

## مقایسه کارایی مدل‌های شبکه عصبی مصنوعی، منطق فازی و جنگل تصادفی در برآورد پارامتر قابلیت انتقال آبخوان دشت ملکان

حسین نوروزی قوشبلاغ<sup>۱</sup>، عطاالله ندیری<sup>۲\*</sup>، اصغر اصغری مقدم<sup>۳</sup>، مریم قره‌خانی<sup>۴</sup>

۱. دانشجوی دکتری هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز، تبریز، ایران

۲. دانشیار هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز، تبریز، ایران

۳. استاد هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز، تبریز، ایران

۴. دانشجوی دکتری هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز، تبریز، ایران

(تاریخ دریافت ۱۳۹۶/۰۵/۳۱؛ تاریخ تصویب ۱۳۹۶/۱۰/۱۴)

### چکیده

آبخوان دشت ملکان به عنوان یکی از آبخوان‌های حوضه دریاچه ارومیه، به مدیریت صحیح کمی و کیفی نیاز دارد. روش‌های مختلفی از جمله انجام آزمون پمپاژ، روش‌های آزمایشگاهی، استفاده از ردیاب‌ها و روش‌های ژئوفیزیکی برای ارزیابی پارامترهای هیدروژئولوژیکی و مدیریت مناسب آبخوان‌ها وجود دارد. هر چند تعبیر و تفسیر داده‌های به‌دست‌آمده از آزمون پمپاژ، بهترین روش تخمین پارامترهای هیدروژئولوژیکی آبخوان است، اما این روش پرهزینه، وقت‌گیر و نتایج آن مختص به مناطق محدودی خواهد بود. با توجه به اینکه مدل‌های هوش مصنوعی توانایی‌هایی در برآورد پارامترهای هیدروژئولوژیکی نشان داده‌اند، در تحقیق حاضر کارایی مدل‌های شبکه‌های عصبی مصنوعی، منطق فازی و جنگل تصادفی در برآورد پارامتر قابلیت انتقال آبخوان دشت ملکان بررسی شده است. پارامترهای ژئوفیزیکی و هیدروژئولوژیکی مرتبط با قابلیت انتقال، از جمله مقاومت عرضی، هدایت الکتریکی، ضخامت آبخوان و هدایت هیدرولیکی به عنوان مهم‌ترین ورودی در این مدل‌ها در نظر گرفته شده است. بر اساس نتایج به‌دست‌آمده از بین مدل‌های شبکه عصبی و فازی و جنگل تصادفی، مدل جنگل تصادفی دقت و توانایی بیشتری در شبیه‌سازی داشته است. نتایج به‌دست‌آمده از شبیه‌سازی (  $AUC = 0.96$ ،  $MSE = 0.01$  و  $R^2 = 0.986$  ) و تعیین مهم‌ترین پارامترهای تأثیرگذار در پیش‌بینی قابلیت انتقال، گویای برتری این مدل نسبت به مدل‌های شبکه‌های عصبی مصنوعی و منطق فازی در بحث پیش‌بینی است.

**کلیدواژگان:** آب‌های زیرزمینی، جنگل تصادفی، دشت ملکان، قابلیت انتقال، هوش مصنوعی.

## مقدمه

به منظور مدیریت منابع آب‌های زیرزمینی و ارائه راه‌کارهای مدیریتی مناسب، شناخت پارامترهای هیدروژئولوژیکی و توصیف کمی از آبخوان امری حیاتی است. با توجه به اهمیت این پارامترها روش‌های مختلفی از جمله روش‌های آزمایشگاهی، شبیه‌سازی آب‌های زیرزمینی، آزمون پمپاژ، استفاده از ردیاب‌ها، روش‌های ژئوفیزیکی و غیره برای اندازه‌گیری این پارامترها توسعه پیدا کرده‌اند [۱-۴]. ارزیابی پارامترهای هیدروژئولوژیکی با روش‌های یادشده نه تنها به شرایط لازم نیاز دارد بلکه دارای محدودیت‌هایی نظیر صرف وقت و هزینه‌های نسبتاً زیاد است. قابلیت انتقال آبخوان (T)، حاصل ضرب هدایت هیدرولیکی (K)، در ضخامت آبخوان (B) است و یکی از پارامترهای مهم آبخوان است که بسیاری از خصوصیات جریان آب در محیط‌های متخلخل به مقدار این پارامتر بستگی دارد. اخیراً، با توسعه مدل‌های هوش مصنوعی در علوم مختلف، روش‌های متنوعی برای تعیین پارامترهای هیدروژئولوژیکی ارائه شده است. مطالعات فراوانی روی مقایسه کارایی روش‌های هوش مصنوعی از جمله روش‌های تخمین‌گری مانند فازی<sup>۱</sup>، شبکه‌های عصبی مصنوعی<sup>۲</sup> و نروفازی<sup>۳</sup> و روش‌های متنوع بهینه‌سازی در برآورد و پیش‌بینی زمانی و مکانی پارامترها و پدیده‌های هیدرولیکی انجام شده است. از جمله می‌توان به مطالعاتی اشاره کرد که با استفاده از روش‌های تخمین‌گری و بهینه‌سازی، هدایت هیدرولیکی را در محیط متخلخل به دست آوردند [۵ و ۶]. تا کنون مطالعات فراوانی روی پیش‌بینی پارامترهای هیدروژئولوژیکی با استفاده از مدل‌های هوش مصنوعی مانند فازی و شبکه‌های عصبی مصنوعی صورت گرفته است [۷-۹]. مدل‌های فازی معمولاً در مدل‌سازی مسائلی که با عدم قطعیت ذاتی همراه هستند، مناسب‌اند [۱۰ و ۱۱] و مدل شبکه‌های عصبی مصنوعی به عنوان یک تقریب‌کننده جهانی می‌تواند در پیش‌بینی بررسی مسائل غیرخطی مفید باشد [۱۰ و ۱۲]. روش‌هایی که تا کنون برای ارزیابی قابلیت انتقال ارائه شده است، به‌طور نسبی از داده‌های کمتری استفاده

کرده‌اند [۱۳]. ابزارهای یادگیری<sup>۴</sup> و روش‌های جدید، با بهره‌گیری از رگرسیون‌های گروهی ارائه شده و در حال توسعه‌اند. جنگل تصادفی<sup>۵</sup> (RF) یکی از انواع روش‌های یادگیری است که از الگوریتم‌های پایه برای پیش‌بینی‌های چندگانه تکراری<sup>۶</sup> استفاده می‌کند [۱۴]. مدل RF با داشتن مزایایی همچون یادگیری روابط غیرخطی، توانایی مقابله با داده‌های پرت و ساختگی، برآورد خطای غیرمرتبط داخلی، اجرای هزاران داده ورودی و داشتن حساسیت کمتر در برابر گیرافتادن در مینیمم محلی، به عنوان یک روش دقیق در پیش‌بینی شناخته شده است. جنگل تصادفی اهمیت نسبی متغیرها را ارزیابی می‌کند و قادر به انتخاب متغیرهای مهم نیز است و در عین حال پارامترسازی آن نسبت به روش‌های دیگر مانند شبکه‌های عصبی، محاسبات ساده‌تری دارد [۱۳]. جنگل تصادفی هزاران متغیر ورودی را پذیراست، همچنین می‌تواند اهمیت نسبی متغیرها را ارزیابی کند و مهم‌ترین متغیر را در پیش‌بینی مدل تشخیص دهد [۱۳]. این روش پتانسیل لازم را به عنوان یک ابزار مدل مکانی برای ارزیابی مسائل زیست‌محیطی و منابع آب داد [۱۳]. در سال ۲۰۱۲، یو و همکارانش [۱۵] عملکرد رگرسیون منطقی، درخت کلاس‌بندی رگرسیونی و RF را بررسی کردند. در این مطالعه مشخص شد که روش RF بهترین عملکرد را در میان روش‌های به کار برده شده دارد. این مدل برای تعیین مناطق آسیب‌پذیر آبخوان دشت ملکان به کار برده شد و نتایج نشان داد مدل مد نظر می‌تواند با شناسایی مناطق مستعد به آلودگی، برای مدیریت و نظارت صحیح آب‌های زیرزمینی استفاده شود [۱۶]. دشت ملکان یکی از دشت‌های حوضه دریاچه ارومیه محسوب می‌شود که متأسفانه به علت وجود برداشت بی‌رویه از منابع آب زیرزمینی به شدت با بحران کاهش سطح آب زیرزمینی مواجه است و با وجود مطالعات فراوانی مانند مطالعات تفصیلی و ژئوفیزیک، که در سال‌های اخیر در این دشت صورت گرفته، هنوز نیاز فراوانی به شناخت پارامترهای هیدروژئولوژیکی برای مدیریت منابع آب دارد. بنابراین، در مطالعه حاضر با استفاده از داده‌های مرتبط با قابلیت انتقال

4. Machine learning  
5. Random Forest  
6. Repeated Multiple Prediction

1. Fuzzy logic  
2. Artificial Neural Networks  
3. Neurofuzzy

هم‌ضخامت، لاگ‌های حفاری و داده‌های ژئوفیزیکی در قسمت‌های بالادست آبخوان، ذرات تشکیل‌دهنده سفره دانه‌درشت بوده و قابلیت انتقال در این مناطق زیاد است و هر چه به سمت دریاچه ارومیه نزدیک می‌شویم رسوبات دانه‌ریزتر می‌شود و قابلیت انتقال نیز کاهش می‌یابد.

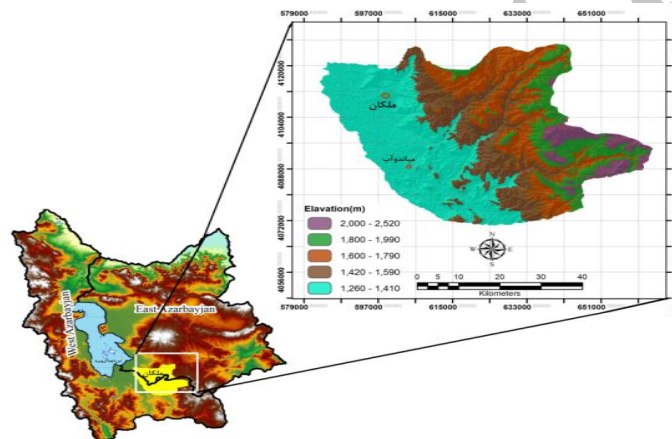
بر اساس مقاطع شکل ۲، آبرفت‌های دوران کواترنری بیشتر به وسیله پوششی از رس شور با مقاومت مخصوص ۳ تا ۱۰ اهم‌متر پوشانده شده است که ضخامت این پوشش گاهی به ۲۰ متر نیز می‌رسد. همچنین، از قسمت شرق آبخوان به سمت غرب، به علت ریزدانه‌شدن رسوبات و افزایش املاح محلول آب‌های زیرزمینی، میزان مقاومت ویژه مواد زمین‌شناسی کاهش می‌یابد.

مقدار این پارامتر به کمک مدل‌های جنگل تصادفی و هوش مصنوعی برآورد شده و سرانجام کارایی این مدل‌ها در امر پیش‌بینی ارزیابی و مقایسه شده است.

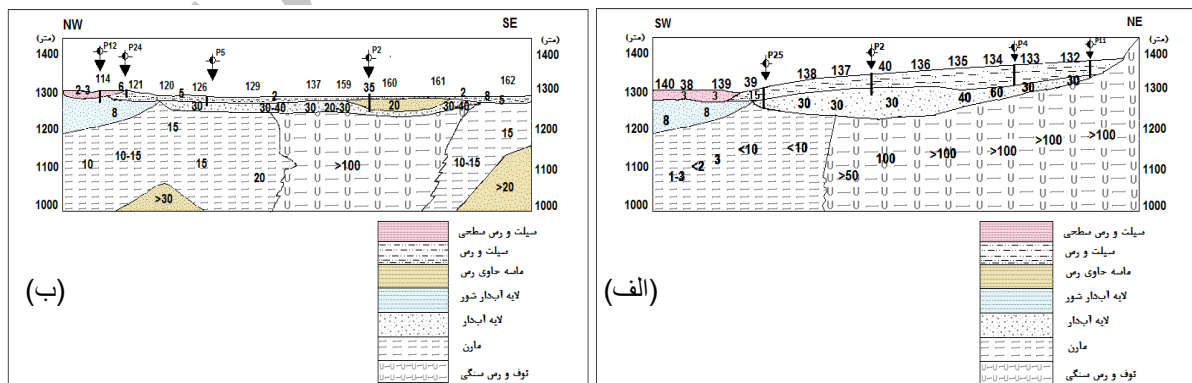
## مواد و روش‌ها

### منطقه مطالعه‌شده

دشت ملکان در جنوب استان آذربایجان شرقی و جنوب شرقی دریاچه ارومیه واقع شده است (شکل ۱) و با وسعتی حدود ۴۵۰ کیلومترمربع، اقلیم سرد و نیمه‌خشک دارد. بر اساس مطالعات ژئوفیزیکی انجام‌شده به تعداد ۱۰۲ سونداژ ژئوالکتریک در راستای ۱۶ پروفایل در دشت ملکان، داده‌های مقاومت ظاهری و نقشه هم‌ضخامت رسوبات آبرفتی به دست آمده است. بر اساس نقشه‌های



شکل ۱. موقعیت منطقه مطالعه‌شده



شکل ۲. مقاطع ژئوالکتریک دشت

(الف) در راستای شمال شرق - جنوب غرب؛ (ب) راستای شمال غرب - جنوب شرق

## روش کار

## معرفی مدل‌های هوش مصنوعی

## شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی یک سیستم پردازشگر اطلاعات توده‌ای هستند که به صورت موازی قرار گرفته‌اند و عملکردی شبیه شبکه عصبی مغز انسان دارند [۱۷]. ساختار شبکه‌های عصبی مصنوعی به وسیله الگوی ارتباطی بین نورون‌ها، روش تعیین وزن‌های ارتباطی و تابع انتقال، معرفی می‌شود. این ساختار معمولاً از لایه ورودی، لایه‌های میانی و لایه خروجی تشکیل شده است. لایه ورودی یک لایه انتقال دهنده و وسیله‌ای برای تهیه کردن داده‌هاست. آخرین لایه یا لایه خروجی شامل مقادیر پیش‌بینی شده به وسیله شبکه است و بنابراین خروجی مدل را معرفی می‌کند. لایه‌های میانی یا مخفی، که از نورون‌های پردازشگر تشکیل شده‌اند، محل پردازش داده‌هاست. تعداد لایه‌های مخفی و تعداد نورون‌ها در هر لایه مخفی به طور معمول به وسیله روش آزمون و خطا مشخص می‌شود. نورون‌های لایه‌های مجاور در شبکه به طور کامل با هم در ارتباط‌اند. به طور کلی، شبکه‌های عصبی به دو نوع شبکه‌های پیشرو<sup>۱</sup> (FFN) و برگشتی<sup>۲</sup> (RNN) تقسیم می‌شوند. در یک شبکه پیشرو (FFN)، نورون‌های پردازشگر در لایه‌های مخفی قرار گرفته‌اند و هر شبکه می‌تواند چندین لایه مخفی و هر لایه مخفی هم می‌تواند چندین نورون داشته باشد. در این شبکه‌ها اطلاعات از ورودی به طرف خروجی حرکت می‌کنند. ورودی‌های هر نورون ممکن است از متغیرهای ورودی یا از خروجی نورون‌های دیگر به دست آید. در شبکه‌های عصبی برگشتی (RNN)، داده‌ها بین نورون‌ها در دو جهت جریان می‌یابند از ورودی به خروجی و کاملاً به عکس. این شبکه از خروجی‌ها به عنوان ورودی‌های جدید صحیح استفاده می‌کند، بنابراین به داده‌ها اجازه برگشت به ورودی را می‌دهد. در تحقیق حاضر، از شبکه پرسپترون چندلایه<sup>۳</sup> (MLP)، که یکی از شبکه‌های پیشرو است، استفاده شده است. در این شبکه‌ها هر نورون یک تابع محرک دارد و هر ورودی  $P(p_1, p_2, \dots, p_n)$  توسط یک بردار وزن  $W(w_1, w_2, \dots, w_n)$  به نورون مد نظر مرتبط می‌شود (شکل ۳). خروجی هر نورون که  $a$  نامیده می‌شود طبق رابطه ۱ محاسبه می‌شود:

$$a = f(P \cdot w + b) \quad (1)$$

در این رابطه؛  $f$ ، تابع فعال‌سازی<sup>۴</sup> و  $b$  مقدار بایاس<sup>۵</sup> هستند.

بنابراین، خروجی یک نورون در یک لایه به سیگنالی که از لایه قبلی دریافت می‌کند، وزن تعیین شده برای آن و نوع تابع تبدیل کننده بستگی دارد. نورون‌ها می‌توانند از توابع فعال‌سازی متفاوتی برای تولید خروجی استفاده کنند. رایج‌ترین آنها می‌توان به توابع لگاریتم سیگموئیدی، تانژانت سیگموئیدی و تابع فعال‌سازی خطی اشاره کرد [۱۸]. روش‌های آموزش در شبکه‌های پرسپترون چندلایه برای آموزش شبکه‌های عصبی توسط الگوریتم‌های ریاضی متفاوت صورت می‌گیرد که از جمله مهم‌ترین آنها الگوریتم انتشار خطا به عقب<sup>۶</sup> (BP)، الگوریتم گرادیان مزدوج<sup>۷</sup> (CG)، الگوریتم لونیبرگ-مارکوارت<sup>۸</sup> (LM) و غیره هستند. از این میان، الگوریتم لونیبرگ-مارکوارت به عنوان کاراترین و سریع‌ترین روش آموزش شبکه‌های عصبی شناخته شده است که جزئیات آموزش و فرایند محاسبه الگوریتم لونیبرگ-مارکوارت توسط انجمن مهندسان عمران آمریکا [۱۹] ارائه شده است. در تحقیق حاضر برای اجرای مدل شبکه‌های عصبی مصنوعی، داده‌های ورودی و هدف به دو دسته آموزش و آزمایش تقسیم شد. به طوری که ۷۰ درصد داده‌ها برای مرحله آموزش و ۳۰ درصد داده‌ها برای مرحله آزمایش در نظر گرفته شده است.

## منطق فازی (FL)

منطق فازی نگرشی چندارزشی به وقایع و رویدادها دارد که این خلاف منطق صریح یا دو ارزشی است که در آن همه چیز یا درست است یا درست نیست. در حقیقت، روش فازی روشی مناسب برای کاهش خطای برآوردی و انسانی نسبت به تئوری‌های قابل اعتماد دیگر است [۲۰ و ۲۱]. داده‌های هیدروژئولوژیکی معمولاً خطای برآوردی ذاتی دارند و داده‌هایی صریح و بدون خطا به حساب نمی‌آید. برای مثال، به دست آوردن قابلیت انتقال از آزمایش پمپاژ خطایی درخور توجه دارد. تابع عضویت تابعی است که توسط آن داده‌های ورودی فازی می‌شوند. یعنی هر ورودی

4. Activation Function

5. Bias

6. Back Propagation Algorithm

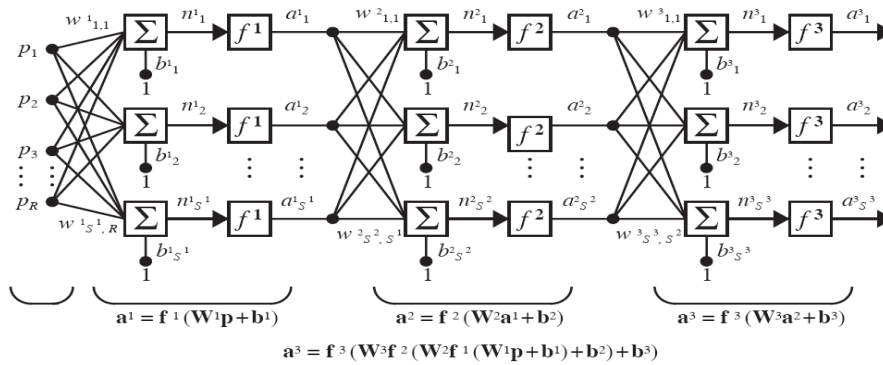
7. Conjugate Gradient Algorithm

8. Levenberg-Marquard

1. Feedforward Network

2. Recurrent / Feedback Network

3. Multilayer perceptron



شکل ۳. ساختار شبکه پرسپترون چندلایه (MLP)

### مدل جنگل تصادفی

یکی از ابزارهای کارآمد در مسائل مربوط به تخمین متغیرهای هدف و یا طبقه‌بندی الگوها، درخت تصمیم است. جنگل تصادفی (RF) یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیش‌بینی مکرر از هر پدیده ترکیب می‌کند. به‌طور کلی، درخت تصمیم منفرد مستعد برازش اضافی<sup>۳</sup> بوده و قدرت تعمیم‌پذیری کمی دارد. هنگام تشکیل یک درخت تصمیم، تغییرات کم در الگوهای یادگیری می‌تواند سبب تغییرات اساسی در ساختار آن درخت شود [۲۲]. RF می‌تواند الگوهای پیچیده را یاد بگیرد و ارتباط غیرخطی بین متغیرهای توضیحی و متغیرهای وابسته را در نظر بگیرد. همچنین، می‌تواند انواع مختلف داده‌ها را به علت نبود توزیع نرمال داده‌های استفاده‌شده، در تجزیه و تحلیل بگنجانند و ترکیب کند. ترکیب درخت‌های تصمیم را روش‌های دسته‌جمعی می‌گویند. یک مجموعه دسته‌جمعی شامل تعدادی یادگیرنده است که به آنها یادگیرنده‌های پایه گفته می‌شود. یادگیرنده‌های پایه معمولاً از داده‌های آموزشی به وسیله یک الگوریتم یادگیری پایه ساخته می‌شوند که این الگوریتم یادگیری پایه می‌تواند درخت تصمیم، شبکه عصبی یا الگوریتم‌های یادگیری دیگر باشد. قابلیت تعمیم یک مجموعه، اغلب قوی‌تر از یادگیرنده‌های پایه است و در واقع روش‌های مجموعه‌ای بیشتر به دلیل داشتن توانایی در تقویت یادگیرنده‌های ضعیف، قابل قبول‌اند [۲۳]. در تحقیق حاضر برای فائق آمدن بر مشکلات یادگیرنده‌های پایه، الگوریتم جنگل تصادفی، که یک روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است، پیشنهاد شده است.

به سیستم فازی به عددی در فاصله صفر تا یک تبدیل می‌شود. توابع عضویت هم برای داده‌های خروجی و هم داده‌های ورودی تعریف می‌شوند. انواع مختلفی توابع عضویت وجود دارد که از جمله آنها، تابع عضویت سیگموئید، گوسی، دابل گوسی، مثلثی، ذوزنقه‌ای و غیره را می‌توان نام برد. به‌طور کلی، در منطق فازی دو بحث مدل‌سازی و دسته‌بندی فازی مطرح است که از مدل‌سازی فازی برای تخمین پارامترهای عددی مانند تخلخل، هدایت هیدرولیکی، تراوایی و قابلیت انتقال استفاده می‌شود. مدل‌سازی فازی به سه روش فازی ممدانی (MFL)، تاگای - سوگنو - کانگ که به‌طور خلاصه روش سوگنو (SFL) نامیده می‌شود و لارسن (LFL) قابل اجراست. تفاوت روش سوگنو با دو روش دیگر در خروجی آنهاست، در روش سوگنو تابع عضویت داده‌های خروجی سیستم فازی به صورت رابطه خطی یا ثابت است که توسط روش دسته‌بندی به دست می‌آید. اولین مرحله در ایجاد یک مدل فازی دسته‌بندی داده‌هاست که بسته به نوع مدل فازی استفاده‌شده می‌توان از روش‌های دسته‌بندی مختلف استفاده کرد. از روش دسته‌بندی کاهشی<sup>۱</sup> برای مدل سوگنو و از روش دسته‌بندی میان مرکز<sup>۲</sup> (FCM) برای روش ممدانی و لارسن استفاده می‌شود. هر مدل فازی شامل سه مرحله اصلی است: الف) فازی کردن داده‌ها: که این عمل با تعریف تابع عضویت انجام می‌شود، ب) ایجاد ارتباط بین ورودی و خروجی: که این کار نیز با یک سری قوانینی مانند اگر-آن‌گاه (if-then) انجام می‌شود و ج) مرحله آخر که مرحله بررسی سیستم، تجمیع نتایج و غیرفازی‌سازی است. در مطالعه حاضر از روش فازی سوگنو استفاده شده است.

3. Over fitting

1. Subtractive  
2. Fuzzy C-Means

## رگرسیون و الگوریتم جنگل تصادفی

در الگوریتم RF برای تشکیل هر درخت، دسته متفاوتی از الگوهای موجود با در نظر گرفتن جایگزینی دوباره هر الگوی منتخب، انتخاب می‌شوند. اندازه این دسته نمونه‌برداری شده، برابر تعداد کل الگوهای موجود خواهد بود. RF در سال ۲۰۰۱ توسط بریمن به عنوان یک روشی که از توسعه جدید درخت‌های تصمیم‌گیری ایجاد شده، ارائه شد که پیش‌بینی چندین الگوریتم منفرد را با هم با استفاده از قوانین مبتنی ترکیب می‌کند. اصول کلی تکنیک‌های آموزش گروهی بر پایه این فرض است که دقت آنها از دیگر الگوریتم‌های آموزشی بیشتر است چون ترکیبی از چند مدل پیش‌بینی، دقیق‌تر از یک مدل است و گروه‌ها قدرت مجموعه‌های منفرد و منحصر به فرد از طبقه‌ها را بیشتر می‌کنند در حالی که ضعف‌های طبقه‌ها را در همان زمان کاهش می‌دهند [۲۴].

درخت رگرسیونی<sup>۱</sup> (RT) مجموعه‌ای از شرایط یا محدودیت‌ها را بیان می‌کند که به صورت سلسله‌مراتبی سازمان یافته‌اند و به حالت متوالی از گره ریشه به سمت پایین رشد می‌کنند و به گره‌های پایانی یا گره‌های برگ می‌رسند [۲۵ و ۲۶]. به منظور به وجود آوردن درخت رگرسیونی از پارتیشن‌بندی بازگشتی و رگرسیون‌های چندگانه استفاده می‌شود. فرایند تصمیم از گره ریشه، در هر گره داخلی، طبق قانون درختی تا زمانی که شرط توقف قبلی تعیین شده به دست آید، تکرار می‌شود. هریک از گره‌های نهایی یا برگ‌ها به یک مدل رگرسیونی ساده که فقط در گره به کار برده می‌شوند، متصل می‌شود. زمانی که فرایند فراخوانی درخت به پایان برسد، هرس کردن یا پرونینگ<sup>۲</sup> با هدف بهبود ظرفیت تعمیم درخت‌ها به وسیله کاهش پیچیدگی ساختار به کار برده می‌شود و تعداد موردها یا نمونه‌ها در گره‌ها می‌تواند به عنوان معیار پرونینگ در نظر گرفته شود. برای جلوگیری از تطابق RTS‌های مختلف، RF تنوع درختان را از طریق درست کردن زیرمجموعه‌های مختلف از داده‌های آموزشی کم می‌کند که به اصطلاح کیسه‌بندی<sup>۳</sup> نامیده می‌شود. اصطلاح Bagging از مخفف Bootstrap Aggregating به دست آمده است [۲۷].

تکنیکی است که از طریق نمونه‌برداری دوباره تصادفی از مجموعه داده‌های اصلی و همراه با جایگزینی، برای ایجاد داده‌های

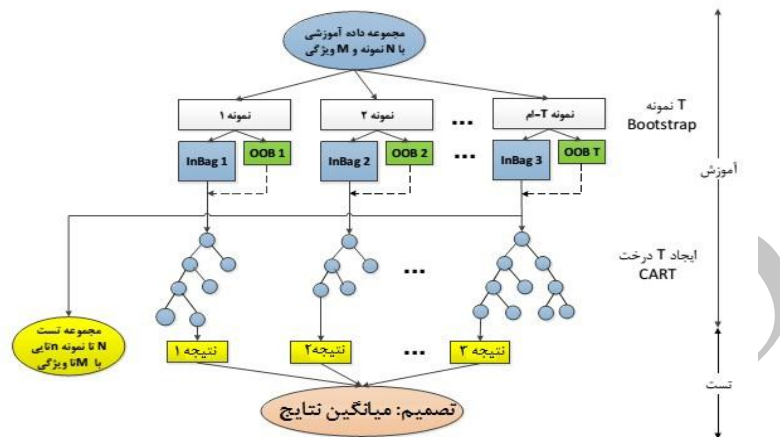
آموزشی استفاده می‌شود و در این مرحله هیچ‌یک از داده‌های انتخاب شده از نمونه‌های ورودی را برای تولید زیرمجموعه بعدی، حذف نمی‌کند و به این ترتیب واریانس نیز کاهش می‌یابد. از این رو، برخی از داده‌ها ممکن است بیش از یکبار در شاخه‌های آموزشی استفاده شود در حالی که برخی از داده‌های دیگر، که در مدل‌سازی مؤثر نیستند، ممکن است هرگز استفاده نشوند. بنابراین، ثبات بیشتری برای مدل به دست می‌آید و مدل را در برابر تغییرات جزئی در داده‌های ورودی قابل اعتمادتر می‌کند و دقت پیش‌بینی آن را افزایش می‌دهد [۱۴]. از سوی دیگر، هنگامی که RF یک درخت رشد ایجاد می‌کند از بهترین متغیرها یا نقاط تقسیم در داخل زیرمجموعه‌های متغیرها استفاده می‌کند که به صورت تصادفی از بین مجموعه‌های کلی متغیرهای ورودی انتخاب می‌کند. در نتیجه، قدرت هر RT منفرد را کاهش می‌دهد و میزان تطابق را کم می‌کند و به این صورت خطای کلی مدل را کاهش می‌دهد [۱۴]. یکی دیگر از ویژگی‌های خوب RF این است که درختان RF، بدون پرونینگ یا هرس کردن رشد می‌کنند و در این روش، آموزش بیش از اندازه بر دقت مدل تأثیری نمی‌گذارد. علاوه بر این، آن دسته از نمونه‌هایی که در آموزش درختان در فرایند Bagging انتخاب نمی‌شوند شامل بخشی از زیرمجموعه‌هایی می‌شوند که الگوهای خارج از کیسه<sup>۴</sup> (OOB) نامیده می‌شوند. این قسمت در روش RF می‌تواند برای ارزیابی عملکرد مدل استفاده شود [۲۷]. به این ترتیب، RF می‌تواند تخمین غیر مرتبط داخلی از خطای تعمیم را محاسبه کند بدون اینکه از زیرمجموعه‌های داده‌های خارجی استفاده کند [۱۴]. روند کلی الگوریتم RF به صورت ساده در شکل ۴ نشان داده شده است. اگر بخواهیم روش RF برای طبقه‌بندی را به طور خلاصه بیان کنیم، به این صورت است که ابتدا T نمونه خودراه‌انداز<sup>۵</sup> از داده آموزشی بیرون کشیده می‌شود و سپس از هر نمونه خودراه‌انداز  $\beta$  یک درخت طبقه‌بندی و رگرسیونی<sup>۶</sup> (CART) هرس نشده ایجاد می‌شود که برای انشعاب در هر گره CART، فقط یک متغیر انتخاب شده به صورت تصادفی استفاده می‌شود. در نهایت، خروجی طبقه‌بندی بر اساس یک نتیجه میانگین، از پیش‌بینی‌های همه درخت‌های منفرد آموزش دیده به دست می‌آید. یک

4. Out of Bag  
5. Bootstrap  
6. Classification and Regression Tree

1. Regression Tree  
2. Pruning  
3. Bagging

می‌برد. درخور یادآوری است که الگوریتم RF به دلیل ارزیابی درونی نتایج هر طبقه‌بندی درختی که در داخل خود دارد و با وزن‌دهی به نتایج هر درخت، می‌تواند نتایج صحیحی را تولید کند.

مجموعه داده خودراهناداز، مجموعه‌ای از نقاط انتخابی به‌طور تصادفی است که از مجموعه داده آموزشی بیرون کشیده شده است. برای اینکه همیشه اندازه نمونه آموزشی اولیه ثابت بماند، مجموعه داده خودراهناداز یک نسخه کپی از نقاط را به کار



شکل ۴. فلوچارت مدل جنگل تصادفی برای رگرسیون

معمول‌ترین روش آن، روش آماری چندمتغیره است که به دلیل کاهش ابعاد داده‌ها، مؤلفه‌های اصلی جایگزین پارامترهای اولیه می‌شوند [۳۱]. در مطالعه حاضر روش RF، که بر اساس درخت‌های گروهی است، ترجیح داده شده است. اولین مرحله در ایجاد مدل پیش‌بینی قابلیت انتقال به وسیله مدل جنگل تصادفی، تنظیم پارامترهای مدل است. به‌منظور تنظیم تعداد درخت‌ها ( $k$ ) به طوری که مقدار خطا همگرا شود و تخمین قابل اعتمادتر باشد، مدلی از یک تا ۱۰۰۰ درخت ساخته شد. با توجه به اینکه با افزایش درخت‌ها میزان خطا کاهش می‌یابد، تعداد ۱۰۰۰ درخت برای فراخوانی مدل استفاده شد. پارامتر  $m$  نیز به وسیله تغییر تعداد متغیرها بین یک و ماکزیمم متغیرهای هر زیرمجموعه، بهینه شد و نتایج مدل به وسیله برآورد خطای OOB ارزیابی شد. علاوه بر این، برای کاهش ابعاد و افزایش دقت و قابل تفسیر بودن مدل، روش FS به کار برده شد و متغیرهای مهم در پیش‌بینی نیز شناسایی شدند. برای فراخوانی مدل پیش‌بینی قابلیت انتقال، داده‌های به‌دست‌آمده از لایه‌های رستری هشت تا متغیر که در بخش قبل توضیح داده شد، برای هر پیکسل از لایه‌های رستری یک مقدار استخراج شد و این مقادیر همراه با متغیرهای هدف به عنوان ورودی مدل استفاده شدند. متغیرهای توضیحی (پیش‌بینی‌کننده) و متغیر پاسخ در مجموعه‌ای از

گرچه اطلاعات زیاد برای مدل‌سازی ممکن است مفید باشند، افزایش تعداد پارامترهای ورودی پیچیدگی‌های اضافی و افزایش زمان محاسبات و مشکلات ابعادی را به سیستم تحمیل می‌کند [۲۸]. برای رسیدن به یک تقسیم بهینه، تصمیم‌ساز<sup>۱</sup> می‌تواند به وسیله مقدار بسیار زیادی از داده‌های مختلف، اطلاعات و شرایطی که در منطقه موجود هستند به سرعت تحت‌الشعاع قرار بگیرد. تعداد زیادی از متغیرهای وابسته به خواص و رفتار سیستم آب زیرزمینی و نیروهای راندگی، می‌توانند توانایی مدل را افزایش دهند. ابعاد زیاد در مجموعه داده‌ها می‌تواند سبب کاهش دقت مدل شود و بنابراین برای جلوگیری از این خطاها و ابعاد زیاد داده‌ها، انتخاب ویژگی<sup>۲</sup> (FS) به کار برده می‌شود. FS یک روش برای انتخاب زیرمجموعه‌های پارامترهای مربوط به آن برای آموزش بهتر مدل است [۲۹]. در مطالعات آب زیرزمینی، از تعداد زیادی متغیرهای وابسته به خصوصیات فیزیکی و شیمیایی آبخوان استفاده می‌شود که برخی از آنها می‌توانند مرتبط و یا غیرمرتبط باشند [۳۰]. FS با زیاد کردن سرعت فرایند آموزش، افزایش قابلیت تعمیم، کاهش اثر از بین رفتن ابعاد و افزایش قابلیت تفسیر، دقت مدل‌های پیش‌بینی را افزایش می‌دهد. روش‌های زیادی برای FS بیان شده است، ولی

1. Decision-Maker  
2. Feature Selection

نورون‌ها در لایه مخفی به منظور کمینه کردن خطا و بیشینه کردن دقت مدل، بهینه‌سازی می‌شوند، بنابراین با استفاده از روش آزمون و خطا، تعداد نورون‌ها در لایه میانی و خروجی به ترتیب سه و یک به دست آمد. تابع سیگموئید تانژانتی هیپربولیک (Tansig)، که به عنوان بهترین تابع برای مدل‌سازی سیستم‌های طبیعی غیرخطی تشخیص داده شده‌اند، به عنوان تابع تبدیل‌کننده نورون‌های پردازشگر در لایه دوم و تابع خطی (Purlin) در لایه سوم استفاده شد. توابع انتقال با آزمون و خطا به گونه‌ای انتخاب شد که میانگین مربعات خطا (MSE) در دو مجموعه آموزش و آزمایش کمینه شد، مقادیر ضریب همبستگی و میانگین مربعات خطا برای مرحله آموزش و آزمایش در جدول ۲ ارائه شده است. در نهایت، به منظور پیش‌بینی قابلیت انتقال هیدرولیکی به کمک شبکه‌های عصبی مصنوعی، شبکه چندلایه پرسپترون با الگوریتم آموزشی لونبرگ-مارکوارت به عنوان کاراترین شبکه شناخته شد. به طوری که با کمترین MSE و بیشترین مقدار R2 به عنوان بهترین مدل شبکه عصبی مصنوعی شناخته شده در صورتی که سایر ساختارهای شبکه عصبی مصنوعی به‌طور میانگین MSE و R2 بیشتری داشتند.

### مدل فازی

نخستین مرحله در این روش دسته‌بندی داده‌ها و تعیین توابع عضویت است که در اینجا از روش کاهشی، که روشی کارا و مفید در دسته‌بندی داده‌ها، تعیین توابع عضویت و قوانین در تحقیقات اخیر بوده است، استفاده شد. در این سیستم استنتاج فازی از یک سری قوانین اگر-آن‌گاه فازی برای برقراری ارتباط بین توابع عضویت داده‌های ورودی و توابع عضویت داده‌های خروجی استفاده شده است. پارامتر اصلی در این روش دسته‌بندی برای تعیین دسته‌ها و قوانین اگر-آن‌گاه شعاع دسته‌بندی است. توابع عضویت با استفاده از دسته‌بندی کاهشی توسط دستور  $\text{genfis2}(\text{datin}, \text{datout}, r)$  تعیین شد که در آن،  $\text{datin}$  داده‌های ورودی،  $\text{datout}$  داده‌های خروجی و  $r$  شعاع دسته‌بندی است که بین صفر تا یک متغیر است. تعریف مقدار بزرگ شعاع دسته‌بندی به تولید تعداد کمی دسته منجر می‌شود و تعریف مقادیر کوچک، تعداد زیادی دسته کوچک تولید خواهد کرد. هر یک از دسته‌ها، نماینده یک تابع عضویت است و هر تابع عضویت نیز نشان‌دهنده یک مجموعه فازی در داده‌ها است که برای استخراج قوانین به کار می‌رود.

بردارهای ویژگی ورودی ترکیب شدند. این بردارها که ورودی الگوریتم RF را تشکیل می‌دهند، به عنوان بردارهای Input-feature شناخته شده است. متغیر پاسخ دوتایی (متغیر توضیحی و متغیر پاسخ) به عنوان مقادیر هدف برای آموزش الگوریتم استفاده شدند.

### داده‌های ورودی مدل‌ها

بر اساس مطالعات ژئوفیزیک انجام‌شده در این دشت داده‌های مقاومت الکتریکی ظاهری و نقشه هم ضخامت رسوبات آبرفتی در دسترس است. مقاومت الکتریکی ظاهری دارای دو مؤلفه اصلی، مقاومت الکتریکی عرضی<sup>۱</sup> ( $R_t$ ) و مقاومت الکتریکی طولی<sup>۲</sup> ( $R_L$ ) است که به ترتیب تغییرات پارامترهای هیدروژئولوژیکی (مانند نفوذپذیری، کیفیت آب و غیره...) عمقی و جانبی را نشان می‌دهد. بنابراین، مقاومت عرضی با مقادیر قابلیت انتقال مرتبط است [۳۲] و از طرفی، با توجه به ارتباط مقادیر مقاومت عرضی با هدایت الکتریکی (EC) می‌توان هدایت الکتریکی را نیز از پارامترهای مرتبط با مقادیر قابلیت انتقال دانست [۳۳]. بنابراین، ورودهای مدل به کاررفته عبارت‌اند از: مقاومت عرضی ( $R_t$ )، هدایت الکتریکی (EC)، ضخامت آبخوان (B)، هدایت هیدرولیکی (K)، گرادیان هیدرولیکی (H)، محیط آبخوان (A)، محیط خاک و شیب. خصوصیات آماری پارامترهای ورودی در جدول ۱ ارائه شده است.

### نتایج و بحث

#### شبکه‌های عصبی مصنوعی

مرحله نخست مدل‌سازی توسط شبکه‌های عصبی مصنوعی مربوط به شناخت بهترین ساختار شبکه‌های عصبی مصنوعی و طراحی شبکه برای برآورد پارامتر قابلیت انتقال هیدرولیکی در منطقه مطالعاتی است. به این منظور، ساختارهای شبکه‌های عصبی مصنوعی که متشکل از شبکه‌های (FFN) و (RNN) است با الگوریتم‌های مختلف از جمله الگوریتم‌های انتشار خطا به عقب (BP)، الگوریتم گرادیان مزدوج (CG) و الگوریتم لونبرگ-مارکوارت (LM) بررسی و مطالعه شدند. پارامترهای مقاومت عرضی ( $R_t$ )، هدایت الکتریکی (EC)، ضخامت آبخوان (B) و هدایت هیدرولیکی (K) به عنوان لایه‌های ورودی شبکه استفاده شدند. با توجه به اینکه تعداد

1. Transversal resistance  
2. Longitudinal resistance



مدل سوگنو تابعی است خطی که بر اساس ورودی‌ها ساخته می‌شود. جدول ۳ نتایج ارزیابی عملکرد مدل‌سازی فازی در مراحل آموزش و آزمایش را نشان می‌دهد. بر اساس نتایج به دست آمده مدل شبکه عصبی مصنوعی با الگوریتم لوبنبرگ-مارکواریت و مدل منطق فازی سوگنو در بین روش‌های هوش مصنوعی بیشترین بازده، کمترین درصد خطا و توانایی زیادی در پیش‌بینی پارامترهای هیدروژنولوژیکی این دشت دارند.

کاهش این پارامتر سبب افزایش دسته‌ها و قوانین می‌شود و افزایش آن به سمت یک موجب کاهش دسته‌ها و قوانین خواهد بود. بازده مدل فازی بستگی به تعیین شعاع بهینه دسته‌بندی دارد که در برآورد قابلیت انتقال مقدار بهینه آن ۰/۹ تعیین شد. توابع عضویت استفاده شده برای مدل‌سازی فازی قابلیت انتقال، تابع عضویت گوسی بود که به داده‌های دسته‌بندی شده برازش داده شد و نیز تابع عضویت خروجی

جدول ۱. خصوصیات آماری پارامترهای ورودی مدل شبکه عصبی مصنوعی

پارامتر	EC	B	Rt	K	T	H
ماکزیمم	۳۰۱۰۶/۱	۶۹/۰	۲۵۲/۴	۷۱/۹	۸۷۵/۶	۰/۰۰۷۸
مینیمم	۷۲۰/۵	۸/۲	۰/۳	۰/۵	۶/۷	۰/۰۰۱۵
میانگین	۴۵۴۶/۷	۲۴/۶	۱۹/۲	۲۱/۲	۳۸۰/۱	۰/۰۰۳۳
واریانس	۲۳۳۸۴۰۰۱	۵۴/۲	۸۲۲/۸	۶۴/۸	۲۵۵۴۲/۴	۱/۵۷E-۰۶
متوسط انحراف معیار	۳۱۹۳/۷	۵/۳	۱۶/۶	۵/۹	۱۳۰/۰	۰/۰۰۱۲

جدول ۲. نتایج ارزیابی عملکرد شبکه عصبی مصنوعی با ساختارهای مختلف در مراحل آموزش و آزمایش

نوع شبکه	FFN						RNN					
	LM		GC		BP		LM		GC		BP	
	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE
الگوریتم آموزش												
معیار ارزیابی												
مرحله آموزش	۰/۹۲۴	۰/۰۴۴	۰/۹۱۴	۰/۰۵۱	۰/۸۲۱	۰/۵۶۴	۰/۸۱۳	۰/۱۶۲	۰/۶۵۱	۰/۷۵۹	۰/۶۸۲	۰/۴۸۶
مرحله آزمایش	۰/۹۳۱	۰/۰۶۵	۰/۸۷۲	۰/۰۸۲	۰/۷۸۷	۰/۶۵۵	۰/۷۸۴	۰/۲۰۱	۰/۶۲۴	۰/۸۷۵	۰/۶۴۳	۰/۵۷۱

جدول ۳. نتایج ارزیابی عملکرد مدل منطق فازی در مراحل آموزش و آزمایش

معیار ارزیابی	R2	MSE
مرحله آموزش	۰/۸۱۵	۰/۰۶۸
مرحله آزمایش	۰/۹۰۶	۰/۰۹۷

جنس رسوبات محیط آبخوان در این قسمت از منطقه است. جدول ۴ سهم هر متغیر<sup>۱</sup> توضیحی، به همراه امتیاز<sup>۲</sup> پیش‌بینی شده برای هر یک از متغیرها را در مدل نشان می‌دهد. بر اساس پیش‌بینی مدل RF، پارامترهای هدایت الکتریکی، محیط آبخوان، گرادبان هیدرولیکی و مقاومت عرضی آبخوان به ترتیب بیشترین تأثیر را در پیش‌بینی پارامتر قابلیت انتقال دارند. یک مدل پیش‌بینی بر اساس سه اصل دقت مدل پیش‌بینی، تعداد متغیرهای لازم و در دسترس بودن آنها و دقت داده‌ها پایدار خواهد بود. مدل RF با توجه به مقدار خطای OOB، پیش‌بینی را انجام می‌دهد و در این روش از داده‌ها برای تست استفاده نمی‌شود. شکل ۶ الف نتایج پیش‌بینی مدل RF (داده‌های محاسباتی) را در مقابل

### مدل جنگل تصادفی

شکل ۵ لایه‌های مهم رستری متغیرهای توضیحی استفاده شده در روش RF را نشان می‌دهد در لایه‌های ایجاد شده توزیع مکانی متغیرهای توضیحی مشاهده می‌شود. با توجه به اینکه آب‌های زیرزمینی قسمت غربی دشت در عمق کم حدود یک تا چهار متری قرار دارد. از طرف دیگر، وجود سازندهای ریزدانه با نفوذپذیری بسیار کم که سبب می‌شود آب بر اثر نیروی کاپیلاری بالا بیاید و در عمق کم واقع شود، همه این‌ها در نهایت موجب تبخیر شدید و افزایش شوری آب زیرزمینی در این قسمت از منطقه می‌شوند [۳۴]. مقاومت عرضی در قسمت‌های شرقی به خصوص در منطقه اطراف شهرستان ملکان و همچنین در قسمت‌های جنوب شرقی منطقه زیاد است که زیاد بودن آن به دلیل بیشتر بودن ضخامت در منطقه مرکزی و شرقی دشت و همچنین به علت

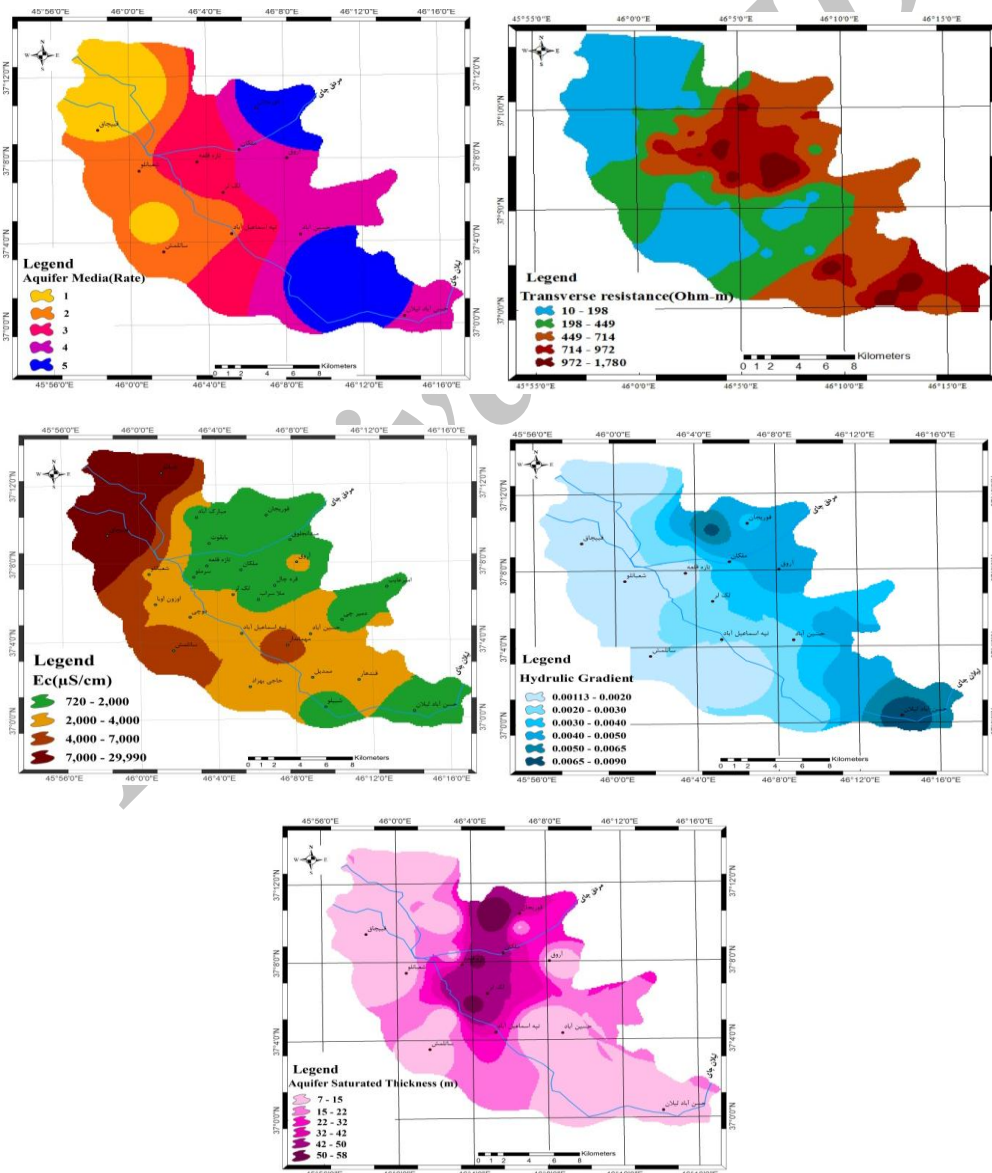
1. Variable importance  
2. Score

برازش اضافی هنگام آموزش با داده‌های دارای نویز بالا وجود دارد، آموزش ببیند [۳۵].

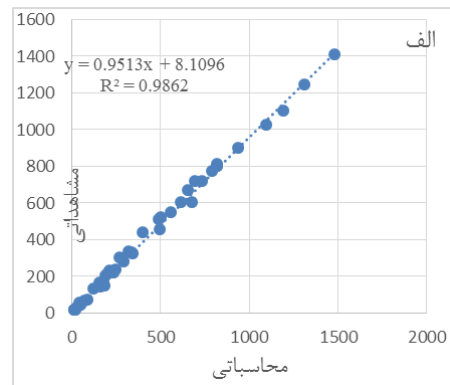
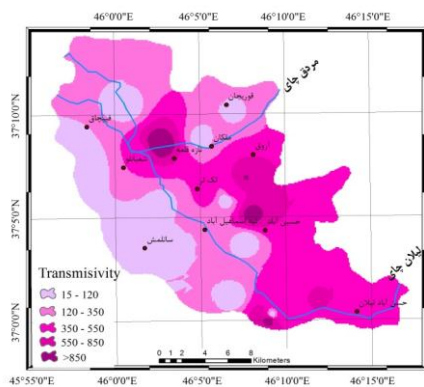
جدول ۴. سهم هر متغیر توضیحی در پیش‌بینی مدل

Variable	Score	Variable Importance
هدایت الکتریکی	۱۰۰	
محیط آبخوان	۷۷/۷۲	
گرادبان هیدرولیکی	۷۷/۳۰	
مقاومت عرضی آبخوان	۷۴/۰۹	
ضخامت اشباع	۲۴/۳۸	
محیط خاک	۲۱	
خاک سطحی	۳/۱۵	
شیب	۲/۶۹	

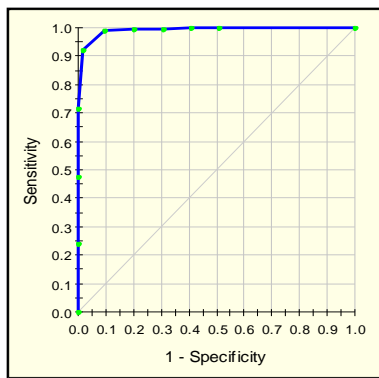
داده‌های مشاهداتی برای پارامتر قابلیت انتقال را نشان می‌دهد. همچنین، توزیع مکانی پارامتر قابلیت انتقال حاصل از مدل RF در شکل ۶- ب نشان داده شده است. مطابق شکل قابلیت انتقال پیش‌بینی‌شده در قسمت‌های ورودی دشت در مناطق شرقی و مرکزی زیاد است و هرچه به سمت خروجی دشت نزدیک می‌شویم، از مقدار قابلیت انتقال کاسته می‌شود. از آنجا که طبقه‌بندی در RF از میانگین‌گیری و شمارش نتایج طبقه‌بندی کننده‌های پایه ساده و متنوع حاصل می‌شود، بنابراین RF قادر است با داده‌های دارای نویز نیز به گونه‌ای صحیح‌تر، در مقایسه با دیگر الگوریتم‌ها از جمله SVM و شبکه‌های عصبی مصنوعی که در آنها احتمال



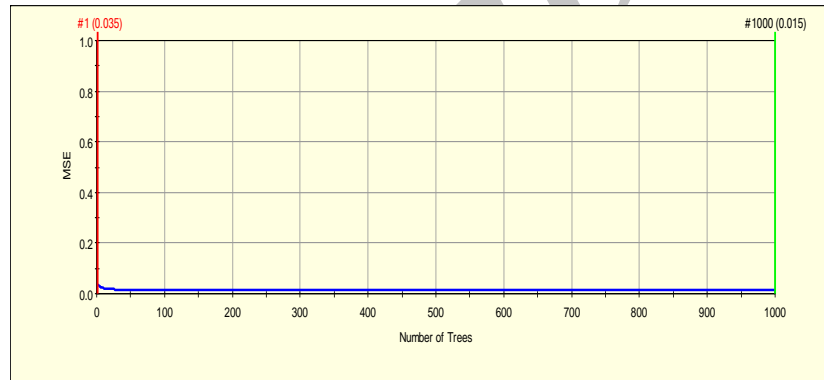
شکل ۵. توزیع مکانی پارامترهای ورودی مدل جنگل تصادفی



شکل ۶. الف) مقایسه نتایج مشاهداتی و محاسباتی پیش‌بینی برای قابلیت انتقال توسط مدل RF برای داده‌های آزمایش، ب) توزیع مقادیر قابلیت انتقال به‌دست‌آمده از مدل RF



ب



الف

شکل ۷. الف) خطای MSE مدل و کاهش آن با افزایش درخت‌های تصمیم؛ ب) منحنی ROC حاصل از مدل پیش‌بینی

در جفت (TPR- FPR) و یک سری از چنین جفت‌هایی برای رسم منحنی‌های ROC استفاده می‌شوند. مساحت زیرمنحنی<sup>۶</sup> (AUC) در منحنی‌های ROC نیز به عنوان یکی از روش‌های ارزیابی خطا به کار برده می‌شود که هرچه میزان AUC به یک نزدیک باشد، مدل صحت بیشتری دارد. مطابق شکل ۷-الف میزان خطای مدل‌سازی با افزایش درخت‌های تصمیم کاهش می‌یابد به گونه‌ای که در درخت تصمیم اولی میزان خطای MSE برابر با ۰/۳۵۶ بوده و هنگامی که تعداد درخت‌های تصمیم ۱۰۰۰ می‌شود، میزان خطا به ۰/۰۱۵ کاهش می‌یابد. همچنین، مقدار AUC برابر با ۰/۹۶ است که نشان‌دهنده دقت زیاد شبیه‌سازی است (شکل ۷-ب). البته، خطای موجود در مدل نهایی به عوامل مختلفی از جمله تعداد محدود داده‌ها، ناهمگنی آبخوان، خطای ذاتی موجود در

پیش‌بینی‌ها به واسطه ترکیب تعداد زیادی از پارامترهای ممکن انجام می‌گیرد و به وسیله میانگین مربعات خطا (MSE) ارزیابی می‌شوند و در نهایت، مدلی که کمترین میزان (MSE) را داشته باشد به عنوان مدل بهینه، برای پیش‌بینی انتخاب می‌شود. روش دیگر تجزیه و تحلیل مدل بر پایه منحنی‌های مشخصه عملکرد سیستم<sup>۱</sup> (ROC) است. منحنی‌های ROC به گونه‌ای مشابه میزان موفقیت هستند که می‌توانند به وسیله میزان مثبت واقعی<sup>۲</sup> (TPR) کنترل شوند. به طور کلی، در منحنی‌های ROC نتایج<sup>۳</sup> FPR (محور X) در مقابل TPR (محور Y) رسم می‌شوند. (TPR) به عنوان حساسیت<sup>۴</sup> و (1 - FPR) به عنوان اختصاصیت<sup>۵</sup> شناخته می‌شوند. هر یک از نتایج

1. Receiver Operating Characteristic
2. True Positive Rate (TPR)
3. False positive rate
4. Sensitivity
5. Specificity

6. Area under Curve

کمک مدل ارائه شده از صرف زمان و هزینه‌های مالی بیشتر جلوگیری کرد.

#### منابع

- [1]. Chow VT. On the determination of transmissibility and storage coefficient from pumping test data. Transactions. American Geophysical Union. 1952; 33(3): 397-404.
- [2]. Cooper H, Jacob C E. A generalized graphical method for evaluation formation constants and summarizing well field history. Transactions, American Geophysical Union. 1946; 27(4): 526-534.
- [3]. Neuman SP. Theory of flow in unconfined aquifers considering delayed response of water table. Journal of Water Resources Research. 1972; 8(4): 1031-1045.
- [4]. Theis CV. The relationship between the lowering of piezometric surface and the rate and duration of discharge of a well using groundwater storage. Transactions, American Geophysical Union. 1935; 16(2): 519-524.
- [5]. Samani N, Gohari-Moghadam M, Safavi AA. A simple neural network model for the determination of aquifer parameters. Journal of Hydrology. 2007; 340(1-2): 1-11.
- [6]. Nadiri AA, Chitsazan N, Frank TC, Moghaddam A. Bayesian Artificial Intelligence Model Averaging for Hydraulic Conductivity Estimation. Journal of Hydrologic Engineering. 2014; 19(3): 520- 532.
- [7]. Chen CH, Lin Z. A committee machine with empirical formulas for permeability prediction. Journal of Computers and Geosciences. 2006; 32: 485-496.
- [8]. Chitsazan N, Nadiri AA, Tsai F. Prediction and structural uncertainty analyses of artificial neural networks using hierarchical bayesian model averaging. Journal of Hydrology. 2015; 528: 52-62.
- [9]. Kadkhodaie A, Amini A. A fuzzy logic approach to estimation hydraulic flow units from well log data: case study from the Ahvaz oilfield in south Iran. Journal of Petroleum Geology. 2009; 32(1): 67-78 67.
- [10]. Kadkhodaie A, Rezaee MR, Rahimpour-Bonab H. A committee neural network for prediction of normalized oil content from well log data: An example from South Pars Gas Field, Persian Gulf. Journal of Petroleum Science and Engineering. 2009a; 65: 23-32.

داده‌های ورودی و حتی داده‌های خروجی بستگی دارد. افزایش دقت داده‌های ژئوفیزیکی و آزمایش پمپاژ در آبخوان‌ها می‌تواند در کاهش خطای موجود در مدل ارائه شده مؤثر باشد.

#### نتیجه‌گیری

با توجه به هزینه‌های زیاد آزمایش‌های صحرایی و عدم قطعیت موجود در پارامترهای هیدروژئولوژیک به خصوص قابلیت انتقال، مدل‌سازی می‌تواند روشی مناسب برای پیش‌بینی و برآورد این پارامترها باشد و در مدیریت و مطالعات منابع آب‌های زیرزمینی استفاده شود. بر اساس نتایج به دست آمده از بین مدل‌های شبکه عصبی مصنوعی با الگوریتم‌های مختلف، مدل شبکه عصبی با الگوریتم لونیگ-مارکوارت بیشترین بازده و کمترین درصد خطا را دارد که نشان‌دهنده قدرت زیاد این روش غیرخطی در تخمین پارامتر مد نظر است. نتایج به دست آمده از مدل شبکه عصبی مصنوعی ارائه شده ( $R^2 = 0/981$  و  $MSE = 0/008$ ) و منطق فازی ( $R^2 = 0/966$  و  $MSE = 0/009$ ) همبستگی زیادی را بین مقادیر مشاهداتی و محاسباتی قابلیت انتقال در آبخوان دشت ملکان نشان دادند. به طور کلی، روش RF با داشتن مزایایی مثل یادگیری روابط غیرخطی و داشتن حساسیت کمتر در برابر گیرافتادن در کمینه محلی، به عنوان یک روش دقیق در مدل‌سازی پیش‌بینی قابلیت انتقال شناخته شد. نتایج صحیح مدل RF در پیش‌بینی قابلیت انتقال و همچنین تعیین پارامترهای مهم در پیش‌بینی توسط این مدل، نشان از برتری مدل ارائه شده نسبت سایر مدل‌ها در بحث پیش‌بینی، است. نتایج به دست آمده از مدل جنگل تصادفی ( $R^2 = 0/96$ ،  $AUC = 0/001$  و  $MSE = 0/986$ ) در پیش‌بینی پارامتر قابلیت انتقال، مقادیر خطای کمتر و همبستگی بیشتری را نسبت به دو مدل دیگر نشان داد که این بیان‌کننده برتری مدل ارائه شده نسبت به دیگر مدل‌هاست. تعیین مهم‌ترین پارامترهای تأثیرگذار در پیش‌بینی قابلیت انتقال، از دیگر مزایای مدل جنگل تصادفی در بحث پیش‌بینی است. این پژوهش از نظر روش‌شناسی می‌تواند راه‌گشای بسیاری از مشکلات موجود در پیش‌بینی پارامترهای هیدروژئولوژیک آبخوان‌ها باشد و با توجه به هزینه‌های زیاد آزمایش‌های صحرایی در اندازه‌گیری پارامتر قابلیت انتقال، می‌توان به

- [11]. Nadiri AA, Asghari Moghaddam A, Tsai F, Fijani E. Hydrogeochemical analysis for Tasuj plain aquifer, Iran. *Journal of Earth System Science*. 2013; 122(4): 1091-1105.
- [12]. Pulido CI, Gutiérrez JC. Improved irrigation water demand forecasting using a soft computing hybrid model. *Journal of Biosystems Engineering*. 2009; 102(2): 202-218.
- [13]. Rodriguez V, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sánchez J.P. An assessment of the effectiveness of a Random Forest classifier for land-cover classification. *ISPRS Journal of Photogram Remote Sens*. 2012d; 67: 9 -104.
- [14]. Breiman L. Random Forests. *Machine Learning*. 2001; 45(1): 5–32.
- [15]. Yoo W, Ference BA, Cote ML, Schwartz A. A Comparison of Logistic Regression, Logic Regression, Classification Tree, and Random Forests to Identify Effective Gene-Gene and Gene-Environmental Interactions. *International Journal of Applied Science and Technology*. 2012; 2(7): 268-274.
- [16]. Norouzi H, Asghari Mogaddam A, Nadiri AA. Determining vulnerable areas of Malekan Plain Aquifer for Nitrate, Using Random Forest method. *Journal of Environmental Studies*. 2015; 41(4): 923-94. [In Persian]
- [17]. Hopfield JJ. Neural network and physical systems with emergent collective computational abilities. *Proc. Nat. Academy of scientists*. 1982; 79: 2554-2558.
- [18]. Demuth H, Beale M. *Neural Network Toolbox User, s Guide, By the Math Works. Inc Version*. 2000; 4: 840pp.
- [19]. ASCE. Task Committee on Application of Artificial Neural Networks in Hydrology, Part I and II. *Journal of Hydrology*. 2000; 5(2): 115-137.
- [20]. Chiu S. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*. 1994; 2(4): 267–278.
- [21]. Nikravesh M, Aminzadeh F. *Soft Computing and Intelligent Data Analysis in Oil Exploration. Part1: Introduction: Fundamentals of Soft Computing*. Elsevier, Berkeley, USA. 2003; pp.744.
- [22]. Quinlan JR. Induction of decision trees. *Journal of Machine Learning*. 1986; 1(1): 81-106.
- [23]. Schapire R. The strength of weak learnability. *Journal of Machine learning*, 1990; 5:197-227.
- [24]. Kotsiantis S, Pintelas P. Combining bagging and boosting. *International Journal of Computational Intelligence*. 2004; 1(4): 324–33.
- [26]. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and regression trees*, Chapman & Hall/CRC, New York. 1984; pp.744.
- [26]. Quinlan JR. *C4.5 programs for machine learning*. San Mateo, CA: Morgan Kaufmann. 1993; 303 pp.
- [27]. Breiman L. Bagging predictors. *Machine Learning*. 1996; 24(2): 123–40.
- [28]. Bellman R. *Dynamic programming*. Mineola, NY: Dover Publications. 2003; 366 pp.
- [29]. Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Res*. 2003; 3: 1157–82.
- [30]. Dixon BA. Case study using support vector machines, neural networks and logistic regression in a GIS to identify wells contaminated with nitrate-N. *Journal of Hydrogeology*. 2010; 17(6): 1507–20.
- [31]. Critto A, Carlon C, Marcomini A. Characterization of contaminated soil and groundwater surrounding an illegal landfill by principal component analysis and kriging. *Journal of Environmental Pollution*. 2003; 122(2): 235–44.
- [32]. Harb N, Haddad K, Farkh S. Calculation of transverse resistance to correct aquifer resistivity of groundwater saturated zones, implications for estimating its hydrogeological properties. *Lebanese science journal*. 2010; 11(1): 105-115.
- [33]. Valcarce RM, Rodríguez WM. Resolution power of well log geophysics in karst aquifers. *Journal of Environmental Hydrology*. 2004; 12: 1-7.
- [34]. Lehmann P, Davis. Evaporation and capillary coupling across vertical textural contrasts in porous media. *Journal of Phys, Rev*. 2009; 80(4): 44-57
- [35]. Chehata N, Guo L, Mallet C. Airborne lidar feature selection for urban classification using random forests. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2009; 39: 207-12.