



Data Mining of 24-Hour Cumulative Precipitation Data in Iran Using Machine Learning: Multilayer Perceptron Neural Network and Decision Tree

Mozaffar Faraji,¹  Majid Rezaei Banafsheh Doragh,^{2*}  Behrouz Sari Sarraf,³ 
Ali Mohammad Khorshid Dust⁴ 

1. PhD Student, Department of Climatology, University of Tabriz, Tabriz, Iran; Email: mozaffar.faraji@tabrizu.ac.ir
2. Professor, Department of Climatology, University of Tabriz, Tabriz, Iran; Email: mrbanafsheh@tabrizu.ac.ir
3. Professor, Department of Climatology, University of Tabriz, Tabriz, Iran; Email: sarisarraf@tabrizu.ac.ir
4. Professor, Department of Climatology, University of Tabriz, Tabriz, Iran; Email: khorshid@tabrizu.ac.ir

Article Info

Article type:
Research Article

Article history:

Received April 06, 2025
Received in revised form May 25, 2025
Accepted May 31, 2025
Available online June 22, 2025

Keywords:

24-hour Cumulative
Precipitation,
Multilayer Perceptron
Neural Network,
Decision Tree,
Iran.

ABSTRACT

Research Topic: This study examines 24-hour cumulative precipitation data in Iran and predicts rainfall amounts over various time periods using data mining and machine learning.

Objective: To develop an accurate model for predicting 24-hour cumulative precipitation in regions of Iran using multilayer neural networks and decision trees to improve hydrological planning and water resource management.

Method: A daily precipitation dataset D was collected from Iranian stations and prepared using normalization. Two machine learning models including MLP with activation function σ and decision tree with entropy criterion were implemented. The models' performance was evaluated and compared with accuracy, precision, and error criteria.

Results: The MLP model demonstrated efficiency in estimating monthly precipitation by minimizing MSE to 0.04. The decision tree analysis classified Iran provinces into seven clusters based on precipitation characteristics; clusters 4 and 7 represent provinces with minimum (including Isfahan, Sistan and Baluchestan, Yazd) and maximum precipitation (including Gilan, Kohgiluyeh and Boyer-Ahmad, Mazandaran), respectively. Linear regression showed a significant effect of the time variable with 0.209 on precipitation variance.

Conclusions: The use of machine learning, especially neural networks, is effective in analyzing hydrological data in Iran and can help improve precipitation forecasting systems.

Cite this article: Faraji, M., Rezaei Banafsheh Doragh, M., Sari Sarraf, B., Khorshid Dust, A. M. (2024). Data Mining of 24-Hour Cumulative Precipitation Data in Iran Using Machine Learning: Multilayer Perceptron Neural Network and Decision Tree. *ECOHYDROLOGY*, 12 (2), 795-811. <http://doi.org/10.22059/IJE.2025.395276.1870>



© Mozaffar Faraji, Majid Rezaei Banafsheh Doragh, Behrouz Sari Sarraf, Ali Mohammad Khorshid Dust.
Publisher: University of Tehran Press.
DOI: <http://doi.org/10.22059/IJE.2025.395276.1870>

Materials and Methods

The data used in this research are hourly and daily information from synoptic stations throughout Iran in the time frame of 1990 to 2020. The selection of this time frame was made considering the availability of high-quality data and appropriate geographical coverage. Before being used in machine learning models, the raw data underwent the necessary pre-processing steps, including the removal of missing values, detection and correction of errors, and data normalization.

Findings

The multilayer perceptron (MLP) neural network, with optimized settings, was able to accurately predict monthly rainfall, reducing the MSE in the test section to 0.04. This indicates the high accuracy of the model in predicting monthly rainfall, and the training time of the model is also within an acceptable range. The MLP neural network with the proposed architecture and settings is introduced as a suitable model for predicting monthly rainfall in different regions of Iran. The use of standard scaling, the hyperbolic tangent activation function, and the MSE error function have helped to improve the model's performance. This study shows that artificial neural networks are powerful tools for modeling and predicting complex phenomena such as rainfall and can be useful in hydrological studies and water resource management. Furthermore, decision tree analysis has been used to identify 24-hour rainfall patterns in the provinces of Iran. The results of this analysis are presented in the form of a tree structure diagram, showing the clustering structure of the provinces based on the average rainfall. These clusters include the provinces of Khuzestan, East Azerbaijan, North Khorasan, Bushehr, Razavi Khorasan, and Alborz with relatively higher rainfall; the provinces of Fars, Markazi, Ardabil, Zanzan, Qazvin, and Hamedan with different rainfall characteristics; the provinces of Tehran, West Azerbaijan, Kermanshah, Lorestan, and Kurdistan with special climatic characteristics; the provinces of Isfahan, Sistan and Baluchestan, and Yazd with significantly lower rainfall; the provinces of Kerman, South Khorasan, Hormozgan, Semnan, and Qom with less rainfall and under the influence of desert and semi-desert conditions; the provinces of Ilam, Chaharmahal and Bakhtiari, and Golestan with rainfall diversity in high altitude areas; and the provinces of Mazandaran, Gilan, and Kohgiluyeh and Boyer-Ahmad with abundant rainfall due to the presence of forests and mountains. This analysis has helped to identify rainfall patterns in different provinces of Iran and can be used to optimize water resource management and predict future rainfall conditions, and serve as a guide for decision-makers in the field of water crisis management and climate change.

Conclusion

In general, the use of machine learning techniques, especially multilayer perceptron (MLP) neural networks and decision tree algorithms, in the analysis of hydrological data related to Iran has effective results and can greatly contribute to the improvement of rainfall prediction systems. MLP neural networks, with high accuracy in predicting monthly rainfall, act as a powerful tool in modeling complex hydrological phenomena. On the other hand, the decision tree algorithm, by identifying rainfall patterns in different provinces, enables the optimization of water resource management and planning for coping with water crises. These results show that by using modern machine learning approaches, a better understanding of rainfall patterns in Iran can be achieved and more effective decisions can be made in the field of water resource management. Combining these two methods with other hydrological models can lead to the creation of more accurate and efficient rainfall prediction systems and play an important role in the sustainable management of the country's water resources.

Data Availability Statement

This article is derived from the doctoral dissertation of Mr. Mozaffar Faraji, which is being conducted at the University of Tabriz.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

The authors are committed to not presenting and falsifying data.

Acknowledgments

The authors express their gratitude to the Meteorological Organization of the country for providing meteorological data, as well as to the esteemed Journal of Ecohydrology of the University of Tehran, the members of the editorial board of the journal, the referees and the respected editor-in-chief. Also, special thanks to the respected scientific and executive director of the journal, Mr. Dr. Mohammad Mirzavand, for his valuable and scientific reviews for this article.

داده‌کاوی داده‌های بارش تجمعی ۲۴ ساعته ایران با استفاده از یادگیری ماشین شبکه عصبی پرسپترون چندلایه و درخت تصمیم

مظفر فرجی،^۱ مجید رضایی بنفشه درق،^{۲*} بهروز ساری صراف،^۳ علی محمد خورشیددوست^۴

۱. دانشجوی دکتری، گروه آب‌وهواشناسی، دانشگاه تبریز، تبریز، ایران. رایانامه: mozaffar.faraji@tabrizu.ac.ir

۲. استاد، گروه آب‌وهواشناسی، دانشگاه تبریز، تبریز، ایران. رایانامه: mrbanafsheh@tabrizu.ac.ir

۳. استاد، گروه آب‌وهواشناسی، دانشگاه تبریز، تبریز، ایران. رایانامه: sarisarraf@tabrizu.ac.ir

۴. استاد، گروه آب‌وهواشناسی، دانشگاه تبریز، تبریز، ایران. رایانامه: khoshid@tabrizu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	موضوع: این مطالعه به بررسی داده‌های بارش تجمعی ۲۴ ساعته در ایران و پیش‌بینی میزان بارش در دوره‌های زمانی مختلف از طریق داده‌کاوی و یادگیری ماشین می‌پردازد.
تاریخ دریافت: ۱۴۰۴/۰۱/۱۷	هدف: توسعه مدلی دقیق برای پیش‌بینی بارش تجمعی ۲۴ ساعته در مناطق ایران با استفاده از شبکه‌های عصبی چندلایه و درخت تصمیم، به منظور بهبود برنامه‌ریزی هیدرولوژیکی و مدیریت منابع آب.
تاریخ بازنگری: ۱۴۰۴/۰۲/۲۵	روش تحقیق: مجموعه داده بارش روزانه D از ایستگاه‌های ایران جمع‌آوری و با نرمال‌سازی آماده شد. دو مدل یادگیری ماشین شامل MLP با تابع فعال‌سازی σ و درخت تصمیم با معیار آنتروپی پیاده‌سازی شد. عملکرد مدل‌ها با معیارهای دقت، صحت و خطا ارزیابی و مقایسه گردید.
تاریخ پذیرش: ۱۴۰۴/۰۳/۱۰	یافته‌ها: مدل MLP در تخمین بارش ماهانه با حداقل سازی MSE به میزان ۰/۰۴ کارایی نشان داد. تحلیل درخت تصمیم، استان‌های ایران را براساس ویژگی‌های بارش به هفت خوشه‌بندی تقسیم کرد؛ خوشه‌های ۴ و ۷ به‌ترتیب استان‌های با کمینه (شامل اصفهان، سیستان و بلوچستان، یزد) و بیشینه بارش (شامل گیلان، کهگیلویه و بویراحمد، مازندران) را نشان داد. رگرسیون خطی تأثیر معنادار متغیر زمان با ۰/۲۰۹ بر واریانس بارش را نشان داد.
تاریخ انتشار: ۱۴۰۴/۰۴/۰۱	نتیجه‌گیری: استفاده از یادگیری ماشین، به‌ویژه شبکه‌های عصبی، در تحلیل داده‌های هیدرولوژیکی ایران مؤثر است و می‌تواند به بهبود سیستم‌های پیش‌بینی بارش کمک کند.
کلیدواژه‌ها: بارش تجمعی ۲۴ ساعته، شبکه عصبی پرسپترون چندلایه، درخت تصمیم، ایران.	

استناد: فرجی، مظفر؛ رضایی بنفشه درق، مجید؛ ساری صراف، بهروز؛ خورشیددوست، علی محمد. داده‌کاوی داده‌های بارش تجمعی ۲۴ ساعته ایران با استفاده از یادگیری

ماشین شبکه عصبی پرسپترون چندلایه و درخت تصمیم. *اکوهیدرولوژی*، ۱۲(۲)، ۷۹۵-۸۱۱.

<http://doi.org/10.22059/IJE.2025.395276.1870>



© مظفر فرجی، مجید رضایی بنفشه درق، بهروز ساری صراف، علی محمد خورشیددوست.

ناشر: انتشارات دانشگاه تهران.

مقدمه

تغییرات آب‌وهوایی اشاره به تغییراتی در میانگین یا تنوع ویژگی‌های آب‌وهوایی دارد که به‌طور مستمر برای دوره‌های طولانی ادامه خواهد داشت. این تغییرات ناشی از عوامل طبیعی و فعالیت‌های انسانی هستند (عساکره و مطلبی‌زاد، ۱۳۹۶). بارش به‌عنوان یکی از عناصر کلیدی چرخه هیدرواقليمی کره زمین، تحت تأثیر روابط پیچیده و غیرخطی بین سامانه اقليم و عوامل محیطی قرار دارد. شناخت این روابط و انجام برنامه‌ریزی‌های محیطی بر مبنای آن‌ها در حالت عادی امری دشوار است و به همین دلیل، طبقه‌بندی داده‌ها و تقسیم اطلاعات به دسته‌های متجانس و کوچک می‌تواند از این منظر مفید واقع شود (شاهبایی کوتنایی و عساکره، ۱۴۰۲). این فرایندها می‌توانند در مدیریت و مهار سیلاب‌ها، طراحی سازه‌های آبی در حوزه‌های آبخیز و نیز مدیریت خشکسالی مؤثر واقع شوند (دهقانی، وفاخواه و بهره‌مند، ۱۳۹۵). پیش‌بینی بارش به‌عنوان نقش اساسی در حفظ، مدیریت و تخصیص منابع آب، تعیین حجم تأسیسات آبی، تأمین نیاز آبی محصولات کشاورزی و همچنین ارزیابی میزان فرسایش و رسوب به شمار می‌رود. برای پیش‌بینی بارش، روش‌های متعددی وجود دارد. یکی از این روش‌ها، استفاده از شبکه‌های عصبی مصنوعی است که براساس اطلاعات اقلیمی و جغرافیایی نظیر دمای حداقل و حداکثر، رطوبت، بارش، طول و عرض جغرافیایی، و ارتفاع از سطح دریا عمل می‌کند (بهرامی، سالاری و امیری، ۱۴۰۲). پیش‌بینی بارش یکی از مسائل مهم در زمینه مدیریت بهینه منابع آب در بخش‌های مختلف نظیر صنعت، شرب و کشاورزی است. توانایی پیش‌بینی بارش می‌تواند به جلوگیری از تلفات و خسارات ناشی از بلایای طبیعی کمک کند (فلاح قاله‌ری و شاکری، ۱۳۹۴). بارش از جمله پارامترهای اقلیمی مهم تأثیرگذار بر رژیم هیدرولوژیکی حوضه‌های آبخیز است. با وجود روش‌های متعدد پیش‌بینی میزان بارش، از جمله مدل‌های سری زمانی و شبکه‌های عصبی مصنوعی، شایان ذکر است که بسیاری از این مدل‌ها بدون در نظر گرفتن مسائل گرمایش جهانی و تغییر اقليم عمل می‌کنند (بابایی حصار و قضاوی، ۱۳۹۴). امروزه انواع مختلف شبکه‌های عصبی مصنوعی در حوزه‌های مختلف علوم جو و اقليم‌شناسی برای اهدافی نظیر طبقه‌بندی، رگرسیون و پیش‌بینی کاربرد دارند (پاک‌دامن، ۱۴۰۱). گسترش سریع استفاده از این شبکه‌ها به‌عنوان مدل‌های تجربی و کارآمد در زمینه‌های مختلف از جمله هواشناسی و اقليم‌شناسی نشان‌دهنده ضرورت و اهمیت بالای مطالعه این مدل‌هاست. پیش‌بینی بارش برای اهداف مختلفی نظیر برآورد سیلاب، خشکسالی، مدیریت حوضه آبریز، و کشاورزی از اهمیت بالایی برخوردار است (قلی‌زاده و دارند، ۱۳۸۹). بارش‌های تجمعی ۲۴ ساعته در مناطق مختلف ایران نقش اساسی در شکل‌دهی به اکوسیستم‌ها، فعالیت‌های کشاورزی و جنبه‌های مختلف حیات انسانی ایفا می‌کنند؛ به‌ویژه این بارش‌ها می‌توانند تأثیرات قابل توجهی بر روی منابع آبی، تغذیه خاک و محیط‌زیست داشته باشند؛ درعین حال، نوسانات شدید بارش ممکن است موجب خسارات اقتصادی و اجتماعی فراوانی بشود؛ از این‌رو، درک دقیق از الگوها و روندهای بارشی به‌ویژه در مقیاس زمانی ۲۴ ساعته، برای برنامه‌ریزی مؤثر و مدیریت بهینه منابع آبی ضروری است. تحقیقات نشان داده‌اند که نوسانات بارش و تغییرات اقلیمی می‌توانند منجر به بروز پدیده‌های طبیعی آسیب‌زا مانند سیلاب‌ها، خشکسالی‌ها و الگوهای نامنظم کشاورزی شوند. بنابراین، پیش‌بینی دقیق بارش‌ها نه تنها اهمیت اقتصادی و اجتماعی دارد، بلکه از نظر محیط‌زیستی نیز در کاهش آسیب‌های ناشی از تغییرات اقلیمی و تأمین امنیت غذایی مؤثر است. پژوهش حاضر، با بهره‌گیری از مدل‌های پیشرفته و روش‌های تحلیلی ریاضی، در پی درک عمیق‌تر الگوهای بارشی در ایران و ارزیابی تأثیرات آن‌ها بر مدیریت منابع آب و پیش‌بینی رخدادهای طبیعی است. به همین منظور، از روش‌هایی نظیر شبکه عصبی پرسپترون چندلایه (MLP)، درخت تصمیم، تحلیل کلاسترینگ، و رگرسیون خطی استفاده شده است. به‌طور مشخص، این مطالعه به دنبال شناسایی الگوهای مشخص بارش در سطح منطقه‌ای، توسعه مدل‌های پیش‌بینی مؤثر برای بارش‌های آینده و ارائه ابزاری کارآمد برای ارزیابی و تحلیل داده‌های بارشی است. از آنجاکه وجود الگوهای قوی می‌تواند در بهبود دقت پیش‌بینی‌ها و مدیریت منابع آب تأثیرگذار باشد، زمینه‌های پژوهشی در این حوزه دارای اهمیت و ضرورت بیشتری است. این پژوهش با استفاده از زبان برنامه‌نویسی R رویکردی نوآورانه برای داده‌کاوی بارش ۲۴ ساعته در ایران ارائه می‌دهد. نوآوری‌های اصلی شامل ترکیب مدل‌های MLP و درخت تصمیم برای افزایش دقت و تفسیرپذیری، به‌کارگیری روش‌های پیش‌پردازش داده پیشرفته نظیر انتخاب ویژگی RFE کاهش ابعاد (PCA/t-SNE) و مدیریت داده‌های ناقص (KNN imputation) است. همچنین، ارزیابی جامع مدل‌ها با معیارهای چندگانه MAE، RMSE، R-squared و... و اعتبارسنجی متقابل، در کنار استفاده از کتابخانه‌های تخصصی R، به توسعه مدل‌های پیش‌بینی بارش دقیق‌تر، پایدارتر و قابل

تفسیرتر منجر شده است. این یافته‌ها به درک الگوهای بارش و مدیریت منابع آب در ایران، به‌ویژه در برنامه‌ریزی مخاطرات بارش‌های شدید کمک می‌کند.

پیشینه پژوهش

در راستای درک عمیق‌تر الگوهای بارش در ایران و ارزیابی کارایی روش‌های آماری و یادگیری ماشین، این پژوهش مروری بر مطالعات پیشین در زمینه تحلیل بارش با استفاده از این روش‌ها انجام داده است. این بخش به بررسی و تحلیل این مطالعات می‌پردازد تا زمینه علمی و جایگاه این تحقیق را در ادبیات پژوهشی مشخص کند. مطالعات پیشین در این حوزه را می‌توان به چند دسته کلی تقسیم کرد. پژوهش‌های پیشین نشان داده‌اند که تغییرات اقلیمی می‌تواند تأثیرات قابل توجهی بر الگوهای بارش داشته باشند. به عنوان نمونه، در مطالعه‌ای که با استفاده از مدل CanESM2 تحت سناریوهای مختلف RCP انجام شده است، شبیه‌سازی تغییرات اقلیمی در ایستگاه کرج نشان داد که میزان بارندگی در آینده نزدیک و دور به طور قابل توجهی کاهش خواهد یافت (ساداتی‌نژاد، سلیمانی ساردو و میرزاوند، ۱۴۰۳). این یافته‌ها با نتایج حاصل از مدل‌های یادگیری ماشین که برای پیش‌بینی پارامترهای اقلیمی مورد استفاده قرار می‌گیرند و دقت این مدل‌ها را در پیش‌بینی مقادیر بارش تأیید می‌کنند، همسو است. این همسویی نشان می‌دهد که استفاده از روش‌های مختلف مدل‌سازی اقلیمی می‌تواند به درک بهتر الگوهای بارش و پیش‌بینی تغییرات آبی کمک کند. برخی پژوهش‌ها به بررسی کاربرد متغیرهای هواشناسی سینوپتیک و روش‌های داده‌کاوی برای پیش‌بینی بارش‌های شدید منطقه‌ای در جنوب غرب ایران پرداخته‌اند، که نتایج نشان داد ترکیب الگوهای آب‌وهوایی و روش‌های یادگیری ماشین می‌تواند به بهبود پیش‌بینی بارش کمک کند و بر اهمیت داده‌کاوی و یادگیری ماشین در تحلیل الگوهای بارش تأکید دارد (شاهقلیان، بذرافشان و ایران‌نژاد، ۱۴۰۱). همچنین، کاربرد رویکرد هوش مصنوعی در مطالعه تأثیر محرک‌های بزرگ‌مقیاس آب‌وهوایی بر بارش بلوچستان پاکستان نشان داد که مدل‌های شبکه عصبی مصنوعی ANN و شبکه عصبی بازگشتی BRNN عملکرد بهتری نسبت به مدل‌های رگرسیون و مؤلفه‌های اصلی در پیش‌بینی بارش ماهانه دارند، که این امر قدرت رویکردهای هوش مصنوعی را در پیش‌بینی بارش برجسته می‌کند (تاجبار، خورشیددوست و جهان‌بخش اصل، ۱۴۰۱). بررسی‌های دیگر بر تأثیر الگوریتم‌های آموزش شبکه‌های عصبی پرسپترون چندلایه بر دقت پیش‌بینی بارش ماهانه ایران متمرکز بوده‌اند و مشخص شده که الگوریتم‌های تنظیم‌کننده بیزی و لونبرگ مارکوف عملکرد بهتری نسبت به سایر روش‌ها داشته‌اند (پاکدامن، ۱۴۰۱). همچنین، شبیه‌سازی عددی تغییرات آب‌وهوایی ایران با استفاده از الگوریتم شبکه عصبی مصنوعی نشان داده است که مدل‌های ارائه‌شده می‌توانند با دقت قابل قبولی مقادیر بارش را پیش‌بینی نمایند (فرجی و همکاران، ۱۴۰۲). در همین راستا، روش‌های یادگیری ماشین برای ایجاد یک سیستم هشدار اولیه برای بارندگی‌های شدید کوتاه‌مدت به کار گرفته شده و نشان داده‌اند که این مدل‌ها می‌توانند به پیش‌بینی بموقع رویدادهای بارش شدید کمک کنند (مون^۱ و همکاران، ۲۰۱۹). در یک مطالعه دیگر، روزهای بارندگی شدید در شبه‌جزیره هند با استفاده از الگوریتم‌های یادگیری ماشین پیش‌بینی شده و نتایج حاکی از آن بود که مدل‌های مختلف یادگیری ماشینی می‌توانند به بهبود دقت پیش‌بینی‌های بارش سنگین کمک کنند (سوبرهمانیام^۲ و همکاران، ۲۰۲۱). همچنین، عدم قطعیت مدل شبکه عصبی در ریزمقیاس‌گردانی HadCM3 با روش فاصله اطمینان بوت استرپ بررسی شده و مشخص گردید که دقت شبکه عصبی در پیش‌بینی دمای بیشینه در تابستان کمتر است و عدم قطعیت بالاتری دارد، که این امر بر اهمیت ارزیابی عدم قطعیت در مدل‌های پیش‌بینی بارش تأکید دارد (قرمز چشمه و همکاران، ۱۳۹۴). در زمینه تغییرات زمانی-مکانی بارش‌های بحرانی در غرب ایران طی سال‌های ۱۹۶۵-۲۰۱۶، مشخص شد که با افزایش عرض جغرافیایی، از مجموع بارش‌های فرین و سهم آن‌ها از بارش سالانه کاسته می‌شود، که این مطالعه الگوهای منطقه‌ای بارش را روشن ساخته است (جهانبخش اصل و همکاران، ۱۳۹۹). در تحلیلی بر مطالعات تغییر اقلیم در ایران، افزایش فرین‌های جوئی به‌عنوان مهم‌ترین پیامد این پدیده در ایران معرفی شده است، که نشان می‌دهد تغییرات اقلیمی باعث تغییر الگوهای آب‌وهوایی و افزایش رویدادهای شدید جوئی شده است (اکبری و صیاد، ۱۴۰۰). درنهایت، نقش عوامل درونی و بیرونی در وردایی دهه‌ای بارش سالانه ایران بررسی شد و مشخص گردید که طی چهار دهه اخیر، تأثیر عوامل بیرونی موجب کاهش پهنه‌های

1. Moon

2. Subrahmanyam

بارشی شده است، که این مطالعه بر اهمیت بررسی عوامل مؤثر بر الگوهای بارش در مقیاس‌های زمانی بزرگ‌تر تأکید دارد (عساکره، مسعودیان و ترک‌ارانی، ۱۴۰۰). مطالعات پیشین به‌وضوح اهمیت استفاده از روش‌های آماری و یادگیری ماشین در تحلیل و پیش‌بینی بارش را نشان می‌دهند. با این حال، پژوهش حاضر با بررسی جامع و ترکیبی روش‌های مختلف یادگیری ماشین و آماری (مانند شبکه عصبی MLP، تحلیل خوشه‌ای و درخت تصمیم)، و تمرکز ویژه بر الگوهای بارش در شمال ایران، رویکردی متفاوت را در پیش گرفته است. این پژوهش سعی دارد با تحلیل دقیق‌تر داده‌های بارش، درک جامع‌تری از الگوهای بارشی در شمال ایران و همچنین توانایی روش‌های مختلف در مدل‌سازی این الگوها ارائه دهد. تغییرات آب‌وهوایی به تغییراتی در میانگین یا تنوع ویژگی‌های آب‌وهوایی اشاره دارد که به‌طور مستمر برای دوره‌های طولانی ادامه خواهد داشت و این تغییرات ناشی از عوامل طبیعی و فعالیت‌های انسانی هستند (عساکره و مطلبی‌زاد، ۱۳۹۶).

مواد و روش‌ها

داده‌های مورد استفاده در این پژوهش، اطلاعات ساعتی و روزانه ایستگاه‌های سینوپتیک سراسر ایران در بازه زمانی ۱۹۹۰ تا ۲۰۲۰ هستند. انتخاب این بازه زمانی با توجه به در دسترس بودن داده‌های با کیفیت و پوشش مناسب جغرافیایی صورت گرفته است. داده‌های خام قبل از استفاده در مدل‌های یادگیری ماشین، مراحل پیش‌پردازش لازم را شامل حذف مقادیر ناقص، تشخیص و اصلاح خطاها و نرمال‌سازی داده‌ها پشت‌سر گذاشته‌اند.

شبکه عصبی پرسپترون چندلایه

شبکه پرسپترون چندلایه MLP یکی از انواع شبکه‌های عصبی است. در این شبکه‌ها عملیات یادگیری و پردازش داده‌ها به‌سمت جلو از سمت لایه ورودی به لایه خروجی صورت می‌گیرد (دستوربان و افخمی، ۲۰۱۱). پرسپترون، تابعی است که ورودی x آن را با ضریب وزن‌های آموزش‌دیده ضرب می‌کند. مقدار خروجی $f(x)$ به‌صورت (رابطه ۱) است. بردار وزن‌ها با ارزش واقعی، b بایاس و X برابر با بردار مقادیر ورودی بارش تجمعی ۲۴ ساعته است. هریک از نرون‌ها شامل یک تابع محرک‌اند. توابع محرک مختلفی وجود دارد که با توجه به ساختار مسئله و نیز ساختار مد نظر برای شبکه عصبی از آن‌ها استفاده می‌شود (سینیکو، ۱۹۸۹). تابع تانژانت هیپربولیک، مقادیر حقیقی را به بازه $[-1, 1]$ نگاشته و به‌صورت (رابطه ۲) تعریف می‌شود.

$$F(x) = \begin{cases} 0 & w \cdot x + b > 0 \\ 1 & \text{Otherwise} \end{cases} \quad (\text{رابطه ۱})$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (\text{رابطه ۲})$$

فرایند پیش‌پردازش داده‌ها پس از جمع‌آوری داده‌ها، مراحل پیش‌پردازش شامل حذف مقادیر گمشده، نرمال‌سازی داده‌ها و آماده‌سازی ویژگی‌ها برای ورود به شبکه عصبی انجام می‌شود. سپس داده‌ها به مجموعه‌های آموزشی و آزمایشی تقسیم می‌گردند تا بتوان صحت و کارایی مدل را بر روی داده‌های مستقل ارزیابی کرد. در فرایند آموزش، تابع هزینه برای اندازه‌گیری فاصله بین خروجی پیش‌بینی‌شده و خروجی واقعی به‌صورت (رابطه ۳) زیر است:

$$C = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (\text{رابطه ۳})$$

$$\delta^{(l)} = \nabla_a C \times \phi'(z^{(l)}) \quad (\text{رابطه ۴})$$

که در آن، N تعداد نمونه‌ها، y_i خروجی واقعی و $\hat{y}_i$ خروجی پیش‌بینی شده است. پس از محاسبه خطا (رابطه ۴)، فرایند پس‌گردانی خطا Backpropagation با استفاده از روش‌های بهینه‌سازی مانند گرادیان کاهشی آغاز می‌شود. در این مرحله، وزن‌ها و بایاس‌های شبکه به‌روزرسانی می‌شوند تا خطا به حداقل برسد. این به‌روزرسانی‌ها براساس محاسبات گرادیان انجام می‌شود. فرایند پس‌انتشاری و به‌روزرسانی وزن‌ها برای به‌روزرسانی وزن‌ها در هر لایه، از گرادیان‌ها و قوانین انتشارات خطا استفاده می‌شود.

گردایان‌های مربوطه در معادلات (رابطه‌های ۵ تا ۶) پایین: $\delta(l)$ خطای لایه l است. $a(l-1)$ خروجی لایه قبلی است. $\phi'(z(l))$ مشتق تابع فعال‌سازی است.

$$\nabla_w C = a^{(l-1)} \times \delta^{(l)} \quad \text{(رابطه ۵)}$$

$$\nabla_b C = \delta^{(l)} \quad \text{(رابطه ۶)}$$

ساختار نهایی شبکه و توابع فعال‌سازی برای خروجی، توابع فعال‌سازی مانند سیگموید (رابطه ۷) و ReLU (رابطه ۸) به کار می‌روند. خروجی نهایی شبکه نیز به صورت رابطه ۹ است:

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad \text{(رابطه ۷)}$$

$$\text{ReLU}(z) = \max(0, z) \quad \text{(رابطه ۸)}$$

$$\hat{y} = a^{(L)} \quad \text{(رابطه ۹)}$$

تابع تحلیل واریانس

در مبحث آزمون برابری میانگین‌ها، اگر تعداد جمعیت‌هایی که می‌خواهیم میانگین آن‌ها را باهم مقایسه کنیم بیش از دو تا باشد، از تابع آنالیز واریانس یک‌طرفه استفاده شود. از این تابع anova برای تشکیل جدول آنالیز واریانس ساده و چندطرفه به فرم کلی (object,...) anova استفاده می‌شود (اسماعیلیان، وحدت و حیدردوست، ۱۳۹۵: ۱۵۰). مقایسه میانگین بارش در ایستگاه‌های مختلف، به تعیین تفاوت احتمالی میانگین‌ها کمک می‌کند. این فرایند از جمع‌آوری داده تا تحلیل آماری (با توجه به سطح معناداری و درجه آزادی)، به اعتبارسنجی مدل و پیش‌بینی دقیق‌تر منجر می‌شود (فرجی و همکاران، ۱۴۰۳).

تابع درخت تصمیم

روشی قابل تفسیر برای حل مسائل طبقه‌بندی و رگرسیون است. این روش سلسله‌مراتبی، براساس ویژگی‌ها تصمیم می‌گیرد و نیازی به فرض‌های خاص درباره توزیع داده‌ها ندارد. قابلیت کار با داده‌های گمشده و تفسیر آسان از مزایای آن است. (راهنمای دسترسی به توابع در زبان برنامه‌نویسی، S). آنتروپی درخت تصمیم به صورت زیر (رابطه ۱۰) است که در آن P_i احتمال برخورد به کلاس i میزان اطلاعات کسب‌شده (گین اطلاعات) برای ویژگی A (رابطه ۱۱) است.

$$H(S) = -\sum_{i=1}^C P_i \log_2 P_i \quad \text{(رابطه ۱۰)}$$

$$IG(S, A) = H(S) - \sum_{v \in A} \frac{|S_v|}{|S|} H(S_v) \quad \text{(رابطه ۱۱)}$$

تابع کلاسترینگ

روشی در یادگیری غیرنظارتی برای خوشه‌بندی داده‌های مشابه است. الگوریتم k-means، با انتخاب نقاط تصادفی به عنوان مراکز اولیه، داده‌ها را به نزدیک‌ترین مرکز تخصیص می‌دهد و مراکز را با میانگین داده‌های تخصیص یافته به‌روزرسانی می‌کند. این فرایند تا زمانی که مراکز تغییر نکنند تکرار می‌شود (راهنمای دسترسی به توابع در زبان برنامه‌نویسی، S). تابع هزینه در روش k-means به صورت زیر (رابطه ۱۲) است. در این معادله x_j نمونه‌های داده است و μ_i مرکز خوشه i است.

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad \text{(رابطه ۱۲)}$$

تابع رگرسیون خطی

در مبحث رگرسیون، ما به دنبال یافتن رابطه‌ای بین متغیر وابسته و متغیرهای مستقل هستیم. این رابطه ممکن است از نوع خطی یا غیرخطی باشد. تابعی که ارتباط بین متغیرهای مستقل و وابسته را بیان می‌کند به تابع رگرسیون موسوم است (اسماعیلیان، ۱۳۸۴). تابع مدل رگرسیون خطی به شکل زیر (رابطه ۱۳) است: در این رابطه، Y مقدار بارش ۲۴ ساعته است $X_1, X_2, \dots, X_n$. ویژگی‌های ورودی مانند زمان β_0 عرض از مبدأ β_i ضرایب مربوط به هر ویژگی و ϵ خطای مدل است.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad \text{(رابطه ۱۳)}$$

یافته‌ها و نتایج تحقیق

نتایج شبکه عصبی پرسپترون لایه

نتایج آموزش و آزمون مدل (جدول ۱) نشان می‌دهد که شبکه عصبی MLP با استفاده از تنظیمات فوق، قادر به پیش‌بینی دقیق بارش ماهانه است. مقدار MSE در بخش آزمون به ۰/۰۴ کاهش یافته که نشان‌دهنده دقت بالای مدل در پیش‌بینی بارش ماهانه است. همچنین، زمان آموزش برای مدل نیز در محدوده قابل قبولی قرار دارد. براساس نتایج به‌دست‌آمده، شبکه عصبی MLP با معماری و تنظیمات پیشنهادی، به‌عنوان یک مدل مناسب برای پیش‌بینی بارش ماهانه در مناطق مختلف ایران معرفی می‌شود. استفاده از مقیاس‌بندی استاندارد برای داده‌ها، تابع فعال‌ساز تانژانت هیپربولیک و تابع خطای MSE، همگی به بهبود عملکرد مدل کمک کرده‌اند. این مطالعه نشان می‌دهد که شبکه‌های عصبی مصنوعی، ابزارهای قدرتمندی برای مدل‌سازی و پیش‌بینی پدیده‌های پیچیده مانند بارش هستند و می‌توانند در مطالعات هیدرولوژیکی و مدیریت منابع آب مورد استفاده قرار گیرند.

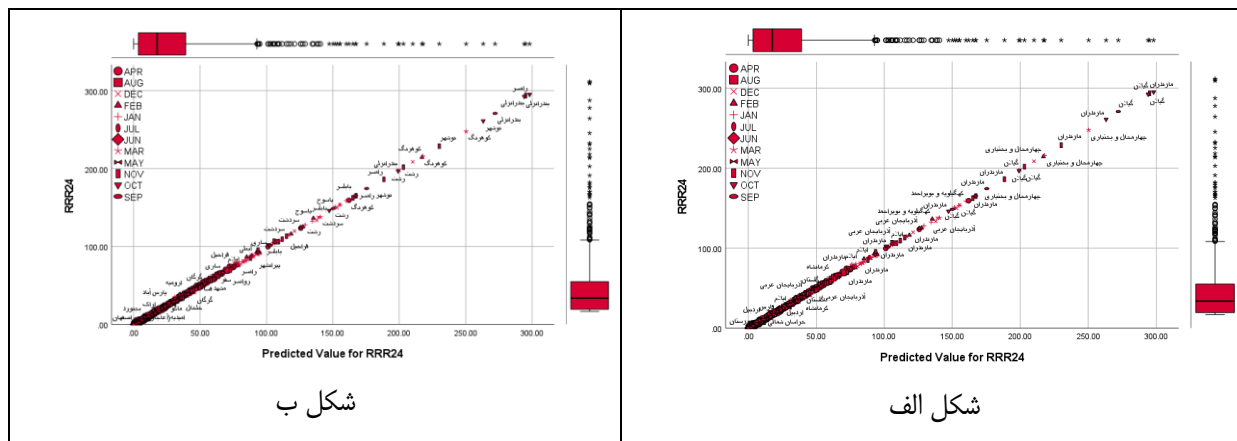
جدول ۱. نتایج مدل شبکه‌های عصبی پرسپترون چندلایه

آموزش	آزمون
مجموع مربعات خطا: ۱۲۷۰۶/۶۰۴	مجموع مربعات خطا: ۶۳۷۱/۱۵۱
میانگین خطای نسبی کل: ۰/۳۳۱	میانگین خطای نسبی کل: ۰/۱۹۴
خطای نسبی برای متغیر وابسته: ۰/۵۲۵	خطای نسبی برای متغیر وابسته: متغیر
قانون زمان متوقف‌شده: یک مرحله متوالی بدون کاهش خطا	خطای نسبی برای متغیر وابسته: بارش تجمعی ۲۴ ساعته
زمان آزمون: ۳۹/۳۸ :۰۰:۰۵	خطای نسبی برای متغیر وابسته: ۰/۰۴۰

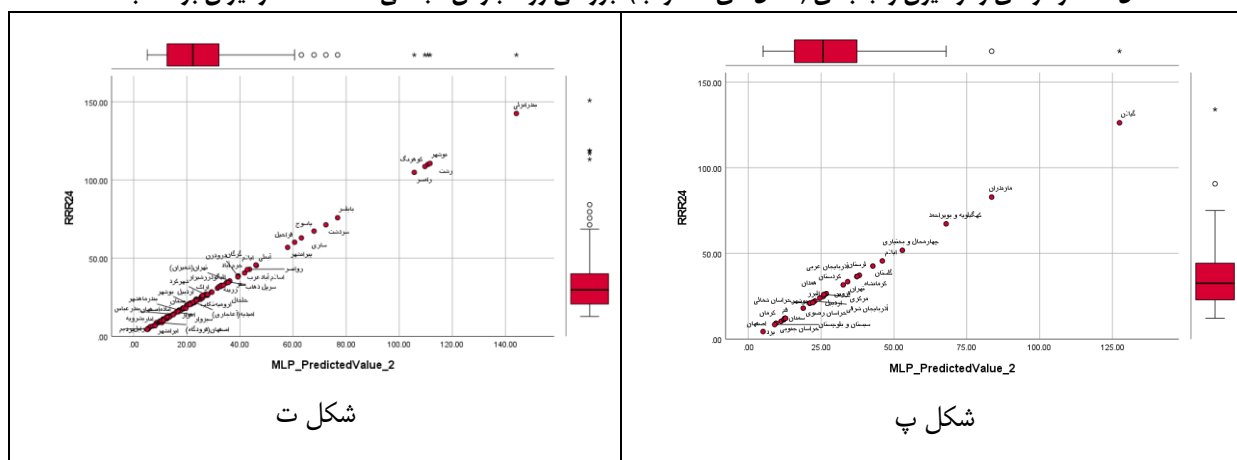
تحلیل مقایسه‌ای بارش واقعی و پیش‌بینی‌شده

این بخش به تحلیل تطبیقی داده‌های بارش واقعی (مشاهده‌شده) و مقادیر پیش‌بینی‌شده توسط مدل می‌پردازد. هدف، درک الگوهای زمانی (ماهانه، فصلی و سالانه) و مکانی بارش، و ارزیابی توانایی مدل در بازتولید این الگوهاست. برای این منظور، از نمودارهای رگرسیون و باکس‌پلات (شکل‌های ۱ تا ۲) استفاده شده است. داده‌های بارش ماهانه از ایستگاه‌های هواشناسی سراسر ایران جمع‌آوری و با پیش‌بینی‌های مدل مقایسه شده‌اند: تحلیل ۱، الگوهای ماهانه و فصلی بارش تجمعی، فصل زمستان (دسامبر تا فوریه)؛ این فصل، به‌ویژه ماه‌های دسامبر، ژانویه و فوریه، بیشترین میزان بارش را در ایران تجربه می‌کند. این پدیده عمدتاً به دلیل فعالیت سامانه‌های مدیترانه‌ای و سیستم‌های پرفشار سیبری در این فصول است که با ورود رطوبت از دریای مدیترانه و دریای سیاه، بارش‌های قابل توجهی را به‌ویژه در نوار شمالی و غرب کشور به همراه دارد. بارش‌ها شامل برف در ارتفاعات و باران در مناطق پست‌تر است. فصل بهار (مارس تا مه)؛ بارش‌ها در ماه‌های مارس و آوریل افزایش قابل توجهی نشان می‌دهند، اما از ماه مه به تدریج کاهش می‌یابند. اوایل بهار، همچنان تحت تأثیر باقی‌مانده فعالیت سامانه‌های زمستانی و جبهه‌های هوایی ناپایدار است. با پیشرفت فصل و افزایش دما، به تدریج الگوهای همرفتی (بارش‌های محلی و رگباری) غالب می‌شوند، اما بارش‌های گسترده‌تر کاهش می‌یابند. فصل تابستان (ژوئن تا سپتامبر)؛ بارش‌ها در این فصل، به‌ویژه در ماه‌های ژوئن، ژوئیه و اوت به حداقل خود می‌رسند. این فصل با استقرار پرفشار آזור بر روی ایران و خاورمیانه مشخص می‌شود که منجر به پایداری جوئی و فرونشینی هوا می‌شود. این شرایط مانع از تشکیل ابرها و بروز بارش‌های گسترده می‌گردد. بارش‌های تابستانی عمدتاً محلی، رگباری و در مناطق کوهستانی است. فصل پاییز (اکتبر تا نوامبر)؛ با شروع فصل پاییز، به‌ویژه در ماه‌های اکتبر و نوامبر، بارش‌ها مجدداً افزایش می‌یابند. با تضعیف پرفشار تابستانی و شروع حرکت به سمت جنوب سامانه‌های مدیترانه‌ای و ورود جبهه‌های سرد به کشور، رطوبت و ناپایداری جوئی افزایش یافته و منجر به آغاز بارش‌های پاییزی می‌شود. نتیجه‌گیری الگوهای زمانی مدل پیش‌بینی‌شده، الگوی ماهانه و فصلی مشابهی را با داده‌های مشاهده‌شده نشان می‌دهد. بیشترین و کمترین مقادیر بارش در ماه‌های یکسان (یعنی مارس تا آوریل و دسامبر تا فوریه برای حداکثر؛ و ژوئن تا اوت برای حداقل) توسط مدل پیش‌بینی شده است. این انطباق، نشان‌دهنده توانایی بالای مدل در بازتولید تغییرات فصلی بارش در ایران است. تحلیل ۲، مقایسه مقادیر بارش تجمعی سالانه و ۲۴ ساعته در سطح

ایستگاه‌های ایران، داده‌های مشاهده‌شده: بیشترین بارش سالانه در ایستگاه بندر انزلی (۱۴۲/۵۵ میلی‌متر) و کمترین آن در یزد (۴/۳۶ میلی‌متر) ثبت شده است. پیش‌بینی مدل: مدل نیز مقادیر بسیار نزدیکی را برای بندر انزلی (۱۴۴/۰۸ میلی‌متر) و یزد (۵/۰۸ میلی‌متر) پیش‌بینی کرده است. تفاوت اقلیمی بین مناطق ساحلی شمال (بندر انزلی با اقلیم پرباران خزر) و مناطق کویری مرکزی (یزد با اقلیم خشک) ایران، دلیل اصلی این تفاوت فاحش در میزان بارش سالانه است. مدل به‌درستی، این تفاوت‌های اقلیمی و مکانی را بازتاب داده است. تحلیل ۳، تحلیل بارش تجمعی ۲۴ ساعته استانی داده‌های واقعی و پیش‌بینی مدل در این بخش، به مقایسه مقادیر بارش تجمعی ۲۴ ساعته بین داده‌های مشاهده‌شده و پیش‌بینی‌های مدل در سطح استان‌ها پرداخته می‌شود تا توانایی مدل در شناسایی و بازتولید الگوهای مکانی بارش‌های شدید بررسی گردد. بالاترین میزان بارش تجمعی ۲۴ ساعته مشاهده‌شده، در استان گیلان با ۱۲۶/۲۴ میلی‌متر ثبت شده است. درمقابل، کمترین میزان بارش ۲۴ ساعته مشاهده‌شده، در استان یزد با ۴/۳۶ میلی‌متر گزارش شده است. مدل نیز برای استان گیلان، بارش تجمعی ۲۴ ساعته‌ای معادل ۱۲۷/۴۲ میلی‌متر را پیش‌بینی کرده است. برای استان یزد نیز، مدل مقدار ۵/۰۸ میلی‌متر را پیش‌بینی نموده است. موقعیت جغرافیایی و اقلیمی: گیلان به دلیل قرار گرفتن در ساحل جنوبی دریای خزر و دامنه شمالی رشته‌کوه البرز، یک منطقه با اقلیم بسیار مرطوب و پربارش است. تأثیر دریای خزر: رطوبت حاصل از تبخیر آب دریای خزر، یک منبع عظیم بخار آب برای منطقه فراهم می‌کند. پدیده‌های اوروگرافیک (کوهساری): هنگامی که توده‌های هوای مرطوب از روی دریای خزر به سمت جنوب حرکت کرده و به دامنه‌های شمالی البرز برخورد می‌کنند، مجبور به صعود می‌شوند. این صعود اجباری هوا باعث سرد شدن، رسیدن به نقطه شبنم و در نتیجه چگالش بخار آب و تشکیل ابرها و بارش‌های شدید اوروگرافیک (به‌ویژه در بازه ۲۴ ساعته) می‌گردد. این پدیده، دلیل اصلی بارش‌های فوق‌العاده زیاد در این منطقه است. استان یزد (بارش حداقلی): یزد در مرکز فلات خشک ایران واقع شده و بخش عمده آن را مناطق بیابانی و نیمه‌بیابانی تشکیل می‌دهند. دوری از منابع رطوبتی: این استان از منابع رطوبتی بزرگ مانند دریاها فاصله زیادی دارد. اثر سایه کوهستان: رشته‌کوه‌های اطراف، مانند زاگرس در غرب و البرز در شمال، اثر سایه کوهستان Rain Shadow Effect بر یزد دارند. به این معنا که رطوبت توده‌های هوا قبل از رسیدن به این منطقه، در دامنه‌های رو به باد این کوه‌ها (مانند دامنه‌های غربی زاگرس) تخلیه شده و هوای خشک به یزد می‌رسد. این عوامل منجر به بارش‌های بسیار اندک و نامنظم در این استان می‌شوند. نتایج مقایسه‌ای نشان می‌دهد که مدل قادر است تفاوت‌های فاحش مکانی در میزان بارش تجمعی ۲۴ ساعته بین مناطق پربارش (مانند گیلان) و مناطق خشک (مانند یزد) را با دقت قابل قبولی پیش‌بینی کند. انطباق بالای مقادیر پیش‌بینی‌شده با داده‌های واقعی در این دو استان با اقلیم کاملاً متفاوت، نشان‌دهنده توانایی مدل در درک و شبیه‌سازی الگوهای پیچیده آب‌وهوایی و هیدرولوژیکی ایران است. این قابلیت برای برنامه‌ریزی مدیریت سیلاب، منابع آب و کشاورزی در مقیاس استانی، بسیار حائز اهمیت است. تحلیل ۴، نمودارهای رگرسیون ترسیم‌شده بین داده‌های واقعی و پیش‌بینی‌شده، همبستگی بالایی را بین این دو مجموعه داده نشان می‌دهند. این همبستگی قوی به این معناست که مدل قادر است تغییرات در بارش واقعی را به خوبی ردیابی کند. نقاط داده‌ای که نزدیک به خط رگرسیون خط $Y=X$ قرار دارند، نشان‌دهنده دقت بالای پیش‌بینی مدل در آن موارد هستند. این یافته، اعتبار مدل را در پیش‌بینی مقادیر بارش در سطوح مختلف تأیید می‌کند. تحلیل مقایسه‌ای نشان می‌دهد که مدل توسعه‌یافته نه تنها قادر به پیش‌بینی دقیق مقادیر بارش تجمعی ۲۴ ساعته است، بلکه توانایی قابل توجهی در بازتولید الگوهای پیچیده ماهانه، فصلی و سالانه بارش و همچنین تفاوت‌های مکانی شدید در سراسر ایران دارد. این نتایج، ارزش عملی مدل را در کاربردهای هیدرولوژیک و مدیریت منابع آب افزایش می‌دهد.



شکل ۱. نمودارهای رگرسیون و جعبه‌ای (شکل‌های الف و ب)، بررسی روند بارش تجمعی ۲۴ ساعته در ایران برحسب ماه



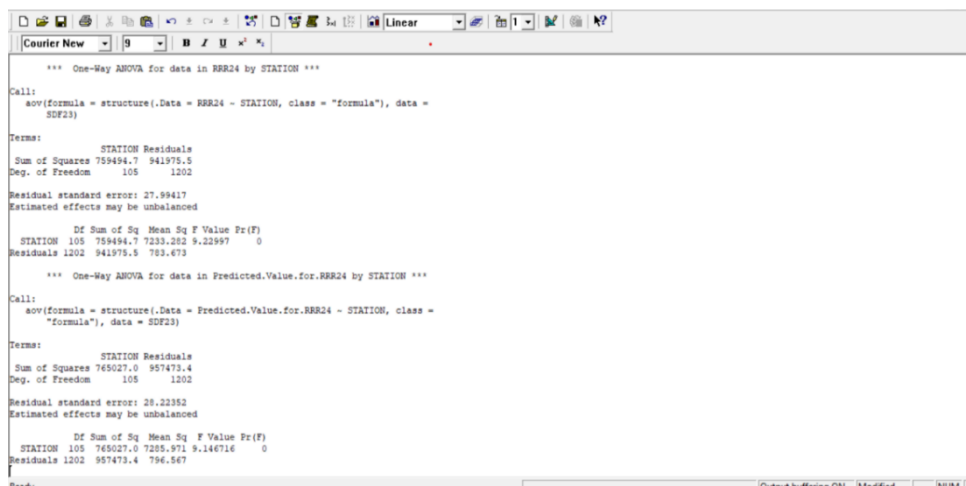
شکل ۲. نمودارهای رگرسیون و جعبه‌ای (شکل‌های پ و ت)، بررسی روند بارش تجمعی ۲۴ ساعته در ایران برحسب سال

نتایج تابع تحلیل واریانس

براساس تحلیل انجام‌شده (جدول ۲ و شکل ۳) مقدار آماره F به‌دست‌آمده و همچنین سطح معنی‌داری P که کمتر از 0.05 گزارش شده است، نشان‌دهنده وجود تفاوت‌های معنادار بین میانگین بارش تجمعی ۲۴ ساعته ایستگاه‌های هواشناسی ایران در دوره مشاهده‌ای و شبیه‌سازی شده می‌باشد. نتایج آزمون آنالیز واریانس یک‌طرفه به‌روشنی مؤید این واقعیت است که میانگین بارش در ایستگاه‌های مختلف تحت بررسی، از نظر آماری به‌طور معناداری با یکدیگر متفاوت است و می‌توان نتیجه گرفت که فرضیه صفر مبنی بر برابری میانگین‌های بارش تجمعی ۲۴ ساعته در ایستگاه‌های مختلف رد می‌شود؛ در نتیجه، این یافته‌ها اهمیت بسزایی در ارزیابی دقت مدل‌های شبیه‌سازی بارش و بررسی کارایی آن‌ها در پیش‌بینی بارش واقعی ایفا می‌کنند. این نتایج، به تحقیقات آینده در زمینه بهبود الگوریتم‌های شبیه‌سازی بارش و تطابق آن‌ها با داده‌های واقعی کمک شایانی خواهد کرد. برای تقویت اعتبار این نتایج، توصیه می‌شود که تحلیل‌های تکمیلی و بررسی‌های بیشتری در زمینه عوامل مؤثر بر بارش و دقت شبیه‌سازی‌ها صورت پذیرد.

جدول ۲. نتایج تحلیل واریانس برای بارش تجمعی ۲۴ ساعته در ایران

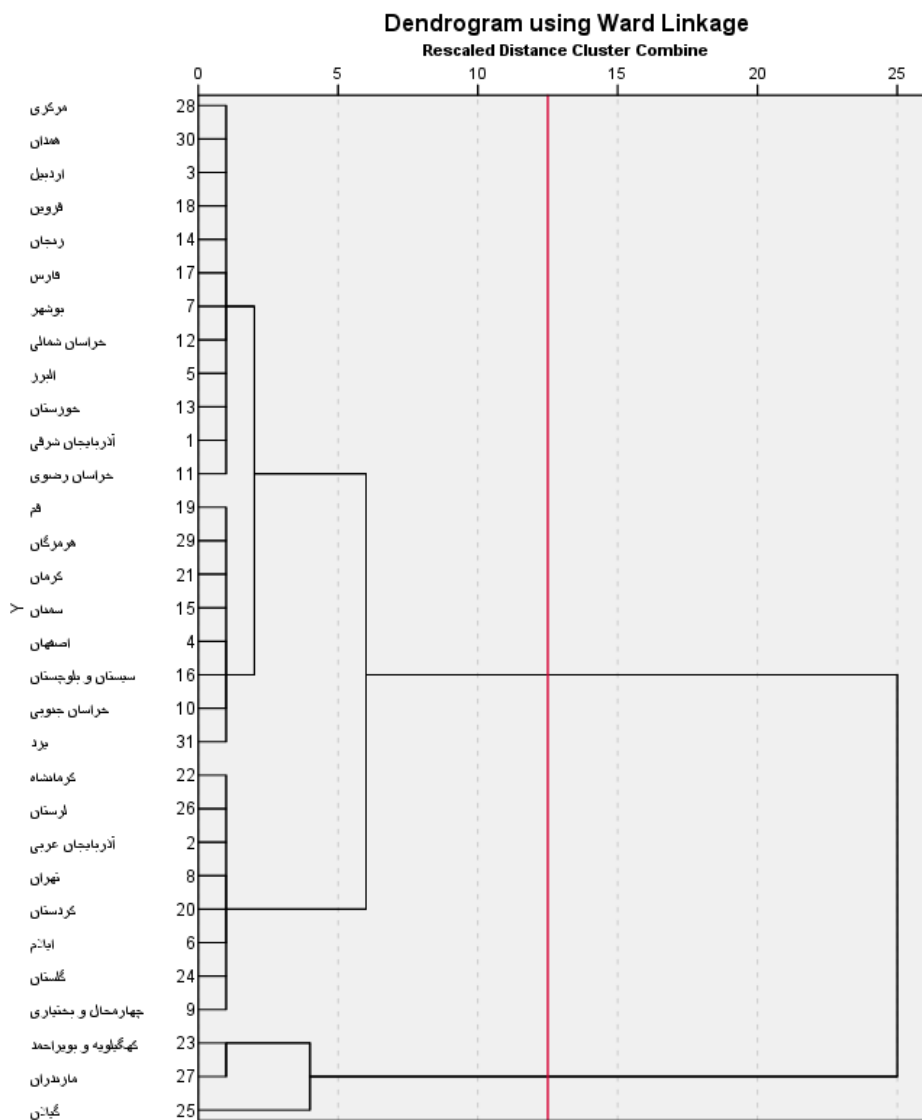
متغیر	درجه آزادی	آماره F	سطح معنی‌داری
بارش ثبت‌شده در ایران	۱۰۵	۹/۲۲	۰/۰۰۰
بارش مدل‌سازی شده در ایران	۱۰۵	۹/۱۴	۰/۰۰۰



شکل ۳. نتایج تحلیل واریانس برای بارش تجمعی ۲۴ ساعته در ایران (راهنمای دسترسی به توابع در زبان برنامه‌نویسی، S)

نتایج تابع کلاسترینگ

تحلیل خوشه‌بندی به‌عنوان یک ابزار مهم در داده‌کاوی و تحلیل داده‌های پیچیده، با هدف شناسایی الگوهای مشابه در داده‌های مربوط به بارش ۲۴ ساعته استان‌های ایران، به کار گرفته شد. نتایج این تحلیل (شکل ۴)، از طریق شکل‌گیری دندروگرام نمایان گردید که نحوه خوشه‌بندی استان‌ها براساس میزان بارش را به تصویر می‌کشد. در نمودار درختی دندروگرام خوشه‌بندی، هر استان در سطح اول، یک خوشه مجزا را تشکیل می‌دهد. این نمای کلی، نشان‌دهنده تنوع بارش‌ها در استان‌های مختلف و تمایز بین آن‌ها است. در این دندروگرام، سطوح مختلف به‌عنوان لایه‌های خوشه‌بندی تعریف شده‌اند که هرکدام بعد خاصی از تشابه بارش‌ها را نمایان می‌سازد. سطح اول: هر استان یک خوشه مستقل را تشکیل می‌دهد و هیچ ارتباطی بین استان‌ها وجود ندارد. سطح دوم: استان‌ها براساس تشابهات بارشی به چهار خوشه کلی گروه‌بندی می‌شوند. این طبقه‌بندی نشان‌دهنده وجود روابط مشابه در میزان بارش استان‌هاست و به ما این امکان را می‌دهد تا استان‌هایی با بارش مشابه را شناسایی کنیم. سطح سوم: در این سطح، استان‌ها به سه طبقه تقسیم می‌شوند که مهر تأییدی بر نشانه‌های مشترک بین آن‌هاست. این استان‌ها دارای ویژگی‌ها و الگوهای بارشی مشابهی هستند که می‌تواند ناشی از شرایط جغرافیایی و اقلیمی مشترک باشد. سطح چهارم: در نهایت، استان‌ها به دو طبقه تقسیم می‌شوند. در این مرحله، استان‌های مازندران، چهارمحال و بختیاری و کهگیلویه و بویراحمد در یک طبقه قرار گرفته‌اند و به لحاظ بارش‌های مشابه با یکدیگر هم‌گروه شده‌اند. در همین راستا، استان گیلان به‌تنهایی در طبقه آخر این سطح‌بندی قرار دارد که نشان‌دهنده ویژگی‌های خاص بارش در این استان است. این تحلیل خوشه‌ای و دندروگرام حاصل، می‌تواند به شناسایی مناطق با خصوصیات بارشی مشابه یاری رسانده و به برنامه‌ریزی بهتر برای مدیریت منابع آب و پیش‌بینی خطرات مرتبط با بارش‌های شدید کمک نماید.

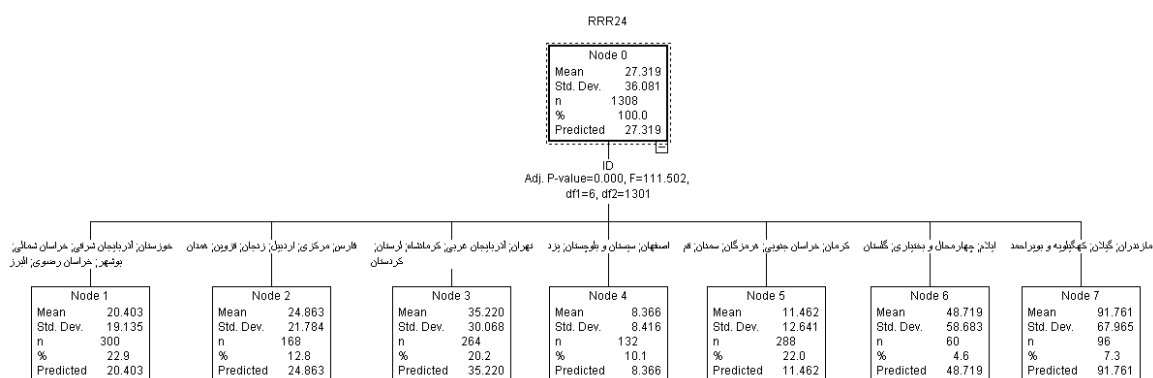


شکل ۴. نمودار دندروگرام خوشه‌بندی بارش استان‌های ایران با استفاده از روش وارد

نتایج تابع درخت تصمیم

تحلیل درخت تصمیم به‌عنوان یکی از الگوریتم‌های پیشرفته یادگیری ماشین، در این پژوهش برای شناسایی الگوهای بارش ۲۴ ساعته در استان‌های ایران به کار گرفته شده است. نتایج این تحلیل (شکل ۵)، در قالب یک نمودار درختی Tree Structure نشان می‌دهد که ساختار خوشه‌بندی استان‌ها براساس میزان میانگین بارش را به تصویر می‌کشد. نمودار درختی به‌گونه‌ای طراحی شده است که متغیر وابسته، یعنی میزان بارش، در گره ریشه Root Node قرار دارد. این نمای ابتدایی حاکی از آن است که بارش به‌عنوان معیاری کلیدی در تعیین وضعیت ملزومات آب‌وهوایی استان‌ها مورد استفاده قرار می‌گیرد. در ادامه، خوشه‌های درخت به ترتیب میزان بارش، به خوشه‌بندی‌های مختلف تقسیم‌بندی می‌شوند که هر خوشه‌بندی نمایانگر گروهی از استان‌ها با ویژگی‌های مشابه در میزان بارش است. خوشه‌بندی ۱: در این سطح، استان‌های خوزستان، آذربایجان شرقی، خراسان شمالی، بوشهر، خراسان رضوی و البرز گروه‌بندی شده‌اند. این گروه از استان‌ها به دلیل شرایط جغرافیایی و اقلیمی خاص، بارش‌های نسبتاً بالاتری را تجربه می‌کنند. خوشه‌بندی ۲: استان‌های فارس، مرکزی، اردبیل، زنجان، قزوین و همدان در این گروه قرار دارند. بارش در این گره نشان‌دهنده ویژگی‌های بارشی متفاوت نسبت به گروه اول است، که ممکن است به عوامل جغرافیایی و ارتفاعی برگردد. خوشه‌بندی ۳: در این سطح، استان‌های تهران، آذربایجان غربی، کرمانشاه، لرستان و کردستان گروه‌بندی شده‌اند. این گروه از استان‌ها به‌طور خاص، خصوصیات اقلیمی خاصی دارند که می‌تواند بر روی بارش‌های فصلی تأثیرگذار باشد. خوشه‌بندی ۴: اصفهان،

سیستان و بلوچستان و یزد در این گروه قرار دارند. میزان بارش در این گروه به‌طور قابل توجهی پایین‌تر است و این امر ناشی از اقلیم خشک و نیمه‌خشک این استان‌هاست. خوشه‌بندی ۵: در این گروه، استان‌های کرمان، خراسان جنوبی، هرمزگان، سمنان و قم قرار گرفته‌اند. بارش‌های نسبت به سایر استان‌ها نسبتاً کمتر است و به نظر می‌رسد تحت‌تأثیر شرایط اقلیمی بیابانی و نیمه‌بیابانی قرار دارند. خوشه‌بندی ۶: ایلام، چهارمحال و بختیاری و گلستان در این سطح دسته‌بندی شده‌اند. این استان‌ها نشان‌دهنده تنوع بارشی در نواحی مرتفع و کوهستانی کشور هستند. خوشه‌بندی ۷: در نهایت، مازندران، گیلان و کهگیلویه و بویراحمد در این گروه قرار دارند. این استان‌ها به‌دلیل وجود جنگل‌ها و کوهستان‌ها، بارش‌های فراوانی را به خود اختصاص داده‌اند. این تحلیل درخت تصمیم به ما این امکان را می‌دهد تا الگوهای بارش در استان‌های مختلف ایران را شناسایی کرده و با استفاده از آن‌ها در مدیریت منابع آب و پیش‌بینی وضعیت بارش‌های آینده بهینه عمل نماییم. نتایج به‌دست‌آمده در این کار، می‌تواند به‌عنوان راهنمایی برای تصمیم‌گیرندگان در زمینه مدیریت بحران‌های آبی و تغییرات اقلیمی مورد استفاده قرار گیرد.



شکل ۵. نمودار درخت تصمیم خوشه‌بندی استان‌های ایران براساس بارش تجمعی ۲۴ ساعته

نتایج تابع رگرسیون خطی

بخش اول: تحلیل رگرسیون خطی (جدول ۳ و شکل ۶)، تحلیل داده‌ها با استفاده از مدل رگرسیون خطی نشان داد که مقدار ضریب تعیین تعدیل‌شده Adjusted R² برابر با ۰/۲۰۹ برآورد شده است. این شاخص به‌روشنی نشان می‌دهد که متغیر پیش‌بین زمانی (ماه‌ها) توانسته است ۲۰/۹ درصد از واریانس متغیر ملاک (مقدار بارش) را تبیین نماید. به عبارت دیگر، ۲۰ درصد از تغییرات مشاهده‌شده در بارش‌ها ناشی از تأثیرات متغیر زمان است. این نتایج به‌ویژه تأکید دارند بر اینکه تغییرات زمانی در سطوح ماهانه به‌عنوان یک عامل کلیدی در روندهای بارش شناخته می‌شوند و همچنین بیانگر اهمیت درک الگوهای بارشی در سطوح ماهانه به‌منظور پیش‌بینی و مدیریت منابع آبی است. بخش دوم: تحلیل واریانس (جدول ۴ و شکل ۶)، نتایج تحلیل واریانس به‌منظور ارزیابی معنی‌داری کلی مدل رگرسیونی به‌دست‌آمده، حاکی از آن است که سطح معنی‌داری p-value کمتر از ۰/۰۵ است. این یافته‌ها نشان می‌دهند که مدل رگرسیونی به‌لحاظ آماری معنی‌دار است و متغیر پیش‌بین مستقل (ماه‌ها) به‌طور معناداری توانسته است تغییرات در مقدار بارش را پیش‌بینی کند. معنی‌داری آماری آزمون تحلیل واریانس F-test نشانه‌ای از وجود یک رابطه معنادار میان متغیر پیش‌بین (ماه‌ها) و متغیر ملاک (بارش) است و تأکید بر این نکته دارد که تغییرات فصلی و ماهانه تأثیر مهمی بر الگوهای بارش در ایران دارند. این نتایج نه‌تنها در جهت درک بهتر الگوهای بارشی و نوسانات آن در سطح ماهانه کمک می‌کنند، بلکه می‌توانند مبنای علمی و عملی برای توسعه راهبردهای مدیریت منابع آب و پیش‌بینی خشکسالی و ترسالی در آینده باشند.

جدول ۳. نتایج رگرسیون خطی برای بخش اول تابع، با متغیر وابسته بارش تجمعی ۲۴ ساعته در ایران

متغیر وابسته	متغیر مستقل	ضریب تعیین تعدیل شده	سطح معنی داری
بارش ثبت‌شده در ایران	سری‌های زمانی ماه‌ها	۰/۲۰۹	۰/۰۰۰

جدول ۴. نتایج رگرسیون خطی برای بخش دوم تابع، با متغیر وابسته بارش تجمعی ۲۴ ساعته در ایران

متغیر وابسته	درجه آزادی	آماره F	سطح معنی داری
بارش ثبت شده در ایران	۱۱	۳۱/۲۴	۰/۰۰۰

راهنما پنجره ها بسته های محاسباتی ویرایش پرونده

```

*** Linear Model ***

Call: lm(formula = RRR24 ~ DATE., data = SDF23, na.action = na.exclude)
Residuals:
    Min       1Q   Median       3Q      Max
-40.23 -15.87  -5.086   3.996   270

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  36.2857   3.0855   11.7601  0.0000
DATE.AUG    -30.9394   4.3635   -7.0904  0.0000
DATE.DEC     6.9940   4.3635    1.6028  0.1092
DATE.FEB     5.0238   4.3635    1.1513  0.2498
DATE.JAN     9.5950   4.3635    2.1989  0.0281
DATE.JUL    -31.2983   4.3635   -7.1727  0.0000
DATE.JUN    -30.0794   4.3635   -6.8934  0.0000
DATE.MAR    11.4104   4.3635    2.6149  0.0090
DATE.MAY   -18.1377   4.3635   -4.1567  0.0000
DATE.NOV     5.7147   4.3635    1.3096  0.1905
DATE.OCT    -11.1166   4.3635   -2.5476  0.0110
DATE.SEP    -24.7623   4.3635   -5.6748  0.0000

Residual standard error: 32.21 on 1296 degrees of freedom
Multiple R-Squared: 0.2096
F-statistic: 31.24 on 11 and 1296 degrees of freedom, the p-value is 0

Analysis of Variance Table

Response: RRR24

Terms added sequentially (first to last)
      Df Sum of Sq Mean Sq F Value Pr(>F)
DATE.  11   356611  32419.15  31.24134    0
Residuals 1296  1344860  1037.70

```

شکل ۶. نتایج حاصل از تحلیل رگرسیون خطی بر روی بخش های اول و دوم تابع، با در نظر گرفتن بارش تجمعی ۲۴ ساعته در ایران به عنوان متغیر وابسته (راهنمای دسترسی به توابع در زبان برنامه نویسی، R)

نتیجه گیری

تحلیل های جامع انجام شده بر روی داده های بارش با بهره گیری از روش های پیشرفته آماری و یادگیری ماشین، دیدگاه های ارزشمندی را در مورد الگوهای بارش در ایران ارائه داد. کاربرد موفقیت آمیز مدل MLP در پیش بینی دقیق بارش ماهانه، با خطای میانگین مربعات 0.04 (MSE)، نشان دهنده توان بالای شبکه های عصبی مصنوعی در مدل سازی پدیده های پیچیده هیدرولوژیکی است. این یافته ها اهمیت استفاده از رویکردهای نوین را در بهبود دقت پیش بینی بارش و مدیریت منابع آب تأیید می کند. همچنین، نتایج تحلیل واریانس یک طرفه تفاوت های آماری معناداری را بین میانگین بارش ایستگاه های مختلف در ایران آشکار ساخت که بر تأثیرگذاری عوامل جغرافیایی و اقلیمی متنوع در پراکنش بارش تأکید دارد. تحلیل کلاسترینگ و درخت تصمیم نیز به شناسایی الگوهای منطقه ای و گروه بندی استان ها براساس ویژگی های بارشی مشابه کمک شایانی کرد. این خوشه بندی ها ابزار مؤثری برای برنامه ریزی هدفمند منابع آب و مدیریت ریسک های مرتبط با بارش های شدید فراهم می آورد. در نهایت، تحلیل رگرسیون خطی با ضریب تعیین تعدیل شده 0.209 (R^2 Adjusted)، نقش زمان را در تبیین واریانس بارش برجسته ساخت و نشان داد که تغییرات فصلی و ماهانه تأثیر قابل توجهی بر الگوهای بارش دارد. به طور کلی، یافته های این مطالعه نه تنها دانش موجود در زمینه مدل سازی و پیش بینی بارش را تعمیق می بخشند، بلکه چارچوبی جامع برای درک الگوهای پیچیده بارش در ایران فراهم می آورند. این بینش ها می توانند مبنایی علمی برای سیاست گذاری های آتی در حوزه مدیریت پایدار منابع آب، کاهش مخاطرات ناشی از تغییرات اقلیمی، و ارتقای تاب آوری سیستم های هیدرولوژیک کشور قرار گیرند.

References

- Akbari, M., & Sayad, V. (2021). Analysis of climate change studies in Iran. *Journal of Natural Geography Research*, 53(1), 37–74. (in Persian)
- Asakereh, H., & Mottalebizad, S. (2017). Comparison of SDSM and artificial neural networks performance in predicting minimum temperature changes (Case study: Urmia station). *Spatial Planning*, 21(4), 140–160. (in Persian)
- Asakereh, H., Masoudian, S. A., & Torkarani, F. (2021). Disentangling the role of internal and external factors in the decadal variability of annual rainfall in Iran during the last four decades (1976–2015). *Journal of Natural Geography Research*, 53(1), 91–107. <https://doi.org/10.22059/jphgr.2021.304776.1007529>
- Babaei Hesar, S., & Ghazavi, R. (2015). Comparison of time series and neural network models with emission scenario results in rainfall prediction. *Water and Soil*, 29(4), 943–953. (in Persian)
- Bahrani, M., Salari, A., Amiri, M. J., & Bahrani, A. (2023). Performance evaluation of artificial neural network in estimating rainfall using climatic and geographic data (Case study: Fars province). *Irrigation and Water Engineering*, 13(3), 121–140. <https://doi.org/10.22125/iwe.2023.168171>
- Cybenko, G. (1989). Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2, 183–192. <https://doi.org/10.1080/19373260802659226>
- Dastorani, M. T., & Afkhami, H. (2011). Application of artificial neural networks on drought prediction in Yazd (Central Iran). *Desert*, 16(1), 39–48. (in Persian)
- Dehghani Navid, V., Vafakhah, M., & Bahrehmand, A. R. (2016). Rainfall-runoff modeling using artificial neural network and adaptive neuro-fuzzy network in Kasilian watershed. *Watershed Management Research Journal*, 7(13), 128–137. (in Persian)
- Esmailian, M. (2005). *Comprehensive SPSS Guide*. Naghoos Publications. (in Persian)
- Esmailian, M., Vahdat, J., & Heidardoust, H. (2016). *R software guide*. University of Mohaghegh Ardabili Publications. (in Persian)
- Fallah Ghalhari, G., & Shakeri, F. (2015). Application of artificial neural networks in winter rainfall forecasting. *Iranian Journal of Watershed Management Science & Engineering*, 9(31), 98–110. (in Persian)
- Faraji, M., Rezaei Banafsheh, M., Sari Sarraf, B., & Khorshiddoost, A. (2023). Numerical simulation of climate change in Iran using artificial neural network algorithm. *Climate Change Researches*, 4(14), 43–64. (in Persian)
- Faraji, M., Rezaei Banafsheh, M., Sari Sarraf, B., & Khorshiddoost, A. (2024). Data mining of 24-hour air temperature in Iran using machine learning multilayer perceptron neural network. *Climate Change Researches*, 5(20), 33–48. <https://doi.org/10.30488/ccr.2024.458168.1216>
- Ghermezchashme, B., Rasouli, A. A., Rezaei Banafsheh, M., Massah Bavani, A. R., & Khorshiddoost, A. M. (2015). Uncertainty assessment of neural network model in downscaling HadCM3 using bootstrap confidence interval method. *Journal of Watershed Engineering and Management*, No. 3, 306–316. (in Persian)
- Gholizadeh, M. H., & Darand, M. (2011). Monthly rainfall forecasting using artificial neural networks (Case study: Tehran). *Journal of Natural Geography Research*, 42(7), 51–63. (in Persian)
- Jahanbakhsh Asl, S., Sari Sarraf, B., Asakereh, H., & Shirmohammadi, S. (2020). Analysis of temporal-spatial changes of critical rainfalls (extreme high) in the west of Iran during the years 1965–2016. *Journal of Environmental Hazards Analysis*, 7(1), 89–106. (in Persian)
- Moon, S. H., Kim, Y. H., Lee, Y. H., & Moon, B. R. (2019). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*, 568, 1042–1054. <https://doi.org/10.1016/j.jhydrol.2018.11.060>
- Pakdaman, M. (2022). The effect of training algorithm type of multilayer perceptron neural network on the accuracy of monthly rainfall forecast of Iran, case study: ECMWF model. *Journal of Earth and Space Physics*, 48(1), 213–226. <https://dor.isc.ac/dor/20.1001.1.2538371.1401.48.1.14.2> (in Persian)
- *R Programming Language. Available functions guide.
- S Programming Language. Available functions guide.

- Sadatinejad, S. J., Soleimani Sardo, F., & Mirzavand, M. (1403). Modeling and Predicting Climatic Parameters Using the CanESM2 Model Under RCP Scenarios (Case Study: Karaj Station). *Ecohydrology Journal*, 11(3), 411-426. doi: 10.22059/ije.2024.382370.1845 (in Persian).
- Shahbahi Kotnaei, A., & Asakereh, H. (2023). Evaluation of fuzzy clustering and artificial neural network methods in spatial zoning of annual rainfall in Iran. *Journal of Water and Soil Science*, 27(1), 17-32. (in Persian)
- Shahgholian, K., Bazrafshan, J., & Irannejad, P. (2023). Synoptic weather variables and data mining methods for predicting regional heavy precipitation over the southwest of Iran. *Theoretical and Applied Climatology*, 147, 401-416. <https://link.springer.com/article/10.1007/s12040-021-01725-9>
- Subrahmanyam, K. V., Ramsenthil, C., Girach Imran, A., Chakravorty, A., Sreedhar, R., Ezhilrajan, E., & Jha, C. S. (2021). Prediction of heavy rainfall days over a peninsular Indian station using the machine learning algorithms. *Journal of Earth System Science*, No. 130, 1-9. <https://doi.org/10.21203/rs.3.rs-3768340/v>
- Tajbar, M., Khoureshid Doost, S., Saeed Jahanbakhsh Asl, S. (2022). Application of Artificial Intelligence Approach in Studying the Impact of Large-Scale Climatic Drivers on Rainfall in Balochistan, Pakistan. *Journal of Geography and Environmental Planning*, 33(87), 1-20 (in Persian).