

پیش‌بینی قابلیت انتقال آبخوان دشت ملکان با استفاده از روش جنگل تصادفی

حسین نوروزی^۱، عطاالله ندیری^{۲*}، اصغر اصغری مقدم^۳، مریم قره‌خانی^۴

تاریخ دریافت: ۹۴/۰۹/۲۴ تاریخ پذیرش: ۹۵/۱۰/۲۵

۱- کارشناسی ارشد هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز

۲- استادیار هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز

۳- استاد هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز

۴- دانشجوی دکتری هیدروژئولوژی، دانشکده علوم طبیعی، دانشگاه تبریز

* مسئول مکاتبات، پست الکترونیکی: Nadiri@tabrizu.ac.ir

چکیده

قابلیت انتقال یکی از پارامترهای مهم در شناسایی خصوصیات آبخوان‌ها می‌باشد و به‌دست آوردن مقادیر و چگونگی توزیع آن توسط مدل‌سازی، برای مدیریت آبخوان‌ها ضروری به‌نظر می‌رسد. برآورد این پارامتر با استفاده از آزمایشات صحرایی مانند آزمایش پمپاژ بسیار پرهزینه و وقت‌گیر است. برای مدیریت مناسب آبخوان دشت ملکان که یکی از قطب‌های کشاورزی شمال‌غرب کشور می‌باشد شناخت پارامترهای هیدروژئولوژیکی، خصوصاً قابلیت انتقال امری لازم و ضروری می‌باشد. در مطالعه حاضر روش جنگل تصادفی (RF) که یک روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است، برای پیش‌بینی قابلیت انتقال آبخوان پیشنهاد شده که تاکنون در این زمینه مورد استفاده قرار نگرفته است. روش RF نسبت به روش‌های دیگر دارای مزایایی مانند دقت پیش‌بینی بالا، توانایی در یادگیری روابط غیرخطی، توانایی بالا در تعیین متغیرهای مهم در پیش‌بینی و ماهیت غیر پارامتری می‌باشد. در مدل RF به‌دلیل اینکه با افزایش درخت‌ها میزان خطا کاهش می‌یابد، بنابراین تعداد ۵۰۰ درخت برای فراخوانی مدل استفاده شد و پس از اجرای مدل نتایج مدل به‌وسیله برآورد خطای خارج از کیسه (OOB) ارزیابی گردید و علاوه بر این، برای کاهش ابعاد و افزایش دقت و قابل تفسیر بودن مدل، روش انتخاب ویژگی (FS) به‌کار برده شد و متغیرهای مهم در پیش‌بینی نیز شناسایی گردید. بر اساس نتایج مدل‌سازی با RF مقادیر MSE و AUC به‌ترتیب برابر با ۰/۰۳۶ و ۰/۹۶، متغیرهای هدایت الکتریکی، محیط آبخوان و گرادیان هیدرولیکی به‌ترتیب به‌عنوان پارامترهای مهم و تأثیرگذار در پیش‌بینی قابلیت انتقال شناسایی شدند. همچنین نتایج قابل قبول مدل RF در پیش‌بینی قابلیت انتقال و تعیین پارامترهای مهم در پیش‌بینی، نشان از مزایای مدل ارائه‌شده نسبت به سایر مدل‌ها در بحث پیش‌بینی، می‌باشد.

واژه‌های کلیدی: آبخوان، جنگل تصادفی، درخت تصمیم، دشت ملکان، قابلیت انتقال

Prediction of Transmissivity of Malikan Plain Aquifer Using Random Forest Method

H Norouzi¹, AA Nadiri^{*2}, A Asghari Mogaddam³, M Gharekhani⁴

Received: 15 December 2015 Accepted: 14 January 2017

1- M.Sc. of Hydrogeology, University of Tabriz, Iran

2- Assit. Prof. of Hydrogeology, University of Tabriz, Iran

3- Prof. of Hydrogeology, University of Tabriz, Iran

4- Ph.D Student of Hydrogeology, University of Tabriz, Iran

*Corresponding Author, Email: Nadiri@tabrizu.ac.ir

Abstract

Transmissivity is an important factor in identifying the characteristics of aquifers, so, estimating its value and distribution by modeling is necessary for aquifer management. Estimation of this parameter using field experiments such as pumping test is costly and time-consuming. For suitable management of Malikan plain, as one of active agricultural areas in north-west of the country, understanding the hydrogeological parameters, such as transmissivity is required. In this study the random forest (RF) algorithm, which is a learning method based on ensemble of decision trees, is proposed for predicting transmissivity that has not been used in this field, yet. The RF technique has advantages over other methods due to having high prediction accuracy, ability to learn nonlinear relationships, non-parametric natural and ability to determine the important variables in the prediction process. Increasing the number of trees decrease the error, so 500 trees were selected to reap high efficiency of the model. The model results were evaluated by OOB error estimating method and in addition, to reduce the dimensions, increase the accuracy and better interpretation of the model process, the FS method was used. The most important variables in the prediction were also identified by the FS method. Based on the results of RF modeling with AUC=0.96 and MSE=0.036, electrical conductivity, aquifer media and hydraulic gradient variables were the most important parameters in predicting transmissivity, respectively. Also the accuracy of RF model and determining the important parameters in transmissivity prediction showed the advantages of this model over other models in prediction issue.

Keywords: Aquifer, Decision trees, Malikan plain, Random forest, Transmissivity.

مقدمه

زمان بیان می‌شود (اصغری مقدم ۱۳۸۹). ضریب قابلیت انتقال (T) از حاصل ضرب ضریب نفوذپذیری مواد تشکیل دهنده لایه آبدار در ضخامت لایه آبدار بدست می‌آید (تاد و مایز ۲۰۰۵). این پارامتر یکی از پارامترهای مهم در شناسایی خصوصیات آبخوان‌ها می‌باشد که بدست آوردن

شناخت پارامترهای هیدروژئولوژیکی آبخوان برای پیش‌بینی رفتار آن امری ضروری می‌باشد. یکی از مفیدترین مفاهیم در مطالعه آبخوان‌ها قابلیت انتقال می‌باشد که عبارت است از مقدار جریان آبی که تحت تأثیر شیب هیدرولیکی واحد از مقطعی با عرض واحد و کل ضخامت سفره عبور می‌کند که بر حسب مربع طول بر

¹.Transmissivity

سالیان اخیر، ابزارهای یادگیری^۲ و روش‌های جدیدی برای حل برخی از مشکلات فوق‌الذکر ارائه شده است و به طور گسترده‌ای مورد استفاده قرار گرفته است. این روش‌های یادگیری جدید، از رگرسیون‌های گروهی در حال پیدایش بهره گرفته‌اند. یکی از انواع روش‌های یادگیری که از الگوریتم‌های پایه برای پیش‌بینی چندگانه تکراری^۳ استفاده می‌کند، جنگل تصادفی^۴ نامیده می‌شود (فریدل و همکاران ۱۹۹۹، برایمن ۲۰۰۱). این روش در طبقه‌بندی پوشش زمین و نیز در مسائل الکترونیک و پزشکی کاربرد فراوان دارد (پال ۲۰۰۵، کوپر و همکاران ۲۰۱۱). همچنین روش RF (پتانسیل لازم به عنوان یک ابزار مدل مکانی برای ارزیابی مسائل زیست محیطی و منابع آب را دارا می‌باشد (بوکر و اسلندر ۲۰۱۲). RF یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیش‌بینی مکرر از هر پدیده ترکیب می‌کند. جنگل تصادفی می‌تواند الگوهای پیچیده را یاد بگیرد و ارتباط غیر خطی بین متغیرهای توضیحی و متغیرهای وابسته را در نظر بگیرد. همچنین می‌تواند انواع مختلف داده‌ها را در تجزیه و تحلیل گنجانده و ترکیب کند که این هم به علت عدم وجود توزیع پیش‌فرض‌ها (توزیع نرمال) درباره داده‌های استفاده شده، می‌باشد. RF هزاران متغیر ورودی را بدون حذف یکی از آن‌ها پذیرا بوده و اجرا می‌کند، همچنین می‌تواند برآوردی از اینکه کدام متغیر در پیش‌بینی مدل، مهم است را تشخیص دهد. RF نسبت به شبکه‌های عصبی مصنوعی در شرایط گیرافتادن در کمینه محلی و داده‌های نویزدار، حساسیت کمتری دارد و می‌تواند تخمین بهتری از پارامترها را داشته باشد. دشت ملکان به عنوان یکی از دشتهای حوضه دریاچه ارومیه، به دلیل فعالیت‌های شدید کشاورزی با بحران کاهش سطح آب زیرزمینی مواجه است و با وجود مطالعات فراوانی مانند مطالعات تفصیلی و ژئوفیزیک که در سال‌های اخیر

مقادیر و چگونگی توزیع آن در مدل‌سازی و مدیریت آبخوان‌ها ضروری است. با توجه به عدم قطعیت ذاتی این پارامتر، مدل‌های مختلفی برای پیش‌بینی دقیق آن قابل استفاده است (کخدایی ایلخچی و همکاران ۲۰۰۹). بدست آوردن مقادیر قابلیت انتقال با استفاده از آزمایشات صحرائی مانند آزمایش پمپاژ به روش‌های کوپر و ژاکوب (۱۹۶۶)، چاو (۱۹۵۲) و تیس (۱۹۳۵) بسیار پرهزینه و وقت‌گیر است. بنابراین به طور وسیعی مدل‌های عددی و فیزیکی مختلفی در پیش‌بینی و بررسی توزیع این پارامتر توسعه یافته‌اند (راس و همکاران ۲۰۰۷). با توجه به برخی محدودیت‌های مدل‌های عددی و فیزیکی مانند وقت‌گیر بودن، پیچیده بودن، پرهزینه بودن و نیاز داشتن به داده‌های زیاد، روش‌های دیگری برای پیش‌بینی قابلیت انتقال ارائه شده‌است. تاکنون مطالعات فراوانی بر روی پیش‌بینی پارامترهای هیدروژئولوژیکی با استفاده از مدل‌های هوش مصنوعی مانند فازی و شبکه‌های عصبی مصنوعی صورت گرفته است (اولاتونجی و همکاران ۲۰۱۱، کخدایی ایلخچی و همکاران ۲۰۰۹، چن و همکاران ۲۰۰۶، چیت سازان و همکاران ۲۰۱۵). مدل‌های فازی عموماً در مدل‌سازی مسائلی که با عدم قطعیت ذاتی همراه هستند، مناسب می‌باشند (پولیدو و همکاران ۲۰۰۴، ندیری و همکاران ۲۰۱۳) و مدل شبکه‌های عصبی مصنوعی به عنوان یک تقریب‌گر جهانی می‌تواند در پیش‌بینی بررسی مسائل غیرخطی مفید باشد (ندیری و همکاران ۲۰۱۴). در پیش‌بینی پارامترهایی مانند قابلیت انتقال که با عدم قطعیت همراه است، عموماً بیش از یک مدل هوش مصنوعی قابل استفاده است و می‌توانند نتایج مشابهی حاصل کنند. هر یک از این مدل‌ها به تنهایی دارای عدم قطعیت و مزایای مربوط به خود است. روش‌هایی که تاکنون برای ارزیابی قابلیت انتقال ارائه شده است، به‌طور نسبی از داده‌های کمتری استفاده کرده‌اند (رودریگز ۲۰۱۲). در حالی که در

2. Machine learning

3. Repeated multiple prediction

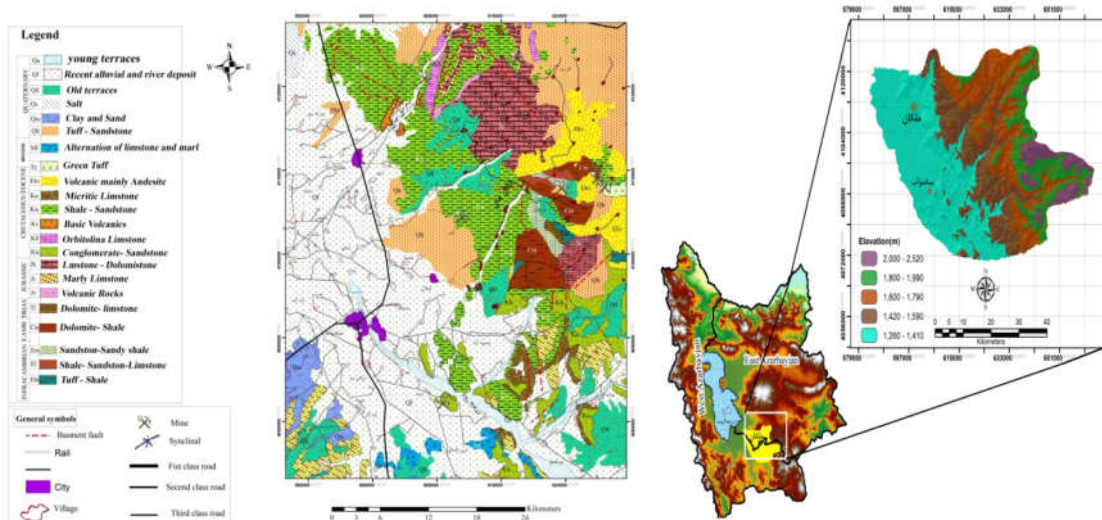
4. Random forest

در بخش شمال غرب دارای پهنه‌های رسی- نمکی هستند. آبخوان دشت از نوع آزاد می‌باشد که اکثراً از پادگانه‌های آبرفتی قدیمی، پادگانه‌های آبرفتی جدید، مخروط افکنه‌ها و رسوبات رودخانه‌ای تشکیل یافته است و مواد اصلی تشکیل دهنده آبخوان، رسوبات ماسه، سیلت و رس می‌باشد. بر اساس نقشه‌های هم ضخامت رسوبات آبرفتی، لاگ‌های حفاری و داده‌های ژئوفیزیکی، در قسمت‌های بالایی دشت ذرات تشکیل دهنده سفره دانه درشت بوده و قابلیت انتقال دشت نیز در این مناطق بالا می‌باشد و به سمت مرکز دشت و نواحی خروجی و به سمت دریاچه ارومیه رسوبات دانه‌ریزتر شده و قابلیت انتقال نیز کاهش می‌یابد. شکل ۳ مقاطع ژئوالکتریک دشت ملکان در راستای شمال‌شرق- جنوب‌غرب و راستای شمال‌غرب- جنوب‌شرق را نشان می‌دهد. طبق بررسی‌ها، کاهش مقاومت ویژه از شمال‌شرق به سمت جنوب‌غرب به علت عدم گسترش رسوبات دانه درشت و مقاوم و افزایش مواد محلول آب‌های زیرزمینی می‌باشد.

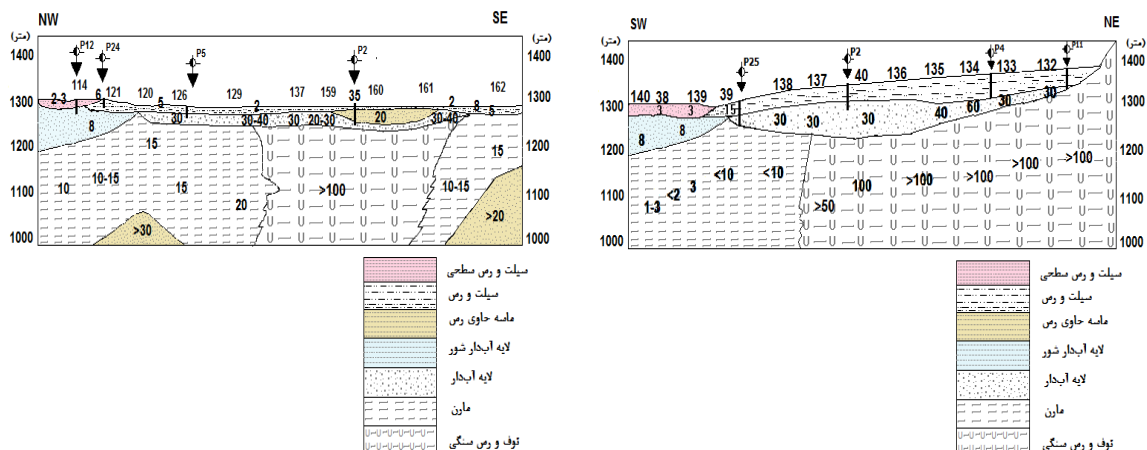
در این دشت صورت گرفته هنوز نیاز مبرم به شناخت پارامترهای هیدروژئولوژیکی برای مدیریت منابع آب آن دارد. لذا در این تحقیق سعی گردید که با استفاده از داده‌های مرتبط با قابلیت انتقال اقدام به برآورد این پارامتر شود. در این مطالعه با توجه به خصوصیات قابلیت انتقال، از روش RF برای پیش‌بینی و بررسی این پارامتر استفاده شد.

مواد و روش‌ها

دشت ملکان با وسعتی تقریباً ۴۵۰ کیلومتر مربع در جنوب استان آذربایجان شرقی و در جنوب‌شرق دریاچه ارومیه واقع شده است (شکل ۱) و جزء بخش زمین‌ساختاری البرز - آذربایجان محسوب می‌شود. این حوضه بر اساس روش تجربی آمبرژه (۱۹۵۲) و با استفاده از آمار ۱۲ ساله ایستگاه هواشناسی ملکان دارای اقلیم سرد و نیمه خشک می‌باشد. بر اساس نقشه زمین‌شناسی ۱:۱۰۰۰۰۰ مراغه منطقه ملکان دارای سازندهای زمین‌شناسی مختلفی می‌باشد. همانطور که در شکل ۲ نشان داده شده است، بخش اعظمی از مساحت منطقه در بخش غربی مربوط به رسوبات آبرفتی دوره کواترنر است و بخش کمی از آن‌ها



شکل ۱- موقعیت منطقه مورد مطالعه. شکل ۲- نقشه زمین‌شناسی منطقه مورد مطالعه (علوی و شهرابی ۱۳۵۸).



(ب)

(الف)

شکل ۳- مقاطع ژئوالکتریک دشت ملکان. (الف) راستای شمال‌شرق - جنوب‌غرب، (ب) راستای شمال‌غرب - جنوب‌شرق.

روش‌شناسی

یکی از ابزارهای کارآمد مورد استفاده در مسائل مربوط به تخمین متغیرهای هدف و یا طبقه‌بندی گیاهها، درخت تصمیم است. یک درخت تصمیم، فضای ورودی را به مجموعه‌هایی از نواحی مجزا تقسیم می‌کند و یک مقدار پاسخ را به هر ناحیه اختصاص می‌دهد. در حالت ساده، این پاسخ در مسائل رگرسیون می‌تواند بر اساس میانگین مقادیر هدف مرتبط با گیاههای قرار گرفته در هر ناحیه تعیین شود و یا به عبارتی پاسخ اختصاص یافته به هر ناحیه براساس میانگین مقادیر هدف، متناظر با گیاههای یادگیری قرارگرفته در هر ناحیه است. به طور کلی درخت تصمیم منفرد مستعد برازش اضافی^۶ بوده و قدرت تعمیم‌پذیری کمی دارد. در هنگام تشکیل یک درخت تصمیم، تغییر کوچکی در گیاههای یادگیری

می‌تواند باعث تغییرات اساسی در ساختار آن درخت گردد (کوینلن ۱۹۸۶). ترکیب درخت‌های تصمیم را جلوگیری می‌کند. یک مجموعه دسته جمعی شامل تعدادی یادگیرنده است که به آنها یادگیرنده‌های پایه گفته می‌شود. یادگیرنده‌های پایه معمولاً به وسیله یک الگوریتم یادگیری پایه، از داده آموزشی ساخته می‌شوند که این الگوریتم‌های یادگیری پایه می‌تواند درخت تصمیم، شبکه‌های عصبی مصنوعی یا الگوریتم‌های یادگیری دیگر باشد. قابلیت تعمیم یک مجموعه، اغلب قوی‌تر از یادگیرنده‌های پایه است. در واقع روش‌های مجموعه‌ای بیشتر به دلیل توانایی که در تقویت یادگیرنده‌های ضعیف دارند، مورد قبول هستند (اسچاپیر ۱۹۹۰). از این رو به یادگیرنده‌های پایه، یادگیرنده‌های ضعیف نیز گفته می‌شود که در این تحقیق برای فائق آمدن بر مشکلات یادگیرنده‌های پایه، الگوریتم RF که یک روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است، استفاده شد.

- ۵. Decision trees
- 6. Over fitting
- 7. Regression tree
- 8. Root node
- 9. Terminal root

رگرسیون و الگوریتم جنگل تصادفی (RF)

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 (کوزی‌انتیس و پینتلس ۲۰۰۴). درخت رگرسیونی^۷ (RT) مجموعه‌ای از شرایط یا محدودیت‌ها را بیان می‌کند که به صورت سلسله مراتبی سازمان یافته‌اند و به حالت متوالی از گره ریشه^۸ به سمت پایین رشد می‌کنند و به گره‌های پایانی^۹ یا گره‌های برگ^{۱۰} می‌رسند (برایمن و همکاران ۱۹۸۴، کوپینلن ۱۹۹۳). به منظور ایجاد درخت رگرسیونی از جزءبندی بازگشتی و رگرسیون‌های چندگانه استفاده می‌شود. فرایند تصمیم در هر گره داخلی از گره ریشه، طبق قانون درختی تکرار می‌شود تا زمانی که شرط توقف قبلی تعیین شده بدست آید. هر یک از گره‌های نهایی یا برگ‌ها به یک مدل رگرسیونی ساده که فقط در گره بکار برده می‌شوند، متصل می‌شود. زمانی که فرایند فراخوانی درخت به

پایان برسد، هرس کردن یا پرونینگ^{۱۱} به منظور بهبود ظرفیت تعمیم درخت‌ها به وسیله کاهش پیچیدگی ساختار بکار برده می‌شود. تعداد نمونه‌ها در گره‌ها می‌تواند به عنوان معیار پرونینگ در نظر گرفته شود. برای جلوگیری از تطابق RTs های مختلف، RF تنوع درختان را از طریق ایجاد زیرمجموعه‌های مختلف از داده‌های آموزشی، کاهش می‌دهد که اصطلاحاً کیسه‌بندی^{۱۲} نامیده می‌شود. کیسه‌بندی تکنیکی است که برای ایجاد داده‌های آموزشی استفاده می‌شود که این کار از طریق نمونه‌برداری مجدد تصادفی از مجموعه داده‌های اصلی و همراه با جایگزینی انجام می‌شود. در این مرحله هیچ‌کدام از داده‌های انتخاب شده از نمونه‌های ورودی برای تولید زیرمجموعه بعدی، حذف نمی‌شود و بدین ترتیب واریانس نیز کاهش می‌یابد. از این‌رو برخی از داده‌ها ممکن است بیش از یکبار در شاخه‌های آموزشی استفاده شود درحالی‌که برخی از داده‌های دیگر که در مدل‌سازی موثر نیستند هرگز استفاده نشوند. بنابراین ثبات بیشتری برای مدل بدست می‌آید و مدل را در برابر تغییرات جزئی در داده‌های ورودی قابل اعتمادتر می‌سازد و دقت پیش‌بینی آن را افزایش می‌دهد (برایمن ۲۰۰۱). از سوی دیگر RF برای ایجاد یک درخت رشد، از بهترین متغیرها یا نقاط تقسیم در داخل زیرمجموعه‌های متغیرها استفاده می‌کند که به صورت تصادفی از مجموعه‌ای از متغیرهای ورودی انتخاب شده‌اند. بنابراین قدرت هر RT منفرد را کاهش داده و میزان تطابق را پایین می‌آورد و بدین صورت خطای کلی مدل را کاهش می‌دهد (برایمن ۲۰۰۱). این روش یک الگوریتم پیچیده می‌باشد که برای بهبود

11. Pruning
12. Bagging

10. Leaf node

K تولید شده گروهی، مربوط به هر درخت برابر $\hat{y}_1 = h_1(x), \hat{y}_2 = h_2(x), \dots, \hat{y}_k = h_k(x)$ که \hat{y}_k خروجی درخت Kام می‌باشد. برای به دست آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود. خطای پیش‌بینی نیز براساس نمونه‌های خارج از کیسه طبق فرمول ۱ محاسبه می‌گردد.

$$MSE \square MSE^{OOB} = n^{-1} \sum_{i=1}^n [\hat{y}(x_i) - y_i]^2 \quad [1]$$

در رابطه فوق $\hat{y}(x_i)$ نتایج محاسباتی، y_i نتایج مشاهداتی و n تعداد کل مشاهدات است و MSE میزان خطای بین مقادیر مشاهداتی و محاسباتی را نشان می‌دهد. روش RF برای طبقه‌بندی به طور خلاصه، بدین صورت است که: در ابتدا T نمونه خودراه‌انداز از داده آموزشی بیرون کشیده می‌شود و سپس از هر نمونه خودراه‌انداز β یک درخت طبقه‌بندی و رگرسیون (CART)^{۱۴} هرس نشده ایجاد می‌شود که برای انشعاب در هر گره CART، تنها یکی از M ویژگی انتخاب شده به صورت تصادفی استفاده می‌شود.

در نهایت، خروجی طبقه‌بندی براساس یک نتیجه میانگین از پیش‌بینی‌های تمام درخت‌های منفرد آموزش دیده، بدست می‌آید. یک مجموعه داده خود راه‌انداز، مجموعه‌ای از نقاط انتخابی تصادفی است که از مجموعه داده آموزشی بیرون کشیده شده است (دودا و همکاران ۲۰۱۱). برای این که همیشه اندازه نمونه آموزشی اولیه ثابت بماند، مجموعه داده خودراه‌انداز یک نسخه کپی از نقاط را به کار می‌برد. لازم به ذکر است که الگوریتم RF به دلیل ارزیابی

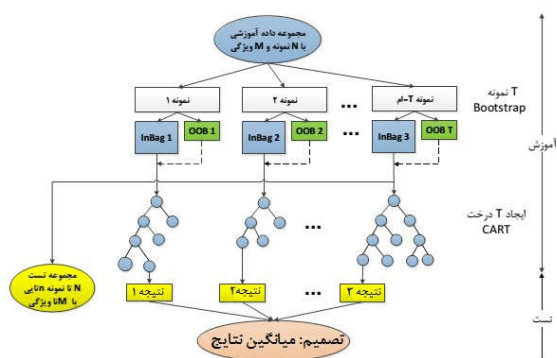
دادن یادگیری ماشین رده‌بندی و مدل‌های پرسرعتی بر حسب پایداری و دقت رده‌بندی استفاده می‌شود. این روش همچنین واریانس را کاهش داده و به دوری از برازش اضافی کمک می‌کند. اگر چه این روش در درخت تصمیم به کار می‌رود اما می‌تواند در هر نوع مدل استفاده شود. کیسه‌بندی یک حالت مخصوص از روند مدل میانگین می‌باشد. یکی دیگر از ویژگی‌های خوب RF این است که درختان RF بدون پرونینگ یا هرس کردن رشد می‌کند و در این روش آموزش بیش از اندازه بر دقت مدل تأثیری نمی‌گذارد که آن را از دیدگاه محاسباتی سبک‌تر می‌سازد. علاوه بر این آن دسته از نمونه‌هایی که در آموزش درختان در فرایند کیسه‌بندی انتخاب نمی‌شوند شامل بخشی از زیر مجموعه‌هایی می‌شوند که الگوهای خارج از کیسه^{۱۳} (OOB) نامیده می‌شوند. این قسمت در روش RF، می‌تواند برای ارزیابی عملکرد مدل مورد استفاده قرار گیرد (پیترز و همکاران ۲۰۰۷). به این ترتیب RF می‌تواند تخمین غیرمرتبط داخلی از خطای تعمیم را محاسبه کند بدون اینکه از زیرمجموعه‌های داده‌های خارجی استفاده کند (برایمن ۲۰۰۱). روند کلی الگوریتم RF بصورت ساده در شکل ۵ نشان داده شده است. در این روش بردار تصادفی Θ_k که مستقل از بردارهای تصادفی $\Theta_1, \dots, \Theta_{k-1}$ بوده، برای درخت K ام، تولید می‌شود. همچنین همه‌ی بردارها دارای توزیع یکسانی می‌باشند. درخت رگرسیونی با استفاده از مجموعه داده‌های آموزش و Θ_k رشد می‌کند و نتیجه‌ی مجموعه درخت‌های K برابر، $\{h_1(x), h_2(x), \dots, h_k(x)\}$ می‌باشد که در اینجا $h_k(x) = h(x, \theta_k)$ ، $x = \{x_1, x_2, \dots, x_p\}$ می‌باشد. این بردارها یک بردار ورودی P بعدی می‌باشند که یک جنگل را تشکیل می‌دهند. خروجی‌های

14. Classification and regression tree

15. Feature selection

13. Out of bag

تعداد درخت ها (k) به طوریکه مقدار خطا همگرا شود و تخمین قابل اعتمادتر باشد، مدل از ۱ تا ۵۰۰ درخت ساخته شد. به دلیل اینکه با افزایش درخت‌ها میزان خطا کاهش می‌یابد، بنابراین تعداد ۵۰۰ درخت برای فراخوانی مدل استفاده شد. پارامتر m نیز به روش آزمون و خطا بهینه شد. نتایج مدل به وسیله برآورد خطای OOB ارزیابی گردید. علاوه بر این، برای کاهش ابعاد و افزایش دقت و قابل تفسیر بودن مدل، روش FS بکار برده شد و متغیرهای مهم در پیش‌بینی نیز شناسایی شدند. برای فراخوانی مدل پیش‌بینی قابلیت انتقال، مقادیر هر پیکسل از لایه‌های رستری ۸ متغیر ورودی توضیح داده شده استخراج شده و با هم و با متغیر هدف که همان قابلیت انتقال می‌باشد، استفاده شد. متغیرهای توضیحی (پیش‌بینی کننده) و متغیر پاسخ باهم در مجموعه‌ای از بردارهای ویژگی ورودی ترکیب شدند. این بردارها ورودی الگوریتم RF را تشکیل دادند و به عنوان بردارهای Input-feature شناخته شدند. متغیر پاسخ دوتایی (متغیر توضیحی و متغیر پاسخ) به عنوان مقادیر هدف برای آموزش الگوریتم استفاده شدند.



شکل ۴: فلوچارت RF برای رگرسیون (اقتباس از گیو و همکاران ۲۰۱۱).

درونی نتایج هر طبقه‌بندی درختی که در داخل خود دارد و با وزندهی به نتایج هر درخت، می‌تواند نتایج صحیحی را تولید کند. تعداد زیادی از متغیرهای وابسته به خواص و رفتار سیستم آب زیرزمینی و نیروهای راندگی، می‌توانند توانایی مدل را افزایش دهند. این ابعاد بالا در مجموعه‌ی داده‌ها می‌تواند باعث کاهش دقت مدل شود (بلمن ۲۰۰۳). برای جلوگیری از این خطاها و ابعاد بالای داده‌ها، انتخاب ویژگی^{۱۰} (FS) بکار برده می‌شود. FS یک روشی برای انتخاب زیرمجموعه‌های پارامترهای مربوطه برای آموزش بهتر مدل می‌باشد (گایون و الیسف ۲۰۰۳). 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 برخی غیرمرتبط باشند (دیکسون ۲۰۰۹). FS با بالابردن 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. روش معمول FS، روش آماری چندمتغیره است که به علت کاهش ابعاد داده‌ها، مولفه‌های اصلی جایگزین پارامترهای اولیه می‌شوند (کریتو و همکاران ۲۰۰۳) که در مطالعه حاضر روش RF را که بر اساس درخت‌های گروهی است ترجیح داده شده است. در این روش با افزایش درختان، خطای تعمیم همیشه همگرا می‌شود، از این رو آموزش زیاد در این روش مشکلی ندارد و تعداد درختان زمانی که خطا همگرا شود، می‌تواند ثابت باشد. اولین مرحله در ایجاد مدل پیش‌بینی قابلیت انتقال تنظیم پارامترهای مدل است. به منظور تنظیم

منطقه مطالعاتی

در دشت ملکان تعداد ۱۰۲ عدد سونداژ ژئوالکتریک در راستای ۱۶ پروفایل در رسوبات آبرفتی اجرا شده است (بی‌نام ۱۳۸۵). همزمان با مطالعات ژئوفیزیک در سال ۱۳۸۵، برای محاسبه قابلیت انتقال آبخوان دشت ملکان از نتایج ۳۶ حلقه چاه پمپاژ که به روش پمپاژ با دبی ثابت و پمپاژ با افت پله‌ای توسط مهندسی مشاور فرسپند آب (بی‌نام ۱۳۸۵) و کارشناسان شرکت سهامی آب منطقه‌ای استان آذربایجان شرقی انجام گرفته است، استفاده شد.

با توجه به اینکه در روش پمپاژ به روش افت پله خطای بالایی مقادیر بدست آمده محتمل است، لذا مقادیر قابلیت انتقال بدست آمده با لاگ‌های زمین‌شناسی مقایسه و در صورت لزوم تصحیح گردیده است. همچنین ضریب ذخیره که نمایانگر عکس‌العمل سطح آب مخزن زیرزمینی در اثر تغذیه یا تخلیه است، به طور متوسط حدود ۲/۲ درصد می‌باشد (بی‌نام ۱۳۸۵). مقاومت الکتریکی ظاهری دارای دو مؤلفه اصلی مقاومت الکتریکی عرضی (R_i) و مقاومت الکتریکی طولی (R_L) می‌باشد که به ترتیب تغییرات پارامترهای هیدروژئولوژیکی (مانند

نفوذپذیری، کیفیت آب و غیره) عمقی و جانبی را نشان می‌دهد لذا مقاومت الکتریکی عرضی با مقادیر قابلیت انتقال مرتبط است (مایلت ۱۹۴۷، هارب و همکاران ۲۰۱۰) و همچنین مقادیر هدایت الکتریکی (EC) را نیز می‌توان از پارامترهای مرتبط با مقادیر قابلیت انتقال دانست (اولانتجی و همکاران ۲۰۰۴). لذا بر این اساس ورودی مدل‌های به کار رفته عبارت از مقاومت الکتریکی عرضی، ضخامت اشباع رسوبات آبرفتی (B)، هدایت الکتریکی (EC)، گرادیان هیدرولیکی، شیب سطح زمین و محیط خاک، آبخوان و خاک سطحی و بردار هدف مدل قابلیت انتقال می‌باشد که پس از نرخ‌بندی ورودی‌ها و بردار هدف، استفاده شد. جنس محیط آبخوان، خاک و خاک سطحی بر میزان قابلیت انتقال تاثیر بسزایی دارند، طوری که در مواد دانه درشت‌تر میزان قابلیت انتقال بالاتر از مواد دانه‌ریز می‌باشد. با توجه به اینکه این پارامترها کیفی می‌باشند لذا برای کمی کردن آنها از رتبه بندی استفاده شد، به نحوی که به مواد دانه درشت‌تر نرخ بیشتری نسبت به مواد دانه ریز اختصاص داده شد، جدول ۱ خصوصیات آماری برخی از پارامترهای ورودی را نشان می‌دهد.

جدول ۱- برخی از خصوصیات آماری پارامترهای ورودی در مدل.

پارامتر	واحدها	میانگین	بیشینه	کمینه	واریانس	انحراف معیار
هدایت الکتریکی	$\mu S cm^{-1}$	۷۸۰۶/۹	۳۰۱۰۰	۷۲۰	$5/06 E + 07$	۷۱۱۲/۳
گرادیان	-	۰/۰۰۳۳	۰/۰۰۷۸	۰/۰۰۱۵	$1/07 E - 06$	۰/۰۰۱۲
شیب زمین	%	۹/۱۰	۵۸/۵۹	۱/۲۱	۵۴/۳۰	۷/۳۶
ضخامت اشباع	m	۱۸/۵۶	۵۳/۹۸	۷/۳۹	۶۷/۰۳	۸/۱۸
مقاومت عرضی	Ωm	۲۹۳/۵۵	۱۰۴۵/۸۹	۷/۴۸	۷۶۷۱۷	۲۷۶/۹۸
قابلیت انتقال	$m^2 day^{-1}$	۴۵۸/۶۲	۱۴۳۳/۶۶	۱۵۶/۱۴	۲۲۷۶۱	۱۵۰/۸۷
محیط آبخوان	-	۲/۴۸	۶	۱	۳/۱۸	۱/۷۸
محیط خاک	-	۲/۷۹	۵	۱	۱/۳۳	۱/۱۵
خاک	-	۱/۹۹	۵	۱	۲/۴۸	۱/۵۷

بالایی دشت، ذرات تشکیل دهنده سفره دانه درشت بوده و قابلیت انتقال دشت نیز در این مناطق بالا می‌باشد و به سمت مرکز دشت و نواحی خروجی و نیز به سمت دریاچه ارومیه رسوبات دانه‌ریزتر شده و قابلیت انتقال نیز کاهش می‌یابد. همچنین بالا بودن مقدار قابلیت انتقال در قسمت مرکزی دشت به دلیل بالا بودن ضخامت رسوبات آبرفتی در این قسمت می‌باشد. هدایت الکتریکی آب زیرزمینی در قسمت‌های شمال‌غرب و غرب منطقه بسیار زیاد می‌باشد (شکل ۵-ز). همچنین با برداشت آب از چاه‌ها، جریان آب زیرزمینی به سمت بالادست آبخوان برقرار می‌شود و بنابراین جریان آب شور به قسمت‌های بالادست آبخوان نیز توسعه پیدا می‌کند. مقاومت عرضی در قسمت‌های شرقی و جنوب‌شرقی منطقه بالا می‌باشد که بالا بودن آن به دلیل بیشتر بودن ضخامت در منطقه مرکزی و شرقی دشت و همچنین وابسته به جنس محیط آبخوان می‌باشد (شکل ۵ الف). جدول ۲ ارتباط مقاومت ویژه دشت مورد نظر با جنس آبخوان را بیان می‌کند که هرچه جنس آبخوان و دانه‌بندی آن درشت باشد، مقاومت ویژه از مقدار زیادی برخوردار است و در مناطق نزدیک به دریاچه ارومیه که ریزدانه می‌باشند مقاومت ویژه پایین می‌باشد و به تبع آن‌ها میزان قابلیت انتقال نیز دارای مقدار کمتری می‌باشد.

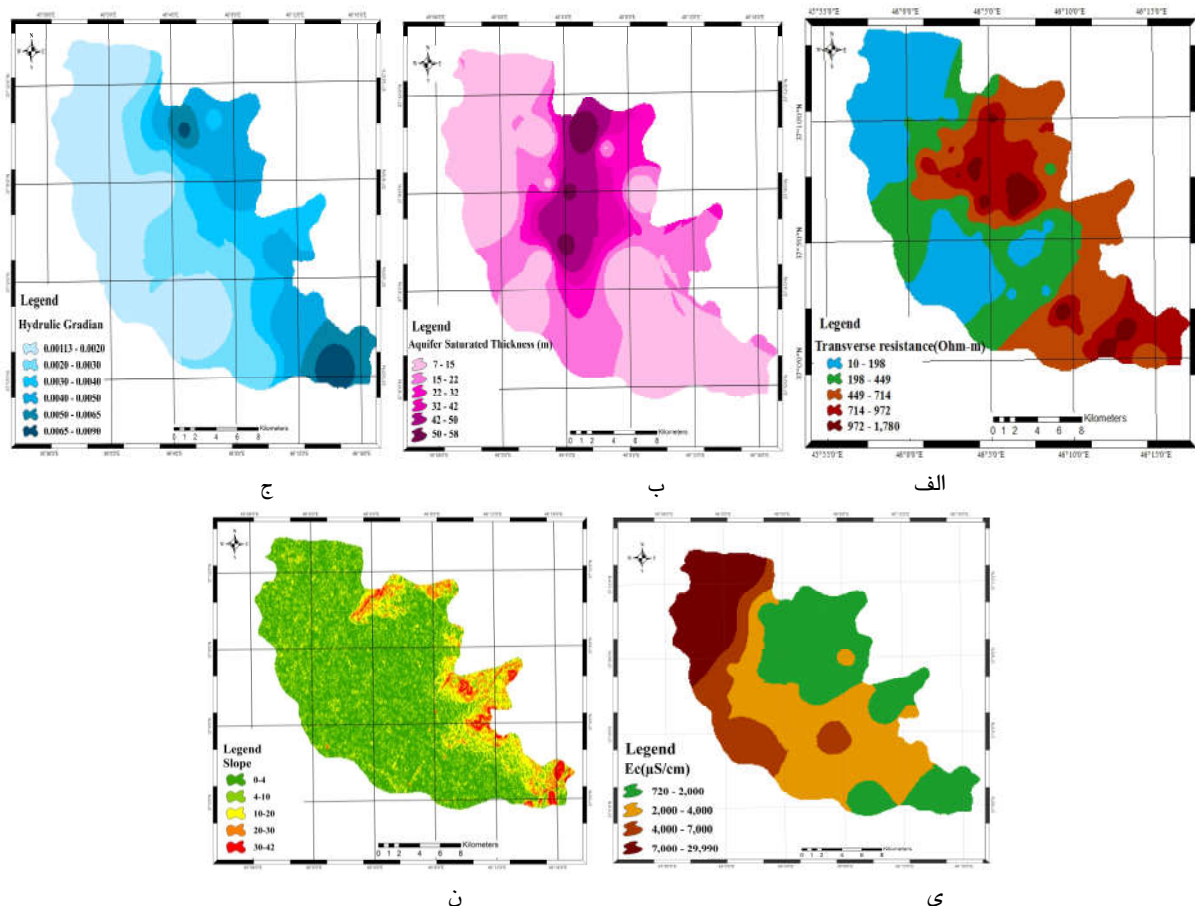
سپس داده‌های ورودی و بردار هدف به رستر تبدیل شده و سپس وارد مدل شدند. ۳۰ درصد داده‌ها برای مرحله آزمایش کنار گذاشته شد و ۷۰ درصد آنها برای مرحله مدل‌سازی استفاده شد. با داشتن عمق سنگ کف و عمق سطح آب زیرزمینی، ضخامت اشباع آب زیرزمینی برای منطقه بدست آمد و در نهایت با استفاده از هدایت الکتریکی و ضخامت اشباع منطقه، مقدار قابلیت انتقال محاسبه گردید.

نتایج و بحث

شکل ۵ لایه‌های رستری متغیرهای توضیحی استفاده شده در روش RF را نشان می‌دهد که در لایه‌های ایجاد شده، توزیع مکانی متغیرهای توضیحی مشاهده می‌گردد. مقدار گرادیان هیدرولیکی که مسئول حرکت آب زیرزمینی می‌باشد، برای منطقه مورد مطالعه، از طریق تقسیم متوسط عرض مقطع بر فاصله ارتفاعی بین دو منحنی هم‌تراز مقطع در سال ۱۳۸۵ به دست آمد و بدون بعد می‌باشد. در دشت ملکان این متغیر در نقاط مختلف متفاوت می‌باشد و به‌طور کلی در ورودی‌های دشت میانگین $0/006$ و در خروجی‌ها نیز در حدود $0/001$ می‌باشد (شکل ۵ ج). بر اساس نقشه‌های هم‌ضخامت رسوبات آبرفتی، لاگ‌های حفاری و داده‌های ژئوفیزیکی در قسمت‌های

جدول ۲. مقاومت ویژه لایه‌های زیرزمینی دشت ملکان.

مقاومت ویژه ظاهری (Ωm)	شرح
بیش از ۵۰	رسوبات دانه درشت، کنگلومرا و توده‌های آذرین
۲۰-۵۰	رسوبات آبرفتی آب دار حاوی آب شیرین
۱۰-۲۰	رسوبات رسی و مارنی
کمتر از ۵	رسوبات دانه ریز دارای آب شور



شکل ۵- توزیع مکانی پارامترهای استفاده شده در مدل RF: الف) مقاومت عرضی، ب) ضخامت اشباع آبخوان، ج) گرادیان هیدرولیکی، د) هدایت الکتریکی، ی) هدایت الکتریکی، ن) شیب.

روش، از داده‌ها برای آزمایش استفاده نمی‌گردد (چهارتا و همکاران ۲۰۰۹). شکل (۶-الف) نتایج پیش‌بینی مدل RF را در مقابل داده‌های مشاهداتی برای قابلیت انتقال نشان می‌دهد که مطابقت داده‌های محاسباتی و مشاهداتی ملاحظه می‌گردد، همچنین توزیع مقادیر قابلیت انتقال حاصل از مدل RF در شکل (۶-ب) نشان داده شده است که بر اساس شکل مربوطه قابلیت انتقال در قسمت‌های ورودی دشت در مناطق شرقی و مرکزی بالا می‌باشد و به سمت خروجی دشت از مقدار قابلیت انتقال کاسته می‌گردد. از آنجا که RF طبقه‌بندی را با میانگین‌گیری و

جدول ۳ اهمیت هر متغیر^{۱۶}، همراه با میزان امتیاز^{۱۷} پیش‌بینی شده برای هرکدام از متغیرها را در مدل نشان می‌دهد که به‌عنوان یکی از خروجی‌های مدل می‌باشد. بر اساس پیش‌بینی مدل RF، متغیرهای هدایت الکتریکی، محیط آبخوان، گرادیان هیدرولیکی و مقاومت عرضی آبخوان به‌ترتیب بیشترین تأثیر را در پیش‌بینی قابلیت انتقال داشتند. مدل RF با استفاده از مقدار خطای OOB، پیش‌بینی را انجام می‌دهد و در این

16. Variable importance

17. Score

۵۰۰ می‌گردد، میزان خطا مساوی ۰/۰۲۱ می‌شود. همچنین در شکل (۷-ب) مقدار AUC برابر با ۰/۹۶ می‌باشد که نمایانگر دقت بالای مدل‌سازی می‌باشد. خطای موجود در مدل نهایی به عوامل مختلفی از آن جمله تعداد محدود داده‌ها، ناهمگنی آبخوان، خطای ذاتی موجود در 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 (ندیری و همکاران ۱۳۹۳). یکی از مهمترین مزیت این روش نسبت به روشهای قبلی ارائه شده توسط چن و لی (۲۰۰۶) و کدخدایی ایلخچی و امینی (۲۰۰۹) که از روش‌های هوش مصنوعی برای تخمین قابلیت انتقال استفاده کرده‌اند را می‌توان به وزندهی این روش به ورودی‌ها و پارامترهای مرتبط دانست. به طوری که در این روش در حین پیش‌بینی مقادیر قابلیت انتقال می‌توان اهمیت هر یک از پارامترهای مرتبط را نیز تعیین کرد که در روش‌های قبلی ارائه شده این قابلیت وجود ندارد.

جدول ۳- اهمیت هر متغیر در پیش‌بینی قابلیت انتقال.

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

شمارش نتایج طبقه‌بندی‌کننده‌های پایه ساده و متنوع انجام می‌دهد بنابراین قادر است داده‌های دارای نویز را نیز به طور صحیح‌تری آموزش بدهد. پیش‌بینی‌ها از ترکیب تعدادی زیادی از پارامترهای ممکن ناشی می‌شود که به وسیله میانگین مربعات خطا^{۱۸} (MSE) ارزیابی می‌گردند. مدلی که کمترین MSE را داشته باشد به عنوان مدل بهینه، برای پیش‌بینی انتخاب می‌گردد. روش دیگر ارزیابی مدل برپایه‌ی منحنی‌های عملکرد سیستم^{۱۹} (ROC) می‌باشد. منحنی‌های ROC به نحوی مشابه روش‌های ارزیابی مدل عمل می‌کنند. در حقیقت مقادیر بدست آمده از مدل برای داده‌های آموزش و آزمایش که به صورت تصادفی انتخاب شده‌اند، براساس روابط موجود برای روش ROC استاندارد شده و مقادیر مرحله آموزش به عنوان TPR^{۲۰} یعنی مقادیر صحیح در محور Y رسم و مقادیر آزمایشی انتخاب شده به صورت تصادفی، به عنوان FPR^{۲۱} در محور X به صورت متقابل قرار می‌گیرند و سطح زیرمنحنی که به دست می‌آید بیانگر دقت مدل است. هرچه مساحت زیر منحنی^{۲۲} (AUC) در منحنی‌های ROC به یک نزدیک باشد، مدل از صحت بیشتری برخوردار است. همانطور که در شکل (۷-الف) مشاهده می‌گردد میزان خطای مدل‌سازی با افزایش درخت‌های تصمیم کاهش می‌یابد، به نحوی که در درخت تصمیم اولی میزان خطای MSE برابر با ۰/۴۶۶ بوده و هنگامی که تعداد درخت‌های تصمیم

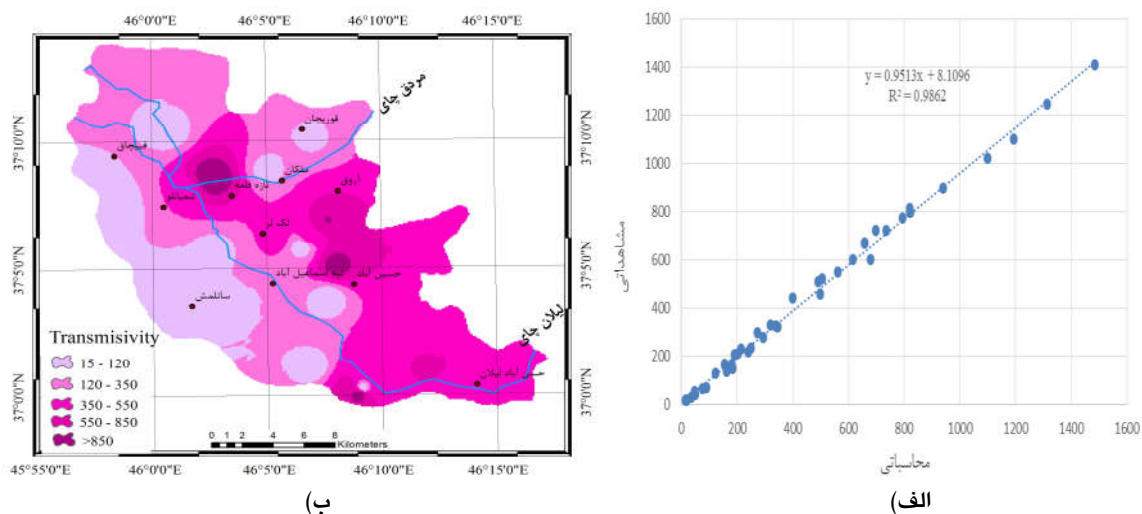
18. Mean square error

19. Receiver operating characteristic

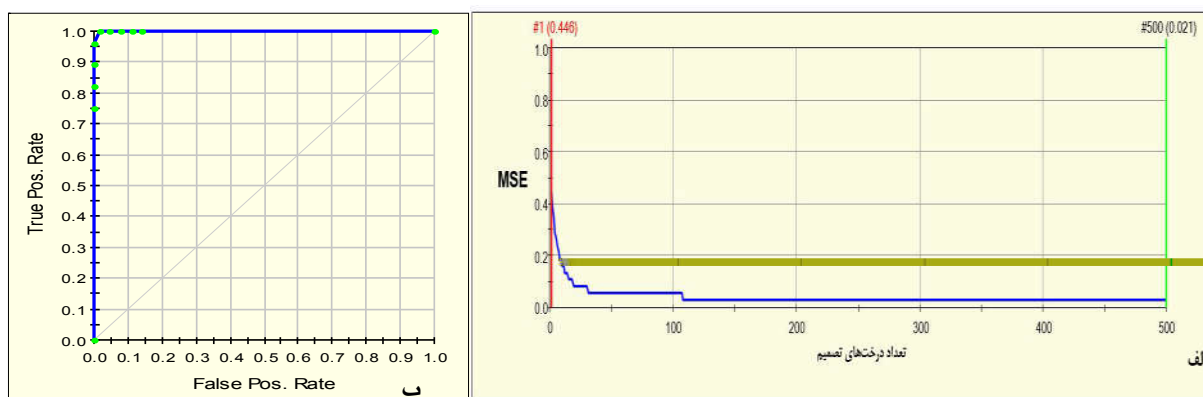
20. True positive rate (TPR)

21. False positive rate

22. Area under curve



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



شکل ۷- الف) خطای MSE مدل و کاهش آن با افزایش درخت‌های تصمیم، ب) منحنی ROC حاصل از مدل 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، می‌باشد. بر اساس نتایج مدل‌سازی RF با MSE و AUC به ترتیب برابر با ۰/۳۶ و ۰/۹۶، متغیرهای هدایت الکتریکی، محیط آبخوان و گرادیان هیدرولیکی به ترتیب به عنوان پارامترهای مهم و تأثیرگذار در پیش‌بینی قابلیت انتقال شناسایی شدند. هدایت

نتیجه‌گیری کلی

با توجه به هزینه‌های بالای آزمایش‌های صحرایی و عدم قطعیت موجود در پارامترهای هیدروژئولوژیک به خصوص قابلیت انتقال، مدل RF ارائه شده می‌تواند روشی مناسب برای پیش‌بینی و برآورد این پارامترها باشند. روش RF با داشتن مزایایی مثل یادگیری روابط غیرخطی و داشتن حساسیت کمتر در برابر گیر افتادن در کمینه محلی، به عنوان یک روش دقیق در مدل‌سازی پیش‌بینی

الکتریکی با تأثیر بر روی مقاومت عرضی و همچنین بسیار مهم در پیش‌بینی قابلیت انتقال می‌باشد. با داشتن رابطه معکوس با قابلیت انتقال، متغیری

منابع مورد استفاده

اصغری مقدم الف، ۱۳۸۹. اصول شناخت آب‌های زیرزمینی. انتشارات دانشگاه تبریز.
علوی نائینی م و شهرابی م، ۱۳۵۸. نقشه زمین‌شناسی مراغه، مقیاس ۱:۱۰۰۰۰۰، سازمان زمین‌شناسی و اکتشاف معدن کشور. تهران.

بی‌نام، ۱۳۸۵. تهیه بیلان و چرخه آب در محدوده مطالعاتی ملکان. شرکت آب منطقه‌ای آذربایجان شرقی، تبریز.
ندیری ع، اصغری مقدم ا، عبقری ه، کلانتری اسکویی ع، حسین پور ع و حبیب زاده ا، ۱۳۹۳. مدل منطق فازی در تخمین قابلیت انتقال آبخوان‌ها. مجله دانش آب و خاک، شماره ۱، دوره ۲۴، صفحه‌های ۲۱۹ تا ۲۳۳.

Bellman R, 2003. Dynamic Programming. Dover Publications 366 p. Mineola, New York.

Booker DJ, Snelder TH, 2012. Comparing methods for estimating flow duration curves at ungauged sites. Journal of Hydrology 435: 78–94.

Breiman L (2001) Random Forests. Machine Learning 45(1): pp. 5–32.

Breiman L, Friedman JH, Olshen RA, Stone CJ, 1984. Classification and Regression Trees. Chapman & Hall/CRC, New York.

Chehata N, Guo L, Mallet C, 2009. Airborne lidar feature selection for urban classification using random forests. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39: 207-12.

Chen CH, Lin ZS, 2006. A committee machine with empirical formulas for permeability prediction. Journal of Computers and Geosciences 32: 485–496.

Chitsazan N, Nadiri AA, Tsai F, 2015. Prediction and structural uncertainty analyses of artificial neural networks using hierarchical bayesian model averaging. Journal of Hydrology 528: 52-62.

Chow VT, 1952. On the determination of transmissibility and storage coefficient from pumping test data. Transactions, American Geophysical Union 33: 397-404.

Cooper HH, Jacob CE, 1946. A generalized graphical method for evaluation formation constants and summarizing well field history. Transactions, American Geophysical Union 27: 526-534.

Critto A, Carlon C, Marcomini, A, 2003. Characterization of contaminated soil and groundwater surrounding an illegal landfill by principal component analysis and kriging. Journal of Environmental Pollution 122(2): 235–44.

Dixon B.A, 2009. Case study using support vector machines, neural networks and logistic regression in a GIS to identify wells contaminated with nitrate-N. Journal of Hydrogeology 17(6): 15–20.

Duda, RO, Hart PE, Stork DG, 2011. Pattern Classification. 2nd Edition. John Wiley & Sons, New York.

Emberger L, 1952. Sur le quotient pluviothermique. C.R. Sciences, 234: 2508-2511.

Friedl MA, Brodley CE, Strahler AH, 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Trans Geoscience Remote Sense 37(2): 969–77.

Guo L, Chehata N, Mallet C, Boukir S, 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. Journal of Photogramm Remote Sensing 66(1): 56–66.

Guyon I, Elisseeff A, 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3: 1157–82.

Harb N, Haddad K, Farkh S, 2010. Calculation of transverse resistance to correct aquifer resistivity of groundwater saturated zones: implications for estimating its hydrogeological properties. Journal of Lebanese Science 11(1): 105-115.

Kadkhodaie-Ilkhechi A, Amini A, 2009. Journal of Petroleum Geology, 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 32(1): 67-78 67.

- Kadkhodaie-Ilkhchi A, Rezaee MR, Rahimpour-Bonab H, (2009a) 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* 65: 23-32.
- Kadkhodaie-Ilkhchi A, Rezaee MR, Rahimpour-Bonab H, Chehrazi A, 2009b. Petro physical data prediction from seismic attributes using committee inference system. *Journal of Computers & Geosciences* 35: 314–330.
- Ko B, Gim J, Nam J, 2011. Image classification based on ensemble features and random forest. *Electronics Letters* 47: 638-9.
- Kotsiantis S, Pintelas P, 2004. Combining bagging and boosting. *Journal of Computational Intelligence* 1(4): 324–33.
- Maillet R, 1947. The fundamental equations of electrical prospecting. *Journal of Geophysics* 12: 529-556.
- Nadiri AA, Asghari Moghaddam A, Tsai F, Fijani E, 2013. Hydrogeochemical analysis for Tasuj plain aquifer, Iran. *Journal of Earth System Science* 122(4): 1091-1105.
- Nadiri AA, Chitsazan N, Tsai F, Asghari Moghaddam A, 2014. Bayesian artificial intelligence model averaging for hydraulic conductivity estimation. *Journal of Hydrological Engineering* 19(3): 520-523.
- Olatunji SO, Selamat A, Abdulraheem A, 2011. Modeling the permeability of carbonate reservoir using type-2 fuzzy logic systems. *Journal of Computers in Industry* 62: 147–163.
- Pal M, 2005. Random Forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26(1): 217–22.
- Peters J, Baets BD, Verhoest NEC, Samson R, Degroev S, Becker P D, 2007. Random Forests as a tool for Eco hydrological distribution modelling. *Journal of Ecology Modeling* 207(2–4): 304–18.
- Pulido CI, Gutiérrez JC, 2009. Improved irrigation water demand forecasting using a soft computing hybrid model. *Journal of Biosystems Engineering* 102(2): 202-218.
- Quinlan JR, 1986. Induction of decision trees. *Journal of Machine Learning* 1(1): 81-106.
- Quinlan JR, 1993. C4.5 programs for machine learning. San Mateo, 303 pp, Morgan Kaufmann. *San Mateo, CA, California*.
- Rodriguez VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sánchez JP, 2012d. An assessment of the effectiveness of a Random Forest classifier for land-cover classification. *Journal of Photogram Remote Sens* 67: 9 -104.
- Ross J, Ozbek M, 2007. Hydraulic conductivity estimation via fuzzy. *Journal of Mathematical Geology* 39(8): 765-780.
- Schapire R, 1990. The strength of weak learnability. *Journal of Machine learning* 5: 197-227.
- Theis CV, 1935. The relationship between the lowering of piezo metric surface and the rate and duration of discharge of a well using groundwater storage. *Transactions, American Geophysical Union* 16: 519-524.
- Todd DK, Mays LW, 2005. *Groundwater Hydrology*, 3rd, John Wiley and Son's Publishers, p 636.