

استفاده از الگوریتم‌های یادگیری ماشین برای مدل‌سازی منابع آب زیرزمینی در دشت

موسیان استان ایلام

نازنین سلیمی؛ گروه مهندسی طبیعت (مرتع و آبخیزداری)، دانشکده کشاورزی، دانشگاه ایلام، ایلام، ایران.
مرزبان فرامرزی^۱؛ گروه مهندسی طبیعت (مرتع و آبخیزداری)، دانشکده کشاورزی، دانشگاه ایلام، ایلام، ایران.
محسن توکلی؛ گروه مهندسی طبیعت (مرتع و آبخیزداری)، دانشکده کشاورزی، دانشگاه ایلام، ایلام، ایران.
حسن فتحی‌زاد؛ گروه مهندسی طبیعت (مرتع و آبخیزداری)، دانشکده کشاورزی، دانشگاه ایلام، ایلام، ایران.

پذیرش نهایی: ۱۴۰۲/۰۸/۲۳

دریافت مقاله: ۱۴۰۱/۱۰/۲۷

چکیده

امروزه میزان برداشت از آبهای زیرزمینی بیش از میزان تغذیه آبهای زیرزمینی است که این عامل باعث افت شدید سطح سفره‌های آب زیرزمینی شده است. مراتع و جنگل‌ها بعنوان اصلی‌ترین مکان‌های تغذیه سفره‌های آب زیرزمینی محسوب می‌شوند، درحالی‌که بیشترین برداشت از این منابع در کاربری کشاورزی انجام می‌شود. هدف اصلی از پژوهش حاضر استفاده از الگوریتم‌های یادگیری ماشین شامل رگرسیون جنگل تصادفی و تابع آنتروپی شانون برای مدل‌سازی منابع آب زیرزمینی در مراتع نیمه‌خشک غرب ایران می‌باشد. برای این هدف، ابتدا لایه‌های اطلاعاتی شامل: درجه شیب، جهت شیب، ارتفاع از سطح دریا، فاصله از گسل، شکل شیب، فاصله از آبراهه، فاصله از جاده، بارندگی، لیتولوژی و کاربری اراضی تهیه شد. پس از تعیین وزن پارامترها با استفاده از تابع آنتروپی شانون و سپس تعیین طبقات آن‌ها، در محیط سامانه‌های اطلاعات جغرافیایی، از ترکیب وزن پارامترها و طبقات آن‌ها نقشه نهایی مناطق دارای پتانسیل منابع آب زیرزمینی مدل‌سازی گردید. بعلاوه، برای اجرای مدل جنگل تصادفی از نرم‌افزار R ۳.۵.۱ و بسته randomForest استفاده شد. در تحقیق حاضر از اعتبارسنجی ضربدری k-fold برای صحت‌سنجی مدل‌ها استفاده گردید. به منظور ارزیابی کارایی مدل‌های جنگل تصادفی و آنتروپی شانسون برای پتانسیل‌یابی منابع آب زیرزمینی، از شاخص‌های آماری MAE، RMSE و R^2 استفاده شد. نتایج ارزیابی نشان داد که مدل جنگل تصادفی با دقت (MAE: ۲٫۸۵، RMSE: ۳٫۴۱، $R^2 = ۰٫۸۲۵$) دارای دقت بالاتری نسبت به مدل آنتروپی شانون با دقت (MAE: ۳٫۳۴، RMSE: ۴٫۳۶، $R^2 = ۰٫۷۲۷$) می‌باشد. یافته‌های مدل جنگل تصادفی نشان داد که قسمت زیادی از منطقه مورد مطالعه دارای پتانسیل متوسط (۲۶۹۵۴/۲۲ هکتار) و مساحت خیلی اندکی (۲۰۵/۶۱ هکتار) بدون پتانسیل آب زیرزمینی می‌باشد. از طرفی، نتایج مدل آنتروپی شانون نشان داد که قسمت اعظمی از منطقه مورد مطالعه دارای پتانسیل متوسط (۲۴۶۳۳/۰۵ هکتار) و مساحت خیلی اندکی (۱۵۰۲/۱۲ هکتار) بدون پتانسیل آب زیرزمینی می‌باشد.

واژه‌های کلیدی: پتانسیل آب زیرزمینی، مدل جنگل تصادفی، آنتروپی شانون، دشت موسیان.

مقدمه

در دهه‌های اخیر در بسیاری از نقاط دنیا به دلیل گسترش جمعیت و صنعتی شدن، آب‌های زیرزمینی به عنوان یکی از مهم‌ترین منابع طبیعی شناخته شده‌اند (Jothibas and Anbazhagan, ۲۰۱۶) و نقش حیاتی در توسعه پایدار کشاورزی، جامعه و اقتصاد ایفا می‌کنند (Pande et al., ۲۰۲۱). وجود آب زیرزمینی در هر نقطه از کره زمین، نتیجه برهم‌کنش بین عوامل مختلف اقلیمی، زمین‌شناسی، هیدرولوژیکی، فیزیوگرافی و اکولوژیکی است (Naghbi et al., ۲۰۱۷). به عنوان مثال، هر گونه تغییر کاربری از مرتع به کاربری‌های دیگر مانند کشاورزی و جنگل‌های دست کاشت باعث افت سفره‌های آب زیرزمینی می‌شود (Yaghobi et al., ۲۰۱۹). حرکت آب زیرزمینی درون زمین تحت تاثیر عوامل مختلفی از قبیل سنگ‌شناسی، توپوگرافی، ساختار زمین‌شناسی، شکاف‌ها (تراکم، اندازه و ارتباط)، تخلخل ثانویه، تغذیه آب زیرزمینی، الگوی زهکشی، شکل زمین، پوشش زمین و شرایط اقلیمی است (Oh et al., ۲۰۱۱; Pourtaghi and Pradhan, ۲۰۰۹; Pourghasemi, ۲۰۱۴). بی‌گمان با در نظر گرفتن عوامل موثر می‌توان به تهیه نقشه‌ها و پیش‌بینی پتانسیل آب زیرزمینی با استفاده از روش‌های مختلف در هر منطقه به منظور مدیریت صحیح این منابع با ارزش پرداخت. از طرفی دیگر، وقوع خشکسالی‌های متناوب و طولانی و تغییرات اقلیمی از عوامل اصلی کمبود آب بویژه منابع آب سطحی است که فشار مضاعفی را بر منابع آب زیرزمینی وارد می‌کند (Halder et al., ۲۰۲۰).

با توجه به افزایش تقاضا برای منابع آب شیرین، دو ابزار سامانه‌های اطلاعات جغرافیایی (GIS) و سنجش از دور (RS) مناسب توسط مهندسين و برنامه‌ریزان برای مدیریت کارآمد استفاده از منابع آب زیرزمینی مورد استفاده قرار می‌گیرد (Oh et al., ۲۰۱۱). سیستم اطلاعات جغرافیایی (GIS) می‌تواند در فرآیند تصمیم‌گیری در مدیریت منابع آب مورد استفاده قرار گیرد زیرا GIS یک ابزار قدرتمند است (Nampak et al., ۲۰۱۴). محققان متعددی ترکیبی از ابزارهای GIS و سنجش از دور (RS) را برای ارزیابی نقشه‌برداری بالقوه آب‌های زیرزمینی در مقیاس متوسط تا منطقه‌ای استفاده کرده‌اند. با پیشرفت تکنیک‌ها و نرم‌افزارهای RS و GIS، نقشه‌برداری از پتانسیل آب زیرزمینی به روشی آسان و کارآمد تبدیل شده است (Tweed et al., ۲۰۰۷). با این حال، در کشورهای در حال توسعه و آستانه مانند ایران، فقدان شدید مجموعه داده‌های دقیق و کامل وجود دارد، به همین دلیل، داده‌های سنجش از دور برای درک وضعیت آب زیرزمینی و متعاقباً مدیریت آن ضروری است.

حرکت آب زیرزمینی درون زمین تحت تاثیر عوامل مختلفی از قبیل سنگ‌شناسی، توپوگرافی، ساختار زمین‌شناسی، شکاف‌ها (تراکم، اندازه و ارتباط)، تخلخل ثانویه، تغذیه آب زیرزمینی، الگوی زهکشی، شکل زمین، پوشش زمین و شرایط اقلیمی است (Pourtaghi et al., ۲۰۱۴). بی‌گمان با در نظر گرفتن عوامل موثر می‌توان به تهیه نقشه‌ها و پیش‌بینی پتانسیل آب زیرزمینی با استفاده از روش‌های مختلف در هر منطقه به منظور مدیریت صحیح این منابع با ارزش پرداخت. وقوع خشکسالی‌های متناوب و طولانی و نوسان زیاد آب و هوا از عوامل اصلی کمبود آب بویژه منابع آب سطحی است که فشار مضاعفی را بر منابع آب زیرزمینی وارد می‌کند. براساس اجرای پژوهش‌های بیشتر در زمینه پتانسیل یابی، اکتشاف و بهره‌برداری بهینه از منابع آب زیرزمینی ضروری به نظر می‌رسد. روش‌های متداول مورد استفاده برای تعیین مناطق دارای پتانسیل آب زیرزمینی اصولاً براساس مطالعات و بازدیدهای صحرائی است (Ganapuram et al., ۲۰۰۹). با توجه به اینکه

اجرای آزمایش‌های اکتشافی منابع آب زیرزمینی نیازمند صرف زمان و هزینه بسیار زیادی است، استفاده از تکنیک‌های جدید از جمله سامانه‌های اطلاعات جغرافیایی و سنجش از دور در این زمینه ضروری به نظر می‌رسد. نقشه‌های پتانسیل آب زیرزمینی می‌تواند مورد استفاده مدیران و برنامه‌ریزان به منظور سیاست‌گذاری و برنامه‌ریزی‌های خرد و کلان کشور قرار گیرد. بر همین اساس تعیین مناطقی که دارای پتانسیل آب زیرزمینی هستند از ابزارهای مهم در حفاظت، مدیریت و بهره‌برداری از منابع آب است (Moghaddam et al., ۲۰۱۵).

از این نظر، محققان زیادی از روش‌ها و مدل‌های مختلف استفاده کرده‌اند برای مثال، خوش‌تیانت و همکاران (۲۰۱۹) از روش‌های *weights of evidence, weighting factor, and statistical index*، هو و همکاران (Hou et al., ۲۰۱۸) از شاخص آماری، شاخص آنتروپی و مدل‌های قطعیت عوامل، نهو و همکاران (Nhu et al., ۲۰۲۰) از روش‌های *Novel Ensemble Bivariate and Multivariate Models* برای پتانسیل‌یابی آب زیرزمینی استفاده کرده‌اند. اخیراً، برخی محققان از الگوریتم‌های داده‌کاوی مانند *BRT*^۲، *KNN*^۳، *LDA*^۴، *QDA*^۵ و *RF*^۶ برای پتانسیل‌یابی آب زیرزمینی استفاده کرده‌اند (Tien Bui et al., ۲۰۱۶; Naghibi et al., ۲۰۱۶).

منطقه مورد مطالعه در این تحقیق در شهرستان دهلران در جنوب استان ایلام واقع شده است. طبق مطالعات قبلی این منطقه در کلاس خشک اقلیمی واقع شده است (Yaghoobi et al., ۲۰۱۹) که دارای متوسط بارش کم و متوسط دمای سالانه بالایی می‌باشد. از طرفی، نظر به اهمیت موضوع و خشکسالی‌های به وقوع پیوسته در کشور و منطقه، شناسایی منابع آب زیرزمینی و مدیریت صحیح این منابع به عنوان راهکار توسعه‌ی جامع و پایدار امری حائز اهمیت است. بنابراین، هدف اصلی از مطالعه حاضر، مدل‌سازی منابع آب زیرزمینی با استفاده از الگوریتم‌های یادگیری ماشین در مناطق خشک غرب ایران می‌باشد. الویت‌بندی فاکتورهای موثر بر مناطق مستعد آب زیرزمینی و همچنین بررسی کارایی الگوریتم‌های یادگیری ماشین، سنجش از دور و سامانه‌های اطلاعات جغرافیایی در مدل‌سازی منابع آب زیرزمینی از موضوعات دیگر این تحقیق می‌باشد.

داده‌ها و روش کار

الف) قلمرو جغرافیایی مورد مطالعه

منطقه مورد مطالعه قسمتی از دشت موسیان است با وسعت ۴۲۸۷۵ هکتار، بین طول‌های جغرافیایی "۲۰' ۱۴' ۴۷ تا" ۳۸' ۱۲' ۴۷° و عرض‌های جغرافیایی "۲۱' ۳۲' ۳۲ تا" ۱۴' ۳۷' ۳۲ در غرب ایران قرار گرفته است (شکل ۱). میانگین ارتفاع، بارندگی و دما منطقه مورد مطالعه به ترتیب ۱۳۴ متر، ۲۳۵/۲ میلی‌متر و ۲۶/۴ درجه سانتیگراد است. اقلیم این منطقه بر اساس طبقه‌بندی دمارتن داری اقلیم خشک است (de Martonne ۱۹۲۶). دمای میانگین خاک در عمق پنج سانتیمتری ۲۹/۴ و در عمق ده سانتی متری ۲۸/۸ واحد است.

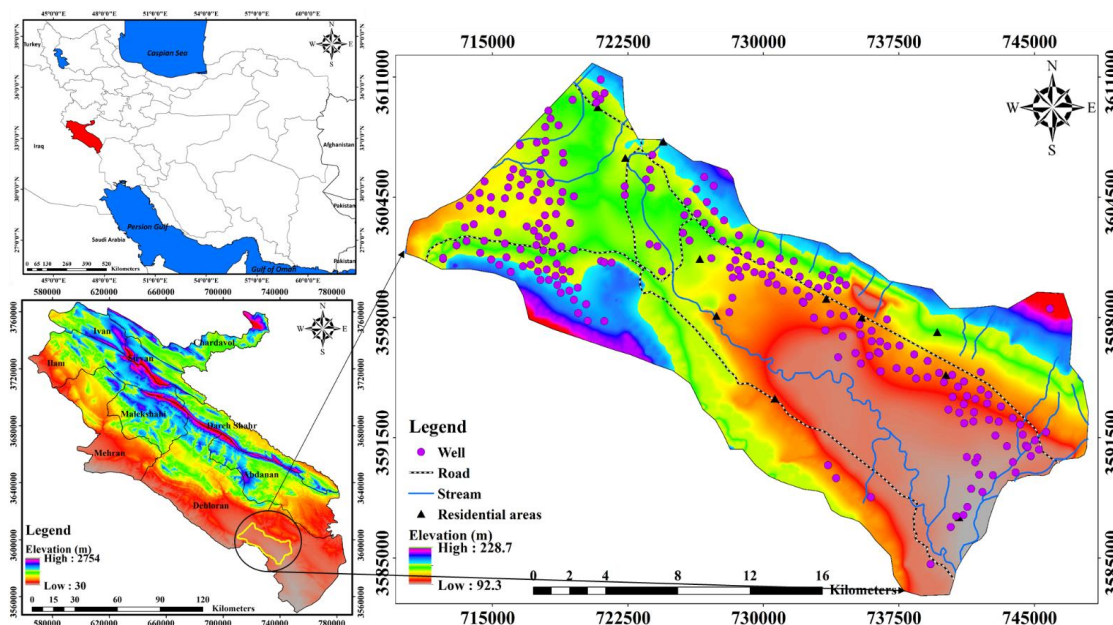
^۲ Boosted Regression Trees

^۳ K nearest neighbor

^۴ Linear discriminant analysis

^۵ Quadric discriminant analysis

^۶ Random forests



شکل ۱- موقعیت چاه‌های بهره‌برداری، جاده‌ها، رودخانه‌ها و مناطق مسکونی در منطقه مورد مطالعه روی نقشه‌های استان ایلام و کشور ایران (لایه Digital Elevation Model منطقه، استخراج شده از نقشه توپوگرافی با قدرت تفکیک ۳۰ متر)

ب) داده‌های مورد نیاز

برای انجام مطالعات این تحقیق دو دسته داده لازم می‌باشد:

۱- داده‌های مربوط به چاه‌های منطقه مورد مطالعه

این داده‌ها از شرکت آب منطقه‌ای ایلام تهیه گردید. اطلاعات مربوط به ۲۳۶ حلقه چاه در منطقه مطالعاتی دشت موسیان استخراج گردید. از این آمار تمامی آمار موجود دبی، سطح سفره و موقعیت جغرافیایی چاه‌ها مورد استفاده قرار گرفت (شکل ۱).

۲- تهیه لایه‌های اطلاعاتی مورد نیاز

لایه‌های اطلاعاتی شامل مدل رقومی ارتفاع، درصد شیب، جهت شیب، کاربری اراضی، زمین‌شناسی، خاکشناسی، شاخص رطوبت توپوگرافی، تعادل سلولی، شکل زمین، تجمع جریان، گرادیان، شاخص قدرت جریان، شاخص ناهمواری زمین، حرارت روزانه انجماد می‌باشند.

۳- سپس داده‌ها و لایه‌های موجود، در مدل‌های آنتروپی شانون و جنگل تصادفی مورد استفاده و تجزیه و تحلیل قرار گرفت و در نهایت نقشه‌های پتانسیل آب زیرزمینی در هر یک از مدل‌ها استخراج گردید.

ج) روش تحقیق

به منظور اجرای پژوهش حاضر، نخست فاکتورهای توپوگرافی از قبیل درصد شیب، جهت شیب، ارتفاع، شکل زمین، شاخص ناهمواری زمین و غیره استخراج گردید. سپس فاکتورهای مرتبط با آب مانند شاخص رطوبت توپوگرافی، تجمع جریان، شاخص قدرت جریان، و غیره، فاکتورهای زمین‌شناسی، خاکشناسی (بافت خاک) و کاربری اراضی تهیه شد. همه

لایه‌های اطلاعاتی مورد استفاده در این تحقیق شامل درجه شیب، جهت شیب، شکل شیب، طول شیب، طبقه ارتفاعی، شاخص رطوبت توپوگرافی، شاخص قدرت جریان، شاخص ناهمواری زمین، تجمع جریان، گرادیان، زمین‌شناسی و کاربری اراضی و غیره با توجه به سابقه پژوهش‌های مختلف صورت گرفته در جهان و ایران (Naghibi et al, ۲۰۱۴; Manap et al, ۲۰۱۶; Rahmati et al, ۲۰۱۱; Ozdemir, ۲۰۱۵; et al, ۲۰۱۵) به عنوان عوامل موثر بر پتانسیل آب زیرزمینی از روی نقشه-های پایه (توپوگرافی، زمین‌شناسی و...) موجود با استفاده از الحاقیه‌های مختلف در محیط سامانه‌های اطلاعات جغرافیایی رقومی و تهیه شد. به این منظور با استفاده نقشه مدل رقومی ارتفاعی (Digital Elevation Model: DEM) فاکتورهای درجه شیب، جهت شیب، شکل شیب، طبقه ارتفاعی و شاخص رطوبت توپوگرافی از آن استخراج گردید. لایه اطلاعاتی کاربری اراضی از تصاویر ماهواره ای لندست استخراج شد. نقشه زمین‌شناسی منطقه تحقیق نیز از سازمان مربوطه تهیه و در محیط GIS رقومی و لایه‌های زمین‌شناسی، شکل زمین، شاخص ناهمواری زمین از آن استخراج شد. لایه‌های مربوط به آبراهه نیز از نقشه توپوگرافی منطقه تهیه شد. اطلاعات مربوط به سفره‌های آب زیرزمینی با استفاده از چاه‌های پیژومتری و بهره برداری ثبت شده توسط شرکت آب منطقه‌ای استخراج گردید. برای تهیه نقشه لایه خاک از سایت <https://www.isric.org/explore/soilgrids> استفاده شده است (Hengli et al., ۲۰۱۷). SoilGrids سیستمی برای نقشه‌برداری خودکار خاک است که بر اساس پیشرفته‌ترین روش‌های پیش‌بینی مکانی است.

• مدل آنتروپی شانون

تئوری آنتروپی را نخستین بار آنتروپی شانون به صورت کمی بیان کرد (Shannon, ۲۰۰۱). آنتروپی، یکی از رویکردهای مدیریتی است که به منظور برخورد با، بی‌ثباتی و عدم قطعیت‌های موجود در یک سیستم استفاده می‌شود (Pourghasemi et al, ۲۰۱۲). شاخص آنتروپی مقیاسی از مقدار برابری است که در آن گروه‌ها به طور مساوی و یکنواخت در بین واحدهای سازمانی توزیع شده اند (Massey & Nancy, ۱۹۹۸). رابطه یک به یکی بین کمیت و مقدار آنتروپی سیستم و درجه بی‌نظمی وجود دارد که اصول بولتزمن نامیده می‌شود، که برای نمایش وضعیت ترمودینامیک یک سیستم مورد استفاده قرار می‌گیرد. آنتروپی در حقیقت بیانگر آن است که چگونه از بین عوامل موثر یک هدف، می‌توان مهمترین عوامل را تخمین زد یا به عبارتی متغیرهایی که بیشترین تاثیر را در رخداد یک واقعه دارند برای ما مشخص می‌نماید. در تحقیق حاضر، با استفاده از روش آنتروپی ابتدا نقشه‌های موجود که در مراحل قبل تهیه شدند توسط یکسری فرمول مورد بررسی قرار گرفت. به این منظور هر نقشه به طور جداگانه به طبقات مختلف تقسیم شدند و با استفاده از فرمول‌ها بررسی لازم روی آن‌ها انجام شد و سپس در مرحله آخر یک ضریب برای هر لایه تهیه گردید. به منظور تهیه نقشه پتانسیل آب زیرزمینی روابط زیر با استفاده از مدل آنتروپی شانون انجام گرفت (Jafari et al., ۲۰۱۳; Al-Abadi et al, ۲۰۱۶):

$$P_{ij} = \frac{a}{b} \quad \text{رابطه (۱)}$$

$$(P_{ij}) = \frac{P_{ij}}{\sum_{j=1}^{S_j} P_{ij}} \quad \text{رابطه (۲)}$$

$$H_j = -\sum_{i=1}^{S_j} (P_{ij}) \log_2 (P_{ij}) \cdot j = 1. \dots . n \quad \text{رابطه (۳)}$$

$$H_{jmax} = \log_2(S_j) \quad \text{رابطه (۴)}$$

$$(P_{ij}) = \frac{P_{ij}}{\sum_{j=1}^n P_{ij}} \cdot I(0.1), j = 1, \dots, n \quad \text{رابطه (۵)}$$

$$W_j = I_j P_{ij} \quad \text{رابطه (۱)}$$

که در آن a و b تعداد پیکسل و تعداد چاه، (P_{ij}) چگالی احتمال (احتمال وجود چاه در هر عامل و کلاس مربوطه) هستند، H_j و H_{jmax} به ترتیب مقدار انترپی و ماکزیمم انترپی را نشان می‌دهد، همچنین I_j ضریب اطلاعاتی و W_j مقدار وزن حاصل را برای کلیه عوامل نشان می‌دهد. مقدار W_j بین صفر و یک است که هر چه این ارزش به عدد یک نزدیک باشد حساسیت بیشتر را نشان می‌دهد. مقادیر مربوط به هر عامل در جدول (۱) در بخش نتایج آورده شده است. بعد از اعمال ضرایب مربوط به هر عامل در نرم‌افزار ArcGIS ۱۰٫۲ نقشه نهایی پتانسل آب زیرزمینی استخراج گردید.

• مدل جنگل تصادفی (RF)

در این روش یک سوم داده‌های اصلی در ایجاد هر درخت استفاده نمی‌شوند که تحت عنوان نمونه‌های بیرون از سبد^۷ (OOB) نامیده می‌شوند. چون این نمونه‌ها در آموزش درخت استفاده نشده اند در تست مورد استفاده قرار می‌گیرند و بنابراین نیازی به اعتبار سنجی مستقل در این روش نمی‌باشد. تکنیک RF همچنین اهمیت متغیرها در مدل سازی را نیز تعیین می‌کند. RF به دو روش میانگین کاهش دقت^۸ و میانگین کاهش ضریب جینی^۹ اهمیت متغیرها را نشان می‌دهد. در روش میانگین کاهش دقت که در بیش‌تر مطالعات استفاده شده، مقادیر درست متغیرها با مقادیر که به طور تصادفی برای هر درخت تولید شده است جایگزین می‌شود و اثر این تغییر را روی طبقه بندی اندازه گیری می‌کند، اگر این جایگزینی اثری روی خطای اندازه گیری نداشته باشد اهمیت آن متغیر کم است و اگر مقدار خطای آن اندازه گیری افزایش یابد آن متغیر مهم می‌باشد (Breiman and Cutler, ۲۰۰۴).

هدف از ماشین یادگیری پیدا کردن یک تقریب مفید از تابعی است که اساس آن پیش بینی رابطه بین متغیرهای کمکی ورودی و نتایج مورد نظر است (Hastie et al, ۲۰۰۱). در این مطالعه متغیرهای ورودی کمکی شامل زمین شناسی، کاربری اراضی، پارامترهای کیفی آب زیرزمینی، شاخص‌های پوشش گیاهی، شاخص‌های شوری خاک، مشتقات حاصل از مدل رقومی ارتفاعی، فاصله از معادن، فاصله از جاده، فاصله از رودخانه، فاصله از نقاط مسکونی، بارندگی و جمعیت می‌باشند و به عنوان متغیر ورودی برای پیش‌بینی شوری، کیفیت، SAR، PH و فلزات سنگین خاک استفاده شدند. اطلاعات خاک و متغیرهای محیطی به فرمت csv تبدیل و برای انجام فرآیند مدل‌سازی در نرم افزار KNIME Analytics Platform و R آماده شدند.

^۷. Out of Bag

^۸. Mean Decrease in Accuracy

^۹. Mean Decrease in Gini Coefficient

نقشه برداری رقومی توسعه یک مدل عددی یا آماری از رابطه بین متغیرهای محیطی و خصوصیات خاک است. برای ایجاد این ارتباط می‌توان از انواع مدل‌ها استفاده نمود. جنگل تصادفی (Random Forest)، یک روش یادگیری ترکیبی برای دسته بندی رگرسیون می‌باشد که پیشرفت‌های قابل ملاحظه‌ای را در تکنولوژی داده‌کاوی و علوم خاک ارائه داده است. در الگوریتم RF برای تشکیل هر درخت، دسته متفاوتی از الگوهای موجود با در نظر گرفتن جایگزینی دوباره‌ی هر الگوی انتخاب شده، انتخاب می‌شوند. اندازه این دسته انتخاب شده، برابر تعداد کل الگوهای موجود خواهد بود. RF در سال ۲۰۰۱ توسط برایمن به عنوان یک روشی از توسعه جدید درخت‌های تصمیم‌گیری، ارائه شد که پیش‌بینی چندین الگوریتم منفرد را با هم و با استفاده از قوانین مبتنی ترکیب می‌کند. اصول کلی تکنیک‌های آموزش گروهی بر پایه‌ی این فرض است که دقت آنها از دیگر الگوریتم‌های آموزشی بالاتر است. چون ترکیبی از چند مدل پیش‌بینی، دقیقتر از یک مدل می‌باشد و گروه‌ها در حالی که نقاط ضعف طبقه‌ها را کاهش می‌دهد، همزمان قدرت مجموعه‌های منفرد و منحصر بفرد طبقه‌ها را نیز بیشتر می‌کنند (Kotsiantis and Pintelas, ۲۰۰۴).

رویکرد جنگل تصادفی مبتنی بر روش‌های جدید ترکیب اطلاعات است که در آن تعداد زیادی درخت تصمیم ایجاد گردیده، سپس تمام درختان با هم برای پیش‌بینی ترکیب می‌گردند (Cutler et al., ۲۰۰۷). زمانی که متغیرهای پیشگو و هدف مشخص شدند، جنگل‌های تصادفی با رویاندن یک درخت تصمیم شبیه CART شروع می‌شود. این درخت از چندین لحاظ با درخت CART استاندارد متفاوت است. پارامترهای کلیدی برای مدل جنگل تصادفی، تعداد درختان و تعداد متغیرهای پیشگو می‌باشند. براساس این دو پارامتر، درخت تصمیم به بزرگترین اندازه ممکن رویانده و بدون هرس رها می‌شود (۲۰۱۲; Rodriguez-Galiano et al., ۲۰۰۶; Allouche et al., ۲۰۰۶). برای اجرای مدل جنگل تصادفی از نرم‌افزار R ۳.۱ و بسته randomforest (Breiman, ۲۰۰۱) استفاده شد.

• واسنجی^۱ و اعتبار سنجی^{۱۱} مدل

اعتبارسنجی ضربدری (Cross Validation) یکی از مهمترین تکنیک‌هایی است که به طور گسترده توسط دانشمندان علم داده مورد استفاده قرار می‌گیرد. مسائله‌ای که در هنگام کار با مدل‌های یادگیری ماشین (machine learning model) وجود دارد این است که کاربر نمی‌تواند کارایی مدل خود را تا زمانی که کارایی آن را بر روی یک مجموعه داده مستقل تست نکرده است بدست آورد (مجموعه داده‌ای که برای آموزش training مدل یادگیری ماشین مورد استفاده قرار نگرفته باشد). اعتبار سنجی ضربدری برای کمک به تخمین کارایی مدل یادگیری ماشین می‌باشد. اعتبارسنجی ضربدری k-fold یکی از عمومی‌ترین انواع اعتبار سنجی ضربدری می‌باشد که به طور گسترده‌ای در یادگیری ماشین مورد استفاده می‌باشد. اعتبار سنجی ضربدری k-fold با استفاده از گام‌های زیر انجام می‌شود.

- مجموعه داده اصلی را به k زیر مجموعه مساوی تقسیم و هر زیر مجموعه یک fold نامیده می‌شود (f_1, f_2, \dots, f_k). فولد f_i را به عنوان مجموعه اعتبارسنجی در نظر گرفته می‌شود و همه $k-1$ فولد باقی مانده را در مجموعه آموزشی اعتبار سنجی ضربدری Cross validating training set نگاه داشته می‌شوند.

^۱. Calibration

^{۱۱}. Validation

۲- مدل یادگیری ماشین خود با استفاده از مجموعه آموزشی اعتبارسنجی ضربدری آموزش داده می‌شود و سپس دقت و کارایی مدل با استفاده از اعتبارسنجی نتایج پیش‌بینی در مقابل مجموعه اعتبارسنجی Validation sert محاسبه می‌شود.

۳- دقت و کارایی مدل یادگیری ماشین با استفاده از بدست آوردن میانگین همه دقت‌ها و صحت‌های بدست آمده از همه موارد k از اعتبارسنجی ضربدری تخمین زده می‌شود. در روش اعتبارسنجی ضربدری k-fold، همه ورودی‌ها در مجموعه داده آموزشی برای آموزش و همچنین اعتبارسنجی validating به کار می‌رود، هر ورودی تنها یک بار برای اعتبارسنجی مورد استفاده قرار می‌گیرد.

نکته اینکه معمولا برای k عدد ۱۰ در نظر گرفته می‌شود، اما این یک قانون نیست و k می‌تواند هر عددی در نظر گرفته شود.

• ارزیابی مدل‌های تهیه نقشه آب زیرزمینی

جهت ارزیابی کارایی مدل جنگل تصادفی و آنتروپی شانون برای پتانسیل‌یابی منابع آب زیرزمینی، از شاخص‌های آماری MAE، RMSE و R^2 استفاده شد. روابط ۱ تا ۳ آنها را نشان می‌دهد.

$$MAE = \frac{\sum_{i=1}^n |X_i - Y_i|^2}{n} \quad \text{رابطه ۱}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad \text{رابطه ۲}$$

$$R^2 = \frac{\sum_{i=1}^n (X_i^* - Y_i^*)^2}{\sum_{i=1}^n (X_i - Y_i)^2} \quad \text{رابطه ۳}$$

که X_i ، Y_i ، X_i^* و Y_i^* به ترتیب مربوط به مقادیر اندازه‌گیری شده و برآورد شده، میانگین مقادیر اندازه‌گیری شده و برآورد شده می‌باشد.

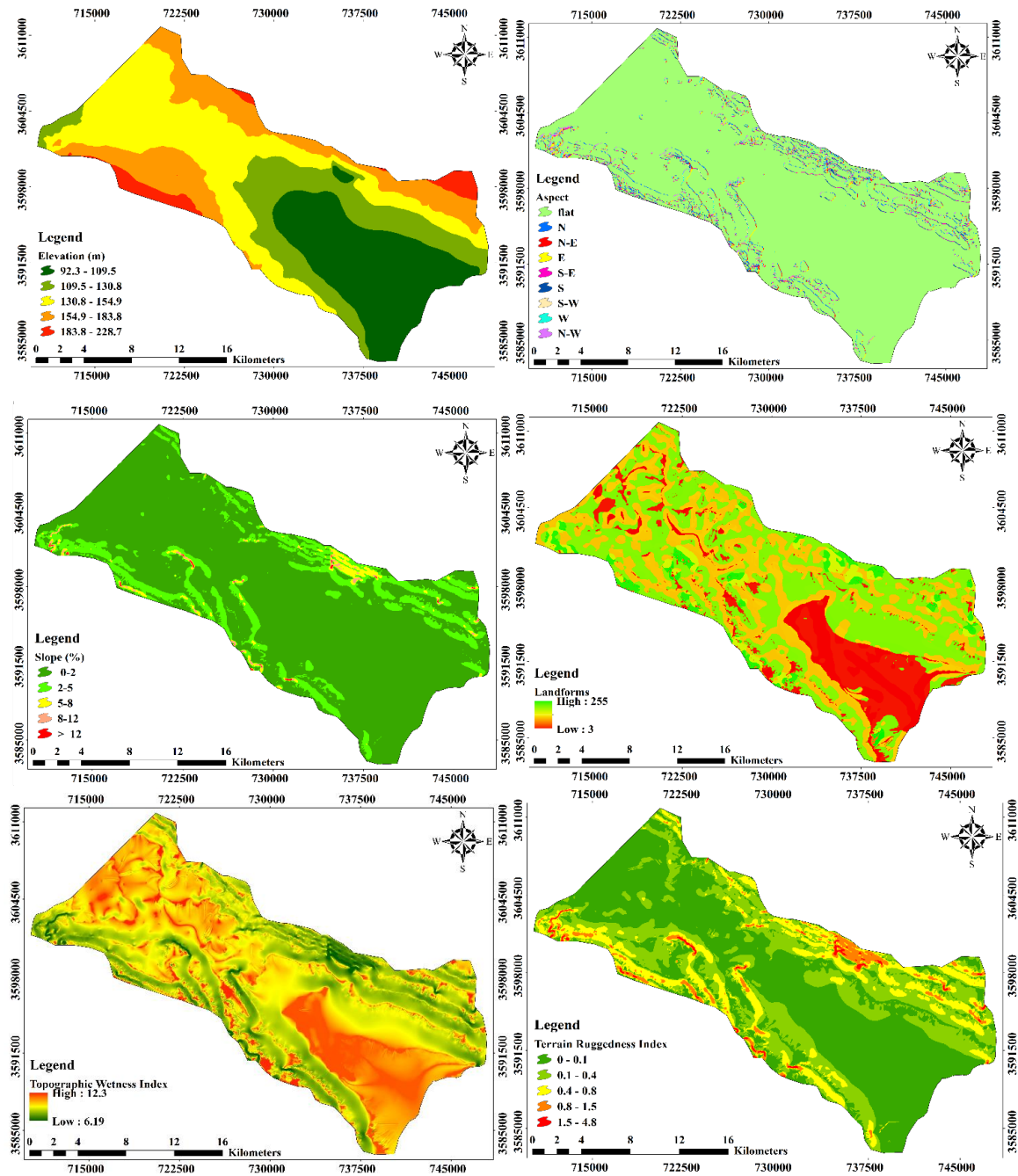
شرح و تفسیر نتایج

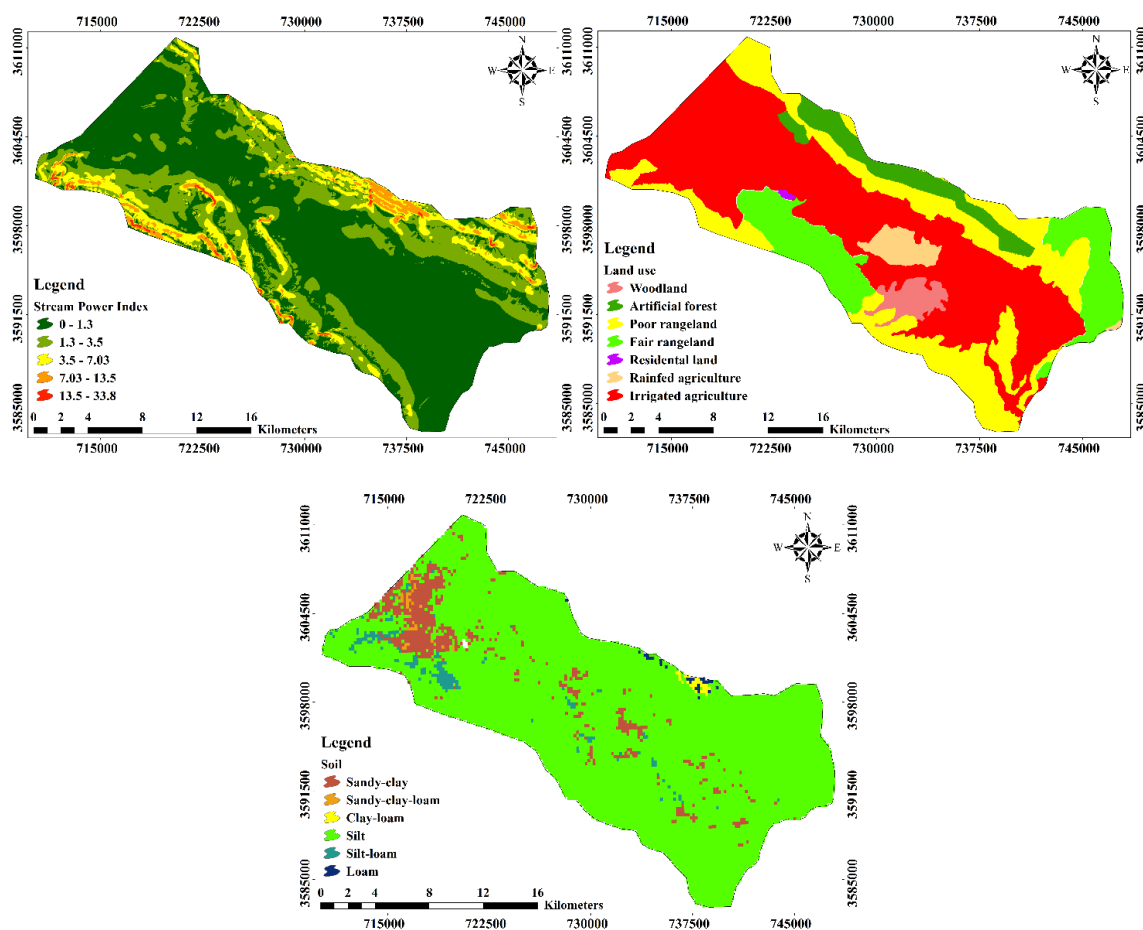
مدل‌سازی منابع آب زیرزمینی با استفاده از الگوریتم‌های یادگیری ماشین از جدیدترین روش‌های حال حاضر بحساب می‌آید. در این بخش، ابتدا خروجی لایه‌های مورد استفاده در مدل‌ها نمایش داده می‌شود. سپس نتایج مربوط به مدل‌سازی منابع آب زیرزمینی با استفاده از روش‌های آنتروپی شانون و جنگل تصادفی در منطقه مورد مطالعه ارائه می‌گردد. همچنین، ارزیابی دقت مدل‌ها به منظور انتخاب مدل برتر نشان داده می‌شود.

• نقشه‌های رستری

تعداد نه لایه رستری شامل ارتفاع از سطح دریا، جهت شیب، درصد شیب، شکل زمین، شاخص رطوبت توپوگرافی، شاخص ناهمواری زمین، شاخص قدرت جریان، کاربری اراضی و خاکشناسی (شکل ۲) برای ورودی مدل‌ها تهیه گردید. نقشه‌ها

نشان داد که بیشتر منطقه مورد مطالعه، درصد شیب پایین (۰-۲ درصد) و ارتفاع پایین هستند و همچنین از نظر جهت جغرافیایی به منطقه هموار طبقه‌بندی می‌شوند (شکل ۲).





شکل ۲. نقشه لایه‌های مورد استفاده در این تحقیق

مدل‌سازی منابع آب زیرزمینی در روش آنتروپی ابتدا نقشه‌های موجود که در مراحل قبل تهیه شدند توسط یکسری فرمول مورد بررسی قرار می‌گیرند به اینصورت که هر نقشه به طور جداگانه به طبقات مختلف تقسیم شده و فرمول‌ها روی آن‌ها بررسی انجام می‌دهند و در آخر یک ضریب تهیه می‌گردد (Naghibi et al, ۲۰۱۵, Rahmati et al, ۲۰۱۶). Pourghasemi et al, ۲۰۱۲. در محیط GIS هر یک از ضریب‌های وزن نهایی در نقشه‌ی مربوط به خود ضرب گردید و در آخر همه‌ی نقشه‌ها با هم جمع گردید و نقشه‌ی که بدست آمد همان نقشه پتانسیل آب زیرزمینی می‌باشد که بر اساس پتانسیل کلاسه بندی گردید (Al-Abadi et al, ۲۰۱۱; Al-Abadi, Malekian and Azarinvand, ۲۰۱۵). نتایج حاصل از ارتباط بین پتانسیل آب زیرزمینی، عوامل مؤثر و حد اثر هر یک از آنها با استفاده از مدل آنتروپی (۲۰۱۵). نتایج حاصل از ارتباط بین پتانسیل آب زیرزمینی، عوامل مؤثر و حد اثر هر یک از آنها در جدول (۱) ارائه شده است. براساس نتایج حاصل، طبقه ارتفاعی متوسط (۱۶۰-۱۳۰ متر) بیشترین تعداد چاه (۱۱۴ چاه) را دارد و در طبقه پایین تر تعداد چاه‌ها کاهش می‌یابد. همچنین بیشترین تعداد چشمه‌ها در کلاس جهت شیب جنوب است، که نشان دهنده‌ی زیاد بودن پتانسیل آب زیرزمینی در جهت مذکور می‌باشد

و کمترین آن در مناطق با جهت جنوب غرب است. بررسی عوامل دیگر نشان داد که بیشترین تعداد وقوع چشمه به ترتیب مربوط به کلاس‌های درجه شیب ۰ تا ۲ درجه، طبقه‌ی ۱۰/۱ تا ۱۱ از عامل شاخص رطوبت توپوگرافی، طبقه‌ی ۰ تا ۱ از عامل شاخص ناهمواری زمین، خاکشناسی سیلتی، کاربری اراضی زراعت آبی و باغات، زمین شناسی ذخایر تراسی، طبقه‌ی ۸ تا ۱۸۶ متر و تجمع جریان است.

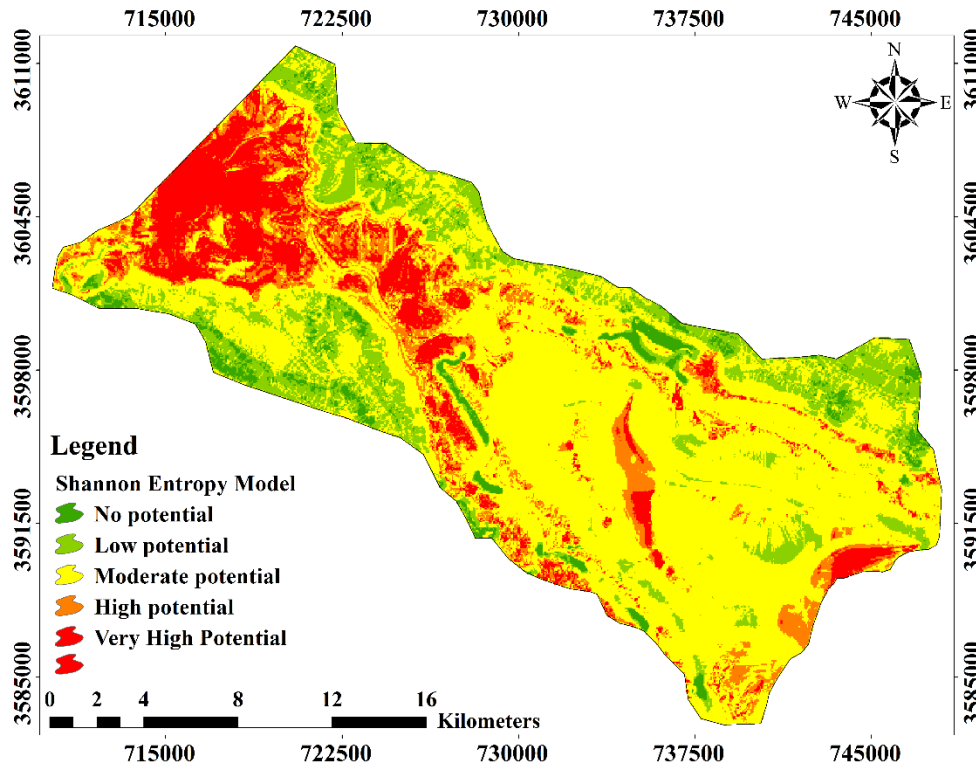
همچنین براساس نتایج ارائه شده در جدول وزن نهایی هر عامل (W_j) مشخص شد که عوامل خاکشناسی، شکل زمین، شیب، تعادل سلولی، تجمع جریان، کاربری، زمین شناسی، شاخص ناهمواری زمین، شاخص قدرت جریان، گرادبان، شاخص رطوبت توپوگرافی، جهت شیب، مدل رقومی ارتفاع به ترتیب بیشترین وزن (۰/۹۳۱، ۰/۵۷۷، ۰/۵۲۳، ۰/۲۱۲، ۰/۱۸۸، ۰/۱۵۸، ۰/۱۵۱، ۰/۱۲۷، ۰/۱۱۸، ۰/۰۹۷، ۰/۰۵۵، ۰/۰۴۴، ۰/۰۳۹) را به خود اختصاص داده اند.

جدول ۱- نتایج حاصل از ارتباط بین پتانسیل آب زیرزمینی، عوامل مؤثر و حد اثر هر یک از آنها با استفاده از مدل آنتروپی شانون

عامل	طبقه	a (تعداد پیکسل)	a (%)	b (تعداد جاه)	b (%)	Pij	(Pij)	Hj	Hjmax	Ij	Wij
مدل رقومی ارتفاع	<۱۰۰	۷۳۳۴۴	۰/۱۵	۲۲	۰/۰۹	۰/۶	۰/۱۳	۲/۲۲	۲/۳۲۲	۰/۰۴۴	۰/۰۳۹
	۱۳۰-۱۰۰	۱۲۴۸۵۴	۰/۲۶	۶۱	۰/۲۶	۱	۰/۲۲				
	۱۶۰-۱۳۰	۱۴۹۰۲۷	۰/۳۱	۱۱۴	۰/۴	۱/۲۹	۰/۲۹				
	۱۹۰-۱۶۰	۸۶۵۴۵	۰/۱۹	۳۱	۰/۳۱	۱/۱	۰/۲۴				
	>۱۹۰	۴۳۲۷۲	۰/۰۹	۸	۰/۰۴	۰/۴۴	۰/۰۹				
شیب	۲-۰	۴۵۶۳۸۲	۰/۹۵	۲۰۹	۰/۸۸	۰/۹۲	۰/۱۳	۱/۴۱۷	۲/۳۲۲	۰/۳۸۹	۰/۵۲۳
	۵-۲	۱۷۲۰۸	۰/۰۴	۲۴	۰/۱	۲/۵	۰/۳۸				
	۸-۵	۲۲۵۷	۰/۰۰۶	۳	۰/۰۲	۳/۳	۰/۴۹				
	۱۲-۸	۱۱۰۶	۰/۰۰۳	۰	۰	۰	۰				
	>۱۲	۸۹	۰/۰۰۱	۰	۰	۰	۰				
جهت شیب	شمال	۴۶۱۹۰	۰/۰۹	۳۶	۰/۱۵	۱/۶۶	۰/۱۸	۳/۳	۳/۱۷	۰/۰۴	۰/۰۴۴
	شمال غرب	۶۰۵۹۸	۰/۱۳	۱۱	۰/۰۶	۰/۴۶	۰/۰۶				
	غرب	۴۴۴۷۸	۰/۰۹	۱۴	۰/۰۷	۰/۷	۰/۰۹				
	جنوب غرب	۴۳۸۶۲	۰/۰۹	۱۰	۰/۰۵	۰/۵	۰/۰۶				
	جنوب شرق	۷۹۱۶۰	۰/۱۶	۴۷	۰/۲	۱/۲۵	۰/۱۴				
	جنوب	۸۶۵۷۰	۰/۱۸	۶۰	۰/۲۶	۱/۴	۰/۱۷				
	شرق	۳۵۸۹۷	۰/۰۸	۱۷	۰/۱	۱/۲۵	۰/۱۴				
شمال شرق	۳۱۵۴۶	۰/۰۸	۱۸	۰/۱۳	۱/۶۲۵	۰/۱۹					
شاخص رطوبت توپوگرافی	۶/۸-۲/۵	۴۰۳۰۵	۰/۰۸	۷	۰/۰۳	۰/۳۷	۰/۰۸	۲/۱۸۱	۲/۳۲۲	۰/۰۶۱	۰/۰۵۵
	۸/۹-۵/۳	۱۱۹۱۹۳	۰/۲۵	۶۰	۰/۲۵	۱	۰/۲۲				
	۹/۱۰-۳/۱	۱۲۵۵۷۰	۰/۲۶	۷۰	۰/۳	۱/۱۵	۰/۲۵				
	۱۰/۱۱-۱	۱۰۴۹۷۴	۰/۲۲	۷۶	۰/۳۲	۱/۴۵	۰/۳۲				
	۱۲-۱۱/۳	۸۷۰۰۰	۰/۱۹	۲۳	۰/۱	۰/۵۲	۰/۱۳				
شاخص ناهمواری زمین	۰ - ۱	۲۴۹۹۱۴	۰/۵۲	۱۳۲	۰/۵۶	۱/۰۷	۰/۳۴	۱/۸۵۶	۲/۳۲۲	۰/۲	۰/۱۲۷
	۲-۱	۱۴۴۶۲۲	۰/۳	۸۲	۰/۳۵	۱/۱۶	۰/۳۶				

				۰/۲	۰/۶۱	۰/۰۸	۱۹	۰/۱۳	۶۳۴۲۷	۳-۲	
				۰/۱	۰/۳۳	۰/۰۱	۳	۰/۰۳	۱۵۵۸۸	۴-۳	
				۰	۰	۰	۰	۰/۰۲	۳۴۹۱	۴-۴/۸	
۰/۹۳۱	۰/۳۶۱	۲/۵۸۵	۱/۶۵۲	۰/۰۵	۰/۸۱	۰/۷۷	۱۸۱	۰/۹۵	۴۵۶۰۱۱	سیلت	خاکشناسی
				۰/۴۳	۶/۶۶	۰/۱۶	۳۷	۰/۰۲۴	۱۱۵۸۲	شنی رسی	
				۰/۳۹	۶	۰/۰۶	۱۵	۰/۰۱	۴۸۳۴	سیلنتی لومی	
				۰/۱۳	۲	۰/۰۱	۳	۰/۰۰۵	۲۶۸۰	شنی رسی لومی	
				۰	۰	۰	۰	۰/۰۰۲	۱۲۵۹	رسی لومی	
				۰	۰	۰	۰	۰/۰۰۹	۶۷۶	لومی	
۰/۱۱۸	۰/۱۷	۲/۳۲۲	۱/۹۳	۰/۳	۱/۰۷	۰/۶۲	۱۴۷	۰/۵۸	۲۷۷۰۸۲	۷-۰	شاخص قدرت جریان
				۰/۳	۱/۱	۰/۳۲	۷۶	۰/۲۹	۱۳۷۶۷۵	۱۴-۷	
				۰/۱۳	۰/۴۲	۰/۰۴۲	۱۰	۰/۱	۴۶۴۵۶	۲۱-۱۴	
				۰/۲۷	۰/۹	۰/۰۱۸	۳	۰/۰۲	۱۳۸۴۵	۲۸-۲۱	
				۰	۰	۰	۰	۰/۰۱	۱۹۸۴	۳۳-۲۸	
۰/۵۷۷	۰/۵۷۱	۱/۵۸۵	۰/۶۸	۰/۱۸	۰/۵۳	۰/۱	۲۵	۰/۱۹	۸۸۹۹۱	۳-۰	شکل زمین
				۰/۸۲	۲/۵	۰/۹	۲۱۱	۰/۳۶	۱۷۲۵۷۶	۱۲۹-۳	
				۰	۰	۰	۰	۰/۴۵	۲۱۵۴۷۵	۲۵۵-۱۲۹	
۰/۱۵۱	۰/۱۸۹	۱	۰/۸۱۱	۰/۲۵	۰/۴	۰/۲	۳۲	۰/۱۴۶۵	۶۲۸۱	بختیاری	زمین شناسی
				۰/۱۵	۱/۲	۰/۳	۱۶	۰/۲۲۲۷	۹۵۴۷	آغاچاری	
				۰/۱۶	۲/۸	۰/۷	۱۸۸	۰/۶۳۰۸	۲۷۰۴۷	ذخایر تراسی	
۰/۱۸۸	۰/۵۹۴	۲/۳۲۲	۰/۹۴۲	۰/۶۴	۱/۰۱۶	۰/۹۹۶	۲۳۵	۰/۹۸	۴۷۱۱۷۱	۱۸۶-۸	تجمع جریان
				۰/۳۶	۰/۵۷۱	۰/۰۰۴	۱	۰/۰۰۷	۳۷۲۴	۳۶۴-۱۸۶	
				۰	۰	۰	۰	۰/۰۰۷	۹۶۹	۵۵۱-۲۶۴	
				۰	۰	۰	۰	۰/۰۰۵	۷۰۴	۷۲۹-۵۵۱	
				۰	۰	۰	۰	۰/۰۰۱	۴۷۴	۹۰۰-۷۲۹	
۰/۲۱۲	۰/۵۷۲	۲/۳۲۲	۰/۹۹۳	۰/۵۵	۱/۰۲	۰/۹۵	۲۲۵	۰/۹۳	۴۴۵۸۲۰	<۱	تعادل سلولی
				۰/۴۵	۰/۸۳	۰/۰۵	۱۱	۰/۰۶	۲۹۳۲۰	۲-۱	
				۰	۰	۰	۰	۰/۰۰۲	۱۳۰۳	۳-۲	
				۰	۰	۰	۰	۰/۰۰۴	۳۰۶	۴-۳	
				۰	۰	۰	۰	۰/۰۰۴	۲۹۳	>۴	
۰/۱۵۸	۰/۳۳۹	۲/۸۰۷	۱/۸۵۶	۰	۰	۰	۰	۰/۰۲	۱۲۶۰۹	بیشه زار و درختچه زار	کاربری اراضی
				۰/۲۰	۰/۶۷	۰/۰۴۲	۱۰	۰/۰۶	۲۹۷۰۰	زراعت آبی و باغات	
				۰/۴۷	۱/۵۴	۰/۷۴۱	۱۷۵	۰/۴۸	۲۳۰۰۰۰	جنگل دست کاشت	
				۰/۰۴	۰/۱۴	۰/۰۰۴	۱	۰/۰۲	۱۲۸۸۱	زراعت دیم	
				۰/۲۳	۰/۷۲	۰/۱۹۵	۴۶	۰/۲۵	۱۲۱۹۷۴	مراتع کم متراکم	
				۰/۰۳	۰/۱۲	۰/۰۱۸	۴	۰/۱۴	۶۸۹۹۸	مراتع نیمه متراکم	
				۰	۰	۰	۰	۰/۰۰۴	۸۸۰	مناطق مسکونی	

پس از تعیین وزن پارامترها با روش شاخص آنتروپی و طبقات آن ها، در محیط نرم افزار ArcGIS ۱۰٫۲ از ترکیب وزن پارامترها و طبقات آن ها نقشه نهایی پتانسیل منابع آب زیرزمینی حاصل گردید. قسمت اعظمی از منطقه مورد مطالعه دارای پتانسیل متوسط می باشد و مساحت خیلی اندکی بدون پتانسیل می باشد. همچنین مساحت هریک از محدوده های دارای پتانسیل مختلف منطقه مورد مطالعه بر اساس مدل آنتروپی در جدول (۲) آمده شده است.



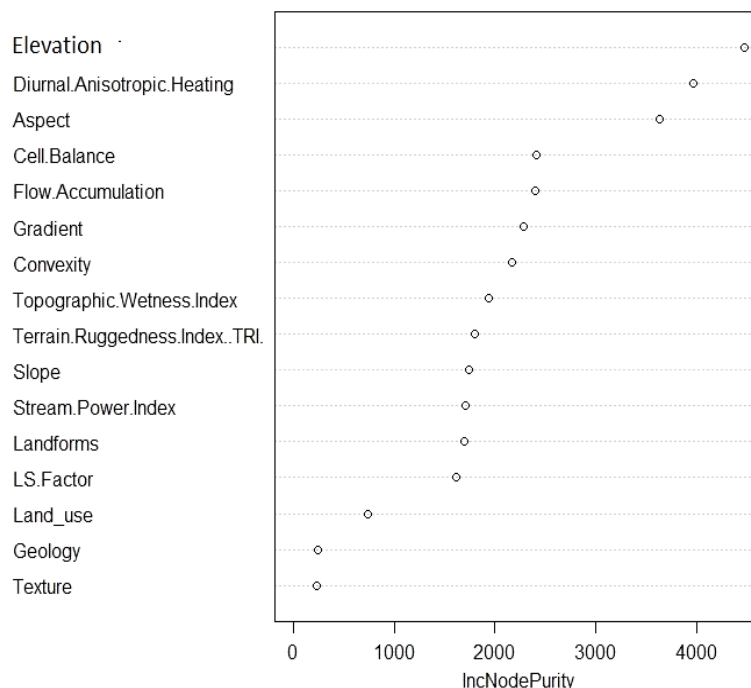
شکل ۳- نقشه پتانسیل یابی منابع آب زیرزمینی توسط مدل آنتروپی شانون

جدول ۲- مساحت هریک از محدوده های دارای پتانسیل مختلف منطقه مورد مطالعه (مدل آنتروپی شانون)

Potential	Area (ha)
No potential	۱۵۰۲٫۱۲
Low potential	۶۳۸۰٫۲۰
Moderate potential	۲۴۶۳۳٫۰۵
High potential	۵۶۶۷٫۰۴
Very High Potential	۴۶۹۲٫۶۰

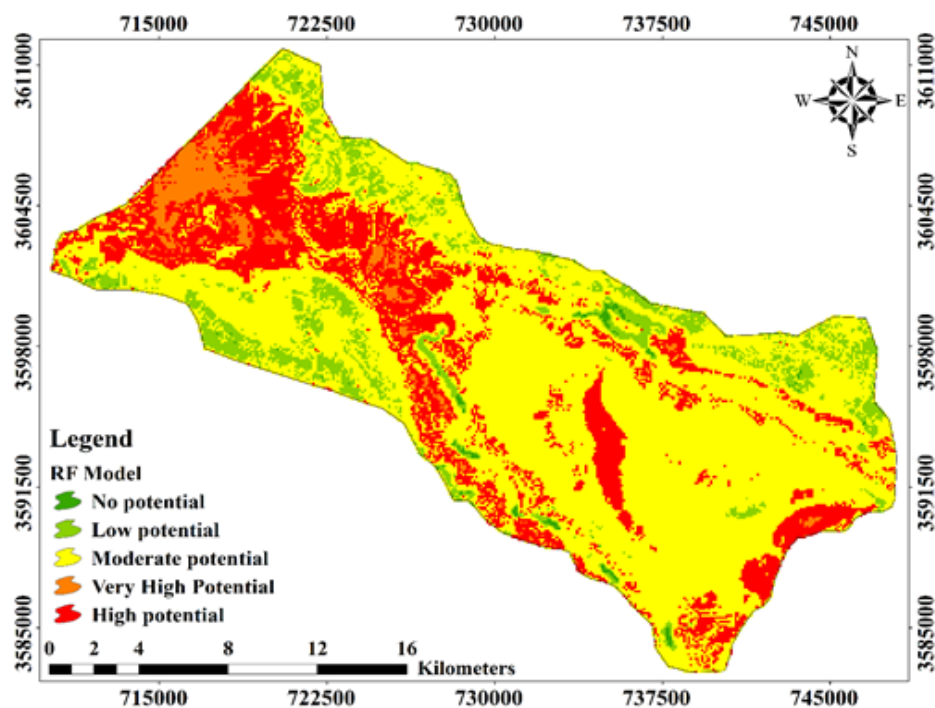
مدل جنگل تصادفی روش دیگری بود که به منظور پتانسیل یابی آب زیرزمینی از داده های آب زیرزمینی به عنوان متغیر هدف و داده های کمکی به عنوان داده پیش بینی کننده استفاده شد. در این تحقیق ۱۶ عامل مؤثر شامل درجه شیب، جهت شیب، ارتفاع از سطح دریا، فاصله از گسل، شکل شیب، فاصله از آبراهه، فاصله از جاده، بارندگی، لیتولوژی و کاربری اراضی

مورد استفاده قرار گرفتند که در پژوهش‌های نقیبی و همکاران (Naghibi et al, ۲۰۱۷) و رحمتی و همکاران (Rahmati et al, ۲۰۱۶) به آن توجه شده است. آنالیز حساسیت مدل جنگل تصادفی در شکل (۴) آورده شده است. بر اساس نتایج این مدل عوامل ارتفاع و کاربری اراضی بیشترین تاثیر را بر پتانسیل آب زیرزمینی در منطقه مطالعاتی دارند.



شکل ۴- نمودار آنالیز اهمیت عامل مؤثر در مدل جنگل تصادفی به منظور تعیین پتانسیل آب زیرزمینی

پس از تعیین عوامل مؤثر در پتانسیل‌یابی آب زیرزمینی، نقشه پتانسیل آب زیرزمینی منطقه‌ی تحقیق بر اساس مدل جنگل تصادفی تهیه و در شکل (۵) ارائه شده است. همچنین مساحت هریک از محدوده‌های دارای پتانسیل مختلف منطقه مورد مطالعه بر اساس مدل جنگل تصادفی در جدول (۳) آمده شده است. بر این اساس بخش عظیمی از منطقه دارای پتانسیل متوسط و مساحت اندکی بدون پتانسیل بوده است.



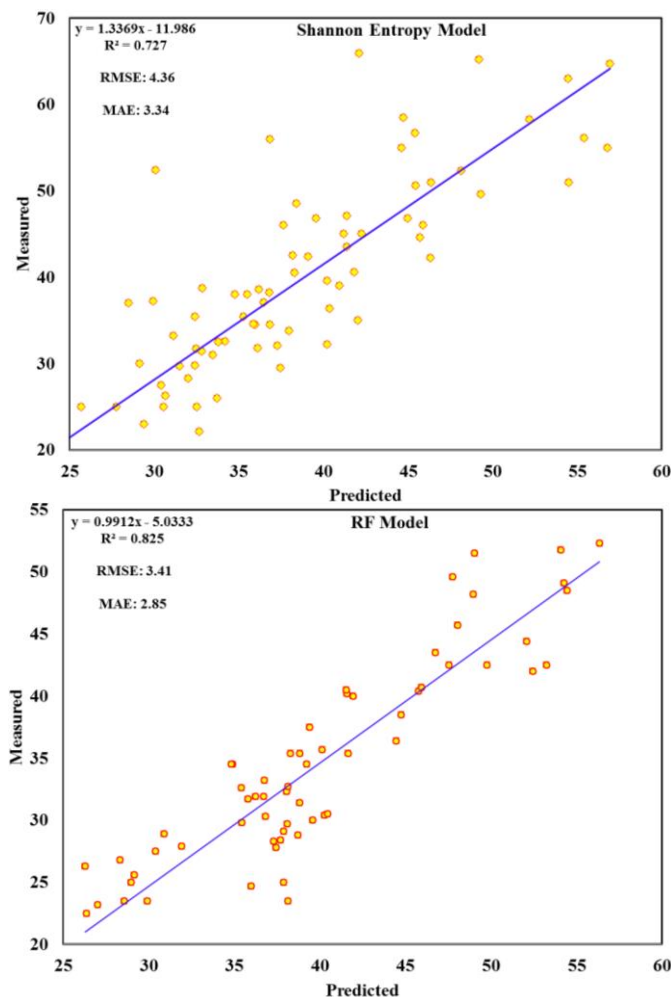
شکل ۵- نقشه پتانسیل یابی منابع آب زیرزمینی توسط مدل جنگل تصادفی

جدول ۳- مساحت هریک از محدوده‌های دارای پتانسیل مختلف منطقه مورد مطالعه (مدل جنگل تصادفی)

Potential	Area (ha)
No potential	۲۰۵.۶۱
Low potential	۳۸۵۷.۴۳
Moderate potential	۲۶۹۵۴.۲۲
High potential	۱۰۱۰۸.۴
Very High Potential	۱۷۳۰.۸۹

• ارزیابی مدل‌های تهیه نقشه آب زیرزمینی

نتایج ارزیابی مدل‌های استفاده شده با استفاده از نمودار همبستگی بین مقدار پیش‌بینی شده و اندازه‌گیری شده را نشان می‌دهد شکل (۶). نتایج نشان داد که مدل جنگل تصادفی با دقت (MAE: ۲,۸۵, RMSE: ۳,۴۱, $R^2 = ۰,۸۲۵$) دارای دقت بهتری نسبت به مدل آنتروپی شانون با دقت (MAE: ۳,۳۴, RMSE: ۴,۳۶, $R^2 = ۰,۷۲۷$) می‌باشد.



شکل ۶- نتایج ارزیابی مدل‌های جنگل تصادفی و آنترپی شانون در تهیه نقشه‌ی پتانسیل آب زیرزمینی

ارزیابی و اولویت بندی عوامل موثر روی اکوسیستم‌های طبیعی برای مدیریت و استفاده پایدار از منابع آب و زمین در مناطق خشک و نیمه‌خشک که بهره‌برداری از آن‌ها به سرعت در حال تغییر است، ضروری است (Al-Abadi et al., ۲۰۱۹; Yaghoobi et al., ۲۰۱۶). نتایج نشان داد که مدل جنگل تصادفی (RF) صحت بالایی جهت پتانسیل یابی آب زیرزمینی در منطقه مورد مطالعه دارد. مدل جنگل تصادفی که از مدل‌های برتر یادگیری ماشین است در مطالعات مختلف جز مدل‌های با صحت بالا برای مدل‌سازی مطالعات محیطی به حساب می‌آید (Al-Abadi et al., ۲۰۱۶; Breiman, ۲۰۰۱, Breiman et al. ۲۰۰۴; Cutler et a., ۲۰۰۷; Thanh et al., ۲۰۲۳; Madani and Niyazi, ۲۰۲۳). همچنین در مطالعه حاضر، دقت مدل جنگل تصادفی توسط نمودار همبستگی بین داده‌های پیش‌بینی شده و اندازه‌گیری شده تعیین گردید، که مقدار آن معادل ($R^2=0.82$) می‌باشد. بر اساس نقشه پهنه‌بندی مدل جنگل تصادفی مشخص شد که بیشترین مساحت حساسیت در کلاس پتانسیل متوسط می‌باشد (۶۲/۹۳٪) و کلاس بدون پتانسیل کمترین مساحت را در بر گرفته است (۰/۴۸٪). این نتایج نشان دهنده پتانسیل نسبتاً بالای سفره‌های آب زیرزمینی در منطقه مورد مطالعه می‌باشد. این در حالیست که برخی مناطق پتانسیل خیلی پایین و پایینی برای سفره‌های آب زیرزمینی دارند (Al-

(Abadi et al., ۲۰۱۶). در این تحقیق از بین تمامی عوامل مؤثر که شامل درجه شیب، جهت شیب، ارتفاع از سطح دریا، فاصله از گسل، شکل شیب، فاصله از آبراهه، فاصله از جاده، بارندگی، لیتولوژی و کاربری اراضی می‌باشد، عوامل ارتفاع و کاربری اراضی بر اساس مدل جنگل تصادفی بیشترین تاثیر را بر پتانسیل آب زیرزمینی در منطقه مطالعاتی دارند. در پژوهش‌های قبلی مانند نقیبی و همکاران (Naghbi et al., ۲۰۱۷) و رحمتی و همکاران (Rahmati et al., ۲۰۱۶) بیشتر این عوامل برای در آنالیز حساسیت مدل جنگل تصادفی بکار برده شده است. در مطالعه حاضر، بر اساس نتایج مدل جنگل تصادفی عوامل ارتفاع از سطح دریا، کاربری اراضی و جهت شیب بیشترین تاثیر را بر پتانسیل آب زیرزمینی در منطقه مورد مطالعه دارند. در تحقیقی که توسط رضوی ترمه و همکاران (Razavi-Termeh et al., ۲۰۱۹) انجام گرفت، نتایج مدل جنگل تصادفی نشان داد که پارامترهای فاصله از رودخانه و جهت شیب مهمترین عوامل مؤثر روی پتانسیل آب زیرزمینی بودند.

در تحقیق حاضر، مدل آنتروپی شانون نیز برای نقشه نهایی پتانسیل منابع آب زیرزمینی بر اساس وزن پارامترها مختلف استفاده گردید. نتایج این مدل نشان داد که حدود ۵۷/۴۵ درصد از منطقه مورد مطالعه دارای پتانسیل متوسط می‌باشد و مساحت بدون پتانسیل حدود ۳/۵ درصد می‌باشد. اگرچه، در مطالعه حاضر مدل آنتروپی شانون صحت کمتری نسبت به مدل جنگل تصادفی نشان داد اما در برخی مطالعات دیگر ترکیب مدل آنتروپی شانون با مدل‌های دیگر مانند الگوریتم پشتیبانی تصمیم نیز دارای صحت قابل قبولی بود (Halder et al., ۲۰۲۳).

نتیجه‌گیری

در تحقیق حاضر، از الگوریتم‌های یادگیری ماشین شامل آنتروپی شانون و جنگل تصادفی برای مدل‌سازی منابع آب زیرزمینی در مراتع خشک غرب ایران استفاده شد. نتایج ارزیابی مدل‌ها نشان داد که روش جنگل تصادفی (RMSE: ۲.۸۵, MAE: ۳.۴۱, $R^2 = ۰.۸۲۵$) دارای دقت بالاتری نسبت به مدل آنتروپی شانون (RMSE: ۴.۳۶, $R^2 = ۰.۷۲۷$, MAE: ۳.۳۴) می‌باشد. نقشه تهیه شده با استفاده از مدل جنگل تصادفی نشان داد که بیشترین مساحت منطقه مورد مطالعه مربوط به پتانسیل منابع آب زیرزمینی در کلاس متوسط می‌باشد (۶۲/۹۳٪) و کلاس بدون پتانسیل کمترین مساحت را در بر گرفته است (۰/۴۸٪). همچنین، بر اساس نتایج این مدل عوامل ارتفاع و کاربری اراضی بیشترین تاثیر را بر پتانسیل آب زیرزمینی در منطقه مطالعاتی دارند. الگوریتم جنگل تصادفی با انعطاف‌پذیری فوق‌العاده برای تحلیل وقایع طبیعی، قادر است ابهامات ایجاد شده در ذهن انسان، محیط و عدم قطعیتی که همواره در قضاوت انسان وجود دارد مدل‌سازی و تحلیل نماید. لذا نتایج حاصل، توجه مدیران و مسئولان امر در اتخاذ تصمیمات آتی مدیریت منابع آب و همچنین بهره‌برداری از آنها در منطقه مورد مطالعه را می‌طلبد.

منابع

- Abd Manap, M., Nampak, H., Pradhan, B., Lee, S., Sulaiman, W.N.A. and Ramli, M.F. ۲۰۱۴. Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arabian Journal of Geosciences*, 7(۲): ۷۱۱-۷۲۴.
- Al-Abadi, A.M. ۲۰۱۷. Modeling of groundwater productivity in northeastern Wasit Governorate, Iraq using frequency ratio and Shannon's entropy models. *Applied Water Science*, 7(۲): ۶۹۹-۷۱۶.
- Al-Abadi, A.M., Al-Temmeme, A.A. and Al-Ghanimy, M.A. ۲۰۱۶. 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(۳): ۲۶۵-۲۸۳.
- Allouche, O., Tsoar, A. and Kadmon, R. ۲۰۰۶. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, 43(۶): ۱۲۲۳-۱۲۲۶.
- Breiman, L. and Cutler, A. ۲۰۰۴. Random Forests, URL: http://www.stat.berkeley.edu/users/breiman.RandomForests/cc_papers.htm.
- Breiman, L. ۲۰۰۱. Random forests. *Machine learning*, 45(۱): ۵-۳۲.
- Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. and Lawler, J.J. ۲۰۰۷. Random forests for classification in ecology. *Ecology*, 88(۱۱): ۲۷۸۳-۲۷۹۲.
- De Martonne, E. M. ۱۹۲۶. L'indice d'aridité. *Bulletin de l'Association de géographes français*, 3(۹): ۳-۵.
- Friedman, J., Hastie, T. and Tibshirani, R., ۲۰۱۰. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(۱): ۱.
- Ganapuram, S., Kumar, G.V., Krishna, I.M., Kahya, E. and Demirel, M.C. ۲۰۰۹. Mapping of groundwater potential zones in the Musi basin using remote sensing data and GIS. *Advances in Engineering Software*, 40(۷): ۵۰۶-۵۱۸.
- Halder, S., Roy, M. B., & Roy, P. K. ۲۰۲۰. Analysis of groundwater level trend and groundwater drought using Standard Groundwater Level Index: A case study of an eastern river basin of West Bengal, India. *SN Applied Sciences*, 2(۳): ۱-۲۴.
- Halder, S., Roy, M. B., Roy, P. K., & Sedighi, M. (۲۰۲۳). Groundwater vulnerability assessment for drinking water suitability using Fuzzy Shannon Entropy model in a semi-arid river basin. *Catena*, 229, ۱۰۷۲۰۶.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B. and Guevara, M.A. ۲۰۱۷. SoilGrids۲۵۰m: Global gridded soil information based on machine learning. *PLoS one*, 12(۲): e۰۱۶۹۷۴۸.
- Hou, E., Wang, J. and Chen, W. ۲۰۱۸. A comparative study on groundwater spring potential analysis based on statistical index, index of entropy and certainty factors models. *Geocarto international*, 33(۷): ۷۵۴-۷۶۹.
- Jaafari, A., Najafi, A., Pourghasemi, H. R., Rezaeian, J., & Sattarian, A. ۲۰۱۴. GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology*, ۱۱: ۹۰۹-۹۲۶.
- Jothibasu, A. and Anbazhagan, S. ۲۰۱۶. Modeling groundwater probability index in Ponnaiyar River basin of South India using analytic hierarchy process. *Modeling Earth Systems and Environment*, 2(۳): ۱۰۹.
- Khoshtinat, S., Aminnejad, B., Hassanzadeh, Y. and Ahmadi, H. ۲۰۱۹. Application of GIS-based models of weights of evidence, weighting factor, and statistical index in spatial modeling of groundwater. *Journal of Hydroinformatics*, 21(۵): ۷۴۵-۷۶۰.
- Madani, A., & Niyazi, B. (۲۰۲۳). Groundwater Potential Mapping Using Remote Sensing and Random Forest Machine Learning Model: A Case Study from Lower Part of Wadi Yalamlam, Western Saudi Arabia. *Sustainability*, 15(۳), ۲۷۷۲.
- Malekian, A. and Azarnivand, A. ۲۰۱۶. Application of integrated Shannon's entropy and VIKOR techniques in prioritization of flood risk in the Shemshak watershed, Iran. *Water Resources Management*, 30(۱): ۴۰۹-۴۲۵.

- Massey, D.S. and Denton, N.A., ۱۹۸۸. The dimensions of residential segregation. *Social forces*, **67**(۲):۲۸۱-۳۱۵.
- Moghaddam, D.D., Rezaei, M., Pourghasemi, H.R., Pourtaghie, Z.S. and Pradhan, B. ۲۰۱۵. Groundwater spring potential mapping using bivariate statistical model and GIS in the Taleghan watershed, Iran. *Arabian Journal of Geosciences*, **8**(۲): ۹۱۳-۹۲۹.
- Molden, D. (۲۰۱۳). *Water for food water for life: A comprehensive assessment of water management in agriculture*: Routledge.
- Naghibi, S. A., Ahmadi, K. and Daneshi, A. ۲۰۱۷. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, **31**(۹): ۲۷۶۱-۲۷۷۵.
- Naghibi, S.A., Pourghasemi, H.R. and Dixon, B. ۲۰۱۶. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental monitoring and assessment*, **188**(۱): ۴۴.
- Naghibi, S.A., Pourghasemi, H.R., Pourtaghi, Z.S. and Rezaei, A. ۲۰۱۵. Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. *Earth Science Informatics*, **8**(۱): ۱۷۱-۱۸۶.
- Nampak, H., Pradhan, B. and Abd Manap, M. ۲۰۱۴. Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology*, **۵۱۳**: ۲۸۳-۳۰۰.
- Nhu, V.H., Rahmati, O., Falah, F., Shojaei, S., Al-Ansari, N., Shahabi, H., Shirzadi, A., Górski, K., Nguyen, H. and Ahmad, B.B. ۲۰۲۰. Mapping of Groundwater Spring Potential in Karst Aquifer System Using Novel Ensemble Bivariate and Multivariate Models. *Water*, **12**(۴): ۹۸۵.
- Oh, H.J., Kim, Y.S., Choi, J.K., Park, E. and Lee, S. ۲۰۱۱. GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology*, **399**(۳-۴): ۱۵۸-۱۷۲.
- Ozdemir, A. ۲۰۱۱. GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *Journal of Hydrology*, **411**(۳-۴): ۲۹۰-۳۰۸.
- Pande, C. B., Moharir, K. N., Panneerselvam, B., Singh, S. K., Elbeltagi, A., Pham, Q. B., ... & Rajesh, J. ۲۰۲۱. Delineation of groundwater potential zones for sustainable development and planning using analytical hierarchy process (AHP), and MIF techniques. *Applied Water Science*, **11**(۱۲), ۱۸۶.
- Pourghasemi, H.R., Mohammady, M. and Pradhan, B. ۲۰۱۲. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena*, **۹۷**: ۷۱-۸۴.
- Pourtaghi, Z.S. and Pourghasemi, H.R. ۲۰۱۴. GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. *Hydrogeology Journal*, **22**(۳): ۶۴۳-۶۶۲.
- Pradhan, B. (۲۰۰۹). Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. *Open Geosciences*, **1**(۱): ۱۲۰-۱۲۹.
- Rahmati, O., Pourghasemi, H.R. and Melesse, A.M. ۲۰۱۶. 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**:۳۶۰-۳۷۲.
- Razavi-Termeh, S. V., Sadeghi-Niaraki, A., & Choi, S. M. (۲۰۱۹). Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. *Water*, **11**(۸): ۱۵۹۶.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. ۲۰۱۲. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **۶۷**: ۹۳-۱۰۴.
- Shannon, C.E., ۲۰۰۱. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, **5**(۱):۳-۵۵.

- Thanh, N. N., Chotpantarat, S., Ha, N. T., & Trung, N. H. (۲۰۲۳). Determination of conditioning factors for mapping nickel contamination susceptibility in groundwater in Kanchanaburi Province, Thailand, using random forest and maximum entropy. *Environmental Geochemistry and Health*, ۱-۲۰.
- Tien Bui, D., Le, K.T.T., Nguyen, V.C., Le, H.D. and Revhaug, I. ۲۰۱۶. Tropical forest fire susceptibility mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, using GIS-based kernel logistic regression. *Remote Sensing*, **8**(۴): ۳۴۷.
- Tweed, S.O., Leblanc, M., Webb, J.A. and Lubczynski, M.W. ۲۰۰۷. Remote sensing and GIS for mapping groundwater recharge and discharge areas in salinity prone catchments, southeastern Australia. *Hydrogeology Journal*, **15**(۱): ۷۵-۹۶.
- Yaghobi, S., Faramarzi, M., Karimi, H., & Sarvarian, J. ۲۰۱۹. Simulation of land-use changes in relation to changes of groundwater level in arid rangeland in western Iran. *International Journal of Environmental Science and Technology*, **16**(۳): ۱۶۳۷-۱۶۴۸.