



Assessment of Machine Learning Models in GIS For Predicting Groundwater in Semi-Arid Regions of Eastern Iran

Mobin Eftekhari^{1✉} | Ali Haji Elyasi² | Seyed Ahmad Eslaminezhad³

1. Ph.D. Student, Water Engineering Department, University of Birjand, Birjand, Iran.
2. Ph.D. Student, Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran.
3. MSc Graduate of Tehran University and GIS expert of Gas Company of South Khorasan Province, Birjand, Iran..

✉Corresponding Author: Mobineftekhari@yahoo.com

Received:
13 February 2024

Accepted:
15 March 2024

Published:
19 March 2024

Keywords:

*Birjand Plain,
predictive maps,
random forest,
decision tree,
support vector machine.*

Extended abstract

Introduction

Groundwater is recognized as a vital and critical resource worldwide, playing a key role in providing water for agriculture, industry, livestock, and human communities. In Iran, due to climate change and over-exploitation of groundwater resources, many areas have become dry, leading to a serious water shortage in various provinces of the country. In previous research, researchers have investigated machine learning methods to identify the best model for predicting groundwater potential. The main objectives of this research include identifying high-performance machine learning models for prediction with minimal available data. This research is distinguished by the use of advanced machine learning methods based on the Random Subspace Classifier algorithm. In this study, machine learning models such as DT, RF, and SVM were used to predict potential groundwater zones in the Birjand Plain region, which represents an innovation in current research.

Method and Material

The present research method is analytical-descriptive and is based on the defined applied objectives. All data processing processes were performed in the Arc GIS and Google Earth Engine software environments, and quantitative calculations and method development were performed in Python.

Cite this article: Eftekhari, M., Haji Elyasi, A. & Eslaminezhad, S.A (2024). Assessment of machine learning models in GIS for predicting groundwater in semi-arid regions of eastern Iran. *Journal of Aquifer and Qanat Title*, 4 (2), 49-66. DOI: <http://doi.org/10.22077/jaaq.2024.7282.1062>



Copyright: © 2024 by the authors. Licensee Journal of Aquifer and Qanat. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

This study consists of five main steps: collecting groundwater observation well data in the study area to define dependent variables, extracting spatial criteria that directly affect groundwater potential to define independent variables, using the LSSVM algorithm to determine effective criteria, predicting potential groundwater zones using machine learning approaches (SVM, RF, DT), and evaluating and comparing the performance of machine learning methods.

Method and Material

The present research method is analytical-descriptive and is based on the defined applied objectives. All data processing processes were performed in the Arc GIS and Google Earth Engine software environments, and quantitative calculations and method development were performed in Python.

This study consists of five main steps: collecting groundwater observation well data in the study area to define dependent variables, extracting spatial criteria that directly affect groundwater potential to define independent variables, using the LSSVM algorithm to determine effective criteria, predicting potential groundwater zones using machine learning approaches (SVM, RF, DT), and evaluating and comparing the performance of machine learning methods.

Case Study

The Birjand plain, which is part of the Lut Desert watershed, is located 485 kilometres south of Mashhad in geographical coordinates and is located in the northern part of the Baqeran Heights. This area is limited by longitudes between $41^{\circ}58'$ and $46^{\circ}59'$ east and latitudes between $35^{\circ}32'$ and $8^{\circ}33'$ north (Fig 1).

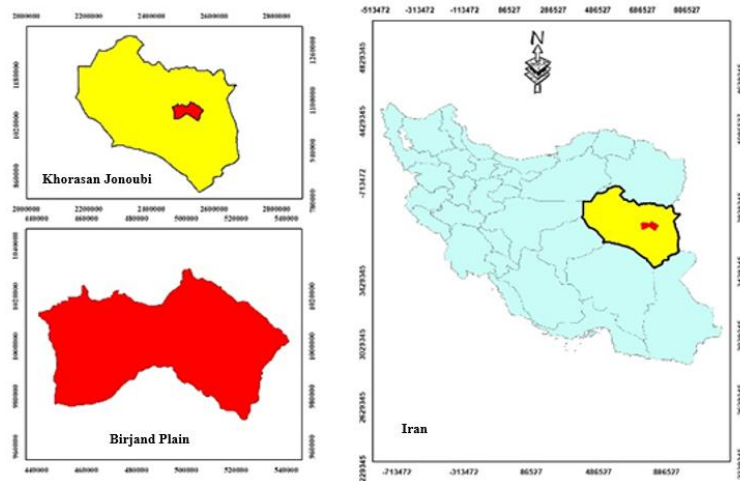


Fig 1. The location of South Khorasan province and Birjand plain.

Methodology

In this research, 17 different criteria including hydrological, topographic, geological, and environmental criteria such as precipitation, normalized difference vegetation index (NDVI), topographic wetness index (TWI), stream power index (SPI), terrain ruggedness index (TRI), topographic position index (TPI), digital elevation model (DEM), slope angle, flow accumulation, slope aspect, surface curvature, distance to fault, distance to road, distance to a river, lithology, soil texture, and land use were extracted to determine the groundwater potential in the study area. Raster map criteria with 30-meter precision have been defined. SRTM satellite images with 30-meter resolution are used to create the DEM layer. Then, slope angle, slope direction, and surface curvature layers are generated based on the DEM layer. The land use layer of the study area was extracted from Landsat 8 satellite images through the Google Earth Engine platform. Considering the feature of Google Earth Engine that provides images in time series, 162 images were received as an average in the second 6-month period of 1398 in this research, and image processing operations were performed on them and the land uses of the study area were identified. The soil texture map of the study area was extracted using Sentinel-2 optical images and Sentinel-1 radar data at a scale of 1:250,000. Soil texture parameters are determined based on soil properties (physical, chemical, and biological). The data related to groundwater wells in the Birjand plain, including the number of wells, their location, and the groundwater level or water table in the second half of 1398, were obtained from the South Khorasan Regional Water Company (Fig 2).

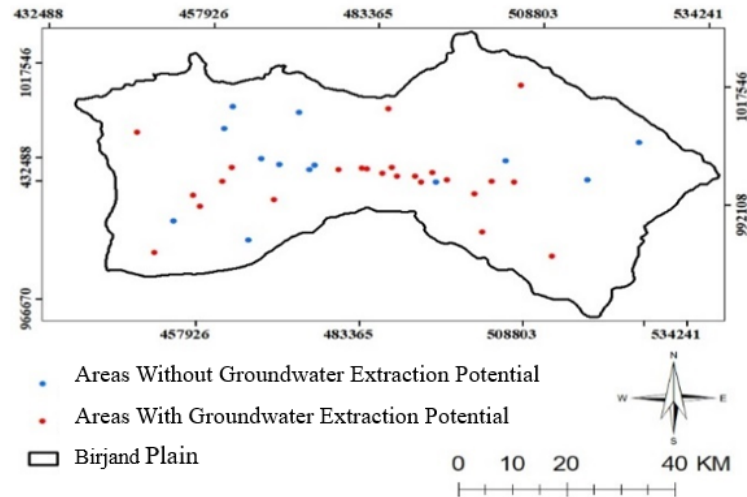


Fig 2. Classification of observation wells in Birjand Plain based on the average per capita harvest of 0.5 litres per second.

In total, 37 groundwater wells were identified in the study area, which was randomly divided into two parts: 70% for the training dataset and 30% for the validation dataset. This ratio between the training dataset and the validation dataset is recommended by most researchers focused on assessing the sensitivity of natural hazards. In the present study, a groundwater yield of 5/0 litres per second was used as a threshold value for selecting groundwater.

Results

In this study, groundwater potential maps were predicted using RF, DT and SVM models and using training and validation datasets. The predicted potential map was classified into five classes using the natural break classification method, including very low potential, low potential, medium potential, high potential and very high potential (Fig 3).

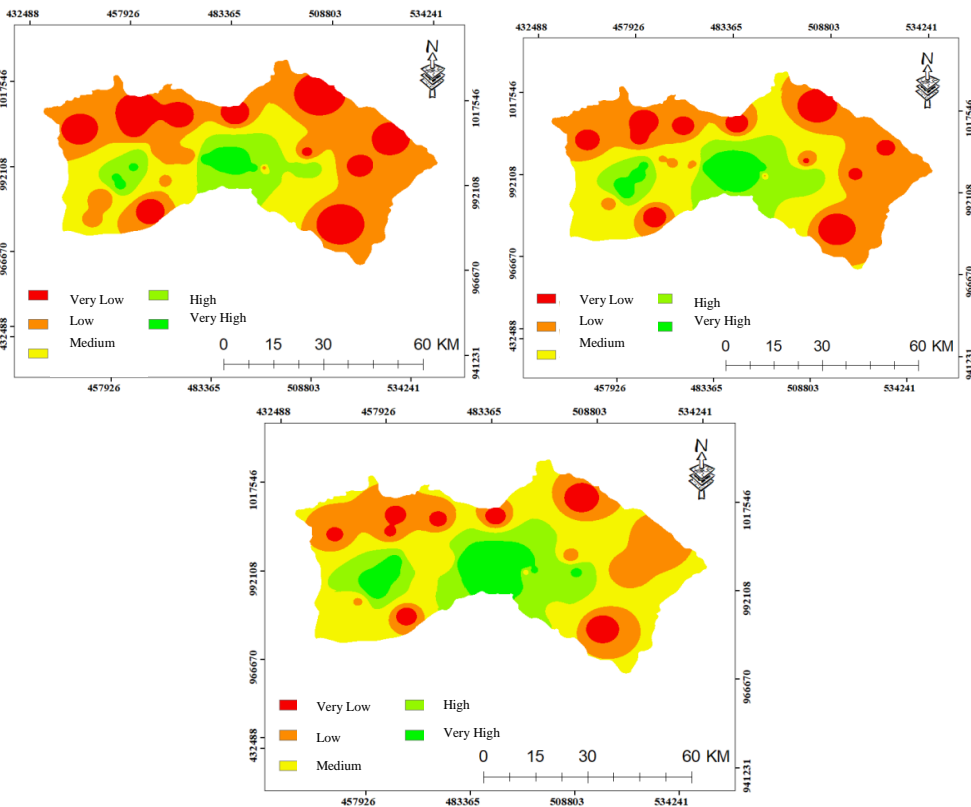


Fig 3. The prediction map of the underground water potential of the study area is based on the DT, RF & SVM models.

The validation and prediction capabilities of the developed hybrid models for predicting groundwater potential were evaluated using the training and validation datasets in Table 1. Fig 4 also shows the ROC curve for the training and validation datasets of the DT, RF and SVM models.

Table 1. Comparison of performance of hybrid models using training and validation data sets

| Models | R ² | RMSE |
|--------|----------------|-------|
| SVM | 0.81 | 1.205 |
| DT | 0.89 | 0.705 |
| RF | 0.86 | 1.055 |

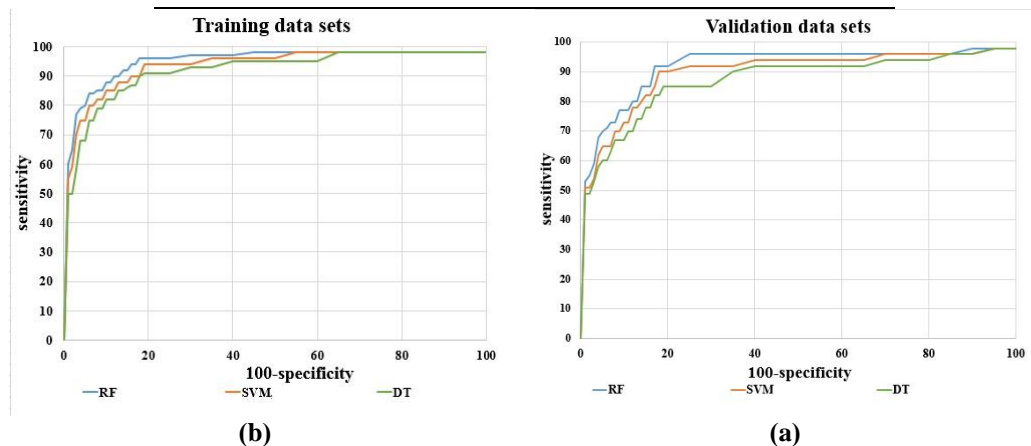


Fig 4. ROC curve and AUC value for the proposed research methods (a) training dataset and (b) validation dataset.

Conclusion

The results of this study indicated that all 17 criteria used have significance greater than zero, and therefore all were employed for constructing groundwater potential models. The effective selection of these 17 criteria enhanced the prediction accuracy of groundwater potential by reducing noise and overfitting for the training dataset. This outcome suggests that the chosen criteria are highly suitable for predicting groundwater potential in the investigated area and might be beneficial in similar regions.

The performance of the developed models was evaluated using standard statistical criteria (PPV, NPV, SST, SPC, RMSE, and ROC). The analysis results revealed that all the newly developed hybrid models possess good prediction capability. However, the DT model exhibited the best performance in accurately predicting potential groundwater zones, followed by the RF and SVM models for precise mapping of potential groundwater zones. This model can also be utilized in other regions for accurate mapping and proper identification of potential groundwater zones, which will contribute to effective groundwater management of an area.

Therefore, it can be concluded that the developed hybrid models based on spatial models and machine learning techniques demonstrate a good capability for accurate groundwater potential assessment of a region with minimal data.



ارزیابی مدل‌های یادگیری ماشین در GIS جهت پیش‌بینی آب زیرزمینی مناطق نیمه خشک شرق ایران

مبین افتخاری^۱ | علی حاجی الیاسی^۲ | سید احمد اسلامی نژاد^۳

۱. دانشجوی دکترای منابع آب، گروه علوم و مهندسی آب، دانشگاه بیرجند، بیرجند، ایران.
۲. دانشجوی دکتری، گرایش مهندسی آب و سازه‌های هیدرولیکی، دانشکده‌ی مهندسی عمران، دانشکدگان فنی، دانشگاه تهران، تهران، ایران.
۳. دانش‌آموخته کارشناسی ارشد دانشگاه تهران و کارشناس GIS شرکت گاز، استان خراسان جنوبی، بیرجند، ایران.

✉ نویسنده مسئول: Mobineftekhari@yahoo.com

چکیده

پیش‌بینی پتانسیل آب‌های زیرزمینی جهت توسعه و برنامه‌ریزی سیستماتیک منابع آب بسیار حیاتی است. هدف اصلی این تحقیق، توسعه مدل‌های یادگیری ماشینی از جمله جنگل تصادفی (RF)، درخت تصمیم (DT) و ماشین بردار پشتیبان (SVM) برای پیش‌بینی مناطق دارای پتانسیل آب زیرزمینی در دشت بیرجند است. بنابراین، برای انجام این مطالعه، داده‌های ژئوهیدرولوژیکی مربوط به ۳۷ چاه آب زیرزمینی (شامل تعداد و موقعیت چاه‌ها و سطح آب زیرزمینی) و ۱۷ معیار هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی مورد استفاده قرار گرفت. روش انتخاب ویژگی از طریق کمترین مربعات ماشین بردار پشتیبان جهت تعیین معیارهای مؤثر برای بهبود عملکرد الگوریتم‌های یادگیری ماشین به کار گرفته شد. در نهایت، نقشه‌های پیش‌بینی پتانسیل آب زیرزمینی با استفاده از مدل‌های DT، RF و SVM تهیه شدند و عملکرد این مدل‌ها با استفاده از سطح زیر منحنی (AUC) و سایر شاخص‌های آماری مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل DT (AUC=0.89) توانایی پیش‌بینی بسیار بالایی برای پتانسیل آب زیرزمینی در منطقه مورد مطالعه دارد و معیار ارتفاع به عنوان مهم‌ترین عامل در پیش‌بینی پتانسیل آب زیرزمینی در این منطقه شناخته شد. نتایج این مطالعه می‌تواند به عنوان راهنمایی برای تصمیم‌گیری و برنامه‌ریزی مناسب در استفاده بهینه از منابع آب زیرزمینی مورد استفاده قرار گیرد.

تاریخ دریافت: ۱۴۰۲/۱۱/۲۴

تاریخ پذیرش: ۱۴۰۲/۱۲/۲۵

تاریخ انتشار: ۱۴۰۲/۱۲/۲۹

کلیدواژه‌ها:

دشت بیرجند،

نقشه‌های پیش‌بینی،

جنگل تصادفی،

درخت تصمیم،

ماشین بردار پشتیبان.

مقدمه

اطمینان (CF) و مدل تابع باور شواهد (EBF) بود (Razavi Termeh et al., 2019).

۱۵ معیار مؤثر که شامل معیارهای توپوگرافی، هیدرولوژی، زمین‌شناسی و محیطی برای مدل‌سازی انتخاب شده‌اند، نتایج نشان داده‌اند که مقدار AUC با استفاده از روش‌های FR-RF، EBF-RF، CF-RF، FR-LMT، EBF-LMT و CF-LMT به ترتیب ۰/۹۲۷، ۰/۹۲۴، ۰/۹۱۷، ۰/۹۰۶، ۰/۸۸۵ و ۰/۸۳ بوده است. چن و همکاران با استفاده از ترکیب سه الگوریتم داده‌کاوی RF، بیز ساده (NB) و ماشین بردار پشتیبان (SVM) با الگوریتم ژنتیک (GA) نقشه پتانسیل آب زیرزمینی در شهرستان ووقی کشور چین را تهیه کرده‌اند (Chen et al., 2019). برای پیاده‌سازی، ۱۶ عامل متعدد مانند 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 شده‌اند. نتایج نشان داده‌اند که مدل‌های ترکیبی بهینه شده از نظر دقت و کارایی (مقدار AUC) نسبت به مدل‌های اصلی برتر هستند. پراساد و همکاران با استفاده از سه مدل یادگیری ماشین شامل مدل RF، درخت رگرسیون تقویت‌شده (BRT) و مدل SVM نقشه پیش‌بینی مناطق بالقوه آب زیرزمینی در سواحل غربی هند را تهیه کرده‌اند (Prasad et al., 2020).

برای دستیابی به این اهداف، ۱۴ معیار مهم برای تأثیر بر آب زیرزمینی مانند ارتفاع، شیب، جهت شیب، طول شیب، انحنای نیمرخ، انحنای سطح، شاخص رطوبت توپوگرافی، فاصله تا رودخانه، فاصله تا جاده، سنگ‌شناسی، جنس خاک، کاربری زمین، شاخص نرمال شده تفاوت پوشش گیاهی و بارندگی استخراج شده‌اند. نتایج نشان می‌دهند که مدل RF (AUC=۰/۹۴) دقت بیشتری نسبت به مدل‌های BRT و SVM در پیش‌بینی مناطق بالقوه آب زیرزمینی دارد. چن و همکاران مدل‌های یادگیری ماشین شامل آدابوست (AB)، مدل بگینگ (BA)، مدل زیرفضای تصادفی (RS)، مدل دگینگ (DA) و مدل جنگل

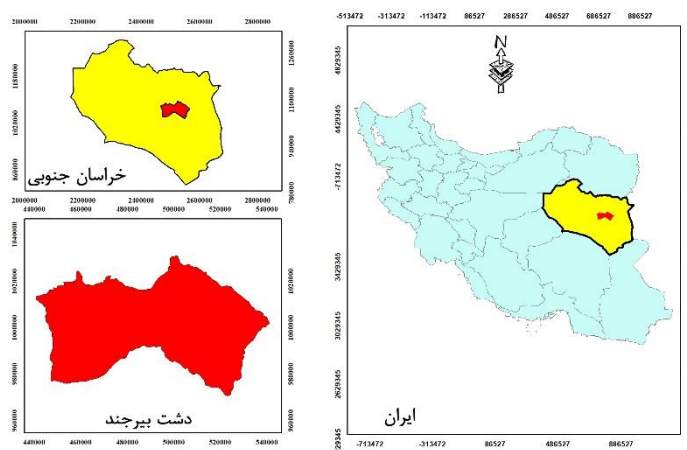
آب زیرزمینی به عنوان یکی از منابع حیاتی و بحرانی در سراسر جهان شناخته شده است که نقش مهمی در تأمین آب مورد نیاز برای کشاورزی، صنعت، دامداری و جوامع انسانی دارد (Foster et al., 2013). در ایران، به دلیل تغییرات اقلیمی و برداشت بی‌رویه از منابع آب زیرزمینی، بسیاری از مناطق خشک شده‌اند و این موضوع منجر به کمبود جدی آب در استان‌های مختلف کشور شده است (Abbaspour et al., 2009). پیدایش و حرکت آب زیرزمینی تحت تأثیر عوامل متعددی قرار دارد، از جمله 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 (Deepika et al., 2013). برای شناسایی مناطق بالقوه آب زیرزمینی، استفاده از روش‌هایی که به مدیران کمک می‌کنند تا از این منابع به صورت مؤثرتری استفاده کنند، ضروری است (Tashayo et al., 2020). این روش‌ها برای توسعه آینده، مدیریت و جلوگیری از کاهش منابع آب زیرزمینی مورد نیاز می‌باشند. بسیاری از مطالعات از فناوری‌های سیستم اطلاعات مکانی (GIS) و سنجش از دور (RS)، برای ارزیابی پتانسیل آب‌های زیرزمینی استفاده کرده‌اند. برای مثال، رحمتی و همکاران برای تهیه نقشه پتانسیل آب زیرزمینی در حوضه آبریز شهرستان مهران از مدل‌های جنگل تصادفی (RF) و حداکثر آنتروپی (ME) در بستر GIS استفاده کرده‌اند. معیارهایی مانند ارتفاع، زاویه شیب، جهت شیب، انحنای سطح، تراکم زهکشی، فاصله تا رودخانه، شاخص رطوبت توپوگرافی، کاربری زمین، سنگ‌شناسی و بافت خاک برای پیاده‌سازی استفاده شده‌اند. نتایج اعتبارسنجی نشان داد که مدل ME (AUC=۰/۹۱) 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 (LMT) با مدل‌های آماری نسبت فرکانس (FR)، ضریب

روش‌های پیشرفته یادگیری ماشین مبتنی بر الگوریتم طبقه‌بندی‌کننده زیرفضای تصادفی متمایز است. در این مطالعه، مدل‌های یادگیری ماشین مانند RF، DT و SVM برای پیش‌بینی مناطق بالقوه آب‌های زیرزمینی در منطقه دشت بیرجند استفاده شده‌اند که نشان‌دهنده نوآوری در تحقیقات فعلی است.

مواد و روش‌ها

۱- منطقه مورد مطالعه

دشت بیرجند، که بخشی از حوضه آبریز کویر لوت است، در مسیر جغرافیایی ۴۸۵ کیلومتر جنوب مشهد واقع شده و در بخش شمالی ارتفاعات باقران قرار دارد. این منطقه در طول‌های جغرافیایی بین $58^{\circ}41'$ تا $59^{\circ}46'$ شرقی و عرض‌های جغرافیایی بین $32^{\circ}35'$ تا $33^{\circ}08'$ شمالی محدود می‌شود (شکل ۱).



شکل ۱- موقعیت مکانی استان خراسان جنوبی و دشت بیرجند

Fig 1- The location of South Khorasan province and Birjand plain

حداقل ارتفاع ۱۱۷۲ متر در خروجی دشت (منطقه فدشک) است.

تهیه لایه‌های اطلاعاتی منطقه مورد مطالعه

عوامل مؤثر بر پتانسیل آب‌های زیرزمینی

در این تحقیق، ۱۷ معیار مختلف از جمله معیارهای هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی مانند بارندگی، شاخص نرمال شده تفاوت پوشش گیاهی (NDVI)، شاخص رطوبت توپوگرافی (TWI)، شاخص توان آبراهه (SPI)، شاخص ناهمواری زمین (TRI)، شاخص موقعیت توپوگرافی (TPI)، مدل رقومی ارتفاعی (DEM)، زاویه شیب، جریان تجمعی، جهت شیب،

چرخشی (ROF) را توسعه داده‌اند و با استفاده از الگوریتم J48DT برای پیش‌بینی مکانی پتانسیل آب زیرزمینی (در فصل بهار) در شهرستان ووقی کشور چین استفاده کرده‌اند (Chen et al., 2020). برای پیاده‌سازی روش‌های پیشنهادی، از ۱۳ معیار مؤثر شامل معیارهای توپوگرافی، هیدرولوژی و محیطی استفاده شده‌اند. آن‌ها نتیجه گرفتند که مدل ترکیبی ROF-J48DT (AUC=۷/۷۹) دقت بیشتری در

پیش‌بینی مکانی پتانسیل آب زیرزمینی عمل دارد. در تحقیقات موجود، محققان روش‌های یادگیری ماشین را برای تعیین بهترین مدل در زمینه پیش‌بینی ظرفیت آب‌های زیرزمینی مورد بررسی قرار داده‌اند. اهداف اصلی این تحقیق شامل شناسایی مدل‌های یادگیری ماشین با کارایی بالا برای پیش‌بینی با کمترین داده‌های موجود است. این پژوهش با توجه به استفاده از

حوضه آبریز بیرجند با وسعت $3406/72$ کیلومتر مربع است، که از آن سهم دشت و ارتفاعات به ترتیب $900/63$ و $2506/9$ کیلومتر مربع را تشکیل می‌دهند. رودخانه شاهرود با طول ۱۰۸ کیلومتر از شرق به غرب، دشت را زهکشی می‌کند و آب را به کویر لوت می‌ریزد. این رودخانه در طول سال به جز مواقع خشک سیلابی، به تأمین نیازهای کشاورزی، صنعتی و شرب شهرستان بیرجند کمک می‌کند (Eftekhari et al., 2019). با توجه به بارش‌های فصلی و مربوط بودن داده‌های این تحقیق به نیمه دوم سال ۱۳۹۸، جریان آب در رودخانه شاهرود مشاهده شده است. حداکثر ارتفاع حوضه از سطح دریا ۲۷۲۹ متر در ارتفاعات باقران و

تعادل منطقه را به هم زده و بریدگی‌های عمودی به وجود می‌آورند. در واقع جاده‌ها به مثابه سطوح نفوذناپذیر که بر تولید رواناب اثر زیادی دارند، بر میزان نفوذپذیری نیز مؤثرند (Eslaminezhad et al., 2022). در دشت بیرجند تعداد محدودی پیزومتر واقع شده است که تعدادی از آن‌ها در مناطق شهری و در سایر مراتع قرار دارند. همان طور که ذکر 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. لایه فاصله تا گسل نیز از نقشه زمین‌شناسی در مقیاس ۱:۱۰۰۰۰۰ استخراج شده است.

لایه‌ی کاربری اراضی منطقه مورد بررسی از طریق تصاویر ماهواره‌ای لندست ۸ از طریق پلتفرم گوگل ارث انجین استخراج شده است. با توجه به ویژگی گوگل ارث انجین که تصاویر را به صورت سری زمانی در اختیار قرار می‌دهد، در این تحقیق ۱۶۲ تصویر به صورت میانگین‌گیری شده در بازه ۶ ماهه دوم سال ۱۳۹۸ دریافت شد و عملیات پردازش تصاویر بر روی آنها صورت گرفت و کاربری‌های منطقه مطالعاتی مشخص شدند. کاربری‌های مختلف می‌توانند در میزان نفوذ آب، نفوذ، فرسایش خاک و تبخیر و تعرق دخیل باشند. به عنوان مثال، در کاربری‌های شهری نفوذی کمتری وجود دارد (به دلیل آسفالت خیابان‌ها); بنابراین، هر چه فاصله از جاده بیشتر باشد، نفوذ آب در سطح زمین به دلیل عدم وجود آسفالت بیشتر است و تأثیر به سزایی بر ظرفیت آب‌های زیرزمینی دارد. علاوه بر این، باندهای ۴ و ۵ تصاویر ماهواره‌ای لندست ۸ برای تولید لایه NDVI استفاده شده‌اند (Sanaeinejad et al., 2014). برای محاسبه شاخص میانگین بارش سالیانه، از چهار ایستگاه سینوپتیک در منطقه مورد بررسی استفاده شده است. رطوبت خاک و اشباع سطح نیز شاخص رطوبت توپوگرافی (TWI) را نشان

انحنای سطح، فاصله تا گسل، فاصله تا جاده، فاصله تا رودخانه، سنگ‌شناسی، جنس خاک و کاربری اراضی برای تعیین پتانسیل آب‌های زیرزمینی در منطقه مورد بررسی استخراج شده‌اند. این معیارها بر اساس تحقیقات قبلی در مناطق آبرفتی، نظرات کارشناسان با تجربه و بازدیدهای میدانی در منطقه مورد بررسی اعمال شده‌اند (Eslaminezhad et al., 2022).

معیارهای مورد استفاده در نقشه رستری با دقت ۳۰ متری تعریف شده‌اند. برای ساخت لایه DEM از تصاویر ماهواره‌ای SRTM با دقت تفکیک ۳۰ متر استفاده می‌شود. پس از آن، لایه‌های زاویه شیب، جهت شیب و انحنای سطح بر اساس لایه DEM تولید می‌شوند. ارتفاع سطح تأثیر قابل توجهی بر شرایط آب‌وهوایی دارد و به طور غیرمستقیم بر روی ظرفیت آب‌های زیرزمینی تأثیر می‌گذارد. به عنوان مثال، مناطق با ارتفاع کمتر شیب کمتری دارند و به همین دلیل نفوذ آب بیشتر است (Uuemaa et al., 2020).

رواناب‌ها به طور مستقیم با شیب در ارتباط هستند و کاهش شیب منجر به افزایش نفوذ آب‌های زیرزمینی می‌شود. جهت شیب، جهت غالب شیب و جهت سیستم زهکشی را نشان می‌دهند. زیرا مدت زمان تابش آفتاب یا سایه نقش مهمی در تعیین میزان رطوبت خاک دارد. نقشه تهیه شده انحنای سطح آبراهه، در واقع انحنای مورفولوژیکی و توپوگرافی منطقه را نمایش می‌دهد که بر روی تغییر جریان در سطح تأثیر می‌گذارد (Xiong et al., 2014).

رودخانه‌ها و سرچشمه‌ها در مناطق نیمه‌خشک به عنوان منابع اصلی تغذیه آب‌های زیرزمینی عمل می‌کنند. بنابراین، فاصله یکی از عناصر مهم هیدرولوژی است که بر روی ظرفیت آب‌های زیرزمینی تأثیر دارد (Yenehun et al., 2020). جاده‌ها به عنوان سطوح طبیعی و به عبارت دیگر شیب منطقه، تعادل شیب را تعیین می‌کنند و بریدگی‌های عمودی ایجاد می‌کنند. در واقع، جاده‌ها به عنوان سطوح نفوذناپذیر که بر تولید رواناب تأثیر زیادی دارند، بر نفوذپذیری نیز تأثیر می‌گذارند. لایه‌های فاصله تا رودخانه و فاصله تا جاده با استفاده از لایه توپوگرافی در مقیاس ۱:۵۰۰۰۰ تهیه شده‌اند. جاده‌ها، وضعیت طبیعی و به عبارت دیگر شیب

بیشترین امتیاز است و گروه چهارم کمترین امتیاز را دارد.

نقشه جنس خاک منطقه مورد بررسی با استفاده از تصاویر نوری سنتینل ۲ و داده‌های راداری سنتینل ۱ در مقیاس ۱:۲۵۰۰۰۰ استخراج شده است. پارامترهای جنس خاک بر اساس خصوصیات خاک (فیزیکی، شیمیایی و بیولوژیکی) تعیین می‌شوند و اغلب برای برآورد ویژگی‌های هیدرولیکی خاک استفاده می‌شود. محتوای رس تأثیر زیادی در خاصیت هیدرولیکی خاک دارد و می‌توان از نظر ظرفیت نگهداری آب در خاک آن را تفسیر کرد. جنس خاک به عنوان توزیع ذرات تشکیل‌دهنده 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 ۲).

معیار جریان تجمعی نشان‌دهنده میزان جریانی است که از سلول‌های بالادستی به سلول موردنظر ختم می‌شود و با استفاده از 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 موردنظر بیشتر است (Miralles et al., 2019).

جمع‌آوری چاه‌های مشاهداتی آب زیرزمینی

داده‌های مربوط به چاه‌های آب زیرزمینی دشت بیرجند، شامل تعداد چاه‌ها، موقعیت مکانی آن‌ها و سطح ایستابی آب زیرزمینی یا تراز آب در نیمه دوم سال ۱۳۹۸ است که از شرکت سهامی آب منطقه‌ای استان خراسان جنوبی اخذ شده‌اند (شکل ۳). در این مطالعه، برخی از چاه‌ها پیژومتر دارند که برای اندازه‌گیری سطح آب زیرزمینی استفاده می‌شوند. با توجه به وضعیت بحرانی دشت بیرجند، کنترل‌های آب بر روی چاه‌های کشاورزی نصب شده‌اند تا برداشت آب زیرزمینی را مشخص کنند. با استفاده از این اطلاعات، پتانسیل آب زیرزمینی دشت بیرجند پیش‌بینی شده است. اطلاعات ژئوهیدرولوژیکی ۳۷ حلقه چاه آب زیرزمینی مربوطه به نیمه دوم سال ۱۳۹۸ می‌باشند که

می‌دهند که بر اساس رابطه (۱) محاسبه می‌شود (Li et al., 2018).

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

که A_s مساحت حوضه (m^2/m) و β زاویه شیب (درجه) است. شاخص توان آبراهه^۱ (SPI) نیز بیانگر توان فرساینده‌گی جریان بوده که با درجه شیب و مساحت حوضه آبریز رابطه مستقیم دارد و مطابق رابطه (۲) محاسبه می‌شود:

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

که A_s و β به ترتیب نشان‌دهنده مساحت حوضه (m^2/m) و زاویه شیب (درجه) است. شاخص موقعیت توپوگرافی^۲ (TPI)، اختلاف ارتفاع هر سلول با میانگین ارتفاعات سلول‌های همسایه آن را نشان می‌دهد که طبق رابطه (۳) محاسبه می‌شود:

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

که Z_0 ارتفاع سلول موردنظر و $\frac{\sum_{i=1}^n Z_i}{n}$ میانگین ارتفاع سلول‌های همسایه است. شاخص ناهمواری زمین^۳ (TRI)، در واقع نشان‌دهنده اختلاف ارتفاع عنوان بین سلول‌های مجاور در یک شبکه ارتفاعی می‌باشد که طبق رابطه (۴) محاسبه می‌گردد:

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

که \max و \min به ترتیب بزرگ‌ترین و کوچک‌ترین مقادیر ارتفاع سلول‌های همسایه هستند. وضعیت سنگ شناسی و خاک حاصل از منطقه نیز تعیین‌کننده شدت و ظرفیت نفوذپذیری خاک بوده و رواناب سطحی را نیز تحت تأثیر قرار می‌دهد (Ferreira et al., 2015).

نقشه زمین‌شناسی کشور ایران در مقیاس ۱:۱۰۰۰۰۰ (۳۹ واحد سنگ‌شناسی) تهیه شده است. نقشه وضعیت سنگ‌شناسی منطقه (جدول ۱) بر اساس میزان تخلخل و نفوذپذیری به چهار گروه اصلی تقسیم‌بندی شده است: آبرفتی / شن و ماسه بادی، کوارتزیت آب‌وهوای معتدل، کوارتزیت شیبست، گرانیت سنگ خارا زبر شیبست متورق. گروه اول با نفوذپذیری بالا دارای

1- Stream power index

2- Topographic Position Index

3- Terrain ruggedness index

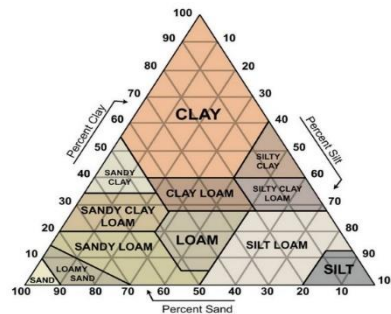
نیمه‌خشک هستند. از این رو فاصله تا رودخانه‌ها، یکی از عناصر هیدرولوژی اصلی است که بر روی پتانسیل آب‌های زیرزمینی تأثیر می‌گذارد.

با توجه به بارش‌های فصلی، جریان آب در رودخانه شاهرود جریان داشته است. در واقع رودخانه‌ها، سرچشمه اصلی تغذیه آب‌های زیرزمینی در مناطق

جدول ۱. گروه‌های زمین‌شناسی و واحدهای زمین‌شناسی مربوط به آن‌ها

Table 1. Geological groups and their related geological units.

| گروه اصلی (Main group) | آبرفتی / شن و ماسه بادی (Alluvium/wind sand) | کوارتزیت آب‌وهوای معتدل (Moderate climate quartzite) | کوارتزیت شیست (Quartzite schist) | گرانیت سنگ خارا زبر شیست متورق (Moturgh schist coarse barbed granite) |
|--------------------------------|--|--|--|---|
| واحد زمین‌شناسی (Geology unit) | Qft2, Od.av, Qft1, Qcf, Qbv, pgr, QCF, Pld.asv | OMgr-di, E11, E1m, Ed.av, 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 |

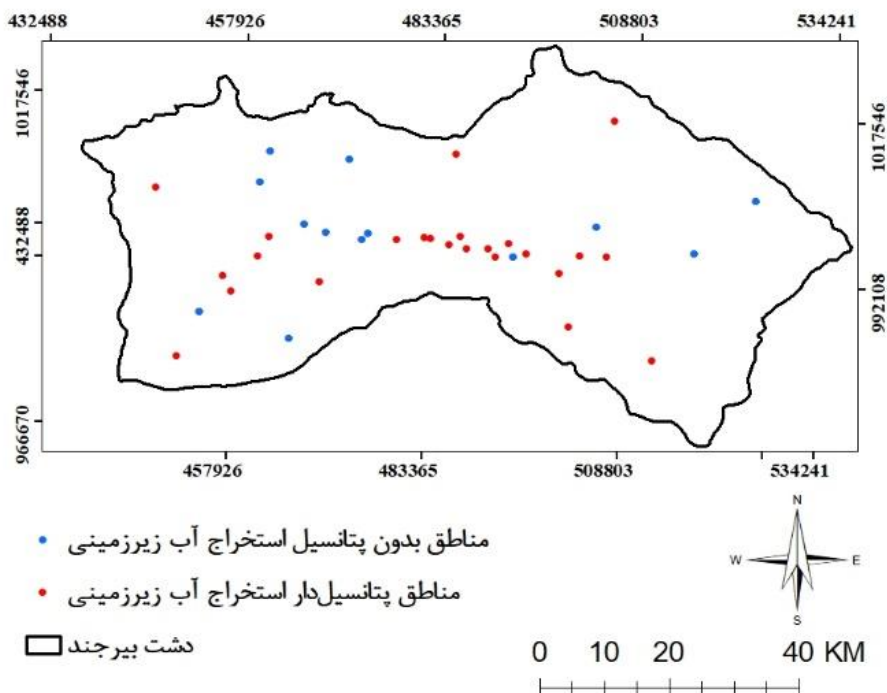


شکل ۲. طبقه بندی USDA جهت دسته‌بندی لایه جنس خاک.

Fig 2. USDA classification to classify the soil type.

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 تغییرات سیاست‌های آبی وزارت نیرو مبنی بر کاهش و عدم برداشت، متوسط مقادیر برداشت در بازه‌های مختلف در چاه‌های گوناگون مورد بررسی قرار گرفته است. 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 مشاهداتی موجود در دشت بیرجند، ۲۴ چاه با عملکرد آب زیرزمینی بیشتر از ۰/۵ لیتر بر ثانیه به عنوان کلاس "مناطق پتانسیل‌دار استخراج آب زیرزمینی" و مابقی چاه‌ها با عملکرد آب زیرزمینی کمتر از ۰/۵ لیتر بر ثانیه به عنوان کلاس "مناطق بدون پتانسیل استخراج آب زیرزمینی" برای فرآیند مدل‌سازی در نظر گرفته شدند (شکل ۳).

در مجموع، ۳۷ چاه آب زیرزمینی در منطقه مورد مطالعه شناسایی شده‌اند که به طور تصادفی به دو بخش ۷۰ درصد مجموعه داده‌های آموزشی و ۳۰ درصد مجموعه داده‌های اعتبارسنجی تقسیم شده‌اند. این نسبت بین مجموعه داده‌های آموزشی و مجموعه داده‌های اعتبارسنجی توسط اکثر محققان متمرکز بر ارزیابی حساسیت مخاطرات طبیعی توصیه شده است (Eslaminezhad et al., 2022). 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.



شکل ۳. دسته‌بندی چاه‌های مشاهداتی دشت بیرجند بر اساس متوسط سرانه برداشت ۰/۵ لیتر بر ثانیه.

Fig 3. Classification of observation wells in Birjand Plain based on the average per capita harvest of 0.5 litres per second.

۱۷ معیار شرطی و b فاصله از مبدأ ابر صفحه می‌باشد. رگرسیون منطقی یک روش آماری است که برای برقراری رابطه بین متغیرهای مستقل و وابسته استفاده می‌شود. در این مطالعه، این روش برای پیش‌بینی پتانسیل آب زیرزمینی با استفاده از مجموعه‌ای از معیارهای شرطی اعمال شده است (Guido et al., 2006). که در آن P احتمال پتانسیل آب زیرزمینی و Z ضریب خطی است که از $(-)$ تا $(+)$ متغیر می‌باشد. دامنه P بین ۰ و ۱ بوده به طوری که هرچه به ۱ نزدیک‌تر باشد، احتمال پتانسیل آب زیرزمینی بیشتر خواهد بود. b_0 ضریب ثابت مدل، n تعداد معیارهای شرطی (متغیرهای مستقل)، b_n ضریب رگرسیون منطقی و x_n معیارهای شرطی است (روابط ۸ و ۹).

$$y_i = [w \cdot \Phi(x_i) + b] = 1 - e_i \quad (5)$$

$$w = \sum_{i=1}^n \alpha_i y_i \Phi(x_i) \quad (6)$$

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i y_i k(x, x_i) + b \right] \quad (7)$$

$$P = \frac{1}{1 + e^{-z}} \quad (8)$$

$$Z = \log \text{it}(p) = \ln \left(\frac{p}{1-p} \right) = b_0 + b_1 x_1 + \dots + b_n x_n \quad (9)$$

روش‌های مدل‌سازی

روش انتخاب ویژگی کمترین مربعات ماشین بردار پشتیبان (LSSVM)

برای ارائه یک نقشه پتانسیل آب زیرزمینی با دقت بالاتر، انتخاب مدل مناسب تنها یکی از مراحل است. علاوه بر این، ارزیابی کیفیت معیارهای شرطی (متغیرهای مستقل) نیز ضروری است. در برخی موارد، معیارهای شرطی ممکن است به دلیل تأثیر نویز بر توانایی پیش‌بینی، نقشی بی‌ثمر ایفا کنند (Hussein et al., 2017). بنابراین، شناسایی و حذف معیارهای شرطی با توانایی پیش‌بینی کم یا صفر یکی از مراحل حیاتی قبل از پیاده‌سازی مدل‌های یادگیری ماشین است. چارچوب کلی الگوریتم LSSVM به شرح روابط ۵ الی ۷ است.

که در آن x_i نمونه آموزشی n م، y_i بردار متغیر وابسته (چاه آب زیرزمینی و چاه بدون پتانسیل استخراج آب زیرزمینی)، e_i خطای رگرسیون، γ یک ترم ثابت مثبت، W^T ماتریس معکوس ماتریس وزن اختصاص داده شده به هر یک از معیارهای شرطی چاه آب زیرزمینی، Φ تابع برازش، k تابع هسته (کرنل)، a_i بردار ورودی‌ها که شامل

الگوریتم جنگل تصادفی

عملکرد خوبی دارد، اما ممکن است به بیش‌برازش و حساسیت به تغییرات کوچک در داده‌ها مبتلا شود (Hilario et al., 2006). در کل، الگوریتم درخت تصمیم به عنوان یک ابزار قدرتمند در تصمیم‌گیری مسائل پیچیده شناخته می‌شود.

الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) یک فناوری نوین در حوزه یادگیری ماشین است که از روش‌های آماری غیر پارامتریک و نیمه‌نظارت شده بهره می‌برد. (Dong et al., 2021). این روش در سال ۱۹۹۲ میلادی توسط بوسر و همکاران براساس تئوری یادگیری آماری معرفی شد. طی سال‌های پس از آن، آنها تئوری ابرصفحه بهینه را به عنوان یک طبقه‌بندی‌کننده خطی معرفی کردند و طبقه‌بندی‌کننده‌های غیرخطی را با استفاده از توابع کرنل معرفی نمودند (Yaman & Cengiz, 2021). اصول اساسی ماشین بردار پشتیبان که امروزه به عنوان یک روش معتبر شناخته می‌شود، به نتایج کارهای بوسر و همکاران بازمی‌گردد و در نهایت، گسترش ماشین بردار پشتیبان بر اساس رگرسیون نیز در سال ۱۹۹۵ توسط وپنیک حاصل شد. (Sansone et al., 2013). این روش به منظور حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار دارند، از مدل طبقه‌بندی ماشین بردار پشتیبان استفاده می‌کند و مدل رگرسیون بردار پشتیبان در حل مسائل پیش‌بینی به کار می‌رود (Bhavsar & Panchal, 2012). ویژگی اصلی این روش، توانایی بالا در استفاده از تعداد نمونه‌های آموزشی کمتر و در عین حال دستیابی به دقت بالاتر نسبت به سایر روش‌ها است (Shao & Lunetta, 2012).

روش‌های ارزیابی عملکرد مدل

شاخص‌های آماری

برای ارزیابی عملکرد و دقت مدل‌های ترکیبی توسعه‌یافته، باید از معیارهای آماری متنوعی استفاده کنیم. این معیارها شامل معیارهایی مانند PPV (Precision Value)، NPV (Negative Predictive Value)، SST (Separation Strength) و SPC (Specificity) هستند. در این مطالعه، PPV به احتمال داشتن پتانسیل آب زیرزمینی بیش از ۵ لیتر بر ثانیه در یک پیکسل اشاره دارد، که نتایج آزمایش مثبت هستند.

روش جنگل تصادفی یک رویکرد نوین و قدرتمند در حوزه فناوری داده‌کاوی است که بهبودهای چشمگیری در این زمینه ارائه داده است. تکنیک جنگل تصادفی به عنوان یک مدل گسترش‌یافته از روش طبقه‌بندی و رگرسیون درختی CART توسعه یافته است (Ziegler & König, 2014). به عبارت دیگر، جنگل تصادفی یک روش یادگیری مبتنی بر دسته‌های درختی است. این روش قادر به یادگیری الگوهای پیچیده و در نظر گرفتن ارتباطات غیرخطی بین متغیرهای توضیحی و وابسته است (Louppe, 2014). فرآیند آموزش درخت با یک فرآیند تکرارپذیر شروع می‌شود که از گره ریشه شروع شده و در گره‌های نهایی (برگ‌ها) ختم می‌شود. سپس نمونه جدیدی انتخاب شده و برای آموزش یک درخت دیگر استفاده می‌شود. هنگامی که درخت کامل شد، یک مجموعه از قوانین تصمیم‌گیری برای تخمین داده جدید استخراج می‌شود (Genuer et al., 2020). روش جنگل تصادفی نسبت به روش‌های دیگر دارای مزایایی از قبیل دقت پیش‌بینی بالا، توانایی در یادگیری روابط غیرخطی، توانایی بالا در تعیین متغیرهای مهم در پیش‌بینی و ماهیت غیر پارامتری است (Matin et al., 2018).

الگوریتم درخت تصمیم

الگوریتم درخت تصمیم یکی از قوی‌ترین و پرکاربردترین الگوریتم‌های یادگیری ماشین است که در داده‌های دسته‌بندی و رگرسیون استفاده می‌شود (Charbuty & Abdulazeez, 2021). این الگوریتم از ساختار درختی برای نمایش تصمیمات و ترکیب آنها بر اساس ویژگی‌های ورودی بهره می‌برد و به‌سادگی قابل تفسیر است (Barros et al., 2011). ساختار درخت شامل گره‌های ریشه، داخلی و برگ‌ها است که به ترتیب داده‌ها را بر اساس ویژگی‌های مختلف تقسیم می‌کنند (Patel & Prajapati, 2018). هر گره درخت تصمیم بر اساس یک ویژگی انتخابی تصمیم می‌گیرد و فرآیند تصمیم‌گیری از ریشه شروع شده و با تقسیم مکرر داده‌ها به گره‌های فرعی ادامه می‌یابد (Zhu et al., 2018). روند آموزش درخت تصمیم شامل انتخاب بهینه ویژگی، تقسیم داده‌ها بر اساس این ویژگی، و تکرار این فرآیند تا ایجاد گره‌های برگ می‌شود (Rai et al., 2016). این الگوریتم با اینکه به‌سادگی قابل تفسیر است و در داده‌های بدون تباهی

$$SST = \frac{TP}{TP + FN} \quad (12)$$

$$SPC = \frac{TN}{TN + FP} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{pr} - x_{ac})^2} \quad (14)$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (15)$$

که در آن x_{pr} و x_{ac} 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 تحقیق، منحنی مشخصه عملکرد سیستم (ROC Curve) برای ارزیابی عملکرد مدل‌ها با توجه به نرخ‌های مثبت واقعی و مثبت کاذب بر روی محور Y و محور X استفاده شده است. این روش از معیار Area Under the Curve (AUC) برای ارزیابی کمی استفاده می‌کند که میزان حساسیت محور X را نسبت به محور Y نشان می‌دهد. منحنی ROC یکی از روش‌های کارآمد در پیش‌بینی سیستم‌ها است که دقت مدل را به صورت کمی برآورد می‌کند. در این روش، محدوده مقادیر AUC بین ۰/۵ تا ۱ قرار دارد و مقادیر نزدیک‌تر به ۱ نشان‌دهنده دقت بالاتر مدل هستند.

نتایج و بحث

آماده‌سازی معیارها

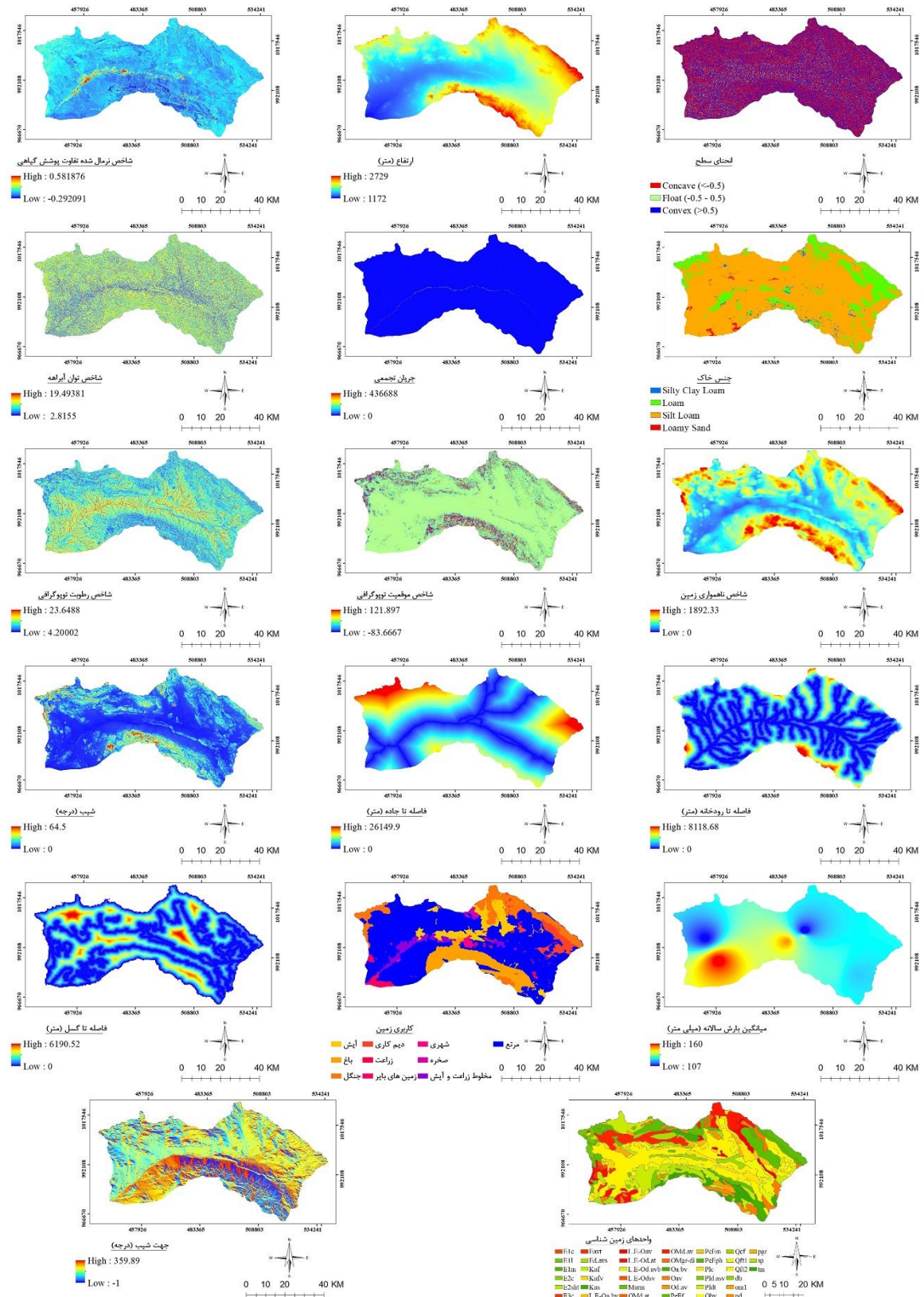
در این تحقیق، از ۱۷ معیار مختلف از حوزه‌های هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی برای پیش‌بینی پتانسیل آب زیرزمینی استفاده شده است. برای تولید نقشه‌های تأثیر این عوامل بر پتانسیل آب زیرزمینی که در شکل ۴ نشان‌دهنده، از روش درون‌یابی کریجینگ استفاده شده است.

در مقابل، NPV به احتمال داشتن پتانسیل آب زیرزمینی کمتر از ۵ لیتر بر ثانیه در یک پیکسل اشاره دارد، که نتایج آزمایش منفی هستند. SST نشان‌دهنده نسبت چاه‌های آب زیرزمینی با پتانسیل بیش از ۵ لیتر بر ثانیه است که به درستی طبقه‌بندی شده‌اند، در حالی که SPC نشان‌دهنده نسبت دسته‌بندی صحیح چاه‌های بدون پتانسیل آب زیرزمینی است. شاخص ACC نسبت چاه‌های آب زیرزمینی با پتانسیل ۵ لیتر بر ثانیه و چاه‌های بدون پتانسیل (پیکسل‌های آب زیرزمینی کمتر از ۵ لیتر بر ثانیه) را نشان می‌دهد.

RSME یا Root Mean Square Error میزان خطای میانگین بین مقادیر پیش‌بینی شده و واقعی چاه‌های آب زیرزمینی را نشان می‌دهد و مقادیر بالاتر این معیار نشان‌دهنده عملکرد بهتر مدل‌ها است. این معیار بر اساس چهار نوع پیامد احتمالی محاسبه می‌شود: مثبت واقعی (True Positive - TP)، منفی واقعی (True Negative - TN)، مثبت کاذب (False Positive - FP)، و منفی کاذب (False Negative - FN). TP نشان می‌دهد که به درستی به عنوان چاه‌هایی با پتانسیل بالا طبقه‌بندی شده‌اند، در حالی که FP تعداد چاه‌هایی را نشان می‌دهد که به نادرستی به عنوان چاه‌های با پتانسیل بالا طبقه‌بندی شده‌اند. علاوه بر این، TN و FN تعداد چاه‌هایی را نشان می‌دهند که به طور صحیح و نادرست به عنوان چاه‌های بدون پتانسیل طبقه‌بندی شده‌اند. R^2 Squared (یا ضریب تعیین) یک معیار آماری در یک مدل رگرسیونی است که نسبت واریانس در متغیر وابسته را تعیین می‌کند که می‌تواند توسط متغیر مستقل توضیح داده شود. به عبارت دیگر، ضریب تعیین نشان می‌دهد که چقدر داده‌ها با مدل رگرسیونی (خوبی برازش) مطابقت دارند.

$$NPV = \frac{TN}{TN + FN} \quad (10)$$

$$PPV = \frac{TP}{TP + FP} \quad (11)$$



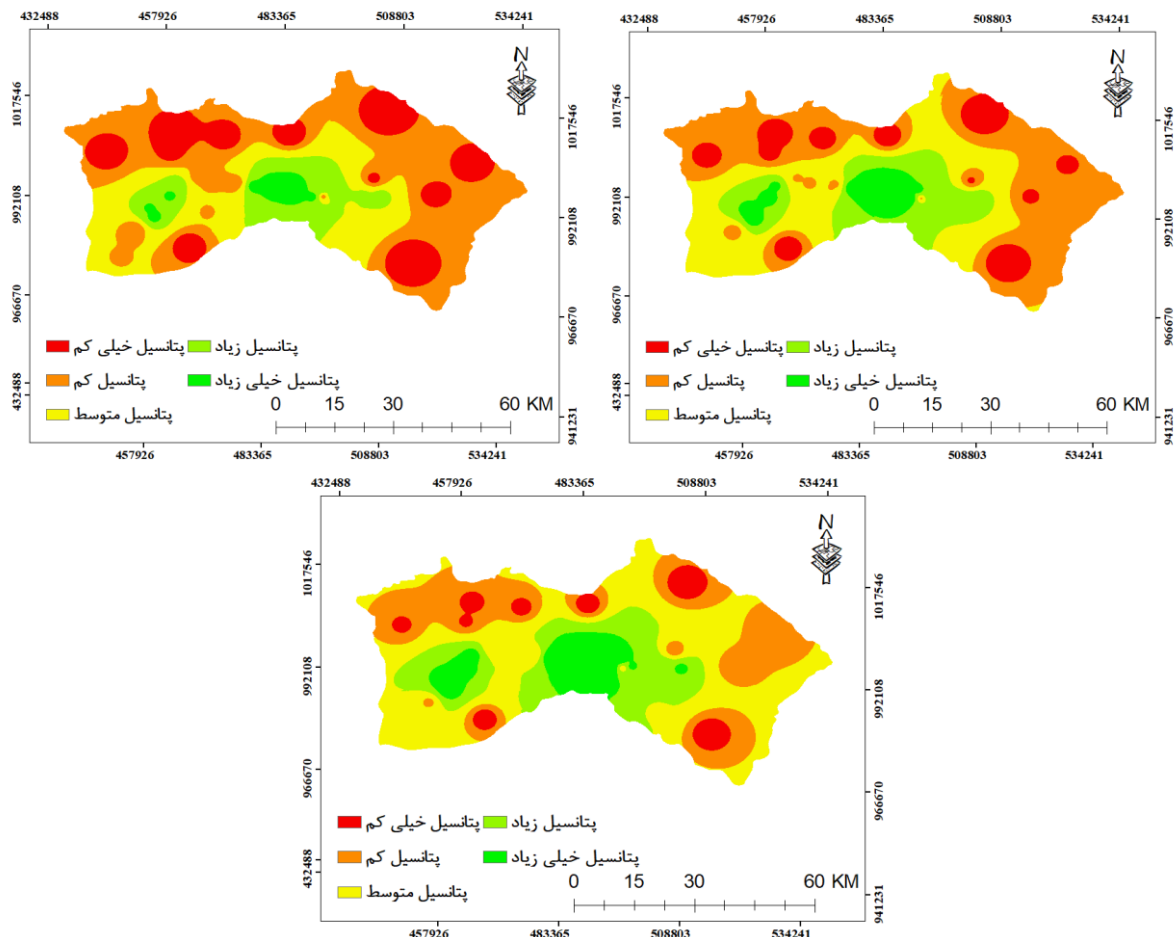
شکل ۴. معیارهای هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی مؤثر بر پتانسیل آب زیرزمینی.

Fig 4. Hydrological, topographical, geological and environmental criteria affecting the potential of underground water.

شکست طبیعی در پنج کلاس شامل پتانسیل خیلی کم، پتانسیل کم، پتانسیل متوسط، پتانسیل زیاد و پتانسیل خیلی زیاد طبقه‌بندی گردیده است. شکل ۵ نقشه پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه را نشان می‌دهد که با استفاده از مدل‌های مورد نظر پیش‌بینی شده‌اند.

تهیه نقشه پیش‌بینی پتانسیل آب زیرزمینی

در این تحقیق، نقشه‌های پیش‌بینی مناطق بالقوه آب زیرزمینی در منطقه مورد مطالعه با استفاده از مدل‌های DT، RF، SVM و با استفاده از مجموعه داده‌های آموزشی و اعتبارسنجی پیش‌بینی شده‌اند. نقشه میزان پتانسیل پیش‌بینی شده با استفاده از روش طبقه‌بندی



شکل ۵. نقشه پیش‌بینی پتانسیل آب‌های زیرزمینی منطقه مورد مطالعه بر اساس مدل‌های DT، RF و SVM (به ترتیب از نقشه سمت راست بالا).

Fig 5. The prediction map of the underground water potential of the study area is based on the DT, RF & SVM models.

مقایسه قابلیت پیش‌بینی مدل‌ها

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

همچنین هر یک از کلاس‌های پتانسیل آب زیرزمینی پیش‌بینی شده توسط مدل‌های DT، RF و SVM نشان داده شده است. نتایج نشان می‌دهد که در مدل DT، سه کلاس با پتانسیل آب زیرزمینی خیلی زیاد، زیاد و متوسط، درصد بیشتری را در منطقه مورد مطالعه نسبت به کلاس‌های مشابه در مدل‌های SVM و RF تحت پوشش قرار می‌دهد.

جدول ۲. مقایسه عملکرد مدل‌های DT، SVM و RF.

Table 2. Comparison of performance of hybrid models using training and validation data sets.

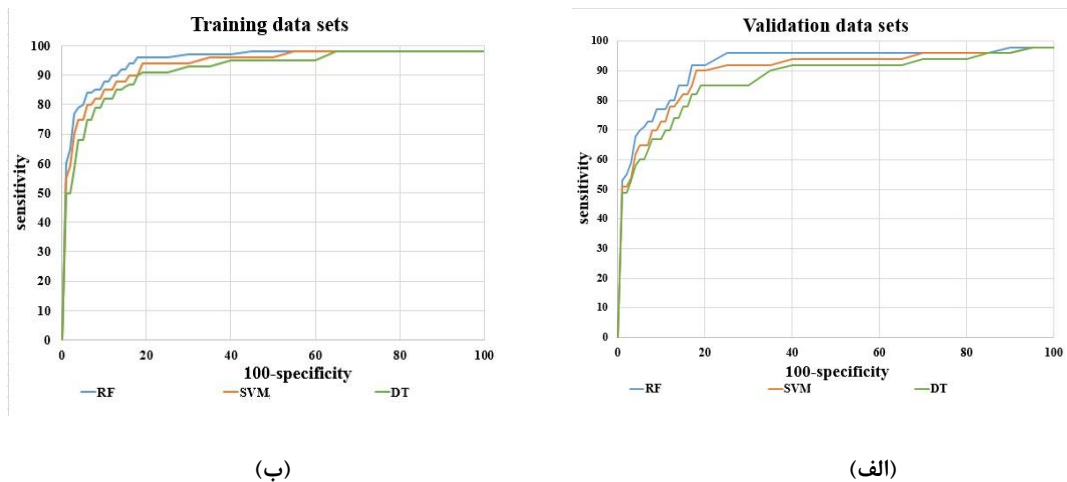
| مدل‌ها | R ² | RMSE |
|--------|----------------|-------|
| SVM | ۰٫۸۱ | ۱٫۲۰۵ |
| DT | ۰٫۸۹ | ۰٫۷۰۵ |
| RF | ۰٫۸۶ | ۱٫۰۵۵ |

DT، RF و SVM نشان می‌دهد. در کل، این منحنی نشان‌دهنده قابلیت مدل‌ها در پیش‌بینی ظرفیت آب زیرزمینی با استفاده از داده‌های آموزشی است و داده‌های اعتبارسنجی نشان‌دهنده دقت پیش‌بینی مدل در این مجموعه است. بر اساس شکل ۸، مدل DT در مجموعه داده‌های آموزشی با بالاترین مقدار AUC برتر است و سپس مدل‌های RF و SVM در ترتیب قرار دارند. شایان ذکر است همه مدل‌ها قابلیت پیش‌بینی خوبی دارند، اما مدل DT برای پیش‌بینی ظرفیت آب زیرزمینی در منطقه مورد بررسی بهترین عملکرد را دارد.

بر اساس جدول ۲ نیز به ترتیب روش DT، RF و SVM جهت پیش‌بینی نتایج قابل قبولی را ارائه می‌دهند. مقادیر R² و RMSE نیز برای مدل‌های DT، RF و SVM به ترتیب برابر است با ۰٫۸۹، ۰٫۸۶ و ۰٫۸۱ و ۰٫۷۰۵، ۱٫۰۵۵ و ۱٫۲۰۵ می‌باشد.

اعتبارسنجی مدل‌های پیش‌بینی پتانسیل آب زیرزمینی

شکل ۶ نمایش داده شده منحنی ROC را برای مجموعه‌های داده‌های آموزشی و اعتبارسنجی مدل‌های



شکل ۶. منحنی ROC و مقدار AUC برای روش‌های پیشنهادی تحقیق (الف) مجموعه داده‌های آموزشی (ب) مجموعه داده‌های اعتبارسنجی.

Fig 6. ROC curve and AUC value for the proposed research methods (a) training dataset and (b) validation dataset.

نتیجه‌گیری

زیرزمینی استفاده کرد که به مدیریت مناسب آب زیرزمینی یک منطقه کمک خواهد کرد. بنابراین، می‌توان نتیجه گرفت که مدل‌های ترکیبی توسعه‌یافته براساس مدل‌های مکانی و روش‌های یادگیری ماشین قابلیت خوبی در ارزیابی پتانسیل آب زیرزمینی صحیح یک منطقه با کمترین داده‌ها دارند.

منابع

- Abbaspour, K. C., Faramarzi, M., Ghasemi, S. S., & Yang, H. (2009). Assessing the impact of climate change on water resources in Iran. *Water resources research*, 45(10).
- Barros, R. C., Basgalupp, M. P., De Carvalho, A. C., & Freitas, A. A. (2011). A survey of evolutionary algorithms for decision-tree induction. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(3), 291-312.
- Bhavsar, H., & Panchal, M. H. (2012). A review on support vector machine for data classification. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(10), 185-189.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- Deepika, B., Avinash, K., & Jayappa, K. S. (2013). Integration of hydrological factors and demarcation of groundwater prospect zones: insights from remote sensing and GIS techniques. *Environmental earth sciences*, 70, 1319-1338.
- Dong, H., Yang, L., & Wang, X. (2021). Robust semi-supervised support vector machines with Laplace kernel-induced correntropy loss functions. *Applied Intelligence*, 51, 819-833.
- 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.
- Eslaminezhad, S. A., Eftekhari, M., Azma, A., Kiyangfar, R., & Akbari, M. (2022). Assessment of flood susceptibility prediction based on optimized tree-based machine learning models. *Journal of Water and Climate Change*, 13(6), 2353-2385.
- Ferreira, C. S. S., Walsh, R. P. D., Steenhuis, T. S., Shakesby, R. A., Nunes, J. P. N., Coelho, C. O. A., & Ferreira, A. J. D. (2015). Spatiotemporal variability of hydrologic soil properties and the implications for overland flow and land management in a peri-urban Mediterranean catchment. *Journal of Hydrology*, 525, 249-263.
- Foster, S., Chilton, J., Nijsten, G. J., & Richts, A. (2013). Groundwater—a global focus on the ‘local

ارزیابی مناسب پتانسیل آب زیرزمینی یک منطقه برای استفاده بهینه آن با استفاده از فناوری و روش‌های جدید مطلوب است. هدف اصلی این تحقیق ارائه مدل‌های یادگیری ماشین جمعی مبتنی بر الگوریتم زیرفضای تصادفی برای پیش‌بینی پتانسیل آب زیرزمینی است. بنابراین، در این تحقیق، سه مدل ترکیبی یادگیری ماشین به نام‌های DT، RF و SVM برای پیش‌بینی پتانسیل آب زیرزمینی در دشت بیرجند توسعه یافته‌اند. محققان قابلیت و کاربرد روش‌های یادگیری ماشین را در پیش‌بینی مکانی پتانسیل آب‌های زیرزمینی مناطق مختلف اثبات کرده‌اند. در این تحقیق، از ۱۷ معیار هیدرولوژی، توپوگرافی، زمین‌شناسی و محیطی برای پیش‌بینی پتانسیل آب زیرزمینی در دشت بیرجند استفاده شده است.

روش انتخاب ویژگی کمترین مربعات ماشین بردار پشتیبان (LSSVM) برای تعیین معیارهای مؤثر در افزایش عملکرد الگوریتم‌های یادگیری ماشین استفاده شده است. نتایج این تحقیق نشان داده‌اند که همه ۱۷ معیار مورد استفاده در این مطالعه دارای شایستگی بیشتر از صفر هستند، بنابراین تمام این معیارها برای ساخت مدل‌های پتانسیل آب زیرزمینی استفاده شده‌اند. انتخاب مؤثر این ۱۷ معیار باعث شد که دقت پیش‌بینی پتانسیل آب زیرزمینی را با کاهش نویز و تناسب بیش از حد برای مجموعه داده‌های آموزشی افزایش دهد. این نتیجه نشان می‌دهد که معیارهای انتخاب شده برای پیش‌بینی پتانسیل آب زیرزمینی در منطقه مورد بررسی بسیار مناسب هستند و ممکن است در مناطق مشابه نیز مفید باشند. عملکرد مدل‌های توسعه‌یافته با استفاده از معیارهای آماری استاندارد (SPC، SST، NPV، PPV، ROC و RMSE) ارزیابی شده است. نتایج تجزیه و تحلیل نشان داده‌اند که همه مدل‌های ترکیبی جدید توسعه‌یافته قابلیت پیش‌بینی خوبی دارند، اما مدل DT در پیش‌بینی دقیق مناطق بالقوه آب زیرزمینی بهترین عملکرد را نشان داده است و پس از آن مدل‌های RF و SVM برای نگاشت دقیق مناطق بالقوه آب زیرزمینی مورد استفاده قرار دارند. در مناطق دیگر نیز می‌توان از این مدل برای نقشه‌برداری دقیق و شناسایی صحیح مناطق پتانسیل آب

- Xiong, L., Tang, G., Yan, S., Zhu, S., & Sun, Y. (2014). Landform-oriented flow-routing algorithm for the dual-structure loess terrain based on digital elevation models. *Hydrological Processes*, 28(4), 1756-1766.
- Yaman, A., & Cengiz, M. A. (2021). The Effects of Kernel Functions and Optimal Hyperparameter Selection on Support Vector Machines. *Journal of New Theory*, (34), 64-71.
- Yenehun, A., Nigate, F., Belay, A. S., Desta, M. T., Van Camp, M., & Walraevens, K. (2020). Groundwater recharge and water table response to changing conditions for aquifers at different physiography: The case of a semi-humid river catchment, northwestern highlands of Ethiopia. *Science of The Total Environment*, 748, 142243.
- Zhu, F., Tang, M., Xie, L., & Zhu, H. (2018). A classification algorithm of CART decision tree based on MapReduce attribute weights. *International Journal of Performability Engineering*, 14(1), 17.
- Ziegler, A., & König, I. R. (2014). Mining data with random forests: current options for real-world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(1), 55-63.
- resource'. *Current opinion in environmental sustainability*, 5(6), 685-695.
- Genuer, R., Poggi, J. M., Genuer, R., & Poggi, J. M. (2020). *Random forests*. Springer International Publishing.33-55.
- Guido, J. J., Winters, P. C., & Rains, A. B. (2006). *Logistic regression basics*. MSc University of Rochester Medical Center, Rochester, NY, 21.
- Hilario, M., Kalousis, A., Pellegrini, C., & Müller, M. (2006). Processing and classification of protein mass spectra. *Mass spectrometry reviews*, 25(3), 409-449.
- Hussein, A. A., Govindu, V., & Nigusse, A. G. M. (2017). Evaluation of groundwater potential using geospatial techniques. *Applied Water Science*, 7, 2447-2461.
- Li, H., Zhao, X., Gao, X., Ren, K., & Wu, P. (2018). Effects of water collection and mulching combinations on water infiltration and consumption in a semiarid rainfed orchard. *Journal of Hydrology*, 558, 432-441.
- Louppe, G. (2014). *Understanding random forests: From theory to practice*. arXiv preprint arXiv:1407.7502.
- Matin, S. S., Farahzadi, L., Makaremi, S., Chelgani, S. C., & Sattari, G. H. (2018). Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest. *Applied Soft Computing*, 70, 980-987.
- Miralles, P., Church, T. L., & Harris, A. T. (2012). Toxicity, uptake, and translocation of engineered nanomaterials in vascular plants. *Environmental science & technology*, 46(17), 9224-9239.
- Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree-based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74-78.
- Rai, K., Devi, M. S., & Guleria, A. (2016). Decision tree-based algorithm for intrusion detection. *International Journal of Advanced Networking and Applications*, 7(4), 2828.
- Sanaeinejad et al. (2014), Wheat yield estimation using Landsat images and field observation: A case study in Mashhad. *J. of Plant Production*, 20 (4).
- Sansone, M., Fusco, R., Pepino, A., & Sansone, C. (2013). Electrocardiogram pattern recognition and analysis based on artificial neural networks and support vector machines: a review. *Journal of Healthcare Engineering*, 4, 465-504.
- Tashayo, B., Honarbakhsh, A., Akbari, M., & Eftekhari, M. (2020). Land suitability assessment for maize farming using a GIS-AHP method for a semi-arid region, Iran. *Journal of the Saudi Society of Agricultural Sciences*, 19(5), 332-338.
- Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., & Kmoch, A. (2020). Vertical accuracy of freely available global digital elevation models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote Sensing*, 12(21), 3482.