

Predicting Groundwater Potential Areas Using Hybrid Artificial Intelligence Methods (Case Study: Birjand Plain)

MOBIN EFTEKHARI¹, SEYED AHMAD ESLAMINEZHAD², ALI HAJI ELYASI³, MOHAMMAD AKBARI^{4*}

1. Department of Civil Engineering, Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

2. Department of Surveying Engineering, Faculty of Surveying Engineering and Spatial Information, University of Tehran, Tehran, Iran.

3. Department of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran.

4. Department of Civil Engineering, University of Birjand, Birjand, Iran.

(Received: June. 11, 2021- Revised: Sep. 4, 2021- Accepted: Sep. 25, 2021)

ABSTRACT

Groundwater is one of the most valuable resources for communities, agriculture, and industry. In the present study, three new artificial intelligence models, including Modified Real AdaBoost (MRAB), Bagging model (BA), and Rotation Forest model (RF), have been developed by the Functional Tree Base Classifier (FT) model to predict groundwater potential in Birjand plain area. Therefore, for implementation, geo-hydrological data of 37 groundwater wells and ten factors of topography, hydrology, and geology were used. The performance of these models was evaluated using the area under the curve (AUC) and other statistical indicators. The results showed that although all the hybrid models developed in this study increased the prediction accuracy, MRAB-FT model (AUC = 0.742) has higher accuracy in predicting potential groundwater areas in Birjand plain. Accurate mapping of groundwater potential areas while maintaining a balance between consumption and operation will help feed the aquifer for optimal use of groundwater resources.

Keywords: Groundwater Potential, Artificial Intelligence, Semi-Arid Areas.

پیش‌بینی مناطق بالقوه آب زیرزمینی با استفاده از روش‌های هوش مصنوعی ترکیبی (مطالعه موردی: دشت بیرجند)

مبین افتخاری^۱، سید احمد اسلامی‌نژاد^۲، علی حاجی الیاسی^۳، محمد اکبری^{۴*}

۱. مهندسی عمران آب و سازه‌های هیدرولیکی و عضو باشگاه پژوهشگران جوان و نخبگان، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران.

۲. گروه مهندسی نقشه‌برداری، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران، تهران، ایران.

۳. گروه مهندسی عمران، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.

۴. گروه مهندسی عمران، دانشگاه بیرجند، بیرجند، ایران.

(تاریخ دریافت: ۱۴۰۰/۳/۲۱ - تاریخ بازنگری: ۱۴۰۰/۶/۱۳ - تاریخ تصویب: ۱۴۰۰/۷/۳)

چکیده

آب‌های زیرزمینی یکی از مهم‌ترین منابع با ارزش برای استفاده جوامع، کشاورزی و صنایع هستند. در مطالعه حاضر، سه مدل هوش مصنوعی جدید شامل مدل آدابوست واقعی بهبود یافته (MRAB)، مدل بگینگ (BA) و مدل جنگل چرخشی (RF) توسط مدل طبقه‌بندی‌کننده پایه درخت عملکردی (FT) برای پیش‌بینی مناطق بالقوه آب‌های زیرزمینی در منطقه دشت بیرجند توسعه داده شده‌اند. لذا جهت پیاده‌سازی، داده‌های ژئوهیدرولوژیکی ۳۷ حلقه چاه آب زیرزمینی و ۱۰ عامل توپوگرافی، هیدرولوژی و زمین‌شناسی مورد استفاده قرار گرفت. عملکرد این مدل‌ها با استفاده از سطح زیر منحنی (AUC) و سایر شاخص‌های آماری مورد ارزیابی قرار گرفت. نتایج نشان داد که هر چند تمامی مدل‌های ترکیبی توسعه داده شده در این تحقیق دقت پیش‌بینی را افزایش دادند، اما مدل MRAB-FT (AUC=۰/۷۴۲) دقت بالاتری را در پیش‌بینی مناطق بالقوه آب‌های زیرزمینی در منطقه دشت بیرجند دارد. تهیه نقشه دقیق از مناطق بالقوه آب زیرزمینی، با حفظ تعادل بین مصرف و بهره‌برداری، به تغذیه مناسب آبخوان برای استفاده بهینه از منابع آب زیرزمینی کمک خواهد کرد.

واژه‌های کلیدی: پتانسیل آب زیرزمینی، هوش مصنوعی، مناطق نیمه خشک.

مقدمه

برای بهره‌برداری از آب‌های زیرزمینی استفاده می‌شوند. این روش‌ها مناسب هستند، اما نیازمند منابع، زمان و تخصص لازم و کافی هستند (Jha et al., 2010; Chen et al., 2020).

در سال‌های اخیر، روش‌های مبتنی بر سامانه اطلاعات جغرافیایی (GIS) و سنجش از دور^۲ (RS) برای تهیه نقشه پتانسیل آب‌های زیرزمینی مورد استفاده قرار گرفته‌اند. (Nampak et al., 2014) از تابع باور شواهد^۳ (EBF) و رگرسیون منطقی^۴ (LR) برای پیش‌بینی مکانی پتانسیل آب‌های زیرزمینی در حوضه لنگات کشور مالزی در بستر GIS استفاده کردند. جهت پیاده‌سازی روش‌های پیشنهادی، ۱۲ عامل توپوگرافی، هیدرولوژی و زمین‌شناسی شامل ارتفاع، زاویه شیب، انحنای سطح، شاخص قدرت جریان، شاخص رطوبت توپوگرافی، تراکم زهکشی، سنگ‌شناسی، تراکم خطواره، کاربری زمین، شاخص تغییرات پوشش گیاهی، جنس خاک و بارندگی استفاده شدند. سطح زیر منحنی^۵ (AUC) برای

با توجه به اثر تغییر آب‌وهوای جهانی، وابستگی به آب‌های سطحی می‌بایست کاهش پیدا کند و آب‌های زیرزمینی به طور سیستماتیک جهت استفاده بهینه در مصارف عمومی، کشاورزی و صنایع شناسایی شوند (Oh et al., 2011). پیش‌بینی پتانسیل آب‌های زیرزمینی می‌تواند به عنوان ابزاری جهت توسعه و برنامه‌ریزی سیستماتیک منابع آب تعریف شود (Díaz-Alcaide and Martínez-Santos, 2019). این رویکرد می‌تواند نواحی یک حوضه آبریز یا یک منطقه را براساس احتمال وقوع آب‌های زیرزمینی طبقه‌بندی کند (Jha et al., 2010). پیش‌بینی مناطق بالقوه آب‌های زیرزمینی در یک منطقه به ویژگی‌های سنگ‌ها، خاک‌ها و ساختارهای زمین‌شناسی آن منطقه بستگی دارد (Jha et al., 2010). تکنیک‌های حفاری زمین‌شناسی و اکتشافی قبلاً

* نویسنده مسئول: moakbari@birjand.ac.ir

1 Geographic information system
2 Remote sensing
3 Evidential belief function
4 Logistic regression
5 Area Under the Curve

قرار داد. هم‌چنین مقدار سطح زیر منحنی برای روش AHP، ۷۵/۴۵ به دست آمد که نشان‌دهنده دقت پیش‌بینی بالای این مدل در تعیین مناطق بالقوه آب‌های زیرزمینی در منطقه مورد مطالعه بود. (Chen *et al.* (2019). از ترکیب سه الگوریتم داده کاوی جنگل تصادفی (RAF)، نایو بیس^۹ (NB) و ماشین بردار پشتیبانی^{۱۰} (SVM) با الگوریتم ژنتیک^۸ (GA) جهت تهیه نقشه پتانسیل چشمه آب زیرزمینی در حوضه آبریز طالقان استان البرز از مدل آماری دو متغیره نسبت فرکانس^{۱۱} (FR) در بستر GIS استفاده کردند. جهت پیاده‌سازی از عوامل مؤثر بر آب زیرزمینی مانند زاویه شیب، جهت شیب، ارتفاع، شاخص رطوبت توپوگرافی، شاخص قدرت جریان، انحنا سطح، فاصله از رودخانه، فاصله از جاده، فاصله از گسل، سنگ‌شناسی، کاربری زمین و جنس خاک استفاده کردند. نتایج اعتبارسنجی نشان داد که مدل FR (AUC=۷۵/۹۹) دقت پیش‌بینی نسبتاً خوبی را جهت پیش‌بینی پتانسیل چشمه آب زیرزمینی در حوضه آبریز طالقان دارد. (Naghibi *et al.* (2016). از الگوریتم‌های جنگل تصادفی^۲ (RAF)، رگرسیون طبقه‌بندی درخت تصمیم^۳ (CART) و درخت رگرسیون تقویت شده^۴ (BRT) جهت تهیه نقشه‌های بالقوه چشمه‌های آب زیرزمینی در حوضه آبریز کوه‌رنگ استان چهارمحال بختیاری استفاده کردند. جهت پیاده‌سازی روش‌های پیشنهادی، عوامل مؤثر بر پتانسیل آب زیرزمینی شامل زاویه شیب، جهت شیب، ارتفاع، شاخص رطوبت توپوگرافی، انحنا سطح، انحنا نیم‌رخ، فاصله از رودخانه، فاصله از گسل، سنگ‌شناسی، کاربری زمین، تراکم زهکشی و تراکم گسل استخراج شدند. سطح زیر منحنی برای مدل‌های BRT و CART و RF به ترتیب ۰/۸۱۰۳، ۰/۷۸۷۰ و ۰/۷۱۱۹ محاسبه شد که نشان‌دهنده دقت بالای مدل BRT در پیش‌بینی مناطق بالقوه چشمه‌های آب زیرزمینی در منطقه مورد مطالعه بود. هدف از مطالعه Kumar and Krishna (2018) تعیین مناطق بالقوه آب‌های زیرزمینی با استفاده از فرآیند سلسله مراتبی تحلیلی (AHP)^۵ در منطقه رامگار کشور هند بود. نقشه‌های نهایی پتانسیل آب‌های زیرزمینی در در منطقه مورد مطالعه در ۵ دسته پتانسیل عالی، بسیار خوب، خوب، متوسط و ضعیف طبقه‌بندی شدند. نتایج مطالعه نشان داد که مناطق با پتانسیل عالی، بسیار خوب و خوب به ترتیب ۱۴۸/۳، ۳۷۳/۶۶ و ۴۳۸/۸۶ کیلومتر مربع از منطقه مورد مطالعه را پوشش داد، در حالی که منطقه با پتانسیل آب زیرزمینی ضعیف، ۱۸۰/۰۵ کیلومتر مربع را پوشش

رخ پیش‌بینی روش‌های EBF و LR به ترتیب ۷۸٪ و ۷۲٪ محاسبه شد که نشان‌دهنده دقت و کارایی بالای روش EBF در پیش‌بینی مکانی پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه بود. (Moghaddam *et al.* (2015). جهت تهیه نقشه پتانسیل چشمه آب زیرزمینی در حوضه آبریز طالقان استان البرز از مدل آماری دو متغیره نسبت فرکانس^{۱۱} (FR) در بستر GIS استفاده کردند. جهت پیاده‌سازی از عوامل مؤثر بر آب زیرزمینی مانند زاویه شیب، جهت شیب، ارتفاع، شاخص رطوبت توپوگرافی، شاخص قدرت جریان، انحنا سطح، فاصله از رودخانه، فاصله از جاده، فاصله از گسل، سنگ‌شناسی، کاربری زمین و جنس خاک استفاده کردند. نتایج اعتبارسنجی نشان داد که مدل FR (AUC=۷۵/۹۹) دقت پیش‌بینی نسبتاً خوبی را جهت پیش‌بینی پتانسیل چشمه آب زیرزمینی در حوضه آبریز طالقان دارد. (Naghibi *et al.* (2016). از الگوریتم‌های جنگل تصادفی^۲ (RAF)، رگرسیون طبقه‌بندی درخت تصمیم^۳ (CART) و درخت رگرسیون تقویت شده^۴ (BRT) جهت تهیه نقشه‌های بالقوه چشمه‌های آب زیرزمینی در حوضه آبریز کوه‌رنگ استان چهارمحال بختیاری استفاده کردند. جهت پیاده‌سازی روش‌های پیشنهادی، عوامل مؤثر بر پتانسیل آب زیرزمینی شامل زاویه شیب، جهت شیب، ارتفاع، شاخص رطوبت توپوگرافی، انحنا سطح، انحنا نیم‌رخ، فاصله از رودخانه، فاصله از گسل، سنگ‌شناسی، کاربری زمین، تراکم زهکشی و تراکم گسل استخراج شدند. سطح زیر منحنی برای مدل‌های BRT و CART و RF به ترتیب ۰/۸۱۰۳، ۰/۷۸۷۰ و ۰/۷۱۱۹ محاسبه شد که نشان‌دهنده دقت بالای مدل BRT در پیش‌بینی مناطق بالقوه چشمه‌های آب زیرزمینی در منطقه مورد مطالعه بود. هدف از مطالعه Kumar and Krishna (2018) تعیین مناطق بالقوه آب‌های زیرزمینی با استفاده از فرآیند سلسله مراتبی تحلیلی (AHP)^۵ در منطقه رامگار کشور هند بود. نقشه‌های نهایی پتانسیل آب‌های زیرزمینی در در منطقه مورد مطالعه در ۵ دسته پتانسیل عالی، بسیار خوب، خوب، متوسط و ضعیف طبقه‌بندی شدند. نتایج مطالعه نشان داد که مناطق با پتانسیل عالی، بسیار خوب و خوب به ترتیب ۱۴۸/۳، ۳۷۳/۶۶ و ۴۳۸/۸۶ کیلومتر مربع از منطقه مورد مطالعه را پوشش داد، در حالی که منطقه با پتانسیل آب زیرزمینی ضعیف، ۱۸۰/۰۵ کیلومتر مربع را پوشش

9 AdaBoost

10 Bagging

11 Random subspace

12 Dagging

13 Rotation forest

1 Frequency ratio

2 Random Forest

3 Classification and Regression Tree

4 boosted regression tree

5 Analytic hierarchy process

6 Naïve byes

7 Support vector machine

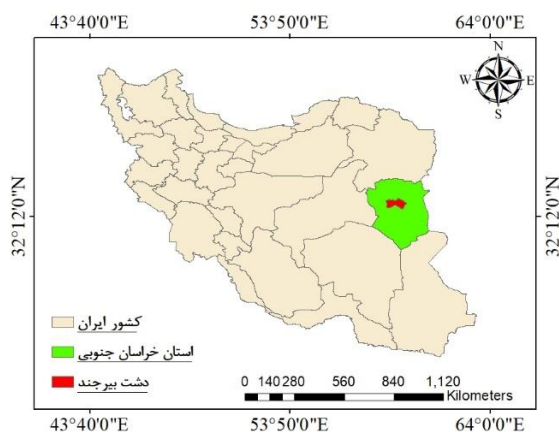
8 Genetic Algorithm

استفاده قرار گیرند، مطلوب است.

مواد و روش‌ها

شرح ناحیه مورد مطالعه

دشت بیرجند، یکی از محدوده‌های مطالعاتی حوضه آبریز کویر لوت می‌باشد که مختصات جغرافیایی آن بین طول‌های 58° و 41° تا 59° و $46'$ شرقی و عرض‌های 32° و $33'$ تا $8'$ شمالی قرار گرفته است. مساحت کلی محدوده مطالعاتی بیرجند $3406/72 \text{ km}^2$ است که از این مقدار سهم دشت و ارتفاعات به- ترتیب $900/63 \text{ km}^2$ و $2506/09 \text{ km}^2$ می‌باشد. رودخانه شاهرود به طول 108 کیلومتر از شرق به غرب، دشت را زهکشی کرده و به کویر لوت می‌ریزد. این رودخانه در تمام سال به جز مواقع سیلابی خشک است، لذا اکثر آب موردنیاز کشاورزی، صنعت و شرب شهرستان بیرجند از سفره آب زیرزمینی تأمین می‌شود. حداکثر ارتفاع حوضه از سطح دریا 2729 متر در ارتفاعات باقران و حداقل ارتفاع نیز 1172 متر در خروجی دشت (منطقه فدشک) می‌باشد. منطقه مورد مطالعه در شکل (۱) نشان داده شده است.



شکل ۱- موقعیت کلی دشت بیرجند

عوامل موثر بر پتانسیل آب‌های زیرزمینی

عوامل توپوگرافی

عوامل توپوگرافی مانند زاویه شیب^۸، جهت شیب^۹، ارتفاع^{۱۰} و انحنای سطح^{۱۱} عوامل مهمی برای پیش‌بینی پتانسیل آب‌های زیرزمینی هستند. چون این عوامل بر تجمع آب سطحی زمین و در نتیجه نفوذ آب در زمین تأثیر می‌گذارند. در این مطالعه، برای تهیه مدل رقومی ارتفاع زمین^{۱۲} (DEM) با قدرت تفکیک 30 متر از تصاویر ماهواره‌ای SRTM^{۱۳} استفاده شد و سپس لایه‌های زاویه

اگرچه تعدادی از الگوریتم‌های ترکیبی یادگیری ماشین برای پیش‌بینی پتانسیل آب‌های زیرزمینی استفاده شده‌اند، اما هیچ روشی در پیش‌بینی دقیق پتانسیل آب زیرزمینی کامل نیست. بر اساس نتایج تحقیقات پیشین، روش‌های هوش مصنوعی AB، BA و RF دارای دقت و کارایی بالایی در پیش‌بینی مناطق بالقوه آب‌های زیرزمینی هستند. مدل آدابوست واقعی بهبود یافته^۱ (MRAB)، بهبود یافته و اصلاح‌شده مدل آدابوست است که قابلیت فراوانی در یادگیری ماشین به‌منظور افزایش توانایی پیش‌بینی الگوریتم‌های طبقه‌بندی‌کننده دارد (Freund and Schapire, 1997). هم‌چنین مزیت مدل BA این است که از طریق نمونه‌گیری با جایگزین کردن داده‌های آموزشی، دقت تخمین‌های مورد استفاده را افزایش می‌دهد (Breiman, 1996). علاوه بر این مزیت اصلی الگوریتم RF این است که هم تغییرپذیری در داده‌ها و هم دقت خوشه‌بندی به‌طور هم‌زمان توسط این الگوریتم افزایش می‌یابد (Xia et al., 2013). هدف از این پژوهش تهیه نقشه مناطق بالقوه آب‌های زیرزمینی با استفاده از روش‌های نوین یادگیری ماشین مبتنی بر مدل طبقه‌بندی‌کننده پایه درخت عملکردی (FT) است. بنابراین، این مطالعه از مطالعات قبلی متمایز است، زیرا در این مطالعه از مدل‌های ترکیبی یادگیری ماشین مبتنی بر درخت یعنی MRAB-FT، BA-FT و RF-FT برای پیش‌بینی مناطق بالقوه آب‌های زیرزمینی در منطقه دشت بیرجند استفاده شده است که نوآوری تحقیق حاضر نیز می‌باشد. دلیل استفاده از الگوریتم FT این است که الگوریتم موردنظر یک درخت چندمتغیره است و از یک تابع رگرسیون لجستیک برای تخصیص وزن به نمونه‌های مجموعه داده آموزشی و پیش‌بینی برگ‌های تابعی استفاده می‌کند که این موضوع سبب کاهش خطا و بهبود دقت مدل‌های MRAB، BA و RF خواهد شد. در نهایت، معیار سطح زیر منحنی (AUC) و معیارهای آماری شامل مقدار پیش‌بینی‌شده مثبت^۲ (PPV)، مقدار پیش‌بینی‌شده منفی^۳ (NPV)، حساسیت^۴ (SST)، دقت^۵ (ACC)، تشخیص‌پذیری یا ویژگی^۶ (SPC)، کاپا (Kappa) و خطای مربع میانگین ریشه^۷ (RMSE) جهت بررسی دقت و کارایی مدل‌های پیشنهادی به کار گرفته شدند.

8 Slope
9 Aspect
10 Elevation
11 Plane Curvature
12 Digital elevation model
13 Shuttle Radar Topography Mission

1 Modified Real AdaBoost
2 Positive Predictive Value
3 Negative Predictive Value
4 Sensitivity
5 Accuracy
6 Specificity
7 Root mean square error

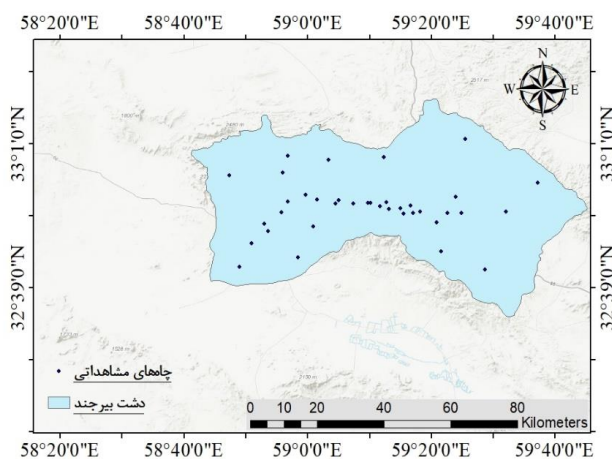
زبر شیبست متورق تقسیم‌بندی شده است که گروه اول نفوذپذیری بالای آن دارای بیشترین امتیاز و گروه چهار کمترین امتیاز تعلق می‌گیرد. نقشه سنگ‌شناسی منطقه مورد مطالعه با استفاده از نقشه زمین‌شناسی کشور در مقیاس ۱:۱۰۰۰۰۰ تهیه شد. جهت تهیه نقشه فاصله تا گسل (استخراج از بانک اطلاعاتی گسل‌های کشور) نیز از ابزار فاصله اقلیدسی نرم‌افزار ArcMap استفاده شد. همچنین نقشه جنس خاک منطقه مورد مطالعه در مقیاس ۱:۲۵۰۰۰۰ از Google Earth Engine اخذ گردید.

کاربری زمین

کاربری‌های مختلف می‌توانند در میزان نفوذ آب، نفوذ، فرسایش خاک و تبخیر و تعرق دخیل باشند. برای مثال در کاربری‌های شهری تقریباً هیچ نوع نفوذی وجود ندارد (به دلیل آسفالت خیابان‌ها و ...). در حالی که مراتع بسیار خوب می‌توانند موجب نفوذ مقادیر زیاد آب در زمین شوند. از تصویر ETM⁺ برای تهیه نقشه کاربری زمین منطقه مورد مطالعه استفاده شد.

نقشه موجودی چاه‌های آب زیرزمینی

در این مطالعه، ۳۷ نمونه آب زیرزمینی از چاه‌ها و پیزومترها در طول سال ۱۳۹۳ تا ۱۳۹۸ از طریق شرکت سهامی آب منطقه ای استان خراسان جنوبی جمع‌آوری شدند که موقعیت ۳۷ نمونه مورد نظر در شکل (۲) نشان داده شده است.



شکل ۲- موقعیت چاه‌های مشاهداتی در منطقه مورد مطالعه

آزمایش پمپاژ واقعی است که در محل برای ارزیابی پتانسیل آب زیرزمینی توسط شرکت سهامی آب منطقه ای استان خراسان جنوبی انجام شده است.

روش تحقیق

روند اجرایی این تحقیق در شکل (۳) نشان داده شده است.

شیب، جهت شیب و انحنای سطح از لایه DEM ایجاد شدند.

عوامل هیدرولوژی

عوامل هیدرولوژیکی مانند بارش و فاصله تا رودخانه نقش مهمی در پیش‌بینی پتانسیل آب‌های زیرزمینی دارند. در این مطالعه، نقشه بارش (بر حسب میلی‌متر) با استفاده از داده‌های متوسط بارندگی سالانه از ایستگاه‌های بارندگی واقع در بخشی از دشت بیرجند ایجاد شد. لایه رودخانه نیز طبق بانک اطلاعاتی رودخانه های کشور استخراج شد و سپس به کمک ابزار فاصله اقلیدسی در نرم‌افزار Arc Map رستری شد.

عوامل زمین‌شناسی

زمین‌شناسی یکی از عوامل مهم در تشکیل آبخوان‌ها می‌باشد. وجود جریان آب‌های زیرزمینی به ویژگی‌های واحدهای سنگی مانند تخلخل و نفوذپذیری سنگ بستگی دارد (Chen et al. 2019; Guru et al. 2017; Miraki et al. 2019; Park et al. 2014). در این مطالعه، نقشه‌های موضوعی مرتبط با زمین‌شناسی یعنی نقشه سنگ‌شناسی، جنس خاک و نقشه فاصله تا گسل برای پیش‌بینی پتانسیل آب‌های زیرزمینی استفاده شدند. نقشه سنگ‌شناسی منطقه بر اساس میزان تخلخل و نفوذپذیری به چهار گروه اصلی شامل ۱- آبرفتی/شن و ماسه بادی، ۲- کوارتزیت آب وهوای معتدل، ۳- کوارتزیت شیبست، ۴- گرانیت سنگ خارا

جهت تجزیه و تحلیل، داده‌های چاه به نسبت ۷۰٪ برای آموزش و ۳۰٪ برای آزمایش تقسیم شدند. در مطالعه حاضر، از بازدهی آب زیرزمینی ۰/۵ لیتر بر ثانیه به عنوان یک مقدار آستانه برای انتخاب آب زیرزمینی استفاده شده است. این مقدار آستانه بر اساس شرایط محلی آب زیرزمینی و نظرات کارشناسان انتخاب شده است. داده‌های عملکرد آب زیرزمینی این چاه‌ها بر اساس

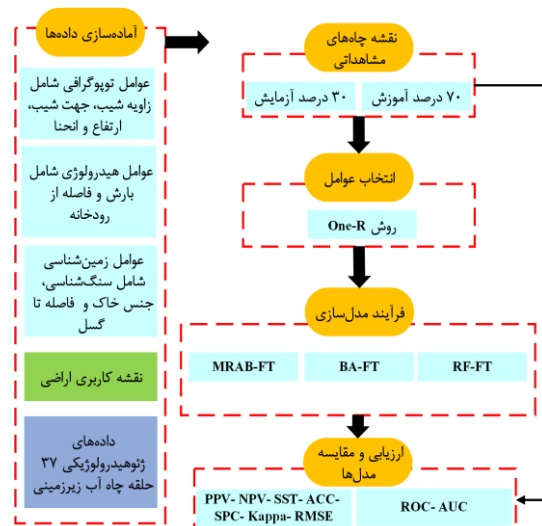
(Schapire, 1997). در مدل MRAB، پس از نمونه برداری نقاط آموزشی، یک پایگاه داده آموزشی برای یادگیری مدل ایجاد می‌شود (Freund et al., 1999; Friedman et al., 2000). در مرحله بعد، با استفاده از یک الگوریتم پایه (در این مطالعه الگوریتم FT)، یک وزن یا احتمال به نمونه‌هایی در مجموعه داده‌های آموزشی اختصاص داده می‌شود. مدل MRAB مجدداً وزن‌های اختصاص داده شده را برای نمونه‌های مختلف مجموعه داده‌های آموزشی که به چندین مجموعه داده آموزشی فرعی تقسیم شده‌اند، وزن‌گذاری می‌کند تا خطای مدل را کاهش دهد و مدلی با دقت پیش‌بینی بالاتر را براساس اکثریت آرا انتخاب کند (Freund et al., 1999).

مدل بگینگ (BA)

مدل BA براساس مفاهیم ترکیب^۲ است که جهت ارزیابی دقت تخمین‌های مورد استفاده در روش‌های داده‌کاوی از طریق نمونه‌گیری با جایگزین کردن داده‌های آموزشی به کار گرفته می‌شود (Breiman, 1996). در این روش فرض بر این است که مجموعه داده آموزشی نمونه‌ای از ارتباطات مورد مطالعه است و انواع سناریوهای قابل تحقق جامعه را می‌توان از این مجموعه داده شبیه‌سازی کرد. هنگامی که یک نمونه جدید وارد هر یک از خوشه‌ها می‌شود، از توافق اکثریت برای شناسایی کلاس موردنظر استفاده می‌شود. در واقع مدل BA از خوشه‌های مشابه استفاده کرده و نتایج آن‌ها را با یک رأی اکثریت ساده ترکیب می‌کند. در مدل BA، هنگامی که خوشه‌های پایه بخشی از الگوریتم‌های یادگیری ناپایدار هستند (مانند درخت تصمیم یا شبکه عصبی)، نتایج خوبی حاصل می‌شود، به طوری که تغییرات کوچک در داده‌های تجربی منجر به تغییرات عمده در مدل ایجاد شده توسط آن الگوریتم می‌شود (He et al., 2019; Hong et al., 2018).

مدل جنگل چرخشی (RF)

در مدل RF، ویژگی‌ها با استفاده از الگوریتم تحلیل مؤلفه‌های اصلی^۳ (PCA) برای آموزش هر یک از خوشه‌های اصلی چرخش می‌کنند (Rodriguez et al., 2006). به طور خاص، برای ساخت مجموعه داده آموزشی برای هر یک از خوشه‌های پایه، مجموعه ویژگی F به K زیرمجموعه طبقه‌بندی می‌شود و الگوریتم PCA برای هر یک از زیرمجموعه‌ها اعمال می‌شود. تمام مؤلفه‌های اصلی در مجموعه داده باقی می‌مانند تا تنوع اطلاعات حفظ شود. بنابراین، K محور چرخش قرار داده شده و به عنوان ویژگی‌های جدید برای آموزش خوشه‌بندی پایه استفاده می‌شود (Xia et al.,



شکل ۳- فرآیند اجرایی تحقیق

طبق شکل (۳)، در این مطالعه، داده‌های ژئوهیدرولوژیکی ۳۷ حلقه چاه آب زیرزمینی و ۱۰ عامل توپوگرافی، هیدرولوژی و زمین‌شناسی منطقه مورد بررسی قرار گرفته است. در این مطالعه، بازدهی آب زیرزمینی بیش از ۰/۵ لیتر بر ثانیه به عنوان یک آستانه برای تعریف دو کلاس (آب زیرزمینی و غیر آب زیرزمینی) جهت انتخاب و ساخت مجموعه داده‌های آموزشی (۷۰ درصد) و آزمایشی (۳۰ درصد) استفاده شد. سپس روش One-R برای انتخاب مهم‌ترین عوامل به عنوان پارامترهای ورودی در مدل‌سازی استفاده شد. روش One-R یک الگوریتم ساده است که از همبستگی‌های آماری بین یک متغیر خروجی و مجموعه‌ای از عوامل ورودی انتخابی استفاده می‌کند تا اهمیت عوامل را براساس متوسط شایستگی^۱ (AM) رتبه‌بندی می‌کند (Holte, 1993). در نهایت روش FT به عنوان یک طبقه‌بندی‌کننده پایه با مدل‌های یادگیری ماشین MRAB، BA و RF ترکیب شد. توانایی مدل‌ها با استفاده از مجموعه داده‌های آموزشی (عملکرد / خوبی برازش) و داده‌های آزمایشی (دقت پیش‌بینی) و توسط چندین شاخص آماری به نام‌های PPV، NPV، SST، ACC، SPC، Kappa و RMSE ارزیابی شدند. در نهایت نقشه‌های تولید شده با استفاده از سطح زیر منحنی (AUC) مورد ارزیابی قرار گرفتند. جهت پیاده‌سازی و تهیه نقشه‌های خروجی از نرم‌افزارهای Matlab، ArcMap و Google Earth Engine استفاده شده است.

مدل آدابوست واقعی بهبود یافته (MRAB)

مدل MRAB بهبود یافته و اصلاح‌شده مدل آدابوست است که قابلیت فراوانی در یادگیری ماشین به منظور افزایش توانایی پیش‌بینی الگوریتم‌های طبقه‌بندی‌کننده دارد (Freund and

از اهمیت ویژه‌ای برخوردار است. در مطالعه حاضر، از روش ORAE^۱ استفاده شده است. این رویکرد اهمیت عوامل را براساس متوسط شایستگی^۲ (AM) رتبه‌بندی می‌کند. روش One-R یک الگوریتم ساده است که از همبستگی‌های آماری بین یک متغیر خروجی و مجموعه‌ای از عوامل ورودی انتخابی استفاده می‌کند (Holte, 1993). این روش براساس یک قانون برای هر ویژگی در مجموعه داده آموزشی انجام می‌شود. زمانی که خطا کوچک‌تر باشد، برای فرآیند مدل‌سازی بهتر است. در این روش، تمام عوامل به طور مستقل با توجه به اهمیت آن‌ها برای توسعه مدل‌ها طبقه‌بندی می‌شوند.

روش‌های اعتبارسنجی

در این مطالعه، ۷۰٪ از چاه‌های مشاهداتی برای آموزش مدل‌ها و ۳۰٪ مابقی برای آزمایش مدل‌ها استفاده شدند. عملکرد و دقت پیش‌بینی مدل‌های ترکیبی توسعه‌یافته باید با استفاده از معیارهای آماری مختلف ارزیابی شود (Abedini et al., 2018; Nguyen et al., 2019; Shirzadi et al., 2017; Pham et al., 2019). معیارهای آماری مورد استفاده در این مطالعه عبارتند از: PPV, NPV, SST, SPC, ACC, Kappa و RMSE. در این مطالعه، PPV احتمال این است که یک پیکسل دارای پتانسیل آب زیرزمینی بیش از ۰/۵ لیتر بر ثانیه به نام "آب زیرزمینی" باشد که نتایج آزمایش مثبت هستند؛ در حالی که، NPV احتمال این است که یک پیکسل دارای پتانسیل آب زیرزمینی کم‌تر از ۰/۵ لیتر بر ثانیه به نام "غیر آب زیرزمینی" باشد که نتایج آزمایش منفی هستند. SST نشان‌دهنده نسبت چاه‌های آب زیرزمینی (با پتانسیل بیش از ۰/۵ لیتر بر ثانیه) است که به درستی طبقه‌بندی شده‌اند، در حالی که SPC نشان‌دهنده نسبت دسته‌بندی صحیح چاه‌های غیر آب زیرزمینی است. شاخص ACC بیانگر نسبت چاه‌های آب زیرزمینی با پتانسیل ۰/۵ لیتر بر ثانیه و چاه‌های غیر پتانسیل (پیکسل‌های آب زیرزمینی کمتر از ۰/۵ لیتر بر ثانیه) می‌باشد. شاخص Kappa قابلیت اطمینان هر مدل را ارزیابی می‌کند. RSME تفاوت بین مقادیر پیش‌بینی شده و واقعی چاه‌های آب زیرزمینی می‌باشد که مقادیر بالاتر نشان‌دهنده عملکرد بالاتر مدل‌ها می‌باشد. همه این معیارها براساس ماتریس درهم‌ریختگی^۳ حاصل می‌شوند که متشکل از چهار نوع پیامد احتمالی شامل مثبت واقعی^۴ (TP)، منفی واقعی^۵ (TN)، مثبت کاذب^۶ (FP) و منفی اشتباه^۷ (FN) برای محاسبه معیارهای آماری است. TP نسبت تعداد چاه‌هایی است که به درستی به عنوان یک چاه با پتانسیل بالا طبقه‌بندی می‌شوند

مزیتم الگوریتم RF این است که هم تغییرپذیری در داده‌ها و هم دقت خوشه‌بندی به طور همزمان توسط این الگوریتم افزایش می‌یابد. استخراج ویژگی برای هر یک از خوشه‌های اصلی، تنوع داده‌ها و حفظ تمام مؤلفه‌های اصلی دقت خوشه‌بندی را افزایش می‌دهد (Xia et al., 2013). روند اصلی ساخت یک مدل RF به شرح زیر است: (۱) ویژگی‌های مجموعه داده آموزشی به طور تصادفی به K زیرمجموعه تقسیم می‌شوند، (۲) الگوریتم PCA به هر زیرمجموعه اعمال می‌شود، (۳) ماتریس چرخش با توجه به توالی ویژگی‌های اصلی مجدداً تراز می‌شود، (۴) الگوریتم طبقه‌بندی‌کننده FT براساس مجموعه داده‌های دوران یافته آموزش داده می‌شود، (۵) نتایج طبقه‌بندی‌کننده‌های مختلف با یکدیگر ترکیب می‌شوند (۶) به هر پیکسل از مجموعه داده آموزشی یک برچسب کلاس اختصاص داده می‌شود و سپس تمام پیکسل‌های منطقه آموزش داده می‌شوند.

درخت عملکردی (FT)

الگوریتم طبقه‌بندی‌کننده پایه FT یک درخت چندمتغیره است که از ترکیبی از صفات در برگ و/یا در گره‌های داخلی استفاده می‌کند و همچنین می‌تواند برای ساخت درخت‌های چندمتغیره مورد استفاده قرار گیرد (Gama, 2004). تفاوت عمده بین الگوریتم FT و مدل‌های سلسله مراتبی سنتی این است که الگوریتم FT از یک تابع رگرسیون لجستیک برای تقسیم به گره‌های داخلی کارکردی و پیش‌بینی برگ‌های تابعی استفاده می‌کند، در حالی که مدل‌های سلسله مراتبی سنتی از ورودی استفاده می‌کنند. ابتدا $X = (x_1, x_2, \dots, x_n)$ به‌عنوان متغیرهای مؤثر بر پتانسیل آب زیرزمینی و $Y = y_i$ به‌عنوان دسته آب زیرزمینی و غیر آب زیرزمینی در نظر گرفته می‌شود. سپس یک تابع تشخیص بیز خطی برای ساخت مدل $p = p(x)$ انتخاب می‌شود که احتمال توزیع دسته‌های آب‌های زیرزمینی و غیر آب‌های زیرزمینی است. سپس x_i با متغیر جدید تأثیرگذار بر آب‌های زیرزمینی توسعه داده می‌شود تا مجموعه داده جدید ایجاد شود. در نهایت، متغیرهای مؤثر بر آب‌های زیرزمینی از بین مجموعه داده اصلی و تمام مجموعه داده‌های جدید تولید شده برای ساخت درخت طبقه‌بندی انتخاب می‌شوند (Hong et al., 2018).

روش انتخاب ویژگی One-R

قابلیت پیش‌بینی هر عامل در بهره‌وری آب زیرزمینی به‌طور کلی متفاوت است؛ از این رو، ترتیب اهمیت آن در مرحله مدل‌سازی

5 True negative
6 False positive
7 False negative

1 One-R Attribute Evaluation
2 average merit
3 confusion matrix
4 True positive

(بیش از ۰/۵ لیتر بر ثانیه)، در حالی که FP نسبت تعداد چاه‌هایی است که به درستی به عنوان یک چاه بدون پتانسیل طبقه‌بندی می‌شوند (بیش از ۰/۵ لیتر بر ثانیه). علاوه بر این، TN و FN تعداد چاه‌هایی هستند که به طور صحیح و نادرست به عنوان چاه با پتانسیل بالا (بیش از ۰/۵ لیتر بر ثانیه) و یک چاه با پتانسیل غیر بالقوه (بیش از ۰/۵ لیتر بر ثانیه) طبقه‌بندی می‌شوند. معیارهای آماری ذکر شده در بالا را می‌توان به صورت روابط زیر محاسبه کرد:

رابطه (۱) $PPV = \frac{TP}{TP + FP}$

رابطه (۲) $NPV = \frac{TN}{TN + FN}$

رابطه (۳) $SST = \frac{TP}{TP + FN}$

رابطه (۴) $SPC = \frac{TN}{TN + FP}$

رابطه (۵) $ACC = \frac{TP + TN}{TN + FN + TP + FP}$

رابطه (۶) $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{pr} - x_{ac})^2}$

نتایج و بحث

آماده‌سازی داده‌ها

همان‌طور که ذکر شد، در این مطالعه، از ۱۰ عامل توپوگرافی (زاویه شیب، جهت شیب، ارتفاع و انحنای سطح)، هیدرولوژی (بارش و فاصله تا رودخانه)، زمین‌شناسی (سنگ‌شناسی، جنس خاک و فاصله تا گسل) و کاربری زمین جهت پیش‌بینی پتانسیل آب زیرزمینی استفاده گردید. برای تولید نقشه‌های عوامل ذکر شده (شکل ۴)، از درون یابی کریجینگ استفاده شده است، زیرا این روش دارای حداقل خطا است (Eftekhari and Akbari, 2020; Eftekhari et al., 2021).

اهمیت‌دهی عوامل مؤثر بر پتانسیل آب زیرزمینی با استفاده از

روش One-R

بر اساس جدول (۱)، نتایج نشان می‌دهد که ارتفاع بیش‌ترین مقدار متوسط شایستگی یعنی ۵۸/۹۷۳ را دارد که نشان‌دهنده مهم‌ترین نقش این عامل بر روی بهره‌وری آب زیرزمینی در منطقه مورد مطالعه است و پس از آن، به ترتیب نزولی، معیارهای فاصله تا رودخانه، کاربری اراضی، زاویه شیب، سنگ‌شناسی، فاصله تا گسل، انحنای سطح، جنس خاک، بارش و جهت شیب مشاهده می‌شود که تمام عوامل در منطقه مورد مطالعه تأثیر مثبتی بر بهره‌وری آب زیرزمینی دارند. از این رو تمام این عوامل باید در مدل‌های پیشنهادی در نظر گرفته شوند.

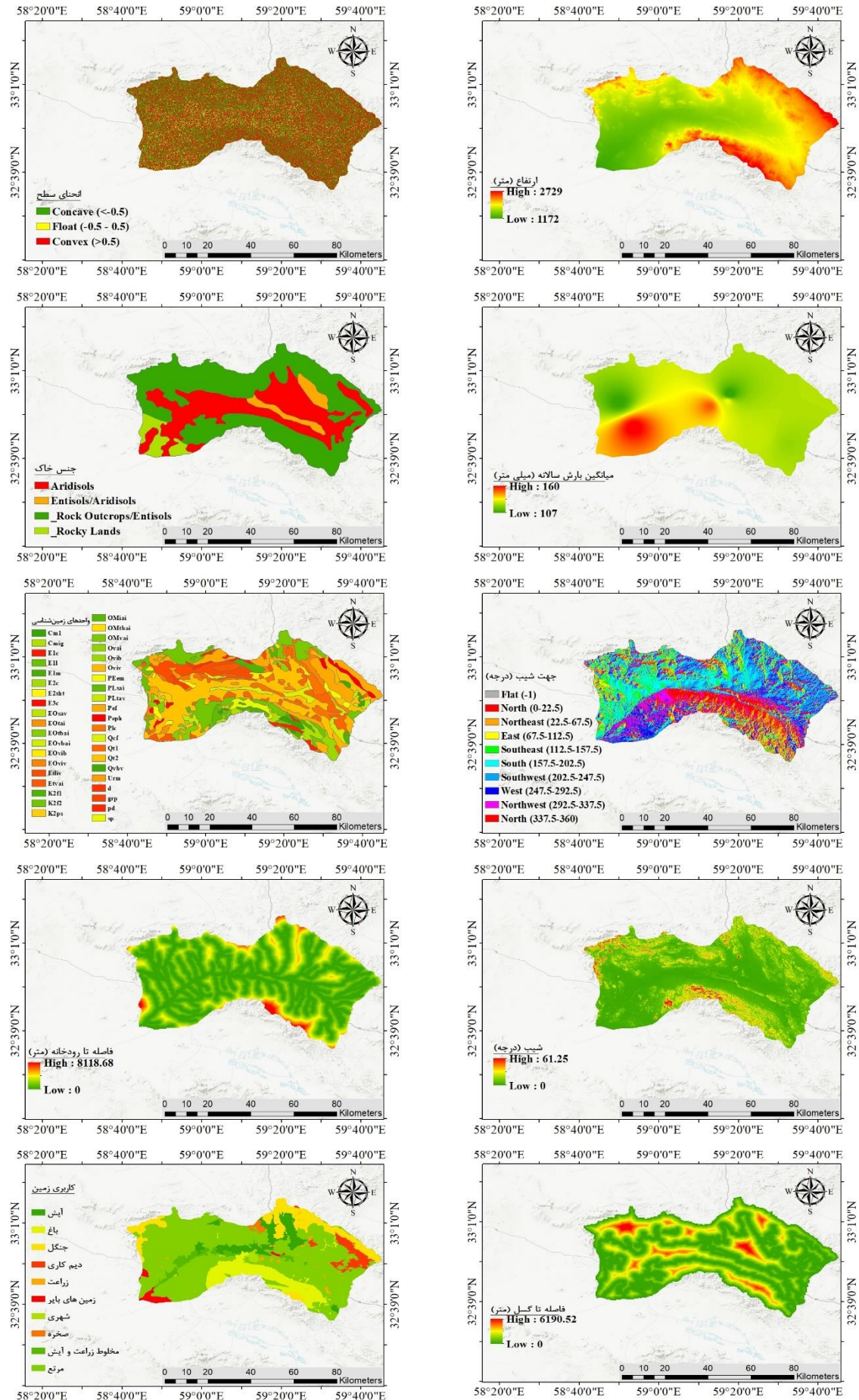
که در آن x_{ac} و x_{pr} به ترتیب به عنوان مقادیر واقعی پیش‌بینی شده حاصل از مدل‌سازی و مشاهده واقعی و n به عنوان تعداد کل چاه‌های مورد استفاده در مجموعه داده‌ها تعریف می‌شوند.

منحنی مشخصه عملکرد سامانه (ROC)

در این مطالعه منحنی مشخصه عملکرد سامانه^۱ (ROC) برای ارزیابی عملکرد مدل‌ها به ترتیب با نرخ‌های مثبت واقعی و نرخ مثبت کاذب بر روی محور Y و محور X استفاده شده است (Gorsevski et al., 2006). منطقه تحت منحنی ROC یعنی AUC عملکرد مدل را به صورت کمی تعیین می‌کند (Kamal et

جدول ۱- اهمیت عوامل توپوگرافی، هیدرولوژی و زمین‌شناسی با استفاده از روش انتخاب ویژگی One-R

ردیف	عوامل	متوسط شایستگی (AM)	انحراف معیار
۱	ارتفاع	۵۸/۹۷۳	± ۴/۲۷۵
۲	فاصله تا رودخانه	۵۵/۸۰۴	± ۳/۲۵۳
۳	کاربری اراضی	۵۳/۵۹۷	± ۴/۰۳۱
۴	زاویه شیب	۵۱/۸۹۴	± ۲/۲۹۱
۵	سنگ‌شناسی	۵۰/۵۴۲	± ۵/۹۲۲
۶	فاصله تا گسل	۵۰/۵۰۶	± ۴/۲۳۶
۷	انحنای سطح	۴۹/۵۱۶	± ۱/۱۵۹
۸	جنس خاک	۴۹/۲۱۳	± ۱/۱۰۱
۹	بارش	۴۸/۸۱۹	± ۴/۹۲۵
۱۰	جهت شیب	۴۸/۳۶۲	± ۳/۶۹۸



شکل ۴- عوامل توپوگرافی، هیدرولوژی و زمین‌شناسی مؤثر بر پتانسیل آب زیرزمینی

مقایسه قابلیت پیش‌بینی مدل‌ها

یافته برای پیش‌بینی پتانسیل آب زیرزمینی با استفاده از مجموعه داده‌های آموزشی (جدول ۲) و داده‌های آزمایشی (جدول ۳) ارزیابی شدند.

قابلیت‌های صحت‌سنجی و پیش‌بینی مدل‌های ترکیبی توسعه

جدول ۲- عملکرد مدل‌ها با استفاده از مجموعه داده‌های آموزشی

شماره	پارامترها	MRAB-FT	BA-FT	RF-FT	FT
۱	TP	۲۷	۲۶	۲۷	۲۴
۲	TN	۲۹	۳۰	۲۳	۲۰
۳	FP	۸	۱۰	۸	۱۱
۴	FN	۷	۸	۱۳	۱۶
۵	PPV (%)	۷۷/۱۴	۷۲/۲۲	۷۷/۱۴	۶۸/۵۷
۶	NPV (%)	۸۰/۵۵	۷۸/۹۴	۸۰/۵۵	۵۵/۵۵
۷	SST (%)	۷۹/۴۱	۷۶/۴۷	۶۷/۵۰	۶۰/۰۰
۸	SPF (%)	۷۸/۳۷	۷۵/۰۰	۷۴/۱۹	۶۴/۵۱
۹	ACC (%)	۷۶/۷۱	۷۵/۶۷	۷۰/۴۲	۶۱/۹۷
۱۰	Kappa	۰/۶۷۰	۰/۶۷۰	۰/۵۳۹	۰/۴۰۷
۱۱	RMSE	۰/۳۸۱۷	۰/۳۸۵	۰/۴۰۰۴	۰/۴۵۲۳

جدول ۳- عملکرد مدل‌ها با استفاده از مجموعه داده‌های آزمایشی

شماره	پارامترها	MRAB-FT	BA-FT	RF-FT	FT
۱	TP	۱۲	۱۱	۱۲	۱۲
۲	TN	۱۶	۱۶	۱۵	۱۴
۳	FP	۷	۸	۷	۷
۴	FN	۴	۴	۵	۶
۵	PPV (%)	۶۳/۱۶	۵۷/۸۹	۶۳/۱۶	۶۳/۱۶
۶	NPV (%)	۸۰/۰۰	۸۰/۰۰	۷۵/۰۰	۷۰/۰۰
۷	SST (%)	۷۵/۰۰	۷۳/۳۳	۷۰/۵۹	۶۶/۶۷
۸	SPC (%)	۶۹/۵۷	۶۶/۶۷	۶۸/۱۸	۶۶/۶۷
۹	ACC (%)	۷۱/۷۹	۶۹/۲۳	۶۹/۲۳	۶۶/۶۷
۱۰	Kappa	۰/۴۳۳	۰/۳۹۹	۰/۴۰۷	۰/۳۳۲
۱۱	RMSE	۰/۴۵۵۳	۰/۴۵۷۴	۰/۴۷۲۲	۰/۴۸۷۱

دست آمده است. در حالی که کم‌ترین مقدار NPV (۵۵/۵۵) است، اما این مقادیر کم‌تر از مدل‌های ترکیبی هستند. در واقع می‌توان گفت که تمام مدل‌های ترکیبی می‌توانند مناسب‌تر از سایر مدل‌ها در پیش‌بینی پتانسیل آب زیرزمینی باشند. بنابراین می‌توان نتیجه گرفت که مدل MRAB-FT مدل بهتری است که از طبقه‌بندی کننده پایه و دیگر مدل‌های ترکیبی مورد استفاده، عملکرد مناسب‌تری دارد. هم چنین جدول (۳) قابلیت پیش‌بینی مدل‌های ترکیبی توسعه یافته با استفاده از مجموعه داده‌های آزمایشی را نشان می‌دهد.

نتایج نشان داد که بالاترین مقدار PPV (۶۳/۱۶) توسط مدل‌های MRAB-FT، RF-FT و FT و سپس مدل ترکیبی BA-FT (۵۷/۸۹) به دست آمده است. بالاترین مقدار NPV توسط

طبق جدول (۲)، نتایج مجموعه داده‌های آموزشی نشان داد که مقادیر PPV (%), NPV (%), SST (%), SPC (%), Kappa و RMSE برای مدل ترکیبی MRAB-FT به ترتیب ۷۷/۱۴، ۸۰/۵۵، ۷۹/۴۱، ۷۸/۳۷، ۷۶/۷۱، ۰/۶۷۰ و ۰/۳۸۱۷ محاسبه شدند. برای مدل ترکیبی BA-FT، این مقادیر به ترتیب ۷۲/۲۲، ۷۸/۹۴، ۷۴/۴۷، ۷۵/۰۰، ۷۵/۶۷، ۰/۶۷۰ و ۰/۳۸۵ به دست آمدند. در حالی که در مورد مدل ترکیبی RF-FT، بالاترین مقادیر اندازه‌گیری آماری پارامترها عبارتند از: PPV (۷۷/۱۴) و پس از آن SPC (۷۴/۱۹)، ACC (۷۰/۴۲)، SST (۶۷/۵۰) و NPV (۶۳/۸۸) قرار دارند. با این حال، شاخص‌های آماری Kappa و RMSE برای این مدل به ترتیب ۰/۴۰۰۴ و ۰/۵۳۹ محاسبه شدند. در مدل طبقه‌بندی کننده پایه FT، نتایج نشان داد که بالاترین مقدار برای PPV (۶۸/۵۷) به

اساس، میزان پتانسیل پیش‌بینی شده به پنج کلاس شامل پتانسیل خیلی کم، پتانسیل کم، پتانسیل متوسط، پتانسیل زیاد و پتانسیل خیلی زیاد طبقه‌بندی گردید. علت استفاده از روش طبقه‌بندی شکست طبیعی در این است که طبقه‌بندی بر اساس گروه‌بندی‌های طبیعی موجود در ذات داده‌ها انجام می‌شود و نقاط شکستگی بین دسته‌ها به گونه‌ای معین می‌شود که به بهترین وجه ممکن مقادیر مشابه در یک گروه جای داده شوند و تفاوت بین کلاس‌ها بیشینه شود (Chen et al., 2020). شکل (۵) نقشه پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه را نشان می‌دهد که با استفاده از مدل‌های مورد نظر پیش‌بینی شده‌اند.

نقشه مناطق بالقوه آب زیرزمینی منطقه مورد مطالعه نشان می‌دهد که پتانسیل آب در مناطق مرکزی منطقه مورد مطالعه، به دلیل ارتفاع و زاویه شیب کمتر، بیشتر از سایر مناطق است. همچنین در شکل (۶)، درصد هر یک از کلاس‌های پتانسیل آب زیرزمینی پیش‌بینی شده توسط مدل‌های MRAB-FT، BA-FT، RF-FT و FT نشان داده شده است. نتایج نشان داد که در مدل MRAB-FT، ۲ کلاس با پتانسیل خیلی زیاد و زیاد، درصد بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های BA-FT، RF-FT و FT تحت پوشش قرار می‌دهد. همچنین در مدل RF-FT، کلاس با پتانسیل متوسط، درصد بیشتری را در منطقه مورد مطالعه نسبت به کلاس مشابه در مدل‌های MRAB-FT، BA-FT و FT دربر گرفته است. علاوه بر این در مدل FT، ۲ کلاس با پتانسیل خیلی کم و کم، درصد بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های MRAB-FT، BA-FT و RF-FT دربر گرفته است.

عملکرد مدل‌های توسعه‌یافته با روش منحنی ارزیابی ROC شد. شکل (۷) منحنی ROC را برای مجموعه داده‌های آموزشی (منحنی نرخ موفقیت^۲ (SRC)) و داده‌های آزمایشی (منحنی نرخ پیش‌بینی^۳ (PRC)) مدل‌ها نشان می‌دهد. شکل (۷-الف) نشان می‌دهد که مدل ترکیبی BA-FT دارای بالاترین مقدار AUC (۰/۸۹۸) و پس از آن مدل‌های MRAB-FT (AUC= ۰/۸۵۸)، RF-FT (AUC= ۰/۸۴۰) و FT (AUC= ۰/۷۶۰) قرار دارند. به طور کلی، قابلیت مدل‌ها برای پیش‌بینی پتانسیل آب زیرزمینی توسط مجموعه داده‌های آزمایشی ارزیابی می‌شود. بر این اساس شکل (۷-ب) نشان می‌دهد که مدل ترکیبی MRAB-FT قوی‌تر بوده و دارای بالاترین دقت (AUC= ۰/۷۴۲) نسبت به مدل‌های BA-FT (AUC= ۰/۷۱۴)، RF-FT (AUC= ۰/۷۳۶) و FT (AUC= ۰/۶۴۷) می‌باشد. بنابراین، می‌توان نتیجه گرفت که اگر

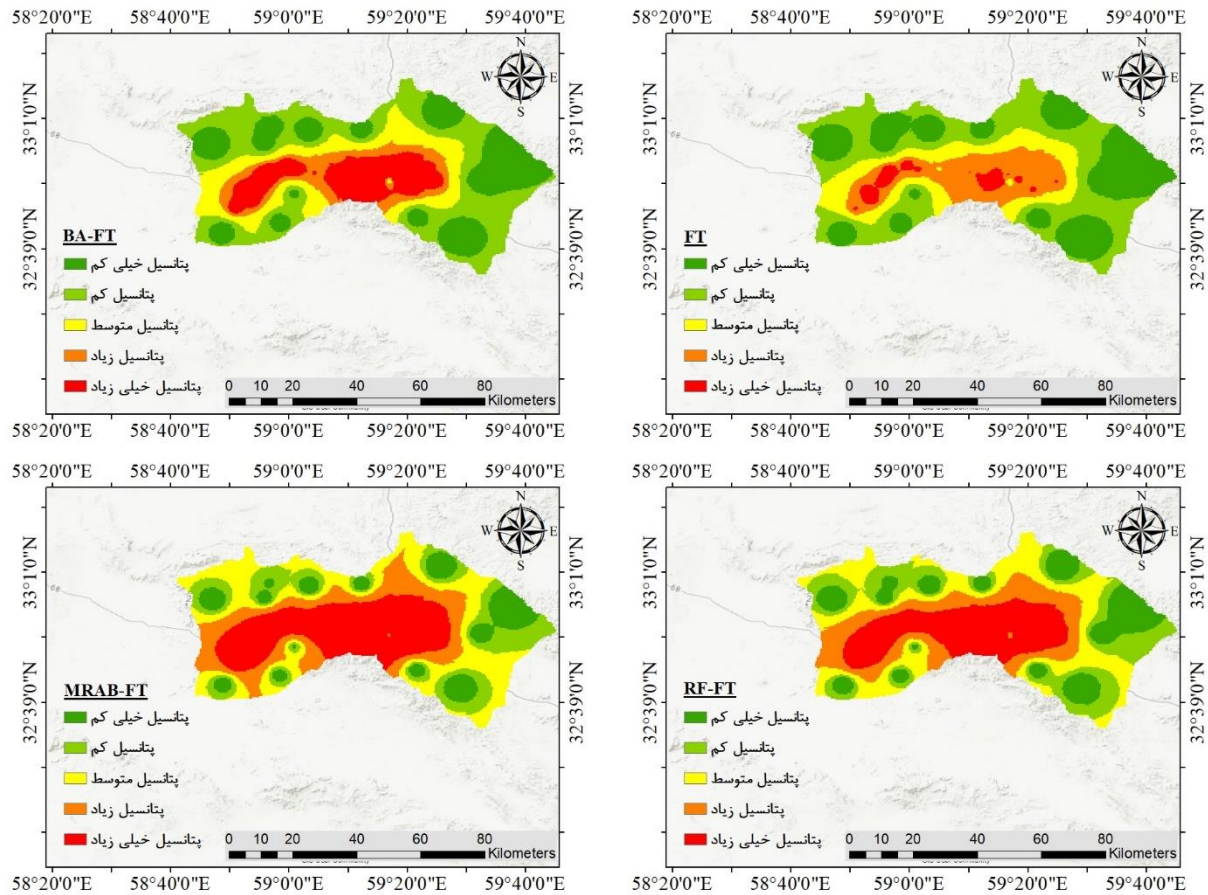
مدل‌های ترکیبی MRAB-FT و BF-FT (۰/۸۰/۰۰) محاسبه شده است، در حالی که RF-FT (۰/۷۵/۰۰) و مدل‌های FT (۰/۷۰/۰۰) کم‌ترین مقدار NPV را دارند. بالاترین مقدار SST (۰/۷۵/۰۰) توسط مدل ترکیبی MRAB-FT و سپس مدل‌های BA-FT (۰/۷۳/۳۳)، RF-FT (۰/۷۰/۵۹) و FT (۰/۶۶/۶۷) به دست آمده است. بنابراین، می‌توان مشاهده کرد که مدل ترکیبی MRAB-FT قابلیت پیش‌بینی بالاتری نسبت به مدل‌های دیگر دارد که به طور نادرستی (۰/۷۵) چاه‌های آب زیرزمینی را با پتانسیل بیش از ۰/۵ لیتر بر ثانیه (به عنوان پتانسیل بالا) در منطقه مورد مطالعه طبقه‌بندی شده‌اند. بیش‌ترین مقدار SPC (۰/۶۹/۵۷) در مدل ترکیبی MRAB-FT و کم‌ترین مقدار آن برای FT و BA-FT (۰/۶۶/۶۷) به دست آمد. این بدان معنی است که مدل ترکیبی MRAB-FT قابلیت پیش‌بینی بهتری نسبت به دیگر مدل‌ها در طبقه‌بندی صحیح (۰/۶۹/۵۷) چاه‌های آب زیرزمینی با پتانسیل کم‌تر از ۰/۵ لیتر بر ثانیه (به عنوان پتانسیل پایین) در منطقه مورد مطالعه دارد. مدل گروهی MRAB-FT نیز بالاترین مقدار ACC (۰/۷۱/۷۹) را دارد و پس از آن مدل‌های گروهی BA-FT و RF-FT (۰/۶۹/۲۳) و مدل FT (۰/۶۶/۶۷) قرار دارند. نتایج نشان داد که مدل MRAB-FT قادر به طبقه‌بندی صحیح ۷۱/۷۹٪ از چاه‌های آب زیرزمینی دارای پتانسیل بیشتر از ۰/۵ لیتر بر ثانیه (به عنوان پتانسیل بالا) و همچنین کم‌تر از ۰/۵ لیتر بر ثانیه (به عنوان پتانسیل پایین) در منطقه مورد مطالعه است. مقدار Kappa برای مدل ترکیبی MRAB-FT (۰/۴۳۳) بیش‌ترین مقدار را دارد و پس از آن مدل‌های RF-FT (۰/۴۰۷)، BA-FT (۰/۳۹۹) و FT (۰/۳۳۲) قرار دارند. لذا می‌توان نتیجه گرفت که مقدار RMSE برای مدل MRAB-FT (۰/۴۵۵۳) کم‌ترین مقدار است و پس از آن مدل‌های BA-FT (۰/۴۵۷۴)، RF-FT (۰/۴۷۲۲) و FT (۰/۴۸۷۱) قرار دارند (جدول ۳). بنابراین، برای مجموعه داده‌های آزمایشی، مدل MRAB-FT بالاترین توان پیش‌بینی را دارد و پس از آن مدل‌های BA-FT، RF-FT و FT قرار دارند.

تهیه نقشه پیش‌بینی پتانسیل آب‌های زیرزمینی

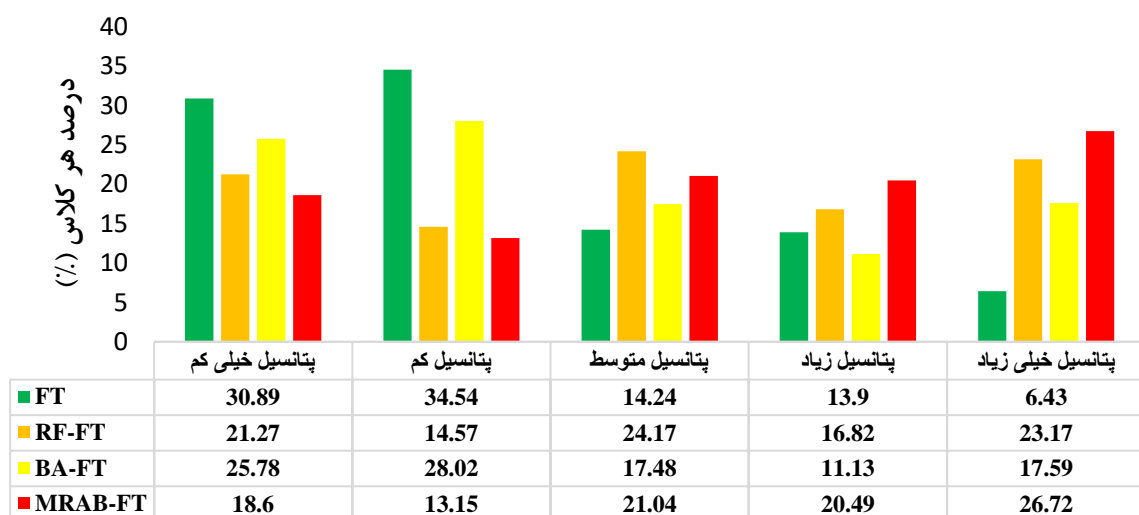
در این مطالعه، نقشه‌های پیش‌بینی مناطق بالقوه آب زیرزمینی در منطقه مورد مطالعه بر اساس مدل‌های MRAB-FT، BA-FT، RF-FT و FT و با استفاده از مجموعه داده‌های آموزشی (۷۰ درصد) و آزمایشی (۳۰ درصد) برآورد شدند. نقشه میزان پتانسیل پیش‌بینی شده با استفاده از روش طبقه‌بندی شکست طبیعی^۱ به کلاس‌های مختلف در نرم‌افزار Arc Map طبقه‌بندی شد. بر این

زیرزمینی در منطقه مورد مطالعه بهتر است.

چه تمام مدل‌های ترکیبی به کار رفته قدرت پیش‌بینی خوبی دارند، اما عملکرد مدل MRAB-FT برای پیش‌بینی پتانسیل آب

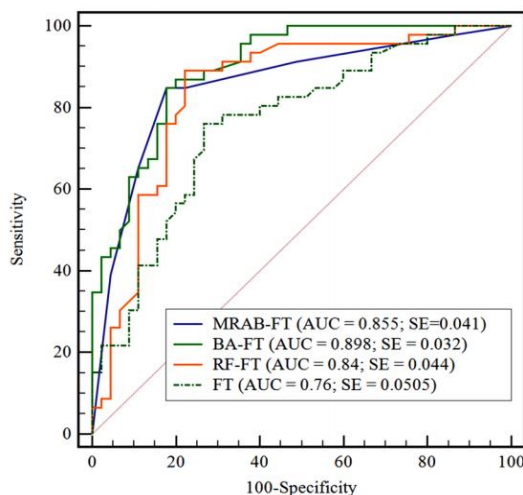


شکل ۵- نقشه پیش‌بینی پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه بر اساس مدل‌های FT و RF-FT, BA-FT, MRAB-FT

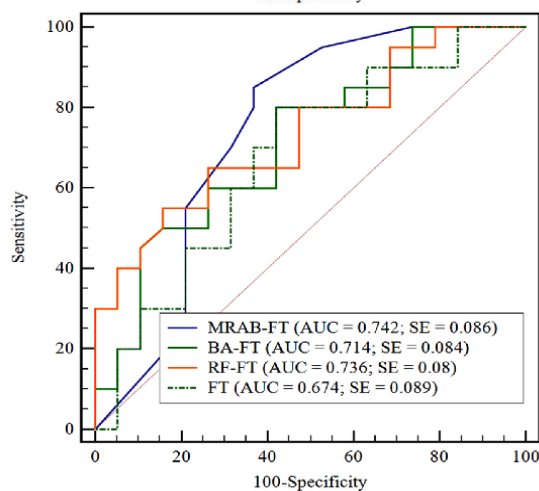


■ FT ■ RF-FT ■ BA-FT ■ MRAB-FT

شکل ۶- درصد هر یک از کلاس‌های پتانسیل آب‌های زیرزمینی پیش‌بینی شده بر اساس مدل‌های FT و RF-FT, BA-FT, MRAB-FT



(الف)



(ب)

شکل ۷- منحنی ROC و مقدار AUC برای روش‌های پیشنهادی تحقیق (الف) مجموعه داده‌های آموزشی (ب) مجموعه داده‌های آزمایشی

نتیجه‌گیری

ارزیابی شد. نتایج تجزیه و تحلیل نشان داد که تمام مدل‌های ترکیبی جدید توسعه یافته قابلیت پیش‌بینی خوبی را نشان داده‌اند، اما عملکرد مدل MRAB-FT در پیش‌بینی دقیق مناطق بالقوه آب زیرزمینی بهترین است و پس از آن مدل‌های RF-FT، BA-FT و FT برای نگاشت دقیق مناطق بالقوه آب زیرزمینی قرار دارند. در مناطق دیگر نیز این مدل را می‌توان برای نقشه برداری دقیق و شناسایی صحیح مناطق پتانسیل آب زیرزمینی به کار برد که به مدیریت مناسب آب زیرزمینی یک منطقه کمک خواهد کرد. می‌توان نتیجه گرفت که مدل‌های ترکیبی توسعه یافته بر اساس مدل‌های مکانی و روش‌های یادگیری ماشین قابلیت خوبی در ارزیابی پتانسیل آب زیرزمینی صحیح یک منطقه با کمترین داده‌ها دارد.

"هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد"

ارزیابی مناسب پتانسیل آب زیرزمینی یک منطقه برای استفاده بهینه آن با اتخاذ فن‌آوری و روش‌های جدید مطلوب است. در این مطالعه، سه مدل ترکیبی جدید هوش مصنوعی یا یادگیری ماشین به نام‌های MRAB-FT، BA-FT، RF-FT بر اساس طبقه‌بندی کننده پایه FT جهت پیش‌بینی پتانسیل آب زیرزمینی در دشت بیرجند توسعه داده شده است. محققان قابلیت و کاربرد روش‌های یادگیری ماشین و مدل‌های هوش مصنوعی ترکیبی را در پیش‌بینی مکانی پتانسیل آب‌های زیرزمینی مناطق مختلف اثبات کرده‌اند (Miraki et al., 2019; Nhu et al., 2020). برای انتخاب عوامل مناسب برای مدل‌سازی از رویکرد One-R استفاده شد. عملکرد مدل‌های توسعه یافته با استفاده از معیارهای آماری استاندارد (ROC و RMSE, Kappa, SPC, SST, NPV, PPV)

REFERENCES

Abedini, M., Ghasemian, B., Shirzadi, A., Shahabi, H., Chapi, K., Pham, B. T. and Tien Bui, D. (2019). A novel hybrid approach of bayesian logistic regression and its ensembles for landslide

susceptibility assessment. *Geocarto International*, 34(13), 1427-1457.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

- Chen, W., Tsangaratos, P., Ilija, I., Duan, Z. and Chen, X. (2019). Groundwater spring potential mapping using population-based evolutionary algorithms and data mining methods. *Science of The Total Environment*, 684, 31-49.
- Chen, W., Zhao, X., Tsangaratos, P., Shahabi, H., Ilija, I., Xue, W. and Ahmad, B. B. (2020). Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. *Journal of Hydrology*, 583, 124602.
- Díaz-Alcaide, S. and Martínez-Santos, P. (2019). Advances in groundwater potential mapping. *Hydrogeology Journal*, 27(7), 2307-2324.
- Eftekhari, M. and Akbari, M. (2020). Development of DRASTIC Method Considering Land Use to Analyze the Potential of Aquifer Pollution in Semi-Arid Regions. *Environment and Water Engineering*, 6(4), 345-359. (In Farsi)
- Eftekhari, M., Eslaminezhad, S., Haji Elyasi A. and Akbari, M. (2021). Geostatistical Evaluation with Drinking Groundwater Quality Index (DGWQI) in Birjand Plain Aquifer. *Environment and Water Engineering*; 7(2), 268-279. (In Farsi)
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters* 27, 861-874.
- Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.
- Freund, Y., Schapire, R. and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society for Artificial Intelligence*, 14(771-780), 1612.
- Friedman, J., Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of statistics*, 28(2), 337-407.
- Gama, J. (2004). Functional trees. *Machine Learning*, 55(3), 219-250.
- Gorsevski, P. V., Gessler, P. E., Foltz, R. B. and Elliot, W. J. (2006). Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Transactions in GIS*, 10(3), 395-415.
- He, Q., Xu, Z., Li, S., Li, R., Zhang, S., Wang, N. and Chen, W. (2019). Novel entropy and rotation forest-based credal decision tree classifier for landslide susceptibility modeling. *Entropy*, 21(2), 106.
- Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1), 63-90.
- Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A. X. and Kazakis, N. (2018). Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Science of the total Environment*, 621, 1124-1141.
- Jha, M. K., Chowdary, V. M. and Chowdhury, A. (2010). Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *Hydrogeology journal*, 18(7), 1713-1728.
- Kumar, A. and Krishna, A. P. (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., Hong, S. M. and Jung, H. S. (2018). GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. *Geocarto international*, 33(8), 847-861.
- Miraki, S., Zanganeh, S. H., Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H. and Pham, B. T. (2019). Mapping groundwater potential using a novel hybrid intelligence approach. *Water resources management*, 33(1), 281-302.
- Moghaddam, D. D., Rezaei, M., Pourghasemi, H. R., Pourtaghie, Z. S. and Pradhan, B. (2015). Groundwater spring potential mapping using bivariate statistical model and GIS in the Taleghan watershed, Iran. *Arabian Journal of Geosciences*, 8(2), 913-929.
- Naghibi, S. A., Pourghasemi, H. R. and Dixon, B. (2016). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental monitoring and assessment*, 188(1), 1-27.
- Nampak, H., Pradhan, B. and Abd Manap, M. (2014). Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology*, 513, 283-300.
- Nguyen, P. T., Tuyen, T. T., Shirzadi, A., Pham, B. T., Shahabi, H., Omidvar, E. and Bui, D. T. (2019). Development of a novel hybrid intelligence approach for landslide spatial prediction. *Applied Sciences*, 9(14), 2824.
- Nhu, V. H., Rahmati, O., Falah, F., Shojaei, S., Al-Ansari, N., Shahabi, H. and Ahmad, B. B. (2020). Mapping of groundwater spring potential in karst aquifer system using novel ensemble bivariate and multivariate models. *Water*, 12(4), 985.
- Oh, H. J., Kim, Y. S., Choi, J. K., Park, E. and Lee, S. (2011). GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology*, 399(3-4), 158-172.
- Pham, B. T., Prakash, I., Dou, J., Singh, S. K., Trinh, P. T., Tran, H. T. and Bui, D. T. (2020). A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto International*, 35(12), 1267-1292.
- Rodriguez, J. J., Kuncheva, L. I. and Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10), 1619-1630.
- Shirzadi, A., Bui, D. T., Pham, B. T., Solaimani, K., Chapi, K., Kavian, A. and Revhaug, I. (2017). Shallow landslide susceptibility assessment using a novel hybrid intelligence

- approach. *Environmental Earth Sciences*, 76(2), 60.
- Tien Bui, D., Shahabi, H., Omidvar, E., Shirzadi, A., Geertsema, M., Clague, J. J. and Lee, S. (2019). Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sensing*, 11(8), 931.
- Xia, J., Du, P., He, X. and Chanussot, J. (2013). Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters*, 11(1), 239-243.