

Introducing a Hybrid Model Based on Certainty Factor Statistical and Bagging Methods for Discovering Groundwater

SEYED VAHID RAZAVI TERMEH¹, MAJID RAHIMZADEGAN^{2*}

1. Department of Geodesy & Geomatics, Khajeh Nasir Toosi University of Technology, Tehran, Iran.

2. Department of Water Resources, Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran.

(Received: May. 29, 2019- Revised: July. 20, 2019- Accepted: July. 27, 2019)

ABSTRACT

Due to climate change and growth of urban communities, the need for groundwater and exploration of these resources are increasing. Therefore, the purpose of this study was to provide a groundwater potential mapping using the geographic information system (GIS) in a region located in Booshehr plain using an ensemble of certainty factor (CF) method with Bagging data mining method. For this purpose, in the first step, 339 well locations were identified in the study area, of which 238 wells (70%) were randomly selected as training points and 101 wells (30%) were selected as validation points. In the next step, 15 factors affecting groundwater such as altitude, slope angle, slope direction, slope length, plan curvature, profile curvature, topographic wetness index, distance from fault, fault density, distance from river, drainage density, rainfall, lithology, Soil and land cover were prepared in ArcGIS 10.3 and Saga GIS software. The spatial relationship between the effective parameters and the location of the wells was determined using a CF model. These weights were used to implement the Bagging model. In order to validate the accuracy of the ensemble model, the RMSE and MAE indices were used. Also, in order to validate the accuracy of the maps, ROC and AUC were used. The results of this study showed that the values of RMSE and MAE indices for training and validation are equal to 0.247, 0.162, 0.256 and 0.169 respectively. The evaluation results of the ROC curve indicated that the AUC was 86.2 and 94.8%, respectively, for CF models and the ensemble of CF model with the Bagging model.

Keyword: Groundwater level potential, certainty factor (CF) method, bagging method, Geographic information system (GIS).

ارائه مدلی مبنی بر ترکیب روش آماری عامل اطمینان و روش بگینگ به منظور اکتشاف آب زیرزمینی

سید وحید رضوی ترمه^۱، مجید رحیم زادگان^{۲*}

۱. دانشکده مهندسی نقشه برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران.

۲. گروه مدیریت منابع آب، دانشکده مهندسی عمران، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران.

(تاریخ دریافت: ۱۳۹۸/۳/۸ - تاریخ بازنگری: ۱۳۹۸/۴/۲۹ - تاریخ تصویب: ۱۳۹۸/۵/۵)

چکیده

با توجه به تغییرات اقلیمی و رشد جوامع شهری، نیاز به آب زیرزمینی و اکتشاف این منابع رو به افزایش است؛ بنابراین هدف از تحقیق حاضر، تهیه نقشه پتانسیل سطح آب زیرزمینی با استفاده از سیستم اطلاعات مکانی (GIS) در منطقه‌ای واقع در دشت بوشهر با استفاده از ترکیب روش آماری عامل اطمینان با روش داده‌کاوی بگینگ است. بدین منظور در گام اول، ۳۳۹ موقعیت چاه در منطقه مورد مطالعه مشخص گردید و به صورت تصادفی، ۲۳۸ چاه (۷۰ درصد) به عنوان نقاط آموزشی و ۱۰۱ چاه (۳۰ درصد) به عنوان نقاط اعتبارسنجی تعیین گردید. در گام بعد، ۱۵ عامل تأثیرگذار بر تجمع آب زیرزمینی مانند ارتفاع، زاویه شیب، جهت شیب، طول شیب، انحنای سطح، انحنای آبراهه، شاخص رطوبت توپوگرافی، فاصله از گسل، تراکم گسل، فاصله از رودخانه، تراکم آبراهه، بارندگی، لیتولوژی، پوشش اراضی و نوع خاک در نرم‌افزار ArcGIS 10.3 و Saga GIS تهیه گردید. رابطه مکانی بین پارامترهای مؤثر و موقعیت چاه‌ها با استفاده از مدل عامل اطمینان مشخص گردید و به منظور پیاده‌سازی مدل بگینگ از این وزن‌ها استفاده شد. به منظور ارزیابی دقت مدلی ترکیبی از شاخص‌های ضریب تعیین، RMSE و MAE استفاده شد و همچنین به منظور ارزیابی دقت نقشه‌های تهیه‌شده از منحنی تشخیص عملکرد نسبی (ROC) و سطح زیر آن (AUC) استفاده گردید. نتایج حاصل از ارزیابی نشان می‌دهد که مقادیر شاخص‌های ضریب تعیین، RMSE و MAE برای داده‌های آموزشی و اعتبارسنجی به ترتیب برابر ۷۶ درصد، ۰/۲۴۷، ۰/۱۶۲، ۷۲/۵، ۰/۲۵۶ و ۰/۱۶۹ است. نتایج ارزیابی منحنی ROC نشان می‌دهد که سطح زیر منحنی به ترتیب ۸۶/۲ و ۹۴/۸ درصد برای مدل‌های عامل اطمینان و ترکیب مدل عامل اطمینان با مدل داده‌کاوی بگینگ است.

واژه‌های کلیدی: پتانسیل سطح آب زیرزمینی، مدل عامل اطمینان، مدل بگینگ، سیستم اطلاعات مکانی (GIS)

مقدمه

با شرایط اقلیمی خشک و نیمه‌خشک است. در این شرایط ۶۵ درصد مصرف آب در کشور از منابع آب‌های زیرزمینی تهیه می‌گردد (Rahmati et al, 2015). اخیراً بسیاری از مناطق ایران به علت تغییرات آب و هوایی و برداشت بی‌رویه منابع آب زیرزمینی خشک شده‌اند که منجر به کمبود جدی آب در بسیاری از استان‌های کشور شده است (Osati et al, 2014). اولین عوامل تأثیرگذار بر آب زیرزمینی بهره‌برداری بیش‌ازحد یا مفرط آب زیرزمینی و منفی شدن بیلان است و همچنین عوامل ثانویه در ارتباط با توپوگرافی، لیتولوژی، سازندهای زمین‌شناسی، تراکم شکستگی‌ها، تخلخل ثانویه، توزیع سفره آب زیرزمینی، شیب، الگوی زه‌کشی، وضعیت آب و هوایی و روابط متقابل آن‌ها آب زیرزمینی را تحت تأثیر قرار می‌دهد (Oh et al, 2011). به دلیل تعدد عوامل مؤثر بر تجمع آب زیرزمینی، ارزیابی آب زیرزمینی و شناسایی مناطق بالقوه آن امری پیچیده است و هنوز چالشی بزرگ برای مدیران منابع آب است. روش‌های سنتی مورد استفاده برای شناسایی، ترسیم و نقشه مناطق پتانسیل سطح آب

آب‌های زیرزمینی یکی از مهم‌ترین و حیاتی‌ترین منابع طبیعی است که در سازندهای زمین‌شناسی زیرسطحی در منطقه بحرانی زمین ذخیره می‌شود (Fitts, 2002). آب زیرزمینی به‌عنوان یک منبع آب برای مصارف خانگی، صنعتی، کشاورزی و سایر کاربردها استفاده می‌شود (Ayazi et al, 2010; Nampak et al, 2014). تقاضای روزافزون آب برای برآورده کردن نیازها و توسعه‌های انسانی، فشار زیادی بر این منابع آب شیرین تحمیل کرده است (Greenbaum, 1992; Lee et al, 2012; Khosravi et al, 2018). منابع آب شیرین شامل آب‌های سطحی، آب‌های زیرزمینی و یخچال‌های طبیعی است که در ایران بیشترین استفاده از منابع آب شیرین مربوط به آب‌های زیرزمینی است (Rahmati et al, 2015). مشکلات مربوط به آب‌های زیرزمینی در بیشتر بخش‌های مناطق گرمسیری و نیمه گرمسیری که تراکم جمعیت و توسعه اقتصادی بالایی دارند، شدید است. کشور ایران با میانگین بارش ۲۵۰ میلی‌متر بارش سالانه یکی از کم‌آب‌ترین کشورهای جهان

بگینگ در منطقه‌ای واقع در دشت بوشهر است.

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه در استان بوشهر در جنوب شرقی ایران واقع شده است. مساحت منطقه مورد مطالعه ۲۶۹۶ کیلومتر مربع است و قسمت زیادی از منطقه مورد مطالعه، دشت با ارتفاع کم است. ارتفاع منطقه مورد مطالعه بین ۳ تا ۱۴۹۰ متر بالای سطح دریا است. میانگین دمای سالانه در این منطقه، ۲۴ درجه سانتی‌گراد است و بالاترین دمای آن مربوط به فصل تابستان با دمای ۵۰ درجه سانتی‌گراد است. میزان بارندگی در این منطقه پایین است به طوری که میانگین سالانه بارندگی کمتر از ۲۵۵ میلی‌متر است و بیشتر باران‌ها در ماه آذر و فروردین اتفاق می‌افتد. بارندگی در بهار و پاییز کوتاه ولی با شدت زیاد است، در حالی که در فصل زمستان به طور نامنظم و پراکنده است. میانگین رطوبت سالانه در این منطقه بالای ۷۱ درصد است. منطقه مورد مطالعه در شکل (۱) نشان داده شده است.

نقشه پراکنش چاه

اطلاعات مربوط به چاه‌های آب زیرزمینی از اداره مدیریت منابع آب استان بوشهر تهیه گردید. بر اساس پژوهش‌های پیشین و گزارش‌های اداره آب تنها داده‌های چاه‌های آب زیرزمینی با پتانسیل آبدهی بالا (بیش از ۱۱ مترمکعب در ساعت) مورد استفاده قرار گرفت (Nampak *et al*, 2014). علاوه بر پتانسیل آبدهی بالای ۱۱ مترمکعب در ساعت، از معیارهای میانگین PH ۶/۹ و میانگین هدایت الکتریکی ۴۹۵ میکرو موس بر سانتی‌متر استفاده شد. در این تحقیق، این چاه‌ها به صورت تصادفی به دو گروه آموزش و اعتبارسنجی تقسیم شد. هفتاد درصد چاه‌ها یعنی ۲۳۸ چاه برای گروه آموزش و سی درصد باقیمانده یعنی ۱۰۱ چاه برای گروه اعتبارسنجی در نظر گرفته شد. موقعیت چاه‌های گروه آموزش و گروه اعتبارسنجی در شکل ۱ ارائه شده است.

معیارهای تأثیرگذار بر آب زیرزمینی

انتخاب معیارهای تأثیرگذار بر آب زیرزمینی با توجه به تحقیقات گذشته در داخل و خارج از کشور (Rahmati *et al*, 2015; Naghibi *et al*, 2018; Khosravi *et al*, 2018) و لایه‌های اطلاعاتی مختلف مانند ارتفاع، زاویه شیب، جهت شیب، طول شیب، شاخص رطوبت توپوگرافی (TWI)، انحنا سطح، انحنا آبراهه، کاربری اراضی، لیتولوژی، نوع خاک، فاصله از رودخانه، فاصله از گسل، تراکم آبراهه، تراکم گسل و بارندگی

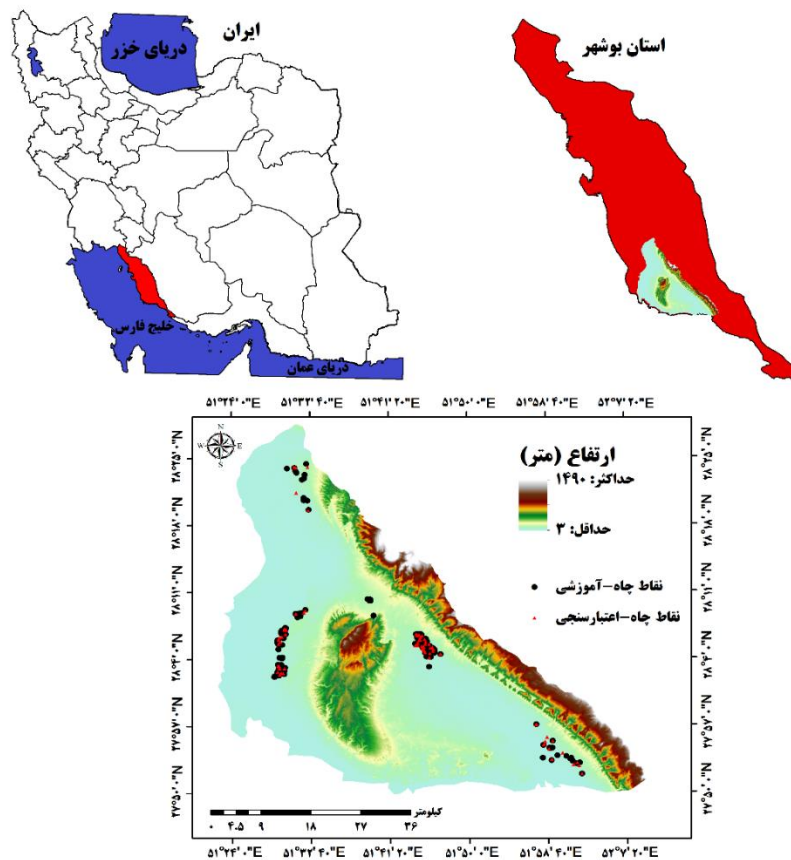
زیرزمینی عمدتاً بر پایه بررسی‌های زمینی با استفاده از ابزارهای زمین‌شناسی و هیدرولوژیکی هستند که عموماً گران و زمان‌بر هستند (Israil *et al*, 2006; Jha *et al*, 2010; Mukherjee *et al*, 2018). سیستم اطلاعات جغرافیایی و سنجش از دور در زمینه تهیه نقشه پتانسیل آب‌های زیرزمینی، در مقایسه با روش‌های سنتی که بسیار وقت‌گیر، کم‌دقت و پرهزینه هستند، دارای کارایی بالایی بوده و باعث افزایش دقت و سرعت پژوهش‌ها در زمینه ارزیابی آب‌های زیرزمینی شده‌اند (Jha *et al*, 2007). نقشه پتانسیل سطح آب زیرزمینی با استفاده از فناوری سیستم اطلاعات جغرافیایی و سنجش از دور، تاکنون با روش‌های مختلفی مانند نسبت فراوانی (Mair & Elkadi, 2015)، رگرسیون لجستیک (Naghibi *et al*, 2015)، وزن شواهد (Tahmassebpour *et al*, 2016)، آنتروپی (Razandi *et al*, 2015) و عامل اطمینان (Naghibi *et al*, 2015) تهیه شده است. با توسعه سریع فن‌آوری اطلاعات و پایگاه داده، الگوریتم‌های داده‌کاوی مانند شبکه‌های عصبی مصنوعی (Lee *et al*, 2018)، جنگل‌های تصادفی (Golkarian *et al*, 2018)، ماشین‌های بردار پشتیبان (Naghibi *et al*, 2017) و درخت‌های تصمیم‌گیری (Lee and Lee, 2015) برای تهیه نقشه پتانسیل سطح آب‌های زیرزمینی استفاده شده است. در این راستا Chen *et al* (2018a)، به بررسی ترکیب روش‌های وزن شواهد با روش داده‌کاوی درخت تابعی پرداختند که نتایج آن‌ها نشان‌دهنده عملکرد بسیار خوب روش‌های ترکیب داده‌کاوی در تهیه نقشه پتانسیل سطح آب زیرزمینی است. Lee *et al* (2019) به مقایسه روش آماری نسبت فراوانی و روش داده‌کاوی BCT به منظور تهیه نقشه پتانسیل سطح آب زیرزمینی پرداختند که بر طبق نتایج آن‌ها روش‌های داده‌کاوی از دقت بالاتری نسبت به روش‌های آماری برخوردارند. Pham *et al* (2019) از روش‌های ترکیبی داده‌کاوی برای تهیه نقشه پتانسیل آب زیرزمینی در بخشی از هند استفاده کردند. Kordestani *et al* (2018) از روش داده‌کاوی BRT به منظور تهیه نقشه پتانسیل سطح آب زیرزمینی استفاده کردند که نتایج آن‌ها نشان‌دهنده دقت بالاتر مدل‌های ترکیبی نسبت به مدل‌ها به تنهایی است.

یکی از مسائل مهم در الگوریتم‌های داده‌کاوی، پیش‌پردازش و آماده‌سازی داده‌ها و انتخاب بهترین داده ورودی به الگوریتم‌ها جهت بالاتر بردن دقت است (Zhang *et al*, 2003). تاکنون تحقیقات اندکی از ترکیب روش‌های آماری و الگوریتم‌های داده‌کاوی به منظور تهیه نقشه پتانسیل سطح آب زیرزمینی استفاده کرده‌اند (Kordestani *et al*, 2018). از این رو، هدف و نوآوری این تحقیق تهیه نقشه پتانسیل سطح آب زیرزمینی با استفاده از ترکیب مدل آماری عامل اطمینان با مدل داده‌کاوی

انتخاب و در محیط GIS آماده گردید.

در ابتدا مدل رقومی ارتفاعی منطقه مورد مطالعه با استفاده از تصاویر ماهواره‌ای ASTER تهیه شد و معیارها ارتفاع، زاویه شیب، جهت شیب، انحنای سطح، انحنای توپوگرافی، طول شیب و شاخص رطوبت توپوگرافی در نرم‌افزار ArcGIS 10.3 و Saga از آن استخراج گردید. طبقات ارتفاعی مختلف باعث ایجاد شرایط اقلیمی متفاوت شده و در نتیجه باعث ایجاد خاک و نوع

پوشش گیاهی متفاوت می‌گردد (Jothibasu and Anbazhagan, 2016). شیب، اطلاعات لازم را در مورد فرآیندهای زمین‌شناسی و ژئودینامیکی فعال در مقیاس منطقه‌ای ارائه می‌دهد و رواناب سطحی و سرعت نفوذ تحت تأثیر شیب قرار می‌گیرند (Mogaji et al, 2015; Razavi Termeh et al 2017). جهت شیب وابسته به جهت عمده است و بر روی میزان دریافت بارش و نوع پوشش گیاهی تأثیرگذار است (Ercanoglu and Gokceoglu, 2002).



شکل ۱- منطقه مورد مطالعه و موقعیت چاه‌ها

(Moore et al, 1991). این شاخص مطابق رابطه (۲) محاسبه می‌گردد.

$$TWI = \ln \left(\frac{A_s}{\tan \beta} \right) \quad (\text{رابطه ۲})$$

که در آن: A_s و β به ترتیب نشان‌دهنده‌ی مساحت یک حوضه مشخص (m^2) و زاویه شیب (برحسب درجه) است. بارندگی منبع اصلی آب در چرخه هیدرولوژی و مهم‌ترین عامل تأثیرگذار در آب‌های زیرزمینی منطقه است. به‌منظور تهیه نقشه بارندگی از آمار بارندگی ایستگاه‌های منطقه مطالعاتی در یک دوره آماری ۳۰ ساله و از روش درون‌یابی کریجینگ در نرم‌افزار ArcGIS 10.3 استفاده شد. فاصله از رودخانه‌ها با استفاده از نقشه توپوگرافی ایجاد شد، درحالی‌که فاصله از گسل‌ها با استفاده از نقشه زمین‌شناسی منطقه مورد مطالعه ایجاد شد. شبکه زه‌کشی

انحنای سطح بر روی همگرایی و واگرایی جریان تأثیر می‌گذارد و انحنای مقطع مطابق با جهت حداکثر شیب است و عمدتاً بر روی سرعت جریان در سطح تأثیرگذار است (Al-Abadi et al, 2016). طول شیب که ترکیبی از طول شیب (L) و سرازیری شیب (S) است، مطابق رابطه (۱) محاسبه می‌گردد (Moore and Burch, 1986).

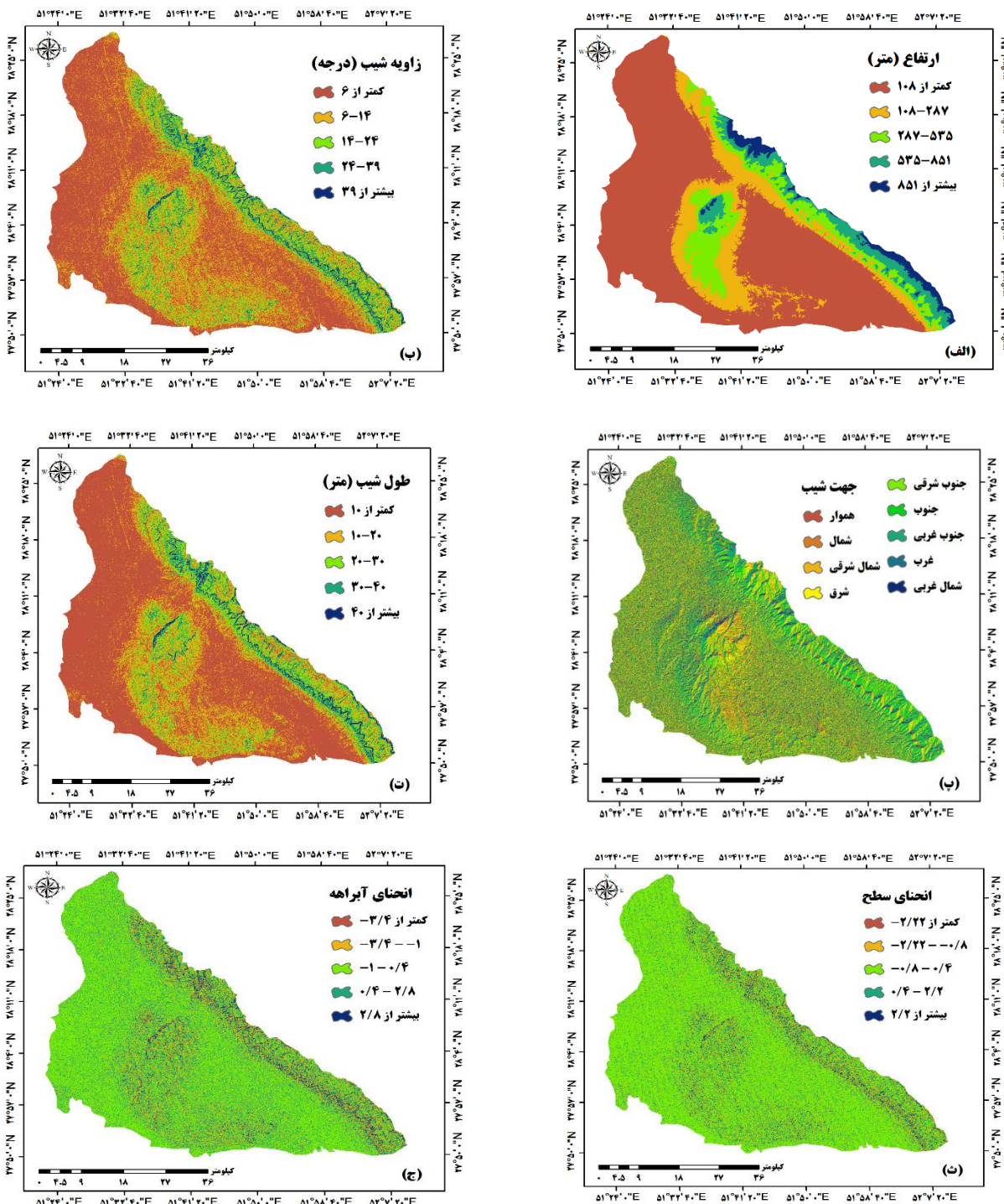
$$LS = \left(\frac{B_s}{22.13} \right)^{0.6} \left(\frac{\sin \alpha}{0.0896} \right)^{1.3} \quad (\text{رابطه ۱})$$

که در آن: α گرادیان شیب برحسب درجه و B_s مساحت حوضه آبخیز (m^2) است.

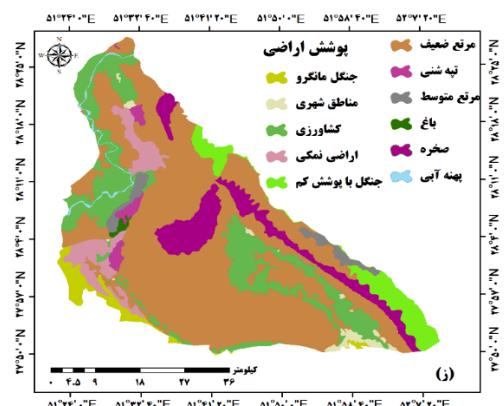
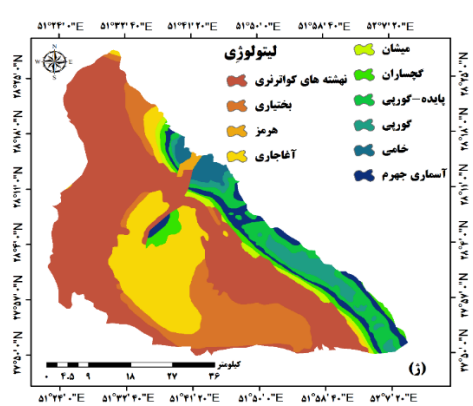
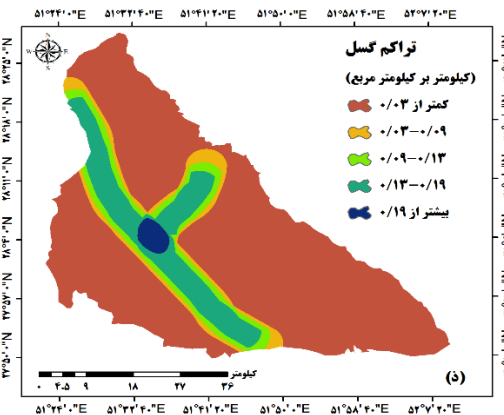
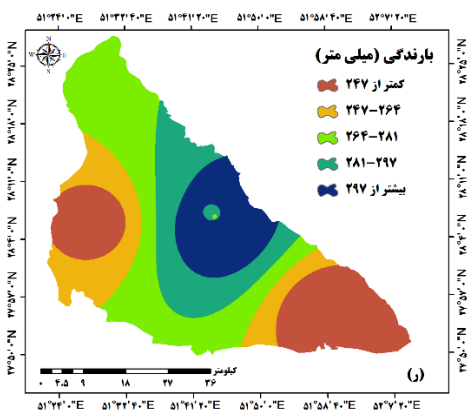
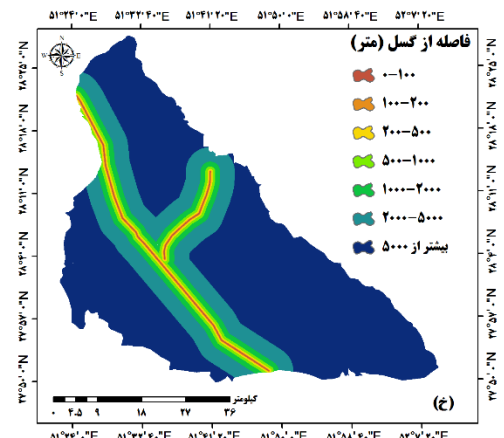
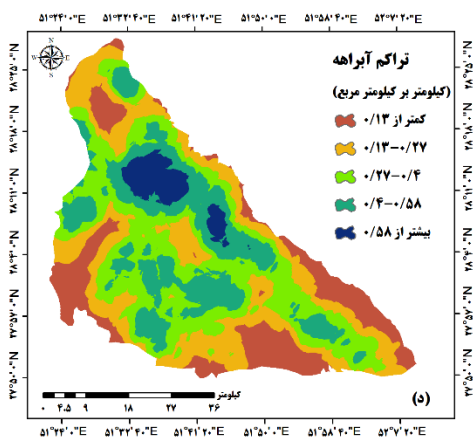
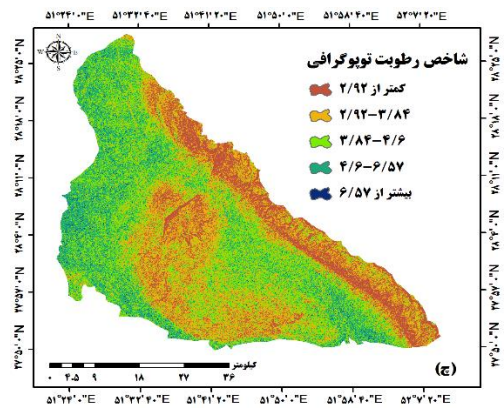
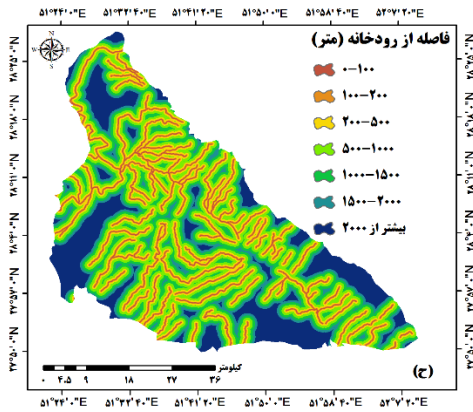
شاخص رطوبت توپوگرافی معمولاً برای محاسبه کنترل توپوگرافی در فرآیندهای هیدرولوژیکی استفاده می‌شود و نقش مهمی در رطوبت خاک، جریان آب زیرزمینی و ثبات شیب دارد

منابع طبیعی استان بوشهر در مقیاس ۱:۱۰۰۰۰۰ تهیه شد. انواع خاک نقش مهمی در مقدار آب ایفا می‌کنند که می‌تواند به سازنده‌ای زیرسطحی نفوذ کند و از این رو نفوذ آب زیرزمینی را تحت تأثیر قرار دهد. لیتولوژی تخریل خاک و نفوذپذیری آب را کنترل می‌کند که به‌نوبه خود بر ذخیره ویژه آب‌های زیرزمینی تأثیر می‌گذارد. نقشه لیتولوژی در مقیاس ۱:۱۰۰۰۰۰ از نقشه زمین‌شناسی ایران تهیه گردید. معیارهای تأثیرگذار بر آب زیرزمینی در شکل (۲) نشان داده شده است.

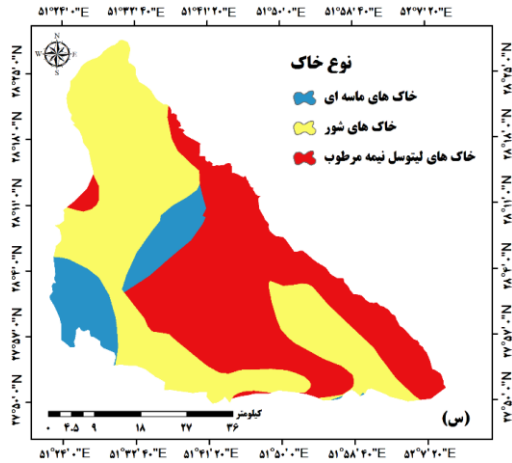
به لیتولوژی وابسته است و شاخص مهمی از سرعت نفوذ را فراهم می‌کند. با تقسیم‌بندی طول کل رودخانه‌ها در یک حوضه زه‌کشی به مساحت کل حوضه زه‌کشی، تراکم آبراهه به دست می‌آید. تراکم گسل، شکلی از خطواره‌هایی است که بر ذخیره و حرکت آب‌های زیرزمینی تأثیر می‌گذارد و با استفاده از نقشه‌های توپوگرافی تهیه گردید. پوشش اراضی اطلاعات لازم را در مورد نفوذ، رطوبت خاک، آب‌های زیرزمینی، آب‌های سطحی و غیره فراهم می‌کند. نقشه کاربری اراضی منطقه مورد مطالعه از اداره



شکل ۲- لایه‌های اطلاعاتی تأثیرگذار بر آب زیرزمینی (الف) ارتفاع (ب) زاویه شیب (پ) جهت شیب (ت) طول شیب (ث) انحنای سطح (ج) انحنای آبراهه



ادامه شکل ۲- لایه‌های اطلاعاتی تأثیرگذار بر آب زیرزمینی (چ) شاخص رطوبت توپوگرافی (ح) فاصله از رودخانه (خ) فاصله از گسل (د) تراکم آبراهه (ذ) تراکم گسل (ز) بارندگی (ز) پوشش اراضی (ز) لیتولوژی



ادامه شکل ۲- لایه‌های اطلاعاتی تأثیرگذار بر آب زیرزمینی (س) نوع خاک

مقادیر منفی نشان‌دهنده کاهش قطعیت است. مقادیر نزدیک به صفر نشان می‌دهد که احتمال اولیه، بسیار مشابه احتمال شرطی است و هیچ اطلاعاتی درباره قطعیت وقوع رخداد به دست نمی‌دهد.

مقادیر عامل اطمینان برای همه معیارهای تأثیرگذار با همپوشانی و محاسبه نسبت چاه آب زیرزمینی محاسبه می‌شود، سپس مقادیر عامل اطمینان تمام پارامترها در ۱۵ معیار تهیه آب زیرزمینی با استفاده از رابطه (۴) تعیین می‌شود. در نتیجه، مقادیر عامل اطمینان برای همه‌ی معیارهای تأثیرگذار، به صورت جفتی با یکدیگر تلفیق می‌شوند. ترکیب دو لایه X و Y به خاطر اطلاعات متفاوت دو لایه، به عنوان پارامتر Z در معادله تعیین می‌شود. (رابطه ۴)

$$z = \begin{cases} X + Y - XY, & X, Y \geq 0 \\ X + Y + XY, & X, Y \leq 0 \\ \frac{X + Y}{1 - \min(|X| + |Y|)}, & \text{otherwise} \end{cases}$$

ترکیب جفتی (با استفاده از قاعده ادغام (معادله ۴) مکرراً انجام می‌شود تا زمانی که تمام لایه‌های عامل اطمینان به محاسبه شاخص پتانسیل سطح آب زیرزمینی اضافه گردند.

مدل بگینگ

الگوریتم بگینگ در سال ۱۹۹۶، توسط برایمن به منظور بهبود دقت طبقه‌بندی و تعمیم الگوهای داده معرفی شد (Breiman, 1996). این الگوریتم، از گروهی طبقه‌بندی‌کننده‌های بر مبنای درخت تشکیل شده است. این روش یک متا الگوریتم است که برای بهبود دادن یادگیری ماشین رده‌بندی و مدل‌های پسرقتی بر اساس پایداری و دقت رده‌بندی است. این روش همچنین واریانس را کاهش داده و از افتادن در بیش برآزش جلوگیری می‌کند (Bui et al, 2016). اگرچه این روش معمولاً در درخت تصمیم به کار می‌رود، اما می‌تواند در هر نوع مدل استفاده شود.

روش‌های مورد استفاده

شرط هم خطی

بعد از تعریف عوامل مؤثر، ضریب شرطی باید برای مسئله هم خطی مورد ارزیابی قرار گیرد. هم خطی زمانی اتفاق می‌افتد که دو یا چند عامل غیرمستقل بسیار به هم وابسته باشند. به منظور تعیین اینکه پارامترهای تأثیرگذار بر آب زیرزمینی مستقل از یکدیگر هستند و می‌توانند در مدل‌سازی شرکت کنند از شاخص‌های ضریب تحمل (Tolerance) و عامل تورم واریانس (VIF) استفاده شد (Khosravi et al, 2018). در مواردی که مقدار شاخص عامل تورم واریانس کمتر از ۱۰ و شاخص ضریب تحمل بیشتر از ۰/۱ باشد، تمامی پارامترها می‌توانند در مدل‌سازی شرکت کنند.

مدل عامل اطمینان

مدل عامل اطمینان مدلی است که به طور گسترده توسط محققین مختلف در تهیه نقشه حساسیت خطر مورد استفاده قرار گرفته است (Kanungo et al, 2011; Devkota et al, 2013). مدل عامل اطمینان یکی از توابع مطلوبیت ممکن به منظور رسیدگی به مشکل ترکیب لایه‌های اطلاعاتی مختلف و عدم قطعیت و ناهمگونی داده‌های ورودی است (Devkota et al, 2013). نقشه پتانسیل سطح آب زیرزمینی با استفاده از مدل عامل اطمینان با استفاده از رابطه (۳) محاسبه می‌گردد.

$$\begin{cases} cf = \frac{pp_a - pp_s}{pp_a(1 - pp_s)} & \text{if } pp_a > pp_s \\ cf = \frac{pp_a - pp_s}{pp_s(1 - pp_a)} & \text{if } pp_a < pp_s \end{cases} \quad (\text{رابطه ۳})$$

که در آن، cf عامل اطمینان، pp_a احتمال شرطی وجود تعدادی از چاه‌ها در یک کلاس و pp_s احتمال اولیه یا پیشین وجود کل چاه‌ها در منطقه مورد مطالعه است. دامنه عامل اطمینان بین -۱ و +۱ است، مقادیر مثبت نشان‌دهنده افزایش قطعیت و

عمودی (محور Y) است که به صورت روابط ۷ و ۸ محاسبه می‌گردد (Razavi Termeh et al, 2017).

$$X = 1 - \frac{TN}{TN+FP} \quad (\text{رابطه ۷})$$

$$Y = \frac{TP}{TP+FN} \quad (\text{رابطه ۸})$$

نتایج

نتایج شرط هم خطی

نتایج به‌دست‌آمده از شرط هم خطی در جدول (۱) نشان داده شده است. بر اساس نتایج، تمام پارامترها مقادیر VIF کمتر از ۱۰ و تحمل بیشتر از ۰/۱ دارند و می‌توانند در مدل‌سازی شرکت کنند.

جدول ۱- نتایج شرط هم خطی

ضریب تحمل	عامل تورم واریانس	معیار
۱/۴۲۵	۰/۷۰۲	انحنای سطح
۳/۶۹۵	۰/۲۷۱	طول شیب
۱/۴۲۳	۰/۷۰۳	پوشش اراضی
۲/۲۳۸	۰/۴۴۷	لیتولوژی
۱/۱۳۲	۰/۸۸۳	فاصله از گسل
۱/۱۹۶	۰/۸۳۶	تراکم آبراهه
۱/۱۰۲	۰/۹۰۸	تراکم گسل
۱/۰۸۹	۰/۹۱۹	جهت شیب
۱/۱۷۸	۰/۸۴۹	بارندگی
۱/۱۷۱	۰/۸۵۴	فاصله از رودخانه
۴/۲۷۷	۰/۲۳۴	شیب
۱/۰۲۵	۰/۹۷۶	نوع خاک
۲/۸۵۳	۰/۳۵	شاخص رطوبت توپوگرافی
۲/۵۷۸	۰/۳۸۸	ارتفاع
۱/۶۲۴	۰/۶۱۶	انحنای آبراهه

نتایج مدل عامل اطمینان

در جدول (۲)، نتایج روش آماری عامل اطمینان ارائه شده است. بر اساس نتایج حاصل از روش عامل اطمینان، در پارامتر ارتفاع، کلاس کم‌تر از ۱۰۸ متر (۰/۳۶۱) بیش‌ترین ارتباط را با وجود آب‌های زیرزمینی داشته است درحالی‌که ارتفاع بیش‌تر از ۲۸۷ متر امتیاز منفی کسب کرده است. ارتفاعات بالا رواناب بیش‌تری دارند درحالی‌که ارتفاعات کم‌تر تغذیه و نفوذپذیری بیش‌تری دارند. بر اساس پارامتر شیب، کلاس شیب کمتر از ۶ درجه بیش‌ترین وزن را به خود اختصاص داده است که برابر ۰/۳۴۵ است. نتایج نشان می‌دهد که شیب کمتر ارتباط قوی با آب‌های زیرزمینی از خود نشان داده است درحالی‌که شیب بیش‌تر تأثیری کم‌تری بر وقوع آب زیرزمینی دارند و دلیل آن افزایش رواناب در شیب‌های بالا است. در پارامتر جهت شیب، جهت شمال با وزن ۰/۳۲۶ بیش‌ترین وزن را به خود اختصاص داده است که به دلیل دریافت میزان

بگینگ یک حالت مخصوص از روند میانگین است. یک مجموعه آموزشی استاندارد D با اندازه n را فرض کنید، بگینگ توسط نمونه‌گیری به‌طور یکنواخت و با جایگزینی مثال‌ها از D، m مجموعه آموزشی جدید D_i با اندازه $n \geq n'$ تولید می‌شود. نمونه‌گیری با جایگزینی این امکان را می‌دهد که بعضی از مثال‌ها امکان تکرار در هر D_i را داشته باشند. درنهایت این روند برای کل داده‌ها انجام می‌شود و کلاس داده در i های مختلف، مقادیر مختلفی می‌گیرد که سپس بر مبنای حداکثر تعداد مقادیر محاسبه شده، کلاس داده مشخص می‌گردد (Bui et al, 2016).

ارزیابی نتایج

خطای پیش‌بینی به‌عنوان روشی کمی، تفاوت بین مقادیر مشاهده شده و برآورد شده را تعریف می‌کند که برای تعیین دقت مدل‌های انجام شده استفاده می‌گردد. در این تحقیق، به‌منظور ارزیابی دقت مدل ترکیبی از شاخص‌های خطای جذر میانگین مربعات (RMSE) و خطای میانگین مطلق (MAE) و ضریب تعیین به‌صورت روابط ۵ و ۶ استفاده شده است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{N}} \quad (\text{رابطه ۵})$$

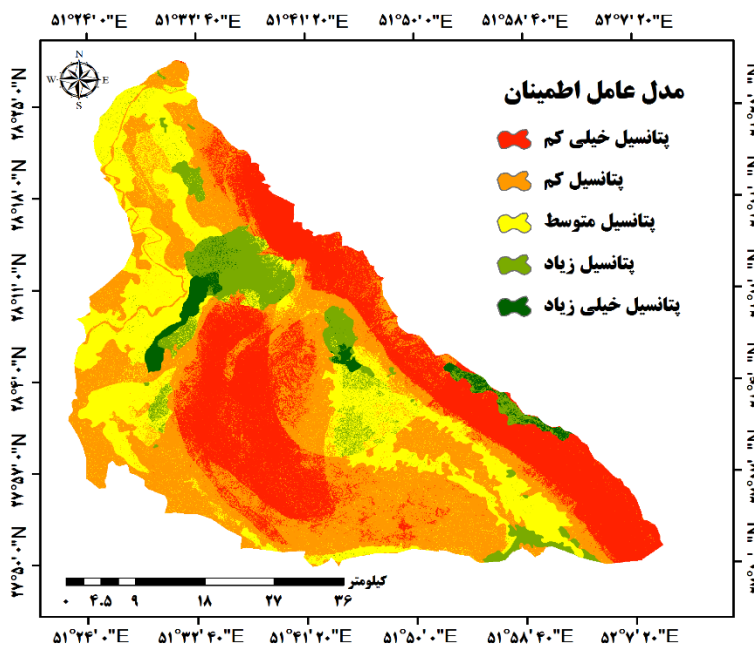
$$MAE = \frac{\sum_{i=1}^n |y_i - \bar{y}_i|}{N} \quad (\text{رابطه ۶})$$

که در آن، N تعداد کل داده‌های آموزشی، y_i مقادیر مشاهده شده و \bar{y}_i مقادیر پیش‌بینی شده است.

در این تحقیق از منحنی ROC برای ارزیابی نقشه‌های پتانسیل سطح آب زیرزمینی استفاده شد. منحنی ROC یک فن علمی است که برای توصیف کارایی روش‌های احتمالاتی و قطعی استفاده می‌شود و از طریق ترسیم رابطه بین نرخ مثبت واقعی (حساسیت) روی محور X و نرخ مثبت کاذب (۱ - ویژگی) بر محور Y شکل می‌گیرد. منحنی ROC می‌تواند به منحنی‌های نرخ موفقیت و منحنی‌های نرخ پیش‌بینی مطابق با منحنی‌های نرخ موفقیت و منحنی‌های نرخ پیش‌بینی مطابق با داده‌های مورد استفاده تقسیم شود. علاوه بر بررسی روند نمودار طبقه مورد نظر، سطح زیر آن نمودار نیز محاسبه می‌گردد. این مساحت بیانگر احتمال این است که یک مقدار انتخاب شده به‌طور تصادفی، صحیح طبقه‌بندی می‌شود و هرچه بیشتر باشد، قابلیت اطمینان روش یاد شده را نشان می‌دهد (Pourghasemi et al, 2013). در این شاخص ارزیابی مقادیری که به‌درستی به طبقه مورد نظر اختصاص یافته‌اند (TP)، مقادیری که به‌درستی به طبقه مورد نظر اختصاص نیافته‌اند (TN)، مقادیری که به نادرستی به طبقه مورد نظر اختصاص یافته‌اند (FP) و مقادیری که به نادرستی به طبقه مورد نظر اختصاص نیافته‌اند (FN) مورد استفاده قرار می‌گیرند. این منحنی شامل دو نمودار افقی (محور x) و نمودار

است (۰/۲۰۶). در پارامتر تراکم آبراهه، در روش عامل اطمینان بیشترین وزن مربوط به کلاس بیشتر از ۰/۵۸ با مقدار ۰/۷۶۱ است. با توجه به نتایج پارامتر تراکم گسل، کلاس ۰/۱۳-۰/۹ بیشترین وزن را به خود اختصاص داده است (۰/۴۰۹). در پارامتر بارندگی، کلاس بیشتر از ۲۹۷ میلی‌متر دارای بیشترین وزن (۰/۶۳۲) است. با توجه به پارامتر لیتولوژی، سازنده نهشته‌های کواترنری بیشترین وزن را به خود اختصاص داده است (۰/۴۸۹). با توجه به نتایج پارامتر پوشش اراضی، کلاس مرتع متوسط بیشترین وزن (۰/۹۰۴) را به خود اختصاص داده است و دارای بیشترین تأثیر را در وقوع آب زیرزمینی است که می‌تواند به علت نفوذ آب حاصل از آبیاری مناطق کشاورزی به سیستم آب زیرزمینی و شارژ آبخوان باشد. در پارامتر خاک، کلاس خاک‌های ماسه‌ای بیشترین وزن را به خود اختصاص داده است (۰/۳۴۸).

بارش و رطوبت بیش‌تر نسبت به جهات دیگر در کنترل ایجاد آب‌های زیرزمینی نقش بیش‌تری داشته است. بر اساس پارامتر انحنای سطح، کلاس ۰/۴ - ۰/۸ - بیشترین وزن را به خود اختصاص داده (۰/۱۸۹) و بیشترین تأثیر بر وقوع آب زیرزمینی دارد. با توجه به پارامتر انحنای آبراهه، کلاس ۰/۴ - ۱ - بیشترین وزن را به خود اختصاص داده است که برابر ۰/۲۸۳ است. در پارامتر طول شیب، کلاس ۱۰ - ۰ - متر بیشترین تأثیر بر وقوع آب زیرزمینی را نشان می‌دهد (۰/۳۳۶). با توجه به نتایج شاخص رطوبت توپوگرافی، کلاس ۶/۵۷ - ۴/۶۹ بیشترین وزن را به خود اختصاص داده است که مقدار آن برابر ۰/۳۹ است. بر اساس پارامتر فاصله از رودخانه، کلاس ۱۰۰ - ۰ - متر بیشترین وزن را به خود اختصاص داده است (۰/۴۲۹). در پارامتر فاصله از گسل، کلاس ۲۰۰۰ - ۱۰۰۰ متر بیشترین وزن را به خود اختصاص داده



شکل ۳- نقشه پتانسیل سطح آب زیرزمینی حاصل از مدل عامل اطمینان

مدل ترکیبی، وزن‌های به‌دست‌آمده از مدل‌های نسبت عامل اطمینان به‌عنوان ورودی مدل بگینگ در نظر گرفته می‌شود. در این تحقیق، مدل بگینگ در نرم‌افزار داده‌کاوی WEKA پیاده‌سازی گردید. نتایج مربوط به ضریب تعیین، RMSE و MAE مدل ترکیبی برای داده‌های آموزشی و اعتبارسنجی در جدول (۳) نشان داده‌شده است که مقادیر RMSE برای داده‌های آموزشی و اعتبارسنجی به ترتیب ۰/۲۴۷ و ۰/۲۵۶ و مقادیر شاخص MAE برای داده‌های آموزشی و اعتبارسنجی به ترتیب ۰/۱۶۲ و ۰/۱۶۹ است. همچنین مقدار ضریب تعیین برای داده‌های آموزشی و اعتبارسنجی به ترتیب برابر ۷۶ و ۷۳/۵ درصد است.

پس از تلفیق نقشه‌های به‌دست‌آمده با روش عامل اطمینان در رابطه (۴)، نقشه پتانسیل سطح آب زیرزمینی در نرم‌افزار ArcGIS بر اساس روش طبقه‌بندی شکست طبیعی (Natural breaks) به پنج طبقه پتانسیل خیلی کم، کم، متوسط، زیاد و خیلی زیاد تهیه گردید (شکل ۳). در روش شکست طبیعی، پس از تعیین تعداد طبقات، نرم‌افزار با یک الگوریتم محاسباتی، سعی در به حداقل رساندن اختلاف بین داده‌ها در هر طبقه و به حداکثر رساندن اختلاف بین طبقات می‌کند.

نتایج مدل ترکیبی

به‌منظور تهیه نقشه پتانسیل سطح آب زیرزمینی با استفاده از

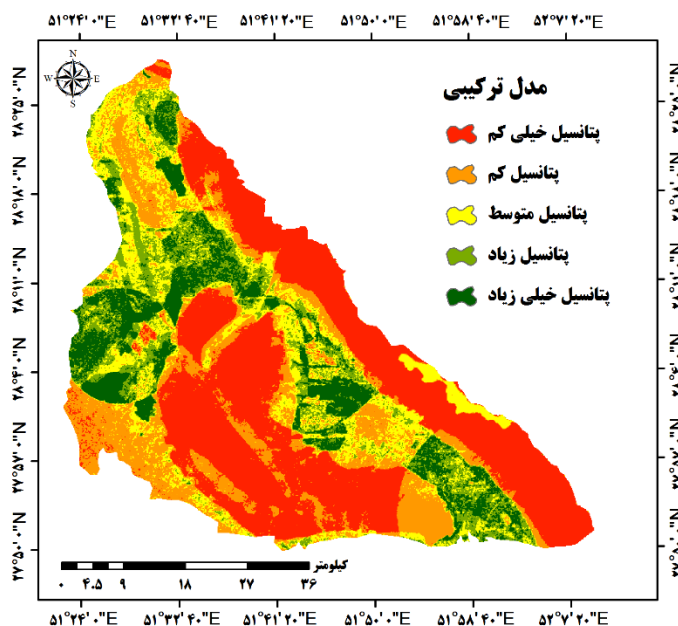
جدول ۲- رابطه مکانی بین هر معیار مؤثر و محل چاه‌ها با استفاده از مدل عامل اطمینان

عامل اطمینان	تعداد چاه	تعداد پیکسل	معیار	عامل اطمینان	تعداد چاه	تعداد پیکسل	معیار
۰/۳۴۵ -۰/۱۸۲ -۱ -۰/۹۴ -۱	۱۸۴ ۵۳ ۰ ۱ ۰	۳۴۰۲۷۶ ۱۸۳۰۱۶ ۸۵۶۹۸۸ ۴۷۵۱۳۰ ۱۵۴۸۲۷	شیب (درجه) کمتر از ۱۰۸ ۲۸۷-۱۰۸ ۵۳۵-۲۸۷ ۸۵۱-۵۳۵ بیشتر از ۸۵۱	۰/۳۶۱ ۰/۰۱۶۴ -۱ -۱ -۱	۲۳۷ ۱ ۰ ۰ ۰	۴۲۷۵۷۶ ۱۰۸۳۵۷ ۷۳۶۷۸۶ ۴۲۱۴۴۷ ۲۰۲۳۰۱	ارتفاع (متر) کمتر از ۱۰۸ ۲۸۷-۱۰۸ ۵۳۵-۲۸۷ ۸۵۱-۵۳۵ بیشتر از ۸۵۱
-۰/۷۹۴ -۰/۹۰۸ ۰/۲۸۳ -۰/۰۸۵ -۱	۱ ۳ ۱۶۷ ۶۷ ۰	۱۳۷۴۷۲ ۹۲۴۵۶۹ ۳۳۷۹۲۳ ۲۰۶۷۸۴ ۲۱۰۷۵۵	انحنای آبراهه (۱۰۰/متر) کمتر از ۳/۴ -۳/۴-۱ -۰-۱/۴ ۰/۲-۴/۸ بیشتر از ۲/۸	-۱ -۰/۳۹۱ ۰/۱۸۹ -۰/۲۷۷ -۱	۰ ۱۵ ۱۸۵ ۳۸ ۰	۱۴۶۹۶۳ ۶۹۵۶۱۹ ۴۲۳۳۱۷ ۱۴۸۴۶۵ ۱۵۹۴۶۵	انحنای سطح (۱۰۰/متر) کمتر از ۲/۲۲ -۲/۲۲-۰/۱۸ -۰/۰-۸/۴ ۰/۲-۴/۲ بیشتر از ۲/۲
-۰/۹۷۵ -۰/۵۳۳ ۰/۳۷۴ ۰/۳۹ -۱	۱ ۳۱ ۱۲۰ ۸۶ ۰	۱۱۵۲۸۱ ۱۸۷۴۹۷ ۲۱۲۰۵۹ ۱۴۷۸۸۷ ۹۲۶۱۷	شاخص رطوبت توپوگرافی (متر) کمتر از ۲/۹۲ ۲/۳-۹۲/۸۴ ۳/۴-۸۴/۶۹ ۴/۶-۶۹/۵۷ بیشتر از ۶/۵۷	۰/۳۳۶ -۰/۵۵۷ -۱ -۰/۹۳۷ -۱	۲۱۶ ۲۱ ۰ ۱ ۰	۴۰۴۹۲۴ ۱۳۴۰۸۷ ۷۴۰۵۹۷ ۴۵۳۴۲۰ ۱۳۵۷۳۹	طول شیب (متر) ۱۰-۰ ۲۰-۱۰ ۳۰-۲۰ ۴۰-۳۰ بیشتر از ۴۰
-۰/۴۵۷ -۱ -۱ -۱ ۰/۲۰۶ ۰/۱۱۷ ۰/۰۳۳	۱ ۰ ۰ ۰ ۱۹ ۵۰ ۱۶۸	۵۲۰۴۷ ۴۶۴۰۶ ۱۴۰۸۰۲ ۲۲۶۴۷۷ ۴۲۵۵۲۵ ۱۲۴۵۹۴ ۴۵۸۳۱۸	فاصله از گسل (متر) ۱۰۰-۰ ۲۰۰-۱۰۰ ۵۰۰-۲۰۰ ۱۰۰۰-۵۰۰ ۲۰۰۰-۱۰۰۰ ۵۰۰۰-۲۰۰۰ بیشتر از ۵۰۰۰	۰/۴۲۹ ۰/۱۷۵ ۰/۲۷۹ -۰/۲۷۱ -۰/۴۹۹ -۰/۷۴ ۰/۳۵	۲۶ ۱۸ ۶۰ ۴۴ ۲۰ ۶ ۶۴	۴۱۸۵۴۰ ۴۱۸۹۰۳ ۱۲۲۰۹۹ ۱۷۰۴۱۳ ۱۱۲۸۶۹ ۵۶۳۵۰۹ ۱۱۷۳۳۸	فاصله از رودخانه (متر) ۱۰۰-۰ ۲۰۰-۱۰۰ ۵۰۰-۲۰۰ ۱۰۰۰-۵۰۰ ۱۵۰۰-۱۰۰۰ ۲۰۰۰-۱۵۰۰ بیشتر از ۲۰۰۰
۰/۰۵۴ -۱ ۰/۴۰۹ -۰/۱۸۷ -۱	۱۹۰ ۰ ۲۴ ۲۴ ۰	۵۰۶۹۰۷ ۳۳۱۰۷۲ ۳۹۹۷۳۵ ۸۳۳۵۲۳ ۸۴۷۵۱	تراکم گسل کمتر از ۰/۰۳ ۰/۰-۰۳/۰۹ ۰/۰-۰۹/۱۳ ۰/۰-۱۳/۱۹ بیشتر از ۰/۱۹	۰/۱۹۵ -۰/۵۵۹ -۰/۶۰۲ ۰/۳۵۶ ۰/۷۶۱	۵۲ ۲۷ ۲۸ ۸۲ ۴۹	۱۱۸۶۹۴ ۱۷۴۰۸۵ ۱۹۹۸۲۴ ۱۴۹۷۷۷ ۳۳۱۳۹۷	تراکم آبراهه کمتر از ۰/۱۳ ۰/۰-۱۳/۲۷ ۰/۰-۲۷/۴ ۰/۰-۴/۵۸ بیشتر از ۰/۵۸
-۰/۲۱۹ -۰/۶۵۶ ۰/۲۱۶ -۰/۱۰۸ -۰/۱۶۲ ۰/۲۷۸ ۰/۰۷۳ -۰/۰۴۵ ۰/۳۲۶	۱۳ ۸ ۳۳ ۲۵ ۳۰ ۴۷ ۳۶ ۲۷ ۱۹	۴۷۰۴۰۰ ۶۵۶۸۹۵ ۷۳۰۰۴۶ ۷۹۱۴۲۹ ۱۰۱۱۵۳ ۹۵۷۸۴۲ ۹۴۱۸۳۶ ۷۹۸۸۵۱ ۳۶۱۰۴۵	جهت شیب هموار شمال شمال شرق شرق جنوب شرق جنوب جنوب غرب غرب شمال غرب	۰/۴۸۹ -۰/۶۷۲ -۰/۹۱۱ -۱ -۱ -۱ -۱ -۱ -۱ -۱	۲۲۲ ۱۳ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰	۳۲۰۲۷۰ ۱۱۲۱۳۵ ۹۵۹۵۳۹ ۱۲۳۰۸۴ ۲۲۱۸۳۶ ۲۸۸۲۸۵ ۳۸۹۲۳۸ ۲۵۸۴۶۵ ۱۲۵۴۶۴ ۳۰۴۶۲	لیتولوژی نهشته‌های کواترنری آغاچاری بختیاری میشان گچساران آسماری جهرم پایده-گورپی گورپی خامی هرمز
۰/۵۴۵ ۰/۰۰۰۰۳ ۰/۰۰۰۰۲۵ ۰/۰۰۰۰۲۹ ۰/۶۳۲	۱۱۹ ۲ ۲۱ ۷ ۸۹	۱۵۳۰۶۷ ۱۰۴۳۸۲ ۲۰۹۷۲۶ ۱۱۴۲۹۴ ۹۲۶۸۲۰	بارندگی (میلی‌متر) ۲۴۷-۰ ۲۶۴-۲۴۷ ۲۸۱-۲۶۴ ۲۹۷-۲۸۱ بیشتر از ۲۹۷	-۱ ۰/۸۲۳ ۰/۵۸۳ ۰/۳۲ -۱ -۰/۶۸ ۰/۷۶۹ ۰/۹۰۴ -۱ -۱	۰ ۱۶ ۱۰۰ ۱۸ ۰ ۴۱ ۱۴ ۴۹ ۰ ۰	۱۸۰۰۵۷ ۷۹۶۳۶ ۱۱۷۴۸۳ ۳۴۵۴۰۹ ۳۵۴۱۸۱ ۳۷۳۰۶۷ ۹۱۲۷۵ ۱۳۲۷۶۲ ۱۸۱۶۸ ۵۷۴۱۶۲ ۳۹۲۲۶	پوشش اراضی جنگل مانگرو مناطق شهری کشاورزی اراضی نمکی جنگل با پوشش کم مرتع ضعیف تپه شنی مرتع متوسط باغ صخره پهنه آبی
۰/۳۴۸ -۱ -۱	۲۳۸ ۱ ۰	۴۲۶۸۶۵ ۱۴۰۰۸۵ ۹۳۹۰۴۱	نوع خاک خاک‌های ماسه‌ای خاک‌های شور خاک‌های لیتوسل نیمه مرطوب	-۱ -۱ -۱	۰ ۰ ۰	۱۸۱۶۸ ۵۷۴۱۶۲ ۳۹۲۲۶	

پس از آموزش مدل بگینگ در ترکیب با روش عامل اطمینان، فرآیند مدل سازی به کل منطقه مورد مطالعه تعمیم داده شد و نقشه پتانسیل سطح آب زیرزمینی در نرم افزار ArcGIS بر اساس روش طبقه بندی Natural breaks به پنج طبقه پتانسیل خیلی کم، کم، متوسط، زیاد و خیلی زیاد تهیه گردید (شکل ۴).

جدول ۳- نتایج ارزیابی مدل ترکیبی

معیار	داده آموزش	داده اعتبارسنجی
RMSE	۰/۲۴۷	۰/۲۵۶
MAE	۰/۱۶۲	۰/۱۶۹
ضریب تعیین	۷۶ درصد	۷۳/۵ درصد

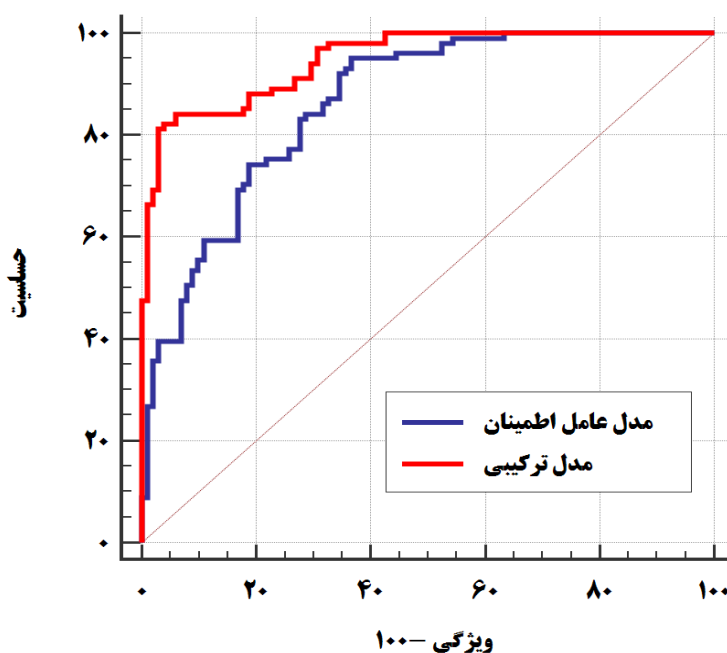


شکل ۴- نقشه پتانسیل سطح آب زیرزمینی حاصل از مدل ترکیبی

اطمینان و مدل ترکیبی عامل اطمینان-بگینگ به ترتیب دارای دقت های ۰/۸۶۲ و ۰/۹۴۸ می باشند (جدول ۴). نتایج نشان دهنده دقت بالای ترکیب مدل های آماری با مدل های داده کاوی در تهیه نقشه پتانسیل سطح آب زیرزمینی است.

ارزیابی نتایج و بحث

نمودار ROC مربوط به مدل عامل اطمینان و مدل ترکیبی در شکل (۵) نشان داده شده است. نتایج نشان می دهد که مدل عامل



شکل ۵- منحنی ROC نقشه های پتانسیل سطح آب زیرزمینی با استفاده از مدل عامل اطمینان و مدل ترکیبی

جدول ۴- مساحت زیر منحنی تشخیص عملکرد نسبی به همراه استاندارد و سطح اطمینان آن

مدل	سطح زیر منحنی	خطای استاندارد	سطح اطمینان	حد پایین	حد بالا	فواصل اطمینان در سطح ۰/۹۵
عامل اطمینان	۰/۸۶۲	۰/۰۲۵۱	۰/۰۰۰	۰/۸۰۷	۰/۹۰۷	
عامل اطمینان-بگینگ	۰/۹۴۸	۰/۰۱۳۶	۰/۰۰۰	۰/۹۰۸	۰/۹۷۴	

نتیجه‌گیری

آب‌های زیرزمینی یکی از مهم‌ترین منبع طبیعی قابل اطمینان و پایدار برای منابع آب در سراسر جهان است. با این وجود، بهره‌برداری بیش‌ازحد از آب‌های زیرزمینی منجر به تخلیه سفره‌های آب زیرزمینی در سراسر جهان شده است. توسعه پایدار و مدیریت این منابع حیاتی تحت شرایط تغییر شرایط محیطی، یکی از چالش‌های عمده قرن بیست و یکم است. روش‌های سنتی اکتشاف آب زیرزمینی، هزینه‌بر است، بنابراین با استفاده از روش‌های مبتنی بر GIS می‌توان مناطق بالقوه آب زیرزمینی را با صرف هزینه و زمان کمتری اکتشاف کرد. در روش‌های سنتی، متخصصین اکتشاف آب زیرزمینی بدون اطلاع از مناطق بالقوه آب زیرزمینی و بر اساس فرضیات و ساختار سازنده‌ها به بررسی مناطق دارای پتانسیل آب زیرزمینی می‌پردازند درحالی‌که به کل منطقه دسترسی نداشته و همچنین ممکن است در بسیاری مواقع نتایج موفقیت‌آمیزی از اکتشاف به دست نیاورند. تهیه نقشه پتانسیل سطح آب زیرزمینی با استفاده از سیستم اطلاعات جغرافیایی، نقشه‌هایی که دارای پتانسیل بیشتر آب زیرزمینی هستند در اختیار متخصصین قرار داده که به راحتی و بدون صرف زمان و همچنین دوری از فرضیات به اکتشاف آب زیرزمینی بپردازند. از این رو هدف از این تحقیق تهیه نقشه پتانسیل سطح آب زیرزمینی با استفاده از مدل آماری عامل اطمینان و ترکیب آن با مدل داده‌کاوی بگینگ در دشت بوشهر است. در ابتدا نقشه پراکنش چاه شامل ۳۳۹ محل وقوع چاه در منطقه مورد مطالعه تعیین و سپس ۱۵ معیار تأثیرگذار بر آب زیرزمینی در محیط GIS آماده گردید. بدین منظور از ۳۳۹ محل وقوع چاه، ۲۳۸ محل (۷۰ درصد) به‌عنوان داده‌های آموزشی جهت مدل‌سازی و ۱۰۱ محل (۳۰ درصد) جهت ارزیابی با استفاده از منحنی تشخیص عملکرد نسبی مورد استفاده قرار گرفت. دقت‌های به‌دست‌آمده از مدل‌های عامل اطمینان و ترکیب عامل اطمینان با روش بگینگ به ترتیب ۸۶/۲ درصد و ۹۴/۸ درصد است. همچنین نتایج ارزیابی مدل‌های موردنظر نشان‌دهنده دقت بالاتر مدل ترکیبی نسبت به هر یک از مدل‌ها به‌تنهایی است؛ بنابراین به‌عنوان پیشنهاد می‌توان گفت که نقشه‌های مذکور، می‌تواند نقش به‌سزایی در تصمیم‌گیری‌های آبی، مدیریت آب‌های زیرزمینی در منطقه مورد مطالعه داشته باشد.

تاکنون روش‌های آماری متعددی به‌منظور تهیه نقشه پتانسیل سطح آب زیرزمینی به کار گرفته شده است، اما به دلیل نیاز به دقت بیشتر در تعیین مناطق بالقوه از روش‌های داده‌کاوی و ترکیب آن با روش‌های آماری استفاده می‌شود. از طرفی، قسمت عمده‌ای در فرآیند مدل‌سازی روش‌های داده‌کاوی مربوط به بحث آماده‌سازی و پردازش داده‌ها است. از این رو در این تحقیق از ترکیب فاکتور اطمینان با مدل بگینگ برای تهیه نقشه آب زیرزمینی در دشت بوشهر استفاده شده است. بر طبق نتایج به‌دست‌آمده از منحنی ROC، مدل ترکیبی با سطح زیر منحنی ۰/۹۴۸ دارای دقت بالاتری نسبت به مدل عامل اطمینان به‌تنهایی (۰/۸۶۲) است.

تفاوت اساسی بین روش‌های آماری و روش‌های داده‌کاوی بر اساس فرض‌ها و یا طبیعت داده‌هایی که پردازش می‌شوند. به‌عنوان یک قانون کلی فرض‌های فن‌های آماری بر این اساس است که توزیع داده‌ها مشخص است که بیشتر موارد فرض بر این است که توزیع نرمال است و در نهایت درستی یا نادرستی نتایج نهایی به درست بودن فرض اولیه وابسته است. در مقابل روش‌های داده‌کاوی از هیچ فرض در مورد داده‌ها استفاده نمی‌کند و همین مورد باعث تفاوت‌هایی بین این دو روش می‌شود (Clapcott *et al.*, 2013). از دیگر مزیت‌های روش‌های داده‌کاوی نسبت به روش‌های آماری، درجه‌هایی که داده‌ها ناقص باشند و یا با یکدیگر متناقض باشند، روش‌های داده‌کاوی بسیار بهتر عمل می‌کنند. چراکه در این موارد به نحوی داده‌ی مفقودشده بر اساس الگوی موجود در داده‌ها بازیابی می‌شود. این در حالی است که در روش‌های آماری، مشکل نبود برخی پارامترها در برخی داده‌ها، منجر به بلااستفاده شدن آن داده می‌شود. با افزایش تعداد پارامترها، روش‌های آماری توانایی یافتن الگوها را از دست می‌دهند و به علت ماهیت خطی خود از کشف روابط غیرخطی و پیچیده بین متغیرها عاجز هستند؛ اما روش‌های داده‌کاوی طوری طراحی شده‌اند که می‌توانند روابط مرکب و پیچیده بین چندین پارامتر را در پایگاه داده کشف کنند (Chen *et al.*, 2018b). بر طبق نتایج به‌دست‌آمده ترکیب روش‌های آماری با الگوریتم‌های داده‌کاوی باعث افزایش دقت در تهیه نقشه پتانسیل سطح آب زیرزمینی شده است که با تحقیقات (Naghbi *et al.*, 2014; Nampak *et al.*, 2017) مطابقت دارد.

REFERENCES

- Al-Abadi, A. M., Al-Temmeme, A. A., & Al-Ghanimy, M. A. (2016). A GIS-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al Al-Gharbi–Teeb areas, Iraq. *Sustainable Water Resources Management*, 2(3), 265-283.
- Ayazi, M. H., Pirasteh, S., Arvin, A. K. P., Pradhan, B., Nikouravan, B., & Mansor, S. (2010). Disasters and risk reduction in groundwater: Zagros Mountain Southwest Iran using geoinformatics techniques. *Disaster Adv*, 3(1), 51-57.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- Bui, D. T., Ho, T. C., Pradhan, B., Pham, B. T., Nhu, V. H., & Revhaug, I. (2016). GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences*, 75(14), 1101.
- Chen, W., Shahabi, H., Shirzadi, A., Li, T., Guo, C., Hong, H. Ma, M. (2018a). A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. *Geocarto International*, 33(12), 1398-1420.
- Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M. Niu, C. (2018b). GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Science of the Total Environment*, 634, 853-867.
- Clapcott, J., Goodwin, E., & Snelder, T. (2013). Predictive Models of Benthic Macroinvertebrate Metrics. Prepared for Ministry for the Environment. Retrieved from
- Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., ... & Althuwaynee, O. F. (2013). Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Natural hazards*, 65(1), 135-165.
- Ercanoglu, M., & Gokceoglu, C. (2002). Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. *Environmental Geology*, 41(6), 720-730.
- Fitts, C. R. (2002). *Groundwater science*: Elsevier.
- Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS. *Environmental monitoring and assessment*, 190(3), 149.
- Greenbaum, D. (1992). Structural influences on the occurrence of groundwater in SE Zimbabwe. *Geological Society, London, Special Publications*, 66(1), 77-85.
- Israil, M., Al-Hadithi, M., & Singhal, D. C. (2006). Application of a resistivity survey and geographical information system (GIS) analysis for hydrogeological zoning of a piedmont area, Himalayan foothill region, India. *Hydrogeology Journal*, 14(5), 753-759.
- Jha, M. K., Chowdary, V. M., & Chowdhury, A. (2010). Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *Hydrogeology journal*, 18(7), 1713-1728.
- Jha, M. K., Chowdhury, A., Chowdary, V. M., & Peiffer, S. (2007). Groundwater management and development by integrated remote sensing and geographic information systems: prospects and constraints. *Water Resources Management*, 21(2), 427-467.
- Jothibasu, A., & Anbazhagan, S. (2016). Modeling groundwater probability index in Ponnaiyar River basin of South India using analytic hierarchy process. *Modeling Earth Systems and Environment*, 2(3), 109.
- Kanungo, D. P., Sarkar, S., & Sharma, S. (2011). Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. *Natural hazards*, 59(3), 1491.
- Khosravi, K., Panahi, M., & Tien Bui, D. (2018). Spatial Prediction of Groundwater Spring Potential Mapping Based on Adaptive Neuro-Fuzzy Inference System and Metaheuristic Optimization.
- Kordestani, M. D., Naghibi, S. A., Hashemi, H., Ahmadi, K., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeology Journal*, 1-14.
- Lee, S., & Lee, C.-W. (2015). Application of decision-tree model to groundwater productivity-potential mapping. *Sustainability*, 7(10), 13416-13432.
- Lee, S., Hong, S.-M., & Jung, H.-S. (2018). GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. *Geocarto International*, 33(8), 847-861.
- Lee, S., Hyun, Y., & Lee, M.-J. (2019). Groundwater Potential Mapping Using Data Mining Models of Big Data Analysis in Goyang-si, South Korea. *Sustainability*, 11(6), 1678.
- Lee, S., Kim, Y. S., & Oh, H. J. (2012). Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping. *Journal of Environmental Management*, 96(1), 91-105.
- Mair, A., & El-Kadi, A. I. (2013). Logistic regression modeling to assess groundwater vulnerability to contamination in Hawaii, USA. *Journal of contaminant hydrology*, 153, 1-23.
- Mogaji, K., Lim, H., & Abdullah, K. (2015). Regional prediction of groundwater potential mapping in a multifaceted geology terrain using GIS-based Dempster–Shafer model. *Arabian Journal of*

- Geosciences*, 8(5), 3235-3258.
- Moore, I. D., Grayson, R., & Ladson, A. (1991). Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*, 5(1), 3-30.
- Moore, I., & Burch, G. (1986). Sediment transport capacity of sheet and rill flow: application of unit stream power theory. *Water Resources Research*, 22(8), 1350-1360.
- Mukherjee, S., Aadhar, S., Stone, D., & Mishra, V. (2018). Increase in extreme precipitation events under anthropogenic warming in India. *Weather and climate extremes*, 20, 45-53.
- Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, 31(9), 2761-2775.
- Naghibi, S. A., Pourghasemi, H. R., & Abbaspour, K. (2018). A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theoretical and applied climatology*, 131(3-4), 967-984.
- Naghibi, S. A., Pourghasemi, H. R., Pourtaghi, Z. S., & Rezaei, A. (2015). Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. *Earth Science Informatics*, 8(1), 171-186.
- Nampak, H., Pradhan, B., & 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.
- Oh, H. J., Kim, Y. S., Choi, J. K., Park, E., & Lee, S. (2011). GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology*, 399(3-4), 158-172.
- Osati, K., Koeniger, P., Salajegheh, A., Mahdavi, M., Chapi, K., & 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(2), 169-183.
- Pham, B. T., Jaafari, A., Prakash, I., Singh, S. K., Quoc, N. K., & Bui, D. T. (2019). Hybrid computational intelligence models for groundwater potential mapping. *Catena*, 182, 104101.
- Pourghasemi, H., Moradi, H., & Aghda, S. F. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, 69(1), 749-779.
- Rahmati, O., Samani, A. N., Mahdavi, M., Pourghasemi, H. R., & Zeinivand, H. (2015). Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arabian Journal of Geosciences*, 8(9), 7059-7071.
- Razandi, Y., Pourghasemi, H. R., Neisani, N. S., & Rahmati, O. (2015). Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. *Earth Science Informatics*, 8(4), 867-883.
- Razavi Termeh, S., Mesgari, M., kazemi, K. (2017). Evaluation and comparison of frequency ratio, statistic index and entropy methods for groundwater potential mapping using GIS (Case Study: Jahrom Township). *Iranian journal of Ecohydrology*, 4(3), 725-736 (in farsi).
- Tahmassebipoor, N., Rahmati, O., Noormohamadi, F., & Lee, S. (2016). Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arabian Journal of Geosciences*, 9(1), 79.
- Termeh, S. V. R., Kornejady, A., Pourghasemi, H. R., & Keesstra, S. (2018). Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment*, 615, 438-451.
- Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied artificial intelligence*, 17(5-6), 375-381.