificador bastante simples, este é muito usado em análises exploratórias de dados (MUDA *et al.*, 2011).

O classificador Comitê Aleatório – *Random Committee* – consiste em um conjunto de classificadores binários, escolhidos aleatoriamente que procedem sua aprendizagem com o mesmo conjunto de dados e produzem um resultado de classificação. A predição final é uma média das predições geradas pelos classificadores individualmente (CHOU *et al.*, 2009).

O classificador Floresta Aleatórias – *Random Forest* – consiste na criação de diversos classificadores do tipo árvore de decisão (QUINLAN, 1994) e a combinação entre essas “árvores” com o objetivo de obter uma classificação mais acurada e com maior estabilidade (BREIMAN, 2001).

O classificador *Support Vector Machine* (SVM) é um algoritmo de aprendizado de máquina supervisionado cujo seu processo de classificação consiste em plotar, para cada amostra utilizada, um ponto no espaço “n-dimensional”, sendo n a quantidade de atributos analisados, e os valores desse atributos usados como

coordenadas. A aprendizagem objetiva encontrar um hiperplano que diferencie as amostras entre as classes utilizadas (HEARST *et al.*, 1998).

O classificador Zero-R consiste em classificar todas as amostras como pertencentes a classe com maior número de amostras (moda). Este classificador é usado como referência para avaliação de outros classificadores, sendo que a performance de classificação de outros algoritmos não podem ser inferior a aquelas obtidas pelo Zero-R (WITTEN; FRANK, 2005).

O algoritmo *Linear Regression* Utiliza a regressão linear como técnica para predição, utilizando o critério de Akaike para a seleção do modelo (AKAIKE, 1974; WITTEN; FRANK, 2005).

Diante do exposto, as técnicas de inteligência artificial vem se tornando uma grande aliada no ambiente agrícola, desde a detecção de doenças, sua severidade e principalmente, no auxílio a tomada de decisão dos profissionais envolvidos nesse meio (KAUNDAL; KAPOOR; RAGHAVA, 2006; MCQUEEN *et al.*, 1995; MOHANTY; HUGHES; SALATHÉ, 2016; SINGH *et al.*, 2016).

Dessa forma, diversos esforços veem sendo feitos para unir técnicas de espectrometria e inteligência artificial com a finalidade de prever ou classificar doenças em plantas e também seus estresses abióticos de forma a obter maior assertividade ou mesmo mais agilidade na detecção de possíveis patógenos (HUANG *et al.*, 2007; JONES; JONES; LEE, 2010; LIRA *et al.*, 2007).

A MFB do eucalipto é uma doença muito preocupante para os produtores de mudas florestais, com isso, todo o esforço para desenvolver técnicas que auxiliem e abreviem a detecção da doença ou mesmo lancem um alerta sobre a infecção deve ser considerado de grande valia, ademais, a utilização de técnicas avançadas para o cotidiano dos profissionais resultam em celeridade e assertividade nas decisões.

O objetivo geral do presente trabalho foi avaliar o potencial do uso de modelos computacionais de aprendizado de máquina para discriminação de plantas com sintomas da MFB do eucalipto causado por *Xanthomonas spp.*

Este objetivo geral pode ser subdividido em 4 objetivos específicos, sendo eles:

a) Discriminar entre plantas sintomáticas e plantas sadias utilizando algoritmos de aprendizado de máquinas e reflectância foliar espectral;

- b) Definir o período de latência (em dias) de uma planta doente;
- c) Realizar a detecção pré sintomática da MFB com algoritmos computacionais;
- d) *Predizer a severidade da MFB com base na reflectância foliar espectral.*

CAPÍTULO 1: COMPARAÇÃO ENTRE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA DISCRIMINAÇÃO DE MUDAS DE EUCALIPTO INOCULADAS OU NÃO COM *XANTHOMONAS SPP.*

João Ricardo Favan; José Raimundo de Souza Passos

RESUMO

A inteligência artificial vem ganhando cada vez mais espaço no meio agrícola e florestal, buscando solucionar problemas diversos com base em aprendizado a partir de dados anteriormente coletados. O setor florestal brasileiro tem uma importância significativa para a economia, fornecendo matéria-prima para diversas indústrias como papel e celulose, carvão, moveleiras, entre outras. Para estabelecer uma floresta com qualidade e eficiência é necessário obter mudas de boa qualidade com alto padrão fitossanitário. A mancha foliar bacteriana (MFB) do eucalipto causado por *Xanthomonas spp.* é uma das doenças mais importantes para essa cultura e sem o devido controle pode causar diversos prejuízos. Para a produção dessa assinatura a reflectância espectral foliar é a técnica para mensurar a energia refletida em uma folha, criando padrões de respostas que podem ser associados a diversos fenômenos que ocorrem na planta. A inteligência artificial é responsável pelo desenvolvimento de diversos modelos de aprendizagem para resolução de vários tipos de problemas. As técnicas de seleção de atributos elegem os atributos mais importantes para a modelagem do problema e os algoritmos de classificação são utilizados para ajustar o modelo de acordo com os atributos selecionados. Este trabalho testou combinações entre diversas técnicas de seleção de atributos e algoritmos de classificação para discriminação de mudas de eucalipto controle ou inoculadas. Mudas de *Eucalyptus grandis* x *Eucalyptus urophylla* foram inoculadas com *Xanthomonas spp.*, e a reflectância espectral foliar foram medidas durante oito dias consecutivos. Os padrões de reflectância foram analisados por 80 modelos de aprendizado de máquina, compostos pela combinação de dez técnicas de seleção de atributos e oito algoritmos de classificação. Os modelos foram submetidos a 1000 ensaios cada um sendo avaliados pelo percentual de classificações corretas. Os modelos testados obtiveram porcentagens de acertos em torno de 70% para classificação entre mudas sadias e mudas doentes. Para classificação em dias após a inoculação todos os modelos apresentaram médias de percentuais de

classificações corretas abaixo de 50%. Foi possível observar que os modelos testados permitem que seja feita uma detecção pré sintomática da doença. Os modelos testados são aptos para classificação entre plantas saudáveis e plantas doentes, assim como possibilitam uma detecção pré sintomática da doença, demonstrando interessantes formas de utilização como ferramenta de apoio ao especialista.

Palavras-chave: *Multilayer Perceptron*. Detecção pré sintomática de doença. Bacteriose. Eucalipto.

ABSTRACT

Brazilian forestry sector has a significant importance for economy, supplying raw material for several industries such as pulp and paper, coal, furniture and others. Establish a quality forest is necessary obtain seedlings of good quality with a high phytosanitary standard. Bacterial leaf spot of eucalyptus caused by *Xanthomonas spp.* is one of the most important diseases for this crop, and can cause many damages. Leaf spectral reflectance is technique to measure the energy reflected in leaf, creating patterns of responses that can be associated with several phenomena that occur in plant. Artificial intelligence has been gaining more space in agriculture and forestry, and several learning models have been developed to solve various types of problems. Attributes selection techniques choose the most important attributes for problem modeling and the classification algorithms are used to fit the model according to the selected attributes. Objective was to test combinations between different attributes selection techniques and classification algorithms for discrimination between control and inoculated eucalyptus seedlings. Hybrid clone seedlings of *Eucalyptus grandis* x *Eucalyptus urophylla* were inoculated with *Xanthomonas spp.* while another part of the seedlings were not submitted to inoculation. Spectral reflectance of these leaves were taken during eight consecutive days. The reflectance patterns were analyzed by 80 machine learning models, composed by the combination of ten attribute selection techniques and eight classification algorithms. Models were submitted to 1000 tests each being evaluated by the percentage of correct classifications. Tested models obtained correct percentages of around 70% for classification between healthy and inoculated seedlings. Classification of latency period, all models presented averages of correct classification percentages below 50%. It was possible to observe that the models

tested allow a presymptomatic detection of the disease. models tested are suitable for classification between healthy and diseased plants, as well as enable a presymptomatic detection of the disease, demonstrating interesting forms of use as a tool to support the specialist.

Keywords: *Multilayer Perceptron*. Presymptomatic detection of disease. Eucalyptus. Bacterial leaf Spot.

1.1 Introdução

O setor de florestas plantadas tem grande importância para a economia do Brasil. No ano de 2017 obteve uma receita bruta de 73,8 bilhões de Reais e 1,1% de participação no PIB nacional. A área de florestas plantadas totalizam 7,84 milhões de hectares e seus produtos são destinados principalmente para o segmento industrial de papel e celulose, siderurgia e carvão vegetal, painéis de madeira e pisos laminados, etc (IBA, 2018). Dentre toda a área de florestas plantadas, as florestas de *Eucalyptus spp.* se destacam ocupando 5,67 milhões de hectares, sendo esta a principal cultura florestal no Brasil (IBA, 2017).

Para o sucesso na condução de uma floresta plantada é importante ter um manejo apropriado, assim como, a utilização de mudas de boa qualidade e com alto padrão fitossanitários, fornecidas por viveiros idôneos que se preocupam com o padrão das mudas e que estas sejam livres de doenças.

A mancha foliar bacteriana (MFB) é uma das principais doenças do eucalipto, principalmente quando se tratando de produção de mudas em viveiros. Estima-se que esta doença tenha causado perdas na ordem de 7,5 milhões de dólares entre os anos de 2003 e 2008 (ALFENAS *et al.*, 2009).

Os principais sintomas dessa doença são lesões úmidas do tipo anasarcas, com ocorrência internervurais, angulares, concentradas ao longo da nervura principal, nas margens das folhas ou distribuídas pelo limbo foliar. Com a evolução da doença, as manchas se tornam necróticas e causam a deformação da folha. Em estágios mais avançados há um aumento do número de lesões e essas adquirem um aspecto ressecado, podendo conter orifícios no centro das lesões. Em seu ápice, ocorre uma intensa desfolha devido a precoce senescência das folhas infectadas

(ALFENAS *et al.*, 2009; AUER; SANTOS; NETO, 2011; MAFIA; TEIXEIRA; FERREIRA, 2017).

Diversos agentes etiológicos são associados a MFB do eucalipto tanto no Brasil como em diversos outros países, no entanto, o gênero *Xanthomonas* spp. tem maior ocorrência para esta doença do eucalipto (FERRAZ *et al.*, 2018).

A MFB do eucalipto é uma doença de alta importância para os produtores de mudas florestais, com isso, todo o esforço para desenvolver técnicas que auxiliem e abreviem o diagnóstico da doença ou mesmo lancem um alerta sobre a possível infecção deve ser considerado de grande valia.

A radiação solar ao atingir a planta, interage com a mesma em três frações, a primeira é absorvida pela folha e utilizado nos processos fotossintéticos, a segunda fração é refletida pelas folhas em um fenômeno chamado reflexão, a terceira fração sofre o processo de transmissão, adentrando as camadas da folha e suas estruturas (CUNHA, 2004; MOREIRA, 2011). Dessa forma, as propriedades espectrais das folhas são funções de sua composição química, morfológica e estrutura interna (COLWELL, 1974).

O comportamento espectral da vegetação se manifesta de forma distinta em três regiões do espectro eletromagnético, sendo eles, a região do visível (400 nm a 700nm) que é considerada a faixa de radiação fotossinteticamente ativa, a região do infravermelho próximo (700 nm a 1300 nm) onde a radiação interage com a estrutura interna resultando em padrões referentes a estrutura celular da folha, e a região do infravermelho médio (1300 nm a 3000 nm) onde são percebidos os padrões de respostas para o conteúdo de água na folha (CUNHA, 2004; LAW; WARING, 1994; MOREIRA, 2011; ODUM, 1983).

A reflectância espectral foliar pode ser considerado como uma resposta da planta a diversos fenômenos biofísicos e bioquímicos que ocorrem em sua estrutura, dessa forma, pode ser associada a um padrão de resposta, denominado assinatura espectral, que pode variar para um dado binômio patógeno-hospedeiro, demonstrando que esta planta está sadia ou não, de acordo com suas respostas espectrais em determinadas faixas do espectro eletromagnético. Técnicas similares veem sendo usadas em trabalho de sensoriamento remoto para monitorar plantações infectadas (FRANKE; MENZ, 2007; MAHLEIN *et al.*, 2012).

O aprendizado de máquina pode ser definido como o estudo e construção de algoritmos que podem “aprender” com base em seus erros e fazer previsões sobre dados. (SIMON, 2013).

Os modelos de aprendizado de máquinas podem ser construídos por uma técnica de seleção de atributos e um algoritmo de aprendizagem. A técnica de seleção de atributos é utilizada com o objetivo de eleger os atributos mais relevantes para o processo de aprendizagem (LANGLEY, 1994; REZENDE; MARCACINI; MOURA, 2011). Os algoritmos de classificação tem por objetivo classificar as amostras recebidas entre as classes pré-determinadas, sendo que cada um deles segue uma estratégia própria para alcançar esse objetivo (MITCHELL, 1997).

Pode ser destacado como técnicas de seleção de atributos as seguintes técnicas, *Classifier Attribute Evaluation* (WITTEN; FRANK, 2005); *Classifier Subset Evaluation* (WITTEN; FRANK, 2005); *Information Gain Attribute Evaluation* (KAREGOWDA; MANJUNATH; JAYARAM, 2010); *Gain Ratio Attribute Evaluation* (KAREGOWDA; MANJUNATH; JAYARAM, 2010); A técnica de Análise de Componentes Principais (PCA) (WOLD; ESBENSEN; GELADI, 1987); *Relief Attribute Evaluation* (KIRA; RENDELL, 1992; WITTEN; FRANK, 2005); *Symmetrical Uncert* (APPAVU *et al.*, 2011) e *Wrapper* (KOHAVI; JOHN, 1997).

Pode ser destacados como algoritmos de classificação a rede bayesiana (MEYER-BAESE; SCHMID, 2014); O classificador J48 (QUINLAN, 1994); O *Perceptron* Multicamadas (GARDNER; DORLING, 1998), conjuntamente com o método de aprendizagem *Backpropagation* (RUMELHART, 1986); O classificador OneR (HOLTE, 1993; MUDA *et al.*, 2011); O classificador Comitê Aleatório (CHOU *et al.*, 2009); O classificador Floresta Aleatórias (BREIMAN, 2001); O classificador *Support Vector Machine* (SVM) (HEARST *et al.*, 1998); O classificador Zero-R (WITTEN; FRANK, 2005). Sendo esses classificadores presentes na literatura com maior potencial de utilização para o presente trabalho.

1.1.1 Objetivo

Avaliar a combinação entre técnicas de seleção de atributos e algoritmos de classificação para discriminação de plantas de eucalipto com sintomas e sem sintomas de MFB por *Xanthomonas spp.* Permitindo identificar os comprimentos de onda em uma banda do infravermelho que possam estar associados a doença.

1.2 Material e Métodos

Foram utilizadas 288 mudas do clone IPB02 (híbrido de *Eucalyptus grandis* e *Eucalyptus urophylla*), no estado fenológico A (95 dias de idades), cedidas pelo viveiro florestal Piraflo, situado em Holambra II – Paranapanema, São Paulo. Essas mudas foram mantidas com irrigação e nutrição padronizada conforme recomendação da cultura (VALERI; CORRADINI, 2000).

As mudas foram separadas em 12 bandejas com 24 mudas. O conjunto de 6 bandejas compõem cada um dos tratamentos, denominados não inoculado (C) e inoculado (B).

O isolado PATFLO38 de *Xanthomonas spp.*, utilizado no preparo da suspensão bacteriana, pertence à coleção de culturas do Laboratório de Patologias Florestais da FCA/UNESP de Botucatu, e é mantido preservado em 30% glicerol (v/v) a -80° C, e foi cultivado em meio de cultura Nutriente-Sacarose-Ágar (NSA), este, consistindo em meio Nutriente-Ágar (NA) (SCHAAD; JONES; CHUN, 2001) acrescido de 5g L⁻¹ de sacarose, incubado a 28° C / 48 h. A suspensão bacteriana na concentração de 108 UFC mL⁻¹ foi ajustada em espectrofotômetro (OD_{600nm}=0,1) e utilizada na pulverização das mudas no tratamento inoculado.

O tratamento inoculado compreende ao conjunto de mudas que foram pulverizadas com suspensão bacteriana em suas folhas nas faces abaxial e adaxial até o ponto de escorrimento. Por outro lado, o tratamento controle corresponde às mudas que foram pulverizadas com água destilada também em ambas as faces das folhas até o ponto de escorrimento.

As mudas foram mantidas a temperatura constante de 25° C, umidade relativa do ar controlada em 80% e fotoperíodo de 12 horas.

As avaliações foram realizadas 24 (D1), 48 (D2), 72 (D3), 96 (D4), 120 (D5), 144 (D6), 168 (D7) e 192 (D8) horas após a inoculação. Para cada avaliação 6 mudas de cada tratamento foram sorteadas, sem a reposição. Para cada muda, foram coletadas aleatoriamente 4 folhas e estas acondicionadas em saco plástico identificado e mantidas em recipiente térmico até o momento da avaliação.

A reflectância espectral foliar das plantas de eucalipto, em porcentagem, foram obtidas utilizando um espectroradiômetro da marca Ocean Optics modelo

Flame-NIR, com faixa de leitura de 900 a 1700 nm e sensibilidade de 5,51 nm, por meio do software Ocean View®, do mesmo fabricante do equipamento (OCEAN OPTICS, 2018). Sendo o equipamento calibrado diariamente antes do início das leituras.

O conjunto de dados utilizado foi composto pela média da reflectância das quatro folhas analisadas de cada planta, sendo utilizado como atributos os 128 comprimentos de onda mensurado pelo espectroradiômetro, no intervalo de 900nm a 1700nm.

Foram avaliados 64 modelos de classificação, sendo cada um deles uma combinação entre um algoritmo de aprendizagem e uma técnica de seleção de atributos.

As técnicas de seleção de atributos avaliadas foram ClassifierAttributeEval-1R (WITTEN; FRANK, 2005), ClassifierAttributeEval-J48 (WITTEN; FRANK, 2005), ClassifierSubsetEval-1R (WITTEN; FRANK, 2005), ClassifierSubsetEval-J48 (WITTEN; FRANK, 2005), GainRatioAttributeEval (KAREGOWDA; MANJUNATH; JAYARAM, 2010), InfoGainAttributeEval (KAREGOWDA; MANJUNATH; JAYARAM, 2010), OneRAttributeEval (WITTEN; FRANK, 2005), PCA (WOLD; ESBENSEN; GELADI, 1987), ReliefAttributeEval (KIRA; RENDELL, 1992), SymmetricalUncertAttributeEval (APPAVU *et al.*, 2011) e WrapperSubsetEval-J48 (KOHAVI; JOHN, 1997). As técnicas que selecionam um subconjunto de atributos para classificação, tiveram todos os atributos selecionados utilizados no processo de aprendizagem, por outro lado, as técnicas que avaliam cada atributo ranqueando-os em ordem de importância, tiveram selecionados os 12 melhores atributos para o processo de aprendizagem.

Os algoritmos de aprendizado utilizados foram Rede bayesiana (BayesNet) (MEYER-BAESE; SCHMID, 2014), Árvore de decisão (J48) (QUINLAN, 1994), Multilayer Perceptron (MLP) (GARDNER; DORLING, 1998), OneR (HOLTE, 1993), Random Committee (CHOU *et al.*, 2009), Random Forest (BREIMAN, 2001), Support Vector Machine (SVM) (HEARST *et al.*, 1998), ZeroR (WITTEN; FRANK, 2005).

Para a classificação entre plantas inoculadas e controle foram utilizadas as assinaturas espectrais das plantas colhidas em todo o período do experimento. Para

classificação do período de latência, foram utilizadas somente as plantas do grupo inoculado. Para a classificação entre plantas inoculadas e controle para cada dia do período de latência foram utilizadas todas as plantas coletadas, no entanto, foi testado apenas o classificador com maior porcentagem de acertos do experimento de classificação entre plantas do grupo controle (C) e inoculado (B).

Cada uma das combinações entre o algoritmo de classificação e a técnica de seleção de atributos, tanto para a classificação entre infectado ou controle, quanto para a determinação do período de latência, foram avaliados por meio da porcentagem de classificações corretas do modelo, sendo as avaliações repetidas 1000 vezes cada, sempre com seleção aleatória de amostras, utilizando 70% das amostras para aprendizado e 30% das amostras para validação (JAMES *et al.*, 2013), utilizando o software WEKA (WITTEN; FRANK, 2005).

1.3 Resultados e Discussão

1.3.1 Discriminação entre plantas sadias e plantas inoculadas com

***Xanthomonas* spp.**

As técnicas de seleção de atributos utilizadas conjuntamente com os algoritmos de classificação entre plantas sadias (C) e inoculadas (B) com *Xanthomonas* spp. destacaram os diversos comprimentos de onda utilizados, como é mostrado na Tabela 1.

Tabela 1 - Quantidade de atributos e comprimentos de onda (nm) selecionados para cada técnica de seleção de atributos testadas com dados da reflectância foliar espectral (%) de eucalipto submetidos a infecção com *Xanthomonas* spp.

Técnica de Seleção	Quantidade de atributos	Comprimentos de onda (nm)
ClassifierAE-1R	12	1380, 1408, 1414, 1477, 1483, 1488, 1494, 1523, 1529, 1569, 1575, 1685
ClassifierSubsetEval-1R	1	1534
ClassifierSubsetEval-J48	2	1667 e 1685
OneRAE	12	1408, 1431, 1414, 1494, 1483, 1471, 1403, 1523, 1488, 1540, 1437, 1557
PCA	2	V1 e V2 (Autovalores)
ReliefFAE	12	1408, 1414, 1420, 1425, 1431, 1437, 1443, 1448, 1667, 1673, 1679, 1685
WrapperSubsetEval-1R	2	1471 e 1477
WrapperSubsetEval-J48	2	1374 e 1685

A classificação das mudas de eucalipto entre infectadas ou controle foi feita utilizando diversos algoritmos de aprendizado de máquinas, sendo que o algoritmo de *Perceptron* de Multi Camadas (MLP) obteve por quatro vezes o maior percentual de classificações corretas, como mostra a Tabela 2.

Tabela 2 - Média (desvio padrão entre parênteses) do percentual de classificações corretas para curvas de reflectância espectral de mudas de eucaliptos, submetidas e não submetidas à inoculação de *Xanthomonas* spp, para cada algoritmo de classificação (coluna) utilizando cada uma das técnicas de seleção de atributos (linha)

Seleção de Atributos	Bayes Net	J48	MLP	OneR	R. Commite	R. Forest	SVM	ZeroR
ClassifierAE-1R	60,13 (7,21)	62,77 (7,02)	68,83 (7,41)	51,95 (7,71)	56,88 (8,31)	59,02 (7,90)	55,31 (7,94)	49,17 (0,86)
ClassifierSubsetEval-1R	49,17 (0,86)	48,75 (1,95)	51,62 (6,74)	53,64 (8,26)	55,59 (7,85)	55,59 (7,85)	51,76 (6,35)	49,17 (0,86)
ClassifierSubsetEval-J48	60,13 (7,19)	63,72 (6,46)	67,18 (7,30)	53,32 (8,50)	59,72 (8,54)	60,32 (8,33)	57,18 (7,79)	49,17 (0,86)
OneRAE	49,16 (0,88)	48,42 (2,55)	51,60 (7,69)	52,57 (8,18)	49,20 (8,50)	46,28 (8,30)	51,92 (7,86)	49,17 (0,86)
PCA	49,10 (1,13)	49,10 (2,44)	51,45 (7,77)	50,65 (8,41)	51,05 (8,78)	51,80 (8,16)	52,28 (6,47)	49,17 (0,86)
ReliefFAE	60,96 (6,67)	62,03 (6,74)	69,28 (7,49)	50,65 (7,55)	57,39 (7,95)	58,86 (7,76)	54,91 (8,03)	49,17 (0,86)
WrapperSubsetEval-1R	49,16 (0,88)	48,62 (2,34)	52,29 (6,70)	51,90 (8,40)	45,38 (8,28)	44,53 (815)	53,39 (6,29)	49,17 (0,86)
WrapperSubsetEval-J48	60,14 (7,19)	63,54 (6,82)	70,06 (7,20)	51,75 (8,61)	61,88 (8,79)	63,12 (8,31)	57,34 (8,17)	49,17 (0,86)

O algoritmo de MLP (*Perceptron* Multi Camadas) combinado com a técnica de seleção WrapperSubsetEval-J48 obteve a maior porcentagem de acerto, com 70,06%, seguido pelo mesmo algoritmo com a técnica ReliefFAE com 69,28% e o mesmo algoritmo com a técnica ClassifierAE-1R com 68,83%. Mostrando que o algoritmo MLP consegue obter maiores porcentagens de classificações corretas para o presente estudo.

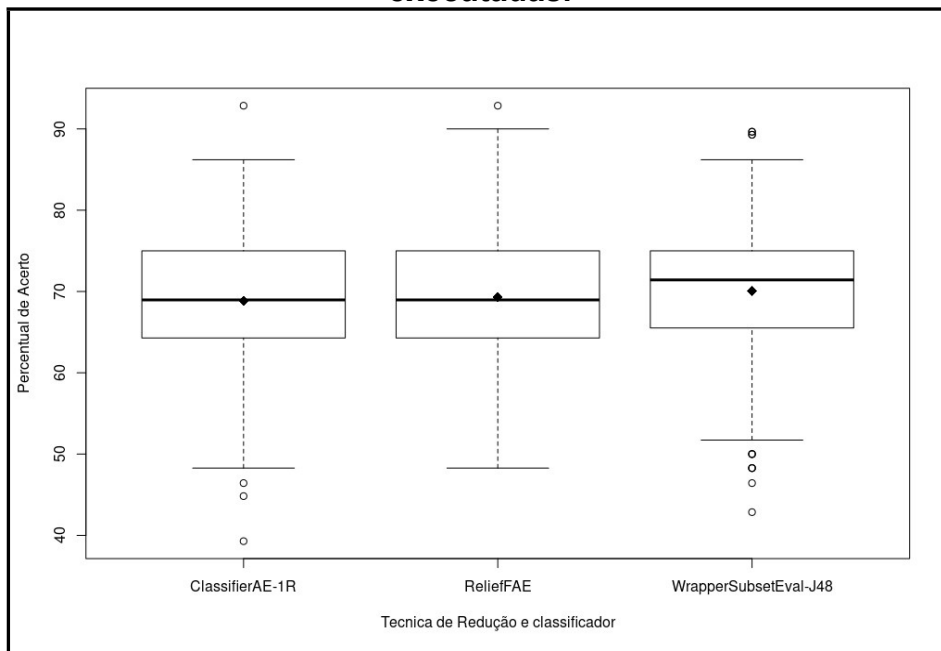
Os modelos baseados em MLP obtiveram as maiores porcentagens de classificações corretas com 4 combinações nesse estudo, os modelos baseados em SVM obtiveram maior porcentagem de classificação, 2 combinações, e os modelos baseados em J48, OneR e *Random Forest*, obtiveram maiores porcentagens de classificações corretas uma única vez.

Pode ser percebido que o algoritmo MLP consegue modelar melhor o problema investigado quando comparado com os demais algoritmos testados. O trabalho de Tamouridou *et al.*, (2018) mostra o uso da mesma técnica para identificação da doença causada por *Microbotryum silybum* para controle de plantas indesejadas. Por outro lado, o trabalho de Odabas *et al.*, (2017) mostrou o uso do

mesmo algoritmo para estimar o índice de concentração de clorofila em plantas de alface. Demonstrando a grande diversidade de problemas que podem ser modelados utilizando o algoritmo MLP.

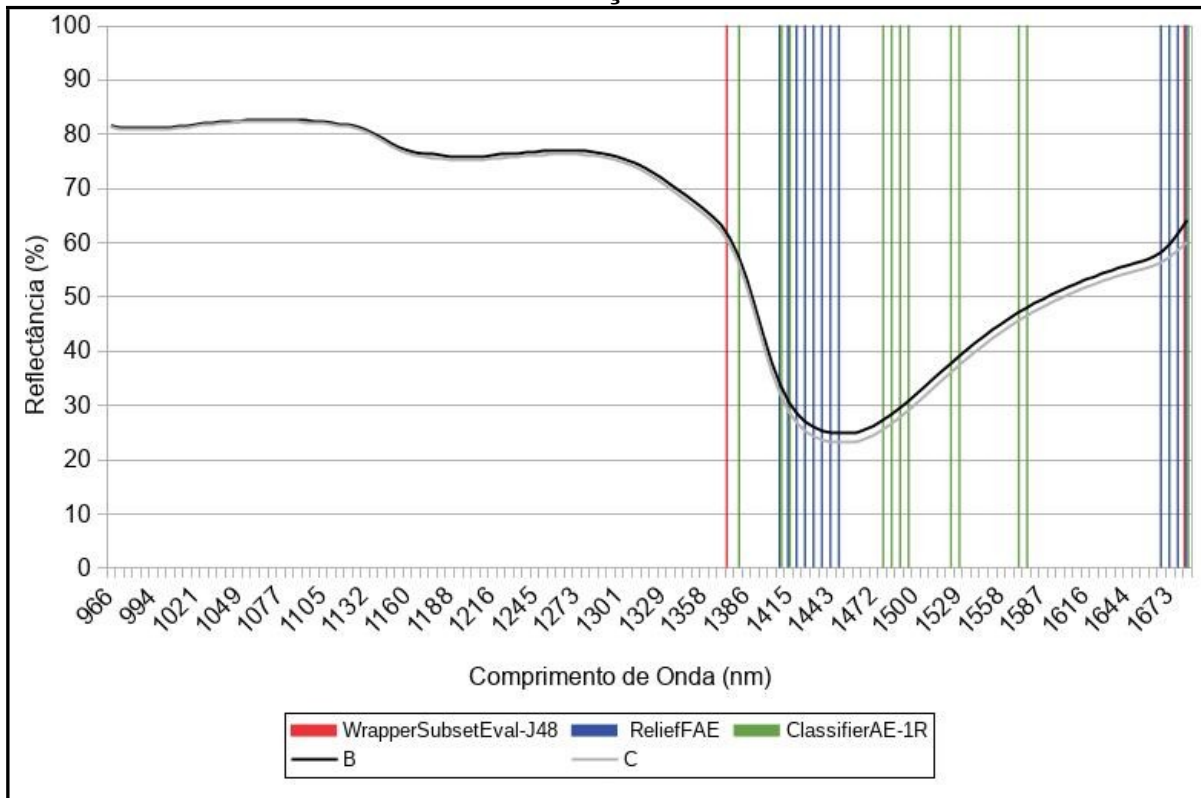
Os demais modelos de classificação testados performaram porcentagens de acertos entre 45% e 63% de classificações corretas, diferenciando-se da pontuação obtida pelos modelos destacados com as 3 maiores porcentagens de classificações corretas, que obtiverem resultados entre 68% e 70% de classificações corretas, como mostrado na Figura 2.

Figura 2 - Boxplot das porcentagens de acertos para as três melhores técnicas de seleção, utilizando algoritmo de Perceptron de multicamadas, para as mil rodadas de testes executadas.



As técnicas de seleção de atributos destacam os comprimentos de onda mais importantes para que o algoritmo de classificação possa ter resultados mais assertivos. As técnicas WrapperSubsetEval-J48, ReliefFAE e ClassifierAE-1R selecionaram comprimentos de onda diversos, como já mostrado na Tabela 1, no entanto, os comprimentos de onda selecionados sempre estão acima de 1350 nm, região onde há maior distanciamento da curva de reflectância das plantas doentes da curva de reflectância das plantas saudáveis, como Mostra a Figura 3.

Figura 3 - Curva média da reflectância foliar espectral de mudas de *Eucalyptus grandis* x *E. urophylla* submetidos (B) e não submetidos (C) a infecção com *Xanthomonas spp.* destacado os comprimentos de onda (nm) elegidos pelas técnicas de seleção de atributos



Para Moreira (2011) o intervalo entre 1300nm até 3000nm compreende a região do infravermelho médio, associado a absorção de água da planta, onde a reflectância na assinatura espectral está ligado ao conteúdo relativo de água das folhas analisadas. Outros pontos importantes selecionados estão no intervalo de 1100nm e 1300nm do comprimento de onda, que, para Mahlein *et al.*, (2012) são pontos da assinatura espectral cuja reflectância exprimem comportamentos da estrutura celular da folha.

A infiltração das células bacterianas no mesófilo da folha do hospedeiro ocorre pois bactérias que causam manchas foliares respondem aos estímulos químicos em seus pontos de entradas como estômatos, hidatódios e lenticelas para penetrar e se multiplicar nos espaços intracelulares (GONÇALVES *et al.*, 2008).

Dessa forma, essas alterações provocadas pelas *Xanthomonas spp* causam diversas alterações em suas estruturas celulares que são refletidas em sua assinatura espectral, sendo possível verificar pontos de divergências nos valores de

reflectância, para os tratamentos analisados, entre os comprimentos de onda de 1000 nm e 1200 nm.

Essa diferenciação dos valores de reflectância para a região de comprimento de onda citado, embora presente nesse estudo, não foram apontadas por nenhuma das técnicas de seleção estudadas, demonstrando que esses pontos são capazes de detectar mudanças na estrutura celular das folhas inoculadas, no entanto, a maior diferenciação na assinatura espectral está presente na faixa do infravermelho médio.

Na região do Infravermelho médio, pode utilizar a técnica de reflectância espectral para analisar o conteúdo de água das folhas, pois nessa região, são observados os pontos de absorção de água (SOUSA; RIBEIRO; PONZONI, 1996). No presente trabalho foi observado estes pontos de absorção de água, principalmente entre os comprimentos de onda 1350nm e 1550nm, sendo que esta região, teve comprimentos de ondas selecionados por todas técnicas de seleção de atributos.

O processo de infecção promovido por fitopatógenos, os quais pertencem ao grupo sintomático conhecido como “mancha”, causam efeitos sobre a translocação de água na planta devido a alterações na taxa de transpiração e por consequência mudanças no potencial hídrico (LEITE; PASCHOLATE, 2018). Dessa forma, este efeito pode estar relacionado as alterações na assinatura espectral das plantas doentes, no entanto, se faz necessários testes específicos para comprovação dessas informações.

Os três modelos com maiores porcentagens de classificações corretas selecionaram diversos comprimentos de onda nessa região, sugerindo que as alterações contidas nesses comprimentos de ondas são suficientes para distinguir entre as plantas inoculadas com *Xanthomas* spp e as plantas controle.

A técnica de Perceptron de Multi Camadas conjuntamente com a seleção de atributos pela técnica de WrapperSubSetEval-J48 apresentaram os melhores percentuais de classificações corretas dentre os modelos testados devido a sua assertividade na escolha dos comprimentos de onda que foram utilizados como parâmetros de decisão, sendo que estes pontos, contemplam a região da assinatura espectral referente ao conteúdo de água das folhas.

O trabalho de Griffel *et al.* (2018) que objetivou determinar se plantas de batatas estavam infectadas com Potato Virus Y (Potyviridae, PVY) utilizando as assinaturas espectrais dessas plantas e classificando com a técnica de Support Vector Machine, a qual alcançou 89,8% de acertos, o que se considera valores altos, diferentemente dos valores encontrados para esta técnica no presente trabalho, que variou entre 51% e 58% de classificações corretas.

Zhu *et al.*, (2017) utilizou as técnicas de aprendizado de máquina Back Propagation Neural Network, Extreme Learning Machine e Least Square Support Vector Machine e assinaturas espectrais (380 – 1023 nm) para detecção de folhas doentes de tabaco, conseguindo valores de acurácia acima de 80% para todas as técnicas testadas.

Os trabalhos citados acima demonstram a utilização das técnicas utilizadas no presente trabalho e sua efetividade na classificação de diversas doenças foliares, trazendo a luz que essas técnicas tem grande potencial de utilização no meio agroflorestal.

1.3.2 Classificação das mudas em dias após a inoculação

Para a classificação do tempo em dias após a inoculação as técnicas de seleção de atributos utilizadas resultaram na seleção dos comprimentos de onda apresentado na Tabela 3.

Tabela 3 - Quantidade de atributos e comprimentos de onda (nm) selecionados para cada técnica de seleção de atributos testadas com dados da reflectância foliar espectral de folhas de eucaliptos submetidos a infecção com *Xanthomonas spp.* para classificação em dias após inoculação

Técnica de Seleção	Quantidade de Atributos	Comprimentos de onda (nm)
ClassifierAE-1R	12	1149, 1154, 1177, 1182, 1188, 1194, 1199, 1205, 1210, 1216, 1679, 1685
ClassifierAE-J48	12	1289, 1295, 1363, 1374, 1380, 1627, 1650, 1656, 1661, 1667, 1679, 1685
ClassifierSubsetEval-1R	1	1673
ClassifierSubsetEval-J48	3	1374, 1386, 1685
GainRatioAE	12	1261, 1272, 1278, 1284, 1289, 1295, 1301, 1312, 1329, 1334, 1679, 1685
InfoGainAE	12	1082, 1087, 1093, 1099, 1110, 1115, 1121, 1126, 1132, 1188, 1679, 1685
OneRAE	12	1177, 1194, 1199, 1205, 1210, 1216, 1222, 1227, 1233, 1244, 1673, 1679
PCA	2	V1 e V2 (Autovalores)
ReliefFAE	12	977, 1182, 1216, 1222, 1289, 1295, 1301, 1306, 1312, 1673, 1679, 1685
SymmetricalUncertAE	12	1082, 1087, 1093, 1099, 1110, 1115, 1121, 1126, 1132, 1188, 1679, 1685
WrapperSubsetEval-J48	4	1391, 1627, 1679, 1685

A combinação entre técnicas de seleção de atributos e algoritmos de classificação resultaram em 80 modelos que tiveram como parâmetro de avaliação a porcentagem de classificações corretas, conforme apresentado na Tabela 4

Tabela 4 - Média (desvio padrão entre parênteses) do percentual de classificações corretas para curvas de reflectância espectral de mudas de eucaliptos, submetidas à inoculação de *Xanthomonas spp*, para cada algoritmo de classificação (coluna) utilizando cada uma das técnicas de seleção de atributos (linha) para a classificação em dias após a inoculação

Seleção de Atributos	Bayes Net	J48	MLP	OneR	R. Commite	R. Forest	SVM	ZeroR
ClassifierAE-1R	27,16 (7,55)	31,92 (10,30)	23,58 (8,83)	8,04 (2,06)	28,05 (10,23)	29,30 (10,29)	23,28 (9,43)	8,04 (2,06)
ClassifierAE-J48	26,27 (7,85)	36,99 (11,08)	42,88 (11,65)	8,04 (2,06)	33,17 (11,06)	34,84 (10,62)	19,96 (9,24)	8,04 (2,06)
ClassifierSubsetEval-1R	17,12 (6,06)	25,89 (8,94)	20,34 (8,32)	8,04 (2,06)	18,98 (8,84)	18,97 (8,84)	12,54 (7,09)	8,04 (2,06)
ClassifierSubsetEval-J48	22,27 (6,15)	38,66 (12,03)	46,37 (11,63)	8,04 (2,06)	37,63 (11,20)	37,39 (11,03)	18,61 (9,04)	8,04 (2,06)
GainRatioAE	26,93 (7,21)	33,08 (10,16)	28,62 (10,14)	8,04 (2,06)	30,03 (10,32)	30,59 (10,43)	23,13 (9,24)	8,04 (2,06)
InfoGainAE	26,00 (7,09)	30,59 (10,15)	23,43 (8,96)	8,04 (2,06)	26,31 (10,15)	26,26 (10,09)	22,24 (9,28)	8,04 (2,06)
OneRAE	23,50 (6,79)	29,76 (9,78)	23,05 (9,21)	8,04 (2,06)	23,96 (9,52)	26,11 (9,65)	22,03 (9,27)	8,04 (2,06)
PCA	11,86 (3,42)	22,53 (6,59)	20,98 (6,22)	19,16 (5,12)	21,15 (6,67)	23,07 (6,43)	17,52 (5,46)	10,51 (0,57)
ReliefFAE	26,63 (7,55)	32,69 (10,08)	26,41 (9,92)	8,04 (2,06)	29,53 (10,25)	30,43 (10,28)	23,13 (9,48)	8,04 (2,06)
SymmetricalUAE	26,00 (7,09)	30,59 (10,15)	23,43 (8,96)	8,04 (2,06)	26,31 (10,15)	26,26 (10,09)	22,24 (9,28)	8,04 (2,06)
WrapperSubsetEval-J48	25,81 (7,78)	42,64 (12,58)	49,38 (12,20)	8,04 (2,06)	38,15 (11,53)	35,59 (10,96)	20,17 (8,82)	8,04 (2,06)

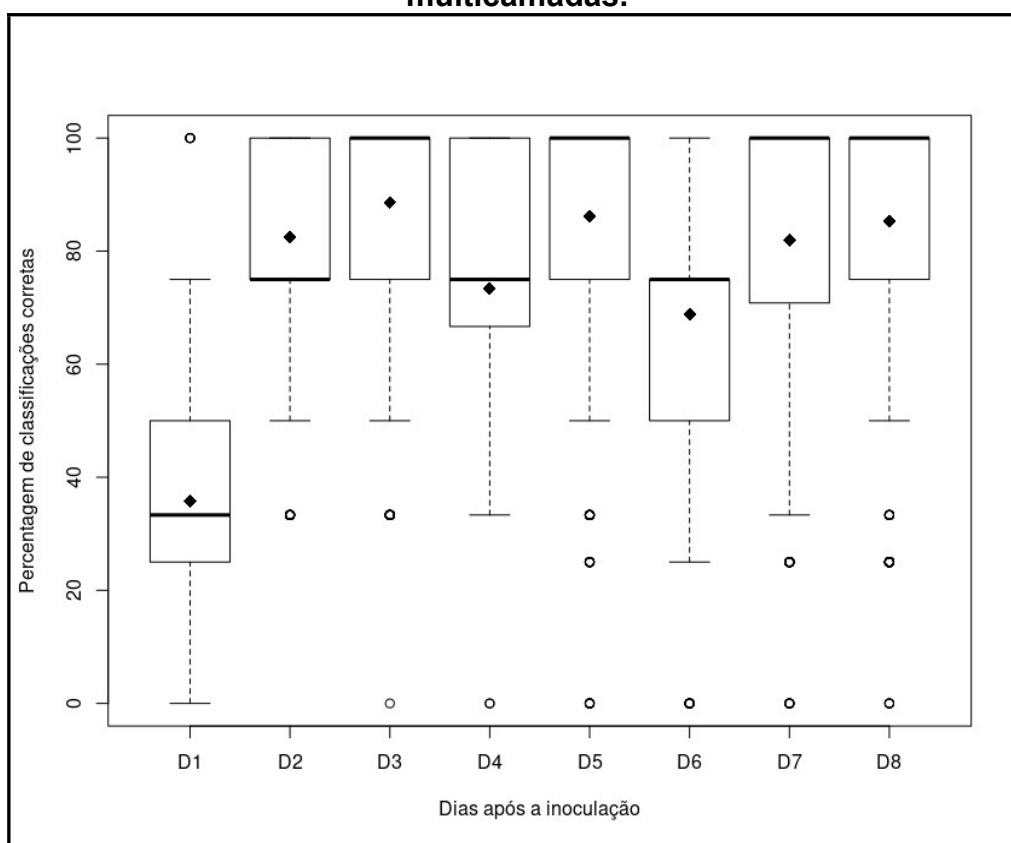
Dentre todos os modelos testados, aqueles com maior média de porcentagem de classificações corretas foram as combinações de MLP com WrapperSubsetEval-J48, com média de 49,37%, MLP com ClassifierSubsetEval-J48 com média de 46,36% e MLP com ClassifierAttributeEval-J48 com média de 42,88% de classificações corretas.

Foi percebido que os modelos testados, não alcançam percentagens de classificações corretas acima de 50% o que pode ser considerado como um resultado insatisfatório para a resolução do problema modelado, dessa forma, outros modelos e outras técnicas de investigação devem ser utilizados para resolução desse tipo de classificação

1.3.3 Detecção pré sintomática da doença

O algoritmo com melhor desempenho na classificação entre plantas inoculadas e controle – MLP com WrapperSubsetEval-J48 – foi testado para classificar as plantas entre saudáveis e doentes, no entanto, o *dataset* foi separada por dias após a inoculação, a fim de perceber a partir de qual dia após a inoculação seria possível ter uma boa classificação entre plantas doentes e saudáveis. A Figura 3 apresenta o resultado da classificação para cada dia de experimento.

Figura 4 - Boxplot das porcentagens de acertos entre plantas não inoculadas e inoculadas com *Xanthomonas spp.* em folhas de clone híbrido de *Eucalyptus grandis* x *E. urophylla* para cada dia após a inoculação utilizando algoritmo de Perceptron de multicamadas.



Ao primeiro dia após a inoculação (D1) a média de classificações corretas é baixa, em torno de 36%, mostrando que não há uma boa discriminação entre plantas saudáveis e doentes com este modelo. No entanto, no segundo e terceiros dias, há um aumento significativo no percentual de classificações corretas, em torno de 83%

e 89% mostrando que o modelo testado pode classificar corretamente entre as plantas controle e inoculadas com alta taxa de acertos.

No quarto (D4) e sexto (D6) dias houve um decréscimo da percentagem de classificações corretas, obtendo valores respectivamente de 74% e 69%, valores estes ainda considerados aceitáveis, mas ainda sim, abaixo da média obtida nos outros dias testados.

O quinto (D5), sétimo (D7) e oitavo (D8) dias apresentam novamente um aumento no percentual médio de classificações corretas, obtendo respectivamente 84%, 82% e 85%, mostrando assim que nesses últimos dias há uma retomada nos acertos obtidos pelo modelo.

Ferraz *et al.*, (2018) Mostraram em seu trabalho que os primeiros sintomas da MFB do eucalipto causada por *Xanthomonas spp.* surgiram no sétimo dia após a inoculação, tempo este considerado correto, por outros autores, dependendo das condições de plantio e manejo da cultura (AUER; SANTOS; NETO, 2011; GONÇALVES *et al.*, 2008).

Zhu *et al.* (2017) apresentaram o uso de técnicas de *machine learning* como *back propagation neural network* para detecção pré sintomática de Tobacco mosaic Virus (TMV) em fumo, conseguindo detectar a doença em 48 horas após inoculação. Mesmo não havendo nenhuma relação do patógeno ou do hospedeiro trabalhado nesse artigo, foi demonstrado como técnicas de inteligência artificial vem sendo empregadas em uma tentativa de antecipar a detecção de diversas doenças.

A detecção pré sintomática de doenças utilizando reflectância espectral e inteligências artificial pode ser encontrado em diversos cenários, como o trabalho de Rumpf *et al.* (2010) que utilizou *Support Vector Machine* para detecção de mancha foliar de *Cercospora*, mancha de ferrugem e oídio em cana-de-açúcar. VEYS *et al.* (2019) utilizou a técnica de *Support Vector Machine* para a detecção precoce de mancha de *Pyrenopeziza brassicae* em *Brassica napus L.* Neste trabalho, a detecção se deu ao terceiro dia, com acurácia de 82%.

Dessa forma, metodologias de detecção precoce de doenças veem sendo criadas, utilizando inclusive imagens de veículos aéreos não tripulados (VANT) para observação em grandes talhões (SANDINO *et al.*, 2018)

Isto posto, a possibilidade da detecção da MFB do eucalipto causada por *Xanthomonas spp*, aos dois dias após a infecção, e com resultados melhores aos três dias após a infecção, indica que o modelo testado no presente trabalho apresenta grande potencial de aplicação da metodologia utilizada, sendo esta complementada com estudos posteriores.

1.4 Conclusão

De todas as combinações entre técnicas de seleção de atributos e algoritmos de classificação testadas, o algoritmo de *Multilayer Perceptron* (MLP) obteve os melhores resultados, sendo que utilizando a seleção de atributos por WrapperSubsetEval-J48, obteve a maior porcentagem de classificações corretas para mudas sadias e infectadas com *Xanthomonas spp*.

Esta técnica de seleção de atributos elegeu como comprimentos de onda mais importantes para classificação os comprimentos 1374 nm e 1685 nm, demonstrando que estes são suficientes para classificar entre mudas sadias e doentes.

Os modelos testados para classificação de dias após a inoculação apresentaram resultados inferiores aos considerados recomendados para uma boa classificação, dessa forma, se faz necessário maior aprimoramento dos modelos para testes futuros.

O modelo MLP, utilizando os comprimentos de ondas 1374 nm e 1685 nm possibilitaram a detecção pré sintomática da doença, com resultados considerados altos a partir do segundo dia, e apresentando seus melhores resultados três dias após a inoculação

Referências

ALFENAS, A. C. *et al.* **Clonagem e Doenças do Eucalipto**. 2 ed. ed. Viçosa, MG: Editora UFV, 2009.

APPAVU, S. *et al.* **Bayes Theorem and Information Gain Based Feature Selection for Maximizing the Performance of Classifiers**. (N. Meghanathan, B. K. Kaushik, D. Nagamalai, Eds.) *Advances in Computer Science and Information Technology. Anais...: Communications in Computer and Information Science*. Springer Berlin Heidelberg, 2011

AUER, C. G.; SANTOS, Á. F. DOS; NETO, J. R. **Mancha foliar bacteriana em plantios de eucalipto na região Sul do Brasil**. Colombo, PR: Embrapa, 2011.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 1 out. 2001.

CHOU, T. *et al.* **Ensemble of machine learning algorithms for intrusion detection**. 2009 IEEE International Conference on Systems, Man and Cybernetics. **Anais...** In: 2009 IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. out. 2009

COLWELL, J. E. Vegetation canopy reflectance. **Remote Sensing of Environment**, v. 3, n. 3, p. 175–183, 1 jan. 1974.

CUNHA, C. A. H. DA. **Relação entre comportamento espectral, índice de área foliar e produção de matéria seca em capim Tanzânia submetido a diferentes níveis de irrigação e doses de nitrogênio**. text—[s.l.] Universidade de São Paulo, 5 out. 2004.

FERRAZ, H. G. M. *et al.* *Xanthomonas axonopodis* pv. *eucalyptorum* pv. nov. Causing Bacterial Leaf Blight on Eucalypt in Brazil. **The Plant Pathology Journal**, v. 34, n. 4, p. 269–285, ago. 2018.

FRANKE, J.; MENZ, G. Multi-temporal wheat disease detection by multi-spectral remote sensing. **Precision Agriculture**, v. 8, n. 3, p. 161–172, 1 jun. 2007.

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, p. 2627–2636, 1 ago. 1998.

GONÇALVES, R. C. *et al.* Etiology of bacterial leaf blight of eucalyptus in Brazil. **Tropical Plant Pathology**, v. 33, n. 3, p. 180–188, jun. 2008.

GRIFFEL, L. M.; DELPARTE, D.; EDWARDS, J. Using Support Vector Machines classification to differentiate spectral signatures of potato plants infected with Potato Virus Y. **Computers and Electronics in Agriculture**, v. 153, p. 318–324, 1 out. 2018.

HEARST, M. A. *et al.* Support vector machines. **IEEE Intelligent Systems and their Applications**, v. 13, n. 4, p. 18–28, jul. 1998.

HOLTE, R. C. Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. **Machine Learning**, v. 11, n. 1, p. 63–90, 1 abr. 1993.

IBA. **Relatório 2017**. [s.l.] Industria Brasileira de Arvores, 2017. Disponível em: <<https://www.iba.org/datafiles/publicacoes/pdf/iba-relatorioanual2017.pdf>>. Acesso em: 15 out. 2018.

IBA. **Sumário Executivo 2018**. [s.l.] Industria Brasileira de Arvores, 2018. Disponível em: <<https://www.iba.org/datafiles/publicacoes/relatorios/digital-sumarioexecutivo-2018.pdf>>. Acesso em: 13 fev. 2019.

JAMES, G. *et al.* (EDS.). **An introduction to statistical learning: with applications in R**. New York: Springer, 2013.

KAREGOWDA, A. G.; MANJUNATH, A. S.; JAYARAM, M. A. COMPARATIVE STUDY OF ATTRIBUTE SELECTION USING GAIN RATIO AND CORRELATION BASED FEATURE SELECTION. **International Journal of Information Technology and Knowledge Management**, v. 2, n. 2, p. 7, 2010.

KIRA, K.; RENDELL, L. A. A Practical Approach to Feature Selection. In: SLEEMAN, D.; EDWARDS, P. (Eds.). . **Machine Learning Proceedings**. San Francisco (CA): Morgan Kaufmann, 1992. p. 249–256.

KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. **Artificial Intelligence, Relevance**. v. 97, n. 1, p. 273–324, 1 dez. 1997.

LANGLEY, P. **Selection of Relevant Features in Machine Learning.**: Fort Belvoir, VA: Defense Technical Information Center, 1 nov. 1994. Disponível em: <<http://www.dtic.mil/docs/citations/ADA292575>>. Acesso em: 13 fev. 2019.

LAW, B. E.; WARING, R. H. Remote Sensing of Leaf Area Index and Radiation Intercepted by Understory Vegetation. **Ecological Applications**, v. 4, n. 2, p. 272–279, 1994.

LEITE, B.; PASCHOLATE, S. ALTERAÇÕES FISIOLÓGICAS EM PLANTAS DOENTES. In: AMORIM, L.; BERGAMIN FILHO, A.; REZENDE, J. A. M. (Eds.). . **Manual de Fitopatologia**. 5. ed. [s.l.] Agronômica Ceres, 2018. v. 1p. 573.

MAFIA, R. G.; TEIXEIRA, L. P.; FERREIRA, M. A. Produção de mudas de eucalipto sob dois sistemas de irrigação e a ocorrência da bacteriose foliar (*Xanthomonas axonopodis* Vauterin). **Scientia Forestalis**, v. 45, n. 116, 1 dez. 2017.

MAHLEIN, A.-K. *et al.* Recent advances in sensing plant diseases for precision crop protection. **European Journal of Plant Pathology**, v. 133, n. 1, p. 197–209, 1 maio 2012.

MEYER-BAESE, A.; SCHMID, V. Chapter 6 - Statistical and Syntactic Pattern Recognition. In: MEYER-BAESE, A.; SCHMID, V. (Eds.). . **Pattern Recognition and Signal Analysis in Medical Imaging**. 2. ed. Oxford: Academic Press, 2014. p. 151–196.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto**. 4 Ed ed. Viçosa, MG: Editora UFV, 2011.

MUDA, Z. *et al.* **Intrusion detection based on k-means clustering and OneR classification**. 2011 7th International Conference on Information Assurance and Security (IAS). **Anais...** In: 2011 7TH INTERNATIONAL CONFERENCE ON INFORMATION ASSURANCE AND SECURITY (IAS). dez. 2011

OCEAN OPTICS. **OceanView 1.6.7Ocean Optics**, 2018. Disponível em: <<https://oceanoptics.com/product/oceanview/>>. Acesso em: 13 fev. 2019

- ODABAS, M. S. *et al.* Multilayer Perceptron Neural Network Approach to Estimate Chlorophyll Concentration Index of Lettuce (*Lactuca sativa* L.). **Communications in Soil Science and Plant Analysis**, v. 48, n. 2, p. 162–169, 19 jan. 2017.
- ODUM, E. P. **Ecologia**. 1 ed ed. Rio de Janeiro: Editora Interamericana, 1983. v. 1
- QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Mateo, CA: Morgan Kaufmann, 1994.
- REZENDE, S. O.; MARCACINI, R. M.; MOURA, M. F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação**, n. 7, p. 15, 2011.
- RUMELHART, D. E. Learning internal representations by error propagation. In: RUMELHART, D. E.; MCCLELLAND, J. L. (Eds.). . **Parallel distributed processing: explorations in the microstructure of cognition**. Cambridge: MIT Press, 1986. v. 1p. 318–362.
- RUMPF, T. *et al.* Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. **Computers and Electronics in Agriculture**, v. 74, n. 1, p. 91–99, 1 out. 2010.
- SANDINO, J. *et al.* Aerial Mapping of Forests Affected by Pathogens Using UAVs, Hyperspectral Sensors, and Artificial Intelligence. **Sensors**, v. 18, n. 4, p. 944, abr. 2018.
- SCHAAD, N. W.; JONES, J. B.; CHUN, W. (EDS.). **Laboratory guide for identification of plant pathogenic bacteria**. 3rd ed ed. St. Paul, Minn: American Phytopathological Society, 2001.
- SIMON, P. **Too big to ignore: the business case for big data**. Hoboken, New Jersey: John Wiley & Sons, Inc, 2013.
- SOUSA, C. L. DE; RIBEIRO, M. C.; PONZONI, F. J. Influência Do Tempo E Do Tipo De Armazenamento Na Reflectância Espectral De Folhas De *Eucalyptus Grandis* “EX-SITU”. **Revista Árvore**, v. 20, n. 2, p. 255–265, 1996.
- TAMOURIDOU, A. A. *et al.* Spectral Identification of Disease in Weeds Using Multilayer Perceptron with Automatic Relevance Determination. **Sensors**, v. 18, n. 9, p. 2770, set. 2018.
- VALERI, S. V.; CORRADINI, L. Fertilização em viveiros para produção de mudas de eucaliptos e pinus. In: GONÇALVES, J. L. DE M.; BENEDETTI, V. (Eds.). . **Nutrição e fertilização florestal**. 1. ed. Piracicaba: UPEF, 2000. p. 427.
- VEYS, C. *et al.* Multispectral imaging for presymptomatic analysis of light leaf spot in oilseed rape. **Plant Methods**, v. 15, n. 1, p. 4, 24 jan. 2019.
- WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2nd ed ed. Amsterdam; Boston, MA: Morgan Kaufman, 2005.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2. ed. [s.l: s.n.].

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists. v. 2, n. 1, p. 37–52, 1 ago. 1987.

ZHU, H. *et al.* Hyperspectral Imaging for Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-learning Classifiers. **Scientific Reports**, v. 7, n. 1, p. 4125, 23 jun. 2017.

**CAPÍTULO 2: MODELOS COMPUTACIONAIS E ESTATÍSTICOS PARA A
PREDIÇÃO DA SEVERIDADE DA MANCHA FOLIAR CAUSADA POR
XANTHOMONAS SPP. EM CLONE HÍBRIDO DE *EUCALYPTUS GRANDIS* X
*EUCALYPTUS UROPHYLLA***

João Ricardo Favan; José Raimundo de Souza Passos

RESUMO

A mancha foliar bacteriana (MFB) do eucalipto causada por *Xanthomonas spp.* é uma das principais doenças da eucaliptocultura no Brasil, causando perdas de milhares de reais anualmente. A forma tradicional de quantificação da doença feita por um observador com auxílio de uma escala diagramática esta sujeita a diversas fontes de erros, por isso, métodos alternativos utilizando imagens digitais e reflectância foliar espectral veem surgindo com promessas de alta acurácia e menor termo de avaliação. O presente trabalho teve por objetivo avaliar diversos modelos computacionais baseado em aprendizado de máquinas e modelos estatísticos de regressão linear múltipla para a predição da severidade da MFB do eucalipto. Mudanças de clone híbrido de *Eucalyptus grandis* x *Eucalyptus urophylla* foram submetidos a inoculação de *Xanthomonas spp.* A reflectância foliar espectral dessas mudas foram tomadas durante oito dias consecutivos. Os padrões de reflectância foram analisados por 35 modelos baseados em aprendizado de máquina e dois modelos de regressão linear múltipla. Os modelos foram submetidos à 1000 ensaios cada um sendo avaliados pelo coeficiente de correlação para os modelos de aprendizado de máquinas e pelo coeficiente de determinação ajustado para os modelos estatísticos. Os modelos baseados em aprendizado de máquina obtiveram coeficiente de correlação abaixo de 0,35; enquanto os modelos estatísticos resultaram em coeficientes de determinação ajustado acima de 55%. Nenhum dos modelos testados obtiveram resultados considerados elevados, no entanto, os modelos baseados em regressão linear múltipla obtiveram valores superiores aos modelos baseados em aprendizado de máquinas.

Palavras-chave: Aprendizado de máquina. Regressão linear múltipla. Máquinas de vetores de suporte.

ABSTRACT

Bacterial leaf spot of eucalyptus caused by *Xanthomonas spp.* is one of the main disease of eucalyptus crops in Brazil, causing losses annually. Traditional form of quantification of the disease are made by an observer with a diagrammatic scale is subject to several sources of errors. Alternative methods using digitizing images and spectral leaf reflectance promises high accuracy and shorter evaluation time. Present work objectives evaluate several computational models based on machine learning and statistical models of multiple linear regression for the prediction of the severity of the eucalyptus bacterial spot. Hybrid clone seedlings of *Eucalyptus grandis* x *Eucalyptus urophylla* were inoculated with *Xanthomonas spp.* The spectral leaf reflectance of these seedlings were taken during eight consecutive days. The reflectance patterns were analyzed by 35 models based on machine learning and two models of multiple linear regression. Models were submitted to 1000 tests each being evaluated by the correlation coefficient for the machine learning models and by the adjusted determination coefficient for the statistical models. Models based on machine learning obtained correlation coefficient below 0.35; while the statistical models resulted in adjusted determination coefficients upper to 55%. No one models obtained high results, however, the models based on multiple linear regression obtained better values than models based on machine learning.

Keywords: Machine Learning. multiple linear regression. Support vector machines.

2.1 Introdução

O setor de florestas plantadas brasileiro tem grande importância para a economia nacional. No ano de 2017 obteve uma receita bruta de 73,8 bilhões de Reais e 1,1% de participação no PIB nacional. A área de florestas plantadas totalizam 7,84 milhões de hectares e seus produtos são destinados principalmente para o segmento industrial de papel e celulose, siderurgia e carvão vegetal, painéis de madeira e também a indústria moveleira (IBA, 2018). Dentre toda a área de florestas plantadas, as florestas de *Eucalyptus spp.* se destacam ocupando 5,67 milhões de hectares, sendo esta a principal cultura florestal no Brasil (IBA, 2017).

Para o sucesso na condução de uma floresta plantada é importante ter um manejo apropriado, assim como, a utilização de mudas de boa qualidade e com alto

padrão fitossanitários, fornecidas por viveiros idôneos que se preocupam com o padrão das mudas e que estas sejam livres de doenças.

A mancha foliar bacteriana (MFB) do eucalipto é uma doença bastante preocupante para os produtores de mudas florestais, com isso, todo o esforço para desenvolver técnicas que auxiliem e abreviem o diagnóstico da doença ou mesmo lancem um alerta sobre a possível infecção deve ser considerado de grande valia.

A MFB é uma das principais doenças do Eucalipto, principalmente quando se tratando de produção de mudas em viveiros. Estima-se que esta doença tenha causado perdas na ordem de 7,5 milhões de dólares entre os anos de 2003 e 2008 (ALFENAS *et al.*, 2009).

Os principais sintomas dessa doença são lesões úmidas do tipo anasarcas, com ocorrência internervurais, angulares, concentradas ao longo da nervura principal, nas margens das folhas ou distribuídas pelo limbo foliar. Com a evolução da doença, as manchas se tornam necróticas e causam a deformação da folha. Em estágios mais avançados há um aumento do número de lesões e essas adquirem um aspecto ressecado, podendo conter orifícios no centro das lesões. Em seu ápice, ocorre uma intensa desfolha devido a precoce senescência das folhas infectadas (ALFENAS *et al.*, 2009; AUER; SANTOS; NETO, 2011; MAFIA; TEIXEIRA; FERREIRA, 2017).

Diversos agentes etiológicos são associados a MFB do eucalipto tanto no Brasil como em diversos outros países, no entanto, o gênero *Xanthomonas spp.* tem maior ocorrência para esta doença do eucalipto (FERRAZ *et al.*, 2018).

A quantificação da severidade de doenças sempre foi alvo de diversos estudos, inicialmente, eram feitas com observadores com e sem apoio de uma escala diagramática (PARKER; SHAW; ROYLE, 1995), posteriormente, passou-se a utilizar imagens digitais e software como o QUANT (VALE; FERNANDES FILHO; LIBERATO, 2003) e mais recentemente, veem se adotando o uso de reflectância foliar espectral associado ao uso de algoritmos computacionais e modelos estatísticos (BOCK *et al.*, 2010; HUANG *et al.*, 2007).

A radiação solar ao atingir a planta, interage com a mesma em três frações, a primeira é absorvida pela folha e utilizado nos processos fotossintéticos, a segunda fração é refletida pelas folhas em um fenômeno chamado reflexão, a terceira fração sofre o processo de transmissão, adentrando as camadas da folha e suas estruturas (CUNHA, 2004; MOREIRA, 2011). Dessa forma, as propriedades espectrais das folhas são funções de sua composição química, morfológica e estrutura interna (COLWELL, 1974).

O comportamento espectral da vegetação se manifesta de forma distinta em três regiões do espectro eletromagnético, sendo eles, a região do visível (400 nm a 700nm) que é considerada a faixa de radiação fotossinteticamente ativa, a região do infravermelho próximo (700 nm a 1300 nm) onde a radiação interage com a estrutura interna resultando em padrões referentes a estrutura celular da folha, e a região do infravermelho médio (1300 nm a 3000 nm) onde são percebidos os padrões de respostas para o conteúdo de água na folha (CUNHA, 2004; LAW; WARING, 1994; MOREIRA, 2011; ODUM, 1983).

A reflectância espectral foliar pode ser considerado como uma resposta da planta a diversos fenômenos biofísicos e bioquímicos que ocorrem em sua estrutura, dessa forma, pode ser associada a um padrão de resposta, denominado assinatura espectral, que pode variar para um dado binômio patógeno-hospedeiro, demonstrando que esta planta está sadia ou não, de acordo com suas respostas espectrais em determinadas faixas do espectro eletromagnético. Técnicas similares veem sendo usadas em trabalho de sensoriamento remoto para monitorar plantações infectadas (FRANKE; MENZ, 2007; MAHLEIN *et al.*, 2012).

O aprendizado de máquina é considerado como o estudo e construção de algoritmos que podem “aprender” com base em seus erros e fazer previsões sobre dados. Esses algoritmos constroem modelos a partir das amostras direcionadas como entradas a fim de fazer previsões ou decisões guiadas por esses dados em vez de seguir instruções previamente programadas (SIMON, 2013).

Os modelos de aprendizado de máquinas são formados por um algoritmo de aprendizagem, que pode objetivar a classificação de amostras, a predição de valores e o agrupamento de conjuntos semelhantes, e, podem contar com uma técnica de seleção de atributos. A técnica de seleção de atributos é utilizada com o

objetivo de selecionar os atributos mais relevantes para o processo de aprendizagem (LANGLEY, 1994).

As técnicas de seleção de atributos buscam encontrar os atributos mais importantes ou com maior relação com o fenômeno estudado, de forma a eliminar os atributos que não sejam relevantes ou que possam causar redução na efetividade do modelo desenvolvido (REZENDE; MARCACINI; MOURA, 2011). Em destaque pode ser citada as técnicas *Classifier Subset Evaluation* (WITTEN; FRANK, 2005); Análise de Componentes Principais (WOLD; ESBENSEN; GELADI; 1987) e *Wrapper* (KOHAVI; JOHN, 1997)

Os algoritmos de predição tem por objetivo prever o valor de determinado atributo baseado nos valores dos demais atributos apresentados, cada algoritmo busca uma estratégia diferente para alcançar tal objetivo (MITCHELL, 1997). Sendo destacado no presente trabalho os algoritmos *Linear Regression* (AKAIKE, 1974; WITTEN; FRANK, 2005); O *Perceptron* Multicamadas (GARDNER; DORLING, 1998), conjuntamente com o método de aprendizagem *Backpropagation* (RUMELHART, 1986); *Support Vector Machine* (Shevade *et al.*, 2000; Smola; Schölkopf, 2004); Floresta Aleatórias (BREIMAN, 2001) e Comitê Aleatório (CHOU *et al.*, 2009);

2.1.1 Objetivo

Avaliar a performance de modelos computacionais baseados em aprendizado de máquinas e de modelos estatísticos para a predição da severidade da mancha foliar bacteriana do eucalipto causada por *Xanthomonas spp.* a partir de dados de reflectância foliar espectral.

2.2 Material e métodos

Foram utilizadas 144 mudas do clone IPB02 (híbrido de *Eucalyptus grandis* e *Eucalyptus urophylla*), no estado fenológico A (95 dias de idades), cedidas pelo viveiro florestal Piraflores, situado em Holambra II – Paranapanema, São Paulo. Essas mudas foram mantidas com irrigação e nutrição padronizada conforme recomendação da cultura (VALERI; CORRADINI, 2000). As mudas utilizadas no experimento foram separadas em 6 bandejas com 24 mudas cada uma.

As mudas foram pulverizadas com suspensão bacteriana em suas folhas nas faces abaxial e adaxial até o ponto de escorrimento. O isolado PATFLO38 de *Xanthomonas spp.*, utilizado no preparo da suspensão bacteriana, pertence à coleção de culturas do Laboratório de Patologia Florestais da FCA/UNESP de Botucatu, e foi mantido preservado em 30% glicerol (v/v) a -80° C, e foi cultivado em meio de cultura Nutriente-Sacarose-Ágar (NSA), este, consistindo em meio Nutriente-Ágar (NA) (SCHAAD; JONES; CHUN, 2001) acrescido de 5g L⁻¹ de sacarose, incubado a 28° C / 48 h. A suspensão bacteriana na concentração de 10⁸ UFC mL⁻¹ foi ajustada em espectrofotômetro (OD_{600nm}=0,1) e utilizada na pulverização das mudas do presente trabalho.

As bandejas com as mudas foram acondicionadas na Câmara de Inoculação do Departamento de Proteção Vegetal da FCA/UNESP de Botucatu, sendo este ambiente mantido a temperatura constante de 25° C, umidade relativa do ar controlada em 80% e fotoperíodo de 12 horas.

A partir de 24 horas após a inoculação das mudas, iniciou-se o período de avaliação. Durante 8 dias consecutivos 6 mudas foram sorteadas, sem a reposição. Para cada muda, eram coletadas aleatoriamente 4 folhas e estas acondicionadas em saco plástico identificado e mantidas em recipiente térmico até o momento da avaliação.

Anteriormente a avaliação, foi reduzida a umidade superficial de cada folha utilizando-se toalhas de papel e estas folhas foram coladas em um gabarito de papel sulfite contendo a identificação da folha no experimento.

A reflectância espectral foliar das plantas de eucalipto, em porcentagem, foram obtidas utilizando um espectroradiômetro da marca Ocean Optics modelo Flame-NIR, com faixa de leitura de 900nm a 1700nm e sensibilidade de 5,51nm, por meio do software Ocean View®, do mesmo fabricante do equipamento (OCEAN OPTICS, 2018). Sendo o equipamento calibrado diariamente antes do início das leituras, conforme orientação do fabricante.

Foi obtida imagem digital de cada folha utilizando uma câmera fotográfica da marca Canon, modelo EOS Rebel T5i Full HD 18MP, equipada com lentes 18-55mm, f/3,5-5,6; Devidamente travada em um tripé e posicionada a 50 cm do alvo.

A severidade da doença bacteriana foi avaliada através da razão entre o número de pústulas observadas na folha e sua área foliar (cm²). O número de pústulas de cada folha foi obtido por meio de contagem visual com auxílio de um microscópio digital 2MP LED USB com ampliação de 500X. A área foliar foi obtida a partir da imagem digital da folha, utilizando o software de processamento de imagem ImageJ® (SCHNEIDER; RASBAND; ELICEIRI, 2012).

O conjunto de dados utilizado foi composto pela reflectância espectral (%) das 192 folhas analisadas, sendo utilizado como atributos cada comprimento de onda (nm) mensurado pelo espectralradiômetro e o valor de severidade (número de pústulas/cm²) calculado.

As técnicas de seleção de atributos procuram definir os comprimentos de onda mais importantes para a predição da severidade. Foram utilizadas as técnicas de seleção de atributos *Classifier Subset Evaluation* (WITTEN; FRANK, 2005), com os classificadores de atributos *Linear Regression* (Classifier-LR) (WITTEN; FRANK, 2005), *M5 Rules* (Classifier-M5) (HOLMES; HALL; PRANK, 1999) e *Support Vector Machine* para regressão (Classifier-SMOReg) (SMOLA; SCHÖLKOPF, 2004), Análise de componentes principais (PCA) (WOLD; ESBENSEN; GELADI, 1987) e a técnica *Wrapper Subset Evaluation* (KOHAVI; JOHN, 1997), com os classificadores de atributos *Linear Regression* (Wrapper-LR) (WITTEN; FRANK, 2005), *M5 Rules* (Wrapper-M5) (HOLMES; HALL; PRANK, 1999) e *Support Vector Machine* para regressão (Wrapper-SMOReg) (SMOLA; SCHÖLKOPF, 2004).

Os algoritmos de predição buscam calcular a severidade da mancha bacteriana do eucalipto a partir dos comprimentos de ondas selecionados. Foram utilizados os algoritmos de aprendizado de máquina *Linear Regression* (*Linear Regression*) (WITTEN; FRANK, 2005), *Multilayer Perceptron* (MLP) (RUMELHART, 1986), *Support Vector Machine* para regressão (SMOReg) (SMOLA; SCHÖLKOPF, 2004), *Random Forest* (R Forest) (BREIMAN, 2001) e *Random Committee* (R Committee) (CHOU *et al.*, 2009).

Cada um dos 35 modelos de predição de severidade da mancha foliar bacteriana do eucalipto causado por *Xanthomonas spp.* foram compostos pela combinação de uma das 7 técnicas de seleção de atributos apresentadas e um dos 5 algoritmos de aprendizagem de máquinas apresentados.

Cada modelo foi testado mil vezes, sendo que em cada rodada eram selecionados aleatoriamente 70% das amostras do conjunto de dados para o processo de aprendizado, denominado conjunto de aprendizado e os 30% restantes eram utilizadas para validação do modelo, denominado conjunto de teste, utilizando a metodologia de *bootstrap* para seleção aleatória das amostras (JAMES *et al.*, 2013).

Os modelos foram avaliados segundo o coeficiente de correlação de Pearson entre os valores de severidade preditas pelo modelo e o valor de severidade calculado para cada amostra do conjunto de teste. Todo esse processo foi feito utilizando o software *Waikato Environment for Knowledge Analysis* (WEKA) desenvolvido pela Universidade de Waikato em Hamilton, na Nova Zelândia (WITTEN; FRANK, 2005).

Foram ajustados modelos de regressão linear múltipla entre a severidade (no. pústulas/cm²) e os comprimentos de onda (nm), com a opção de redução dos atributos (comprimentos de onda) pelo método *stepwise* com o critério de seleção de entrada no modelo $p < 0,05$. Os dois modelos ajustados diferem com relação à variável resposta:

- a) Considerando a variável dependente como a severidade:

$$y_i = \sum_{i=1}^n \beta_i \lambda_i + e_i \quad (1)$$

- b) Considerando a variável dependente como logaritmo natural da severidade adicionado de uma unidade:

$$(\ln(y+1))_i = \sum_{i=1}^n \beta_i \lambda_i + e_i \quad (2)$$

em que,

y_i é a severidade (no. pústulas/cm²); λ_i são os comprimentos de onda (nm) selecionados pelo método *stepwise*; β_i são os coeficientes estimados de regressão linear múltipla; e_i é o componente aleatório, com $e_i \sim N(0, \sigma_e^2)$.

Para a seleção do melhor modelo, foram utilizados os seguintes diagnósticos: valor-p do teste F do quadro ANAVA; coeficiente de determinação ajustado p/ graus de liberdade (%); critério de Akaike corrigido (AIC_c); resíduos padronizados de Pearson e normal Q-Q plot.

2.3 Resultados e Discussão

2.3.1 Modelagem computacional com aprendizado de máquinas

As técnicas de seleção de atributos utilizadas destacaram comprimentos de ondas diferentes para predição de severidade de *Xanthomonas spp.* em mudas do clone híbrido de *Eucalyptus grandis* x *Eucalyptus urophylla*, como é mostrado na Tabela 5.

Tabela 5 - Quantidade de atributos e comprimentos de onda (nm) selecionados para cada técnica de seleção de atributos testadas com dados da reflectância foliar espectral de folhas de eucaliptos submetidos a infecção com *Xanthomonas spp.* para predição da severidade da doença.

Técnica de seleção de atributos	Quantidade de atributos	Comprimentos de onda (nm)
ClassifierSE-M5Rules	8	1166, 1380, 1403, 1408, 1569, 1598, 1650 e 1656
ClassifierSE-LR	4	1087, 1443, 1667 e 1685
ClassifierSE-SMOReg	7	977, 988, 993, 1289, 1301, 1632 e 1656
PCA	2	V1 e V2 (Autovalores)
WrapperSE-M5Rules	2	1408 e 1425
WrapperSE-LR	7	1087, 1099, 1121, 1205, 1295, 1465 e 1667
WrapperSE-SMOReg	16	971, 977, 999, 1026, 1032, 1082, 1138, 1149, 1278, 1289, 1295, 1301, 1306, 1317, 1656 e 1661

A predição da severidade da folha com base em sua assinatura espectral foi feita utilizando diversos algoritmos de predição, sendo que os algoritmos *Linear Regression* obteve o maior valor de coeficiente de correlação quatro vezes, enquanto o algoritmo SMOReg obteve três vezes, como mostrado na Tabela 6.

Tabela 6 - Média (desvio padrão entre parênteses) do coeficiente de correlação de Pearson para curvas de reflectância espectral de folhas de clone híbrido de *Eucalyptus grandis* x *E. urophylla*, submetidas à inoculação de *Xanthomonas spp.*, para cada algoritmo de predição (coluna) utilizando cada uma das técnicas de seleção de atributos (linha) para a predição da severidade da doença

	Linear Regression	MLP	SMOReg	Random Forest	Random Committee
Classifier-LR	0,27 (0,10)	0,23 (0,09)	0,26 (0,11)	0,11 (0,13)	0,09 (0,13)
Classifier-M5	0,22 (0,09)	0,22 (0,09)	0,25 (0,10)	0,22 (0,14)	0,18 (0,14)
Classifier-SMOReg	0,25 (0,10)	0,26 (0,09)	0,31 (0,09)	0,06 (0,12)	0,08 (0,13)
PCA	0,23 (0,10)	0,19 (0,08)	0,22 (0,11)	0,09 (0,12)	0,06 (0,13)
Wrapper-LR	0,27 (0,10)	0,23 (0,09)	0,25 (0,10)	0,11 (0,13)	0,09 (0,12)
Wrapper-M5	0,19 (0,09)	0,14 (0,09)	0,18 (0,09)	0,05 (0,10)	0,02 (0,11)
Wrapper-SMOReg	0,26 (0,10)	0,26 (0,09)	0,32 (0,09)	0,05 (0,11)	0,06 (0,13)

Os dois maiores valores de coeficiente de correlação de Pearson obtidos no presente trabalho foram alcançados pelo algoritmo SMOReg, sendo que o maior valor foi obtido conjuntamente com a técnica Wrapper-SMOReg resultando em 0,32 e, o segundo, associado com a seleção por Classifier-SMOReg, com 0,31. Demonstrando que esses dois modelos oriundo das combinações entre técnicas de seleção e algoritmos de predição são capazes de obter valores melhores do que os outros modelos testados.

Os demais modelos testados obtiveram coeficientes de correlação abaixo de 0,30; sendo que esses valores foram considerados como uma correlação desprezível, enquanto os dois maiores valores, foram considerados correlações fracas por apresentar valores entre 0,30 e 0,50; no presente estudo, esses valores estão mais próximos do limite inferior estabelecido (MUKAKA, 2012).

Dessa forma, nenhum dos modelos testados pode ser considerado adequado para a predição da severidade da MFB do eucalipto causada por *Xanthomonas spp.* a partir da assinatura foliar espectral.

A severidade para MFB do eucalipto causada por *Xanthomonas spp.* é mais acentuada com períodos maiores de infecção aos testados neste trabalho. Neves *et al.* (2014) avaliaram a severidade de *Xanthomonas Axonopodis* em clone híbrido de *Eucalyptus grandis x E. urophylla* demonstrando que 20 dias após a infecção, a severidade das folhas oscilou entre 3% e 17% de área foliar lesionada, conforme os estágios fenológicos das folhas. Ao passo que Silva *et al.* (2018) avaliou o efeito de nutrientes na severidade da mesma doença em clones de *Eucalyptus sp.*, no entanto, a avaliação foi feita aos trinta dias após a inoculação, e obteve valores entre 1% e 54% de área foliar lesionada.

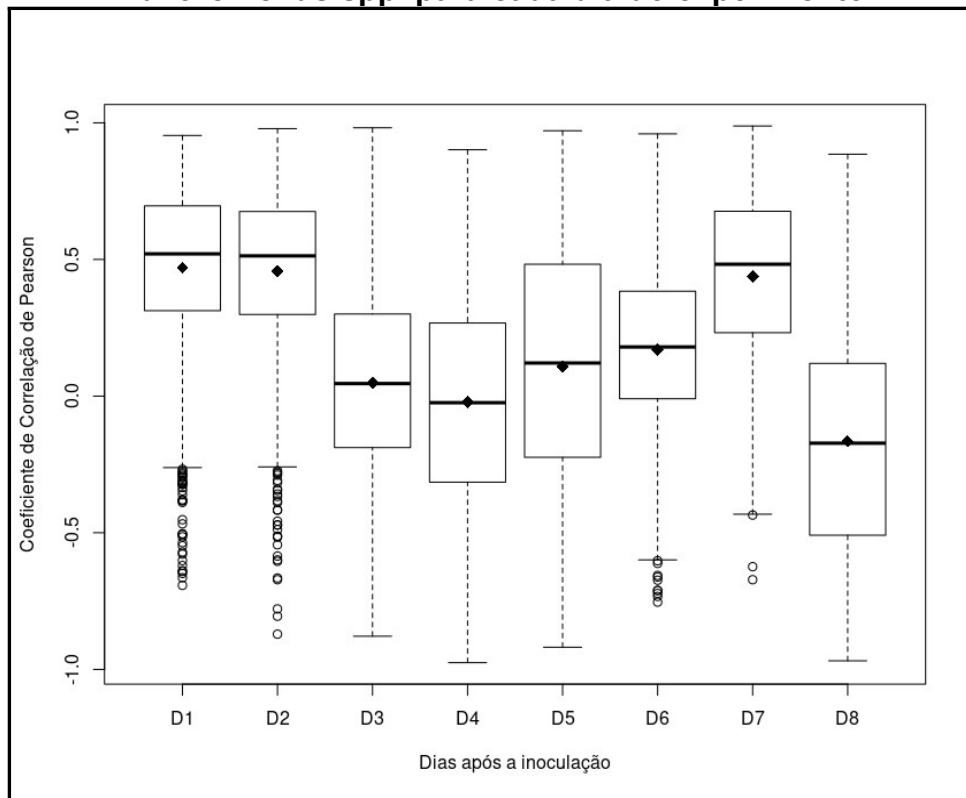
Os resultados de severidade mensurados no presente trabalho, foram obtidos até oito dias após a inoculação, o que demonstra que este período não foi suficiente para atingir valores relevantes de severidade da doença nas folhas. Caso o presente trabalho fosse estendido até o vigésimo dia, ou posterior, possivelmente a avaliação da severidade a partir de algum dos modelos propostos obteria resultados superiores aos encontrados.

A severidade da MFB causada por *Xanthomonas spp.* em tomate (*Solanum lycopersicon*) foi estimada usando reflectância, com avaliação feita aos 28 dias após a inoculação, os resultados mostram fortes correlações entre os valores preditos e os analisados pelos especialistas, mostrando este método interessante para a avaliação da doença (BORGES *et al.*, 2016).

Marin *et al.* (2018) correlacionam índices de vegetação obtidos a partir de sensores multiespectrais de satélite com a severidade de mancha bacteriana em café (*Coffea arabica*) no estado de Minas Gerais, obtendo valores de correlação entre 0,52 e 0,76, mostrando grande potencial da metodologia para monitorar plantações de café afetadas pela doença.

A simulação para predição da severidade da MFB do eucalipto causada por *Xanthomonas spp.* a partir da assinatura espectral foliar também foi analisada em cada um dos dias separadamente, os resultados são apresentados na Figura 5.

Figura 5 - Boxplot dos coeficientes de correlação de Pearson entre a severidade real e a severidade predita pelo algoritmo *Support Vector Machine* para folhas de clone híbrido de *Eucalyptus grandis* x *E. urophylla*, inoculadas com *Xanthomonas* spp. para cada dia do experimento



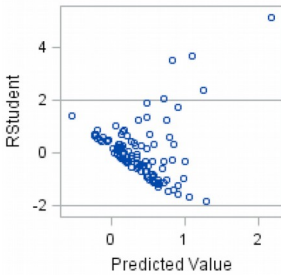
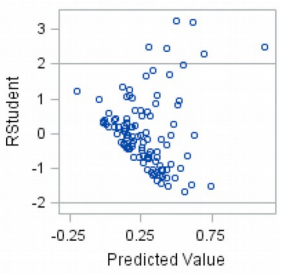
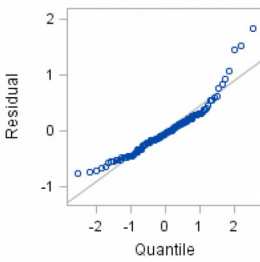
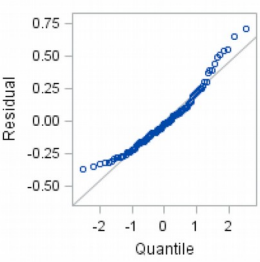
Para qualquer um dos dias analisados, o coeficiente de correlação médio está sempre abaixo de 0,50, valor considerado como correlação fraca. Embora 1 e 2 dias após a inoculação (D1 e D2) apresentem valores superiores aos demais, ainda são valores baixos para considerar a correlação entre os valores preditos e os valores reais.

Após os dois primeiros dias, há uma oscilação entre os valores de -0,02 a 0,17; reestabelecendo o valor de 0,44 no sétimo dia (D7). Dessa forma, não é possível considerar algum dos dias analisados como mais favorável para a mensura da severidade com base na assinatura espectral foliar, uma vez que estes valores não podem ser considerados como uma correlação forte entre o modelo testado e os valores reais.

2.3.2 Modelagem estatística com modelos de regressão linear múltipla

Na Tabela 7 estão apresentados os diagnósticos de ajustes dos modelos de regressão linear múltipla para a severidade (número de pústulas/cm²) em função dos comprimentos de ondas (nm).

Tabela 7 - Diagnósticos de ajustes dos modelos de regressão linear múltipla para a severidade (no. pústulas/cm²) em função dos comprimentos de ondas (nm)

	Modelos de regressão $y = \text{severidade (no. pústulas/cm}^2\text{)}$ R_i 's: reflectâncias foliares (%) associadas a um dado comprimento de onda (nm) selecionado ⁽¹⁾	
Nomenclatura	Modelo A	Modelo B
Modelos de regressão linear múltipla	$y_i = \sum_{i=1}^n \beta_i R_i + e_i$	$(\ln(y+1))_i = \sum_{i=1}^n \beta_i R_i + e_i$
Valor-p (Teste F ANOVA) ⁽²⁾	$p < 0,0001$	$p < 0,0001$
Coefficiente de determinação ajustado p/ graus de liberdade (%) R_{aj}^2 ⁽³⁾	56,21	66,42
AIC_c ⁽⁴⁾	-11,88	-163,83
Resíduos padronizados de Pearson ⁽⁵⁾		
Q-Q Normal Plot		

⁽¹⁾ comprimentos de onda (nm) selecionados: 1082, 1121, 1284, 1301, 1317, 1329, 1386, 1431, 1465 e 1667;

⁽²⁾ hipóteses: $H_0: \beta_i = 0, i \in \{1, \dots, n\}$ vs $H_1: \beta_i \neq \beta_j; i \neq j; i, j \in \{1, \dots, n\}$;

⁽³⁾ $R_{aj}^2 = R^2 - \frac{1-R^2}{n-2}$; $R^2 = \text{coeficiente de determinação} = \frac{SQReg}{SQTotal}$;

$$^{(4)} AIC_c = AIC + \frac{2p^2 + 2p}{n - p - 1}; AIC = 2p - 2 \ln(\hat{L}); p = \text{número de parâmetros do modelo}; n =$$

número de observações; \hat{L} = valor máximo da função de verossimilhança;

⁽⁵⁾ resíduos obtidos dos valores observados e esperados padronizados versus os valores preditos;

⁽⁶⁾ normal Q-Q plot: gráfico entre os quantis teóricos e amostrais.

Os modelos de regressão linear múltipla mostraram um bom ajuste como pode ser observado pelos diagnósticos apresentados na Tabela 9: Os *Q-Q normal plots*, os resíduos padronizados de Pearson estão dispostos em sua maioria no intervalo de -2 a 2.

Os valores-p dos testes F mostraram que há significância nos modelos testados e o coeficiente de determinação ajustado apresentaram valores superiores a 55%, sendo que o Modelo B apresentou o valor de 66%, resultando em um bom poder preditivo do modelo. Além desses critérios, optou-se pela seleção do Modelo B pois este apresentou menor valor de AIC_c (AKAIKE, 1974) quando comparado ao Modelo A.

Embora ambos os modelos de regressão analisados apresentem valores de coeficientes de determinação ajustados próximos a 60%, indicando melhor desempenho na predição da severidade, quando comparado aos modelos baseados em aprendizado de máquinas. Tais resultados não invalidam os modelos computacionais, somente demonstram que esses devem ser aprimorados para trabalhos posteriores.

Cabe ressaltar que a partir dos modelos propostos, se torna possível estimar a severidade da mancha foliar bacteriana do eucalipto causado por *Xanthomonas ssp.* utilizando a reflectância foliar espectral, condicionado a presença sintomática na folha. Assim como foi mostrado em trabalhos com *Xanthomonas perforans* em tomate (*Lycopersicon esculentum*) (JONES; JONES; LEE, 2010) e *Cnaphalocrocis medinalis* em arroz (*Oryza sativa L.*) (HUANG *et al.*, 2012).

Dessa forma, o modelo B (Tabela 7) ajustado foi descrito com a seguinte equação:

$$\ln(y+1) \text{ estimado} = 2,09(R1082) - 2,94(R1121) + 1,95(R1284) - 1,23(R1301) - 2,98(R1317) - 3,04(R1329) + 0,32(R1386) + 0,38(R1431) - 0,41(R1465) - 0,11(R1667)$$

em que, y é a severidade (no. pústulas/cm²); \ln é o logaritmo natural; R_i 's são os comprimentos de onda (nm) selecionados pelo método *stepwise*.

2.3.3 Considerações Adicionais

A severidade da doença alvo do presente trabalho foi observada até o oitavo dia após a inoculação, este período pode ser considerado curto para o desenvolvimento dos sintomas da MFB do eucalipto de forma mais severa, explicando assim os valores não elevados tanto para o coeficiente de correlação de Pearson, utilizado nos modelos de aprendizado de máquinas, quanto para os valores do coeficiente de determinação ajustado para os modelos de regressão linear múltipla.

Neves *et al.* (2014) propõem que os maiores incrementos de severidades são observados entre dez e quinze dias após a inoculação, de acordo com os experimentos realizados com *Xanthomonas axonopodis* em clone híbrido de *Eucalyptus grandis* x *E. urophylla*. No entanto, este aumento da severidade depende da dinâmica entre o patógeno, o hospedeiro e o ambiente (GONÇALVES *et al.*, 2008).

WEKA é um software destinado para trabalhos envolvendo técnicas de mineração de dados e aprendizado de máquinas, sendo largamente utilizados em trabalho acadêmicos de diversas áreas (IJAQ *et al.*, 2019; LIU *et al.*, 2019; PIMENTEL *et al.*, 2019). No entanto, o software apresentou algumas ressalvas, principalmente se tratando de pré-requisitos relacionados a modelos estatísticos. Um exemplo seria o cálculo do coeficiente de correlação de Pearson que foi empregado sem previamente ser realizado o teste de normalidade dos dados, assim como, a não apresentação do valor-p referente ao teste de hipóteses da nulidade da correlação.

Dessa forma, nenhum dos modelos baseados em aprendizado de máquinas testados obtiveram resultados eficientes em prever a severidade da MFB do eucalipto. Os modelos baseados em regressão linear múltipla obtiveram resultados mais relevantes e com maior poder de predição da severidade da doença, sendo eles mais indicados para essa finalidade.

Não obstante, ressalta-se que a construção de modelos de inteligência artificial deve ser feita com parcimônia, buscando sempre a integração entre as áreas da matemática, computação e estatística, em uma equipe multidisciplinar.

2.4 Conclusão

Os modelos de aprendizado de máquinas testados obtiveram valores baixos de coeficiente de correlação de Pearson, pressupondo a normalidade dos dados, entre os valores predito e os valores reais da severidade da MFB do eucalipto, não se mostrando eficientes para essa finalidade.

O modelo de regressão linear múltipla considerando a variável dependente como logaritmo natural da severidade adicionado de uma unidade obteve os melhores resultados para estimar a severidade da doença, sendo este modelo indicado pelo presente estudo.

2.5 Referências

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, dez. 1974.

ALFENAS, A. C. *et al.* **Clonagem e Doenças do Eucalipto**. 2 ed. ed. Viçosa, MG: Editora UFV, 2009.

AUER, C. G.; SANTOS, Á. F. DOS; NETO, J. R. **Mancha foliar bacteriana em plantios de eucalipto na região Sul do Brasil**. Colombo, PR: Embrapa, 2011.

BOCK, C. H. *et al.* Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. **Critical Reviews in Plant Sciences**, v. 29, n. 2, p. 59–107, 10 mar. 2010.

BORGES, D. L. *et al.* Detecting and grading severity of bacterial spot caused by *Xanthomonas* spp. in tomato (*Solanum lycopersicon*) fields using visible spectrum images. **Computers and Electronics in Agriculture**, v. 125, p. 149–159, 1 jul. 2016.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 1 out. 2001.

CHOU, T. *et al.* **Ensemble of machine learning algorithms for intrusion detection**. 2009 IEEE International Conference on Systems, Man and Cybernetics. **Anais...** In: 2009 IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. out. 2009

COLWELL, J. E. Vegetation canopy reflectance. **Remote Sensing of Environment**, v. 3, n. 3, p. 175–183, 1 jan. 1974.

CUNHA, C. A. H. DA. **Relação entre comportamento espectral, índice de área foliar e produção de matéria seca em capim Tanzânia submetido a diferentes níveis de irrigação e doses de nitrogênio**. text—[s.l.] Universidade de São Paulo, 5 out. 2004.

FERRAZ, H. G. M. *et al.* Xanthomonas axonopodis pv. eucalyptorum pv. nov. Causing Bacterial Leaf Blight on Eucalypt in Brazil. **The Plant Pathology Journal**, v. 34, n. 4, p. 269–285, ago. 2018.

FRANKE, J.; MENZ, G. Multi-temporal wheat disease detection by multi-spectral remote sensing. **Precision Agriculture**, v. 8, n. 3, p. 161–172, 1 jun. 2007.

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, p. 2627–2636, 1 ago. 1998.

GONÇALVES, R. C. *et al.* Etiology of bacterial leaf blight of eucalyptus in Brazil. **Tropical Plant Pathology**, v. 33, n. 3, p. 180–188, jun. 2008.

HOLMES, G.; HALL, M.; PRANK, E. **Generating Rule Sets from Model Trees**. (N. Foo, Ed.) Advanced Topics in Artificial Intelligence. **Anais...**: Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1999

HUANG, J. *et al.* Hyperspectral detection of rice damaged by rice leaf folder (Cnaphalocrocis medinalis). **Computers and Electronics in Agriculture**, v. 82, p. 100–107, 1 mar. 2012.

HUANG, W. *et al.* Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. **Precision Agriculture**, v. 8, n. 4, p. 187–197, 1 out. 2007.

IBÁ. **Relatório 2017**. [s.l.] Industria Brasileira de Arvores, 2017. Disponível em: <<https://www.iba.org/datafiles/publicacoes/pdf/iba-relatorioanual2017.pdf>>. Acesso em: 15 out. 2018.

IBA. **Sumário Executivo 2018**. [s.l.] Industria Brasileira de Arvores, 2018. Disponível em: <<https://www.iba.org/datafiles/publicacoes/relatorios/digital-sumarioexecutivo-2018.pdf>>. Acesso em: 13 fev. 2019.

IJAQ, J. *et al.* A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. **BMC Bioinformatics**, v. 20, n. 1, p. 14, 8 jan. 2019.

JAMES, G. *et al.* (EDS.). **An introduction to statistical learning: with applications in R**. New York: Springer, 2013.

JONES, C. D.; JONES, J. B.; LEE, W. S. Diagnosis of bacterial spot of tomato using spectral signatures. **Computers and Electronics in Agriculture**, v. 74, n. 2, p. 329–335, 1 nov. 2010.

KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. **Artificial Intelligence, Relevance**. v. 97, n. 1, p. 273–324, 1 dez. 1997.

LANGLEY, P. **Selection of Relevant Features in Machine Learning.**: Fort Belvoir, VA: Defense Technical Information Center, 1 nov. 1994. Disponível em: <<http://www.dtic.mil/docs/citations/ADA292575>>. Acesso em: 13 fev. 2019.

LAW, B. E.; WARING, R. H. Remote Sensing of Leaf Area Index and Radiation Intercepted by Understory Vegetation. **Ecological Applications**, v. 4, n. 2, p. 272–279, 1994.

LIU, Y. *et al.* Identification of Coenzyme-Binding Proteins with Machine Learning Algorithms. **Computational Biology and Chemistry**, 28 jan. 2019.

MAFIA, R. G.; TEIXEIRA, L. P.; FERREIRA, M. A. Produção de mudas de eucalipto sob dois sistemas de irrigação e a ocorrência da bacteriose foliar (*Xanthomonas axonopodis* Vauterin). **Scientia Forestalis**, v. 45, n. 116, 1 dez. 2017.

MAHLEIN, A.-K. *et al.* Recent advances in sensing plant diseases for precision crop protection. **European Journal of Plant Pathology**, v. 133, n. 1, p. 197–209, 1 maio 2012.

MARIN, D. B. *et al.* Multispectral radiometric monitoring of bacterial blight of coffee. **Precision Agriculture**, 11 dez. 2018.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto**. 4 Ed ed. Viçosa, MG: Editora UFV, 2011.

MUKAKA, M. A guide to appropriate use of Correlation coefficient in medical research. **Malawi Medical Journal: The Journal of Medical Association of Malawi**, v. 24, n. 3, p. 69–71, set. 2012.

NEVES, D. A. *et al.* Favorable conditions for *Xanthomonas axonopodis* infection in Eucalyptus spp. **Tropical Plant Pathology**, v. 39, n. 6, p. 428–433, dez. 2014.

OCEAN OPTICS. **OceanView 1.6.7Ocean Optics**, 2018. Disponível em: <<https://oceanoptics.com/product/oceanview/>>. Acesso em: 13 fev. 2019

ODUM, E. P. **Ecologia**. 1 ed ed. Rio de Janeiro: Editora Interamericana, 1983. v. 1

PARKER, S. R.; SHAW, M. W.; ROYLE, D. J. The reliability of visual estimates of disease severity on cereal leaves. **Plant Pathology**, v. 44, n. 5, p. 856–864, 1995.

PIMENTEL, A. *et al.* Coronary Heart Disease Prognosis Using Machine-Learning Techniques on Patients With Type 2 Diabetes Mellitus. **Chronic Illness and Long-Term Care: Breakthroughs in Research and Practice**, p. 198–217, 2019.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Mateo, CA: Morgan Kaufmann, 1994.

REZENDE, S. O.; MARCACINI, R. M.; MOURA, M. F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação**, n. 7, p. 15, 2011.

RUMELHART, D. E. Learning internal representations by error propagation. In: RUMELHART, D. E.; MCCLELLAND, J. L. (Eds.). **Parallel distributed processing:**

explorations in the microstructure of cognition. Cambridge: MIT Press, 1986. v. 1p. 318–362.

SCHAAD, N. W.; JONES, J. B.; CHUN, W. (EDS.). **Laboratory guide for identification of plant pathogenic bacteria.** 3rd ed ed. St. Paul, Minn: American Phytopathological Society, 2001.

SCHNEIDER, C. A.; RASBAND, W. S.; ELICEIRI, K. W. NIH Image to ImageJ: 25 years of image analysis. **Nature Methods**, v. 9, p. 671–675, 28 jun. 2012.

SHEVADE, S. K. *et al.* Improvements to the SMO algorithm for SVM regression. **IEEE Transactions on Neural Networks**, v. 11, n. 5, p. 1188–1193, set. 2000.

SILVA, A. G. *et al.* Severity of *Xanthomonas axonopodis* leaf blight on eucalypt cuttings under different levels of nutrients. **Forest Pathology**, v. 48, n. 2, p. e12412, 2018.

SIMON, P. **Too big to ignore: the business case for big data.** Hoboken, New Jersey: John Wiley & Sons, Inc, 2013.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, n. 3, p. 199–222, ago. 2004.

VALE, F. X. R.; FERNANDES FILHO, E. I.; LIBERATO, J. R. **QUANT: a software for plant disease severity assessment.** : 8. In: INTERNATIONAL CONGRESS OF PLANT PATHOLOGY. Abstract Christchurch, New Zealand: Australasian Plant Pathology Society, 2003

VALERI, S. V.; CORRADINI, L. Fertilização em viveiros para produção de mudas de eucaliptos e pinus. In: GONÇALVES, J. L. DE M.; BENEDETTI, V. (Eds.). . **Nutrição e fertilização florestal.** 1. ed. Piracicaba: UPEF, 2000. p. 427.

WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques.** 2nd ed ed. Amsterdam ; Boston, MA: Morgan Kaufman, 2005.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists. v. 2, n. 1, p. 37–52, 1 ago. 1987.

CONSIDERAÇÕES FINAIS

As técnicas de aprendizado de máquinas são consideradas viáveis para a discriminação entre plantas com sintomas de MFB do eucalipto, a partir da assinatura espectral das plantas. O modelo baseado em *Multilayer Perceptron* associado à técnica de seleção de atributos *Wrapper* com classificador J48, obteve a maior porcentagem de classificações corretas para o problema analisado, demonstrando que os comprimentos de ondas mais importantes para esta finalidade são o 1374 nm e 1685 nm.

O modelo de classificação acima mencionado possibilitou boa identificação das plantas doentes, de forma a discriminá-las das plantas sadias 48 horas após a inoculação, com alta porcentagem de classificações corretas.

Os modelos testados para classificação de dias após a inoculação apresentaram resultados inferiores aos considerados recomendados para uma boa classificação, dessa forma, se faz necessário maior aprimoramento dos modelos para testes futuros.

Para predição da severidade da MFB do eucalipto, os modelos de aprendizado de máquinas testados obtiveram valores baixos de coeficiente de correlação de *Pearson* entre os valores predito e os valores reais da severidade da mancha foliar bacteriana do eucalipto, não se mostrando eficientes para essa finalidade.

O modelo de regressão linear múltipla considerando a variável dependente como logaritmo natural da severidade adicionado de uma unidade obteve os melhores resultados para estimar a severidade da doença, sendo este modelo indicado pelo presente estudo.

Cabe ressaltar que para desenvolvimento de modelos tanto para classificação quando para predição de valores, se faz importante a participação de uma equipe multidisciplinar formada pelas áreas da matemática, computação, estatística e fitopatologia, mostrando entrosamento e cuidado com os modelos gerados e suas interpretações.

Nenhuma das técnicas abordadas no presente estudo tem como intenção substituir o profissional especializado, mas sim, servir como uma ferramenta no

auxílio da detecção de doenças e promover maior celeridade nos processos de tomada de decisão.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, dez. 1974.
- ALFENAS, A. C. *et al.* **Clonagem e Doenças do Eucalipto**. 2 ed. ed. Viçosa, MG: Editora UFV, 2009.
- ALPAYDIN, E. **Introduction to Machine Learning: Selected Papers of Lionel W. McKenzie**. Cumberland: MIT Press, The, 2016.
- AL-SADDIK, H. *et al.* Using Image Texture and Spectral Reflectance Analysis to Detect Yellowness and Esca in Grapevines at Leaf-Level. **Remote Sensing**, v. 10, n. 4, p. 618, abr. 2018.
- APPAVU, S. *et al.* **Bayes Theorem and Information Gain Based Feature Selection for Maximizing the Performance of Classifiers**. (N. Meghanathan, B. K. Kaushik, D. Nagamalai, Eds.) *Advances in Computer Science and Information Technology. Anais...: Communications in Computer and Information Science*. Springer Berlin Heidelberg, 2011
- AUER, C. G.; SANTOS, Á. F. DOS; NETO, J. R. **Mancha foliar bacteriana em plantios de eucalipto na região Sul do Brasil**. Colombo, PR: Embrapa, 2011.
- BEDENDO, I. P. Manchas Foliares. In: AMORIM, L.; REZENDE, J. A. M.; BERGAMO FILHO, A. (Eds.). **Manual de Fitopatologia**. 4. ed. Piracicaba: Editora Agronomica Ceres, 2011. v. 1p. 704.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 1 out. 2001.
- CHOU, T. *et al.* **Ensemble of machine learning algorithms for intrusion detection**. 2009 IEEE International Conference on Systems, Man and Cybernetics. **Anais...** In: 2009 IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. out. 2009
- CUNHA, C. A. H. DA. **Relação entre comportamento espectral, índice de área foliar e produção de matéria seca em capim Tanzânia submetido a diferentes níveis de irrigação e doses de nitrogênio**. text—[s.l.] Universidade de São Paulo, 5 out. 2004.
- DAMASCENO, V. F. F. *et al.* Comparison of two methods to elaborate and validate a diagrammatic scale for quantifying the severity of *Cylindrocladium* leaf spot on Eucalyptus. **Summa Phytopathologica**, v. 40, n. 3, p. 248–255, set. 2014.
- FOELKEL, C. E. B. Eucalipto no Brasil, história de pioneirismo. **Revista Visão Agrícola**, v. 1, n. 4, p. 4, 2005.
- FRANKE, J.; MENZ, G. Multi-temporal wheat disease detection by multi-spectral remote sensing. **Precision Agriculture**, v. 8, n. 3, p. 161–172, 1 jun. 2007.

- 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, p. 2627–2636, 1 ago. 1998.
- GONÇALVES, R. C. *et al.* Etiology of bacterial leaf blight of eucalyptus in Brazil. **Tropical Plant Pathology**, v. 33, n. 3, p. 180–188, jun. 2008.
- HEARST, M. A. *et al.* Support vector machines. **IEEE Intelligent Systems and their Applications**, v. 13, n. 4, p. 18–28, jul. 1998.
- HOLTE, R. C. Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. **Machine Learning**, v. 11, n. 1, p. 63–90, 1 abr. 1993.
- HUANG, W. *et al.* Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. **Precision Agriculture**, v. 8, n. 4, p. 187–197, 1 out. 2007.
- IBA. **Relatório 2017**. [s.l.] Industria Brasileira de Arvores, 2017. Disponível em: <<https://www.iba.org/datafiles/publicacoes/pdf/iba-relatorioanual2017.pdf>>. Acesso em: 15 out. 2018.
- IBA. **Sumário Executivo 2018**. [s.l.] Industria Brasileira de Arvores, 2018. Disponível em: <<https://www.iba.org/datafiles/publicacoes/relatorios/digital-sumarioexecutivo-2018.pdf>>. Acesso em: 13 fev. 2019.
- JONES, C. D.; JONES, J. B.; LEE, W. S. Diagnosis of bacterial spot of tomato using spectral signatures. **Computers and Electronics in Agriculture**, v. 74, n. 2, p. 329–335, 1 nov. 2010.
- KAREGOWDA, A. G.; MANJUNATH, A. S.; JAYARAM, M. A. Comparative study of attribute selection using gain ratio and correlation based feature selection. **International Journal of Information Technology and Knowledge Management**, v. 2, n. 2, p. 7, 2010.
- KAUNDAL, R.; KAPOOR, A. S.; RAGHAVA, G. P. Machine learning techniques in disease forecasting: a case study on rice blast prediction. **BMC Bioinformatics**, v. 7, n. 1, p. 485, 3 nov. 2006.
- KIRA, K.; RENDELL, L. A. A Practical Approach to Feature Selection. In: SLEEMAN, D.; EDWARDS, P. (Eds.). **Machine Learning Proceedings**. San Francisco (CA): Morgan Kaufmann, 1992. p. 249–256.
- KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. **Artificial Intelligence, Relevance**. v. 97, n. 1, p. 273–324, 1 dez. 1997.
- KRATZ, D.; WENDLING, I. PRODUÇÃO DE MUDAS DE *Eucalyptus dunnii* EM SUBSTRATOS RENOVÁVEIS. **FLORESTA**, v. 43, n. 1, p. 125–136, 24 abr. 2013.
- LIRA, M. M. S. *et al.* **Combining Multiple Artificial Neural Networks Using Random Committee to Decide upon Electrical Disturbance Classification**. 2007 International Joint Conference on Neural Networks. **Anais...** In: 2007 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS. ago. 2007

MAFIA, R. G.; TEIXEIRA, L. P.; FERREIRA, M. A. Produção de mudas de eucalipto sob dois sistemas de irrigação e a ocorrência da bacteriose foliar (*Xanthomonas axonopodis* Vauterin). **Scientia Forestalis**, v. 45, n. 116, 1 dez. 2017.

MAHLEIN, A.-K. *et al.* Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. **Plant Methods**, v. 8, n. 1, p. 3, 24 jan. 2012a.

MAHLEIN, A.-K. *et al.* Recent advances in sensing plant diseases for precision crop protection. **European Journal of Plant Pathology**, v. 133, n. 1, p. 197–209, 1 maio 2012b.

MCQUEEN, R. J. *et al.* Applying machine learning to agricultural data. **Computers and Electronics in Agriculture**, v. 12, n. 4, p. 275–293, 1 jun. 1995.

MEYER-BAESE, A.; SCHMID, V. Chapter 6 - Statistical and Syntactic Pattern Recognition. In: MEYER-BAESE, A.; SCHMID, V. (Eds.). . **Pattern Recognition and Signal Analysis in Medical Imaging**. 2. ed. Oxford: Academic Press, 2014. p. 151–196.

MOHANTY, S. P.; HUGHES, D. P.; SALATHÉ, M. Using Deep Learning for Image-Based Plant Disease Detection. **Frontiers in Plant Science**, v. 7, 2016.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto**. 4 Ed ed. Viçosa, MG: Editora UFV, 2011.

MUDA, Z. *et al.* **Intrusion detection based on k-means clustering and OneR classification**. 2011 7th International Conference on Information Assurance and Security (IAS). **Anais...** In: 2011 7TH INTERNATIONAL CONFERENCE ON INFORMATION ASSURANCE AND SECURITY (IAS). dez. 2011

ODUM, E. P. **Ecologia**. 1 ed ed. Rio de Janeiro: Editora Interamericana, 1983. v. 1

PLANTSYSTEMATICS.ORG. **PlantSystematics.org**. Disponível em: <<http://plantsystematics.org/>>. Acesso em: 9 jul. 2019.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Mateo, CA: Morgan Kaufmann, 1994.

REZENDE, S. O.; MARCACINI, R. M.; MOURA, M. F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação**, n. 7, p. 15, 2011.

RUMELHART, D. E. Learning internal representations by error propagation. In: RUMELHART, D. E.; MCCLELLAND, J. L. (Eds.). . **Parallel distributed processing: explorations in the microstructure of cognition**. Cambridge: MIT Press, 1986. v. 1p. 318–362.

SIMON, P. **Too big to ignore: the business case for big data**. Hoboken, New Jersey: John Wiley & Sons, Inc, 2013.

SINGH, A. *et al.* Machine Learning for High-Throughput Stress Phenotyping in Plants. **Trends in Plant Science**, v. 21, n. 2, p. 110–124, 1 fev. 2016.

VALENZUELA, P. L.; LÓPEZ-GARCÍA, J. Detección de árboles dañados por plaga en bosques de *Abies religiosa* en la Reserva de la Biosfera Mariposa Monarca, mediante fotografías aéreas infrarroja. **Investigaciones Geográficas**, v. 0, 31 mar. 2017.

WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2nd ed ed. Amsterdam; Boston, MA: Morgan Kaufman, 2005.

WITTEN, I. H.; FRANK, E. **Data Mining: Pratical Machine Learning Tools and Techniques**. 2. ed. [s.l: s.n.].

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists. v. 2, n. 1, p. 37–52, 1 ago. 1987.