

APRENDIZADO DE MÁQUINA E REGRESSÃO LINEAR NA ESTIMATIVA DO VOLUME DE *Eucalyptus* NA AMAZÔNIA ORIENTAL

MACHINE LEARNING AND LINEAR REGRESSION ON *Eucalyptus*' VOLUME ESTIMATES IN EASTERN BRAZILIAN AMAZON

Ernandes Macedo da Cunha Neto¹, Marks Melo Moura¹, Emmanoella Costa Guaraná Araujo¹, Gabriel Mendes Santana¹, Ana Paula Dalla Corte¹, Carlos Roberto Sanquetta¹

¹Universidade Federal do Paraná, Curitiba, Paraná, Brasil, – netomacedo878@gmail.com, marksmoura@yahoo.com.br, manugarana@gmail.com, gsantanaflorestal@gmail.com, anapaulacorte@gmail.com & carlossanquetta@gmail.com

RESUMO

O crescimento da área florestada trouxe a necessidade de aprimorar técnicas que proporcionem estimativas mais precisas e acuradas do volume de árvores individuais. Neste contexto, técnicas de aprendizagem de máquina foram aplicadas para este fim, contudo, ainda não foi testado o uso combinado dessas técnicas (*Ensemble*). Assim, a hipótese deste estudo é que o *Ensemble* proporciona estimativas mais precisas e acuradas do volume que os métodos convencionais de estimativa e aprendizado de máquina isoladamente. Com isso, objetivou-se avaliar a estimativa do volume individual do *Eucalyptus* spp. utilizando técnicas isoladas e conjuntas de aprendizado de máquina. O estudo foi realizado em duas fazendas em Paragominas, Pará. Foram cubadas 90 árvores, distribuídas em 4 clones, com 6 anos idade e 1.111 indivíduos.ha⁻¹. Utilizaram-se 70% das árvores para o ajuste e 30% para a validação. Aplicaram-se quatro métodos distintos de estimação: regressão linear (Modelo de Schumacher-Hall), máquinas de vetores de suporte (SVM), redes neurais artificiais (RNA) e *Ensemble* (SVM, RNA e modelos lineares generalizados - GLM) com distribuição gaussiana). O *Ensemble* resultou em maior coeficiente de determinação, enquanto a RNA no menor. O erro padrão da estimativa variou entre 8,21 % e 12,99 %, para as técnicas *Ensemble* e RNA, respectivamente. Os modelos preditivos apresentaram generalização acurada e precisa, com RMSE entre 7,01 % e 13,17 %. Schumacher-Hall e *Ensemble* apresentaram leve tendência em subestimar o volume, já SVM e RNA superestimaram. Os modelos de Schumacher-Hall e *Ensemble* foram os mais precisos e acurados à predição do volume para os dados utilizados.

PALAVRAS-CHAVE: **Ensemble, RNA, Schumacher-Hall, SVM.**

ABSTRACT

The forest stands growth, require techniques that provide more accuracy on the individual tree's volume estimates. In this context, machine learning techniques were applied for this purpose, however, these technique's combined (*Ensemble*) use has not yet been tested. Thus, this study's hypothesis is that *Ensemble* provides most accuracy volume estimates than conventional methods of estimation and machine learning techniques by itself. Thus, the aim was to evaluate the *Eucalyptus* spp. individual volume estimates using isolated and joint machine learning techniques. The study was conducted on two farms in Paragominas, Pará. 90 trees were cubed, distributed in 4 clones, with 6 years age and 1,111 ha⁻¹ individuals. 70% of the trees were used for fit and 30% for validation. Four estimation methods were applied: linear regression (Schumacher-Hall Model), support vector machines (SVM), artificial neural networks (ANN) and *Ensemble* (SVM, ANN and generalized linear models - GLM with Gaussian distribution). *Ensemble* resulted the highest coefficient of determination, while RNA obtained the lowest. The standard estimative error varied between 8.21% and 12.99% for *Ensemble* and ANN techniques, respectively. The predictive models presented accurate generalization, with RMSE between 7.01% and 13.17%. Schumacher-Hall and *Ensemble* showed a slight tendency to underestimate the volume, while SVM and ANN overestimated. The Schumacher-Hall and *Ensemble* models were the most accuracy on volume estimates in this database.

KEYWORDS: **Ensemble, ANN, Schumacher-Hall, SVM.**

INTRODUÇÃO

O investimento em florestas equiâneas vem crescendo em função das barreiras legais e necessidade de redução da utilização das florestas naturais. Com isso, é preciso que o estoque de crescimento em campo seja obtido de forma fidedigna, com objetivo de auxiliar a gestão florestal e a tomada de decisão, promovendo uma atividade sustentável e economicamente viável. Contudo, apesar de existirem diversas pesquisas sobre métodos de modelagem preditiva de atributos florestais (ELITH et al., 2006; GARZÓN et al., 2006; WERE et al., 2015; VENDRUSCULO et al., 2016; GAO et al., 2018; CUNHA NETO et al., 2018, 2019a, 2019b;), ainda é crescente a necessidade de aprimorar técnicas que proporcionem estimativas mais precisas e acuradas para estimar o volume de árvores individuais.

Estimativas confiáveis requerem equações compatíveis com as condições de sítio a serem inventariadas, levando em consideração a espécie, classe de idade, espaçamento, e se tratando de plantios do gênero *Eucalyptus* deve-se considerar inclusive o tipo de clone produzido (BINOTI et al., 2014). Os plantios de *Eucalyptus* são amplamente utilizados pelo rápido crescimento, adaptabilidade a diversas condições edafoclimáticas e versatilidade no uso de seus produtos e subprodutos (SMETHURST et al., 2020), o que ressalta a necessidade de grandes volumes de estudos direcionados para o gênero. Assim, os modelos preditivos buscam gerar estimadores que podem ser utilizados com diferentes conjuntos de dados, sob mesmas condições, para prever fenômenos físicos (COSTA FILHO et al., 2019).

A Aprendizagem de Máquina (*Machine Learning*) foi desenvolvida por meio de algoritmos advindos dos avanços computacionais e na área florestal, os quais podem ser utilizados para previsões. Seu uso está relacionado à capacidade de trabalho com dados imperfeitos, bem como a redução do impacto de outliers (WITTEN; FRANK, 2005). De forma geral, o uso de Aprendizagem de Máquina para esse tipo de análise, tem gerado resultados consistentes para as diversas aplicações (GAO et al., 2018; PANDIT et al., 2018).

Dentre as técnicas de aprendizagem de máquina, destaca-se as Redes Neurais artificiais (RNA), devido apresentar resultados precisos e acurados na previsão de variáveis biométricas florestais (BINOTI et al., 2014; MARTINS et al., 2016a, 2016b; CUNHA NETO et al., 2019b). Para a variável volume, em muitos casos, esses algoritmos apresentam resultados mais robustos e precisos quando

comparados aos modelos de regressão tradicionais (BINOTI et al., 2014; GORGENS et al., 2014).

Outra técnica que merece destaque é a Máquina de vetores de suporte (SVM), a qual se baseia na minimização de riscos, favorecendo a acurácia nas estimativas (GIGOVIĆ et al., 2019). Apesar disso, existem poucos estudos que avaliem o emprego dessa ferramenta nas ciências florestais, tais como: para estimativa do volume de *Acacia mangium* (CORDEIRO et al., 2015), previsão de biomassa de florestas tropicais (MONTAÑO et al., 2017), e relação hipsométrica em plantios de *Pinus taeda* (COSTA FILHO et al., 2019). Além dessas técnicas, a aprendizagem de máquina possibilita unir vários algoritmos em um único modelo, denominado de classificador *Ensemble* (MARQUES et al., 2012), o qual possui maior precisão e/ou robustez dos modelos (OZA & TUMER, 2008). Porém, assim como a SVM, pouco se conhece sobre o potencial do *Ensemble* nas previsões florestais.

Como método de classificação, o *Ensemble* demonstrou ser eficaz, com maior acurácia em relação aos modelos independentes (MINOWA, 2008; ENGLER et al., 2013; BUI et al., 2019; GIGOVIĆ et al., 2019; PENG et al., 2020). Com base no potencial dessa técnica, é imprescindível testá-la na estimação de variáveis florestais, a fim de verificar se há ganhos em precisão e acurácia no *Ensemble* em relação aos modelos que utilizam uma única técnica preditiva.

Neste contexto, a hipótese deste estudo é que combinação de técnicas de aprendizado de máquinas proporcionam estimativas mais precisas e acuradas do volume que os métodos convencionais de estimativa e aprendizado de máquina isoladamente. Para tanto, este trabalho teve como objetivo avaliar a estimativa do volume individual de *Eucalyptus* spp. utilizando técnicas isoladas e conjuntas de aprendizado de máquina.

MATERIAL E MÉTODOS

Caracterização de área e coleta de dados

O estudo foi realizado no bloco Piquiá, correspondente as fazendas Piquiá e Formosa, pertencentes a empresa Transportadora Floresta do Araguaia – TFA, localizadas no município de Paragominas, Pará, na Amazônia Oriental brasileira. O município possui clima do tipo Aw (apresentando estação mais seca no inverno) segundo a classificação climática de Köppen, com precipitação média anual de 1.800 mm e temperatura média de 26 °C (ALVARES et al., 2013).

Nessa fazenda foram cubadas 90 árvores, as quais estavam distribuídas em 4 clones, com 6 anos idade e

densidade média de 1.111 indivíduos.ha⁻¹, a compreender: C1 = *Eucalyptus brassiana* S.T. Blake X *Eucalyptus grandis* W. Hill ex Maiden, C2 = *Eucalyptus* sp., C3 = *Eucalyptus grandis* W. Hill ex Maiden x *Eucalyptus* sp. e C4 = *Eucalyptus urophylla* S.T. Blake.

A cubagem rigorosa foi realizada com base no método de Smalian, de tal maneira que as árvores foram seccionadas em 0,1 m; 0,5 m; 1,30 m; 2,0 m; 4,0 m e assim sucessivamente até o diâmetro igual ou inferior a 4,0 cm. O volume foi determinado a partir da soma dos volumes das secções.

Ajuste dos modelos

Para o ajuste dos modelos utilizou-se 70 % das árvores, totalizando 63 indivíduos. Foram ajustados quatro modelos preditivos, por quatro métodos distintos de estimação: regressão linear por mínimos quadrados ordinários (MQO), máquinas de vetores de suporte (SVM), redes neurais artificiais (RNA) e *Ensemble*, o qual agrega várias técnicas preditivas em um único modelo. No *Ensemble* foi utilizado SVM, RNA e modelos lineares generalizados (GLM) com distribuição gaussiana.

Em MQO aplicou-se o modelo de Schumacher e Hall (1933), enquanto nas técnicas de aprendizado de máquina, utilizou-se o volume como variável de saída, assim como diâmetro à altura do peito (DAP) e altura como variável de entrada. Os dados foram processados em linguagem R (R CORE TEAM, 2019), com auxílio dos pacotes *Caret* (KUHNS et al., 2019), *Neuralnet* (FRITSCH et al., 2019) e *CaretEnsemble* (DEANE-MAYER & KNOWLES, 2016).

Para SVM variou-se o parâmetro sigma de 0,05 a 1,05, com intervalos de 0,05, enquanto o parâmetro C variou entre 0,25 e 2,5, com intervalos de 0,25. Para a RNA, utilizou-se o algoritmo de aprendizado *Resilient Propagation RPROP+*, pelo qual foram testadas 10 configurações de neurônios na camada oculta, variando entre 1 a 10 de 1 em 1. Os parâmetros de busca da SVM e RNA do *Ensemble* foram os mesmos das técnicas isoladas.

Avaliação e validação do ajuste dos modelos

A avaliação do ajuste dos modelos foi realizada com base nas médias estatísticas: coeficiente de determinação (R²), erro padrão da estimativa (Syx) e critério de seleção de Akaike (AKAIKE, 1974), além da análise gráfica dos resíduos. No treinamento das técnicas de aprendizado de máquina foi realizada a validação cruzada, por *K-fold*, com 10 *folds* e 5 repetições, a fim de evitar *underfitting* e *overfitting*.

Para a validação, aplicou-se o modelo nas 27 árvores que não participaram do ajuste e avaliou-se a precisão e acurácia com as medidas: raiz quadrada do erro médio (RMSE), viés e análise gráfica dos resíduos. Além disso, foi aplicado o teste *t*, ao nível de 5 % de significância, a fim de verificar se há semelhança estatística entre os volumes cubados e estimados. O modelo com melhor desempenho, frente a essas métricas de ajuste e validação foi o selecionado.

RESULTADOS E DISCUSSÃO

As árvores apresentam distribuição simétrica para as variáveis diâmetro, altura e volume, com diâmetro médio de 16,86 cm, altura média de 22,30 m e volume médio de 0,2794 m³ (Figura 1).

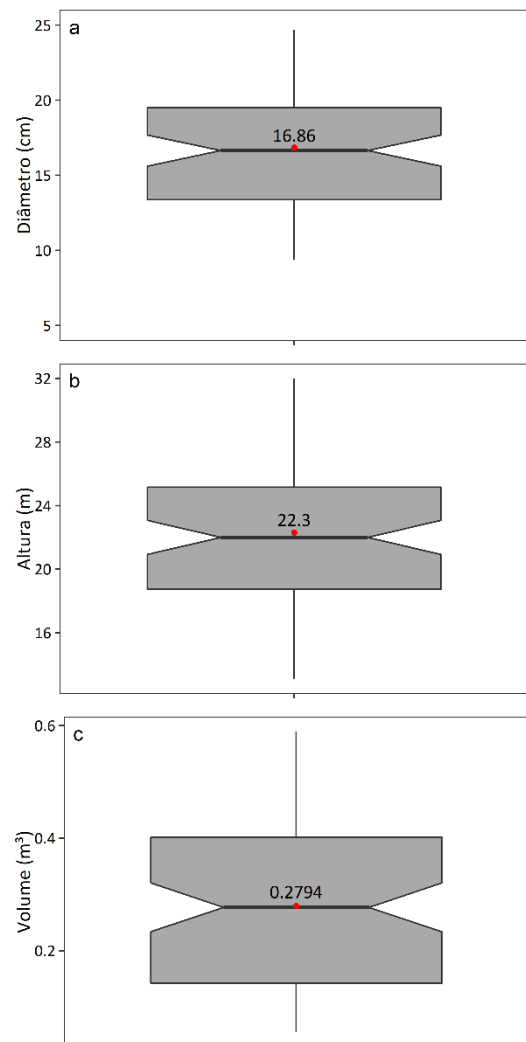


Figura 1. Boxplot das variáveis dendrométricas dos indivíduos de eucalipto na Amazônia Oriental. O ponto vermelho representa a média, enquanto a linha em negrito representa a mediana.

Os coeficientes da regressão foram significativos, denotando a importância das variáveis independentes para a variável dependente. A melhor configuração de SVM apresentou Sigma = 0,25 e C = 1,05. A RNA com melhor desempenho para os dados utilizados possui 4 neurônios na camada oculta. No *Ensemble*, a SVM selecionada obteve os mesmos parâmetros Sigma e C da SVM isolada, enquanto a RNA desta técnica possui 5 neurônios na camada oculta. Assim como o método MQO, o GLM do *Ensemble* obteve significância nos seus coeficientes (Tabela 1).

Tabela 1. Parâmetros dos modelos aplicados na estimativa do volume de *Eucalyptus* na Amazônia Oriental.

Parâmetros	MQO	SVM	RNA	Ensemble
P1	-10,0584*	0,25	2	0,25
P2	1,6143*	1,05	4	1,05
P3	1,3295*	-	1	2
P4	-	-	-	5
P5	-	-	-	1
P6	-	-	-	-0,4217*
P7	-	-	-	0,0236*
P8	-	-	-	0,0136*

P1 é β_0 para MQO; Sigma para SVM e *Ensemble*; quantidade de neurônios da camada de entrada para RNA, P2 é β_1 para MQO; C para SVM e *Ensemble*; quantidade de neurônios da camada oculta para RNA, P3 é β_2 para MQO; quantidade de neurônios da camada de saída para RNA; quantidade de neurônios da camada de entrada para *Ensemble*, P4 é a quantidade de neurônios da camada oculta, P5 é a quantidade de neurônios da camada de saída, P6 é o β_0 da GLM do *Ensemble*, P7 é o β_1 da GLM do *Ensemble*, P8 é o β_2 da GLM do *Ensemble*, * indica que coeficiente apresenta significância.

A técnica *Ensemble* demonstrou maior coeficiente de determinação, enquanto a RNA obteve o menor valor dessa estatística. O erro padrão da estimativa variou entre 8,21 % e 12,99 %, para as técnicas *Ensemble* e RNA, respectivamente. Com base no critério de informação de Akaike, o modelo que mais aproveitou a informação disponível nas variáveis independentes foi o *Ensemble*, seguido da SVM e MQO, respectivamente (Tabela 2).

Ainda na Tabela 2, observa-se o desempenho da técnica tradicional de ajuste (MQO) em comparação às técnicas com procedimentos mais robustos e dependentes computacionalmente (SVM, RNA e *Ensemble*), a qual, mostrou-se equivalente estatisticamente as demais. Inúmeros artigos têm comparado o modelo de Schumacher e Hall com outros modelos e técnicas para o ajuste de equações para o *Eucalyptus* sp. (REGO et al., 2019; ALMEIDA et al., 2020; MATTOS et al., 2020; MENDES et al., 2020) e outras espécies florestais (GARCIA-ESPINOZA & VALERIO, 2019; HERNÁNDEZ-RAMOS et al., 2019;

VILLAVICENCIO-GUTIÉRREZ et al., 2020), os autores citados obtiveram melhores estimativas na utilização deste modelo, demonstrando sua superioridade, o que corrobora com os resultados deste trabalho.

Tabela 2. Medidas estatísticas do ajuste e validação dos modelos aplicados na estimativa do volume de *Eucalyptus* spp. na Amazônia Oriental.

Medida estatística	Modelo			
	MQO	SVM	RNA	Ensemble
R ²	0,964	0,964	0,899	0,978
Syx (%)	10,659	9,359	12,988	8,207
AIC	-275,528	-289,330	-246,072	-306,664
RMSE (m ³)	0,021	0,020	0,037	0,019
RMSE (%)	7,525	7,335	13,171	7,014
Viés (m ³)	0,0021	-0,0020	-0,0036	0,0019
Viés (%)	0,576	-0,741	-1,515	0,873
teste t	0,037 ^{ns}	0,048 ^{ns}	0,101 ^{ns}	0,056 ^{ns}

R² é o coeficiente de determinação, Syx é o erro padrão da estimativa, AIC é o critério de informação de Akaike, RMSE é a raiz quadrada do erro médio, ns é não significativo pelo teste T ao nível de 5 % de significância.

Um fator determinante no ajuste de um modelo é sua aplicabilidade, ou seja, a utilização deste em práticas cotidianas, sem que seja necessário grande dispêndio de tempo e técnicas avançadas, neste ponto em específico, o modelo de Schumacher e Hall é superior aos demais, pois, os parâmetros ajustados podem ser aplicados sem grandes empecilhos em campo, denotando que este modelo é simples e possui acurácia e precisão nas estimativas (Tabela 2).

Os modelos preditivos apresentaram generalização acurada e precisa, com RMSE entre 7,01% e 13,17%. No método MQO foi observada uma leve tendência em subestimar o volume (Viés = 0,57%), já SVM e RNA superestimaram, enquanto *Ensemble* subestimou. Pelo teste t ao nível de 5% de significância, todas as técnicas estimaram volumes estatisticamente semelhantes aos volumes mensurados na cubagem rigorosa (Tabela 2).

Neste sentido, o método MQO, *Ensemble* e SVM obtiveram RMSE semelhantes, no entanto o *Ensemble* auferiu o melhor resultado dentre os modelos testados, denotando que este método é uma ferramenta preditiva eficaz no âmbito da aprendizagem de máquina, pois, potencializa os resultados, tornando-os mais precisos e acurados (ZHANG & SUGANTHAN, 2014; BUI et al., 2019; GIGOVIC et al., 2019).

Para MQO os resíduos foram mais dispersos que as outras técnicas, principalmente no ajuste do modelo, havendo a presença de um ponto influente com dispersão

superior a 3 desvios padrões. Para SVM e *Ensemble*, os resíduos foram homocedásticos no ajuste e validação do modelo, enquanto a RNA demonstrou tendência para estimar o volume das árvores com até 0,3 m³ (Figura 1a).

Em todas as técnicas verificou-se precisão nas estimativas do volume, com dispersão próxima a correlação perfeita (Figura 1b – linha tracejada), com destaque para SVM e *Ensemble* pela proximidade da dispersão dos pontos com a reta média (Figura 1b). Observa-se nesta figura que o modelo MQO e *Ensemble* foram semelhantes em comparação às retas vermelhas e a linha tracejada (correlação perfeita entre as variáveis), os demais modelos apresentaram pequenos desvios nas

árvores de menor e maior volume.

Contraopondo esses resultados, Almeida et al. (2020) obtiveram estimações mais precisas utilizando redes neurais artificiais em comparação com modelos não lineares e SVM. No entanto os autores utilizaram para o treinamento da rede 818 árvores com diferentes idades e tamanhos, e por esse motivo a rede foi mais acurada na predição do volume. Isto deve-se ao fato que quanto mais informação for disponibilizada para o treinamento da RNA, melhor será sua compreensão do fenômeno, conseqüentemente suas estimativas serão mais precisas e acuradas.

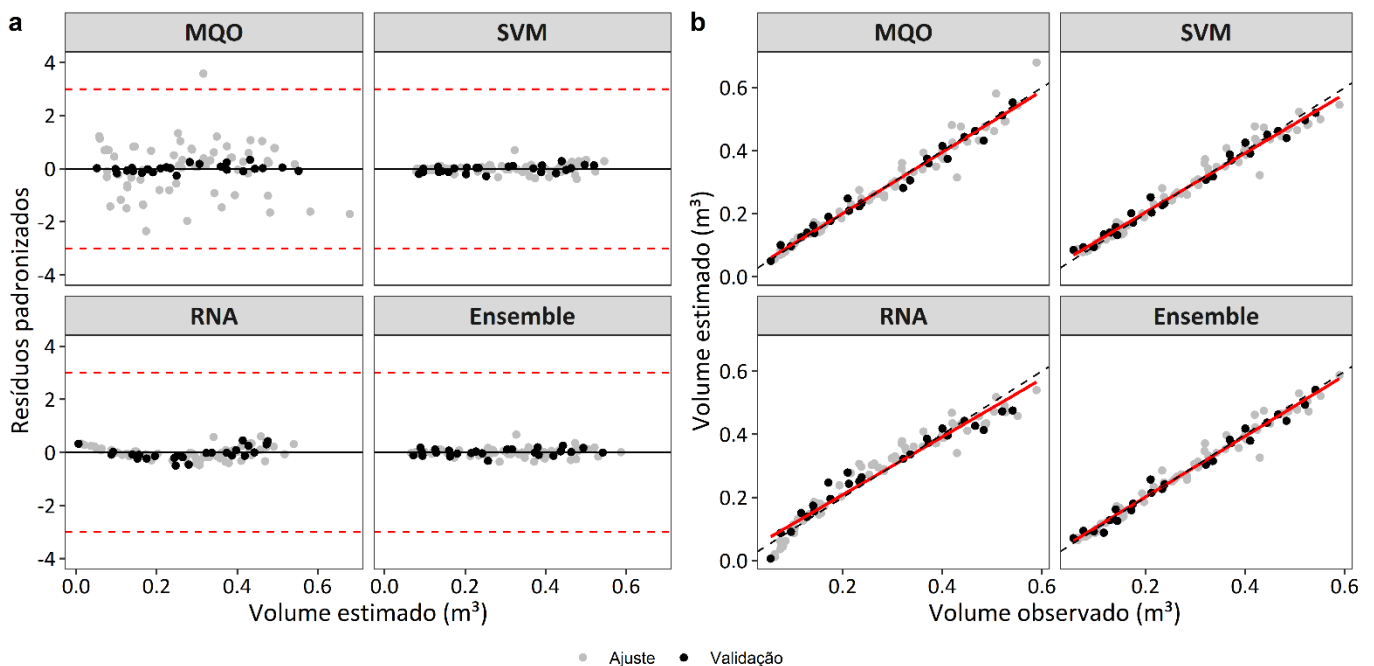


Figura 2. Dispersão dos resíduos e volumes dos indivíduos de *Eucalyptus* spp. na Amazônia Oriental.

CONCLUSÕES

As técnicas utilizadas são satisfatórias para estimativas do volume do *Eucalyptus* sp., as quais devem ser testadas para estimar outras variáveis, assim como em outras espécies florestais.

Os modelos de Schumacher-Hall e *Ensemble* são os mais precisos e acurados à predição do volume para os dados utilizados. Com destaque para Schumacher-Hall, pois utiliza técnicas mais simples de predição.

A quantidade de dados utilizada para o treino das redes neurais artificiais não foi suficiente para resultar em estimativas acuradas e precisas do volume estimado.

AGRADECIMENTOS

Os autores são gratos a empresa Transportadora Floresta do Araguaia (TFA), e ao Jóilson Rosa, por disponibilizar suas florestas para a realização deste estudo.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.19, n.6, p.716-723, 1974.
- ALMEIDA, M.R.D. et al. Ajuste de modelos volumétricos de espécies comerciais no município de Ipixuna (PA). *Revista em Agronegocio e Meio Ambiente*, v.13, n.1, p.259-278, 2020.
- ALVARES, C.A. et al. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v.22, n.6, p.711-728, 2013.
- BINOTI, D.H.B.; BINOTI, M.L.M.S.; LEITE, H.G. Configuração de redes neurais artificiais para estimação do volume de árvores. *Revista Ciência da Madeira - RCM*, v.5, n.1, p.58-67, 2014.

- BINOTI, M.L.M.S. et al. Redes neurais artificiais para estimar o volume de árvores. **Revista Arvore**, v.38, n.2, p.283-288, 2014.
- BUI, D.T. et al. New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. **Forests**, v.10, n.9, p. 743-764, 2019.
- CORDEIRO, M.A. et al. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, v.35, n.83, p.255-261, 2015.
- COSTA FILHO, S.V.S. et al. Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. **Ciência Florestal**, v.29, n.4, p.1501-1515, 2019.
- CUNHA NETO, E.M. et al. Redes neurais artificiais e regressão na estimativa da altura em povoamento experimental misto e equiâneo. **Caderno de ciências agrárias**, v.10, n.3, p.60-68, 2018.
- CUNHA NETO, E.M. et al. Modelo de Kozak e redes neurais artificiais na estimativa do fuste de eucalipto. **Revista de Engenharia e Tecnologia**, v.11, n.3, p.150-158, 2019a.
- CUNHA NETO, E.M. et al. Estimativa do afilamento do fuste de indivíduos de eucalipto por meio de técnicas de inteligência artificial. **BIOFIX Scientific Journal**, v.4, n.2, p.166-171, 2019b.
- DEANE-MAYER, Z.A.; KNOWLES, J.E. **caretEnsemble: ensembles of caret models**. R package, 2016. Disponível em: <<https://cran.r-project.org/package=caretEnsemble>>
- ELITH, J. et al. Novel methods improve prediction of species' distributions from occurrence data. **Ecography**, v.29, n.2, p.129-151, 2006.
- ENGLER, R. et al. Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. **Forest Ecology and Management**, v.310, p.64-73, 2013.
- FRITSCH, S.; GUENTHER, F.; WRIGHT, M.N. **neuralnet: training of neural networks**. R package, 2019. Disponível em: <<https://cran.r-project.org/package=neuralnet>>
- GAO, Y. et al. Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. **Remote Sensing**, v.10, n.4, p.627-648, 2018.
- GARCIA-ESPINOZA, G.G.; VALERIO, P.D. Precisión de la fórmula de Pressler para la estimación del volumen de árboles en pie de *Pinus montezumae* Lamb. en Nuevo San Juan Parangaricutiro, Michoacán. **Ciencia Nicolaita**, v.78, p.87-101, 2019.
- GARZÓN, M.B. et al. Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. **Ecological Modelling**, v.197, n.3-4, p.383-393, 2006.
- GIGOVIĆ, L. et al. Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia's Tara National Park. **Forests**, v.10, n.5, p.408-428, 2019.
- GORGENS, E.B. et al. Influência da arquitetura na estimativa de volume de árvores individuais por meio de redes neurais artificiais. **Revista Arvore**, v.38, n.2, p.289-295, 2014.
- HERNÁNDEZ-RAMOS, A. et al. Models to estimate biomass of *Euphorbia antisyphilitica* Zucc. in six townships of Coahuila. **Madera y Bosques**, v.25, n.2, p.1-13, 2019.
- KUHN, M. et al. **caret: classification and regression training**. R package, 2019. Disponível em: <<https://cran.r-project.org/package=caret>>
- MARQUES, A.; GARCIA, A.; SANCHEZ, J. Exploring the behaviour of base classifiers in credit scoring ensembles. **Expert Systems with Applications**, v.39, n.11, p.10244-10250, 2012.
- MARTINS, E.D.R. et al. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. **Revista Brasileira de Ciências Agrárias**, v.11, n.1, p.33-38, 2016a.
- MARTINS, E.R. et al. Configuração de redes neurais artificiais para estimação da altura total de árvores de eucalipto. **Revista Brasileira de Ciências Agrárias**, v.11, n.2, p.117-123, 2016b.
- MATTOS, E.M. et al. Variation in canopy structure, leaf area, light interception and light use efficiency among *Eucalyptus* clones. **Forest Ecology and Management**, v.463, p.1-12, 2020.
- MENDES, T.R.S. et al. Use of aerial image in the estimation of volume and biomass of *Eucalyptus* sp. forest stand. **Australian Journal of Crop Science**, v.14, n.2, p.286-294, 2020.
- MINOWA, Y. Verification for generalizability and accuracy of a thinning-trees selection model with the ensemble learning algorithm and the cross-validation method. **Journal of Forest Research**, v.13, n.5, p.275-285, 2008.
- MONTAÑO, R.A.N.R. et al. Artificial intelligence models to estimate biomass of tropical forest trees. **Polibits**, v.56, n.1, p.29-37, 2017.
- OZA, N.C.; TUMER, K. Classifier ensembles: select real-world applications. **Information Fusion**, v.9, n. 1, p.4-20, 2008.
- PANDIT, S.; TSUYUKI, S.; DUBE, T. Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data. **Remote Sensing**, v.10, n. 4, p.601-618, 2018.
- PENG, L. et al. Combining GF-2 and RapidEye satellite data for mapping mangrove species using ensemble machine-learning methods. **International Journal of Remote Sensing**, v.41, n.3, p.813-838, 2020.
- R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2019.
- REGO, A.M. et al. Modelos de equações volumétricas para um povoamento de eucalipto situado em Codó (MA). **Revista Ibero-**

Americana de Ciências Ambientais, v.10, n.4, p.1-8, 2019.

SCHUMACHER, F.X.; HALL, F. Logarithmic expression of timber-tree volume. **Journal of Agriculture Research**, v.47, p.719-734, 1933.

SMETHURST, P.J. et al. Generalized model for plantation production of *Eucalyptus grandis* and hybrids for genotype-site-management applications. **Forest Ecology and Management**, v.469, p.118-164, 2020.

VENDRUSCULO, L.G. et al. Height prediction of *Tectona grandis* trees by mixed effects modelling and artificial neural networks. **International Journal of Current Research**, v.8, n.12, p.43189-43195, 2016.

VILLAVICENCIO-GUTIÉRREZ, E.E.; MENDOZA-MORALES, S.; GONZALEZ, J.M. Modelo para predecir biomasa foliar seca de *Litsea parvifolia* (Hemsl.) Mez. **Revista Mexicana de Ciencias Forestales**, v.11, n.58, p.112-133, 2020.

WERE, K. et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. **Ecological Indicators**, v.52, p.394-403, 2015.

WITTEN, I.H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2.ed. San Francisco, 2005.

ZHANG, L.; SUGANTHAN, P.N. Random forests with ensemble of feature spaces. **Pattern Recognition**, v.47, n.10, p.3429-3437, 2014.

Recebido em 30-07-2021

Aceito em 28-08-2021