



مرکز تحقیقات آموزش کشاورزی و منابع طبیعی اصفهان

پژوهش‌های آب‌بخیزداری

شاپا: ۲۰۳۸-۲۹۸۱



مادان تحقیقات، آموزش و ترویج کشاورزی

برآورد احتمال رخداد آلودگی نیترات آب زیرزمینی با استفاده از الگوریتم تقویت شیب شدید

مه‌دی هاشمی*^۱، علی دسترنج^۲

- ۱- استادیار بخش تحقیقات حفاظت خاک و آب‌بخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی اصفهان، سازمان تحقیقات، آموزش و ترویج کشاورزی، اصفهان، ایران
- ۲- استادیار بخش تحقیقات حفاظت خاک و آب‌بخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی خراسان رضوی، سازمان تحقیقات، آموزش و ترویج کشاورزی، مشهد، ایران

چکیده مبسوط

مقدمه و هدف

آب‌های زیرزمینی به‌عنوان یک منبع حیاتی آب شیرین، نقش اساسی در تأمین نیازهای شرب، کشاورزی و صنعتی در بسیاری از مناطق خشک و نیمه‌خشک جهان دارند. با این حال، تشدید فعالیت‌های انسانی و صنعتی، منجر به افزایش آلودگی این منابع ارزشمند شده است. در این راستا، آلودگی ناشی از نیترات، به‌دلیل حلالیت و تحرک زیاد در آب، به‌عنوان یکی از مهم‌ترین تهدیدها برای سلامت انسان و بوم‌سازگان‌های آبی شناخته‌شده است. مصرف آب آلوده به نیترات می‌تواند منجر به بروز بیماری‌های گوناگونی از جمله متهموگلوبینمی (سندرم نوزاد آبی) در نوزادان و حتی برخی از سرطان‌ها در بزرگسالان شود. افزون بر این، ورود نیترات به آب‌های سطحی می‌تواند به پدیده اوتروفیکاسیون و ازبین‌رفتن زیست‌بوم‌های آبی منجر شود. با توجه به اهمیت موضوع و ضرورت حفاظت از منابع آب زیرزمینی، این پژوهش با هدف توسعه یک چارچوب یکپارچه و جامع برای برآورد احتمال رخداد آلودگی آب‌های زیرزمینی، به‌ویژه با تمرکز بر آلاینده نیترات، در منطقه دشت لنجان در استان اصفهان ایران انجام شد. با استفاده از این چارچوب و بهره‌گیری از رویکردهای نوین مدل‌سازی و تحلیل فضایی، مناطق مستعد به آلودگی شناسایی شد که به ارائه راهکارهای مدیریتی مؤثر برای کاهش خطرات ناشی از آلودگی آب‌های زیرزمینی کمک خواهد کرد. از نتایج این پژوهش می‌توان به‌عنوان الگویی برای برنامه‌ریزی‌های آتی در زمینه مدیریت پایدار منابع آب و حفاظت از سلامت جامعه استفاده کرد.

نوع مقاله: پژوهشی

*مسئول مکاتبات: m_hashemi@areeo.ac.ir

استناد: هاشمی، م.، دسترنج، ع. ۱۴۰۵. برآورد احتمال رخداد آلودگی نیترات آب زیرزمینی با استفاده از الگوریتم تقویت شیب شدید. پژوهش‌های آب‌بخیزداری. ۳۹(۱): ۶۳-۴۵.

شناسه دیجیتال: 10.22092/wmrj.2025.370560.1631

تاریخ دریافت: ۱۴۰۴/۰۶/۰۸، تاریخ بازنگری: ۱۴۰۴/۰۶/۲۱، تاریخ پذیرش: ۱۴۰۴/۰۶/۳۱، تاریخ انتشار: ۱۴۰۵/۰۱/۰۱
پژوهش‌های آب‌بخیزداری، سال ۱۴۰۵، دوره ۳۹، شماره ۱، شماره پیاپی ۱۵۰، بهار ۱۴۰۵، صفحه‌های ۴۵ تا ۶۳

© نویسندگان

ناشر: مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس



مواد و روش‌ها

در این پژوهش، داده‌های مربوط به غلظت نیترات منابع آب زیرزمینی به‌طور دقیق بررسی شد. از این‌رو، اطلاعات حیاتی از ۱۰۲ حلقه چاه در دشت لنجان استان اصفهان جمع‌آوری شد. هر یک از این چاه‌ها نماینده‌ای از وضعیت نیترات در سفره‌های آب زیرزمینی منطقه مطالعه‌شده بود. برای تحلیل این حجم از داده‌ها و استخراج الگوهای پنهان در آن‌ها، از مدل تقویت شیب شدید استفاده شد. این مدل به‌دلیل توانایی زیاد در شناسایی روابط پیچیده و غیرخطی میان متغیرها و نیز عملکرد و دقت قابل قبول در پیش‌بینی، انتخاب شد. افزون بر داده‌های غلظت نیترات، ده عامل مهم محیطی و انسانی که به‌طور بالقوه بر آلودگی نیترات در آب‌های زیرزمینی اثرگذار بودند، شناسایی و در مدل تحلیلی استفاده شدند. این عوامل شامل شیب زمین، بلندی، تراکم زهکشی، شاخص رطوبت پستی‌بلندی، رده خاک، فاصله از آبراهه، نوع سنگ‌شناسی منطقه و کاربری زمین بودند. با گنجاندن این هشت عامل در مدل تقویت شیب شدید، امکان شناسایی مهم‌ترین عامل‌های مؤثر بر آلودگی نیترات و همچنین برآورد مکانی اندازه احتمال آلودگی نیترات در آب‌های زیرزمینی فراهم شد.

نتایج و بحث

نتایج این پژوهش به‌خوبی بیانگر اثربخشی و کارایی تقویت شیب شدید در برآورد آلودگی نیترات در آب‌های زیرزمینی بود. دقت کلی این مدل ۰/۸۶ به‌دست آمد که با بهره‌گیری از آن وضعیت آلودگی منطقه مطالعه‌شده به‌خوبی تفکیک شد. افزون بر این، دیگر معیارهای ارزیابی عملکرد مدل نیز بیانگر دقت زیاد آن در شناسایی صحیح مناطق آلوده و غیرآلوده بود؛ به‌طوری که اندازه سطح زیر منحنی ROC برابر با ۰/۸۵ بود. همچنین، اندازه بازیابی مدل ۰/۸۰ به‌دست آمد که بیانگر آن بود ۸۰٪ از کل مناطق آلوده واقعی با استفاده از این مدل به‌درستی شناسایی شده است. سرانجام، آماره F1-score که یک معیار ترکیبی از دقت و بازیابی بود، ۰/۸۳ به‌دست آمد که بیانگر تعادل خوب میان این دو معیار و عملکرد کلی قابل اعتماد مدل بود. تحلیل حساسیت مدل نشان داد که اثر برخی از متغیرهای ورودی بر روند برآورد مکانی آلودگی نیترات آب زیرزمینی، زیاد بود. در میان ده عامل محیطی و انسانی بررسی‌شده، متغیرهای بارش (۲۱٪) و تغییرات بلندی (۱۸٪) به‌عنوان اثرگذارترین و مهم‌ترین متغیرها در تعیین الگوی مکانی آلودگی نیترات شناخته شدند. این یافته‌ها بیانگر اهمیت ویژگی‌های طبیعی و زمین‌ریخت‌شناختی منطقه در مهار انتشار و تجمع نیترات در آب‌های زیرزمینی بود که می‌توان از آنها به‌عنوان الگوی مفیدی برای پژوهش‌های آتی و تدوین راهبردهای مدیریتی بهره برد.

نتیجه‌گیری و پیشنهادها

یکی از دستاوردهای مهم این پژوهش، تولید نقشه‌های خطر بود که در آن به‌خوبی مناطق با خطر زیاد آلودگی نیترات در بخش مرکزی دشت مطالعه‌شده، شناسایی شد. پیشنهاد می‌شود مدیران منابع آب و برنامه‌ریزان شهری و روستایی از این نقشه‌ها به‌عنوان ابزاری ارزشمند برای انجام اقدامات پیشگیرانه در مناطق حساس، بهره ببرند. شایان ذکر است، نقش فعالیت‌های انسانی در افزایش خطر آلودگی نیترات با همپوشانی معنادار مناطق پرخطر با کاربری زمین‌های کشاورزی، قاطعانه تأیید شد. بر اساس این یافته پیشنهاد می‌شود برای حفاظت از منابع آب زیرزمینی و مدیریت پایدار فعالیت‌های کشاورزی از کودهای نیتروژنی به‌شکل بهینه استفاده شود.

واژگان کلیدی

آب زیرزمینی، احتمال رخداد، دشت لنجان، نیترات، XGBoost

مقدمه

از دهه ۱۹۶۰ تاکنون، افزایش چشمگیر مصرف کودهای شیمیایی (یواس دی ای ۲۰۲۳) در رشد تولید جهانی غذا، نقش مهمی داشته است. با این حال، پیامدهای زیست‌محیطی مصرف این کودها نیز قابل توجه است. در این راستا، افزایش نشت نیتروژن به منابع آب زیرزمینی، موجب شده نیترات (NO_3) به یکی از آلاینده‌های مهم سفره‌های آب زیرزمینی تبدیل شود. مهم‌ترین نگرانی بهداشتی برای نیترات، بیماری متهموگلوبینمی (نوزاد آبی) است (ورد و همکاران ۲۰۱۸)، که در نوزادان با غلظت‌های بیش از ۵۰ میلی‌گرم در لیتر بروز می‌کند و در غلظت‌های بسیار زیاد نیز بر بزرگسالان اثرگذار است (سازمان جهانی بهداشت ۲۰۲۲). پژوهش‌های کنونی، مصرف بلندمدت نیترات را با برخی از انواع سرطان مرتبط دانسته‌اند (پیکتی و همکاران ۲۰۲۲؛ ورد و همکاران ۲۰۱۸). افزون بر خطرات مستقیم، نیترات شاخص مهمی برای ارزیابی کیفیت کلی آب است. زیرا، وجود آن می‌تواند نشان‌دهنده وجود دیگر آلاینده‌های کشاورزی مانند آفت‌کش‌ها نیز باشد. به این دلیل، نیترات یکی از سه شاخص مهم در کنار شوری و pH برای سنجش کیفیت آب زیرزمینی در اهداف توسعه پایدار سازمان ملل به‌شمار می‌آید (یوان ای پی ۲۰۲۳). با توجه به گستردگی آلودگی نیترات و خطرات مرتبط با آن، اهمیت شناسایی مناطقی که غلظت نیترات در آنها زیاد است در مدیریت و کاهش این مخاطرات زیاد است. شبکه‌های پایش حتی با تراکم نسبی، معمولاً شامل چاه‌ها و ایستگاه‌هایی هستند که فاصله زیادی از یکدیگر دارند. از این رو، شناخت توزیع مکانی نیترات در آب زیرزمینی با کمبود داده و خلأ مواجه است (کواتی و همکاران ۲۰۲۵). در دهه‌های پیشین، روش‌های مختلفی از جمله مدل‌های مکانیکی با نیاز داده‌ای گسترده برای پر کردن این خلأها به کار گرفته شده‌اند (الماسری و کالوارچی ۲۰۰۷؛ کرید و بند ۱۹۹۸؛ آلر ۱۹۸۷) و پهنه‌بندی آسیب‌پذیری با وزن‌دهی عامل‌های

اثرگذار نیز با هدف برآورد احتمال آلودگی و تعیین مناطق حساس، استفاده شده‌اند (آلر و همکاران ۱۹۸۷؛ آراوزو ۲۰۱۷). در سال‌های گذشته، به روش‌های مختلف یادگیری ماشین به دلیل قابلیت کاربرد در مقیاس‌های گسترده و توانایی در تحلیل روابط پیچیده مرتبط با سنج‌های کیفیت آب زیرزمینی، توجه فزاینده‌ای شده است. این روش‌ها به‌ویژه در پیش‌بینی غلظت نیترات در آب‌های زیرزمینی کاربرد مؤثری دارند. در این راستا، می‌توان به نتایج بررسی مدل وایازی لجستیک برای ۱۷ سفره آب اصلی ایالات متحده (گوردک و چی ۲۰۱۲)، سه مدل یادگیری ماشین درختان وایازی تقویت‌شده، تجزیه تفکیک‌کننده چند متغیره و ماشین بردار پشتیبان در دشت لنجانان استان اصفهان (ساجدی‌حسینی و همکاران ۲۰۱۸)، مدل‌های K-نزدیکترین همسایه، ماشین بردار پشتیبان و جنگل تصادفی در دشت اندیمشک-دزفول استان خوزستان (رحمتی و همکاران ۲۰۱۹)، مدل‌های جنگل تصادفی، ماشین بردار پشتیبان، شبکه عصبی و ماشین تقویت‌شده با شیب در ایالت کارولینای شمالی (مسیه و همکاران ۲۰۱۹)، مدل جنگل تصادفی در آلمان (نول و همکاران ۲۰۲۰)، الگوریتم تقویت شیب شدید در سراسر ایالات متحده (رنسوم و همکاران ۲۰۲۲)، مدل‌های جنگل تصادفی، درخت وایازی تقویت‌شده و وایازی لجستیک در هند (سرکر و همکاران ۲۰۲۲)، مدل‌های شبکه عصبی عمیق، تقویت شیب شدید و وایازی خطی چندگانه در شمال ایران (غلامی و بویج ۲۰۲۲)، مدل‌های جنگل تصادفی معمولی و وزن‌دار، مدل جمعی تعمیم‌یافته با استفاده از LOESS و K-نزدیکترین همسایه در دشت لنجانان اصفهان (حسینی و همکاران ۲۰۲۳) اشاره کرد که همگی این پژوهش‌ها بیانگر قابلیت زیاد یادگیری ماشین در مدل‌سازی و تحلیل داده‌های پیچیده محیطی

نیترا ت ارائه داد و موجب ارتقای درک این پدیده در مقیاس آبخوان شد. این دانش قابلیت تعمیم و کاربرد در مناطق فراتر از محدوده مطالعه شده را نیز دارد.

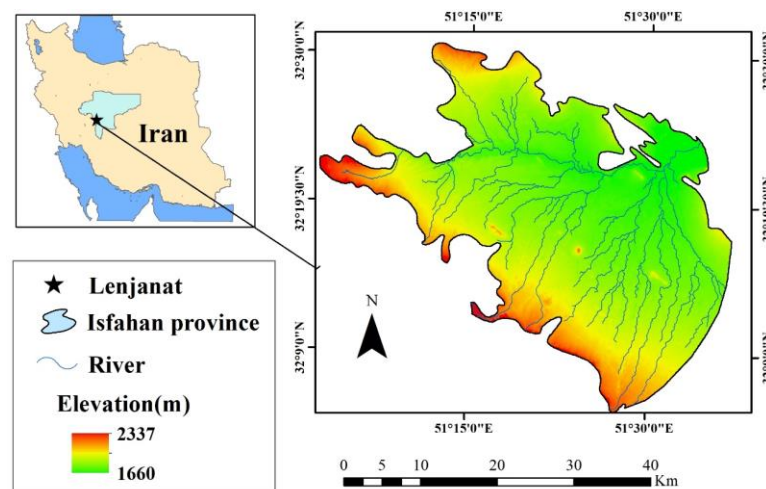
مواد و روش‌ها

معرفی منطقه مطالعه شده

دشت لنجان در استان اصفهان، در مرکز ایران است که با مساحت تقریبی ۱۱۸۰ کیلومتر مربع، به عنوان منطقه پژوهشی انتخاب شد (شکل ۱). این دشت به وسیله کوه‌های آهکی احاطه شده و بلندی آن میان ۱۶۶۰ تا ۲۳۳۷ متر از سطح دریا متغیر است. بلندی‌های بخش‌های غربی و شمالی این دشت بیشتر است. اقلیم دشت لنجان خشک و سرد است و میانگین سالانه بارندگی در دوره ۱۹۷۱ تا ۲۰۱۴ حدود ۱۶۰ میلی‌متر و بیشتر در فصل زمستان بوده است. بیشترین اندازه بارش در بلندی‌های غرب و شمال دشت متمرکز است و از غرب به شرق کاهش می‌یابد، در حالی که دمای هوا در همین جهت افزایش می‌یابد. میانگین ماهانه دما در گرم‌ترین ماه (مرداد) ۲۷ درجه سانتی‌گراد و در سردترین ماه (دی) ۴ درجه سانتی‌گراد ثبت شده است (ناصری و همکاران ۲۰۱۲؛ رحیمی و همکاران ۲۰۲۴).

در زمینه آلودگی نیترا ت است. از این رو، بهره‌گیری از توانایی زیاد این مدل‌ها در مناطقی که ناشناخته است و یا احتمال آلودگی دارد حائز اهمیت است.

در پژوهش‌های پرشماری کاربرد الگوریتم تقویت شیب شدید برای بررسی آلودگی نیترا ت در آب زیرزمینی بررسی شده است (رنسوم و همکاران ۲۰۲۲، غلامی و بویج ۲۰۲۲، لیانگ و همکاران ۲۰۲۴، آلام و همکاران ۲۰۲۵). نتایج این پژوهش‌ها نیز بیانگر کاربرد موفقیت آمیز این مدل در تحلیل و پیش‌بینی غلظت نیترا ت است. در این راستا، در این پژوهش نیز از الگوریتم تقویت شیب شدید که یکی از روش‌های پیشرفته در یادگیری ماشین به شمار می‌آید، برای تولید نقشه‌های حساسیت نیترا ت بهره گرفته شد از بیشترین ظرفیت شبکه‌های پایش موجود استفاده شود. از این نقشه‌ها می‌توان در شناسایی مناطق نیازمند اقدامات پیشگیرانه نشت نیترا ت، بهره برد. افزون بر این، با تحلیل مدل، روابط میان غلظت نیترا ت و عامل‌های زمین‌شناسی، پستی بلندی، خاک، اقلیم و کاربری زمین بررسی شد تا عامل‌های مؤثر در آلودگی شناسایی شوند. برخلاف نتایج پژوهش‌های پیشین که صرفاً به پیش‌بینی اشاره داشته‌اند، این رویکرد، با بهره‌گیری از پیشرفت‌های کنونی در تفسیر مدل‌های یادگیری ماشین، دیدگاهی عمیق‌تر نسبت به فرآیندهای مؤثر در آلودگی



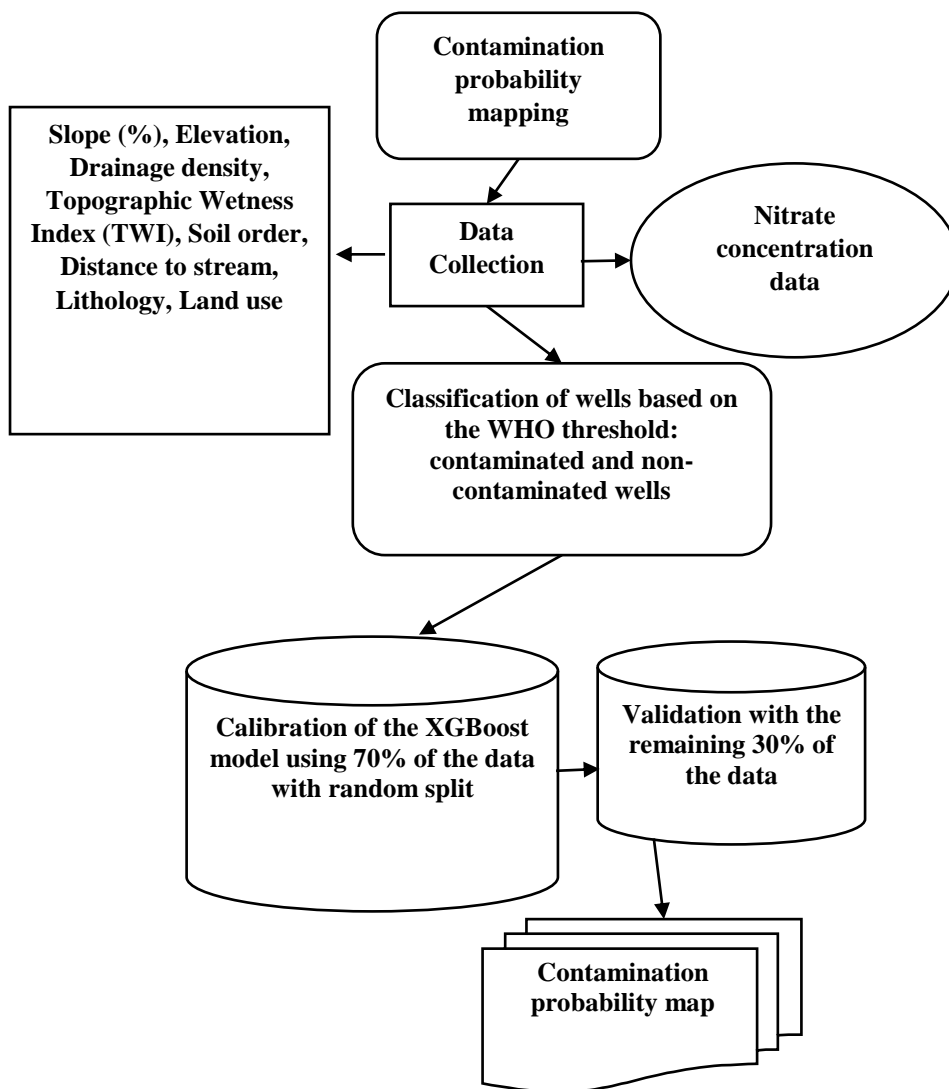
شکل ۱- موقعیت جغرافیایی منطقه مطالعه شده.

Figure 1- Geographical location of the study area.

رویکرد پژوهش

شد و ۷۰٪ داده‌ها به‌شکل تصادفی برای آموزش مدل و ۳۰٪ باقی‌مانده برای اعتبارسنجی آن به‌کار گرفته شد. خروجی مدل به‌شکل نقشه احتمال آلودگی ارائه شد که از آن می‌توان در شناسایی مناطق نیازمند اقدامات پیشگیرانه بهره برد. همچنین، با تحلیل مدل، روابط میان متغیرهای وابسته و مستقل بررسی شد تا عوامل مؤثر در آلودگی شناسایی شوند و درک عمیق‌تری از فرآیندهای مؤثر در مقیاس آبخوان فراهم آید.

شرح گام به گام روش ارزیابی احتمال آلودگی آب‌های زیرزمینی در دشت لنجان در این پژوهش در شکل ۲ نشان داده شده است. ابتدا داده‌های اثرگذار بر آلودگی نیترات (شامل عامل‌های زمین‌شناسی، پستی‌بلندی، خاک، اقلیم و کاربری زمین) به‌عنوان متغیرهای پیش‌بینی‌کننده، گردآوری شدند. سپس، داده‌های نیترات چاه‌های نمونه‌برداری شده به‌عنوان متغیر وابسته در نظر گرفته شدند. برای مدل‌سازی، از الگوریتم تقویت شیب شدید (XGBoost) استفاده



شکل ۲- روند گام به گام پژوهش.

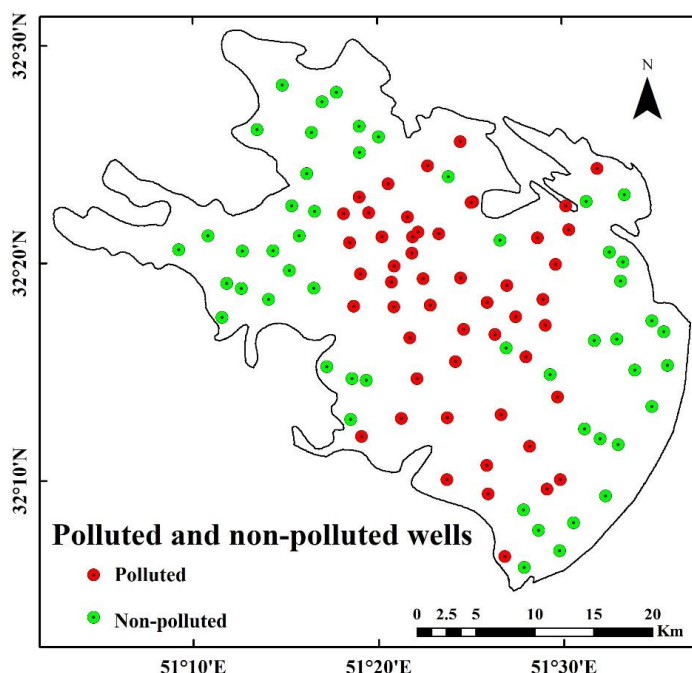
Figure 2- Step-by-step process of the study.

داده‌های استفاده‌شده

نقشه پراکنش آلودگی نیترات آب زیرزمینی

در این پژوهش، به منظور مدل‌سازی و ارزیابی خطر نیترات در آب‌های زیرزمینی، داده‌های مربوط به غلظت نیترات از ۱۰۲ حلقه چاه با پراکنش مکانی مناسب در دشت لنجان، از شرکت مدیریت منابع آب ایران گردآوری شد. این داده‌ها مربوط به دوره فروردین ۱۳۹۶ بودند. پس از گردآوری داده‌های غلظت نیترات، بر اساس دستورالعمل‌های سازمان بهداشت جهانی (WHO ۲۰۱۱)، آستانه‌ای معادل ۵۰ میلی‌گرم بر لیتر به عنوان حد مرزی برای تفکیک چاه‌های آلوده و غیرآلوده در نظر گرفته شد. بر این اساس، چاه‌هایی با غلظت نیترات بیش

از ۵۰ میلی‌گرم بر لیتر به عنوان چاه‌های آلوده و چاه‌هایی با غلظت کمتر از این اندازه به عنوان چاه‌های غیرآلوده طبقه‌بندی شدند. سپس، چاه‌های آلوده با کد ۱ و چاه‌های غیرآلوده با کد ۰ مشخص شدند. این طبقه‌بندی به عنوان متغیر هدف در مدل‌سازی باینری استفاده شد، به طوری که مدل با تفکیک دو طبقه آلوده و غیرآلوده آموزش داده شد تا احتمال آلودگی نیترات در دیگر نقاط نیز پیش‌بینی شود. نقشه پراکنش چاه‌های آلوده و غیرآلوده از دیدگاه نیترات در دشت لنجان در شکل ۳ نشان داده شده است.



شکل ۳- نقشه پراکنش چاه‌های آلوده و غیرآلوده از دیدگاه نیترات در دشت لنجان.

Figure 3- Distribution map of polluted and non-polluted wells in terms of nitrate in the Lanjanat Plain.

ویژگی‌های اثرگذار بر آلودگی آب زیرزمینی

نقش عامل‌های محیطی و انسانی در اندازه آلودگی آب‌های زیرزمینی مهم است. برای ایجاد نقشه احتمال رخداد آلودگی آب‌های زیرزمینی، در مجموع ده عامل موثر بر آلودگی آب زیرزمینی که در شکل ۴ نشان داده شده است، استفاده شد. این

عامل‌ها شامل درصد شیب، بلندی، تراکم زهکشی، شاخص رطوبت پستی‌بلندی، بارش، رده خاک، فاصله از آبراهه، سنگ‌شناسی، پوشش گیاهی و کاربری زمین است (گاد و همکاران ۲۰۲۴؛ ساجدی حسینی و همکاران ۲۰۱۸). نقش عامل بلندی (شکل a) و شیب زمین (شکل b)

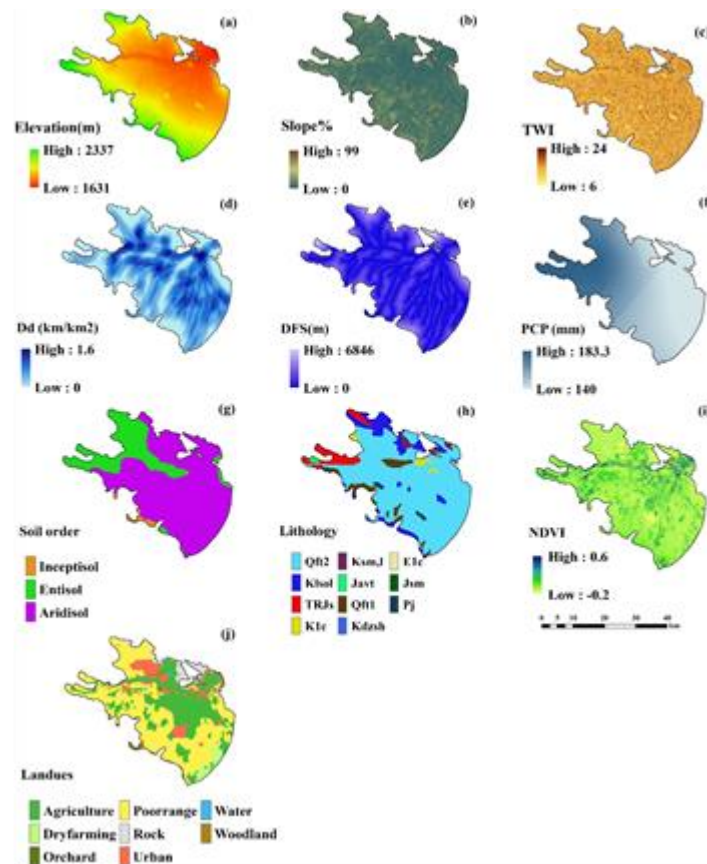
بارش (شکل f۴) با افزایش نفوذ آب و شست‌وشوی نیترات، خطر آلودگی آب زیرزمینی را افزایش می‌دهد (ونگ و همکاران ۲۰۲۲). در این پژوهش، از آمار بلندمدت ۳۰ ساله (۹۹-۱۳۷۰) ایستگاه‌های باران‌سنجی شرکت مدیریت منابع آب ایران و ایستگاه‌های سینوپتیک سازمان هواشناسی کشور استفاده شد و میانگین بارش با استفاده از روش درون‌یابی معکوس وزنی فاصله‌ای محاسبه شد. رده‌های خاک (شکل g۴) بسته به بافت و نفوذپذیری بر کیفیت آب زیرزمینی به‌طور قابل توجهی اثرگذارند.

رده‌های خاک منطقه مطالعه‌شده از موسسه تحقیقات خاک و آب دریافت شد. این رده‌ها شامل اینسپتیسول، انتیسول و اریدیسول بود. خاک‌های سبک و شنی با نفوذپذیری زیاد مانند انتیسول و اینسپتیسول، آلودگی بیشتری را انتقال می‌دهند، در حالی که خاک‌های سنگین و رسی با خاصیت نفوذپذیری کمتر، مانع مناسبی برای گسترش آلاینده‌ها هستند. از سوی دیگر، شرایط سنگ‌شناسی (شکل h۴) منطقه نیز مهم بود. جنس سنگ‌ها و ویژگی‌های زمین‌شناسی نقش مهمی در مهار فرسایش‌پذیری و انتقال آلودگی به سفره‌های آب زیرزمینی دارند (شوکلا و همکاران ۲۰۲۰). در این پژوهش، از نقشه سازمان زمین‌شناسی ایران با مقیاس ۱:۱۰۰۰۰۰ استفاده شد. بیشتر مساحت دشت به‌وسیلهٔ مخروط‌افکنه‌های دامنه‌ای با بلندی کم و رسوبات تراس‌های دره‌ای پوشیده‌شده است. تمام متغیرهای ورودی به‌شکل نقشه‌های رستری و با اندازه پیکسل ۳۰ در ۳۰ متر در محیط ArcGIS 10.5 آماده شد.

که از عامل‌های محیطی و طبیعی هستند در اندازه آلودگی آب‌های زیرزمینی مهم است. به‌طوری که در مناطق بلند و با شیب تند، رواناب سطحی سریع‌تر تخلیه شده و فرصت نفوذ آلاینده‌ها به آبخوان کاهش می‌یابد، در حالی که در زمین‌هایی با بلندی کم و شیب ملایم، اندازه نفوذ بیشتر و احتمال آلودگی سفره‌های زیرزمینی افزایش می‌یابد.

شاخص رطوبت پستی‌بلندی (شکل c۴) نیز نشان‌دهنده اندازهٔ تجمع رطوبت در سطح زمین است، به‌طوری که اندازه‌های زیاد بیانگر شرایط مساعد برای نفوذ و افزایش احتمال آلودگی است (حسینی و همکاران ۲۰۲۳). این شاخص با استفاده از مدل ارتفاعی رقومی با اندازه پیکسل در محیط SAGA GIS محاسبه شد. تراکم زهکشی و فاصله از رود (شکل d۴، e۴) به‌ترتیب با استفاده از ابزارهای فاصله اقلیدسی و تراکم خط در ArcGIS محاسبه شدند. تراکم زیاد شبکه زهکشی باعث انتقال سریع رواناب و کاهش فرصت نفوذ می‌شود، در حالی که تراکم کم منجر به باقی‌ماندن طولانی‌تر آب در سطح و افزایش خطر آلودگی آب‌های زیرزمینی می‌شود (ساجدی حسینی و همکاران ۲۰۱۸).

فاصله از آبراهه نیز یکی از عامل‌های مؤثر بر آلودگی آب‌های زیرزمینی است. زمین‌های مجاور با آبراهه‌ها به‌دلیل شست‌وشوی سریع آلاینده‌ها و انتقال مستقیم آن‌ها به داخل آبخوان، ظرفیت آلودگی بیشتری دارند. از سوی دیگر، با افزایش فاصله از آبراهه، زمان کافی برای جذب، تجزیه و کاهش غلظت آلاینده‌ها در خاک فراهم می‌شود و به این دلیل شدت آلودگی آب زیرزمینی در این مناطق کمتر است (حسینی و همکاران ۲۰۲۳).



شکل ۴- عوامل مؤثر بر آلودگی آب‌های زیرزمینی در دشت لنجان: (الف) بلندی، (ب) شیب، (ج) شاخص رطوبت پستی‌بلندی، (د) تراکم زهکش، (ه) فاصله از آبراهه، (و) بارش، (ز) رده خاک، (ح) سنگ‌شناسی، (ط) شاخص پوشش گیاهی بهنجارشده، (ی) کاربری زمین.

Figure 4- Factors affecting groundwater pollution in the Lenjanat Plain: (a) Elevation, (b) Slope, (c) Topographic Wetness Index (TWI), (d) Drainage density (Dd), (e) Distance from streams (DFS), (f) Precipitation (PCP), (g) Soil order, (h) Lithology, (i) Normalized Difference Vegetation Index (NDVI), (j) Land use.

مدیریت منابع آب ایران دریافت شد. این نقشه شامل کاربری کشاورزی، مرتع، شهری، صخره‌ای، دیم‌کاری، باغ، بیشه و تالاب بود که زمین‌های کشاورزی به دلیل مصرف کود و سموم شیمیایی، منابع اصلی نیترات و فسفات هستند. از سوی دیگر، مناطق صنعتی آلاینده‌های شیمیایی و فلزات سنگین تولید می‌کنند و مناطق شهری نیز به دلیل فاضلاب‌های انسانی می‌توانند موجب آلودگی میکروبی و شیمیایی شوند، در حالی که

پوشش گیاهی با جذب نیتروژن از خاک، نقش مؤثری در کاهش نشت نیترات دارد (اولا و همکاران ۲۰۱۴). در این پژوهش، شاخص پوشش گیاهی بهنجارشده (شکل ۱۴) برای ماه ژوئن سال ۲۰۱۷ (منطبق با سال داده‌های نیترات) با استفاده از سامانه گوگل ارث انجین و تصویرهای دریافتی از سنسور ۲، محاسبه شد. نوع کاربری زمین (شکل ۱۴) بر آلودگی تأثیر مستقیمی دارد. نقشه کاربری زمین منطقه مطالعه‌شده از شرکت

4 Normalized Difference Vegetation Index

(۲۰۲۴). تابع زیان XGBoost معمولاً به شکل مجموع اختلاف میان پیش‌بینی‌های مدل و اندازه‌های واقعی است و با استفاده از رابطه‌های ۱ و ۲ محاسبه می‌شود.

$$(y_i - \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (1)$$

$$\Omega(f) = Y \times T + \frac{1}{2} \lambda \times \sum_{j=1}^T w_j^2 \quad (2)$$

T: تعداد گره‌ها، (\hat{y}_i) : اندازه‌های پیش‌بینی شده، (y_i) : اندازه‌های واقعی، (w_i) : وزن هر گره است.

وزن هر گره همراه با سنج‌های منظم‌سازی (λ, γ) تعیین‌کننده پیچیدگی درخت و اندازه جریمه برای وزن برگ‌ها است. این الگوریتم درخت را به شکل مرحله‌ای و تدریجی می‌سازد. در هر مرحله، بهترین ویژگی‌ها و نقاط تقسیم انتخاب می‌شوند تا تابع زیان در همان مرحله به حداقل برسد. برای پیدا کردن بهترین تقسیم، الگوریتم تمام نقاط ممکن را بررسی می‌کند (لوو و همکاران ۲۰۲۴). در هر مرحله از آموزش، از نزول شیب برای به‌روزرسانی سنج‌های مدل و کاهش تابع زیان بهره‌گرفته می‌شود. به‌طور خاص، بسط دوم تیلور برای تقریب تابع زیان و محاسبه اندازه‌های به‌روزرسانی سنج‌ها به‌کار می‌رود. افزون بر این، XGBoost از چند راهبرد بهینه‌سازی مانند نمونه‌گیری از ستون‌ها و ردیف‌ها و ارزیابی اهمیت ویژگی‌ها استفاده می‌کند تا عملکرد، پایداری و قابلیت تفسیر مدل بهبود یابد (دالیوال و همکاران ۲۰۲۲). این ویژگی‌ها باعث می‌شوند که XGBoost در مواجهه با داده‌های بزرگ، پراکنده و با ابعاد گسترده، عملکرد مؤثر و قابل اطمینانی ارائه دهد.

مدل‌سازی آلودگی نیترات آب زیرزمینی

در این پژوهش مدل‌سازی از نوع طبقه‌بندی بود. عامل‌های محیطی مؤثر بر آلودگی نیترات به‌عنوان متغیرهای مستقل و وضعیت آلودگی چاه‌ها (آلوده یا غیرآلوده) به‌عنوان متغیر هدف در نظر گرفته شدند. پیش از مدل‌سازی، به‌منظور اطمینان از نبودن هم‌خطی شدید میان متغیرهای مستقل، شاخص تورم پراکنش (VIF) برای هر یک از متغیرها محاسبه شد.

کاربری‌های طبیعی مانند مراتع و جنگل‌ها کمترین نقش را در آلودگی دارند.

الگوریتم تقویت شیب شدید (XGBoost)

این الگوریتم یکی از پیشرفته‌ترین روش‌های یادگیری ماشین مبتنی بر درخت تصمیم تقویتی (GBDT) است که بر اساس اصول یادگیری تقویتی تجمعی توسعه‌یافته است (چن و همکاران ۲۰۲۳). با ترکیب هم‌زمان تقویت شیب و سازوکارهای منظم‌سازی، امکان ساخت مدل‌های پیش‌بینی دقیق و قابل اعتماد به‌وسیله این الگوریتم فراهم می‌آید. در فرآیند تقویت یا بوستینگ، چندین مدل ضعیف (معمولاً درخت‌های تصمیم کم‌عمق) به شکل مرحله‌ای آموزش داده می‌شوند، به طوری که با هر مدل جدید می‌توان خطاهای پیش‌بینی مدل‌های قبلی را اصلاح کرد (واینس و همکاران ۲۰۲۵). در این پژوهش، بخش منظم‌سازی الگوریتم با مهار نرخ یادگیری و تعیین وزن اهمیت ویژگی‌ها پیاده‌سازی شد. این اقدام ضمن جلوگیری از بیش‌برازش، توانایی تعمیم مدل را ارتقا می‌دهد (ژانگ و همکاران ۲۰۲۲) همچنین، این الگوریتم با استفاده از توسعه تیلور مرتبه دوم، امکان تقریب‌زدن و بهینه‌سازی توابع زیان پیچیده و غیرخطی را فراهم می‌آورد، به گونه‌ای که محاسبات، بهینه و کارآمد انجام شوند (ایمانی و همکاران ۲۰۲۵). ساختار الگوریتم بر پایه درخت‌های طبقه‌بندی و وایازی است و این طراحی موجب شده عملکرد XGBoost در حل مسائل طبقه‌بندی و وایازی بسیار مؤثر باشد (دینگ و همکاران ۲۰۲۰). در هر مرحله از آموزش، الگوریتم یک درخت جدید ایجاد می‌کند تا سنج‌های مجموع تابع زیان و منظم‌سازی به حداقل برسد. سنج منظم‌سازی با تنظیم وزن‌های مربوطه، تعادل میان پیچیدگی مدل و توانایی تعمیم‌دهی را حفظ می‌کند، در حالی که دقت کلی مدل همچنان پایدار باقی می‌ماند (لوو و همکاران

1 Extreme Gradient Boosting

2 Gradient Boosting Decision Tree

3 Gradient Boosting

4 Variance Inflation Factor

است، شاخص دقت که بیانگر نسبت نمونه‌های پیش‌بینی شده درست به کل نمونه‌های پیش‌بینی شده مثبت است (رابطه ۳)، شاخص بازخوانی که بیانگر نسبت نمونه‌های واقعی آلوده که درست شناسایی شده‌اند (رابطه ۴) و آماره F1-score که بیانگر میانگین هارمونیک دقت و بازخوانی است (رابطه ۵)، ارزیابی شد. با این شاخص‌ها می‌توان امکان بررسی جامع صحت پیش‌بینی و شناسایی نمونه‌های آلوده را فراهم آورد. محدوده‌های قابل قبول شاخص‌ها، شامل اندازه‌های بیش از ۰/۷ برای AUC، دقت، بازخوانی و آماره F1-score بود که نشان‌دهنده عملکرد مطلوب مدل است.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

نتایج و بحث

نتایج بررسی هم‌خطی میان متغیرها

در این پژوهش، نتایج بررسی هم‌خطی نشان داد تمام اندازه‌های VIF به‌طور قابل‌توجهی کمتر از آستانه بحرانی ۱۰ بود (موسوی و همکاران ۲۰۲۱). بیشترین اندازه مربوط به متغیر تراکم زهکشی (۲/۶۹) بود و این اندازه همچنان در محدوده قابل قبول بود (جدول ۱). این نتایج مؤید آن بود که میان متغیرهای مستقل هم‌خطی چندگانه وجود ندارد، از این‌رو، می‌توان از تمام متغیرها در مدل پایانی استفاده کرد.

ارزیابی نتایج مدل‌سازی

برای تعیین احتمال رخداد آلودگی آب‌های زیرزمینی، از مدل یادگیری ماشین تقویت شیب شدید استفاده شد. این مدل با تعداد ده متغیر ورودی و داده‌های نیترات و اسنجی و اعتبارسنجی شد. فرایند واسنجی تا دستیابی به بیشترین اندازه AUC انجام شد.

سپس، الگوریتم XGBoost در محیط نرم‌افزار R و با استفاده از کتابخانه caret اجرا شد (کوهن و همکاران ۲۰۲۰). برای توسعه مدل، از روش تقسیم تصادفی داده‌ها به نسبت ۷۰٪ برای آموزش و ۳۰٪ برای اعتبارسنجی استفاده شد. سنجه‌های مهم مدل شامل سنجه nrounds که مشخص‌کننده تعداد تکرارهای الگوریتم بوستینگ است (هر تکرار یک درخت جدید به مدل اضافه می‌کند)، سنجه ژرفای بیشینه درخت تصمیم‌گیری (max depth) که مهارکننده پیچیدگی مدل است. ژرفای زیاد می‌تواند الگوهای پیچیده‌تری را یاد بگیرد اما خطر بیش‌برازش را افزایش می‌دهد. سنجه نرخ کاهش (eta) ضریبی است که تأثیر هر درخت جدید را کاهش می‌دهد و باعث یادگیری تدریجی‌تر و پایدارتر می‌شود. سنجه gamma آستانه‌ای برای تقسیم‌گره‌ها است. اگر کاهش تابع زیان کمتر از این اندازه باشد، تقسیم انجام نمی‌شود. این موضوع پیچیدگی مدل را مهار می‌کند. سنجه colsample_bytree درصدی از ویژگی‌ها (ستون‌ها) را مشخص می‌کند که برای ساخت هر درخت انتخاب می‌شوند و این باعث کاهش هم‌وابستگی درخت‌ها و افزایش تعمیم‌پذیری می‌شود. سنجه min_child_weight حداقل مجموع وزن نمونه‌ها در یک گره را تعیین می‌کند تا آن گره قابل تقسیم باشد. اندازه بیشتر این سنجه از تقسیم‌های ضعیف جلوگیری می‌کند. سنجه subsample درصدی از داده‌های آموزشی را مشخص می‌کند که برای ساخت هر درخت استفاده می‌شود و نمونه‌گیری تصادفی از داده‌ها باعث افزایش مقاومت مدل در برابر بیش‌برازش می‌شود (کوهن و همکاران ۲۰۲۰). سنجه‌های مدل با استفاده از جست‌وجوی تصادفی و اعتبارسنجی متقابل^۱ بهینه‌سازی شدند.

ارزیابی عملکرد مدل

در این پژوهش، عملکرد مدل با استفاده از شاخص مساحت زیر منحنی مشخصه عملکرد گیرنده (ROC) که بیانگر توانایی تمایز میان نمونه‌های آلوده و غیرآلوده

3 Precision
4 Recall

4 Cross-validation
2 Receiver operating characteristic

جدول ۱- نتایج تحلیل هم‌خطی چندگانه میان متغیرهای مستقل با استفاده از عامل تورم پراکنش.

Table 1- Results of multicollinearity analysis of independent variables using the Variance Inflation Factor (VIF).

Factor	VIF
Drainage density	2.69
Distance from stream	2.50
Soil order	1.98
Precipitation	1.87
Elevation	1.63
Land use	1.37
Slope	1.33
Lithology	1.22
NDVI	1.16
Topographic Wetness Index	1.14

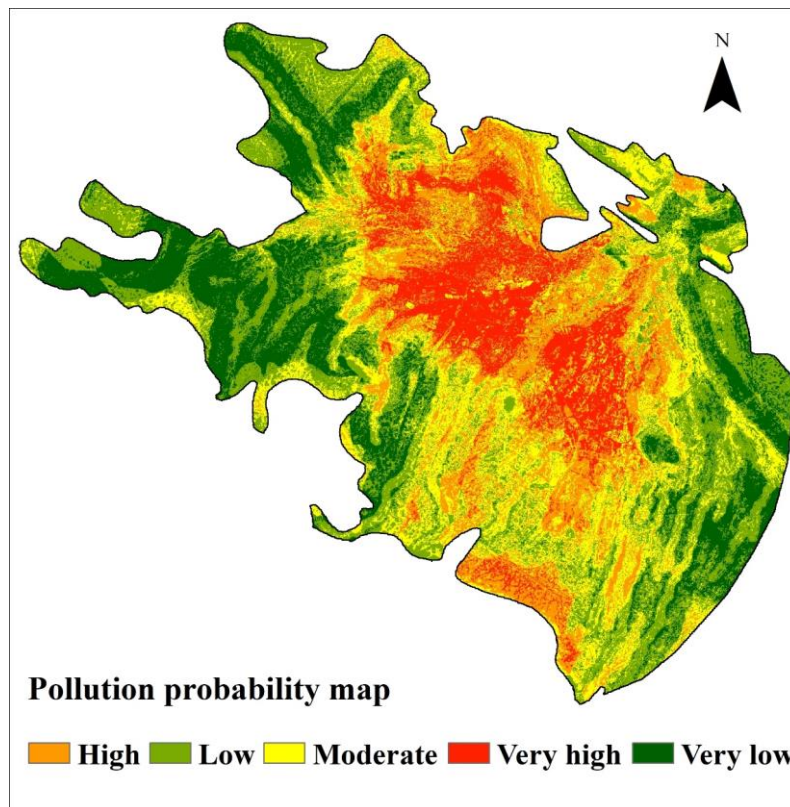
۶۳٪)، از دیگر مدل‌های وایزای بهتر بود. نتایج پژوهش آلام و همکاران (۲۰۲۵) نیز مؤید آن بود که عملکرد مدل XGBoost در منطقه بینچوان چین عالی بود. این نتایج بیانگر توانایی قابل توجه مدل تقویت شیب شدید در پیش‌بینی‌های کاربردی و تصمیم‌گیری‌های مبتنی بر داده است و می‌توان از آن به‌عنوان ابزار سومندی برای شناسایی نمونه‌های مهم در زمینه پژوهش بهره برد. در پژوهش‌های پرشماری، عملکرد مطلوب این مدل در مقایسه با دیگر مدل‌های یادگیری ماشین گزارش شده است (غلامی و بویج ۲۰۲۲؛ لیانگ و همکاران ۲۰۲۴؛ آلام و همکاران ۲۰۲۵). نتایج این پژوهش‌ها بیانگر توانایی مدل تقویت شیب شدید در تحلیل داده‌های پیچیده و غیرخطی محیط‌زیستی است. بر اساس نقشه احتمال آلودگی به‌دست آمده در این پژوهش مشخص شد که مناطق قابل توجهی از مساحت دشت لنجان‌ات خطر آلودگی آب زیرزمینی دارند (شکل ۵). بر این اساس، ۱۱/۶٪ از سطح منطقه در طبقه "احتمال آلودگی خیلی زیاد" و ۱۹/۶٪ در طبقه "احتمال آلودگی زیاد" بود. این ارقام بیانگر آن است که حدود یک سوم از دشت در معرض خطر زیاد یا بسیار زیاد است، که از دیدگاه مدیریت منابع آب زیرزمینی و سلامت عمومی، زنگ هشدار جدی تلقی می‌شود. با مقایسه نقشه آلودگی (شکل ۵) با موقعیت کاربری‌های زمین (شکل ۴)، واضح است که مناطق با خطر زیاد و بسیار زیاد عمدتاً با کاربری کشاورزی، هم‌پوشانی دارند.

پس از دستیابی به بیشترین عملکرد، نقشه احتمال رخداد آلودگی آب‌های زیرزمینی تولید و انتخاب شد (شکل ۵). عملکرد مدل در پیش‌بینی نقشه احتمال رخداد آلودگی در جدول ۲ ارائه شده است. عملکرد مدل پیشنهادی در پیش‌بینی آلودگی با دقت کلی ۸۶٪، نشان‌دهنده توانایی زیاد مدل در دسته‌بندی صحیح داده‌ها بود. اندازه سطح زیر منحنی ROC برابر با ۰/۸۵ و بازیابی مدل ۰/۸ بود. این موضوع بیانگر آن بود که ۸۰٪ از نمونه‌ها به درستی شناسایی شده‌اند. افزون بر این، آماره F1-score، برابر با ۰/۸۳ مؤید آن بود که مدل تعادل مناسبی میان دقت و حساسیت داشته و از نظر عملکرد کلی قابل اعتماد است. نتایج پژوهش‌های پیشین نیز مؤید توانمندی‌های مدل XGBoost هستند، هرچند اغلب بر کاربردهای وایزای آن تأکید داشته‌اند. در این راستا، غلامی و بویج (۲۰۲۲) غلظت نیترات در دشت مازندران را برآورد کردند و دریافتند عملکرد مدل XGBoost با ضریب تبیین ۸۶٪ در مقایسه با شبکه عصبی عمیق و وایزای چندگانه بهتر بود. رنسوم و همکاران (۲۰۲۲) گزارش کردند که عملکرد XGBoost (با ضریب تبیین ۵۶٪) از مدل آدابوست (با ضریب تبیین ۴۰٪) بهتر بود، هرچند در مقایسه با جنگل تصادفی (با ضریب تبیین ۶۲٪) و وایزای تقویت‌شده شیب (با ضریب تبیین ۶۳٪) ضعیف‌تر بود. لیانگ و همکاران (۲۰۲۴) نیز نشان دادند برای پیش‌بینی توزیع غلظت نیترات در آب‌های زیرزمینی چین عملکرد XGBoost (با ضریب تبیین

جدول ۲- ارزیابی کارایی مدل‌سازی آلودگی آب‌های زیرزمینی با مدل XGBoost.

Table 2- Evaluation of the efficiency of groundwater pollution modeling with the XGBoost Model.

AUC	Precision	Recall	F1-score
0.85	0.86	0.80	0.83



شکل ۵- نقشه احتمال آلودگی نیترات آب‌های زیرزمینی مبتنی بر مدل XGBoost.

Figure 5- Groundwater nitrate contamination probability map based on the XGBoost Model.

اهمیت متغیرهای ورودی

نتایج الگوریتم تقویت شیب شدید نشان داد بیشترین اهمیت در میان متغیرهای ورودی مربوط به متغیرهای بارش (۲۱٪) و بلندی (۱۸٪) بود (جدول ۳). این یافته‌ها هم از دیدگاه آماری معنادار بودند و هم از دیدگاه فرآیندهای زمین‌آب‌شیمیایی (هیدروژئوشیمیایی) قابل تبیین هستند. بارش به‌عنوان عامل اصلی تغذیه آب زیرزمینی، نقش دوگانه‌ای در آلودگی نیتراتی دارد. از یک‌سو افزایش بارندگی موجب نفوذ بیشتر آب به درون خاک و سفره‌های زیرزمینی می‌شود، که این فرآیند می‌تواند به شست‌وشوی نیترات‌های تجمع‌یافته در لایه‌های سطحی خاک منجر

شود. از سوی دیگر، در مناطق کشاورزی که مصرف کودهای نیتروژنه رایج است، بارش شدید می‌تواند باعث انتقال سریع نیترات به ژرفای خاک و سرانجام به منابع آب زیرزمینی شود. این یافته با نتایج ونگ و همکاران (۲۰۲۲) مبنی بر اینکه بارش‌های فصلی نقش مهمی در افزایش غلظت نیترات در آب‌های زیرزمینی به‌ویژه در مناطقی با نفوذپذیری زیاد و پوشش گیاهی محدود دارند، هماهنگی دارد. متغیر بلندی نیز به‌عنوان یک متغیر پستی‌بلندی، بر الگوهای جریان سطحی و زیرسطحی تأثیرگذار است. مناطق با بلندی بیشتر معمولاً شیب‌های تندتر دارند که می‌تواند منجر به رواناب سریع‌تر و کاهش زمان تماس آب با خاک شود.

به‌طور غیرمستقیم به پوشش گیاهی، رطوبت، خاک و فرآیندهای انتقال آلاینده مرتبط است (ژانگ و همکاران ۲۰۱۸؛ بوون ۲۰۱۲). نتایج این پژوهش نشان داد که پس از بارش و بلندی، متغیرهای فاصله از آبراهه، تراکم زهکشی، NDVI و شیب زمین در رتبه دوم اهمیت بودند (جدول ۳).

این موضوع ممکن است از نفوذ مؤثر آب به لایه‌های زیرین جلوگیری کرده و در نتیجه بر اندازه انتقال نیترات به سفره‌های زیرزمینی تأثیرگذار باشد. از سوی دیگر، مناطق با بلندی کم و شیب ملایم‌تر، مستعد تجمع آب و نفوذ بیشتر هستند، که می‌تواند خطر آلودگی را افزایش دهد. از این رو، اهمیت متغیر بلندی در مدل هم به ویژگی‌های فیزیکی زمین مربوط است و هم

جدول ۳- اهمیت و وزن متغیرهای مهم در تحلیل احتمال رخداد آلودگی نیترات.

Table 3- Importance and weights of key variables in nitrate contamination probability analysis.

Variables	Importance
Precipitation	0.21
Elevation	0.18
Distance from streams	0.13
NDVI	0.13
Drainage density	0.13
Slope	0.11
Topographic Wetness Index	0.05
Lithology	0.03
Land use	0.03
Soil order	0.01

خاک در برخی نواحی ناقص یا با دقت محدود بودند. این محدودیت‌ها مشابه مشکلاتی است که در پژوهش‌هایی با داده‌های محدود چاه‌های خصوصی نیز قابل مشاهده است که باعث افزایش خطا در مدل‌سازی می‌شود (آگی‌مانگ و همکاران ۲۰۱۷). این محدودیت‌ها می‌توانند نبودن قطعیت مدل‌های پیش‌بینی را افزایش داده و قابلیت تعمیم نتایج به دیگر مناطق یا بازه‌های زمانی را کاهش دهند. از این رو، به منظور بهبود صحت و قابلیت اعتماد تحلیل‌ها، گردآوری داده‌های با دوره زمانی بیشتر و با پوشش مکانی گسترده‌تر و پایش پیوسته چاه‌ها برای ثبت نوسانات فصلی ضروری است.

استفاده از اندازه کد مصرفی به عنوان یک متغیر ورودی در منطقه مطالعه شده امکان‌پذیر نبود. استفاده از چنین ویژگی مهمی دقت مدل‌سازی را افزایش می‌داد.

محدودیت‌های پژوهش

یکی از محدودیت‌های اساسی این پژوهش، کمبود و ناقص بودن داده‌های پایه بود. داده‌های نیترات فقط از ۱۰۲ چاه و در بازه‌ای زمانی محدود جمع‌آوری شدند. از این رو، تغییرات فصلی و نوسانات کوتاه‌مدت کیفیت آب زیرزمینی لحاظ نشد. این موضوع مشابه محدودیت منتج از داده‌های نسبتاً ناپیوسته در مدل‌سازی نوسانات نیترات است که غنای فرایندهای زیست‌زمین‌شیمیایی را کاهش می‌دهد (ویلیکینز و همکاران ۲۰۲۳). در روش پژوهش فرض شد که سامانه در شرایط پایدار است و انتقال آلودگی از آبخوان در نظر گرفته نشد. از این رو، ممکن است پیوند موجود میان آلودگی و نقاط نمونه‌برداری از آب زیرزمینی در سمت پایین‌دست را «قطع» کند (شو و گومز-هرناندز ۲۰۱۶). افزون بر این، داده‌های محیطی مرتبط با عامل‌های مؤثر بر کیفیت آب، مانند کاربری زمین و ویژگی‌های

نتیجه‌گیری و پیشنهادها

جایگزینی روش‌های آبیاری سنتی با سامانه‌های نوین مانند آبیاری قطره‌ای، راهکارهای مؤثری برای کاهش نفوذ نیترات به سفره‌های آب زیرزمینی ارائه داد. ارزیابی اقتصادی-زیست‌محیطی این جایگزینی‌ها، همراه با تحلیل سیاست‌های حمایتی و آموزشی برای کشاورزان، می‌تواند به طراحی برنامه‌های اجرایی قابل‌انکسار منجر شود. تلفیق چارچوب ارزیابی خطر با سامانه‌های اطلاعات جغرافیایی و داده‌های سنجش‌ازدور نیز امکان پایش منطقه‌ای و تصمیم‌گیری هوشمند را فراهم می‌آورد. در نتیجه، طراحی شاخص‌های ترکیبی برای سنجش پایداری منابع آب زیرزمینی و ارائه ابزارهای تصمیم‌گیری چندمعیاره، می‌تواند به ارتقای کیفیت مدیریت منابع آب در مناطق خشک و نیمه‌خشک کمک شایانی کند.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ‌گونه تضاد منافی در راستای نگارش و انتشار مطالب و نتایج این پژوهش ندارند.

دسترسی به داده‌ها

داده‌ها و نتایج استفاده‌شده در این پژوهش، با مکاتبه با نویسنده مسئول در اختیار مخاطب قرار خواهد گرفت.

مشارکت نویسندگان

نویسنده اول: مفهوم‌سازی، انجام تحلیل‌های نرم‌افزاری/آماري، نگارش نسخه اولیه مقاله
نویسنده دوم: راهنمایی، ویرایش و بازبینی مقاله، بررسی نتایج

ارزیابی خطر آلودگی آب‌های زیرزمینی ابزار مهم و مؤثری برای مدیریت منابع آب‌های زیرزمینی به‌ویژه در مناطق خشک و نیمه‌خشک به‌شمار می‌آید. این پژوهش، چارچوبی برای ارزیابی خطر آلودگی آب‌های زیرزمینی در دشت لنجان ارائه داد. نتایج نشان داد که احتمال رخداد آلودگی در بخش مرکزی منطقه بیشتر است. با توجه به تحلیل نقشه کاربری زمین، مشخص شد که مناطق با خطر زیاد و بسیار زیاد عمدتاً با کاربری کشاورزی هم‌پوشانی دارند. این یافته‌ها مؤید آن است که فعالیت‌های انسانی، به‌ویژه کشاورزی و مدیریت نامناسب کودها، نقش مهمی در افزایش خطر آلودگی آب‌های زیرزمینی دارند. نتایج این پژوهش نشان داد که گسترش ارزیابی خطر آلودگی آب‌های زیرزمینی با استفاده از رویکرد پیشنهادی، امکان‌پذیر و قابل اعتماد است. برای مدیریت و مهار کیفیت آب‌های زیرزمینی در منطقه، کاهش مصرف کودهای نیتروژنی در آبیاری اهمیت ویژه‌ای دارد. از این‌رو، پیشنهاد می‌شود سامانه آبیاری قطره‌ای جایگزین آبیاری غرقابی شود تا از شسته شدن نیترات خاک جلوگیری شود و اثرات منفی کشاورزی بر کیفیت آب کاهش یابد. یافته‌های این پژوهش برای ارزیابی خطر آلودگی آب‌های زیرزمینی در مقیاس منطقه‌ای قابل اعتماد است. از این‌رو، پیشنهاد می‌شود تا تصمیم‌گیری‌های مدیریتی در زمینه حفاظت از منابع آب زیرزمینی و پیشگیری از آلودگی در مناطق نیمه‌خشک از نتایج ارزشمند آن بهره‌برد. در این راستا، پیشنهاد می‌شود در پژوهش‌های آینده به توسعه مدل‌های پیش‌بینی جابه‌جایی آلودگی با بهره‌گیری از الگوریتم‌های یادگیری ماشین و داده‌های مکانی توجه شود تا در آینده مناطق پرخطر با دقت بیشتری شناسایی شوند. همچنین، پیشنهاد می‌شود با بررسی و تحلیل سناریوهای مدیریتی کاهش مصرف کود و

فهرست منابع

- Agyemang, ABA. 2017. Vulnerability assessment of groundwater to NO₃ contamination using GIS, DRASTIC model and geostatistical analysis. East Tennessee State University. p.13830146.
- Alam SK, Li P, Rahman M, Fida M, Elumalai V. 2025. Key factors affecting groundwater nitrate levels in the Yinchuan Region, Northwest China: Research using the eXtreme Gradient Boosting (XGBoost) model with the SHapley Additive exPlanations (SHAP) method. *Environmental Pollution*. 364(1): p.125336. <https://doi.org/10.1016/j.envpol.2024.125336>
- Aller L, Bennett T, Lehr J, Petty RJ, Hackett G. 1987. DRASTIC: A standardized system for evaluating ground water pollution potential using hydrogeologic settings. US Environmental Protection Agency. Washington, DC, 455.
- Almasri MN, Kaluarachchi JJ. 2007. Modeling nitrate contamination of groundwater in agricultural watersheds. *Journal of Hydrology*. 343(3-4):211–229. <https://doi.org/10.1016/j.jhydrol.2007.06.016>
- Arauzo M, 2017. Vulnerability of groundwater resources to nitrate pollution: A simple and effective procedure for delimiting Nitrate Vulnerable Zones. *Science of the Total Environment*. 575:799–812. <https://doi.org/10.1016/j.scitotenv.2016.09.139>
- Beven KJ, Alcock RE. 2012. Modelling everything everywhere: A new approach to decision-making for water management under uncertainty. *Freshwater Biology*. 57:124–132. <https://doi.org/10.1111/j.1365-2427.2011.02592.x>
- Chen C, Yin C, Wang Y, Zeng J, Wang S, Bao Y, Liu X. 2023. XGBoost-based machine learning test improves an accuracy of hemorrhage prediction among geriatric patients with long-term administration of rivaroxaban. *BMC Geriatrics*. 23(1): p. 418. <https://doi.org/10.1186/s12877-023-04049-z>
- Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A. 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*. 651:2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Covatti G, Li KY, Podgorski J, Winkel LH, Berg M. 2025. Nitrate contamination in groundwater across Switzerland: Spatial prediction and data-driven assessment of anthropogenic and environmental drivers. *Science of the Total Environment*. 973(10): p.179121. <https://doi.org/10.1016/j.scitotenv.2025.179121>
- Creed IF, Band LE. 1998. Exploring functional similarity in the export of Nitrate-N from forested catchments: A mechanistic modeling approach. *Water Resources Research*. 34(11):3079–3093. <https://doi.org/10.1029/98WR02102>
- Dhaliwal JK, Panday D, Saha D, Lee J, Jag Adamma S, Schaeffer S, Mengistu A. 2022. Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Computers and Electronics in Agriculture*. 199: p.107107. <https://doi.org/10.1016/j.compag.2022.107107>
- Gad M, Gaagai A, Agrama AA, El-Fiqy W F, Eid MH, Szűcs P, Ibrahim H. 2024. Comprehensive evaluation and prediction of groundwater quality and risk indices using quantitative approaches, multivariate analysis, and machine learning models: An exploratory study. *Heliyon*. 10(17): p. e36606. <https://doi.org/10.1016/j.heliyon.2024.e36606>
- Gholami V, Booij MJ. 2022. Use of machine learning and geographical information system to predict nitrate concentration in an unconfined aquifer in Iran. *Journal of Cleaner Production*. 360: p. 131847. <https://doi.org/10.1016/j.jclepro.2022.131847>
- Gurdak JJ, Qi SL. 2012. Vulnerability of recently recharged groundwater in principle aquifers of the United States to nitrate contamination. *Environmental Science and Technology*. 46(11):6004–6012. <https://doi.org/10.1021/es300688b>
- Hosseini FS, Choubin B, Bagheri-Gavkosh M, Karimi O, Taromideh F, Mako C. 2023. Susceptibility assessment of groundwater nitrate contamination using an ensemble machine learning approach. *Groundwater*. 61(4):510–516. <https://doi.org/10.1111/gwat.13258>
- Hosseini FS, Choubin B, Bagheri-Gavkosh M, Karimi O, Taromideh F, Mako C. 2023. Susceptibility assessment of groundwater nitrate contamination using an ensemble machine learning approach. *Groundwater*. 61(4):510–516. <https://doi.org/10.1111/gwat.13258>
- Imani M, Beik Mohammadi A, Arabnia HR. 2025. Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. *Technologies*. 13(3): p. 88. <https://doi.org/10.3390/technologies13030088>
- Interagency Agricultural Projections Committee. 2023. USDA Agricultural Projections to 2032. p. 119. <https://doi.org/10.22004/ag.econ.348090>

- International Agency for Research on Cancer, IARC. 2010. Ingested nitrate and nitrite, and cyanobacterial peptide toxins. International Agency for Research on Cancer. <https://cir.nii.ac.jp/crid/1970867909768062125>
- Knoll L, Breuer L, Bach M. 2020. Nationwide estimation of groundwater redox conditions and nitrate concentrations through machine learning. *Environmental Research Letters*. 15(6):p.064004. <https://doi.org/10.1088/1748-9326/ab7d5c>
- Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B. Team RC, 2020. Package 'caret'. *The R Journal*. 22(7):1-48.
- Liang Y, Zhang X, Gan L, Chen S, Zhao S, Ding J, Wulue K, Yang H. 2024. Mapping specific groundwater nitrate concentrations from spatial data using machine learning: A case study of chongqing, China. *Heliyon*. 10(6): p.e27867. <https://doi.org/10.1016/j.heliyon.2024.e27867>.
- Luo S, El X, Li X. 2024. Data Preprocessing Method for Landslide Displacement Prediction Based on XG Boost. In 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS). IEEE. pp. 745-750. <https://doi.org/10.1109/DDCLS61622.2024.10606761>
- Messier KP, Wheeler DC, Flory AR, Jones RR, Patel D, Nolan BT, Ward M H. 2019. Modeling groundwater nitrate exposure in private wells of North Carolina for the Agricultural Health Study. *Science of The Total Environment*. 655:512-519. <https://doi.org/10.1016/j.scitotenv.2018.11.022>
- Mosavi A, Sajedi Hosseini F, Choubin B, Goodarzi M, Dineva AA, Rafiei Sardooi E. 2021. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resources Management*. 35(1):23-37. <https://doi.org/10.1007/s11269-020-02704-3>
- Naseri HR, Key HZ, Nakhai M. 2012. The impact of natural and human factors on water quality in Lenjanat Plain, Isfahan. *Geosciences*. 22(85):173-186. (In Persian).
- Ouedraogo I, Vanclooster M. 2016. A meta-analysis and statistical modelling of nitrates in groundwater at the African scale. *Hydrology and Earth System Sciences*. 20(6):2353-2381. <https://doi.org/10.5194/hess-20-2353-2016>
- Picetti R, Deeney M, Pastorino S, Miller MR, Shah A, Leon DA, Green R. 2022. Nitrate and nitrite contamination in drinking water and cancer risk: A systematic review with meta-analysis. *Environmental Research*. 210:p.112988. <https://doi.org/10.1016/j.envres.2022.112988>
- Rahimi D, Bashirian F, Nourbakhsh A. 2024. Assessment of climate change impacts on water resources (Lenjanat Sub-basin). *Natural Geography*. 64(17): 63-78. (In Persian).
- Rahmati O, Choubin B, Fathabadi A, Coulon F, Soltani E, Shahabi H, Mollaeifar E, Tiefenbacher J, Cipullo S, Ahmad BB, Bui DT. 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. *Science of the Total Environment*. 688:855-866. <https://doi.org/10.1016/j.scitotenv.2019.06.320>
- Ransom KM, Nolan BT, Stackelberg PE, Belitz K, Fram MS. 2022. Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States. *Science of the Total Environment*. 807:p.151065. <https://doi.org/10.1016/j.scitotenv.2021.151065>
- Sajedi-Hosseini F, Malekian A, Choubin, B, Rahmati O, Cipullo S, Coulon F, Pradhan B. 2018. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Science of the Total Environment*. 644:954-962. <https://doi.org/10.1016/j.scitotenv.2018.07.054>
- Sarkar S, Mukherjee A, Gupta SD, Bhanja SN, Bhattacharya A. 2022. Predicting regional-scale elevated groundwater nitrate contamination risk using machine learning on natural and human-induced factors. *Acs Es and T Engineering*. 2(4):689-702. <https://doi.org/10.1021/acsestengg.1c00360>
- Shukla S, Saxena A. 2021. Appraisal of groundwater quality with human health risk assessment in parts of Indo-Gangetic alluvial plain, North India. *Archives of Environmental Contamination and Toxicology*. 80(1):55-73. <https://doi.org/10.1007/s00244-020-00771-6>
- Ullah S, Zhang H, Heathwaite AL, Heppell C, Lansdown K, Binley A, Trimmer M. 2014. Influence of emergent vegetation on nitrate cycling in sediments of a groundwater-fed river. *Biogeochemistry*. 118(1):121-134. <https://doi.org/10.1007/s10533-013-9909-2>
- United Nations Environment Programme, UNEP. 2023. An introduction to SDG indicator 6.3.2: proportion of bodies of water with good ambient water quality. <https://communities.unep.org/display/sdg632/Documents+and+Materials>
- Wang ZJ, Yue FJ, Lu J, Wang YC, Qin CQ, Ding H, Xue LL, Li SL. 2022. New insight into the response and transport of nitrate in karst groundwater to rainfall events. *Science of the Total Environment*. 818:p.151727. <https://doi.org/10.1016/j.scitotenv.2021.151727>

Ward MH, Jones RR, Brender JD, De Kok TM, Weyer PJ, Nolan BT, Van Breda SG. 2018. Drinking water nitrate and human health: an updated review. *International Journal of Environmental Research and Public Health*. 15(7):1557.

<https://doi.org/10.3390/ijerph15071557>

Wiens M, Verone-Boyle A, Henscheid N, Podichetty JT, Burton J. 2025. A tutorial and use case example of the extreme gradient boosting (XGBoost) artificial intelligence algorithm for drug development applications. *Clinical and Translational Science*. 18(3):p.e70172.

<https://doi.org/10.1111/cts.70172>

Wilkins B, Johns T, Mager S. 2024. Nitrate-nitrogen dynamics in response to forestry harvesting and climate variability: Four years of UV nitrate sensor data in a shallow, gravel aquifer. *EGUsphere*. 2024:1–27.

<https://doi.org/10.5194/egusphere-2024-964>

World Health Organization (WHO). 2011. *Guidelines for Drinking-water Quality*. 38(4):1–108.

World Health Organization (WHO). 2022. *Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda*. World Health Organization. p. 494.

Xu T, Gómez-Hernández JJ. 2016. Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. *Water Resources Research*. 52(8):6587–6595. <https://doi.org/10.1002/2016WR019111>

Zhang P, Jia Y, Shang Y. 2022. Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*. 18(6):p.15501329221106935

Zhang W, Wang W, Zhou D, Zhang, R, Goh ATC, Hou Z. 2018. Influence of groundwater drawdown on excavation responses—A case history in Bukit Timah granitic residual soils. *Journal of Rock Mechanics and Geotechnical Engineering*. 10(5):856-864. <https://doi.org/10.1016/j.jrmge.2018.04.006>



Estimating Groundwater Nitrate Contamination Probability Using Extreme Gradient Boosting

Mehdi Hashemi^{1*}, Ali Dastranj²

1- Assistant Professor, Soil conservation and watershed management research department, Isfahan Agricultural and Natural Resources Research and Education Center, AREEO, Isfahan, Iran

2- Assistant Professor, Soil conservation and watershed management research department, Khorasan Razavi Agricultural and Natural Resources Research and Education Center, AREEO, Mashhad, Iran

Extended Abstract

Introduction and Goal

Groundwater, as a vital source of fresh water, plays a fundamental role in supplying drinking, agricultural, and industrial needs in many arid and semi-arid regions of the world. However, increased human and industrial activities have led to the exacerbation of pollution in these valuable resources. In this regard, nitrate pollution, due to its high solubility and mobility in water, is recognized as one of the most serious threats to human health and aquatic ecosystems. The consumption of nitrate-contaminated water can lead to various diseases, including methemoglobinemia (blue baby syndrome) in infants and even some cancers in adults. Furthermore, the entry of nitrates into surface waters can result in eutrophication and the degradation of aquatic ecosystems. Given the importance of the issue and the necessity of protecting groundwater resources, this research was conducted with the aim of developing an integrated and comprehensive framework for estimating the probability of groundwater contamination, especially with a focus on nitrate contaminant, in the Lenjanat Plain region located in Isfahan Province, Iran. Using this framework and employing advanced modeling and spatial analysis approaches, areas prone to contamination were identified, which will help in providing effective management solutions to reduce the risks associated with groundwater contamination. The results of this research can serve as a basis for future planning in the sustainable management of water resources and the protection of community health.

Materials and Methods

In this study, data related to the nitrate concentration in groundwater sources were carefully examined. Therefore, crucial information was collected from 102 wells in the Lenjanat Plain of Isfahan Province. Each of these wells represented the nitrate status in the groundwater aquifers of the studied region. To analyze this large volume of data and extract hidden patterns, the Extreme Gradient Boosting model was used. This model was chosen due to its high capability in identifying complex and non-linear relationships between variables, as well as its acceptable prediction precision.

Article Type: Research Article

***Corresponding Authors' E-mail:** m_hashemi@areeo.ac.ir

Citation: Hashemi, M., Dastranj, A. 2026. Estimating Groundwater Nitrate Contamination Probability Using Extreme Gradient Boosting. Watershed Management Research. 39(1): 45-63.

DOI: 10.22092/WMRJ.2025.370560.1631

Received: 30 August 2025, **Received in revised form:** 12 September 2025, **Accepted:** 22 September 2025

Published online: 21 March 2026

Watershed Management Research, Vol. 39, No.1, Ser. No: 150, Spring 2026, pp. 45-63.

Publisher: Fars Agricultural and Natural Resources and Education Center

© Author(s)



In addition to nitrate concentration data, ten key environmental and anthropogenic factors potentially influencing nitrate contamination in groundwater were identified and incorporated into the analytical model. These factors included slope, elevation, drainage density, topographic wetness index, soil order, and distance from streams, lithology, and land-use. By integrating these eight factors into the Extreme Gradient Boosting model, it was possible to identify the most significant factors affecting nitrate contamination and also to spatially predict the probability of nitrate contamination in groundwater.

Results and Discussion

The results of this study clearly demonstrated the effectiveness and efficiency of the Extreme Gradient Boosting in predicting nitrate contamination in groundwater. The overall accuracy of this model was 0.86 which allowed the contamination status of the studied area to be well distinguished. In addition, other performance evaluation criteria of the model also indicated its high accuracy in correctly identifying contaminated and uncontaminated areas; with the area under the ROC curve was equal to 0.85. Moreover, the model recall was found to be 0.80, indicating that 80% of all the real contaminated areas were correctly identified using this model. Finally, the F1-score statistic, which is a combined measure of precision and recall, with a value of 0.83, indicates a good balance between these two measures and the overall reliable performance of the model. The sensitivity analysis of the model revealed that the effect of certain input variables on the spatial estimation of nitrate contamination in groundwater was significant. Among the ten environmental and anthropogenic factors examined, precipitation (21%) and elevation changes (18%) were identified as the most influential and important variables in determining the spatial pattern of nitrate contamination. These findings highlight the importance of natural and geomorphological characteristics of the region in controlling the dispersion and accumulation of nitrates in groundwater and can serve as a useful guide for future studies and the development of targeted management strategies.

Conclusion and Suggestions

One of the important achievements of this study was the production of hazard maps that clearly identified areas with high risk of nitrate contamination in the central part of the studied plain. It is recommended that water resource managers and urban and rural planners use these maps as a valuable tool for taking preventive measures in sensitive areas. Notably, the role of human activities in increasing the risk of nitrate contamination was strongly confirmed by the significant overlap of high-risk areas with agricultural land-use. Based on these findings, it is suggested that nitrogen fertilizers be used optimally for the protection of groundwater resources and the sustainable management of agricultural activities.

Keywords: Groundwater, Lenjanat Plain, nitrate, occurrence probability, XGBoost

Article Type: Research Article

Conflicts of interest

The authors of this article declared no conflict of interest regarding the authorship or publication of this article.

Data Availability Statement

The datasets are available upon a reasonable request to the corresponding author.

Authors' Contribution

Author 1: Conceptualization, Software, Formal analysis, Writing - original draft preparation

Author 2: Conceptualization, Evaluating the result, Manuscript editing