

RESEARCH

Open Access



Innovative deep learning artificial intelligence applications for predicting relationships between individual tree height and diameter at breast height

İlker Ercanlı

Abstract

Background: Deep Learning Algorithms (DLA) have become prominent as an application of Artificial Intelligence (AI) Techniques since 2010. This paper introduces the DLA to predict the relationships between individual tree height (ITH) and the diameter at breast height (DBH).

Methods: A set of 2024 pairs of individual height and diameter at breast height measurements, originating from 150 sample plots located in stands of even aged and pure Anatolian Crimean Pine (*Pinus nigra* J.F. Arnold ssp. *pallasiana* (Lamb.) Holmboe) in Konya Forest Enterprise. The present study primarily investigated the capability and usability of DLA models for predicting the relationships between the ITH and the DBH sampled from some stands with different growth structures. The 80 different DLA models, which involve different the alternatives for the numbers of hidden layers and neuron, have been trained and compared to determine optimum and best predictive DLAs network structure.

Results: It was determined that the DLA model with 9 layers and 100 neurons has been the best predictive network model compared as those by other different DLA, Artificial Neural Network, Nonlinear Regression and Nonlinear Mixed Effect models. The alternative of 100 # neurons and 9 # hidden layers in deep learning algorithms resulted in best predictive ITH values with root mean squared error (RMSE, 0.5575), percent of the root mean squared error (RMSE%, 4.9504%), Akaike information criterion (AIC, -998.9540), Bayesian information criterion (BIC, 884.6591), fit index (FI, 0.9436), average absolute error (AAE, 0.4077), maximum absolute error (max. AE, 2.5106), Bias (0.0057) and percent Bias (Bias%, 0.0502%). In addition, these predictive results with DLAs were further validated by the Equivalence tests that showed the DLA models successfully predicted the tree height in the independent dataset.

Conclusion: This study has emphasized the capability of the DLA models, novel artificial intelligence technique, for predicting the relationships between individual tree height and the diameter at breast height that can be required information for the management of forests.

Keywords: Artificial intelligence, Prediction, Deep learning algorithms, Individual tree height

Correspondence: ilkerercanli@karatekin.edu.tr
Çankırı Karatekin University, Faculty of Forestry, 18200 Çankırı, Turkey



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

The significant components of forest inventory, which is the first phase of forest planning, are the measurement of the individual tree heights (ITH) and the diameter at breast height (DBH). These individual tree attributes are used to predict total and merchant volume and biomass, forest site index, especially for uneven-aged stand, and also these attributes have the roles of significant input and independent variable in yield and growth models (Vanclay 1994; Kv and Hui, 1999). The measurements of the individual tree heights are more difficult and time consuming than those of DBH (Huang et al. 1992; Martin and Flewelling 1998) and so the ITH of all trees in sampling units cannot be measured in forest managements (Loetsch et al. 1973; Van Laar and Akça 2007). The ITH, whose could not be measured in forest inventory applications, can be predicted by the stand height curves which show the statistical relationships between the ITH and DBH (Avery and Burkhart 1983; Van Laar and Akça 2007).

In forest biometric studies, the empirical relationships between the ITH and DBH are represented by the statistical equations and these relations are modelled by using Nonlinear Regression Models (NLRM) owing to the sigmoid or “S” shaped trend to be evident to these ITH and DBH relations (Wykoff et al. 1982; Huang et al. 1992; Robinson and Wykoff 2004). In forest areas, various stand growing conditions with stand age, site quality and stand stocking have significant effects on the relationships between the ITH and DBH. Thus, the models with only an independent variable such as DBH can remain incapable in successful and effective predicting these relations. In this regard, different statistical prediction techniques have been proposed and used in modelling these relationships between the ITH and DBH that were sampled from different stand growing structures. Ferguson and Leech (1978), Krumland and Wensel (1988), Larsen and Hann (1987) and Parresol (1992) proposed an approach which comprises the prediction of parameter values of these regression models separately for different stand structures at first phase and subsequently developed linear regression models for the relationships between the parameters of this regression model and some stand attributes such as stand age, site index and stocking index at second phase. As a more common approach, the multivariate nonlinear regression models which comprise various stand attributes such as stand basal area, site index, stand age or stocking index in addition to the DBH were developed by various studies such as Huang et al. (2000), Sharma and Zhang (2004), Temesgen and Gadow (2004), Dorado et al. (2005), Trincado et al. (2007), Adame et al. (2008), Paulo et al. (2011). These multivariate ITH models with supplemental stand attributes are also called as “*generalized height-*

diameter models”. Nanos et al. (2004) analyzed the spatial pattern of the height models and offered the “geostatistical” modelling.

Another statistical modelling technique that has been used widely to predict the ITH in forestry literature is the Nonlinear Mixed Effect Regression Modelling Approach. This regression modeling technique has been frequently used in modeling empirical relationships between the ITH and DBH, because hierarchically correlated data with clustered and hierarchically sample plots that have been measured to develop the ITH-DBH models may cause serious fitting problem in modeling these relations (Dorado et al. 2006; Sharma and Parton 2007). These hierarchical data structures can be evident in the sample plots measured from the stands with different growing structures owing to different stand site quality, stocking and stand age (Calama and Montero 2004; Budhathoki et al. 2008). These highly correlated data violated the assumption of independence of data which is one of the basic assumptions in developing the regression models. The violation of this assumption is called as “*autocorrelation*” or “*serial correlation*” (Littell et al. 1996; Lappi 1997). The usage of approaches of nonlinear regression models, especially for the hierarchical data structures, causes biased predictions of the confidence intervals of model parameters in regression models (Searle et al. 1992; Grégoire et al. 1995). This situation negatively affects the reliability of the results of the regression models and as a result incorrect results can be obtained in the height predictions. Especially, in forest studies including the development of the ITH models, the Nonlinear Mixed Effect (NLME) Regression Models have been commonly proposed and used as a solution to deal with this “autocorrelation” problem (Calama and Montero 2004; Mehtätalo 2004; Lynch et al. 2005; Dorado et al. 2006; Sharma and Parton 2007; Trincado et al. 2007; Adame et al. 2008; Budhathoki et al. 2008; Crecente-Campo et al. 2010).

Beside these statistical modeling techniques with the NLRM and NLME, Artificial Neural Networks (ANN), which is a part of Artificial Intelligence (AI), have become popular as another modelling methodology for predicting the individual tree and stand yield and growth. Especially, significant studies related to the ANN have been conducted since the beginning of 2000s. Numerous prediction models based on AI, especially ANNs, have been developed for modeling various individual tree and stand attributes such as tree volume (Diamantopoulou 2005a, 2006; Özçelik et al. 2008; Diamantopoulou and Milios 2010; Özçelik et al. 2010; Soares et al. 2011; Miguel et al. 2016), tree taper (Diamantopoulou 2005b; Leite et al. 2011; Nunes and Görgens 2016), tree height (Diamantopoulou and Özçelik 2012; Özçelik et al. 2013), tree mortality (Hasenauer et al. 2001), survival model (Guan and Gertner 1991),

regeneration establishment and height growth (Hase-nauer and Kindermann 2002), bark volume (Diamantopoulou 2005a), biomass prediction (Özçelik et al. 2017), basal area and volume increment growth model (Ashraf et al. 2013). In addition to many ANN studies, Deep Learning Algorithms (DLA) stand out as another prominent AI technique. Although there are a number of significant ANN studies predicting the yield and growth of tree and stand in forestry literature, the DLA models seem to be an innovative technique in front of forest biometrics since 2010. Especially, DLA can be successfully used in analyzing the data clouds (structures which consist of millions or billions of data) and in data mining. The DLA models are basically multi-layer ANN models with at least 3 hidden layers, and this artificial intelligence technique tries to approach the learning and decision-making capacity of the human brain to a certain extent with its complex structure that can contain 5–10 or tens of layers and hundreds and thousands of neurons. Although there is a certain number of studies consisting of ANN for forest yield and growth predictions in today's literature, the modelling studies with DLA models are in the beginning phase. With the development of computer systems which consist of highly effective graphic processing units, DLA models become more applicable and accessible in today's world. Its application examples such as the diagnosis of plant illnesses and plant specification have been conducted in agriculture areas (Lee et al. 2015; Mohanty et al. 2016; Sladojevic et al. 2016; Carranza-Rojas et al. 2017; Sun et al. 2017; Ferentinos 2018; Ubbens et al. 2018). Furthermore, there is a need to address the evaluation of new AI techniques for investigating the capability and obtainability in predicting tree and forest attributes that have been important in forest management applications. According to the knowledge of the forest biometric studies including growth and yield models, no studies have been achieved to develop the DLA models to predict individual tree attributes, especially tree height and so the issue of the capability of DLA in predicting tree attributes has been uncertain and needs to be clarified. By widespread of AI techniques such as the DLA models, these scientific evaluations based on the comparative methods have been received remarkable interest and require further modelling studies in forest literature. In this study, it is aimed to evaluate the capability of the usability of the DLA models in predicting empirical relationships between the ITH and DBH as a leading and innovative application. To that end, (1) the DLA models in order to predict relationships between the ITH and DBH measured from stands with different growing structures were trained, (2) the success status of these predictions obtained from DLA models was compared with those of nonlinear regression (NLRM) models,

nonlinear mixed effect regression (NLME) and artificial neural network (ANN) models, and (3) the ideal and optimal DLA model structure in the prediction of the ITH was decided by comparing the DLA network structures with various numbers of layers and neurons alternatives. Thus, this study presents scientific the clarification about the issue of whether DLA models can be evaluated as an alternative technique for statistical methods in predicting the individual tree height.

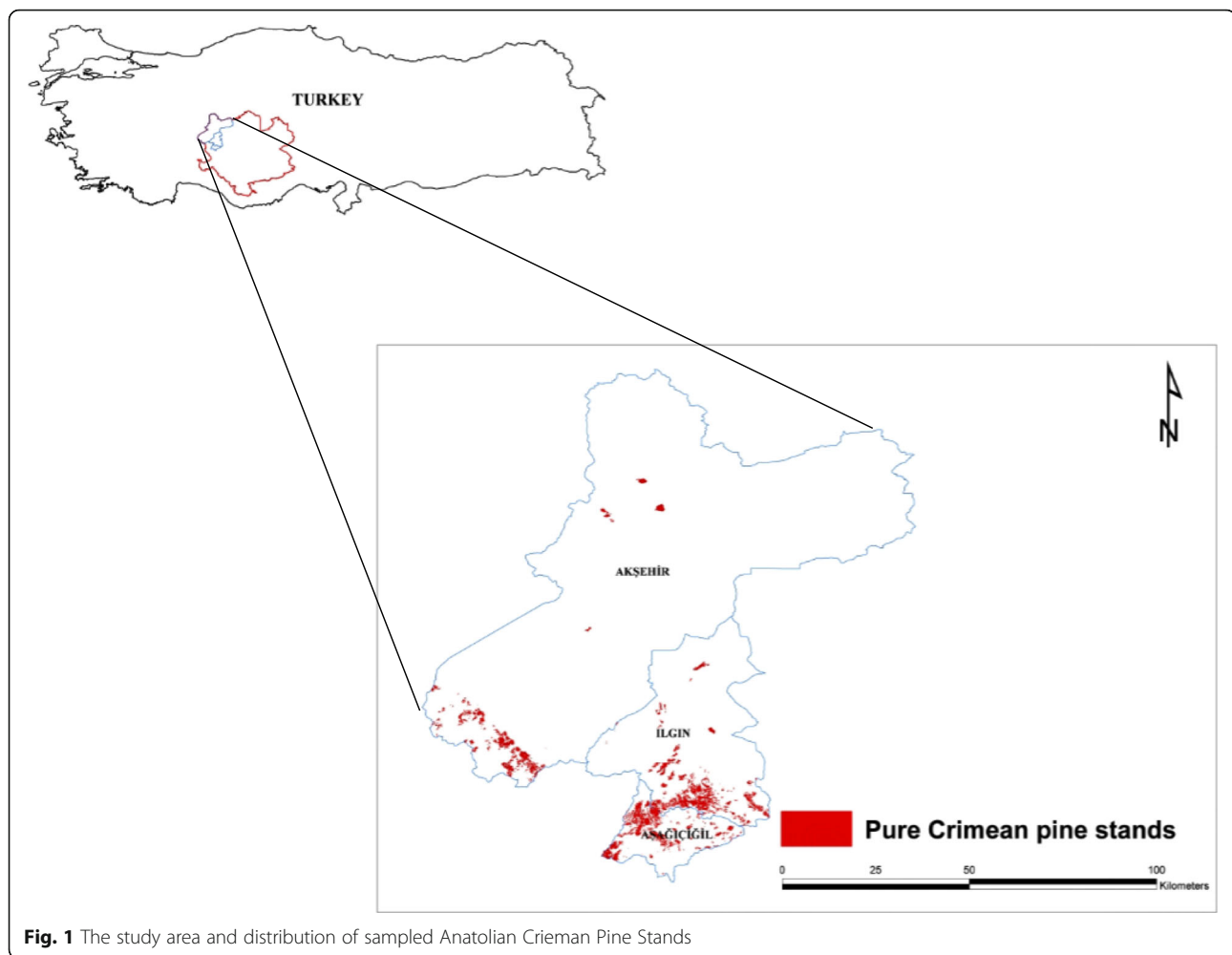
Materials and methods

Materials

In this study, the research material was the data obtained from temporary 150 sample plots as a result of measuring of stands of even aged and pure Anatolian Crimean Pine (*Pinus nigra* J.F. Arnold ssp. *pallasiana* (Lamb.) Holmboe) in Konya Forest Enterprise. The studied Pure Anatolian Crimean Pine Stands covered Akşehir, Ilgın and Aşağıcigil Forest district areas (Fig. 1). This tree species is the most common and dominant species in this region, and so this species of Anatolian Crimean Pine was selected to model relationships between the ITH and DBH as the particular tree species. The characteristic of these studied stands is even aged and pure forest stands with the dominant species of Anatolian Crimean Pine. The altitudes of studied area varied from 250 to 1050 m and the slope ranged between 5% and 60%. The areas studied were characterized geomorphologically as being high mountainous land with moderate and steep slopes. The mean annual temperature is between -5.8°C and 24.8°C , respectively. The climatic regime is a typical a semi-arid continental climate characterized by hot, dry summers and cold, snowy winters. Most of the region usually has low precipitation throughout the year. The mean annual rainfall varies from 400 to 850 mm with a relatively homogeneous precipitation.

These sample plots were selected by random sampling in terms of different stand age, site quality and density. The sample plots were in the shape of a circle and their size varied from 400 to 800 m² depending on the structure of the stand. At each sample plot, DBH was measured to 0.1 cm precision using calipers at every living tree with a DBH > 8 cm. Individual Tree height (ITH) was measured in a subset of trees, selecting two-three trees for each of 4 cm diameter class using Blume–Less Altimeter (0.1 m precision). In addition, the ITH and DBH measurements were obtained from dominant and co-dominant trees, which were selected based on the 100 dominant and co-dominant highest trees per unit area (e.g. Four highest trees in a 0.04-ha plot).

Totally, 2024 pairs of height-diameter measurements were obtained by the measurement, which realized in these sample plots. These data were divided into two groups randomly in order to use in the training of DLA and ANN models and in developing NLRM and NLME models (1st



group data set) and in validating ITH predictions obtained by these methods (2nd group data set). There are approximately 85% (1720 sample trees) of the total data in the 1st group and approximately 15% (304 sample trees) of the total data in the 2nd group. Various statistical information related to the data is provided in Table 1.

Methods

Nonlinear regression models (NLRM)

In order to model the empirical relationships between the ITH and DBH obtained from the stands in different growing structure, various regression models including various stands attributes, further to DBH, have been proposed and used (Huang et al. 1992; Fang and Bailey 1998; Peng 1999; Temesgen and Gadow 2004). Peng et al. (2001) expressed that some model attributes such as the number of parameters, the biological explanation and the validity of model prediction of these models. The model which was chosen to model the relationships between the ITH and DBH possesses some mathematical characteristics such as (i) monotonic increment (ii) inflection point and (iii) horizontal

asymptote (Peng et al. 2001). Therefore, seven commonly used functions (M1, M2, M3, M4, M5, M6 and M7) were selected to model the relationships between the ITH and DBH and develop generalized height–diameter model (Table 2). These tested ITH-DBH functions which were proposed by Meyer (1940) modified by Cañadas et al. (1999) (M1), Loetsch et al. (1973) modified by Cañadas et al. (1999) (M2), Prodan (1965) modified by Tomé (1989) (M3), Hui and Kv (1993) (M4), Soares and Tomé (2002) (M5), Richards (1959) modified by Sharma and Parton (2007) (M6) and Schnute (1981) modified by Dorado et al. (2006) (M7), have desirable characteristics such as asymptotic with inflection point models. They are biologically reasonable and can provide biological growth curves. These ITH-DBH functions were chosen owing to their desired properties and to commonly preferred in numerous studies modeling relationships between ITH and DBH.

The nonlinear functions were fitted using the 1st group data set (1720 trees). Based on the Nonlinear Least Squares (NLS), which uses the Levenberg-Marquardt algorithm, the parameters of ITH-DBH functions was

Table 1 Summary statistics of the sample trees originated for 1st group and 2nd group data set

Variables		Min.	Max.	Mean	Std. Deviation
Diameter at breast height (<i>d</i> , cm)	1st group data set	10.00	67.00	29.41	7.69
	2nd group data set	11.00	63.00	29.01	9.84
Total tree height (<i>h</i> , m)	1st group data set	5.00	18.00	11.26	2.35
	2nd group data set	5.00	17.00	10.84	2.51
Dominant diameter (<i>d</i> ₀ , cm)	1st group data set	15.00	67.00	34.21	8.51
	2nd group data set	15.00	63.00	33.29	9.67
Dominant height (<i>H</i> ₀ , m)	1st group data set	6.00	18.00	12.08	2.26
	2nd group data set	6.00	18.00	11.60	2.44
Quadratic mean diameter (<i>D</i> _g , cm)	1st group data set	11.16	51.91	23.85	6.65
	2nd group data set	11.16	51.91	23.49	7.46
Number of trees of sample plots per ha (<i>N</i> , #·ha ⁻¹)	1st group data set	112.50	1575.00	588.62	303.49
	2nd group data set	112.50	1575.00	581.69	293.47
Basal area of sample plots per ha (BA, m ² ·ha ⁻¹)	1st group data set	2.37	55.09	23.34	9.12
	2nd group data set	2.37	45.03	21.87	8.60

predicted by using NLS package available in the R statistical environment (R Development Core Team 2018).

Nonlinear mixed effect (NLME) regression models

To deal with this “autocorrelation” problem originating from the hierarchical data structures, a Nonlinear Mixed Effect (NLME) modeling procedure was applied to the best predictive height–diameter model by simultaneously predicting both fixed and random parameters. Different from the NLRM, the model parameters of the NLME are divided into two groups as fixed effects and random effect parameters in its model structure. While the fixed effect parameter reveals ITH trend which is common to overall stands, the random effect parameter represents the variance between the stands and defines the variability in the relationships between the ITH and DBH along various stands (Lappi 1997; Calama and Montero 2004; Mehtätalo 2004; Dorado et al. 2006; Crecente-Campo et al. 2010).

The procedure of NLME package available in the R statistical environment, which is based on the Maximum Likelihood method, was used to obtain the parameter predictions of NLME that presents the best predictive height–diameter model. To decide the best predictive random-fixed parameter alternative for model structure based on NLME, the ITH-DBH models including one, two or three random parameters were fitted and compared based on some statistical comparison criteria. The adaptive Gaussian quadrature was used in the computation of the integral over the random effects as described by Pinheiro and Bates (2000). Furthermore, this NLME procedure was performed assuming the homogenous within-tree variance and uncorrelated residuals.

Artificial neural network models

Being as Artificial Intelligence (AI) prediction technique, Artificial Neural Network (ANN) based on the Feed Forward Backprop (FFB) and Cascade Correlation (CC) training algorithms with training function of Levenberg-

Table 2 The nonlinear ITH-DBH functions tested by this study

Function number	Mathematical form	References
(M1)	$ITH = 1.3 + (H_0 - 1.3) \frac{(1 - e^{-b_0 d})}{(1 - e^{-b_0 d_0})}$	Meyer (1940) modified by Cañadas et al. (1999)
(M2)	$ITH = 1.3 + (b_0 (\frac{1}{d} - \frac{1}{d_0}) + (\frac{1}{H_0 1.3})^2)^{-2}$	Loetsch et al. (1973) modified by Cañadas et al. (1999)
(M3)	$ITH = 1.3 + (H_0 - 1.3) (1 + b_0 (H_0 - 1.3) (\frac{1}{d} - \frac{1}{d_0}))^{-1}$	Prodan (1965) modified by Tomé (1989)
(M4)	$ITH = 1.3 + b_0 H_0^{b_1} d^{b_2} H_0^{b_3}$	Hui and Kv (1993)
(M5)	$ITH = H_0 (1 + (b_0 + b_1 H_0 + b_2 Dg) e^{b_3 H_0}) (1 - e^{-\frac{b_4 d}{H_0}})$	Soares and Tomé (2002)
(M6)	$ITH = 1.3 + (b_0 H_0^{b_1}) (1 - e^{-b_2 (\frac{d}{H_0})^{b_3}})^{b_4}$	Richards (1959) modified by Sharma and Parton (2007)
(M7)	$ITH = (1.3^{b_0} + (H_0^{b_0} - 1.3^{b_0}) \frac{(1 - e^{-b_1 d})}{(1 - e^{-b_1 d_0})})^{\frac{1}{b_0}}$	Schnute (1981) modified by Castedo Dorado et al. (2006)

*b*_i: regression parameters to be predicted by model

Marquardt were used to model the relationship between ITH and DBH. These training algorithms including the Feed Forward Backprop (FFB) and Cascade Correlation (CC) have commonly been used to predict tree and forest attributes in forest literature. The reason for choosing these training functions from different training algorithms is its intensive use in forestry. When training the ANN models with FFB and CC algorithms, the individual tree height values, ITH were predicted as target variable. In these ANN models, DBH and the best predictor variables selected from preliminary analyses including a trial and error procedure using different combinations of these stand attributes, such as basal area, number of trees of sample plots, quadratic mean diameter, the dominant DBH and ITH of the sample plot, were used as input variables. The standard ANN models can include three layers such as input layer, hidden layer and output layer. Especially, the activation functions including Hyperbolic tangent sigmoid (tan-sig), logistic sigmoid function (log-sig) and linear function (Pure-lin) connect these network layers in ANN structures. These activation function alternatives have significant effects on fitting performance of neural network. In this study, alternatives including some activation functions in the connection between input, hidden and output layers were compared to decide the best predictive one: (A1) tan-sig function between input layer and hidden layer and tan-sig function between hidden layer and output layer, (A2) tan-sig function between input layer and hidden layer and log-sig function between hidden layer and output layer, (A3) tan-sig function between input layer and hidden layer and pure-lin function between hidden layer and output layer, (A4) log-sig function between input layer and hidden layer and log-sig function between hidden layer and output layer, (A5) log-sig function between input layer and hidden layer and tan-sig function between hidden layer and output layer, (A6) log-sig function between input layer and hidden layer and pure-lin function between hidden layer and output layer, (A7) pure-lin function between input layer and hidden layer and pure-lin function between hidden layer and output layer, (A8) pure-lin function between input layer and hidden layer and log-sig function between hidden layer and output layer and (A9) pure-lin function between input layer and hidden layer and tan-sig function between hidden layer and output layer. Other important parameter of the network structure is the number of neurons in hidden layer. Thus, some alternatives for the number of neurons which ranged from 1 to 100; 1, 2, 3, ..., 20, 30, 50, 70, 90 and 100 number of neurons were compared to select the best predictive neuron alternative in this study. As a result, a total of 900 network alternatives including 100 number neurons and 9 transfer function alternatives ($100 \times 9 = 900$

alternatives) based on the Feed Forward Backprop (FFB) and Cascade Correlation (CC) training algorithms, totally 1800 alternatives for FFB and CC-ANN models, were trained and used to obtain the ITH predictions. Being as other significant parameters for ANN structure, the value of 3000 for epochs, the value of 1×10^{-10} for performance goal, the value of 1×10^{-10} for Minimum performance gradient and 1×10^{-8} for epsilon gave the best predictive results to train these FFB and CC-ANN models in the preliminary of this study and so, these parameters were used to obtain the ITH predictions and to compare with those by other predictions methods such as NLRM, NLME and DLA models. These network trainings for 1800 network alternatives for FFB and CC-ANN models were carried out using *newff* syntax for the feed-forward backpropagation network and *newcf* syntax for the cascade-forward backpropagation network coded in MATLAB software (MATLAB 2014).

Deep learning algorithms

Deep Learning Algorithm (DLA) models are an artificial intelligence technique which has remained on the agenda since 2010. The DLA has shown quite successful results in various applications such as image classification, video analysis, speech recognition, natural language learning process in recent time. The Artificial Neural Network (ANN) models which is another Artificial Intelligence (AI) type have been usually developed to the input layer, hidden layer (two hidden layers in some cases), output layers in its model structure. However, the DLA models have a quite complex structure comprising many (5, 10 or even tens of) hidden layers compared with ANN models. Especially, the use of Graphics Processing Units (GPU) of computers in the training of DLA models provides this DLA model to be more accessible and usable with effective and successful results in various applications, especially visual and speech recognition by modern day computer technologies, that have not been seen in the history of humanity. In addition to all these successful and efficient use of DLA models in computer systems, the use of DLA models in forestry applications, specifically some tree and forest attributes prediction practices, has been quite limited.

As the calculations in applications of DLAs are quite complex and intense, obtaining the predictions for tree and forest attributes by DLA model requires intensive use of computer software. Despite some DLA applications and platforms were developed in various languages, the H₂O package (R Development Core Team 2018), which has been coded in R software language, becomes prominent in terms of its characteristics such as user-friendly, the ability of finding a solution and comparison of different network alternatives. Basically, the H₂O package, which operates on the R software platform, is

an open source coded artificial intelligence library and comprises different artificial intelligence applications such as “Generalized Linear Models”, “Gradient Boosting Machines”, “Random Forests”, “Deep Neural Networks (Deep Learning)”, “Stacked Ensembles”, “Naive Bayes”, “Cox Proportional Hazards”, “K-Means”, “PCA” and “Word2Vec” (R Development Core Team 2018).

In this study, the H₂O package was used to train the network models based on the DLA models which predicted the individual tree height values, ITH (target variable). In order to determine the input variables in DLA model structure, the trial and error method were used by comparing some alternatives including various independent variables such as the DBH and stand attributes similar to variable determination method in ANN models. Also, the network parameters such as number of layers, number of neuron and type of transfer function are important attributes that affect the success of prediction results in obtaining the predictions with the DLAs in these training DLAs. From various transfer functions, the “Rectifier” function was selected as a transfer function in DLAs’ structure owing to its successful fitting results in our preliminary analyses. The H₂O package uses the adaptive learning rate algorithm (ADADELTA) in the trainings of DLA (Zeiler 2012). The rho describes the rate of ADADELTA and epsilon expresses learning rate for DLA models. In the present study, the value of 0.999 for rho and 1 × 10⁻⁸ for epsilon were used to train DLA models. Also, the value of 1000 for the epochs, the number of iterations to be accepted in training networks, was used in the training of DLA models, since the best predictive results have been obtained with 1000 in various neural network studies. As a training algorithm, the Gradient Descent Function with the Gaussian distribution model based on the Mean Squared Error function type was used.

In addition to these parameters of DLAs, the number of hidden layers and the number of neurons in these hidden layers are network parameters that need special attention in training DLA models. In training DLA models, 8 numbers of hidden layers (with 3, 4, 5, 6, 7, 8, 9 and 10 layers) starting from 3 layers (that is the minimum number of layers of DLA) to 10 layers and 10 different neuron alternatives ranging from 10 to 100 by increasing 10 at each step (10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 neurons) were considered as important network parameters. Thus, 80 different DLA models, 8 different numbers of hidden layers and 10 different numbers of neuron alternatives, were trained to obtain the predictions of ITH.

The K-Fold Cross Validation method was used in trainings of DLA models, because this method “cross validation up to k number” may reduce “overfitting errors” in obtaining the predictions by the network models. In this study,

the value of “cross validation up to k number” was applied on the basis of as (nfolds = 10) with the “nfolds” parameter of the H₂O.ai Team package (R Development Core Team 2018).

Comparison criteria

In this study, various statistical fitting criterion values were used to compare and evaluate the predictions of ITH that were obtained by the NLRM, NLME, FFB-ANN and CC-ANN and DLA models. These fitting criteria are (1) average absolute error (AAE), (2) the maximum absolute error (max. AE), (3) the root mean squared error (RMSE), (4) % root mean squared error (RMSE%), (5) the average Bias (Bias), (6) % average Bias (Bias%), (7) the fit index (FI), (8) Akaike’s information criterion (AIC) and (9) Bayesian information criterion (BIC). These criteria are calculated as follows:

$$r = \left(\frac{\sum_{i=1}^n [(ITH_i - \overline{ITH}_i) \cdot (\widehat{ITH}_i - \overline{ITH}_i)]}{\sqrt{\sum_{i=1}^n (ITH_i - \overline{ITH}_i)^2 \cdot \sum_{i=1}^n (\widehat{ITH}_i - \overline{ITH}_i)^2}} \right) \tag{1}$$

$$AAE = \sum_{i=1}^n |ITH_i - \widehat{ITH}_i| / n \tag{2}$$

$$max.AE = \max(|ITH_1 - \widehat{ITH}_1|, \dots, |ITH_n - \widehat{ITH}_n|) \tag{3}$$

$$RMSE = \sqrt{\sum_{i=1}^n (ITH_i - \widehat{ITH}_i)^2 / (n-k)} \tag{4}$$

$$RMSE\% = \left(\left[\sqrt{\sum_{i=1}^n (ITH_i - \widehat{ITH}_i)^2 / (n-k)} \right] / \overline{ITH}_i \right) \cdot 100 \tag{5}$$

$$Bias = \sum_{i=1}^n (ITH_i - \widehat{ITH}_i) / n \tag{6}$$

$$Bias\% = \left(\left[\sum_{i=1}^n (ITH_i - \widehat{ITH}_i) / n \right] / \overline{ITH}_i \right) 100 \tag{7}$$

$$FI = \frac{\sum_{i=1}^n (ITH_i - \widehat{ITH}_i)}{\sum_{i=1}^n (ITH_i - \overline{ITH}_i)} \tag{8}$$

$$AIC = \ln(RMSE) + 2k \tag{9}$$

$$BIC = \ln(RMSE) + \ln(k) \tag{10}$$

where, ITH_i is the measured individual total height value in the sample plot (observed values), \overline{ITH}_i is the average of observed individual total height values, \widehat{ITH}_i is the predicted individual total height value obtained by NLRM, NLME, FFB-ANN and CC-ANN and DLA models, k is the number of inputs or independent variable in the prediction methods, ln is the natural logarithm with the base of the mathematical constant e. From these fitting criterion values, it is desired that the

fit index (FI), which is between 0 and 1, should be as close to 1 as possible. Smaller values of other criterion values indicate that better predictive ITH are obtained. In order to evaluate all these ten criteria together, Relative Rank Methods proposed by Poudel and Cao (2013) were used and the values of relative rank were calculated to these all prediction methods with NLRM, NLME, FFB-ANN and CC-ANN and DLA models. Especially, after the calculation of the rank values, the prediction method with the smallest rank value was chosen as the best predictive method in ITH prediction from these methods.

This study has carried out two-level comparisons in evaluating many prediction models including NLME models including one (five alternatives), two (ten alternatives) and three (ten alternatives) random parameters (non-convergence fitting results for four and five random parameter alternatives) for the best predictive function from seven ITH-DBH functions tested, 900 # FFB-ANN models and 900 # CC-ANN models including 100 number neurons and 9 transfer function alternatives and 80 # DLA models including 8 different numbers of hidden layers and 10 different numbers of neuron alternatives. This two-stage evaluation process was carried out to determine the best predictive one from different prediction methods: (1) Firstly, the performance of ITH predictions obtained by the NLME (different random and fixed effect parameter alternatives), DLA (80 different models), FFB-ANN (900 different models) and CC-ANN (900 different models) methods were compared based on the Relative Rank Values proposed by Poudel and Cao (2013) within each prediction methods, (2) in the second stage, the best predictive model alternative at each AI model level such as DLA, FFB-ANN and CC-ANN, and NLME including various random and fixed effect parameters were compared with those by NLR. Thus, it has been possible to evaluate about 1900 various model alternatives obtained by various modeling techniques such as NLRM, NLME, DLA, FFB-ANN and CC-ANN and to determine the best predictive model.

The validation of prediction methods

The ITH predictions obtained by NLRM, NLME, FFB-ANN and CC-ANN and DLA models were further evaluated by using independent data including 304 trees which were not used in the development of regression models, NLRM and NLME, and in the trainings of FFB-ANN and CC-ANN and DLA models. After obtaining the ITH predictions with these methods with NLRM, NLME, FFB-ANN and CC-ANN and DLA models, the validation processes related to these prediction methods were applied by using the “Equivalence” test which is prominent in recent model evaluation processes. The evaluation process was carried out using notably an

independent data set to assess whether these ITH predictions obtained by the DLA models have the problem of “overfitting”, which this problem frequently occurs in the predictions of various tree attribute obtained by AI models. In this evaluation process, two one-sided test strategy (TOST) was used to test the equality of slopes (b_1) to $1 \pm 10\%$ and the equality of intercepts (b_0) to $\bar{y} \pm 10\%$. The predictions of the confidence intervals for these parameters were obtained by using a nonparametric bootstrap procedure, described in Robinson et al. (2005) and Robinson and Froese (2004), in which the number of bootstrap replicates was specified as 1000. These equivalence test procedures for different prediction methods were performed by using “Regression-based TOST using bootstrap, equiv. boot” function of the “equivalence” package in the R statistical environment (Development Core Team, 2018).

Results

In first level comparisons, the best predictive models from NLME, FFB-ANN, CC-ANN and DLA model alternatives were selected based on the Relative Rank Methods proposed by Poudel and Cao (2013) and these best predictive models of different prediction methods with NLME, FFB-ANN, CC-ANN and DLA were evaluated with NLRM models in second level comparison. As input variables in these AI models, the stand attributes including diameter at breast height (cm), the dominant height (h_0 , m), and (cm) (Dg) gave best predictive fitting results in FFB-ANN, CC-ANN models. Nevertheless, DLA models including diameter at breast height (d , cm), the dominant height (h_0 , m) and dominant diameter (d_0 , cm) as predictor variables resulted in best predictive ITH from various input variables.

As a result of the second level comparison, the fitting criteria of AAE, max. AE, RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for various prediction methods with NLRM, NLME, FFB-ANN, CC-ANN and DLA models are given in Table 3. The relative rank values (Poudel and Cao 2013) related to these goodness-of-fit criteria values and total relative rank values were shown in Table 4. In these fitting criterion values, RMSE ranged from 0.5575 to 0.8306, RMSE% ranged from 4.9504% to 7.3750%, AIC ranged from -998.9540 to -313.3060 , BIC ranged from 884.6591 to 1570.3072, FI ranged from 0.8749 to 0.9436, AAE ranged from 0.4077 to 0.6170, max. AE ranged from 2.3696 to 4.4859, Bias ranged from -0.0006 to -0.2695 and Bias% ranged from -0.0050 to -2.3927% . From various ITH-DBH functions tested, the function of Soares and Tomé (2002), M5, gave the best predictive fitting results a RMSE value of 0.7621, RMSE% value of 6.7672%, the AIC value of -461.2447 , BIC value of 1422.3685, FI of 0.8946 values, AAE value of 0.6132, max. AE value of 3.8927, Bias value of -0.005

Table 3 The goodness-of-fit statistics *r*, AAE, max. AE, RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for the best predictive DLA models with best predictive number of neuron alternative according to each hidden layer choices, the ITH-DBH functions based on NLRM, M5 based on NLME with *f* random, FFB-ANN and CC-ANN

AI alternatives and regression models	The number of hidden layers	RMSE	RMSE% (%)	AIC	BIC	FI	AAE	Max. AE	Bias	Bias% (%)
50 # neuron	3	0.6416	5.6974	-757.2286	1126.3846	0.9253	0.5051	2.3653	0.0701	0.6225
90 # neuron	4	0.6110	5.4257	-841.2650	1042.3481	0.9323	0.4620	2.3596	0.0391	0.3469
90 # neuron	5	0.5942	5.2766	-889.1954	994.4178	0.9359	0.4474	2.9231	-0.0111	-0.0988
100 # neuron	6	0.5859	5.2021	-913.6424	969.9708	0.9377	0.4346	2.5950	0.0447	0.3973
70 # neuron	7	0.5940	5.2747	-889.8224	993.7907	0.9360	0.4531	2.4365	0.0053	0.0468
90 # neuron	8	0.5798	5.1482	-931.5577	952.0554	0.9390	0.4336	2.3822	0.0263	0.2339
100 # neuron	9	0.5575	4.9504	-998.9540	884.6591	0.9436	0.4077	2.5106	0.0057	0.0502
80 # neuron	10	0.5892	5.2315	-903.9603	979.6528	0.9370	0.4286	2.5746	0.0512	0.4546
M1 NLRM		0.8193	7.2746	-336.8843	1546.7288	0.8782	0.6058	4.4859	-0.2465	-2.1891
M2 NLRM		0.8224	7.3027	-330.2519	1553.3612	0.8773	0.6102	4.3628	-0.2609	-2.3169
M3 NLRM		0.8306	7.3750	-313.3060	1570.3072	0.8749	0.6170	4.2243	-0.2695	-2.3927
M4 NLRM		0.7647	6.7905	-455.3473	1428.2659	0.8939	0.6064	4.0474	-0.0020	-0.0174
M5 NLRM		0.7621	6.7672	-461.2447	1422.3685	0.8946	0.6132	3.8927	-0.0005	-0.0047
M6 NLRM		0.7922	7.0343	-394.6698	1488.9433	0.8862	0.6254	4.1908	-0.0022	-0.0192
M7 NLRM		0.8158	7.2440	-344.1317	1539.4815	0.8793	0.6070	4.4574	-0.2540	-2.2551
NLME for M5 with <i>f</i> random		0.7073	6.2807	-589.5670	1294.0461	0.9092	0.5769	3.4091	-0.0006	-0.0050
FFB-ANN based on A3 activation function alternative and 85 # neuron		0.7160	6.3576	-568.6345	1314.9787	0.9070	0.5775	2.5863	-0.0175	-0.1557
CC-ANN based on A3 activation function alternative and 73 # neuron		0.7110	6.3132	-580.6896	1302.9236	0.9083	0.5638	2.9023	-0.0144	-0.1282

Table 4 The relative rank values of *r*, AAE, max. AE, RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for the best predictive DLA models with best predictive number of neuron alternative according to each hidden layer choices, the ITH-DBH functions based on NLRM, M5 based on NLME with *f* random, FFB-ANN and CC-ANN

AI alternatives and regression models	The number of hidden layers	RMSE	RMSE% (%)	AIC	BIC	FI	AAE	Max. AE	BIAS	BIAS% (%)	Σ
50 # neuron	3	6.854	6.854	7.698	7.698	6.057	9.495	1.051	5.916	5.916	57.538
90 # neuron	4	4.725	4.725	5.370	5.370	4.136	5.734	1.000	3.723	3.723	38.504
90 # neuron	5	3.556	3.556	4.042	4.042	3.121	4.467	6.035	1.748	1.748	32.315
100 # neuron	6	2.973	2.973	3.364	3.364	2.625	3.342	3.104	4.124	4.124	29.991
70 # neuron	7	3.541	3.541	4.024	4.024	3.108	4.963	1.687	1.335	1.335	27.559
90 # neuron	8	2.550	2.550	2.868	2.868	2.270	3.256	1.202	2.824	2.824	23.211
100 # neuron	9	1.000	1.000	1.000	1.000	1.000	1.000	2.349	1.362	1.362	11.073
80 # neuron	10	3.203	3.203	3.632	3.632	2.820	2.825	2.921	4.580	4.580	31.395
M1 NLRM		19.213	19.213	19.347	19.347	19.065	18.283	20.000	18.380	18.380	171.228
M2 NLRM		19.433	19.433	19.530	19.530	19.325	18.666	18.900	19.397	19.397	173.612
M3 NLRM		20.000	20.000	20.000	20.000	20.000	19.261	17.662	20.000	20.000	176.923
M4 NLRM		15.419	15.419	16.064	16.064	14.735	18.337	16.081	1.101	1.101	114.322
M5 NLRM		15.237	15.237	15.900	15.900	14.535	18.936	14.699	1.000	1.000	112.445
M6 NLRM		17.330	17.330	17.745	17.745	16.878	20.000	17.363	1.116	1.116	126.623
M7 NLRM		18.973	18.973	19.146	19.146	18.782	18.395	19.745	18.905	18.905	170.971
M5 NLME with <i>f</i> random		11.425	11.425	12.345	12.345	10.499	15.764	10.378	1.003	1.003	86.185
FFB-ANN based on A3 activation function alternative and 85 # neuron		12.027	12.027	12.925	12.925	11.117	15.814	3.025	2.201	2.201	84.263
CC-ANN based on A3 activation function alternative and 73 # neuron		11.679	11.679	12.591	12.591	10.759	14.623	5.849	1.983	1.983	83.736

and Bias% value of -0.0047% . This best predictive ITH-DBH function, M5, was selected to apply NLME procedure including some random-fixed evaluations based on the relative rank values. From these different random-fixed effect parameter alternatives, the NLME model of M5 with one random parameters, f parameter, and four fixed effect parameters, resulted in the best predictive fitting statistics with a RMSE value of 0.7073, RMSE% value of 6.2807%, the AIC value of -589.567 , BIC value of 1294.0461, FI of 0.9092 values, AAE value of 0.5769, max. AE value of 3.4091, Bias value of -0.0006 and Bias% value of -0.005% . According to these results obtained with NLRM and NLME prediction models, FFB-ANN and CC-ANN models have partially improved the prediction success. In ANN models, CC-ANN based on A3 activation function alternative and 85 # neuron and FFB-ANN based on A3 activation function alternative and 73 # neuron gave the best predictive fitting results with a RMSE values of 0.7110 and 7160, RMSE% values of 6.3132% and 6.3576%, the AIC values of -580.6896 and -568.6347 , BIC values of 1302.9236 and 1314.9787, FI values of 0.9083 and 0.9070, AAE values of 0.5638 and 0.5775, max. AE values of 2.9023 and 2.5863, Bias values of -0.0144 and -0.0175 and Bias% values of -0.1282% and -0.1557% , respectively. Nonetheless, the DLA models showed better predictive performance in explaining the variation in ITH and resulted lower total ranks (ranging from 11.073 to 57.538) than those by NLRM, NLME, FFB-ANN, CC-ANN (ranging from 83.736 to 176.923). In Tables 3 and 4, the results of DLA models with the best predictive number of neuron alternative according to each hidden layer choices from 80 various DLA models was presented. On the basis of the total relative rank values for these prediction methods, the DLA model structure with 9 hidden layers and 100 neurons showed the better predictive results in the prediction of the ITH than those by the other prediction models (Table 4). This DLA model structure alternative has a significant predictive ability with a RMSE value of 0.5575, RMSE% value of 4.9504%, the AIC value of -998.9540 , BIC value of 884.6591, FI of 0.9436 values, AAE value of 0.4077, max. AE value of 2.5106, Bias value of 0.0057 and Bias% value of 0.0502% compared to other prediction models.

Figure 1 showed the relationships obtained between observed and predicted height values by network models including (a) the M5 based on NLRM, (b) the M5 based with f random on NLME, (c) FFB-ANN based on A3 activation function alternative and 85 # neuron, (d) CC-ANN based on A3 activation function alternative and 73 # neuron, (e) DLA with 100 # neurons in nine hidden layers. When these graphs were examined (Fig. 2), it is seen that the best predictive DLA network model (DLA with 100 # neurons in nine hidden layers) evidenced more correlated

relationships between predicted and measured values around the 1:1-line than those for other prediction models with NLRM, NLME, FFB-ANN and CC-ANN. Thus, ITH predictions which were obtained by this best predictive DLA network model more precise than those of other prediction methods including NLRM, NLME, FFB-ANN and CC-ANN. These graphical results about predictive ability of this best predictive DLA network model were propped with the relationships between these residual and prediction values which were presented in Fig. 2. This graph (Fig. 3) presented random trend of residual around zero and no important relations, suggesting that there is no serious failure of homoscedasticity, violations of the assumption of constant variance, for those by this best predictive DLA model. For a further analysis of residuals of the best predictive DLA, NLRM, NLME, FFB-ANN and CC-ANN models, Fig. 4 presented the plot of residuals against lagged residuals by (a) the M5 based on NLRM, (b) the M5 based on NLME with f random, (c) the FFB-ANN based on A3 activation function alternative and 85 # neuron, (d) the CC-ANN based on A3 activation function alternative and 73 # neuron, (e) DLA with 100 # neurons in nine hidden layers. This plot shows a significant autocorrelation in residuals from the ITH prediction by NLRM of M5 function. A moderate improvement was obtained in predictions with the NLME of M5 function including f random parameters. This improvement about autocorrelation quite clearly obtained by this best predictive DLA model, give no trends in the lag-residuals, suggesting that no-autocorrelation problem was the case for the height predictions by this network model (Fig. 3e).

In this study, it was pointed out the effect of alternatives for different numbers of hidden layers and neurons on the fitting ability of the ITH predictions and so judged the ideal and optimal DLA model structure in these predictions. The results related to this evaluation are presented as the average fitting criteria of RMSE, RMSE%, AIC, BIC, FI and AAE according to the alternatives for different numbers of layers and neurons in Tables 5 and 6. When the changes of these fitting criterion values according to the alternatives for the number of layers and neurons, it was seen that there was a progress in the criteria values from the 3rd layer to 8th layer generally; however, there was a worsening in 7th, 9th and 10th layers in these prediction success values. On the other hand, it is seen that the increase in the number of neurons causes a general improvement in these fitting criteria, which except the number of 50 and 90 neurons.

The present study validated the NLRM, NLME, FFB-ANN, CC-ANN and DLA models to the independent data set by using "Equivalence" test and the results related to this test were shown in Table 7. Consistent with these analysis results, the h_0 hypothesis which pronounce that "the constant is different from 10.8421 cm

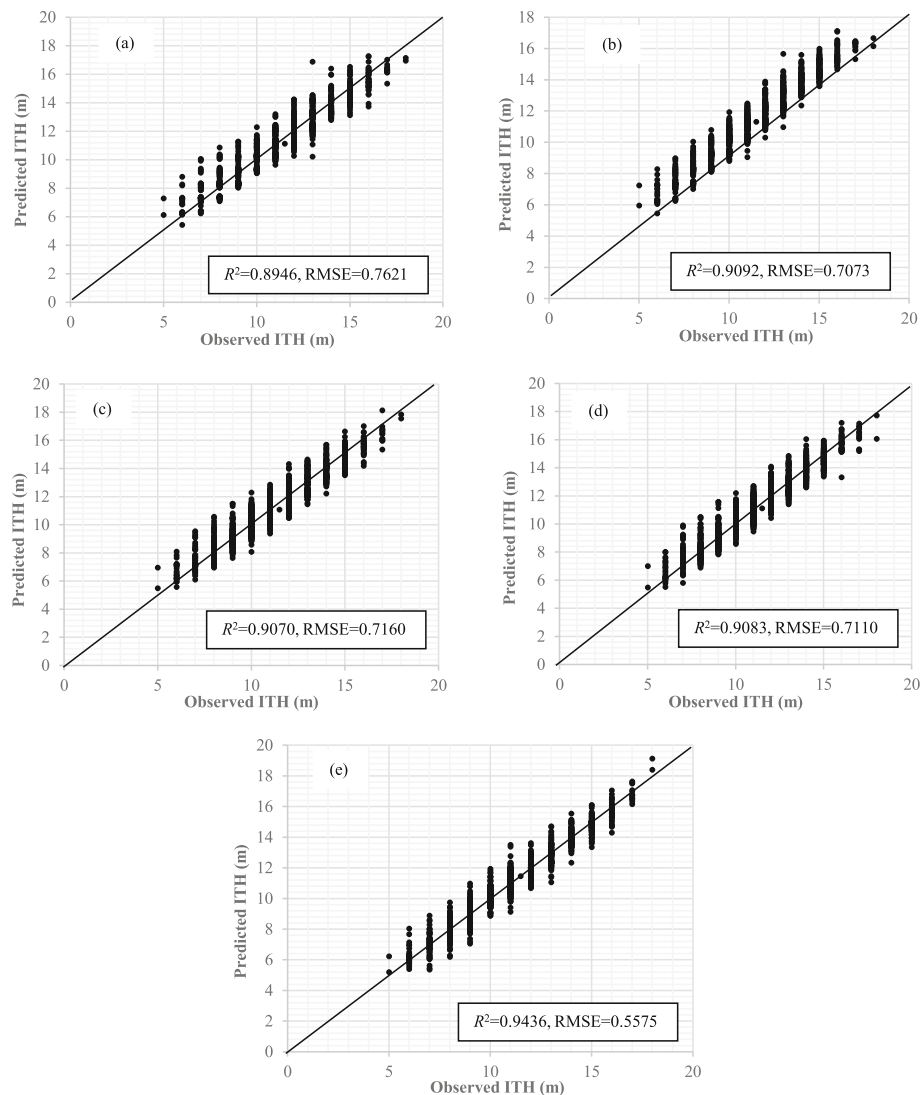


Fig. 2 The relationships between the observed and predicted ITH values obtained by (a) the M5 based on NLRM, (b) the M5 based with f random on NLME, (c) FFB-ANN based on A3 activation function alternative and 85 # neuron, (d) CC-ANN based on A3 activation function alternative and 73 # neuron, (e) DLA with 100 # neurons in nine hidden layers

(the average observed ITH values) and the slope coefficient (b_1) (except the DLA model with 8 layers) is different from 1" has been rejected. Thus, it can be concluded that the aforementioned DLA models (except the DLA with 8 layers) can be accepted and used statistically 95% in the ITH predictions of the stands in the study areas. Also, the fitting criteria values related to the prediction obtained with different DLA models from these 304 trees are shown in Table 8.

Discussion

This study is the first attempt to model individual tree height-diameter relationships by using Deep Learning Algorithms (DLA) that have been another application of Artificial Intelligence Techniques. The main topic of this

research is the question whether the DLA model, as an alternative, will offer predictive results compared as the classical regression models, which have been in use for many years in modelling the growth of trees, and ANN models, another type of AI technique. In addition, various network alternatives were evaluated to determine the optimal network structure based on the statistical criteria and, for this purpose, 80 different DLA models were trained by using the data collected from different forest stands. When considering the evaluation results based on the Relative Rank Methods (Poudel and Cao 2013) seen in Tables 3 and 4, these DLA models offer better statistical performance than those by the NLRM, NLME, FFB-ANN, CC-ANN and DLA models in the predictions of tree heights. Especially, the DLA network

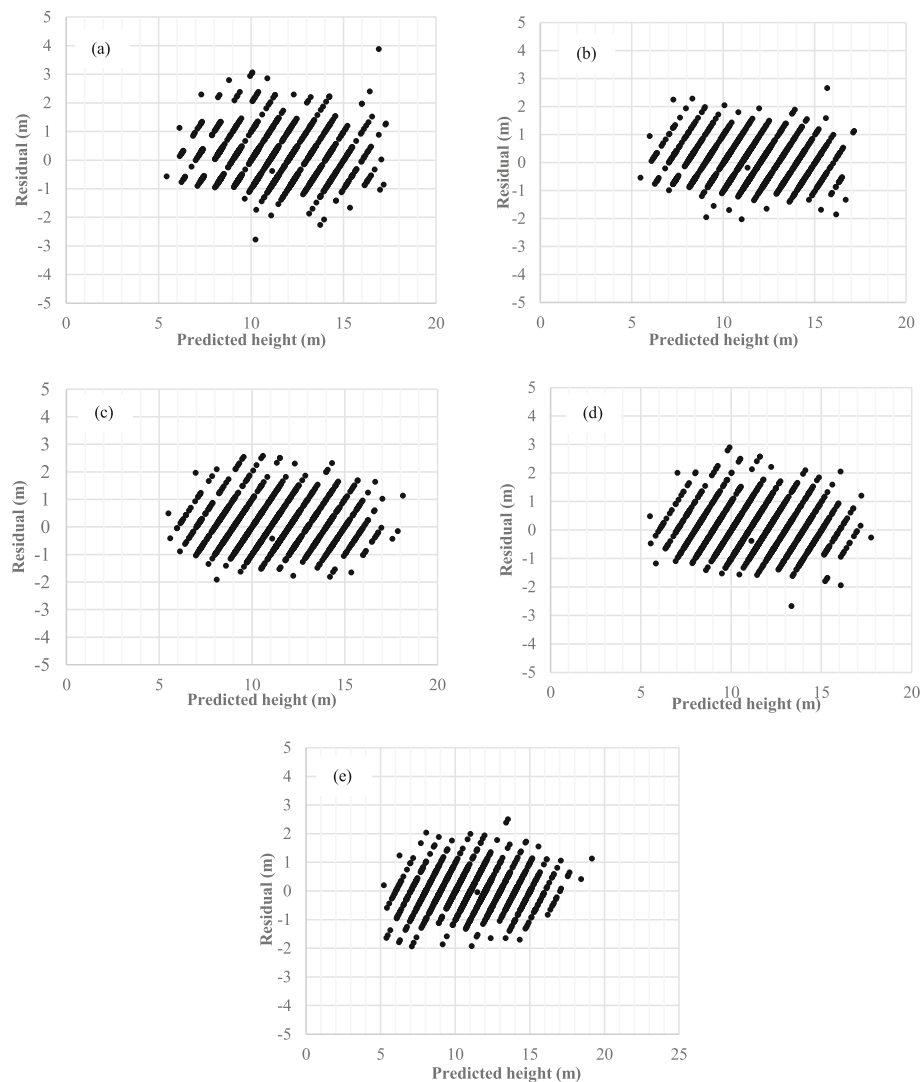


Fig. 3 The relationships between predicted (x-axis) and Residuals ITH (y-axis) obtained by the best predictive deep learning network models: **a** the M5 based on NLRM, **b** the M5 based on NLME with f random, **c** FFB-ANN based on A3 activation function alternative and 85 # neuron, **d** CC-ANN based on A3 activation function alternative and 85 # neuron, **e** DLA with 100 # neurons in nine hidden layers

model with 9 layers and 100 neurons resulted in the best predictive tree heights in this study. This DLA network model gave significant improvement in the values of RMSE, AIC, BIC, FI, AAE, max. AE with the rates of 26.85%, 116.58%, 37.80%, 5.48%, 33.52%, 35.51%, respectively, compared as those of NLRM.

Considering the predictive capability of ITH obtained by these DLA models, it can be observed that the DLA model with 9 layers and 100 neurons produced higher prediction precisions than those by the NLRM, NLME and FFB-ANN and CC-ANN (Fig. 2), which this DLA model gave the tree height predictions that were very close to the observed ones. Also, the graphical analysis of the scatter plot of the residuals against to predicted heights (Fig. 3) shows a uniform distribution around zero with approximately

constant variance, indicating that the homoscedastic model provides a good representation of the data. Moreover, this uniform and random distribution in the errors obtained by the DLA model with 9 layers and 100 neurons is more distinct (Fig. 2e). When considering the residuals against lagged residuals obtained by the DLA model with 9 layers and 100 neurons and others (Fig. 4), it is seen that this DLA network model provides no trends in the lag-residuals (Fig. 4e) and more desirable qualities for auto-correlation problem than those by the NLR model. Based on all these results with fitting performance criteria, it is concluded that the DLA network models, especially the network model with 9 layers and 100 neurons, have been considered as an alternative prediction method to traditional regression models such as NLRM or NLME and

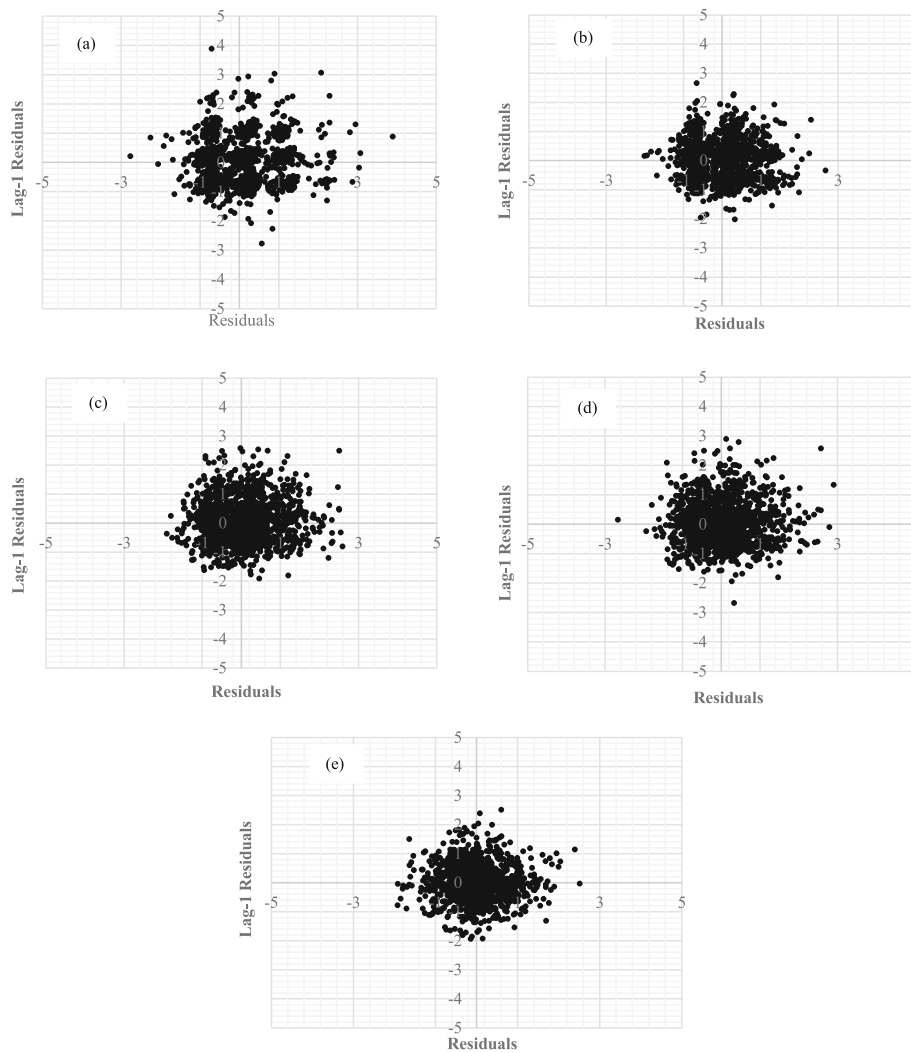


Fig. 4 The plot of residuals against lagged residuals obtained from **a** the M5 based on NLRM, **b** the M5 based on NLME with *f* random, **c** FFB-ANN based on A3 activation function alternative and 85 # neuron, **d** CC-ANN based on A3 activation function alternative and 73 # neuron, **e** DLA with 100 # neurons in nine hidden layers

Table 5 The average of fitting criteria of RMSE, RMSE%, AIC, BIC, FI and AAE according to the number of hidden layers

The alternatives for the number of hidden layers	RMSE	RMSE% (%)	AIC	BIC	FI	AAE
3	0.6823	6.0586	-653.5176	1230.0955	0.9153	0.5424
4	0.6606	5.8658	-708.9420	1174.6712	0.9207	0.5134
5	0.6546	5.8124	-726.2995	1157.3136	0.9219	0.5109
6	0.6326	5.6170	-784.7394	1098.8737	0.9271	0.4858
7	0.6394	5.6778	-765.4140	1118.1991	0.9256	0.4888
8	0.6225	5.5271	-812.0883	1071.5248	0.9295	0.4762
9	0.6286	5.5814	-795.3290	1088.2841	0.9281	0.4803
10	0.6401	5.6840	-763.9056	1119.7075	0.9254	0.4866

Table 6 The average of fitting criteria of RMSE, RMSE%, AIC, BIC, FI and AAE according to the number of Neurons

The alternatives for the number of neurons	RMSE	RMSE% (%)	AIC	BIC	FI	AAE
10	0.7182	6.3773	-564.3337	1319.2794	0.9063	0.5763
20	0.6704	5.9532	-682.1601	1201.4530	0.9184	0.5254
30	0.6581	5.8435	-714.6754	1168.9378	0.9213	0.5164
40	0.6473	5.7474	-744.2704	1139.3427	0.9238	0.4994
50	0.6532	5.8004	-727.2888	1156.3244	0.9225	0.5068
60	0.6414	5.6953	-759.3048	1124.3084	0.9252	0.4942
70	0.6219	5.5218	-812.3812	1071.2320	0.9297	0.4735
80	0.6157	5.4675	-829.2568	1054.3563	0.9311	0.4649
90	0.6195	5.5007	-819.3154	1064.2977	0.9302	0.4701
100	0.6051	5.3732	-859.8078	1023.8053	0.9334	0.4536

other AI technique including FFB-ANN and CC-ANN to model individual tree height-diameter relationships. In this research area about modeling height-diameter relations, Brandao (2007), de A Silva et al. (2008), Diamantopoulou and Özçelik (2012) and Özçelik et al. (2013) compared Artificial Neural Network models with NLR for predicting tree heights, and these studies found that the ANN is superior to NLRM in terms of many statistical criteria. Similarly, Lee et al. (2015), Mohanty et al. (2016), Sladojevic et al. (2016), Carranza-Rojas et al. (2017), Sun et al. (2017), Ferentinos (2018) and Ubbens et al. (2018) successfully used the DLA to determine plant disease diagnosis in agriculture applications. Beyond all these studies including the ANN models in forestry and the DLA models in the agriculture area, this study presents a first DLA model for predicting the relationships between individual tree height and diameter at breast height that

have been an important individual tree measurement in forest inventory. When evaluated the results obtained by the present study, it is seen that the DLA models which are a leading and innovative artificial intelligence technique can be used as an alternative method for regression models whose applications has started in the 1940s such as Metzler (1940), Samuelson (1942), Tintner (1944) and which have problems in providing various statistical assumptions mentioned in many studies nowadays. Although the regression models have provided a certain extent successful fitting results for predicting the relationships between ITH and DBH, the DLA models stand out with some important and attractive features: (1) its strong nonlinear modeling capability without predetermined any statistical functions and (2) no assumptions needed for independence, normal distribution, and homoscedasticity of residuals; and multicollinearity among variables, and

Table 7 The results of equivalence tests for the best predictive DLA of the number of neuron alternatives regarding the numbers of hidden layer and M5 function based on NLRM, M5 based on NLME with f random, FFB-ANN, CC-ANN

The number of neuron alternatives and regression models	The number of hidden layers	b_0 limits		Bootstrap b_0 limits			H_0 : not Equivalent	b_1 limits		Bootstrap b_1 limits			H_0 : not Equivalent
		lower	upper	lower	upper			lower	upper	lower	upper		
50 # neuron	3	9.7579	11.9263	10.7578	10.9326		Rejected	0.9000	1.1000	0.90224	0.9871		Rejected
90 # neuron	4	9.7579	11.9263	10.7458	10.9343		Rejected	0.9000	1.1000	0.9112	0.9895		Rejected
90 # neuron	5	9.7579	11.9263	10.7492	10.9279		Rejected	0.9000	1.1000	0.9227	0.9886		Rejected
100 # neuron	6	9.7579	11.9263	10.7494	10.9343		Rejected	0.9000	1.1000	0.9136	0.9892		Rejected
70 # neuron	7	9.7579	11.9263	10.7530	10.9229		Rejected	0.9000	1.1000	0.9228	0.9941		Rejected
90 # neuron	8	9.7579	11.9263	10.7431	10.9346		Rejected	0.9000	1.1000	0.8774	0.9829		Not rejected
100 # neuron	9	9.7579	11.9263	10.7558	10.9357		Rejected	0.9000	1.1000	0.9015	0.9761		Rejected
80 # neuron	10	9.7579	11.9263	10.7523	10.9339		Rejected	0.9000	1.1000	0.9178	0.9999		Rejected
M5 function based on NLRM		9.7579	11.9263	10.7621	10.9238		Rejected	0.9000	1.1000	0.9851	1.0562		Rejected
NLME with f random		9.7579	11.9263	10.7655	10.9135		Rejected	0.9000	1.1000	0.9849	1.0591		Rejected
FFB-ANN		9.7579	11.9263	10.7400	10.9397		Rejected	0.9000	1.1000	0.9049	1.0184		Rejected
CC-ANN		9.7579	11.9263	10.7490	10.9354		Rejected	0.9000	1.1000	0.8686	1.0118		Not rejected

Table 8 The goodness-of-fit statistics of DLA in validation data set

The number of neuron alternatives and regression models	The number of hidden layers	RMSE	RMSE% (%)	AIC	BIC	FI	AAE	Max. AE	Bias	Bias% (%)
50 # neuron	3	0.8314	7.6684	-50.1267	277.8514	0.8907	0.6544	3.7280	-0.0905	-0.8351
90 # neuron	4	0.8426	7.7715	-46.0660	281.9122	0.8878	0.6546	3.3615	0.0377	0.3475
90 # neuron	5	0.7787	7.1825	-70.0252	257.9530	0.9042	0.6320	2.7641	-0.0126	-0.1162
100 # neuron	6	0.8631	7.9606	-38.7586	289.2195	0.8823	0.6878	3.1312	-0.0114	-0.1048
70 # neuron	7	0.7800	7.1940	-69.5381	258.4401	0.9038	0.6219	2.8616	-0.0406	-0.3740
90 # neuron	8	0.8985	8.2869	-26.5451	301.4330	0.8724	0.6741	5.5280	0.0256	0.2360
100 # neuron	9	0.8291	7.6466	-50.9918	276.9863	0.8914	0.6501	2.9782	-0.0481	-0.4433
80 # neuron	10	0.8613	7.9438	-39.4010	288.5772	0.8828	0.6808	2.9433	-0.0209	-0.1926

spatial and longitudinal autocorrelations in data. In this respect, as an alternative to traditional regression models, the use of DLA models for predicting these ITH-DBH relations and other possible tree and forest attributes can be highlighted.

In addition to the satisfactory findings by the DLA to training data, another issue that should be considered is the analysis of the fitting ability in the simulation data group, especially later uses of the trained model, which were not used in the training process. In the simulation of Artificial Intelligence (AI) models to other forest areas or new measured data, the predictive performance may substantially decrease and the “overfitting” problem may occur in the AI applications. In this regard, the analysis of the success status in independent data is an issue which should be given particular importance in the evaluation of the applicability of AI models. In this study, the DLA models were evaluated in terms of “overfitting” problem by using “Equivalence” test in the independent data. When the “Equivalence” test in Table 7 and fitting criteria in Table 8 are evaluated, it is seen that the DLA models provided acceptable results for these independent data and produced the fitting criteria similar to those of training data. These better predictive results obtained for these independent data compared as those for the training data set suggest that the DLA models may not have a problem of “overfitting”. These predictive results of the DLA, especially for independent data, with no “overfitting” problem can be explained by the fact that the DLA models were trained with the appropriate number of iterations to represent successfully the relationships in the data, which detailed information were provided by Ruder (2017). In this regard, the determination of ideal and optimum DLA has a significant effect on not only increasing the predictive ability of DLA models, but also overcoming the “overfitting” problem during the simulation of independent data in the trainings of DLA models.

In this study, various alternatives with the number of layers and neurons included in network structure were

compared and evaluated to decide the optimal network structure for DLA models, because another issue that should be considered in studies about DLA models was the determination of the optimal network structure. While a significant improvement in fitting criteria can be seen in the average of these criteria from 3 to 8 (3, 4, 5, 6, 7 and 8) number of layers, thus upgrading could not be observed in the average success criteria in 7, 9 and 10 layers (Table 5). With respect to increase in number of neurons, from 10 to 100 numbers of neurons, consistent progress in the average success criteria was obtained in general (Table 6). This worsening in the success criteria depending on 7, 9, and 10 numbers of layers can be explained by the failure of a DLA model structure to represent and model the height-diameter relations, owing to unsatisfactory solution of parameter values related to a DLA model structure which is complicated by excessive increase of the number of layers. On the other hand, when the change in the success criteria due to the increase in the number of neurons is evaluated, it can be explained that the complex model structure that was formed with the increase in the number of neurons in DLA model structure, even with 100 neurons, do not cause a data representation failure in the parameter values. However, another issue that should be considered is the interaction of layer and neuron number changes in the DLA model structure. When these interactive changes of the number of layers and neurons are evaluated, the best predictive results were obtained with 8 layers (Table 5), nevertheless; the best results predictions were obtained with the DLA model which has 9 layers and 100 neurons due to the mutual interaction of numbers of layers and neurons (Table 3). As it is seen in this study, the number of both layers and neurons, if possible, the other parameters of the DLA model structure, should be evaluated together to decide the ideal and optimal DLA model structures and these evaluations can be carried out by comprising the mutual interaction of these factors. These preliminary findings about the number of layers and neurons for a DLA model structure which were

firstly obtained by present study are important results that will make significant contributions to future DLA studies.

Besides the predictive ability of the DLA models in predicting individual tree height-diameter relationships, some features restricting the applicability of these models should also be taken into consideration while evaluating the applicability of DLA models. In general, the regression models where the equation structures and parameter values can be given together are preferred in modelling studies. Also, the DLA models, which consist of tens of layers and neurons, can have the model structures which comprise hundreds even thousands of weight values. In this regard, it will not be possible to give the equation structure of the DLA that has many weight values and to use the applications such as excel, etc. Thus, the applications of DLA models are only possible with the support of various computer software and programs, which gives insight clearly that it will not be very difficult given that we live in the computer era. Especially, the R software platform, which becomes prominent with its applications and usage nowadays, will allow the forest planners initially and other various applicators to use the DLA and various AI models. The applications of DLA models, which were trained by various researchers and applicators, should be prepared in R platform, which is free and open for all, and shared with various stakeholders and other users in forest management.

The study provided the R syntax file of the best predictive DLA network model with 9 layers and 100 neurons as the supplementary file and the downloadable link from Google Drive Link (<https://drive.google.com/open?id=1ewzoB0-0G89rZLkKHVqdkFSLhMjnR9JP>) so that other forest practitioners can use this best predictive DLA model, which similar applications were applied for validation data of 304 trees in this study. This DLA model can be downloaded and can be used by future forest practitioners to obtain the ITH predictions for other tree species in other parts of the world. In the use of these best predictive DLA models in other species and areas, it is an important requirement that the tree species and area for future use are similar to the study area in which the species and data included in this study. As this present study has shown by training the DLA models and providing R syntax codes of the best predictive DLA models, artificial intelligence studies should provide more innovative network tools for different users, as well as including comparisons with other classical methods. This study provides a presentation of R syntax code file for artificial intelligence models to give the opportunity to other forest practitioners to use artificial intelligence model developed in this study.

The data in this study were limited in the sample size of this study is 2024, of which 1720 were used for training, and so the effectiveness and success of artificial

intelligence models in modeling big data may not have been obtained sufficiently, or a limited number of data may have negative effects on iteration success. However, while data pools in the forest growth and yield modeling studies such as this study remain limited the sample size, data analysis which may consist of millions or even millions of data, also called as big data, may be involved in applications such as forestry image processing such as Hamdi et al. (2019), Fricker et al. (2019) and Sylvain et al. (2019). In the analysis of forestry image processing data based on big data, the effectiveness of deep learning techniques will be even more apparent.

This study has introduced innovative Deep Learning Algorithms (DLA), being as another application of Artificial Intelligence Techniques, which were resulted in superior fitting statistics compared as conventional regression models. The weakness of this study is that the fitting results are obtained by modeling only one species form pure stands. However, the future applications of DLA models need to be realized for mixed stands or uneven forest stands. Thus, the acceptability of the results for the DLA models will become even more apparent and the availability of other models can be achieved. However, more scientific studies are needed to compare DLA models with other convenient models. As an artificial intelligence technique, the present study is a preliminary step and contribution to the evaluation process regarding the future usability of deep learning technique and its scientific acceptability.

Conclusion

We have been experiencing the fourth Industrial Revolution with the proliferation of the use of artificial intelligence nowadays and the evaluation of the Deep Learning Algorithms, one of the Artificial Intelligence Techniques that has come up since 2010, stands out as an important requirement in forest yield and growth modelling studies. This paper presents the DLA models, as innovative prediction technique, to predict the relationships between individual tree heights and diameter at breast height, which are an important growth parameter of trees and so, the usability and capability of the DLA were evaluated based on some fitting criteria in both training and simulation datasets. The fitting results obtained by the DLA models underlined that the DLA models can be assessed as an alternative prediction method for traditional regression models to obtain individual tree heights in forest inventory. This paper introduces the abilities of the DLA models that have been a novel neural network model in the field of Artificial Intelligence to predict the individual tree heights from the diameter at breast height measured in the sample plots. Besides predictive applications of the DLA models in modeling tree height-diameter relations in this study, the fitting ability and usability of the DLA models should be evaluated

in predicting the other individual tree attributes such as tree volume, taper and growth and so stand attributes such as stand volume, basal area, biomass and carbon. It is confronted as an important need that the realization of different studies related to the evaluation of the DLA models being as novel Artificial Intelligence Application, which found a place newly in the forestry literature, as an alternative for conventional statistical methods in predicting various stands and individual tree attributes.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40663-020-00226-3>.

Additional file 1 Appendix 1. R syntax code.

Abbreviations

AAE: the Average Absolute Error; ADADELTA: the Adaptive Learning Rate Algorithm; AI: the Artificial Intelligence; AIC: Akaike's Information Criterion; ANN: the Artificial Neural Networks; Bias: the average Bias; Bias%: percent of average Bias; BIC: Bayesian Information Criterion; DBH: the diameter at breast height; DLA: Deep Learning Algorithms; Ft: the Fit Index; ITH: the individual total height; max. AE: the Maximum Absolute Error; NLME: the Nonlinear Mixed Effect; NLRM: the Nonlinear Regression Models; NLS: the Nonlinear Least Squares; RMSE: the Root Mean Squared Error; RMSE%: percent of Root Mean Squared Error; TOST: Two One-Sided Test Strategy

Acknowledgements

Not Applicable.

Authors' contributions

İlker ERCANLI contributed the manuscript as supervising the work, conception, administrative, design, material support, critical revision and statistical analysis. The author(s) read and approved the final manuscript.

Authors' information

İlker ERCANLI is an associated professor at Çankırı Karatekin University, Forest Faculty, Department of Forest Yield Studies. He is a forest biometrician, focused on forest growth modeling.

Funding

Not Applicable.

Availability of data and materials

Available on request.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The author declare that they have no competing interests.

Received: 3 August 2019 Accepted: 2 March 2020

Published online: 03 April 2020

References

- Adame P, del Río M, Canellas I (2008) A mixed nonlinear height–diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). *Forest Ecol Manag* 256:88–98
- Ashraf MI, Zhao Z, Bourque CP-A, MacLean DA, Meng F-R (2013) Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology. *Can J For Res* 43:1162–1171
- Avery TE, Burkhart HE (1983) *Forest measurements*. McGraw-Hill Education, USA
- Brandao FG (2007) Estimativa da altura total de eucalyptus sp. utilizando lógica fuzzy e neuro fuzzy. Dissertation, Universidade Federal de Lavras
- Budhathoki CB, Lynch TB, Guldin JM (2008) A mixed-effects model for the dbh–height relationship of shortleaf pine (*Pinus echinata* mill.). *South J Appl Forest* 32:5–11
- Calama R, Montero G (2004) Interregional nonlinear height diameter model with random coefficients for stone pine in Spain. *Can J For Res* 34:150–163
- Cañadas N, García C, Montero G (1999) Relación altura-diámetro para *Pinus pinea* L. en el Sistema Central. Congreso de Ordenación y Gestión Sostenible de Montes, Santiago de Compostela, pp 139–153
- Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A (2017) Going deeper in the automated identification of herbarium specimens. *BMC Evol Biol* 17:181
- Crecente-Campo F, Tome M, Soares P, Dieguez-Aranda U (2010) A generalized nonlinear mixed-effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. *Forest Ecol Manag* 259:943–952
- de A Silva RM, Brandão FG, Baleiro GB, Valentim FL, de Mendonça AR, Pires DM (2008) Fuzzy and neuro-fuzzy estimates of the total height of eucalyptus trees. Proceedings of the 2008 ACM Symposium on Applied Computing. ACM, pp 1772–1776
- Development Core Team R (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Diamantopoulou M, Özçelik R (2012) Evaluation of different modeling approaches for total tree-height estimation in Mediterranean region of Turkey. *Forest Syst* 21:383–397
- Diamantopoulou MJ (2005a) Artificial neural networks as an alternative tool in pine bark volume estimation. *Comput Electron Agr* 48:235–244
- Diamantopoulou MJ (2005b) Predicting fir trees stem diameters using artificial neural network models. *South Afr For J* 205:39–44
- Diamantopoulou MJ (2006) Tree-bole volume estimation on standing pine trees using cascade correlation artificial neural network models. *Agr Eng Int CIGR J* 8:1–14
- Diamantopoulou MJ, Milios E (2010) Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. *Biosyst Eng* 105:306–315
- Dorado FC, Anta MB, Parresol BR, González JGÁ (2005) A stochastic height-diameter model for maritime pine ecoregions in Galicia (northwestern Spain). *Ann For Sci* 62:455–465
- Dorado FC, Diéguez-Aranda U, Anta MB, Rodríguez MS, von Gadow K (2006) A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. *Forest Ecol Manag* 229:202–213
- Fang Z, Bailey R (1998) Height–diameter models for tropical forests on Hainan Island in southern China. *Forest Ecol Manag* 110:315–327
- Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. *Comput Electron Agr* 145:311–318
- Ferguson I, Leech J (1978) Generalized least squares estimation of yield functions. *For Sci* 24:27–42
- Fricker GA, Ventura JD, Wolf JA, North MP, Davis FW, Franklin J (2019) A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. *Remote Sens* 11:2326
- Grégoire TG, Schabenberger O, Barrett JP (1995) Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can J For Res* 25:137–156
- Guan BT, Gertner G (1991) Modeling red pine tree survival with an artificial neural network. *For Sci* 37:1429–1440
- Hamdi ZM, Brandmeier M, Straub C (2019) Forest damage assessment using deep learning on high resolution remote sensing data. *Remote Sens* 11:1976
- Hasenauer H, Kindermann G (2002) Methods for assessing regeneration establishment and height growth in uneven-aged mixed species stands. *Forestry* 75:385–394
- Hasenauer H, Merkl D, Weingartner M (2001) Estimating tree mortality of Norway spruce stands with neural networks. *Adv Environ Res* 5:405–414
- Huang S, Price D, Titus S (2000) Development of ecoregion-based height–diameter models for white spruce in boreal forests. *Forest Ecol Manag* 129: 125–141
- Huang S, Titus SJ, Wiens DP (1992) Comparison of nonlinear height–diameter functions for major Alberta tree species. *Can J For Res* 22:1297–1304
- Hui G, Gadow KV (1993) Zur Entwicklung von Einheitshöhenkurven am Beispiel der Baumart *Cunninghamia lanceolata*. *Allgemeine Forst-und Jagdzeitung* 164:218–220
- Krumland BE, Wensel LC (1988) A generalized height-diameter equation for coastal California species. *West J Appl For* 3:113–115

- Kv G, Hui GY (1999) Modelling Forest Development. Kluwer Academic Publishers, Dordrecht
- Lappi J (1997) A longitudinal analysis of height/diameter curves. *For Sci* 43:555–570
- Larsen DR, Hann DW (1987) Height-diameter equations for seventeen tree species in Southwest Oregon. Forest Research Laboratory, College of Forestry, Oregon State University
- Lee SH, Chan CS, Wilkin P, Remagnino P (2015) Deep-plant: Plant identification with convolutional neural networks. 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, pp 452–456
- Leite HG, da Silva MLM, Binoti DHB, Fardin L, Takizawa FH (2011) Estimation of inside-bark diameter and heartwood diameter for *Tectona grandis* Linn. Trees using artificial neural networks. *Eur J Forest Res* 130:263–269
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS system for mixed models. SAS Institute Inc., Cary
- Loetsch F, Zöhrer F, Haller KE (1973) Forest Inventory, Volume II. BLV Verlagsgesellschaft München Bern Wien, München
- Lynch TB, Holley AG, Stevenson DJ (2005) A random-parameter height-dbh model for cherrybark oak. *South J Appl For* 29:22–26
- Martin FC, Flewelling JW (1998) Evaluation of tree height prediction models for stand inventory. *West J Appl For* 13:109–119
- MATLAB (2014) MATLAB and Statistics Toolbox. Release 2014b. The MathWorks, Inc., Natick
- Mehtätalo L (2004) A longitudinal height–diameter model for Norway spruce in Finland. *Can J For Res* 34:131–140
- Metzler LA (1940) The Assumptions Implied in Least Squares Demand Techniques. *Rev Econ Stat* 22:138–149
- Meyer HA (1940) A mathematical expression for height curves. *J For* 38:415–420
- Miguel EP, Mota FCM, Téo SJ, Nascimento RGM, Leal FA, Pereira RS, Rezende AV (2016) Artificial intelligence tools in predicting the volume of trees within a forest stand. *Afr J Agric Res* 11:1914–1923
- Mohanty SP, Hughes DP, Salathé M (2016) Using deep learning for image-based plant disease detection. *Front Plant Sci* 7:1419
- Nanos N, Calama R, Montero G, Gil L (2004) Geostatistical prediction of height/diameter models. *Forest Ecol Manag* 195:221–235
- Nunes MH, Görgens EB (2016) Artificial intelligence procedures for tree taper estimation within a complex vegetation mosaic in Brazil. *PLoS One* 11: e0154738
- Özçelik R, Diamantopoulou MJ, Brooks JR, Wiant HV Jr (2010) Estimating tree bole volume using artificial neural network models for four species in Turkey. *J Environm Manag* 91:742–753
- Özçelik R, Diamantopoulou MJ, Crecente-Campo F, Eler U (2013) Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecol Manag* 306:52–60
- Özçelik R, Diamantopoulou MJ, Eker M, Gürlevik N (2017) Artificial neural network models: an alternative approach for reliable aboveground pine tree biomass prediction. *For Sci* 63:291–302
- Özçelik R, Diamantopoulou MJ, Wiant HV Jr, Brooks JR (2008) Comparative study of standard and modern methods for estimating tree bole volume of three species in Turkey. *Forest Prod J* 58:73
- Parresol BR (1992) Baldcypress height–diameter equations and their prediction confidence intervals. *Can J For Res* 22:1429–1434
- Paulo JA, Tome J, Tome M (2011) Nonlinear fixed and random generalized height-diameter models for Portuguese cork oak stands. *Ann Forest Sci* 68: 295–309
- Peng C, Zhang L, Liu J (2001) Developing and validating nonlinear height–diameter models for major tree species of Ontario's boreal forests. *North J Appl For* 18:87–94
- Peng CH (1999) Nonlinear height-diameter models for nine boreal forest tree species in Ontario. Forest Research Report, Ontario Forest Research Institute, p 28
- Pinheiro J, Bates D (2000) Mixed-Effects Models in S and S-PLUS. Springer-Verlag, New York
- Poudel KP, Cao QV (2013) Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. *For Sci* 59:243–252
- Prodan M (1965) Holzmesslehre. Sauerlaender's Verlag, Frankfurt am Maine
- Richards F (1959) A flexible growth function for empirical use. *J Exp Bot* 10:290–301
- Robinson AP, Duursma RA, Marshall JD (2005) A regression-based equivalence test for model validation: shifting the burden of proof. *Tree Physiol* 25:903–913
- Robinson AP, Froese RE (2004) Model validation using equivalence tests. *Ecol Model* 176:349–358
- Robinson AP, Wykoff WR (2004) Imputing missing height measures using a mixed-effects modeling strategy. *Can J For Res* 34:2492–2500
- Ruder S (2017) An overview of multi-task learning in deep neural networks. arXiv preprint, arXiv:1706.05098
- Samuelson PA (1942) A note on alternative regressions. *Econometrica* 10(1):80–83
- Schnute J (1981) A versatile growth model with statistically stable parameters. *Can J Fish Aquat Sci* 38:1128–1140
- Searle S, Casella G, McCulloch CJINY (1992) Variance components. Wiley, New York
- Sharma M, Parton J (2007) Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecol Manag* 249: 187–198
- Sharma M, Zhang SY (2004) Height–diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. *Scand J Forest Res* 19:442–451
- Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D (2016) Deep neural networks based recognition of plant diseases by leaf image classification. *Comput Intel Neurosci*. <https://doi.org/10.1155/2016/3289801>
- Soares FAA, Flôres EL, Cabacinha CD, Carrijo GA, Veiga ACP (2011) Recursive diameter prediction and volume calculation of eucalyptus trees using multilayer perceptron networks. *Comput Electr Agric* 78:19–27
- Soares P, Tomé M (2002) Height–diameter equation for first rotation eucalypt plantations in Portugal. *Forest Ecol Manag* 166:99–109
- Sun Y, Liu Y, Wang G, Zhang H (2017) Deep learning for plant identification in natural environment. *Comput Intel Neurosci*. <https://doi.org/10.1155/2017/7361042>
- Sylvain J-D, Drolet G, Brown N (2019) Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS J Photogr Remote Sens* 156:14–26
- Temesgen H, Gadow KV (2004) Generalized height–diameter models—an application for major tree species in complex stands of interior British Columbia. *Eur J Forest Res* 123:45–51
- Tintner G (1944) An application of the variate difference method to multiple regression. *Econometrica* 12(2):97–113
- Tomé MMB (1989) Modelação do crescimento de árvore individual em povoamentos de *Eucalyptus globulus* Labill. (1ª rotação) Região centro de Portugal, p. 277
- Trincado G, VanderSchaaf CL, Burkhart HE (2007) Regional mixed-effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations. *Eur J Forest Res* 126:253–262
- Ubbens J, Cieslak M, Prusinkiewicz P, Stavness I (2018) The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Method* 14:6
- Van Laar A, Akça A (2007) Forest mensuration. Springer Science & Business Media, Netherlands
- Vanclay JK (1994) Modelling forest growth and yield: applications to mixed tropical forests. School Environm Sci Manag Papers, p 537
- Wykoff WR, Crookston NL, Stage AR (1982) User's guide to the stand prognosis model. Gen. Tech. Rep. INT-133. US Department of Agriculture, Forest Service, Intermountain Forest Range Experiment Station, Ogden
- Zeiler MD (2012) ADADELTA: an adaptive learning rate method. *ArXiv-Machine Learning*

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
