



مدل سازی و پیش بینی شوری آب زیرزمینی دشت قزوین با رویکرد داده محور و هوش مصنوعی

نیما مقدم^۱، افشین اشرف زاده^{۲*}

۱. دانشجوی کارشناسی ارشد مهندسی منابع آب گروه مهندسی آبیاری و ابادانی دانشکدهگان کشاورزی و منابع طبیعی دانشگاه تهران

۲. دانشیار گروه مهندسی آبیاری و ابادانی دانشکدهگان کشاورزی و منابع طبیعی دانشگاه تهران

۱- چکیده

در مطالعه حاضر که باهدف مدل سازی کیفیت آب های زیرزمینی انجام شد، روشی برای محاسبه ساختار فضایی داده ها در فرآیند مدل سازی پیشنهاد شده است. این روش فاصله بین نقاط مشاهده و نقطه تخمین را به عنوان یکی از ورودی های مدل در برمی گیرد. مدل های GBR، GPR، KNN، MLP، SVM و RF مورد استفاده قرار گرفتند و مدل ها با استفاده از داده های کیفیت آب زیرزمینی به دست آمده به ترتیب از ۳۰۵۸ چاه برای ۴ چاه مشاهده ای نزدیک به هدف، ۲۷۲۴ چاه برای ۵ چاه مشاهده ای نزدیک به هدف، ۲۰۸۰ چاه برای ۶ چاه مشاهده ای نزدیک به هدف، ۱۳۶۴ چاه برای ۷ چاه مشاهده ای نزدیک به هدف و در نهایت از ۶۳۱ چاه برای ۸ چاه مشاهده ای نزدیک به هدف در منطقه ای در کل استان قزوین در شمال غرب ایران آموزش و آزمایش شدند دو مجموعه داده مجزا شامل اطلاعات ۶ ماهه اول سال و ۶ ماهه دوم سال استفاده شد. میانگین خطا (R-squared)، MAE و ضریب اصلاح پیرسون و RMSE برای مدل هایی با چهار، پنج، شش، هفت و هشت چاه همسایه نشان دهنده عملکرد رضایت بخش مدل جنگل تصادفی بود. این مدل در هر دو مرحله آموزش و آزمون عملکرد بسیار خوبی داشته و کمترین خطا و بالاترین همبستگی را نشان می دهد. اگرچه پیچیدگی و زمان اجرای این مدل ممکن است بیش تر باشد، اما دقت بالای آن این معایب را جبران می کند. نتایج عددی حاصل از مدل سازی نشان داد که مدل جنگل تصادفی (RF) در اغلب سناریوهای مورد بررسی عملکرد برتری نسبت به سایر مدل ها داشته است. برای مثال، در حالت استفاده از چهار چاه مجاور، این مدل در مرحله آزمون به مقدار $R^2=0/915$ ، ضریب همبستگی $0/957$ ، مقدار $MAE \approx 156$ و $RMSE \approx 326$ دست یافت. همچنین در حالت پنج چاه مجاور نیز مدل RF مقدار $R^2=0/935$ و $RMSE \approx 316$ را ثبت کرد که بیانگر پایداری و دقت بالای آن است. مدل فرآیند گاوسی (GPR) نیز به عنوان دومین مدل دقیق، در دو دوره زمانی شش ماهه به مقادیر $R^2 \approx 0.904$ تا $0/934$ و $RMSE \approx 350$ تا 318 دست یافت و عملکردی نسبتاً پایدار از خود نشان داد. در مقابل، مدل های KNN و GBR علیرغم ارائه نتایج قابل قبول، خطاهای بالاتری نسبت به RF و GPR داشتند. به طور نمونه، مقدار RMSE مدل KNN در سناریوی چهار چاه مجاور حدود 407 و برای مدل GBR حدود 461 محاسبه شد که نشان دهنده ضعف نسبی این مدل ها در بازسازی ساختار فضایی هدایت الکتریکی است. به طور کلی، میانگین مقادیر MAE، RMSE، R-squared و ضریب همبستگی پیرسون نشان داد که مدل RF در هر دو مرحله آموزش و آزمون کمترین خطا و بالاترین میزان همبستگی را ارائه کرده است. این مدل در بهترین حالت به R^2 بالاتر از $0/95$ در مرحله آموزش و RMSE کمتر از 270 دست یافت که نشان دهنده توانایی بالای آن در مدل سازی الگوهای مکانی-فضایی هدایت الکتریکی در آبخوان منطقه است. اگرچه پیچیدگی محاسباتی این مدل نسبتاً بالاتر است، اما دقت زیاد و پایداری نتایج، این محدودیت را به طور کامل جبران می کند.

کلمات کلیدی: مدل سازی، آب های زیرزمینی، چاه مشاهده ای، هوش مصنوعی

۲- مقدمه

نیازسنجی ثابت کرده است که یکی از نیازهای حیاتی انسان دسترسی به آب آشامیدنی است. به همین دلیل بشر اقدامات گسترده ای را انجام داده است تا این نیاز را تأمین کند، از جمله حفر چاه، سدسازی، قنات و سامانه های آبیاری. اما در حال حاضر، مخالفت های زیادی وجود دارد که باید به آن ها توجه کرد. اخیراً با شاهد بودن تغییرات آب و هوایی و گرمایش جهانی، هشدارهای زیادی به ما داده شده است. افزایش جمعیت نیز باعث افزایش نگرانی ها شده است. در عصر حاضر با پیشرفت علم و فناوری، به دنبال راه حل های نوین برای مدیریت منابع آبی و بهبود کیفیت آب هستیم. مدل سازی پارامترهای کیفیت آب یکی از چالش های اساسی است که توسط چندین مطالعه و پژوهش بررسی شده است. از این روش های



علمی جدید می‌توان برای حفظ و بهبود منابع آبی استفاده کرد، اما باید با دقت و آگاهی انجام شود تا به نتایج مطلوب دست‌یابیم و آینده بهتری برای نسل‌های آینده فراهم کنیم (Parmar & Bhardwaj, 2014; Zhou, 2020).

منابع آب زیرزمینی یکی از مهم‌ترین و کاربردی‌ترین منابع آب هستند که در توسعه پایدار فعالیت‌های اجتماعی، اقتصادی و فرهنگی یک منطقه، به‌ویژه در مناطق خشک و نیمه‌خشک، نقش بسیار مهمی دارند. از این رو، شناخت صحیح و بهره‌برداری اصولی از منابع آب زیرزمینی امری بسیار حیاتی است. عدم شناخت صحیح می‌تواند به کاهش دبی چاه‌ها و قنات، تغییر الگوی جریان آب زیرزمینی و پیشروی جبهه‌های آب‌شور و تداخل آب‌شور منجر شود (Mahmoudpour et al., 2020). برای آگاهی از وضعیت منابع آب زیرزمینی و مدیریت بهینه آن‌ها، باید بررسی دقیقی از نوسانات و شبیه‌سازی پارامترهای کمی و کیفی آب زیرزمینی انجام گیرد (Naderianfar et al., 2011). مدیریت منابع آب زیرزمینی ابتدا نیازمند شناخت عملکرد سطح آب زیرزمینی در شرایط طبیعی و سپس پیش‌بینی تغییرات آن است. بهترین روش برای شناخت رفتارهای یک سیستم آب زیرزمینی انجام یک سری تحقیقات بلندمدت برای هر منطقه خاص است، اما با توجه به محدودیت بودجه‌های تحقیقاتی، عملی‌سازی این روش‌ها ممکن نیست. از این رو، از ابزارهایی مانند شبیه‌سازی و مدل‌سازی برای به دست آوردن نتایج قابل قبولی استفاده می‌شود (Askari & Egdermezhad, 2022).

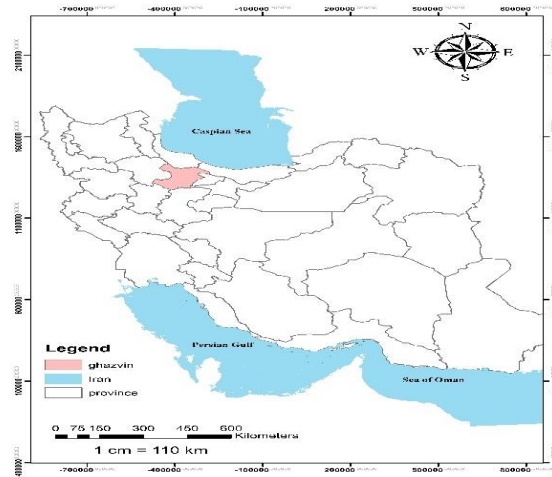
از جمله فن‌های پیشرفته کامپیوتری برای شبیه‌سازی و درک روابط بین پارامترهای کمی و کیفی آب زیرزمینی، استفاده از مدل شبکه‌های عصبی مصنوعی است. این شبکه‌ها از رفتارهای مغز انسان الهام می‌گیرند و برای پردازش اطلاعات، آموزش و یادگیری به کار می‌روند. همچنین، تلفیق فناوری شبکه‌های عصبی مصنوعی با الگوریتم‌های بهینه‌سازی، نتایج قابل قبولی را در مدل‌سازی سامانه‌های پیچیده غیرخطی در مسائل هیدرولوژی و مدیریت منابع آب نشان داده است. این روش‌ها توسط محققان در حوزه‌های مختلف گزارش شده و بهبودهای چشم‌گیری در این زمینه ایجاد کرده‌اند (Luk et al., 2001). عاقلی و همکاران، به مدل‌سازی وابستگی مکانی داده‌های باکیفیت با استفاده از ورودی‌های مبتنی بر فاصله و یک مدل مبتنی بر داده پرداختند. آن‌ها در این مطالعه که باهدف مدل‌سازی کیفیت آب‌های زیرزمینی انجام شد، روشی برای محاسبه ساختار فضایی داده‌ها در فرآیند مدل‌سازی پیشنهاد کردند، این روش فاصله بین نقاط مشاهده و نقطه تخمین را به‌عنوان یکی از ورودی‌های مدل در برمی‌گیرد. نتایج مطالعه آن‌ها نشان داد که هرچه ترکیب تعداد چاه‌های مطالعاتی و فاصله آن‌ها افزایش یابد امکان برآورد دقیق‌تر پارامتر موردبررسی آن‌ها که رسانایی الکتریکی بوده بیش‌تر شده است (Agheli et al., 2024). امامی و همکاران از الگوریتم‌های انتخاباتی ژنتیک و مدل شبکه عصبی مصنوعی به منظور مدل‌سازی تر از آب زیرزمینی دشت میاندواب استفاده کردند. بر اساس نتایج به‌دست‌آمده، الگوریتم انتخاباتی، نسبت به دو روش شبکه عصبی مصنوعی و الگوریتم ژنتیک، نتایج بهتری در پیش‌بینی تراز سطح آب زیرزمینی ارائه نموده است (Emami et al., 2018). رجائی و پور اصلان به‌پیش‌بینی زمانی و مکانی تراز آب زیرزمینی دشت داورزن با استفاده از روش‌های پرسپترون چندلایه شبکه عصبی و کریجینگ پرداختند داده‌های ورودی شامل سری زمانی آب زیرزمینی به مدت ۸ سال بود. نتایج مطالعه آنان نشان داد که مدل شبکه عصبی مصنوعی، نتایج قابل قبولی برای پیش‌بینی مکانی و زمانی تراز آب زیرزمینی ارائه داد (Rajaei & Pouraslan, 2015). محمدی و همکاران کارایی الگوریتم هیبریدی ازدحام ذرات در شبیه‌سازی تر از سطح ایستابی آبخوان دشت اردبیل را بررسی کردند. نتایج این پژوهش حاکی از عملکرد مناسب این الگوریتم با مجذور میانگین مربعات خطا ۰/۴۱۷ بود (Mohammadi et al., 2017). پایش آب‌های زیرزمینی نه‌تنها شالوده مدیریت آب‌های زیرزمینی را تشکیل می‌دهد، بلکه امکان ارزیابی نرخ تغذیه، تعادل آب و تحقیقات مربوطه را نیز فراهم می‌کند، همچنین به‌عنوان ابزاری ارزشمند برای ایجاد سیاست‌ها در خصوص منابع آب زیرزمینی عمل می‌کند (Van Tuan et al., 2023; Mahmoudpour et al., 2023). علی‌رغم پیشرفت‌های قابل‌توجه در یادگیری ماشین برای مبحث آب زیرزمینی و مدل‌سازی کیفیت، روش‌های داده محور قبلی به دلیل توزیع نامنظم چاه‌های نظارتی، همبستگی فضایی ذاتی در داده‌ها را نادیده گرفته‌اند و یا از مدل‌هایی با تعداد کم‌تر و دقت پایین‌تر استفاده کرده‌اند که چالش‌هایی را برای مدل‌های یادگیری ماشین سنتی در مدیریت و حکمرانی آب بدون ساختار و داده‌های متناقض ایجاد می‌کنند.

۳- مواد و روش‌ها

۳-۱- منطقه مطالعاتی



این مطالعه بر روی سفره‌های زیرزمینی استان قزوین تمرکز دارد. استان قزوین یکی از استان‌های شمال غربی ایران است که مرکز آن شهر قزوین می‌باشد. این استان از شمال به استان گیلان، از غرب به استان زنجان، از جنوب به استان همدان و قم و از شرق به استان البرز محدود می‌شود. استان قزوین دارای اقلیم معتدل کوهستانی با تابستان‌های معتدل و زمستان‌های سرد است. ارتفاعات البرز و جلگه قزوین از جمله ویژگی‌های این استان در زمینه جغرافیا هستند. (شکل ۱ الف)

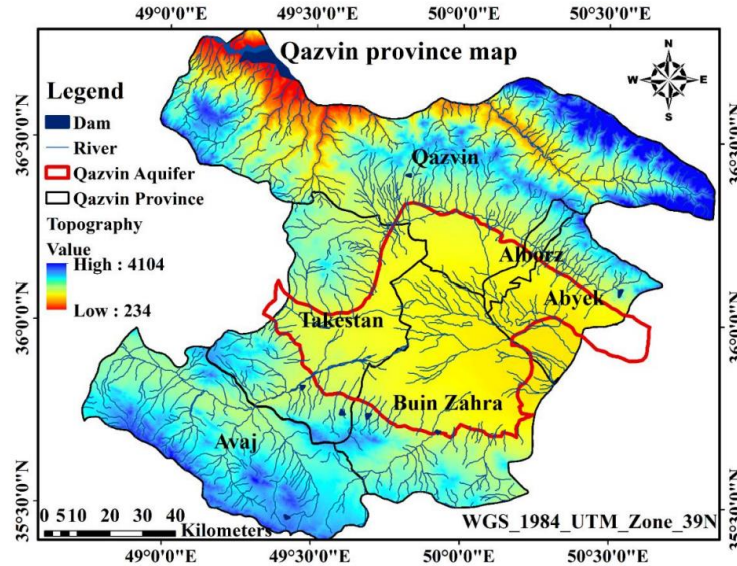


ویپر اسد

شکل ۱ الف) نقشه ایران و موقعیت استان قزوین در آن

جدول ۱- پارامترهای کیفی آب زیرزمینی دشت قزوین

	K+	Na+	Mg+2	Ca+2	So4-2	Cl-1	Hco3-1	CO3-2	pH	TDS	EC
	meq/l	meq/l	meq/l	meq/l	meq/l	meq/l	meq/l	meq/l		Mg/l	ds/m
Maximum	1	63.2	23.9	41.46	31.67	59.94	20.16	1.6	10.3	9.84	1.6
Minimum	0.01	0.03	0.12	0.33	0.02	0.08	0.29	0.1	5.6	0.11	0.1
Average	0.07	7.33	3.42	4.28	4.81	6.19	4.06	0.56	7.77	1.52	0.56
Standard deviation	0.07	3.33	1.47	1.75	2.94	2.83	1.22	0.28	0.43	0.56	0.35



شکل ۱ ب) توپوگرافی استان قزوین (Janbaz Fotamy et al., 2023).

این استان به دلیل موقعیت جغرافیایی خود دارای تنوع آب و هوایی و دمایی خاصی است که آن را برای کشاورزی و فعالیت‌های دامی مناسب می‌سازد. قزوین دارای آب‌وهوای متنوعی است که از مناطق کوهستانی و سردسیر در شمال و شمال غربی تا مناطق نیمه‌خشک و معتدل در جنوب و مرکز استان متغیر است. میانگین دمای سالانه در قزوین حدود ۱۳ درجه سانتی‌گراد است. در استان قزوین فصل تابستان معمولاً گرم و خشک با دمای میانگین حدود ۲۵ درجه سانتی‌گراد و زمستان‌ها سرد با دمای میانگین حدود ۲ درجه سانتی‌گراد می‌باشد (Haghshenas Gatabi et al., 2015; Banihashemi et al., 2021; Barzoo et al., 2022).

میانگین بارش سالانه استان قزوین تا سال ۲۰۲۳ حدود ۲۹۰ میلی‌متر بوده است. این میزان در نقاط مختلف استان متغیر است و از حدود ۲۱۰ میلی‌متر در بخش‌های شرقی تا بیش از ۵۵۰ میلی‌متر در ارتفاعات شمال شرقی متغیر است. در سال‌های اخیر، میزان بارندگی در این استان نسبت به میانگین بلندمدت کاهش یافته و در برخی سال‌ها این میزان حدود ۳۸/۵ درصد کمتر از میانگین بوده است (Hedayati Dezfouli & Kakavand, 2007).

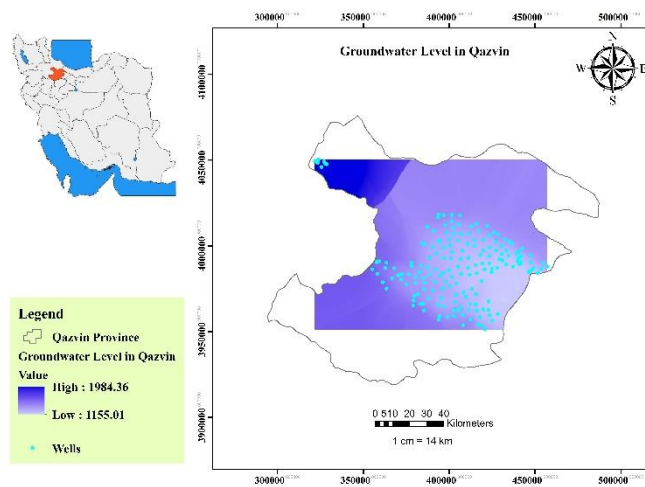
استان قزوین دارای دشت‌های حاصلخیز بسیاری است که عمدتاً در مناطق مرکزی و جنوبی استان واقع شده‌اند که در (شکل ۱ ب) به صورت نوار سبز رنگ مشخص گردیده. دشت قزوین یکی از مهم‌ترین و بزرگ‌ترین دشت‌های کشاورزی ایران است. خاک‌های این دشت‌ها عمدتاً لومی و لومی-رسی هستند که برای کشاورزی بسیار مناسب‌اند. محصولات عمده کشاورزی استان شامل گندم، جو، سیب‌زمینی، چغندر قند، انگور و انواع میوه‌ها و سبزی‌ها است. قزوین دارای منابع آب زیرزمینی مهمی است که از سفره‌های آب زیرزمینی تغذیه می‌شود. باین‌حال، به دلیل مصرف زیاد آب در بخش کشاورزی و صنعتی، سطح آب زیرزمینی در برخی مناطق استان به‌طور قابل‌توجهی کاهش یافته است (Faraji et al., 2017). دشت قزوین، مهم‌ترین دشت کشاورزی استان است که به دلیل خاک‌های غنی و منابع آب مناسب، مرکز اصلی تولید محصولات کشاورزی در استان است. دشت بوبین‌زهر، یکی دیگر از دشت‌های حاصلخیز استان است که به تولید محصولات زراعی و باغی اختصاص دارد. در این تحقیق ۱۰ تا از نزدیک‌ترین چاه‌ها به هر چاه مشاهده‌ای در ماه فوریه از سال ۱۹۹۹ تا ۲۰۱۸ در نظر گرفته شد. استان قزوین دارای انواع مختلف سفره‌های آب زیرزمینی است که به دلیل ویژگی‌های زمین‌شناسی و اقلیمی متنوع این منطقه به وجود آمده‌اند. این سفره‌ها شامل:

۱. سفره‌های آزاد، این نوع سفره‌ها در مناطق مختلف استان قزوین به‌ویژه در دشت قزوین و دشت بوبین‌زهر وجود دارند. سفره‌های آزاد به دلیل نفوذپذیری بالا و سطح آب زیرزمینی که مستقیماً با جو ارتباط دارد، به راحتی از بارش‌ها تغذیه می‌شوند.
۲. سفره‌های محصور، این سفره‌ها در بخش‌های عمیق‌تر زمین قرار دارند و توسط لایه‌های غیرقابل نفوذ مانند رس و شیل محصور شده‌اند. این نوع سفره‌ها نیز در مناطق مختلف قزوین یافت می‌شوند و معمولاً آب با فشار بالاتری در آن‌ها ذخیره می‌شود.
۳. سفره‌های نیمه محصور، این نوع سفره‌ها ترکیبی از ویژگی‌های سفره‌های آزاد و محصور را دارند. لایه‌های نیمه نفوذپذیر می‌توانند باعث جریان آب به صورت جانبی و عمودی شوند. این سفره‌ها نیز در برخی مناطق استان وجود دارند و به دلیل ویژگی‌های خاص خود، در مدیریت منابع آب زیرزمینی اهمیت دارند.



۴. سفره‌های کارستی، این سفره‌ها در مناطق سنگ آهکی و گچی که تحت تأثیر فرآیندهای انحلال قرار گرفته‌اند، وجود دارند. قزوین به دلیل داشتن مناطق کارستی، دارای سفره‌های آب زیرزمینی کارستی است که به دلیل حفره‌ها و کانال‌های طبیعی ایجاد شده در سنگ‌ها، می‌تواند ذخیره آب قابل توجهی داشته باشد. کیفیت آب زیرزمینی در این سفره‌ها به‌طور کلی برای کاربردهای مختلف مناسب است، اگرچه بسته به موقعیت و عمق سفره می‌تواند متفاوت باشد (Moghaddam, 2024).

استان قزوین دارای چندین هزار چاه آب است که به‌منظور مصارف کشاورزی، صنعتی و شرب حفر شده‌اند. بر اساس گزارش‌های رسمی، تعداد چاه‌های مجاز و غیرمجاز استان قزوین بیش از ۶,۰۰۰ حلقه چاه است که برخی از این چاه‌ها غیرمجاز و بدون پروانه هستند و باعث افت سطح آب زیرزمینی شده‌اند. میزان آبدهی چاه‌ها بستگی به عوامل مختلفی نظیر عمق چاه، نوع سفره آب زیرزمینی، شرایط جغرافیایی و اقلیمی منطقه دارد. به‌طور متوسط، میزان آبدهی چاه‌ها در دشت قزوین حدود ۲۰ لیتر در ثانیه بوده است، اما این میزان در سال‌های اخیر به دلیل کاهش سطح آب زیرزمینی و خشک‌سالی‌های پی‌درپی کاهش یافته است. برخی چاه‌ها با آبدهی کمتر از ۱۰ لیتر در ثانیه و برخی با آبدهی بیشتر از ۳۰ لیتر در ثانیه نیز وجود دارند. (شکل ۲) نشان‌دهنده تغییرات در هد هیدرولیک آب زیرزمینی در منطقه مورد مطالعه را نشان می‌دهد، این شکل یک نمای کلی از جهت جریان آب زیرزمینی را ارائه می‌دهد، که در آن بخشی از آب زیرزمینی از شمال شرقی به سمت مرکز و بخش عمده دیگر در از جنوب به سمت غرب و شمال غرب امتداد دارد (Masoudi et al., 2024).



شکل ۲) تراز هیدرولیکی آب زیرزمینی در سراسر سفره آب زیرزمینی در استان قزوین، ایران

منابع آب زیرزمینی استان قزوین به دلیل برداشت بی‌رویه و کاهش بارندگی در سال‌های اخیر با مشکلات جدی مواجه شده‌اند. افت سطح آب زیرزمینی و کاهش کیفیت آب از جمله چالش‌های اصلی در این زمینه هستند. برنامه‌های مدیریت منابع آب شامل کنترل برداشت‌ها، توسعه سامانه‌های آبیاری مدرن، جمع‌آوری چاه‌های غیرمجاز و افزایش بهره‌وری آب در بخش کشاورزی از اقدامات مهمی است که در دستور کار قرار دارد. استان قزوین به علل چالش‌های مدیریت آب‌های زیرزمینی از جمله کاهش سطح آب زیرزمینی، آلودگی منابع آب به دلیل کشاورزی بی‌رویه و شور شدن آب‌های زیرزمینی، که در آن وجود دارد به‌عنوان منطقه مطالعاتی انتخاب شد. از طرفی استان قزوین دارای منابع آبی زیرزمینی متنوع و بسیار مهمی هست که می‌تواند به انجام تحقیقات در زمینه آب‌های زیرزمینی کمک کند. وجود چاه‌ها، چشمه‌ها و آب‌های زیرزمینی با ویژگی‌های مختلف این امکان را ایجاد می‌کند که تحقیقات دقیق‌تر و جامع‌تری انجام شود.

۲-۳- داده‌ها

پایش آب‌های زیرزمینی استان قزوین شامل ۱۷۹ حلقه چاه مشاهده‌ای است. (شکل ۱ ب) شرکت مدیریت منابع آب ایران دو بار در سال نمونه‌های آب را از این چاه‌ها جمع‌آوری می‌کند. نمونه‌های جمع‌آوری شده برای اندازه‌گیری پارامترهای کیفی مختلف مانند هدایت الکتریکی، pH و غلظت کاتیون‌ها و آنیون‌های مختلف آنالیز می‌شوند. جدول ۱ آمار خلاصه‌ای را برای آنیون‌های اصلی، کاتیون‌های اصلی، مواد جامد محلول کل و هدایت الکتریکی مشاهده‌شده در منطقه مورد مطالعه از سال ۱۹۹۹ تا

۲۰۱۸ ارائه می‌دهد.



انتخاب استان قزوین به عنوان منطقه مطالعه به دلیل محدودیت منابع آب سطحی مانند رودخانه‌ها و چشمه‌ها و وابستگی شدید بخش کشاورزی به منابع آب زیرزمینی صورت گرفت. وسعت اراضی کشاورزی این استان نیازمند حجم بالای آب است و آب زیرزمینی به عنوان منبعی پایدار و قابل اتکا نقش حیاتی در تأمین این نیاز دارد. در بسیاری از مناطق، پایداری آب زیرزمینی نسبت به منابع سطحی آن را به گزینه‌ای مناسب‌تر برای آبیاری تبدیل کرده است. بهره‌برداری اصولی از این منبع می‌تواند ضمن حفظ رونق کشاورزی، خطر کمبود آب در دوره‌های خشکسالی را کاهش دهد.

هدایت الکتریکی (EC) به عنوان یکی از شاخص‌های کلیدی کیفیت آب زیرزمینی در کشاورزی مطرح است. مقدار EC بیانگر میزان مواد محلول در آب بوده و افزایش آن نشان‌دهنده تجمع نمک‌ها و مواد معدنی است که می‌تواند کیفیت آب را کاهش دهد. در حالی که مقادیر مناسب EC بیانگر وجود مواد مغذی مورد نیاز گیاهان است، افزایش بیش از حد آن منجر به شوری آب و خاک، کاهش بازدهی محصولات و اختلال در جذب مواد مغذی توسط گیاهان خواهد شد. بنابراین پایش و کنترل EC برای مدیریت پایدار منابع آب زیرزمینی و تضمین پایداری کشاورزی در استان قزوین ضروری است.

پایگاه داده و دوره ۲۰۱۸-۱۹۹۹ را پوشش می‌دهد. این شامل مقادیر مشاهده‌شده پارامترهای کیفیت آب زیرزمینی برای دو نیمسال اول و دوم سال است. از آنجایی که حجم آب استخراج‌شده از سفره‌های زیرزمینی استان قزوین به دلیل نیاز آبی کشاورزی بیشتر در تابستان بین زمستان و تابستان متفاوت است، در داده‌های مشاهده‌شده تقسیم‌بندی صورت گرفت. این منجر به دو مجموعه داده مجزا شد. مجموعه داده‌ها تحت یک کاوش کامل داده‌ها قرار گرفتند تا یکپارچگی و مناسب بودن آن‌ها برای تجزیه و تحلیل تضمین شود. این شامل بررسی داده‌ها برای ناهنجاری‌ها یا روندها با استفاده از نرم‌افزار آماری SPSS و زبان برنامه‌نویسی پایتون بود. این مرحله برای شناسایی و رسیدگی به هر نقطه داده شدید یا نامنظم که می‌توانست در مراحل جمع‌آوری یا ورود داده‌ها رخ داده باشد، ضروری بود. از روش زیر برای تهیه مجموعه داده‌های لازم برای آموزش و آزمایش مدل‌ها استفاده شد. در ابتدا، فاصله اقلیدسی بین هر چاه رصدی و همه چاه‌های دیگر در شبکه نظارت با استفاده از مختصات UTMx و UTM_y در دسترس چاه‌های مشاهده تعیین شد. پس از آن، یک ماتریس ۳۱۷×۳۱۷ برای نشان دادن این تفاوت‌ها ایجاد شد.

برای هر چاه رصدی در شبکه مانیتورینگ، ۱۰ حلقه چاه نزدیک‌تر شناسایی شد. داده‌های نیمسال‌های اول و دوم طی دوره ۱۹۹۹ تا ۲۰۱۸ استخراج و برای تشکیل نخستین الگو (یک ردیف داده در مجموعه) تحلیل گردید. در این الگو، مقدار هدایت الکتریکی ثبت‌شده در چاه اصلی به عنوان متغیر وابسته در نظر گرفته شد، در حالی که مقادیر هدایت الکتریکی ۱۰ چاه مجاور و فاصله آن‌ها از چاه اصلی به عنوان متغیرهای مستقل تعریف شدند. این روند برای تمامی چاه‌های رصدی در سال نخست دوره آماری تکرار شد و در نتیجه، ۳۱۷ الگو برای مجموعه داده نیمسال اول ایجاد گردید. سپس این فرآیند برای کل دوره ۱۹۹۹ تا ۲۰۱۸ بازاجرای شد تا مجموعه‌ای جامع شامل الگوهای ۱۰×۳۱۷ برای نیمسال اول فراهم شود. مجموعه داده مشابهی نیز برای نیمسال دوم به همین روش تولید شد.

با توجه به اینکه هدف اصلی مطالعه، مدل‌سازی ساختار فضایی هدایت الکتریکی آب‌های زیرزمینی در منطقه مورد بررسی بود، هر الگو در مجموعه داده (مانند مجموعه نیمسال اول) بازتاب‌دهنده بخشی از این ساختار فضایی است. فرض بر این است که شبکه‌های عصبی مصنوعی با دریافت نمونه‌های متعدد قادر به شبیه‌سازی مؤثر این ساختار خواهند بود. همچنین، با توجه به احتمال تفاوت ساختار فضایی بین نیمسال اول و دوم، دو مجموعه داده مستقل تهیه و برای هر یک مدلی جداگانه توسعه داده شد.

^۱GPR یک روش قدرتمند برای پیش‌بینی و رگرسیون است که بر اساس فرآیندهای گاوسی عمل می‌کند. این الگوریتم از توزیع‌های احتمالی برای پیش‌بینی استفاده می‌کند و توانایی مدل‌سازی عدم قطعیت را دارد. یکی از مزایای اصلی GPR این است که می‌تواند تابع‌های پیچیده و غیرخطی را با دقت بالا مدل‌سازی کند (Bai & Tahmasebi, 2023; Williams & Rasmussen, 2006).

^۲GBR یک روش تقویتی^۳ است که از مدل‌های پایه‌ای ضعیف مانند درخت‌های تصمیم استفاده می‌کند و به تدریج مدل‌های جدیدی را به مجموعه اضافه می‌کند تا خطاهای مدل‌های قبلی را اصلاح کند. این الگوریتم معمولاً برای مسائل رگرسیون و طبقه‌بندی استفاده می‌شود و به دلیل دقت بالایی آن در بسیاری از کاربردهای عملی محبوب است (Schulz et al., 2018; Friedman, 2001).

¹ Gaussian Process Regression

² Gradient Boosting Regression

³ boosting



⁴KNN یک الگوریتم ساده و مؤثر برای طبقه‌بندی و رگرسیون است. در این روش، یک نقطه داده جدید با استفاده از نزدیک‌ترین همسایه‌های آن در فضای ویژگی طبقه‌بندی یا پیش‌بینی می‌شود. تعداد همسایه‌ها (k) معمولاً به صورت یک پارامتر انتخاب می‌شود و تأثیر زیادی بر دقت مدل دارد (Natekin & Knoll, 2013; Wu & Wu, 2020).

RF⁵ یک روش مجموعه‌ای⁶ است که از مجموعه‌ای از درخت‌های تصمیم برای انجام طبقه‌بندی یا رگرسیون استفاده می‌کند. هر درخت در جنگل به طور مستقل آموزش داده می‌شود و نتایج آن‌ها با هم ترکیب می‌شود تا پیش‌بینی نهایی حاصل شود. این روش معمولاً دقت بالایی دارد و به طور گسترده در کاربردهای مختلف استفاده می‌شود (Chen et al., 2022; Zhu & Zeng, 2020; Moghaddam et al., 2025).

⁷MLP یک نوع شبکه عصبی مصنوعی است که از لایه‌های چندگانه پرسپترون⁸ برای یادگیری و پیش‌بینی استفاده می‌کند. این شبکه‌ها قادر به یادگیری توابع پیچیده و غیرخطی هستند و به طور گسترده‌ای در مسائل مختلفی مانند طبقه‌بندی، رگرسیون و تشخیص الگو استفاده می‌شوند (Fattahi et al., 2021; Tikhmarine et al., 2020).

به طور کلی مدل‌های شبکه‌های عصبی چندلایه (MLP) در زمینه هیدرولوژی برای وظایف مختلفی مانند مدل‌سازی رواناب باران، پیش‌بینی خشک‌سالی، مدل‌سازی سطح زیرزمینی و دریاچه‌ها، پیش‌بینی رسوبات و تحلیل کیفیت آب به کار گرفته شده‌اند (Lee, 2022; Karbasi, 20203; Fan, et al, 2023).

یکی از مزایای مدل‌های MLP، قابلیت یادگیری از حجم بزرگی از داده‌ها و تعمیم‌پذیری خوب به نمونه‌های ناشناخته است. آن‌ها می‌توانند با متغیرهای ورودی مختلفی از جمله داده‌های هواشناسی، ویژگی‌های خاک و رکورد‌های هیدرولوژی تاریخی کنار بیایند. این مدل‌ها می‌توانند الگوها و ارتباطات در داده‌ها را یاد بگیرند، آن‌ها را قادر می‌سازند که پیش‌بینی‌های دقیقی انجام داده و برای تحلیل هیدرولوژیک ارزشمند باشند. قبلاً ذکر شد که آب زیرزمینی آبی است که از زمین استخراج می‌شود و یکی از منابع مهم تأمین آب در شرایط خشک‌سالی هست، لذا تمام این راه‌کارهای با کمک هوش مصنوعی تأمین‌کننده سهولت بیشتر استخراج آب زیرزمینی خواهد بود (Van Thieu et al., 2023; Moghaddam, 2024; Moghaddam & Salar, 2024).

در مقالات هانون و همکاران⁹ (۲۰۲۱)، و نوردین و همکاران¹⁰ (۲۰۲۱)، به طور جامع روش‌های هوش مصنوعی در مدل‌سازی کیفیت آب زیرزمینی را مرور نمودند. آن‌ها به مرور مقالات بسیاری که بین سال‌های ۲۰۰۱ تا ۲۰۲۱ منتشر شده‌اند متمرکز شدند و بر ویژگی‌ها و توانایی‌های این روش‌های هوش مصنوعی تأکید داشتند. یافته‌ها نشان می‌دهد که با وجود ضعف‌های خاص، روش‌های خوب ساخته شده از هوش مصنوعی قادر به پیش‌بینی مؤثر کیفیت آب زیرزمینی در انواع سفره‌های زیرزمینی هستند. در مطالعه ما، مدل MLP بر اساس کارایی اثبات شده‌اش در مدل‌سازی فرآیندهای پیچیده هیدرولوژی انتخاب شد. مدل‌های MLP برای ضبط روابط غیرخطی بین متغیرهای ورودی و پارامترهای هیدرولوژی مناسب هستند و ارزشمند برای مدل‌سازی کیفیت آب زیرزمینی می‌باشد (Hanoon et al., 2021; Che Nordin, 2021; Ansari & Akhoondzadeh, 2020).

۴- نتایج و بحث مدل‌های اجرا شده

در این بخش، نتایج آزمایش و اعتبارسنجی مدل‌های هوش مصنوعی به کار گرفته شده مورد بررسی قرار می‌گیرد. یکی از اهداف این تحقیق، ارزیابی تأثیر چاه‌های مجاور بر مقدار هدایت الکتریکی چاه هدف بوده است. بدین منظور، مقادیر هدایت الکتریکی (EC) در ۱۰ چاه مجاور اندازه‌گیری شد و فاصله هر یک از آن‌ها تا چاه هدف به عنوان ورودی مدل مورد استفاده قرار گرفت. مدل‌ها با استفاده از ۷۵ درصد داده‌های موجود آموزش داده شدند و ۲۵ درصد باقی‌مانده برای مرحله آزمایش به کار گرفته شد. از این بخش به بعد، نمودارهای ارائه شده، پراکندگی نتایج تخمین مدل‌ها و میزان دقت آن‌ها را نمایش می‌دهند. نتایج به طور جداگانه برای مدل‌هایی گزارش می‌شود که در آن‌ها ۴، ۵، ۶، ۷ و ۸ چاه مجاور در نظر گرفته شده‌اند.

در ادامه، ابتدا نتایج مربوط به مدل با در نظر گرفتن ۴ چاه مجاور ارائه می‌شود.

جدول ۲- مقایسه عملکرد مدل‌های ۴، ۵، ۶، ۷، ۸ حلقه چاه نزدیک در ۶ ماهه‌های اول و دوم سال

⁴ K-Nearest Neighbors

⁵ Random Forest

⁶ ensemble

⁷ Multi-Layer Perceptron

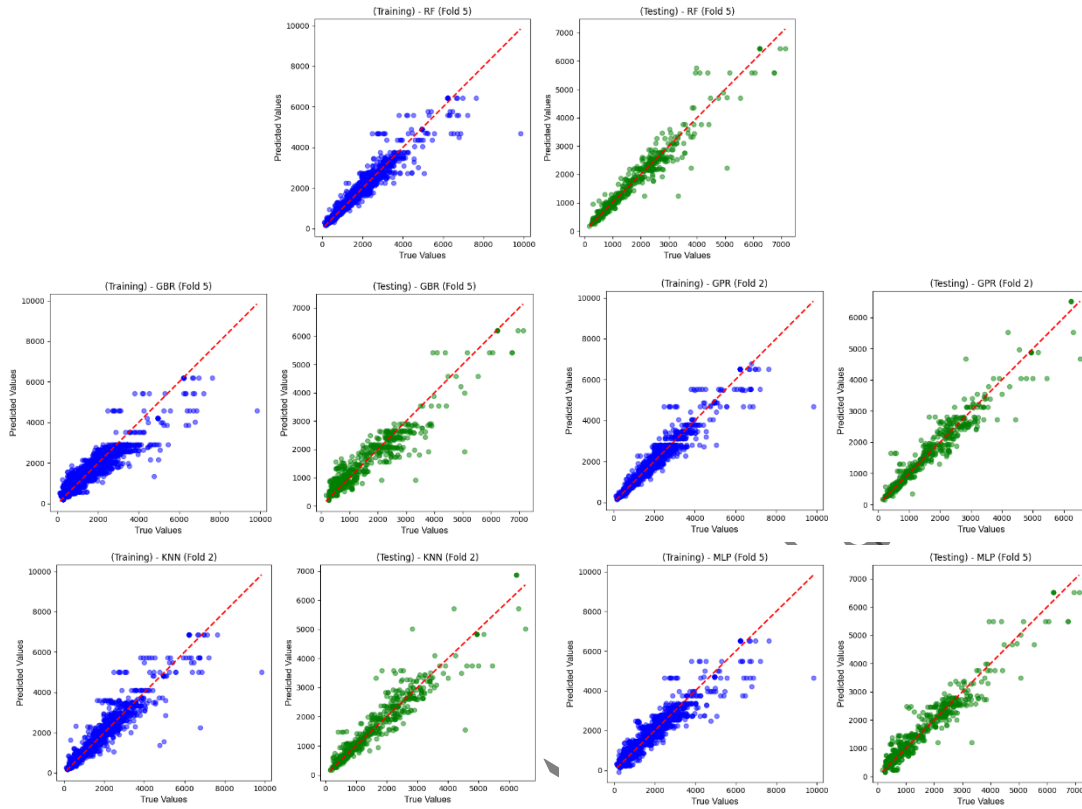
⁸ neurons

⁹ Hanoon et al

¹⁰ Che Nordin et al

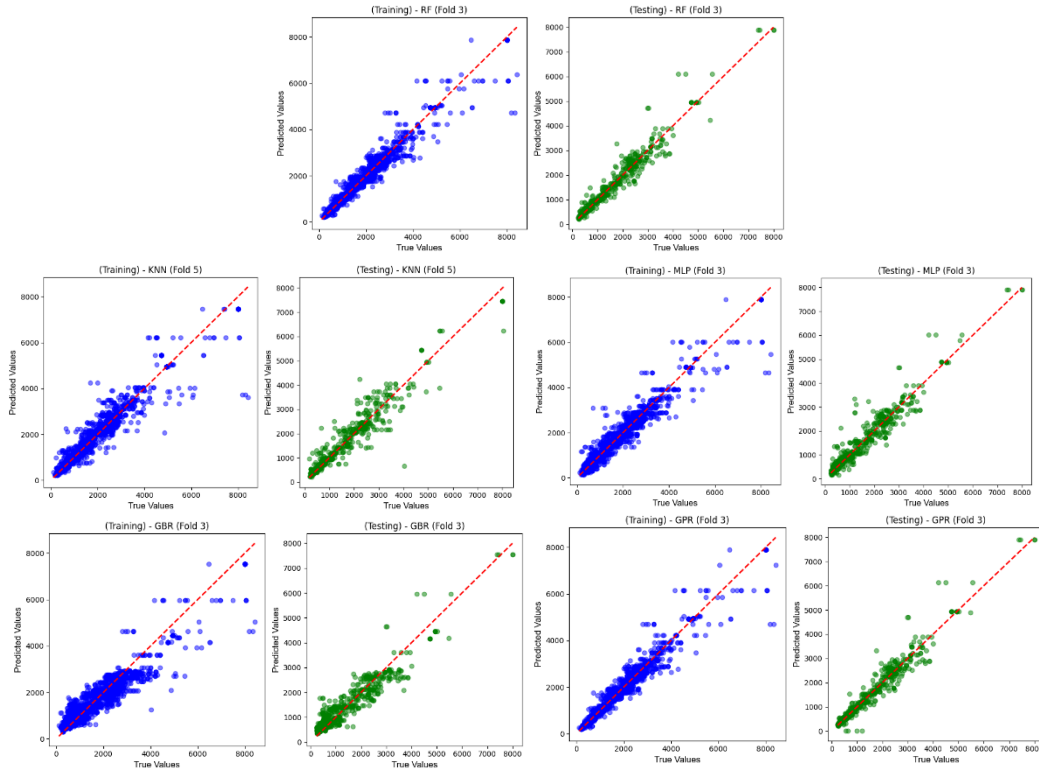


Model	R-squared	Pearson correlation	MAE	RMSE	R-squared	Pearson correlation	MAE	RMSE	Wells
	(Testing)	(testing)	(Testing)	(Testing)	(Training)	(Training)	(Training)	(Training)	
MLP	0.873855761	0.935458731	234.2302694	401.4935318	0.905924227	0.952160621	208.4991646	350.8140232	4WFirst
KNN	0.872505221	0.93689659	198.1963206	406.9572488	0.89615763	0.949562431	173.9116161	367.4299735	
RF	0.915414152	0.957368654	155.8033001	326.3417902	0.942467967	0.970824048	135.0378161	274.4581239	
GBR	0.835625719	0.918657566	294.2649975	461.1418344	0.867026313	0.935537322	275.2508975	417.6322367	
GPR	0.904200194	0.951790591	157.5258192	349.5533518	0.945114432	0.97216743	130.8630756	268.1533679	
MLP	0.873855761	0.935458731	234.2302694	401.4935318	0.905924227	0.952160621	208.4991646	350.8140232	4WSecond
KNN	0.872505221	0.93689659	198.1963206	406.9572488	0.89615763	0.949562431	173.9116161	367.4299735	
RF	0.915414152	0.957368654	155.8033001	326.3417902	0.942467967	0.970824048	135.0378161	274.4581239	
GBR	0.835625719	0.918657566	294.2649975	461.1418344	0.867026313	0.935537322	275.2508975	417.6322367	
GPR	0.904200194	0.951790591	157.5258192	349.5533518	0.945114432	0.97216743	130.8630756	268.1533679	
MLP	0.906163834	0.952992273	190.2349697	359.7736106	0.934077305	0.966587089	166.8246904	302.1278605	5WFirst
KNN	0.861477653	0.932287347	207.4170343	436.5750703	0.895591487	0.948786035	178.9636412	379.3436258	
RF	0.915801299	0.957891634	160.0150398	340.2879657	0.942177768	0.970670038	137.654738	282.9696737	
GBR	0.867484668	0.9355358	268.2839747	427.7779961	0.893544579	0.948747803	250.121005	384.003974	
GPR	0.903692867	0.951551454	162.4779578	361.7848881	0.944554749	0.971881321	133.4008377	277.0731185	
MLP	0.918134744	0.958405808	180.2265057	353.7549092	0.946723209	0.973128859	153.0237222	287.1930086	5WSecond
KNN	0.871644291	0.935380331	207.711333	443.8696933	0.916602414	0.957859208	168.7589357	359.3208103	
RF	0.934922418	0.967095796	148.6528325	316.2795661	0.952902619	0.976194669	127.4101156	269.9814702	
GBR	0.881285938	0.942029107	282.7124203	428.1931013	0.902353613	0.953374673	262.0262696	388.7970737	
GPR	0.933597326	0.966451901	146.7428737	317.7368121	0.955497241	0.977494986	122.872939	262.4565861	
MLP	0.883911726	0.94156607	195.712522	418.9245854	0.93856061	0.968835974	163.1430729	311.5694506	6WFirst
KNN	0.8643339	0.934176744	211.3524256	456.156165	0.907141002	0.954049156	179.7892435	383.0677407	
RF	0.916242656	0.957463893	165.3668708	359.8945837	0.94241059	0.970797522	141.0829831	301.645121	
GBR	0.890631707	0.945645658	244.1749527	413.0121551	0.916721953	0.959481898	222.1998247	362.7748401	
GPR	0.906126591	0.95248461	168.0409251	379.2806415	0.945976294	0.972610571	136.0288775	292.1130268	
MLP	0.909185566	0.955002743	195.3600487	394.4529868	0.945585766	0.97248522	166.0108278	309.7637	6WSecond
KNN	0.888381902	0.943894217	205.3088714	443.3074197	0.909767496	0.955119504	175.7715166	398.593709	
RF	0.931718691	0.965797187	153.4589533	344.8071327	0.952730113	0.976118077	130.5004024	288.7373718	
GBR	0.901383102	0.951713171	253.7238999	414.814687	0.924210913	0.963369634	231.6169044	365.7314627	
GPR	0.929860399	0.965001086	151.8254632	349.7128	0.955865741	0.977682896	125.7763093	279.0364394	
MLP	0.895678968	0.947355114	184.1728298	398.1921501	0.923507752	0.961015973	158.7059467	342.5876525	7WFirst
KNN	0.798779213	0.910994594	228.5520917	545.6428855	0.858912105	0.937708423	193.4169848	463.5459623	
RF	0.899697169	0.949728462	168.6563213	390.2625415	0.927577709	0.963119353	142.8324768	333.3564959	
GBR	0.895456978	0.947464162	201.5573842	398.1633972	0.919632564	0.959392583	179.5264691	351.14814	
GPR	0.892612594	0.946226241	169.5168766	403.163604	0.92949208	0.964098398	138.4151691	328.8756317	
MLP	0.895678968	0.947355114	184.1728298	398.1921501	0.923507752	0.961015973	158.7059467	342.5876525	7WSecond
KNN	0.798779213	0.910994594	228.5520917	545.6428855	0.858912105	0.937708423	193.4169848	463.5459623	
RF	0.899697169	0.949728462	168.6563213	390.2625415	0.927577709	0.963119353	142.8324768	333.3564959	
GBR	0.895456978	0.947464162	201.5573842	398.1633972	0.919632564	0.959392583	179.5264691	351.14814	
GPR	0.892612594	0.946226241	169.5168766	403.163604	0.92949208	0.964098398	138.4151691	328.8756317	
MLP	0.826172719	0.912881999	259.5234715	542.4176813	0.88695363	0.941813452	211.1966972	453.7440229	8WFirst
KNN	0.726600038	0.898033492	317.875873	682.3643697	0.777108259	0.914378996	277.3436508	635.4189937	
RF	0.838741344	0.920909384	239.2140559	526.047036	0.888667655	0.942692611	199.3999061	450.0860262	
GBR	0.839919516	0.921225141	245.3056186	524.3930698	0.890882103	0.943917006	202.9426841	445.6858504	
GPR	0.827864973	0.915401182	235.5742559	543.2867012	0.892554545	0.944737143	189.924158	442.2702565	
MLP	0.84938249	0.928105085	224.8396736	487.9222244	0.902731424	0.950154948	173.5337773	406.0312687	8WSecond
KNN	0.721139734	0.892183888	315.8907191	664.2698723	0.805186127	0.915962314	254.8661183	573.8160002	
RF	0.862973223	0.934735055	207.9093099	469.9617812	0.901867167	0.949705112	171.6081757	407.9053833	
GBR	0.866253763	0.935979052	210.1878015	463.5149262	0.905422314	0.951597477	172.3586583	400.3591924	
GPR	0.838414102	0.92328085	214.1785001	505.2193836	0.906761951	0.952232812	162.5919365	397.5308084	



شکل ۴-الف- نتایج ۶ ماهه اول سال مربوط به ۴ چاه همسایگی

بررسی‌های جدول ۲ و نتایج نشان می‌دهد که مدل‌های RF و GBR بهترین پاسخ‌های ممکن را برای ما داده‌اند. در جدول ۲ بخش ۴ حلقه چاه، مدل جنگل تصادفی (RF) عملکرد کلی بهتری از مدل GPR ارائه می‌دهد. در مدل RF، ضریب تعیین تست برابر با 0.9154 و ضریب همبستگی پیرسون 0.9515 است که نشان‌دهنده دقت و پایداری بالا در پیش‌بینی می‌باشد و RMSE آن نیز $(344/04)$ نسبتاً پایین است. در مقابل، مدل GPR با R-squared برابر با 0.904 و کمترین MAE $(159/59)$ عملکرد مناسبی دارد؛ اما RMSE بالای آن $(528/12)$ نشان می‌دهد که در برخی موارد خطاهای بزرگی رخ داده است. به‌طور خلاصه، اگرچه هر دو مدل از نظر دقت قابل‌قبول هستند، مدل RF با تعادل بهتر بین دقت و پایداری، گزینه مطلوب‌تری به شمار می‌آید.



شکل ۴-ب- نتایج ۶ ماهه دوم سال مربوط به ۴ چاه همسایگی

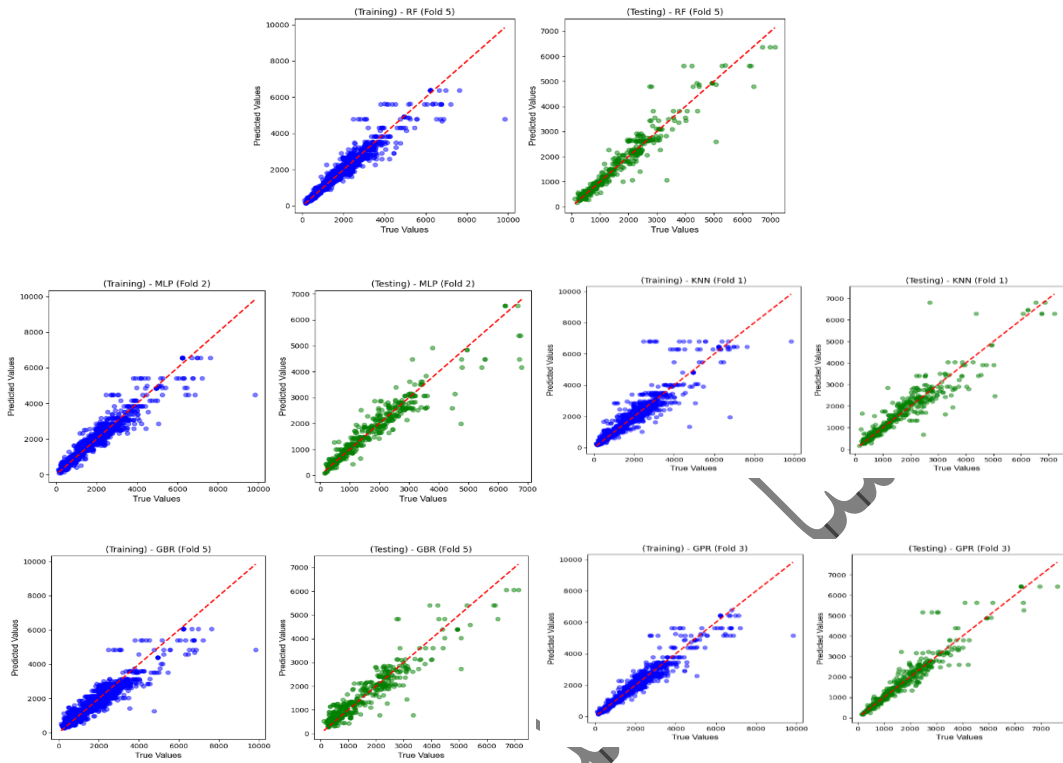
۴-۱- تحلیل آمارهای عملکردی برای ۴ حلقه چاه اطراف

در این تحقیق، عملکرد مدل‌های مختلف یادگیری ماشین برای پیش‌بینی داده‌های ۶ ماهه اول و دوم سال مورد بررسی قرار گرفته است. مدل‌های مورد استفاده شامل RF (Random Forest)، KNN، MLP، GBR و GPR بودند و معیارهای ارزیابی شامل R-squared، ضریب همبستگی پیرسون، MAE و RMSE در دو بخش آموزش و آزمون بودند. نتایج بررسی جداول نشان می‌دهد که در هر دو بازه زمانی، مدل RF (Random Forest) بهترین عملکرد را داشته است. این مدل در ۶ ماهه اول سال مقدار R-squared برابر ۰/۹۱۵۴ در آزمون و ۰/۹۴۲۵ در آموزش را نشان داده که از سایر مدل‌ها بالاتر است. همچنین، مقدار ضریب همبستگی پیرسون در آزمون برابر ۰/۹۵۷۳ و در آموزش ۰/۹۷۰۸ بوده که نشان‌دهنده ارتباط قوی بین داده‌های واقعی و پیش‌بینی شده است. علاوه بر این، میزان خطاهای MAE ۱۵۵/۸۰ و RMSE ۳۲۶/۳۴ در آزمون، کمترین مقدار در میان مدل‌های بررسی شده است. در ۶ ماهه دوم سال نیز مدل RF همچنان بهترین عملکرد را داشته و مقادیر R-square، ضریب همبستگی پیرسون و میزان خطاهای MAE و RMSE تقریباً مشابه با ۶ ماهه اول سال است. سایر مدل‌ها مانند GPR و KNN نیز عملکرد نسبتاً خوبی داشته‌اند اما دقت آن‌ها به اندازه مدل RF بالا نبوده است. با مقایسه عملکرد مدل‌ها در دو دوره ۶ ماهه، مشخص می‌شود که نتایج ۶ ماهه اول و دوم تقریباً یکسان هستند و تفاوت محسوسی بین آن‌ها مشاهده نمی‌شود. با این حال، در برخی موارد مانند مقدار RMSE و MAE تغییرات جزئی دیده می‌شود که می‌تواند ناشی از شرایط محیطی، تغییرات فصلی، و توزیع داده‌های آموزشی و تستی در دو دوره مختلف باشد. در بسیاری از مطالعات محیطی، معمولاً در ۶ ماهه دوم سال به دلیل کاهش بارندگی، افزایش تبخیر و تغییر در میزان آب‌های سطحی و زیرزمینی، دقت مدل‌ها اندکی کاهش پیدا می‌کند، اما در این بررسی، این کاهش دقت چندان محسوس نبوده است (Luk et al., 2001). در نتیجه، مدل RF در هر دو بازه زمانی بهترین عملکرد را داشته و از لحاظ کلی، تفاوت محسوسی بین دقت مدل‌ها در ۶ ماهه اول و دوم سال مشاهده نمی‌شود.

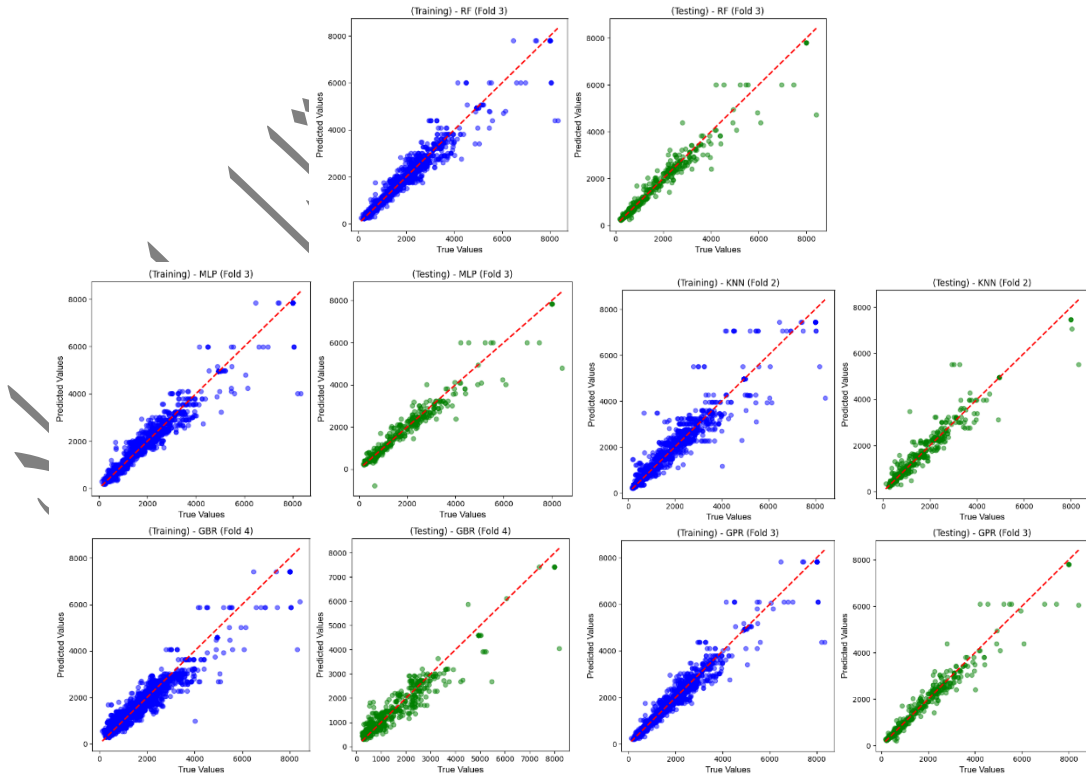


تحقیقات منابع آب ایران

انجمن علوم و مهندسی منابع آب ایران



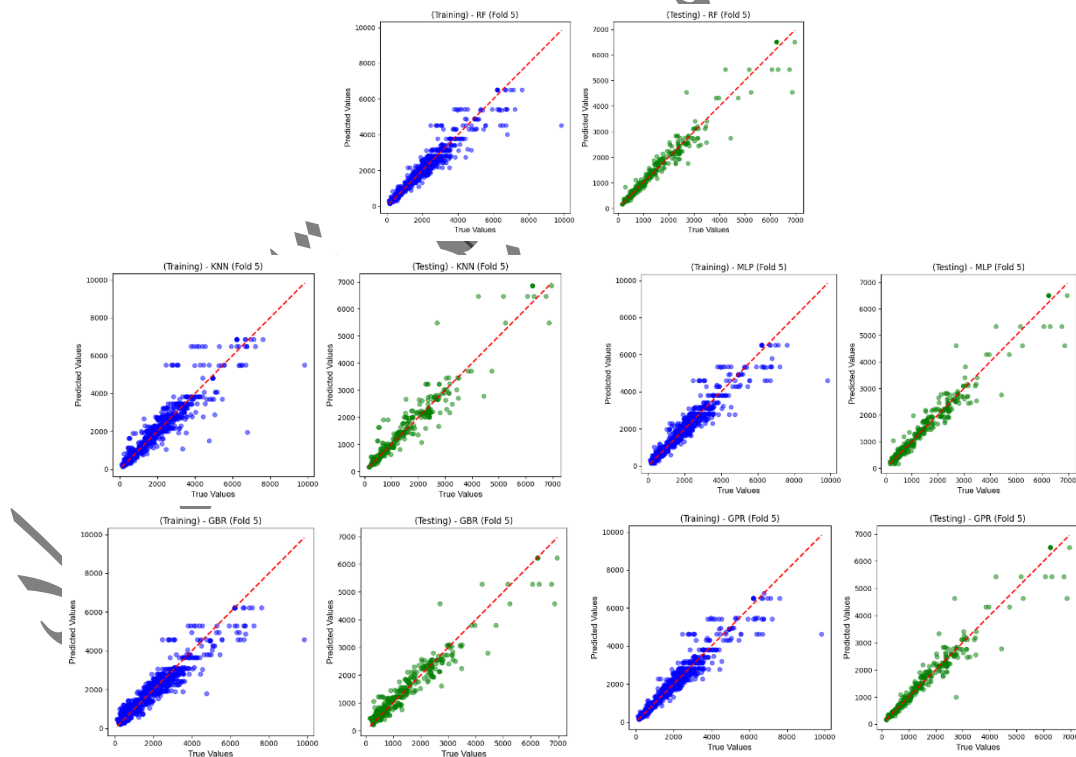
شکل ۵-الف - نتایج ۶ ماهه اول سال مربوط به ۵ چاه همسایگی



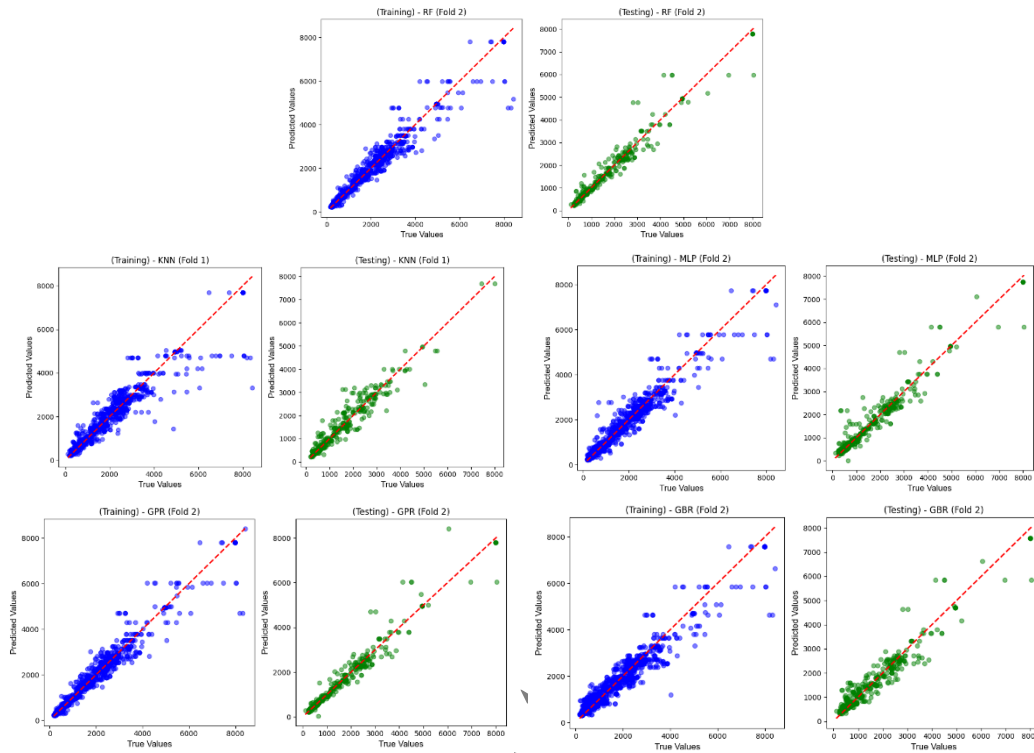
شکل ۵-ب - نتایج ۶ ماهه دوم سال مربوط به ۵ چاه همسایگی

۴-۲- تحلیل آمارهای عملکردی برای ۵ حلقه چاه اطراف

جدول ۲ عملکرد پنج مدل یادگیری ماشین MLP، KNN، RF، GBR، GPR را در پیش‌بینی عملکرد چاه‌ها در ۶ ماه اول سال بررسی می‌کند. معیارهای ارزیابی شامل R-squared، همبستگی پیرسون، MAE و RMSE برای داده‌های آزمون و آموزش هستند. در فاز آزمون، مدل RF با مقدار R-squared برابر ۰/۹۱۵۸ بهترین عملکرد را دارد، در حالی که مدل GPR بالاترین مقدار همبستگی پیرسون (۰/۹۵۱۶) را نشان می‌دهد. همچنین، مدل RF کمترین MAE برابر ۱۶۰/۰۱ یعنی یکی از کمترین مقادیر، همچنین RMSE برابر ۳۴۰/۲۸ را دارد که نشان‌دهنده دقت مناسب پیش‌بینی‌های آن است. در فاز آموزش، مدل GPR با مقدار R-squared برابر ۰/۹۴۴۶ و RMSE برابر ۲۷۷/۰۷ بهترین عملکرد را دارد. در جدول ۲ که مربوط به ۶ ماه دوم سال است، همان مدل‌ها بررسی شده‌اند. در فاز آزمون، مدل RF عملکرد بهتری نسبت به دوره قبلی داشته و مقدار R-squared آن به ۰/۹۳۴۹ افزایش یافته است. هم‌زمان، مدل GPR همچنان بالاترین مقدار همبستگی پیرسون (۰/۹۶۶۴) را دارد و مدل RF دارای کمترین مقدار MAE ۱۴۸/۶۵ و RMS ۳۱۶/۲۷ است. در فاز آموزش نیز، مدل GPR با مقدار R-squared برابر ۰/۹۵۵۴ و RMSE برابر ۲۶۲/۴۶ بهترین عملکرد را دارد. با مقایسه دو جدول، مشخص می‌شود که مدل‌ها در نیمه دوم سال عملکرد بهتری دارند. مقدار R-squared مدل RF از ۰/۹۱۵۸ به ۰/۹۳۴۹ در فاز آزمون افزایش یافته که نشان‌دهنده بهبود دقت پیش‌بینی‌ها در این دوره است. همچنین، مقدار R-squared مدل GPR از ۰/۹۰۳۶ به ۰/۹۳۳۵ و مقدار RMSE آن از ۳۶۱/۷۸ به ۳۱۷/۷۳ کاهش یافته است. این بهبودها نشان می‌دهند که دقت مدل‌ها در ۶ ماه دوم افزایش پیدا کرده است. علاوه بر این، رتبه‌بندی مدل‌ها در هر دو دوره ثابت باقی‌مانده است، به طوری که مدل‌های RF و GPR در هر دو دوره بهترین عملکرد را داشته‌اند، در حالی که مدل GBR ضعیف‌ترین نتایج را ارائه داده است. در نتیجه، می‌توان گفت که مدل‌ها در ۶ ماه دوم سال عملکرد بهتری دارند. این مسئله می‌تواند ناشی از تغییرات فصلی، دسترسی به داده‌های بهتر، یا بهبود توانایی تعمیم مدل‌ها در طول زمان باشد.



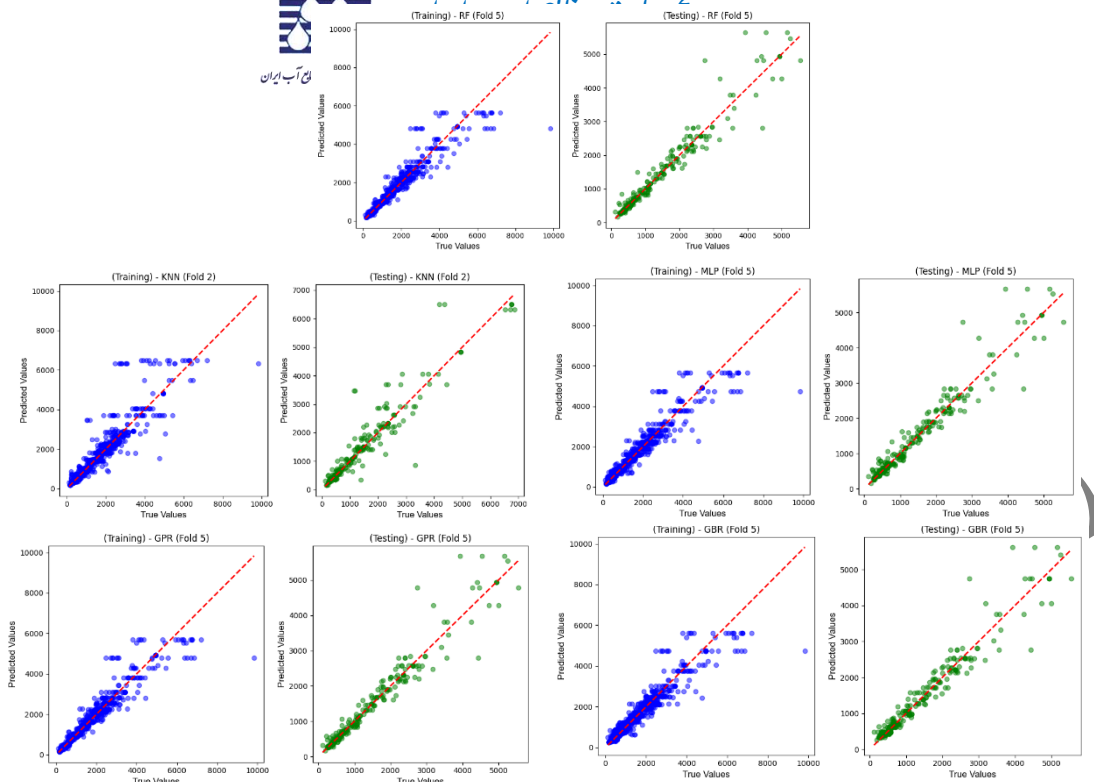
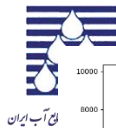
شکل ۶-الف- نتایج ۶ ماهه اول سال مربوط به ۶ چاه همسایگی



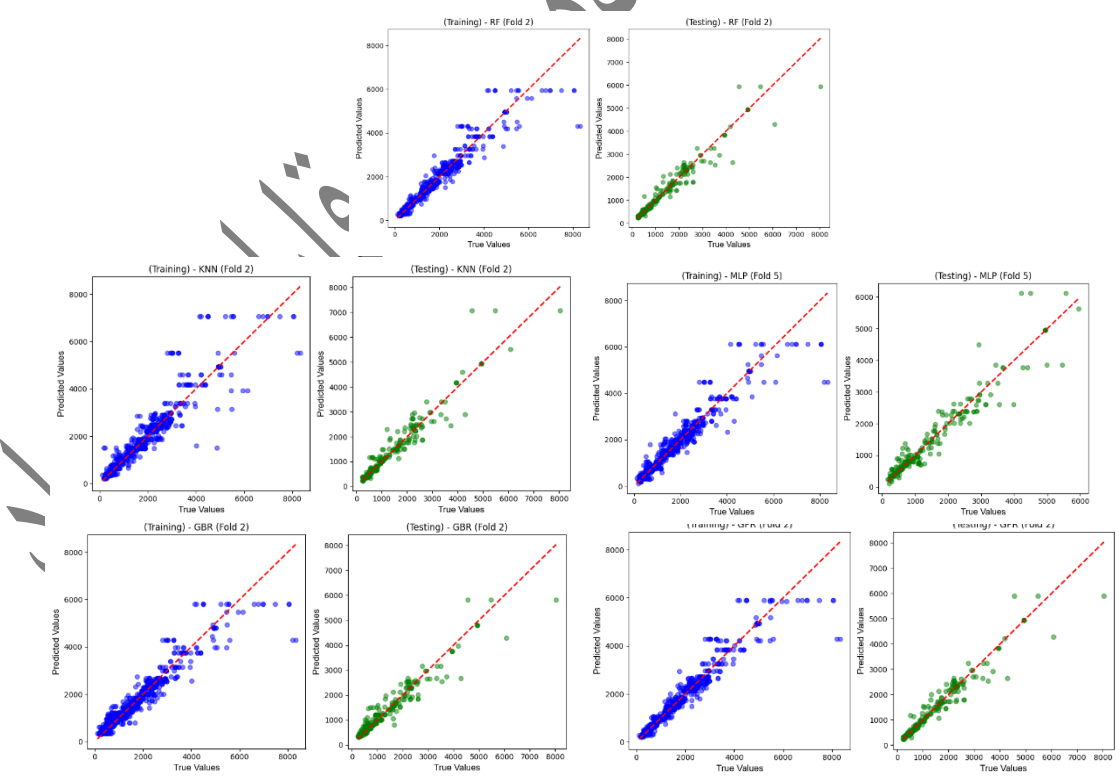
شکل ۶-ب- نتایج ۶ ماهه دوم سال مربوط به ۶ چاه همسایگی

۳-۴- تحلیل آمارهای عملکردی برای ۶ حلقه چاه اطراف

بررسی عملکرد مدل‌های یادگیری ماشین در جدول ۲ نشان می‌دهد که برای ۶ ماهه اول سال، مدل جنگل تصادفی (RF) با R^2 برابر با ۰/۹۱۶۴ در آزمون و ۰/۹۴۲۴ در آموزش، عملکرد بهتری نسبت به سایر مدل‌ها داشته است. همچنین این مدل دارای کمترین مقدار MAE ۱۶۵/۳۷ و $RMSE$ ۳۵۹/۸۹ در داده‌های آزمون است که نشان‌دهنده دقت بالایی آن است. مدل GPR نیز عملکرد بسیار خوبی داشته و در برخی معیارها نزدیک به RF است، اما در مجموع برتر به نظر می‌رسد. در ۶ ماهه دوم سال، عملکرد مدل‌ها بهبود یافته و مدل RF همچنان بهترین عملکرد را نشان می‌دهد، زیرا R^2 آن برابر با ۰/۹۳۱۷ در آزمون و ۰/۹۵۲۷ در آموزش است. همچنین MAE ۱۵۳/۴۶ و $RMSE$ ۳۴۴/۸۰ در داده‌های آزمون برای این مدل، بهترین مقادیر در بین تمام مدل‌ها هستند. مدل GPR نیز در این دوره عملکرد بسیار نزدیکی به RF دارد، اما همچنان RF دقت بهتری ارائه می‌دهد. در مجموع، مقایسه مدل‌ها در جدول ۲ نشان می‌دهد که مدل RF در هر دو دوره ۶ ماهه بهترین عملکرد را داشته است و می‌توان آن را بهترین مدل در بین مدل‌های مقایسه شده دانست. با این حال، باید توجه داشت که دقت مدل‌ها در ۶ ماهه دوم سال نسبت به ۶ ماهه اول بهبود یافته است. این بهبود می‌تواند به دلیل وجود داده‌های آموزشی بیشتر، توزیع بهتر داده‌ها، تغییرات محیطی تأثیرگذار یا تغییرات فصلی در پدیده‌های مربوط به سطح آب زیرزمینی باشد.



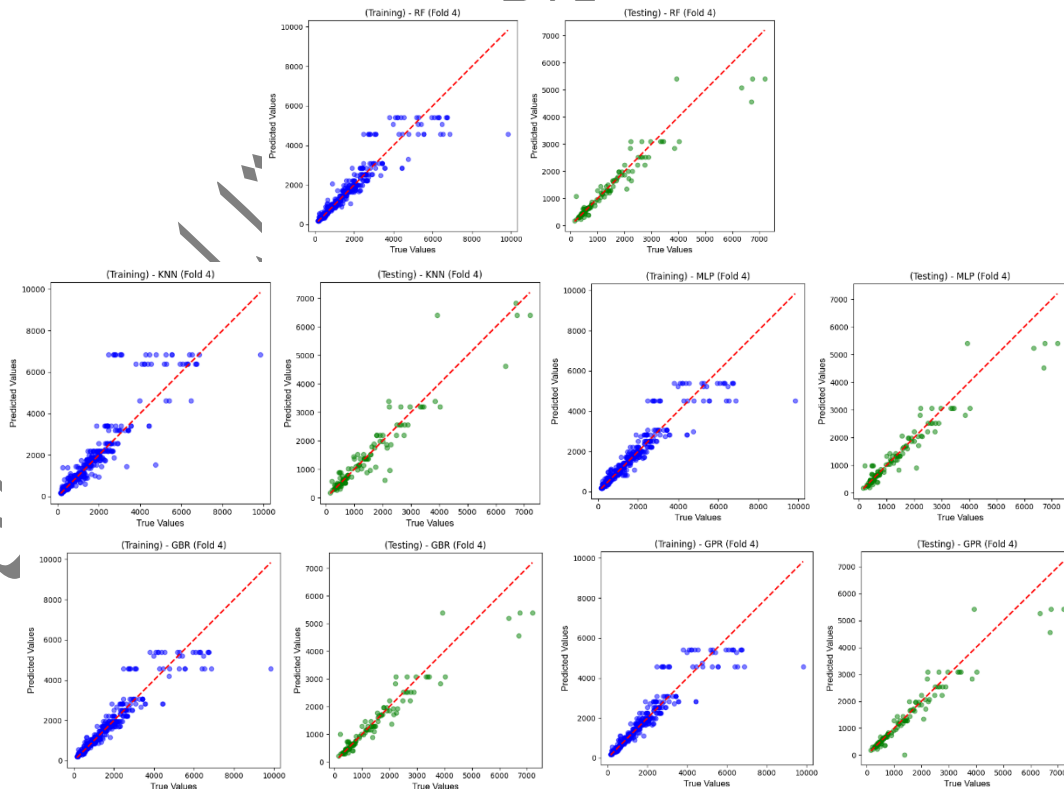
شکل ۷-الف- نتایج ۶ ماهه اول سال مربوط به ۷ چاه همسایگی



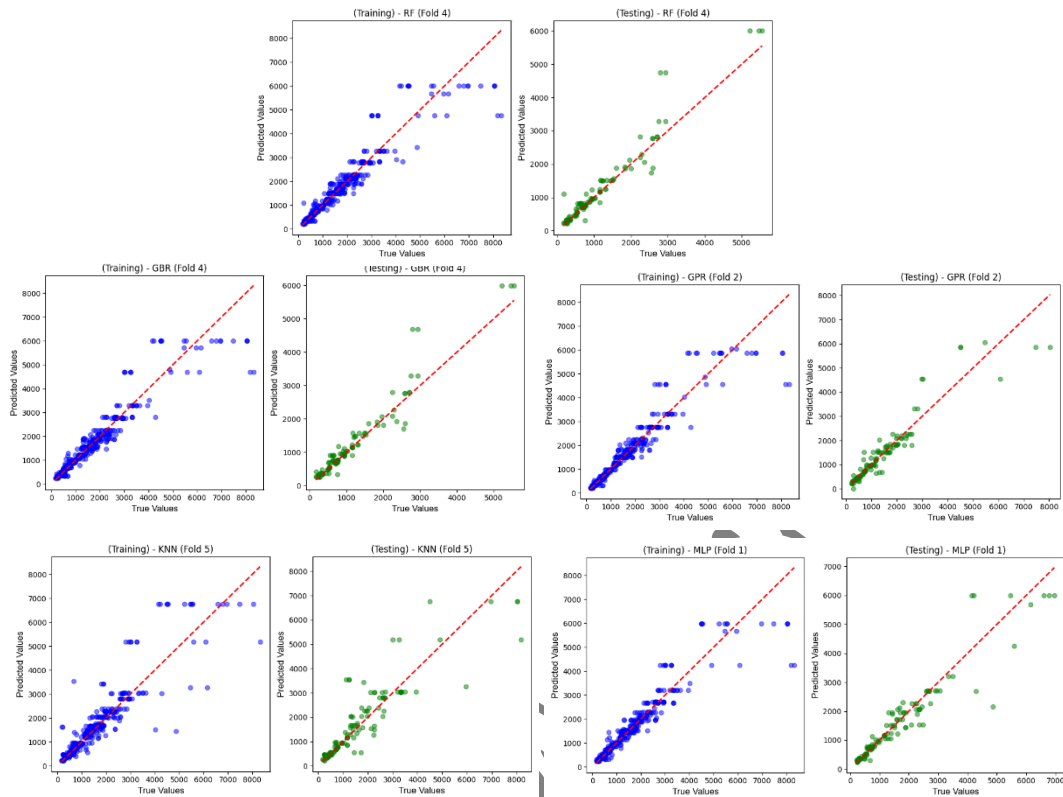
شکل ۷-ب- نتایج ۶ ماهه دوم سال مربوط به ۷ چاه همسایگی

۴-۴- تحلیل آمارهای عملکردی برای ۷ حلقه چاه اطراف

تحلیل عملکرد مدل‌های مختلف در پیش‌بینی سطح آب زیرزمینی در دو نیمه اول و دوم سال بر اساس جدول ۲ ارائه شده نشان می‌دهد که مدل‌های یادگیری ماشین در هر دو بازه عملکرد متفاوتی داشته‌اند. مقایسه مقادیر R^2 ، همبستگی پیرسون، MAE و RMSE نشان می‌دهد که در نیمه اول سال، مدل RF بالاترین مقدار R^2 را در داده‌های آزمون دارد (۰.۸۹۹۷) و مدل MLP نیز عملکرد نسبتاً مشابهی را نشان داده است (۰.۸۹۵۷). با این حال، در نیمه دوم سال، مقدار R^2 در اکثر مدل‌ها کاهش یافته است، به طوری که مدل RF همچنان بهترین عملکرد را داشته اما مقدار R^2 آن اندکی کاهش پیدا کرده (۰.۸۹۶۰). در مقابل، مدل KNN که در نیمه اول مقدار R^2 کمتری نسبت به سایر مدل‌ها داشت، در نیمه دوم بهبودی نسبی داشته است. در مورد همبستگی پیرسون، مقادیر در نیمه اول معمولاً بیشتر از نیمه دوم هستند که نشان‌دهنده پراکندگی کمتر داده‌ها و قابلیت پیش‌بینی بالاتر در این بازه زمانی است. این مسئله می‌تواند ناشی از تغییرات فصلی و پایداری بیشتر شرایط هیدرولوژیکی در نیمه اول سال باشد. در نیمه دوم سال، کاهش میزان همبستگی در برخی مدل‌ها نظیر GPR و KNN مشاهده می‌شود که احتمالاً بیانگر افزایش نوسانات داده‌ها و تغییر در الگوهای رفتاری سطح آب زیرزمینی در این بازه است. از نظر میزان خطا، مقادیر MAE و RMSE در نیمه دوم معمولاً بیشتر از نیمه اول هستند، به این معنی که مدل‌ها در نیمه دوم دقت کمتری داشته و میزان خطای پیش‌بینی افزایش یافته است. به ویژه مدل‌های MLP و KNN در نیمه دوم مقدار RMSE بالاتری دارند که نشان می‌دهد این مدل‌ها در نیمه دوم توانایی پیش‌بینی ضعیف‌تری نسبت به نیمه اول دارند. با این حال، مدل RF همچنان دقت بهتری نسبت به سایر مدل‌ها نشان داده و مقدار RMSE آن در نیمه دوم کمتر از دیگر مدل‌هاست. مدل GPR نیز بهبود نسبی در نیمه دوم داشته، زیرا مقدار RMSE و MAE آن در مقایسه با نیمه اول کاهش یافته است. به طور کلی، این نتایج نشان می‌دهند که عملکرد مدل‌های پیش‌بینی در نیمه دوم سال به دلیل افزایش نوسانات داده‌ها و تأثیرات فصلی دچار افت شده است. این کاهش دقت می‌تواند به عواملی نظیر تغییرات بارندگی، تبخیر، و تغییر در میزان برداشت آب از چاه‌ها مرتبط باشد. برای بهبود عملکرد مدل‌ها در نیمه دوم سال، پیشنهاد می‌شود از متغیرهای اضافی که تغییرات فصلی را بهتر توصیف می‌کنند استفاده شود. همچنین می‌توان روش‌های بهینه‌سازی مدل را بررسی کرد تا میزان خطا در نیمه دوم کاهش یابد.



شکل ۸-الف- نتایج ۶ ماهه اول سال مربوط به ۸ چاه همسایگی



شکل ۸-ب- نتایج ۶ ماهه دوم سال مربوط به ۸ چاه همسایگی

۴-۵-۴- تحلیل آمارهای عملکردی برای ۸ حلقه چاه اطراف

۱-۴-۵- تحلیل عملکرد مدل‌ها در ۶ ماهه نخست سال

در این بازه زمانی، مدل‌های مختلف از جمله MLP، KNN، RF، GBR و GPR مورد بررسی قرار گرفته‌اند. بر اساس شاخص R-squared که میزان تطابق پیش‌بینی‌ها با داده‌های واقعی را نشان می‌دهد، مدل GBR (0.8399) و RF (0.8387) بهترین عملکرد را در مجموعه‌ی آزمون داشته‌اند. همچنین مقدار RMSE که نشان‌دهنده‌ی میزان خطای ریشه میانگین مربعات است، برای این دو مدل به ترتیب ۵۲۴/۳۹ و ۵۲۶/۰۴ بوده که در مقایسه با سایر مدل‌ها کمتر است، بنابراین این مدل‌ها دقت بهتری در پیش‌بینی داشته‌اند. در همین حال، مدل MLP نیز با مقدار $R^2=0.8261$ و RMSE برابر ۵۴۲.۴۱ عملکرد قابل قبولی نشان داده است. از طرفی، مدل KNN با $R^2=0.7266$ و مقدار $RMSE=682.36$ دارای کمترین دقت در میان مدل‌های مقایسه شده است.

۲-۴-۵- تحلیل عملکرد مدل‌ها در ۶ ماهه دوم سال

در این بازه زمانی، به طور کلی، عملکرد مدل‌ها بهبود یافته است. بهترین مدل‌ها از نظر مقدار R^2 همچنان GBR (0.8662) و RF (0.8629) هستند که مقادیر R^2 بیشتری در مقایسه با نیمه‌ی نخست سال داشته‌اند. در این بازه نیز مدل KNN با مقدار $R^2=0.7211$ ضعیف‌ترین عملکرد را نشان داده است. شاخص RMSE در این دوره نیز برای مدل‌های برتر نسبت به ۶ ماهه نخست کاهش یافته است، به ویژه مدل‌های GBR و RF که مقدار ۴۶۳.۵۱ و ۴۶۹.۹۶ را به خود اختصاص داده‌اند، که نشان‌دهنده‌ی پیش‌بینی دقیق‌تر نسبت به نیمه‌ی نخست سال است.

۳-۴-۵- مقایسه کلی عملکرد مدل‌ها در کل سال



با توجه به نتایج دو جدول، مدل‌های GBR و RF به‌عنوان دقیق‌ترین مدل‌ها در هر دو نیمه‌ی سال شناسایی شده‌اند. مدل GBR با داشتن مقادیر بالاتر R^2 و مقادیر کمتر RMSE در هر دو دوره عملکرد مطلوب‌تری نسبت به سایر مدل‌ها نشان داده است. مدل MLP نیز عملکرد نزدیکی به این دو مدل داشته، اما در هر دو دوره مقدار RMSE بیشتری داشته که دقت پایین‌تری را نشان می‌دهد. از سوی دیگر، مدل KNN در هر دو بازه ضعیف‌ترین مدل بوده و به نظر می‌رسد برای این نوع داده‌ها چندان مناسب نیست.

۴-۵-۴- مقایسه کیفیت داده‌های نیمه‌ی نخست و دوم سال

با مقایسه‌ی عملکرد مدل‌ها در ۶ ماهه اول و دوم سال، مشاهده می‌شود که مقدار R^2 برای تمام مدل‌ها در نیمه‌ی دوم سال افزایش یافته است. این موضوع نشان می‌دهد که در نیمه‌ی دوم سال، مدل‌ها توانایی بیشتری در یادگیری روابط میان متغیرها داشته‌اند و داده‌ها احتمالاً ساختار منظم‌تری داشته‌اند. همچنین، مقدار RMSE در نیمه‌ی دوم سال کمتر از نیمه‌ی اول بوده که نشان‌دهنده‌ی دقت بالاتر پیش‌بینی‌ها در این بازه است. این مسئله می‌تواند به عوامل مختلفی مانند کاهش نویز در داده‌ها، تغییرات فصلی، یا وجود الگوهای مشخص‌تر در نیمه‌ی دوم سال مرتبط باشد. در نتیجه، می‌توان گفت داده‌های نیمه‌ی دوم سال برای مدل‌سازی و پیش‌بینی مناسب‌تر بوده‌اند.

۵- نتایج کلی پژوهش

در بررسی یکپارچه عملکرد مدل‌های یادگیری ماشین برای پیش‌بینی داده‌های آب زیرزمینی در چاه‌های همسایگی با تعدادهای مختلف (۴، ۵، ۶، ۷ و ۸) در نیمه‌های اول و دوم سال، نتایج حاکی از آن است که مدل Random Forest در اکثر سناریوها عملکرد برتری از خود نشان داده است. در داده‌های مربوط به ۴ حلقه چاه همسایگی، مدل RF با داشتن بالاترین مقدار R^2 در مرحله آزمون و آموزش و همچنین کمترین مقادیر خطاهای MAE و RMSE نسبت به سایر مدل‌ها مانند MLP، KNN، GBR و GPR انتخاب شده است. مشابه این روند در نتایج ۵ و ۶ چاه همسایگی نیز مشاهده می‌شود؛ به‌طوری‌که اگرچه مدل GPR در برخی از شاخص‌ها مانند ضریب همبستگی پیرونی نتایج خوبی ارائه داده است، عملکرد کلی RF در هر دو دوره ۶ ماهه به‌واسطه پایداری و دقت بالای پیش‌بینی‌ها به‌عنوان بهترین مدل برجسته می‌شود. در سناریوی ۷ چاه همسایگی، اگرچه RF همچنان در رتبه اول باقی‌مانده است، اما کاهش اندکی در مقادیر R^2 و افزایش مقادیر خطا به‌ویژه در نیمه دوم سال نشان‌دهنده تأثیر نوسانات بیشتر داده‌ها است. در نهایت، در مجموعه ۸ چاه همسایگی نیز نتایج به‌گونه‌ای است که هرچند مدل GBR در برخی از شاخص‌ها عملکرد نزدیکی به RF داشته، اما ترکیب نتایج نشان می‌دهد که RF و GBR به‌عنوان دقیق‌ترین مدل‌ها شناخته شده‌اند؛ همچنین مشاهده می‌شود که در نیمه دوم سال، به دلیل بهبود ساختار داده‌ها و کاهش نویز، دقت پیش‌بینی‌ها نسبت به نیمه نخست افزایش یافته است.

یافته‌های این پژوهش با نتایج گزارش شده در مطالعات پیشین هم‌خوانی دارد. عاقلی و همکاران^{۱۱} نشان دادند که مدل‌های مبتنی بر فاصله، به‌ویژه الگوریتم‌های یادگیری ماشین مانند RF، توانایی بالایی در پیش‌بینی پارامترهای کیفی آب زیرزمینی دارند و افزایش تعداد نقاط مجاور موجب بهبود دقت مدل می‌شود. امامی و همکاران^{۱۲} نیز گزارش کردند که GPR و الگوریتم‌های تکاملی در مدل‌سازی رسانایی الکتریکی عملکرد دقیق‌تری نسبت به MLP دارند؛ نتیجه‌ای که با یافته‌های حاضر سازگار است، زیرا در این تحقیق نیز MLP عملکردی ضعیف‌تر از RF و GPR نشان داد.

رجائی و پوراصلان^{۱۳}، بیان کردند که مدل‌های شبکه عصبی و روش‌های هوشمند نسبت به مدل‌های کلاسیک توانایی بیش‌تری در بازسازی الگوهای زمانی-مکانی آب زیرزمینی دارند. مقایسه این نتایج با مطالعه حاضر نشان می‌دهد که استفاده از داده‌های چاه‌های مجاور (۴ تا ۸ حلقه) موجب می‌شود مدل‌های RF و GPR ساختار فضایی هدایت الکتریکی را بهتر از سایر مدل‌ها بازنمایی کنند. این نتیجه با یافته‌های محمدی و همکاران نیز همسو است؛ آن‌ها نشان دادند که RF در شرایط پیچیده آبخوان، حتی با داده‌های ناقص یا پرنویز، پایداری بیشتری نسبت به MLP دارد.

به‌طور کلی، نتایج این تحقیق در امتداد پژوهش‌های گذشته قرار می‌گیرد و تأیید می‌کند که الگوریتم‌های مبتنی بر درخت تصمیم و مدل‌های فرآیند گاوسی، به دلیل توانایی در مدیریت روابط غیرخطی و ساختار فضایی داده‌ها، نسبت به مدل‌های مبتنی بر شبکه عصبی کارایی بالاتری در پیش‌بینی EC در آبخوان‌های نیمه‌پیچیده دارند.

¹¹ Agheli, E. et al. (2024).

¹² Emami. Er al. (2018).

¹³ Rajaei, T., & Poursalan, F. (2015).

- Agheli E, Ashrafzadeh A, Janatrostami S (2024) Modeling the spatial dependence of quality data using distance-based inputs and a data-driven model. *Groundwater for Sustainable Development* 24:101061. <https://doi.org/10.1016/j.gsd.2023.101061>.
- Ansari M, Akhoondzadeh M (2020) Mapping water salinity using Landsat-8 OLI satellite images (Case study: Karun basin located in Iran). *Advances in Space Research* 65:1490–1502. <https://doi.org/10.1016/j.asr.2019.12.007>.
- Askari J, Egdernezhad A (2022) Groundwater modeling using artificial intelligence methods (Case study: Dezful-Andimeshk plain). *Journal of Research in Environmental Health* 8:160–171. <https://doi.org/10.22038/jreh.2022.61396.1457>
- Bai T, Tahmasebi P (2023) Graph neural network for groundwater level forecasting. *Journal of Hydrology* 616:128792. <https://doi.org/10.1016/j.jhydrol.2022.128792>.
- Banihashemi SM, Eslamian SS, Nazari B (2021) Prediction of local alterations in temperature and precipitation caused by climate change and drought investigation using SPI and SPEI indices in Qazvin plain, Iran. *Journal of Water and Soil Science* 25:25–44. <https://doi.org/10.47176/jwss.25.2.41271>.
- Barzoo F, Ramazani-Etedali H, Kaviani A (2022) Prediction and analysis of minimum and maximum temperature and precipitation data of Qazvin synoptic station (2021–2100) using LARS-WG model and DKRZ database. *Nivar Journal*. [10.30467/nivar.2022.347939.1220](https://doi.org/10.30467/nivar.2022.347939.1220).
- Chen Y, Hu Y, Li H (2022) A novel weighted K-nearest neighbor algorithm for imbalanced data classification. *IEEE Access* 10:34402–34410. <https://doi.org/10.1109/ACCESS.2022.3156996>.
- Che Nordin NF, Mohd NS, Koting S, Ismail Z, Sherif M, El-Shafie A (2021) Groundwater quality forecasting modelling using artificial intelligence: a review. *Groundwater for Sustainable Development* 14:100643. <https://doi.org/10.1016/j.gsd.2021.100643>.
- Emami S, Chooapan Y, Parsa J (2018) Modeling the groundwater level of the Miandoab Plain using artificial neural network method and election and genetic algorithms. *Iranian Journal of Ecohydrology* 5:1175–1189. <https://doi.org/10.22059/ije.2018.258644.889g>
- Fan J, Zhao G, Mu X, Lu A, Tian P, Gao P, Sun W (2023) Effects of cascading reservoirs on streamflow and sediment load with machine learning reconstructed time series in the upper Yellow River basin. *Catena* 225:107008. <https://doi.org/10.1016/j.catena.2023.107008>.
- Faraji Z, Kaviani A, Ashrafzadeh A (2017) Assessment of GRACE satellite data for estimating the groundwater level changes in Qazvin province. *Iranian Journal of Ecohydrology* 4:463–476. <https://doi.org/10.22059/ije.2017.61482>.
- Fattahi P, Ashrafzadeh A, Pirmoradian N, Vazifedoust M (2021) Integrating IHACRES with a data-driven model to investigate the possibility of improving monthly flow estimates. *Water Supply* 22:360–371. <https://doi.org/10.2166/ws.2021.267>.
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics*:1189–1232. <https://www.jstor.org/stable/2699986>
- Haghshenas Gatabi R, Mohseni B, Razaghian H, Asadi Telouki A (2015) Identification and investigation of climate change in Qazvin province with emphasis on temperature and precipitation indices. *International Conference on Environment and Natural Resources*. <https://civilica.com/doc/551342>.
- Hanoon MS, Ahmed AN, Fai CM, Birima AH, Razzaq A, Sherif M, Sefelnasr A, El-Shafie A (2021) Application of artificial intelligence models for modeling water quality in groundwater: comprehensive review, evaluation and future trends. *Water Air Soil Pollution* 232:411. <https://doi.org/10.1007/s11270-021-05311-z>.
- Hedayati Dezfouli A, Kakavand R. Climatic zoning of Qazvin province. *Iran Meteorological Organization (IRIMO)*. https://nivar.irimo.ir/article_13201.html.



- Janbaz Fotamy M, Kholghi M, Abdeh Kolahchi A, Roostaei M (2023) The performance of the evidence weighting in GIS for determining the effective factors on the land subsidence in Qazvin Plain. *Iran-Water Resources Research* 19:118–135. [20.1001.1.17352347.1402.19.3.8.7](https://doi.org/10.117352347.1402.19.3.8.7).
- Kang LS, Hamm SY, Cheong JY, Jeon HT, Park JH (2020) Groundwater monitoring system and groundwater policy in relation to unified water resource management in Korea. *Water Policy* 22:211–222. <https://doi.org/10.2166/wp.2020.171>.
- Karbasi M, Jamei M, Malik A, Kisi O, Yaseen ZM (2023) Multi-steps drought forecasting in arid and humid climate environments. *Agricultural Water Management* 281:108210. <https://doi.org/10.1016/j.agwat.2023.108210>.
- Lee EH (2022) Runoff prediction of urban stream based on the discharge of pump stations using improved multi-layer perceptron applying new optimizers combined with a harmony search. *Journal of Hydrology* 615:128708. <https://doi.org/10.1016/j.jhydrol.2022.128708>.
- Luk KC, Ball JE, Sharma A (2001) An application of artificial neural network for rainfall forecasting. *Mathematical and Computer Modelling* 33:683–693. [https://doi.org/10.1016/S0895-7177\(00\)00272-7](https://doi.org/10.1016/S0895-7177(00)00272-7).
- Mahmoudpour H, Janatrostami S, Ashrafzadeh A (2020) Qualitative assessment of the coastal plain of Talesh using the modified DRASTIC vulnerability method. *JWSS-IUT* 24:97–118. [10.47176/jwss.24.3.40842](https://doi.org/10.47176/jwss.24.3.40842).
- Mahmoudpour, H., Janatrostami, S., Ashrafzadeh, A. (2023). Optimal Design of Groundwater Quality Monitoring Network Using Aquifer Vulnerability Map. *Water Resources Management*. 37, 797–818. <https://doi.org/10.1007/S11269-022-03404-W>
- Masoudi MJ, Ashrafzadeh A, Khaledian M, Janatrostami S (2024) Assessment of groundwater quality for agricultural purposes in Qazvin Province, northwestern Iran. *Environmental and Sustainability Indicators* 24:100528. <https://doi.org/10.1016/j.indic.2024.100528>.
- Mohammadi B, Biazar M, Asadi E (2017) Efficiency of particle swarm hybrid algorithm in simulating water level (Case study: Ardabil plain aquifer). *Journal of Rainwater Reservoir Systems* 5:77–87. <https://doi.org/10.1007/s11269-022-03404-W>.
- Moghaddam N., Hadizadeh H. & Ashrafzadeh A. Data-driven analysis for estimating and predicting well salinity using artificial intelligence algorithms. *Discov Appl Sci* (2025) . [10.1007/s42452-025-08021-x](https://doi.org/10.1007/s42452-025-08021-x)
- Moghaddam N (2024) Managing water resources to mitigate drought impacts due to climate change in the United States with a focus on the state of California. *Iran-Water Resources Research*. [10.22034/iwrr.2024.448106.2748](https://doi.org/10.22034/iwrr.2024.448106.2748).
- Moghaddam N (2024) Analysis of underground water quality in Hashtgerd region using AHP weighting method. *Journal of Water and Energy Engineering* 3. [Journal of Water and Energy Engineering](https://doi.org/10.1007/s11269-022-03404-W).
- Moghaddam N, Salar Khorasani SM (2024) Investigating the vulnerability of carbonate (karst) and alluvial aquifers. [International Conference on Applied Research in Science and Engineering](https://doi.org/10.1007/s11269-022-03404-W).
- Naderianfar M, Ansari H, Ziaie A, Davary K (2011) Evaluating the groundwater level fluctuations under different climatic conditions in the Neyshabour basin. *Irrigation and Water Engineering* 1:22–37. https://www.watexjournal.ir/article_69922.html?lang=fa.
- Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. *Frontiers in Neurobotics* 7. <https://doi.org/10.3389/fnbot.2013.00021>.
- Parmar KS, Bhardwaj R (2014) Water quality management using statistical analysis and time-series prediction model. *Applied Water Science*. <https://doi.org/10.1007/S13201-014-0159-9/FIGURES/3>
- Rajaei T, Pouraslan F (2015) Spatiotemporal Groundwater Level Forecasting in Davarzan Plain. *Hydrogeomorphology* 2:1–19. [20.1001.1.23833254.1394.2.4.1.4](https://doi.org/10.1007/s11269-022-03404-W).
- Rahnamayami E, Ghahreman N (2007) Analytical perspective on climate change, precipitation, and temperature in Qazvin province. 4th National Conference on Watershed Management.
- Schulz E, Speekenbrink M, Krause A (2018) A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology* 85:1–16. DOI: [10.1016/j.jmp.2018.03.001](https://doi.org/10.1016/j.jmp.2018.03.001)



- Tikhamarine Y, Souag-Gamane D, Ahmed AN, Sammen SS, Kisi O, Huang YF, El-Shafie A (2020) Rainfall–runoff modelling using improved machine learning methods. Journal of Hydrology 589:125133. <https://doi.org/10.1016/j.jhydrol.2020.125133>.
- Van Thieu N, Deb Barma S, Van Lam T, Kisi O, Mahesha A (2023) Groundwater level modeling using augmented artificial ecosystem optimization. Journal of Hydrology 617:129034. <https://doi.org/10.1016/j.jhydrol.2022.129034>.
- Van Tuan P, Zhou Y, Stigter T, Van Tuc D, Hai DH, Vuong BT (2023) Design of preliminary groundwater monitoring networks for the coastal Tra Vinh province in Mekong Delta, Vietnam. Journal of Hydrology: Regional Studies 47:101425. <https://doi.org/10.1016/j.ejrh.2023.101425>.
- Williams CK, Rasmussen CE (2006) Gaussian processes for machine learning. [MIT press](https://www.mitrepress.com/).
- Wu Q, Wu Y (2020) An efficient K-nearest neighbor search algorithm based on spatial indexing. IEEE Access 8:158203–158213. <https://doi.org/10.1109/ACCESS.2020.3020715>.
- Zhou Y (2020) Real-time probabilistic forecasting of river water quality under data missing situation. Journal of Hydrology 589:125164. <https://doi.org/10.1016/J.JHYDROL.2020.125164>
- Zhu J, Zeng Y (2020) Random forest-based prediction and analysis of cancer immunotherapy outcome. IEEE Access 8:148830–148837. <https://doi.org/10.1109/ACCESS.2020.3016156>

دینار استادی
غیر نشده، نماینده،
فاطمیل استناد

Groundwater Salinity Modeling and Prediction in the Qazvin Plain using a Data-Driven and Artificial Intelligence Approach

Nima Moghaddam¹, Afshin Ashrafzadeh^{2,*}

Master's student in Water Resources Engineering, Faculty of Agriculture and Natural Resources, University of Tehran.

Associate Professor, Department of Irrigation and Reclamation Engineering, Faculty of Agriculture and Natural Resources, University of Tehran, Tehran, Iran.

1. Abstract

In the current study, conducted with the aim of modeling groundwater quality, a method is proposed to calculate the spatial structure of data within the modeling process. This method incorporates the distance between observation points and the estimation point as one of the model inputs. GBR, GPR, KNN, MLP, SVM, and RF models were utilized. The models were trained and tested using groundwater quality data obtained from: 3058 wells for 4 neighboring observation wells, 2724 wells for 5 neighboring observation wells, 2080 wells for 6 neighboring observation wells, 1364 wells for 7 neighboring observation wells, and finally from 631 wells for 8 neighboring observation wells, in a region covering the entire Qazvin province in northwestern Iran. Two separate datasets comprising information from the first six months and the second six months of the year were used. The Mean Absolute Error (MAE), R-squared (R^2), Pearson correlation coefficient, and Root Mean Square Error (RMSE) for models with four, five, six, seven, and eight neighboring wells indicated the satisfactory performance of the Random Forest (RF) model. This model performed exceptionally well in both training and testing phases, exhibiting the lowest error and highest correlation. Although the complexity and execution time of this model may be higher, its high accuracy compensates for these drawbacks. Numerical results from the modeling showed that the Random Forest (RF) model outperformed other models in most of the examined scenarios. For instance, in the scenario using four adjacent wells, during the testing phase, this model achieved an $R^2 = 0.915$, a correlation coefficient of 0.957, an $MAE \approx 156$, and an $RMSE \approx 326$. Furthermore, in the scenario with five adjacent wells, the RF model recorded an $R^2 = 0.935$ and an $RMSE \approx 316$, demonstrating its stability and high precision. The Gaussian Process Regression (GPR) model, as the second most accurate model, achieved R^2 values ranging from approximately 0.904 to 0.934 and RMSE values from approximately 350 to 318 across the two six-month periods, showing relatively stable performance. In contrast, the KNN and GBR models, despite providing acceptable results, had higher errors compared to RF and GPR. For example, the RMSE value for the KNN model in the four-adjacent-well scenario was calculated at around



407, and for the GBR model around 461, indicating the relative weakness of these models in reconstructing the spatial structure of electrical conductivity. Overall, the average values of MAE, RMSE, R-squared, and Pearson correlation coefficient demonstrated that the RF model offered the lowest error and highest correlation in both training and testing stages. In the best-case scenario, this model achieved an R^2 higher than 0.95 in the training phase and an RMSE less than 270, indicating its high capability in modeling the spatiotemporal patterns of electrical conductivity in the regional aquifer. Although the computational complexity of this model is relatively higher, the substantial accuracy and stability of the results fully compensate for this limitation.

Keywords: Modeling, Groundwater, Observation Well, Artificial Intelligence.

ویراستاری نشده، غیر قابل استناد