

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*

COMPUTATIONAL AND STATISTICAL MODELS FOR PREDICTION OF SEVERITY OF LEAF SPOT CAUSED BY *XANTHOMONAS SPP.* IN *EUCALYPTUS GRANDIS X E. UROPHYLLA* SEEDING

João Ricardo Favan¹

Renata Bruna dos Santos Coscolin²

André Stefanini Jim³

Rafaela Lanças Gomes⁴

José Raimundo de Souza Passos⁵

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 está sujeita a diversas fontes de erros, por isso, métodos alternativos utilizando imagens digitais e reflectância foliar espectral vêm surgindo com promessas de alta acurácia e menor termo de avaliação. O presente trabalho teve por objetivo avaliar diversos modelos computacionais baseados 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 submetidas à 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 a 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.

¹ Tecnólogo em Informática, Doutor em Ciências Florestais, Professor na Faculdade de Tecnologia Shunji Nishimura de Pompéia, Rua Shunji Nishimura, 605, Distrito Industrial, CEP: 17580-000. Pompeia, São Paulo, rickfavan@gmail.com

² Engenheira Agrônoma. Doutora em Agronomia. Professora na Faculdade de Tecnologia Shunji Nishimura de Pompéia,

³ Engenheiro Agrônomo. Doutor em Ciência Florestal. Departamento de Bioestatística. Instituto de Biociências de Botucatu (UNESP).

⁴ Bióloga. Mestranda em Botânica. Departamento de Botânica. Instituto de Biociências de Botucatu (UNESP).

⁵ Engenheiro Florestal. Doutor em Agronomia. Professor do Departamento de Bioestatística. Instituto de Biociências de Botucatu (UNESP).

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.

Key words: Machine Learning. multiple linear regression. Support vector machines

¹ Tecnólogo em Informática, Doutor em Ciências Florestais, Professor na Faculdade de Tecnologia Shunji Nishimura de Pompéia, Rua Shunji Nishimura, 605, Distrito Industrial, CEP: 17580-000. Pompeia, São Paulo, rickfavan@gmail.com

² Engenheira Agrônoma. Doutora em Agronomia. Professora na Faculdade de Tecnologia Shunji Nishimura de Pompéia,

³ Engenheiro Agrônomo. Doutor em Ciência Florestal. Departamento de Bioestatística. Instituto de Biociências de Botucatu (UNESP).

⁴ Bióloga. Mestranda em Botânica. Departamento de Botânica. Instituto de Biociências de Botucatu (UNESP).

⁵ Engenheiro Florestal. Doutor em Agronomia. Professor do Departamento de Bioestatística. Instituto de Biociências de Botucatu (UNESP).

1. INTRODUÇÃO

O setor de florestas plantadas brasileiro tem grande importância para a economia nacional. No ano de 2017, esse setor obteve uma receita bruta de 73,8 bilhões de Reais e 1,1% de participação no PIB nacional. A área de florestas plantadas totaliza 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 a principal cultura florestal no Brasil (IBA, 2017).

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 trata 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 à 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, vem 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).

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 (COLWELL, 1974). 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 vêm 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 (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 podem 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 têm 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 destacados 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);

O objetivo do presente trabalho foi 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. 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). Essas mudas foram mantidas com irrigação e nutrição padronizada (VALERI; CORRADINI, 2000) e 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 (OD600nm=0,1) e utilizada na pulverização das mudas do presente trabalho.

As bandejas com as mudas foram acondicionadas em Câmara de Inoculação 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, foram avaliadas durante 8 dias consecutivos, selecionadas 6 mudas, 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.

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).

Foram obtidas imagens digitais 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 (cm^2). 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 espectroradiômetro e o valor de severidade (número de pústulas/ cm^2) 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) (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. O modelo A considera a variável dependente como a severidade; o modelo B considera a variável dependente como logaritmo natural da severidade adicionado de uma unidade, como mostra a Figura 1.

Figura 1. Modelos de regressão linear utilizados para estimar a severidade da mancha foliar bacteriana do eucalipto, com a variável dependente a severidade (modelo A) e a variável dependente o logaritmo natural da severidade adicionado de uma unidade (modelo B)

MODELO A	MODELO B
$y_i = \sum_{i=1}^n \beta_i \lambda_i + e_i$	$(\ln(y+1))_i = \sum_{i=1}^n \beta_i \lambda_i + e_i$

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; resíduos padronizados de Pearson e normal Q-Q plot.

3. RESULTADOS E DISCUSSÃO

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 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 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

Fonte: Próprio Autor, 2019.

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 2.

Tabela 2: 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)

Fonte: Próprio Autor, 2019.

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 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 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).

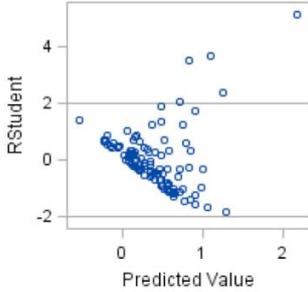
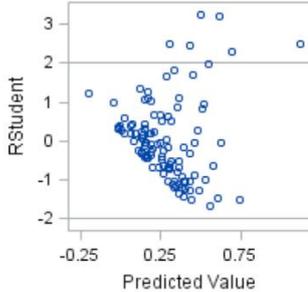
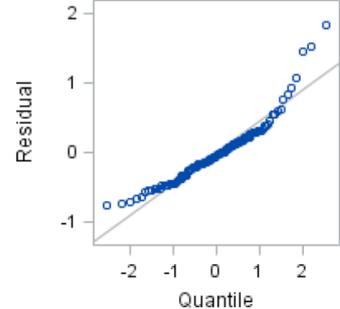
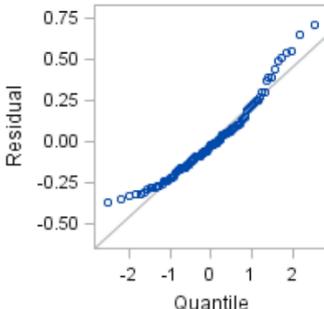
Marin et al. (2018) correlacionam índices de vegetação obtidos a partir de sensores multiespectrais de satélite com a severidade de mancha bacteriana em café (*Coffea arabica*) no

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. Para qualquer dia, o coeficiente de correlação médio está sempre abaixo de 0,50, valor considerado como correlação fraca, embora 1 e 2 dias após a inoculação apresentaram valores superiores aos demais, ainda são valores baixos para considerar como relevantes.

Na Tabela 2, 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 2: 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 (1)	
Nomenclatura	Modelo A	Modelo B
Valor-p (Teste F ANOVA)	p < 0,0001	p < 0,0001
Coefficiente de determinação ajustado p/ graus de liberdade (%) (2)	56,21	66,42
AIC_C(3)	-11,88	-163,83
Resíduos padronizados de Pearson (4)		
Q-Q Normal Plot (5)		

Fonte: Próprio Autor, 2019.

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

(2) coeficiente de determinação (R^2)

(3) critério de Akaike corrigido

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

(5) 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 2: 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 AICC (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, torna-se 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 equação apresentada na Figura 2.

Figura 2. Equação do modelo B ajustado 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.

$\ln(y+1) = 2.09 R_{1082} - 2.94 R_{1121} + 1.95 R_{1284} - 1.23 R_{1301} - 2.98 R_{1317} - 3.04 R_{1329} + 0.32 R_{1386} + 0.38 R_{1431} - 0.41 R_{1465} - 0.11 R_{1667}$
 Fonte: Próprio Autor, 2019.

A severidade da doença alvo do presente trabalho foi observada até o oitavo dia após a inoculação. Esse 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).

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 (IJAQ et al., 2019; LIU et al., 2019; PIMENTEL et al., 2019), buscando sempre a integração entre as áreas da matemática, computação e estatística, em uma equipe multidisciplinar.

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.

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.

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.

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.

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.