ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TREE VOLUME MODELLLING AND VALIDATION USING MACINE LEARNING APPROACHES FOR FOREST RESERVE

[1]SITI HAJAR MOHD MUSHAR, [2*]SHARIFAH SAKINAH SYED AHMAD, [3]FAUZIAH KASMIN, [4]NUR HAJAR ZAMAH SHARI

[1]Department of Intelligent Computing and Analytics, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, Durian Tunggal, 76100 Malacca, Malaysia

[2]Deputy Dean (Student Development), Assoc. Prof. Dr., Department of Intelligent Computing and Analytics, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, Durian Tunggal, 76100 Malacca, Malaysia

[3]Senior Lecturer, Dr., Department of Intelligent Computing and Analytics, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, Durian Tunggal, 76100 Malacca, Malaysia

[4]Senior Research Officier, Forestry and Environment Division, Forest Research Institute Malaysia (FRIM), Kepong, 52109 Kuala Lumpur, Malaysia

E-mail: [1]ctha_jar@yahoo.com.my, [2*]sakinah@utem.edu.my, [3]fauziah@utem.edu.my, [4]hajar@frim.gov.my

ABSTRACT

Forest plays an important role in national growth as the forestry and logging activities contribute 5.6% to the Malaysia GDP of the agricultural sector in 2018. A precise value of tree volume estimation highly affects forest management and administration. The forest management and administration framework are designed based on the evaluation of the forest, including its current volume; therefore this strongly supports the need for a precise tree volume estimation. Tree volume can be expressed either in terms of the total cubic volume of a tree or in terms of the total cubic volume of an area. However, this paper is going to focus on the volume estimation technique for an individual tree. Analysis of the literature found that the commonly used method in estimating the tree volume is regression, however, growth in the information technology has driven the use of machine learning techniques. The state-of-the-art highlighted that machine learning not only has a high capability in developing a robust model but also able to overcome the regression analysis problem such as overfitting of the data. Numerous comparison studies on the application of machine learning in forest modelling can be found but there are discrepancies of analysis among scholars. Therefore, this paper will perform tree volume estimation by using regression and four machine learning techniques which are artificial neural network (ANN), epsilon-Support Vector Regression (ε-SVR), k-Nearest Neighbor (k-NN) and random forest (RF). The precision and accuracy of the volume model will be verified by using the root mean square error (RMSE) and standard deviation (SD). The result and analysis of this study seem to be consistent with other research which found that machine learning techniques perform better than regression as the ANN is the best modelling technique for dipterocarp and non-dipterocarp datasets while for all species dataset, ε-SVR records the highest accuracy.

Keywords: *Machine Learning, Regression, Volume Model, Tree Volume*

1. INTRODUCTION

Generally, forest covered around 30.6% of the world terrestrial surface [1]. Forest is very important and there is a considerable amount of literature on the importance of forest that can be found. Henry et al. [2], Kuyah et al. [3], Xia et al. [4] & Mugasha et al. [5] emphasize that forest is one of the important components in the terrestrial ecosystem and plays major roles in regulating the climate and mitigating natural disasters has been emphasized. It has been conclusively known that forest is the largest carbon pools with an estimated amount of more than 650 billion tons included carbon content of biomass, dead wood and litter, as well as in the soil [6]. Carbon dioxide will be used up by the green plant for the photosynthesis process. Hence, the absorption of carbon dioxide will reduce the amount of carbon dioxide gas in the atmosphere and later will indirectly reduce the risk of thinning ozone layer and other natural disasters as well. Besides as a carbon sinker, forest plays a pivotal role in the national growth as the forest products or any potentially tradable forest activities is not only going to benefit the rural society, but to the country as a whole [4], [7].

Malaysia is a tropical country that characterized by moist tropical forest and statistics released by the Department of Statistics Malaysia (DOSM), forestry and logging contribute 5.6% to the GDP of the agricultural sector in 2018. Japan, US, India, Korea and China are among countries that have been importing Malaysia's major timber product. The urge in having a precise value of tree volume is highlighted by JR. and Wood [8] & Shari et al. [9] where it plays a vital role in the forest management and administration.

Tree volume can be expressed either in terms of total cubic volume of a tree or in terms of total cubic volume of an area. However, this paper is going to focus on the volume estimation technique for an individual tree. In a forest mensuration book written by Jr. et al. [10], the volume of an individual tree can be estimated by using three methods, which are graphical methods and integration, water displacement method and regression. Most of the research effort in the area of tree volume estimation apply the regression technique. However, growth in the information technology field has slowly transformed the current modeling technique of volume model development to machine learning.

It has been claimed by Parresol [11] that the common issue in the forest modelling is the values of the dependent variable is directly proportional to the variation in the error. This is contradicts with the initial assumption of least square regression whereby errors are independent and normally distributed with mean 0 and constant variance, $X \sim N(0,1)$. This may due to the fact that the real tree data is noisy in nature and probable of having high variability or non-normal distribution in the data is an important issue that needs to be handled in environmental modelling. Swingler [12] suggests the use of machine learning (ML) to overcome this problem. Generally, ML is part of artificial intelligence that ease the data analytics process as it automates the construction process of model analytics. There are several types of ML technique and the application of ML is not only limited to regression and classification, but it can also be used for clustering and anomaly detection.

Diamantopoulou et al. [13] point out that ML is favorable due to its capability in automating the detection of hidden data patterns and so modelling it. ANN has been claimed by Diamantopoulou et al. [13] as a reliable branch of ML for the application of forest modelling. In addition, the application of ANN is not only widely used in forestry, but numerous study can also be found in the healthcare field [14]–[17]. Generally, ANN is a system which inspired from the central nervous system of the human brain where it involves computerization of a simple mathematical function before being distributed parallelly [18], [19]. A vast application of ANN is found in forestry such as the biomass and volume prognosis [20]–[23], as well as the designation of model for forest growth and forest mortality [24], [25]. Besides that, the application of ANN can also be found in tree volume estimation study as in the work of Lacerda et al. [21] & Tavares Júnior et al. [26]. In terms of the accuracy of volume estimation, there are different opinions among scholars as Lacerda et al. [21] discover that ANN estimates the volume better than regression while Tavares Júnior et al. [26] is on the contrary.

Other than ANN, another dominant approach for regression type problem is support vector regression (SVR). Support vector machine (SVM) was proposed by Vapnik–Chervonenkis around 1960s and it can be categorized into two, which are support vector classification (SVC) and support vector regression (SVR) [27]. Briefly, SVR is a prediction tool which been introduced by Vapnik, Steven Golowich, and Alex Smola in 1992 and it uses the application of machine learning theory in automate the evasion of any data overfitting for high accuracy of the regression prediction [27], [28]. There are two types of SVR, which are epsilon-Support Vector Regression (ε-SVR) and nu-Support Vector Regression (μ-SVR). In a comparison study by Diamantopoulou et al. [13] reports that the ε-SVR performs better than the regression.

Apart from SVR, k-Nearest Neighbor (k-NN) has been classified as the top 10 of the most common practiced algorithm in data mining [29]. k-NN is a kind of supervised learning algorithm and non-parametric learning technique as described by Hechenbichler and Schliep [30]. Meanwhile, Faceli et al. [31] defined k-NN as the simplest ML algorithm as it based on the classification or estimation of a specific attribute. Generally, there are two types of k-NN, which are k-NN regression and k-NN classification. It has been identified in the study of Li et al. [32] that the application of k-NN comprises of three important components which are a set of labeled objects, a distance or similarity metrics for the calculation of object distance and the value of k (number of nearest neighbor). Friedman et al. [33] point out that k-NN's forte is dealing with low-dimensional problems. On top of that, k-NN functioning reasonably well for the modelling of complex relationships [34] and a vital tool in estimating the missing values in databases [35]. Despite all of its advantages, Schikowski et al. [34] report that k-NN lacks of statistical properties in comparison with the traditional regression method. In a recent study by Souza et al. [36] recommended the use of k-NN upon the condition where the classical regression method or any other simpler method does not perform well. However, Montano et al. [37] found that artificial intelligence models show better accuracy than the classic allometric models in a comparison study of biomass estimation by using artificial intelligence models and classic allometric models. In the study of Sanquetta et al. [22], two artificial intelligence models which are ANN and k-NN are tested and the result shows that k-NN performs the best in estimating the volume estimation of *Cryptomeria japonica* logs. Conversely, Schikowski et al. [34] report that k-NN has the least accuracy in modelling of stem volume of several machine learning techniques.

In the year 2001, Breiman [38] has devised a supervised learning algorithm named random forest (RF). RF used the application of ensemble learning method in classifying and regressing that makes it arguably one of the most powerful statistical learning methods as claimed by Roy and Larocque [39]. Even RF does not require assumptions on its data distribution, it is said to have high capability in modelling large number of predictive variables without overtraining [40]. Yue et al. [41] point out that RF has a high accuracy level. It has been proven in a comparison study of several ML and conventional regression techniques by Yue et al. [41] that RF suits for a study that set out for a model that precisely estimates the prediction. In a study of Schikowski et al. [34], the accuracy of RF in modelling the stem volume is tested against the other two ML techniques and the result shows that RF performs better k-NN but lower than ANN. Unlike Yue et al. [41] & Schikowski et al. [34], Rex et al. [42] identify that linear regression is more accurate than RF in estimating the aboveground biomass (AGB). Apart from accuracy, RF is said to have the capability in dealing with data that is high in noise and outliers. This is supported by the noise immunity result performed by Yue et al. [41] that RF performs the best than the other seven techniques.

This study, therefore, set out to explore and assess the performance of tree volume estimation by using regression and machine learning techniques. With regards to this, real data corresponding to measurements originating from one of the Malaysia forest reserved is used.

2. MATERIAL AND METHODS

This section presents the methodology for tree volume estimation along with the background information regarding the used regression, ANN, RF, k-NN and ε-SVR models. It begins with the data description including the measurement procedure then followed by further elaboration on the application of this dataset from the point view of statistical technique until the assessment of the tree volume estimation by using regression and machine learning.

2.1 Data Description

This study adopts selective sampling methods. The number of trees was selected based on diameter classes and species group. The real tree data was collected at the compartment 37, Cherul Forest Reserve, Terengganu and the measurement were carried out on both, the standing and felled trees. Trees that are dead, dying, broken or even the bad form was excluded from measurements. Similarly, trees of a diameter at breast height (DBH) less than 15 cm. For the standing trees, handheld Laser Criterion 400 was used as an instrument to measure the diameter and height of the tree. Meanwhile, for the measurement on felled trees were taken by the forest officer by using the caliper and measuring tape before the log is transferred to the logger's lorry.

Measurements on the merchantable length and stump height was taken and recorded. The measurement of the DBH at 1.3m above ground level was recorded. Besides, the total height was also measured from predetermined height of ground till the crown point. Other than DBH and total height, the stem diameter (overbark) were measured from the stump height up to first main branch at an interval of 2m. The volume of the tree were then measured by using Huber's formula, a commonly-used method in Europe and some other tropical countries as well [43], [44]. The formula for Huber's is as follows:

$$V_i = f * L * (dM)^2 \qquad (1)$$

where:

V_i = Log volume at *i*th (m³)
L = Length of log (m)
f = 0.00007854 (metric units)
dM = Diameter at the mid-length log end (cm)

The total tree sample data collected is 265, however, after going through the preprocessing stage, the number of tree sample data is 241. One of the aims of the study was to produce volume equation that not only as in general, but also for dipterocarp and non-dipterocarp species. A total of 241 tree sample which is then subdivided into its tree group which either dipterocarp or non-dipterocarp. It has been analyzed that out of 241 tree samples, 201 samples were dipterocarps, while 40 other samples were non-dipterocarps. Briefly, dipterocarps are family of the hardwood and the coverage of dipterocarps at the emergent layer of the lowland forest in Peninsular Malaysia is about 57%. Meanwhile, non-dipterocarp trees are all the other trees that make up the rest of the jungle.

Table 1: Summary Statistics for Fitting and Testing Datasets

Variable	Fitting data				Testing data			
	Mean	Min.	Max.	SD	Mean	Min.	Max.	SD
	All species (169 trees)				All species (72 trees)			
D (cm)	52.2	12.8	98.0	17.0	53.1	23.6	129.6	18.0
H (m)	17.0	10.7	30.0	3.9	17.3	6.5	25.9	3.7
	Dipterocarp species (28 trees)				Dipterocarp species (12 trees)			
D (cm)	61.7	25.8	98.0	20.0	66.6	32.8	129.6	26.9
H (m)	20.4	13.3	30.0	4.6	17.1	11.9	20.9	2.6
	Non-dipterocarp species (141 trees)				Non-dipterocarp species (60 trees)			
D (cm)	49.3	23.6	86.9	14.4	52.66	12.8	99.4	17.26
H (m)	16.4	6.5	28.2	3.5	17.3	11.2	25.9	3.75

The data in each group was then subdivided according to its diameter class. The rule of thumb was applied to each group species where 70% of the data is for training and 30% is for testing. In order to avoid bias, the rule of thumb was also applied to each diameter class. The testing dataset is totally independent from the fitting procedure for any types of constructed models including the regression and machine learning techniques and it is only used for the evaluation of each respective constructed models. Therefore, Table 1 shows summary statistics of three different groups which all species, dipterocarps species and non-dipterocarps species.

2.2 Regression Analysis

As a result of extensive research on volume table development, it has been found that there are lists of regression model which commonly-used by researchers. Each dataset will be fitted into volume models as listed below :

1. $V = b_o + b_1 D$
2. $V = b_o + b_1 D + b_2 D^2$
3. $V = b_o + b_1 D^2$

where:

V = Tree volume (m³)
D = Diameter (m)
H = Log length (m)
b_i = Regression coefficients

2.3 Support Vector Regression Models (SVRs)

Applying the ε-SVR algorithm of SVR theory, the input data is first mapped onto an m-dimensional feature space using nonlinear mapping and then a linear model is constructed in this feature space. The goal is to find a function f(x) among the pairs of the training data (input x_i, target y_i) = (x_i, y_i) without considering these pairs that show deviation from the SV larger than ε deviation. According to this principle, the error band of the function f(x) lies in the interval [-ε, ε]. In this paper, the ε-SVR modelling was performed by using "e1071" package in the R software which enables the parameters of the ε-SVR models to be optimized and trained.

2.4 Artificial Neural Network (ANN)

ANNs are comprised of one or more processing units called 'artificial neurons' or 'perceptrons' [45]. Perceptrons of an ANN are interconnected with one another by a series of weighted connections. The perceptrons of an ANN, depending on the system being replicated, are arranged in layers, with each perceptron of the preceding layer having a weighted connection with each neuron of the proceeding layer. In the process of ANN training to replicate a system, a training data set is fed through the network. Each perceptron processes the input data or input signal from either the input layer or the preceding perceptrons. The final layer of the ANN produces an output signal. The weights and structure of the network are altered in a manner depending on the specific training algorithm.

2.5 k-Nearest Neighbor (k-NN)

k-NN comprises of three important components which are a set of labeled objects, a distance or similarity metrics for the calculation of object distance and the value of k (number of nearest neighbor) [32]. In this paper, the k-NN modelling was performed by using "caret" package in the R software which enables the search of optimal fit hyperparameters k. This include the process of k-fold validation, repeated k-fold cross-validation, leave-one out cross validation and bootstrap. The trainControl function was used in the training process of k-NN model of 3 repeats of 10-folds cross validation. The values of k were defined by manual grid by the tuneGrid argument. Therefore, there is 450 k-NN models to be validated as there is 3x10-CV *folds* scheme for k 1:15. R-squared and root mean square error (RMSE) are then used as the performance measure to test the k-NN model in each subset k disjoint.

2.6 Random Forest (RF)

Random forest is an ensemble learning method whereby the prediction will be based on a combination of multiple smaller models to achieve a high predictive power and more generalized. "randomForest" package in the R software was used in performing the modelling by random forest algorithm.

2.7 Statistical Criteria for Model Evaluation

Furnival's Index (FI) is chosen in this study instead of other common statistical test is to avoid the problem that will arise in the data analysis phase, due to the inclusion of transformations of the dependent variable and weighted regression into the analysis. Meanwhile, the standard error (SE) is required in the calculations of Furnival's Index. Thus, the formula of Furnival's Index is as the following:

$$FI = [f'(V)]^{-1} * s \quad (2)$$

where:

FI	=	Furnival's Index
s	=	Residual standard error from the fitted regression
$[f'(V)]^{-1}$	=	Geometric mean

Precision is the consistency measure of the prediction as defined by Walther and Moore [46]. The limitation that exists either in the measurement or estimation technique which been used at different times and under several circumstances is said to give effect to the preciseness level [10]. The standard deviation (SD) is classified by Jr. et al. [10] as one of the precision measurement. Therefore, the precision is estimated by using the following formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(V_i - \bar{V})^2}{(n-1)}} \quad (3)$$

where:

V_i	=	Log volume at *i*th (m^3)
$\bar{V}$	=	Mean volume (m^3)
n	=	Number of sample

Accuracy represents the closeness level of the estimation from the true value. Mean square error (MSE) is the most common accuracy measure because it comprises a combination concept of bias and precision [46]. This can be seen where an accurate estimator would have a small value of variance and the least bias in the prediction. Meanwhile, RMSE is the standard deviation of the prediction error and it can be expressed by using these formulae:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(V_i - \bar{V})^2} \quad (4)$$

where:

V_i = Log volume at *i*th (m³)
$\bar{V}$ = Mean volume (m³)
n = Number of sample

3. RESULTS

The result of the regression coefficient, RMSE and SD for all species, dipterocarp and non-dipterocarp species were summarized as in

Table *2*. The table contains the results for both, the fitting and testing datasets

Table *2* displays the regression analysis result for all species, dipterocarp and non-dipterocarp dataset. From the table, fitting of 169 number of trees, volume model 1 and 2 records an equal yet the lowest value of FI, which is 1.3550. Poor fit of the model is implies from a large value of FI, therefore, the lower the better. Meanwhile, looking on dipterocarp species dataset, volume model 3 is the best-fitted model with the value of 2.0820. For the non-dipterocarp species, volume model 2 is the best-fitted value as it record the lowest FI value.

The completion of volume equation modelling process will then be followed by model validation. Model validation will use the remaining of the dataset that did not participate in the regression analysis for the purpose of biases avoidance. The performance of the model validation was then evaluated from the value of SD and RMSE.

Table *2* also summarize the result obtained from the statistical test.

Result of the statistical test attained by each dataset is quite revealing in several ways. The consistency of the prediction can be seen from SD value. Then, the RMSE gives a representation of how far the estimated volume deviates from the true volume. Apparently, from the statistical test attained by all species dataset, volume model 1 recorded the smallest value of the SD and RMSE with the value of 1.2907 and 1.9864 respectively. On the other hand, for the dipterocarp species dataset, volume model 1 also recorded the smallest value in the RMSE and SD. For the case of non-dipterocarp species dataset, the lowest value of RMSE and SD is recorded by volume model 2 with the value of 1.3980 and 1.6105 respectively. Hence, it could conceivably be hypothesized that volume model 1 is the best volume model for all species, dipterocarp and non-dipterocarp as RMSE and SD is all about the accuracy and precision of the model and lower in the RMSE and SD value indicates higher accuracy and precision level.

Table 2: Results of Regression Analysis and Statistical Test for Fitting and Testing Data

Model	Fitting data				Testing data	
	β_0	β_1	β_2	FI	RMSE	SD
	All species (169 trees)				All species (72 trees)	
Vol Model 1	-2.8658	0.1120		1.3550	**1.2907**	**1.9864**
Vol Model 2	0.2031	-0.0181	0.0012	1.3550	1.7553	2.6214
Vol Model 3	-0.2408	0.0011		1.3980	1.6763	2.5367
	Dipterocarp species (28 trees)				Dipterocarp species (12 trees)	
Vol Model 1	-3.7566	0.1334		2.1900	**1.8680**	**3.5948**
Vol Model 2	1.7472	-0.0842	0.0019	2.1040	4.0170	5.9939
Vol Model 3	-0.4645	0.0012		2.0820	3.1552	5.1079
	Non-dipterocarp species (141 trees)				Non-dipterocarp species (60 trees)	
Vol Model 1	-2.3691	0.0990		1.0600	**1.3980**	**1.6105**
Vol Model 2	-1.3247	0.0526	0.0005	1.0590	1.4290	1.7130
Vol Model 3	-0.1076	-0.0010		1.0610	1.4890	1.8213

Table 3: Comparison of Result between Regression and Machine Learning Techniques

	All species	Dipterocarp	Non-dipterocarp

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANN	1.2639	**1.6750**	**0.6377**
SVR	**0.8076**	2.1627	0.8762
RF	0.9273	2.3368	0.8774
k-NN	0.9876	2.2842	0.9134
Regression	1.2907	1.8680	1.3980

In order to meet the objective of the study, the accuracy assessment of the constructed volume model was conducted. Besides than modelling the tree volume by using the regression, the tree volume modelling by using machine learning techniques was also performed. The accuracy assessment results of five different modelling techniques, which are regression, ANN, SVR, RF and k-NN are as shown in Table 3.

It can be seen from the numeric value in Table 3 that the accuracy of the regression method for all species and dipterocarp group data is in between the four machine learning techniques. For all species data, SVR recorded the lowest value of RMSE as compared to ANN, RF, k-NN and regression. In short, RMSE represents the distance between the estimated volume and true volume, therefore, the volume estimation using regression deviates by 1.2907 from its true volume. This is contradicting with the dipterocarp group data whereby, SVR is the method that results in the highest value of RMSE of 2.1627. Meanwhile, the lowest value of RMSE was recorded by ANN. For non-dipterocarp data, the accuracy assessment is the same pattern with the dipterocarp group data as regression records the highest RMSE value.

4. CONCLUSIONS

The findings of this study will be of assistance and interest to the researchers in assessing the accuracy of tree volume modelling via regression and machine learning techniques. There is a limited number of studies that compare tree volume modelling via regression against several machine learning techniques as it can be widely found that researchers normally compare at most two machine learning techniques. Taken together, this study examines the accuracy of the most powerful tools in analyzing natural and physical sciences data, regression analysis, and machine learning techniques.

An analysis of regression techniques shows that volume model 1 is the best for all species, dipterocarp and non-dipterocarp as it records the highest accuracy level. In modelling, one of the essential elements is the capability in managing the regression analysis problems such as overfitting of data. Regards to the fact that the machine learning technique is not only highly capable in developing a reliable but also a robust model that can deal with complex environmental problems, modelling via machine learning technique are then be chosen. The results of this study are in accord with several comparison studies of modelling between regression and machine learning whereby the performance of the machine learning is better than regression for all species, dipterocarp and non-dipterocarp dataset.

ACKNOWLEDGEMENTS

The authors express their deepest gratitude to Universiti Teknikal Malaysia Melaka (UTeM) for supporting this research PJP/2019/FTMK(68)/S01678.

REFRENCES:

[1] Food and Agriculture Organization (FAO), *Global Forest Resources Assessment 2015: How are the World's Forests Changing?*, 2nd ed. Rome, Italy: Food and Agriculture Organization of the United Nations, 2016.

[2] M. Henry *et al.*, "Estimating Tree Biomass of Sub-Saharan African Forests: A Review of Available Allometric Equations," *Silva Fenn*, vol. 45, no. 3B, pp. 477–569, 2011.

[3] S. Kuyah *et al.*, "Allometric Equations for Estimating Biomass in Agricultural Landscapes: I. Aboveground Biomass," *Agr Ecosyst Env.*, vol. 158, pp. 216–224, 2012.

[4] Z. S. Xia, W. S. Zeng, S. Zhu, and H. Z. Luo, "Construction of Tree Volume Equations for Chinese Fir Plantations in Guizhou Province, Southwestern China," *For. Sci. Pract.*, vol. 15, no. 3, pp. 179–185, 2013.

[5] W. A. Mugasha *et al.*, "Allometric Models for Prediction of Above- and Belowground Biomass of Trees in the Miombo Woodlands of Tanzania," *For. Ecol Manag*, vol. 310, pp. 87–101, 2013.

[6] Food and Agriculture Organization of the United Nations (FAO), "Global Forest

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Resources Assessment 2010," 2010.

[7] Food and Agriculture Organization of the United Nations (FAO), "State of the World's Forests," Rome, Italy, 1997.

[8] H. V. W. JR. and G. B. Wood, "Modern Methods of Estimating Tree and Log Volume," in *IUFRO Conference Proceedings*, 1993, p. 9.

[9] N. H. Z. Shari, W. M. S. W. Ahmad, S. Musa, W. R. W. Mohd., and H. Ismail, "Development of Local Volume Table for Second Growth Forests Using Standing Tree Measurements," *Malaysian For.*, vol. 73, no. 2, pp. 163–170, 2010.

[10] J. A. K. Jr., M. J. Ducey, T. W. Beers, and B. Husch, "Determination of Tree Volume, Weight and Biomass," in *Forest Mensuration*, 5th ed., Wiley-Blackwell, 2016, pp. 135–183.

[11] B. R. Parresol, "Modeling Multiplicative Error Variance: An Example Predicting Tree Diameter from Stump Dimensions in Baldcypress.," *For. Sci.*, vol. 39, pp. 670–679, 1993.

[12] K. Swingler, *Applying Neural Networks: A Practical Guide*. Great Britain: Morgan Kaufman Publishers Inc., 2001.

[13] M. J. Diamantopoulou, R. Özçelik, and H. Yavuz, "Tree-bark Volume Prediction via Machine Learning: A Case Study Based Onblack Alder's Tree-bark Production," *Comput. Electron. Agric.*, vol. 151, pp. 431–440, 2018.

[14] D. A. Roffman *et al.*, "Development and Validation of a Multiparameterized Artificial Neural Network for Prostate Cancer Risk Prediction and Stratification," *JCO Clin. Cancer Iformatics*, vol. 2, pp. 1–10, 2018.

[15] A. Catic, L. Gurbeta, A. Kurtovic-Kozaric, S. Mehmedbasic, and A. Badnjevic, "Application of Neural Networks for Classification of Patau, Edwards, Down, Turner and Klinefelter Syndrome based on First Trimester Maternal Serum Screening Data, Ultrasonographic Findings and Patient Demographics," *BMC Med. Genomics*, vol. 11, no. 1, p. 19, 2018.

[16] I. O. Ahmed, B. A. Ibraheem, and Z. A. Mustafa, "Detection of Eye Melanoma Using Artificial Neural Network," *J. Clin. Eng.*, vol. 43, no. 1, pp. 22–28, 2018.

[17] P. F. J. Tsai *et al.*, "Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients using Artificial Neural Network," *J. Healthc. Eng.*, 2016.

[18] R. Özçelik, M. J. Diamantopoulou, J. R. Brooks, and H. V. Wiant, "Estimating Tree Bole Volume using Artificial Neural Network Models for Four Species in Turkey," *J. Environ. Manag.*, vol. 91, no. 3, pp. 742–753, 2010.

[19] L. P. Reis *et al.*, "Prognosis on the Diameter of Individual Trees on the Eastern Region of the Amazon using Artificial Neural Networks," *For. Ecol. Manag*, vol. 382, pp. 161–167, 2016.

[20] S. Yang, Q. Feng, T. Liang, B. Liu, W. Zhang, and H. Xie, "Modeling Grassland Above-ground Biomass based on Artificial Neural Network and Remote Sensing in the Three-River Headwaters Region," *Remote Sens. Environ.*, vol. 204, pp. 448–455, 2018.

[21] T. H. S. Lacerda, C. D. Cabacinha, C. A. Araújo Júnior, R. D. Maia, and K. W. de S. Lacerda, "Artificial Neural Networks for Estimating Tree Volume in the Brazilian Savanna," *Cerne*, vol. 23, no. 4, 2017.

[22] C. R. Sanquetta, L. R. Piva, J. Wojciechowski, A. P. Corte, and A. B. Schikowski, "Volume Estimation of Cryptomeria Japonica Logs in Southern Brazil using Artificial Intelligence Models," *South. For. a J. For. Sci.*, vol. 80, no. 1, pp. 29–36, 2017.

[23] A. A. Vahedi, "Artificial Neural Network Application in Comparison with Modeling Allometric Equations for Predicting Above-ground Biomass in the Hyrcanian Mixed-beech Forests of Iran," *Biomass Bioenergy*, vol. 88, pp. 66–76, 2016.

[24] D. Zhang, L. Zhang, Q. Ye, and H. Ruan, "Robust Learning-based Prediction for Timber-volume of Living Trees," *Comput. Electron. Agric.*, vol. 136, pp. 97–110, 2017.

[25] S. J. S. S. da Rocha *et al.*, "Artificial Neural Networks: Modeling Tree Survival and Mortality in the Atlantic Forest Biome in Brazil," *Sci. Total Environ.*, vol. 645, pp. 655–661, 2018.

[26] I. D. S. Tavares Júnior *et al.*, "Artificial Neural Networks and Linear Regression Reduce Sample Intensity to Predict the Commercial Volume of Eucalyptus Clones," *Forests*, vol. 10, no. 3, p. 268, 2019.

[27] J. Langhammer and J. Cesák, "Applicability of a Nu-Support Vector Regression Model for the Completion of Missing Data in Hydrological Time Series," *Water*, vol. 8, no. 12, p. 560, 2016.

[28] D. Basak, S. Pal, and D. C. Patranabis, "Support Vector Regression," *Neural Inf. Process. Rev.*, vol. 11, no. 10, pp. 203–224, 2007.

[29] X. Wu *et al.*, "Top 10 Algorithms in Data Mining," *Knowl. Inf. Syst.*, vol. 14, pp. 1–37, 2008.

[30] K. Hechenbichler and K. Schliep, "Weighted k-Nearest Neighbor Techniques and Ordinal Classification," Munich, 399, 2004.

[31] K. Faceli, G. Joao, A. C. . P. L. F. de Carvalho, and A. C. Lorena, *Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina*. Rio de Janeiro: LTC, 2011.

[32] C. Li, Z. Qiu, and C. Liu, "An Improved Weighted K-Nearest Neighbor Algorithm for Indoor Positioning," *Wirel. Pers. Commun.*, vol. 96, no. 2, pp. 239–2251, 2011.

[33] J. Friedman, T. Hastie, and R. Tibshirani, "Overview of Supervised Learning," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York: Springer, 2009, pp. 9–41.

[34] A. B. Schikowski, A. P. D. Corte, M. S. Ruza, C. R. Sanquetta, and R. A. N. R. Montaño, "Modeling of Stem Form and Volume through Machine Learning," *nais da Acad. Bras. Ciências*, vol. 90, no. 4, pp. 3389–3401, 2018.

[35] B. N. I. Eskelson, H. Temesgen, V. Lemay, T. M. Barrett, N. L. Crookston, and A. T. Hudak, "The Roles of Nearest Neighbor Methods in Imputing Missing Data in Forest Inventory and Monitoring Databases," *Scand J Res*, vol. 24, pp. 235–246, 2009.

[36] D. V. Souza *et al.*, "k-Nearest Neighbor Regression in the Estimation of Tectona Grandis Trunk Volume in the State of Pará, Brazil," *J. Sustain. For.*, pp. 1–14, 2019.

[37] R. A. N. R. Montano, C. R. Sanquetta, J. Wojciechowski, E. Mattar, A. P. D. Corte, and E. Todt, "Artificial Intelligence Models to Estimate Biomass of Tropical Forest Trees," *Polibits*, vol. 56, pp. 29–37, 2017.

[38] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 1, pp. 5–32, 2001.

[39] M.-H. Roy and D. Larocque, "Robustness of Random Forests for Regression," *J. Nonparametr. Stat.*, vol. 24, pp. 993–1006, 2012.

[40] A. M. Prasad, L. R. Iverson, and A. Liaw, "Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction," *Ecosystems*, vol. 9, no. 2, pp. 181–199, 2006.

[41] J. Yue, H. Feng, G. Yan, and Z. Li, "A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass using Near-Surface Spectroscopy," *Remote Sens.*, vol. 10, no. 1, p. 66, 2018.

[42] F. E. Rex *et al.*, "Comparison between Random Forest and Linear Regression for Tropical Forest Aboveground Biomass Estimation," in *XIX Brazilian Symposium on Remote Sensing*, 2019.

[43] H. S. Grave, *Forest Mensuration*. New York: John Wiley & Sons, Inc., 1906.

[44] R. Ozçelik, H. V. W. JR., and J. R. Brooks, "Estimating Log Volumes of Three Tree Species in Turkey by Six Formulae," *For. Prod J J.*, vol. 56, no. November/December, pp. 84–86, 2006.

[45] N. B. Karayiannis and A. N. Venetsanopoulos, *Artificial Neural Networks*. Springer, 1993.

[46] B. A. Walther and J. L. Moore, "The Concepts of Bias, Precision and Accuracy, and Their Use in Testing the Performance of Species Richness Estimators, with a Literature Review of Estimator Performance," *Ecography (Cop.)*, vol. 28, no. 6, pp. 815–829, 2005.