Anadolu Orman Araştırmaları Dergisi
Anatolian Journal of Forest Research

http://dergipark.gov.tr/ajfr

An Application of R Software Model Based on Deep Learning Algorithms to Provide Future Use of Other Forest Practitioner for Predicting Individual Tree Height

İ. Ercanlı

Çankırı Karatekin University, Department of Forest Engineering, 18200, ÇANKIRI, TURKEY

ARTICLE INFO

Recieved : October 13rd, 2019
Accepted : December 30th, 2019

*e-mail of corresponding author: ilkerercanli@karatekin.edu.tr

ABSTRACT

In this study, the artificial intelligence models based on Deep Learning Algorithms were developed to model the relationships between the individual tree total heights (ITH) and diameter at breast heights (DBH) with the stand variables. The H20 package with an h2o.deeplearning function, which have been coded in R software language, was used to train these DLA models and obtain the ITH predictions. To determine best predictive input variables, various input variable alternatives were evaluated based on the statistical fitting criteria. From these fitting statistics for the training data set, the DLA model, which includes the input variables with the DBH, dominant diameter (cm), dominant height (m), number of trees per hectare and basal area (m^2/ha) resulted in the best predictive statistics with a RMSE value of 0.7173, RMSE% value of 4.5986, the AIC value of -291.3037, BIC value of 1158.4564, FI of 0.9785 values, AAE value of 0.4311, Bias value of 0.0438 and Bias% value of 0.2805. Similar to the fitting statistics in training data, the DLA model which includes the input variables with the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare and basal area (m^2/ha) for the validation data set gave the best predictive statistics with a RMSE value of 1.8217, RMSE% value of 10.2151, the AIC value of 99.9615, BIC value of 331.3772, FI value of 0.8334, AAE value of 1.2051, Bias value of -0.0985 and Bias% value of -0.5521. The R software platform, which is free and open for all, was used to train these DLA models, and also this network model was shared with various stakeholders and other users in forest management. Thus, besides the modeling studies including the comparison of various network models with classical regression models, the opportunity to share other forest practitioner to use artificial intelligence model developed in this study can be achieved by downloading this best predictive DLA model from Google Drive Link (https://drive.google.com/open?id=1OD9HVFqrur8bXQOgt2Rprn5vfhZEHq1f).

Keywords: Individual Tree Height, Deep Learning Algorithms, R software, H20 package.

Tek Ağaç Boylarını Tahmin Eden Derin Öğrenme Algoritması Temelli R Yazılım Modellerinin Diğer Kullanıcılar ile Paylaşılmasına İlişkin Bir Uygulama

ÖZ

Bu çalışmada, meşcere özellikleri ile birlikte ağaçların göğüs çapları ile boyları arasındaki ilişkileri modellemek üzere Derin Öğrenme Algoritmalarına dayanan Yapay Zeka Modelleri geliştirilmiştir. Tek ağaç boylarının elde edilmesinde ve Derin Öğrenme Algoritmalarının eğitilmesinde, R yazılım dili ile kodlanmış h2o.deeplearning fonkisyonunu iceren H20 paketi kullanmıştır. Tahminlerin elde edilmesinde en başarılı olan girdi değişkenlerinin belirlenmesinde, çeşitli istatistiksel başarı kriterlerine dayanarak değişik girdi değişkeni alternatifleri karşılaştırlmıştır. Eğitim seti için elde edilen başarı kriterleri değerlendirldiğinde; göğüs çapı, dominant çapı, dominant boyu, meşcere ağaç sayısını ve meşcere göğüs yüzeyini giriş değişkeni olarak iceren Derin Öğrenme Algoritma Modeli en başarılı sonuçları vermiştir (RMSE = 0.7173, RMSE% = 4.5986, the AIC = -291.3037, BIC = 1158.4564, FI = 0.9785, AAE = 0.4311, Bias = 0.0438, Bias% = 0.2805). Eğitim verilerine benzer olarak, bağımsız verilerle yapılan analiz sonuçlarına göre Derin Öğrenme Algoritma Modeli en başarılı sonuçlar elde edilmiştir (RMSE = 1.8217, RMSE% = 10.2151, AIC

Citation:

Ercanlı, İ., 2019. An Application of R Software Model Based on Deep Learning Algorithms to Provide Future Use of Other Forest Practitioner for Predicting Individual Tree Height. Anatolian Journal of Forest Research 5(2): 117-126.

= 99.9615, BIC = 331.3772, FI = 0.8334, AAE = 1.2051, Bias = -0.0985, Bias% = -0.5521). Bu derin Öğrenme Algoritma Modellerinin eğitiminde, ücretsiz ve herkesin kullanımına acık R yazılım platformu kullanılmış ve en başarılı yapay zeka modeli çeşitli paydaşlar ve orman amenajmanı planlayıcılarına sunulmuştur. Böylece, klasik regresyon modelleri ile değişik yapay sinir ağ modelleriyle karşılaştırlmasını iceren çalışmaların ötesinde, Google Drive Linki (https://drive.google.com/open?id=1OD9HVFqrur8bXQOgt2Rprn5vfhZEHq1f) olarak sunulan R dosyası indirilerek bu çalışmada geliştirilen yapay zeka modelinin diğer orman mühendisliği kullanıcıları ile paylaşılma fırsatı sağlanmıştır.

Anahtar Kelimeler: Tek Ağaç Boyu, Derin Öğrenme Algoritması, R yazılımı, H20 Paketi.

1. Introduction

Individual tree heights are important individual tree measurements used for total and merchantable volume predictions, growth and yield modeling, and site index predictions (Clutter et al., 1983; Van Laar and Akça, 2007). Obtaining individual tree total heights (ITH) is complex, tedious, and time-consuming in forest inventory application. Therefore, the ITH is generally measured in a limited number of trees in sample plots (Van Laar and Akça, 2007; von Gadow and Hui, 1999). The ITH of other trees which are not measured in forest inventory can be predicted by using statistical models including easy-to-measure DBH variable. In forest inventory practices, many statistical equations, which have been developed to model the relationships between tree height and diameter at breast height by using the regression analysis have been used to estimate these individual tree heights from DBH measurements (Martin and Flewelling, 1998; Huang et al., 1992).

These statistical ITH-DBH equations have been developed by using various statistical regression model forms from nonlinear regression models to nonlinear mixed effect models. These statistical regression models have been used and evaluated to acquire accurate height predictions which are considered as effective forest inventory tools for the prediction of height by using DBH as the predictor variable (Nanos et al., 2004). Beyond these statistical regression models, of which their limitation due to statistical assumptions have been criticized continuously, Artificial Intelligence Techniques have become their alternatives as they need no statistical assumption and their flexibility in modeling nonlinear relations. Deep Learning Algorithms (DLA), which is an application of Artificial Intelligence (AI) Techniques, have become more pronounced since 2010. In recent years, Deep Learning Algorithms (DLA) have offered an advanced modeling approach with innovative modeling abilities to obtain the predictions of some of tree and stand attributes. However, the structure of Artificial Neural Networks (ANNs) with one or two hidden layers may be insufficient in modeling complex nonlinear data structure. DLAs can comprise more layers, especially greater number of hidden layers, which makes them more efficient modeling techniques. The network models based on the DLAs contain their advanced computational systems based on the Graphical Processing Units (GPU) embedded processors. The DLAs have been introduced first in 2010s, in agriculture, plant disease diagnosis, and the 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). Forest modeling literature suggest that more studies are needed evaluating DLAs' performance to predict tree and stand attributes.

The comparison of artificial intelligence models with traditional regression models has been frequently carried out in previous modeling studies. The studies with the AI models have showed significant improvements in the prediction of forest and tree attributes compared to those by traditional regression models. However, the question of how these developed AI models can be used by other forest practitioner has always been ignored. The issue about these AI models will be put into practice appears will be unlikely. Thus, the most important problem in these AI models is that artificial intelligence models, consisting of a large number of parameters, hundreds or even thousands of parameters, cannot be given in page-bound journal publications. The best effective solutions to this problem can be achieved with open source R software. Especially, the R software platform, which becomes prominent with its applications and usage nowadays, will allow the forest planners and other forest practitioners to use the DLA and other AI models. The DLA models, which have trained by various researchers and forest practitioners, should be prepared in R software platform, which is free and open at free of charge. Subsequently, these AI models can be shared and presented to various forest practitioners and other users in forest management. By the way, the presentation of the DLA models, which were developed by R software, as additional files in journal publications will be great opportunity to predict tree and stand variables by the forest planners and other forest practitioners. This study aims providing a network model based on Deep learning Algorithms, which predicts relationship between ITH and DBH and other stand variables and this model can be used forest practitioners, as well.

2. Materials and Methods

2.2. Material

In this study, 124 temporary sample plots were used as the research material to predict the relationship between ITH and DBH and other stand variables. These sample plots were obtained from the forest management inventory of Turkish General Directorate of Forest in even-aged and pure Oriental beech (*Fagus orientalis* Lipsky) stands located in the Kestel forests, Northwestern Turkey. These inventory plots were sampled to represent various stand conditions such as site quality, age, and stand density of the even-aged Oriental beech stands. The size of circular sample plots ranged from 400 to 800 m^2 to include a minimum of 30–35 trees in sample plots. In each sample plot, DBH was measured to 0.1 cm precision using calipers for every living tree with a DBH > 8 cm. The individual tree total heights (ITH) were measured on a subset of trees created by selecting two-three trees for each 4 cm diameter class using Blume–Leiss Altimeter (0.1 m precision). In addition to the tree level measurements, the number of stems per hectare (N/ha), stand basal area (m^2/ha), quadratic mean diameter (cm), dominant diameter (cm), and dominant height (m) were calculated as part of the plot level information for each sample plot. Dominant height and diameter were calculated by averaging the height and diameter of the dominant and co-dominant trees.

In total, 1057 pairs of height–diameter measurements in 124 sample plots were used to analyze the relationships among ITH, DBH, and stand variables, and to predict the relationship between ITH and DBH and other stand variables. Approximately 85% (907 trees in 104 sample plots) were used to train the DLA models, and the remaining 150 trees in 20 sample plots were allocated for the evaluation of performance of these DLA models in different validation data, which were not used in training these DLA models.

2.2. Methods

2.2.1. Deep learning algorithms and training codes

In this study, the network models based on the Deep Learning Algorithms (DLAs), which is a type of Artificial Intelligence, were used to obtain the ITH predictions from the predictor variables including the DBH and the best predictive stand variables. 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 variables. The stand variables used in these alternatives as the input variables including the number of stems per hectare (N/ha), stand basal area (m^2/ha), dominant diameter (cm), and dominant height (m), quadratic mean diameter (cm) were evaluated to determine the best predictive input stand variables in the model structure of DLA models. These input variable alternatives are (A1) the DBH, dominant diameter (cm) and dominant height (m), (A2) the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare and basal area (m2/ha), (A3) the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare, basal area (m2/ha) and quadratic mean diameter (cm) (A4) the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare, basal area (m2/ha), quadratic mean diameter (cm) and stand density and (A5) the DBH, dominant height (m), number of trees in hectare, basal area (m2/ha) and stand density.

From various artificial intelligence software, the H_20 package (R Development Core Team, 2018), which have been coded in R software language, was used to train these DLA models and obtain the ITH predictions. H_2O package includes an h_2o.deeplearning function, which was coded on Java and is suitable to train multilayer feedforward deep learning neural networks. The h_2o.deeplearning function in H_2O package comprised by R was used to train these DLA models. This h_2o.deeplearning function provides multi-layer feedforward neural network models, which comprise the supervised training protocol to predict ITH from the DBH and the best predictive stand variables.

The DLA models, such as other AI models, require effective determination of network architecture including the number of hidden layers, 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 h_2o.deeplearning function utilizes adaptive learning rate algorithm ADADELTA (Zeiler, 2012), which includes the learning rate annealing and momentum training, enabling fast converge of the DLAs (H_2O.ai Team, 2018). The rho describes the degree of ADADELTA and epsilon outlines learning rate strengthening during preliminary training. The value of 0.999 for rho and 1x10-8 for epsilon were used to train DLAs. Also, the value of 1000 for the epochs, the number of iterations to be carried out in training networks, was applied in the training of these DLA models.

In addition to these network structure parameters, the number of hidden layers, type of activation function and the number of neurons in the hidden layers are significant attributes of network including. In some of our preliminary analyses, Tanh and Maxout activation functions resulted in extremely poor predictions of ITH. Therefore, rectifier function was selected as an activation function to train the DLA models. As other network parameters, the number of 9 hidden layers and 100 neurons gave the best predictive results for ITH in in training DLA models. Thus, the DLA model structure with 9 layers and 100 neurons were chosen to train these DLA models and obtain the ITH predictions.

To overcome the overfitting problem, which may be noticeable in that the training and validation datasets differ substantially in their fitting statistics, the cross-validation method was used by regulating some parameters in h_2o.deeplearning function in $H_2$0 package. When this cross-validation method was applied to the DLA models, the K=10-fold cross-validation resampling techniques with nfolds=10 in h_2o.deeplearning function were used in this study.

In training this DLA model with 9 layers and 100 neurons to predict the ITH, the $H_2$0 package coded on the R software were implemented as shown below.

To train DLA models, install "H20" package from "The Comprehensive R Archive Network" site by using following R code:

> *install.package("h2o")*

After installing this package, this package must be activated:

> *library(h2o)*

Establish the IP and PORT parameters in H2O package on your local machine:

> h2o.init()

Load your data to H2o from your local machine:

> yourmodelingdata <- as.h2o(dataonyourlocalmachine)

To train DLA model with 9 layers and 100 neurons, these h2o.deeplearning functions with "rectifier" activation function, 100 number of neurons, 9 number of layers and "Gaussian of distribution function was used:

> DLA9L100N <- h2o.deeplearning(x = 2:4, y = 5, training_frame = yourmodelingdata, epoch=1000, nfolds=10, distribution="gaussian", variable_importance=T, hidden=c(100, 100, 100, 100, 100, 100, 100, 100, 100), activation = "Rectifier")

To obtain the ITH predictions from this trained DLA model, following R code should be used:

>predictDLA9L100N<- as.data.frame(h2o.predict(DLA9L100N, yourmodelingdata))

To use this DLA models and then obtain the ITH prediction for other data sets from this DLA models, this trained DLA models should be saved in a local machine:

> model_path <- h2o.saveModel(object = DLA9L100N, path=getwd(), force = TRUE)

2.2.2. R codes of network model for future use of other forest practitioner

In this study, besides training of DLA models, it is aimed to present these trained DLA models to other forest practitioners. For this purpose, the best predictive DLA model, which was trained by using "H2o" package of R software in the first stage of this study, was saved to present other forest practitioners. This network model as R syntax file can be downloaded from Google Drive Link, which was presented in the discussion section of this study. This DLA model can be downloaded and used to obtain the ITH predictions of other studied forest areas or other validation data set by other forest practitioner. In this study, this saved network model was used to predict the ITH of 150 trees in 20 sample plots, which were reserved as the validation data set for the evaluation of the best predictive DLA model. In later usage of this DLA model to predict the ITH, this saved model can be loaded with the R codes shown in the steps below:

1.) activating H2O package on a local machine using all available cores for deep learning Algorithms:

>libray(h2o)

2.) establishing the IP and PORT parameters in H2O package on a local machine*:*

>h2o.init()

3.) loading data to H2o from a local machine

>yourdata <- as.h2o(dataonyourlocalmachine)

4.) loading the best predictive DLA model with 9 layers and 100 neurons trained in this study from a local machine, which this DLA model may be downloaded from the Google Drive Link of this study:

>DLAmodel<-h2o.loadModel("C:/Users/GROWTH/Desktop/the best predictive DLA model with 9 layers and 100 neurons")

5.) predicting a individual tree height by using input variables including diameter at breast height (cm), the dominant height (m) (h_0) and dominant diameter (cm) (d_0) from the best predictive DLA model in this study.

>predicts<-as.data.frame(h2o.predict(DLAmodel, yourdata))

6.) loading xlsx package to obtain ITH predictions

>library(xlsx)

7.) saving ITH predictions to *a local machine as Excel worksheet*

>write.xlsx(predicts, file = "predictions.xlsx")

2.2.3. Nonlinear regression model

In this study, the Schnute (1981)'s model, the nonlinear regression model as the traditional prediction method, which presented successful results for predicting ITH, was compared with the DLA models. This Schnute (1981)'s model, which is the most flexible and adaptable function for modeling this relation (Bredenkamp and Gregoire, 1988; Lei, 1998), was used to model the relationships between the ITH and DBH with the dominant diameter (cm) and dominant height (m). Ercanlı (2015) predicted the parameters of Schnute (1981)'s model by using NLIN procedure available in SAS/STAT® 9 software (SAS Institute Inc., 2004). The Schnute (1981)'s 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, h_i is the observed height of the *i*th tree in the sample plots, d_i is DBH of the *i*th tree the sample plots, H_0 is the dominant height of the sample plot, D_0 is the dominant diameter of the sample plot.

2.2.4. Criteria of comparisons for different DLA models

To compare different DLA models including input variable alternatives including various stand attributes and the Schnute (1981)'s nonlinear regression model, eight fitting criteria were used to choose the best predictive model. The following criteria were considered: (1) average absolute error (AAE), (2) the root mean squared error (RMSE), (3) percent root mean squared error (percent RMSE), (4) the average Bias (Bias), (5) percent average Bias (shortly, percent Bias), (6) the fit index (FI), (7) Akaike's information criterion (AIC), and (8) Bayesian information criterion (BIC) to compare prediction performance of DLAs. These criteria are calculated as follows:

$$AAE = \sum_{i=1}^{n}\left|h_i - \hat{h}_i\right|/n \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(h_i - \hat{h}_i)^2/(n-k)} \quad (3)$$

$$RMSE\% = \left(\left[\sqrt{\sum_{i=1}^{n}(h_i - \hat{h}_i)^2/(n-k)}\right]/\bar{h}_i\right)\cdot 100 \quad (4)$$

$$Bias = \sum_{i=1}^{n}\left(h_i - \hat{h}_i\right)/n \quad (5)$$

$$Bias\% = \left(\left[\sum_{i=1}^{n}\left(h_i - \hat{h}_i\right)/n\right]/\bar{h}_i\right)100 \quad (6)$$

$$FI = 1 - \frac{\sum_{i=1}^{n}\left(h_i - \hat{h}_i\right)^2}{\sum_{i=1}^{n}(h_i - \bar{h}_i)^2} \quad (7)$$

$$AIC = n ln(RMSE) + 2k \quad (8)$$

$$BIC = n ln(RMSE) + n ln(k) \quad (9)$$

Where, $\hat{h}_i$ is the predicted height of the *i*th tree in the sample plots, h_i is the observed height of the *i*th tree in the sample plots, $\hat{h}_i$ is the average of observed height values, k is the number of independent variables in the models. Smaller values of AAE, RMSE, RMSE%, Bias, Bias%, AIC, BIC and higher values of FI indicate better prediction performance of the models. To evaluate these eight criteria together and their general predictive ability for the ITH, the relative rank values, which were proposed by Poudel and Cao (2013) for DLAs model were calculated. The model with lowest sum of relative rank was recognized as the best predictive ITH model.

3. Results and Discussion

For training data set, Table 1 presents the fit statistics of AAE RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for the DLA models including different input variable alternatives and the nonlinear regression model of Schnutte (1981). The relative rank values (Poudel and Cao, 2013) for this DLA models and Schnutte (1981)'s nonlinear model is given in Table 2. From these fitting statistics for training data set, the DLA model which includes the input variables with the DBH, dominant diameter (cm), dominant height (m), number of trees per hectare and basal area (m^2/ha) in A2 input alternative resulted in the best predictive statistics with a RMSE value of 0.7173, RMSE% value of 4.5986, the AIC value of -291.3037, BIC value of 1158.4564, FI of 0.9785 values, AAE value of 0.4311, Bias value of 0.0438 and Bias% value of

0.2805. When compared the fitting statistics of the DLA models with those by nonlinear regression model of Schnutte (1981), the DLA models provided important improvements in fit statistics for the prediction of the ITH.

Table 1. The goodness-of-fit statistics of AAE RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for the DLA models including different input variable alternatives and nonlinear regression model of Schnutte (1981) for training data set.

The DLA models and regression model	RMSE	Percent RMSE	AIC	BIC	FI	AAE	Bias	Percent Bias
Nonlinear Regression	1.8009	11.5445	539.5550	1529.9963	0.8639	1.0517	0.1767	1.1324
A1	1.4334	9.1887	332.5461	1322.9874	0.9138	0.8607	0.1825	1.1697
A2	**0.7173**	**4.5986**	**-291.3037**	**1158.4564**	**0.9785**	**0.4311**	**0.0438**	**0.2805**
A3	0.6383	4.0919	-395.1821	1217.9437	0.9830	0.3724	0.0349	0.2239
A4	0.7879	5.0508	-202.2373	1548.7032	0.9741	0.4415	0.0751	0.4813
A5	0.8562	5.4889	-130.7782	1318.9820	0.9693	0.5028	0.1673	1.0727

AAE: average absolute error, RMSE: the root mean squared error, percent RMSE: percent of root mean squared error, Bias: the average Bias, percent Bias: percent average Bias, FI: the fit index, AIC: Akaike's information criterion and BIC: Bayesian information criterion.

Table 2. The rank values for the DLA models including different input variable alternatives and nonlinear regression model of Schnutte (1981) for training data set.

The DLA models and regression model	RMSE	Percent RMSE	AIC	BIC	FI	AAE	Bias	Percent Bias	∑
Nonlinear Regression	6.000	6.000	6.000	5.760	6.000	6.000	5.803	5.803	47.36
A1	4.419	4.419	3.468	3.108	3.905	4.594	6.000	6.000	35.914
A2	**1.340**	**1.340**	**2.963**	**1.000**	**1.189**	**1.432**	**1.300**	**1.300**	**11.864**
A3	1.000	1.000	4.234	1.762	1.000	1.000	1.000	1.000	11.996
A4	1.643	1.643	1.874	6.000	1.373	1.509	2.361	2.361	18.765
A5	1.937	1.937	1.000	3.057	1.574	1.960	5.487	5.487	22.439

AAE: average absolute error, RMSE: the root mean squared error, percent RMSE: percent of root mean squared error, Bias: the average Bias, percent Bias: percent average Bias, FI: the fit index, AIC: Akaike's information criterion and BIC: Bayesian information criterion.

Besides the evaluation of prediction performance of DLA models for training data set, the prediction ability of these DLA models in different validation data sets, which are not used in training stage is an important evaluation process in modeling studies. In Table 3 and 4, the goodness-of-fit statistics for the DLA models and the nonlinear regression model of Schnutte (1981) were presented for the validation data set. For these validation data set, DLA models gave better predictive results likewise in the training data than those by the nonlinear regression model of Schnutte (1981). Similar to the fitting statistics in training data, the DLA model which includes the input variables with the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare and basal area (m^2/ha) for the validation data set gave the best predictive statistics with a RMSE value of 1.8217, RMSE% value of 10.2151, the AIC value of 99.9615, BIC value of 331.3772, FI of 0.8334 values, AAE value of 1.2051, Bias value of -0.0985 and Bias% value of -0.5521.

Table 3. The goodness-of-fit statistics of AAE RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC for the DLA models including different input variable alternatives and nonlinear regression model of Schnutte (1981) for validation data set.

The DLA models and regression model	RMSE	Percent RMSE	AIC	BIC	FI	AAE	Bias	Percent Bias
Nonlinear Regression	2.0838	11.6853	116.1324	274.9242	0.7790	1.3685	0.1544	0.8656
A1	2.2050	12.3645	124.6062	283.3980	0.7526	1.3565	0.0562	0.3150
A2	**1.8217**	**10.2151**	**99.9615**	**331.3772**	**0.8334**	**1.2051**	**-0.0985**	**-0.5521**
A3	2.0416	11.4482	119.0568	375.8208	0.7922	1.2391	-0.1357	-0.7607
A4	2.1077	11.8192	125.8403	403.7268	0.7801	1.2857	-0.0931	-0.5221
A5	2.1354	11.9743	123.7959	355.2115	0.7711	1.1778	0.4651	2.6078

AAE: average absolute error, RMSE: the root mean squared error, percent RMSE: percent of root mean squared error, Bias: the average Bias, percent Bias: percent average Bias, FI: the fit index, AIC: Akaike's information criterion and BIC: Bayesian information criterion.

Table 4. The rank values for the DLA models including different input variable alternatives and nonlinear regression model of Schnutte (1981) for validation data set

The DLA models and regression model	RMSE	Percent RMSE	AIC	BIC	FI	AAE	Bias	Percent Bias	∑
Nonlinear Regression	4.420	4.420	4.124	1.000	4.365	6.000	2.201	2.201	28.73
A1	6.000	6.000	5.762	1.329	6.000	5.684	1.000	1.000	32.775
A2	**1.000**	**1.000**	**1.000**	**3.191**	**1.000**	**1.715**	**1.517**	**1.517**	**11.940**
A3	3.869	3.869	4.689	4.917	3.548	2.608	1.972	1.972	27.443
A4	4.731	4.731	6.000	6.000	4.299	3.828	1.452	1.452	32.493
A5	5.092	5.092	5.605	4.117	4.854	1.000	6.000	6.000	37.760

AAE: average absolute error, RMSE: the root mean squared error, percent RMSE: percent of root mean squared error, Bias: the average Bias, percent Bias: percent average Bias, FI: the fit index, AIC: Akaike's information criterion and BIC: Bayesian information criterion.

The Fig.1 shows the relationships obtained between residual values and predicted ITH values by the nonlinear regression model of Schnutte (1981) and the best predictive DLA model including the input variables with the DBH, dominant diameter (cm), dominant height (m), number of trees per hectare and basal area (m^2/ha) and the nonlinear regression model of Schnutte (1981) for training data and validation data. When these residuals from this DLA model is compared with those from nonlinear model, the residuals associated with the DLA model are closer to the X axis and have a more collective trend. Thus, it can suggest that the DLA model provides more precise and effective ITH predictions than those by the nonlinear regression model of Schnutte (1981).

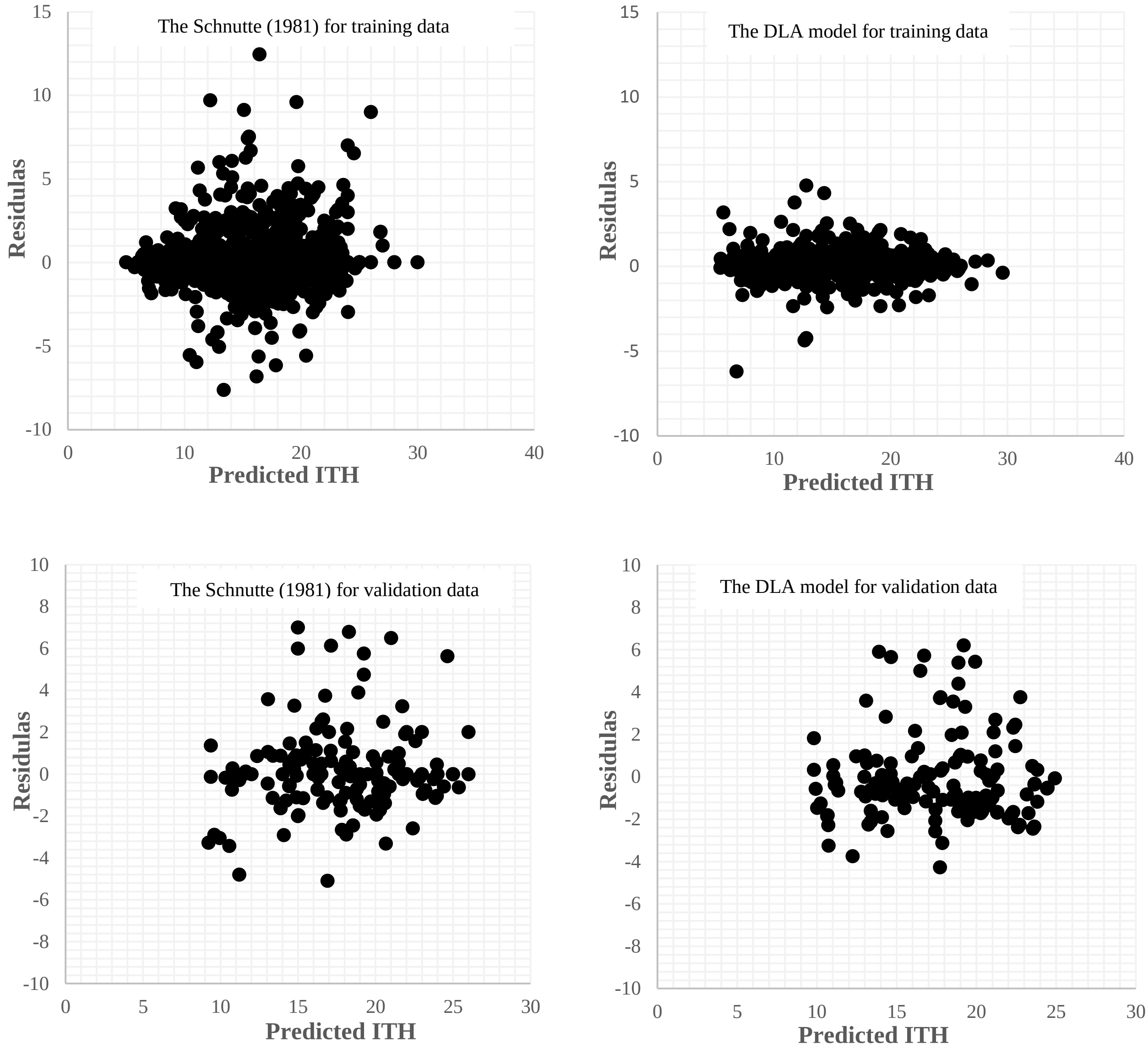

Figure 1. Relationships between predicted (x-axis) and residuals ITH (y-axis) obtained by the best predictive deep learning network model including the input variables with the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare and basal area (m2/ha) and the nonlinear regression model of Schnutte (1981) for training data and validation data.

In this study, besides the use of deep learning algorithms, which is an innovative technique in forest and tree growth model studies, the best predictive network model based on the DLA is presented in the Google Drive Link, which other forest practitioner can download and use this network model to predict the ITH sampled from their forest areas. In many artificial intelligence modeling studies, it is limited to the comparison of various network models with classical regression models, however the presentation of these artificial intelligence models has been generally ignored for future usage of other forest practitioner for predicting individual tree height. In this respect, this study stands out from other artificial intelligence studies by providing a tool that can be used by forest practitioner to obtain ITH predictions for their forest areas.

Before presenting an artificial intelligence module, it is necessary to train this network model. In the first stage of this study, various artificial intelligence models based on deep learning algorithms were trained by using by using "H_20" package which was codded in R software. The five input variable alternatives including various stand attributes were evaluated based on the statistical criteria with AAE, RMSE, RMSE%, Bias, Bias%, FI, AIC and BIC and their integrated assessment calculated according to Poudel and Cao (2013)'s relative rank values. The input variable alternatives considered are: (A1) the DBH, dominant diameter (cm) and dominant height (m), (A2) the DBH, dominant diameter (cm), dominant height (m), number of trees per hectare and basal area (m2/ha), (A3) the DBH, dominant diameter (cm), dominant height (m), number of trees in hectare, basal area (m^2/ha) and quadratic mean diameter (cm), (A4) the DBH, dominant diameter (cm), dominant height (m), number of trees per hectare, basal area (m2/ha), quadratic mean diameter (cm) and stand density and (A5) the DBH, dominant height (m), number of trees in hectare, basal area (m2/ha) and stand density. From these alternatives, the DLA model with A2 input alternative presented the best predictive statistics with a RMSE value of 0.7173, RMSE% value of 4.5986, the AIC value of -291.3037, BIC value of 1158.4564, FI of 0.9785 values, AAE value of 0.4311, Bias value of 0.0438 and Bias% value of 0.2805. Thus, this best predictive DLA model accounted for over 97 % of the total variance in ITH predictions.

In addition to the success of artificial intelligence models in training data, the evaluation process needs special attention in comparisons of predictive ability of these network models for the validation data. For this purpose, the predictive DLA model with A2 input variable alternative was used to obtain the ITH predictions of 150 trees that were not used in the training of network models. The most important problem encountered in artificial intelligence models is that the network model, which is quite predictive for training data, provides highly unsuccessful and irrelevant predictions for the validation data or independent data. This problem is called the “overfitting” problem in artificial modeling studies. However, the best predictive DLA model which was trained in this study gave significant account with over 83% in the total variance of ITH predictions. Therefore, we can say that this “overfitting” problem remains quite limited in the ITH predictions by the network model, because if this problem was excessive, this DLA network model would have been quite unsuccessful for the ITH predictions for validation data.

In this study, the R syntax file of the best predictive DLA model with A2 input variable alternative can be downloaded so that other forest practitioner can use this best predictive DLA model, which a similar application was applied for validation data of 150 trees in this study. The best predictive DLA network model with 9 layers and 100 neurons as the Google Drive Link were presented in this study. This best predictive DLA model from Google Drive (https://drive.google.com/open?id=1OD9HVFqrur8bXQOgt2Rprn5vfhZEHq1f) can be downloaded, and so other forest practitioner can use this best predictive DLA model, which some similar applications was applied for validation data of 304 trees in this study. This DLA model can be downloaded and used by others to obtain the ITH predictions. As this present study have shown by training the DLA models and providing R syntax codes of the best predictive DLA models, artificial intelligence studies should provide more proper 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.

4. Conclusions

This study presented the artificial intelligence models based on the deep learning algorithms predicting the individual tree heights (ITH) from the diameter at breast heights and some stand variable for Oriental beech (*Fagus orientalis*) in Kestel, Northwestern Turkey.In future studies, it is necessary to investigate the success of the deep learning algorithms to test their performance in predicting of forest stand characteristics such as stand volume, basal area, biomass and carbon, and individual tree variables such as diameter, height, volume and growth. As an innovative prediction technique, previous studies about artificial intelligence models offering successful predictive results have underlined to be an important requirement for further studies on the evaluation of the artificial intelligence models based on the deep learning algorithms as an alternative to classical statistical methods in the predictions of stand and single tree characteristics. In addition, it will be important to evaluate whether the artificial intelligence models provide estimates of growth laws for single tree and stand which are important in forestry practices, as well as in terms of success in accordance with statistical model success criteria.

Acknowledgements

The authors thank Head of Forest Management and Planning Department, General Directorate of Forestry, Republic of Turkey, for providing data used in this study.

References

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A., 2017. Going deeper in the automated identification of Herbarium specimens. BMC Evolutionary Biology 17, 181.

Clutter, J. L., Fortson, J. C., Pienaar, L. V., Brister, G. H., Bailey, R. L., 1983. Timber management: a quantitative approach. John Wiley & Sons, Inc.

Ercanlı, İ., 2015. Nonlinear mixed effect models for predicting relationships between total height and diameter of oriental beech trees in Kestel, Turkey. – Revista Chapingo Serie Ciencias Forestales y del Ambiente 21(1): 185-202.

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture 145, 311-318.

Gadow, K. V. and Hui, G. Y., 1999. Modelling Forest Development, Kluwer Academic Publishers, Dordrect, 213 s.

Huang, S., Titus, S. J., Wiens, D. P., 1992. Comparison of nonlinear heightdiameter functions for major Alberta tree species. Canadian Journal of Forest Research, 22, 1297–1304.

H2O.ai Team., 2018. H2O GitHub Repository. URL: https://github.com/h2oai

Lee, S. H., Chan, C. S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks. In, 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 452-456.

Martin, F. C., Flewelling, J. W., 1998. Evaluation of tree height prediction models for stand inventory. Western Journal of Applied Forestry, 13, 109–119.

Mohanty, S. P., Hughes, D. P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science 7, 1419.

Nanos, N., Calama, R., Montero, G., Gil, L., 2004. Geostatistical prediction of height-diameter models. Forest Ecology and Management, 195, 221-235.

Poudel, K. P., Cao, Q. V., 2013. Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. Forest Science 59: 243-252.

R Core Team, (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna., Austria.: https://www.R-project.org

SAS Institute Inc., 2004. SAS/IML 9.3 User's Guide. Cary, USA. – SAS Institute Inc., Cary, NC, USA

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence Neuroscience 2016.

Schnute, J., 1981. A versatile growth model with statistically stable parameters. Canadian Journal of Fisheries Aquatic Sciences 38, 1128-1140.

Sun, Y., Liu, Y., Wang, G., Zhang, H., 2017. Deep learning for plant identification in natural environment. Computational Intelligence Neuroscience 2017.

van Laar, A., Akça, A., 2007. Forest mensuration: in Managing Forest Ecosystems, Dordrecht, The Netherlands: Springer. 383 p.

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.

Zeiler M.D., 2012. ADADELTA: an adaptive learning rate method. ArXiv-Machine Learning.