

## مقایسه تکنیک‌های داده‌کاوی آنتروپی شانون و الگوریتم جنگل تصادفی در تهیه نقشه پتانسیل آب زیرزمینی جهرم

کریم سلیمانی<sup>۱\*</sup>، فاطمه علی‌دادگان‌فرد<sup>۲</sup>، حمیدرضا پورقاسمی<sup>۳</sup>

تاریخ دریافت: ۱۳۹۶/۹/۱۸

تاریخ پذیرش: ۱۳۹۸/۴/۱۵

### چکیده

منابع آب زیرزمینی همواره یکی از مهم‌ترین و مطمئن‌ترین منابع آبی در مناطق خشک و نیمه‌خشک است و به‌دست‌آوردن آب از این منابع نسبت به اقلیم‌های دیگر، اهمیت ویژه‌ای دارد. بنابراین هدف از تحقیق حاضر، تهیه نقشه پتانسیل آب زیرزمینی در شهرستان جهرم، استان فارس با استفاده از روش‌های آنتروپی شانون و الگوریتم جنگل تصادفی و مقایسه دقت آن‌هاست. بدین منظور با توجه به اطلاعات و داده‌های زمین‌شناسی، هیدرولوژیکی، ساختاری و توپوگرافی، نقشه‌های مورد نیاز تحقیق در محیط ArcGIS 10.3 و SAGA با فرمت رستری تهیه شدند. سپس با استفاده از مدل آنتروپی شانون و مدل جنگل تصادفی، وزن هریک از عوامل مؤثر در نرم‌افزار R محاسبه و در نهایت نقشه پتانسیل آب زیرزمینی تهیه شد. پس از تهیه نقشه‌های پتانسیل چشمه با استفاده از روش‌های مذکور، برای ارزیابی نتایج از منحنی تشخیص عملکرد نسبی (ROC) استفاده شد. سطح زیر منحنی (AUC) به‌دست‌آمده از منحنی تشخیص عملکرد نسبی، نشان‌دهنده دقت ۸۳/۹۰ و ۷۶/۲۰ درصد به‌ترتیب برای مدل‌های آنتروپی شانون و جنگل تصادفی است. نتایج به‌دست‌آمده نشان‌دهنده دقت بالای مدل آنتروپی شانون نسبت به مدل جنگل تصادفی است.

**واژه‌های کلیدی:** آب زیرزمینی، آنتروپی شانون، سامانه اطلاعات جغرافیایی، جنگل تصادفی، داده‌کاوی.

۱. استاد، گروه آبخیزداری، دانشکده کشاورزی و منابع طبیعی ساری؛ solaimani2001@yahoo.co.uk

۲. دانش‌آموخته کارشناسی ارشد سنجش از دور و GIS، مؤسسه آموزش عالی آبان هراز، آمل

۳. دانشیار، بخش مهندسی منابع طبیعی و محیط زیست، دانشکده کشاورزی دانشگاه شیراز

## مقدمه

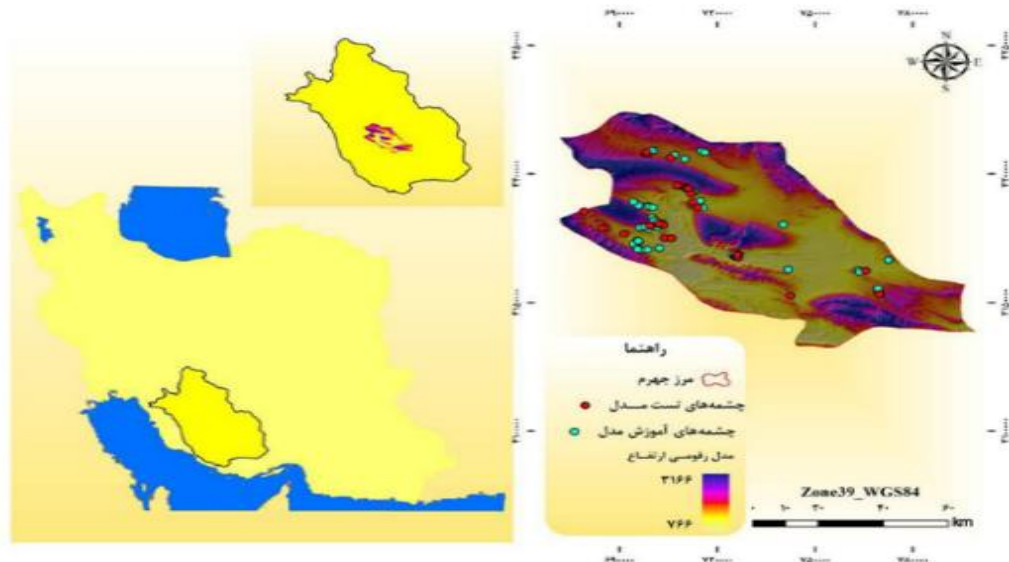
فلات ایران به‌ویژه نواحی مختلف ایران از اقلیم‌های متفاوتی تشکیل شده است، در حالی که در نواحی جنوبی آن، در تابستان گرما به ۵۰ درجه بالای صفر می‌رسد، نوار مدیترانه‌ای خزر، آب‌وهوای مرطوب را از سر می‌گذرانند. در این میان یکی از مهم‌ترین شاخص‌های تفاوت اقلیم‌ها مسئله آب است. امروزه تأمین آب به‌منظور تحقق اهداف توسعه پایداری، یکی از مهم‌ترین دغدغه‌ها و چالش‌ها در بیشتر کشورهای جهان است (ذبیحی و همکاران، ۲۰۱۵). به همین دلیل، تعیین مناطق دارای پتانسیل آب زیرزمینی از ابزارهای مهم در حفاظت، مدیریت و بهره‌برداری از منابع آب زیرزمینی به‌شمار می‌رود (همان). رحمتی و همکاران (۲۰۱۶) در پژوهشی به ارزیابی پتانسیل آب زیرزمینی مهران با مدل آنتروپی و الگوریتم جنگل تصادفی پرداختند. نتایج حاصل از این تحقیق، بیانگر عملکرد دقت بسیار خوب مدل آنتروپی با دقت ۹۱ درصد نسبت به مدل جنگل تصادفی با دقت ۸۵ درصد است. ذبیحی و همکاران (۲۰۱۵) با مدل آنتروپی شانون و جنگل تصادفی در دشت بجنورد خراسان شمالی به تهیه نقشه پتانسیل آب زیرزمینی پرداختند؛ نتایج ارزیابی دقت روش‌های مورد استفاده به‌منظور پهنه‌بندی پتانسیل آب زیرزمینی با استفاده از منحنی عملکرد نسبی ۸۵/۵۵، بیانگر دقت خیلی خوب مدل آنتروپی شانون نسبت به جنگل تصادفی با ۷۶/۹۵ درصد است. همچنین بر اساس مقایسه با مدل جنگل تصادفی با آنتروپی شانون لایه‌های فاصله از آبراهه، لیتولوژی، کاربری اراضی و طبقات ارتفاعی بیشترین تأثیر را بر پتانسیل آب زیرزمینی منطق مورد مطالعاتی داشته است. نقیبه و همکاران (۲۰۱۴) از مدل نسبت فراوانی و آنتروپی شانون به‌منظور تهیه نقشه پتانسیل آب زیرزمینی در حوضه آبخیز مغان در استان خراسان رضوی استفاده کردند. نتایج ایشان نشان داد که مدل آنتروپی شانون با سطح زیر منحنی برابر با ۹۱/۲۱ درصد کارایی بالایی داشته است. پورتنقی و پورقاسمی (۲۰۱۴) به ارزیابی پتانسیل آب زیرزمینی با استفاده از نقاط چشمه در شهرستان بیرجند پرداختند و نشان دادند که روش نسبت فراوانی با مقدار

سطح زیر منحنی AUC برابر ۷۵/۶۹ درصد بهترین عملکرد را در مقایسه با مدل‌های دیگر داشته است. اواده<sup>۱</sup> (۲۰۰۹) برای پتانسیل یابی آب زیرزمینی در منطقه شمال شرق اردن، از لایه‌های ارتفاع، خاک، ژئومورفولوژی، زمین‌شناسی، تراکم گسل و با روش AHP استفاده کردند. این منطقه به ۵ کلاس از خیلی فقیر تا خیلی خوب تقسیم شده است. ایتیشری<sup>۲</sup> و همکارانش (۲۰۱۳) در تحقیقی با استفاده از فرایند تحلیل شبکه آب زیرزمینی در منطقه یونائو در کشور هند را پتانسیل یابی کردند. نتایج این تحقیق نشان داد حدود ۱۵ درصد از مساحت منطقه، پتانسیل خوب و خیلی خوب از نظر آب زیرزمینی دارد. هدف از انجام این تحقیق، ارزیابی کارایی روش مدل آنتروپی و الگوریتم جنگل تصادفی در تهیه نقشه پتانسیل و وقوع چشمه در شهرستان جهرم و آزمایش قابلیت‌های این مدل‌ها در نقشه‌برداری نواحی با پتانسیل چشمه‌های آب زیرزمینی و تعیین مهم‌ترین فاکتورهای تأثیرگذار در وقوع چشمه منطقه مورد مطالعه است.

## مواد و روش کار

## منطقه مورد مطالعه

شهرستان جهرم با مساحت آن ۵۷۶۸ کیلومتر مربع، یکی از شهرستان‌های تابع استان فارس است که در نیمه جنوبی استان واقع شده است. ارتفاع بیشینه و کمینه منطقه مورد مطالعه به ترتیب ۳۱۶۶ و ۷۶۶ است. همچنین میانگین بارش سالیانه در شهرستان جهرم ۱۵۰ تا ۲۰۰ میلی‌متر است. بر اساس گزارش‌های آب منطقه‌ای شهر جهرم که در سال آبی ۲۰۰۶ تا ۲۰۰۹ انجام گرفته، در منطقه مطالعاتی ۵۰۰۰ حلقه چاه عمیق و نیمه‌عمیق و ۱۰۳ دهانه چشمه وجود دارد. بنا بر آخرین آمار در شهرستان جهرم طی سال ۲۰۰۹، مجموع برداشت از منابع آب زیرزمینی ۴۳۲ میلیون مترمکعب محاسبه شده است (شکل ۱).



شکل (۱): منطقه مورد مطالعه

Figure (1): The study area region

در محدوده مطالعاتی برای تهیه لایه لیتولوژی از نقشه

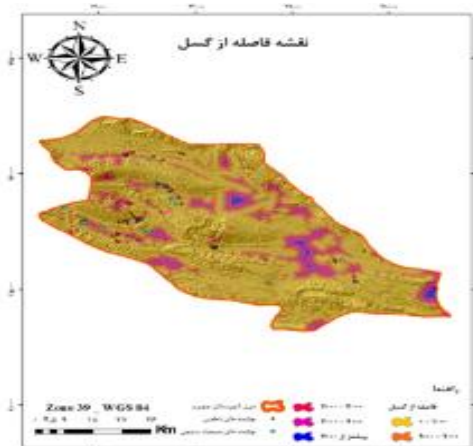
زمین‌شناسی ۱:۱۰۰۰۰۰ استفاده شد و پس از زمین مرجع شدن و رقومی شدن در محیط نرم‌افزاری ArcGIS لایه مذکور استخراج گردید (شکل ۲).

## روش کار

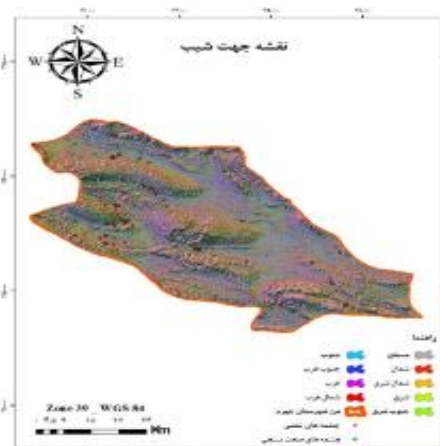
در ابتدا مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر از سازمان نقشه‌برداری کشور تهیه، و لایه‌های درجه شیب، ارتفاع و شکل شیب به‌طور مستقیم از آن در نرم‌افزار ArcGIS استخراج شد. نقشه کاربری اراضی از اداره کل منابع طبیعی استان فارس و در مقیاس ۱:۱۰۰۰۰۰ تهیه گردید. لایه‌های مربوط به آبراهه، یعنی تراکم آبراهه و فاصله از آن، با استفاده از نقشه توپوگرافی منطقه و DEM حوضه به کمک ابزار توابع تحلیلی هیدرولوژی در ArcGIS 10.3 تهیه شد. این لایه از واحد طول آبراهه در واحد کیلومتر مربع با  $(\text{Km}/\text{Km}^2)$  تهیه شد. نقشه فاصله و تراکم گسل منطقه با استفاده از نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ سازمان زمین‌شناسی و اکتشافات معدنی کشور در نرم‌افزار ArcGIS تهیه شد. فاکتورهای وابسته به آب مانند TWI نیز به‌طور مستقیم از مدل ارتفاعی رقومی در نرم‌افزار (SAGA-GIS, ArcGIS) مطابق رابطه (۱) محاسبه شد (مور<sup>۱</sup> و همکاران، ۱۹۹۱؛ جعفری و همکاران، ۲۰۱۴).

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

الف



ب



ج



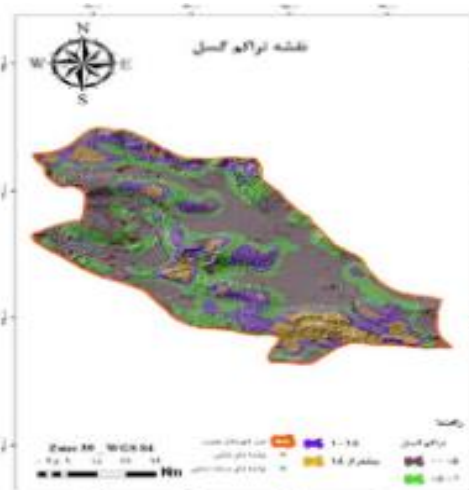
د



ذ



ر



شکل (۲): نقشه عوامل مؤثر مورد استفاده برای مدل سازی  
 Figure (2): The maps of conditioning factors for modeling

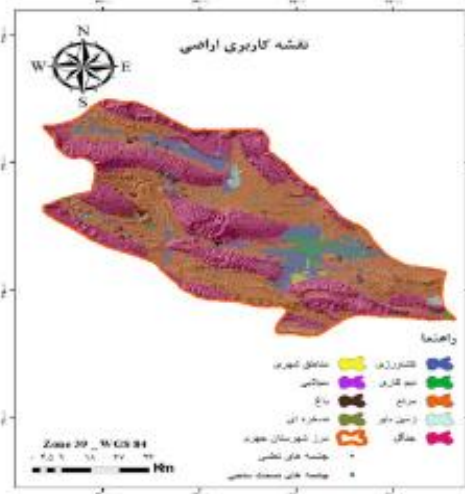
ز



س



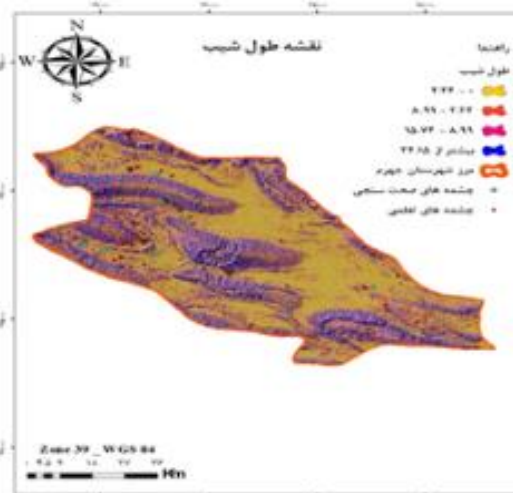
ش



ک



گ



ل



ادامه شکل (۲)

### مدل آنتروپی شانون

موجود در اطلاعات مشخصی است. آنتروپی نخستین بار توسط شانون (۱۹۴۸) بیان شد، که می‌توان آن را با بی‌نظمی معادل دانست. هرچه نظم سیستمی بیشتر شود، آنتروپی آن

آنتروپی در علوم فیزیک و در نظریه سیستم‌ها به معنای (بی‌نظمی) و در علوم ارتباطات و اطلاعات بیانگر اطمینان

دسته‌ای از درخت‌های تصمیم است و در حال حاضر یکی از بهترین الگوریتم‌های یادگیری است (ابراهیم‌خانی و همکاران، ۲۰۱۱). این الگوریتم یکی از ابزارهای کارآمد مورد استفاده در مسائل مربوط به تخمین متغیرهای هدف و یا طبقه‌بندی الگوها، درخت تصمیم است. در ساخت هر درخت تصمیم، یک استراتژی پاریشن‌بندی بازگشتی بالا به پایین استفاده می‌شود (رحمان، ۲۰۰۸). مدل پیش‌بینی‌کننده (RF)، بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوطه استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را انجام می‌دهند. درختان تصادفی با در نظر گرفتن بردار ورودی، آن را با هر درخت در جنگل طبقه‌بندی کرده و خروجی، برچسب‌های کلاسی هستند که از اکثریت آرا دریافت شده است.

### ارزیابی و مقایسه مدل‌ها

به منظور ارزیابی مدل‌های تهیه‌شده، از منحنی تشخیص عملکرد نسبی (ROC) استفاده شده است (پورقاسمی و همکاران، ۲۰۱۲؛ پورتنقی و همکاران، ۲۰۱۴). منحنی ROC یک نمایش گرافیکی از موازنه بین نرخ خطای منفی و مثبت برای هر مقدار احتمالی از برش‌هاست (پورقاسمی و همکاران، ۲۰۱۴). شاخص عملکرد نسبی، منحنی‌ای است که مؤلفه قائم و افقی آن که از ماتریس مقایسه با تعریف حد آستانه بین صفر و یک به دست می‌آیند، محاسبه می‌شوند.

### نتایج و بحث

نتایج حاصل از ارتباط بین پتانسیل آب زیرزمینی، برای ۱۳ عامل مؤثر انتخاب شده و با استفاده از مدل آنتروپی شانون در منطقه مورد مطالعه محاسبه و نشان داده شد. بر اساس نتایج به دست آمده، طبقه ارتفاعی پایین (۱۲۴۶-۷۷۶) بیشترین مقدار چشمه را در خود جای داده است بنابراین هرچه اختلاف ارتفاع نسبت به میانگین ارتفاع محل‌های تخلیه کمتر باشد، مطلوبیت بهتری در پتانسیل‌یابی آب زیرزمینی دارا خواهد بود و اینکه مناطق با پتانسیل بالا بیشتر منطبق بر ارتفاعات پایین به علت شیب کمتر نسبت به ارتفاعات، کم بودن تراکم زهکشی و تراس‌های آبرفتی و مخروط‌افکنه‌ها در این طبقه و

کاهش می‌یابد و برعکس، کاهش نظم باعث افزایش آنتروپی می‌شود. شانون تئوری آنتروپی را به صورت کمی بیان کرد (شانون، ۱۹۴۸). آنتروپی، یکی از رویکردهای مدیریتی است که به منظور برخورد با بی‌نظمی، بی‌ثباتی و عدم یقین‌های موجود در یک سیستم استفاده می‌شود (پورقاسمی و همکاران، ۲۰۱۲). در آنتروپی، متغیرهایی با حداکثر تأثیر در وقوع یک رخداد، تعیین و مشخص می‌شوند. به منظور پهنه‌بندی و ترسیم نقشه با استفاده از مدل آنتروپی شانون از رابطه‌های (۲) تا (۶) در این تحقیق استفاده شد (بدناریک<sup>۲</sup> و همکاران، ۲۰۱۰؛ کنستانتین<sup>۳</sup> و همکاران، ۲۰۱۱؛ پورقاسمی و همکاران، ۲۰۱۲).

$$(P_{ij}) = P_{ij} / \sum_{j=1}^{S_j} P_{ij} \quad (2)$$

$$H_j = \sum_{j=1}^{S_j} (P_{ij}) \log_2(P_{ij}), j = 1, \dots, n \quad (3)$$

$$H_{j \max} = \log_2 S_j, S_j - \text{Number of Class} \quad (4)$$

$$W_{ij} = I_j P_{ij} \quad (6)$$

که در آن،  $P_{ij}$  تراکم چشمه در هر طبقه،  $(P_{ij})$  احتمال وجود چشمه در هر عامل و کلاس‌های مرتبط با آن،  $H_j$  و  $H_{j \max}$  مقادیر آنتروپی و ماکزیمیم آنتروپی،  $I_j$  ضریب اطلاعات و  $W_{ij}$  وزن نهایی هر عامل است. بعد از تعیین وزن نهایی هر عامل و ضرب آن در کلاس‌های عامل مذکور، یعنی مقدارهای  $P_{ij}$  مربوط به هر عامل و کلاس‌های آن بر اساس تقسیم تعداد چشمه بر تعداد چشمه بر تعداد پیکسل‌های کلاس‌های عوامل (جدول ۱)، نقشه‌های وزنی با هم جمع شده و نقشه نهایی پتانسیل آب زیرزمینی تهیه شد. طبقه‌هایی که حساسیت بیشتری دارند، وزن بیشتری به خود اختصاص می‌دهند (پورقاسمی و همکاران، ۲۰۱۲؛ نقیبی و همکاران، ۲۰۱۴؛ کنستانتین و همکاران، ۲۰۱۱).

### مدل جنگل تصادفی

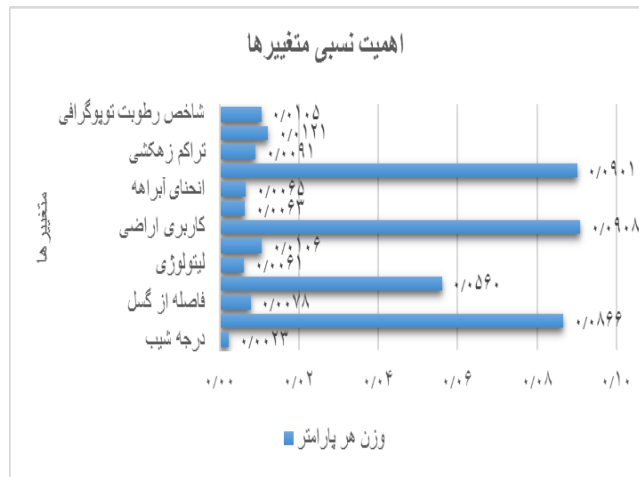
این الگوریتم را نخستین بار لئوبرایمن و آدل کاتلر ایجاد کردند و توسعه دادند. الگوریتم جنگل تصادفی مبتنی بر

1. Shannon
2. Bednarik
3. Constantin

زیرا مناطق با صخره‌های مقاوم و یا مواد زیرخاک نفوذپذیر، تراکم زهکشی کمتری دارند و همچنین بیش از ۶۰ درصد از سازنده‌های منطقه مورد مطالعه از دوره سنوزوئیک است. عملکرد فعالیت‌های تکتونیکی موجب گسترش گسل‌ها و سیستم درزه و شکستگی‌ها، در این پهنه‌های آهکی، شرایط مناسبی را برای ذخیره و حرکت آب فراهم آورده است. مناطق با پوشش گیاهی متراکم‌تر، مستعد نفوذ آب به داخل خاک با توجه به رابطه مثبت بین توانایی نفوذ و تراکم پوشش هستند. همچنین بر اساس نتایج ارائه شده در جدول (۲) (Wij) مشخص شد که عوامل لیتولوژی، کاربری اراضی، تراکم گسل، شاخص رطوبت توپوگرافی، ارتفاع، فاصله از گسل، فاصله از رودخانه، درجه شیب، انحنای آبراهه، انحنای سطح، تراکم زهکشی، جهت شیب، طول شیب به ترتیب بیشترین وزن (۰/۱۷، ۸/۶۵، ۰/۳۳۶، ۰/۳۲۲، ۰/۳۰۸، ۰/۳۰۱، ۰/۲۲۲، ۰/۱۵۹، ۰/۱۲۵، ۰/۱۰۹، ۰/۰۵۰، ۰/۰۲۵، ۰/۰۰۲) را به خود اختصاص دادند.

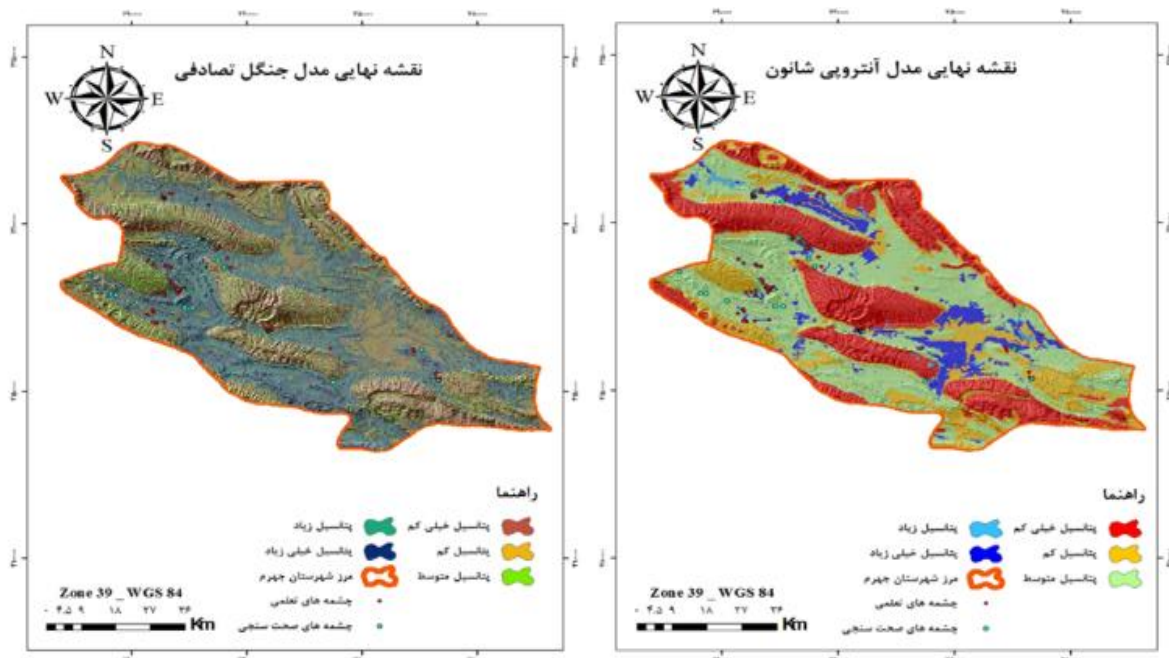
در نهایت، نقشه پتانسیل آب زیرزمینی منطقه مورد مطالعه بر اساس مدل آنتروپی شانون تهیه و در شکل (۵) نشان داده شده است. نتایج بررسی اهمیت هریک از فاکتورهای مؤثر به دست آمده از مدل جنگل تصادفی در شکل مشهود است. بر این اساس، مهم‌ترین عوامل تأثیرگذار در این مدل به ترتیب، فاصله از آبراهه، کاربری اراضی، تراکم گسل و ارتفاع است. همچنین پارامترهایی مانند جهت شیب، لیتولوژی، انحنای سطح، انحنای آبراهه به ترتیب کمترین تأثیر بر وقوع چشمه در منطقه مطالعاتی داشتند. در شکل (۳) این نتایج ارائه شده است. نقشه پتانسیل آب زیرزمینی بر اساس مدل آنتروپی شانون و الگوریتم جنگل تصادفی در محیط نرم‌افزار ArcGIS به پنج کلاس با پتانسیل خیلی کم (بدون پتانسیل)، پتانسیل کم، پتانسیل متوسط، پتانسیل زیاد و پتانسیل خیلی زیاد طبقه‌بندی و تهیه شدند. شکل (۴) این نقشه‌ها را به ترتیب برای مدل آنتروپی شانون و الگوریتم جنگل تصادفی نشان می‌دهد. نقشه پتانسیل متوسط در بخش غربی منطقه مورد مطالعه قرار گرفته است که احتمال وقوع چشمه را در این طبقه نشان می‌دهد.

در نتیجه نفوذ بیشتر و امکان ورود آب زیرزمینی از طریق درز و شکاف‌ها رو به این سطح از طبقه ارتفاعی فراهم می‌کند. بیشترین تعداد چشمه‌ها در مناطق مسطح است و کمترین آن در کلاس جهت شمال غرب است. همچنین نتایج نشان می‌دهد که فاصله صفر تا ۲۰۰ متری از آبراهه، بیشترین چشمه را در خود جای داده است. بررسی عوامل دیگر نشان داده که بیشترین تعداد وقوع چشمه به ترتیب مربوط به کلاس درجه شیب صفر تا ۵ درجه است؛ مناطقی که دارای شیب پایینی هستند آب را برای مدت طولانی حفظ می‌کنند، این امر باعث نفوذ یا تغذیه بیشتر آب می‌شود، ولی نواحی با شیب بالا دارای مقادیر زیادی رواناب هستند و مقدار نفوذ در آن‌ها کم است (رحمان، ۲۰۰۸). طول شیب صفر تا ۲/۲۴، تراکم زهکشی ۱/۲۸ تا ۲/۵۶، شاخص رطوبت توپوگرافی مربوط به کلاس ۱۲ تا ۱۶، انحنای سطح مسطح و انحنای آبراهه مسطح، فاصله از گسل کمتر از ۱۰۰۰ متر و تراکم گسل مربوط به کلاس ۰/۵ تا ۱، لیتولوژی Qft2 بیشترین مقدار وقوع چشمه را نشان می‌دهند. بر اساس نتایج جدول (۱) می‌توان گفت تراکم زهکشی یک منطقه با پتانسیل آب زیرزمینی نسبت عکس داشته، به طوری که تراکم بیشتر نشان‌دهنده جریان بیشتر آب به صورت سطحی و نفوذ کمتر، در نتیجه کم بودن آب زیرزمینی است. وجود رابطه معکوس بین فاصله از گسل و تعداد چشمه نیز از دیگر نتایج پژوهش حاضر است. این نتیجه بدان معناست که هرچه فاصله از گسل کمتر باشد، تعداد چشمه بیشتر است. دلیل این امر، تراکم شکستگی است که معیاری برای خردشدگی سازند می‌باشد؛ بدین معنی که تعداد درزه و شکستگی که به‌ازای هر واحد طول در یک منطقه ظاهر می‌شوند، مدنظر است. افزایش تراکم درزه و گسل‌ها به‌طور کلی، نقش مؤثری در نفوذ و انتقال‌دهی آب زیرزمینی دارد. نتایج پژوهش حاضر در خصوص ارتباط پتانسیل آب زیرزمینی با لیتولوژی نشان داد که در دوره سنوزوئیک در معیار زمین‌شناسی دارای بیشترین تأثیر بر وقوع چشمه منطقه مورد مطالعه دارد. تعداد چشمه در سازند OMr، Qft2، MuPlaj به ترتیب بیشترین تأثیر را در وقوع چشمه داشته و بقیه سازندها تأثیر چندانی بر پتانسیل آب زیرزمینی ندارند؛



شکل (۳): اهمیت نسبی عوامل مؤثر بر وقوع چشمه در شهرستان جهرم با مدل جنگل تصادفی

Figure (3): Relative importance of effective factors on spring occurrence in Jahrom with Random Forest model



شکل (۴): نقشه نهایی پتانسیل آب زیرزمینی منطقه مورد مطالعه مدل آنتروپی شانون (راست) و جنگل تصادفی (چپ)

Figure (4): Final map of groundwater potential based-on Shannon Entropy (right) and Random Forest (left)

تأمین کننده آب چشمه منطقه شمالی است. همچنین شکل (۵) نشان می دهد در مدل الگوریتم جنگل تصادفی، مناطق با پتانسیل زیاد در بخش شمال غربی، دارای پتانسیل آب زیرزمینی هستند. در این منطقه، رودخانه سیمکان که از شاخه های اصلی رودخانه قره آغاج است، با جریان دائمی و دبی پایه کم جریان دارد. پوش گیاهی عمده آن، جنگلی و کشاورزی است. بر اساس نتایج نقشه های پتانسیل آب زیرزمینی هر دو مدل، می توان دلیل تفاوت های آن را در ساختار و عدم قطعیت های موجود در مدل سازی دانست که

پتانسیل آب زیرزمینی بر اساس این دو مدل مذکور تفاوت هایی با هم دارند. بر اساس مدل آنتروپی شانون، مناطق با آبدهی بیشتر چشمه های این قسمت بین دامنه صفر تا ۵ درجه و از نظر ساختار از نوع زهکشی- درز و شکافی هستند. طبقه بعدی که می توان از لحاظ آب زیرزمینی نام برد مناطق با پتانسیل خیلی زیاد در شمال شهرستان جهرم است. در این مناطق پوشش گیاهی، جنگل امیر سالار خفر یکی از عوامل مؤثر در پتانسیل آب زیرزمینی است. سازند آسماری به عنوان مهم ترین عامل تغذیه منابع آب زیرزمینی و همچنین



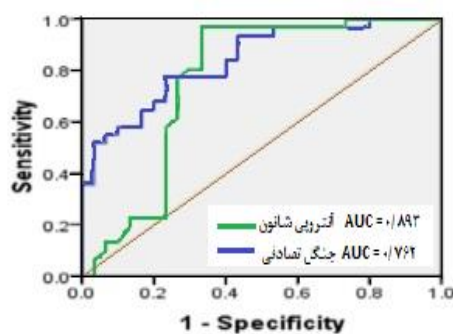
## نتیجه‌گیری

هدف اصلی این پژوهش نیز استفاده از مدل آنتروپی شانون و تکنیک الگوریتم جنگل تصادفی به‌عنوان تکنیکی در روش داده‌کاوی به‌عنوان صحت و اعتماد به نقشه پتانسیل آب زیرزمینی در منطقه مورد مطالعه شهرستان جهرم است. با توجه به ویژگی‌های هیدرولیکی، زمین‌شناسی، اقلیمی و ژئومورفولوژیکی منطقه و اهداف پژوهش، مهم‌ترین عوامل مؤثر بر نفوذپذیری و تغذیه سفره‌های زیرزمینی مد نظر قرار گرفت. بدین منظور ۱۳ عامل به‌عنوان عوامل مؤثر در پتانسیل آب زیرزمینی تهیه شد. در مدل آنتروپی شانون لیتولوژی، کاربری اراضی تراکم گسل بیشترین تأثیر و طول شیب، درجه شیب و تراکم زهکشی کمترین تأثیر را در وقوع چشمه در منطقه مورد مطالعه نشان دادند. همچنین فاصله از رودخانه، کاربری اراضی و تراکم گسل بیشترین و لیتولوژی، انحنای سطح و انحنای آبراهه کمترین تأثیر در وقوع چشمه در الگوریتم جنگل تصادفی در شهرستان جهرم داشتند. تعداد بسیاری از چشمه‌های آزمایشی در ناحیه با پتانسیل متوسط و زیاد قرار گرفته و چشمه‌های واقع شده در کلاس پتانسیل نامناسب و کم، ناچیز هستند (زندى و همکاران، ۲۰۱۵). صحت‌سنجی نقشه‌های به‌دست‌آمده با ارزیابی ROC برای تهیه پهنه‌بندی حساسیت پتانسیل آب زیرزمینی مؤید این مطلب است که مدل آنتروپی شانون با ارزیابی ۰/۸۳۹ درصد نسبت به الگوریتم جنگل تصادفی با ارزیابی ۰/۷۶۲ درصد دقت مطلوب‌تری داشته است و می‌تواند مورد آزمون و بررسی قرار گیرد. همچنین با توجه به نقشه‌های موجود در مدل‌های مذکور، مناطق شمالی و غربی محدوده مطالعاتی نسبت به سایر مناطق، از پتانسیل آب زیرزمینی بیشتری برخوردارند و این می‌تواند به‌دلیل انطباق این مناطق با مناطقی باشد که شیب پایین، ارتفاع پایین و پوشش گیاهی فراوان دارد؛ همچنین لیتولوژی این قسمت در بخش بالایی مربوط به رسوبات آبرفتی کواترنری (مخروط‌افکنه‌های جدید) است که ضخامت کمی داشته و نقش تغذیه‌سازند سخت را بر عهده دارد (تبرمایه و واعظی هیر، ۲۰۱۶). با توجه به سطح زیر منحنی به‌دست‌آمده از این پژوهش، نتایج

نقش اهمیت نسبی عوامل اثرگذار بر وقوع چشمه توسط همان ملاک‌ها و معیارهای ارزیابی در مرحله مدل‌سازی انجام می‌شود؛ زیرا بعضی از این عوامل ممکن است در یک مدل نقش بسیار بالایی داشته باشند در حالی که برای یک مدل دیگر چندان مفید نباشند و بالعکس (تین بوی<sup>۱</sup> و همکاران، ۲۰۱۵). همان‌گونه که در مدل آنتروپی شانون، عوامل تأثیرگذار بر وقوع چشمه لیتولوژی، کاربری اراضی و تراکم گسل بودند که بیشترین وزن را به خود اختصاص دادند، در مدل الگوریتم جنگل تصادفی فاصله از رودخانه، کاربری اراضی، تراکم گسل و ارتفاع بیشترین وزن را گرفتند. همچنین بر اساس شکل (۵) مدل آنتروپی شانون نسبت به مدل جنگل تصادفی ارزیابی و صحت بیشتری را برای پتانسیل‌یابی آب زیرزمینی منطقه مورد مطالعه داشته است.

## ارزیابی صحت نقشه‌های پتانسیل آب زیرزمینی

به‌منظور ارزیابی صحت‌سنجی مدل‌های تهیه‌شده، از منحنی تشخیص عملکرد نسبی ROC استفاده شد (پورقاسمی و همکاران، ۲۰۱۳). در این پژوهش از داده‌های صحت‌سنجی برای ارزیابی دقت مدل‌ها استفاده شد. در شکل (۵) استفاده از منحنی ROC برای مدل آنتروپی شانون نشان می‌دهد که مقدار AUC برابر ۰/۸۳۹ درصد بوده است که در مقایسه با مدل الگوریتم جنگل تصادفی که مقدار AUC آن برابر با ۰/۷۶۲ درصد است، عملکرد بهتری داشته و این مقدار در طبقه خیلی خوب قرار می‌گیرد.



شکل (۵): منحنی ROC و سطح زیر منحنی مربوط به نقشه پتانسیل آب زیرزمینی مدل آنتروپی شانون و جنگل تصادفی

Figure (5): ROC curve and area under curve related to groundwater potential map of Entropy and Random forest model

این تحقیق با مطالعات دیگر پژوهشگران از قبیل ذبیحی و همکاران، رحمتی و همکاران که پیش تر در پیشینه تحقیق به آن اشاره شد، همخوانی دارد. پژوهشگران مذکور روش های داده کاوی جنگل تصادفی و آنتروپی شانون را به عنوان روش مناسب در تهیه نقشه پتانسیل آب زیرزمینی در مناطق مختلف پیشنهاد کرده بودند و نتایج به دست آمده از این پژوهش نیز بیانگر نظر آنهاست.

## منابع

1. Awawdeh. M. M., Al-Mohammed. M., 2009. Integrated GIS and remote sensing for mapping groundwater potential zones in Tulul al Ashaqif Highlands, NE Jordan, International Symposium Geotunis 2009, 16 -20 December.
2. Bednarik, M., Magulova, B., Matys. M., Marschalko, M., 2010. Landslide susceptibility assessment of the Kralovany–Liptovsky Mikulaš railway case study, Physics and Chemistry of the Earth, vol. 35, pp.162–171.
3. Constantin, M., Bednarik, M., Jurchescu, M., Vlaicu, M., 2011. Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania), Environmental Earth Science, vol. 63, pp. 397–406.
4. Ebrahim Khani, S., Afzali, M., Shokouhi A., 2011, Prediction and Study of Road Accident Factors Using Data Mining Algorithms, Zanjan Police Knowledge Quarterly, No. 1, Year One: 127-111. (In Persian).
5. Etishree A., Rajat A., Garg R.D., Garg, P.K., 2013. Delineation of groundwater potential zone: An AHP/ANP approach. Journal of Earth System Science. 122(3): 887-898.
6. Jaafari, A., Najafi, A., Pourghasemi, H., Rezaeian, J., Sattarian, A., 2014. 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, 11(4), 909-926.
7. 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.
8. Naghibi, S., A., Pourghasemi, H. R., Pourtaghi Z., S., Rezaei, A., 2014. Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran, Journal of Earth Science, vol. 8 (1), pp. 171-186.
9. Pourghasemi, H., Moradi, H., Aghda, F., 2013, Preparation of Landslide Sensitivity Map Using Adaptive Neuro-Fuzzy Inference System in Northern Tehran, Journal of Earthquake Research, No. 10: 63-78. (In Persian).
10. Pourghasemi, H., R., Mohammady, M., Pradhan, B., 2012. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran, Catena, vol. 97, pp. 71–84.
11. Pourghasemi, H., R., Moradi, H., R., Fatemi A., S., M., Gokceoglu, C., Pradhan, B., 2014. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi criteria evaluation models (North of Tehran, Iran), Arabian Journal of Geoscience, vol. 7, pp. 1857-1878.
12. Pourtaghi, Z., S., Pourghasemi, H., R., 2014. GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province. Hydrogeology Journal, vol. 22, pp. 643-662.
13. Rahman. A., 2008. A GIS based DRASTIC model for assessing groundwater vulnerability in shallow aquifer in Aligarh, India, Applied Geography, 28, pp:32–53.
14. Rahmati, O., Pourghasemi, H., R., Melesse, M., A., 2016. Application of GIS-Based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. Catena 137(2016) 360 – 372.
15. Shannon, Claude. 1948, A mathematical theory of communication, Bulletin System, Technology Journal, vol. 27, pp. 379–423.
16. Tarbamayeh, M., Vaezi Hir, A., 2016. Groundwater potential in hard formations by application of AHP and SAW multi criteria methods (case study: Almaneh watershed), Journal of Water and Soil (Agricultural Sciences and Technology), Volume 30, Number 5, Azar-1395, pp. 1477-1. (In Persian)

17. Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2015. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, <http://dx.doi.org/10.1007/s10346-015-0557-6>.
18. Zabihi M., Pourghasemi, H., Behzadfar, M., 2015. Preparation of Groundwater Potential Map Using Shannon Entropy Models and Random Forest in Bojnourd Plain. *Journal of Ecohydrology* Volume 2, summer 2015, pp. 221-232. (In Persian).
19. Zandi, S., Solaimani, K., Zandi, J., 2015. Preparing Map of Areas with the Potential of Groundwater Springs Using Logistic Regression Method in the Environment 4. Case Study: Mountainous of watershed, Kordestan. *Journal of watershed management* Vol. 6 / No. 12 / Autumn and Winter of 2015. (In Persian).

## Comparison of Shannon Entropy Data Mining Techniques and Random Forest Algorithm to Preparing Underground Water Potential Map of Jahrom

Karim Solaimani<sup>1\*</sup>, Fatemeh Alidadganfard<sup>2</sup>, Hamid Reza Pourghasemi<sup>3</sup>

Received: 09/12/2017

Accepted: 06/07/2019

---

### Expanded abstracts

**Introduction:** Iran's plateau, especially in different regions of Iran, is composed of different climates, while in its southern areas summer temperatures reach 50 degrees above zero, the Caspian Mediterranean region is running wet weather. One of the most important indicators of climate difference is the issue of water. Today, water supply is one of the most important concerns to realize the goals of sustainable development and challenges are in most countries in the world (Zabihi et al., 2015). For this reason, the identification of areas with groundwater potential is one of the important tools for conservation, management and exploitation of groundwater resources (Zabihi et al., 2016). In an exploratory study, Mehran groundwater potential assessment was conducted with the entropy model and random forest log algorithm (Rahmati, et al. 2016).

**Materials and Methods:** Jahrom is one of the cities of Fars, located in the southern part of this province with an area of 5,768 km<sup>2</sup>. The maximum and minimum altitudes of the study area are 3166 and 766, respectively. These altitudes are part of the Zagros folding, and for most of the mountains, they are located in the southern part of the Southern Zagros Mountains, and its branches are surrounded by the city. In the present study, 13 variables including geological layers, slope, tilt direction, elevation, distance from fault, fault density, distance from the waterway, water congestion, land use, topography humidity index Surface curvature, curvature of the waterways,

---

1. Professor, Dept of Watershed Management, University of Agricultural and Natural Resources of Sari; solaimani2001@yahoo.co.uk

2. M.Sc Graduated of RS and GIS, Aban Hazard Institute of Higher Education, Amol

3. Associate Prof., Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University  
DOI: 10.22052/deej.2018.7.24.25

slope length using the experiences of experts and researchers in the surveys have been used in the study area. At first, a digital elevation model with a spatial resolution of 10 meters from the mapping organization of the country was prepared and the slope, height and gradient layers were directly extracted from the ArcGIS software. The map of distance and density of the region fault was prepared using ArcGIS software using 1:100000 Geological maps of Iran Geological Survey and Mining Exploration. Water dependent factors such as TWI were calculated directly from the digital elevation model in software (SAGA-GIS, ArcGIS) according to Equation 1 (Moore et al., 1991; Jafari et al., 2014). In the entropy model, variables with the maximum effect on the occurrence of an event are determined. (Bodarik et al., 2010; Constantin et al., 2011; Pourghasemi et al., 2012).

**Results:** In order to evaluate the produced models, the relative recognition function curve (ROC) has been used (Pourghasemi et al., 2012; Porruthi et al., 2014). The ROC curve is a graphical representation of the equilibrium between the negative and positive error rates for any possible cutoffs (Pourghasemi et al., 2013). The relative performance index is a curve whose vertical and horizontal components are calculated from the comparison matrix with the definition of the threshold between zero and one. Shannon's entropy model shows that the AUC was 83.9%, which is better than classical forestry algorithm model, whose AUC is equal to 76.20%, and this level is very good.

**Discussion and conclusion:** Considering the experimental springs located in the potential springs of the spring, it reveals the fact that with the increase in the potential spring levels of the spring, there has been a significant increase in the density of springs in these areas, while experimental springs have not played a role in the modeling. Many of the pilot springs are located in a medium and high potential area, and springs located in an inappropriate and low potential class are insignificant (Zandi et al., 1394). Validation of the maps obtained by assessing the ROC for the zoning of groundwater potential susceptibility confirms that Shannon entropy model with a 0.839 percent estimation compared to the random forest log algorithm with an estimated 762 percent accuracy has a better accuracy.

**Keywords:** Groundwater, Shannon Entropy, Geographic Information System, Random forest.