

## Modeling Daily and Monthly Rainfall in Tabriz using Ensemble Learning Models and Decision Tree Regression

Mohammad Taghi Sattari<sup>1\*</sup>, Raziieh Bagheri<sup>2</sup>, Kimia Shirini<sup>3</sup>, Pouya Allahverdipour<sup>4</sup>

<sup>1</sup> Associate Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, Email: mtsattar@gmail.com

<sup>2</sup> M.Sc student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, Email: bagheri.gnmrm@gmail.com

<sup>3</sup> Ph.D Student, Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, Email: k.shirini@tabrizu.ac.ir

<sup>4</sup> Ph.D Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, Email: a.pouya@tabrizu.ac.com

### Article Info

**Article type:**  
Research Full Paper

**Article history:**  
Received: 2023-12-31  
Accepted: 2024-4-12

**Keywords:**  
Artificial Neural Networks  
Ensemble Learning  
Extra Trees  
Precipitation  
Random Forest

### ABSTRACT

Precipitation is one of the most important components of hydrology and meteorology, and the prediction of its values are important in various fields, such as agriculture and the environment. Considering that the occurrence of precipitation and its amount depend on many factors, the modeling of precipitation has many complications. In this research, the meteorological data of Tabriz synoptic station including minimum, maximum and average temperature, relative humidity, air pressure, maximum wind speed and precipitation in the period of 1986-2020 were used. The machine learning and ensemble learning methods including Multi Layer Perceptron (MLP), Random Forest (RF), Ada Boost (AB), Gradient Boost (GB), Extra Trees (ET) and Decision Tree Regression (DTR) models were used for rainfall modeling. 70% of the data was used for training and 30% for testing the models. The statistical criteria of Coefficient of Correlation (R), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Kling-Gupta Efficiency (KGE) were used to evaluate the models. According to the results, on daily scale the MLP model with  $R=0.993$ ,  $RMSE=0.184$  mm,  $MAE=0.184$  mm and  $KGE=0.82$ , and the ET model with  $R=0.986$ ,  $RMSE=0.324$  mm,  $MAE=0.324$  mm and  $KGE=0.75$ , respectively, and on monthly scale the MLP model with  $R=0.999$ ,  $RMSE=0.153$  mm,  $MAE=0.222$  mm  $KGE=0.88$ , and the ET model with  $R=0.981$ ,  $RMSE=0.266$  mm,  $MAE=0.197$  mm and  $KGE=0.71$ , respectively, have the highest accuracy. Overall, the results show that machine learning and ensemble learning models perform well in predicting daily and monthly rainfall.

**Cite this article:** Sattari, M.T., Bagheri, R., Shirini, K., Allahverdipour, P. (2024). Modeling Daily and Monthly Rainfall in Tabriz using Ensemble Learning Models and Decision Tree Regression. Journal of the Climate Change Research, 5 (18), 31-48.





## مدل‌سازی بارش روزانه و ماهانه تبریز با استفاده از مدل‌های یادگیری جمعی و رگرسیون درخت تصمیم

محمدتقی ستاری<sup>۱\*</sup>، راضیه باقری<sup>۲</sup>، کیمیا شیرینی<sup>۳</sup>، پویا الله‌وردی‌پور<sup>۴</sup>

<sup>۱</sup> دانشیار، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران، رایانامه: mtsattar@gmail.com

<sup>۲</sup> دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران، رایانامه: bagheri.gnmrm@gmail.com

<sup>۳</sup> دانشجوی دکتری، گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، تبریز، ایران، رایانامه: k.shirini@tabrizu.ac.ir

<sup>۴</sup> دانشجوی دکتری، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران، رایانامه: a.pouya@tabrizu.ac.com

### چکیده

### اطلاعات مقاله

بارش یکی از مهم‌ترین مولفه‌های هیدرولوژی و هواشناسی است که پیش‌بینی مقادیر آن در زمینه‌های مختلفی همچون کشاورزی، شرب، صنعت، بهداشت، اقلیم و محیط‌زیست از اهمیت ویژه‌ای برخوردار است. با توجه به اینکه وقوع، نوع و میزان بارش به عوامل بسیار زیادی بستگی دارد، مدل‌سازی و پیش‌بینی مقدار بارش همواره دارای پیچیدگی‌ها و چالش‌های فراوان بوده است. در این پژوهش از داده‌های پارامترهای هواشناسی ایستگاه همدیدی تبریز شامل دماهای کمینه، بیشینه و میانگین، رطوبت نسبی میانگین، فشار هوای میانگین، سرعت باد بیشینه و بارش در مقیاس‌های زمانی روزانه و ماهانه در یک دوره ۳۵ ساله (۱۹۸۶-۲۰۲۰) استفاده شد. از روش‌های یادگیری ماشین و یادگیری جمعی شامل شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، جنگل تصادفی (RF)، آدابوست (AB)، تقویت گرادیان (GB)، درختان اضافی (ET) و مدل رگرسیون درخت تصمیم (DTR) برای مدل‌سازی بارش استفاده شد. بدین منظور ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد از داده‌ها نیز برای آزمون مدل‌ها در نظر گرفته شدند. برای ارزیابی مدل‌های مورد استفاده، از معیارهای آماری ضریب همبستگی (R)، جذر میانگین مربعات خطاها (RMSE)، میانگین خطای مطلق (MAE) و کلینگ گوپتا (KGE) استفاده شد. مطابق نتایج این تحقیق در مقیاس روزانه مدل MLP با  $R=0/993$ ،  $RMSE=0/184$  میلی‌متر،  $MAE=0/184$  میلی‌متر و  $KGE=0/82$  و مدل ET با  $R=0/986$ ،  $RMSE=0/324$  میلی‌متر،  $MAE=0/324$  میلی‌متر و  $KGE=0/75$  و در مقیاس ماهانه به ترتیب مدل MLP با  $R=0/999$ ،  $RMSE=0/153$  میلی‌متر،  $MAE=0/222$  میلی‌متر و  $KGE=0/88$  و مدل ET با  $R=0/981$ ،  $RMSE=0/266$  میلی‌متر،  $MAE=0/197$  میلی‌متر و  $KGE=0/71$  از بیشترین دقت و کمترین خطا برخوردارند. به‌طور کلی نتایج این پژوهش نشان‌دهنده عملکرد بسیار مناسب مدل‌های یادگیری ماشین و یادگیری جمعی در مدل‌سازی بارش در هر دو مقیاس زمانی روزانه و ماهانه است.

نوع مقاله:  
مقاله کامل علمی

تاریخ دریافت: ۱۴۰۲/۱۰/۱۰

تاریخ پذیرش: ۱۴۰۳/۰۱/۲۴

### واژه‌های کلیدی:

بارش  
جنگل تصادفی  
درختان اضافی  
یادگیری جمعی  
شبکه‌های عصبی مصنوعی

**استناد:** ستاری، محمدتقی؛ باقری، راضیه؛ شیرینی، کیمیا؛ الله‌وردی‌پور، پویا. (۱۴۰۳). مدل‌سازی بارش روزانه و ماهانه تبریز با استفاده از مدل‌های یادگیری جمعی و رگرسیون درخت تصمیم. نشریه پژوهش‌های تغییرات آب و هوایی، ۵ (۱۸)، ۴۸-۳۱.



## مقدمه

بارش یکی از مهم‌ترین مولفه‌های هیدرولوژیکی و هواشناسی بوده و پیش‌بینی آن برای اهداف مختلفی مانند مدیریت سیلاب، خشکسالی، رواناب، رسوب، مدیریت حوضه آبریز و فعالیت‌های کشاورزی اهمیت فراوانی دارد. امروزه روش‌های یادگیری ماشین<sup>۱</sup> و سیستم‌های هوشمند در حوزه‌های مختلف به کمک انسان آمده است که با توجه به مشاهدات و داده‌های ورودی، آموزش دیده و داده‌های جدید را پیش‌بینی می‌کنند. یکی از مهم‌ترین ویژگی‌های الگوریتم‌های یادگیری ماشین (که به اصطلاح روش‌های جعبه سیاه<sup>۲</sup> نامیده می‌شوند) برخلاف روش‌های کلاسیک، عدم وابستگی به نوع داده‌های ورودی و ماهیت آنها است. از طرفی انتخاب داده‌ها در این نوع مدل‌ها به صورت تصادفی بوده و تأثیر هرگونه عامل نایستایی در سری داده‌ها از بین می‌رود. این امر مزیت و قابلیت مهم این مدل‌ها محسوب می‌شود. انواع مدل‌های درختی و الگوریتم درخت تصمیم<sup>۳</sup> (DT)، ماشین بردار پشتیبان<sup>۴</sup> (SVM)، شبکه‌های عصبی مصنوعی<sup>۵</sup> (ANNs) و روش‌های یادگیری جمعی<sup>۶</sup> در زمینه علوم مهندسی بسیار مهم و کارآمد هستند. با توجه به اینکه بارش به عوامل بسیار زیادی مانند دما، رطوبت، فشار هوا و غیره بستگی دارد و از روابط خطی تبعیت نمی‌کند، پیش‌بینی بارش همواره با پیچیدگی‌ها و چالش‌های فراوان همراه بوده است. اما امروزه محققان با استفاده از روش‌های یادگیری ماشین می‌توانند بارش را با دقت بالایی مدل‌سازی و پیش‌بینی کنند.

خلیلی و همکاران (۱۳۸۸) با استفاده از مدل شبکه‌های عصبی از نوع پرسپترون چندلایه<sup>۷</sup> (MLP)، میزان بارش ایستگاه سینوپتیک مشهد را با دقت قابل قبولی مدل‌سازی کردند. قلی‌زاده و دارند (۱۳۸۹)

میزان بارش ماهانه تهران را با استفاده از مدل شبکه عصبی مصنوعی مدل‌سازی کردند. حلبیان و دارند (۱۳۹۱) با استفاده از داده‌های بارش ماهانه ایستگاه همدید اصفهان در بازه آماری ۲۰۰۹-۱۹۵۱ به مدت ۵۹ سال و مدل ANN بارش ماهانه را پیش‌بینی کرده و نتیجه گرفتند که این مدل به‌خوبی رابطه غیرخطی بین مقادیر ماهانه بارش را پیش‌بینی می‌کند. ستاری و نهرین (۱۳۹۲) پارامترهای هواشناسی ایستگاه‌های اهر و جلفا در استان آذربایجان شرقی را به‌عنوان ورودی مدل‌های هوشمند شامل ANN، برنامه‌ریزی ژنتیک<sup>۸</sup> (GP) و مدل درختی M5 در نظر گرفته و نتیجه گرفتند که به‌دلیل ارائه روابط خطی ساده و قابل فهم توسط مدل درختی M5، این روش به‌عنوان روشی کاربردی و جایگزین برای محاسبه حداکثر بارش روزانه می‌تواند مورد استفاده قرار گیرد. امیدوار و نبوی‌زاده (۱۳۹۳) طی دوره مشترک آماری ۲۳ ساله (۲۰۱۲-۱۹۸۹) با استفاده از مدل ANN بارش روزانه ایستگاه‌های کرمان، بافت و میانه جیرفت را پیش‌بینی کردند. مظفری و همکاران (۱۳۹۴) از داده‌های پنج پارامتر اقلیمی شامل میانگین بارش ماهانه، میانگین رطوبت نسبی، میانگین حداکثر دما، میانگین حداقل دما و میانگین سرعت باد در دوره آماری ۴۰ ساله (۲۰۱۰-۱۹۷۰) استفاده کرده و با استفاده از مدل ترکیبی شبکه عصبی و موجک<sup>۹</sup> به پیش‌بینی بارش ماهانه ایستگاه سینوپتیک کرمانشاه پرداختند و نتیجه گرفتند که روش موجک-شبکه عصبی نسبت به روش شبکه عصبی ساده دقت بیشتری دارد. عبدالله پورآزاد و همکاران (۱۳۹۵) با استفاده از قوانین درختی M5 و مدل شبکه‌های عصبی مصنوعی المانی<sup>۱۰</sup> (ENN) به پیش‌بینی جریان روزانه رودخانه اهرچای پرداخته و نتیجه گرفتند که مدل ENN دقت بیشتری نسبت به روش M5 دارد ولی روش M5 با توجه به ارائه قوانین کارآمد و ساده اگر-آنگاه و روابط خطی ساده برای پیش‌بینی جریان و نیز

<sup>1</sup> Machine Learning

<sup>2</sup> Black Box

<sup>3</sup> Decision Tree

<sup>4</sup> Support Vector Machines

<sup>5</sup> Artificial Neural Networks

<sup>6</sup> Ensemble Learning

<sup>7</sup> Multi-Layer Perceptron

<sup>8</sup> Genetic Programming

<sup>9</sup> Wavelet

<sup>10</sup> Elman Neural Network

زیادی برخوردار است، به خصوص اینکه در مقیاس‌های زمانی روزانه و ماهانه انجام شود. بنابراین هدف از این پژوهش مدل‌سازی و پیش‌بینی بارش روزانه و ماهانه ایستگاه همدیدی تبریز با استفاده از مدل‌های هوشمند شامل DT، MLP، و روش‌های یادگیری جمعی شامل جنگل تصادفی<sup>۲</sup> (RT)، تقویت گرادیان<sup>۳</sup> (GB)، آدابوست<sup>۴</sup> (AB)، درختان اضافی<sup>۵</sup> (ET) و مقایسه عملکرد آن‌ها و تعیین دقیق‌ترین مدل در پیش‌بینی بارش روزانه و ماهانه است.

### مواد و روش‌ها

شکل ۱ فلوجارت مراحل انجام این تحقیق را نشان می‌دهد.

#### منطقه مورد مطالعه و داده‌های مورد استفاده:

تبریز مرکز استان آذربایجان شرقی با مساحتی حدود ۱۲۰۰ کیلومتر مربع، در موقعیت ۴۶ درجه و ۲۴ دقیقه طول شرقی و ۳۲ درجه و ۱۲ دقیقه عرض شمالی و در حوضه آبریز دریاچه ارومیه واقع شده است و ارتفاع آن از سطح دریا حدود ۱۳۶۱ متر است. تبریز از جهت شمال به کوه عینالی، از جهت جنوب و شرق به دامنه‌های کوه سهند و از جهت غرب به جلگه تبریز و دریاچه ارومیه محدود شده است. مطابق با اقلیم‌نمای دمارتن، دشت تبریز با دمای متوسط سالانه ۱۳/۶۱ درجه سانتی‌گراد و میانگین بارش سالانه ۲۵۵/۵۵ میلی‌متر دارای اقلیم نیمه‌خشک است. در این پژوهش از داده‌های هواشناسی ایستگاه همدیدی تبریز در بازه ۳۵ ساله (۱۹۸۶-۲۰۲۰) استفاده شده است. شکل ۲ موقعیت ایستگاه همدیدی تبریز را نشان می‌دهد.

تعداد پارامتر ورودی مورد نیاز کمتر، می‌تواند به‌عنوان یک روش جایگزین مناسب به کار گرفته شود. مهتابی و همکاران (۱۳۹۷) بارش روزانه ایستگاه همدیدی اصفهان را با استفاده از مدل‌های هوشمند شامل SVM، k-نزدیک‌ترین همسایگی<sup>۱</sup> (KNN)، ANN و DT و با استفاده از داده‌های مختلف هواشناسی مدل‌سازی کردند. بر اساس نتایج این پژوهش، بهترین نتایج در بخش صحت‌سنجی به ترتیب مربوط به سناریوی یک روز قبل (با پارامتر حداقل رطوبت نسبی) و سناریوی چهار روز قبل (با پارامتر دمای حداکثر) بود. شهیدی و همکاران (۱۳۹۸) به پیش‌بینی بارش ماهانه ایستگاه سینوپتیک بیرجند با به‌کارگیری الگوریتم درخت M5 و نزدیک‌ترین همسایگی (KNN) پرداخته و نتیجه گرفتند که در تمامی سناریوهای تعریف شده، مدل درختی M5 نسبت به مدل KNN توانایی بیشتری در پیش‌بینی بارش ماهانه این ایستگاه دارد. شرفی و همکاران (۱۳۹۹) از دو مدل هوشمند GP و SVM برای پیش‌بینی بارش ماهانه استان اردبیل استفاده کردند و نتایج نشان داد که عملکرد هر دو مدل خوب و تقریباً یکسان بوده ولی با توجه به ارزیابی‌های انجام شده مدل رگرسیون بردار پشتیبان با ضریب همبستگی برابر با ۰/۹۹۹ عملکرد بهتری دارد. سینگ و همکاران برای تخمین میانگین سیلاب سالانه در حوضه‌های آبریز هند از مدل درختی M5 استفاده کردند (Singh et al., 2010). مکینک و همکاران با پیش‌بینی بارش فصل بهار در حوضه ویکتوریای استرالیا با استفاده از ANN و رگرسیون چندگانه نشان دادند که مدل ANN، عملکرد بهتر و دقت بیشتری در پیش‌بینی بارش دارد (Mekanik et al., 2013).

بررسی پیشینه تحقیق نشان می‌دهد که استفاده از روش‌های نوین با هدف مدل‌سازی و پیش‌بینی پارامترهای مختلف و به‌ویژه بارش از اهمیت بسیار

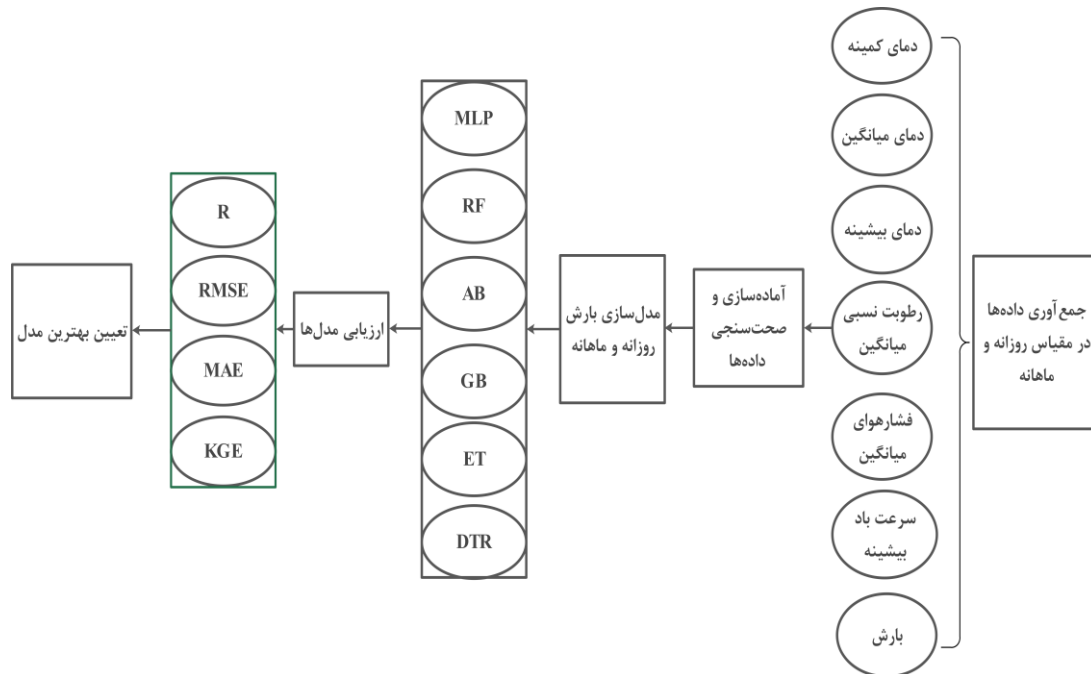
<sup>2</sup> Random Forest

<sup>3</sup> Gradient Boost

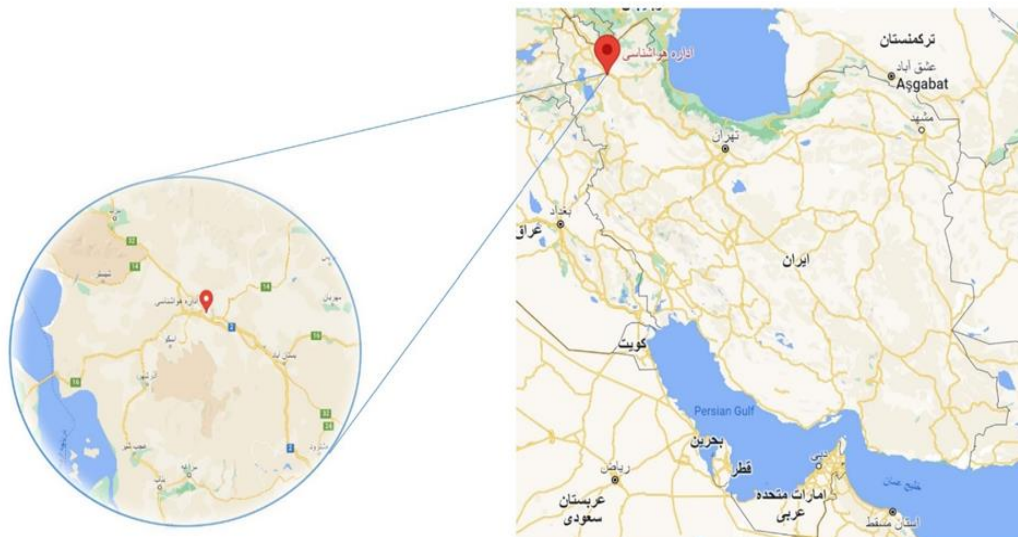
<sup>4</sup> Ada Boost

<sup>5</sup> Extra Trees

<sup>1</sup> K-Nearest Neighbors



شکل ۱: فلوجارت مراحل انجام تحقیق




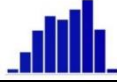
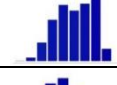
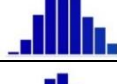
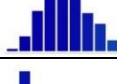
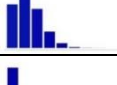

شکل ۲: موقعیت ایستگاه همدیدی تبریز

مطالعه‌ی هیدرولوژی و منابع آب که در آن از داده‌های هواشناسی و هیدرولوژی استفاده می‌شود، صحت‌سنجی داده‌ها است. واضح است که بدون اطمینان از صحت و کیفیت داده‌ها، نمی‌توان از آن‌ها برای استخراج نتایج بعدی استفاده کرد (الله‌پوردی پور و همکاران، ۱۴۰۲). در این پژوهش داده‌های مورد


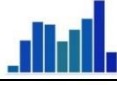
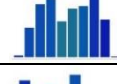

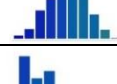
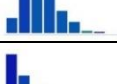
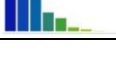
داده‌های هواشناسی مورد استفاده در این پژوهش شامل پارامترهای دمای کمینه ( $^{\circ}\text{C}$ )، دمای بیشینه ( $^{\circ}\text{C}$ )، دمای میانگین ( $^{\circ}\text{C}$ )، رطوبت نسبی میانگین ( $\%$ )، فشار هوای میانگین (hPa) و سرعت باد بیشینه (m/s) به‌عنوان پارامترهای پیش‌بینی‌کننده و میزان بارش (mm) به‌عنوان پارامتر وابسته است. لازم‌هی هر

استفاده قبل از انجام فرآیند مدل‌سازی مورد بررسی آماری و صحت‌سنجی قرار گرفت و همچنین نسبت به تکمیل داده‌های گمشده و حذف داده‌های پرت اقدام شد.

جدول ۱: مشخصات آماری داده‌های بارش در مقیاس روزانه

گمشده	میان	میانگین	بیشینه	کمینه	توزیع	پارامتر
۶۳۸ (۵٪)	۶/۶	۵/۸۵	۲۲	-۳۲/۲		دمای کمینه (°C)
۳۲۷ (۳٪)	۱۸/۲	۱۷/۲۴	۴۰	-۹/۸		دمای بیشینه (°C)
۰ (۰٪)	۱۱/۵	۱۱/۰۳	۲۹/۶	-۱۸/۷		دمای میانگین (°C)
۰ (۰٪)	۱۰۱۳/۸۷	۱۰۱۴/۶۸	۱۰۴۴/۲۴	۹۹۳/۶۲		فشار هوای میانگین (hPa)
۱ (۰٪)	۵۸	۵۹/۰۷	۹۸/۷۵	۱۶/۲۵		رطوبت نسبی میانگین (/)
۰ (۰٪)	۲/۷۵	۳/۰۱	۱۵/۶۲۵	۰		سرعت باد بیشینه (m/s)
۰ (۰٪)	۰	۰/۷۹	۳۴	۰		بارش (mm)

جدول ۲: مشخصات آماری داده‌های بارش در مقیاس ماهانه

گمشده	میان	میانگین	بیشینه	کمینه	توزیع	پارامتر
۰ (۰٪)	۵/۸۱	۵/۴۶	۱۸/۱۳	-۱۴/۲۱		دمای کمینه (°C)
۰ (۰٪)	۱۸/۱۴	۱۷/۰۴	۳۳/۴۲	-۳/۵۹		دمای بیشینه (°C)
۰ (۰٪)	۱۱/۴۳	۱۰/۹۹	۲۵/۸۴	-۸/۶۳		دمای میانگین (°C)
۰ (۰٪)	۱۰۱۴/۶۶	۱۰۱۴/۷۰	۱۰۲۹/۴۵	۱۰۰۱/۹۶		فشار هوای میانگین (hPa)
۰ (۰٪)	۵۹/۵۷	۵۹/۰۹	۷۸/۳۷	۳۵/۱۴		رطوبت نسبی میانگین (/)
۰ (۰٪)	۶/۲۵	۶/۵۳	۱۵/۶۳	۲/۶۳		سرعت باد بیشینه (m/s)
۰ (۰٪)	۲۰/۰۱	۲۳/۹۴	۱۱۹/۵۹	۰		بارش (mm)

این متغیرها به دو یا چند دسته تقسیم کرد. مشخصه‌هایی که به درجه خلوص گره‌های فرزند حاصل (یعنی تناسب با شرایط هدف) مربوط می‌شوند، برای انتخاب متغیرهای ورودی بالقوه مختلف استفاده می‌شوند. هنگام ساخت یک درخت تصمیم باید قوانین توقف اعمال شود تا از پیچیده شدن بیش از حد مدل جلوگیری شود. در برخی شرایط، قوانین توقف به خوبی کار نمی‌کنند که در این حالت یک راه جایگزین برای ساخت یک مدل درخت تصمیم این است که ابتدا یک درخت بزرگ را رشد دهیم و سپس با حذف گره‌هایی که اطلاعات اضافی بی‌نیازی را ارائه می‌دهند، آن را به اندازه مطلوب هرس کنیم (Song and Ying, 2015). در روش درخت تصمیم، تقسیم‌بندی گره‌ها بر اساس بهره اطلاعاتی<sup>۳</sup> (IG) محاسبه می‌شود. برای انتخاب متغیر شکننده و تقسیم‌شونده در فرآیند رشد درخت استفاده می‌شود. همگنی نمونه‌ها در یک گره با شاخصی به نام آنتروپی<sup>۴</sup> تعیین می‌شود. برای محاسبه IG ابتدا باید شاخص آنتروپی محاسبه شود. اگر متغیر هدف دارای C مقدار متفاوت باشد، آنگاه آنتروپی S وابسته به کلاس C طبق رابطه ۱ به دست می‌آید (Han et al., 2012).

$$\text{Entropy (S)} = - \sum_{i=1}^C p_i \log_2 p_i \quad \text{رابطه (۱)}$$

در رابطه ۱ پارامتر  $p_i$  به صورت نسبتی از S تعریف می‌شود که متعلق به کلاس i است. IG در واقع نشان‌دهنده میزان کاهش در آنتروپی است. خلوص داده‌ها در یک گزینه توسط آنتروپی ارائه شده و تاثیر یک متغیر در کلاس‌بندی با استفاده از IG تعیین می‌شود. مقدار IG مربوط به متغیر A طبق رابطه ۲ محاسبه می‌شود (Han et al., 2012).

از ۷۰ درصد داده‌ها به عنوان داده‌های آموزش<sup>۱</sup> و از ۳۰ درصد داده‌ها به عنوان داده‌های صحت‌سنجی<sup>۲</sup> استفاده شد. در مقیاس روزانه از داده‌های مربوط به یک روز قبل و در مقیاس ماهانه از داده‌های مربوط به یک ماه قبل برای پیش‌بینی میزان بارش استفاده شد. مشخصات آماری داده‌های مورد استفاده در مقیاس‌های زمانی روزانه و ماهانه به ترتیب در جدول‌های ۱ و ۲ ارائه شده است.

### شبکه‌های عصبی مصنوعی پرسپترون‌های

**چندلایه (MLP):** شبکه‌های عصبی مصنوعی به عنوان یکی از تکنیک‌های اصلی سیستم هوش مصنوعی بر اساس سیستم عصبی و مغز انسان توسعه داده شده است. در این تکنیک دینامیک با پردازش داده‌های مشاهداتی، ساختمان روابط و وابستگی حاکم بر پدیده مورد نظر، یادگیری و تعیین می‌شود. سپس مدل آموزش‌یافته برای پیش‌بینی رخداد‌های آتی این پدیده مورد استفاده قرار می‌گیرد (Jain et al., 1999). به طور معمول این شبکه‌ها شامل مجموعه‌ای از واحدهای حسی (نورون‌های پایه) می‌باشند که تشکیل‌دهنده لایه‌های ورودی، یک یا چند لایه پنهان و یک لایه خروجی می‌باشند. سیگنال ورودی در خلال شبکه و در مسیری رو به جلو به صورت لایه به لایه منتشر می‌شود. این نوع شبکه با عنوان پرسپترون چندلایه (MLP) نامیده می‌شود (Haykin, 1999).

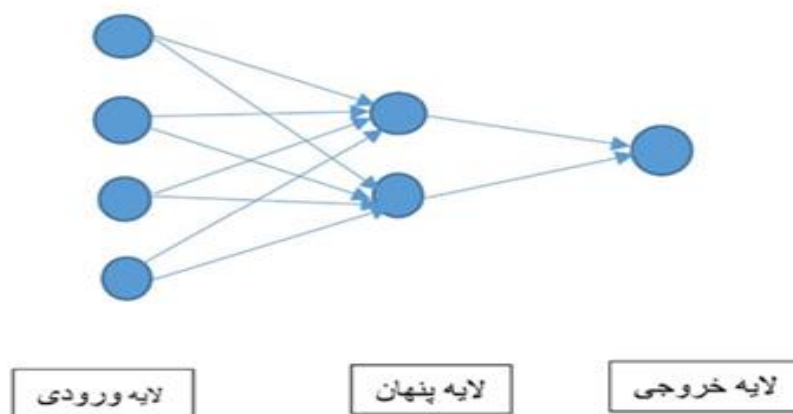
**درخت تصمیم (DT):** درختان تصمیم برای اولین بار در دهه ۱۹۶۰ معرفی شدند و یکی از مؤثرترین روش‌ها برای داده‌کاوی هستند، زیرا استفاده از آن‌ها آسان، بدون ابهام و حتی در صورت وجود مقادیر از دست رفته ممکن و عملی است. در این مدل‌ها هر دو متغیر گسسته و پیوسته می‌توانند به عنوان متغیرهای مستقل یا متغیرهای هدف مورد استفاده قرار گیرند. هنگام ساخت مدل ابتدا باید مهم‌ترین متغیرهای ورودی را شناسایی کرد و سپس بهترین آن‌ها را در گره ریشه و در گره‌های داخلی بعدی بر اساس وضعیت

<sup>3</sup> Information Gain

<sup>4</sup> Entropy

<sup>1</sup> Training Data

<sup>2</sup> Testing Data



شکل ۳: شبکه عصبی مصنوعی چند لایه (MLP)

پارامتر تنظیم بستگی دارند، یک برآورد داخلی از خطای تعمیم داشته و می‌توانند به‌طور مستقیم برای مسائل با ابعاد بالا استفاده شوند و همچنین به‌راحتی می‌توان این مدل‌ها را به‌صورت موازی اجرا کرد (Cutler et al., 2012).

**درختان اضافی (Extra-Trees):** مدل‌های درخت تصادفی<sup>۱</sup> (RT) و RF دو مجموعه روش بسیار مشابه هستند و اشتراکات زیادی دارند. هر دو این مجموعه مدل‌ها از تعداد زیادی درخت تصمیم تشکیل شده‌اند که تصمیم نهایی با در نظر گرفتن پیش‌بینی هر درخت به‌دست می‌آید. هر دو نوع مدل رویه و فرایند رشد درخت یکسانی دارند که در مسائل طبقه‌بندی با اکثریت رای و در مسائل رگرسیونی با میانگین حسابی عمل مدل‌سازی را پیش می‌برند. علاوه بر این، هنگام انتخاب بخش‌های هر گره، هر دو به‌طور تصادفی زیرمجموعه‌ای از ویژگی‌ها را انتخاب می‌کنند. مدل RT داده‌های ورودی را با جایگزینی زیرنمونه می‌گیرد، در حالی که مدل‌های ET از کل نمونه اصلی استفاده می‌کنند. روش RF تقسیم بهینه را انتخاب می‌کند در حالی که ET این انتخاب را به‌طور تصادفی انجام می‌دهد. با این حال، پس از انتخاب نقاط تقسیم، این دو الگوریتم بهترین‌ها را از بین همه زیرمجموعه ویژگی‌ها انتخاب می‌کنند. بنابراین مدل‌های ET فرآیند تصادفی‌سازی را اضافه می‌کند، اما همچنان بهینه‌سازی دارد؛ این تفاوت‌ها باعث کاهش واریانس

$$IG(A) = \text{Entropy}(S) - \sum_{j=1}^v \frac{|D_j|}{|D|} \text{Entropy}(D_j) \quad (2)$$

در رابطه ۲ پارامتر  $v$  تعداد اعضای مربوط به متغیر  $A$  و  $D_j$  قسمتی از داده‌های اولیه‌ای است که در کلاس مربوط به  $V_j$  قرار دارد. پارامتر  $|D|$  نیز نشان‌دهنده اندازه داده‌های کلاس  $D$  است.

**روش‌های یادگیری جمعی (Ensemble Learning):** استفاده از خرد جمعی در حل مسائل، نسبت به جواب یک شخص نتایج بهتری ارائه می‌دهد، همین‌طور تخمین گروهی از مدل‌ها اغلب نسبت به تخمین یک مدل نتایج بهتری دارد. این نوع مدل‌سازی‌ها، به‌عنوان روش‌های یادگیری جمعی شناخته می‌شوند (Géron, 2019).

**جنگل تصادفی (Random Forest):** جنگل تصادفی یک روش یادگیری یکپارچه است که بر روی درخت تصمیم برای بهبود عملکرد پیش‌بینی و تعمیم، با ترکیب درخت‌های تصمیم‌گیری چندگانه ساخته شده‌است. در مسئله طبقه‌بندی، RF طبقه پیش‌بینی نهایی را از طریق مکانیزم رای‌گیری تعیین می‌کند. در مسائل رگرسیونی، مدل‌ها RF با در نظر گرفتن میانگین مقادیر پیش‌بینی‌شده همه درخت‌های تصمیم، پیش‌بینی نهایی را تعیین می‌کنند (Yin, 2023). از نظر محاسباتی، مدل‌های RF بسیار کارآمد هستند، زیرا به‌طور طبیعی هم رگرسیون و هم طبقه‌بندی (چند طبقه) را مدیریت می‌کنند، برای آموزش و پیش‌بینی نسبتاً سریع هستند، فقط به یک یا دو

<sup>1</sup> Random Tree

را برای به حداقل رساندن این گرادیان آموزش می‌دهد، سپس پیش‌بینی‌های مدل جدید به مجموعه اضافه می‌شود (Géron., 2019). در این پژوهش کلیه مدل‌ها و الگوریتم‌های مورد استفاده در محیط برنامه‌نویسی پایتون اجرا شده است.

**معیارهای ارزیابی:** عملکرد مدل‌های مورد استفاده در این پژوهش با استفاده از معیارهای آماری ضریب همبستگی (R)، جذر میانگین مربعات خطاها (RMSE)، میانگین خطای مطلق (MAE) و کلینگ گوپتا (KGE) ارزیابی شده است که این معیارها به ترتیب در روابط ۱ تا ۴ ارائه شده‌اند.

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \times \sum_{i=1}^N (P_i - \bar{P})^2}} \quad \text{رابطه (۱)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad \text{رابطه (۲)}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |P(i) - O(i)| \quad \text{رابطه (۳)}$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad \text{رابطه (۴)}$$

در روابط ۱ تا ۳ پارامتر  $O(i)$  مقادیر مشاهداتی (واقعی)،  $P(i)$  مقادیر شبیه‌سازی شده،  $\bar{O}$  میانگین مقادیر مشاهداتی،  $\bar{P}$  میانگین مقادیر شبیه‌سازی شده و  $N$  تعداد مشاهدات است. در رابطه ۴ پارامتر  $r$  ضریب همبستگی بین داده‌های مشاهداتی و شبیه‌سازی شده،  $\alpha$  نسبت بین میانگین داده‌های مشاهداتی و میانگین داده‌های شبیه‌سازی شده و  $\beta$  نسبت بین انحراف معیار استاندارد داده‌های مشاهداتی و انحراف استاندارد داده‌های شبیه‌سازی شده است. مقادیر  $R$  و  $KGE$  بیشتر و  $MAE$  و  $RMSE$  کم‌تر، نشان‌دهنده دقت بیشتر و عملکرد بهتر مدل می‌باشد.

**دیاگرام تیلور:** دیاگرام تیلور نموداری است که برای مقایسه چندین خروجی مدل‌سازی در یک نمودار استفاده می‌شود. در این نمودار میزان خطا، همبستگی و همچنین انحراف معیار خروجی‌های مدل‌ها با مقادیر واقعی مقایسه می‌شود. این نمودار خلاصه‌ای واضح و

می‌شود. از نظر فرآیند محاسباتی و در نتیجه زمان اجرا، الگوریتم ET سریع‌تر است. این الگوریتم در زمان صرفه‌جویی می‌کند زیرا کل روش یکسان است، اما به‌طور تصادفی نقطه تقسیم را انتخاب می‌کند و نقطه بهینه را محاسبه نمی‌کند. به این دلایل درختان بسیار تصادفی<sup>۱</sup> یا به‌طور خلاصه درختان اضافی نامیده شده است (Geurts et al., 2006).

**تقویت سازگار (Ada Boost):** در روش تقویت سازگار، هر مدل به دنبال برطرف کردن خطای مدل‌های قبل است و تمرکز بر روی نمونه‌هایی است که مدل قبلی نتوانسته پیش‌بینی کند. در این فرآیند تمام نمونه‌ها در مرحله آموزش یک وزن مختص به خود خواهند داشت که این وزن‌ها بر اساس خطای مدل‌ها به دست می‌آیند. وزن نمونه‌های آموزشی اهمیت آن‌ها را مشخص می‌کند و مدل‌ها به کمک این وزن‌ها اهمیت نمونه‌ها و اینکه کدام نمونه‌ها در مراحل قبلی توسط مدل‌های یادگیری به‌درستی طبقه‌بندی یا پیش‌بینی نشده‌اند را متوجه می‌شوند تا تمرکز خود را روی آن‌ها گذاشته و سعی کنند تا همه یا قسمتی از آن‌ها را درست پیش‌بینی کنند. بدین ترتیب روش (AB) توان دیگر مدل‌های ضعیف را تقویت می‌کند تا بتوانند یک مسئله را دقیق‌تر حل کنند (Géron, 2019).

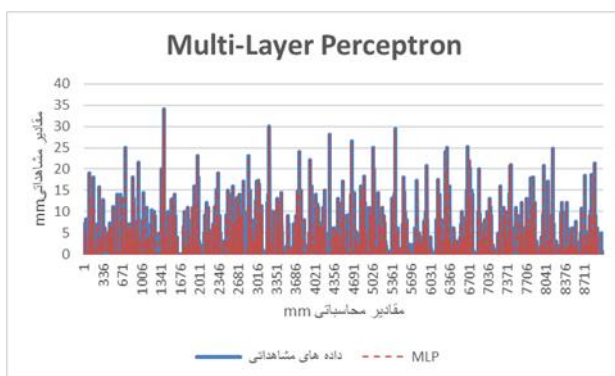
**تقویت گرادیان (Gradient Boost):** تفاوت بین روش‌های AB و GB در عملکردی است که با مقادیر پیش‌بینی نشده در مرحله قبلی انجام می‌دهند. برخلاف روش AB که وزن نمونه‌ها را در هر مرحله تغییر می‌دهد، روش GB سعی می‌کند که پیش‌بینی‌کننده جدید را با خطاهای باقی‌مانده در مرحله قبل تطبیق دهد. به این صورت که در هر مرحله، مدل جدیدی برای به حداقل رساندن تابع خطا مانند میانگین مربعات خطای مدل قبلی با استفاده از گرادیان آموزش داده می‌شود. در هر تکرار، الگوریتم گرادیان تابع خطا را با توجه به پیش‌بینی‌های مجموعه فعلی محاسبه می‌کند و سپس یک مدل ضعیف جدید

<sup>1</sup> Extremely Randomized Tree

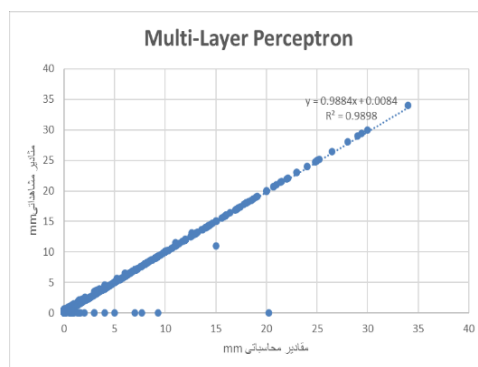
در این پژوهش، از پارامترهای دمای بیشینه، دمای کمینه، دمای میانگین، فشار هوای میانگین، رطوبت نسبی میانگین و سرعت باد بیشینه در مقیاس‌های زمانی روزانه و ماهانه به‌عنوان ورودی مدل‌سازی و پیش‌بینی بارش روزانه و ماهانه استفاده شده است. در شکل‌های ۴ تا ۹ نمودارهای پراکنش و سری زمانی مقادیر مشاهداتی بارش روزانه در مقابل مقادیر محاسباتی یا پیش‌بینی شده با استفاده از سه مدل برتر در مرحله آموزش ارائه شده است.

سریع از میزان مطابقت الگو را نشان می‌دهد و به‌سادگی می‌توان میزان دقت یک مدل شبیه‌سازی سیستم را مشخص کرد. این نمودارها به‌ویژه در ارزیابی شایستگی‌های نسبی مدل‌ها و تعیین عملکرد کلی مدل‌ها مورد بررسی کاربرد دارند. بنابراین برای ارائه مقایسه بصری بهتری از نتایج به‌دست‌آمده از مدل‌سازی از دیاگرام تیلور استفاده می‌شود ( Taylor, 2001).

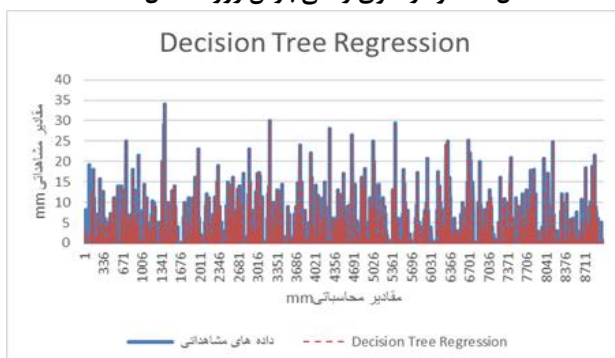
### نتایج و بحث



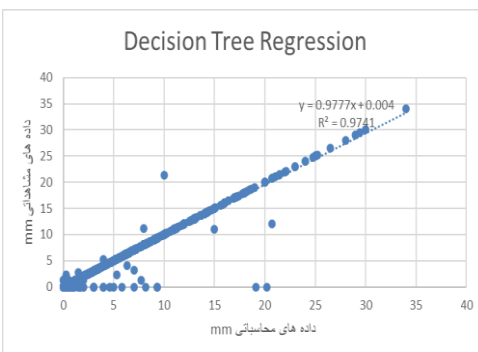
شکل ۵: نمودار سری زمانی بارش روزانه مدل MLP



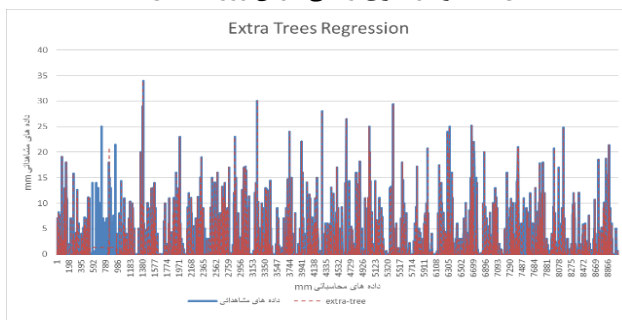
شکل ۴: نمودار پراکنش بارش روزانه مدل MLP



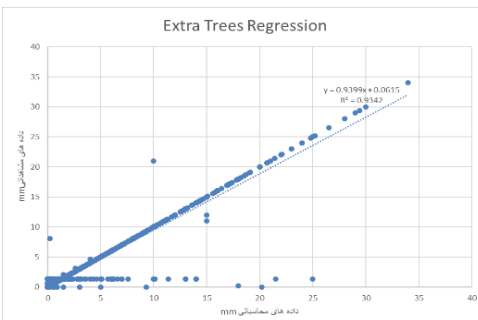
شکل ۷: نمودار سری زمانی بارش روزانه مدل DTR



شکل ۶: نمودار پراکنش بارش روزانه مدل DTR



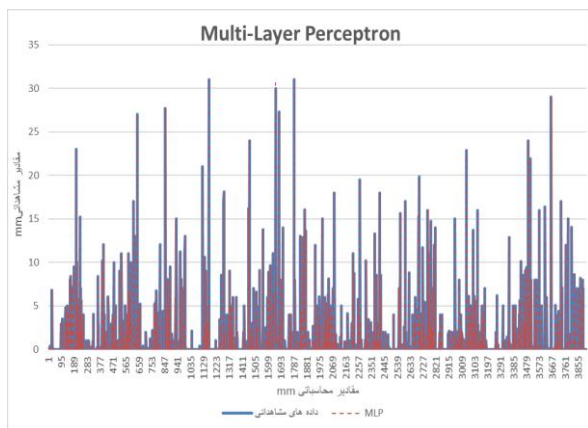
شکل ۹: نمودار سری زمانی بارش روزانه مدل ETR



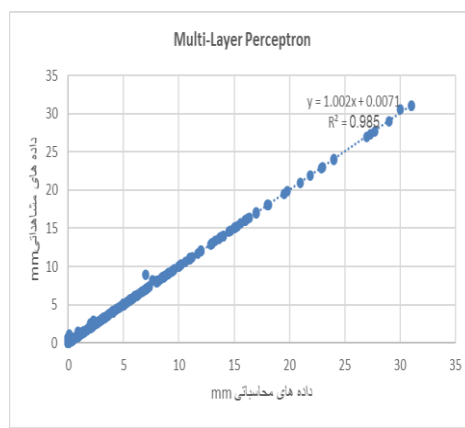
شکل ۸: نمودار پراکنش بارش روزانه مدل ETR

محاسباتی یا پیش‌بینی شده با استفاده از سه مدل برتر در مرحله صحت‌سنجی را نشان می‌دهند.

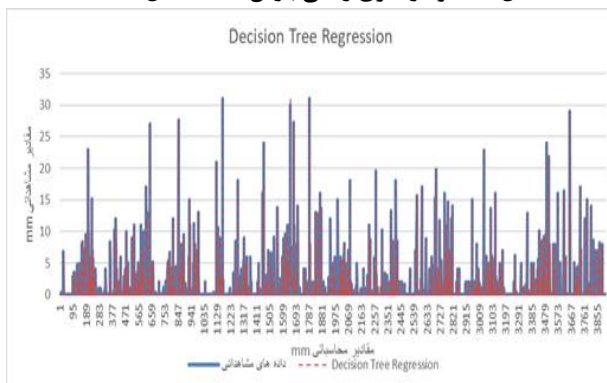
شکل‌های ۱۰ تا ۱۵ نیز نمودارهای پراکنش و سری زمانی مقادیر مشاهداتی بارش روزانه در مقابل مقادیر



شکل ۱۱: نمودار سری زمانی بارش ماهانه مدل MLP



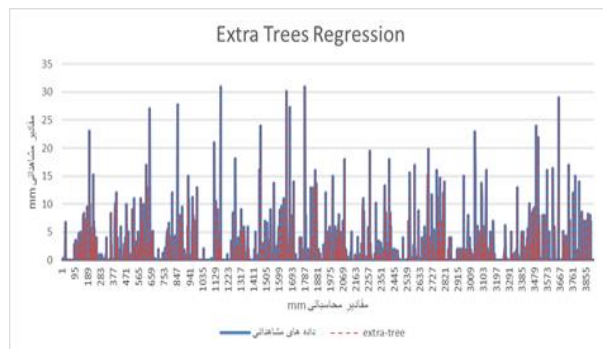
شکل ۱۰: نمودار پراکنش بارش ماهانه مدل MLP



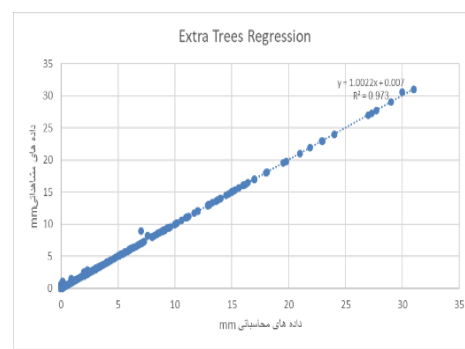
شکل ۱۳: نمودار سری زمانی بارش ماهانه مدل DTR



شکل ۱۲: نمودار پراکنش بارش ماهانه مدل DTR



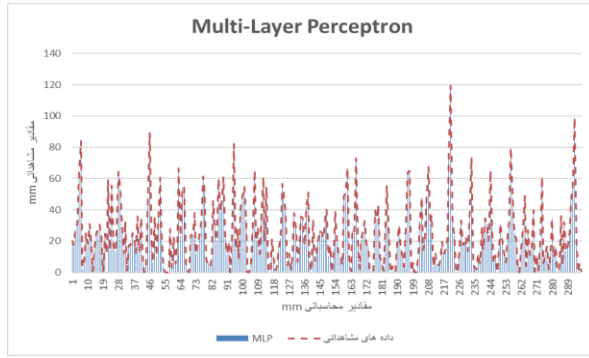
شکل ۱۵: نمودار سری زمانی بارش ماهانه مدل ETR



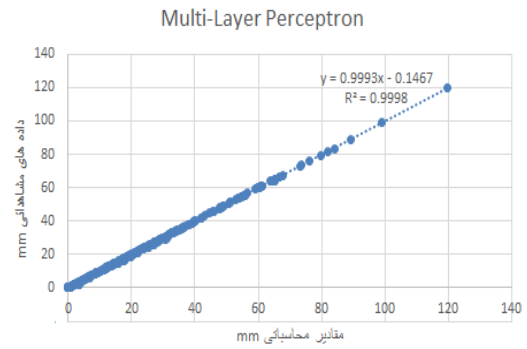
شکل ۱۴: نمودار پراکنش بارش ماهانه مدل ETR

با دقت بالایی شبیه‌سازی کرده است. شکل‌های ۱۶ تا ۲۱ نمودارهای پراکنش و سری زمانی مقادیر مشاهداتی بارش در مقیاس ماهانه در مقابل مقادیر محاسباتی یا پیش‌بینی شده با استفاده از سه مدل برتر در مرحله آموزش را نشان می‌دهند.

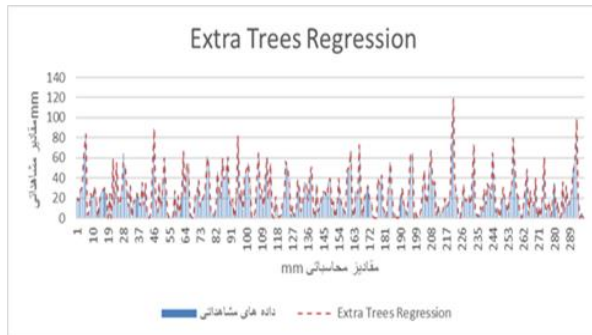
مطابق شکل‌های ۱۰ تا ۱۵ مشاهده می‌شود که داده‌های بارش مشاهداتی و شبیه‌سازی شده با استفاده از هر سه مدل MLP، ETR و DTR میزان انطباق بسیار بالایی دارند. البته مدل MLP نسبت به سایر مدل‌ها دقیق‌تر بوده است و مقادیر بارش روزانه را



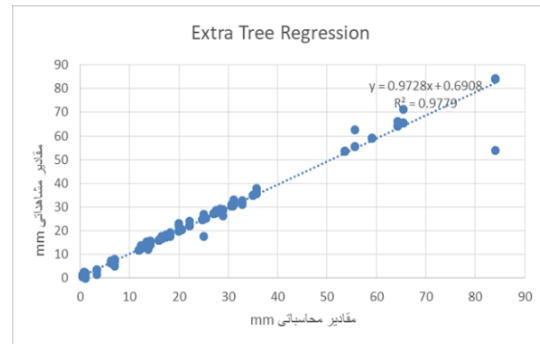
شکل ۱۷: نمودار سری زمانی بارش ماهانه مدل MLP



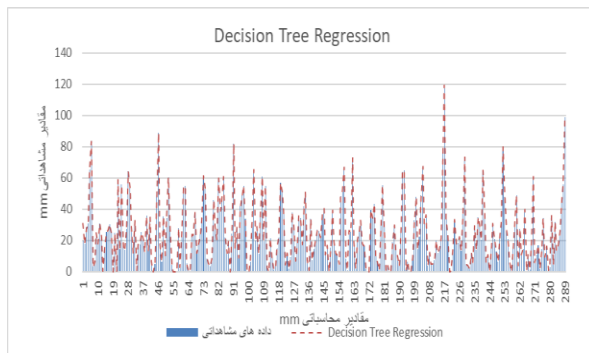
شکل ۱۶: نمودار پراکنش بارش ماهانه مدل MLP



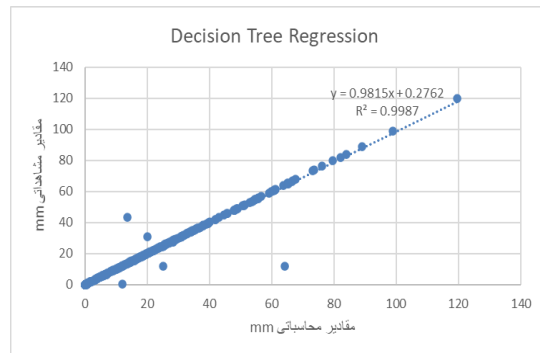
شکل ۱۹: نمودار سری زمانی بارش ماهانه مدل ETR



شکل ۱۸: نمودار پراکنش بارش ماهانه مدل ETR



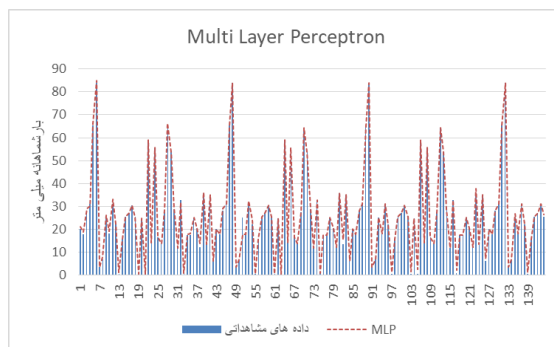
شکل ۲۱: نمودار سری زمانی بارش ماهانه مدل DT



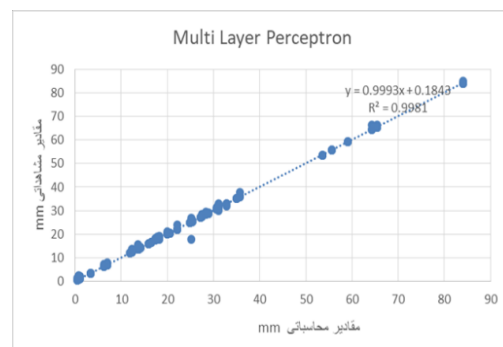
شکل ۲۰: نمودار پراکنش بارش ماهانه مدل DT

استفاده از سه مدل برتر در مرحله صحت‌سنجی را نشان می‌دهند.

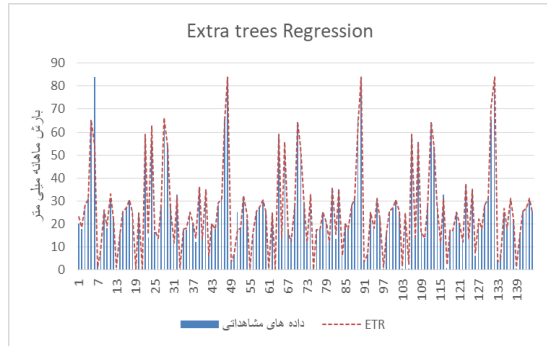
همچنین شکل‌های ۲۲ تا ۲۷ نمودارهای پراکنش و سری زمانی مقادیر مشاهداتی بارش در مقیاس ماهانه در مقابل مقادیر محاسباتی یا پیش‌بینی شده با



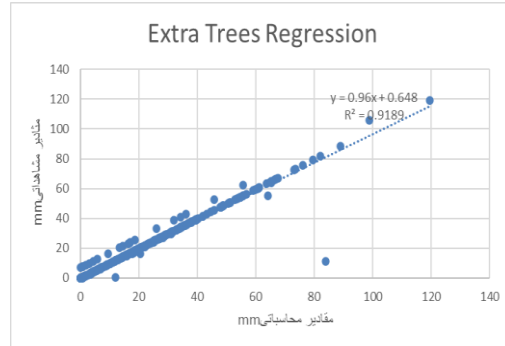
شکل ۲۳: نمودار سری زمانی بارش ماهانه مدل MLP



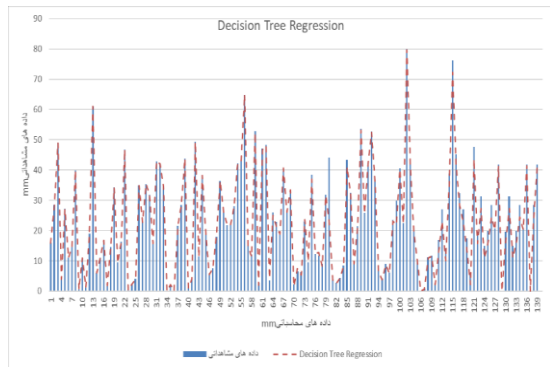
شکل ۲۲: نمودار پراکنش بارش ماهانه مدل MLP



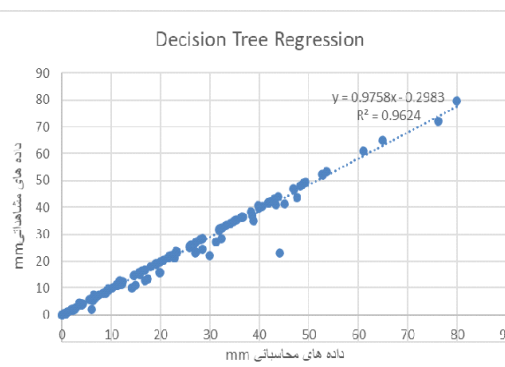
شکل ۲۵: نمودار سری زمانی بارش ماهانه مدل ETR



شکل ۲۴: نمودار پراکنش بارش ماهانه مدل ETR



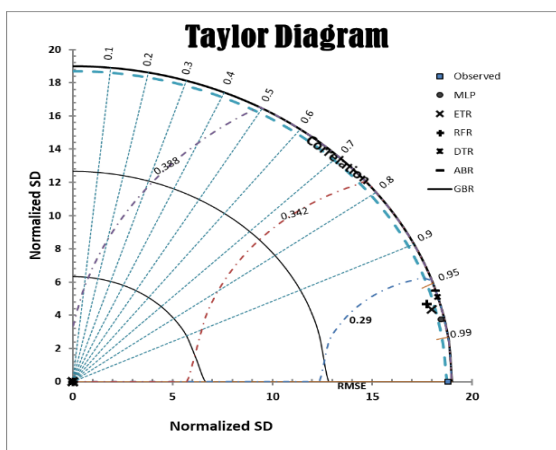
شکل ۲۷: نمودار سری زمانی بارش ماهانه مدل DT



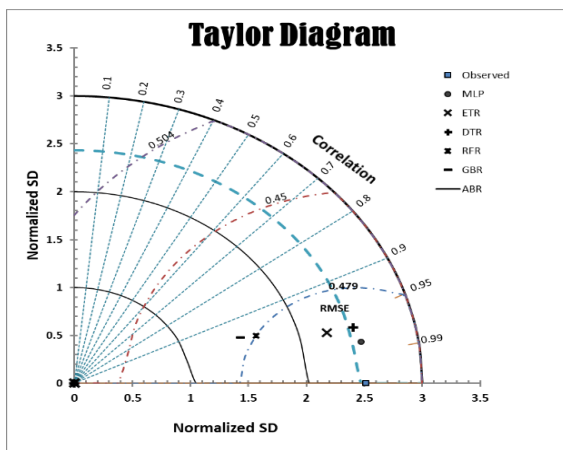
شکل ۲۶: نمودار پراکنش بارش ماهانه مدل DT

است و مقادیر بارش روزانه را با همبستگی بالایی شبیه‌سازی کرده است. برای مقایسه بصری نتایج مدل‌های مورد استفاده در این پژوهش دیاگرام تیلور در هر دو مقیاس زمانی روزانه و ماهانه در شکل‌های ۲۸ و ۲۹ رسم شده است.

مطابق شکل‌های ۲۲ تا ۲۷ مشاهده می‌شود که داده‌های بارش شبیه‌سازی شده با استفاده از هر سه مدل برتر دارای انطباق بسیار بالایی با داده‌های بارش مشاهده‌ای است. مطابق نمودارهای پراکنش مشاهده می‌شود که همانند مقیاس روزانه، در مقیاس ماهانه نیز مدل MLP نسبت به سایر مدل‌ها دقیق‌تر بوده



شکل ۲۹: دیاگرام تیلور برای بارش ماهانه



شکل ۲۸: دیاگرام تیلور برای بارش روزانه

دیاگرام تیلور مربوط به نتایج مدل‌سازی در مقیاس ماهانه را نشان می‌دهد، عملکرد همه مدل‌ها تقریباً مشابه بوده اما با توجه به اینکه SD مربوط به مقادیر محاسبه شده با استفاده از مدل MLP نزدیک به SD مربوط به داده‌های مشاهداتی است و همچنین مقادیر R بیشتر و RMSE کمتری نسبت به سایر مدل‌های مورد بررسی دارد، بهترین عملکرد مربوط به این مدل می‌باشد. در جدول‌های ۳ و ۴ نتایج ارزیابی مدل‌های مورد بررسی در پیش‌بینی بارش به ترتیب در مقیاس‌های زمانی روزانه و ماهانه در مراحل آموزش و صحت‌سنجی ارائه شده است.

مطابق شکل ۲۸ که دیاگرام تیلور مربوط به نتایج مدل‌سازی در مقیاس روزانه را نشان می‌دهد، مدل MLP نسبت به سایر مدل‌های مورد بررسی، دارای کمترین مقدار RMSE و بیشترین مقدار R می‌باشد و بنابراین در مقیاس روزانه بهترین عملکرد را دارد. همچنین همه مدل‌ها به جز مدل MLP و DTR انحراف معیار (SD) کمتری نسبت به مقادیر مشاهداتی داشته‌اند، ولی SD این دو مدل نزدیک به SD مربوط به مقادیر مشاهداتی است و در نتیجه این دو مدل بهترین عملکرد را در مدل‌سازی بارش در مقیاس روزانه از خود نشان داده‌اند. طبق شکل ۲۹ نیز که

جدول ۳- نتایج مدل‌ها در پیش‌بینی بارش در مقیاس روزانه در مراحل آموزش و صحت‌سنجی

KGE	صحت‌سنجی			KGE	آموزش			مدل
	RMSE (mm)	MAE (mm)	R		RMSE (mm)	MAE (mm)	R	
۰/۵۹	۰/۵۴۶	۰/۵۳۷	۰/۹۱۴	۰/۵۱۰	۰/۴۰۰	۰/۴۲۱	۰/۹۳۴	Ada Boost Regression
۰/۶۱	۰/۵۰۴	۰/۴۰۱	۰/۹۴۸	۰/۷۱۱	۰/۶۵۰	۰/۳۲۸	۰/۹۷۲	Gradient Boost Regression
۰/۶۶	۰/۴۵۰	۰/۴۵۰	۰/۹۵۳	۰/۶۴۰	۰/۵۸۰	۰/۳۳۴	۰/۹۶۵	Random Forest Regression
۰/۷۴	۰/۴۷۹	۰/۴۶۱	۰/۹۸۵	۰/۸۷۰	۰/۳۶۰	۰/۳۴۷	۰/۹۶۷	Decision Tree Regression
۰/۷۵	۰/۳۲۴	۰/۳۲۴	۰/۹۸۶	۰/۸۱۰	۰/۳۹۰	۰/۲۹۵	۰/۹۸۷	Extra Trees Regression
۰/۸۲	۰/۱۸۴	۰/۱۸۴	۰/۹۹۳	۰/۹۱۰	۰/۱۰۳	۰/۱۶۱	۰/۹۹۵	Multi-Layer Perceptron

جدول ۴- نتایج مدل‌ها در پیش‌بینی بارش در مقیاس ماهانه در مراحل آموزش و صحت‌سنجی

KGE	صحت‌سنجی			KGE	آموزش			مدل
	RMSE (mm)	MAE (mm)	R		RMSE (mm)	MAE (mm)	R	
۰/۴۱	۰/۳۸۸	۰/۲۹۸	۰/۹۵۱	۰/۷۶	۰/۱۶۰	۰/۲۸۰	۰/۹۶۰	Ada Boost Regression
۰/۳۹	۰/۸۵۴	۰/۶۵۰	۰/۸۸۴	۰/۵۱	۰/۸۵۴	۰/۷۸۰	۰/۹۰۰	Gradient Boost Regression
۰/۶۳	۰/۲۹۰	۰/۲۱۵	۰/۹۵۰	۰/۶۶	۰/۱۹۴	۰/۲۶۰	۰/۹۷۰	Random Forest Regression
۰/۶۵	۰/۳۴۲	۰/۲۵۶	۰/۹۷۲	۰/۷۷	۰/۲۳۹	۰/۱۸۹	۰/۹۸۹	Decision Tree Regression
۰/۷۱	۰/۲۶۶	۰/۱۹۷	۰/۹۸۱	۰/۸۵	۰/۱۷۰	۰/۱۹۷	۰/۹۹۹	Extra Trees Regression
۰/۸۸	۰/۲۲۲	۰/۱۵۳	۰/۹۹۹	۰/۹۴	۰/۱۰۸	۰/۱۴۰	۰/۹۹۹	Multi-Layer Perceptron

میلی‌متر و  $KGE=0/82$  و مدل ET با  $R=0/986$ ،  $RMSE=0/324$  میلی‌متر،  $MAE=0/324$  میلی‌متر و  $KGE=0/75$  نسبت به سایر مدل‌ها عملکرد بهتری داشته است. در مقیاس ماهانه نیز به ترتیب مدل MLP با  $R=0/999$ ،  $RMSE=0/153$  میلی‌متر،  $MAE=0/222$  میلی‌متر و  $KGE=0/88$  و مدل ET با  $R=0/981$ ،  $RMSE=0/266$  میلی‌متر،  $MAE=0/197$

مطابق جدول‌های ۳ و ۴ مشاهده می‌شود که همه مدل‌های مورد بررسی در این پژوهش از دقت بالایی در مدل‌سازی بارش برخوردارند. در هر دو مقیاس زمانی روزانه و ماهانه به ترتیب مدل‌های MLP، ET و DTR نسبت به مدل‌های دیگر دقت بیشتر و خطای کمتری دارند. در مقیاس روزانه مدل MLP با  $R=0/993$ ،  $RMSE=0/184$  میلی‌متر،  $MAE=0/184$

### نتیجه‌گیری

در این پژوهش با استفاده از داده‌های پارامترهای هواشناسی ایستگاه همدیدی تبریز شامل دماهای کمینه، بیشینه و میانگین، رطوبت نسبی میانگین، فشار هوای میانگین و سرعت باد بیشینه در دوره ۲۰۲۱-۱۹۸۶ نسبت به مدل‌سازی بارش در مقیاس‌های زمانی روزانه و ماهانه اقدام شد. بدین منظور از روش‌های یادگیری ماشین شامل شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، روش‌های یادگیری جمعی جنگل تصادفی (RF)، آدابوست (AB)، تقویت گرادیان (GB)، درختان اضافی (ET) و مدل رگرسیون درخت تصمیم (DTR) استفاده شد. نتایج این پژوهش نشان داد که در هر دو مقیاس زمانی روزانه و ماهانه همه مدل‌های مورد بررسی قابلیت بسیار مناسبی دارند. به ترتیب مدل‌های MLP و ET نسبت به سایر مدل‌ها بهترین عملکرد را در هر دو مقیاس روزانه و ماهانه نشان دادند. نتایج این پژوهش نشان‌دهنده عملکرد بسیار مناسب مدل‌های یادگیری ماشین و یادگیری جمعی در مدل‌سازی بارش است. در حالت کلی اگر چه دقت این مدل‌ها در مدل‌سازی بارش روزانه و ماهانه بر اساس سایر پارامترهای هواشناسی بالاست، اما استفاده از این روش‌ها به دانش بالا و تخصص در زمینه استفاده از روش‌های داده‌کاوی نیاز دارد. با توجه به دقت بالای مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه و رگرسیون درخت تصمیم و روش‌های یادگیری جمعی مانند درختان اضافی در پیش‌بینی بارش روزانه و ماهانه، استفاده از این روش‌ها در مباحث مرتبط توصیه می‌شود. همچنین بررسی مقادیر حدی و تاثیر روزها و ماه‌های خشک و تر در تحقیقات آتی پیشنهاد می‌شود.

میلی‌متر و  $KGE=0.71$  از بیشترین دقت و کمترین خطا برخوردار است. این نتایج مشابه پژوهش صالحوند و همکاران (۱۳۹۹) است که در بررسی عوامل مؤثر بر بارش در حوضه کارون نتیجه گرفتند که به دلیل رفتار غیرخطی بارش، استفاده از مدل MLP نتایج مطلوب‌تری دارد. همچنین گوش و همکاران با بررسی مدل‌های یادگیری جمعی برای پیش‌بینی بارش نتیجه گرفتند که استفاده از مدل‌های یادگیری جمعی عملکرد کلی مدل‌ها را بهبود می‌بخشد و از بیش برآزشی و کم‌برآزشی جلوگیری می‌کند (Ghosh et al., 2023). مشابه این نتایج در پژوهش گبره نیز حاصل شده است که با استفاده از مدل‌های یادگیری جمعی مبتنی بر یادگیری ماشین ترکیبی برای مدل‌سازی فرایند بارش-رواناب در حوضه رودخانه کاتار در کشور اتیوپی، دقت و کارایی این مدل‌ها و مدل‌های ترکیبی درخت رگرسیون تقویت شده را نشان داد (Gebre, 2023).

با توجه به مقادیر معیارهای ارزیابی خطای مدل‌ها، به جز مدل GB (طبق هر دو معیار RMSE و MAE) و مدل MLP (طبق معیار RMSE)، سایر مدل‌ها در مدل‌سازی بارش در مقیاس ماهانه خطای کمتری نسبت به مقیاس روزانه داشته‌اند. کم بودن خطای مدل‌سازی در مقیاس ماهانه نسبت به روزانه می‌تواند به دلیل پراکنش کمتر و در نتیجه انحراف معیار کمتر داده‌های ماهانه نسبت به داده‌های روزانه باشد که این مساله موجب بهبود عملکرد مدل‌ها و خطای کمتر آن‌ها در مقیاس ماهانه شده است. هدف از انجام این تحقیق بررسی نتایج مدل‌سازی به صورت یکپارچه بوده است. بنابراین کل سری داده‌های روزانه (به‌طور مشابه در مقیاس ماهانه) به صورت یک سری زمانی متوالی در نظر گرفته شده است. بررسی مقادیر حدی و تاثیر روزها و ماه‌های خشک و تر می‌تواند در تحقیقات آتی مد نظر باشد.

- شبکه‌های عصبی مصنوعی المانی (ENN). علوم و مهندسی آبخیزداری ایران. ۱۰(۳۳): ۱۱-۱۸.
۱۰. قلی‌زاده، م. ح. و دارند، م. ۱۳۸۹. پیش‌بینی بارش ماهانه با استفاده از شبکه‌های عصبی مصنوعی (مورد: تهران). پژوهش‌های جغرافیای طبیعی (پژوهش‌های جغرافیایی). ۴۲(۷۱): ۶۳-۵۱.
۱۱. مظفری، غ. شفیعی، ش. و همتی، ح. ۱۳۹۴. پیش‌بینی بارش ماهانه ایستگاه سینوپتیک کرمانشاه با استفاده از مدل ترکیبی شبکه عصبی و موجک. پژوهش‌های حفاظت آب و خاک. ۲۲(۶): ۱۵۲-۱۳۵.
۱۲. مهتابی، ق. تاران، فرشید. و مظفری، س. ۱۳۹۷. پیش‌بینی وقوع بارش روزانه با استفاده از داده‌های هواشناسی روزهای قبل (مطالعه موردی: شهر اصفهان). جغرافیای طبیعی. ۱۱(۳۹): ۱۱۴-۹۹.
13. Cutler, A., Cutler, D.R. and Stevens, J.R. 2012. Random forests. Ensemble machine learning: Methods and applications. Springer. 157-175.
14. Gebre, G. 2023. Application of hybrid machine learning-based ensemble techniques for rainfall-runoff modeling. Earth Science Informatics. 16(3): 2475-2495.
15. Géron, A. 2022. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc."
16. Geurts, P., Ernst, D. and Wehenkel, L. 2006. Extremely randomized trees. Machine learning, 63: 3-42.
17. Ghosh, S., Gourisaria, M.K., Sahoo, B. and Das, H. 2023. A pragmatic ensemble learning approach for rainfall prediction. Discover Internet of Things. 3(1):13.
18. Han, J. Kamber, M. and Pei, J. 2012. Data mining: concepts and techniques. 3th Ed. Morgan Kaufmann Publishers, San Francisco, California, United States of America, 744 pp.
19. Haykin, S. 1999. Neural Networks: A Comprehensive Foundation. 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ.
20. Jain, S. K., Das, A., and Srivastava, D. K. 1999. Application of ANN for reservoir inflow prediction and operation. Journal of Water Resources
- منابع
۱. الله‌ویردی پور، پ. و ستاری، م. ت. ۱۴۰۲. مقایسه عملکرد روش کلاسیک رگرسیون خطی چندگانه و روش‌های داده‌کاوی نوین در مدل‌سازی بارش سالانه (مطالعه موردی: شهر اهواز). مدل‌سازی و مدیریت آب و خاک. ۳(۲): ۱۴۲-۱۲۵.
۲. امیدوار، ک. و نبوی زاده، م. ۱۳۹۳. پیش‌بینی بارش روزانه استان کرمان با شبکه‌های عصبی مصنوعی (مطالعه موردی: کرمان، بافت و میانده جیرفت). جغرافیا و توسعه ناحیه‌ای. ۱۲(۲): ۲۱۴-۱۹۷.
۳. پورصالحی، ف. شهیدی، ع. و خاشعی سیوکی، ع. ۱۳۹۸. مقایسه‌ی مدل‌های درخت تصمیم M5 و الگوریتم نزدیک‌ترین همسایگی K در پیش‌بینی بارش ماهانه (مطالعه موردی: ایستگاه سینوپتیک بیرجند). آبیاری و زهکشی ایران. ۱۳(۵): ۱۲۹۳-۱۲۸۳.
۴. حلبیان، ا.ح. و دارند، م. ۱۳۹۱. پیش‌بینی بارش اصفهان با استفاده از شبکه‌های عصبی مصنوعی. تحقیقات کاربردی علوم جغرافیایی (علوم جغرافیایی). ۱۲(۲۶): ۶۳-۴۷.
۵. خلیلی، ن. خدانشناس، س. ر. داوری، ک. و موسوی بایگی، م. ۱۳۸۹. پیش‌بینی بارش روزانه با استفاده از شبکه‌های عصبی مصنوعی (مطالعه موردی: ایستگاه سینوپتیک مشهد). پژوهش‌های آبخیزداری. ۲۳(۴): ۷-۱۵.
۶. ستاری، م. ت. و نهرین، ف. ۱۳۹۲. پیش‌بینی مقادیر حداکثر بارش روزانه با استفاده از سیستم‌های هوشمند و مقایسه آن با مدل درختی M5؛ مطالعه موردی ایستگاه‌های اهر و جلفا. آبیاری و آب ایران. ۴(۲): ۹۸-۸۳.
۷. شرفی، م. صمدیان فرد، س. و هاشمی، س. ۱۳۹۹. پیش‌بینی بارش ماهانه با استفاده از مدل‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان. سامانه‌های سطوح آبگیر باران. ۸(۴): ۷۱-۶۳.
۸. صالحوند، ا. گندمکار، ا. و فتاحی، ا. ۱۳۹۹. پیش‌آگاهی بلند مدت بارش با استفاده از سیستم شبکه‌ی عصبی مطالعه‌ی موردی: حوضه‌ی کارون بزرگ. تحقیقات کاربردی علوم جغرافیایی. ۲۰(۵۹): ۹۷-۸۱.
۹. عبدالله پورآزاد، م. ستاری، م. ت. و میرعباسی نجف آبادی، ر. ۱۳۹۵. پیش‌بینی جریان روزانه رودخانه اهرچای با استفاده از مدل قوانین M5 و مقایسه آن با

23. Song, Y.Y. and Ying, L.U. 2015. Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry. 27(2): 130.
24. Taylor, K. E. 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of geophysical research, Atmospheres, 106(D7): 7183-7192.
25. Yin, L. 2023. Forecasting Sector Rotation of A-share Market Using LSTM and Random Forest. Advances in Economics, Management and Political Sciences. 49: 109-123.
21. Mekanik, F., Imteaz, M.A., Gato-Trinidad, S. and Elmahdi, A. 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. Journal of Hydrology. 503: 11-21.
22. Singh, K.K., Pal, M. and Singh, V.P. 2010. Estimation of mean annual flood in Indian catchments using backpropagation neural network and M5 model tree. Water resources management. 24: 2007-2019.

