



Investigating Capabilities of Machine Learning Techniques in Forecasting Daily Using Some Meteorological Data and Normalized Difference Snow Streamflow Index (Case Study: Latian and Navroud Basins)

Mahboobeh Fallah¹, Hosseinali Bahrami², Hossein Asadi³

1. Department of Soil Science, Faculty of Agriculture, University of Tarbiat Modares, Tehran, Iran, Email:

mahboobeh.fallah@modares.ac.ir

2. Corresponding Author, Department of Soil Science, Faculty of Agriculture, University of Tarbiat Modares, Tehran, Iran,

Email: bahramih@modares.ac.ir

3. Department of Soil Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, Email:

ho.asadi@ut.ac.ir

Article Info	ABSTRACT
Article type: Research Article	Accurate prediction of streamflow is crucial for water resources management, especially for the prediction of floods and soil erosion. In the current study, the capability of three machine learning (ML) methods, including Support Vector Regression (SVR), Artificial Neural Network with Backpropagation (ANN-BP), and Gradient Boosting Regression (GBR) was investigated using meteorological observations and MODIS snow cover data to forecast daily streamflow in two different basins, namely Latian and Navroud. For model development, four major predictors, including daily rainfall (P), maximum temperature (T_{max}), minimum temperature (T_{min}), and the Normalized Difference Snow Index (NDSI) from the MODIS satellite were used from 2000 to 2018. The performance of these models was evaluated using statistical indices. Simulation results revealed that all models presented satisfactory results in simulating daily streamflow using meteorological predictors, and the efficiency of all applied models was improved when the NDSI index was applied as an additional predictor. The best performance was observed in GBR among all studied models in both basins, whereas SVR revealed the lowest performance in forecasting streamflow for both validation and calibration steps in most cases. In general, the simulation results demonstrated higher accuracy in Latian basin than Navroud basin in both calibration and validation steps. The best performance among all models was observed using GBR with $R = 0.85$, $NS=0.72$, and $RMSE = 3.43 \text{ m}^3/\text{s}$ using the NDSI index in Latian basin indicating the significant effect of snowmelt on streamflow generation in snowmelt-dominated regions.
Article history:	
Received: Feb. 11, 2022	
Received: May. 28, 2022	
Accepted: May. 29, 2022	
Published online: July. 23, 2022	
Keywords: Artificial Neural Networks, Machine Learning Model, Gradient Boosting Regression, Support Vector Regression.	

Cite this article: Fallah, M., Bahrami, H. A., & Asadi, H. (2022). Investigating Capabilities of Machine Learning Techniques in Forecasting Daily Streamflow Using Some Meteorological Data and Normalized Difference Snow Index (Case Study: Latian and Navroud Basins). *Iranian Journal of Soil and Water Research*, 53 (5), 1127-1144.

© The Author(s).

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/ijswr.2022.338986.669207>



بررسی قابلیت‌های رویکرد یادگیری ماشین در پیش‌بینی جریان سطحی روزانه با استفاده از برخی داده‌های هواشناسی و شاخص تفاضلی نرمال شده برف (مطالعه موردی: حوضه آبخیز لتیان و ناورود)

محبوبه فلاح^۱، حسینعلی بهرامی^۲، حسین اسدی^۳

۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران. ایمیل: mahboobeh.fallah@modares.ac.ir

۲. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران. ایمیل: bahramih@modares.ac.ir

۳. گروه علوم و مهندسی خاک، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، تهران، ایران. ایمیل: ho.asadi@ut.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	پیش‌بینی دقیق جریان سطحی برای مدیریت منابع آب به ویژه پیش‌بینی سیل و فرسایش خاک بسیار مهم است. در مطالعه حاضر، قابلیت سه روش یادگیری ماشین (ML) شامل رگرسیون بردار پشتیبان (SVR)، شبکه عصبی مصنوعی با پس انتشار خطا (ANN-BP) و رگرسیون تقویت گرادیان (GBR) با استفاده از داده‌های هواشناسی و پوشش برف سنجنده MODIS برای پیش‌بینی جریان سطحی روزانه در دو حوضه مختلف لتیان و ناورود بررسی شد. برای توسعه مدل، چهار متغیر اصلی شامل باران روزانه (P)، دمای حداکثر (T_{max})، دمای حداقل (T_{min}) و شاخص تفاضلی نرمال شده برف (NDSI) از سنجنده MODIS در طول سال‌های ۱۳۷۹-۱۳۹۷ استفاده شد. کارایی این مدل‌ها با استفاده از شاخص‌های آماری مورد ارزیابی قرار گرفت. نتایج شبیه‌سازی نشان داد که همه مدل‌ها نتایج رضایت‌بخشی را در شبیه‌سازی جریان سطحی روزانه با استفاده از متغیرهای هواشناسی به عنوان پارامترهای ورودی مدل‌ها ارائه کردند. همچنین، کارایی همه مدل‌های ML مورد مطالعه، زمانی که شاخص NDSI به عنوان متغیر تخمین‌گر در شبیه‌سازی اعمال شد، بهبود یافت. بهترین کارایی را در بین تمام مدل‌های مورد مطالعه در هر دو حوضه، مدل GBR نشان داد. مدل SVR پایین‌ترین کارایی را در پیش‌بینی جریان سطحی روزانه برای هر دو مرحله آموزش و اعتبارسنجی در اکثر موارد نشان داد. به‌طور کلی، نتایج شبیه‌سازی در حوضه لتیان نسبت به حوضه ناورود در هر دو مرحله آموزش و اعتبارسنجی بهتر بود و نسبت به دو مدل دیگر، بهترین کارایی در مدل GBR با ضریب همبستگی ($R=0/85$)، ضریب کارایی نش-ساتکلیف ($NS=0/72$) و جذر میانگین مربعات خطا ($RMSE=3/43 m^3/s$) با استفاده از شاخص NDSI در حوضه لتیان مشاهده شده است که نشان‌دهنده تاثیر زیاد ذوب برف در ایجاد جریان سطحی در مناطق برف‌خیز است.
تاریخ دریافت: ۱۴۰۰/۱۱/۲۲	
تاریخ بازنگری: ۱۴۰۱/۳/۷	
تاریخ پذیرش: ۱۴۰۱/۳/۸	
تاریخ انتشار: ۱۴۰۱/۵/۱	
واژه‌های کلیدی:	
شبکه‌های عصبی مصنوعی،	
مدل یادگیری ماشین،	
رگرسیون بردار پشتیبان،	
رگرسیون تقویت گرادیان.	

استناد: فلاح، محبوبه؛ بهرامی، حسینعلی؛ اسدی، حسین (۱۴۰۱). بررسی قابلیت‌های رویکرد یادگیری ماشین در پیش‌بینی جریان سطحی روزانه با استفاده از برخی داده‌های هواشناسی و شاخص تفاضلی نرمال شده برف (مطالعه موردی: حوضه آبخیز لتیان و ناورود). *مجله تحقیقات آب و خاک ایران*، ۵۳ (۵)، ۱۱۲۷-۱۱۴۴.

DOI: <http://doi.org/10.22059/ijswr.2022.338986.669207>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

مقدمه

پیش‌بینی دقیق بلندمدت جریان سطحی و رواناب در تحقیقات هیدرولوژی به عنوان یک چالش مهم برای دانشمندان و مهندسان در دهه اخیر مورد توجه قرار گرفته است که برای مدیریت منابع آب به ویژه برای پیش‌بینی سیل و فرسایش خاک بسیار مهم است. شبیه‌سازی دقیق دبی جریان به دلیل کمبود داده، پیچیدگی، غیرخطی بودن و ویژگی‌های تصادفی فرآیندهای هیدرولوژیکی، یک کار دشوار و چالش برانگیز است (Adnan et al., 2019; Meng et al., 2019; Niu et al., 2019). به دلیل اهمیت حیاتی پیش‌بینی رواناب و جریان سطحی خروجی از حوضه برای مدیریت منابع آب و کاهش حوادث شدید، روش‌های مختلفی برای شبیه‌سازی به صورت ماهانه و روزانه به طور گسترده پیشنهاد شده است (Wang et al., 2019; Adnan et al., 2019). این روش‌ها را می‌توان به روش‌های پیش‌بینی کوتاه مدت که با زمان‌های پیش‌بینی ساعتی (Xiang and Demir, 2020; Xiang et al., 2020) تا روزانه (Cheng et al., 2020) و روش‌های پیش‌بینی بلندمدت که با مقیاس هفتگی (Kashid et al., 2020)، ماهانه (Keteklahijani et al., 2020; Cheng et al., 2020) و حتی سالانه (Ni et al., 2010) سروکار دارند، تقسیم کرد. با این حال، کمبود داده‌های اقلیمی و مسائلی همچون غیرقابل دسترس بودن داده در برخی مناطق، به عنوان ممانعی برای تحقیقات ارزیابی یکپارچه مؤثر، توسعه مدل‌سازی و شبیه‌سازی مناسب و همچنین توصیه سیاست و روش‌های انعطاف‌پذیر، به ویژه برای کشورهای در حال توسعه عمل می‌کند (He et al., 2021).

در دهه‌های اخیر، تصاویر و داده‌های ماهواره‌ای به همراه سایر فناوری‌ها به طور گسترده‌ای برای جبران کمبودهای ایستگاه‌های هواشناسی مورد استفاده قرار گرفته‌اند. به عنوان مثال، داده‌های کمی بارش تولید شده توسط سنجش از دور طیف وسیعی از مشاهدات را با وضوح مکانی و زمانی بالا پوشش می‌دهد که می‌تواند به عنوان منبع خوبی از داده‌ها برای اهداف شبیه‌سازی و مدل‌سازی استفاده شود. برف به عنوان یکی از شکل‌های بارش، نقش مهمی در چرخه هیدرولوژی و متعاقباً پیش‌بینی و شبیه‌سازی دبی جریان خروجی حوضه به ویژه در مناطق کوهستانی ایفا می‌کند. از آنجایی که تعیین و تقسیم انواع بارش مانند باران و برف تأثیر قابل توجهی بر مدل‌سازی و شبیه‌سازی هیدرولوژی دارد، بسیاری از محققین با استفاده از سیستم اطلاعات جغرافیایی^۱ (GIS) و تصاویر ماهواره‌ای مانند طیف‌سنج تصویربرداری با وضوح متوسط^۲ (MODIS) بر اساس شاخص‌هایی مانند شاخص تفاضلی نرمال شده برف^۳ (NDSI) به عنوان یک شاخص عددی که پوشش برف را بر روی اراضی نشان می‌دهد، برای مدل‌سازی جریان سطحی حاصل از ذوب برف استفاده کردند (Steele et al., 2017; Shahabi et al., 2014; Bennett et al., 2019).

روش‌های شبیه‌سازی جریان سطحی را نیز می‌توان با توجه به نوع مدل مورد استفاده به دو گروه دسته‌بندی کرد: مدل‌های هیدرولوژیکی پایه فیزیکی و مدل‌های داده محور که بر اساس تجزیه و تحلیل داده‌های مشاهداتی هستند (Ha et al., 2021). مدل‌های هیدرولوژیکی پایه فیزیکی قادر به شبیه‌سازی فرآیندهای هیدرولوژیکی حوضه با جزئیات زیاد می‌باشند. با این حال برای استفاده از این گونه مدل‌ها، حجم بسیاری از داده‌های مشاهداتی مورد نیاز است. به همین دلیل، با وجود دقت خوب این مدل‌ها در شبیه‌سازی جریان خروجی از حوضه‌ها، کاربرد آنها در بسیاری از مناطق به دلیل فقدان داده‌های هیدرولوژیکی دقیق محدود شده است (Cheng et al., 2006). در عوض، مدل‌های داده محور می‌تواند جایگزینی برای مدل‌های ذکر شده باشد، زیرا به داده‌های کمتری برای دستیابی به کارایی بهتر در مقایسه با روش‌های پایه فیزیکی نیاز دارند (Liu et al., 2017).

مدل‌های یادگیری ماشین^۴ (ML) که به عنوان مدل‌های داده محور شناخته می‌شوند، می‌توانند با حداقل داده به نتایج قابل قبولی، دست یابند (Xiang et al., 2020). چندین کاربرد موفقیت‌آمیز از مدل‌های ML برای شبیه‌سازی و پیش‌بینی جریان سطحی، سیل و رواناب خروجی از حوضه در دهه گذشته گزارش شده است (Konapala et al., 2020; Malik et al., 2020a; Tongal and Boojiz, 2018). این مدل‌ها از نظر محاسباتی سریع، دقیق و قابل اعتماد بوده و برای شبیه‌سازی هیدرولوژیکی بسیار مناسب هستند. در سال‌های اخیر از روش‌های داده‌کاوی مانند مدل‌های ML نه تنها برای پیش‌بینی دبی جریان خروجی حوضه، بلکه در چندین کاربرد غیرخطی و پیچیده مانند پیش‌بینی تابش خورشیدی (Prasad et al., 2020)، خشکسالی (Malik et al., 2020; Khan et al., 2020; Ali et al., 2019; Dikshit et al., 2020) (Ebrahimi and Rajaei, 2020b) et al., 2020b) مدل‌سازی باران-رواناب (Hosseini and Mahjouri, 2016)، پیش‌بینی سطح آب زیرزمینی (Mokari et al., 2021; Roy et al., 2020) (Ali et al., 2020; Yu et al., 2017) و تخمین تعلق-تبخیر (Ali et al., 2020; Yu et al., 2017) استفاده

^۱ Geographic Information System (GIS)

^۲ Moderate Resolution Imaging Spectroradiometer (MODIS)

^۳ Normalized Difference Snow Index (NDSI)

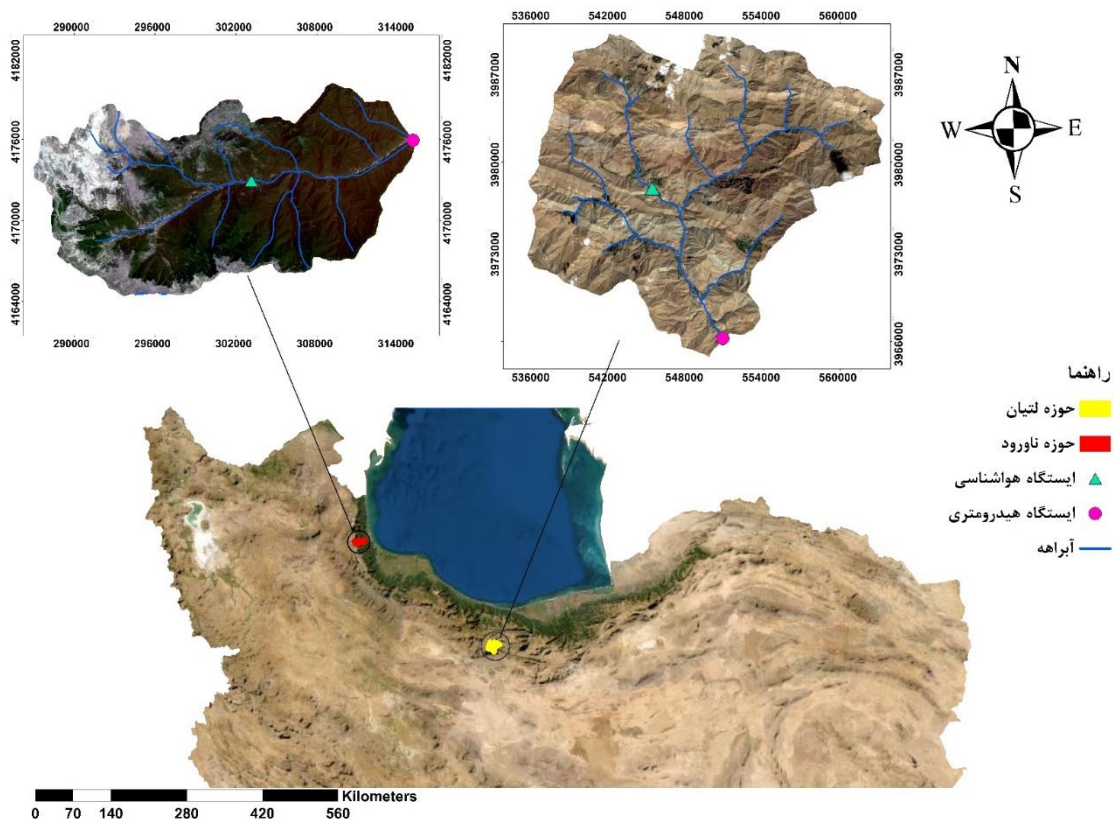
^۴ Machine Learning (ML)

شده است که کارایی و قابلیت مدل‌های ML را در زمینه‌های مختلف نشان می‌دهد. روابط پیچیده‌ای بین بارندگی-جریان سطحی در مناطق کوهستانی وجود دارد و همچنین، دستیابی به داده‌های دقیق و مورد نیاز در این مناطق دشوار است. از این رو، مدل‌های ML ممکن است یک راه جایگزین برای پیش‌بینی و شبیه‌سازی جریان سطحی در این مناطق، به دلیل حداقل داده‌های مورد نیاز و زمان توسعه و پردازش نسبتاً سریع باشد. با توجه به اهمیت شبیه‌سازی دقیق جریان خروجی حوضه آبخیز در مطالعات هیدرولوژی و با توجه به مطالعات بسیار اندک در کاربرد مدل‌های ML در مطالعات هیدرولوژی در ایران، به ویژه به همراه داده‌های تصاویر ماهواره‌ای و سنجش از دور، این پژوهش به بررسی قابلیت شبیه‌سازی دبی جریان خروجی توسط سه مدل پرکاربرد ML در دو حوضه آبخیز لتیان (البرز جنوبی در استان تهران) و ناورود (البرز شمالی در استان گیلان) پرداخته است. همچنین در این پژوهش با استفاده از سنجنده MODIS، سطوح پوشیده شده برف با استفاده از شاخص NDSI در طول دوره مطالعه مشخص گردید و به عنوان متغیر پیش‌بینی کننده تکمیلی برای بهبود قابلیت‌های پیش‌بینی این مدل‌ها به کار برده شد. با توجه به کوهستانی بودن حوضه‌های مورد مطالعه و اهمیت فراوان برف به عنوان عنصر مهم در شبیه‌سازی جریان سطحی خروجی حوضه، به خصوص در حوضه‌های کوهستانی، پژوهش حاضر تلاش نموده است تا علاوه بر نشان دادن تاثیر بارش برف در میزان دبی جریان خروجی حوضه به ویژه در حوضه‌های کوهستانی مورد مطالعه، مناسب‌ترین مدل برای شبیه‌سازی جریان خروجی در این حوضه‌ها را ارائه کند.

مواد و روش‌ها

مناطق مورد مطالعه

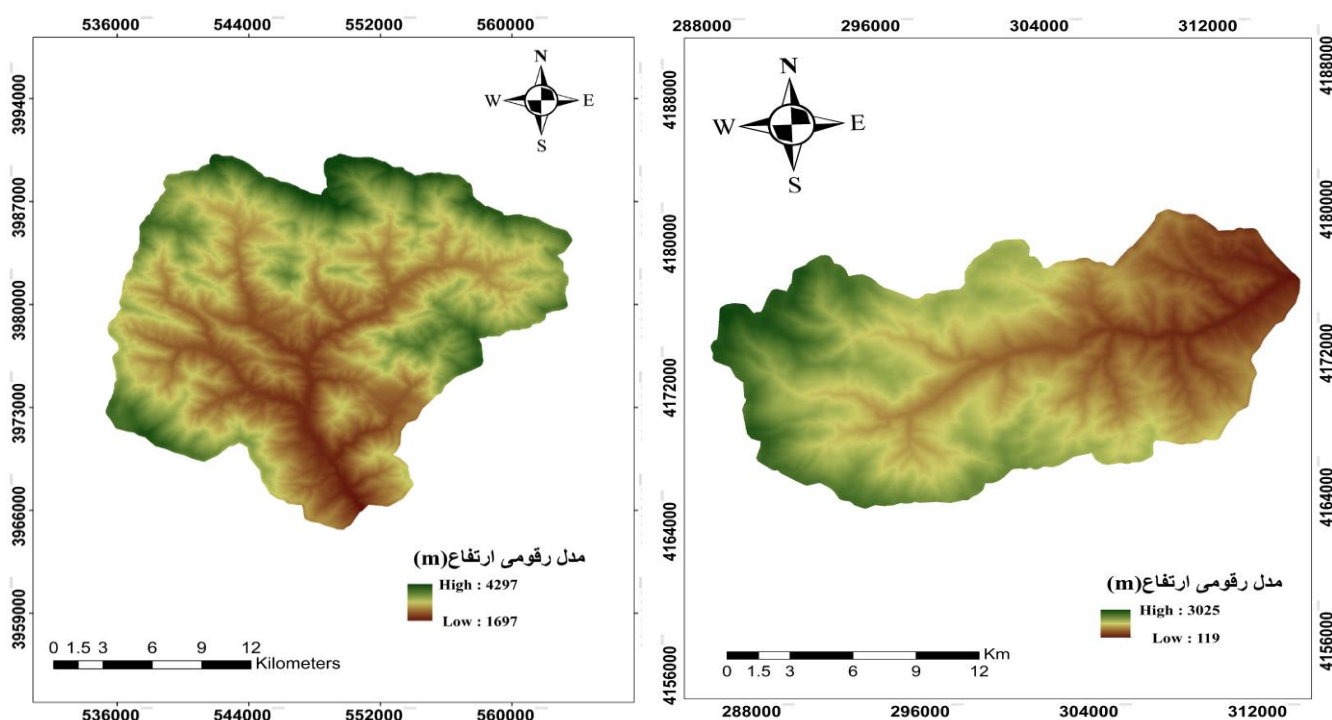
این پژوهش در دو حوضه آبخیز لتیان و ناورود با ویژگی‌های متفاوت از نظر مساحت، توپوگرافی و رژیم‌های هیدرولوژیکی-اقلیمی در ایران انجام شد. موقعیت مناطق مورد مطالعه در شکل ۱ ارائه شده است.



شکل ۱. موقعیت حوضه‌ها و ایستگاه‌های مورد مطالعه

حوضه لتیان در قسمت جنوبی رشته کوه البرز مرکزی در شمال استان تهران قرار دارد و رودخانه جاجرود مهم‌ترین رودخانه این حوضه است. شکل ۱ موقعیت حوضه‌های آبخیز لتیان و ناورود در ایران را با شبکه آبراهه‌ها، ایستگاه‌های هواشناسی و هیدرومتری مورد استفاده در این مطالعه را

نشان می‌دهد. این حوضه بین طول‌های جغرافیایی $51^{\circ} 19'$ تا $51^{\circ} 34'$ شرقی و عرض‌های جغرافیایی $35^{\circ} 42'$ تا $35^{\circ} 55'$ شمالی قرار دارد که تا ایستگاه هیدرومتری رودک در خروجی حوضه به مساحت حدود 440 کیلومتر مربع می‌باشد. حوضه‌های لتیان 25 درصد آب شرب پایتخت ایران را تامین می‌کنند (Mahjouri and Kerachian, 2011). همچنین این حوضه آب مورد نیاز کشاورزی و انرژی برق تهران و حومه را تامین می‌کند. نقشه‌های مدل رقمی ارتفاع^۱ (DEM) هر دو حوضه آبخیز با قدرت تفکیک $12/5$ متر با استفاده از تصویر سنجنده ALOS PALSAR تهیه شده است. حداقل و حداکثر ارتفاع این حوضه به ترتیب 1697 متر و 4297 متر می‌باشد. میانگین ارتفاع حوضه لتیان 2995 متر و میانگین شیب حوضه حدود 47 درصد است (شکل ۲). به طور کلی آب و هوای حوضه لتیان متأثر از آب و هوای البرز جنوبی می‌باشد. این حوضه کوهستانی است و بر اساس اقلیم نمای دومارتن دارای اقلیم نیمه خشک سرد، نیمه مرطوب سرد و اقلیم ارتفاعات کوهستانی است. نظام بارندگی در این حوضه، مدیترانه‌ای است. متوسط دمای سالانه حوضه لتیان 10 درجه سلسیوس و میانگین بارش سالانه این حوضه 340 میلی‌متر است. بیش‌ترین میزان بارش از دی تا اسفند رخ می‌دهد و نزدیک به 50 درصد بارش سالانه برف است که نشان‌دهنده برف‌خیز بودن این منطقه می‌باشد. در حوضه لتیان چهار نوع کاربری اصلی وجود دارد که شامل کشاورزی دیم و آبی، مرتع، جنگل، باغات می‌شود و مراتع در حدود 93 درصد کاربری‌ها را شامل می‌شود (Ahmadi et al., 2012). شکل ۳ نقشه‌های شاخص تفاضل نرمال شده پوشش گیاهی^۲ (NDVI) و پوشش گیاهی حوضه‌های لتیان و ناورود را نشان می‌دهد. بیشتر بخش‌های حوضه لتیان دارای پوشش گیاهی کم یا بدون پوشش گیاهی هستند (شکل ۳).



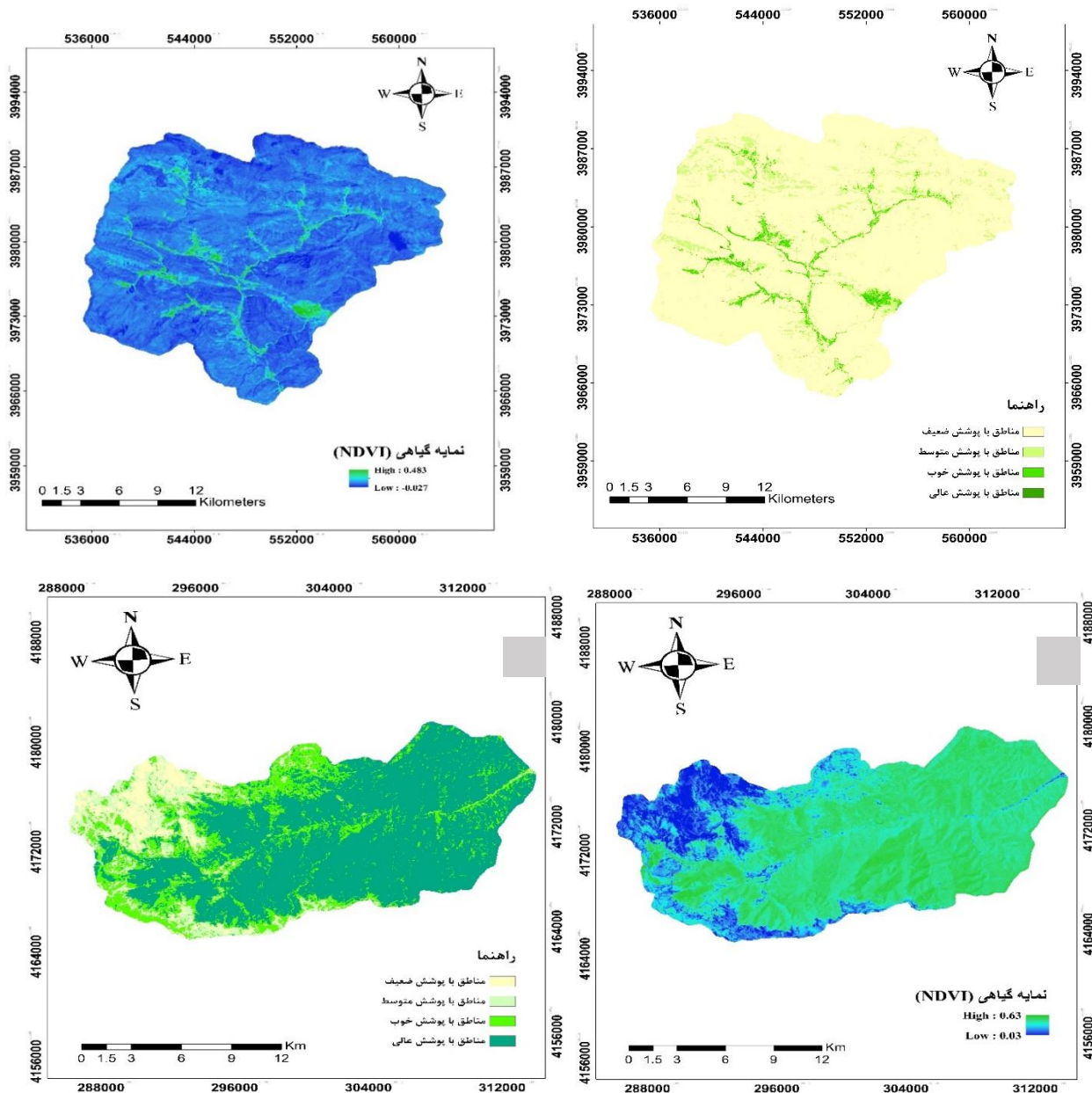
شکل ۲. نقشه DEM حوضه لتیان (چپ) و حوضه ناورود (راست) برحسب متر

حوضه آبخیز معرف ناورود یکی از حوضه‌های مهم غرب استان گیلان در محدوده شهرستان تالش است که این حوضه بین طول‌های جغرافیایی $48^{\circ} 35'$ تا $48^{\circ} 54'$ شرقی و عرض‌های جغرافیایی $37^{\circ} 36'$ تا $37^{\circ} 45'$ شمالی قرار دارد که تا ایستگاه هیدرومتری خرگیل در خروجی حوضه به مساحت حدود 290 کیلومتر مربع می‌باشد. موقعیت حوضه مورد مطالعه در شکل ۱ ارائه شده است. حداقل و حداکثر ارتفاع این حوضه به ترتیب 119 متر و 3025 متر می‌باشد. میانگین ارتفاع حوضه ناورود 1436 متر است (شکل ۲). این حوضه از دو تپ اراضی کوهستانی و تپ جلگه‌های فوقانی تشکیل شده است و به لحاظ پوشش گیاهی دارای پوشش گیاهی خوبی بوده و جزئی از جنگل‌های مرطوب ایران است و در ارتفاعات دارای پوشش مرتعی است که از 2000 متری فاقد جنگل بوده و برف‌گیر است. بیش از 80 درصد حوضه ناورود دارای پوشش گیاهی متوسط تا متراکم است که به دلیل بارش زیاد و شرایط آب و هوایی در این منطقه است و بر

^۱ Digital Elevation Model (DEM)

^۲ Normalized Difference Vegetation Index

اساس نقشه NDVI، اکثر نقاط این حوضه تراکم بالا و فراوانی پوشش گیاهی سبز را نشان می‌دهد (شکل ۳). حوضه آبخیز ناورود به دلیل قرار گرفتن در دامنه شمالی البرز و نیز موقعیت کوهستانی آن، دارای بارش‌های نسبتاً خوبی در طول سال می‌باشد. میانگین بارش سالانه در کل حوضه ۹۸۳ میلی‌متر بوده و اقلیم منطقه از روش دومارتن اصلاح شده در ارتفاعات، مرطوب سرد و در پایین دست و مناطق جلگه‌ای خیلی مرطوب است.



شکل ۳. نقشه‌های پوشش گیاهی و NDVI حوضه لتیان (الف، ب) و حوضه ناورود (ج، د)

متغیرهای ورودی استفاده شده در شبیه‌سازی مدل‌های ML

در تحقیق حاضر، دبی روزانه خروجی حوضه‌ها از سال ۱۳۷۹ تا ۱۳۹۷ از شرکت‌های آب منطقه‌ای استان تهران و گیلان اخذ شد. سه پارامتر اقلیمی از باران روزانه (P)، حداکثر دمای روزانه (T_{max}) و حداقل دمای روزانه (T_{min}) از دو ایستگاه هواشناسی در حوضه‌های لتیان و ناورود برای طول دوره مورد مطالعه به دست آمد. همچنین با توجه به اهمیت برف به عنوان یک منبع بسیار مهم و حیاتی برای تأمین آب و داشتن مزیت‌های فراوان این نوع از بارش نسبت به انواع بارش‌های شناخته شده در منطقه کوهستانی و نیز نقش آن در ایجاد جریان خروجی در حوضه‌های آبخیز به خصوص در مناطق کوهستانی، مناسب بودن محصولات پوشش برف MODIS به عنوان یک متغیر پیش‌بینی‌کننده تکمیلی برای بهبود قابلیت‌های شبیه‌سازی روش‌های ML، برای هر دو حوضه آبخیز مورد بررسی قرار گرفت. شاخص

NDSI به عنوان یک شاخص مفید برای تخمین پوشش برف از داده‌های بازتابی (Rakwatin et al., 2008) Aqua/Terra MODIS می‌باشد که به عنوان یک الگوریتم استخراج اتوماتیک برف به همراه مجموعه‌ای از آستانه‌ها به کار می‌رود. از این رو، برای این تحقیق، NDSI از تصاویر ماهواره ای MODIS از سال ۱۳۷۹ تا ۱۳۹۷ در نظر گرفته شد. NDSI شاخصی هست که از نسبت اختلاف بازتابندگی باند مرئی (band4)، باند ۴ سنجنده MODIS و فروسرخ میانی (band6) باند ۶ سنجنده MODIS تقسیم بر مجموع بازتابندگی در این دو باند به دست می‌آید (برای Aqua MODIS از باندهای ۴ و ۷ برای محاسبه NDSI استفاده می‌شود). شاخص NDSI با داده‌های MODIS بر اساس رابطه ۱ محاسبه شد (Hall et al., 1995).

$$\text{NDSI} = \frac{\text{MODISBand 4} - \text{MODISBand 6}}{\text{MODISBand 4} + \text{MODISBand 6}} \quad (\text{رابطه ۱})$$

شاخص NDVI از معروفترین، ساده‌ترین و کاربردی‌ترین شاخص‌هایی است که در زمینه مطالعات پوشش گیاهی شناخته شده است جهت تعیین فاکتور پوشش گیاهی در حوضه مورد مطالعه، نقشه NDVI حوضه با استفاده از تصویر سنجنده OLI ماهواره لندست ۸ بر اساس رابطه ۲ به دست آمد (Weier and Herring, 2011).

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} = \frac{\text{Band 5} - \text{Band 4}}{\text{Band 5} + \text{Band 4}} \quad (\text{رابطه ۲})$$

که در آن NIR: بازتاب در باند فروسرخ نزدیک و RED: انعکاس در باند سرخ است. همچنین تمامی نقشه‌های GIS در ArcMap 10.4 تولید شده است. در جدول شماره ۱، مشخصات محصولات ماهواره‌ای مورد استفاده ارائه شده است.

جدول ۱. مشخصات محصولات ماهواره‌ای مورد استفاده

نوع محصول	نوع سنجنده	قدرت تفکیک زمانی (روز)	قدرت تفکیک مکانی (متر)	زمان برداشت داده
نقشه رقمی ارتفاع (DEM)	ALOS PALSAR	۴۶	۱۲/۵	۲۰۱۰/۰۶/۲۸
نقشه پوشش گیاهی	ماهواره OLI	۱۶	۳۰	۲۰۱۰/۰۹/۲۶
NDVI و	LANDSAT 8			۲۰۱۶/۰۵/۲۴
نقشه پوشش برف (NDSI)	MODIS	روزانه	۵۰۰	۲۰۱۷/۰۸/۱۰

مدل‌های ML مورد مطالعه

رگرسیون بردار پشتیبان^۱ (SVR)

ماشین بردار پشتیبان^۲ (SVM) یکی از الگوریتم‌های نظارت شده یادگیری ماشین بر پایه تئوری یادگیری آماری است که Boser و همکاران در سال ۱۹۹۲ میلادی معرفی کردند (Boser et al., 1992). اولین کاربرد این مدل در مسائل مدل‌سازی بارش - رواناب توسط Dibike et al. (2001) ارائه شده است. در واقع مدل SVM، مدل کارآمدی است که به منظور رسیدن به یک جواب بهینه از اصل استقرایی کمینه‌سازی خطای ساختاری بهره می‌جوید (Bilandi et al., 2014). SVM یک الگوریتم نظارت شده است که همزمان خطاهای برآورد و پیچیدگی مدل را به حداقل می‌رساند (Benimam et al., 2020). SVR به صورت گسترده در مسائل رگرسیونی و پیش‌بینی کاربرد دارد.

شبکه عصبی مصنوعی با پس انتشار خطا^۳ (ANN-BP)

ANN-BP از ساده‌ترین و پر کاربردترین نوع شبکه عصبی است که در خصوص تخمین پارامترهای مجهول، مناسب ارزیابی شده است. این نوع شبکه دارای یک لایه ورودی (Input Layer)، یک یا چند لایه مخفی (Hidden Layer) و یک لایه خروجی (Output)

^۱ Support Vector Regression (SVR)

^۲ Support Vector Machine (SVM)

^۳ Artificial Neural Network with Backpropagation (ANN-BP)



Layer) است. پارامترهای ورودی در لایه اول و پارامترهای خروجی در لایه سوم قرار دارند. در هر لایه تعدادی گره^۱ (نرون) وجود دارد که از طریق وزن‌ها به گره‌های بعدی متصل می‌شوند. تعداد نرون‌ها در لایه ورودی وابسته به تعداد پارامترهای ورودی و تعداد نرون در لایه خروجی وابسته به تعداد پارامتر خروجی است. شمار لایه‌های مخفی و شمار گره‌ها در هر لایه مخفی به طور معمول به وسیله روش آزمون و خطا مشخص می‌شود و تابع ضابطه خاصی نیست (Garmdareh et al., 2019).

الگوریتم تقویت گرادیان^۲ (GBR)

GBR یک الگوریتم قدرتمند ML است که از درخت‌های تصمیم‌گیری استفاده می‌کند. در واقع، GBR یک مدل پیش‌بینی کننده قدرتمند با استفاده از ترکیبی از چندین مدل یادگیری ضعیف ایجاد می‌کند و به صورت تناوبی برای ایجاد یک مدل نهایی قوی ساخته شده است. این الگوریتم با آموزش یک درخت تصمیم‌گیری آغاز می‌شود که برای هر یک از مشاهدات یک وزن برابر با آن تعیین می‌شود. پس از ارزیابی درخت اول، وزن مشاهداتی که طبقه‌بندی آنها سخت است افزایش می‌یابد و وزن کمتر به مشاهداتی داده می‌شود که طبقه‌بندی آنها راحت‌تر است. بنابراین درخت دوم بر این داده‌های وزنی رشد می‌کند. هدف از این کار این است که پیش‌بینی درخت اول را بهبود بخشد. سپس خطای طبقه‌بندی درخت دوم را محاسبه کرد و یک درخت جدید را برای پیش‌بینی بازمانده‌های اصلاح شده آماده می‌شود (Tyrallis et al., 2021; Avand et al., 2020).

توسعه مدل و متغیرهای ورودی آن

در این پژوهش، همان‌طور که پیش‌تر ذکر شد از ۳ مدل SVR، ANN-BP و GBR برای شبیه‌سازی جریان سطحی ناشی از ذوب برف و باران به صورت روزانه در دو حوضه با شرایط مختلف آب و هوایی و کاربری اراضی استفاده شد. مقادیر اولیه متغیرهای ورودی باران روزانه (P)، حداکثر دما (T_{max}) و حداقل دما (T_{min}) بود. کارایی مدل‌ها در شبیه‌سازی و پیش‌بینی جریان سطحی خروجی از حوضه‌ها، با حضور و عدم حضور شاخص NDSI همراه با سه متغیر ذکر شده، بررسی شد. برای این منظور ابتدا مقادیر روزانه متغیرهای ورودی اصلی (P, T_{max}, T_{min}) به عنوان ورودی مدل (متغیر پیش‌بینی کننده اصلی) وارد شد و سپس، همان‌طور که قبلاً ذکر شد، برای بررسی مناسب بودن شاخص NDSI، به عنوان یک متغیر پیش‌بینی کننده مکمل برای بهبود قابلیت‌های شبیه‌سازی روش‌های ML ذکر شده، این شاخص به همراه سه متغیر ورودی اصلی دیگر استفاده شد. به این منظور، برای استخراج سطح پوشش برف در زمان‌های مختلف از شاخص NDSI از تصاویر سنجنده MODIS، (<https://lpdaacsvc.cr.usgs.gov/appears/>) با گام زمانی روزانه به مدت ۱۹ سال (۱۳۷۹ تا ۱۳۹۷) استفاده شد. برای آموزش مدل‌ها از داده‌های سال‌های ۱۳۷۹ تا ۱۳۹۰ و برای اعتبارسنجی از داده‌های ۱۳۹۰ تا ۱۳۹۷ استفاده شد. علاوه بر متغیرهای ورودی، از یک تأخیر ۱ تا ۱۰ روزه برای همه متغیرهای ورودی استفاده شد و سپس معنی‌دارترین ترکیب متغیرها انتخاب شد. در مطالعه حاضر، پارامترهای بهینه مدل‌های SVR، ANN-BP و GBR با استفاده از اعتبارسنجی متقاطع ۱۰ برابری^۳ با به حداقل رساندن RMSE برای دوره آموزشی تعیین شد. سپس از پارامترهای بهینه انتخاب شده هر مدل برای پیش‌بینی دبی جریان برای دوره اعتبارسنجی استفاده شد. متغیر خروجی تمام مدل‌های مورد بررسی دبی نرمال شده ایستگاه هیدرومتری خروجی حوضه‌ها (Q_t) می‌باشد. معادله (۳) و (۴) به ترتیب ساختار ورودی انتخاب شده برای حوضه لتیان را در عدم حضور شاخص NDSI و با استفاده از شاخص NDSI نشان می‌دهند. معادلات (۵) و (۶) به ترتیب ساختار ورودی انتخاب شده برای حوضه ناورود، مجدداً در عدم حضور شاخص NDSI و با استفاده از شاخص NDSI را نشان می‌دهند. پیاده‌سازی کلیه الگوریتم‌ها در این تحقیق در محیط نرم افزار Python 3.10.1 انجام شد.

$$Q_t = f(P_{t-10}, T_{max,t-1}, T_{max,t-7}, T_{max,t-8}, T_{max,t-9}, T_{max,t-10}, T_{min,t-1}, T_{min,t-2}, T_{min,t-7}, T_{min,t-8}, T_{min,t-9}, T_{min,t-10}, T_{ave,t-1}, T_{ave,t-2}, T_{ave,t-3}, T_{ave,t-4}, T_{ave,t-8}, T_{ave,t-9}, T_{ave,t-10}) \quad (۳)$$

$$Q_t = f(P_{t-10}, T_{max,t-1}, T_{max,t-2}, T_{max,t-7}, T_{min,t-8}, T_{min,t-9}, T_{min,t-10}, T_{ave,t-1}, T_{ave,t-2}, T_{ave,t-3}, T_{ave,t-4}, T_{ave,t-8}, T_{ave,t-9}, T_{ave,t-10}, M_{snow,t-1}, M_{snow,t-2}, M_{snow,t-3}, M_{snow,t-10}) \quad (۴)$$

$$Q_t = f(P_{t-1}, P_{t-2}, P_{t-4}, P_{t-5}, P_{t-7}, P_{t-8}, P_{t-9}, P_{t-10}, T_{max,t-4}, T_{max,t-5}, T_{max,t-6}, T_{max,t-7}, T_{max,t-8}, T_{max,t-9}, T_{max,t-10}, T_{min,t-3}, T_{min,t-9}, T_{min,t-10}) \quad (۵)$$

^۱ Neurons

^۲ Gradient Boosting Regression (GBR)

^۳ 10-Fold Cross-Validation

$$Q_t = f(P_{t-1}, P_{t-2}, P_{t-4}, P_{t-5}, P_{t-7}, P_{t-8}, P_{t-9}, P_{t-10}, Tmax_{t-4}, Tmax_{t-5}, Tmax_{t-6}, Tmax_{t-7}, Tmax_{t-8}, Tmax_{t-9}, Tmax_{t-10}, Tmin_{t-3}, Tmin_{t-9}, Tmin_{t-10}, Msnow_{t-1}, Msnow_{t-2}, Msnow_{t-6}, Msnow_{t-7}) \quad (۶)$$

جدول ۲. پارامترهای بهینه شده هر مدل ML برای حوضه‌های لتیان و ناورود در شرایط بدون برف و با برف: آستانه تحمل (ϵ)، پارامتر ساختاری

(γ) و پارامترهای پنالتی (C)

GBR	ANN-BP	SVR	حوضه	
learning_rate = 0.075 n_estimators = 80 min_samples_split = 60 min_samples_leaf = 10 max_depth = 10 max_features = 6	neurons=8	$\gamma = 0.001$ C = 10000 $\epsilon = 0.001$	لتیان	بدون داده برف
learning_rate = 0.075 n_estimators = 80 min_samples_split = 10 min_samples_leaf = 30 max_depth = 5 max_features = 3	neurons=5	$\gamma = 0.001$ C = 334.2 $\epsilon = 0.001$	ناورود	
learning_rate = 0.050 n_estimators = 80 min_samples_split = 10 min_samples_leaf = 10 max_depth = 10 max_features = 4	neurons=6	$\gamma = 0.001$ C = 10000 $\epsilon = 0.001$	لتیان	با داده برف
learning_rate = 0.075 n_estimators = 80 min_samples_split = 60 min_samples_leaf = 10 max_depth = 5 max_features = 8	neurons=1	$\gamma = 0.001$ C = 334.2 $\epsilon = 0.445$	ناورود	

ارزیابی کارایی مدل‌های ML

به منظور ارزیابی کارایی هر یک از الگوریتم‌های ML مورد استفاده، از شاخص‌های آماری مانند جذر میانگین مربعات خطا (RMSE)، ضریب همبستگی^۲ (R) و ضریب کارایی نش-سانتکلیف^۳ (NS) استفاده گردید. روابط این معیارها به شکل زیر هستند.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (۷)$$

$$R = \frac{(\sum_{i=1}^n O_i P_i - \frac{1}{n} \sum_{i=1}^n O_i \sum_{i=1}^n P_i)}{(\sum_{i=1}^n O_i^2 - \frac{1}{n} (\sum_{i=1}^n O_i)^2) (\sum_{i=1}^n P_i^2 - \frac{1}{n} (\sum_{i=1}^n P_i)^2)} \quad (۸)$$

$$NS = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \quad (۹)$$

در این روابط، P_i مقادیر پیش‌بینی شده، O_i مقادیر مشاهده شده، \bar{O}_i میانگین مقادیر مشاهده شده و n تعداد داده می‌باشد. هر چه مقدار RMSE کمتر و به صفر نزدیکتر، و مقادیر R و NS به یک نزدیک باشند، نشان‌دهنده نزدیکتر بودن مقادیر مشاهداتی و پیش‌بینی شده به یکدیگر و دقیق‌تر بودن شبیه‌سازی و کارایی بیشتر مدل‌ها در هر مرحله است.

۱ Root Mean Square Error (RMSE)

۲ Correlation Coefficient (R)

۳ Nash-Sutcliffe efficiency (NS)

نتایج و بحث

پارامترهای بهینه هر مدل در جدول ۲ نشان داده شده است. نتایج آماری همه مدل‌ها برای آموزش و اعتبارسنجی در مقیاس روزانه به ترتیب در جدول ۳ و جدول ۴ خلاصه شده است. در مرحله آموزش برای هر دو حوضه، مدل GRB کارایی بهتری نسبت به مدل‌های SVR و ANN-BP زمانی که هیچ داده پوشش برفی استفاده نشده، نشان داد (جدول ۳). در هر دو حوضه آبخیز ناورود و لتیان، ضعیف‌ترین کارایی مربوط به مدل SVR نسبت به دو مدل دیگر بود. بنابراین، در مرحله آموزش، دقت مدل‌ها بر اساس شاخص آماری RMSE، R و NS برای هر دو حوضه در عدم حضور شاخص NDSI به صورت $GBR > ANN-BP > SVR$ رتبه‌بندی شدند.

همچنین، نتایج جدول ۳ نشان داد که کارایی تمام مدل‌ها در مرحله آموزش با استفاده از داده‌های سنجش سطح برف از تصاویر ماهواره MODIS در مقایسه با شرایطی که هیچ داده پوشش برفی استفاده نشده بود، افزایش یافته است. مجدداً، مدل GBR از دو مدل دیگر در هر دو حوضه آبخیز لتیان و ناورود کارایی بهتری داشت، در حالی که ضعیف‌ترین دقت در مدل SVR ($RMSE=5/01 \text{ m}^3/s$) و ANN-BP ($NS=0/56$, $R=0/75$, $RMSE=17/25 \text{ m}^3/s$) و GBR ($NS=0/31$, $R=0/58$, $RMSE=17/25 \text{ m}^3/s$) به ترتیب در حوضه لتیان و ناورود دیده شد.

جدول ۳. نتایج مرحله آموزش شبیه‌سازی جریان سطحی در مقیاس روزانه با استفاده از مدل‌های ML مورد مطالعه

حوضه آبخیز مورد مطالعه	روش ML	بدون اطلاعات پوشش برف			با استفاده از اطلاعات پوشش برف		
		NS	R	RMSE	NS	R	RMSE
حوضه آبخیز لتیان	SVR	0/56	0/75	5/01	0/37	0/66	5/98
	ANN-BP	0/74	0/86	3/81	0/69	0/83	4/19
	GBR	0/94	0/97	1/85	0/88	0/94	2/61
حوضه آبخیز ناورود	SVR	0/34	0/60	17/1	0/26	0/54	19/46
	ANN-BP	0/31	0/58	17/25	0/49	0/68	15/21
	GBR	0/71	0/84	11/8	0/51	0/72	14/2

نتایج آماری همه مدل‌ها برای مرحله اعتبارسنجی در مقیاس روزانه در جدول ۴ ارائه شده است. با توجه به این جدول، GBR در بیش‌تر موارد، بهترین کارایی را در بین سه مدل برای مرحله اعتبارسنجی همانند مرحله آموزش در هر دو حوضه نشان داد. با این وجود، این مدل در حوضه آبخیز لتیان نسبت به حوضه ناورود دقت بهتری را نشان داد. همچنین، مدل SVR در مرحله اعتبارسنجی همانند مرحله آموزش، کم‌ترین کارایی و دقت را در حوضه آبخیز لتیان ($NS=0/48-0/36$, $R=0/72-0/64$, $RMSE=5/19-4/68 \text{ m}^3/s$) در هر دو شرایط با استفاده و یا بدون استفاده شاخص NDSI نشان داد (جدول ۴). در حالی که در حوضه ناورود شرایط بدین صورت نبود. همانطور که در جدول ۴ مشاهده می‌شود، همه مدل‌ها بهترین کارایی خود را در شبیه‌سازی دبی جریان با استفاده از شاخص NDSI در مقایسه با شرایطی که هیچ داده پوشش برفی در آن اعمال نشده بود، داشته‌اند. در حوضه آبخیز لتیان، کارایی بهتری ($RMSE=3/43 \text{ m}^3/s$)، نسبت به دو مدل دیگر هنگامی که از شاخص NDSI استفاده شده، از خود نشان داد.

این در حالیست که در حوضه ناورود در مرحله اعتبارسنجی، مدل GBR در مقایسه با سایر شرایط (به همراه و بدون داده‌های پوشش برف) الگوی معکوس از خود نشان داد و کمترین کارایی خود ($NS=0/42$, $R=0/67$, $RMSE=19/80 \text{ m}^3/s$) را در شبیه‌سازی جریان سطحی در این حوضه با استفاده از شاخص NDSI نشان داد. در حوضه لتیان در مرحله اعتبارسنجی همانند مرحله آموزش، زمانی که از NDSI به همراه سایر متغیرهای اصلی برای پیش‌بینی جریان سطحی استفاده می‌شود مدل SVR پایین‌ترین کارایی ($RMSE=4/68 \text{ m}^3/s$) را در شبیه‌سازی داشت. همه مدل‌ها بهترین کارایی خود را در حوضه لتیان در مقایسه با حوضه ناورود داشتند (جدول ۳ و جدول ۴). علاوه بر این، بهبود کارایی در شبیه‌سازی جریان سطحی پس از استفاده از NDSI به عنوان متغیر پیش‌بینی‌کننده تکمیلی در حوضه لتیان به وضوح بیشتر از حوضه ناورود بود که دلیل آن می‌تواند به دلیل فیزیوگرافی و ویژگی‌های فیزیکی حوضه لتیان باشد که یک حوضه برف‌خیز است و جریان سطحی حاصل از ذوب برف منبع اصلی و مهم جریان سالانه است (Hay et al., 2000). بنابراین، متغیرهای ورودی P ، T_{max} ، T_{min} و تأخیرهای زمانی آنها، به تنهایی برای نشان دادن رابطه بین ذوب برف و جریان سطحی کافی به نظر نمی‌رسد (Tongal and Booij, 2018).

جدول ۴. نتایج مرحله اعتبارسنجی شبیه‌سازی جریان سطحی در مقیاس روزانه با استفاده از مدل‌های ML مورد مطالعه

بدون اطلاعات پوشش برف			با استفاده از اطلاعات پوشش برف			روش ML	حوضه آبخیز مورد مطالعه
NS	R	RMSE	NS	R	RMSE		
۰/۴۸	۰/۷۲	۴/۶۸	۰/۳۶	۰/۶۴	۵/۱۹	SVR	حوضه آبخیز لتیان
۰/۶۰	۰/۸۰	۴/۱۴	۰/۴۵	۰/۷۰	۴/۸۵	ANN-BP	
۰/۷۲	۰/۸۵	۳/۴۳	۰/۵۸	۰/۷۷	۴/۲۲	GBR	
۰/۳۶	۰/۶۸	۱۹/۰۲	۰/۳۰	۰/۶۶	۱۹/۸۶	SVR	حوضه آبخیز ناورود
۰/۴۹	۰/۷۱	۱۶/۷۰	۰/۱۲	۰/۴۶	۲۱/۶۵	ANN-BP	
۰/۴۲	۰/۶۷	۱۹/۸۰	۰/۴۱	۰/۶۴	۱۸/۸۹	GBR	

شکل ۴ و ۵ به ترتیب جریان سطحی خروجی مشاهده شده و شبیه‌سازی شده از حوضه‌های لتیان و ناورود توسط مدل‌های SVR، ANN-BP و GBR را در مقیاس روزانه نشان می‌دهد. همه هیدروگراف‌ها کارایی همه مدل‌ها را در طول دوره‌های آموزشی و اعتبارسنجی نشان می‌دهند، در حالی که نمودارهای پراکندگی نتایج دوره اعتبارسنجی را برای همه مدل‌ها نشان می‌دهد. نمودارهای پراکندگی در شکل ۴ نشان می‌دهد که همه مدل‌ها کارایی خوبی در شبیه‌سازی جریان سطحی برای حوضه لتیان دارند و با توجه به مقادیر RMSE و R، دقت همه مدل‌ها در شبیه‌سازی جریان سطحی با استفاده از شاخص NDSI افزایش یافته است. در بین هر سه مدل، مدل SVR مقدار جریان سطحی را بالخصوص در نرخ دبی جریان خروجی بالاتر و عدم استفاده از شاخص NDSI، کمتر تخمین زد.

الگوی مشابهی را می‌توان در شکل ۵ مشاهده کرد که همه مدل‌ها در حوضه ناورود، جریان سطحی را در جریان‌های زیاد کمتر از حد معمول برآورد می‌کنند. نمودار پراکندگی مدل ANN-BP نشان می‌دهد که این مدل تمایل دارد دبی جریان خروجی را در جریان‌های پایین بیش‌تر از حد معمول برآورد کند و سپس مانند دو مدل دیگر، دبی جریان را در جریان‌های زیاد کمتر از حد معمول برآورد کند که این دلیل اصلی کارایی ضعیف این مدل برای حوضه ناورود در مقیاس روزانه در مقایسه با مدل‌های SVR و GRB است. با وجود کارایی مدل‌های ANNs در بسیاری از شبیه‌سازی‌های هیدرولوژیکی، (Meng et al., 2019) بیان کرد که مدل ANNs برای فرآیندهای پیچیده هیدرولوژیکی به علت برخورداری از محدودیت‌هایی از جمله کندی کارایی آن در برخی مسائل، یافتن حداقل‌های محلی به جای حداقل‌های سراسری و بیش‌برازش داده‌ها، مناسب نیستند. این یافته تعمیم‌پذیری ضعیف مدل ANN-BP با کارایی پایین‌تر نسبت به SVR و ELM در هنگامیکه که هیچ محصول برفی استفاده نشده است را نشان می‌دهد که با برخی از مطالعات انجام شده در این راستا مطابقت دارد (Parisouj et al., 2020; Shortridge et al., 2016).

اگرچه مدل ANN-BP کمترین کارایی را در حوضه آبخیز ناورود در مرحله آزمون یا اعتبارسنجی بدون استفاده از شاخص NDSI را دارد (R=۰/۴۶ و RMSE=۲۱/۶۵ m³/s)، اما این مدل در همین حوضه، بهترین دقت (R=۰/۷۱ و RMSE=۱۶/۷۰ m³/s) را در بین هر سه مدل با کاربرد شاخص NDSI به عنوان یک متغیر تخمین‌گر تکمیلی، از خود نشان داد. در مطالعه‌ای به منظور شبیه‌سازی جریان سطحی روزانه برای حوضه‌های با ویژگی‌های فیزیکی مختلف، نشان داده شد که مدل ANNs در مقیاس روزانه کارایی بهتری نسبت به مدل SVR، خود رگرسیون^۱ (AR) و همچنین سیستم استنتاج فازی-عصبی تطبیقی^۲ (ANFIS) داشته است (Hadi and Tombul, 2018). نتایج این تحقیق، همراستا با نتایج مطالعه حاضر در مورد کارایی بهتر مدل ANN-BP در اکثر شرایط در مقایسه با مدل SVR در پیش‌بینی جریان سطحی روزانه بود.

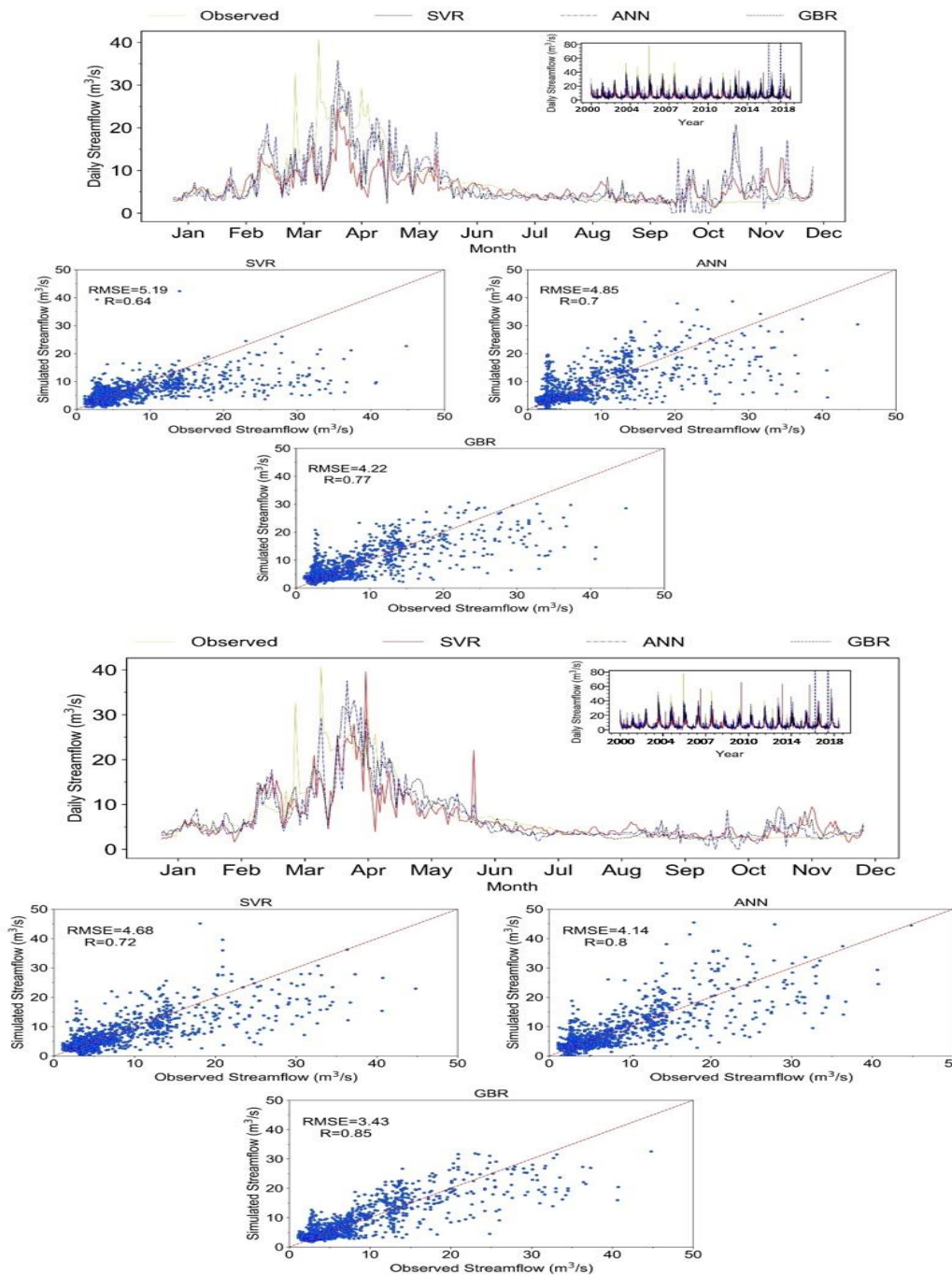
نمودارهای پراکندگی در شکل ۴ و شکل ۵ نشان می‌دهد که همه مدل‌ها کارایی بهتری در حوضه لتیان در مقایسه با حوضه ناورود داشتند و در بیش‌تر موارد، مدل GBR برای هر دو حوضه نسبت به دو مدل SVR و ANN-BP شبیه‌سازی بهتری را ارائه می‌دهد. این موضوع برتری مدل GRB را نسبت به مدل‌های ANN-BP و SVR در شبیه‌سازی جریان سطحی در حوضه‌های مورد مطالعه نشان داد. بنابراین، با توجه به مقایسه نتایج، دقت و کارایی مدل‌ها برای هر دو حوضه‌های مورد مطالعه در مقیاس روزانه به صورت ANN > GBR > SVR است.

در مطالعه‌ای که توسط Liao et al. (2020) انجام شد از مدل GBR به عنوان مدل شبیه‌ساز برتر برای بهبود دقت پیش‌بینی جریان

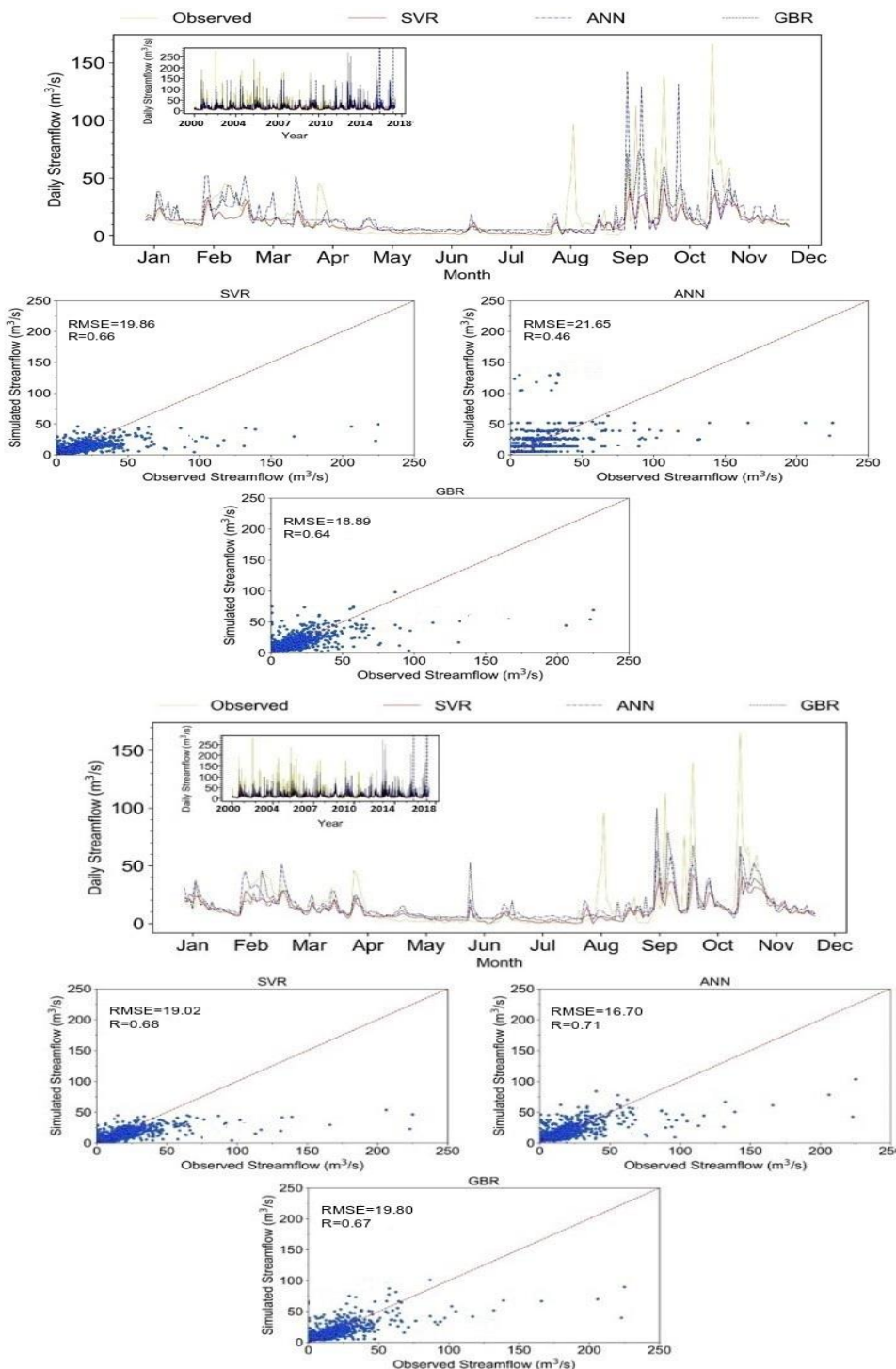
^۱ Autoregressive (AR)

^۲ Adaptive Network-based Fuzzy Inference System (ANFIS)

سطحی به صورت روزانه استفاده شد و نتایج آنها نشان داد که مدل GBR در مقایسه با مدل‌های SVR، ANNs و رگرسیون خطی چندگانه^۱ (MLR) کارایی بهتری داشت، که تطابق خوبی با نتیجه این پژوهش دارد.



شکل ۴. مقایسه جریان سطحی مشاهده شده و شبیه سازی شده حوضه لتیان در مقیاس روزانه بدون استفاده از داده پوشش برف NDSI (الف) و با استفاده از NDSI (ب). دو خط تیره آبی، یک سال از دوره اعتبارسنجی را در هیدروگرافها برای ارائه بزرگتر و بهتر جدا می کنند. نمودارهای پراکندگی ارزیابی دقت را برای دوره اعتبارسنجی نشان می دهد.



شکل ۵. مقایسه جریان سطحی مشاهده شده و شبیه سازی شده حوضه ناورود در مقیاس روزانه بدون استفاده از داده پوشش برف NDSI (الف) و با استفاده از NDSI (ب). دو خط تیره آبی، یک سال از دوره اعتبارسنجی را در هیدروگراف‌ها برای ارائه بزرگتر و بهتر جدا می‌کنند. نمودارهای پراکندگی ارزیابی دقت را برای دوره اعتبارسنجی نشان می‌دهد.

Guo et al. (2021) مطالعه‌ای بر روی قابلیت شبیه‌سازی مدل‌های ML در پیش‌بینی جریان سطحی در تایوان انجام دادند و نتایج آنها نشان داد که مدل GBR نسبت به مدل SVR، ANNs، رگرسیون جنگل تصادفی^۱ (RFR)، کارایی مطلوب‌تری داشت که با نتایج این مطالعه در مورد کارایی بهتر مدل سازگار است. نتایج مشابهی در مورد دقت و کارایی بهتر مدل GBR در مقایسه با سایر مدل‌های ML توسط برخی از محققان گزارش شده است (Zhang et al., 2019; Venkatesan and Mahindrakar, 2019).

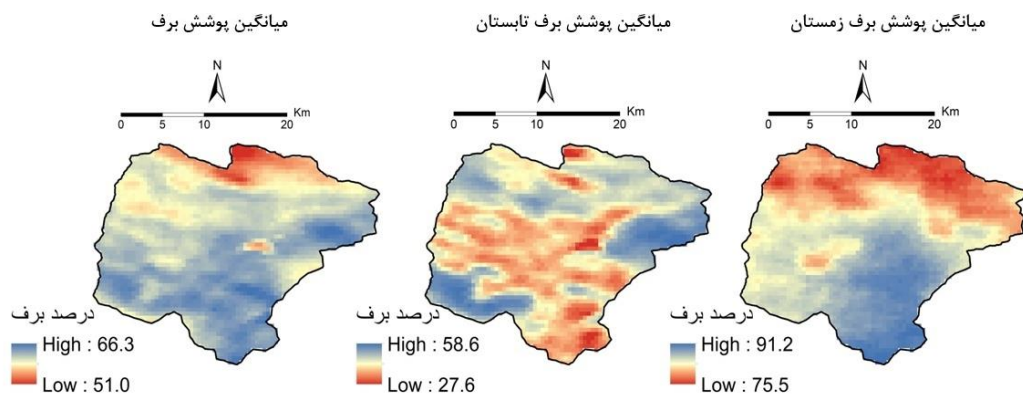
^۱ Random Forest Regression (RFR)

کارایی و دقت تمام مدل‌ها برای شبیه‌سازی جریان سطحی به طور قابل توجهی برای حوضه لتیان در مقایسه با حوضه ناورود به ویژه با استفاده از محصولات پوشش برفی بهبود یافته است (جدول ۳ و جدول ۴). یک رابطه قوی بین کارایی مدل‌های ML در شبیه‌سازی جریان و ویژگی‌های حوضه آبخیز، مانند اقلیم، توپوگرافی و خاک، هیدرولوژی، زمین‌شناسی و پوشش گیاهی وجود دارد (Konapala et al., 2020). علاوه بر این، یکی از دلایل کارایی بهتر مدل‌ها در حوضه لتیان با استفاده از محصول پوشش برف، به این دلیل است که این حوضه در ارتفاعات بالا قرار داشته و منطقه برف‌خیز محسوب می‌شود و جریان سطحی حاصل از ذوب برف، سهم بسزایی در دبی جریان خروجی سالانه ایفا می‌کند که به نوبه خود منجر به بهبود شبیه‌سازی جریان سطحی در صورت استفاده از داده‌های پوشش برف به عنوان تخمین‌گر می‌شود (Hay et al., 2000; Tongal and Booi, 2018).

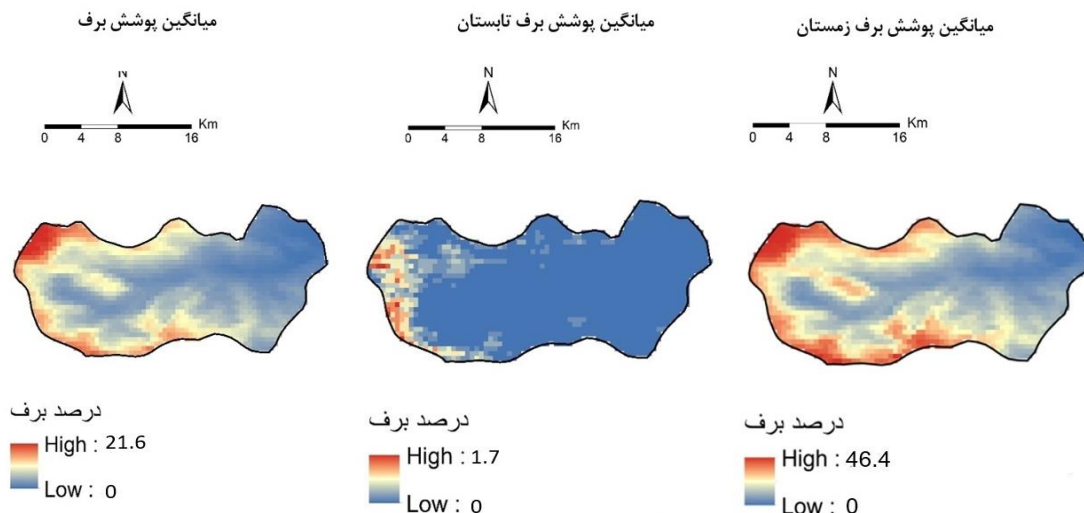
همانطور که در شکل ۲ و ۶ نشان داده شده است، بیشتر قسمت‌های حوضه لتیان در ارتفاعات بالا قرار دارند و میانگین پوشش برف آن ۵۱-۶۶ درصد است در حالی که بیشتر قسمت‌های حوضه ناورود را مناطق با ارتفاع کم تشکیل می‌دهند و میانگین پوشش برف به طور قابل توجهی کمتر از حوضه لتیان می‌باشد (شکل ۶). در واقع بارش برف، بخش قابل توجهی از بارش کل در حوضه لتیان در مقایسه با حوضه ناورود می‌باشد و در نتیجه، دقت و کارایی مدل‌ها در شبیه‌سازی دبی جریان خروجی به طور قابل توجهی با استفاده از NDSI به عنوان یک تخمین‌گر تکمیلی به ویژه در حوضه لتیان بهبود یافت.

به عبارت دیگر، دلیل اصلی کارایی بهتر مدل‌ها در حوضه لتیان در مقایسه با حوضه ناورود می‌تواند به این دلیل باشد که این حوضه در ارتفاعات بالاتر با توپوگرافی بیشتر و همچنین پوشش گیاهی کمتری بوده است (شکل ۲ و ۳). همان‌طور که در شکل ۳ مشاهده می‌شود، بیشتر قسمت‌های حوضه ناورود دارای پوشش گیاهی متوسط تا متراکم هستند و تمامی عوامل ذکر شده باعث می‌شود که جریان سطحی در حوضه لتیان در مقایسه با حوضه ناورود که دارای بارندگی بیشتر و در نتیجه میزان دبی بیشتر بوده، با سرعت بالاتر و تاخیر کمتری به خروجی حوضه برسد.

(الف)



(ب)



شکل ۶. پوشش برف حوضه لتیان (الف) و حوضه ناورود (ب)

نتیجه‌گیری

پژوهش حاضر قابلیت پیش‌بینی سه مدل شامل SVR، ANN-BP و GBR را برای شبیه‌سازی جریان سطحی در مقیاس روزانه در دو حوضه آبخیز لتیان و ناورود با ویژگی‌های اقلیمی، پوشش اراضی و هیدرولوژیکی مختلف در ایران بررسی کرد. کارایی مدل‌ها در شبیه‌سازی و پیش‌بینی جریان سطحی با وجود و عدم وجود داده‌های پوشش برف بررسی شد. نتایج نشان داد که کارایی همه مدل‌ها با استفاده از شاخص NDSI از سنجنده MODIS در مقایسه با شرایطی که در آن هیچ مجموعه داده پوشش برفی استفاده نشده بود، افزایش یافته است. GBR بهترین کارایی را در بین سه مدل ML برای مراحل آموزش و اعتبارسنجی در هر دو حوضه نشان داد، تنها استثنا در حوضه ناورود بوده که در آن مدل ANN-BP در مرحله اعتبارسنجی با استفاده از NDSI کارایی بهتری را داشته است. ضعیف‌ترین کارایی در شبیه‌سازی دبی جریان خروجی با استفاده از SVR برای هر دو مرحله آزمون و اعتبارسنجی در اکثر موارد مشاهده شد. کارایی تمام مدل‌ها برای پیش‌بینی جریان سطحی خروجی به طور قابل توجهی برای حوضه آبخیز لتیان در مقایسه با حوضه ناورود، به‌ویژه با استفاده از محصولات پوشش برفی، بهبود یافته است. دلیل کارایی بهتر مدل‌ها در حوضه لتیان می‌تواند به این دلیل باشد که این حوضه در ارتفاعات بالاتر با توپوگرافی بیشتر و پوشش گیاهی کمتر نسبت به حوضه ناورود واقع شده است. همه این عوامل باعث می‌شود که جریان سطحی در حوضه لتیان، با وجود بارندگی بیشتر و در نتیجه میزان دبی خروجی بیشتر در حوضه ناورود، با سرعت بیشتر و تاخیر کمتری به خروجی حوضه برسد. علاوه بر این، دلیل دیگر برای کارایی بهتر مدل‌ها در حوضه لتیان بویژه با استفاده از محصول پوشش برف، می‌تواند به این دلیل باشد که این منطقه به علت قرار گرفتن در ارتفاعات بالا، برف‌خیز می‌باشد و جریان سطحی ناشی از ذوب برف سهم بسزایی در دبی خروجی از حوضه دارد که باعث می‌شود هنگامیکه از NDSI به عنوان یک تخمین‌گر اضافی استفاده می‌کنیم شبیه‌سازی جریان سطحی در حوضه لتیان نسبت به حوضه ناورود بهبود قابل توجهی پیدا کند (Hay et al., 2000; Tongal and Booij, 2018).

در نهایت، یافته‌های مطالعه حاضر نشان داد که همه مدل‌ها نتایج رضایت بخشی را در شبیه‌سازی جریان سطحی در حوضه‌های آبخیز مورد مطالعه از خود نشان دادند. در میان این مدل‌ها، مدل GBR کارایی دقیق‌تری نسبت به SVR و ANN-BP برای شبیه‌سازی جریان سطحی در مقیاس روزانه برای هر دو حوضه آبخیز داشته است و همچنین می‌توان نتیجه‌گیری کرد که محصولات پوشش برف ماهواره‌ای مانند سنجنش سطح پوشش برف با شاخص NDSI از تصاویر سنجنده MODIS به عنوان یک متغیر تکمیلی می‌تواند قابلیت‌های شبیه‌سازی روش‌های ML را، به ویژه برای حوضه‌های برف‌خیز بهبود بخشد. اکثر رویکردهای هیدرولوژیکی به متغیرهای قابل دسترسی مانند برخی از داده‌های آب و هوایی مانند بارش و دما به عنوان ورودی برای شبیه‌سازی و مدل‌سازی جریان نیاز دارند در حالیکه استفاده از برخی داده‌های تکمیلی مانند نقشه پوشش برف می‌تواند دقت شبیه‌سازی را افزایش دهد. در نتیجه، کیفیت داده‌ها، ساختار و نوع مدل‌های مورد استفاده در شبیه‌سازی، نقش اساسی در انتخاب بهترین مدل در یک منطقه دارد. نتایج مطالعه حاضر نشان داد که تکنیک‌های مدل‌های داده محور، با استفاده از داده‌های قابل دسترس مانند بارش، دما و پایش ماهواره‌ای پوشش برف می‌توانند نتایج قابل اعتمادی را برای شبیه‌سازی جریان سطحی به عنوان ابزاری کاربردی برای تصمیم‌گیرندگان، به‌ویژه در حوضه‌هایی با محدودیت داده، ارائه دهد.

سپاس‌گزاری

بدین وسیله از شرکت‌های آب منطقه‌ای استان تهران و گیلان برای ارائه داده‌های روزانه ایستگاه هیدرومتری و داده‌های هواشناسی برای هر دو حوضه آبخیز مورد مطالعه تشکر و قدردانی می‌گردد. همچنین از ناسا برای ارائه داده‌های ماهواره MODIS قدردانی می‌گردد.

"هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد"

REFERENCES

- Ali, Z., Hussain, I., Faisal, M., Nazir, M.H., Hussain, T., Shad, M.Y., Shoukry, M.A. and Gani, S. (2017). Forecasting drought using multilayer perceptron artificial neural network model. *Advances in Meteorology*, 1-9.
- Ali, M., Prasad, R., Xiang, Y. and Yaseen, Z. (2020). Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. *Journal of Hydrology*, 584, 124647.
- Ali, M., Deo, R.C., Maraseni, T. and Downs, N.J. (2019). Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *Journal of Hydrology*,



576, 164–184.

- Avand, M., Janizadeh, S. and Jafari, F. (2020). Evaluating the Efficiency of Machine Learning Models in Preparing Flood Probability Mapping. *Degradation and Rehabilitation of Natural Land*. 1(1):19-32. (In Farsi)
- Adnan, R.M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B. and Kisi, O. (2019). Daily streamflow prediction using optimally pruned extreme learning machine. *Journal of Hydrology*, 577, 123981.
- Ahmadi, H., Malekian, A. and Abedi, R. (2012). The most appropriate statistical method for estimating suspended sediment of Jajrud River (Case study: Rudak station of Jajrud watershed). *Journal of Environmental Erosion Research*, 2, 88-78.
- Bennett, K., Cherry, J., Balk, B. and Lindsey, S. (2019). Using MODIS estimates of fractional snow cover area to improve streamflow forecasts in interior Alaska. *Hydrology and Earth System Sciences*, 23, 2439-2459.
- Boser B.E., Guyon I.M. and Vapnik V.N. (1992). A training algorithm for optimal margin classifiers. In D.Haussler, editor, *5th Annual ACM Workshop on COLT*, pp. 144 -152, Pittsburgh, PA.
- Bilandi M., Khashei siouki E. and Sadeghi Tabas S. (2014). Daily rainfall-runoff modeling with Least Square Support Vector Machine (LS-SVM). *Journal of Soil and Water Conservation Research*, 6, 293 -304. (In Farsi)
- Benimam H., Si -Moussa C., Laidi M. and Hanini S. (2020). Modeling the activity coefficient at infinite dilution of water in ionic liquids using artificial neural networks and support vector machines. *Neural Comput. Appl.* 32 (12), 8635 –8653.
- Cheng, C., Zhao, M.Y., Chau, K. and Wu, X.Y. (2006). Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *Journal of Hydrology*, 316, 129-140.
- Cheng, M., Fang, F., Kinouchi, T., Navon, I.M. and Pain, C.C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590, 125376.
- Dikshit, A., Pradhan, B. and Alamri, A.M. (2020). Temporal Hydrological Drought Index Forecasting for New South Wales, Australia Using Machine Learning Approaches. *Atmosphere*, 11(6), 585.
- Dibike, Y.B., Velickov, S., Solomatine, D. and Abbott, M.B. (2001). Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15(3), 208-216.
- Ebrahimi, H. and Rajaei, T. (2017). Simulation of groundwater level variations using wavelet combined with neural network, linear regression and support vector machine. *Global and Planetary Change*, 148, 181–191.
- Guo, W.D., Chen, W.B., Yeh, S.H., Chang, C.H. and Chen, H. (2021). Prediction of River Stage Using Multistep-Ahead Machine Learning Techniques for a Tidal River of Taiwan. *Water*, 13(7), 920
- Garmdareh, E., Vafakhah, M. and Eslamian, S. (2019). Assessment the Performance of Support Vector Machine and Artificial Neural Network Systems for Regional Flood Frequency Analysis (A Case Study: Namak Lake Watershed). *Journal of Water and Soil Science*. 23. 351-366. (In Farsi).
- Hall, D.K., Riggs, G.A. and Salomonson, V.V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectro-radiometer data. *Remote Sensing of Environment*, 54(2), 127–140.
- Ha, S., Liu, D. and Mu, L. (2021). Prediction of Yangtze River stream flow based on deep learning neural network with El Niño–Southern Oscillation. *Scientific Reports*, 11, 11738.
- Hosseini, S.M. and Mahjouri, N. (2016). Integrating Support Vector Regression and a geo-morphologic Artificial Neural Network for daily rainfall-runoff modeling. *Applied Soft Computing*, 38, 329–345.
- He, S., Gu, L., Tian, J., Deng, L., Yin, J., Liao, Z., Zeng, Z., Shen, Y. and Hui, Y. (2021). Machine Learning Improvement of Stream Flow Simulation by Utilizing Remote Sensing Data and Potential Application in Guiding Reservoir Operation. *Sustainability*, 13, 3645.
- Hadi, S.J. and Tombul, M. (2018). Forecasting daily stream flow for basins with different physical characteristics through data-driven methods. *Water Resources Management*, 32, 3405–3422.
- Hay, L.E., Wilby, R.L. and Leavesley, G.H. (2000). A comparison of delta changes and downscaled GCM scenarios for three mountainous basins in the United States 1. *JAWRA Journal of the American Water Resources Association*, 36, 387-397
- Kashid, S.S., Ghosh, S. and Maity, R. (2020). Stream flow prediction using multi-site rainfall obtained from hydroclimatic teleconnection. *Journal of Hydrology*, 395, 23–38.
- Keteklahijani, V.K., Alimohammadi, S. and Fattahi, E. (2019). Predicting changes in monthly streamflow to Karaj dam reservoir, Iran, in climate change condition and assessing its uncertainty. *Ain Shams Engineering Journal*, 10, 669–679.

- Khan, N., Sachindra, D.A., Shahid, S., Ahmed, K., Shiru, M.S. and Nawaz, N. (2020). Prediction of droughts over Pakistan using machine learning algorithms. *Advances in Water Resources*, 139, 103562.
- Konapala, G., Kao, S.C., Painter, S. and Lu, D. (2020). Machine learning assisted hybrid models can improve stream flow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, 15(10), 104022.
- Lin, J.Y., Cheng, C.T. and Chau, K.W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), 599-612.
- Liu, Y., Sang, Y.F., Li, X., Hu, J. and Liang, K. (2017). Long-term stream flow forecasting based on relevance vector machine model. *Water*, 9(1), 9.
- Liao, S., Liu, Z., Liu, B., Cheng, C., Jin, X. and Zhao, Z. (2020). Multistep-ahead daily inflow forecasting using the ERA-Interim reanalysis data set based on gradient-boosting regression trees. *Hydrology and Earth System Sciences*, 24, 2343–2363.
- Meng, E., Huang, S., Huang, Q., Fang, W., Wu, L. and Wang, L. (2019). A robust method for non-stationary stream flow prediction based on improved EMD-SVM model. *Journal of Hydrology*, 568, 462-478.
- Malik, A., Tikhamarine, Y., Souag-Gamane, D., Kisi, O. and Pham, Q.B. (2020a). Support vector regression optimized by meta-heuristic algorithms for daily stream flow prediction. *Stochastic Environmental Research and Risk Assessment*, 34, 1755–1773.
- Malik, A., Kumar, A., Salih, S.Q., Kim, S., Kim, N.W., Yaseen, Z.M. and Singh, V.P. (2020b). Drought index prediction using advanced fuzzy logic model: Regional case study over Kumaon in India. *PLoS ONE*, 15 (5), e0233280.
- Mahjouri, N. and Kerachian, R. (2011). Revising river water quality monitoring networks using discrete entropy theory: The Jajrood River experience. *Environmental Monitoring and Assessment*, 175, 291-302.
- Mokari, E., DuBois, D., Samani, Z., Mohebzadeh, H. and Djaman, K. (2021). Estimation of daily reference evapotranspiration with limited climatic data using machine learning approaches across different climate zones in New Mexico. *Theoretical and Applied Climatology*, 1-13.
- Nolin, A.W. and Liang, S. (2002). Progress in bidirectional reflectance modeling and applications for surface particulate media: snow and soils. *Remote Sensing Reviews*, 18, 307–342.
- Niu, W.J., Feng, Z.K., Zeng, M., Feng, B.F., Min, Y.W. and Cheng, C. (2019). Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. *Applied Soft Computing*, 105589.
- Ni, Q., Wang, L., Ye, R., Yang, F. and Sivakumar, M. (2010). Evolutionary modeling for stream flow forecasting with minimal datasets: a case study in the West Malian River, China. *Environmental Engineering Science*, 27,377–385.
- Prasad, R., Ali, M., Xiang, Y. and Khan, H. (2020). A double decomposition-based modelling approach to forecast weekly solar radiation. *Renewable Energy*, 152, 9–22.
- Parisouj, P., Mohebzadeh, H. and Lee, T. (2020). Employing Machine Learning Algorithms for Streamflow Prediction: A Case Study of Four River Basins with Different Climatic Zones in the United States. *Water Resources Management*, 34, 4113–4131.
- Roy, D.K., Barzegar, R., Quilty, J. and Adamowski, J. (2020). Using ensembles of adaptive neuro-fuzzy inference system and optimization algorithms to predict reference evapotranspiration in subtropical climatic zones. *Journal of Hydrology*, 591, 125509.
- Rakwatin, P., Takeuchi, W. and Yasuoka, Y. (2008). Restoration of Aqua MODIS band 6 using histogram matching and local least squares fitting. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 613–627.
- Steele, C., Dialesandro, J., James, D., Elias, E., Rango, A. and Bleiweiss, M. (2017). Evaluating MODIS snow products for modelling snowmelt runoff: Case study of the Rio Grande headwaters. *International Journal of Applied Earth Observation and Geoinformation*, 63, 234-243.
- Shahabi, H., Khezri, S., Ahmad, B.B. and Musa, T. (2014). Application of moderate resolution imaging spectroradiometer snow cover maps in modeling snowmelt runoff process in the central Zab basin, Iran. *Journal of Applied Remote Sensing*, 8(1), 084699.
- Shortridge, J.E., Guikema, S.D. and Zaitchik, B.F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, 20, 2611-2628.
- Tyralis, H., Papacharalampous, G. and Langousis, A. (2021). Super ensemble learning for daily stream flow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms. *Neural Computing and Applications*, 33, 3053–3068.



- Tongal, H. and Booi, M.J. (2018). Simulation and forecasting of stream flows using machine learning models coupled with base flow separation. *Journal of hydrology*, 564, 266-282.
- Venkatesan, E. and Mahindrakar, A.B. (2019). Forecasting floods using extreme gradient boosting – A new approach. *International Journal of Civil Engineering and Technology*, 10, 1336-1346.
- Wang, L., Li, X., Ma, C. and Bai, Y. (2019). Improving the prediction accuracy of monthly stream flow using a data-driven model based on a double-processing strategy. *Journal of Hydrology*, 573, 733-745.
- Weier, J. and Herring, D. (2011). Measuring Vegetation (NDVI & EVI)” (<http://earthobservatory.nasa.gov/Features/MeasuringVegetation/>).
- Xiang, Z. and Demir, I. (2020). Distributed long-term hourly stream flow predictions using deep learning—a case study for State of Iowa. *Environmental Modelling & Software*, 131, 10476.
- Xiang, Z., Yan, J. and Demir, I. (2020). A rainfall-runoff model with LSTM-based sequenceto-sequence learning. *Water Resources Research*, 56(1), e2019WR025326.
- Yu, P., Yang, T., Chen, S., Kuo, C. and Tseg, H. (2017). Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology*, 552, 92–104.
- Zhang, H., Yang, Q., Shao, J., and Wang, G. (2019). Dynamic Stream Flow Simulation via Online Gradient-Boosted Regression Tree. *Journal of Hydrologic Engineering*, 24(10), 0401.