



**Journal of Environmental Studies**  
**Vol. 48, No. 4, Winter 2023**

**Journal Homepage:** [www.Jes.ut.ac.ir](http://www.Jes.ut.ac.ir)  
**Print ISSN: 1025-8620**      **Online ISSN 2345-6922**

**Comparison of Random Forest Models, Support Vector Machine and Multivariate Linear Regression for Biodiversity Assessment in the Hyrcanian Forests**

**Document Type**  
Research Paper

**Received**  
June 05, 2022

**Accepted**  
December 14, 2022

**Mahmoud Bayat<sup>1\*</sup>, Reza Akhavan<sup>1</sup>, Sahar Heidari<sup>2</sup>, Seydeh Kosar Hamidi<sup>3</sup>**

1 Department of forest, Research Institute of Forests and Rangelands, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran

2 Department of Environment Faculty of Natural Resources, College of Agriculture and natural resources, University of Tehran, Karaj, Iran

3 Department of Forestry, Faculty of Natural Resources, Sari Agriculture Sciences and Natural Resource University, Sari, Iran

DOI: [10.22059/JES.2022.345264.1008336](https://doi.org/10.22059/JES.2022.345264.1008336)

**Abstract**

Biodiversity is an important structural feature of dynamic and complex forest ecosystems. One of the most challenging and important issues in assessing the structure of forest ecosystems is understanding the relationship between biodiversity and environmental factors. Hyrcanian Forests are considered a biodiversity hotspot in the world and have special and unique features that have led to an emphasis and importance of biodiversity conservation in these forests. The aim of this study was to investigate the effect of biotic and abiotic factors on the diversity and richness of tree species in Hyrcanian Forests from the west of Gilan province to the east of Golestan province. For this purpose, using 655 fixed sample plots (0.1 hectare), the diversity of trees in 3 provinces in the northern Iran from east to west of the Caspian Sea was analyzed. A combination of non-parametric models including random forest (RF) and support vector machine (SVM) and linear regression models were used to investigate the relationship between tree diversity and biotic and abiotic factors. Biotic and abiotic variables included the number of trees per hectare, diameter, respectively. Basal area (BA), Basal Area in Largest tree (BAL), slope, aspect and elevation. Evaluation statistics including the coefficient of determination, RMSE and percentage RMSE error showed that the random forest model was the best model to determine the relationship between biodiversity and environmental factors and has suitable accuracy for determining biodiversity changes in the northern forests of Iran.

**Keywords:** Altitude, Tree Species Diversity, Hyrcanian Forests, Linear Regression.

\* Corresponding Author:

Email: [mbayat@rifr-ac.ir](mailto:mbayat@rifr-ac.ir)

## Extended Abstract

### Introduction

Hyrceanian Forests are one of the most important resources in the world and biodiversity is a key factor in their role. The forests of northern Iran are one of the richest forest ecosystems in the temperate regions of the world in terms of plant species diversity. There are many species of trees and shrubs naturally in these forests. Topographic factors (altitude, slope and aspect) have the greatest impact on plant diversity in these forests, because the emergence of vegetation is the result of collisions and interactions between vegetative elements and physical environmental factors. Biodiversity is an important structural feature in dynamic and complex forest ecosystems that is influenced by a variety of factors in determining its changes. The use of modeling and statistical principles helps to understand these changes. However, one of the most challenging issues in modeling forest ecosystems understands the relationship between biodiversity and environmental factors.

Machine learning technologies using a set of statistical and modeling approaches, their application to identify patterns and hidden relationships in the database has been proven; also with large data technology and high-performance, computing methods a new one has been created for analysis in natural resources. The potential of using these programs for the multi-factor forecasting and processes makes this computer-based tool a powerful approach to forest modeling and management. The aim of this study was to investigate the effect of the biotic and abiotic factors on the diversity of trees in Hyrcanian Forests in northern Iran. For this purpose, a combination of non-parametric models and parametric models has been used. This study is unique in that the selected forest areas are very comprehensive and cover almost the entire Hyrcanian forest from west to east.

### Materials and Methods

Forest areas have been selected to cover all Hyrcanian Forests from west to east and the sample plots are well distributed throughout the forest (Figure 1). In this study, 15,988 trees were measured in 655 fixed circular sample plots (0.1 ha). In all habitats, the same method was used to collect data. A rectangular census network with dimensions of  $200 \times 150$  m was established in the forest area. Diameter and species type of all living trees with a diameter of more than 7.5 cm were measured using a caliper ruler. Other characteristics of the sample plots such as slope, aspect and altitude, distance and azimuth between the trees in each sample plot relative to the center of the sample plot were measured and recorded. Due to the mix and heterogeneous forest stands and in order to prepare the height curve in each sample plot, the tree closest to the center of the sample plot and the thickest tree in terms of height and diameter were measured.

### Discussion of results

In this study, two machine learning methods as well as statistical analysis were used to model the biodiversity of trees in Hyrcanian forests.

The results of the linear multivariate regression showed that the mean of the Shannon index was 0.67 (SE = 0.38) and ranged from 0 to 1.64. The following shows the relationship between Shannon diversity index with physiographic factors and biotic factors (diameter, basal area and basal area of the thickest trees per square meter per hectare) which were significant.

$$H = 0.7 + 0.004/0 (BA) - 0.005/0 (BAL)$$

In this regard,  $R^2 = 0.019$ , the mean square root of the error was 0.58 and the mean residual deviation (Bias) was 10.97.

In the decision machine vector model, the results showed that the RBF gamma nucleus has the lowest RMSE in both training and evaluation models.

Shannon index using the random forest method had  $R^2 = 0.55$ , with RMSE of 0.33, BIAS of 0.05, relative RMSE of 49.86% and relative BIAS of 7.46%. The result shows modeling with multiple linear regression and a random forest model. As shown in the figure, the stochastic forest model, which was known as the best machine learning model, and the regression model (parametric model), and height and BAL, were recognized as the most important factors determining the diversity of tree species, respectively. The results showed that machine learning methods, especially random forest, are more accurate than support vector machine.

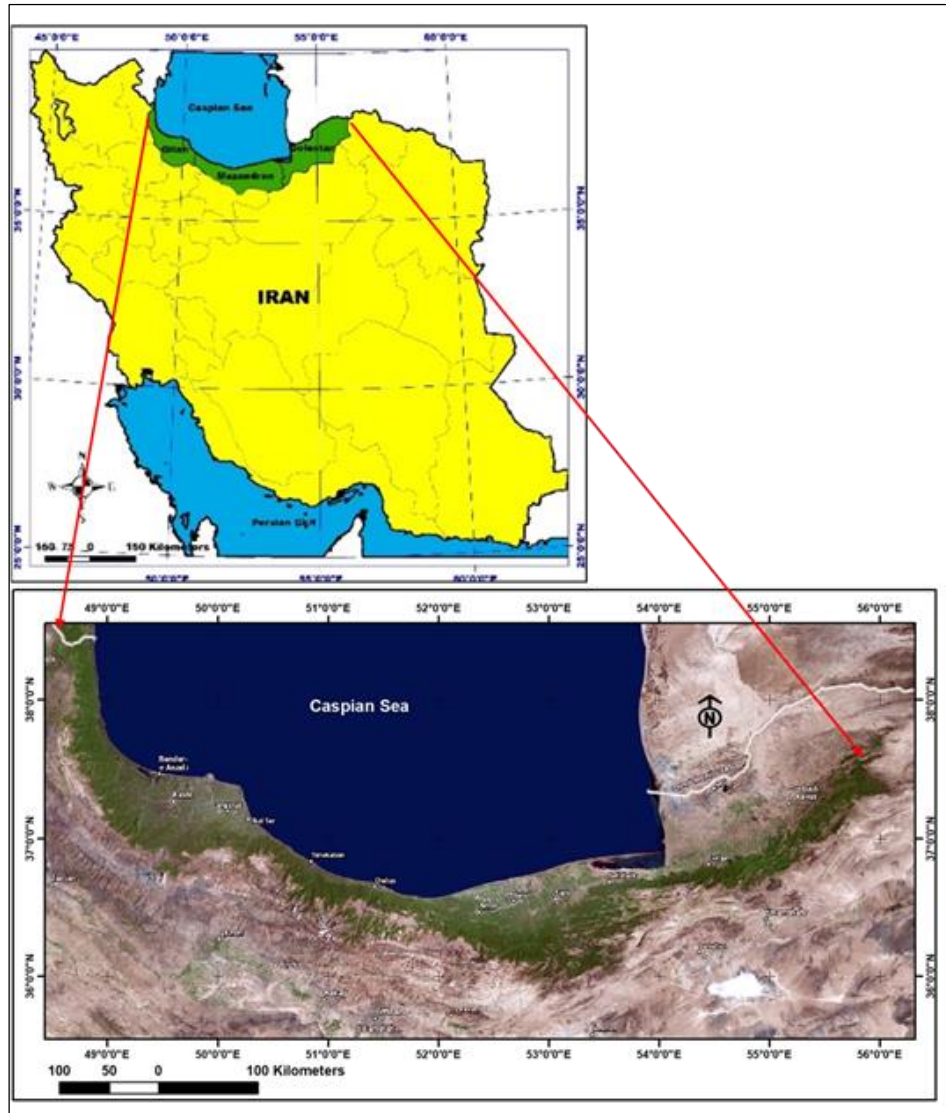


Figure 1. Map of Iran (a); Location of studied forests in northern Iran (b)

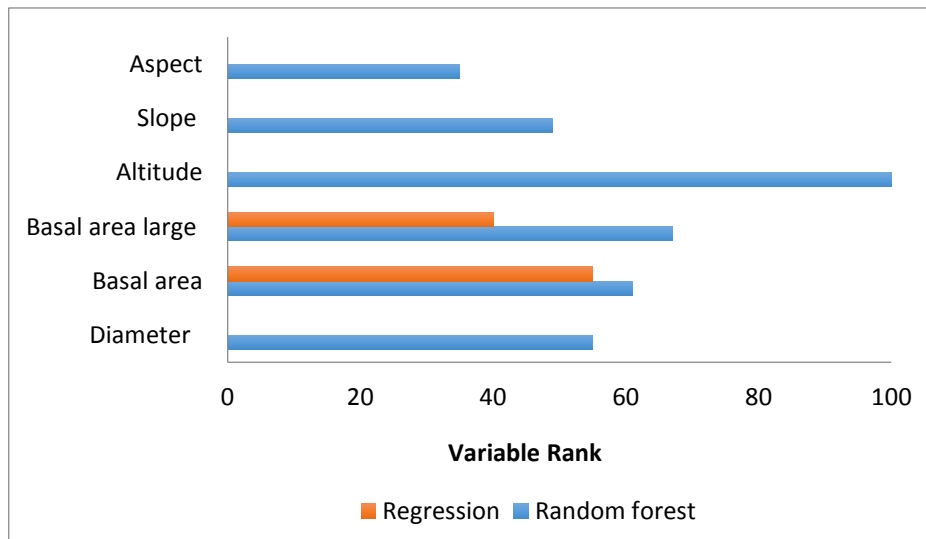


Figure 2- Relative importance of predictor variables in regression and random forest

### **Conclusions**

In fact, with the advent of machine learning methods, the accuracy of solving complex problems in many areas has increased in recent years. Machine learning methods have been widely used to estimate and predict many forest features. In general, although machine learning techniques can often be superior to traditional and regression methods, because they have fewer assumptions about data.

## مقایسه مدل‌های جنگل تصادفی، ماشین بردار تصمیم‌گیری و رگرسیون خطی چندمتغیره در ارزیابی تنوع زیستی جنگل‌های هیرکانی

محمود بیات<sup>۱\*</sup>، رضا اخوان<sup>۱</sup>، سحر حیدری مستعلی<sup>۲</sup>، سیده کوثر حمیدی<sup>۳</sup>

۱ بخش تحقیقات جنگل، مؤسسه تحقیقات جنگل‌ها و مراتع کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران  
۲ گروه محیط زیست، دانشکده منابع طبیعی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.  
۳ گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه کشاورزی و منابع طبیعی ساری، ساری، ایران.

تاریخ پذیرش مقاله: ۱۴۰۱/۰۹/۲۳

تاریخ وصول مقاله: ۱۴۰۱/۰۴/۱۴

### چکیده

تنوع زیستی از ویژگی‌های ساختاری مهم در اکوسیستم‌های جنگلی پویا و پیچیده است. یکی از چالش برانگیزترین و مهمترین موضوعات در ارزیابی ساختار اکوسیستم جنگل، درک رابطه بین تنوع زیستی و عوامل محیطی است. جنگل‌های هیرکانی دارای تنوع زیستی قابل توجهی در سطح جهانی بوده و از ویژگی‌های خاص و منحصر به فرد برخوردار هستند که باعث تاکید و حساسیت بر حفاظت از تنوع زیستی در این جنگل‌ها شده است. هدف از این مطالعه بررسی تأثیر عوامل زنده و غیر زنده بر تنوع و غنای گونه‌ای درختی در جنگل‌های هیرکانی از غرب استان گیلان تا شرق استان گلستان می‌باشد. برای این منظور و جهت رسیدن به این هدف با استفاده از ۶۵۵ قطعه نمونه ثابت (۱/۱ هکتاری) تنوع درختان را در سه استان شمالی کشور از شرق تا غرب دریای خزر مورد تجزیه و تحلیل قرار گرفت. ترکیبی از روش‌های ناپارامتریک شامل، جنگل تصادفی (RF) و ماشین بردار تصمیم‌گیری (SVM) و مدل‌های رگرسیون خطی برای مدل سازی و بررسی رابطه بین تنوع درخت و عوامل زنده و غیر زنده مورد استفاده قرار گرفت. متغیرهای زنده و غیر زنده به ترتیب شامل تعداد درختان در هکتار، قطر برابر سینه، سطح مقطع برابر سینه قطورترین درختان (BAL)، شیب، جهت و ارتفاع از سطح دریا بود. آماره‌های ارزیابی شامل ضریب تعیین، خطای RMSE نشان داد مدل جنگل تصادفی در بین مدل‌های ارایه شده، بهترین مدل برای تعیین رابطه تنوع زیستی و عوامل محیطی بود و از دقت مناسبی برای تعیین تغییرات تنوع زیستی در سطح جنگل‌های شمال کشور برخوردار است.

**کلید واژه‌ها:** ارتفاع سطح دریا، تنوع گونه‌ای درختی، جنگل‌های هیرکانی، رگرسیون خطی

### سراغاز

زیستی یک ویژگی ساختاری مهم در اکوسیستم‌های جنگلی پویا و پیچیده است که عوامل متنوعی در تعیین تغییرات آن تاثیرگذار است (Monarrez-Gonzalez et al., 2020). استفاده از مدلسازی و اصول آماری به درک این تغییرات کمک فراوانی می‌کند. با این حال یکی از چالش برانگیزترین

اکوسیستم‌های جنگلی یکی از مهمترین اکوسیستم‌هایی است که خدمات متنوعی را به انسان ارائه می‌دهند (Zarandian et al., 2016). این اکوسیستمها نقش مهمی در تأمین نیازهای زیستگاهی موجودات مختلف دارند. تنوع

پردازش داده‌ها. ب) انتخاب الگوریتم و پ) انتخاب راه حل‌های ارزیابی می‌باشد (Bourque et al., 2019; Júnior et al., 2020). اگرچه مدل‌های یادگیری ماشین دارای دقت بالاتر و ظرفیت بیشتر در حل مسائل پیچیده و رفتارهای غیرخطی سیستمی هستند، با این وجود این روش‌ها نیاز به آموزش دقیق و همچنین داده‌های تعلیمی مناسب هستند. اما، نظارت طولانی مدت و دقیق می‌تواند هزینه‌بر بوده و جمع‌آوری، ذخیره‌سازی و به روز رسانی اطلاعات با مشکلاتی رو به رو شود. بنابراین، محققان توصیه می‌کنند که برای درک بهتر مکانیسم‌های مختلف محیطی در جنگل و همچنین حل مشکلات مختلف، ترکیبی از مدل‌ها و روش‌های مختلف پارامتریک و یا روش‌های ناپارامتریک و پارامتریک با هم استفاده شود (Liu et al., 2018). توجه به این نکته مهم است که برای همه مطالعات یک روش ثابت جهانی ML وجود ندارد و انتخاب مناسبترین و بهترین روش یا ترکیبی از روش‌ها به ماهیت موضوع و مسائل بستگی دارد. در این زمینه مطالعات مختلفی در ایران و جهان انجام شده است به عنوان مثال (Ahmadi et al., 2020)، با استفاده از داده‌های ماهواره‌ای و روش ماشین یادگیری، خصوصیات جنگل مانند سطح مقطع در هکتار، تعداد و حجم در هکتار را برآورد کردند. آنها نتیجه گرفتند که این مدل‌ها، برای تخمین چنین پارامترهایی مناسب هستند.

Bayat et al (2021a) و Bayat et al (2020)، به ترتیب رویش جنگل در طول یک دوره ده‌ساله و ارتفاع درختان را در جنگل‌های آمیخته و ناهمسال در شمال کشور، با استفاده از مدل‌های یادگیری ماشین پیش‌بینی و مدل‌سازی کردند و عوامل موثر را تعیین نمودند. (Shen et al (2020) از روش یادگیری ماشین و روش‌های آماری جغرافیایی برای پیش‌بینی زیست‌توده روی زمین (AGB) در جنگل‌ها ی چین استفاده کردند و نتیجه گرفتند که روش جنگل تصادفی پیشنهادی یک روش مطمئن و دقیق برای تعیین نقشه AGB در جنگل‌ها ی نیمه گرمسیری با توپوگرافی پیچیده می‌باشد. جنگل‌ها ی هیرکانی به صورت نواری بلند و باریک در در

موضوعات در مدل‌سازی اکوسیستم‌های جنگلی، درک رابطه بین تنوع زیستی و عوامل محیطی است، اگرچه در سالهای اخیر این موضوع مورد توجه تحقیقات بسیاری قرار گرفته و در این زمینه تحقیقات گسترده‌ای انجام شده است (Chu et al., 2019; Yao et al., 2020). اما هنوز نکات نامعلوم زیادی باقی مانده است که ضرورت بر انجام پژوهش‌های بیشتر در این باره بسیار زیاد است. از جمله تجزیه و تحلیل تنوع زیستی در جنگل در ارتباط با عوامل محیطی که می‌تواند اطلاعات زیادی در مورد ساختار جامعه جنگل فراهم کند. تنوع گونه‌ای به طور کلی با افزایش عرض جغرافیایی کاهش می‌یابد (Yao et al., 2020) عوامل موثر بر تنوع گونه‌ای تنها عوامل غیر زنده و محیطی نیستند، بلکه در یک محیط و تحت تاثیر عوامل زنده مانند رقابت تغییر می‌کنند (Roslin et al., 2017). در حقیقت، شرایط زیست‌شناختی بر ترکیب گونه‌ها تأثیر می‌گذارد و اختلافاتی را در تنوع جامعه ایجاد می‌کند (Šímová et al., 2011). چرا که عوامل زیستی مربوط به ساختار جامعه تأثیر مهمی در دسترسی به آشیان اکولوژیک برای گونه‌های همزیست دارند. مطابق با بسیاری از پژوهش‌های انجام شده، ترکیب اکوسیستم‌های جنگلی نتیجه فعل و انفعالات زیست‌شناختی بوده و همراه با تأثیرات زیست محیطی تغییر می‌یابد (Yao et al., 2020). در سالهای اخیر، مدل‌های ناپارامتریک یادگیری ماشین (ML)<sup>۱</sup> برای بررسی تأثیر تغییرات شرایط محیطی بر تنوع زیستی بطور گسترده‌ای مورد استفاده قرار گرفته است. یادگیری ماشین یکی از مهمترین شاخه‌های هوش مصنوعی است که به دلیل مزایای چشمگیر آن در برخی شرایط، برای تجزیه و تحلیل داده‌های جنگل نیز بسیار مورد استفاده قرار می‌گیرد که از آن جمله کاربرد یادگیری ماشین برای بهبود برآوردهای جنگل در سطح محلی، منطقه‌ای و جهانی در این اکوسیستم‌ها است (Bayat et al., 2019; Silva et al., 2019). همچنین از روش‌های ماشین یادگیری در مدیریت جنگل و ارزیابی خطرهای ناشی از عوامل طبیعی به دفعات استفاده شده است. فرآیندهای این مدل‌ها اغلب شامل: الف) انتخاب و پیش

شده‌اند انجام شد. بر اساس آخرین آمار موسسه تحقیقات جنگل‌ها و مراتع کشور، این جنگل‌ها ۱/۶۵ میلیون هکتار مساحت دارند، که ۱/۱٪ کشور و ۱۵٪ کل جنگل‌های ایران را تشکیل می‌دهد. (Mirakhorlo., 2016) گونه‌های اصلی درختان در جنگل‌های هیرکانی عبارتند از راش، ممرز، بلوط بلندمازو، آفر پلت، افرا شیردار، نمدار، توسکای بیلاقی و قشلاقی (Borque & Bayat, 2015) از نظر تقسیم جهانی، این جنگل‌ها، جنگل‌های پهن‌برگ خزان‌کننده هستند که دارای آب و هوای مرطوب نیمه‌مدیترانه‌ای هستند اما تعداد کمی از درختان سوزنی برگ نیز در این جنگل‌ها یافت می‌شوند که سرخدار و ارس از مهمترین این گونه‌ها هستند. میانگین بارندگی سالانه از ۶۰۰ میلی متر در شرق تا ۲۰۰۰ میلی‌متر در غرب جنگل‌ها (آستارا) متغیر است. روش جنگل‌شناسی نزدیک به طبیعت روش غالب مدیریتی در جنگل‌های شمال ایران است که باعث ایجاد جنگل‌های آمیخته و ناهمسال شده است. (شکل ۱).

در هر رویشگاه، یک شبکه آماربرداری مستطیل شکل به ابعاد ۲۰۰ × ۱۵۰ متر در منطقه جنگل ایجاد شد.

قطر برابر سینه و نوع گونه همه درختان زنده که قطری بیشتر از ۷/۵ سانتی‌متر داشتند با استفاده از خطکش دوبازو اندازه‌گیری شدند. سایر مشخصات قطعات نمونه از قبیل شیب، جهت غالب و ارتفاع از سطح دریا فاصله و آزیموت بین درختان در هر قطعه نمونه نسبت به مرکز قطعه نمونه اندازه‌گیری و ثبت شد. به دلیل آمیخته و ناهمگن بودن توده‌های جنگلی و به منظور تهیه منحنی ارتفاع در هر قطعه نمونه، نزدیکترین درخت به مرکز قطعه نمونه و قطورترین درخت از نظر ارتفاع و قطر اندازه‌گیری شدند.

جداول ۱ و ۲ آمار توصیفی داده‌های استفاده شده در مدل تنوع گونه‌ای درختی را در هشت رویشگاه جنگلی نشان می‌دهد.

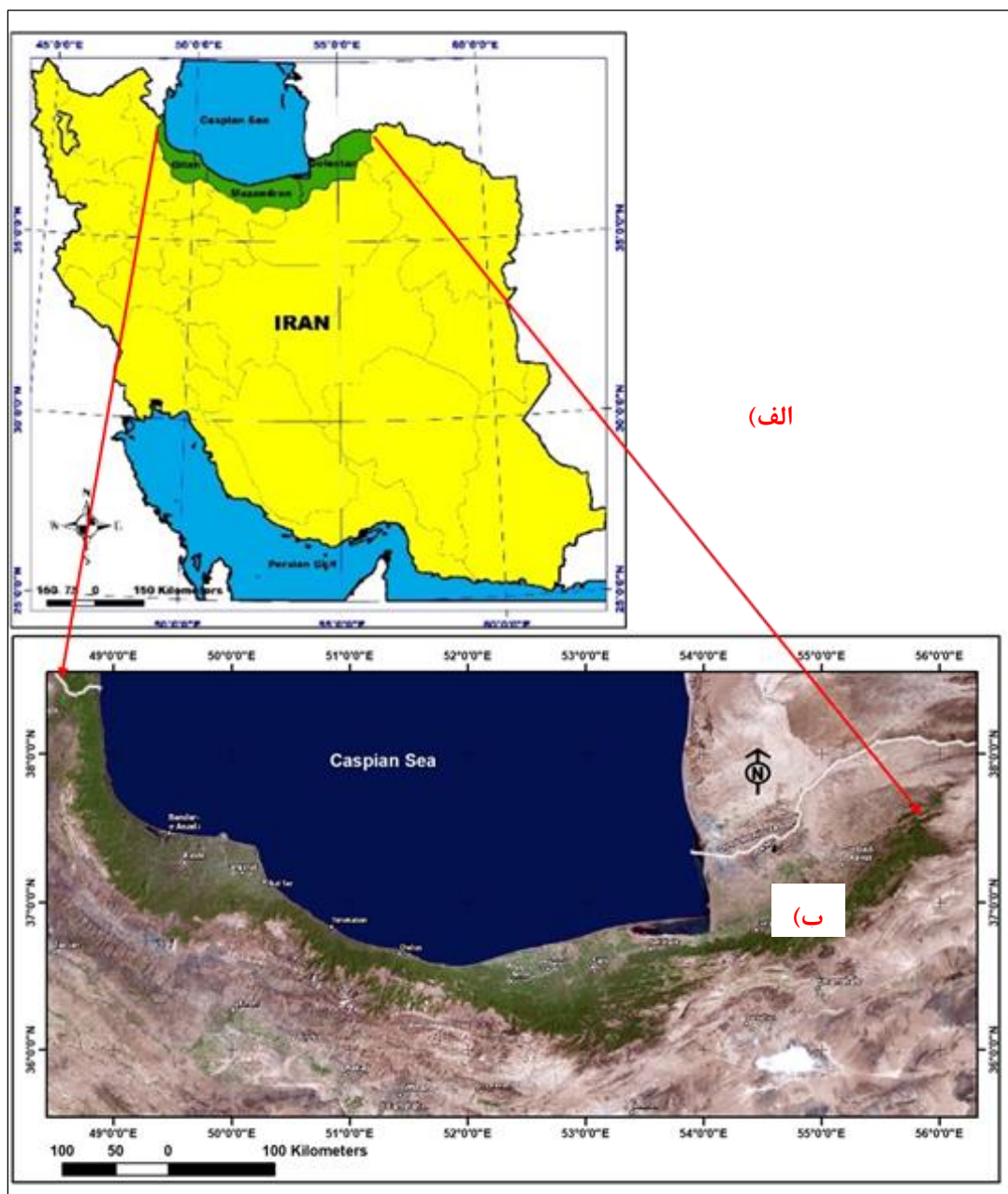
امتداد مرزهای جنوبی دریای خزر و در امتداد دامنه‌های شمالی کوه البرز واقع شده‌اند. جنگل‌های هیرکانی با بیش از ۳۲۰۰ گونه گیاهی دارای تنوع گونه‌ای قابل توجهی در سطح جهانی هستند. این منطقه با حدود ۴۴٪ گیاهان آوندی در ایران از اهمیت ویژه‌ای برخوردار است. در این جنگل‌ها حدود ۵۰۰ گونه گیاهی بومی ایران وجود دارد. همه این ویژگی‌ها بر لزوم محافظت از تنوع زیستی این منطقه تأکید دارند (Mirakhorlo., 2016).

هدف از این مطالعه بررسی عوامل تأثیر گذار زنده و غیرزنده بر تنوع درختان جنگلی در جنگل‌های هیرکانی در شمال ایران است. برای این منظور ترکیبی از مدل‌های ناپارامتریک (مدل‌های یادگیری ماشینی) و مدل‌های رگرسیونی استفاده شده است. مدل‌های ناپارامتریک شامل RF و SVM بوده و سعی شده است تا مهمترین متغیرهای زنده شامل میانگین قطر برابر سینه در قطعه نمونه (DBH)، سطح مقطع قطورترین درختان به متر مربع در هکتار (BAL)، و سطح مقطع درختان به متر مربع در هکتار (BA) و تعداد درخت در هکتار و متغیرهای غیر زنده شامل شیب، جهت و ارتفاع در نظر گرفته و وارد مدل شوند. این مطالعه از آن جهت منحصر به فرد است که مناطق جنگلی انتخاب شده بسیار جامع بوده و تقریباً کل جنگل هیرکانی را از غرب به شرق پوشش می‌دهد. در حالی که مطالعات گذشته در مقیاس کوچک و فقط شامل بخش محدودی از این جنگل‌ها را از نظر تنوع زیستی مطالعه کرده است. همچنین برای اولین بار چندین روش مدلسازی، شامل مدل رگرسیونی و مدل‌های یادگیری ماشین با هم مقایسه شده و بهترین مدلی که کارایی مناسبی در مدلسازی تنوع زیستی داشت انتخاب شد.

## مواد و روش‌ها

### منطقه مورد بررسی

این تحقیق در کل جنگل‌های شمال ایران که از آستارا در شمال غربی ایران تا گیلداغی گرگان در شمال شرقی ایران گسترش یافته و در سه استان گیلان، مازندران و گلستان واقع



شکل ۱- (الف) نقشه ایران؛ (ب) موقعیت جنگل‌های هیرکانی در شمال ایران

### متغیرهای محیطی (غیر زنده)

#### فاکتورهای فیزیوگرافیک

عوامل فیزیوگرافی شامل ارتفاع، شیب و جهت تأثیر قابل توجهی بر پوشش گیاهی یک منطقه دارند و ارتفاع عامل تعیین کننده‌ای در تعیین دما و رژیم آب و هوایی یک منطقه است. در یک ارتفاع مشخص، شیب و جهت شیب بر نور تابیده شده در منطقه تأثیر می‌گذارند و باعث تشکیل خرد اقلیم‌های موجود در جنگل می‌شود (Tiwari et al., 2020).

### متغیرهای زنده

میانگین قطر برابر سینه به سانتیمتر در قطعه نمونه:

یکی از مهمترین عوامل برای مدل سازی رشد و عملکرد جنگل است (Sharma & Parton, 2007). اندازه‌گیری این متغیر نسبتاً آسان بوده بنابراین داده‌های مرتبط با آن تقریباً فراوان و در دسترس است. (Chen et al., 2020).



مدیریت جنگل است و تعداد نسبی درختان یک منطقه یا یک توده را نشان می‌دهد. این شاخص نمادی از ساختار توده‌های جنگلی است و از این نظر بسیار مهم است (Oyebade et al., 2020). تعداد در هکتار توسط رابطه ۲ محاسبه می‌شود:

$$N = \frac{\bar{n}}{a} \quad \text{رابطه ۲}$$

که،  $N$  تعداد درختان در هکتار است،  $\bar{n}$  تعداد متوسط درختان در قطعه‌ها و  $a$  سطح زمین‌ها به هکتار است.

### شاخص شانون وینر

شاخص‌های تنوع گونه‌ای در حقیقت ترکیبی از غنای گونه و یکنواختی است. شاخص Shannon Weiner دو مقدار غنای گونه و یکنواختی را در یک مقدار جمع می‌کند. در این مطالعه، از شاخص‌های شانون وینر برای بررسی تنوع گونه‌ها استفاده شد. در این تحقیق ۱۵۹۸۸ درخت در ۶۵۵ قطعه اندازه‌گیری و گونه‌ها شناسایی شدند و شاخص شانن بر اساس این داده‌ها برآورد شده است. نحوه محاسبه این شاخص به شرح زیر، از طریق رابطه ۳ است:

$$H = -\sum \pi_i \ln(\pi_i) \quad \text{رابطه ۳}$$

که  $H$  شاخص شانون وینر است و  $\pi_i$  فراوانی نسبی نوع  $i$ th است.

### سطح مقطع $^2$ ( $m^2ha^{-1}$ )

سطح مقطع درختان به متر مربع در هکتار: این معیار در واقع به عنوان مجموع سطح مقطع درختان در قطرهای اندازه‌گیری شده و در هر هکتار بیان می‌شود. سطح مقطع در واحد سطح شاخص مهمی در مدیریت و اکولوژی جنگل است زیرا با حجم در هکتار توده‌های جنگلی در ارتباط است.

### سطح مقطع بزرگترین درختان $^3$ ( $m^2ha^{-1}$ )

این شاخص یکی از مهمترین شاخص‌های تعیین رقابت در جنگل می‌باشد و از آنجا که به نور در دسترس درختان در ارتباط می‌باشد، از اهمیت زیادی برخوردار است. میزان نور موجود در درختان کوچک با افزایش BAL کاهش می‌یابد. (Burkhardt & Tomé, 2012). این شاخص از طریق رابطه ۱ محاسبه می‌شود:

$$BAL = \frac{\pi}{4} \cdot \sum_{j=1}^n (TF_j \cdot DBH_j^2) \quad \text{رابطه ۱}$$

که  $DBH_j > DBH_i$  به عنوان مثال، تمام درختان بزرگتر از درخت مورد نظر.

### تعداد درختان در هکتار و نوع گونه‌ها

این شاخص یکی از شاخص‌های مهم در اندازه‌گیری و

جدول ۱- مشخصات متغیرهای زنده در قطعه‌های نمونه ثابت جنگل در هر سه استان مورد مطالعه

شاخص شانون	BAL	سطح مقطع	قطر برابر سینه	متغیر	استان
۰/۶	۲۴/۴	۲۹/۴	۳۶/۳	میانگین	گیلان
۱/۶	۵۲/۹	۵۰/۹	۶۷/۸	حداکثر	
۰	۹	۷/۴	۱۷/۶	حداقل	
۱/۱	۲۳/۲	۳۰/۴	۲۵/۸	میانگین	گلستان
۱/۶	۴۳	۶۲/۴	۴۵/۳	حداکثر	
۰/۳	۱۱/۷	۱۶/۸	۱۶/۵	حداقل	
۰/۶۶	۱۷/۱	۲۱/۷	۳۹/۳	میانگین	مازندران
۱/۵۶	۱۱۲/۷	۱۱۲/۸	۱۷۵	حداکثر	
۰	۰/۱	۰/۶	۱۳/۹	حداقل	

جدول ۲- مشخصات متغیرهای غیرزنده در قطعه‌های نمونه ثابت جنگل در هر سه استان مورد مطالعه

استان	متغیر	شیب	جهت	ارتفاع از سطح دریا
گیلان	میانگین	۴۵/۵	۲/۵	۱۳۶۳/۳
	حداکثر	۸۵	۷	۱۸۵۱
	حداقل	۵	۱	۸۶۵
گلستان	میانگین	۲۶/۶	۱/۴	۱۵۵۷
	حداکثر	۵۰	۷	۱۶۸۶
	حداقل	۵	۱	۱۴۵۳
مازندران	میانگین	۲۹/۸۳	۲/۸۵	۱۰۸۹/۴۲
	حداکثر	۱۰۰	۹	۱۶۵۰
	حداقل	۰	۰	۷۰

### آنالیز آماری

به منظور مقایسه هشت منطقه مورد مطالعه، از نظر شاخص‌های تنوع‌زیستی، پس از آزمون نرمال و همگنی واریانس داده‌ها، به ترتیب از آزمون تجزیه واریانس یک طرفه و آزمون توکی و دانکن برای مقایسه کلی بین گروه‌ها و مقایسه میانگین استفاده شد. کلیه محاسبات فوق با استفاده از نرم افزار SPSS 22 انجام شده است که نتایج آن در جداول ۲ و ۳ ارائه شده است.

### روش‌های مدل‌سازی

#### مدل ماشین بردار تصمیم‌گیری<sup>۴</sup>

الگوریتم ماشین بردار پشتیبان یکی از روش‌های یادگیری ماشین بوده که نتایج آن وابسته به تنظیم درست و صحیح پارامترهای ورودی مدل است (Eshaghi & Shataei, 2016). این روش، یک روش آماری ناپارامتری و نظارت شده بوده که در واقع یک نوع طبقه بندی باینری است (Mountrakis et al, 2011). در این مطالعه، چهار نوع هسته از نوع خطی، چند جمله‌ای، و سیگموئید و کرنل تابع پایه شعاعی<sup>۵</sup> که توسط  $\gamma$  در رابطه ۴ تعریف شده، مورد بررسی قرار گرفتند (Vafaei et al, 2018). پارامترهای هسته شامل ظرفیت (c)، گاما ( $\gamma$ ) و اپسیلون ( $\epsilon$ ) هستند. جستجوی شبکه مشخص شامل محدوده‌ای از ۱ تا ۵۰ ظرفیت برابر با دامنه

متغیرهای ورودی است. نرخ اپسیلون از ۰/۱ تا ۰/۵ متفاوت است. در داخل، SVM مدل را برای بهینه‌سازی محاسبه می‌کند. دقت این نتایج بهینه‌سازی توسط پارامترهای ظرفیت و اپسیلون کنترل می‌شود.

$$\text{رابطه ۴} \quad \gamma > 0k (xi, x) = \exp(-\gamma |xi - x|^2)$$

#### مدل جنگل تصادفی<sup>۶</sup>

مدل جنگل تصادفی یک الگوریتم یادگیری ماشین نظارت شده است که مبتنی بر درخت تصمیم‌گیری بوده که در آن تعداد زیادی درخت تصمیم‌گیری ایجاد و درختان برای پیش‌بینی، ترکیب می‌گردند. این روش، روشی جدید و قدرتمند است که پیشرفت‌های خوبی در داده‌کاوی و مدل‌سازی ارائه داده (Jafarian & Karegar, 2017) و در حال حاضر یکی از بهترین الگوریتم‌های یادگیری است. این مدل طبقه‌بندی با دقت بالا را برای بسیاری از مجموعه‌های داده انجام می‌دهد (Lee et al, 2017).

بر اساس داده‌های ورودی، ۱۱ گره ایجاد و از ۶۶۴ درخت اولیه برای آموزش و آزمایش استفاده شد تا گرافی تولید شود که میانگین میزان خطای مربع به ازای هر درخت را نشان دهد. در این مدل، یکی از اصلی‌ترین پارامترهایی که باید تنظیم شود متغیر  $k$  برای هر گره است. ساده‌ترین راه برای تعیین  $k$  محاسبه ریشه مربع تعداد کل متغیرهای مستقل

مورد استفاده قرار گرفتند. از ضریب تبیین ( $R^2$ ) ریشه دوم میانگین خطا (RMSE) ضریب اریبی (روابط ۶، ۷ و ۸) برای ارزیابی مدل‌های ارایه شده استفاده شد.

$$R^2 = \frac{\text{Regression SS}}{\text{Total SS}} = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad \text{رابطه ۶}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{رابطه ۷}$$

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}) \quad \text{رابطه ۸}$$

$y_i$  تنوع درختی که تخمین زده می‌شود، میانگین تنوع مشاهده شده درخت است و  $y_i$  میانگین تنوع درختی تخمینی است.

### نتایج

جدول زیر نتایج تجزیه واریانس بین هشت رویشگاه جنگلی را نشان می‌دهد و همانطور که از عدد Sig مشاهده می‌شود، از نظر شاخص تنوع شانون تفاوت معنی‌داری بین مناطق وجود دارد.

جدول ۳- نتایج تجزیه واریانس بین هشت منطقه جنگلی مورد مطالعه

Sig.	F	میانگین مربعات	درجه آزادی	مجموع مربعات	
**./...	۱۶/۱۱۷	۲/۰۹۵	۷	۱۴/۶۶۳	بین گروه‌ها
		۰/۱۳۰	۶۵۶	۸۵/۲۵۹	درون گروه‌ها
			۶۶۳	۹۲/۹۲۲	جمع

\*\*معنی‌داری در سطح ۹۹ درصد

$$\text{رابطه ۹} \quad H = 0.70 + 0.04(BA) - 0.05(BAL)$$

که  $H$  شاخص تنوع شانون است،  $BA$  سطح مقطع توده (به متر مربع در هکتار) و  $BAL$  سطح مقطع قطورترین درختان است (به متر مربع در هکتار). در این رابطه  $R^2 = 0.019$ ، میانگین ریشه مربع خطای  $0.58$  و میانگین انحراف باقیمانده (Bias)  $10.97$  بود.

است (رابطه ۵)، به عنوان مثال  $k \leq \sqrt{n}$ ، که  $n$  تعداد متغیرهای ورودی است.

$$\hat{Y}_n = (fX_n) = \frac{1}{K} \sum_{k=1}^K T_{L(\theta_n)} \quad \text{رابطه ۵}$$

هر درخت برآورده کننده  $TL\theta$  به یک وکتور یا خط تصادفی  $\theta$  وابسته بوده که نشان دهنده دسته بندی نمونه از مجموعه آموزشی  $L$  است. برآورد کننده نهایی  $f$ ، اکثریت رای یا میانگین برای همه درختان است.

### مدل رگرسیون خطی چندگانه

مدل رگرسیونی چند متغیره خطی<sup>۷</sup> با هدف ایجاد یک رابطه خطی بین متغیرهای مستقل از جمله شیب، جهت، ارتفاع، آزمون و  $BAL$ ،  $Ba$ ،  $dbh$  و تعداد درخت و متغیر وابسته (شاخص تنوع شانون) به کار گرفته شد.

### مدلسازی و ارزیابی

داده‌های برداشت شده از عرصه به طور تصادفی به داده‌های آموزش و تست تقسیم شدند به صورتی که ۷۰ درصد داده‌ها برای تعلیم و مابقی برای تست یا ارزیابی مدل

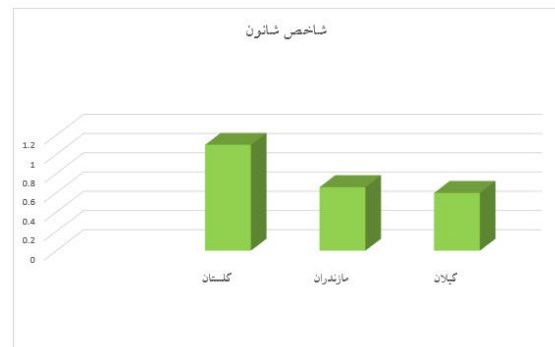
در شکل ۲، بیشترین شاخص تنوع شانون در گلستان (۱/۱) و کمترین مربوط به جنگل گیلان (۰/۶) است.

### نتایج رگرسیون چند متغیره خطی

میانگین شاخص شانون  $0.67$  ( $SE = 0.38$ ) و در محدوده ۰ تا  $1/64$  بود. معادله ۹ رابطه بین شاخص تنوع شانون با فاکتورهای فیزیوگرافی و عوامل زنده (قطر، سطح مقطع سطح مقطع قطورترین درختان به متر مربع در هکتار) را نشان می‌دهد که معنی‌دار بودند.

### مدل ماشین بردار پشتیبان

نتایج نشان داد که هسته گامای RBF در هر دو مدل آموزش و ارزیابی دارای کمترین RMSE است (جدول ۴) مجموع مشخصات وزنی هر یک از بردارهای پشتیبانی متغیر به شرح زیر است: ارتفاع (۱۰۶/۰۸)، جهت (۶۰/۰۱)، شیب (۵۱/۵۰)، قطر (۲۹/۶۸)، سطح مقطع (۱۳/۵۸) و BAL (۸/۱۸).



شکل ۲- میانگین شاخص تنوع زیستی شانون در قطعات نمونه دائمی برای هر استان

جدول ۴- ارزیابی SVM ها بر اساس انواع مختلف هسته

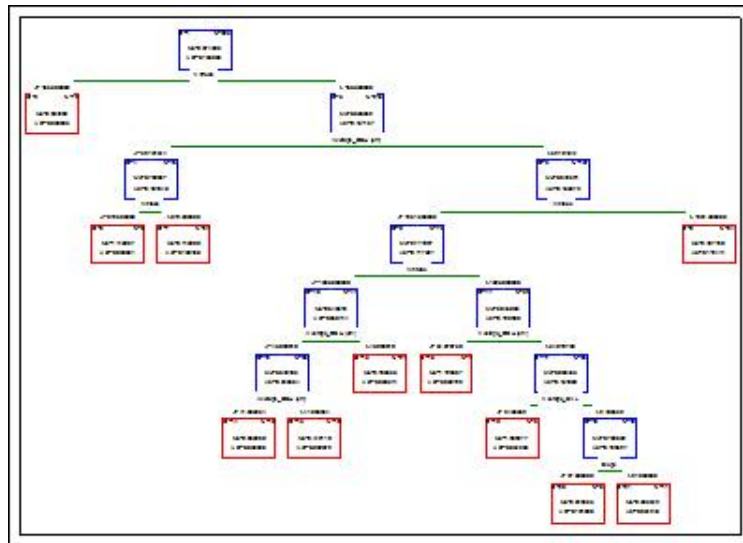
نوع کرنل	خطی		Polynomial		کرنل پایه شعاعی گاما		سیگموئید
	آموزش	ارزیابی	آموزش	ارزیابی	آموزش	ارزیابی	
R <sup>2</sup>	۰/۳۰	۰/۳۰	۰/۳۱	۰/۳۲	۰/۳۲	۰/۳۴	۰/۱۵
RMSE	۰/۴۰	۰/۴۳	۰/۴۵	۰/۴۳	۰/۳۶	۰/۳۸	۰/۵۵
BIAS	۰/۰۸	۰/۰۷	۰/۰۸	۰/۰۵	۰/۰۲	۰/۰۱۹	۰/۱۴
%RMSE	۵۹/۷۰	۶۴/۱۷	۶۷/۱۶	۶۴/۱۷	۵۳/۶۷	۵۵/۴۶	۸۲/۰۸
%BIAS	۱۱/۶۴	۱۰/۴۴	۱۱/۹۴	۷/۴۶	۳/۱۶	۲/۸۹	۲۰/۸۹
گاما	-	-	۰/۱۶	۰/۱۶	۰/۱۶	۰/۱۶	۰/۱۶
اپسیلون	۰/۱	۰/۱	۰/۱	۰/۱	۰/۱	۰/۱	۰/۱
ظرفیت	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰

### مدل جنگل تصادفی

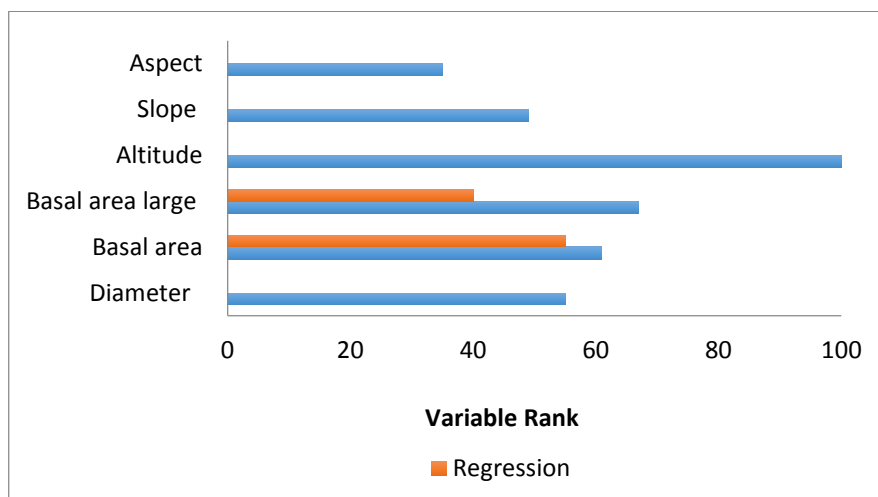
شاخص شانون با استفاده از روش جنگل تصادفی دارای  $R^2=0/55$ ، با  $RMSE=0/33$  متر مکعب در هکتار،  $BIAS=0/05$ ،  $RMSE$  نسبی  $49/86\%$  و  $BIAS$  نسبی  $7/46\%$  بود. درخت تصمیم‌گیری ابزاری قدرتمند و مفید برای طبقه‌بندی و پیش‌بینی است. درخت حاصل (شکل ۳) شامل گره‌های زیادی است که توسط شاخه‌ها به هم متصل شده‌اند. در شکل ۸ هر گره حاوی اطلاعاتی در مورد تعداد نمونه‌ها (N) در آن گره و توزیع مقادیر متغیر وابسته است. موارد موجود در گره ریشه همه مشاهدات موجود در مجموعه آموزش است. مهمترین متغیرهای تصمیم‌گیری در این مدل به ترتیب، ارتفاع با اهمیت ۱،۰ سطح مقطع قطورترین درختان در هکتار با اهمیت ۰/۶۷، سطح مقطع در هکتار با اهمیت ۰/۶۱ قطر برابر سینه با اهمیت ۰/۵۵، شیب

با اهمیت ۰/۴۹ و جهت با اهمیت ۰/۳۵ می‌باشد. همچنین، بر اساس نتایج به دست آمده، مدل‌سازی با هفت متغیر در هر گره ( $k=8$ ) منجر به حداقل RMSE شد و به عنوان K بهینه انتخاب شد.

شکل ۳ نتیجه مدل‌سازی با رگرسیون خطی چندگانه و مدل جنگل تصادفی را نشان می‌دهد. همانطور که در شکل نشان داده شده است، مدل جنگل تصادفی که به عنوان بهترین مدل یادگیری ماشین شناخته شد و مدل رگرسیون (مدل پارامتریک) بوده و ارتفاع و BAL، به ترتیب به عنوان مهمترین عوامل تعیین‌کننده تنوع گونه‌ای درختی شناخته شدند.



شکل ۴. درخت تصمیم حاصل از جنگل تصادفی



شکل ۵. اهمیت نسبی متغیرهای پیش بینی کننده در رگرسیون و جنگل تصادفی (بهترین مدل یادگیری ماشین)

مقایسه ضرایب تعیین، RMSE و RMSE نسبی جنگل تصادفی منجر به بهترین مدل‌ها شده است (جدول ۵) روش‌های مختلف برآورد نشان می‌دهد که روش‌های

جدول ۵. مقایسه آماری متناسب با روش‌های انجام شده در این مطالعه

مدل	R <sup>2</sup>	RMSE	%RMSE
رگرسیون خطی چندگانه	۰/۰۱۹	۰/۵۸	۸۶/۵۶
مدل ماشین پشتیبان بردار	۰/۳۴	۰/۳۸	۵۵/۴۶
مدل جنگل تصادفی	۰/۵۵	۰/۳۳	۴۹/۸۶

### بحث و نتیجه گیری

جلوگیری از سیل و فرسایش، تولید چوب، فرصت‌های گردشگری، تفریح، سلامتی و رفاه از جمله این خدمات هستند. تهدید و چالش اصلی تنوع زیستی در جنگل‌های

جنگل‌های هیرکانی طیف گسترده‌ای از خدمات با ارزش اکوسیستم‌ها را دارا می‌باشند. از زیبایی، ترسیب کربن،

که می‌توان این کاهش را ناشی از بهره‌برداری شدید این جنگل‌ها دانست. که منجر به کاهش تنوع زیستی شده استروند کاهش تنوع از غرب به شرق ثابت بود. در این تحقیق برای مدلسازی از روش‌های ناپارامتریک (ماشین یادگیری)، شامل، SVM و RF، استفاده شد. همچنین، برای مقایسه بین مدل‌های ماشین یادگیری و مدل‌ها و روش‌های پارامتریک، از روش رگرسیون خطی چند متغیره استفاده شده است. از بین مدل‌های یادگیری ماشین، مدل RF با  $R^2 = 0.55$  و  $RMSE = 0.33$  بهترین مدل انتخاب شد. مدل SVM با  $R^2$  برابر با  $0.34$  داشت. همچنین، مدل رگرسیون دارای  $0.19$   $R^2$  و  $RMSE$  برابر با  $0.58$  بود. بنابراین، همانطور که مشاهده می‌شود، مدل RF بهترین مدل در بین مدل‌های یادگیری ماشین بود که با نتایج بسیاری از محققان در این زمینه همخوانی داشت. همانطور که نتایج تجزیه و تحلیل حساسیت در شکل ۹ نشان می‌دهد، فاکتور ارتفاع مهمترین عامل موثر بر تنوع زیستی در مدل انتخاب شده (RF) است، دومین عامل تأثیرگذار BAL است. BAL مهمترین عامل برای مدل رگرسیون نیز بود. عوامل فیزیوگرافی به ویژه عامل ارتفاع تأثیر مهمی بر شاخص تنوع زیستی در جنگل‌ها، دارند. مطالعات مشابه نشان داده است که عوامل فیزیوگرافی به عنوان عواملی موثر بر شاخص غنای گونه‌ای و تنوع زیستی گونه‌ها هستند. همانطور که نتایج نشان می‌دهد، فاکتور ارتفاع عامل اصلی تأثیرگذار در مدل‌های انتخاب شده (RF) و رگرسیون) بود. این متغیر عامل قابل توجهی است که بر تنوع گونه تأثیر می‌گذارد (Woldu et al, 2020; Kebede et al., 2013; Cui and Zheng, 2016). مشابه آن (Kebede et al., 2013) نیز نتیجه گرفت که جهت، دامنه و تنوع ارتفاعی در مناظر اتیوپی وجود انواع مختلف پوشش گیاهی و تنوع را تحت تأثیر قرار داده است. (Zhang et al. (2013) نیز نتیجه گرفتند ارتفاع یکی از مهمترین عوامل موثر بر توزیع جامعه و تنوع گونه‌ای در Beijing، چین بود. نتایج مطالعات مختلف نشان می‌دهد که ارتفاعات میانی به طور کلی دارای بالاترین شاخص غنای گونه‌ای و تنوع زیستی گونه‌ای هستند

هیرکانی از بین رفتن زیستگاه و تکه تکه شدن ناشی از تبدیل جنگل‌ها به زمین‌های کشاورزی و شهری، خانه‌های شخص، سد سازی و غیره استبا توجه به وجود این چالش‌های مهم و خطرات ناشی از بین رفتن تنوع زیستی در جنگل‌های هیرکانی، نیاز به یک تحقیق جامع و کامل در ارتباط با تنوع زیستی و غنای گونه‌ای بود (Zarandian et al., 2016). در این مطالعه، شاخص تنوع زیستی شانون وینر برای سه استان شمالی کشور محاسبه شد. این شاخص از  $1/1$  در جنگل‌های استان گلستان تا  $0.67$  در جنگل‌های استان گیلان متغیر بود. آزمون تجزیه واریانس نشان داد که از نظر شاخص تنوع شانون اختلاف معنی داری بین سه استان وجود دارد، با توجه به نتایج آزمون توکی و دانکن، در استان گلستان بیشترین اختلاف را با سایر مناطق جنگلی دارد. همانطور که نتایج نشان می‌دهد، تنوع زیستی در جنگل‌های هیرکانی از غرب به شرق (از غرب استان گیلان تا شرق استان گلستان) اغلب روند نزولی دارد. بر اساس تحقیقات، یکی از عوامل مهم در ایجاد این روند افزایش دما بوده که با افزایش دما و درجه حرارت از غرب به شرق تنوع زیستی نیز افزایش پیدا نمی‌کند که این روند بر خلاف رطوبت و بارندگی است که همانطور که از داده‌های هواشناسی مشاهده می‌شود، میزان بارندگی سالانه از  $1345$  میلی‌متر در آستارا در استان گیلان در غربی ترین نقطه تا  $524$  میلی‌متر در استان گلستان در شرقی‌ترین نقطه متغیر است. علاوه بر این، عوامل دیگری نیز وجود دارد که باعث کاهش شاخص‌های تنوع زیستی در یک منطقه جنگلی می‌شود، که یکی از مهمترین آنها ارتفاع از سطح دریا، است. تنوع گونه‌ها با افزایش ارتفاع کاهش می‌یابد. اما استثنائاتی در این روند وجود دارد که دلیل غلبه گونه‌های نور پسند مانند بلوط در این است. از آنجا بلوط که گونه‌ای نور پسند است (بر خلاف راش که تحمل سایه دارد)، تنوع گونه‌ای بیشتری نسبت به جنگل‌هایی دارد که راش در آنها غالب است علیرغم اینکه در استان گیلان انتظار داشتن شاخص تنوع بالا است ولی این شاخص نسبت به دو استان دیگر از وضعیت نامناسب تری برخوردار است

دقت بیشتر)، برتری مدل RF را می‌توان به ماهیت یادگیری ماشین نسبت داد که می‌تواند به راحتی اطلاعات متغیرهای محیطی را در مدل‌ها گنجانده و دانش را مستقیماً از داده‌ها استخراج کند بدون اینکه فرضیات از پیش تعریف شده پدیده مورد بررسی قرار گیرد تا کیفیت در مدل‌ها به شکل قابل توجهی بهبود یابد. (Bayat et al. 2019) در حقیقت، با پیدایش و گسترش روش‌های یادگیری ماشین، در سال‌های اخیر دقت حل مسائل پیچیده در بسیاری از زمینه‌ها افزایش یافته است. روش‌های یادگیری ماشین به طور گسترده‌ای برای تخمین و پیش‌بینی بسیاری از ویژگی‌های جنگل استفاده شده که در این مطالعه، از دو الگوریتم یادگیری ماشین و همچنین تجزیه و تحلیل آماری برای مدل‌سازی تنوع زیست درختان در جنگل‌های هیرکانی استفاده شد. نتایج نشان داد که روش‌های یادگیری ماشین، به ویژه جنگل تصادفی، دقیق‌تر از سایر روش‌ها هستند. بر اساس نتایج بدست آمده از مدل RF، ارتفاع و BAL به ترتیب به عنوان مهمترین عوامل تعیین‌کننده تغییرات تنوع‌زیستی درختی شناخته شدند. به طور کلی، اگرچه تکنیک‌های یادگیری ماشین اغلب می‌توانند برتر از روش‌های سنتی و رگرسیونی باشند؛ چرا که پیش‌فرض‌های کمتری در مورد داده‌ها دارند (Westreich et al., 2010).

### تشکر و قدردانی

این مقاله حاصل از طرح پژوهشی با عنوان پایش، ارزیابی غنای گونه‌ای درختی و آرایه مدل‌های مکانی تنوع زیستی جنگل‌های شمال کشور و ارسباران در سطح پلات‌های ثابت، مصوب در موسسه تحقیقات جنگل‌ها و مراتع کشور است و بدینوسیله نویسندگان مقاله، از حمایت‌های انجام شده در انجام این طرح، توسط موسسه مذکور، کمال تشکر و قدردانی را دارد.

### یادداشت‌ها

1. Machine Learning
2. Basal area (BA)
3. Basal Area in Largest tree (BAL)
4. Support Vector Machine (SVM)

که در مطالعه حاضر می‌تواند یکی از دلایل سطح بالای این شاخص در گلستان می‌باشد. به طور کلی، نتایج مطالعات مشابه نشان می‌دهد که در جهات شمالی درجه حرارت پایین‌تر و ذوب شدن یخ‌ها آهسته‌تر است، به ویژه در ارتفاعات بالا، باعث تنوع گونه‌ای کمتر در این جهت‌ها می‌شود. این دامنه‌ها از رطوبت و سرمای بیشتری برخوردار بوده و باعث غلبه گونه راش و در نتیجه کاهش شاخص یکنواختی می‌شوند. اگرچه، ذکر این نکته مهم است که در مناطق مختلف، به دلیل ویژگی‌های آب و هوایی و زمین‌شناسی و موقعیت جغرافیایی، نتایج مختلفی در مورد جهت‌های جغرافیایی مختلف به دست می‌آید، اما معمولاً تنوع در جهت‌هایی با رطوبت و دما بالاتر است. عامل مهم دیگری که بر شاخص تنوع در کنار عوامل طبیعی تأثیر می‌گذارد، عوامل مدیریتی و حفاظتی در جنگل‌ها است. در حقیقت، تفاوت زیاد در شاخص تنوع زیستی بین مناطق جنگلی با شرایط ارتفاعی، محیطی و آب و هوایی، فیزیوگرافی و بیولوژیکی مشابه، می‌تواند ناشی از شیوه مدیریت و شدت بهره‌برداری در این جنگل‌ها نسبت داده شود. بنابراین، به طور کلی، شاخص تنوع شانون در منطقه جنگلی حفاظت شده در مقایسه با مناطق دیگر به طور قابل توجهی بالاتر است، که نشان می‌دهد قطع درختان و فشار انسان در منطقه، منجر به کاهش شدید شاخص‌های تنوع‌زیستی می‌باشد. نتایج نشان داد توانایی مدل‌ها و روش‌های یادگیری ماشین برای برآورد دقیق‌تر شاخص تنوع‌زیستی در مناطق جنگلی و شناسایی متغیرهای مهم تأثیرگذار بر آن به عنوان مثال، ارتفاع و BAL از مدل رگرسیونی بالاتر است (Bayat et al., 2021b). اگرچه نمی‌توان گفت که همیشه مدل‌های ML از مدل‌های رگرسیونی برتر هستند، اما نتایج ما در این تحقیق نشان داد که مدل‌های ماشین یادگیری ضریب تعیین بالاتر و RMSE پایین‌تر، نسبت به مدل‌های رگرسیون از دقت بیشتری در نشان دادن تغییرات تنوع زیستی برخوردار هستند. این نتایج، برتری روش RF نسبت به رگرسیون را نشان داد (از جهت

5. RBF  
6. Random Forest model
7. Multiple Linear Regression (MLR)

## فهرست منابع

- Abedi, T., Hoseini, S.A., & Naghdi, R. (2020). Study of relationship between soil mechanical characteristic and landslide in forest road route (Case Study: Chafroud Watershed Guilan Prov.). *Journal of Watershed Management, 1*, 17-29 (In Persian).
- Bayat, M., Pukkala, T., Namiranian, M., & Zobeiri, M. (2013). Productivity and optimal management of the uneven-aged hardwood forests of Hyrcania. *European Journal of Forest Research, 132*(5-6), 851-864.
- Bayat, M., Noi, P.T., Zare, R., & Bui, D.T. (2019). A semi-empirical approach based on genetic programming for the study of biophysical controls on diameter-growth of *Fagus orientalis* in northern Iran. *Remote Sensing, 11*(14), 1-19.
- 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), 1-18.
- Bayat, M., Bettinger, P., Heidari, S., Hassani, M., 2021a, Ten-year estimation of *Fagus orientalis* Lipsky increment in natural forests: A Comparison of artificial neural networks model, multiple linear regression and actual increment. *2021 Forestry: An International Journal of Forest Research 94*(4), pp. 598-609
- Bayat, M., Burkhart, H., Namiranian, M., Hamidi, S.K., Heidari, S. and Hassani, M., 2021b. Assessing Biotic and Abiotic Effects on Biodiversity Index Using Machine Learning. *Forests* 2021, 12, 461.
- Bourque, C.P.A., & Bayat, M. (2015). Landscape variation in tree species richness in northern Iran forests. *PLoS ONE, 10*(4), 1-17.
- Bourque, C.P.A., Bayat, M., & Zhang, C. (2019). An assessment of height-diameter growth variation in an unmanaged *Fagus orientalis*-dominated forest. *European Journal of Forest Research, 138*(4), 607-621.
- Burkhart, H.E., & Tomé, M. (2012). Modeling forest trees and stands. *Springer Science & Business Media, 9*, 1-18.
- Chen, J., Yang, H., Man, R., Wang, W., Sharma, M., Peng, C., Parton, J., Zhu, H. & Deng, Z., (2020). Using machine learning to synthesize spatiotemporal data for modelling DBH-height and DBH-height-age relationships in boreal forests. *Forest Ecology and Management, 466*, 1-15.
- Chu C., Lutz J.A., Král K., Vrška T., Yin X., Myers J.A., et al.. (2019). Direct and indirect effects of climate on richness drive the latitudinal diversity gradient in forest trees. *Ecology Letter, 22* (2), 245 - 255.
- Cui, W., & Zheng, X.-X. (2016). Spatial heterogeneity in tree diversity and forest structure of evergreen broadleaf forests in southern China along an altitudinal gradient. *Forests, 7*, 216.
- Eshaghi, M., & Shataee joybari, S. (2016). Preparation map of Forest Fire Risk Using SVM, RF & MLP Algorithms (Case Study: Golestan National Park, Northeastern Iran), *Journal of Wood and Forest Science and Technology, 23*(4), 1333-154. (In Persian).
- Jafarian, Z., & Kargar, M. (2017). Comparison of Random Forest (RF) and Boosting Regression Tree (BRT) For Prediction of Dominant Plant Species Presence in Polour Rangelands, Mazandaran Province. *Iran. J. Appl. Ecol, 6* (1), 41-55 (In Persian).
- Júnior, I.D.S.T., Torres, C.M.M.E., Leite, H.G., de Castro, N.L.M., Soares, C.P.B., Castro, R.V.O. & Farias, A.A. (2020). Machine learning: Modeling increment in diameter of individual trees on Atlantic forest fragments. *Ecological Indicators, 117*, 25-35



- Hamidi, S.K., de Luis, M., Bourque, C.P.A. Bayat, M et al. Projected biodiversity in the Hyrcanian Mountain Forest of Iran: an investigation based on two climate scenarios. *Biodivers Conserv* (2022). <https://doi.org/10.1007/s10531-022-02470->
- Kebede, M., Yirdaw, E., Luukkanen, O., & Lemenih, M. (2013). Plant community analysis and effect of environmental factors on the diversity of woody species in the moist Afromontane forest of Wondo Genet, South Central Ethiopia. *Biodiversity. Res Conser*, 29, 63.
- Lee Sunmin Kim, J. C., Jung, H. S., Lee, M. J., & Lee, S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomat. Nat. Hazard. Risk*, 8, 1185–1203.
- Liu, Z., Peng, C., Work, T., Candau, J.N., DesRochers, A., & Kneeshaw, D. (2018). Application of machine-learning methods in forest ecology: recent progress and future challenges. *Environmental Reviews*, 26(4), 339-350.
- Monarrez-Gonzalez, J.C., Gonzalez-Elizondo, M.S., Marquez-Linares, M.A., Gutierrez-Yurrita, P.J. & Perez-Verdin, G., (2020). Effect of forest management on tree diversity in temperate ecosystem forests in northern Mexico. *Plos one*, 15(5), 38-50.
- Mountrakis, G., Im, J., & Ogole, C. (2011), Support Vector Machines in Remote Sensing: A Review, *ISPRS. Journal of Photogrammetry and Remote Sensing*, 66, 247-259.
- Oyebade, B.A., Eguakun, F.S., & Duru, B.N. (2020). Tree basal area models and density for selected plantation species in swamp forest zone of Rivers State, Nigeria. *World News of Natural Sciences*, 30(3), 281-296.
- Roslin T., Hardwick B., Novotny V., Petry W.K., Andrew N.R., et al.. (2017). Higher predation risk for insect prey at low latitudes and elevations. *Science*, 356(6339), 742 -744.
- Sharma, M., & Parton, J. (2007). Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology and Management*, 249(3), pp.187-198.
- Shen, W., Wang, J., Ali, A. & Li, M. (2020). Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests.
- Šimová, I., Storch, D., Keil, P., Boyle, B., Phillips, O.L., & Enquist, B.J. (2011). Global species–energy 652 relationship in forest plots: role of abundance, temperature and species climatic 653 tolerances. *Glob. Ecol. Biogeogr*, 20, 842–856.
- Tiwari, O.P., Sharma, C.M., & Rana, Y.S. (2020). Influence of elevation and slope-aspect on diversity, regeneration and structure of some moist temperate forests of Garhwal Himalaya. *Tropical Ecology*, 61(2).278-289.
- Vafaei, S., Soosani, J., Adeli, K., Fadaei, H., Naghavi, H., Pham, T., & Bui, D. (2018). Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). *Remote Sens*, 10 (172), 1-21.
- Westreich, D., Lessler, J., Funk, M.J. (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J. Clin. Epidemiol*, 63, 826–833. <https://doi.org/10.1016/j.jclinepi.2009.11.020>.
- Woldu, G., Solomon, N., Hishe, H., Gebrewahid, H., Gebremedhin, M.A., & Birhane, E. (2020). Topographic variables to determine the diversity of woody species in the enclosure of Northern Ethiopia. *Heliyon*, 6 (1), p.e03121.
- Wood, S. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of Royal Statistical Society*, <https://doi.org/10.1111/j.1467-9868.2007.00646.x>

- Yadegarnejad, S.A., Sagheb-Talebi, Kh., Heidari, H., Moayeri, M.H., & Mortazavi, S.M. (2017). Evaluation of application of the 15d method in Loveh Forest, Golestan Province. *Iranian Journal of Forest and Poplar Research*, 25 (1): 160-171 (In Persian).
- Yao, L., Ding, Y., Xu, H., Deng, F., Yao, L., Ai, X., & Zang, R. (2020). Patterns of diversity change for forest vegetation across different climatic regions-A compound habitat gradient analysis approach. *Global Ecology and Conservation*, p.e011106.
- Zarandian, A., Baral, H., Yavari, A.R., Jafari, H.R., Stork, N.E., Ling, M.A., & Amirnejad, H. (2016). Anthropogenic decline of ecosystem services threatens the integrity of the unique Hyrcanian (Caspian) forests in Northern Iran. *Forests*, 7 (3), p.51.
- Zhang J., Li M., & Nie E. (2014). Pattern of functional diversity along an altitudinal gradient in the Baihua Mountain Reserve of Beijing, China. *Brazilian Journal of Botany*, 1, 37-45.