

## پیش‌بینی و پهنه‌بندی کیفیت آب زیرزمینی با استفاده از مدل‌های سیستم اطلاعات جغرافیایی (GIS) و روش‌های یادگیری ماشین (مطالعه موردی: دشت زاهدان)

هزیر علیم‌رادی<sup>۱</sup>، عین‌اله روحی‌مقدم<sup>۲</sup>، مهسا خالقی<sup>۳\*</sup> و ابوالفضل بامری<sup>۴</sup>

۱- دانشجوی کارشناسی ارشد، گروه مرتع و آبخیزداری، دانشکده آب و خاک، دانشگاه زابل، سیستان و بلوچستان، ایران.

۲- دانشیار، گروه مرتع و آبخیزداری، دانشکده آب و خاک، دانشگاه زابل، سیستان و بلوچستان، ایران.

۳- دکتری آبیاری و زهکشی، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل، سیستان و بلوچستان، ایران.

۴- مربی، گروه مهندسی علوم خاک، دانشکده آب و خاک، دانشگاه زابل، سیستان و بلوچستان، ایران.

\* نویسنده مسئول: [Khaleghi.83@gmail.com](mailto:Khaleghi.83@gmail.com)

تاریخ پذیرش مقاله: ۱۴۰۱/۰۳/۲۵

تاریخ دریافت مقاله: ۱۴۰۰/۰۷/۳۰

### چکیده

روند صعودی جمعیت در چند دهه اخیر، محدودیت منابع آب و بهره‌برداری بیش از حد از سفره‌های زیرزمینی باعث به بارآمدن خسارات غیرقابل جبران کمی و کیفی به آبخوان‌های کشور شده است. در تحقیق حاضر مدل‌های سیستم اطلاعات جغرافیایی (GIS) و تکنیک‌های یادگیری ماشین (ML) با استفاده از متغیرهای کیفی قابل دسترس به منظور پیش‌بینی و پهنه‌بندی شوری و SAR آب زیرزمینی و مقایسه دقت روش‌های مذکور در محدوده‌ی دشت زاهدان مورد ارزیابی قرار گرفت. داده‌های ورودی بر اساس نمونه‌برداری‌های کیفیت آب در سال آبی ۱۳۹۷ از ۵۹ حلقه چاه مشاهداتی بود. بررسی پارامترها نشان داد که در دشت زاهدان، پارامترهای EC، SAR و TDS دارای تغییرپذیری زیاد ( $CV > 41\%$ ) و اسیدیته دارای تغییرپذیری کم ( $CV = 16\%$ ) می‌باشد. نتایج تحلیل‌های زمین‌آماری نشان داد که برای پارامترهای TDS و EC مدل IDW با توان دو و برای پارامترهای pH و SAR روش کریجینگ معمولی با حداقل RMSE بهترین نتیجه را در مرحله آزمون ارائه داد. ارزیابی عملکرد مدل‌های یادگیری ماشین نشان داد که هر سه مدل RF، ANN و SVM با کسب  $R^2$  بالای ۹۰ درصد و مقادیر NRMSE زیر ۱۵ درصد برای همه پارامترها (به استثنای اسیدیته) نتایج قابل‌قبولی از خود نشان دادند. هرچند در مرحله آموزش تخمین‌های بهتری نسبت به مرحله آزمون مشاهده شد. مقایسه مدل‌های مختلف GIS و یادگیری ماشین نیز حاکی از برتری قابل‌توجه مدل‌های یادگیری ماشین در تخمین پارامترهای مورد بررسی می‌باشد. درنهایت می‌توان نتیجه گرفت که در شرایط نبود امکانات برای بررسی میدانی کیفیت آب زیرزمینی، روش‌های داده‌محور جایگزین قابل اطمینانی برای پایش کیفی آب می‌باشد.

**واژه‌های کلیدی:** پایش منابع آب، جنگل تصادفی، داده‌کاوی، سیستان و بلوچستان، هدایت الکتریکی.

### مقدمه

نمی‌باشد (عربی و همکاران، ۱۳۹۳). از این رو یکی از مهم‌ترین عوامل محدودکننده توسعه‌ی کشاورزی در دهه‌های پیش رو را می‌توان محدودیت منابع آب عنوان نموده و منابع آب ناکافی را تهدیدی جدی برای تولید محصولات کشاورزی دانست (حکمت نیا و همکاران، ۱۳۹۹). یکی از مهم‌ترین راهکارها در جهت افزایش کمی و کیفی آب مدیریت صحیح منابع آب می‌باشد. بنابراین رابطه بین مدیریت منابع آب و بررسی کمی و

منابع آب تجدیدشونده ایران کمتر از یک درصد از منابع دنیا می‌باشد. بر طبق پیش‌بینی‌های صورت گرفته جمعیت کشور تا سال ۱۴۱۰ به مرز بالای ۱۰۰ میلیون نفر خواهد رسید. در چنین شرایطی تامین امنیت غذایی این جمعیت در سال نیازمند بیش از ۱۵۰ میلیارد متر مکعب آب می‌باشد که با توجه به ذخایر آبی کشور این حجم از منابع آب در دسترس

آنان نشان داد که روش کوکریجینگ بهترین روش زمین آماری برای پهنه‌بندی کیفی دشت دز فول است. پورقاسمی و همکاران در سال ۱۳۹۴ تهیه نقشه پتانسیل آب زیرزمینی با استفاده از مدل‌های انتروپی شانون و جنگل تصادفی در دشت بجنورد را انجام دادند که نتایج بیانگر دقت بسیار خوب مدل انتروپی شانون در مقایسه با مدل جنگل تصادفی بود. دولت و همکاران (۱۳۹۶) با استفاده از داده‌های ۱۴ چاه مشاهداتی به مدل‌سازی تغییرات متغیرهای کیفی آب زیرزمینی با استفاده از روش‌های ماشین بردار پشتیبان، شبکه عصبی مصنوعی و شبکه عصبی-فازی پرداختند. نتایج حاصل از مدل‌سازی نشان داد که روش شبکه عصبی-فازی برای متغیر EC، SAR و TH با ضریب تبیین بالای ۰/۹۸ عملکرد بهتری نسبت به دو روش دیگر داشته است. نوروزی و همکاران (۱۳۹۶) قابلیت انتقال آبخوان دشت ملکان را با استفاده از روش یادگیری ماشین جنگل تصادفی مورد بررسی قرار دادند. نتایج این پژوهشگران نشان داد که روش جنگل تصادفی در پیش‌بینی قابلیت انتقال آبخوان عملکرد مطلوبی از خود نشان داد. عیسی‌زاده و همکاران (۱۳۹۸) با استفاده از آزمون گاما و مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی پارامترهای کیفی آبخوان دشت گیلان را مورد بررسی قرار دادند. نتایج این پژوهش بیانگر کارایی بهتر روش SVM در مقابل ANN در تخمین پارامترها بر اساس متغیرهای محیطی بود. ریاحی و همکاران (۱۳۹۹) به ارزیابی پتانسیل منابع آب زیرزمینی با استفاده از ترکیب روش‌های داده‌کاوی از جمله روش جنگل تصادفی در منطقه دشت سرخون هرمزگان پرداختند. نتایج این پژوهش نشان داد که هر چهار روش استفاده شده از جمله روش RF دارای دقت عالی (بالای ۹۰ درصد) جهت پیش‌بینی می‌باشد. همچنین نتایج اهمیت فاکتورها نشان داد که معیارهای شیب، ارتفاع و شاخص قدرت جریان دارای بالاترین میزان اهمیت می‌باشد. (Taheri-Tizro et al., 2014) بررسی تغییرات مکانی کیفیت آب زیرزمینی در دشت زیرین-آباد زنجان با استفاده از روش‌های زمین‌آماري پرداختند. بررسی آنان نشان داد که روش کوکریجینگ در مقایسه با روش IDW عملکرد بهتری در بررسی تغییرات مکانی EC و SAR داشته است. (Rahmati et al., 2016) نیز کاربرد مدل‌های جنگل تصادفی و حداکثر انتروپی (ME) را برای تولید

کیفی آن یک رابطه دو طرفه می‌باشد که تقویت یک سو سبب افزایش قدرت مانور در سمت دیگر خواهد شد. دستیابی به بانک اطلاعات مکفی از ویژگی‌های کمی و کیفی منابع و آبخوان‌های موجود مستلزم انجام عملیات پایش کمی و کیفی آبخوان‌هاست (کلانتری و همکاران، ۱۴۰۱؛ رخ شاد و شهیدی، ۱۴۰۰؛ حسنی و همکاران، ۱۳۹۷).

پایش کیفی آب‌های زیرزمینی بسیار پرهزینه و زمان بر بوده و انتخاب روشی که در آن با حداقل پارامترهای هیدروشیمیایی بتوان پیش‌بینی نسبتاً دقیقی از طبقه کیفیت آب برای مصارف مختلف داشت، مهم و ضروری است. روش‌های آماری، ابزاری مناسب و معتبر را برای درک بهتر تغییرات کیفیت آب‌های سطحی، زیرزمینی و زیست محیطی، فراهم می‌کنند (عزیزپور و همکاران، ۱۴۰۰؛ علی پور و همکاران، ۱۳۹۶). با در اختیار داشتن تعداد محدود اندازه‌گیری‌های آزمایشگاهی یا میدانی از یک متغیر ناحیه‌ای با پراکنش مناسب، می‌توان با استفاده از روش‌های مناسب میان‌یابی نظیر روش‌های زمین‌آماري، آن پارامتر را پهنه‌بندی کرد (Delbari et al., 2011). نیاز به استفاده از استراتژی‌های مدیریتی به منظور مقابله با بحران آب و جلوگیری از برداشت‌های بی‌رویه از سفره‌های زیرزمینی از طریق پایش کمی و کیفی مخازن آب‌های زیرزمینی، پژوهشگران را ملزم دانسته که تحقیقات بسیاری را به انجام رسانند.

زارع ابیانه و همکاران (۱۳۹۰) با کاربرد روش شبکه عصبی مصنوعی با سه ترکیب مختلف مقدار نیترات را در دشت همدان-بهار مورد ارزیابی قرار دادند. نتایج بررسی آنان نشان داد هر سه ترکیب این روش داده‌کاوی کارایی بالایی را در تخمین نیترات در این دشت دارا می‌باشد. محمدی و همکاران (۱۳۹۱) نوسانات سطح آب زیرزمینی در زمان و مکان در دشت کرمان را در سال‌های ۱۳۸۵-۱۳۷۵ مورد بررسی قرار دادند. آنها با بررسی تکنیک‌های کریجینگ، لوگ-کریجینگ و IDW با توان‌های ۱ تا ۵ برای میان‌یابی به این نتیجه رسیدند که تکنیک IDW با توان ۵ در ابتدای دوره و IDW با توان ۴ در انتهای دوره از دقت بالاتری برای میان‌یابی عامل سطح آب برخوردار است. ادیب و زمانی (۱۳۹۴) تغییرات مکانی شاخص‌های کیفی آب زیرزمینی دشت دز فول را با استفاده از روش-های زمین‌آمار مورد بررسی قرار دادند. نتایج حاصل از تحقیق

دشت زاهدان با توجه به موقعیت جغرافیایی و قرارگیری در عرض‌های پایین و دور بودن از محدوده‌ی اثر جبهه‌های مدیترانه‌ای نسبت به سایر نقاط کشور از رطوبت کمتری برخوردار بوده و در نتیجه همواره با خطرات کمبود آب مواجه می‌باشد. لذا بررسی دقیق کمیت و کیفیت منابع آب به‌خصوص آب زیرزمینی و اطلاع از وضعیت توزیع پارامترهای کیفی در این منطقه بیشتر از دیگر مناطق کشور ضرورت دارد. اما باید توجه داشت که بررسی کیفی منابع آب امری هزینه‌بر و زمان‌بر بوده و نیازمند تجهیزات آزمایشگاهی و نیروی انسانی می‌باشد، لذا در چنین شرایطی استفاده از روش‌های جایگزین نظیر روش‌های جدید مدل‌سازی بسیار کارآمد می‌باشد. هدف از تحقیق حاضر پیش‌بینی شوری و نسبت جذب سدیم آب زیرزمینی (دو پارامتر اصلی و تعیین کننده طبقه کیفی آب) با استفاده از مدل‌های مختلف GIS و تکنیک‌های یادگیری ماشین شامل جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی مصنوعی و نیز مقایسه عملکرد این مدل‌ها می‌باشد، که بررسی هم زمان این روش‌ها در منطقه مطالعاتی تاکنون مورد توجه قرار نگرفته است.

## مواد و روش‌ها

### مشخصات منطقه مورد مطالعه

منطقه مطالعاتی در تحقیق حاضر دشت زاهدان واقع در استان سیستان و بلوچستان می‌باشد. این دشت با وسعت تقریبی ۲۱۰ کیلومترمربع و ارتفاع ۱۳۸۵ متر از سطح دریای آزاد، در محدوده‌ی جغرافیایی ۶۰ درجه و ۳۹ دقیقه تا ۶۰ درجه و ۵۶ دقیقه طول شرقی و در ۲۹ درجه ۲۳ دقیقه تا ۲۹ درجه و ۳۵ دقیقه عرض شمالی واقع شده است (Sarhadi et al., 2016). میانگین بارش سالانه در این شهر ۷۲ میلی‌متر گزارش شده است. شهر زاهدان در دشتی صاف و هموار با شیبی ملایم در حدود ۰/۳ درصد تا ۰/۵ درصد قرار گرفته و کوه‌های نسبتاً مرتفعی آن را احاطه نموده است. شیب عمومی دشت از سمت جنوب به شمال و شیب جانبی آن از غرب به شرق است (نیکبخت و دلبری، ۱۳۹۲).

### معرفی مدل‌های اجرا شده در مطالعه

آمار مکانی بخشی از علم آمار است که دارای جنبه‌های کاربردی بسیار مهم در تجزیه و تحلیل داده‌های مکانی

نقشه‌های رقومی آب‌های زیرزمینی به‌کار بردند و نتیجه گرفتند که مدل‌های RF و ME تکنیک‌های سودمندی برای تولید نقشه‌های رقومی آب‌های زیرزمینی می‌باشند. Granata et al., (2017) از دو مدل رگرسیون بردار پشتیبان (SVR) و درختان رگرسیونی (RT) به منظور پیش‌بینی شاخص‌های کیفیت فاضلاب بهره بردند. نتایج آنان نشان داد که هر دو مدل از استحکام، قابلیت اطمینان و قابلیت تعمیم بالایی برای پیش‌بینی شاخص‌های کیفی برخوردار بودند. با این وجود، رگرسیون بردار پشتیبان عملکرد بهتری از خود نشان داد. Haghyabi et al., (2018) عملکرد تکنیک‌های شبکه عصبی مصنوعی (ANN) و دستگاه بردار پشتیبانی (SVM) را برای پیش‌بینی اجزای کیفیت آب رودخانه تیره واقع در جنوب غربی ایران مورد مطالعه قرار دادند. بررسی نتایج ارزیابی ANN و SVM نشان داد که هر دو مدل عملکرد مناسبی برای پیش‌بینی اجزای کیفیت آب دارند. Ahmed et al., (2019) با کاربرد چند الگوریتم مختلف یادگیری ماشین و با استفاده از پارامترهای ورودی نظیر اسیدیته، دما، کدورت و ... به طبقه‌بندی کیفی آب پرداختند. این محققین در نهایت اعلام داشتند که روش پرسپترون چند لایه (MLP) بالاترین دقت را در پیش‌بینی و طبقه‌بندی کیفی آب داراست. Bui et al., (2020) در پژوهشی ۴ روش یادگیری مستقل نظیر جنگل تصادفی، M5P، درخت تصادفی و درخت هرس خطای کاهش یافته و ۱۲ الگوریتم ترکیبی استخراج داده را برای ایجاد پیش‌بینی‌های WQI ایران استفاده کردند. نتایج نشان می‌دهد که کلی فرم مدفوع (FC) و مواد جامد کل (TS) بیشترین و کمترین تأثیر را در پیش‌بینی WQI داشتند. به طور کلی متغیرهایی با همبستگی بسیار کم عملکرد ضعیف‌تری را نشان می‌دهند. هیبرید BA-RT از سایر مدل‌ها بهتر عمل کرد ( $R^2 = 0.94$ ). نتایج نشان داد که اکثر الگوریتم‌ها مقادیر WQI را بیش از حد واقعی تخمین زده‌اند. (Bameri and Khaleghi (2021) در پژوهش دیگری به تخمین نوسانات سطح آب زیرزمینی با استفاده از روش‌های مختلف سیستم‌های اطلاعات جغرافیایی و پهنه‌بندی آن با بهترین روش پرداختند. نتایج این تحقیق بیانگر این بود که روش معین وزن‌دهی عکس فاصله با توان دو (IDW2) نسبت به سایر روش‌های زمین آماری تخمین بهتری را حاصل نمود.

از پایگاه داده‌ها و هم برای ایجاد مدل‌های پیش‌بینی استفاده نمود (حسینی و همکاران، ۱۳۹۷). هدف از ماشین‌های یادگیری پیدا کردن یک عملکرد تقریباً مفید است که زمینه پیش‌بینی ارتباط بین متغیرهای ورودی و نتایج موردنظر را نشان دهد (میرکزهی و همکاران، ۱۳۹۶). در این تحقیق از سه مدل یادگیری ماشین شبکه عصبی مصنوعی (ANN)، ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) برای پیش‌بینی پارامترهای کیفی مورد مطالعه استفاده شد.

معماری ANN از یک ساختار غیرخطی تشکیل شده است که توسط یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل شده است. هر لایه تعداد سلول‌های عصبی متفاوتی دارد که با وزن مشخصی به هر نورون در لایه بعدی مرتبط است (Zhu et al., 2018). تعداد نورون‌های موجود در لایه ورودی برابر با تعداد متغیرهای ورودی بوده، در حالی که از یک نورون (برابر با تعداد متغیرهای خروجی) در لایه خروجی استفاده شده است. تاکنون، تعداد لایه‌های پنهان و تعداد نورون‌های موجود در هر لایه پنهان هنوز با قواعد صریح مشخص نشده است، که معمولاً بسته به نظر طراح شبکه و از آزمایش‌های تجربی همراه با روش آزمایش و خطا بدست می‌آید (عجم‌زاده و همکاران، ۱۳۹۶؛ Kardam et al., 2012). در صورت کافی نبودن تعداد نورون‌ها، شبکه قادر نخواهد بود نگاشت دقیقی بین بردارهای ورودی و خروجی ایجاد کند. در خروجی هر نورون از شبکه یک تابع خطی قرار دارد و فرآیند یادگیری در تمام نورون‌ها و لایه‌ها صورت می‌گیرد (عجم‌زاده و همکاران، ۱۳۹۶). در تحقیق حاضر یک لایه پنهان منفرد اتخاذ شد، زیرا یک لایه پنهان به طور کلی برای مقابله با اکثر مشکلات کافی بوده و همزمان از شلوغ شدن بیش از حد اجتناب می‌کند. هنگام آموزش مدل شبکه عصبی، از الگوریتم یادگیری feedforward error-backpropagation استفاده شده است.

ماشین بردار پشتیبان، تکنولوژی جدید یادگیری ماشین است که روش آماری غیرپارامتریک و نیمه‌نظارت شده می‌باشد (رضایی و همکاران، ۱۳۹۷). این روش در سال ۱۹۹۲ میلادی توسط بوسر و همکاران بر پایه تئوری یادگیری آماری معرفی شد. آنها طی سال‌های بعد تئوری ابرصفحه بهینه را به عنوان

می‌باشد. زمین‌آمار شاخه‌ای از علم آمار است که تغییرات متغیر در محیط را به دو مولفه سیستماتیک و تصادفی تقسیم‌بندی می‌نماید (مدنی، ۱۳۷۳). پیش‌بینی توزیع مکانی یک متغیر در روش‌های زمین‌آمار شامل دو مرحله است. مرحله اول محاسبه نیم‌تغییرنمای تجربی و مدل کردن آن با استفاده از یک مدل تئوری مناسب است. سپس لازم است بهترین مدل تئوری بر آن برآزش شود. از مدل‌های معروف سیستم اطلاعات جغرافیایی (GIS) می‌توان به روش‌های زمین‌آمار<sup>۱</sup> کریجینگ و کوکریجینگ و روش مطلق<sup>۲</sup> وزن‌دهی عکس فاصله (IDW) اشاره کرد. کریجینگ یک روش تخمین بر مبنای "میانگین متحرک وزن‌دار" استوار می‌باشد و در مورد آن می‌توان گفت که بهترین تخمین‌گر خطی ناریب است. از مهم‌ترین ویژگی‌های کریجینگ آن است که به ازای هر تخمین زده شده می‌توان خطای مرتبط با آن را محاسبه کرد (غضبان، ۱۳۸۱). همان‌طور که در آمار کلاسیک روش‌های چند متغیره برای تخمین وجود دارد، در زمین‌آمار نیز می‌توان به روش کریجینگ بر اساس همبستگی بین متغیرهای مختلف تخمین زد. روش کوکریجینگ نوعی کریجینگ چند متغیره است که از اطلاعات کمکی برای تخمین بهتر متغیر اصلی استفاده می‌کند. سیستم معادلاتی کوکریجینگ را می‌توان به هر تعداد متغیر ثانویه تعمیم داد. در این تحقیق ضریب همبستگی پیرسون برای پارامترهای کیفی آب زیرزمینی محاسبه شد و از پارامترهایی که داری ضریب همبستگی بالایی با هریک از خصوصیات کیفیت آب بوده به عنوان متغیر کمکی برای پیش‌بینی خصوصیات کیفی استفاده شد. از طرفی در روش‌های مطلق، تفاوت مربوط به نحوه محاسبه وزنی است که به نقاط مشاهده شده اطراف نقطه مجهول داده می‌شود. در روش وزن‌دهی عکس فاصله (IDW) وزن‌ها تنها با توجه به فاصله هر نقطه معلوم نسبت به نقطه مجهول و بدون توجه به نحوه پراکندگی نقاط حول نقطه مورد تخمین، تعیین می‌شوند. به نقاط نزدیکتر وزن بیشتر و به نقاط دورتر وزن کمتری اختصاص داده می‌شود (Isaaks and Srivastava, 1989).

همچنین، ماشین‌های یادگیری که از نسل جدید تکنیک‌های داده‌کاوی به شمار می‌آیند، در دو دهه اخیر توسعه زیادی یافته‌اند و از آنها نیز می‌توان هم برای کشف و استخراج دانش

<sup>2</sup> Deterministic Method

<sup>1</sup> Geostatistic Method

افزار R با به کارگیری بسته نرم‌افزاری random forest آموزش داده شد. به طور کلی در الگوریتم جنگل تصادفی در نرم‌افزار R بایستی دو متغیر تعریف شود، یکی تعداد درختان موجود در جنگل و دیگری تعداد متغیرهای مورد استفاده در هر درخت که به صورت سعی و خطا بدست می‌آید.

### گردآوری داده‌ها و اجرای مدل‌ها

برای اجرای مدل‌سازی GIS و سه مدل یادگیری ماشین از داده‌های حاصل از نتایج آزمایشات میدانی خصوصیات کیفی آب زیرزمینی دشت زاهدان در سال آبی ۱۳۹۷ (مهر ۱۳۹۷- مهر ۱۳۹۸) برداشت شده توسط بخش تحقیقات سازمان آب منطقه‌ای استان سیستان و بلوچستان از ۵۹ حلقه چاه مشاهداتی استفاده شد. خصوصیات کیفی برداشت شده از چاه‌های مشاهداتی شامل شوری، سدیم، منیزیم، پتاسیم، اسیدیت، کلر، سولفات،  $TDS^3$ ،  $SAR^4$ ، سختی کل و بیکربنات بود. پیش‌بینی توزیع مکانی یک متغیر در مدل‌های مختلف GIS شامل دو روش کلی مطلق و زمین آماری می‌باشد که اساس کار هر دو روش تخمین متغیر در نقاط نامعلوم با استفاده از یک روش درون‌یابی مناسب است (حسینی‌پاک، ۱۳۷۷). با این تفاوت که روش‌های زمین‌آماري شامل دو مرحله می‌باشند. مرحله اول محاسبه نیم‌تغییرنمای تجربی و مدل کردن آن با استفاده از یک مدل تئوری مناسب و مرحله دوم استفاده از ضرایب نیم‌تغییرنما در تخمین متغیر در نقاط نامعلوم است. در این مطالعه جهت آنالیزهای واریوگرافی از مدل‌های تئوری کروی، نمایی، خطی و گوسی استفاده شد و بهترین مدل بر اساس بیشترین مقدار  $R^2$  و کمترین مقدار RSS تعیین گردید. برای مشخص نمودن وابستگی مکانی خصوصیت کیفی آب زیرزمینی از نسبت اثر قطعه‌ای به آستانه نیم‌تغییرنما استفاده شده است. مقدار کمتر از ۲۵ درصد و بیشتر از ۷۵ درصد این نسبت به ترتیب وابستگی مکانی قوی و ضعیف را نشان می‌دهد (Cambardella et al., 1994; Zhu and lin, 2010).

روش‌های یادگیری ماشین این تحقیق در محیط نرم‌افزار R اجرا شد. R یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای محاسبات آماری و تحلیل داده است که حاوی محدوده

طبقه‌بندی کننده خطی ارائه داده و طبقه‌بندی کننده‌های غیرخطی را با کمک توابع کرنل معرفی نمودند. اصول اساسی آن چه که امروز به عنوان ماشین بردار پشتیبان شناخته می‌شود، نتایج کارهای بوسر و همکاران است و در نهایت گسترش ماشین بردار پشتیبان بر اساس رگرسیون نیز در سال ۱۹۹۵ توسط وپنیک به نتیجه رسید (صمدیان فرد و پناهی، ۱۳۹۷). از مدل طبقه‌بندی ماشین بردار پشتیبان جهت حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند، استفاده می‌گردد و مدل رگرسیون بردار پشتیبان در حل مسائل پیش‌بینی کاربرد دارد (سیفی و همکاران، ۱۳۹۲؛ صمدیان فرد و پناهی، ۱۳۹۷). ویژگی اصلی این روش، توانایی بالا در استفاده از نمونه‌های آموزشی کمتر و در عین حال رسیدن به دقت بالاتر در مقایسه با سایر روش‌ها است.

روش جنگل تصادفی یک روش جدید و قدرتمند است که پیشرفت‌های قابل ملاحظه‌ای را در تکنولوژی داده‌کاوی ارائه داده است. تکنیک جنگل تصادفی مدل توسعه یافته‌ای از روش طبقه‌بندی و رگرسیون درختی CART می‌باشد. این روش نوعی الگوریتم مجموعه طبقه‌بندی شده است که توسط برایمن و کاتلر پیشنهاد شده است. جوهر الگوریتم RF بهبود الگوریتم درخت تصمیم است (Ma et al., 2016). به عبارتی، روش جنگل تصادفی یک روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است. جنگل تصادفی می‌تواند الگوهای پیچیده را یاد بگیرد و ارتباط غیرخطی بین متغیرهای توضیحی و متغیرهای وابسته را در نظر بگیرد. فرایند آموزش درخت، یک فرایند تکرار پذیر بوده که از گره ریشه شروع شده و به گره‌های نهایی (برگ‌ها) ختم می‌شود. سپس نمونه خودراه‌انداز دیگری برای آموزش درخت دیگری انتخاب می‌شود. هنگامی که درخت کامل شد، مجموعه‌ای از قوانین تصمیم‌گیری که فرایند تصمیم‌گیری داده‌ها را تشریح می‌کند، جهت تخمین داده جدید استفاده می‌گردند. روش RF نسبت به روش‌های دیگر دارای مزایایی مانند دقت پیش‌بینی بالا، توانایی در یادگیری روابط غیرخطی، توانایی بالا در تعیین متغیرهای مهم در پیش‌بینی و ماهیت غیر پارامتری می‌باشد (نوروزی و همکاران، ۱۳۹۶). در این تحقیق، روش جنگل تصادفی با استفاده از داده‌های کیفیت آب زیرزمینی در نرم

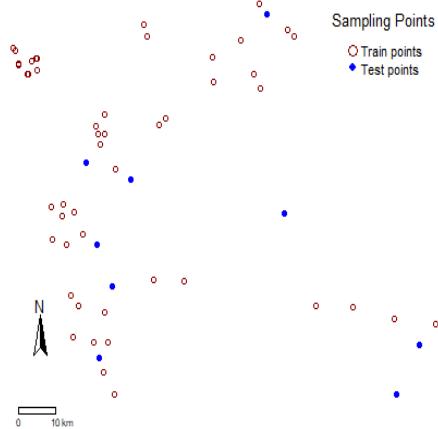
<sup>4</sup> Sodium Adsorption Ratio

<sup>3</sup> Total Dissolved Solids



معرفی گردید. انتخاب داده‌های آموزش و آزمون به صورت تصادفی و توسط نرم افزار انجام شده و بدیهی است که داده‌های یک مرحله در مرحله دیگر مورد استفاده قرار نگرفت. شکل ۱ نقشه پراکنش چاه‌های مشاهداتی در منطقه مطالعاتی و تقسیم‌بندی نقاط نمونه‌برداری را برای دو مرحله آموزش (Train point) و آزمایش (Test point) نشان می‌دهد. مراحل اجرای همه مدل‌های به کار برده شده در این تحقیق (سه مدل مختلف GIS و سه مدل یادگیری ماشین) به صورت فلوجارت در شکل ۲ ارائه شده است.

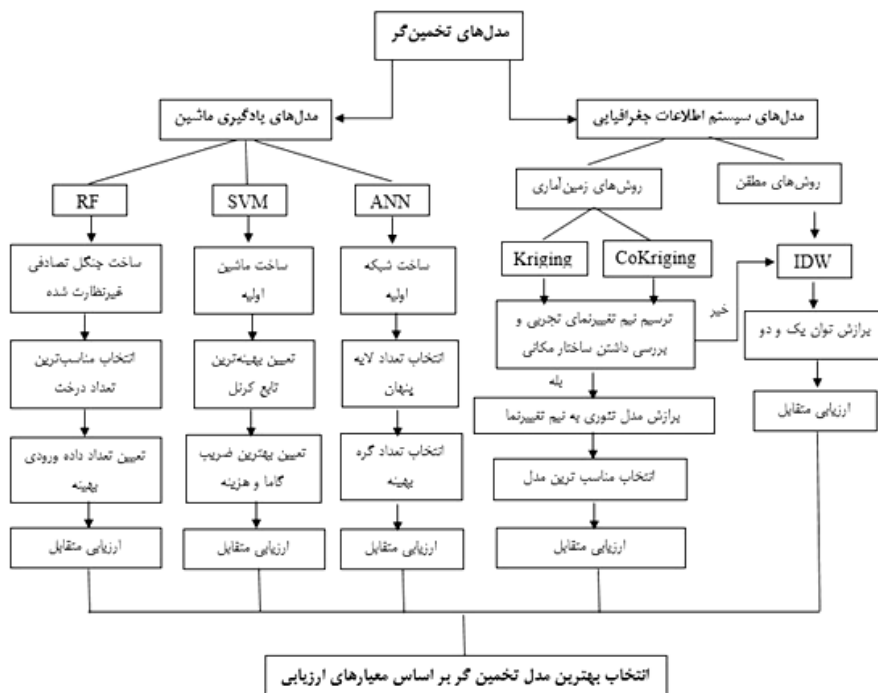
گسترده‌ای از تکنیک‌های آماری از جمله: مدل‌سازی خطی و غیرخطی، آزمون‌های کلاسیک آماری، تحلیل سری‌های زمانی، رده‌بندی، خوشه‌بندی و قابلیت‌های گرافیکی می‌باشد (رستگاریا و صنعتی، ۱۳۹۵). در طی فرایند مدل‌سازی با روش‌های یادگیری ماشین، کل داده‌ها به دو دسته تقسیم‌بندی شدند: داده‌های مرحله آموزش (Training data) و داده‌های مرحله آزمایش (Testing data)، به طوری که در این تحقیق ۸۰ درصد از کل داده‌ها به آموزش مدل اختصاص داده شد و ۲۰ درصد باقیمانده به عنوان داده‌های آزمون به مدل



A

B

شکل ۱- (A) پراکنش چاه‌های مشاهداتی در منطقه و (B) تقسیم‌بندی نقاط نمونه‌برداری برای دو مرحله آموزش و آزمایش.



شکل ۲- فلوجارت مراحل اجرای مدل‌های مختلف تخمین گر پارامترهای کیفی آب زیرزمینی در تحقیق حاضر.

## ارزیابی عملکرد مدل‌ها

به منظور ارزیابی مدل‌های به کار برده شده برای شبیه‌سازی، از چند معیار آماری از جمله ریشه میانگین مربعات خطا (RMSE)، میانگین انحراف خطا (ME)، ضریب همبستگی ( $R^2$ ) و ریشه میانگین مربعات خطا نرمال شده (NRMSE) استفاده شد که به ترتیب در روابط زیر تعریف شده است (امیدوار و همکارن، ۱۳۹۳):

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

$$ME = \bar{P} - \bar{O} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)}{\sum_{i=1}^n (O_i - \bar{O})} \quad (3)$$

$$NRMSE = \frac{RMSE}{O_{max} - O_{min}} \quad (4)$$

که در فرمول‌های مذکور: n تعداد کل نقاط مورد بررسی،  $\bar{O}$  میانگین مقادیر مشاهده شده،  $\bar{P}$  میانگین مقادیر پیش‌بینی شده،  $O_{max}$  حداکثر مقدار مشاهده شده،  $O_{min}$  حداقل مقدار مشاهده شده،  $O_i$  مقدار مشاهده شده در نقطه i ام و  $P_i$  مقدار پیش‌بینی شده با استفاده از مدل در این نقطه است.

## نتایج و بحث

در این پژوهش به منظور تحلیل‌های آماری و مکانی داده‌های کیفی آب زیرزمینی دشت زاهدان شامل هدایت الکتریکی (EC)، اسیدیته (pH)، نسبت جذب سدیم (SAR) و نمک محلول کل (TDS) ابتدا داده‌های اندازه‌گیری شده میدانی مورد کنترل و بازبینی قرار گرفته و داده‌های پرت و ناهمگن

حذف گردید. سپس مدل‌های مختلف GIS و یادگیری ماشین مورد نظر بر روی داده‌های کنترل شده اجرا شد. در نهایت خروجی حاصل از روش‌های مذکور مورد تحلیل و ارزیابی آماری قرار گرفت.

## آمار توصیفی

شاخص‌های آماری پارامترهای کیفی آب زیرزمینی نظیر حداقل، حداکثر، دامنه، میانگین، انحراف معیار، چولگی، کشیدگی و ضریب تغییرات برای کل داده‌ها در جدول ۱ ارائه شده است. طبق طبقه‌بندی (Wilding et al., 1983) اگر ضریب تغییرات داده‌ها کمتر از ۱۵ درصد باشد، داده‌ها دارای تغییرپذیری کم هستند. اگر مقدار این ویژگی بین ۱۵ تا ۳۵ درصد باشد، کلاس تغییرپذیری متوسط و در صورتی که مقدار این آماره بیشتر از ۳۵ درصد باشد، حاکی از تغییرپذیری زیاد داده‌ها است. بر طبق این دسته‌بندی، پارامترهای مورد بررسی در دشت زاهدان به استثنای پارامتر اسیدیته، دارای تغییرپذیری زیاد می‌باشند. با توجه به سازندهای مختلف زمین‌شناسی در منطقه مورد مطالعه و همچنین وجود سیستم‌های مختلف برداشت آب از قبیل آب شرب و مزارع کشاورزی در منطقه، این تغییرپذیری بالا امری مشهود و قابل توجیه می‌باشد. (Antonopoulos et al., 2001) نیز در تحقیق خود تغییرپذیری متوسط تا زیاد را برای پارامترهای کمی و کیفی آب در کشور یونان گزارش نموده و همچنین پارامتر اسیدیته را تنها پارامتر با تغییرپذیری کم عنوان کردند، که با نتایج این پژوهش همسو می‌باشد.

جدول ۱- توصیف آماری خصوصیات کیفی آب زیرزمینی منطقه مورد مطالعه.

Parameter	Min	Max	Range	Mean	SD	Skewness	Kurtosis	CV (%)
EC ( $\mu\text{mhos/cm}$ )	867	15950	15083	5144.20	3650.63	1.07	0.95	70.97
pH	7.30	9.20	1.90	8.14	0.34	0.20	2.08	4.16
SAR ( $\text{mmol/l}$ ) <sup>0.5</sup>	5.08	30.57	25.49	18.12	7.61	-0.36	-1.11	41.99
(mg/L) TDS	563.55	10367.50	9803.95	3343.73	2372.91	1.07	0.95	70.97

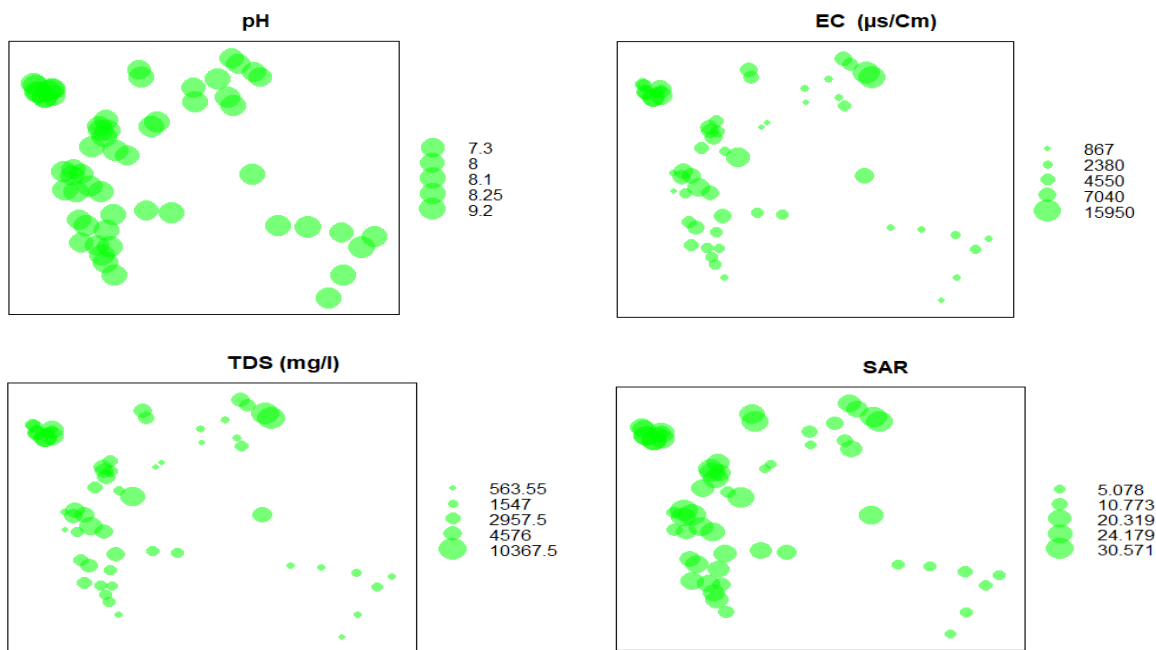
از این آنالیز نرمال بودن داده‌ها است. براساس جدول ۱ و با توجه به اینکه مقادیر چولگی برای پارامترهای اسیدیته و نسبت جذب سدیم در محدوده‌ی یک تا منفی یک قرار گرفته است، بنابراین این دو پارامتر کیفی دارای توزیع نرمال بوده و پارامترهای هدایت الکتریکی و TDS توزیع غیرنرمال داشته و

همان‌طور که در فلوجارت شکل ۲ نشان داده شده است، مدل‌سازی داده‌های مکانی در سیستم GIS ابتدا مستلزم اعمال روش‌های زمین‌آماري می‌باشد. استفاده از روش‌های زمین‌آماري نیازمند بررسی وجود ساختار مکانی در بین داده‌هاست که با آنالیز واریوگرام این امر بررسی می‌شود و شرط استفاده

برای تمامی پارامترهای کیفی مورد بررسی به استثنای اسیدیته (pH)، از شمال و شمال غرب منطقه مورد تحقیق به سمت جنوب شرق آن غلظت‌ها روند کاهشی از خود نشان می‌دهند و مقادیر حداکثر برای پارامترها از چاه‌های حفر شده در بخش شمال و شمال غرب منطقه بدست آمده است. علت این امر احتمالاً برداشت بیشتر از منابع زیرزمینی و نوع مصارف متفاوت به دلیل پیک جمعیتی بالاتر در آن ناحیه می‌باشد.

با لگاریتم‌گیری داده‌ها نرمال شدند (Bameri et al., 2015). جهانشاهی و همکاران (۱۳۹۳) نیز غیر نرمال بودن پارامترهای کیفی آب زیرزمینی نظیر سختی آب، آنیون‌ها و TDS را در تحقیق خود گزارش کردند.

شکل ۳ نقشه پراکنش نقطه‌ای پارامترهای کیفی آب زیرزمینی اندازه‌گیری شده از چاه‌های مشاهداتی در منطقه‌ی مطالعاتی را به وضوح نشان می‌دهد. همانطور که از نقشه‌ها برمی‌آید،



شکل ۳- نقشه پراکنش نقطه‌ای پارامترهای کیفی آب زیرزمینی بر اساس داده‌های مشاهداتی منطقه.

لازم است مناسب‌ترین مدل تئوری بر آن برازش داده شود (نیکبخت و دلبری، ۱۳۹۲). مشخصات بهترین مدل برازش داده شده بر نیم‌تغییرنمای تجربی پارامترهای کیفی آب زیرزمینی در جدول ۲ آمده است.

### آنالیز همبستگی مکانی

به منظور تحلیل‌های زمین‌آماري از ترسیم نیم‌تغییرنمای (واریوگرام) تجربی داده‌ها در نرم‌افزار GS+ استفاده شد. برای دستیابی به یک میانبایی صحیح مهم‌ترین گام برازش مدلی مناسب بر نیم‌تغییرنما است. یعنی قبل از کاربرد نیم‌تغییرنما

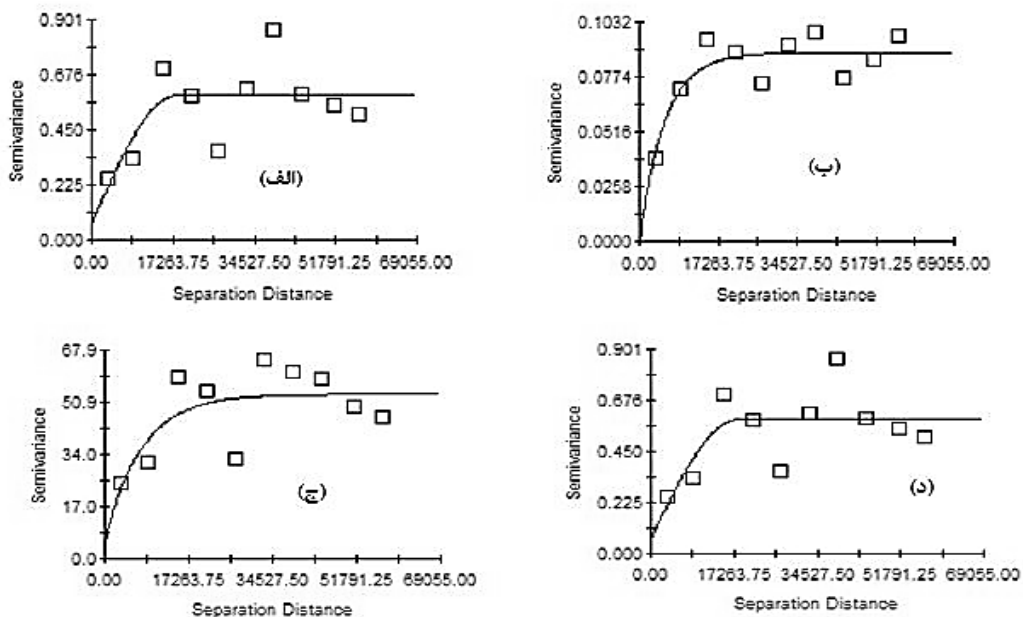
جدول ۲- مشخصات بهترین مدل برازش داده شده بر نیم‌تغییرنمای تجربی پارامترهای کیفی آب زیرزمینی.

Variable	Model	Effective Range (m)	Nugget (Co)	Sill (Co+C)	(Co)/(Co+C)	Spatial Structure	R <sup>2</sup>	RSS
EC	Spherical	17900	0.063	0.59	0.107	Stronge	0.47	0.159
pH	Exponential	16500	0.0001	0.09	0.001	Stronge	0.77	0.0007
SAR	Exponential	22800	5.90	53.61	0.002	Stronge	0.49	906
TDS	Spherical	17900	0.063	0.59	0.107	Stronge	0.47	0.159



نتایج جدول ۲ و بررسی تغییرنمای تجربی نشان داد که پارامترهای مورد ارزیابی دارای ساختار و همبستگی مکانی قوی می‌باشند. بدین صورت که بر اساس نسبت اثر قطعه‌ای بر آستانه  $(Co/Co+C)$  با مقادیر کمتر از ۰/۲۵ برای پارامترهای اسیدیت، نسبت جذب سدیم، نمک محلول کل و هدایت الکتریکی همبستگی مکانی قوی مشاهده می‌گردد. بامری و همکاران (۱۳۹۴) در پژوهش خود برای سدیم و هدایت الکتریکی همبستگی مکانی قوی و برای نسبت جذب سدیم یک همبستگی مکانی متوسط را اعلام کردند. خسروی و همکاران (۱۳۹۸) نیز همبستگی مکانی قوی برای پارامترهای کیفی آب را در مطالعه خود گزارش نمودند که همسو با نتایج این تحقیق است. شکل ۴ بهترین مدل برازش داده شده (نیم تغییرنمای تجربی) بر پارامترهای کیفی مورد مطالعه نشان می‌دهند.

بر اساس مشاهدات بدست آمده از جدول ۲، برای پارامترهای اسیدیت و نسبت جذب سدیم مدل نمایی و برای پارامترهای هدایت الکتریکی و نمک محلول کل مدل کروی بهترین مدل برازش داده شده بودند. طاهری تیزرو و همکاران (۱۳۹۵) برای متغیرهای هدایت الکتریکی، نسبت جذب سدیم و اسیدیت مدل کروی و برای پارامترهای کلر و پتاسیم مدل نمایی را بهترین مدل برازش داده شده در پژوهش خود عنوان نمودند. در مطالعه‌ای دیگر در هند و در غرب شهر دهلی به منظور بررسی کیفیت آب زیرزمینی منطقه، مدل کروی با دامنه تاثیر ۹۱ کیلومتر به عنوان بهترین مدل برازش داده شده بر تغییرات مکانی شوری آب گزارش شده است (Dash et al., 2010). الگوی تغییرات مکانی پارامترهای کیفیت آب از یک دشت به دشت دیگر و از یک منطقه به منطقه ای دیگر با یکدیگر متفاوت بوده و این تفاوت‌ها ناشی از ساختار زمین‌شناسی متفاوت مناطق مورد مطالعه می‌باشد (بامری و همکاران، ۱۳۹۴).



شکل ۴- نیم تغییرنمای تجربی (الف) EC، (ب) pH، (ج) SAR و (د) TDS.

#### ارزیابی عملکرد مدل‌های سیستم اطلاعات جغرافیایی (GIS)

جدول ۳ نتایج ارزیابی روش‌های کریجینگ، کوکریجینگ و وزن‌دهی عکس فاصله را برای پارامتر هدایت الکتریکی نشان می‌دهد. مطابق جدول برای تخمین هدایت الکتریکی از بین ۶ مدل اجرا شده بر اساس شاخص RMSE در مرحله آموزش روش کوکریجینگ ساده (با مقدار ۱۷۱۵/۶۷ میکروموس بر

پس از تجزیه و تحلیل نیم تغییرنما، درونبایی داده‌های کیفی آب زیرزمینی به کمک نرم افزار ArcGIS و به وسیله روش‌های مختلف زمین‌آماري و مطلق انجام گرفت. برای تعیین دقت و صحت روش‌های میانبایی به کاربرده شده از معیارهای ارزیابی آماری استفاده شده و مدل برتر بر اساس مقادیر RMSE کمتر انتخاب گردید.

از روش IDW بدست آوردند که با نتایج این تحقیق هم‌راستا می‌باشد. مقادیر منفی ME در مرحله آزمون بیانگر کم‌برآورد نمودن هدایت الکتریکی توسط مدل‌ها می‌باشد که در تحقیق دیگر محققین نیز گزارش شده است (بامری و همکاران، ۱۳۹۴).

سانتی‌متر) بهترین بازدهی را داشته است. ضریب  $R^2$  نیز برای این روش در مرحله آموزش بالاتر از سایر روش‌ها بود. اما در مرحله آزمون با وجود اختلاف کم بین مقادیر RMSE روش‌های مختلف، مدل IDW با توان دو بهتر از بقیه مدل‌ها عمل نموده است. عینلو و همکاران (۱۳۹۵) در تحقیق خود کمترین میزان RMSE را در مرحله آزمون برای برآورد هدایت الکتریکی

جدول ۳- نتایج ارزیابی مدل‌های GIS در برآورد پارامتر هدایت الکتریکی.

GIS model		Train			Test		
		ME ( $\mu\text{mhos/cm}$ )	RMSE ( $\mu\text{mhos/cm}$ )	$R^2$	ME ( $\mu\text{mhos/cm}$ )	RMSE ( $\mu\text{mhos/cm}$ )	$R^2$
kriging	Ordinary	56.153	3132.06	0.310	-1001.62	1690.74	0.742
	Simple	299.59	3115.96	0.328	-761.35	1617.06	0.742
CoKriging (Covariate: TDS)	Ordinary	71.966	1904.08	0.770	-907.11	1695.33	0.700
	Simple	268.87	1715.67	0.836	-337.60	1702.83	0.609
IDW	Power 1	242.12	3371.25	0.204	-233.87	1515.60	0.672
	Power 2	187.82	3087.42	0.344	-298.57	1504.81	0.709

پایین ( $R^2 > 0/27$ ) هم نشانگر همبستگی ضعیف بین مقادیر مشاهداتی و تخمین زده شده اسیددیده بوده است. جهان‌شاهی و همکاران (۱۳۹۳) در مطالعه خود روش IDW و طاهری تیزرو و همکاران (۱۳۹۵) روش کوکریجینگ را به عنوان روش بهتر برای میانمایی pH گزارش کردند که با این تحقیق همخوانی ندارد. همچنین مدل‌های مورد ارزیابی در مرحله آموزش دچار کم‌برآورد و در مرحله آزمون دچار بیش‌برآورد شدند (به ترتیب بر اساس مقادیر منفی و مثبت شاخص ME).

نتایج ارزیابی روش‌های کریجینگ و وزن‌دهی عکس فاصله برای برآورد اسیددیده آب در جدول ۴ ارائه شده است. لازم به ذکر است که روش کوکریجینگ برای این پارامتر به دلیل عدم همبستگی با دیگر پارامترها و در نتیجه فقدان متغیر کمکی قابل ارزیابی نبوده است. بین روش‌های مورد بررسی بر اساس مقادیر RMSE تفاوت قابل توجهی مشاهده نشد. با این وجود روش کریجینگ معمولی کمترین مقدار RMSE را در هر دو مرحله به خود اختصاص داد. مقادیر ضریب همبستگی

جدول ۴- نتایج ارزیابی مدل‌های GIS در برآورد پارامتر اسیددیده.

GIS model		Train			Test		
		ME	RMSE	$R^2$	ME	RMSE	$R^2$
kriging	Ordinary	-0.011	0.266	0.267	0.046	0.439	0.051
	Simple	-0.017	0.276	0.212	0.004	0.448	0.004
IDW	Power 1	-0.017	0.284	0.171	0.032	0.439	0.043
	Power 2	-0.012	0.296	0.185	0.029	0.449	0.009

و در مرحله آزمون روش کریجینگ معمولی به ترتیب با مقادیر ۴/۷۶۳ و ۲/۶۵۹ بهترین عملکرد را داشتند. ضرایب همبستگی متناظر با این روش‌ها نیز بالاترین مقادیر را نشان می‌دهند (به ترتیب برابر با ۰/۶۶۳ و ۰/۸۴۹). عملکرد خوب مدل کوکریجینگ برای تخمین SAR توسط محققین دیگر نیز

به منظور تحلیل روش‌های تخمین‌گر پارامتر نسبت جذب سدیم، مقادیر شاخص‌های ارزیابی آماری برای ۶ روش میانمایی مورد بررسی در جدول ۵ آمده است. نتایج ارزیابی‌ها نشان می‌دهد که برای این پارامتر کیفی بر اساس مقادیر کمینه RMSE، در مرحله آموزش روش کوکریجینگ معمولی

گزارش شده است (ملکیان و میردشتوان، ۱۳۹۴؛ طاهری تیزرو و همکاران، ۱۳۹۵). درحالیکه ولیزاده و همکاران (۱۳۹۵) روش کریجینگ ساده را به عنوان روش بهتر برای برآورد SAR ذکر نمودند. در اکثر موارد، مقادیر نسبت جذب سدیم توسط مدل‌ها بیش برآورد شده است. بامری و همکاران (۱۳۹۴) نیز در تحقیق خود بیش برآورد SAR توسط مدل کریجینگ معمولی را گزارش نمودند.

جدول ۵- نتایج ارزیابی مدل‌های GIS در برآورد پارامتر نسبت جذب سدیم.

GIS model		Train			Test		
		ME	RMSE	R <sup>2</sup>	ME	RMSE	R <sup>2</sup>
kriging	Ordinary	0.124	6.342	0.324	-0.724	2.659	0.849
	Simple	0.593	6.487	0.308	0.312	3.402	0.759
CoKriging (Covariate: Na)	Ordinary	0.178	4.763	0.663	-0.804	3.279	0.770
	Simple	0.367	4.784	0.662	0.583	4.146	0.647
IDW	Power 1	0.543	6.872	0.210	-0.111	3.00	0.793
	Power 2	0.334	6.530	0.305	0.602	3.423	0.795

جدول ۶ شامل نتایج ارزیابی روش‌های میانبایی برای پارامتر TDS است. بر طبق مندرجات این جدول، برای مرحله آموزش، روش کوکریجینگ ساده با RMSE برابر با ۱۱۱۵/۱۸ میلی گرم بر لیتر بهترین تخمین برای TDS را داشته است. برای مرحله آزمون نیز روش IDW با توان دو با حداقل مقدار RMSE (۹۷۸/۱۳ میلی گرم بر لیتر) بهترین تخمین گر TDS بوده است. عینلو و همکاران (۱۳۹۵) در تحقیق خود گزارش کردند که برای تخمین TDS، روش IDW بهتر از روش کریجینگ عمل نموده است. طاهری تیزرو و همکاران (۱۳۹۵) روش کوکریجینگ را برای برآورد TDS بهتر دانستند.

جدول ۶- نتایج ارزیابی مدل‌های GIS در برآورد پارامتر نمک محلول کل.

GIS model		Train			Test		
		ME (mg/l)	RMSE (mg/l)	R <sup>2</sup>	ME (mg/l)	RMSE (mg/l)	R <sup>2</sup>
kriging	Ordinary	36.499	2035.83	0.310	-651.05	1098.98	0.741
	Simple	194.73	2025.37	0.328	-494.87	1051.09	0.742
CoKriging (Covariate: EC)	Ordinary	46.778	1237.65	0.770	-589.62	1101.97	0.700
	Simple	174.76	1115.18	0.836	-219.44	1106.84	0.609
IDW	Power 1	157.38	2191.31	0.204	-152.02	985.14	0.672
	Power 2	122.09	2006.87	0.344	-194.07	978.13	0.709

### ارزیابی مدل‌های یادگیری ماشین

پارامتر از خود نشان دادند. در مرحله آموزش مدل‌های SVM، ANN و RF به ترتیب با مقادیر RMSE برابر با ۲۱۲/۰۹، ۲۸۴/۹۳ و ۲۹۳/۸۶ و در مرحله آزمون مدل‌های SVM، RF و ANN به ترتیب با مقادیر RMSE برابر با ۲۴۳/۰۵، ۲۲۴/۱ و ۷۴۲ کارایی بهتری از خود نشان دادند. ضرایب همبستگی بالای ۰/۹۵ در هر دو مرحله و در تمامی مدل‌ها بیانگر دقت قابل قبول مدل‌ها و یادگیری خوب و کارآمد آنها با استفاده از الگوریتم آموزشی مورد نظر و داده‌های ارایه شده است.

پس از مدل‌سازی و تحلیل‌های مدل‌های مختلف GIS، تجزیه و تحلیل داده‌های کیفی آب زیرزمینی به کمک نرم افزار R و به وسیله روش‌های مختلف یادگیری ماشین نظیر جنگل تصادفی (RF)، شبکه عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) انجام گرفت. نتایج ارزیابی روش‌های یادگیری ماشین برای پارامتر هدایت الکتریکی (جدول ۷) نشان می‌دهد که هر سه مدل دقت بالایی در تخمین این

کارایی مناسب هر دو مدل ANN و SVM را در تخمین هدایت الکتریکی بر اساس دیگر پارامترهای کیفی اذعان داشتند. هرچند در تحقیق آنان مقادیر خطای حاصل از اجرای مدل ANN کمتر از مدل SVM بود.

میرسنجری و همکاران (۱۳۹۶) نیز در پژوهش خود ضریب همبستگی بالای ۰/۹۸ و RMSE کمتر از ۱/۹ را برای تخمین EC توسط مدل ANN اعلام نمودند که نشان از دقت و عملکرد خوب مدل هوش مصنوعی در تخمین شوری آب زیرزمینی دارد. عیسی زاده و همکاران (۱۳۹۸) نیز در تحقیق خود

جدول ۷- نتایج ارزیابی مدل‌های یادگیری ماشین در برآورد پارامتر هدایت الکتریکی.

ML model	Train			Test		
	bias	RMSE	R <sup>2</sup>	bias	RMSE	R <sup>2</sup>
ANN	0.670	284.93	0.994	-506.01	742.00	0.951
RF	-30.71	293.86	0.995	-3.66	224.10	0.992
SVM	123.03	212.09	0.999	175.35	243.05	0.997

مقادیر مشاهداتی و مقادیر تخمین زده شده نشان می‌دهد که بین متغیر وابسته (pH) و متغیرهای مستقل (سایر پارامترهای کیفی اندازه گیری شده) همبستگی خوبی وجود نداشته و در نتیجه مدل‌ها قادر به تخمین دقیق مقادیر اسیدیته نشده‌اند. Haghyabi et al., (2018) در پژوهش خود مدل SVM را به عنوان مدل دقیق‌تر برای پیش‌بینی اسیدیته اعلام داشتند. همچنین این محققین کارایی ضعیف‌تر مدل‌های اجرا شده (ANN، SVM و GMDH) را در تخمین pH نسبت به سایر پارامترهای کیفی گزارش نمودند. هم‌راستا با تحقیق حاضر، این محققین نیز ضرایب همبستگی پایینی برای هر سه مدل در تخمین pH در مرحله آزمون ( $R^2 > 0/35$ ) بدست آوردند.

نتایج ارزیابی برآورد پارامتر کیفی اسیدیته با کاربرد سه روش یادگیری ماشین در جدول ۸ حاکی از آن است که در مرحله آموزش بر اساس مقادیر RMSE به ترتیب مدل‌های RF، ANN و SVM با مقادیر ۰/۲۲۵، ۰/۲۵۹ و ۰/۱۴ توانستند پیش‌بینی اسیدیته آب زیرزمینی را با دقت بهتری به انجام برسانند. اما در مرحله آزمون این ترتیب دستخوش تغییر شد و با وجود نزدیک بودن نتایج سه مدل، حداقل مقدار RMSE به ترتیب با اجرای مدل‌های ANN، SVM و RF حاصل شد. مقادیر نه چندان خوب ضریب همبستگی در هر دو مرحله اجرای مدل‌ها (به استثنای مدل RF در مرحله آموزش) نشان از همبستگی نامناسب بین مقادیر پیش‌بینی شده و مقادیر مشاهداتی واقعی pH می‌دهد. وجود عدم همبستگی بین

جدول ۸- نتایج ارزیابی مدل‌های یادگیری ماشین در برآورد پارامتر اسیدیته.

ML model	Train			Test		
	bias	RMSE	R <sup>2</sup>	bias	RMSE	R <sup>2</sup>
ANN	-0.0003	0.225	0.466	-0.189	0.408	0.473
RF	0.011	0.140	0.855	0.003	0.456	0.139
SVM	-0.038	0.259	0.308	-0.044	0.416	0.304

همبستگی در همه مدل‌ها نشان از دقت بالای مدل‌ها در تخمین پارامتر نسبت جذب سدیم دارد. El Bilali and Taleb, (2020) در تحقیق خود به ارزیابی مدل‌های مختلف یادگیری ماشین برای پیش‌بینی پارامترهای کیفی آب آبیاری پرداختند. این محققین گزارش کردند که برای تخمین پارامتر نسبت جذب سدیم مدل‌های ANN، RF و SVM به ترتیب با

نتایج ارزیابی مدل‌های یادگیری ماشین برای برآورد پارامتر نسبت جذب سدیم در جدول ۹ آمده است. با توجه به قانون انتخاب مدل کارتر بر اساس حداقل مقدار RMSE و حداکثر مقدار R<sup>2</sup> می‌توان در مرحله آموزش به ترتیب مدل‌های ANN، RF و SVM و در مرحله آزمون به ترتیب مدل‌های SVM، ANN و RF را انتخاب نمود. هرچند مقادیر بالای ضریب

مقادیر حداقل RMSE و حداکثر  $R^2$  دقت بالایی از خود نشان دادند که با نتایج این مطالعه هم‌راستا می‌باشد. میرسنجری و همکاران (۱۳۹۶) نیز در تحقیق خود از مدل ANN به عنوان ابزاری کارا و دقیق در برآورد SAR نام بردند و دقت تخمین این پارامتر را با این روش با ضریب همبستگی بالای ۹۲ درصد گزارش نمودند.

جدول ۹- نتایج ارزیابی مدل‌های یادگیری ماشین در برآورد پارامتر نسبت جذب سدیم.

ML model	Train			Test		
	Bias	RMSE	$R^2$	Bias	RMSE	$R^2$
ANN	-0.005	0.663	0.992	0.484	2.176	0.923
RF	0.025	0.948	0.985	1.575	2.292	0.932
SVM	0.013	1.325	0.970	-0.760	1.846	0.936

نتایج ارزیابی روش‌های یادگیری ماشین برای تخمین پارامتر TDS در جدول ۱۰ آمده است. با توجه به جدول مذکور، هم در مرحله آموزش و هم مرحله آزمون مدل SVM عملکرد بهتری از خود در پیش‌بینی TDS نشان داده است به طوری که حداقل مقدار RMSE (۱۳۷/۸۶ و ۱۵۷/۹۹) به ترتیب در مرحله آموزش و آزمون) و بیشترین مقدار  $R^2$  (۰/۹۹۹ و ۰/۹۹۷) به ترتیب در مرحله آموزش و آزمون) توسط این مدل حاصل شد. پس از این مدل، عملکرد دو مدل RF و ANN در مرحله آموزش بسیار بهم نزدیک بوده (با مقدار RMSE به ترتیب ۱۸۸/۳۵ و ۱۸۲/۵۶) اما در مرحله آزمون مدل RF بسیار بهتر از ANN عمل نموده است (با مقدار RMSE ۱۷۰/۲۴ در مقابل

نتایج ارزیابی روش‌های یادگیری ماشین برای تخمین پارامتر TDS در جدول ۱۰ آمده است. با توجه به جدول مذکور، هم در مرحله آموزش و هم مرحله آزمون مدل SVM عملکرد بهتری از خود در پیش‌بینی TDS نشان داده است به طوری که حداقل مقدار RMSE (۱۳۷/۸۶ و ۱۵۷/۹۹) به ترتیب در مرحله آموزش و آزمون) و بیشترین مقدار  $R^2$  (۰/۹۹۹ و ۰/۹۹۷) به ترتیب در مرحله آموزش و آزمون) توسط این مدل حاصل شد. پس از این مدل، عملکرد دو مدل RF و ANN در مرحله آموزش بسیار بهم نزدیک بوده (با مقدار RMSE به ترتیب ۱۸۸/۳۵ و ۱۸۲/۵۶) اما در مرحله آزمون مدل RF بسیار بهتر از ANN عمل نموده است (با مقدار RMSE ۱۷۰/۲۴ در مقابل

جدول ۱۰- نتایج ارزیابی مدل‌های یادگیری ماشین در برآورد پارامتر نمک محلول کل.

ML model	Train			Test		
	bias	RMSE	$R^2$	bias	RMSE	$R^2$
ANN	0.74	182.56	0.994	-295.84	551.79	0.919
RF	-24.49	188.35	0.995	37.95	170.24	0.989
SVM	79.97	137.86	0.999	113.98	157.99	0.997

و بیش از ۳۰ درصد نشانه ضعیف بودن مدل است. بر این اساس و با توجه به مقادیر ذکر شده در جدول ۱۱، برای هر چهار پارامتر کیفی مورد بررسی کاملاً مشهود است که عملکرد مدل‌های یادگیری ماشین بهتر و دقیق‌تر از روش‌های سیستم اطلاعات جغرافیایی بوده است. همانطور که از مقادیر NRMSE مشخص می‌باشد برای دو پارامتر EC و TDS عملکرد سه روش یادگیری ماشین دقیق و برای پارامتر SAR عملکرد مناسبی داشتند. هرچند در تخمین اسیدیت با وجود عملکرد بهتر نسبت به مدل‌های GIS، روش‌های ماشین عملکرد ضعیفی

#### مقایسه عملکرد مدل‌های اجرا شده

با توجه به این که مقدار RMSE متناسب با واحد اندازه‌گیری متغیر است، مقایسه مقدار آن بین مدل‌های ساخته شده برای دو متغیر با واحدهای متفاوت درست نخواهد بود. لذا مقدار RMSE را به دامنه داده‌های متغیر وابسته تقسیم کرده و آن را RMSE نرمال شده (NRMSE) می‌نامند. این معیار برای مقایسه مدل‌های مختلف مناسب خواهد بود. لازم به ذکر است که NRMSE زیر ۱۰ درصد نشان‌دهنده دقیق بودن مدل، ۱۰-۲۰ درصد مناسب بودن مدل، ۲۰-۳۰ درصد دقت متوسط

داشتند. نتایج نزدیک سه مدل یادگیری ماشین در تخمین پارامترهای کیفی حاکی از قدرت و دقت بالای مدل‌های یادگیری ماشین در پیش‌بینی پارامترهای کیفی آب زیرزمینی می‌باشد (عیسی زاده و همکاران، ۱۳۹۸). در مدل‌های GIS، روش زمین آماری کوکریجینگ نسبت به دو روش دیگر عملکرد بهتر و در حد متوسط (بر اساس NRMSE%) از خود نشان داد.

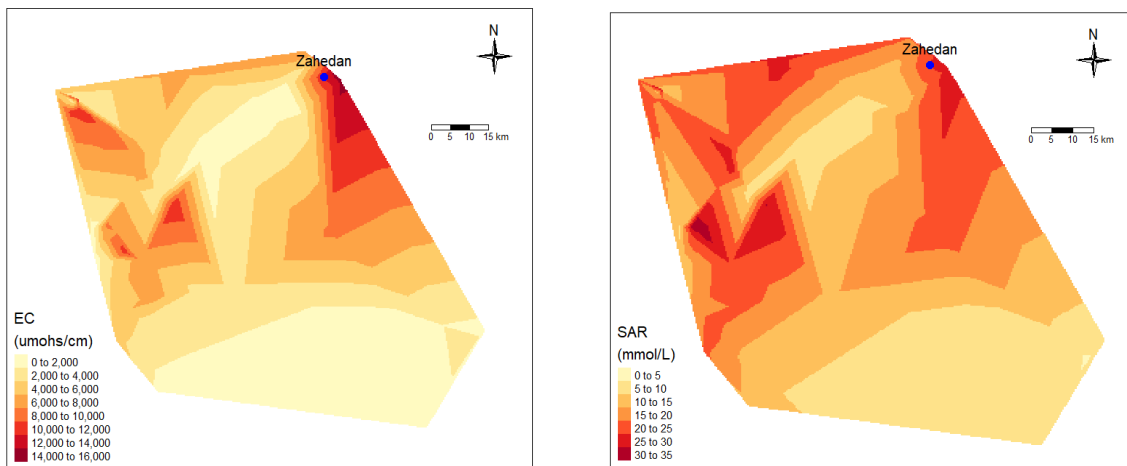
جدول ۱۱- مقایسه عملکرد مدل‌های مختلف در تخمین پارامترهای مورد تحقیق بر اساس NRMSE کل (%).

Parameter	Kriging		CoKriging		IDW		SVM	RF	ANN
	Simple	Ordinary	Simple	Ordinary	Power 1	Power 2			
EC	31.4	31.9	22.7	23.9	32.4	30.5	3.0	3.4	6.8
pH	38.1	37.1	-	-	38.0	39.2	35.5	31.3	33.3
SAR	38.8	35.3	35.1	31.6	38.7	39.1	12.4	12.7	11.1
TDS	31.4	31.9	22.7	23.9	32.4	30.5	3.1	3.7	7.6

می‌گیرد. به طوریکه بر اساس متغیرهای کیفی EC، TDS و SAR می‌توان چنین نتیجه گرفت که آبخوان در وضعیت مناسبی قرار ندارد. تغییرات مکانی کیفیت آب در این دشت نشان می‌دهد که کیفیت آب در نیمه شمالی بخصوص شمال شرقی دشت نسبت به نیمه‌ی جنوبی آن تنزل پیدا کرده است، که علت آن را می‌توان احتمالاً تراکم جمعیت در منطقه‌ی شهر زاهدان و تغییر و افت کیفیت آب به دلیل مصرف بی‌رویه و نامناسب آب زیرزمینی دانست. بر اساس نقشه‌های توزیع مکانی به وضوح مشخص است که بهترین کیفیت آب زیرزمینی متعلق به جنوب دشت بوده و در نتیجه برای مدیریت کیفی باید تمرکز بیشتر را بر قسمت شمالی به خصوص شمال شرق دشت قرار داد و به دنبال راهکارهای مدیریتی برای افزایش کیفیت آب آبخوان بود.

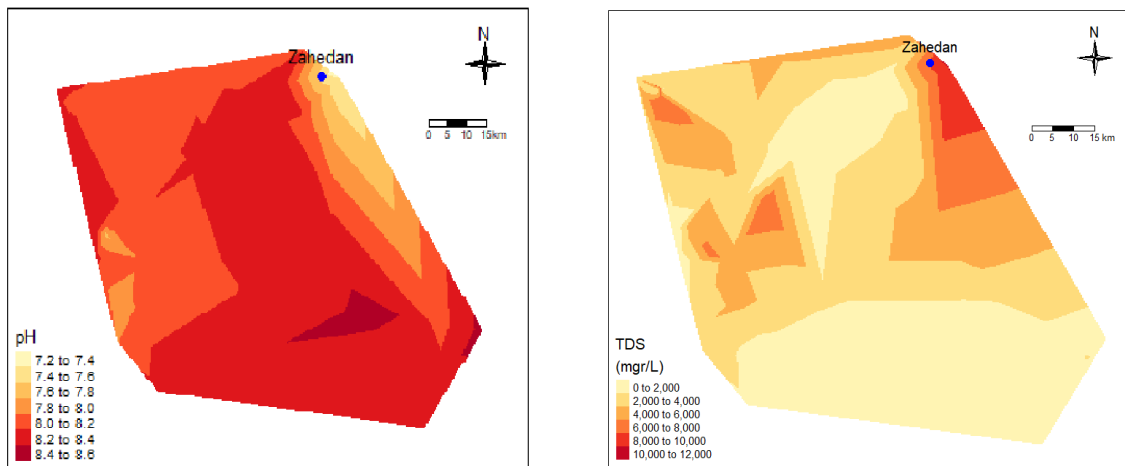
### پهنه‌بندی پارامترهای کیفی بر اساس بهترین مدل تخمین گر

پهنه‌بندی پارامترهای کیفی آب زیرزمینی در دشت زاهدان پس از مدل‌سازی این متغیرهای کیفی با استفاده از مدل‌های زمین آمار و یادگیری ماشین و ارزیابی آن‌ها انجام شد. پهنه‌بندی برای هر پارامتر کیفی بر اساس نتایج بدست آمده از بهترین مدل تخمین گر برای آن پارامتر با داشتن کمترین خطا و همبستگی بالاتر انجام گرفت. بر این اساس، برای متغیر EC و TDS مدل SVM، برای SAR مدل ANN و برای pH مدل RF منتخب و پهنه‌بندی در محیط نرم افزار R انجام شد. شکل (۵) پهنه‌بندی و نحوه‌ی توزیع پارامترهای کیفی مورد بررسی را در منطقه مطالعاتی نشان می‌دهد. نقشه‌های پهنه‌بندی نشان می‌دهد که کیفیت آبخوان دشت زاهدان متناسب با هر یک از پارامترها در وضعیت متفاوتی قرار



شکل ۵- تغییرات مکانی متغیرهای کیفی آب زیرزمینی مورد بررسی در دشت زاهدان.





ادامه شکل ۵- تغییرات مکانی متغیرهای کیفی آب زیرزمینی مورد بررسی در دشت زاهدان.

### نتیجه گیری

برای پارامترهای اسیدیته و نسبت جذب سدیم مدل نمایی و برای پارامترهای هدایت الکتریکی و نمک محلول کل مدل کروی بهترین مدل برازش داده شده بر نیم تغییرنمای تجربی بوده و همه پارامترهای مورد ارزیابی دارای ساختار و همبستگی مکانی قوی بودند. نتایج ارزیابی مدل‌های GIS نشان‌دهنده عملکرد نسبتاً ضعیف این روش‌ها بود. هرچند به طور کلی بر اساس مقادیر NRMSE روش کوکریجینگ عملکرد بهتری نسبت به سایر روش‌ها داشت. نتایج ارزیابی عملکرد مدل‌های یادگیری ماشین مورد مطالعه نیز حاکی از کارایی بالای هر سه مدل در پیش‌بینی سه پارامتر EC، TDS و SAR با ضرایب همبستگی بالای ۹۰ درصد و مقادیر NRMSE زیر ۱۵ درصد می‌باشد که بیانگر دقت بالای مدل‌ها در تخمین پارامترها می‌باشد. هر چند هر سه مدل نتایج قابل قبولی برای برآورد پارامتر اسیدیته از خود نشان ندادند. جهان‌شاهی و همکاران (۱۳۹۳)، بامری و همکاران (۱۳۹۴)، روکی و همکاران (۱۳۹۶)، حسنی و همکاران (۱۳۹۷)، عیسی زاده و همکاران (۱۳۹۸)، (2017) Granata et al., Haghyabi, (2018) et al., (2019) Ahmed et al., (2020) Boi et al., و (2020) El Bilali and Taleb نیز در تحقیقات مشابه، قابلیت خوب مدل‌های زمین‌آماري و روش‌های یادگیری ماشین را در شبیه‌سازی کمی و کیفی منابع آب در مناطق مختلف نشان دادند. در نهایت، می‌توان چنین ادعان داشت که مدل‌های آماری جدید نظیر روش‌های یادگیری ماشین با یک سطح اطمینان قابل قبول می‌تواند در پیش‌بینی پارامترهای شوری و

نظر به اینکه بخش کشاورزی در استان سیستان و بلوچستان مقدار قابل توجهی از منابع آب‌های زیرزمینی را جهت تولید محصولات کشاورزی مصرف می‌کند و بزرگترین مصرف‌کننده‌ی آب در این استان است، بنابراین بررسی وضعیت کمی و کیفی منابع آب زیر زمینی به دلیل تاثیر عمده‌ای که بر کمیت و کیفیت محصولات کشاورزی و متعاقباً سلامت انسان‌ها و محیط زیست دارد، امری ضروری می‌باشد. اطلاع از وضعیت پارامترهای کیفی آبخوان‌ها به مدیریت صحیح منابع آب کمک شایانی خواهد کرد. هرچند، پایش کیفی آب-های زیرزمینی بسیار پرهزینه و زمان بر بوده و انتخاب روشی که در آن با حداقل پارامترهای هیدروشیمیایی بتوان پیش-بینی نسبتاً دقیقی از طبقه کیفیت آب برای مصارف مختلف داشت، مهم و ضروری است. بنابراین در تحقیق حاضر، شش مدل سیستم اطلاعات جغرافیایی (GIS) شامل روش‌های زمین آماری (کریجینگ ساده و معمولی، کوکریجینگ ساده و معمولی) و روش‌های مطلق (IDW با توان یک و دو) و نیز سه مدل مختلف یادگیری ماشین شامل مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی برای پیش-بینی پارامترهای کیفی آب زیرزمینی نظیر هدایت الکتریکی، SAR، اسیدیته و TDS در دشت زاهدان مورد ارزیابی و مقایسه قرار گرفت. روش‌های زمین آماری این تحقیق در محیط نرم افزارهای GS+ و ArcGIS انجام شد. روش‌های یادگیری ماشین نیز در محیط نرم‌افزار R اجرا شد. با توجه به نتایج بدست آمده

حکمت نیا، م.، حسینی، س.م.، و صفدری، م.، ۱۳۹۹. مدیریت منابع آب کشاورزی استان سیستان و بلوچستان از دیدگاه آب مجازی. نشریه علمی پژوهشی مهندسی آبیاری و آب ایران، ۴۱(۳): ۱۴۹-۱۳۷.

خسروی، ی.، زمانی، ع.، و تکین، ف. ز.، ۱۳۹۸. تحلیل مکانی ویژگیهای فیزیکوشیمیایی آبهای زیرزمینی در جنوب و جنوب غربی حوضه آبریز دالکی استان بوشهر. مجله کاوش‌های جغرافیایی مناطق بیابانی، ۷(۲): ۱۴۵-۱۲۵.

دولت کردستانی، م.، نوحه‌گر، ا. و جانی‌زاده، س.، ۱۳۹۶. ارزیابی چند مدل هوش مصنوعی برای پیش بینی کیفیت آب زیرزمینی (مطالعه موردی: دشت گرو). مجله علمی-پژوهشی مهندسی اکوسیستم بیابان، ۶(۱۷): ۴۲-۲۷.

رخ‌شاد، ا.م.، و شهیدی، ع.، ۱۴۰۰. عملکرد شش روش ترکیبی هوشمند در مدل‌سازی کیفیت آب زیرزمینی مطالعه موردی: دشت بافق. نشریه هیدروژئولوژی، ۶(۱): ۱۳۹-۱۲۶.

رستگاریان، م.، و صنعتی، ع. ۱۳۹۵. تعیین واحدهای جریان هیدرولیکی با استفاده از روش جنگل تصادفی برای یکی از مخازن نفتی ایران، ماهنامه علمی- ترویجی اکتشاف و تولید نفت و گاز، ۱۳۵: ۵۹-۵۵.

رضایی، م.، ثامن، ع.م.، و فلاح شمسی، س.ر.، ۱۳۹۷. استفاده از روش‌های پیشرفته یادگیری ماشین در پیش‌فرسایش بادی در جنوب ایران، نشریه پژوهش‌های فرسایش محیطی، ۸(۱): ۵۸-۳۹.

روکی، ر.، آریافر، ا.، و عادل‌نسب، ج.، ۱۳۹۶. بررسی کیفیت آب زیرزمینی آبخوان دشت گناباد خراسان رضوی با استفاده از روش‌های آماری چندمتغیره و هوش مصنوعی، نشریه مهندسی منابع معدنی، ۲(۱): ۶۱-۴۹.

ریاحی، ف.، وقارفر، ح.، دانشکار آراسته، پ.، و کاردان مقدم، ح.، ۱۳۹۹. ارزیابی پتانسیل منابع آب زیرزمینی با استفاده از ترکیب روش‌های داده‌کاوی (منطقه مورد مطالعه: دشت سرخون هرمزگان). نشریه حفاظت منابع آب و خاک، ۹(۳): ۱۲۰-۱۰۵.

نسبت جذب سدیم آب زیرزمینی، که دو متغیر مهم در دسته بندی کیفیت آب می‌باشد، مورد استفاده قرار گیرد و به عنوان یک ابزار توانمند و کارآمد در جهت پایش کیفی و مدیریت بهینه منابع آب زیرزمینی به مدیران ذیربط کمک نماید. از سویی پیشنهاد می‌شود در جهت تکمیل نتایج تحقیق حاضر، در مطالعات آتی کارایی دیگر روش‌های نوین مبتنی بر داده و نیز در گستره‌ی وسیع‌تری از حوضه آبخیز و پارامترهای مورد بررسی مورد تحقیق و ارزیابی قرار گیرد.

## منابع

ادیب، آ. و زمانی، ر.، ۱۳۹۴. بررسی تغییرات مکانی شاخص‌های کیفی آب زیرزمینی دشت دزفول با استفاده از زمین‌آمار، مجله مهندسی منابع آب، ۸(۲۷): ۱۲-۱.

امیدوار، ک.، شفیع، ش.، تقی‌زاده، ز.، و علی‌پور، م.، ۱۳۹۳. ارزیابی کارایی مدل درخت تصمیم در پیش‌بینی بارش ایستگاه سینوپتیک کرمانشاه، نشریه تحقیقات کاربردی علوم جغرافیایی، ۳۴(۳): ۱۱۰-۸۹.

بامری، ا.، پیری، ح.، و گنجی، ف.، ۱۳۹۴. ارزیابی آلودگی آب‌های زیرزمینی دشت بجستان جهت مصارف کشاورزی با استفاده از روش کریجینگ شاخص. مجله پژوهش‌های حفاظت آب و خاک، ۲۲(۱): ۲۲۹-۲۱۱.

پورقاسمی، ح.، ذبیحی، م. و بهزادفر، م.، ۱۳۹۴. تهیه نقشه پتانسیل آب زیرزمینی با استفاده از مدل‌های انترپولی شانون و جنگل تصادفی در دشت بجنورد. مجله اکوهیدرولوژی، ۲(۲): ۲۳۲-۲۲۱.

جهانشاهی، ا.، روحی مقدم، ع.، و دهواری، ع.ح.، ۱۳۹۳. ارزیابی پارامترهای کیفی آب زیرزمینی با استفاده از GIS و زمین‌آمار (مطالعه موردی: آبخوان دشت شهرباک). نشریه دانش آب و خاک، ۲۴(۲): ۱۹۷-۱۸۳.

حسینی، ز.، میرعباسی نجف‌آبادی، ر.، و قاسمی، ا.ر.، ۱۳۹۷. پیش‌بینی کیفیت آب زیرزمینی دشت خانمیرزا با استفاده از روش تصمیم‌گیری درختی. مجله هیدروژئولوژی، ۳(۱): ۱۱۰-۹۹.

حسینی‌پاک، ع.ا.، ۱۳۷۷. زمین‌آمار (ژئواستاتستیک). انتشارات دانشگاه تهران. ۳۰۶ ص.

- زارع ابیانه، ح.، بیات ورکشی، م.، اخوان، س. و محمدی، م.، ۱۳۹۰. تخمین نیترات آب زیرزمینی دشت همدان- بهار با استفاده از شبکه عصبی مصنوعی و اثر تفکیک داده‌ها بر دقت پیش‌بینی. نشریه محیط شناسی، ۵۸: ۱۴۰-۱۲۹.
- سیفی، ا.، میرلطیفی، س.م.، و ریاحی، ح.، ۱۳۹۲. معرفی و کاربرد ماشین بردار پشتیبان حداقل مربعات در برآورد تبخیر تعرق مرجع و تحلیل عدم قطعیت نتایج، مطالعه موردی شهر کرمان، فصلنامه علمی پژوهشی مهندسی آبیاری و آب، ۱۳(۳): ۶۷-۷۹.
- صمدیان فرد، س.، و پناهی، س.، ۱۳۹۷. برآورد تبخیر- تعرق مرجع روزانه با استفاده از روشهای داده کاوی رگرسیون بردار پشتیبان و مدل درختی M5، پژوهشنامه مدیریت حوزه آبخیز، ۹(۱۸): ۱۶۷-۱۵۷.
- طاهری تیزرو، ع.، نوابیان، م.، و بدخشان، آ.، ۱۳۹۵. کاربرد تکنیک زمین آمار برای ارزیابی کیفیت آب زیرزمینی دشت فومنات استان گیلان. نشریه هیدروژئولوژی، ۱(۱): ۴۴-۵۶.
- عجم‌زاده، ع.، ملائی نیا، م.ر.، و قندهاری، ق.، ۱۳۹۶. مقایسه برخی از روش‌های هوش مصنوعی در پیش‌بینی سری-های زمانی روزانه دمای حداقل، حداکثر و بارش ایستگاه سد تنگاب واقع در استان فارس. فصلنامه علمی- پژوهشی فضای جغرافیایی، ۵۹(۳): ۲۰۵-۲۲۸.
- عربی یزدی، ا.، نیکنیا، ن.، مجیدی، ن.، و امامی، ح.، ۱۳۹۳. بررسی امنیت آبی در اقلیم‌های خشک از دیدگاه شاخص رد پای آب (مطالعه موردی: استان خراسان جنوبی). نشریه آبیاری و زهکشی ایران، ۸(۴): ۷۶۴-۷۳۵.
- عزیزپور، ع.، ایزدبخش، م.ع.، شعبانلو، س.، یوسفوند، ف. و رجبی، ا.، ۱۴۰۰. شبیه‌سازی تراز، کلر و بی‌کربنات آب زیرزمینی توسط ماشین آموزش ترکیبی. نشریه هیدروژئولوژی، ۱۱۳(۱): ۹۹-۱۱۳.
- علی‌پور، ع.، رحیمی، ج.، و آذرینوند، ع.، ۱۳۹۶. بررسی کیفیت آب زیرزمینی جهت مصارف شرب و کشاورزی-پیش‌نیازی برای برنامه‌ریزی آمایش سرزمین در مناطق خشک و نیمه خشک ایران. نشریه مرتع و آبخیزداری (منابع طبیعی ایران)، ۷۰(۲): ۴۳۴-۴۲۳.
- عیسی زاده، م.، بی آزار، س.م.، اشرف زاده، ا.، و خانجانی، ر.، ۱۳۹۸. تخمین پارامترهای کیفی آبخوان دشت گیلان با استفاده از آزمون گاما و مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی. فصلنامه علوم و تکنولوژی محیط زیست، ۲۱(۲): ۲۱-۱.
- عینلو، ف.، معافی رابری، ع.، ملکیان، آ.، قضاوی، ر.، و محسنی ساروی، م.، ۱۳۹۵. بررسی کیفیت آب زیرزمینی دشت زنجان از نظر استانداردهای شرب با استفاده از رویکرد زمین آمار، نشریه جغرافیا و برنامه ریزی محیطی، ۶۲(۲): ۱-۱۶.
- غضبان، ف.، ۱۳۸۱. زمین‌شناسی زیست محیطی، چاپ اول، انتشارات دانشگاه تهران، ایران، ۴۱۶ ص.
- کلانتری، ن.، ایرانی اصل، ا.، و محمدی، ه.، ۱۴۰۱. بررسی کمی و کیفی آبخوان باغملک با استفاده از روش تحلیل آماری، روش‌های هیدروشیمیایی و GIS. نشریه هیدروژئولوژی، ۷(۱): ۲۴-۱۱.
- محمدی، ص.، سلاجقه، ع.، مهدوی، م. و باقری، ر.، ۱۳۹۱. بررسی تغییرات مکانی و زمانی سطح آب زیرزمینی دشت کرمان با استفاده از روش زمین آماری مناسب (طی یک دوره آماری ۱۰ ساله، ۱۳۷۵-۱۳۸۵)، فصلنامه تحقیقات مرتع و بیابان ایران، ۱۹(۱): ۷۱-۶۰.
- مدنی، ح.، ۱۳۷۳. مبانی زمین‌آمار، انتشارات دانشگاه امیر کبیر، ۶۵۹ ص.
- ملکیان، آ.، و میردشتوان، م.، ۱۳۹۴. بررسی کیفیت آب زیرزمینی جهت مصارف کشاورزی بر اساس تحلیل‌های زمین‌آمار (مطالعه موردی: دشت هشتگرد استان البرز). مرتع و آبخیزداری، مجله منابع طبیعی ایران، ۶۸(۴): ۸۲۰-۸۰۹.
- میرسنجری، م.م.، محمدیاری، ف.، بصیری، ر. و حمیدی‌پور، ف.، ۱۳۹۶. مدل‌سازی پارامترهای کیفی SAR، EC و TDS در آب زیرزمینی با استفاده از شبکه عصبی مصنوعی (مطالعه موردی: دشت مهران و دهلران). فصلنامه انسان و محیط زیست، ۴۲(۳): ۱۲-۱.
- میرک زهی، خ.، شهریاری، ع.، پهلوان راد، م.ر.، و بامری، ا.، ۱۳۹۶. کاربرد روش درختان تصمیم‌گیری تصادفی در پیش‌بینی کلاس‌های خاک در اراضی با پستی و بلندی

- Delhi, India using geostatistical approach. *Environ Monit Assess.* 167: 599-615.
- Delbari, M., Afrasiab, P., and Loiskandl, W., 2011. Geostatistical analysis of soil texture fractions on the field scale. *Soil Water Research.* 6: 173-189.
- El Bilali, A., and Taleb, A., 2020. Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *Journal of the Saudi Society of Agricultural Sciences*, 19: 439-451.
- Granata, F., Papirio, S., Esposito, G., Gargano, R., and Marinis, G., 2017. Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators. *Water*, 9(105): 1-12.
- Haghiabi, A.H., Nasrolahi, A.H., and Parsaie, A., 2018. Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53 (1): 3-13.
- Isaaks, E.H., and Srivastava, M.R., 1989. *Applied Geostatistics.* (No. 551.72 ISA).
- Kardam, A., Raj, K.R., Arora, J.K., and Srivastava, S., 2012. Artificial neural network modeling for biosorption of Pb (II) ions on nanocellulose fibers. *Bionanoscience*, 2: 153-160.
- Ma, W., Tan, K., and Du, P., 2016. Predicting soil heavy metal based on random forest model. *International Geoscience and Remote Sensing Symposium (IGARSS)*: 4331-4334.
- Rahmati, O., Pourghasemi, H.R., and Melesse, A.M., 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena*, 137: 360-372.
- Sarhadi, M., Nohtani, M., and Reiki, M., 2016. Effect of drought on qualitative and quantitative parameters of Zahedan plain aquifer. *ECOPERSIA*, 4(4): 1541-1554.
- Taheri-Tizro, A., Voudouris, K. and Vahedi, S., 2014. Spatial variation of groundwater quality parameters: a case study from a semiarid region of Iran. *International Bulletin of Water Resources & Development*, 1(3): 1-14.
- Wilding, L.P., Smeck, N. E., and Hall, G.F., 1983. *Pedogenesis and soil taxonomy. I. Concepts and interactions.* Elsevier Publishing Company, 303p.
- Zhu, Q., and Lin, H.S., 2010. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. *Pedosphere*, 20(5): 594-606.
- Zhu, X., Wu, G., Coulon, F., Wu, L., and Chen, D., 2018. Correlating asphaltene dimerization with its molecular structure by potential of mean force calculation and data mining. *Energy Fuel*, 32: 5779-5788.
- کم (مطالعه موردی: شهرستان هیرمند)، نشریه پژوهش های حفاظت آب و خاک، ۲۴(۱): ۸۴-۶۷.
- نوروزی، ح.، ندیری، ع.، اصغری مقدم، ا.، و قره خانی، م.، ۱۳۹۶. پیش بینی قابلیت انتقال آبخوان دشت ملکان با استفاده از روش جنگل تصادفی، نشریه دانش آب و خاک، ۲۷(۲): ۷۵-۶۱.
- نیکبخت، ص.، و دلبری، م.، ۱۳۹۲. برآورد سطح ایستابی آب های زیرزمینی با استفاده از روش های زمین آماری (مطالعه موردی: دشت زاهدان)، نشریه آب و توسعه پایدار، ۱(۱): ۵۶-۴۹.
- ولیزاده، خ.، روستایی، ش.، رحیم پور، ت.، و نخستین روحی، م.، ۱۳۹۵. تعیین مناسب ترین روش زمین آمار در تهیه تغییرات شوری آب های زیرزمینی (مطالعه موردی: دشت شیرامین، استان آذربایجان شرقی). *مجله هیدروژئومورفولوژی*، ۶(۱): ۳۲-۱۷.
- Ahmed, U., Mumtaz, R., Anwa, H., Shah, A.A., Irfan, R., and García-Nieto, J., 2019. Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, 11: 1-14.
- Antonopoulos, V. Z., Papamichail, D. M., and Mitsiou, K. A., 2001. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. *Hydrology and Earth System Sciences* 5 (4): 679-691.
- Bameri, A., and Khaleghi, M., 2021. Monitoring and predicting the groundwater-level fluctuations for better management of groundwater resource in Lowlands Using Geographic Information System (GIS). *Journal of Radar and Optic Remote Sensing*, 4(2):7-16.
- Bameri, A., Khormali, F., Kiani, F., and Dehghani, A.A., 2015. Spatial variability of soil organic carbon in different hillslope positions in Toshan area, Golestan Province, Iran: geostatistical approaches. *Journal of Mountain Science*, 12(6): 1422-1433.
- Bui, D.T., Khosravi, K., Tiefenbacher, J., Nguyen, H., and Kazakis, N., 2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms, *Science of the Total Environment*, 721, 137612.
- Cambardella, C.A., Moorman, T.B., Parkin, T.B., Karlen, D.L., Novak, J.M., Turco, R.F., and Konopka, A.E., 1994. Field-scale variability of soil properties in central Iowa soils. *Soil sci.Soc. Am.* 58: 1501-1511.
- Dash, J.P., Sarangi, A., and Singh, D.K., 2010. Assessment of groundwater pollution in West