



*Research Article*

## Evaluation of tree basal area increment models using machine learning algorithms

Seyedeh Kosar Hamidi<sup>1\*</sup> and Asghar Fallah<sup>2</sup>

<sup>1</sup>Assistant Prof., Dept. of Forestry, Faculty of Natural Resources, Sari Agricultural Sciences and Natural Resource University, Sari, I. R. Iran

<sup>2</sup>Prof., Dept. of Forestry, Faculty of Natural Resources, Sari Agricultural Sciences and Natural Resource University, Sari, I. R. Iran

(Received: 4 November 2024; Accepted: 4 January 2025)

### Abstract

Forest management planning is a critical decision-making tool in forestry, resulting in a plan that outlines anticipated activities, their timing, and control measures to achieve forest management goals. Investigating growth and product models is one of the most important methods for obtaining information about the future state of a forest. In other words, assessing stand growth and yield is a basic prerequisite for forest management planning. Therefore, determining and estimating the basal area increment of trees is crucial for understanding forest dynamics and informing planning and management efforts. Since Hyrcanian forest species are considered among the most valuable, this study aims to investigate basal area increment using machine learning algorithms and model it in the uneven-aged forest of Farim in Mazandaran province. In this study, the basal area increment (BAI) of trees was modeled using Machine Learning (ML) algorithms (Artificial Neural Networks, Support Vector Machine, Random Forest, and Generalized Additive Model) over 10 years. Biometric indices (e.g., diameter, height, basal area, basal area of the largest trees), physiographic factors (aspect, slope, altitude), and climatic variables (temperature, precipitation, evaporation and transpiration) were used as input for model development. The performance of the machine learning algorithms were compared using bias, RMSE, and  $R^2$ . The ANN model, specifically an MLP network with seven hidden layer neurons, achieved the highest accuracy (88%) in predicting basal area increment compared to other models. These results demonstrate the effectiveness of ANN models for accurately modeling basal area increment, making them valuable tools in forest management. The strong performance of the generated models, attributed to their optimal structure (e.g., number of neurons, activation function, and input variables), highlights their stability and generalization capacity across diverse datasets. The potential to improve forest parameter modeling using machine learning techniques, specifically ANN, is crucial for sustainable forest management. Such improvements can enhance the conservation of species composition and the structural characteristics of the forest.

**Keywords:** Artificial Neural Networks, Basal area, Forest planning, Hyrcanian forest, Machine learning.

### 1. Introduction

Hyrcanian forests have an area of about 5,500,000 ha. Of this amount, about 20,000 hectares are located in Azerbaijan and 2 million hectares in Iran. The length of this strip is approximately 800 kilometers, and its width fluctuates between 20 and 70 kilometers. The forests of the north of Iran begin from Astara and continue to Gilidaghi Bojnoord Valley. Its vertical expansion starts at sea level and rises to a maximum height of 2,800 meters. The area of these forests was estimated to be 3.6 million hectares at the beginning of the 21st century.

Hyrcanian forests are mixed and uneven-age forests that are rich in terms of tree and shrub species diversity. Approximately 80 tree species (mainly broad-leaved) and 50 shrub species are naturally found in these forests (Hamidi et al., 2022). Today, forests are considered not only economic drivers but also essential for the survival of other sectors (Vanclay, 1994). Based on this, decision-making, as the essence of management, is an integral part of the duties of managers and planners in natural resource units.

Forest management planning is an important decision-making tool in forestry, and its result is a management plan outlining the anticipated activities, their timing, and control measures to achieve forest management goals within a forest area (Hamidi et al., 2016; Baskent & Keles, 2005). Forest management and planning require information that describes both future and alternative situations in addition to the current state of the forest. Investigating growth and yield models is one of the most important methods for obtaining information about the future state of a forest (Bettinger et al., 2005; Bayat & Hamidi, 2019). In other words, assessing stand growth and yield is a fundamental prerequisite for forest management planning (Hamidi et al., 2023; Bayat et al., 2023; Burkhart, 1990; Vanclay, 1994). Assessing the structure of forest stands by quantifying their different components helps improve the management of various forest ecosystem elements. One of the main reasons for quantifying forest structure is its relationship to the habitat of many plant species, the modeling and reconstruction of the forest based on its structural characteristics, and the fact that forest structure is an important source of information for forest dynamics, aesthetics, and wood production. One of the most important applications of growth models is the study of the dynamics of forest stands (Ghaderi et al., 2013).

In addition to investigating changes in the forest, these studies will make future decision-making easier (Vahedi et al., 2024; Bayat et al., 2022). An example of an individual tree model is the basal area increment (BAI) system proposed by Wykoff in 1986 and 1990, later modified by Pokharel and Froese in 2009, and Hamidi et al., in 2019. Tree diameter is one of the most common features used in individual tree models. The measurement is accurate with little or no error and can be used with other tree traits, such as yield, height (Sharma & Parton, 2007; Vargas-Larreta et al., 2009; Wykoff, 1990), volume (Hai et al., 2011), biomass (IPCC, 2019), and carbon storage (Stavins & Richards, 2005). In addition, tree diameter has been used as a variable for tree age (Pokharel & Froese, 2009). Tree diameter increments or BAI models rely on a relatively fixed set of variables such as tree size and competition (Wykoff, 1986, Hamidi & Fallah., 2024).

Recently, various predictive models have

been used to achieve accurate basal area estimates, including ordinary parametric regression models (Mauya et al., 2015) and nonlinear machine learning models. Several studies have been conducted to evaluate the efficiency of different models in predicting the characteristics of the forest. Some reported that random forest (RF) and regression model (RM) have comparable capabilities in predicting the performance of forest attributes (Görgens et al., 2015), while some researchers concluded that machine learning (ML) outperforms linear regression models in forest attribute prediction (Zhao et al., 2011). Describing non-linear relationships between population characteristics and different biological variables is a challenge for traditional parametric models. In contrast, nonparametric models have the advantage of being more flexible and adept at fitting many functional forms without the need for prior information (Mullainathan & Spiess, 2017). In addition, nonparametric nonlinear models provide a more suitable solution for nonlinear system applications, such as forest systems, and are useful for dealing with nonlinear fitting. However, few studies have considered using machine learning to develop basal area prediction models with basic structural variables, as ML algorithms do not require assumptions about distributions, data transformations, and are robust to outliers. The application of mathematical features and modeling in the ML technique to identify hidden patterns and relationships in the database has been proven (Hamidi et al., 2021). Additionally, with big data technology and high-performance computing, new methods for the analysis of natural resources have been developed (Liakos et al., 2018).

The most widely used machine-learning methods include ANN, SVM, RF, and GAM (Strobl & Forte, 2007). The potential of using these methods for forecasting and multifactorial processes makes these computer-based tools a powerful approach for forest modeling and management (Lohmander, 2017). A study was conducted in the northern forests of eastern Canada using BAI models. The purpose of this study was to create a single BAI model for the main tree species in the aforementioned area using the ML method. One of the most important applications of modeling is to improve forest management. These models examine the rate of growth and its main factors,

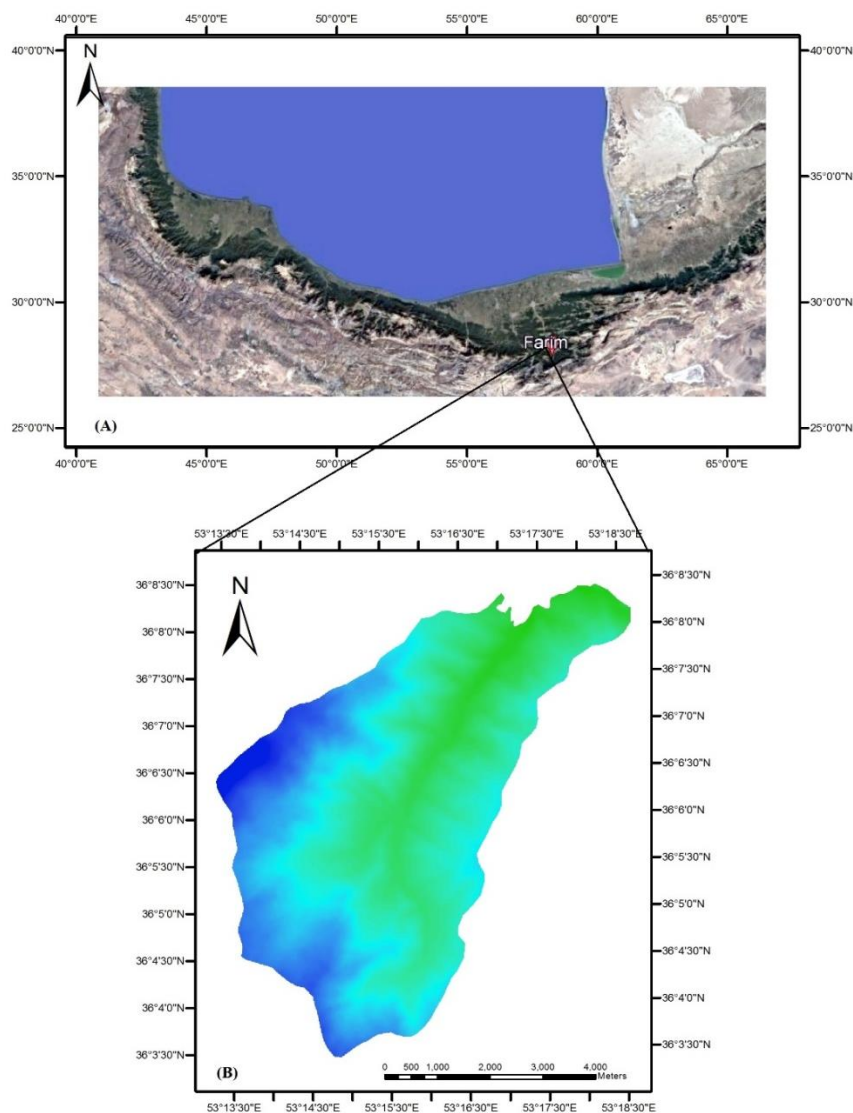
providing the best management option; in some cases, they can also predict the future of the forest. Currently, the effects of this issue on plant species in our country remain unknown, and there is limited knowledge in this field. The purpose of this study was to investigate the factors affecting the growth model of tree basal area. Accordingly, the research questions are: Which of the machine learning methods has better accuracy than the others? and which factor has a significant impact on basal area increment models? The research hypotheses are: 1) The neural network method performs better than other methods. 2) Diameter and height are influential factors in all methods. The results of this research can serve as a valuable tool to assist management decisions, as well as protection and restoration plans for Hyrcanian

forests, ensuring that these decisions and plans are aligned with the conditions of the forests.

## 2. Materials and methods

### 2.1. Case Study

This study was conducted in the Farim forest, located in Mazandaran Province in the north of Iran. (Figure 1). The area covers 2,803 hectares, with altitudes between 782 and 1750 meters above sea level (Hamidi et al., 2019). The annual precipitation is 833 mm and the average annual temperature is 11 °C. The mother rock that forms the soil of the region consists of sandstone and siltstone, accompanied by limestone veins. The soil type is washed brown soil, characterized by an argillic horizon. (Hamidi et al., 2017).



**Figure 1.** Location of the studied area Iran, Mazandaran Province (A), Farim (B)

## 2.2. Research Method

To collect data in this research, an inventory grid (200 m × 150 m) with a random starting point and 313 circular plots of 0.1 ha was regularly implemented in the area. Within each sample plot, the diameter at breast height (DBH) of all living trees with a DBH ≥ 7.5 cm was measured using calipers. The average plot slope, aspect, and terrain altitude were measured using a clinometer and a global positioning system (GPS) unit. The trees measured at the beginning of the period were measured again at the end of the period (10 years later), and the amount of forest growth and the growth model were calculated and used for analysis and modeling (Hamidi et al., 2021). The basal area (BA) of trees and the basal area of larger trees (BAL) are key competitive factors used as independent variables in this study. In addition, biometric factors (diameter, height), physiographic (slope, aspect, and altitude) and climatic factors (temperature, precipitation, evaporation

and transpiration) were used in the model to investigate the effects of these factors on the basal area model (Hamidi et al., 2022). The evaporation and transpiration rates were calculated using Thornthwaite's method (Equations 1 to 3) (Hamidi & Fallah, 2024).

$$ET=16 N_m \left(\frac{10T_m}{I}\right)^a \quad (1)$$

$$I=\left(\frac{T_m}{5}\right)^{1.514} \quad (2)$$

$$a= (675 \times 10^{-9}) I^3 - (771 \times 10^{-7}) I^2 + (179 \times 10^{-4}) I + 0.492 \quad (3)$$

where ET: monthly evaporation and transpiration (mm),

I: annual heat index,

$T_m$ : average monthly temperature in centimeters and

a: a coefficient that is different from one point to another and depends on the heat index.

Table 1 lists the variables used in this study.

**Table 1.** Symbols and units of variables used in this research

| Variables                       | Symbology | Unit                            |
|---------------------------------|-----------|---------------------------------|
| Tree density                    | N         | Tr ha <sup>-1</sup>             |
| Diameter at breast height (DBH) | DBH       | cm                              |
| Total height                    | H         | m                               |
| Basal area                      | BA        | m <sup>2</sup> ha <sup>-1</sup> |
| Basal area of large trees       | BAL       | m <sup>2</sup> ha <sup>-1</sup> |
| Temperature                     | TEMP      | °c                              |
| Precipitation                   | PCP       | mm                              |
| Evapotranspiration              | ET        | mm                              |
| Slope                           | SLP       | %                               |
| Aspect                          | ASP       | -                               |
| Altitude                        | ALT       | m                               |

## 2.3. Machine Learning Models

### 2.3.1. Artificial Neural Networks (ANNs)

ANN includes two common algorithms: Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks. Neural networks based on these two architectures are commonly used in engineering applications (Hamidi et al., 2021). Artificial Neural Networks have great potential because they can recognize complex patterns. In this method, a strong relationship exists between the input or predictor variables and the desired output. Additionally, due to their data processing capabilities, ANNs are less sensitive to errors in input data compared to other models (Rees et al., 2017). In this structure, the neurons in one layer are connected to the neurons in the subsequent layer. This structure forms a

connected network. The number of hidden layers and neurons in each layer was determined through trial and error, and modeling was subsequently performed.

### 2.3.2. Support Vector Machine Algorithm (SVM)

In addition to its solid mathematical foundation in statistical learning theory, the SVM algorithm has demonstrated excellent performance in practical applications (Hamidi et al., 2022). When creating this model, attention should be given to selecting the kernel type and determining the kernel parameters (Shataee et al., 2012). The four kernels used to model this algorithm were linear, polynomial, radial basis, and sigmoid. The kernel parameters include capacity,

gamma, and epsilon (Hamidi et al., 2021). To determine the optimal epsilon and capacity, cross-validation and automatic network search with 10-fold validations and 5000 repetitions were used (Townsend, 2002). To validate the models, mean square error and bias were used. To calculate these two metrics, data from 30% of the sample (test data) were used.

### 2.3.3. Random Forest Algorithm

The random forest algorithm is a modern version of the base tree method (Dietterich, 2000; Breiman, 2001). This algorithm is an extension of the regression and classification trees, first introduced and developed by Breiman in 2001. A decision tree is an ML algorithm that uses a flowchart-like structure for decision-making. Each node represents an attribute or feature, and each branch represents a decision or rule. The algorithm divides the data based on the values of different attributes until it reaches the leaf nodes, which represent the results or predicted values. Decision trees are easy to understand and interpret, suitable for both classification and regression tasks, and can handle various types of data. The first parameter is the number of trees to be predicted. To determine the optimal number of trees, several primary trees were used to generate a graph showing the changes in the mean squared error against the number of trees in the training and evaluation samples. An important parameter in random forests is the number of predictors, K (independent variables), in each node to predict the response (dependent) variable. The square root of the number of independent variables is the simplest way to calculate the value of K (Shataee et al., 2012; Nazariani et al., 2022). By observing the graph of changes in squared errors for both test and training data, the point at which the error rate stabilizes is selected as the optimal number of trees. To estimate the optimal number of predictors, the total number of square roots of the variables, plus and minus ( $\pm 2$ ), was used and the results were evaluated (Hamidi et al., 2021).

### 2.4. Generalized Additive Model (GAM)

One of the important statistical methods developed in the last 40 years is the Generalized Additive Model. This model is a non-parametric method and an extension of generalized linear models, which themselves are extensions of linear models. The Generalized Additive Model

provides an effective method for analyzing data and investigating the relationship between independent variables and responses. Unlike the linear regression model, the Generalized Additive Model allows the data to determine the shape of the response function (Jafarian & Karger, 2016).

All steps were performed using R and STATISTICA software packages. The data collected from the field are randomly divided into training and testing sets, with 70% used for training and the remaining 30% for evaluating the model. Equations 4 to 7 present the evaluation methods of the models.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{est}_i - \text{obs}_i)^2}{n}} \quad (4)$$

$$\text{BIAS} = \frac{\sum_{i=1}^n (\text{est}_i - \text{obs}_i)}{n} \quad (5)$$

$$\text{Relative RMSE} = 100 \times \left( \frac{\text{RMSE}}{\text{mean observation value}} \right) \quad (6)$$

$$\text{Relative BIAS} = 100 \times \left( \frac{\text{BIAS}}{\text{mean observation value}} \right) \quad (7)$$

where  $\text{est}_i$  and  $\text{obs}_i$  are the  $i^{\text{th}}$  estimate and observations, respectively, and  $n$  is the number of observations.

## 3. Results

### 3.1. Growth Modeling Data

Tree diameters ranged from 12.5 to 157 cm. Descriptive statistics of the data used in the growth model are shown in Table 2.

### 3.2. Basal area Growth Model by Machine Learning Models

#### 3.2.1. Artificial Neural Network

In this study, the learning rule type, transfer function, and number of hidden layer neurons were tested for each network. Among the five models obtained from the artificial neural network, Model 2, with an MLP network architecture of 10-7-1, was identified as the best-performing model due to its high  $R^2$  and the lowest RMSE and Bias (Table 3).

The most important predictive values for basal area growth are shown in Figure 3.

### 3.3. Support Vector Machine Model

The choice of kernel affects the efficiency of the SVM algorithm. The results of different kernels in the SVM method showed that the linear kernel produced the best

results (the lowest squared error) among the different kernels (Table 4). The residuals showed a slight positive bias in plots where the basal area increment exceeded 20  $\text{cm}^2\text{ha}^{-1}$  (Figure 4).

**Table 2.** Characteristics of all variables in studied forest

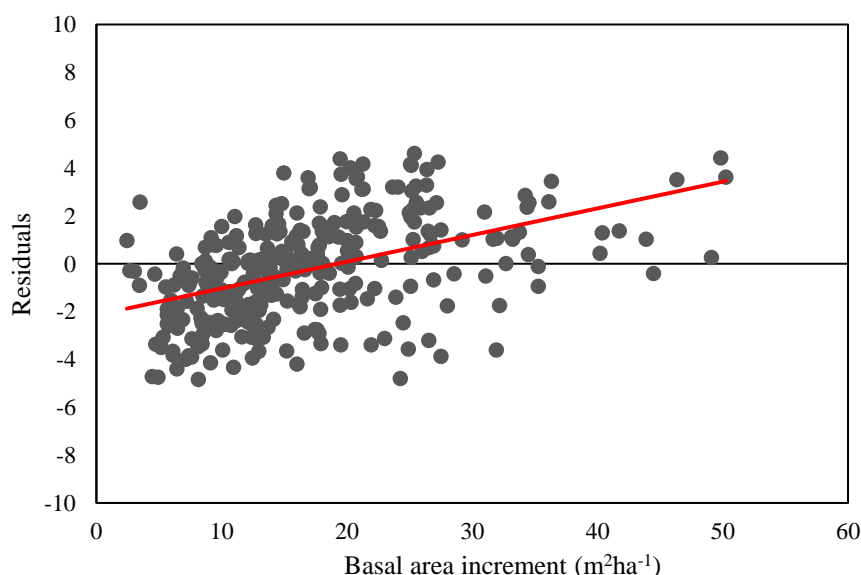
| Variable                          | Mean    | Minimum | Maximum | Standard Deviation | Coefficient of variation |
|-----------------------------------|---------|---------|---------|--------------------|--------------------------|
| N ( $\text{Tr ha}^{-1}$ )         | 154     | 30      | 500     | 74.42              | 48.32                    |
| DBH (cm)                          | 39      | 12.5    | 157     | 22                 | 56.41                    |
| BA ( $\text{m}^2\text{ha}^{-1}$ ) | 17.94   | 3.6     | 49.51   | 4.25               | 23.69                    |
| H (m)                             | 32.01   | 19.89   | 44.88   | 6.34               | 19.80                    |
| SLP (%)                           | 32.29   | 5       | 70      | 13.17              | 40.78                    |
| ALT (m)                           | 1167.94 | 815     | 1650    | 233.21             | 19.96                    |
| TEMP ( $^{\circ}\text{C}$ )       | 14.03   | 12      | 15.89   | 0.95               | 6.77                     |
| PRE (mm)                          | 41.30   | 33.48   | 44.68   | 3.21               | 7.77                     |
| ET (mm)                           | 116.13  | 111.86  | 120.11  | 2.02               | 1.73                     |

**Table 3.** Characteristics of MLP-based ANNs for basal area increment model

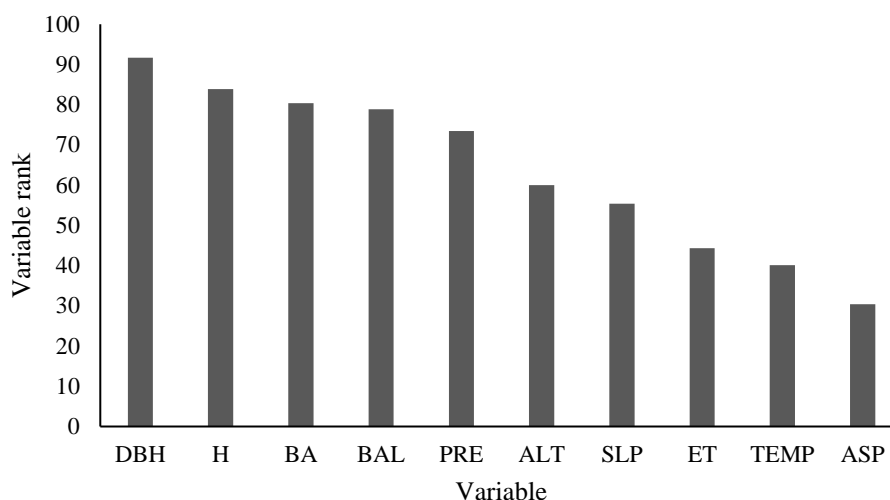
| Index             | 1           | 2                 | 3          | 4           | 5          |
|-------------------|-------------|-------------------|------------|-------------|------------|
| Net. name         | MLP 10-5-1  | <b>MLP 10-7-1</b> | MLP 10-4-1 | MLP 10-12-1 | MLP 10-8-1 |
| R <sup>2</sup>    | 0.86        | <b>0.88</b>       | 0.87       | 0.87        | 0.85       |
| RMSE              | 0.038       | <b>0.028</b>      | 0.037      | 0.032       | 0.039      |
| RMSE%             | 23.025      | <b>16.672</b>     | 22.032     | 19.054      | 23.222     |
| Bias              | 0.0035      | <b>0.0027</b>     | 0.0037     | 0.0035      | 0.004      |
| Bias%             | 2.094       | <b>1.6077</b>     | 2.2032     | 2.084       | 2.381      |
| algorithm         | BFGS 17     | <b>BFGS 15</b>    | BFGS 26    | BFGS 14     | BFGS 11    |
| Error function    | SOS         | <b>SOS</b>        | SOS        | SOS         | SOS        |
| Hidden activation | Exponential | <b>Logistic</b>   | Logistic   | Exponential | Identity   |

BFGS = Broyden–Fletcher–Goldfarb–Shanno (The numbers indicate the optimality of this algorithm in the model).

SOS = Symbiotic Organisms Search.



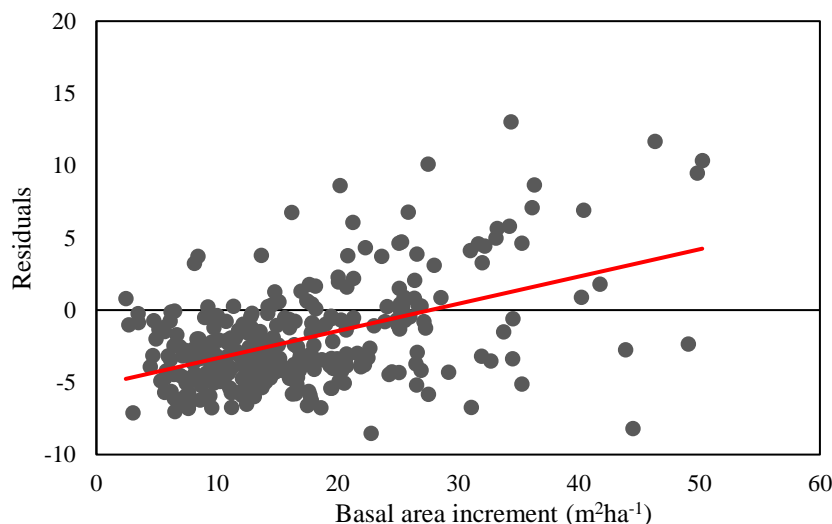
**Figure 2.** Comparison of residuals with observed basal area increment. Estimates are based on an ANN with an MLP-based architecture



**Figure 3.** Relative importance of variables in the Artificial Neural Network model for estimating basal area increment

**Table 4.** Evaluation results of different Support Vector Machine algorithm kernels

| Kernel Type    | Linear | Polynomial | RBF Gamma | Sigmoid |
|----------------|--------|------------|-----------|---------|
| R <sup>2</sup> | 0.83   | 0.80       | 0.84      | 0.67    |
| RMSE           | 0.0439 | 0.049      | 0.042     | 0.0470  |
| BIAS           | 0.0168 | 0.0093     | 0.0012    | 0.0275  |
| RMSE%          | 24.135 | 27.043     | 23.468    | 25.821  |
| BIAS%          | 9.283  | 5.1539     | 0.6802    | 15.1480 |



**Figure 4.** Model residuals against observed basal area increment by SVM

In the support vector method, the sum of the weight vectors for each variable is as follows: BA (63.52), DBH (57.21), BAI (50.11), PCP (49.32), TEMP (37.28), ET (30.65), ALT (27.16), SLP (20.45), and ASP (18.64).

### 3.4. Random Forest

In this method, basal area growth had an R<sup>2</sup> of 0.71, an RMSE of 0.058, a BIAS of 0.0034,

a %RMSE of 34.53, and a %BIAS of 2.02. In this model, the variables importance was determined as follows: DBH (0.99), Height (0.88), BA (0.73), BAL (0.55), precipitation (0.37), altitude (0.35), Evaporation and Transpiration (0.24), temperature (0.23), slope (0.20), and aspect (0.20). Based on the graph of squared error changes with increasing tree numbers in the dataset, the optimal number of

trees for modeling basal area increment was estimated at 202 (Figure 5).

### 3.5. Generalized Additive Model (GAM)

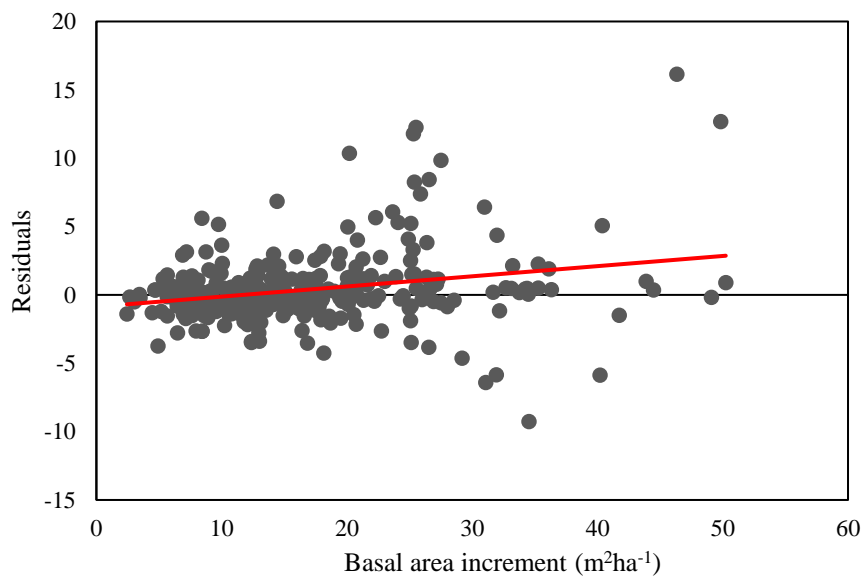
The basal area modeled using the GAM method had an  $R^2$  of 0.69, an RMSE of 0.083, a bias of 0.0059, a %RMSE of 49.42, and a %bias of 3.51. In this model, tree height, basal

area, and larger basal area were identified as influencing factors. The residual dispersion in this method is significantly higher than in other models (Figure 7).

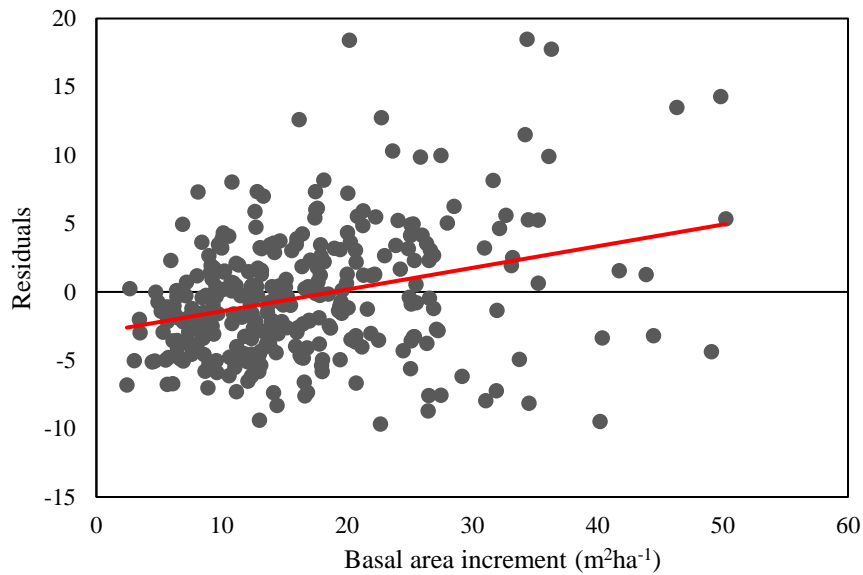
A comparison of the DBH with the annual growth of the basal area showed that the annual growth of the basal area increased as DBH increased (Figure 8).



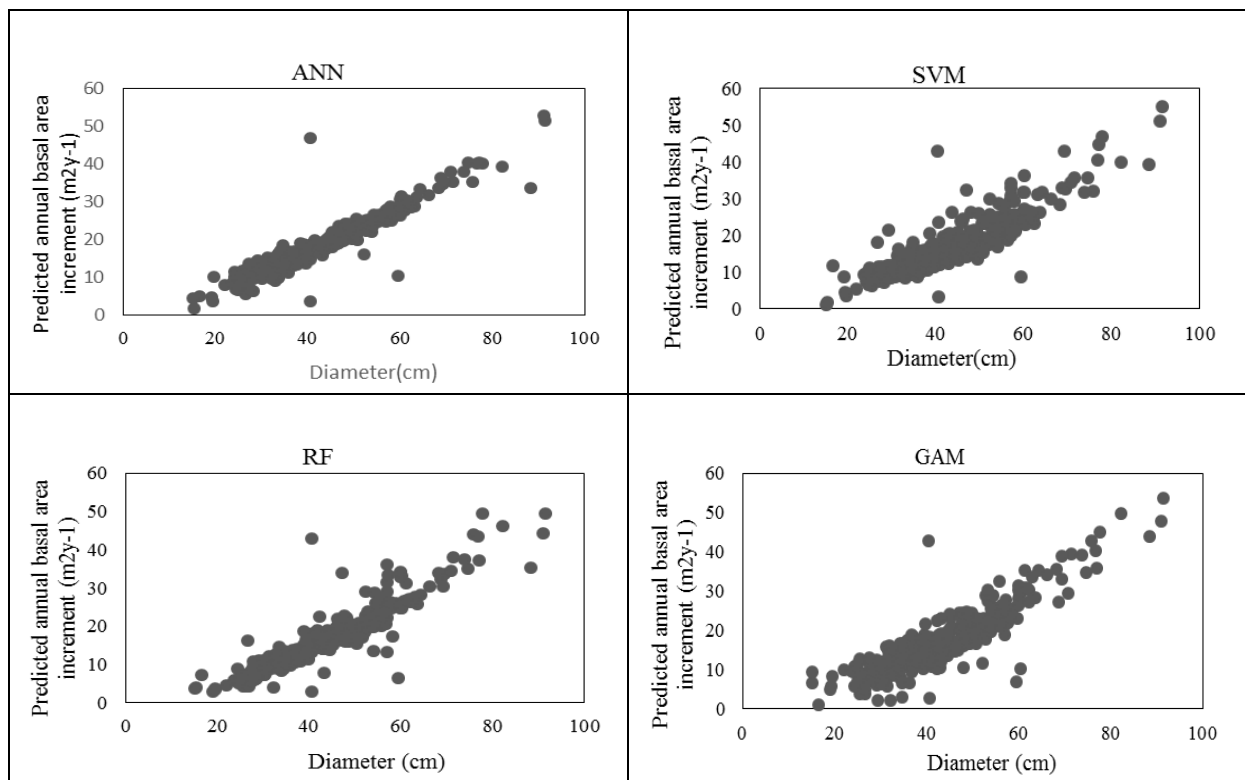
**Figure 5.** Variation in the mean squared error of basal area estimation relative to the number of trees



**Figure 6.** Model residuals against observed basal area increment based on the random forest model



**Figure 7.** Model residuals against observed basal area increment based on the Generalized Additive Model



**Figure 8.** Periodic basal area increment predictions vs. tree diameter

#### 4. Discussion

With scientific progress, forest management should be continuously optimized, evaluated, revised, or validated step by step to move closer to sustainable forest management. Therefore, determining and modeling of growth is one of the most important features in the planning and management of these forest areas, and growth

data can verify the dynamics of forest masses. In addition, models and their applications can be used to assess forest inventory and conduct related studies. In this study, machine-learning methods were used to investigate basal area growth, incorporating biometric factors, physiographic factors (Slope, Aspect, and Altitude), and climatic factors (temperature,

precipitation, evaporation and transpiration) as model inputs.

In the ANN algorithms, a robust model was developed using input variables to evaluate the effect of various factors on basal area increment (Figure 3). In this method, the MLP-based algorithm was used with 10 inputs and 7 neurons in the hidden layer, resulting in a highly accurate model with the lowest RMSE. In the study by Huang et al. (2015), the artificial neural network based on the multilayer perceptron network (MLP) provided a better model than the RBF-based ANN and explained 92% of the variation in the data. Consistent with previous studies (Crecente-Campo et al., 2010, Hamidi et al., 2021, West, 2021), the ANN algorithm in this study outperformed other models, yielding a lower RMSE. In this regard, we can refer to the studies of Lessard et al. (2001), Zhao et al. (2004), Sharma et al. (2019), Yang et al. (2004) and Huy et al. (2021). The results of all studies showed that the use of the model can provide an acceptable estimate of the growth characteristics of the basal area. Careful consideration should be given to selecting the optimal number of hidden layers and transfer functions to minimize error in this model. This is because the optimal number of neurons in the hidden layer depends on the dataset's size and quality. In these models, the analysis is repeated with different numbers of hidden layers until the evaluation criteria indicate that it leads to the best model. The ANN algorithm produced a robust model using input variables to evaluate the effects of tree biometric, physiographic, and climatic factors on basal area increment, with diameter being the most influential factor and aspect the least.

In the SVM method, the choice of kernel is crucial. Accordingly, the input variables are examined and analyzed (Kisi & Kilic, 2016). Among the four kernels tested in this study, the RBF gamma kernel performed slightly better than the linear, polynomial, and sigmoid kernels for the basal area increment model ( $R^2= 0.84$ ). In this method, based on the weight of the vectors used, basal area, diameter, and larger basal area were identified as the most important factors. Similar findings regarding the superior performance of the support vector machine model were reported by Davis & Adjuik (2022). Overall, the results of this study were comparable to those of

previous research. The results of the random forest method were similar to those of the SVM method, further demonstrating the high performance of this model. According to Figure 5, the optimal number of trees for this dataset is 202, determined as the point on the curve where increasing the number of trees no longer reduces the error. In a random forest model, the optimal number of trees depends on the number of predictors. Generally, increasing the number of trees improves model performance. Key influencing factors were identified in this method, with diameter being among the most important. A predictive pattern with reasonable accuracy can be established to estimate the basal area increment. The findings of Vrushali et al. (2014) and Nguyen et al. (2015) highlighted the strong performance of random forest model, which aligns with the present study.

The results of the generalized additive model are comparable to those of the Sigmoid SVM kernels, providing acceptable results consistent with the findings of Bayat et al. (2021). The analysis of the relationship between DBH and annual basal area growth showed that the artificial neural network model provides a better estimate of basal area growth than other models. The explanatory power of the neural network model for the relationship between diameter and annual basal area growth was 0.79, aligning with the findings of Barbosa et al. (2022) and Huy et al. (2021), confirming the model's accuracy. However, this value was lower compared to the studies of Sharma et al. (2019), Zhang et al. (2014), and Moreno et al. (2017), likely due to inclusion of mass, topography, and habitat parameters in their model. While other models performed well in the Hyrkanian forests, artificial neural networks (ANN) proved to be a more practical technique for BAI modeling, allowing for the inclusion of a wider range of variables. This reduces modeling complexity, as the ANN technique incorporates new variables based on biological theory and dynamic ecological processes rather than random correlations (Da Rocha et al., 2018). Additionally, the strong performance of the generated models based on an optimized structure (number of neurons, activation function, and input variables) demonstrates their stability and generalizability.

Future studies on species using the ANN approach will help strengthen these findings, expand their application, and improve the understanding of species dynamics. The potential for improving the description of forest parameters using machine-learning techniques, specifically ANN, for sustainable forest management, focusing on species-specific planning that aligns with the realities of forest ecosystems (Silva et al., 2023; Bayat et al., 2019, 2020), is especially important in mixed and uneven-aged forests. In these forests, accurate predictions are crucial for conserving species composition and structures that define the ecosystem (Hamidi et al., 2021). Additionally, sound intervention strategies based on reliable data can ensure economic returns for forest owners, while also preventing conversion to other land uses (Eisfeld et al., 2019; Stepka et al., 2021).

## 5. Conclusion

The main purpose of measuring forest growth is to provide accurate information for planning, resource management, development, and protection, as these decisions require precise data on forest stock. ML includes various algorithms that aim to improve performance by recognizing patterns in data, allowing them to make accurate decisions with new data. These algorithms recognize patterns

from training data, improving performance based on the results. In other words, these algorithms improve with exposure to new data. In this study, the artificial neural network algorithm outperformed the others with a correlation coefficient of 0.88 and RMSE and BIAS values of 0.028 and 0.0027, respectively. In this model, the biometric variables of trees were more important. It should be noted that various structural factors of the stand, as well as physiographic and climatic conditions, can affect the basal area increment. Providing appropriate models can lead to more reliable estimates of basal area increment for Hyrcanian forest trees, which is crucial for making informed decisions regarding sustainable forest management, development, and planning. A limitation of this study is that, due to the vastness of natural resources, particularly in the northern forests, along with limited budget and time, data collection from all areas under study was not feasible. Furthermore, some areas are difficult to access and inaccessible, meaning data collection from these areas was inevitably omitted. Future research could explore other machine learning algorithms, such as the nearest neighbor algorithm, the CARD algorithm, among others. Moreover, more input variables could be added to the model to examine their effects in the studied forests.

## References

- Adjuik, T.A., & Davis, S.C., (2022). Machine Learning Approach to Simulate Soil CO<sub>2</sub> Fluxes under Cropping Systems. *Agronomy*, 12(1), 197.
- Baskent, E.Z., & Keles, S. (2005). Spatial forest planning: A review. *Ecological modeling*, 188, 145-173.
- Barbosa, L.O., Costa, E.A., Schons, C.T., Finger, C.A.G., Liesenberg, V., & Bispo, P.d.C. (2022). Individual Tree Basal Area Increment Models for Brazilian Pine (*Araucaria angustifolia*) Using Artificial Neural Networks. *Forests*, 13, 1108. <https://doi.org/10.3390/f13071108>.
- Bayat, M., & Hamidi, S. K. (2019). Investigation some of the biotic and abiotic variables effective on the diameter increment of the beech trees at fixed sample plots level by growth models. *Ecology of Iranian Forest*, 7(13), 91-99.
- Bayat, M., Bettinger, P., Heidari, S., Henareh Khalyani, A., Jourgholami, M., & Hamidi, S. K. (2020). Estimation of tree heights in an uneven-aged, mixed forest in northern Iran using artificial intelligence and empirical models. *Forests*, 11(3), 324.
- Bayat, M., Hamidi, S.K., & Sadeghzadeh, M.H. (2019). Investigation some of the Biotic and Abiotic Variables Effective on the Diameter Increment of the Beech Trees at Fixed Sample Plots level by Growth Models. *Ecology of Iranian Forests*. 7(13), 91-99. doi:10.29252/ifej.7.13.91. (In Persian)

- Bayat, M., Knoke, T., Heidari, S., Hamidi, S. K., Burkhart, H., & Jaafari, A. (2022). Modeling tree growth responses to climate change: a case study in natural deciduous mountain forests. *Forests*, *13*(11), 1816.
- Bayat, M., Burkhart, H., Namiranian, M., Hamidi, S. K., Heidari, S., & Hassani, M. (2021). Assessing biotic and abiotic effects on biodiversity index using machine learning. *Forests*, *12*(4), 461. <https://doi.org/10.3390/f12040461>
- Bayat, M., Bettinger, P., Masteali, S. H., Hamidi, S. K., Masood Awan, H. U., & Abolhasani, A. (2023). Recreation potential assessment at Tamarix forest reserves: A method based on multicriteria evaluation approach and landscape metrics. *Forests*, *14*(4), 705.
- Bettinger, P., Gratez, D., & Sessions, J. (2005). A Density- dependent stand-level optimization approach for deriving management prescriptions for interior northwest (USA) landscapes. *Forest Ecology and Management*, *217*(2-3), 171-186.
- Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
- Burkhart, H.E. (1990). Status and future of growth and yield models. In: proc. A symp. On state-of the methodology of forest inventory. *USDA Forest Service*, *283*, 409-414.
- Crecente-Campo, F., Tomé, M., Soares, P., & Diéguez-Aranda, U. (2010). A generalized nonlinear mixed-effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. *Forest Ecology Management*, *259*(5), 943–952. <https://doi.org/10.1016/j.foreco.2009.11.036>.
- da Rocha, S. J. S. S., Torres, C. M. M. E., Jacovine, L. A. G., Leite, H. G., Gelcer, E. M., Neves, K. M., ... & Zanuncio, J. C. (2018). Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. *Science of the total environment*, *645*, 655-661.
- Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, *40*, 139-157.
- Eisfeld, R.D.L., Arce, J.E., Sanquetta, C.R., & Braz, E.M. (2019). Is it forbidden the wood use of *Araucaria angustifolia*? An analysis on the current legal budget. *Floresta*, *50*, 971–982.
- Ghaderi, I., Hassanzad Navroodi, I., & Torkaman, J. (2013). Effect of altitude on annual diameter growth of *Quercus libani* Oliv in Kurdistan province. *Iranian Journal of plant research*, *26*(4), 434-443. (In Persian)
- Görgens, E. B., Montaghi, A., & Rodriguez, L. C. E. (2015). A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Computers and Electronics in Agriculture*, *116*, 221-227.
- Hai, F.Z., Xue, M.S., Zhi, Y.Y., Peng, X., Yan, X., & Hua, T. (2011). August temperature variability in the southeastern Tibetan Plateau since A.D.1385 inferred from tree rings. *PALAEO*, *5*, 703.
- Hamidi, S.K., Fallah, A., Bayat, M., & Hosseini Yekani, S.A. (2017). Determining the Forest Volume Growth using Permanent Sample Plots (Case Study: Farim Forest, Jojadeh District). *Ecology of Iranian Forests* *4*(8), 1–8. (In Persian)
- Hamidi, S.K., Fallah, A., Bayat, M., & Hosseini Yekani, S.A. (2019). Individual Tree Growth Models for Management of Uneven aged and Mixed Hyrcanian Forests (Case Study: Farim Forest). *Iranian Journal of Forest*, *3*(11), 373–386. (In Persian)
- Hamidi, S. K., Zenner, E. K., Bayat, M., & Fallah, A. (2021). Analysis of plot-level volume increment models developed from machine learning methods applied to an uneven-aged mixed forest. *Annals of Forest Science*, *78*, 1-16. <https://doi.org/10.1007/s13595-020-01011-6>
- Hamidi, S. K., Weiskittel, A., Bayat, M., & Fallah, A. (2021). Development of individual tree growth and yield model across multiple contrasting species using nonparametric and parametric methods in the Hyrcanian forests of northern Iran. *European Journal of Forest Research*, *140*(2), 421-434. <https://doi.org/10.1007/s10342-020-01340-1>

- Hamidi, S.K., Fallah, A., Bayat, M., & Hosseini, Y.S.A. (2021). Investigating the diameter and height models of beech trees in uneven age forest of Northern Iran (Case study: Forest Farim). *Ecology of Iranian Forests*, 9(17), 30-40. doi:10.52547/ifej.9.17.30 (In Persian)
- Hamidi, S. K., de Luis, M., Bourque, C. P. A., Bayat, M., & Serrano-Notivoli, R. (2023). Projected biodiversity in the Hyrcanian Mountain Forest of Iran: An investigation based on two climate scenarios. *Biodiversity and Conservation*, 32(12), 3791-3808. <https://doi.org/10.1007/s10531-022-02470-1>
- Hamidi, S.K., Fallah, A., Nazaryani, N. (2023). Modelling the most appropriate vegetation indices under the influence of climatic factors using sentinel 2 images - Case study: Farim Forest. *Scientific-Research Quarterly of Geographical Data (SEPEHR)*, 32(127), 55-76. doi: 10.22131/sepehr, (2023).1983358.2935. (In Persian)
- Hamidi, S.K., & Fallah, A. (2024). Evaluation of the Performance of Climatic Scenarios in the Basal Area Model of Trees (Case Study: the Farim Forest). *Ecology Iranian Forest*, 12(2), 1-14. doi:10.61186/ifej.12.2. (In Persian)
- Huang, H. X., Li, J. C., & Xiao, C. L. (2015). A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. *Expert Systems with Applications*, 42(1), 146-155.
- Huy, B., Poudel, K.P., & Temesgen, H. (2021). Individual tree diameter growth modeling system for Dalat pine (*Pinus dalatensis* Ferré) of the upland mixed tropical forests. *Forest Ecology and Management*, 480, 118612.
- IPCC. (2019). Summary for policymakers. In: Abe-Ouchi, A., Gupta, K., Pereira, J. (Eds.), IPCC Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities.
- Jafarian, Z., & Karger, M. (2016). Modeling the distribution of protective and valuable plant species in the Pleur tourist area using the generalized linear model (GLM) and the generalized aggregate model (GAM). *Geography and Development Quarterly*, 15(4), 117-132.
- Kisi, O., & Kilic, Y. (2016). An investigation on generalization ability of artificial neural networks and M5 model tree in modeling reference evapotranspiration. *Theoretical and applied climatology*, 126, 413-425.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Lessard, V.C., McRoberts, R.E., & Holdaway, M.R. (2001). Diameter growth models using Minnesota forest inventory and analysis data. *Forest Science*, 47(3), 301-310.
- Lohmander, P. (2017) A general dynamic function for the basal area of individual trees derived from a production theoretically motivated autonomous differential equation. *Interdisciplinary Journal of Management Studies*, 10, 917-928.
- Mauya, E. W., Hansen, E. H., Gobakken, T., Bollandsås, O. M., Malimbwi, R. E., & Næsset, E. (2015). Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. *Carbon balance and management*, 10, 1-14. doi: 10.1186/s13021-015-0021-x.
- Moreno, P.C., Palmas, S., Escobedo, F.J., Cropper, W.P., & Gezan, S.A. (2017). Individual-tree diameter growth models for mixed *Nothofagus* second growth forests in southern Chile. *Forests*, 8, 12. 506.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
- Nazariani, N., Fallah, A., Hamidi, S. K., & Varamesh, S. (2022). Estimation of quantitative characteristics of Zagros forests using data mining nonparametric algorithms (case study: Olad

- Ghobad Watershed, Koohdasht, Lorestan). *Forest Research and Development*, 8(3), 249-263. (In Persian)
- Nguyen, T. T., Huang, J. Z., & Nguyen, T. T. (2015). Unbiased Feature Selection in Learning Random Forests for High-Dimensional Data. *The Scientific World Journal*, 2015(1), 471371. DOI: 10.1155/2015/471371
- Pokharel, B., & Froese, R.E. (2009). Representing site productivity in the basal area increment model for FVS-Ontario. *Forest Ecology and Management*, 258, 666-675.
- Rees, F., Doherty, M., Grainge, M. J., Lanyon, P., & Zhang, W. (2017). The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. *Rheumatology*, 56(11), 1945-1961.
- Silva, J.P.M., da Silva, M.L.M., de Mendonça, A.R., da Silva, G.F., de Barros Junior, A.A., da Silva, E.F., Aguiar, M.O., Santos, J.S., & Rodrigues, N.M.M. (2023). Prognosis of forest production using machine learning techniques. *Information Processing in Agriculture*, 10(1), 71-84.
- Sharma, M., & Parton, J. (2007). Height-diameter Equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology Management*, 249, 187-198.
- Sharma, R.P., Stefancik, I., Vacek, Z., & Vacek, S. (2019). Generalized nonlinear mixed-effects individual tree diameter increment models for beech forests in Slovakia. *Forests*, 10(5), 451.
- Shataee, S., Kalbi, S., Fallah, A., & Pelz, D. (2012). Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *International journal of remote sensing*, 33(19), 6254-6280.
- Stavins, Robert N., and Kenneth R. Richards. 2005. "The Cost of U.S. Forest-Based Carbon Sequestration." Arlington, VA: Pew Center on Global Climate Change.
- Stepka, T.F., de Mattos, P.P., Figueiredo Filho, A., Braz, E.M., do Amaral Machado, S. (2021). Growth of *Araucaria angustifolia*, *Cedrela fissilis* and *Ocotea porosa* in different locations within the *Araucaria* Forest in the Southern Brazil. *Forest Ecology Management*, 486, 118941.
- Strobl, R. O., & Forte, F. (2007). Artificial neural network exploration of the influential factors in drainage network derivation. *Hydrological Processes: An International Journal*, 21(22), 2965-2978.
- Townsend, P. A. (2002). Estimating forest structure in wetlands using multitemporal SAR. *Remote sensing of environment*, 79(2-3), 288-304.
- Vahedi, A.A., Fallah, A., Akhavan, R., Nazariani, N., Khatibnia, E., & Hamidi, S.K. (2024). Spatial Analyses for Fine Woody Debris Volume Stock in the Hyrcanian Research Forest of Kheyrood-Kenar. *Ecology Iranian Forest*, 12(1), 73-87. doi:10.61186/ifej.12.1.73. (In Persian)
- Vanclay, J. K. (1994). Sustainable timber harvesting: Simulation studies in the tropical rainforests of north Queensland. *Forest ecology and management*, 69(1-3), 299-320.
- Vargas-Larreta, B., Castedo-Dorado, F., Alvarez-Gonzalez, J.G., Barrio-Anta, M., & Cruz-Cobos, F. (2009). A generalized height-diameter model with random coefficients for uneven-aged stands in El Salto, Durango (Mexico). *Forestry*, 82, 445-462.
- Vrushali, Y., Kulkarni, P., Pradeep, K. (2014). Effective learning and classification using Random Forest algorithm, *International Journal of Engineering and Innovative Technology (IJEIT)*, 3(11), May (2014).
- West, P.W. (2021). Modelling maximum stem basal area growth rates of individual trees of *Eucalyptus pilularis* Smith. *Forest Science*, 67(6), 633– 636. <https://doi.org/10.1093/forsci/xfab047>
- Wykoff, W.R. (1986). Supplement to the user's guide for the stand Prognosis model: Version 5.0. General Technical Report INT-208, USDA Forest Service, *Intermountain Research Station*, Ogden, UT pp: 36.

Wykoff, W.R. (1990). A basal area increment model for individual conifers in the northern Rocky Mountains. *Forest Science*, 36, 1077-1104.

Yang, Y., Monserud, R. A., & Huang, S. (2004). An evaluation of diagnostic tests and their roles in validating forest biometric models. *Canadian Journal of Forest Research*, 34(3), 619-629.

Zhao, D., Borders, B., & Wilson, M. (2004). Individual-tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississippi alluvial valley. *Forest Ecology and Management*, 199(2-3), 307-322.

Zhao, K., Popescu, S., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment*, 115(8), 1978-1996.

Zhang, X., Duan, A., Zhang, J., & Xiang, C. (2014). Estimating Tree Height-Diameter Models with the Bayesian Method. *The Scientific World Journal*, 2014(1), 683691.



## ارزیابی مدل‌های رویش سطح مقطع درخت با استفاده از الگوریتم‌های یادگیری ماشین

سیده کوثر حمیدی<sup>۱\*</sup> و اصغر فلاح<sup>۲</sup>

<sup>۱</sup>استادیار گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران  
<sup>۲</sup>آستاد گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران

(تاریخ دریافت: ۱۴۰۳/۰۸/۱۴؛ تاریخ پذیرش: ۱۴۰۳/۱۰/۱۵)

### چکیده

برنامه‌ریزی مدیریت جنگل یک ابزار مهم تصمیم‌گیری در جنگلداری است و نتیجه آن یک برنامه مدیریتی است که در آن از فعالیت‌های پیش‌بینی شده، زمان بندی و کنترل آنها برای دستیابی به اهداف مدیریت جنگل در یک منطقه جنگلی استفاده می‌شود. بررسی مدل‌های رشد و محصول یکی از مهم‌ترین روش‌ها برای به دست آوردن اطلاعات در مورد وضعیت آینده یک جنگل است. به عبارت دیگر، ارزیابی رشد و عملکرد توده یک پیش‌نیاز اساسی برای برنامه‌ریزی مدیریت جنگل است. بنابراین تعیین و برآورد رویش سطح مقطع درختان برای درک وضعیت و اجرای آن در برنامه‌ریزی و مدیریت جنگل بسیار مهم است. با توجه به اینکه گونه‌های جنگلی هیرکانی به‌عنوان با ارزش‌ترین گونه‌ها محسوب می‌شوند، هدف این پژوهش بررسی رویش سطح مقطع با استفاده از الگوریتم‌های یادگیری ماشین و مدل‌سازی آن‌ها در جنگل ناهمسال فریم در استان مازندران است. در این مطالعه، رویش سطح مقطع درختان با استفاده از الگوریتم‌های یادگیری ماشین (Machine Learning) شبکه عصبی مصنوعی (ANN)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF) و مدل افزایشی تعمیم یافته (GAM) در یک دوره ۱۰ ساله مدل‌سازی شد. اطلاعات زیست‌سنجی (قطر برابر سینه، ارتفاع کل، سطح مقطع درختان، سطح مقطع قطورترین درختان)، فیزیوگرافی (شیب، جهت، ارتفاع از سطح دریا) و اقلیمی (درجه حرارت، بارندگی و تبخیر و تعرق) جنگل به عنوان ورودی برای توسعه مدل استفاده شد. نتایج الگوریتم‌های یادگیری ماشین با یکدیگر مقایسه شد. مدل ANN، به ویژه الگوریتم MLP با هفت نرون لایه پنهان، دقت بالاتری (۸۸ درصد) در پیش‌بینی افزایش سطح مقطع در مقایسه با مدل‌های دیگر به دست آورد. این نتایج نشان می‌دهد که فناوری‌های ANN برای مدل‌سازی رویش سطح مقطع در جنگل‌داری مناسب هستند، همچنین عملکرد خوب مدل‌های تولید شده بر اساس ساختار مناسب (تعداد نورون‌ها، نوع تابع فعال‌سازی و متغیرهای ورودی) نشان‌دهنده پایداری این مدل‌ها و توانایی آن‌ها در ارائه تعمیم است. امکان بهبود توصیف پارامترهای جنگل با استفاده از تکنیک‌های یادگیری ماشینی، مانند ANN، برای مدیریت پایدار جنگل برای افزایش حفاظت از ترکیب گونه‌ها و ساختارهایی که جنگل را مشخص می‌کنند، ضروری است.

**واژه‌های کلیدی:** برنامه‌ریزی جنگل، سطح مقطع، جنگل هیرکانی، شبکه عصبی مصنوعی، یادگیری ماشین.