



دوره ۳۵، شماره ۳، شماره ۱ پی‌اچ، پاییز ۱۴۰۱، صفحه‌های ۴۴-۵۹

شناسه‌ی دیجیتال: 10.22092/WMRJ.2021.356517.1440

مقاله‌ی پژوهشی



پژوهش‌های آبخزرداری

کاربرد الگوریتم‌های درخت وایازی تقویت‌شده، درخت مدل پشتیبان، و جنگل تصادفی برای ارزیابی احتمال آب‌های زیرزمینی

سید احمد اسلامی نژاد

دانش‌آموخته‌ی کارشناسی‌ارشد، گروه مهندسی نقشه‌برداری، دانشکده‌ی مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران، تهران

مبین افتخاری

(نویسنده‌ی مسئول)* دانش‌آموخته‌ی کارشناسی‌ارشد مهندسی نقشه‌برداری، گروه مهندسی عمران آب و سازه‌های هیدرولیکی و عضو باشگاه پژوهشگران جوان و نخبگان، واحد مشهد، دانشگاه آزاد اسلامی، مشهد

محمد اکبری

دانشیار گروه مهندسی عمران، دانشگاه بیرجند

هادی بیات

دانش‌آموخته‌ی کارشناسی‌ارشد، گروه مهندسی نقشه‌برداری، دانشکده‌ی مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران، تهران

وریا برقی

دانش‌آموخته‌ی کارشناسی‌ارشد، گروه مهندسی نقشه‌برداری، دانشکده‌ی مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران، تهران

*رایانامه‌ی نویسنده‌ی مسئول: Mobineftekhari@yahoo.com

تاریخ دریافت: ۸ آبان ۱۴۰۰ تاریخ پذیرش: ۲۷ آذر ۱۴۰۰

چکیده

به دلیل رشد جمعیت و صنعتی‌شدن در منطقه‌های گوناگون جهان، از آب‌های زیرزمینی به‌طور مہارناپذیری بهره‌برداری می‌شود. هدف این پژوهش، ارزیابی احتمال آب‌های زیرزمینی با الگوریتم‌های پیشرفته‌ی یادگیری ماشین و با معیارهای پستی-بلندی، آب‌شناسی، محیطی و زمین‌شناسی است. برای انجام این کار سه الگوریتم پیشرفته‌ی یادگیری ماشین شامل درخت وایازی تقویت‌شده، درخت مدل پشتیبان، و جنگل تصادفی به کار برده شد. داده‌های آب‌شناسی ۳۷ چاه آب زیرزمینی در دشت بیرجند، استان خراسان جنوبی، جمع‌آوری، و با انتخاب تصادفی به کار برده شد. داده‌های آب‌شناسی ۳۷ چاه آب زیرزمینی و اعتبارسنجی تقسیم کرده شد. نقشه‌های احتمال آب زیرزمینی با سه الگوریتم تهیه شد. برای اعتبارسنجی الگوریتم‌های پیش‌بینی احتمال آب زیرزمینی، سطح زیر منحنی و معیارهای آماری نرخ پیش‌بینی‌شده‌ی مثبت، نرخ پیش‌بینی‌شده‌ی منفی، حساسیت، ویژگی، و دقت به کار برده شد. نتیجه نشان داد که درخت مدل پشتیبان ($AUC=0.865$) کارکرد بهتری در پیش‌بینی احتمال آب زیرزمینی منطقه دارد.

واژگان کلیدی: سامانه‌ی اطلاعات جغرافیایی (جی‌آی‌اس)، احتمال آب زیرزمینی، یادگیری ماشین

مقدمه

و همکاران (۲۰۱۹) نشان داد که ترکیب مدل‌های یادگیری ماشین با مدل‌های آماری دقت بسیاری در پیش‌بینی احتمال آب زیرزمینی در دشت بوشهر نتیجه می‌دهد. پراساد و همکاران (۲۰۲۰) روند پیش‌بینی منطقه‌های توانمند آب زیرزمینی در سواحل غربی هند را با سه مدل یادگیری ماشین درخت وایزی تقویت شده، جنگل تصادفی، و ماشین بردار پشتیبان بررسی کردند. نتیجه نشان داد که مدل RF ($AUC=0.94$) دقت بیش‌تری دارد.

در همه‌ی این پژوهش‌ها محققان مدل‌های یادگیری ماشین و فرآیند کار را برای انتخاب بهترین مدل در پیش‌بینی احتمال آب‌های زیرزمینی مقایسه کرده‌اند. اگرچه برخی از الگوریتم‌های ترکیبی یادگیری ماشین برای پیش‌بینی احتمال آب‌های زیرزمینی به کار برده شده‌است، هیچ روشی در پیش‌بینی دقیق کامل نیست. هدف این پژوهش، ارزیابی احتمال آب‌های زیرزمینی با الگوریتم‌های پیش‌رفته‌ی یادگیری ماشین شامل درخت وایزی تقویت شده، درخت مدل پشتیبان، و جنگل تصادفی است. این پژوهش به این دلیل تازگی دارد که درخت مدل پشتیبان را برای ارزیابی احتمال آب‌های زیرزمینی به کار می‌برد. الگوریتم درخت مدل پشتیبان به دلیل کارایی پذیرفتنی در پژوهش‌های مکانی گوناگون مانند حساسیت سیل (فام و همکاران ۲۰۲۰، خسروی و همکاران ۲۰۱۹) و حساسیت زمین‌لغزش (نهو و همکاران ۲۰۲۰) برای این پژوهش انتخاب شد. یکی دیگر از نوآوری‌های این پژوهش کاربرد هم‌زمان ۱۹ معیار آب‌شناسی، پستی‌بلندی، زمین‌شناسی و محیطی است که کم‌تر در پژوهش‌های آب‌های زیرزمینی به کار برده شده‌بود.

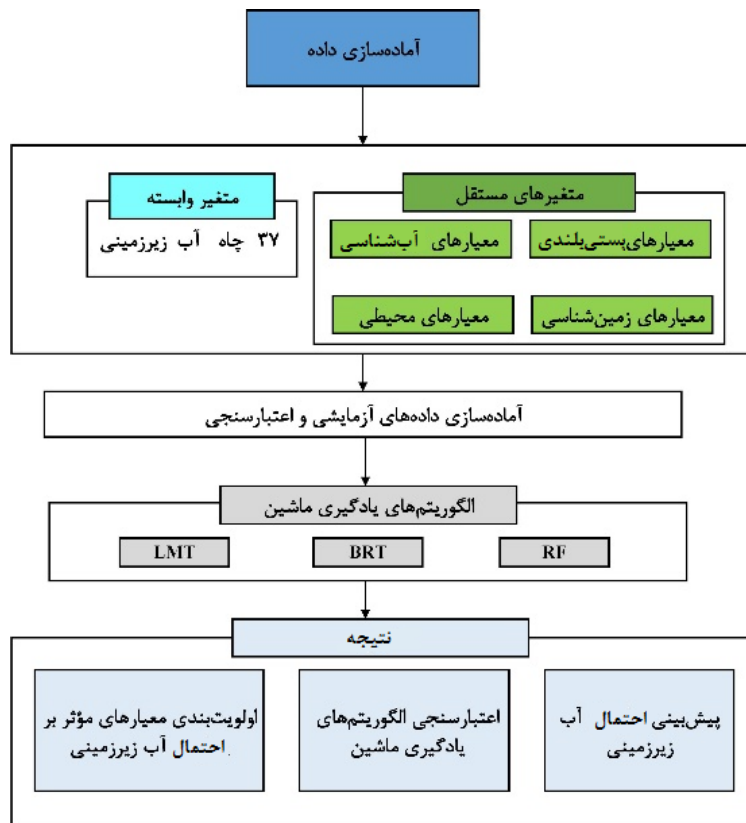
مواد و روش‌ها

روند اجرایی این پژوهش شامل ۵ مرحله‌ی پایه‌ی است (شکل ۱): جمع‌آوری چاه‌های مشاهده‌ی آب زیرزمینی در منطقه برای تعریف کردن متغیر وابسته، استخراج معیارهای مکانی مؤثر بر احتمال آب زیرزمینی برای تعریف کردن متغیرهای مستقل، پیش‌بینی منطقه‌های توانمند آب زیرزمینی بر پایه‌ی سه الگوریتم پیش‌رفته‌ی یادگیری ماشین شامل درخت وایزی تقویت‌شده، درخت مدل پشتیبان، و جنگل تصادفی، تعیین اهمیت معیارهای آب‌شناسی، پستی‌بلندی، زمین‌شناسی، و محیطی در ارزیابی احتمال آب زیرزمینی، و ارزیابی و مقایسه‌ی کارکرد روی کردهای یادگیری ماشین.

سفره‌های آب زیرزمینی در جهان بخش عمده‌ی اندوخته‌های آب شیرین در منطقه‌های خشک و نیمه‌خشک را تشکیل می‌دهند (چزگی و همکاران ۲۰۱۶). کاربرد بلندمدت از اندوخته‌های آب زیرزمینی پی‌آمدهای ناخواسته‌ی مانند کاهش سفره‌های آب زیرزمینی و در نتیجه نشست زمین، نفوذ آب شور و کاهش کیفیت آب را به دنبال دارد. بنابراین، تحقیقات درباره‌ی احتمال آب‌های زیرزمینی در حوزه‌های آبخیز، یا حتی در مقیاس بزرگ‌تر، به مدیران کمک می‌کند تا به شیوه‌ی پایدار از آب‌های زیرزمینی بهره‌برند (ساهو و همکاران ۲۰۱۷، لی و همکاران ۲۰۲۰). به همین دلیل، ارزیابی آب‌های زیرزمینی راه‌برد مفیدی است که ممکن است احتمال بودن آن‌ها را برای کاربرد در هدف‌های بهره‌برداری یا برنامه‌های حفاظتی در منطقه‌های گوناگون تعریف کند (نقیبی و همکاران ۲۰۱۸). برای آگاهی از بهره‌وری آب‌های زیرزمینی چندین شاخص مانند چشمه‌ها، چاه‌ها و قنات‌ها هست (کیم و همکاران ۲۰۱۹، نقیبی و همکاران ۲۰۱۸). بر پایه‌ی نوع منطقه، در این پژوهش بر شاخص چاه‌ها تمرکز شد.

نقشه‌برداری از چاه‌های آب زیرزمینی به تصمیم‌گیرندگان کمک می‌کند تا اندوخته‌های آب زیرزمینی را تشخیص دهند و در زمان‌های ویژه مانند خشک‌سالی از آن‌ها بهره‌گیرند (کرسینی و همکاران ۲۰۰۹). علاوه بر این، شناخت منطقه‌های توانمند آب‌های زیرزمینی به چگونگی کاربرد و برنامه‌های حفاظت از اندوخته‌های آن‌ها کمک می‌کند. در سال‌های گذشته روش‌های زیادی برای تهیه‌ی نقشه‌ی احتمال آب‌های زیرزمینی به کار رفته است. بسیاری از پژوهشگران با در نظر گرفتن طیف گسترده‌ی از معیارهای مؤثر، مدل‌های گوناگونی را بر پایه‌ی سامانه‌ی اطلاعات جغرافیایی و سنجش‌ازدور برای شناخت منطقه‌های توانمند آب‌های زیرزمینی پیاده کرده‌اند. کومار و کریشنا (۲۰۱۸) نشان دادند که اندازه‌ی سطح زیر منحنی برای فرآیند سلسله‌مراتبی تحلیلی $75/45$ است، که نشان‌دهنده‌ی دقت زیاد بینش این مدل در شناخت منطقه‌های توانمند آب‌های زیرزمینی در آبخیز دشت رامگار کشور هند بود. نقیبی و همکاران (۲۰۱۸) نشان دادند که موقعیت چاه‌ها را می‌توان شاخص مناسبی برای احتمال آب زیرزمینی در آبخیز بهشتآباد ایران دانست. علاوه بر این، بلندی و شیب مهم‌ترین معیارهای تاثیرگذار، و سنگ‌شناسی و جهت شیب معیارهایی با تاثیرگذاری کم‌تر دانسته شد. نتیجه‌ی رضوی ترمه

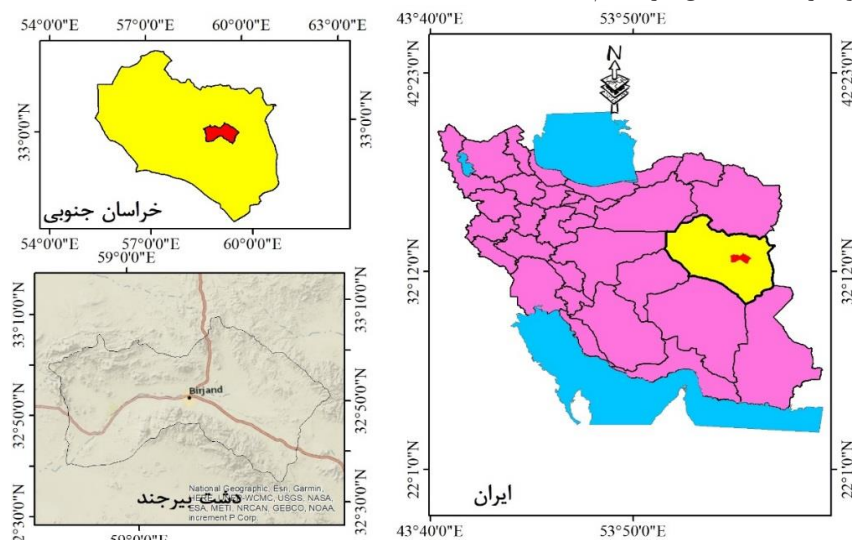
کاربرد الگوریتم‌های درخت وایازی تقویت‌شده، درخت مدل پشتیبان...



شکل ۱- نمودار جریان‌ی روش پژوهش.

است، بنابراین بیش‌تر آب نیازداشته‌ی کشاورزی، صنعت و آشامیدنی شهرستان بیرجند از سفره‌ی آب زیرزمینی تأمین می‌شود (افتخاری و همکاران ۲۰۲۱ a). بیشینه‌ی بلندی حوزه از تراز دریا ۲۷۲۹ متر در بلندی‌های باقران، و کمینه‌ی بلندی ۱۱۷۲ متر در خروجی دشت (منطقه‌ی فدشک) است (افتخاری و همکاران ۲۰۱۹).

دشت بیرجند (58° و $41'$ تا 59° و $46'$ شرقی و 32° و $35'$ تا 33° و $8'$ شمالی) یکی از محدوده‌های پژوهشی آبخیز کویر لوت است (شکل ۲). مساحت کلی آن $3406/72$ کیلومترمربع است که $900/63$ کیلومترمربع آن دشت و $2506/09$ کیلومترمربع آن بلندی‌ها است. رود شاهرود (108 کیلومتر) از شرق به غرب دشت رزه‌کشی می‌کند و به کویر لوت می‌ریزد. این رود در همه‌ی سال جز هنگام سیلاب خشک



شکل ۲- محدوده‌ی دشت بیرجند.

جریان سطحی است (رابطه‌ی ۳) (خسروی و همکاران ۲۰۱۶).

$$LS = \left(\frac{A_s}{22.13} \right)^{0.6} \left(\frac{\sin \beta}{0.0896} \right)^{0.06} \quad 3$$

A_s و β به ترتیب نشان‌دهنده‌ی مساحت حوزه (m²/m) و زاویه‌ی شیب (°) است.

رود سرچشمه‌ی اصلی تغذیه‌ی آب‌های زیرزمینی در منطقه‌های نیمه‌خشک است. از این رو فاصله تا رود یکی از بخش‌های آب‌شناسی اصلی است که بر احتمال آب‌های زیرزمینی تأثیر می‌گذارد (خسروی و همکاران ۲۰۱۶). لایه‌ی فاصله تا رود با نقشه‌های پستی‌بلندی ۱:۵۰۰۰۰ تهیه شد. شاخص توان آبراه (SPI) ^۵ بیانگر توان فرساینده‌ی جریان است که با درجه‌ی شیب و مساحت آبخیز رابطه‌ی مستقیم دارد (رابطه‌ی ۴) (هانگ و همکاران ۲۰۱۸):

$$SPI = A_s \times \tan \beta \quad 4$$

A_s و β به ترتیب نشان‌دهنده‌ی مساحت حوزه (m²/m) و زاویه‌ی شیب (°) است.

شاخص رطوبت پستی‌بلندی (TWI) ^۶ در سنجش اندازه‌ی اثر پستی‌بلندی بر محل، اندازه‌ی منطقه‌ی اشباع‌شده، و تولید روان-آب کاربرد زیادی دارد (رابطه‌ی ۵) (جانسیویز و همکاران ۲۰۱۹):

$$TWI = \ln \left(\frac{A_s}{\tan \beta} \right) \quad 5$$

A_s مساحت حوزه (m²/m)، β زاویه‌ی شیب (°) است. نقشه‌ی شدت بارندگی با چهار ایستگاه هم‌دید در منطقه محاسبه شد. لایه‌ی جریان تجمعی بیان‌کننده‌ی اندازه‌ی جریانی است که از یاخته‌های بالادستی به یاخته‌ی دل خواه ختم می‌شود و به کمک تحلیل‌های مناسب در نرم‌افزار SAGA GIS 7.9 محاسبه می‌شود (کنعانی‌سادات و همکاران ۲۰۱۹).

وجود جریان آب‌های زیرزمینی بستگی به ویژگی‌های واحدهای سنگی مانند تخلخل و نفوذپذیری سنگ بستتر دارد (چن و همکاران ۲۰۱۹). نقشه‌های موضوعی زمین‌شناسی شامل نقشه‌ی سنگ‌شناسی، جنس خاک ۱:۲۵۰۰۰۰ (سازمان جنگل‌ها، مراتع و آبخیزداری کشور، مراجعه‌ی شخصی)، نقشه‌ی

به‌طور کلی بهره‌وری آب‌های زیرزمینی برای آبخوان بستگی به معیارهای گوناگونی دارد. بلندی زمین تأثیر مهمی بر شرایط آب‌وهوایی دارد، و نامستقیم بر احتمال آب‌های زیرزمینی تأثیر می‌گذارد. بلندی‌های کم‌تر شیب کم‌تری هم دارند، و بنابراین اندازه‌ی نفوذ در آن نیز افزایش می‌یابد (التوانه و همکاران ۲۰۱۴). تولید روان‌آب مستقیمانه متناسب با شیب است، و شیب کم‌تر باعث افزایش نفوذ آب‌های زیرزمینی می‌شود (بنجمل و همکاران ۲۰۲۰). جهت شیب، جهت غالب شیب و جهت سامانه‌ی زه‌کشی را نشان می‌دهد، زیرا درازای زمان تابش آفتاب یا درسایه‌بودن نقش مهمی در اندازه‌ی رطوبت خاک دارد (رضوی ترمه و همکاران ۲۰۱۹). نقشه‌ی تهیه‌شده از انحنای سطح آبراه انحنای ریخت‌شناسی و پستی‌بلندی منطقه را نمایش می‌دهد، که بر افزایش یا کاهش جریان در سطح تأثیر می‌گذارد (کاستجه و همکاران ۲۰۲۰). تراکم زه‌کشی نشان‌دهنده‌ی شرایط تغذیه‌ی آب‌های زیرزمینی است، که برای اندازه‌گیری اندازه‌ی نفوذ ناشی از آب‌های سطحی است (مناپ و همکاران ۲۰۱۴).

برای تهیه‌ی مدل رقومی بلندی زمین با قدرت تفکیک ۳۰ متر تصویرهای ماهواره‌ی SRTM ^۱ به‌کار برده شد. لایه‌های زاویه‌ی شیب، جهت شیب، تراکم زه‌کشی، و انحنای سطح از لایه‌ی مدل رقومی بلندی ایجاد شد. موقعیت پستی‌بلندی (TPI) ^۲ اختلاف بلندی هر یاخته با میانگین بلندی‌های یاخته‌های همسایه‌ی آن است (رابطه‌ی ۱) (کنعانی‌سادات و همکاران ۲۰۱۹):

$$TPI = Z_0 - \frac{\sum_{i=1}^n Z_i}{n} \quad 1$$

Z_0 بلندی یاخته، و $\frac{\sum_{i=1}^n Z_i}{n}$ میانگین بلندی یاخته‌های همسایه است.

شاخص ناهمواری زمین (TRI) ^۳ نشان‌دهنده‌ی اختلاف بلندی میان یاخته‌های همسایه در شبکه‌ی بلندی است (رابطه‌ی ۲) (کلانتری و همکاران ۲۰۱۷):

$$TRI = \sqrt{Abs (max^2 - min^2)} \quad 2$$

max و min بزرگ‌ترین و کوچک‌ترین اندازه‌های بلندی یاخته‌های همسایه است.

طول شیب (LS) ^۴ مقیاسی از ظرفیت انتقال بار معلق برای

- 1 - Shuttle Radar Topography Mission
- 2 - Topographic Position Index
- 3 - Terrain ruggedness index
- 4 - Length slope
- 5 - Stream power index
- 6 - Topographic Wetness Index

می‌دهد (یوان و همکاران ۲۰۰۷). شاخص NDVI^v با رابطه‌ی ۶ محاسبه شد.

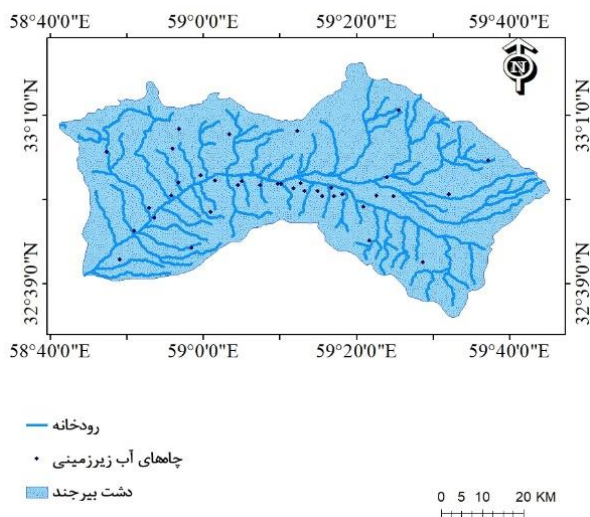
$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad 6$$

NIR و Red به ترتیب نوار فرورسرخ و سرخ ماهواره‌ی لندست ۸ است.

سی و هفت نمونه آب زیرزمینی از چاه‌ها یا پیژومترها از ۱۳۹۳ تا ۱۳۹۸ جمع‌آوری شد (آزمایشگاه کیفی شرکت سهامی آب منطقه‌یی استان خراسان جنوبی) (شکل ۳). داده‌های چاه به نسبت ۷۰٪ برای آموزش و ۳۰٪ برای آزمایش تقسیم شد. بر پایه‌ی شرایط محلی آب زیرزمینی و نظر کارشناسان ۰/۵ L/s از بازدهی آب زیرزمینی اندازه‌ی آستانه برای انتخاب آب زیرزمینی گرفته شد. داده‌های کارکرد آب زیرزمینی این چاه‌ها بر پایه‌ی آزمایش آب کشیدن (پمپاژ) واقعی است، که شرکت سهامی آب منطقه‌یی استان خراسان جنوبی برای ارزیابی احتمال آب زیرزمینی در محل انجام داد.

فاصله تا گسل، و فاصله تا جاده) برای پیش‌بینی احتمال آب‌های زیرزمینی به کار برده شد. گوناگون بودن واحدهای سنگ‌شناسی اختلاف چشمگیری در ناپایداری دامنه نشان می‌دهد. نقشه‌ی سنگ‌شناسی از روی نقشه‌ی زمین‌شناسی کشور ایران در مقیاس ۱:۱۰۰۰۰۰ (۲۶ واحد سنگ‌شناسی) تهیه شد. لایه‌ی فاصله تا جاده و گسل نیز به ترتیب با نقشه‌های پستی‌بلندی ۱:۵۰۰۰۰ و ۱:۱۰۰۰۰۰ تهیه و استخراج شد.

کاربری‌های گوناگون ممکن است در اندازه‌ی نفوذ آب، نفوذ، فرسایش خاک و تبخیر و تعرق نقش داشته‌باشد. برای نمونه در کاربری‌های شهری تقریباً هیچ نفوذی نیست (به دلیل آسفالتی‌بودن خیابان‌ها و ...؛ در حالی که مرتع‌های بسیار خوب ممکن است موجب نفوذ اندازه‌های زیاد آب در زمین شود. تصویر ETM⁺ برای تهیه‌ی نقشه‌ی کاربری زمین به کار برده شد. هنگامی که اندازه‌های شاخص بهنجار شده‌ی تفاوت پوشش گیاهی (NDVI^v) میان ۰/۷۵ و ۱ باشد، نشان‌دهنده‌ی پوشش گیاهی است، در حالی که برای زمین و خاک بایر نزدیک به ۰ است، و اندازه‌های منفی سطح‌های آبی مانند مخزن‌های آب را نشان



شکل ۳- موقعیت چاه‌های مشاهده‌یی در منطقه.

و وایازی پشتیبان به کار می‌گیرد تا احتمال را در گره بررسی کند (رابطه‌ی ۷) (خسروی و همکاران ۲۰۱۸).

$$P(N|x) = \frac{\exp(L_i(x))}{\sum_{i=1}^n \exp(L_i(x))} \quad 7$$

$$L_i(x) = \sum_{i=1}^n a_i(x) + a_0 = 0 \quad 7$$

درخت مدل پشتیبان الگوریتم یادگیری ماشین برای رده‌بندی است که از ترکیب الگوریتم‌های C4.5 و وایازی پشتیبان ساخته شد (خسروی و همکاران ۲۰۱۸)، و ساختاری درختی با تعدادی گره و لایه‌ی درونی دارد. درخت مدل پشتیبان الگوریتم CART را برای حذف درختان پیاده‌سازی می‌کند، و اعتبارسنجی متقابل را اجرا می‌کند تا تعداد تکرارهای الگوریتم را بررسی، و از برازش بیش‌ازحد جلوگیری کند (چن و همکاران ۲۰۱۷). درخت مدل پشتیبان C4.5 را در گره‌ها

آماري R اجرا شد.

کارکرد و دقت پیش‌بینی مدل‌های ترکیبی توسعه‌یافته باید با معیارهای آماری گوناگون ارزیابی شود (نقیبی و همکاران ۲۰۱۸، التوینسه و همکاران ۲۰۱۴). معیارهای آماری به‌کاررفته در این پژوهش نرخ پیش‌بینی‌شده‌ی مثبت (PPV)^۸، نرخ پیش‌بینی‌شده‌ی منفی (NPV)^۹، حساسیت (SST)^{۱۰}، دقت (ACC)^{۱۱} و ویژگی (SPC)^{۱۲} بود. همه‌ی این معیارها بر پایه‌ی چهار نوع پی‌آمد احتمالی شامل مثبت واقعی (TP)^{۱۳}، منفی واقعی (TN)^{۱۴}، مثبت کاذب (FP)^{۱۵} و منفی کاذب (FN)^{۱۶} محاسبه شد. TP نسبت تعداد چاه‌هایی است که به‌درستی چاهی با احتمال زیاد (با احتمال بیش‌تر از $L/S \cdot 0/5$) رده‌بندی می‌شود، در حالی که FP نسبت تعداد چاه‌هایی است که به‌نادرستی چاهی با احتمال زیاد رده‌بندی شد (با احتمال بیش‌تر از $L/S \cdot 0/5$). علاوه بر این، TN و FN تعداد چاه‌هایی است که به‌درستی و نادرستی چاه بی‌احتمال ناتوانمندی (با احتمال کم‌تر از $L/S \cdot 0/5$) رده‌بندی می‌شود (رابطه‌ی ۸ تا ۱۲) (ویتن و همکاران ۲۰۱۱، چاپی و همکاران ۲۰۱۷):

$$PPR = \frac{TP}{TP+FP} \quad 8$$

$$NPR = \frac{TN}{TN+FN} \quad 9$$

$$SST = \frac{TP}{TP+FN} \quad 10$$

$$SPC = \frac{TN}{TN+FP} \quad 11$$

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \quad 12$$

روش سطح زیر منحنی برای ارزیابی کارکرد مدل‌ها به‌ترتیب با نرخ‌های مثبت واقعی و مثبت کاذب بر محور Y و X به‌کار برده شد (تهرانی و همکاران ۲۰۱۳). این روش معیار سطح زیر منحنی را برای ارزیابی کمی به‌کار می‌برد، که ویژگی محور X را در برابر حساسیت محور Y نشان می‌دهد (چاپی و همکاران ۲۰۱۷). منحنی ROC کارآمدترین روش در پیش‌بینی سامانه‌ها

$P(N|x)$ احتمال پسین در گره برگ، و $L_i(x)$ کم‌ترین مربع‌های برازش است.

N تعداد معیارهای شرطی، و α_0 و α_i ضریب‌ها است. برای اجرای الگوریتم تقویت عمیق، Caret و Rweka در نرم‌افزار آماری R با در نظر گرفتن اعتبارسنجی متقاطع ۱۰ گانه اجرا شد.

درخت وایازی تقویت‌شده روشی ناسنجه‌ی است که از الگوریتم CART برای حل کردن دشواری‌های وایازی خطی بهره می‌برد (یوسف و همکاران ۲۰۱۶). در سه راه تحلیل الگوریتم CART ابتدا بیشینه‌ی درختان ساخته می‌شود، سپس با مرحله‌های هرس، بهترین درختان انتخاب می‌شود، و فرآیند اعتبارسنجی متقاطع ادامه می‌یابد. از مهم‌ترین مزیت‌های این روش آن است که می‌تواند داده‌های با حجم زیاد را با سرعت زیاد تحلیل کند، حساسیت کم‌تری به هم‌پوشانی با مدل‌های دیگر رده‌بندی دارد، نیاز به فرض توزیع داده‌ها ندارد، و می‌تواند مؤثرترین و مهم‌ترین عامل‌ها را در رده‌بندی تعیین کند (نقیبی و همکاران ۲۰۱۶). برای اجرای الگوریتم تقویت عمیق، بسته‌های caret و gbm در نرم‌افزار آماری R با در نظر گرفتن اعتبارسنجی متقاطع ۱۰ گانه اجرا شد.

جنگل تصادفی یکی از رایج‌ترین الگوریتم‌های به‌کار رفته برای بررسی دشواری‌های رده‌بندی و پیش‌بینی چندگانه است که حساسیت کمی به چندخطی بودن دارد، و نتیجه‌ی آن از نظر داده‌های ازدست‌رفته و نامتعادل بسیار پایدار است (رضوی ترمه و همکاران ۲۰۱۹). مدل پیش‌بینی‌کننده‌ی RF بر پایه‌ی میانگین‌گیری از نتیجه‌ی به‌دست‌آمده از همه‌ی درخت‌های تصمیم است، و برای بسیاری از مجموعه‌ی داده‌ها، رده‌بندی را با درستی بسیاری انجام می‌دهد (نقیبی و همکاران ۲۰۱۹). در این مدل، درختان تصادفی بردار ورودی را می‌گیرد، و آن را با هر درخت در جنگل رده‌بندی می‌کند. خروجی برچسب‌های رده‌ی است که از بیشینه‌ی رای‌های دریافت شده است. برای رده‌بندی یک شیء جدید بردار ورودی در انتهای هر درخت جنگل تصادفی گذاشته می‌شود، و هر درخت به یک رده‌بندی منجر می‌شود. جنگل به‌دست‌آمده از رده‌بندی که بیش‌ترین رای را در میان همه درخت‌های جنگل داشته باشد، انتخاب می‌شود. این مدل با بسته‌ی randomForest در نرم‌افزار

8 -Positive Predictive Rate

9 -Negative Predictive Rate

10 -Sensitivity

11 -Accuracy

12 -Specificity

13 -True positive

14 -True negative

15 -False positive

16 -False negative

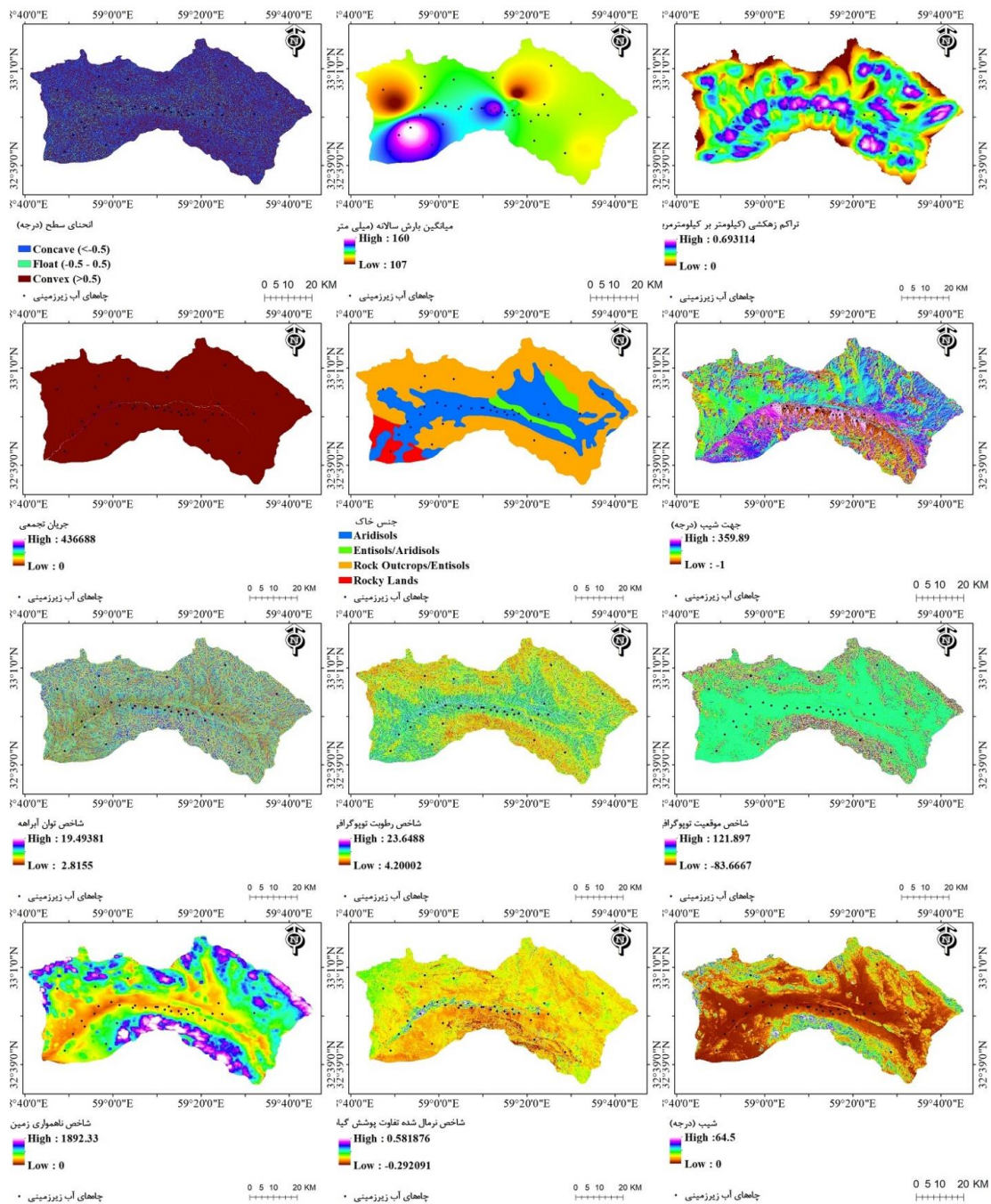
کاربرد الگوریتم‌های درخت و ایازی تقویت‌شده، درخت مدل پشتیبان...

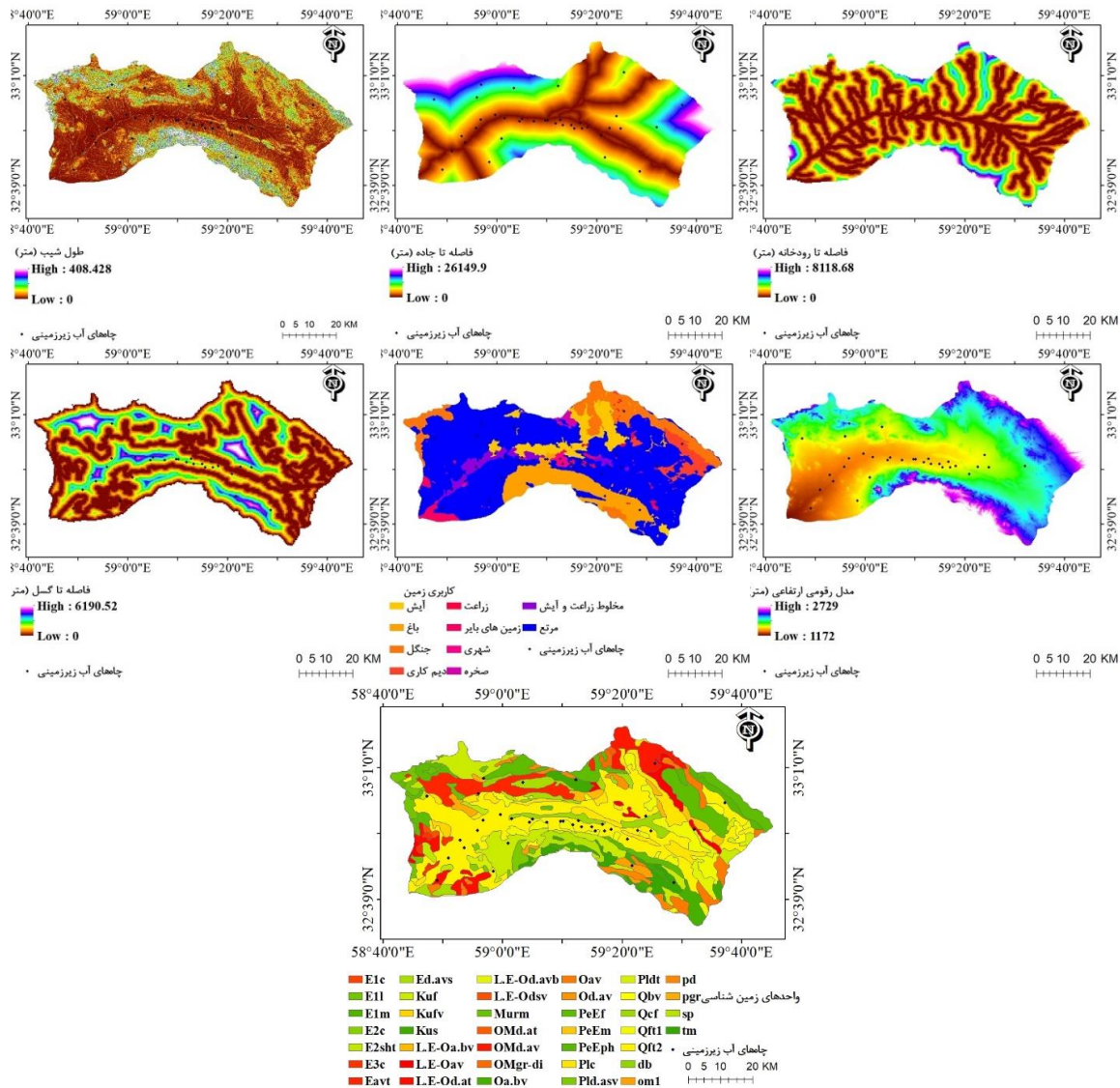
و زمین‌شناسی برای پیش‌بینی احتمال آب زیرزمینی به کار برده شد. برای تولید نقشه‌ها درون‌یابی کریجینگ به کار برده شد (شکل ۴)، زیرا خطای آن کمینه است (اسلامی‌نژاد و همکاران، ۲۰۲۱).

است، که اندازه‌ی دقت مدل را به‌شيوه‌ی کمی برآورد می‌کند. در این روش محدوده‌ی اندازه‌های سطح زیر منحنی ۰/۵ تا ۱ است، و هر چه به ۱ نزدیک‌تر باشد، دقت مدل بیش‌تر است.

نتایج و بحث

در این پژوهش ۱۹ معیار پستی‌بلندی، آب‌شناسی، محیطی





شکل ۴- معیارهای پستی بلندی، آب شناسی، محیطی و زمین شناسی مؤثر بر احتمال آب های زیرزمینی.

پراش (VIF) است. هنگامی که اندازه‌ی TOL^{17} بزرگ تر از $0/1$ و اندازه‌ی VIF^{18} کوچک تر از ۵ باشد، هم خطی چندگانه‌ی بسیاری در میان متغیرهای پیش بینی کننده نیست (جدول ۱) (افتخاری و همکاران b ۲۰۲۱).

نکته‌ی مهمی که باید پیش از پیاده‌سازی الگوریتم‌های یادگیری ماشین انجام شود بررسی مستقل بودن معیارها است. دو سنج‌ه‌ی آماری رایج برای تشخیص هم خطی چندگانه در میان عامل‌های گوناگون، آماره‌ی تحمل (TOL) و عامل تورم

17 - Tolerance
18 - Variance Inflation Factor

کاربرد الگوریتم‌های درخت وایازی تقویت‌شده، درخت مدل پشتیبان...

جدول ۲- مقایسه‌ی کارکرد الگوریتم‌های یادگیری ماشین با مجموعه داده‌های آموزشی و اعتبارسنجی.

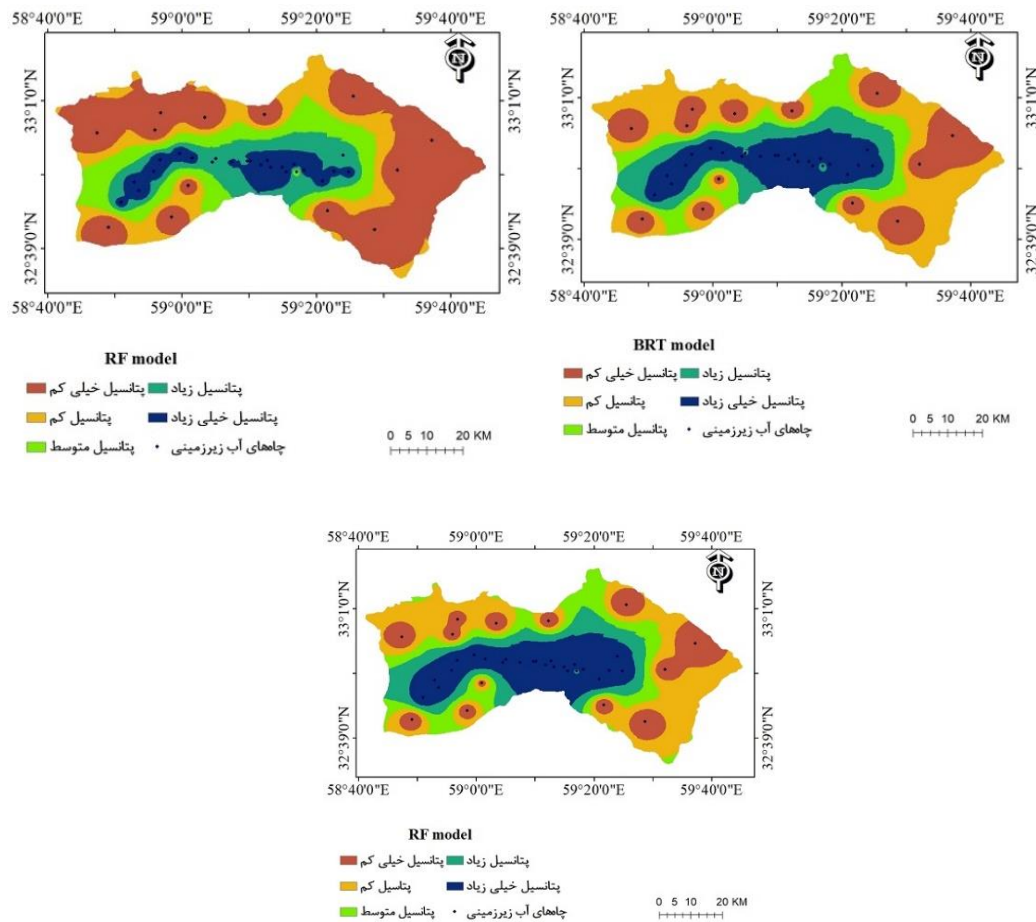
سنجه	LMT		BRT		RF	
	اعتبارسنجی آموزشی	اعتبارسنجی آموزشی	اعتبارسنجی آموزشی	اعتبارسنجی آموزشی	اعتبارسنجی آموزشی	اعتبارسنجی آموزشی
PPV	۰/۸۱۵	۰/۷۹۲	۰/۷۷۷	۰/۷۵۱	۰/۷۴۵	۰/۷۲۳
NPV	۰/۸۴۶	۰/۷۵۳	۰/۷۶۵	۰/۷۲۶	۰/۷۱۳	۰/۶۹۰
SST	۰/۷۸۲	۰/۷۶۱	۰/۷۷۳	۰/۷۶۲	۰/۷۶۵	۰/۷۴۳
SPC	۰/۸۹۲	۰/۸۸۴	۰/۸۱۷	۰/۷۸۲	۰/۸۰۶	۰/۷۷۵
ACC	۰/۸۳۸	۰/۸۱۲	۰/۸۱۶	۰/۷۹۲	۰/۸۰۱	۰/۷۸۵
AUC	۰/۸۹۷	۰/۸۶۵	۰/۸۳۲	۰/۸۱۲	۰/۷۹۵	۰/۷۸۸

جدول ۱- نتیجه‌ی آزمایش هم‌خطی چندگانه‌ی معیارهای به‌کار برده‌شده.

معیار	VIF	TOL
بلندی	۰/۲۱۵	۲/۲۳
شیب	۰/۱۶۳	۱/۳۱
جهت شیب	۰/۱۲۱	۳/۶۵
انحنای سطح	۰/۱۱۱	۲/۵۵
میانگین بارش سالانه	۰/۲۳۱	۴/۲۲
تراکم زه‌کشی	۰/۳۵۱	۴/۴۵
جریان تجمعی	۰/۳۳۳	۳/۳۴
جنس خاک	۰/۲۹۲	۳/۲۶
شاخص توان آبراه	۰/۶۲۳	۱/۲۸
شاخص رطوبت پستی-بلندی	۰/۶۶۶	۱/۳۹
شاخص موقعیت پستی-بلندی	۰/۵۳۲	۲/۶۶
شاخص ناهمواری زمین	۰/۸۱۲	۳/۳۲
شاخص بهنجارشده‌ی تفاوت پوشش گیاهی	۰/۶۱۲	۲/۹۸
طول شیب	۰/۱۹۸	۲/۷۵
فاصله تا جاده	۰/۲۲۱	۳/۶۵
فاصله تا رود	۰/۱۳۶	۱/۵۴
فاصله تا گسل	۰/۴۵۸	۱/۰۲
کاربری زمین	۰/۷۴۲	۱/۴۶
زمین‌شناسی	۰/۳۲۱	۴/۳۶

الگوریتم‌های یادگیری ماشین بر پایه‌ی اعتبارسنجی متقابل ۱۰ گانه بهینه شد. الگوریتم LMT با ۲۵ تکرار ساخته شد (اندازه‌های دقت و کاپا به ترتیب ۰/۸۱۲ و ۰/۷۵). الگوریتم BRT با تعداد ۲۵۰ درخت و ۱۵ گره بهینه‌سازی شد (دقت و کاپا به ترتیب ۰/۷۹۲ و ۰/۷۲). نتیجه‌ی الگوریتم RF نشان داد که با تعیین کردن ۱۰ متغیر در هر گره و ۸۰۰ درخت، اندازه‌های دقت و کاپا به ترتیب ۰/۷۸۵ و ۰/۷۰ بهینه‌سازی شد. نتیجه‌ی اعتبارسنجی الگوریتم‌های یادگیری ماشین در جدول ۲ نشان داده شده‌است. الگوریتم LMT کمی بهتر از BRT و RF احتمال آب‌های زیرزمینی را پیش‌بینی کرد. بر این پایه، برای رده‌بندی کردن جاهای توانمند احتمال آب زیرزمینی، روش LMT بیش‌ترین اندازه‌ی SST را برای مجموعه داده‌های آموزشی و اعتبارسنجی نشان داد (۰/۷۸۲ و ۰/۷۶۱). بنابراین، توان پیش‌بینی مدل ترکیبی LMT بیش‌تر از مدل‌های دیگر است، که به‌نادرستی چاه‌های آب زیرزمینی را با احتمال بیش از ۰/۵ L/S رده‌بندی کرد. روش LMT بهترین اندازه‌ی SPC را برای مجموعه داده‌های آموزشی و اعتبارسنجی نشان داد (۰/۸۹۲ و ۰/۸۸۴)، به این معنی که مدل ترکیبی LMT توانایی صحت‌سنجی و پیش‌بینی بهتری از دیگر مدل‌ها در رده‌بندی درست چاه‌های آب زیرزمینی با احتمال کم‌تر از ۰/۵ L/S در این منطقه داشت. علاوه بر این، بیش‌ترین نرخ پیش‌بینی مثبت و پیش‌بینی منفی برای مجموعه داده‌های آموزشی و اعتبارسنجی در مدل LMT، و پس از آن مدل‌های

BRT و RF بود. مدل LMT بیش‌ترین اندازه‌ی ACC را برای مجموعه داده‌های آموزشی و اعتبارسنجی داشت (۰/۸۳۸ و ۰/۸۱۲). برای این پایه می‌توان گفت که مدل LMT می‌تواند ۸۱٪ از چاه‌های آب زیرزمینی با احتمال بیش‌تر از ۰/۵ L/S، و کم‌تر از ۰/۵ L/S را در منطقه رده‌بندی کند. نتیجه‌ی مجموعه داده‌های آموزشی و اعتبارسنجی نشان داد که توانایی مدل LMT در پیش‌بینی احتمال آب زیرزمینی بهتر از دیگر مدل‌ها بود. برای مجموعه داده‌های آموزشی، بیش‌ترین اندازه‌ی سطح زیر منحنی (۰/۸۹۷) در مدل LMT، و پس از آن در BRT (۰/۸۳۲) و RF (۰/۷۹۵) AUC بود. برای مجموعه داده‌های اعتبارسنجی، مدل LMT قوی‌تر، و دقت آن (۰/۸۶۵) AUC بیش‌تر از مدل‌های BRT (۰/۸۱۲) AUC و RF (۰/۷۸۸) AUC بود. بنابراین، می‌توان نتیجه گرفت که دقت مدل LMT برای پیش‌بینی احتمال آب زیرزمینی در منطقه بهتر است. نتیجه‌ی به‌دست آمده از نقشه‌های احتمال آب‌های زیرزمینی با مدل‌های BRT، LMT و RF در شکل ۵ نشان داده شده‌است. مانند پژوهش پیشین روش رده‌بندی شکست طبیعی^{۱۹} برای رده‌بندی نقشه‌های احتمال آب زیرزمینی، در ۵ رده‌ی خیلی کم، کم، متوسط، زیاد و خیلی زیاد به‌کار برده شد (چن و همکاران ۲۰۲۰، کنعانی‌سادات و همکاران ۲۰۱۹، رحمتی و همکاران ۲۰۱۶).

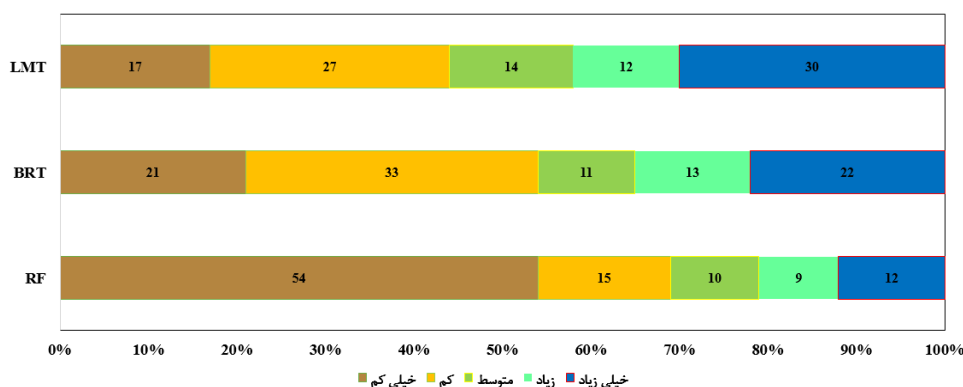


شکل ۵- نقشه‌ی پیش‌بینی منطقه‌های توانمند آب‌های زیرزمینی با الگوریتم‌های پیش‌رفته‌ی یادگیری ماشین.

کاربری زمین به کار برد. در شکل ۶٪ هر یک از رده‌های احتمال آب زیرزمینی پیش‌بینی شده با مدل‌های BRT، LMT و RF نشان داده شده است. در مدل LMT دو رده با احتمال خیلی زیاد و متوسط درصد بیش‌تری از رده‌های همسان در مدل‌های BRT و RF را می‌پوشانند. در مدل BRT دو رده با احتمال زیاد و کم درصد بیش‌تری از رده‌های همسان در مدل‌های LMT و RF دارد. علاوه بر این در مدل RF، رده‌ی با احتمال خیلی کم درصد بیش‌تری از رده‌های همسان در مدل‌های LMT و BRT داشت.

الگوریتم‌های یادگیری ماشین الگوی همسانی از احتمال آب‌های زیرزمینی نشان داد. با این حال، نقشه‌ی پیش‌بینی شده با الگوریتم LMT برای بهره‌برداری از آب‌های زیرزمینی پیشنهاد می‌شود، زیرا کارکرد برتری داشته است. این نقشه نشان می‌دهد که در نزدیک‌ترین جاها به رودها و منطقه‌های پیرامون آن‌ها احتمال بسیاری برای آب‌های زیرزمینی هست (شکل ۵). بخش‌های مرکزی منطقه‌ی پژوهش، به دلیل بلندی و زاویه‌ی شیب کم‌تر، و نزدیکی به رودها احتمال بیش‌تری برای آب‌های زیرزمینی دارد، که با یافته‌های نقیبه و همکاران (۲۰۱۸) و رضوی ترمه و همکاران (۲۰۱۹) هم‌آهنگی دارد. این یافته‌ها را می‌توان برای مدیریت‌کردن اندوخته‌های آب و

کاربرد الگوریتم‌های درخت وایازی تقویت‌شده، درخت مدل پشتیبان...



شکل ۶- درصد هر رده از نقشه‌های پیش‌بینی احتمال آب‌های زیرزمینی با الگوریتم‌های پیش‌رفته‌ی یادگیری ماشین.

تحقیق فلاح و همکاران (۲۰۱۷) هم‌سانی دارد. از میان همه‌ی الگوریتم‌های یادگیری ماشین، معیار NDVI مهم‌ترین معیار در پیش‌بینی احتمال آب زیرزمینی دشت بیرجند شناخته شد. معیار مهم دوم در الگوریتم LMT فاصله تا رودخانه (۹/۴۲) است. در حالی که معیار بلندی (۹/۸۶ و ۹/۲۹) به ترتیب در الگوریتم‌های BRT و RF دومین معیار مهم در مدل‌سازی شناخته شد، که با یافته‌های تحقیق نقیبی و همکاران (۲۰۱۸) هم‌آهنگی دارد. علاوه بر این، نتیجه‌ها نشان داد که معیارهای شیب، جنس خاک، تراکم زه‌کشی، فاصله تا گسل، و کاربری زمین تا حدودی می‌توانند در پیش‌بینی احتمال آب‌های زیرزمینی نقش داشته باشند. می‌توان نتیجه گرفت که تأثیر معیارهای فاصله تا جاده، جریان جمعی و انحنای سطح در میان الگوریتم‌های یادگیری ماشین برای پیش‌بینی احتمال آب‌های زیرزمینی کم‌اهمیت‌ترین بود.

جدول ۳ اهمیت معیارهای پستی‌بلندی، آب‌شناسی، محیطی و زمین‌شناسی را در فرآیند پیش‌بینی احتمال آب زیرزمینی دشت بیرجند نشان می‌دهد، که با الگوریتم‌های گوناگون یادگیری ماشین به دست آمد. اهمیت آن‌ها بر پایه‌ی نتیجه‌ی الگوریتم LMT رده‌بندی شد. بیش‌ترین اندازه‌ی دقت در میان شاخص‌ها در الگوریتم LMT بود. بررسی پدیده‌های سطحی و تأثیرگذار انسانی مانند آبیاری و کشاورزی ممکن است برای اطلاع یافتن از اوضاع اندوخته‌های زیرسطحی مؤثر باشد. از این رو می‌توان گفت که پوشش گیاهی مناسب سبب افزایش یافتن نفوذ آب‌های سطحی و تغذیه‌ی بیش‌تر آب‌های زیرزمینی، و پوشش گیاهی ضعیف سبب افزایش روان‌آب و کاهش نفوذ می‌شود. بنابراین یکی از دلایل‌هایی که معیار NDVI مهم‌ترین عامل در پیش‌بینی احتمال آب زیرزمینی دشت بیرجند شناخته شده است، همین است، که با نتیجه‌ی

جدول ۳- اهمیت معیارهای پستی‌بلندی، آب‌شناسی، محیطی و زمین‌شناسی در ارزیابی احتمال آب‌های زیرزمینی.

معیار	LMT (%)	RF (%)	BRT (%)
شاخص بهنجارشدگی تفاوت پوشش گیاهی	۲۰/۴۵	۲۰/۰۲	۲۹/۴۸
فاصله تا رود	۹/۴۲	۸/۰۶	۷/۶۲
بلندی	۸/۸۹	۹/۲۹	۹/۸۶
شیب	۷/۶۲	۸/۶۳	۸/۳۰
جنس خاک	۶/۸۷	۴/۸۴	۲/۴۹
فاصله تا گسل	۶/۹۲	۴/۶۵	۴/۰۶
تراکم زه‌کشی	۶/۹۱	۴/۲۲	۳/۱۸
کاربری زمین	۵/۹۶	۸/۱۲	۸/۶۸
شاخص توان آبراه	۵/۵۳	۴/۳۰	۳/۴۸
شاخص رطوبت پستی‌بلندی	۴/۹۸	۴/۲۳	۳/۵۹
بارندگی	۳/۶۰	۳/۵۹	۳/۵۹
سنگ‌شناسی	۲/۱۲	۲/۷۵	۳/۳۵
طول شیب	۱/۹۲	۲/۹۳	۲/۰۶
جهت شیب	۱/۸۰	۲/۴۲	۲/۸۶
شاخص موقعیت پستی‌بلندی	۱/۸۱	۳/۶۲	۲/۱۲
شاخص ناهمواری زمین	۱/۵۸	۳/۲۷	۲/۱۰
انحنای سطح	۱/۲۱	۱/۶۳	۱/۱۰
جریان جمعی	۱/۰۶	۲/۲۳	۱/۶۵
فاصله تا جاده	۱/۳۵	۱/۲۰	۰/۴۳

به روشی همسان، پژوهشگران در پژوهش‌های مکانی مانند حساسیت سیل و زمین‌لغزش نیز نشان دادند که الگوریتم LMT دقت بسیاری در مدل‌سازی دارد (فام و همکاران ۲۰۲۰، نهو و همکاران ۲۰۲۰). این نتیجه‌ها تأکید می‌کند که می‌توان از الگوریتم‌های پیش‌رفته‌ی یادگیری ماشین LMT و BRT در پژوهش‌های آب‌های زیرزمینی بهره برد تا مناطقی را که احتمال بسیاری در بهره‌برداری از آب‌های زیرزمینی آن‌ها هست شناسایی کند.

هدف دیگر این پژوهش بررسی اهمیت معیارهای پستی‌بلندی، آب‌شناسی، محیطی و زمین‌شناسی در ارزیابی احتمال آب‌های زیرزمینی بود. مشخص شد که NDVI مهم‌ترین معیار در احتمال‌یابی اندوخته‌های آب زیرزمینی در منطقه، و پس از آن فاصله تا رود، بلندی، و شیب است، که با نتیجه‌ی پژوهش‌های نقیبه و همکاران (۲۰۱۸) و فلاح و همکاران (۲۰۱۷) هم‌خوان است. در منطقه‌های دیگر نیز می‌توان این مدل را برای نقشه‌برداری دقیق و شناسایی صحیح منطقه‌های با احتمال آب زیرزمینی به کار برد، و به مدیریت مناسب آب زیرزمینی کمک کرد. پژوهشگران درجایی که به موقعیت چاه‌ها دسترسی ندارند می‌توانند روی کرده‌های دانش‌محور مانند فرآیند سلسله‌مراتب تحلیلی را به کار ببرند، و وزن عامل‌ها را برای تهیه‌ی نقشه‌های احتمال آب‌های زیرزمینی تعیین کنند.

نتیجه‌گیری

ارزیابی دقیق اندوخته‌های آب‌های زیرزمینی برای دست‌یابی به توسعه‌ی پایدار بسیار مهم است. با این حال، در کشورهای توسعه‌یافته اغلب کم‌بودی کلی در داده‌های دقیق آبی و زمین‌شناسی از آبخوان‌ها و بهره‌وری از آن‌ها هست. هدف اصلی این پژوهش دادن مدل‌های پیش‌رفته‌ی یادگیری ماشین برای پیش‌بینی احتمال آب زیرزمینی بود. بنابراین سه مدل یادگیری ماشین (RF، LMT، BRT) برای پیش‌بینی احتمال آب زیرزمینی در دشت بیرجند به کار برده شد. محققان توانایی و کاربرد روش‌های یادگیری ماشین را در پیش‌بینی مکانی احتمال آب‌های زیرزمینی منطقه‌های گوناگون اثبات کرده‌اند (رضوی ترمه و همکاران ۲۰۱۹، پراساد و همکاران ۲۰۲۰). در این پژوهش ۱۹ معیار آب‌شناسی، پستی‌بلندی، زمین‌شناسی و محیطی برای پیش‌بینی احتمال آب زیرزمینی به کار برده شد. نتیجه‌ی تجزیه و تحلیل نشان داد که همه‌ی مدل‌های ترکیبی جدید توسعه یافته توانایی پیش‌بینی خوبی نشان دادند، اما کارکرد مدل LMT در پیش‌بینی دقیق منطقه‌های توانمند آب زیرزمینی بهترین بود، و پس از آن مدل‌های RF و BRT برای نگاشت دقیق منطقه‌های توانمند آب زیرزمینی بود. دقت بیش‌تر در الگوریتم LMT را می‌توان با این واقعیت مرتبط دانست که از الگوریتم CART بهره می‌برد، و به هرس کردن درختان کمک می‌کند تا آن‌ها را از برآزش بیش‌از حد باز دارد (شومبواتنگ و همکاران ۲۰۱۲).

فهرست منابع

- Althuwaynee OF, Pradhan B, Park HJ, Lee JH. 2020. A novel ensemble bivariate statistical evidential belief functions with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. CATENA, 114 (2): 21–36.
- Benjmel K, Amraoui F, Boutaleb S, Ouchchen M, Tahiri A, Touab A. 2020. Mapping of groundwater potential zones in crystalline terrain using remote sensing, GIS techniques, and multicriteria data analysis (Case of the Ighrem Region, Western Anti-Atlas, Morocco). Water, 12 (1): 471.
- Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui D, Pham BT, Khosravi K. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental Modelling and Software, 95(1): 229–245.
- Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J. 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA, 151: 147–160.
- Chen W, Zhao X, Tsangaratos P, Shahabi H, Ilia I, Xue W, Ahmad BB. 2020. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. Journal of Hydrology, 583(1): 124602.
- Chezigi J, Pourghasemi HR, Naghibi, SA, Moradi HR, Kheirkhah Zarkesh M. 2016. As-

- assessment of a spatial multi-criteria evaluation to site selection underground dams in the Alborz Province, Iran. *Geocarto International*, 31 (1): 628–646.
- Corsini A, Cervi F, Ronchetti F. 2009. Weight of evidence and artificial neural networks for potential groundwater spring mapping: An application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology*, 111 (2): 79–87.
- Cořtache R, Arabameri A, Elkhrahy I, Ghorbanzadeh O, Pham QB. 2021. Detection of areas prone to flood risk using state-of-the-art machine learning models. *Geomatics, Natural Hazards and Risk*, 12(1): 1488–1507.
- Eftekhari M, Eslaminezhad S, Haji Elyasi A, Akbari M. 2021b. Development of DRAS-TIC model using artificial intelligence on the potential of aquifer contamination in semi-arid regions. *Iranian Journal of Ecohydrology*, 8(3): 651–665. (In Persian).
- Eftekhari M, Eslaminezhad SA, Akbari M, DadrasAjirlou Y, Elyasi AH. 2021a. Assessment of the potential of groundwater quality indicators by geostatistical methods in semi-arid regions. *Journal of Chinese Soil and Water Conservation*, 52(3): 158–167.
- Eftekhari M, Madadi K, Akbari M. 2019. Monitoring the fluctuations of the Birjand Plain aquifer using the GRACE satellite images and the GIS spatial analyses. *Watershed Management Research Journal*, 32(4): 51–65. (In Persian).
- Eslaminezhad SA, Omarzadeh D, Eftekhari M, Akbari M. 2021. Development of a data-driven model to predict landslide sensitive areas. *Geographia Technica*, 16(1): 97–112.
- Falah F, Daneshfar M, ghorbaninejad S. 2017. Application of the statistical index model in groundwater potential mapping in the Khorramabad Plain. *Journal of Water and Sustainable Development*, 4(1): 89–98.
- Hong H, Tsangaratos P, Ilia I, Liu J, Zhu AX, Chen W. 2018. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Science of the Total Environment*, 625 (2): 575–588.
- Jancewicz K, Migoń P, Kasprzak M. 2019. Connectivity patterns in contrasting types of tableland sandstone relief revealed by Topographic Wetness Index. *Science of the Total Environment*, 656 (2): 1046–1062.
- Kalantari Z, Ferreira CSS, Walsh RPD, Ferreira AJD, Deštouni G. 2017. Urbanization development under climate change: hydrological responses in a peri-urban Mediterranean catchment. *Land Degradation and Development*, 28 (7): 2207–2221.
- Kanani-Sadat Y, Arabsheibani R, Karimipour F, Nasser M. 2019. A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. *Journal of Hydrology*, 572 (4): 17–31.
- Khosravi K, Melesse AM, Shahabi H, Shirzadi A, Chapi K, Hong H. 2019. Flood susceptibility mapping at Ningdu Catchment, China using bivariate and data mining techniques. In *Extreme Hydrology and Climate Variability*; Elsevier: Amsterdam, The Netherlands, 419–434.
- Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR. 2016. A GIS-based flood susceptibility assessment and its mapping in Iran: A comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural Hazards*, 83(2): 947–987.
- Khosravi K, Pham BT, Chapi K, Shirzadi A,

- Shahabi H, Revhaug I, Prakash I, Tien Bui D. 2018. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment*, 627(1): 744–755.
- Kim JC, Jung HS, Lee S. 2019. Spatial mapping of the groundwater potential of the Geum River basin using ensemble models based on remote sensing images. *Remote Sensing*, 11 (2): 2285.
- Kumar A, Krishna AP. 2018. Assessment of groundwater potential zones in coal mining impacted hard-rock terrain of India by integrating geospatial and analytic hierarchy process (AHP) approach. *Geocarto International*, 33(2): 105–129.
- Lee S, Hyun Y, Lee S, Lee MJ. 2020. Groundwater potential mapping using remote sensing and GIS-based machine learning techniques. *Remote Sensing*, 12 (3): 1200–1213.
- Manap MA, Nampak H, Pradhan B, Lee S, Sulaiman WNA, Ramli, MF. 2014. Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arabian Journal of Geosciences*, 7(2): 711–724.
- Moghaddam DD, Rahmati O, Panahi M, Tufenbacher J, Darabi H, Haghizadeh A, Haghghi AT, Nalivan OA, Tien Bui D. 2020. The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. *Catena*, 187 (4): 104421.
- Naghibi SA, Dolatkordestani M, Rezaei A, Amouzegari P, Heravi MT, Kalantar B, Pradhan B. 2019. Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. *Environmental Monitoring and Assessment*, 191 (3): 1–20.
- Naghibi SA, Moradi Dashtpajardi M. 2017. Evaluation of four supervised learning methods for groundwater spring potential mapping in Khalkhal region (Iran) using GIS-based features. *Hydrogeology Journal*, 25 (4): 169–189.
- Naghibi SA, Pourghasemi HR, Abbaspour K. 2018. A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theoretical and Applied Climatology*, 131 (3): 967–984.
- Nhu VH, Shirzadi A, Shahabi H, Singh SK, Al-Ansari N, Clague JJ, Jaafari A, Chen W, Miraki S, Dou J. 2020. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *International Journal of Environmental Research and Public Health*, 17 (2): 2749.
- Pham BT, Van Phong T, Nguyen HD, Qi C, Al-Ansari N, Amini A, Ho LS, Tuyen TT, Yen HPH, Ly HB. 2020. A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes, and logistic model tree for flash flood susceptibility mapping. *Water*, 12 (1): 239.
- Prasad P, Loveson VJ, Kotha M, Yadav R. 2020. Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience & Remote Sensing*, 57(6): 735–752.

- Razavi-Termeh SV, Sadeghi-Niaraki A, Choi, SM. 2019. Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. *Water*, 11(5): 1596.
- Sahoo S, Munusamy SB, Dhar A, Kar A, Ram P. 2017. Appraising the accuracy of multi-class frequency ratio and weights of evidence method for delineation of regional groundwater potential zones in canal command system. *Water Resour. Manag.*, 31 (4): 4399–4413.
- Shoombuatong W, Hongjaisee S, Barin F, Chaijaruwanich J, Samleerat T. 2012. HIV-1 CRF01_AE coreceptor usage prediction using kernel methods based logistic model trees. *Computers in Biology and Medicine*, 42 (5): 885–889.
- Tehrany MS, Pradhan B, Jebur MN. 2013. Spatial prediction of flood susceptible areas using rule based Decision Tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, 504 (2):69–79.
- Witten IH, Frank E, Mark AH. 2011. Data mining: Practical machine learning tools and techniques. *Acm Sigmod Record*, 31 (1): 76-77.
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM. 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, 13 (3): 839–856.
- Yuan F, Bauer ME. 2007. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sensing of Environment*, 106 (4): 375– 386.



Watershed Management Research

VOL. 35, No.3, Ser. No: 136, Autumn 2022, pp. 44-59

DOI: 10.22092/WMRJ.2021.356517.1440

Research Paper



Using Boosted Regression Tree, Logistic Model Tree, and Random Forest Algorithms to Evaluate the Groundwater Potential

Seyed Ahmad Eslaminezhad

Master of Science (M.Sc.), Department of Surveying Engineering, Faculty of Surveying Engineering and Geospatial Information, University of Tehran, Tehran, Iran

Mobin Eftekhari

(Corresponding Author)* Master of Science (M.Sc.), Civil Engineering, Water and Hydraulic Structures, Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran

Mohammad Akbari

Associate Professor of Civil Engineering Department, University of Birjand

Hadi Bayat

Master of Science (M.Sc.), Department of Surveying Engineering, Faculty of Surveying Engineering and Geospatial Information, University of Tehran, Tehran, Iran

Wrya Barghi

Master of Science (M.Sc.), Department of Surveying Engineering, Faculty of Surveying Engineering and Geospatial Information, University of Tehran, Tehran, Iran

Corresponding Author's Email: Mobineftekhari@yahoo.com

Received: 30 October 2021

Accepted: 18 December 2021

Abstract

Groundwater is exploited uncontrollably due to population growth and industrialization in different parts of the world. The purpose of this study is to evaluate the groundwater potential by advanced machine learning algorithms using topographical, hydrological, environmental, and geological criteria. To do this, three advanced machine learning algorithms were used, including Boosted Regression Tree (BRT), Logistic Model Tree (LMT), and Random Forest (RF). Therefore, for implementation, geo-hydrological data of 37 groundwater wells in Birjand plain of South Khorasan province were collected and randomly selected in a ratio of 70 to 30 were divided into training and validation data sets. Finally, groundwater potential maps were prepared using BRT, LMT, and RF algorithms. In order to validate the groundwater potential prediction algorithms, the area under the curve (AUC) and the statistical criteria of positive predictive rate, negative predictive rate, sensitivity, specificity, and accuracy were used. The results showed that the LMT model (AUC = 0.865) has a better performance than the BRT and RF models in predicting the groundwater potential of the study area.

Keywords: Geographical information system (GIS), groundwater potential, machine learning