RESEARCH ARTICLE

Artificial intelligence with deep learning algorithms to model relationships between total tree height and diameter at breast height

İlker Ercanli (Ercanli, I)[1*]

[1]*Çankırı Karatekin University, Faculty of Forestry, 18200, Çankırı-Turkey.*

Abstract

Aim of study: As an innovative prediction technique, Artificial Intelligence technique based on a Deep Learning Algorithm (DLA) with various numbers of neurons and hidden layer alternatives were trained and evaluated to predict the relationships between total tree height (TTH) and diameter at breast height (DBH) with nonlinear least squared (NLS) regression models and nonlinear mixed effect (NLME) regression models.

Area of study: The data of this study were measured from even-aged, pure Turkish Pine (*Pinus brutia* Ten.) stands in the Kestel Forests located in the Bursa region of northwestern Turkey.

Material and methods: 1132 pairs of TTH-DBH measurements from 132 sample plots were used for modeling relationships between TTH, DBH, and stand attributes such as dominant height (H_o) and diameter (D_o).

Main results: The combination of 100 # neurons and 8 # hidden layer in DLA resulted in the best predictive total height prediction values with Average Absolute Error (0.4188), max. Average Absolute Error (3.7598), Root Mean Squared Error (0.6942), Root Mean Squared error % (5.2164), Akaike Information Criteria (-345.4465), Bayesian Information Criterion (-330.836), the average Bias (0.0288) and the average Bias % (0.2166), and fitting abilities with r (0.9842) and Fit Index (0.9684). Also, the results of equivalence tests showed that the DLA technique successfully predicted the TTH in the validation dataset.

Research highlights: These superior fitting scores coupled with the validation results in TTH predictions suggested that deep learning network models should be considered an alternative to the traditional nonlinear regression techniques and should be given importance as an innovative prediction technique.

Keywords: Prediction; artificial intelligence; deep learning algorithms; number of neurons; hidden layer alternatives.

Introduction

In forest inventories, total tree height (TTH) and diameter at breast height (DBH) (measured at 1.3 m above ground) are significant forest inventory variables that have been used for site index predictions, total and merchantable volume and biomass predictions, carbon budget models, and growth and yield models (Clutter *et al.,* 1983; Van Laar & Akça, 2007). TTHs are generally measured in a subset of trees in sample plots because their field measurements are complex, tedious, and time-consuming. However, DBHs are measured for all sample trees in sample plots because their measurements are simple, easy, and cheap. A high correlation between TTH and DBH has led to the development of regression models to

predict TTH from DBH (Huang *et al.,* 1992; Martin & Flewelling, 1998). Furthermore, some growth simulators such as Prognosis BC (Crookston & Dixon, 2005) and Forest Vegetation Systems, FVS, (Wykoff *et al.,* 1982) have included prediction modules that predict TTHs by using the relationship between TTH and DBH.

To develop the TTH-DBH prediction models as a part of forest inventory, sample plots are sampled from forest stands with different growing conditions such as site quality, stocking, and stand ages. These heterogeneous growing conditions of forest stands have an important effect on the relationships between TTH and DBH, and thus these relations differ from one stand to another with various stand structures. These heterogeneities may result in unexplained variance in TTH predictions (Curtis,

1967). An approach to take into account the TTH variance is to develop generalized or region-wide TTH-DBH models which include some other stand characteristics such as dominant height, quadratic mean diameter, dominant diameter, number of trees per hectare and stand basal area as predictors (Fulton, 1999; Huang *et al.,* 2000; Schröder & González, 2001; Sánchez *et al.,* 2003; Sharma & Yin Zhang, 2004; Temesgen & Gadow, 2004; Castedo-Dorado *et al.,* 2005; Trincado *et al.,* 2007). On the other hand, Nanos *et al.* (2004) analyzed the spatial pattern of variability in height/diameter ratios and proposed geostatistical interpolation of the regression parameters.

Nonlinear mixed effect modeling has been proposed as an alternative method to deal with hierarchical data, representing stands with heterogeneous characteristics. Previous studies including the development of generalized TTH-DBH models have attempted to minimize or decrease the TTH prediction variances between sample plots which are diverse in stand characteristics. In general, clustered and hierarchical data are highly autocorrelated, which mainly induces high inter-stand prediction variance. The autocorrelation in these clustered and hierarchical data may cause serious fitting problems in modeling the TTH–DBH relations. Consequently, the ordinary least squares (OLS) for linear models or nonlinear least squares (NLS) techniques for nonlinear models may have biased predictions of the confidence intervals of model parameters when these autocorrelated data are used in this analysis (West *et al.,* 1984; Gregorie, 1987; Searle *et al.,* 1992; Lappi, 1997). In the linear or nonlinear mixed effect modeling, population-specific fixed-parameter and sampling unit-specific random-parameter are simultaneously predicted by a covariance matrix which is defined in the same model structure (Calama & Montero, 2004). The use of random parameters in model structure enables the estimation of the residual variance of TTH-DBH relationships among clustered or nested sample units from different stands. Also, the NLME models that have been developed for predicting TTH can be calibrated for any locations where data have not been used for constructing the model. The features of the mixed effect modeling technique provide more efficient and accurate TTH predictions from nested and clustered sample units located in different forest stands than either OLS or NLS. Numerous studies such as Lappi (1997); Calama & Montero (2004); Mehtätalo (2004); Castedo-Dorado *et al.* (2006); Trincado *et al.* (2007); Sharma & Parton (2007); Crecente-Campo *et al.* (2010); Mehtätalo *et al.* (2015); Sharma *et al.* (2016); Özçelik *et al.* (2018); Ng'andwe *et al.* (2019); Nong *et al.* (2019); Mugasha *et al.* (2019); Bronisz & Mehtätalo (2020) have used the linear or nonlinear mixed effect technique in modeling TTH-DBH relationships.

In addition to statistical approaches, Artificial Intelligence Techniques (AIT) such as Artificial Neural Networks (ANNs) and Deep Learning Algorithms (DLAs) have provided an innovative modeling approach to obtain the predictions of some tree and stand attributes. Some prediction techniques based on the ANNs have been developed to predict various individual tree and stand attributes (Hasenauer *et al.,* 2001; Diamantopoulou, 2005b, a, 2006; Özçelik *et al.,* 2008; Diamantopoulou & Milios, 2010; Diamantopoulou & Özçelik, 2012; Diamantopoulou *et al.,* 2015; Özçelik *et al.,* 2017; Özçelik *et al.,* 2018). In these studies, ANNs gave better predictive results in the various tree and stand attributes as compared to conventional NLS and NLME models. This predictive ability of ANNs for these tree and stand variables proposed by the previous studies are based on two important characteristics: (1) its strong nonlinear modeling capability without any predetermined statistical functions and (2) no statistical assumptions needed for independence, normal distribution, and homoscedasticity of residuals as well as multicollinearity among variables, and spatial and longitudinal autocorrelations in data.

The network models based on DLAs are another type of artificial intelligence with innovative predictive abilities for modeling forest and tree attributes. ANNs have the general structure of an input layer, a hidden layer (interlayer), and an output layer. This structure of ANNs is limited to several layers including one or two hidden layers, while DLAs can include more layers, especially a greater number of hidden layers, which makes this network model based on the DLAs better predictive modeling than those by ANNs. This is especially true for advanced computational systems such as Graphical Processing Units (GPU) that are directly embedded into the computer processors and have provided highly convenient network models based on DLAs. These network models based on DLAs have been used by several studies in agriculture, plant disease diagnosis, and plant pattern recognition (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). These studies that include the prediction systems based on DLAs have shown significant improvements in the prediction of image processing of plant disease diagnosis and plant pattern recognition. In forest growth and yield modeling studies, there are few studies about the network models based on DLAs to predict various tree and stand variables, which are denoted as significant measurements in forest inventory. Ercanlı *et al.* (2018) trained the network models based on DLAs to model diameter distributions and obtained better predictive results including some reductions with 79.01% in SSE, 54.15% RMSE, 18.49% in AIC and 18.37% in BIC and increase with 30.63% increase in R2 for five-layer deep learning algorithm compared as 3-Parameter Weibull probability density function. Ercanlı (2020) compared the prediction models of nonlinear regression, nonlinear mixed-effect regression, the network models based on the DLAs, and ANNs for tree heights, and showed that the

DLA network model with 9 layers and 100 neurons resulted in the best predictive tree-heights with significant improvement in the values of RMSE, AIC, BIC, FI, AAE, max. AAE with the rates of 26.85%, 116.58%, 37.80%, 5.48%, 33.52%, 35.51%, respectively, compared to those of NLS. In addition to these limited studies which include the network models based on the DLAs, the usability and predictive ability of DLAs with their versatility and high rate for predicting tree and stand attributes remains an unclear and uninvestigated issue. Therefore, the objective of this study is (1) to train the network models based on DLA models to predict the relationships between TTHs from DBHs and (2) to analyze the fitting performance of DLAs to predict these relations and (3) to evaluate these network models with DLAs as alternative prediction techniques compared to the conventional statistical regression models including NLS and NLME, (4) to present the best predictive DLA network model as R syntax which can then be utilized by the future forest practitioners.

Materials and methods

Study Site

The study covers even-aged, pure Turkish Pine (*Pinus brutia* Ten.) stands located in Kestel Forest (40º 00′ 00″– 40º 12′ 10″ N 29º 13′ 00″–29º 21′ 54″ E) of the Bursa
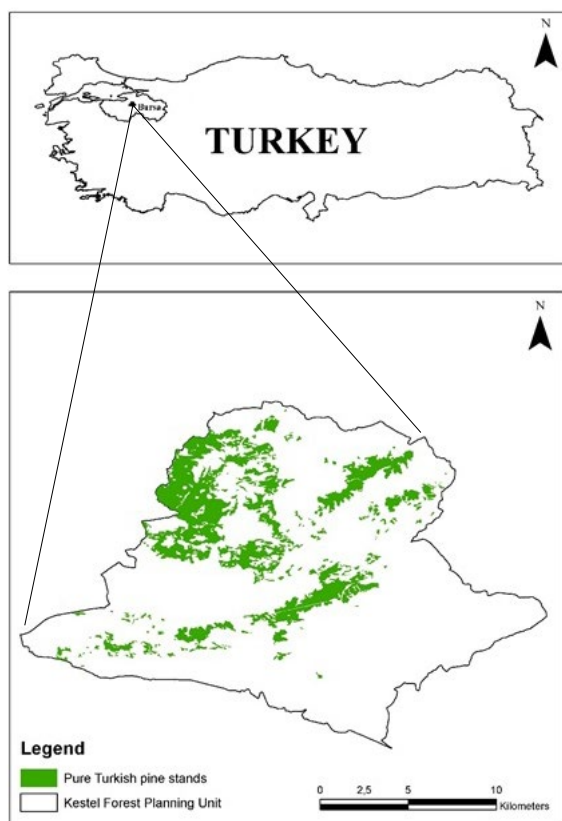

Figure 1. The study area and distribution of sampled Turkish Pine Stands.

Forest District Area in northwestern Turkey (Fig. 1). Kestel Forests are one of the representative distribution areas for natural Turkish Pine stands. The study forests are located on moderate to steeply sloping (15% to 50%) landscapes with elevation ranging from 200 m to 800 m above sea level. The climate is a typical central Anatolian characterized by moderate winters and hot summers. The annual mean temperature is 13.6 ºC and mean daily maximum temperature ranges from 10.1 ºC to 29.4 ºC; mean daily minimum temperature ranges from 3.8 ºC to 17.4 ºC, respectively. Mean annual precipitation ranges from 200 mm to 400 mm; and annual total precipitation are distributed relatively homogeneously throughout the year (Turkey Meteorological Service, 2017).

Data

In the even-aged Turkish Pine stands, 132 sample plots were selected subjectively to represent different stand conditions such as site quality, age, and stand density. These sampled Turkish Pine stands were naturally regenerated and uniformly stocked (60-90% tree layer cover), with no historical evidence of damage caused by fire or storms. The plots were checked carefully for no evidence of intensive silvicultural treatments or clear-cutting. The size of circular sample plots ranged from 400 to 800 m2 to include a minimum of 30–35 trees in a sample plot based on the stand crown closure. At each sample plot, DBH was measured to 0.1 cm precision using calipers on every living tree with a DBH > 8 cm. TTH was measured on a subset of trees, selecting two to three trees for each 4 cm diameter class using Blume–Leiss Altimeter (0.1 m precision). In addition to these TTH measurements in the sub-samples, the TTH of dominant and co-dominant trees was selected based on the 100 dominant and co-dominant highest trees per hectare (e.g. four highest trees in a 0.04-ha plot), were measured. Thus, the dominant diameter (cm) and dominant height (m) were calculated as part of the plot level information for each sample plot. Dominant height (H_o) and diameter (D_o) for a sample plot was calculated by averaging the TTH and DBH of the corresponding dominant and co-dominant trees.

In total, 1132 pairs of TTH-DBH measurements from 132 sample plots were used for modeling relationships between TTH, DBH, and stand attributes such as dominant height (H_o) and diameter (D_o). These data were randomly split into training (also called modeling data sets for NLS and NLME regression models) and the validation data subsets, by using the function SAMPLE in the R statistical environment (R Development Core Team, 2018). Of 1132 trees, approximately 85%, 963 trees in 107 sample plots, were used to train the network models based on the DLAs including various network

Table 1. Summary statistics for sample trees originated from fitting and validation data.

	Number of trees and sample plots	Variables	Min.	Max.	Mean	Std. Deviation
Training	963 trees at	DBH (cm)	8.00	69.00	29.94	11.39
		THT (m)	3.50	25.00	13.31	3.90
dataset	107 plots	D_o (cm)	9.00	69.00	39.93	11.63
		H_o (m)	3.50	25.00	15.56	3.69
Validation	169 trees at 25	DBH (cm)	8.00	66.00	31.99	11.63
		THT (m)	3.00	22.50	13.53	3.90
dataset	plots	D_o (cm)	10.00	69.00	43.05	10.74
		H_o (m)	4.00	25.00	16.14	3.61

TTH: Total tree height (m), *DB*: Diameter at breast height (cm): H_o: Dominant height (m), D_o: Dominant diameter (cm),

architectures and develop NLS and NLME models. The remaining 169 samples from 25 sample plots were used to validate these network models. The summary statistics for model training and simulation data sets are given in Table 1.

Prediction Methods

In this study, the performance of NLS, NLME models, and the Network Models based on the DLAs were evaluated concerning modeling relationships between the TTH and DBH.

Nonlinear Least Squares

The relationships between the TTH and DBH are mainly nonlinear with a typical sigmoidal or S-shaped curve pattern including functional inflection point and an asymptote (Lappi, 1997). The Schnute's (1981) model, which is the most flexible in providing typical patterns and adaptable functions for modeling this relation (Bredenkamp & Gregoire, 1988; Lei, 1998) was selected as the candidate model. The Schnute's (1981) model has the form:

$$h_i = \left(1.3^{b_0} + \left(H_0^{b_0} - 1.3^{b_0} \right) \frac{\left(1 - e^{-b_1 \cdot d} \right)}{\left(1 - e^{-b_1 \cdot D_0} \right)} \right)^{\frac{1}{b_1}} \quad (1)$$

where b_0 and b_1 are model parameters, hi is the observed TTH of the ith tree in the sample plots, di is DBH of the ith tree the sample plots, H_0 is the dominant height of the sample plot, D_0 is the dominant diameter of the sample plot. Based on the Nonlinear Least Squares (NLS), which uses the Levenberg-Marquardt algorithm, the parameters of Schnute's (1981) model were obtained with

NLS function available in the R statistical environment (R Development Core Team, 2018).

To deal with heteroscedasticity problem (i.e. non-homogeneous variance in heights predictions concerning the regressor) in TTH predictions (a characteristic problem in height-diameter models), the residual variances were modeled as a function of DBH using Power function:

$$Var(\varepsilon_{ij}) = \sigma^2 \cdot |DBH_{ij}|^{2\phi} \quad (2)$$

Where DBH_{ij} is the i tree DBH at j sample plot, σ^2 is a scaling factor for the error dispersion of TTH and ϕ is the variance function coefficient (Pinheiro & Bates, 2000).

Nonlinear Mixed Effect Model

The datasets with multiple measurements that were measured at sample plots from various forest stands result in the nested data structure, which is not independent and is highly correlated. Thus, unexplained variance often arises in the TTH-DBH relationships between clustered or nested sample units, which may cause the hypothesis testing and statistical inference to be invalid and produce biased estimates of the confidence intervals for the model parameters for the developed TTH-DBH model (Judge *et al.*, 1982; Gregorie, 1987; Searle *et al.*, 1992).

To overcome this problem originating from the nested data structure, a Nonlinear Mixed Effect (NLME) modeling procedure was applied to the Schnute's (1981) height-diameter model by simultaneously predicting both fixed and random parameters, u_i and v_i. The nonlinear mixed effect modeling approach comprises basic assumptions of a multivariate normal distribution of the residual terms and random-effects parameter, u_i and v_i (Calama & Montero, 2004).

$$u_i \text{ and } v_i \sim N(0, D) \quad \varepsilon_{ij} \sim N(0, R_j) \quad (3)$$

Where D is a $q \times q$ positive-definite variance-covariance matrix, representing the between-plot variance for the random-effects, and R_j is a $q \times q$ matrix representing the within plot variance-covariance. Specifically, the variance-covariance structures were defined by D and R_j matrixes to model random variability existing within and between plots (Calama & Montero, 2004). The D matrix is mutual to all plots and typically assumed to be an unstructured covariance matrix and identical for all plots, and it describes plot variation (Huang *et al.*, 2009). The D matrix with two random parameters, ui and vi, was considered to model the variation between sampling plots, and the variance-covariance matrix structures are defined as follows:

$$D = \begin{bmatrix} \sigma_{u_i}^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_{v_i}^2 \end{bmatrix} \quad (4)$$

where $\sigma_{u_i}^2$ is the variance for the random effect u_i, σ_u^2 is the variance for the random effect v_i, σ_{uv}^2 is the covariance between random effects (Castedo-Dorado *et al.*, 2006).

As explained before, the heteroscedasticity problem was assumed in TTH predictions and the residual variances were modeled as a function of DBH using the Power function for Schnute's (1981) model based on the NLS. Similar to the weighting process in the NLS model, the varPower function, previously formulated in Eq. 2 in NLME package of R programming language by weighting the variance of the residuals with a power of the DBH, was used to obtain the solution for this heteroscedasticity problem (Pinheiro & Bates, 2000).

Taking into account the weighting coefficient in the NLME model structure, the within-plot variance-covariance must be defined by the R_j matrix in a special structure as defined by Calama & Montero (2004), which is a different structure than that described by Castedo-Dorado *et al.* (2006); Paulo *et al.* (2011); Trincado *et al.* (2007). The general explanation of R_j matrix was given by Calama & Montero (2004):

$$R_j(\lambda, b_j, \rho) = \sigma^2 \cdot G_j^{0.5} \cdot I_j \cdot G_j^{0.5} \qquad (5)$$

Where $G_j^{0.5}$ is a $n_j x n_j$ diagonal matrix, Ij is a $nj\,xnj$ identity matrix and σ^2 is a scaling factor for the error dispersion (Grégoire *et al.*, 1995).

By using modeling data including 963 trees in 107 sample plots, the variance components and fixed parameters of Schnute's (1981) height-diameter model were predicted with the NLME package in the R statistical environment (R Development Core Team, 2018). The maximum-likelihood method (ML) was used to fit the nonlinear mixed-effect regression, and the adaptive Gaussian quadrature was used in the computation of the integral over the random effects as described by Pinheiro & Bates (2000).

The Network Models based on Deep Learning Algorithms

As an alternative prediction technique to these statistical regression methods including NLS and NLME, the network models based on the DLAs were used to obtain the TTH predictions from some predictor variables such as the DBH, the dominant height (H_0), and the dominant diameter (D_0). DLAs are a subdivision of ANNs, which can be developed with a greater number (more than three) of hidden layers; although, the architecture of ANNs usually comprises three layers: one input layer, one hidden layer, and one output layer, or two hidden layers for some limited applications. However, DLAs can utilize a greater number of layers, especially hidden layers, with the assistance of the power of particular GPUs. This multi-layer structure provides an effective representation of complex systems (e.g., forest ecosystems). Thus, the network models based on DLAs may define complex and

inter-correlated relationships between different trees and stand attributes, representing different forest areas, better than ANN models.

DLAs utilize great quantities of information and the computational power of GPUs to learn and discover information from data such as images, numerals, and text. To achieve the sophisticated computation processes needed by DLAs, some tools and libraries such as Caffe2, Cognitive Toolkit, MXNet, PyTorch, TensorFlow, and H2O have been developed based on GPU-accelerated operations. Among these applications, H2O is an open-source Artificial Intelligence platform that allows users to utilize Machine Learning techniques such as Naïve Bayes, K-means, PCA, Deep Learning, and Autoencoders (H2O. ai, 2018). The H2O package includes an h2o.deeplearning function, which was coded on Java and is principally suitable to train multi-layer feedforward deep learning neural networks. The h2o.deeplearning function available in R was used to train the network models based on DLAs (H2O.ai, 2018). This h2o.deeplearning function provides multi-layer feedforward neural network models, which comprise a supervised training protocol to predict TTHs from input variables of DBH, D_0, and H_0. In training, these DLAs, DBH, H_0, and D_0 were identified as input variables, and TTHs were identified as the output variable.

The convergence of DLAs critically depends on the architecture of the network, the number of hidden layers, the type of activation function, the number of neurons in hidden layers, and other parameters such as epochs, type of distribution functions, rho, and epsilon. This h2o. deeplearning function uses the adaptive learning rate algorithm ADADELTA (Zeiler, 2012), which includes a combination of learning rate annealing and momentum training, enabling fast converge of the DLAs (H2O.ai, 2018). The rho defines the rate of ADADELTA and epsilon defines the learning rate strength during preliminary training. The value of 0.999 for rho and 1×10^{-8} for epsilon were used to train DLAs. Also, the value of 1000 for the epochs, the number of iterations to be carried in training networks, was used in the training of DLAs, since the best predictive results have been obtained with 1000 in various neural network studies. The h2o.deeplearning function uses the gradient descent and displays distributions of Bernoulli, Multinomial, Poisson, Gamma, Tweedie, Laplace, Huber, and Gaussian. The Gaussian distribution with the Mean Squared Error as training distribution was used in the training of DLAs, since the distribution of our output variable, the TTHs, presented a more similar trend to the Gaussian distribution function than the other distribution functions in the H2O package.

Besides the above-mentioned parameters, the number of hidden layers, type of activation function, and the number of neurons in the hidden layers is also important parameters, affecting the predictive ability of DLAs. In our study, the number of hidden layers ranged from

3 to 10, including the number of hidden layer alternatives of 3, 4, 5, 6, 7, 8, 9, 10. In AI studies, the ANNs with several three and more hidden layers is considered a DLA. We did not analyze more than ten hidden layers since the network model then becomes too complex to converge. The h2o.deeplearning function includes three activation functions such as Tanh, Rectifier, and Maxout. These activation functions define the nonlinear trends in the data. In this study, some preliminary analyses were performed to determine the predictive activation function of choice from these activation functions. These preliminary analyzes included various predictive evaluations between these activation functions. In these analyses, Tanh and Maxout's activation functions resulted in extremely poor predictions of TTH. Therefore, we chose the rectifier function as an alternative activation function to train the DLAs. We tried different numbers of neurons in these hidden layers, another important factor in the architecture of the DLAs, ranging from 10 to 100 by increasing by ten at each step, (10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 number of neurons). Thus, a total of 80 DLAs *(eight alternatives for the number of hidden layers X ten alternatives for the number of neurons in these hidden layers = 80 alternatives)* were trained and used to obtain the predictions for TTH.

"Underfitting" and "overfitting" are principal problems in AIT studies and should be considered when training these network models. The problem of "underfitting", caused by under-accounted prediction variance, may be overcome using nonlinear and flexible fitting abilities of these network models without including any statistical functions. However, the "overfitting" problem may be noticeable in that the training and validation datasets differ substantially in their fitting statistics. The fitting statistics for validation datasets may appear rather poor and even worse as compared to those for training datasets (Srivastava *et al.,* 2014). To deal with this "overfitting" problem, which may be a significant issue for the generalizability of network models, three preventing methods including Lasso (L1) and Ridge (L2) regularizations, convergence-based Early Stopping metrics and cross-validation procedure were used by regulating some parameters in h2o. deeplearning function in the H_20 package. In this study, other preliminary analyses based on various predictive evaluations were carried out to decide the predictive Lasso (L1) and Ridge (L2) regularizations and convergence-based Early Stopping metrics. From these alternatives for the values of regularizations and stopping metrics, preliminary analysis showed superior fitting statistics with 1×10^{-5} of both Lasso (L1) and Ridge (L2) regularizations and stopping metric specified as "RMSE" with 1×10^{-2} of stopping tolerance and 5 of stopping rounds. Also, K=10-fold cross-validation resampling techniques were used by specifying nfolds=10 in h2o.deeplearning function to minimize this "overfitting" problem in the DLAs.

Evaluation Criteria for Prediction Methods

Ten different evaluation criteria that are based on the observed and predicted TTH values obtained by the network models based on the DLAs, NLS, and NLME regression models were used to select the best predictive model from these modeling alternatives. These evaluation criteria include: (1) coefficient of correlation between observed and predicted TTHs (r), (2) average absolute error (AAE), (3) the maximum absolute error (max. AE), (4) the root mean squared error (RMSE), (5) percent root mean squared error (RMSE%), (6) the average Bias (Bias), (7) percent average Bias (Bias%), (8) the fit index (FI), (9) Akaike's information criterion (AIC), and (10) Bayesian information criterion (BIC) to compare the prediction performance of DLAs and the nonlinear regression models. These criteria are calculated as follows:

$$r = \left(\sum_{i=1}^{n} \left[(h_i - \bar{h}_i) \cdot (\hat{h}_i - \bar{\hat{h}}_i) \right] \right) / \sqrt{\sum_{i=1}^{n} (h_i - \bar{h}_i)^2 \cdot \sum_{i=1}^{n} (\hat{h}_i - \bar{\hat{h}}_i)^2} \tag{6}$$

$$AAE = \sum_{i=1}^{n} |h_i - \hat{h}_i| / n \tag{7}$$

$$max. AE = max(|h_i - \hat{h}_i| \ldots \ldots |h_n - \hat{h}_n|) \tag{8}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} (h_i - \hat{h}_i)^2 / (n-k)} \tag{9}$$

$$RMSE\% = \left(\left[\sqrt{\sum_{i=1}^{n} (h_i - \hat{h}_i)^2 / (n-k)} \right] / \bar{h}_i \right) \cdot 100 \tag{10}$$

$$Bias = \sum_{i=1}^{n} (h_i - \hat{h}_i) / n \tag{11}$$

$$Bias\% = \left(\left[\sum_{i=1}^{n} (h_i - \hat{h}_i) / n \right] / \bar{h}_i \right) 100 \tag{12}$$

$$FI = \frac{\sum_{i=1}^{n} (h_i - \hat{h}_i)}{\sum_{i=1}^{n} (h_i - \bar{h}_i)} \tag{13}$$

$$AIC = n ln(RMSE) + 2k \tag{14}$$

$$BIC = n ln(RMSE) + n ln(k) \tag{15}$$

Where, $\hat{h}_i$ is the predicted TTHs of the i_{th} tree in the sample plots by the NLS, NLME and DLA models, h_i is the observed TTHs of the i_{th} tree in the sample plots, $\underline{h}_i$ is the average of observed TTHs of the i_{th} tree in the sample plots, n is the number of trees in the sample plots, k is the number of independent variables in the models. Smaller values of AAE, max. AAE, RMSE, RMSE%, Bias, Bias%, AIC, BIC and higher values of r and FI indicate the better predictive performance of these models. When these evaluation criteria were calculated by using the predicted TTHs from NLME models, the marginal model strategy of NLME models that directly predict the population average (marginal mean) was used for the dataset of 963 trees in 107 sample plots. The reason for using the marginal model strategy for these datasets is that these

NLME models have been developed using TTH-DBH data in the sample plots without sub-sample data. However, the conditional model strategy of NLME models that require the estimation of random parameters based on the BLUP formula by using sub-sample data was used to calculate some evaluation criteria such as RMSE% and FI and for a validation data set.

To jointly evaluate all ten criteria for the general success of the prediction techniques with the NLS, NLME, and DLAs, the relative ranks proposed by Poudel & Cao (2013) were calculated in this study. This relative rank comparison procedure was used because it provides information about relative proximity. The relative rank of the prediction techniques including DLAs, NLS, and NLME is defined as:

$$R_i = 1 + \frac{(k-1) \cdot (S_i - S_{min.})}{(S_{max.} - S_{min.})} \qquad (16)$$

where R_i is the relative rank of the prediction methods (i=1, 2, 3 … k), k is the number of prediction methods with DLAs, NLS, and NLME evaluated. S_i is s the evaluation statistic value of method k, and $S_{min.}$ and $S_{max.}$ are respectively the minimum and maximum value of S_i. In this evaluation system, the best and worst predictive methods have a relative value of 1 and k, respectively. The relative ranking values of the other prediction methods (between the best and worst ones) are expressed as real numbers ranging from 1 to k. Thus, the prediction technique with the lowest sum of rank was accepted as the best predictive technique for predicting TTHs.

Equivalence Tests for Validation of Prediction Models

These prediction techniques with NLS, NLME, and DLAs were further evaluated by the equivalence test procedure (Robinson & Froese, 2004; Robinson *et al.,* 2005) using independent data as a validation dataset (169 trees in 25 sample plots).

When the TTHs of the validation data that includes 169 trees in 25 sample plots were predicted by NLME regression models, the vector of random parameters used as a calibration process was predicted by using the measurements of sub-sample trees to obtain the predictions for specific stands. In this calibration process with the conditional model strategy of the NLME model, random parameters, u_i and v_i, for a given plot were predicted using the best linear unbiased predictors, BLUPs (Lappi, 1991; Mehtätalo, 2004). The comprehensive formula and explanations for components of BLUPs equation can be found in Calama & Montero (2004). For calibration of the conditional model strategy of the NLME model to the validation dataset, the sub-sampling designs and sizes within each plot could be significant to obtain the prediction of TTHs. In this study, the sub-sample alternative based on the TTH of five medium-size trees that can be considered as closest to the quadratic mean diameter at breast-height per plot was used to calculate the random parameters for 169 trees in 25 sample plots because this sub-sample selection alternative gave the best predictive TTHs. Thus, the TTHs of these 169 trees in 25 sample plots were predicted by the conditional model strategy of the NLME models based on the calculation of random parameters. In addition to the calibration of NLME models, the TTHs of the validation dataset were predicted by using Schnute's (1981) height-diameter model with NLS and the network models with DLAs. Then, the equivalency or similarity between the predicted and the observed values were evaluated by the equivalence test (R Development Core Team, 2018).

The equivalence test is commonly applied for the evaluation of various prediction methods in forestry studies, especially in forest growth modeling studies. In performing this test, the magnitude of the region of dissimilarity between observed and predicted TTHs is very important to deciding whether the prediction method can be acceptable or not, depending on the dissimilarity value. This equivalence test involves the null hypothesis of "significant difference" between the predicted and the observed values against an alternative hypothesis of "no difference" between these values. A rejection of the null hypothesis leads to acceptance of the prediction of the TTHs in the specified study forests.

This equivalence test was carried out by regressing the relationships between observed and predicted TTHs (X: observed TTHs and Y: predicted TTHs by DLAs, NLS, and NLME models) and also regressing the regression parameters with the intercepts (b_0) and slopes (b_1) for this relation. The two one-sided test method (TOST) described in Robinson *et al.* (2005) was used to calculate the confidence intervals for the slope and intercept parameters. TOST tests the equality of slopes (b_1) to 1±10% and the equality of intercepts (b_0) to y±10%. The predictions of the confidence intervals for these parameters were obtained by using a nonparametric bootstrap procedure as described by Robinson *et al.* (2005), in which the number of bootstrap replicates was specified as 1000. These equivalence tests procedures for different prediction techniques including NLS, NLME, and DLAs were performed by using the "Regression-based TOST using bootstrap, equiv.boot" function of the "equivalence" package in the R statistical environment (R Development Core Team, 2018).

Results

The parameter estimations and variance components for Schnute's (1981) height-diameter model obtained by NLS and NLME with various random parameter alternatives are presented in Table 2. The fixed parameters of

Table 2. Parameter estimations and variance components for the Schnute (1981)'s height-diameter model obtained by NLS and NLME with various random parameter alternatives

Techniques	Parameter Type	Parameters	Estimate	Standard error	t value	Pr>|t|
NLS ϕ=0.3126	Fixed Parameters	b_0	0.02465	0.00629	3.920	<0.0001
		b_1	1.28508	0.18637	6.895	<0.0001
NLME with random b_0 alternative ϕ=0.353	Fixed Parameters	b_0	0.03930	0.00685	5.733	<0.0001
		b_1	0.91369	0.14787	6.179	<0.0001
	Variance components of random parameter	$\sigma_{u_i}^2$= 0.00066	σ^2= 0.02564			
NLME with random b_1 alternative ϕ=0.4491	Fixed Parameters	b_0	0.06687	0.00704	9.502	<0.0001
		b_1	0.34937	0.13958	2.503	<0.05
	Variance components of random parameter	$\sigma_{u_i}^2$= 0.24179	σ^2= 0.11688			
NLME with random b_0 b_1 alternative ϕ=0.4107	Fixed Parameters	b_0	0.04256	0.00685	6.213	<0.0001
		b_1	0.90629	0.13855	6.541	<0.0001
	Variance components of random parameter	$\sigma_{u_i}^2$= 0.00377	σ_u^2= 1.23969 σ^2= 0.14899 σ_{uv} =-0.965			

NLS: the nonlinear least squares, NLME: the nonlinear mixed effect, ϕ: is the variance function coefficient

these models were found to be significant at the 0.05 probability level (P < 0.05).

The goodness-of-fit statistics of r, AAE, max. AAE, RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for the DLAs, NLS and NLME are given in Table 3 and rank values (Poudel & Cao, 2013) for these prediction methods are given in Table 4. Based on the goodness-of-fit statistics (Table 3 and 4), The NLS resulted in the poorest predicti-

ve ability with higher AAE, max. AAE, RMSE, RMSE%, AIC, BIC, Bias, and Bias%, and lowest r and FI compared to all the other prediction techniques. Especially, the inclusion of random parameters to model structure with the nonlinear mixed effect model provided moderate improvements in fitting statistics. However, the network models based on DLAs with various numbers of neurons and hidden layer alternatives resulted in better fitting statistics

Table 3. Goodness-of-fit statistics of r, AAE, max. AAE, RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC, for the best predictive deep learning algorithms of the number of neuron alternatives regarding the numbers of hidden layer and nonlinear regression and nonlinear regression including various random effect parameter choices.

Number of neuron alternatives and regression models	Number of hidden layers	r	AAE	Max. AAE	RMSE	RMSE %	AIC	BIC	Bias	Bias%	FI
80 # neuron	3	0.948	0.9295	5.583	1.241	9.3297	214.42	229.0	0.0161	0.120	0.8990
100 # neuron	4	0.966	0.7224	4.857	1.011	7.6037	17.437	32.04	-0.0067	-0.0503	0.9329
100 # neuron	5	0.971	0.6606	4.612	0.944	7.0989	-48.7190	-34.1088	-0.0243	-0.1825	0.9415
80 # neuron	6	0.971	0.6599	3.999	0.936	7.0339	-57.5721	-42.9619	-0.0265	-0.1994	0.9426
100 # neuron	7	0.982	0.4855	4.144	0.740	5.5603	-283.9673	-269.357	-0.0569	-0.4275	0.9641
100 # neuron	**8**	**0.984**	**0.4188**	**3.759**	**0.694**	**5.2164**	**-345.4465**	**-330.836**	**0.028**	**0.216**	**0.9684**
70 # neuron	9	0.977	0.5667	4.183	0.838	6.3023	-163.3478	-148.737	0.047	0.356	0.9539
90 # neuron	10	0.978	0.5296	4.177	0.812	6.1022	-194.4066	-179.796	-0.0314	-0.2359	0.9568
Nonlinear Least Squares		0.887	1.2602	6.997	1.789	13.448	566.54	581.1	0.149	1.124	0.7901
NLME with b_0 random		0.932	1.0187	5.630	1.433	10.770	352.70	367.3	0.102	0.773	0.8654
NLME with b_1 random		0.930	0.9413	6.664	1.410	10.596	337.06	351.6	0.166	1.253	0.8697
NLME with b_0 and b_1 random		0.933	0.9565	5.619	1.387	10.423	321.18	335.7	-0.0957	-0.7191	0.8739

r: coefficient of correlation between observed and predicted heights, AAE: average absolute error, max. AAE: the maximum absolute error (max. AE), RMSE: the root mean squared error, RMSE%: percent of root mean squared error, Bias: the average Bias, Bias%: percent average Bias, FI: the fit index, AIC: Akaike's information criterion, and BIC: Bayesian information criterion, NLS: the nonlinear least squares, NLME: the nonlinear mixed effect

Table 4. Rank values for the best predictive deep learning algorithms of the number of neuron alternatives regarding the number of hidden layers and nonlinear regression and nonlinear regression including various random effect parameter choices.

Number of neuron alternatives and regression models	Number of hidden layers	r	AAE	Max. AAE	RMSE	RMSE %	AIC	BIC	Bias	Bias%	FI	Σ
80 # neuron	3	31.70	50.43	34.05	42.066	42.066	31.746	33.1	2.679	2.679	32.86	303.4
100 # neuron	4	16.47	30.38	20.89	24.834	24.834	2.989	5.10	1.208	1.208	17.30	145.2
100 # neuron	5	12.05	24.40	16.44	19.794	19.794	7.556	5.39	3.964	3.964	13.34	126.7
80 # neuron	6	12.05	24.34	5.353	19.146	19.146	8.848	6.65	4.316	4.316	12.85	117.0
100 # neuron	7	2.696	7.460	7.979	4.433	4.433	41.897	38.8	9.075	9.075	2.974	128.8
100 # neuron	**8**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**50.872**	**47.5**	**4.676**	**4.676**	**1.000**	**113.8**
70 # neuron	9	7.182	15.31	8.678	11.841	11.841	24.289	21.6	7.593	7.593	7.661	123.6
90 # neuron	10	6.09	11.72	8.580	9.844	9.844	28.823	26.1	5.077	5.077	6.340	117.5
Nonlinear Least Squares		84.00	82.44	59.68	83.184	83.184	83.148	83.1	23.61	23.618	82.82	688.8
NLME with b_0 random		47.12	51.57	53.65	54.716	54.716	49.649	50.5	26.29	26.298	46.31	460.8
NLME with b_1 random		45.48	59.07	34.90	56.448	56.448	51.932	52.7	16.29	16.290	48.28	437.9
NLME with b_0 and b_1 random		44.49	53.05	34.71	52.985	52.985	47.330	48.2	15.15	15.157	44.37	408.5

(lower total ranks ranging from 113.808 to 303.410) than those by NLS and NLME with b0 and b1 random alternatives, ranging from 408.531 to 688.885. These results suggest that the DLAs outperformed the regression techniques including NLS and NLME with various random parameter alternatives. The DLAs with 100 # neurons and 8 # hidden layer gave significant improvements in the error values of AAE, max. AAE, RMSE, RMSE%, AIC, BIC, Bias, and Bias%, and a noticeable increase in r and FI than those by the NLS and NLME (Table 3 and 4). The DLAs with 100 # neurons and 8 # hidden layer resulted in the best predictive ability with a r-value of 0.9842, AAE value of 0.4188, the Max. AAE value of 3.7598, RMSE value of 0.6942, RMSE% of 5.2164 values, AIC value of -345.4465, BIC value of -330.836, Bias value of 0.0288, Bias% value of 0.2166% and FI value of 0.9684 for TTHs (Table 3).

Fig. 2 and Fig. 3. show the variation of RMSE% and FI of 80 trained DLAs with a various number of neurons and hidden layers. An increased number of hidden layers generally resulted in decreased values for the mean of RMSE% and increased values for FI, except the 7, 9, and 10 # layers for RMSE% and the 7 and 10 # layers for FI (Table S1 [suppl.]). Also, there are no significant differences in RMSE% value for some numbers of hidden layers, namely 6 (Fig. 2). In terms of FI values, lower FI values were obtained for the 3, 4, and 5 numbers of hidden layers than others and more than 6 numbers of hidden layers have similar FI indices (Fig. 3). The network model with 9 # layer gave the best predictive results with an average of fitting criteria including AAE (0.7205), max. AAE (5.3849), RMSE (1.0372), RMSE% (7.7932), r (0.9632), FI (0.9264), and sum of rank (10.614) (Table S1 [suppl.]. Correspondingly, an increase in the num-

ber of neurons from 10 to 100 resulted in a substantial improvement in predictive ability, which was shown by decreased RMSE% and increased FI (Table S2 [suppl.]. Although there is a general increase from the RMSE% value according to the increase in the number of neurons, this change has a fluctuating trend (Fig. 2 and Fig. 3). The network model with 100 # neurons provided the best average predictive results with AAE (0.6205), max. AAE (5.0593), RMSE (0.9184), RMSE% (6.9006), r (0.9719), FI (0.9432), and sum of rank (6.290) (Table S2 [suppl.].

The relationships obtained between observed and predicted TTH values by network models based on the DLAs including (a) 80 # neurons in three hidden layer, (b) 100 # neurons in four hidden layer, (c) 100 # neurons in five hidden layer, (d) 80 # neurons in six hidden layer, (e) 100 # neurons in seven hidden layer, (f) 100 # neurons in ei-
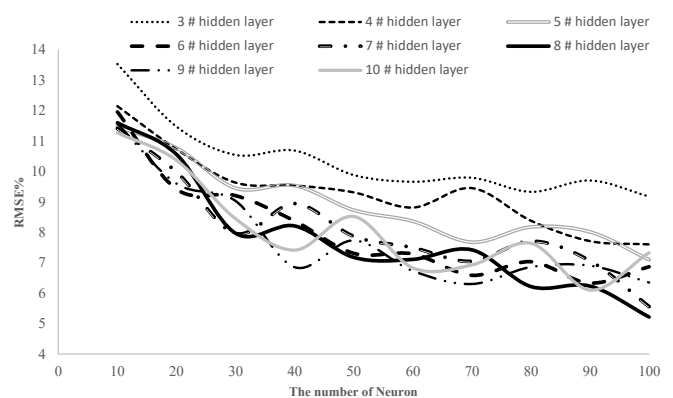

Figure 2. The change of RMSE% of a total of 80 trained Deep Learning Algorithms according to various the number of neurons and hidden layers.
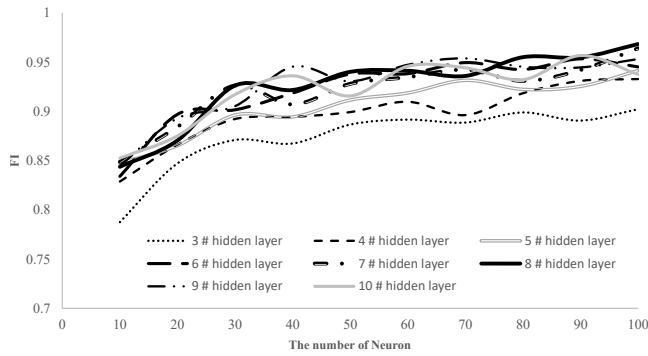
Figure 3. The change of FI of a total of 80 trained Deep Learning Algorithms according to various the number of neurons and hidden layers.

ght hidden layer, (g) 70 # neurons in nine hidden layer, (h) 90 # neurons in ten hidden layer, (i) by the nonlinear regression, (k) by the NLME with b_0 random, (l) by the NLME with b_1 random and (m) by the NLME with b_0 and b_1 random were shown in Fig. 4. As seen in this graph, these network models with DLAs that include various alternatives of the numbers of neurons and hidden layers incline to an angle of 45 degrees with axes. These graphs depict that the predictions of TTH by the network models based on the DLAs have better predictive ability than those of NLS and NLME with random parameter alternatives models, which were shown by more tidy coalescence of predicted and measured values around the 1:1-line.

The results of the equivalence test including predicted bootstrap b0 and b1 limits with RMSE% and FI values for validation data are presented in Table 5. For all prediction techniques with NLS, NLME, and DLAs the null hypothesis of dissimilarity for intercept parameters (b0) was rejected based on equivalence tests, in which

the bootstrap intercept (b0) falls within the equivalent regions, $\overline{y} \pm 10\%$. The null hypothesis of dissimilarity for slope (b1) parameters was rejected by equivalence tests for some DLA models such as 80 # neurons in three hidden layers, 100 # neurons in four hidden layers, 100 # neurons in eight hidden layers, 70 # neurons in nine hidden layers, for which the bootstrap b1 are contained within the equivalent regions, 1±10%. However, other network models, NLS, and NLME models were not validated because these bootstrap slopes (b1) did not fall within these equivalent regions (1±10% for the slope). As a result, the network model based on the DLAs with 100 # neurons in eight hidden layers could be achieved as the best predictive network mode of all the prediction techniques l based on the equivalence tests and related fitting statistics (Table 5).

Discussion

This study investigates whether the network models based on DLAs can be used as alternative techniques to predict the relationships between the total tree height and diameter at breast height for Turkish Pine stands. Many different statistical prediction models have been proposed and used to predict and model this relation in the literature. Besides developing a wide variety of regression models with statistical considerations, artificial intelligence techniques are innovative prediction methods that can be predicated on non-requirement of statistical assumptions with powerful nonlinear modeling ability without using statistical functions. Diamantopoulou & Özçelik (2012) and Özçelik *et al.* (2013) developed the ANN models

Table 5. Results of equivalence tests with RMSE% and FI for the best predictive deep learning algorithms of the number of neuron alternatives regarding the numbers of hidden layer and nonlinear regression and nonlinear regression including various random effect parameter choices

| Number of neuron alternatives and regression models	Number of hidden layers	b_0 limits		Bootstrap b_0 limits			b_1 limits		Bootstrap b_1 limits			RMSE%	FI
		lower	upper	lower	upper	H_0: not Equivalent	lower	upper	lower	upper	H_0: not Equivalent		
80 # neuron3	3	12.2661	15.0218	13.5854	13.9981	Rejected	0.900	1.100	0.9002	1.0306	Rejected	9.404	0.871
100 # neuron	4	12.2661	15.0218	13.5686	13.9781	Rejected	0.900	1.100	0.90297	1.0248	Rejected	8.892	0.885
100 # neuron	5	12.2661	15.0218	13.5696	13.9665	Rejected	0.900	1.100	1.0030	1.1485	Not rejected	9.284	0.875
80 # neuron	6	12.2661	15.0218	13.5666	13.9727	Rejected	0.900	1.100	1.0292	1.1689	Not rejected	9.3689	0.872
100 # neuron	7	12.2661	15.0218	13.5737	13.9798	Rejected	0.900	1.100	0.97507	1.1238	Not rejected	9.826	0.859
100 # neuron	**8**	12.2661	15.0218	13.5802	13.9783	Rejected	0.900	1.100	0.9359	1.0659	Rejected	9.042	0.881
70 # neuron	9	12.2661	15.0218	13.5899	13.97471	Rejected	0.900	1.100	0.9193	1.0511	Rejected	9.180	0.877
90 # neuron	10	12.2661	15.0218	13.5744	13.9994	Rejected	0.900	1.100	0.9608	1.1079	Not rejected	9.733	0.862
Nonlinear Least Squares		12.2661	15.0218	13.2176	13.8244	Rejected	0.900	1.100	0.7759	0.9362	Not rejected	16.052	0.688
NLME with b_0 random		12.2661	15.0218	13.2181	13.8234	Rejected	0.900	1.100	0.7635	0.9227	Not rejected	15.550	0.708
NLME with b_1 random		12.2661	15.0218	13.2044	13.8177	Rejected	0.900	1.100	0.7255	0.8788	Not rejected	16.022	0.689
NLME with b_0 and b_1 random		12.2661	15.0218	13.2293	13.8144	Rejected	0.900	1.100	0.7711	0.9195	Not rejected	15.305	0.717
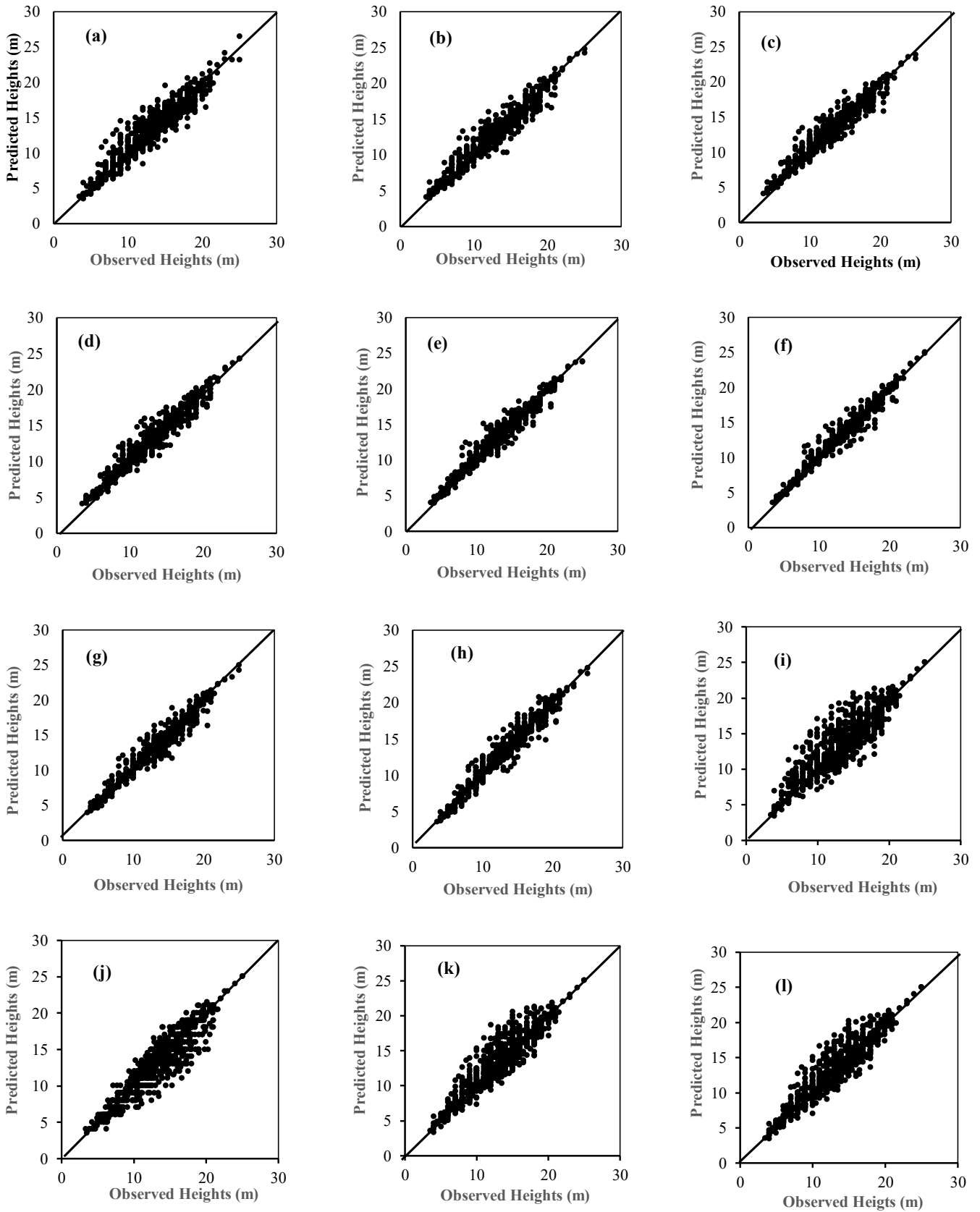
Figure 4. The relationships between observed (x-axis) and predicted heights (y-axis) by the best predictive deep learning network models: (a) 80 # neurons in three hidden layers, (b) 100 # neurons in four hidden layers, (c) 100 # neurons in five hidden layers, (d) 80 # neurons in six hidden layers, (e) 100 # neurons in seven hidden layers, (f) 100 # neurons in eight hidden layers, (g) 70 # neurons in nine hidden layers, (h) 90 # neurons in ten hidden layer, (i) by the nonlinear regression, (j) by the NLME with b_0 random, (k) by the NLME with b_1 random and (l) by the NLME with b_0 and b_1 random.

for predicting tree heights and compared these network models with the nonlinear regression models. Diamantopoulou & Özçelik (2012) showed that the generalized regression neural network models were superior to nonlinear growth functions, providing the highest estimation ability with RMSE% values from 5.69 to 7.13, for all different tree species. Özçelik *et al.* (2013) showed that back-propagation neural network modeling approaches with the diameter variance of each plot gave the most precise results with a 23.7% reduction in RMSE compared with the regression model. In addition to the ANN models predicting tree heights, Ercanlı (2020) compared the ANN models with the network models based on the DLA to predict the tree heights; this study found that the DLA models were superior to both the ANN and nonlinear regression models in terms of various evaluation criteria. In the present study, a total of 80 DLAs with various combinations of hidden layers and neurons were trained and then evaluated by comparing traditional prediction models such as nonlinear regression models and nonlinear mixed-effect regression models including various random effect parameter choices. The DLAs with 100 # neurons and 8 # hidden layers resulted in the best predictive fitting statistics with an RMSE value of 0.6942, RMSE% of 5.2164 values, and FI value of 0.9684. In this regard, this neural network model based on the DLAs accounted for at least 97 (about % 96.84) % of the total variation in tree heights. The network models based on DLAs outperformed the regression techniques, which were evidenced by the evaluation criteria values of AAE, max. AAE, RMSE, RMSE%, Bias, Bias%, AIC, BIC, r, and FI (Table 3 and 4). This best predictive DLA network model with 100 # neurons and 8 # hidden layers resulted in the reductions of AAE, max. AAE, RMSE, RMSE%, Bias and Bias% by 66.767 %, 46.267 %, 61.214 %, 61.211 %, 80.762 %, 80.745 %, and in the increase of r and FI by 10.933 % and 22.567 % as compared to those by NLS. This predictive ability of the DLA network models can be distinguished in Fig. 4 that depicts that the predicted and observed values for the combination of 100 # neurons and 8 hidden layers were highly close to the 1:1-line. Importantly, these improvements in TTH predictions by DLA network models suggest that these DLA network models can be an alternative to nonlinear regression techniques such as NLA and NLME. In addition to the results of DLA prediction of tree heights obtained by the Ercanlı (2020) and the present study, Lee *et al.* (2015), Mohanty *et al.* (2016), Sladojevic *et al.* (2016), Sun *et al.* (2017), Ferentinos (2018) and Ubbens *et al.* (2018) successfully developed prediction systems based on the DLA to specify plant disease diagnosis in agricultural applications. Also, similar improvements in the predictive ability for various individual and stand variables were achieved by those from Hasenauer *et al.* (2001); Diamantopoulou (2005a, 2005b, 2006); Özçelik *et al.* (2008); Diamantopoulou & Milios (2010); Özçelik *et al.* (2010); Leite *et al.* (2011); Ashraf *et al.* (2013); Diamantopoulou *et al.* (2015); Özçelik *et al.* (2017), especially tree height predictions by Diamantopoulou & Özçelik (2012) and Özçelik *et al.* (2013). Unlike those studies which evaluated the ANN models, this present study evaluated the network models based on the DLA for predicting TTH sampled from forest areas. However, the DLA models have been ignored in studies about forestry growth and yield modeling within different artificial intelligence techniques and remained an unstudied prediction technique in forest biometric literature. This study presented a contribution to the evaluation of DLA models to predict the height-diameter relation and is a novel artificial intelligence application.

Significant problems to be considered in the development of various artificial intelligence models involving the ANN or DLA applications are "underfitting" and "overfitting". The "underfitting" problem can be solved by prevailing nonlinear modeling ability of the deep learning network models, which have been verified by numerous studies about forest growth modeling. However, the "overfitting" problem may be the weakness in deep learning network models in simulation and future applications for other datasets. This "overfitting" problem is characterized by the instance that although the network models that are trained by the ANN or DLA techniques give better predictive ability, it may result in poor fitting for validation dataset or another dataset in some studies. In this study, three main methods including Lasso (L1) and Ridge (L2) regularizations, convergence-based Early Stopping metrics, and especially cross-validation were used to overcome the "overfitting" problem in the validation dataset. The "overfitting" problem was checked and evaluated by the equivalence tests with a nonparametric bootstrap procedure for the validation dataset. The results showed that the validation statistics of percent RSME and FI calculated for the validation dataset were close to those calculated for the training dataset (Table 5). Also, the results of equivalence tests for some deep learning network models such as 80 # neurons in three hidden layers, 100 # neurons in four hidden layers, 100 # neurons in eight hidden layers, 70 # neurons in nine hidden layers were validated. These suggest that Lasso (L1) and Ridge (L2) regularizations, convergence-based Early Stopping metrics, and cross-validation were effective in dealing with the "overfitting" problem. However, more scientific research including different datasets and evaluation processes may be required to generalize the results about these regulations and metrics for the "overfitting" problem. In this regard, it may be suggested that these methods for overcoming the "overfitting" problem should be included in training procedures in deep learning network models in further studies.

The results of DLA network models can be further evaluated to decide optimum DLA network architecture from some alternatives for the numbers of hidden layers and the number of neurons in these layers. Especially, the increase in the number of neurons that positively affect fitting statistics for TTH prediction with low values for AAE, max. AAE, RMSE, RMSE%, Bias, Bias%, AIC, and BIC and higher values for r and FI. Thus, the number of neurons can be an important and effective parameter on fitting statistics of the predictions from network models. When the changes of these fitting statistics are joined with an increase in the number of neurons that are evaluated, this change is the fluctuation trend (Fig 2 and Fig. 3). Furthermore, the increase in the number of neurons may result in a more sophisticated network structure, and these network models require a longer iteration time for convergence. Besides the change in the number of neurons, the results concerning the number of hidden layers show that the fit statistics according to the number of the hidden layers had an unstable behavior, independent from the change in the number of neurons (Tables S1 and S2 [suppl.]. For example, the 9 # hidden layer gave the best predictive statistics, while unexpectedly, the 10 # hidden layer gave quite poor-fitting results. Also, differences between some predictive fitting statistics in terms of RMSE% and FI have been achieved at low levels for more than 6 hidden layer numbers (Fig. 2 and Fig. 3). Moreover, equivalence tests show that network models comprising the 3, 4, 7, and 8 # hidden layers were validated, while the rest of the hidden layers were invalidated. These results can be explained by the fact that training with too many hidden layers does not contribute much to the predictive ability of these network models or that the computer hardware where the network models are trained is insufficient in the training of many hidden layer network models. The combinations of the number of hidden layers and neurons may be an important factor that affects the prediction performance of AI techniques. In this regard, the optimum number of hidden layers and neurons should be decided by evaluating fitting statistics and residual values for various alternatives of the numbers of hidden layers with the assistance of the results of the equivalence test.

In addition to evaluating the predictive ability of AI models with various statistical methods, especially regression models, it is important to present various computer syntax files to ensure the usability of these AI prediction systems. Traditional regression models contain a limited number of model parameters so that they can be disseminated directly in publications, whereas AI models can include thousands or even tens of thousands of weight values with the increase in the number of neurons in hidden layers. When going through a translation period the trained network model is run with various software for predicting various tree and stand variables. This is done by downloading these network models from an ad-

ditional file of publications from current applications that can be applied to the regression models by obtaining the parameter values of these models from the publications. As the applications of AI models become more defined in every field, it will not be possible to not utilize this tool regarding forestry growth and yield modeling studies. This study presents the R syntax file of the best predictive DLA network model with 8 layers and 100 neurons as a supplementary file (R syntax source codes for using the best predictive network model based on the DLAs to obtain the predictions of TTH for other datasets are in the Appendix A1 of this study) and the downloadable link can be found from Google Drive Link (https://drive. google.com/open?id=1byX88K9OvnuQMgGv_y3b2ze-Tw7bQO1wN) so that future forest practitioners can use this best predictive DLA model to predict TTH for their specified forest areas. Because the R platform is free and open-source code, this trained network model based on DLAs has been presented on the R platform for various stakeholders and other users in forest management. Therefore, it is possible to use this network model by future forest practitioners to obtain the total tree height predictions for specified forest areas.

Conclusions

In this study, the network models based on DLAs were trained to predict the relationships between TTH and DBH sampled from Kestel Forest. This is a pioneer study of innovative DLAs, a type of AI technique. This study evaluated whether this new AI technique could be an alternative prediction technique to conventional regression models including the nonlinear regression models and nonlinear mixed effect models. The fitting results showed that the deep learning network models resulted in better TTH predictions than those by nonlinear regression models, based on the nonlinear squares, and nonlinear mixed-effect regression models based on different random parameter alternatives. The DLA resulted in considerably better improvements in the predictive ability for TTH than those by NLS and NLME. Also, the best predictive DLA models were validated, while none of the nonlinear regression models were validated, indicating that the DLA proved successful in predicting individual tree TTHs and that it may be preferred over regression models. However, the DLAs should also be validated with different datasets representing different forest areas to improve the portability of the developed models in predicting TTHs. The models should be tested to evaluate their performance in predicting other attributes (e.g. volume, increment, tree taper, etc.) of individual trees and forests. The results showed that highly different layers and neuron combinations should be tested before deciding the most proper structure of the model to predict TTH. Recently

developed alternative computer-based software and platforms of H$_2$0, an open-source AI platform coded in the R core programming language, was successfully used in this study. Thus, these network models may significantly contribute to studies in forest biometry, forest growth and yield modeling, forest planning, and silviculture.

Acknowledgements

References

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. https://doi.org/10.1139/cjfr-2013-0090

Bredenkamp BV, Gregoire TG, 1988. A forestry application of Schnute's generalized growth function. For Sci 34: 790-797.

Bronisz K, Mehtätalo L, 2020. Mixed-effects generalized height-diameter model for young silver birch stands on post-agricultural lands. Forest Ecol Manag 460: 117901. https://doi.org/10.1016/j.foreco.2020.117901

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. https://doi.org/10.1139/x03-199

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. https://doi.org/10.1186/s12862-017-1014-z

Castedo-Dorado F, 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. https://doi.org/10.1051/forest:2005042

Castedo-Dorado F, 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. https://doi.org/10.1016/j.foreco.2006.04.028

Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL, 1983. Timber management: a quantitative approach. John Wiley & Sons, Inc.

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. https://doi.org/10.1016/j.foreco.2009.11.036

Crookston NL, Dixon GE, 2005. The forest vegetation simulator: A review of its structure, content, and applications. Comput Electron Agr 49: 60-80. https://doi.org/10.1016/j.compag.2005.02.003

Curtis RO, 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. For Sci 13: 365-375.

Diamantopoulou M, 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. https://doi.org/10.1016/j.biosystemseng.2009.11.010

Diamantopoulou MJ, 2005a. Artificial neural networks as an alternative tool in pine bark volume estimation. Comput Electron Agr 48: 235-244. https://doi.org/10.1016/j.compag.2005.04.002

Diamantopoulou MJ, 2005b. Predicting fir trees stem diameters using artificial neural network models. South Afr For J 205: 39-44. https://doi.org/10.2989/10295920509505236

Diamantopoulou MJ, 2006. Tree-bole volume estimation on standing pine trees using cascade correlation artificial neural network models. Agr Eng Int Vol. VIII. June: 1-14.

Diamantopoulou MJ, Özçelik R, 2012. Evaluation of different modeling approaches for total tree-height estimation in Mediterranean Region of Turkey. Forest Syst 21: 383-397. https://doi.org/10.5424/fs/2012213-02338

Diamantopoulou MJ, Özçelik R, Crecente-Campo F, Eler Ü, 2015. Estimation of Weibull function parameters for modelling tree diameter distribution using least squares and artificial neural networks methods. Biosyst Eng 133: 33-45. https://doi.org/10.1016/j.biosystemseng.2015.02.013

Ercanlı İ, 2020. Innovative deep learning artificial intelligence applications for predicting relationships between individual tree height and diameter at breast height. For Ecosyst 7: 1-18. https://doi.org/10.1186/s40663-020-00226-3

Ercanlı İ, Bolat F, Yavuz H, 2018. Ormanların Çap Dağılımlarının Modellenmesinde Derin Öğrenme Algoritmalarının Kullanımı: Trabzon ve Giresun Ormanları Doğu Ladini-Sarıçam Karışık Meşcereleri Örneği. Anadolu Orman Araştırmaları Dergisi 4: 122-132.

Ferentinos KP, 2018. Deep learning models for plant disease detection and diagnosis. Comput Electron Agr 145: 311-318. https://doi.org/10.1016/j.compag.2018.01.009

Fulton MR, 1999. Patterns in height-diameter relationships for selected tree species and sites in eastern Texas. Can J For Res 29: 1445-1448. https://doi.org/10.1139/x99-103

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. https://doi.org/10.1139/x95-017

Gregorie TG, 1987. Generalized error structure for forestry yield models. For Sci 33: 423-444.

Hasenauer H, Merkl D, Weingartner M, 2001. Estimating tree mortality of Norway spruce stands with neural networks. Adv Environ Res 5: 405-414. https://doi.org/10.1016/S1093-0191(01)00092-2

Huang S, Meng SX, Yang Y, 2009. Using nonlinear mixed model technique to determine the optimal tree height prediction model for black spruce. Mod Appl Sci 3: 3-18. https://doi.org/10.5539/mas.v3n4p3

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. https://doi.org/10.1016/S0378-1127(99)00151-6

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. https://doi.org/10.1139/x92-172

Judge GG, Hill RC, Griffiths W, Lutkepohl H, Lee TC, 1982. Introduction to the Theory and Practice of Econometrics. John Wiley and Sons Ltd.

Lappi J, 1991. Calibration of height and volume equations with random parameters. For Sci 37: 781-801.

Lappi J, 1997. A longitudinal analysis of height/diameter curves. For Sci 43: 555-570.

Lee SH, Chan CS, Wilkin P, Remagnino P, 2015. Deepplant: Plant identification with convolutional neural networks. In, 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 452-456. https://doi.org/10.1109/ICIP.2015.7350839

Lei Y, 1998 Modelling forest growth and yield of Eucalyptus globulus Labill in central-interior Portugal. In. Universidadede Trás-os-Montes e Alto Douro, Vila Real, Portugal, p. 155.

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 For Res 130: 263-269. https://doi.org/10.1007/s10342-010-0427-7

Martin FC, Flewelling JW, 1998. Evaluation of tree height prediction models for stand inventory. West J Appl For 13: 109-119. https://doi.org/10.1093/wjaf/13.4.109

Mehtätalo L, 2004. A longitudinal height-diameter model for Norway spruce in Finland. Can J For Res 34: 131-140. https://doi.org/10.1139/x03-207

Mehtätalo L, de-Miguel S, Gregoire TG, 2015. Modeling height-diameter curves for prediction. Can J For Res 45: 826-837. https://doi.org/10.1139/cjfr-2015-0054

Mohanty SP, Hughes DP, Salathé M, 2016. Using deep learning for image-based plant disease detection. Front Plant Sci 7: 1419. https://doi.org/10.3389/fpls.2016.01419

Mugasha WA, Mauya E, Njana A, Karlsson K, Malimbwi R, Ernest S, 2019. Height-Diameter Allometry for Tree Species in Tanzania Mainland. Int J For Res 2019. https://doi.org/10.1155/2019/4832849

Nanos N, Calama R, Montero G, Gil L, 2004. Geostatistical prediction of height/diameter models. Forest Ecol Manag 195: 221-235. https://doi.org/10.1016/j.foreco.2004.02.031

Ng'andwe P, Chungu D, Yambayamba AM, Chilambwe A, 2019. Modeling the height-diameter relationship of planted Pinus kesiya in Zambia. Front Plant Sci 447: 1-11. https://doi.org/10.1016/j.foreco.2019.05.051

Nong M, Leng Y, Xu H, Li C, Ou G, 2019. Incorporating competition factors in a mixed-effect model with random effects of site quality for individual tree above-ground biomass growth of Pinus kesiya var. langbianensis. Nz J Forestry Sci 49. https://doi.org/10.33494/nzjfs492019x27x

Özçelik R, Cao QV, Trincado G, Göçer N, 2018. Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in Turkey. Forest Ecol Manag 419: 240-248. https://doi.org/10.1016/j.foreco.2018.03.051

Özçelik R, Diamantopoulou MJ, Brooks JR, Wiant Jr HV, 2010. Estimating tree bole volume using artificial neural network models for four species in Turkey. J Environ Manage 91: 742-753. https://doi.org/10.1016/j.jenvman.2009.10.002

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. https://doi.org/10.1016/j.foreco.2013.06.009

Özçelik R, Diamantopoulou MJ, Eker M, Gürlevık N, 2017. Artificial Neural Network Models: An Alternative Approach for Reliable Aboveground Pine Tree Biomass Prediction. For Sci 63: 291-302. https://doi.org/10.5849/FS-16-006

Özçelik R, Diamantopoulou MJ, Wiant Jr HV, 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.

Paulo JA, Tome J, Tome M, 2011. Nonlinear fixed and random generalized height-diameter models for Portuguese cork oak stands. Ann For Sci 68: 295-309. https://doi.org/10.1007/s13595-011-0041-y

Pinheiro J, Bates D, 2000. Mixed-Effects Models in S and S-PLUS Springer-Verlag, New York. https://doi.org/10.1007/978-1-4419-0318-1

Poudel KP, Cao QV, 2013. Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. For Sci 59: 243-252. https://doi.org/10.5849/forsci.12-001

R Development Core Team, 2018 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. In.

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. https://doi.org/10.1093/treephys/25.7.903

Robinson AP, Froese RE, 2004. Model validation using equivalence tests. Ecol Model 176: 349-358. https://doi.org/10.1016/j.ecolmodel.2004.01.013

Sánchez CAL, Varela JG, Dorado FC, Alboreca AR, Soalleiro RR, González JGÁ, Rodríguez FS, 2003. A height-diameter model for Pinus radiata D. Don in Galicia (Northwest Spain). Ann For Sci 60: 237-245. https://doi.org/10.1051/forest:2003015

Schnute J, 1981. A versatile growth model with statistically stable parameters. Can J Fish Aquat Sci 38: 1128-1140. https://doi.org/10.1139/f81-153

Schröder J, González JGÁ, 2001. Comparing the performance of generalized diameter-height equations for maritime pine in Northwestern Spain. FCUTFJ 120: 18-23. https://doi.org/10.1007/BF02796077

Searle S, Casella G, McCulloch CJINY, 1992. Variance components. A John Wiley & Sons. https://doi.org/10.1002/9780470316856

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. https://doi.org/10.1016/j.foreco.2007.05.006

Sharma M, Yin Zhang S, 2004. Height-diameter models using stand characteristics for Pinus banksiana and Picea mariana. Scand J For Res 19: 442-451. https://doi.org/10.1080/02827580410030163

Sharma R, Vacek Z, Vacek S, 2016. Nonlinear mixed effect height-diameter model for mixed species forests in the central part of the Czech Republic. J For Sci 62: 470-484. https://doi.org/10.17221/41/2016-JFS

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 Neurosc 2016. https://doi.org/10.1155/2016/3289801

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R, 2014. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15: 1929-1958.

Sun Y, Liu Y, Wang G, Zhang H, 2017. Deep learning for plant identification in natural environment. Comput Intel Neurosc 2017. https://doi.org/10.1155/2017/7361042

Temesgen H, Gadow KV, 2004. Generalized height-diameter models-an application for major tree species in complex stands of interior British Columbia. Eur J For Res 123: 45-51. https://doi.org/10.1007/s10342-004-0020-z

Trincado G, VanderSchaaf CL, Burkhart HE, 2007. Regional mixed-effects height-diameter models for loblolly pine (Pinus taeda L.) plantations. Eur J For Res 126: 253-262. https://doi.org/10.1007/s10342-006-0141-7

Turkey Meteorological Service, 2017 Meteotogical Measurements of Kestel station (Bursa, Turkey) of Turkish State Meteorological Service. In, Turkish State Meteorological Service Publications.

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 Methods 14: 6. https://doi.org/10.1186/s13007-018-0273-z

Van Laar A, Akça A, 2007. Forest mensuration. Springer Science & Business Media. https://doi.org/10.1007/978-1-4020-5991-9

West P, Ratkowsky D, Davis A, 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. Forest Ecol Manag 7: 207-224. https://doi.org/10.1016/0378-1127(84)90068-9

Wykooff WR, Crookston NL, Stage AR, 1982. Users's guide to the Stand Prognosis Model. USDA Forest Service, Intermountain Forest and Range Expt. Sta., Gen. Tech. Re INT-133, p 112. https://doi.org/10.5962/bhl.title.109367