

مقاله علمی-پژوهشی

تعیین پتانسیل آب زیرزمینی با استفاده مدل‌های یادگیری ماشین جمعی در بستر GIS (مطالعه موردی: دشت بیرجند)

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

تاریخ دریافت: ۱۴۰۰/۰۷/۰۷ تاریخ پذیرش: ۱۴۰۰/۰۸/۲۴

چکیده

پیش‌بینی پتانسیل آب‌های زیرزمینی جهت توسعه و برنامه‌ریزی سیستماتیک منابع آب بسیار بااهمیت است. هدف اصلی این مطالعه، توسعه مدل‌های یادگیری ماشین جمعی شامل جنگل تصادفی (RF)، رگرسیون منطقی (LR) و بیز ساده (NB) توسط الگوریتم طبقه‌بندی کننده زیرفضای تصادفی (RS)، جهت پیش‌بینی مناطق بالقوه آب زیرزمینی در دشت بیرجند می‌باشد. لذا جهت پیاده‌سازی، داده‌های ژئویدرولوژیکی ۳۷ حلقه چاه آب زیرزمینی (تعداد چاه‌ها، موقعیت مکانی چاه‌ها و تراز آب زیرزمینی یا سطح ایستابی) و ۱۷ معیار هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی مورد استفاده قرار گرفت. روش انتخاب ویژگی کمترین مربعات ماشین بردار پشتیبان (LSSVM) جهت تعیین معیارهای مؤثر به منظور افزایش عملکرد الگوریتم‌های یادگیری ماشین استفاده شد. در نهایت نقشه‌های پیش‌بینی پتانسیل آب زیرزمینی با استفاده از مدل‌های RF-RS، LR-RS و NB-RS تهیه شدند. عملکرد این مدل‌ها با استفاده از سطح زیر منحنی (AUC) و سایر شاخص‌های آماری مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل ترکیبی RF-RS (AUC= ۰/۸۶۷) قابلیت پیش‌بینی بسیار بالایی برای پتانسیل آب زیرزمینی در منطقه مورد مطالعه دارد. همچنین مشخص شد که معیار ارتفاع بیشترین اهمیت را در پیش‌بینی پتانسیل آب زیرزمینی در منطقه مورد مطالعه دارد. نتایج مطالعه حاضر می‌تواند جهت اتخاذ تصمیمات و برنامه‌ریزی مناسب در استفاده بهینه از منابع آب زیرزمینی مفید باشد.

واژه‌های کلیدی: بیز ساده، پتانسیل آب زیرزمینی، جنگل تصادفی، رگرسیون منطقی، زیرفضای تصادفی

مقدمه

بسیاری از استان‌های کشور شده است (Osati et al., 2014; Manap et al., 2013). پیدایش و حرکت آب زیرزمینی متأثر از عوامل بسیاری می‌باشد که به توپوگرافی، سنگ‌شناسی، ساختارهای زمین‌شناسی، شکستگی، تخلخل ثانویه، توزیع سطح سفره آب زیرزمینی، تغذیه آب زیرزمینی، شیب، الگوی زهکشی، اشکال زمین، پوشش زمین، شرایط آب‌وهوایی و روابط متقابل آن‌ها مربوط می‌شود (Oh et al., 2011). برای شناسایی مناطق بالقوه آب زیرزمینی، به کارگیری روش‌هایی که به مدیران در استفاده مؤثرتر از این منابع کمک کنند، ضروری می‌باشد (Rahmati et al., 2015). این روش‌ها برای توسعه آینده، مدیریت و جلوگیری از کاهش منابع آب زیرزمینی مورد نیاز می‌باشند. بسیاری از مطالعات از فن‌آوری‌های سیستم اطلاعات مکانی^۵ (GIS) و سنجش از دور^۶ (RS)، برای ارزیابی پتانسیل آب‌های زیرزمینی استفاده کرده‌اند که در ادامه به برخی از آن‌ها اشاره شده است: رحمتی و همکاران جهت تهیه نقشه پتانسیل آب زیرزمینی

آب زیرزمینی به عنوان یکی از بحرانی‌ترین منابع در سراسر جهان می‌باشد که آب مورد نیاز برای کشاورزی، صنعت، دامداری و جوامع انسانی را تأمین می‌کند (Neshat et al., 2014). اخیراً، بسیاری از مناطق ایران به دلیل تغییرات اقلیمی و برداشت بی‌رویه از منابع آب زیرزمینی موجود خشک شده‌اند که منجر به کمبود جدی آب در

- ۱- دانش‌آموخته کارشناسی ارشد، گروه مهندسی نقشه برداری، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، دانشگاه تهران، تهران، ایران
- ۲- دانش‌آموخته کارشناسی ارشد مهندسی عمران آب و سازه‌های هیدرولیکی و عضو باشگاه پژوهشگران جوان و نخبگان، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران
- ۳- دانشیار گروه مهندسی عمران، دانشگاه بیرجند، بیرجند، ایران.
- ۴- دانش‌آموخته کارشناسی ارشد، گروه مهندسی عمران، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

*- نویسنده مسئول: (Email: Moakbari@birjand.ac.ir)

DOR: 20.1001.1.20087942.1401.16.1.12.9

5- Geographic information system
6- Remote sensing

جاده، سنگ‌شناسی، جنس خاک، کاربری زمین، شاخص نرمال شده تفاوت پوشش گیاهی و بارندگی استخراج شدند. نتایج نشان داد که مدل RF ($AUC=0/94$) دقت بالاتری را نسبت به مدل‌های BRT و SVM در پیش‌بینی مناطق بالقوه آب زیرزمینی دارد. چن و همکاران مدل‌های یادگیری ماشین شامل آدابوست^{۱۲} (AB)، مدل بگینگ^{۱۳} (BA)، مدل زیرفضای تصادفی^{۱۴} (RS)، مدل دگینگ^{۱۵} (DA) و مدل جنگل چرخشی^{۱۶} (ROF) را توسط الگوریتم J48DT برای پیش‌بینی مکانی پتانسیل آب زیرزمینی (در فصل بهار) در شهرستان ووقی کشور چین توسعه دادند (Chen et al., 2020). جهت پیاده‌سازی روش‌های پیشنهادی از ۱۳ معیار مؤثر شامل معیارهای توپوگرافی، هیدرولوژی و محیطی استفاده شد. آن‌ها نتیجه گرفتند که مدل ترکیبی ROF-J48DT ($AUC=79/7$) دقت بالاتری را در پیش‌بینی مکانی پتانسیل آب زیرزمینی دارد.

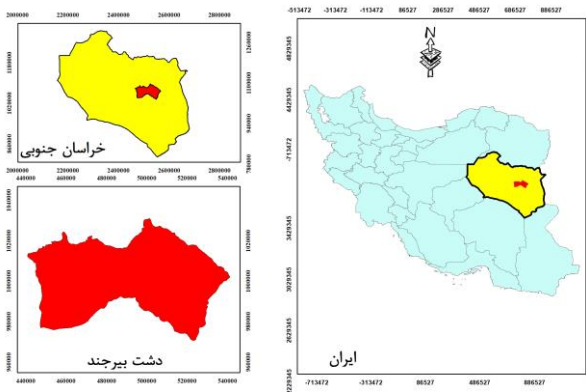
در تمام مطالعات ذکر شده در بالا، محققان مدل‌های یادگیری ماشین را برای انتخاب بهترین مدل در راستای پیش‌بینی پتانسیل آب‌های زیرزمینی مقایسه کرده‌اند. شناسایی مدل‌های یادگیری ماشین مؤثرتر که می‌توانند با حداقل داده‌های میدانی جهت پیش‌بینی پتانسیل آب زیرزمینی مورد استفاده قرار گیرند، مطلوب است. هدف از این پژوهش پیش‌بینی پتانسیل آب زیرزمینی با استفاده از توسعه روش‌های نوین یادگیری ماشین مبتنی بر الگوریتم طبقه‌بندی کننده زیرفضای تصادفی است. بنابراین، این مطالعه از مطالعات قبلی متمایز است، زیرا در این مطالعه از مدل‌های ترکیبی یادگیری ماشین بهبودیافته یعنی RF-RS، LR-RS و NB-RS برای پیش‌بینی مناطق بالقوه آب‌های زیرزمینی در منطقه دشت بیرجند استفاده شده است که نوآوری تحقیق حاضر نیز می‌باشد. دلیل استفاده از الگوریتم RS این است که الگوریتم مورد نظر یک روش فرعی را برای طبقه‌بندی‌های پایه آموزشی ایجاد می‌کند تا نمونه‌های مختلف از فضای مشخصه به جای فضای نمونه استفاده شوند و این موضوع سبب بهبود دقت مدل‌های RF، LR و NB خواهد شد (Pham et al., 2017). در نهایت، معیار سطح زیر منحنی (AUC) و معیارهای آماری شامل مقدار پیش‌بینی شده مثبت^{۱۷} (PPV)، مقدار پیش‌بینی شده منفی^{۱۸} (NPV)، حساسیت^{۱۹} (SST)، دقت^{۲۰} (ACC)، تشخیص‌پذیری یا ویژگی^{۲۱} (SPC) و خطای جذر میانگین مربعات^{۲۲} (RMSE) جهت بررسی دقت مدل‌های پیشنهادی به کار گرفته شدند.

در حوضه آبریز شهرستان مهران از مدل‌های جنگل تصادفی^۱ (RF) و حداکثر آنتروپی^۲ (ME) در بستر GIS استفاده کردند (Rahmati et al., 2016). جهت پیاده‌سازی از معیارهای ارتفاع، زاویه شیب، جهت شیب، انحنای سطح، تراکم زهکشی، فاصله تا رودخانه، شاخص رطوبت توپوگرافی، کاربری زمین، سنگ‌شناسی و بافت خاک استفاده کردند. نتایج اعتبارسنجی نشان داد که مدل ME ($AUC=0/91$) دقت پیش‌بینی نسبتاً خوبی را جهت پیش‌بینی پتانسیل آب زیرزمینی در منطقه مورد مطالعه دارد. هدف از مطالعه رضوی ترمه و همکاران تعیین مناطق بالقوه آب زیرزمینی دشت بوشهر با استفاده از ترکیب مدل‌های یادگیری ماشین RF و درخت مدل لجستیکی^۴ (LMT) با مدل‌های آماری نسبت فرکانس^۵ (FR)، ضریب اطمینان^۶ (CF) و مدل تابع باور شواهد^۷ (EBF) بود (Razavi Termeh et al., 2019). ۱۵ معیار مؤثر شامل معیارهای توپوگرافی، هیدرولوژی، زمین‌شناسی و محیطی جهت مدل‌سازی انتخاب شدند. نتایج نشان داد که مقدار AUC به‌دست‌آمده با استفاده از روش‌های CF-RF، EBF-RF، CF-LMT، FR-RF و EBF-LMT به ترتیب ۰/۹۲۷، ۰/۹۲۴، ۰/۹۱۷، ۰/۹۰۶، ۰/۸۸۵ و ۰/۸۳ بود. چن و همکاران از ترکیب سه الگوریتم داده‌کاوی RF، بیز ساده^۸ (NB) و ماشین بردار پشتیبان^۹ (SVM) با الگوریتم ژنتیک^{۱۰} (GA) جهت تهیه نقشه پتانسیل آب زیرزمینی در شهرستان ووقی کشور چین استفاده کردند (Chen et al., 2019). جهت پیاده‌سازی ۱۶ عامل شامل سنگ‌شناسی، جنس خاک، کاربری زمین، شاخص تغییرات پوشش گیاهی، ارتفاع، زاویه شیب، جهت شیب، انحنای سطح، انحنای نیمرخ، خمیدگی، شاخص قدرت جریان، شاخص انتقال جریان، شاخص رطوبت توپوگرافی، متوسط بارندگی سالانه، فاصله از شبکه رودخانه و فاصله از شبکه جاده استخراج شدند. نتایج نشان داد که مدل‌های ترکیبی بهینه‌شده از نظر دقت و کارایی (مقدار AUC) نسبت به مدل‌های اصلی برتر هستند. پراساد و همکاران جهت پیش‌بینی مناطق بالقوه آب زیرزمینی در سواحل غربی هند از سه مدل یادگیری ماشین شامل مدل RF، درخت رگرسیون تقویت‌شده^{۱۱} (BRT) و مدل SVM استفاده کردند (Prasad et al., 2020). به‌منظور دستیابی به این هدف، ۱۴ معیار تأثیرگذار بر آب زیرزمینی شامل ارتفاع، شیب، جهت شیب، طول شیب، انحنای نیمرخ، انحنای سطح، شاخص رطوبت توپوگرافی، فاصله تا رودخانه، فاصله تا

- 12- AdaBoost
- 13- Bagging
- 14- Random subspace
- 15- Dagging
- 16- Rotation forest
- 17- Positive Predictive Value
- 18- Negative Predictive Value
- 19- Sensitivity
- 20- Accuracy
- 21- Specificity
- 22- Root mean square error

- 1- Random forest
- 2- maximum entropy
- 3- Area Under the Curve
- 4- Logistic model tree
- 5- Frequency ration
- 6- Certainty factor
- 7- Evidential belief function
- 8- Naïve byes
- 9- Support vector machine
- 10- Genetic Algorithm
- 11- Boosted regression tree

(شکل ۲).



شکل ۲- موقعیت مکانی استان خراسان جنوبی و دشت بیرجند

حوضه آبریز بیرجند دارای وسعت $3406/72 \text{ km}^2$ 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 ۱۰۸ 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 (Eftekhari and Akbari, 2020). با توجه به بارش‌های فصلی و مربوط بودن داده‌های این تحقیق به نیمه دوم سال ۱۳۹۸، جریان آب در رودخانه شاهرود جریان داشته است. حداکثر ارتفاع حوضه از سطح دریا ۲۷۲۹ متر در ارتفاعات باقران و حداقل ارتفاع نیز ۱۱۷۲ متر در خروجی دشت (منطقه فدشک) می‌باشد.

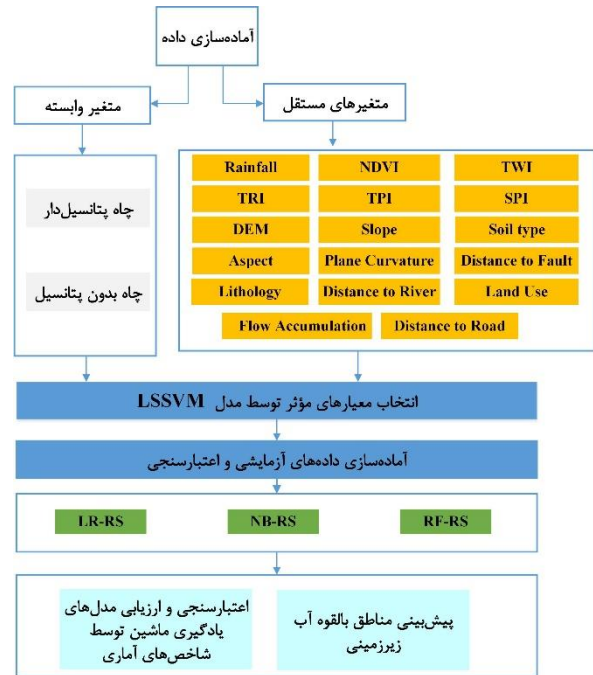
تهیه لایه‌های اطلاعاتی منطقه مورد مطالعه

عوامل مؤثر بر پتانسیل آب‌های زیرزمینی

در این مطالعه، ۱۷ معیار هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی شامل بارندگی^۴، شاخص نرمال شده تفاوت پوشش گیاهی (NDVI)^۵، شاخص رطوبت توپوگرافی (TWI)^۶، شاخص توان آبراهه^۷ (SPI)^۷، شاخص ناهمواری زمین^۸ (TRI)^۸، شاخص موقعیت توپوگرافی^۹ (TPI)^۹، مدل رقومی ارتفاعی (DEM)^{۱۰}، زاویه شیب^{۱۱}،

مواد و روش‌ها

روش انجام پژوهش حاضر، توصیفی-تحلیلی بوده و نوع آن بر اساس هدف کاربردی است. تمامی پردازش‌های مربوط به داده‌ها در محیط‌های نرم‌افزاری Arc GIS، SAGA GIS و Google earth engine و محاسبات کمی و توسعه روش‌ها در محیط برنامه‌نویسی Matlab انجام گرفته است. طبق شکل ۱، روند اجرایی این مطالعه شامل ۵ مرحله اساسی می‌باشد: (۱) جمع‌آوری چاه‌های مشاهداتی آب زیرزمینی در منطقه مورد مطالعه جهت تعریف متغیر وابسته^۱. (۲) استخراج معیارهای مکانی مؤثر بر پتانسیل آب زیرزمینی جهت تعریف متغیرهای مستقل^۲ (۳) استفاده از الگوریتم کمترین مربعات ماشین بردار پشتیبان^۳ (LSSVM) جهت تعیین معیارهای مؤثر. (۴) پیش‌بینی مناطق بالقوه آب زیرزمینی بر اساس توسعه رویکردهای یادگیری ماشین. (۵) ارزیابی و مقایسه عملکرد رویکردهای یادگیری ماشین.



شکل ۱- روند اجرایی تحقیق

منطقه مورد مطالعه

دشت بیرجند بخشی از حوضه آبریز کویر لوت می‌باشد که در فاصله ۴۸۵ km جنوب مشهد و در قسمت شمالی ارتفاعات باقران بین طول‌های جغرافیایی $41^{\circ} 58'$ تا $46^{\circ} 59'$ شرقی و عرض‌های جغرافیایی $35^{\circ} 32'$ تا $33^{\circ} 8'$ شمالی قرار گرفته است

4- Rainfall

5- Normalized difference vegetation index

6-Topographic Wetness Index

7- Stream power index

8- Terrain ruggedness index

9- Topographic Position Index

10- Digital Elevation model

11- Slope

1- dependent variable

2- Independent variable

3- Least Square Support Vector Machine

لایه توپوگرافی در مقیاس ۱:۵۰۰۰۰ تهیه شدند. لایه فاصله تا غسل نیز از نقشه زمین‌شناسی در مقیاس ۱:۱۰۰۰۰۰ استخراج گردید. لایه کاربری اراضی منطقه مورد مطالعه نیز از طریق تصاویر ماهواره‌ای لندست ۸ استخراج شد. 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 ...); لذا هر چه فاصله از جاده بیشتر باشد، میزان نفوذ آب در سطح زمین به دلیل عدم وجود آسفالت بیشتر است و تأثیر به سزایی بر پتانسیل آب‌های زیرزمینی دارد (Hong et al., 2018). هم‌چنین، از باند ۴ و ۵ تصاویر ماهواره‌ای لندست ۸ جهت تولید لایه NDVI استفاده گردید. به‌منظور محاسبه شاخص میانگین بارش سالیانه از چهار ایستگاه سینوپتیک در منطقه مورد مطالعه استفاده شد. هم‌چنین رطوبت خاک و اشباع سطح نشان‌دهنده شاخص رطوبت توپوگرافی (TWI) است که طبق رابطه (۱) محاسبه می‌شود (Jancewicz et al., 2019; Eini et al., 2020):

$$TWI = \ln\left(\frac{A_s}{\tan\beta}\right) \quad (1)$$

که A_s مساحت حوضه (m^2/m) و β زاویه شیب (درجه) است. شاخص توان آبراهه^{۱۱} (SPI) نیز بیانگر توان فرسایشی جریان بوده که با درجه شیب و مساحت حوضه آبریز رابطه مستقیم دارد و مطابق رابطه (۲) محاسبه می‌شود (Hong et al., 2018):

$$SPI = A_s \times \tan\beta \quad (2)$$

که A_s و β به ترتیب نشان‌دهنده مساحت حوضه (m^2/m) و زاویه شیب (درجه) است. شاخص موقعیت توپوگرافی^{۱۲} (TPI)، اختلاف ارتفاع هر سلول با میانگین ارتفاعات سلول‌های همسایه آن را نشان می‌دهد که طبق رابطه (۳) محاسبه می‌شود (Kanani-Sadat et al., 2019):

$$TPI = Z_0 - \frac{\sum_{i=1}^n Z_i}{n} \quad (3)$$

که Z_0 ارتفاع سلول مورد نظر و $\frac{\sum_{i=1}^n Z_i}{n}$ میانگین ارتفاع سلول‌های همسایه است. شاخص ناهمواری زمین^{۱۳} (TRI)، درواقع نشان‌دهنده اختلاف ارتفاع عنوان بین سلول‌های مجاور در یک شبکه ارتفاعی می‌باشد که طبق رابطه (۴) محاسبه می‌گردد (Kalantari et al., 2017):

$$TRI = \sqrt{\text{Abs}(\max^2 - \min^2)} \quad (4)$$

که \max و \min به ترتیب بزرگ‌ترین و کوچک‌ترین مقادیر ارتفاع سلول‌های همسایه هستند. وضعیت سنگ‌شناسی و خاک حاصل از منطقه نیز تعیین‌کننده شدت و ظرفیت نفوذپذیری خاک

جریان تجمعی^۱، جهت شیب^۲، انحنا سطح^۳، فاصله تا غسل^۴، فاصله تا جاده^۵، فاصله تا رودخانه^۶، سنگ‌شناسی^۷، جنس خاک^۸ و کاربری اراضی^۹ برای پیش‌بینی پتانسیل آب‌های زیرزمینی در منطقه مورد مطالعه استخراج شدند. تمامی ۱۷ معیار به کار برده شده (معیارهای هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی) در این پژوهش بر اساس نتایج تحقیقات پیشین در مناطق آبرفتی، نظرات کارشناسان خیره و بازدیدهای میدانی در منطقه مورد مطالعه صورت گرفته است (Nampak et al., 2014; Moghaddam et al., 2015, 2020). هر کدام از این معیارها در قالب نقشه رستری با اندازه پیکسل ۳۰ متر تهیه شدند. به‌منظور تهیه لایه DEM از تصاویر ماهواره‌ای SRTM^{۱۰} با قدرت تفکیک ۳۰ متر استفاده شد و سپس لایه‌های زاویه شیب، جهت شیب و انحنا سطح از لایه DEM ایجاد شدند. ارتفاع، تأثیر مهمی بر روی شرایط آب‌وهوایی دارد و به طور غیرمستقیم بر روی پتانسیل آب‌های زیرزمینی نیز تأثیر می‌گذارد. مثلاً ارتفاعات کمتر، شیب کمتری هم دارند و بنابراین میزان نفوذ نیز افزایش می‌یابد (Chen et al., 2020). تولید رواناب مستقیماً متناسب با شیب است و شیب کمتر باعث افزایش نفوذ آب‌های زیرزمینی می‌شود (Nampak et al., 2014). جهت شیب، جهت غالب شیب و جهت سیستم زهکشی را نشان می‌دهد؛ زیرا مدت‌زمان تابش آفتاب یا سایه، نقش مهمی در تعیین میزان رطوبت خاک دارد (Hong et al., 2018). نقشه تهیه‌شده انحنا سطح آبراهه، در حقیقت انحنا مورفولوژیکی و توپوگرافی منطقه را نمایش می‌دهد که بر روی افزایش یا کاهش جریان در سطح تأثیر می‌گذارد (Razavi Termeh et al., 2019). رودخانه‌ها، سرچشمه اصلی تغذیه آب‌های زیرزمینی در مناطق نیمه‌خشک هستند. از این‌رو فاصله تا رودخانه‌ها، یکی از عناصر هیدرولوژی اصلی است که بر روی پتانسیل آب‌های زیرزمینی تأثیر می‌گذارد (Razavi Termeh et al., 2019). جاده‌ها، وضعیت طبیعی و به‌عبارت‌دیگر شیب تعادل منطقه را به هم زده و بریدگی‌های عمودی به وجود می‌آورند (Moghaddam et al., 2015). درواقع جاده‌ها به‌مثابه سطوح نفوذناپذیر که بر تولید رواناب اثر زیادی دارند، بر میزان نفوذپذیری نیز مؤثرند (Chen et al., 2020). لایه‌های فاصله تا رودخانه و فاصله تا جاده نیز با استفاده از

- 1- Flow Accumulation
- 2- Aspect
- 3- Plane curvature
- 4- Distance to fault
- 5- Distance to road
- 6- Distance to river
- 7- Lithology
- 8- Soil type
- 9- Land use
- 10- Shuttle Radar Topography Mission

11- Stream power index

12- Topographic Position Index

13- Terrain ruggedness index

شامل ۱- آبرفتی/ شن و ماسه بادی، ۲- کوارتزیت آب و هوای معتدل، ۳- کوارتزیت شیست، ۴- گرانیت سنگ خارا زبر شیست متورق تقسیم بندی شده است که گروه اول نفوذپذیری بالای آن دارای بیشترین امتیاز و گروه چهار کمترین امتیاز تعلق می‌گیرد.

بوده و رواناب سطحی را نیز تحت تأثیر قرار می‌دهد (Kanani-Sadat et al., 2019). نقشه سنگ‌شناسی منطقه مورد مطالعه با استفاده از نقشه زمین‌شناسی کشور ایران در مقیاس ۱:۱۰۰۰۰۰ (۳۹ واحد سنگ‌شناسی) تهیه شد. نقشه وضعیت سنگ‌شناسی منطقه (جدول ۱) بر اساس میزان تخلخل و نفوذپذیری به چهار گروه اصلی

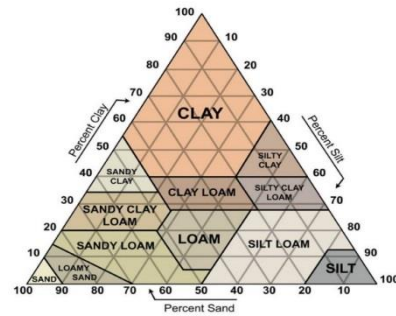
جدول ۱ - گروه‌های زمین‌شناسی و واحدهای زمین‌شناسی مربوط به آن‌ها

گروه اصلی	آبرفتی/ شن و ماسه بادی	کوارتزیت آب و هوای معتدل	کوارتزیت شیست	گرانیت سنگ خارا زبر شیست متورق
واحد زمین‌شناسی	Qft2, Od.av, Qft1, Qcf, Qbv, pgr, QCF, Pld.asv	OMgr-di, E1l, E1m, Ed.av.s, OMd.av, Pld.asv	Oa.bv, Oav, Eavt, Kuf, Kufv, Kus, PeEm, L.E-Odsv	E2c, Od.av, E2sht, E3c, L.E-Oa.bv, PeEph, sp, om1

جمع‌آوری چاه‌های مشاهداتی آب زیرزمینی

داده‌های چاه‌های آب زیرزمینی دشت بیرجند، شامل اطلاعات مربوط به تعداد چاه‌ها، موقعیت مکانی چاه‌ها و تراز آب زیرزمینی یا سطح ایستابی چاه‌ها در نیمه دوم سال ۱۳۹۸ هستند که از شرکت سهامی آب منطقه‌ای استان خراسان جنوبی اخذ شدند (شکل ۴). در این مطالعه یکسری از چاه‌ها دارای پیژومتر بوده که نقش پیژومترها اندازه‌گیری تراز سطح آب زیرزمینی می‌باشد. با توجه به این‌که وضعیت دشت بیرجند وضعیت بحرانی می‌باشد، بر روی اکثر چاه‌های کشاورزی موجود در دشت، توسط شرکت سهامی آب منطقه‌ای کنتور نصب شده تا مقدار برداشت آب زیرزمینی مشخص شود. در این پژوهش از طریق این اطلاعات، پیش‌بینی پتانسیل آب زیرزمینی دشت بیرجند انجام شده است. در مجموع تعداد ۳۷ چاه آب زیرزمینی در منطقه مورد مطالعه شناسایی شدند که به طور تصادفی به دو بخش ۷۰ درصد مجموعه داده آموزشی و ۳۰ درصد مجموعه داده اعتبارسنجی تقسیم شدند. این نسبت بین مجموعه داده‌های آموزشی و مجموعه داده‌های اعتبارسنجی توسط اکثر محققان متمرکز بر ارزیابی حساسیت مخاطرات طبیعی توصیه شده است (Nampak et al., 2014; Chen et al., 2019). 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 (متوسط سرانه برداشت معادل ۰/۵ لیتر بر ثانیه) بر اساس آزمایش پمپاژ واقعی است که در محل برای ارزیابی پتانسیل آب زیرزمینی توسط شرکت آب منطقه‌ای استان خراسان جنوبی انجام شده است. منظور از آزمایش پمپاژ واقعی، متوسط سرانه برداشت است. بر اساس آزمایش‌های پمپاژ دریافت شده رقم برداشت از چاه‌های منطقه مطالعاتی متفاوت بوده است. با توجه به اینکه در سال‌های گذشته برداشت آب‌های زیرزمینی توسط

هم‌چنین نقشه جنس خاک منطقه مورد مطالعه در مقیاس ۱:۲۵۰۰۰۰ با استفاده از تصاویر نوری سنتینل ۲ و داده‌های راداری سنتینل ۱ استخراج شد. پارامترهای جنس خاک بر اساس خصوصیات خاک (فیزیکی، شیمیایی و بیولوژیکی) تعیین می‌شوند و اغلب برای برآورد ویژگی‌های هیدرولیکی خاک استفاده می‌شود. به طور خاص، محتوای رس تأثیر زیادی در خاصیت هیدرولیکی خاک دارد و می‌تواند از نظر ظرفیت نگهداری آب در خاک آن را تفسیر کرد. جنس خاک به عنوان توزیع ذرات تشکیل دهنده خاک و بر اساس اندازه آن تعریف می‌شود. طبق طبقه‌بندی وزارت کشاورزی ایالات متحده (USDA) ذرات با قطر کمتر از ۲ میکرومتر به عنوان رس، با قطر بین ۲ تا ۵۰ میکرومتر به عنوان سیلت و با قطر بین ۰/۰۵ تا ۲ میلی‌متر در کلاس ماسه قرار می‌گیرند (شکل ۳).

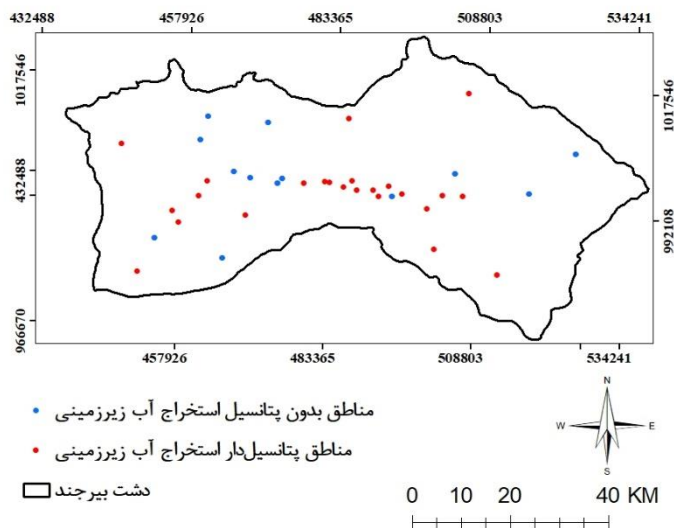


شکل ۳- طبقه‌بندی USDA جهت دسته‌بندی لایه جنس خاک

معیار جریان تجمعی نیز بیان‌کننده میزان جریانی است که از سلول‌های بالادستی به سلول مورد نظر ختم می‌شود و به کمک تحلیل‌های مناسب در نرم‌افزار SAGA GIS محاسبه می‌شود. هر چه مقدار جریان تجمعی سلول بالاتر باشد، نشان‌دهنده این است که میزان جریان نفوذی به سلول مورد نظر بیشتر است (Kanani-Sadat et al., 2019).

مشاهداتی موجود در دشت بیرجند، ۲۴ چاه با عملکرد آب زیرزمینی بیشتر از ۰/۵ لیتر بر ثانیه به‌عنوان کلاس "مناطق پتانسیل دار استخراج آب زیرزمینی" و مابقی چاه‌ها با عملکرد آب زیرزمینی کمتر از ۰/۵ لیتر بر ثانیه به‌عنوان کلاس "مناطق بدون پتانسیل استخراج آب زیرزمینی" برای فرآیند مدل‌سازی در نظر گرفته شدند (شکل ۴).

کشاورزان به‌صورت کامل آزاد بوده، اما اکنون طی سیاست‌های آبی وزارت نیرو مبنی بر کاهش و عدم برداشت می‌باشد، متوسط مقادیر برداشت در بازه‌های مختلف در چاه‌های گوناگون موردبررسی قرار گرفته که این متوسط سرانه برداشت معادل ۰/۵ لیتر بر ثانیه بر اساس داده‌های موجود و نظرات کارشناسی (شرکت سهامی آب منطقه‌ای استان خراسان جنوبی) انتخاب شده است. از بین تعداد چاه‌های



شکل ۴- دسته‌بندی چاه‌های مشاهداتی دشت بیرجند بر اساس متوسط سرانه برداشت ۰/۵ لیتر بر ثانیه

رگرسیون، γ یک ترم ثابت مثبت، w^T ماتریس معکوس ماتریس وزن اختصاص‌داده‌شده به هر یک از معیارهای شرطی چاه آب زیرزمینی، Φ تابع برازش، k تابع هسته (کرنل)، a_i بردار ورودی‌ها که شامل ۱۷ معیار شرطی و b فاصله از مبدأ ابر صفحه می‌باشد.

رگرسیون منطقی (LR)

رگرسیون منطقی یک مدل آماری برای ایجاد رابطه بین متغیرهای مستقل و وابسته می‌باشد که در این پژوهش برای پیش‌بینی پتانسیل آب زیرزمینی توسط مجموعه‌ای از معیارهای شرطی مورد استفاده قرار گرفت (Shahabi et al., 2015). مدل رگرسیون منطقی می‌تواند رابطه بین تابع منطقی $f(z)$ و پتانسیل آب زیرزمینی را به‌صورت زیر بیان کند:

$$p = \frac{1}{1 + e^{-z}} \quad (۸)$$

$$Z = \logit(p) = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + \dots + b_nx_n \quad (۹)$$

که در آن P احتمال پتانسیل آب زیرزمینی و Z ضریب خطی است که از $-\infty$ تا $+\infty$ متغیر می‌باشد. دامنه P بین ۰ و ۱ بوده به‌طوری‌که هرچه به ۱ نزدیک‌تر باشد، احتمال پتانسیل آب زیرزمینی بیشتر خواهد بود. b_0 ضریب ثابت مدل، n تعداد معیارهای شرطی

روش‌های تحقیق

روش انتخاب ویژگی کمترین مربعات ماشین بردار پشتیبان (LSSVM)

برای دستیابی به پیش‌بینی دقیق‌تر نقشه پتانسیل آب زیرزمینی، نه‌تنها انتخاب یک مدل، بلکه کیفیت معیارهای شرطی (متغیرهای مستقل) نیز مهم است (Pradhan, 2013, Guyon and Elisseeff, 2003). این امکان وجود دارد که در فرآیند مدل‌سازی، برخی از معیارهای شرطی به دلیل اثر نویز بر توانایی پیش‌بینی، نقشی بی‌ثمر ایفا کنند؛ بنابراین، شناسایی و حذف معیارهای شرطی با توانایی پیش‌بینی کم یا صفر یکی از مهم‌ترین مراحل قبل از انجام فرآیند پیاده‌سازی مدل‌های یادگیری ماشین است (Bui et al., 2015). چارچوب کلی الگوریتم LSSVM را می‌توان به‌صورت زیر بیان کرد (Tang et al., 2005):

$$y_i = [w \cdot \Phi(x_i) + b] = 1 - e_i \quad (۵)$$

$$w = \sum_{i=1}^n \alpha_i y_i \Phi(x_i) \quad (۶)$$

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i y_i k(x, x_i) + b \right] \quad (۷)$$

که در آن x_i نمونه آموزشی i ام، y_i بردار متغیر وابسته (چاه آب زیرزمینی و چاه بدون پتانسیل استخراج آب زیرزمینی)، e_i خطای

که در آن α و η به ترتیب میانگین و انحراف معیار هستند.

الگوریتم زیرفضای تصادفی (RS)

الگوریتم زیرفضای تصادفی یک رویکرد یادگیری جمعی^۱ است که برای اولین بار توسط هو در سال ۱۹۹۸ توسعه داده شد (Ho, 1998). این روش به‌عنوان یک رویکرد گروهی کارآمد شناخته می‌شود که در آن طبقه‌بندی کننده‌های چندگانه بر روی فضای مشخصه اصلاح‌شده ترکیب، آموزش و اجرا می‌شوند (Pham et al., 2017). این الگوریتم یک روش فرعی را برای طبقه‌بندی‌های پایه آموزشی ایجاد می‌کند و به‌عنوان یک مزیت، نمونه‌های مختلف از فضای مشخصه به‌جای فضای نمونه استفاده می‌شوند (Skurichina and Duin, 2002). مشکل برازش بیش از حد داده‌های آموزشی برای طبقه‌بندی در درختان تصمیم‌گیری با بالاترین دقت را می‌توان با استفاده از الگوریتم زیرفضای تصادفی برطرف کرد (Hosseini et al., 2020)؛ بنابراین، از مجموعه داده‌های آموزشی اصلی، فضاهای فرعی به‌طور تصادفی انتخاب می‌شوند و از طریق ترکیبی از روش‌های رأی‌گیری، نتیجه نهایی ایجاد می‌شود.

روش‌های ارزیابی و مقایسه شاخص‌های آماری

عملکرد و دقت پیش‌بینی مدل‌های ترکیبی توسعه‌یافته باید با استفاده از معیارهای آماری مختلف ارزیابی شود (Chung and Fabbri 1993; Nampak et al. 2014; Rahmati et al., 2018; Althuwaynee et al., 2014). 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؛ درحالی‌که NPV احتمال این است که یک پیکسل دارای پتانسیل آب زیرزمینی کم‌تر از ۰/۵ لیتر بر ثانیه به نام "بدون پتانسیل 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 به‌درستی طبقه‌بندی شده‌اند، درحالی‌که SPC نشان‌دهنده نسبت دسته‌بندی صحیح چاه‌های بدون پتانسیل آب زیرزمینی است. شاخص ACC بیانگر نسبت چاه‌های آب زیرزمینی با پتانسیل ۰/۵ لیتر بر ثانیه و چاه‌های بدون پتانسیل (پیکسل‌های آب زیرزمینی کمتر از ۰/۵ لیتر بر ثانیه) می‌باشد. RSME تفاوت بین مقادیر پیش‌بینی‌شده و واقعی چاه‌های آب زیرزمینی می‌باشد که مقادیر بالاتر نشان‌دهنده عملکرد بالاتر مدل‌ها می‌باشد. همه این معیارها بر اساس چهار نوع پیامد احتمالی شامل مثبت واقعی^۲ (TP)، منفی واقعی^۳

(متغیرهای مستقل)، b_n ضریب رگرسیون منطقی و x_n معیارهای شرطی است.

الگوریتم جنگل تصادفی (RF)

الگوریتم جنگل تصادفی یکی از رایج‌ترین الگوریتم‌های به‌کاررفته برای بررسی مشکلات طبقه‌بندی و پیش‌بینی چندگانه است که حساسیت کمی به چندخطی بودن دارد و نتایج آن از نظر داده‌های از دست‌رفته و نامتعادل نسبتاً پایدار است (De Santana et al., 2018; Quiroz et al., 2018). مدل پیش‌بینی کننده RF بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوطه استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی را با صحت بالایی انجام می‌دهد (De Santana et al., 2018). در این مدل، درختان تصادفی بردار ورودی را گرفته، آن را با هر درخت در جنگل طبقه‌بندی کرده و خروجی برجسب‌های کلاسی هستند که از اکثریت آراء دریافت شده است. برای طبقه‌بندی یک شی جدید، بردار ورودی در انتهای هر یک از درختان جنگل تصادفی قرار می‌گیرد و هر درخت به یک طبقه‌بندی منجر می‌شود. در نهایت جنگل حاصل از طبقه‌بندی که بیشترین رأی را در بین همه درخت‌های جنگل داشته باشد، انتخاب می‌شود.

الگوریتم بیز ساده (NB)

یکی از اهداف اصلی درخت تصمیم‌گیری، ایجاد یک مدل درخت مناسب برای توصیف رابطه بین متغیرهای پیش‌بینی کننده و وابسته می‌باشد (Wang et al., 2006). الگوریتم بیز ساده (NB) بر اساس قضیه بیز کار می‌کند که به نوبه خود بر اساس توزیع مستقل و گسسته سازی مقادیر مشخصه پیوسته با ایجاد یک منحنی احتمال برای هر کلاس در مجموعه داده است (Farid et al., 2014). از جمله مزایای الگوریتم NB می‌توان به سرعت بالای آموزش و طبقه‌بندی این الگوریتم و عدم محدودیت در تعداد مجموعه داده‌های آموزشی برای طبقه‌بندی اشاره کرد (Pham et al., 2015). اگر x_i بردار معیارهای شرطی و y_i بردار متغیر وابسته (چاه آب زیرزمینی و چاه بدون پتانسیل استخراج آب زیرزمینی) باشد، الگوریتم NB به‌صورت زیر محاسبه خواهد شد:

$$y_{NB} = \operatorname{argmax} P(y_i) \prod_{i=1}^{12} P(x_i, y_i) \quad , \quad (10)$$

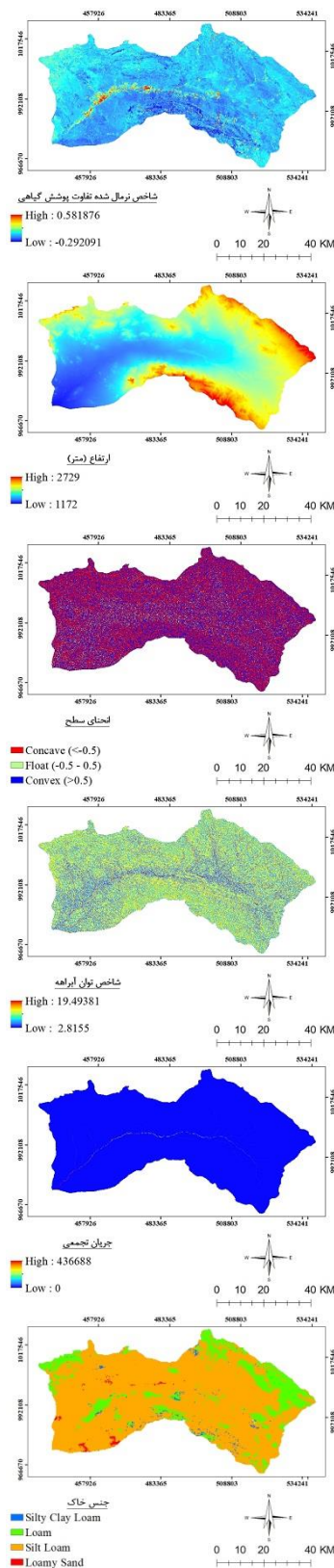
$$y_i = [\text{non well and well}]$$

که $P(y_i)$ احتمال قبلی y_i است که می‌تواند بر اساس نسبت موارد مشاهده‌شده با کلاس خروجی y_i در مجموعه داده آموزشی تخمین زده شود و $P(x_i, y_i)$ احتمال شرطی است که می‌تواند به‌صورت زیر محاسبه شود:

$$P(x_i, y_i) = \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{(x_i - \eta)^2}{2\alpha^2}} \quad (11)$$

- 1- Ensemble
- 2- True positive
- 3- True negative

(۱۴۰۰).



(TN)، مثبت کاذب^۱ (FP) و منفی کاذب^۲ (FN) محاسبه می‌شوند. TP نسبت تعداد چاه‌هایی است که به درستی به عنوان یک چاه با پتانسیل بالا طبقه‌بندی می‌شوند، در حالی که FP نسبت تعداد چاه‌هایی است که به نادرستی به عنوان یک چاه با پتانسیل بالا طبقه‌بندی می‌شوند. علاوه بر این، TN و FN تعداد چاه‌هایی هستند که به‌طور صحیح و نادرست به عنوان چاه بدون پتانسیل طبقه‌بندی می‌شوند.

$$NPV = \frac{TN}{TN + FN} \quad (12)$$

$$PPV = \frac{TP}{TP + FP} \quad (13)$$

$$SST = \frac{TP + FN}{TP + FP} \quad (14)$$

$$SPC = \frac{TN}{TN + FP} \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{pr} - x_{ac})^2} \quad (16)$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (17)$$

که در آن x_{ac} و x_{pr} به ترتیب مقادیر واقعی پیش‌بینی شده حاصل از مدل‌سازی و مشاهده واقعی و n به عنوان تعداد کل چاه‌ها است.

منحنی مشخصه عملکردی سیستم (ROC)

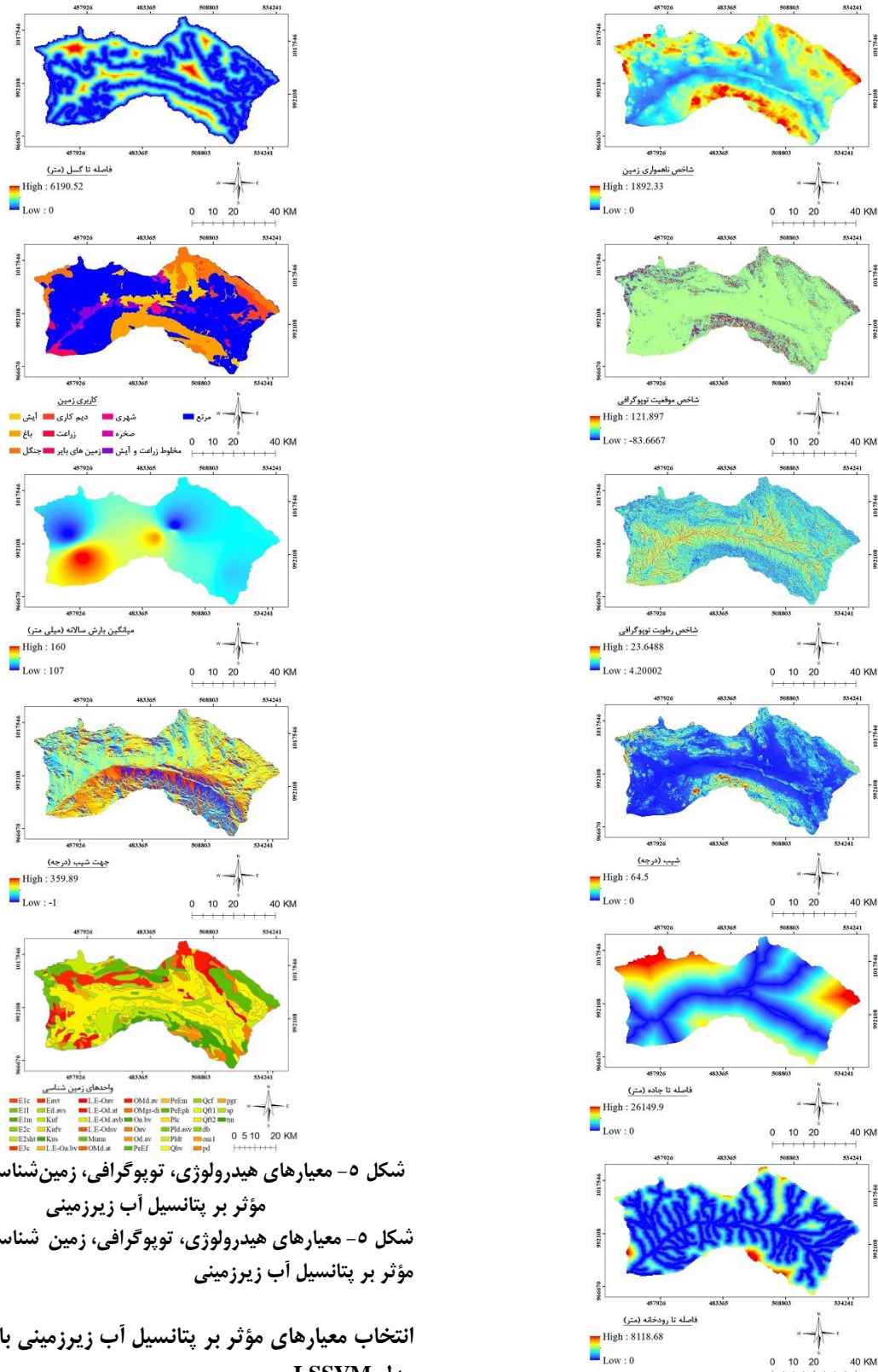
در این مطالعه، منحنی مشخصه عملکرد سیستم^۳ (ROC) برای ارزیابی عملکرد مدل‌ها به ترتیب با نرخ‌های مثبت واقعی و مثبت کاذب بر روی محور Y و محور X استفاده شده است (Tehrany et al., 2013). این روش از معیار AUC برای ارزیابی کمی استفاده می‌کند که ویژگی محور X را در مقابل حساسیت محور Y نشان می‌دهد (Chapi et al., 2017). منحنی ROC کارآمدترین روش در پیش‌بینی سیستم‌هاست که میزان دقت مدل را به‌صورت کمی برآورد می‌کند. در روش ROC، محدوده مقادیر AUC بین ۰/۵ تا ۱ می‌باشد که مقادیر نزدیک‌تر به عدد ۱ نشان‌دهنده دقت بالا می‌باشد.

نتایج و بحث

آماده‌سازی معیارها

همان‌طور که ذکر شد، در این مطالعه، از ۱۷ معیار هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی جهت پیش‌بینی پتانسیل آب زیرزمینی استفاده گردید. برای تولید نقشه‌های عوامل ذکر شده (شکل ۵)، از درون‌یابی کریجینگ استفاده شده است (افتخاری و همکاران،

- 1- False positive
- 2- False negative
- 3- Receiver Operating Characteristic

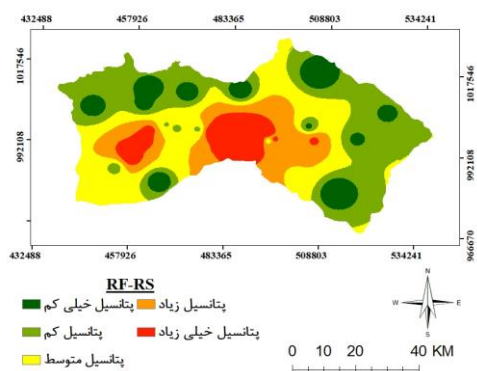
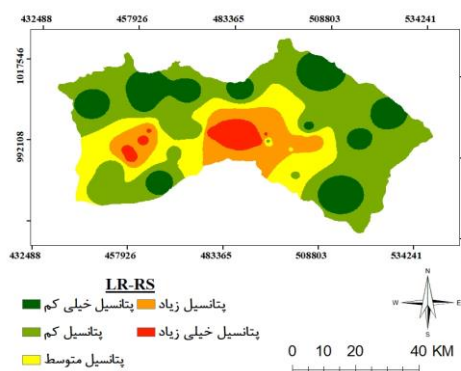
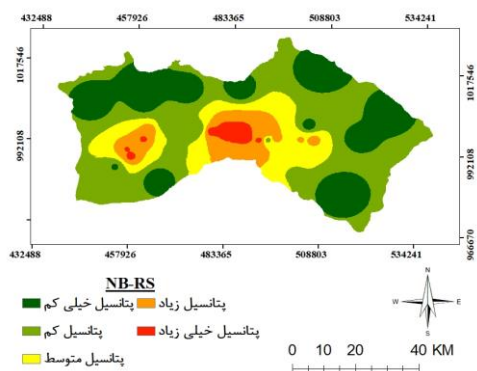


شکل ۵- معیارهای هیدرولوژی، توپوگرافی، زمین شناسی و محیطی مؤثر بر پتانسیل آب زیرزمینی
 شکل ۵- معیارهای هیدرولوژی، توپوگرافی، زمین شناسی و محیطی مؤثر بر پتانسیل آب زیرزمینی

انتخاب معیارهای مؤثر بر پتانسیل آب زیرزمینی با استفاده از مدل LSSVM

انتخاب مهم ترین معیارهای مؤثر بر پتانسیل آب زیرزمینی با استفاده از روش LSSVM در شکل ۶ نشان داده شده است. این

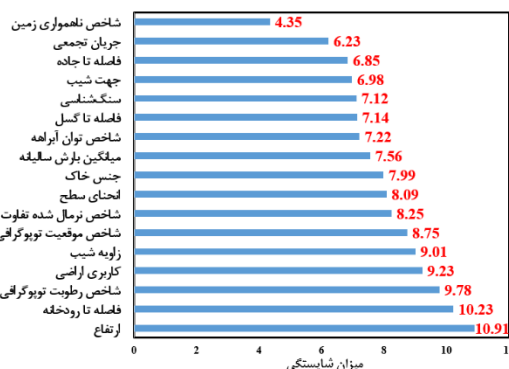
نتایج نشان می‌دهد که معیار ارتفاع با مقدار متوسط شایستگی^۱ (AM) از اهمیت بیشتری برای پیش‌بینی پتانسیل آب زیرزمینی برخوردار است و پس از آن، سایر معیار نیز دارای متوسط شایستگی بیشتر از صفر می‌باشند، لذا تمامی این معیارهای برای پیش‌بینی پتانسیل آب زیرزمینی در مطالعه حاضر در نظر گرفته شدند.



شکل ۷- نقشه پیش‌بینی پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه بر اساس مدل‌های NB-RS، LR-RS، RF-RS

مقایسه قابلیت پیش‌بینی مدل‌ها

قابلیت‌های صحت‌سنجی و پیش‌بینی مدل‌های ترکیبی توسعه یافته برای پیش‌بینی پتانسیل آب زیرزمینی با استفاده از مجموعه داده‌های آموزشی و اعتبارسنجی در جدول ۲ ارزیابی شدند.



شکل ۶- اهمیت‌دهی معیارهای هیدروولوژی، توپوگرافی، زمین‌شناسی و محیطی با استفاده از روش LSSVM

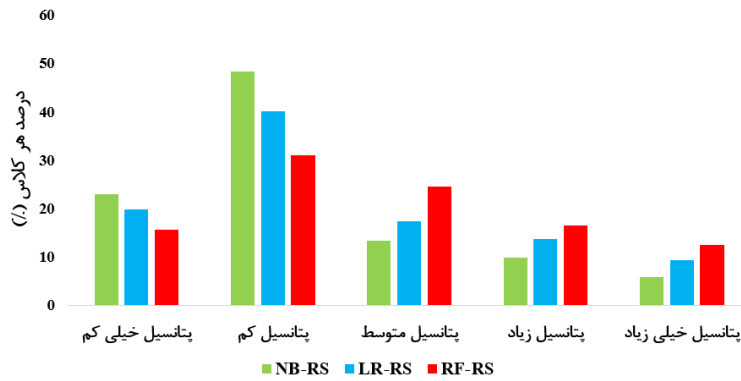
تهیه نقشه پیش‌بینی پتانسیل آب زیرزمینی

در این مطالعه، نقشه‌های پیش‌بینی مناطق بالقوه آب زیرزمینی در منطقه مورد مطالعه بر اساس مدل‌های NB-RS، LR-RS، RF-RS و استفاده از مجموعه داده‌های آموزشی و اعتبارسنجی پیش‌بینی شدند. نقشه میزان پتانسیل پیش‌بینی شده با استفاده از روش طبقه‌بندی شکست طبیعی^۲ در پنج کلاس شامل پتانسیل خیلی کم، پتانسیل کم، پتانسیل متوسط، پتانسیل زیاد و پتانسیل خیلی زیاد طبقه‌بندی گردید. شکل ۷ نقشه پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه را نشان می‌دهد که با استفاده از مدل‌های موردنظر پیش‌بینی شده‌اند.

هم‌چنین، در شکل ۸، درصد هر یک از کلاس‌های پتانسیل آب زیرزمینی پیش‌بینی شده توسط مدل‌های NB-RS، LR-RS و RF-RS نشان داده شده است. نتایج نشان داد که در مدل RF-RS، ۳ کلاس با پتانسیل آب زیرزمینی خیلی زیاد، زیاد و متوسط، درصد بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های NB-RS و LR-RS تحت پوشش قرار می‌دهد. علاوه بر این در مدل NB-RS، ۲ کلاس با پتانسیل آب زیرزمینی خیلی کم و کم، درصد بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های RF-RS و LR-RS دربر گرفته است. در واقع می‌توان نتیجه گرفت که به طور متوسط ۹/۱۶٪ از منطقه مورد مطالعه در کلاس با پتانسیل آب زیرزمینی خیلی زیاد قرار گرفته است.

1- Average merit

2- Natural breaks



شکل ۸- درصد هر یک از کلاس‌های پتانسیل آب‌های زیرزمینی پیش‌بینی شده بر اساس مدل‌های NB-RS، LR-RS، RF-RS

جدول ۲- مقایسه عملکرد مدل‌های ترکیبی با استفاده از مجموعه داده‌های آموزشی و اعتبارسنجی

پارامتر	RF-RS		LR-RS		NB-RS	
	آموزشی	اعتبارسنجی	آموزشی	اعتبارسنجی	آموزشی	اعتبارسنجی
PPV	۰/۸۴۳	۰/۸۱۲	۰/۷۷۱	۰/۷۰۲	۰/۶۶۲	۰/۵۵۳
NPV	۰/۸۷۱	۰/۷۱۲	۰/۷۷۲	۰/۶۷۵	۰/۷۲۶	۰/۵۸۲
SST	۰/۸۷۷	۰/۷۴۱	۰/۷۳۷	۰/۷۰۱	۰/۶۳۳	۰/۵۱۱
SPC	۰/۸۳۱	۰/۸۱۲	۰/۸۱۱	۰/۶۷۱	۰/۷۵۲	۰/۶۱۷
ACC	۰/۸۵۶	۰/۷۶۱	۰/۷۷۲	۰/۶۸۸	۰/۶۹۷	۰/۵۷۱
RMSE	۰/۳۱۱	۰/۳۶۲	۰/۳۸۸	۰/۴۰۸	۰/۴۱۲	۰/۴۲۳

۰/۵ لیتر بر ثانیه در منطقه مورد مطالعه است. نتایج مجموعه داده‌های آموزشی و اعتبارسنجی نشان داد که مدل RF-RS قابلیت بهتری را نسبت به سایر مدل‌ها در پیش‌بینی پتانسیل آب زیرزمینی در منطقه مورد مطالعه دارد. عملکرد بهتر مدل RF می‌تواند به دلیل توانایی آن در مدل‌سازی پایگاه داده‌های بزرگ و توانایی ادغام متغیرهای ورودی زیاد باشد (Rahmati et al., 2015).

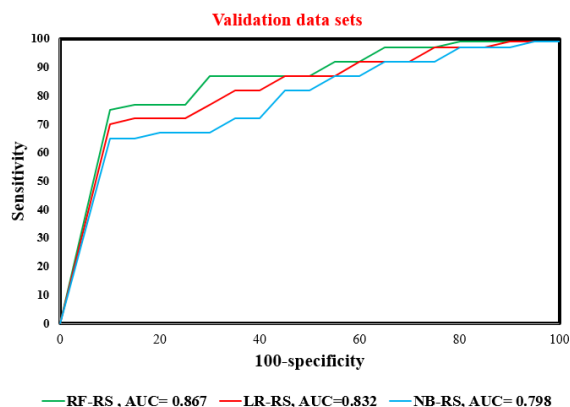
اعتبارسنجی مدل‌های پیش‌بینی پتانسیل آب زیرزمینی

شکل ۹ منحنی ROC را برای مجموعه داده‌های آموزشی و داده‌های اعتبارسنجی مدل‌های RF-RS، LR-RS و NB-RS نشان می‌دهد. به طور کلی، قابلیت مدل‌ها برای پیش‌بینی پتانسیل آب زیرزمینی توسط مجموعه داده‌های آموزشی ارزیابی می‌شود؛ در حالی که مجموعه داده‌های اعتبارسنجی، مهارت پیش‌بینی مدل را نشان می‌دهند. طبق شکل ۹ برای مجموعه داده‌های آموزشی، مدل RF-RS دارای بالاترین مقدار AUC (۰/۹۱۲) می‌باشد و پس از آن مدل‌های LR-RS (AUC= ۰/۸۹۲) و NB-RS (AUC= ۰/۸۵۳) قرار دارند. همچنین، برای مجموعه داده‌های اعتبارسنجی، مدل RF-RS قوی‌تر بوده و دارای دقت بالاتری (AUC= ۰/۸۶۷) نسبت به مدل‌های LR-RS (AUC= ۰/۸۳۲) و NB-RS (AUC= ۰/۷۹۸) قرار دارد.

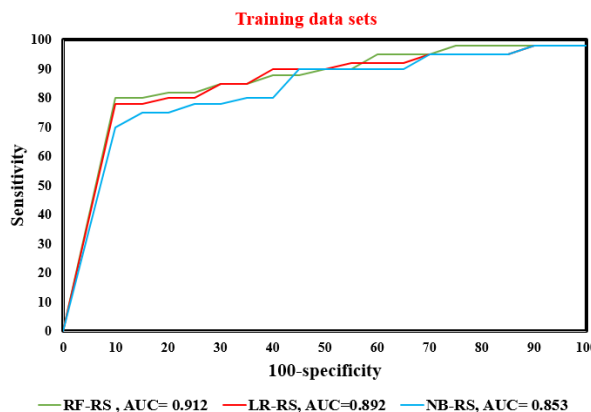
بر این اساس برای طبقه‌بندی پیکسل‌های بالقوه پتانسیل آب زیرزمینی، روش RF-RS بیش‌ترین مقدار SST را برای مجموعه داده‌های آموزشی و اعتبارسنجی نشان می‌دهد (۰/۸۷۷ و ۰/۷۴۱)؛ بنابراین، می‌توان مشاهده کرد که مدل ترکیبی RF-RS قابلیت پیش‌بینی بالاتری نسبت به مدل‌های دیگر دارد که به‌طور نادرستی چاه‌های آب زیرزمینی را با پتانسیل بیش از ۰/۵ لیتر بر ثانیه در منطقه مورد مطالعه طبقه‌بندی کرده است. همچنین، روش RF-RS بهترین مقدار SPC را برای مجموعه داده‌های آموزشی و اعتبارسنجی نشان می‌دهد (۰/۸۳۱ و ۰/۸۱۲). این بدان معنی است که مدل ترکیبی RF-RS قابلیت صحت‌سنجی و پیش‌بینی بهتری را نسبت به دیگر مدل‌ها در طبقه‌بندی صحیح چاه‌های آب زیرزمینی با پتانسیل کم‌تر از ۰/۵ لیتر بر ثانیه در منطقه مورد مطالعه دارد. علاوه بر این، مدل RF-RS دارای بالاترین مقدار نرخ پیش‌بینی مثبت و نرخ پیش‌بینی منفی برای مجموعه داده‌های آموزشی و اعتبارسنجی می‌باشد که پس از آن مدل‌های LR-RS و NB-RS قرار گرفته‌اند. همچنین مدل گروهی RF-RS نیز بالاترین مقدار ACC را برای مجموعه داده‌های آموزشی و اعتبارسنجی دارد (۰/۸۵۶ و ۰/۷۶۱). بر این اساس می‌توان گفت که مدل RF-RS قادر به طبقه‌بندی صحیح ۷۶٪ از چاه‌های آب زیرزمینی دارای پتانسیل بیشتر از ۰/۵ لیتر بر ثانیه و همچنین کم‌تر از

است که این نتایج با یافته‌های تحقیق (Rahmati et al. 2016) و (Chen et al. 2019) مبنی بر برتری روش RF مطابقت دارد.

می‌باشد؛ بنابراین، می‌توان نتیجه گرفت که اگرچه تمام مدل‌های به کاررفته قدرت پیش‌بینی خوبی دارند، اما عملکرد و دقت مدل RF- RS برای پیش‌بینی پتانسیل آب زیرزمینی در منطقه مورد مطالعه بهتر



(ب)



(الف)

شکل ۹- منحنی ROC و مقدار AUC برای روش‌های پیشنهادی تحقیق (الف) مجموعه داده‌های آموزشی (ب) مجموعه داده‌های اعتبارسنجی

می‌دهد که عوامل انتخاب شده برای پیش‌بینی پتانسیل آب زیرزمینی در منطقه مورد مطالعه بسیار مناسب بوده که ممکن است در مناطق مشابه نیز مناسب باشند. عملکرد مدل‌های توسعه‌یافته با استفاده از معیارهای آماری استاندارد (RMSE، SPC، SST، NPV، PPV) و ROC ارزیابی شد. نتایج تجزیه و تحلیل نشان داد که تمام مدل‌های ترکیبی جدید توسعه‌یافته قابلیت پیش‌بینی خوبی را نشان داده‌اند، اما عملکرد مدل RF-RS در پیش‌بینی دقیق مناطق بالقوه آب زیرزمینی بهترین است و پس از آن مدل‌های LR-RS و NB-RS برای نگاشت دقیق مناطق بالقوه آب زیرزمینی قرار دارند. در مناطق دیگر نیز این مدل را می‌توان برای نقشه‌برداری 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 توسعه‌یافته بر اساس مدل‌های مکانی و روش‌های یادگیری ماشین قابلیت خوبی در ارزیابی پتانسیل آب زیرزمینی صحیح یک منطقه با کمترین داده‌ها دارد.

منابع

افتخاری، م.، اسلامی نژاد، س. ا.، حاجی الیاسی، ع. و اکبری، م. ۱۴۰۰. ارزیابی زمین‌آماری با شاخص کیفیت آب زیرزمینی به منظور آسامیدن (DGWQI) در آبخوان دشت بیرجند، نشریه محیط‌زیست و مهندسی آب، ۷(۲): ۲۶۸-۲۷۸

نتیجه‌گیری

ارزیابی مناسب پتانسیل آب زیرزمینی یک منطقه برای استفاده بهینه آن با اتخاذ فن‌آوری و روش‌های جدید مطلوب است. هدف اصلی این مطالعه ارائه مدل‌های یادگیری ماشین جمعی مبتنی بر الگوریتم زیرفضای تصادفی جهت پیش‌بینی پتانسیل آب زیرزمینی است. لذا در این مطالعه، سه مدل ترکیبی یادگیری ماشین به نام‌های RF-RS، LR-RS، NB-RS جهت پیش‌بینی پتانسیل آب زیرزمینی در دشت بیرجند توسعه داده شده است. محققان قابلیت و کاربرد روش‌های یادگیری ماشین جمعی را در پیش‌بینی مکانی پتانسیل آب‌های زیرزمینی مناطق مختلف اثبات کرده‌اند (Razavi Termeh et al., 2019; Chen et al., 2020). در این مطالعه، از ۱۷ معیار هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی جهت پیش‌بینی پتانسیل آب زیرزمینی در دشت بیرجند استفاده گردید. روش انتخاب ویژگی کمترین مربعات ماشین بردار پشتیبان (LSSVM) جهت تعیین معیارهای مؤثر به منظور افزایش عملکرد الگوریتم‌های یادگیری ماشین مورد استفاده قرار گرفت. نتایج نشان داد که تمامی ۱۷ معیار به کاربرده شده در این تحقیق دارای متوسط شایستگی بیشتر از صفر می‌باشند، لذا تمامی این معیارهای برای ساخت مدل‌های پتانسیل آب زیرزمینی در مطالعه حاضر استفاده شدند. انتخاب مؤثر این ۱۷ معیار دقت پیش‌بینی پتانسیل آب زیرزمینی را به وسیله کاهش نویز و تناسب بیش از حد برای مجموعه داده آموزشی افزایش داد. این نشان

- constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20: 832–844.
- Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X. and 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*. 625:575-588.
- Hosseini, F.S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H. and Haghghi, A.T. 2020. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of the total environment*. 711: 135161.
- Jancewicz, K., Migoń, P. and Kasprzak, M. 2019. Connectivity patterns in contrasting types of tablelandsandstone relief revealed by Topographic Wetness Index. *Science of the Total Environment*. 656:1046-1062.
- Kalantari, Z., Ferreira, C.S.S., Walsh, R.P.D., Ferreira, A.J.D. and Destouni, G. 2017. Urbanization development under climate change: hydrological responses in a peri-urban Mediterranean catchment. *Land Degradation & Development*. 28 (7): 2207–2221.
- Kanani-Sadat, Y., Arabsheibani, R., Karimipour, F. and 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*. 572:17-31.
- 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.
- Manap, M.A., Sulaiman, W.N.A., Ramli, M.F., Pradhan, B. and Surip, N. 2013. A knowledge-driven GIS modeling technique for groundwater potential mapping at the upper Langat Basin, Malaysia. *Arabian Journal of Geosciences*. 6:1621–1637
- 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.
- Nampak, H., Pradhan, B. and Manap, M.A. 2014. Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology*. 513: 283–300.
- Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H. 2014. A novel ensemble decision tree-based CHI-squared automatic interaction detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides*. 11:1063–1078.
- Bui, D.T., Pradhan, B., Revhaug, I., Nguyen, D.B., Pham, H.V. and Bui, Q.N. 2015. A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). *Geomat Nat Haz Risk*. 6:243–271.
- Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D., Pham, B.T. and Khosravi, K. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software*. 95:229-245.
- Chen, W., Tsangaratos, P., Ilia, 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., Ilia, 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.
- Chung, C.J.F. and Fabbri, A.G. 1993. The representation of geoscience information for data integration. *Nonrenewable Resources*. 2:122–139.
- De Santana, F.B., de Souza, A.M. and Poppi, R.J. 2018. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 191:454-462.
- Eftekhari, M. and Akbari, M. 2020. Evaluation of the SINTACS-LU model capability in the analysis of aquifer vulnerability potential in semi-arid regions. *Journal of Applied Research in Water and Wastewater*. 7(2):111-119.
- Eini, M., Kaboli, H.S., Rashidian, M. and Hedayat, H. 2020. Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction*: 101687.
- Farid, D.M., Zhang, L., Rahman, C.M., Hossain, M.A. and Strachan, R. 2014. Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*. 41:1937–1946.
- Guyon, I. and Elisseeff, A. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*. 3:1157–1182.
- Ho, T.K. 1998. The random subspace method for

- Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A.N. and Melesse, A.M. 2018. Groundwater spring potential modelling: comprising the capability and robustness of three different modeling approaches. *Journal of Hydrology*. 565:248–261.
- Rahmati, O., Pourghasemi, H.R. and Melesse, A.M. 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran. *Catena*. 137: 360–372.
- Rahmati, O., Samani, A.N., Mahdavi, M., Pourghasemi, H.R. and Zeinivand, H. 2015. Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arabian Journal of Geosciences*. 8: 7059–7071.
- Razavi-Termeh, S.V., Sadeghi-Niaraki, A. and Choi, S.M. 2019. Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. *Water*. 11(8): 1596.
- Shahabi, H., Hashim, M. and Ahmad, B.B. 2015. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. *Environmental Earth Sciences*. 73:8647–8668.
- Skurichina, M. and Duin, R.P. 2002. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis and Applications*. 5:121–135.
- Tang, X., Ou, Z., Su, T., Sun, H. and Zhao, P. 2005. Robust precise eye location by adaboost and svm techniques. In: *International Symposium on Neural Networks*. Springer: 93–98.
- Tehrany, M.S., Pradhan, B. and Jebur, M.N. 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: 69–79.
- Wang, L.M., Li, X.L., Cao, C.H. and Yuan, S.M. 2006. Combining decision tree and naive Bayes for classification. *Knowledge-Based Systems*. 19: 511–515.
- Neshat, A., Pradhan, B., Pirasteh, S. and Shafri, H.Z.M. 2014. Estimating groundwater vulnerability to pollution using a modified DRASTIC model in the Kerman agricultural area, Iran. *Environmental Earth Sciences*. 71: 3119–3131.
- 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: 158–172.
- Osati, K., Koeniger, P., Salajegheh, A., Mahdavi, M., Chapi, K. and Malekian, A. 2014. Spatiotemporal patterns of stable isotopes and hydrochemistry in springs and river flow of the upper Karkheh River basin, Iran. *Isotopes in Environmental and Health Studies*. 50: 169–183.
- Pham, B.T., Bui, D.T., Prakash, I. and Dholakia, M. 2017. Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena*. 149:52–63.
- Pham, B.T., Bui, D. T., Indra, P. and Dholakia, M. 2015. A comparison study of predictive ability of support vector machines and naive bayes tree methods in landslide susceptibility assessment at an area between Tehri Garhwal and Pauri Garhwal, Uttarakhand state (India) using GIS. In: *national symposium on geomatics for digital India and annual conventions of ISG & ISRS, Jaipur (India)*.
- Pradhan, B. 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*. 57: 350–365.
- Prasad, P., Loveson, V.J., Kotha, M. and 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.
- Quiroz, J.C., Mariun, N., Mehrjou, M.R., Izadi, M., Misron, N. and Mohd Radzi, M.A. 2018. Fault detection of broken rotor bar in LS-PMSM using random forests. *Measurement*. 116:273-280.
- Rahmati, O., Naghibi, S.A., Shahabi, H., Bui, D.T.,

Determination of Groundwater Potential Using Ensemble Machine Learning Models in GIS (Case Study: Birjand plain)

S. A. Eslaminezhad¹, M. Eftekhari², M. Akbari^{3*}, A. Haji Elyasi⁴

Received: Sep.29, 2021

Accepted: Nov.15, 2021

Abstract:

Predicting the potential of groundwater is very important for the systematic development and planning of water resources. The main purpose of this study was to develop ensemble machine learning models including random forest (RF), logistic regression (LR) and Naïve Bayes (NB) by random subspace Classifier (RS) algorithm to predict groundwater potential areas in Birjand plain. Therefore, for implementation, geo-hydrological data of 37 groundwater wells (Number of wells, location of wells and groundwater level or Water table) and 17 hydrology, topographic, geological and environmental criteria were used. The least squares support vector machine (LSSVM) feature selection method used to determine the effective criteria to increase the performance of machine learning algorithms. Finally, groundwater potential prediction maps were prepared using RF-RS, LR-RS and NB-RS models. The performance of these models evaluated using the area under the curve (AUC) and other statistical indicators. The results showed that the RF-RS hybrid model (AUC = 0.867) has a very high predictability for groundwater potential in the study area. It was also found that the elevation criterion is most important in predicting the groundwater potential in the study area. The results of the present study can be useful for making appropriate decisions and planning regarding the optimal use of groundwater resources.

Keywords: Groundwater potential, Logistic regression, Naïve Bayes, Random Forest, Random subspace

1- MSc. Graduate, Department of Surveying Engineering, Faculty of Surveying Engineering and Spatial Information, University of Tehran, Tehran, Iran

2- MSc. Graduate, Civil Engineering, Water and Hydraulic Structures, Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran

3- Associate Professor, Department of Civil Engineering, University of Birjand, Birjand, Iran

4- MSc. Graduate, Department of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran

(*- Corresponding Author Email: Moakbari@birjand.ac.ir)